



## Abstract

To improve the protection of wetlands, it is imperative to have a thorough understanding of their structuring elements and of the identification of efficient methods to describe and monitor them. This article uses sophisticated statistical classification, interpolation and error propagation techniques, in order to describe vegetation spatial patterns, map plant community distribution and evaluate the capability of statistical approaches to produce high-quality vegetation maps. The approach results in seven vegetation communities with a known floral composition that can be mapped over large areas using remotely sensed data. The relations between remotely sensing data and vegetation patterns, captured in four factorial axes, were formalized mathematically in multiple linear regression models and used in a universal kriging procedure to reduce the uncertainty in mapped communities. Universal kriging has shown to be a valuable interpolation technique because parts of vegetation variability not explained by the images could be modeled as spatially correlated residuals, increasing prediction accuracy. Differences in spatial dependence of the vegetation gradients evidenced the multi-scale nature of vegetation communities. Cross validation procedures and Monte Carlo simulations were used to quantify the uncertainty in the resulting map. Cross-validation showed that accuracy in classification varies according with the community type, as a result of sampling density and configuration. A map of uncertainty resulted from Monte Carlo simulations displayed the spatial variation in classification accuracy, showing that the quality of classification varies spatially, even though the proportion and arrangement of communities observed in the original map is preserved to a great extent. These results suggested that mapping improvement could be achieved by increasing the number of field observations of those communities with a scattered and small patch size distribution; or by including new digital images as explanatory variables in the model. By comparing the resulting plant community map with a flood duration map, we verified that flooding duration is an important driver of vegetation zonation. We discuss our study in the context of developing a mapping approach that is able to integrate field

**BGD**

7, 6889–6934, 2010

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

point data and high-resolution remote sensing images, providing new basis to map wetland vegetation and allowing its future application in habitat management, conservation assessment and long-term ecological monitoring in wetland landscapes.

## 1 Introduction

Wetland ecosystems are among the habitats most threatened by climatic change, due to their high sensitivity to the hydrological regime (Junk, 2002). They form transitional habitats between aquatic and terrestrial systems and embody different kinds of habitats such as mangroves, peatlands, freshwater swamps and marshes (Mitsch et al., 2009). The ecological importance of these habitats has been recognized worldwide as well as the urgent need to preserve them, as stressed in the Cuiabá Declaration on Wetland elaborated during the 8<sup>o</sup> International Wetlands Conference of INTECOL, Brazil. However, lack of knowledge about the complex natural dynamics of wetlands may lead to arbitrary management decisions (Junk et al. 2006). To improve the protection of wetlands, it is imperative to have a thorough understanding of the structuring elements and of the identification of efficient methods to describe and monitor them.

Vegetation communities have distinct spatial and temporal patterns. Understanding the mechanisms that determine these patterns has been an important issue in ecology for decades (e.g., Connell and Slatyer, 1977; Svenning et al., 2004). Two factors play a key role: spatial interactions in ecological processes (e.g. competition), and environmental factors (e.g. flooding duration) (Tilman, 1988). Ecological processes include interactions between individuals, which may cause particular spatial patterns in the distribution of plants. Spatial variation in environmental factors causes spatial patterns in vegetation communities due to the differences of species requirements. These two factors do not usually operate independently but act together at different spatio-temporal scales (Turner, 1989; Svenning et al., 2004). This multi-scale interaction may lead to complex spatial patterns that are continuously changing (Wagner and Fortin, 2005). Consequently, the ability of distinguishing plant communities that arise

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



from multi-scale ecological processes requires an understanding of the processes and parameters causing the heterogeneity (Turner, 1989).

Classical methods describing vegetation distribution patterns along environmental gradients are based on sampling field plots, often along transects (McIntosh, 1958; Whittaker, 1967). Such an approach yields detailed insights into the vegetation occurrence and vegetation assemblages but does not provide spatially continuous information required to study mechanistic processes and spatial patterns of the landscape (Austin and Smith, 1989). To retrieve such spatially continuous information requires techniques that consider space explicitly (Gardner and Engelhardt, 2008). One of these techniques is remote sensing. By using the spectral signature of different vegetation states, remote sensing enables us to describe spatial and temporal patterns of vegetation in a spatially continuous way (Jensen, 2007). A restriction of this approach is the limited level of detail in attribute information that can be mapped by remote sensing, hindering the detection and identification of many ecologically important properties of vegetation communities, such as floral composition (Chambers et al., 2007).

Whereas field plots and remote sensing data each have their limitations as a source for continuous vegetation maps, is it possible to combine them through a statistical approach (Guisan and Zimmerman, 2000; Ferrier et al., 2002; Pfeffer et al., 2003; Miller et al., 2007). Point-data from field plots and spatially continuous information from remote sensing are here incorporated by means of statistical methods, such as ordination analysis (Jongman et al., 1995) and spatial interpolation techniques such as kriging. In this way, we can make maps representing the spatial distribution of vegetation across large areas that incorporate detailed information on floral composition (Pfeffer et al., 2003). This approach has become increasingly important in ecological studies as it recognizes the influence of spatial correlation in vegetation patterns (Bascompte and Solé, 1996; Turner et al., 2001). In addition, these techniques allow quantifying the uncertainty in mapped vegetation, which is valuable when vegetation maps are used for further quantitative analysis or for calibration and evaluation of mechanistic vegetation models (e.g., Brzeziecki et al., 1993; Guisan and Zimmermann, 2000; Chong et al.,

## BGD

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2001). Here, we will use mapped vegetation (and its uncertainty) to study the effect of flood duration on plant community patterns.

In this study, we integrate field data and remote sensing data through geostatistical methods for a case study in the Pantanal, a 150 000 km<sup>2</sup> floodplain in the center-west part of Brazil. The variability in water depth and flood duration are considered to be the preponderant causes of the high diversity of biological communities and plant zonation patterns found in the area (Junk et al., 1989; Wantzen et al., 2005). In this extensive and pristine wetland floodplain, long-term conservation depends on habitat diversity maintenance (Junk et al., 2006).

The aims of this paper are: (1) to identify plant communities of the Pantanal on key structural and compositional attributes of plant life forms, based on a new data set collected in a field survey; (2) to present a spatial statistical approach based on the integration of field data and remotely sensed data to make accurate predictive maps of vegetation distribution; (3) to evaluate the uncertainties in vegetation classification on the basis of this novel statistical approach; and (4) to investigate relation between flood duration and vegetation zonation.

## 2 Study area

The Pantanal contains a large variety of alluvial ecosystems with different drainage patterns, flooding characteristics, geomorphologic aspects and vegetation types covering about 150 000 km<sup>2</sup> of the upper Paraguay basin (Fig. 1a) (Assine and Soares, 2004). The climate of this region is tropical humid with marked seasonality between winter and summer periods (Köppen, 1948). The summer from November to April is characterized by high temperatures (average day temperature 34 °C) and it is the season with the largest amount of precipitation (Fig. 1b). The precipitation decreases in winter, causing this season to be very dry (de Musis et al., 1997). The water level in the rivers of the Pantanal follows the seasonal trend in the precipitation. Due to the poor surface and subsurface drainage and the smooth, low topography relative to the river

**BGD**

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



level (Alvarenga et al., 1984; Assine and Soares, 2004), large areas of the Pantanal are flooded every summer. Climate oscillations have been shown to be the main cause of the observed multi-year period of cyclic variation in flooding (Junk et al., 2006).

The Pantanal vegetation presents floristic elements of three important morphoclimatic and phytogeographic domains, i.e., Cerrado (Brazilian savanna), Amazonia and Chaco (Ab'Saber, 1988). Savanna vegetation types are dominant physiognomies in the Pantanal (67%), but are not the only one: semideciduous forest, gallery forest, swamp, Chaco, pioneer formations such as monodominant forest of *Vochysia divergens* Pohl (Silva et al., 2000) are the remaining components of the vegetation mosaic. The variability in water depth and flooding duration and the temporal connections and disconnection established between different elements of the landscape by means of the flood pulse are considered the preponderant causes of the high diversity of biological communities in the Pantanal (Junk et al., 1989; Wantzen et al., 2005), dictating where and when plant species with different life strategies and flooding tolerance will appear (Junk et al., 2006). Our study site covers 60 km<sup>2</sup> and is located within a nature reserve in North Pantanal (16°30'–16°44' S and 56°20'–56°30' W) (Fig. 1a). The site is representative of a large part of the Pantanal regarding vegetation and environmental conditions. The fluctuation in water level of the river Cuiabá, which crosses the north part of the studied area, is the main cause of the flooding over the studied area.

### 3 Outline of the approach

Figure 2 shows a diagram with the procedural steps followed to identify vegetation communities, to determine their spatial distribution, and to study their relationship with flooding duration. The first part of the paper addresses the extraction of vegetation communities from high resolution field sampling using factor analyses and clustering (Fig. 2, top-right). Spatially continuous variables were obtained from remotely sensed imagery and a digital elevation model providing spatial information necessary for vegetation mapping (Fig. 2, top-left). These remote sensing and elevation data are related to

**BGD**

7, 6889–6934, 2010

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





Dominant species within each plot, that is, the woody species with the highest biomass or the vine and herbaceous species with the highest coverage degree, were identified and included in vegetation observations and analyses to ensure discrimination between structurally similar but floristically distinct communities.

## 4.2 Sampling scheme and data collection

Field sampling of vegetation was done in 2006 and 2007. A sampling scheme modified from the RAPELD method (c.f., Magnusson et al., 2005) was used here (Fig. 3). The adjusted RAPELD method comprised the establishment of 23 trails of 250 m length distributed over the study site (Fig. 3a). In order to study the effect of flooding duration on vegetation composition and structure, each trail was positioned at a different topographical elevation. Each trail was thus placed along an elevation contour, defined using a tripod-mounted telescope. In order to capture variation in vegetation over short distances, the trails were divided into sampling trails of 50 m length (Fig. 3b), producing a total number of 115 sampling units.

Measurement acquisition and sample dimensions of a sampling unit varied according to life form (Fig. 3c). Herbaceous and vine species were sampled according to the point quadrat method, which is based on point-intercept frequency measurements by plants (Bullock, 1996). Presence or absence of species was recorded at 25 points spaced at 2 m intervals along the sampling trail. The coverage value for a sampling trail was calculated as the proportion of these points being intercepted by the plant. For woody life forms, plots were positioned along the trail (Fig. 3c). The plots have a length equal to the length of the sampling unit along the trail (50 m), and a width depending on the life form size as suggested by the RAPELD method (Fig. 3c). Shrub measurements were taken in plots of 200 m<sup>2</sup> (50 × 4 m); medium-sized tree measurements in plots of 1000 m<sup>2</sup> (50 × 20 m); and large-sized tree measurements in plots of 2000 m<sup>2</sup> (50 × 40). All species found in the plot were identified and trunk diameter and species height were measured for trees and shrubs. For shrub species, diameters were measured for each individual at 5 cm above the soil surface and for tree species

**BGD**

7, 6889–6934, 2010

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





at breast height. These data were also used to calculate some variables describing the vegetation structure. Aboveground biomass of woody individuals was estimated by two different allometric equations for shrubs and trees, respectively. Aboveground woody biomass ( $B_s$ , kg) of shrubs was calculated using the allometric model developed by  
5 Barbosa and Ferreira (2004):

$$B_s = \exp \left( -3.9041 + 0.4658 \ln \left( Cb^2H \right) + 0.0458 \left( \ln \left( Cb^2H \right) \right)^2 \right) \quad (1)$$

with,  $Cb$  is circumference at the ground height (cm) and  $H$  the species height (m).

Biomass ( $B_t$ , Kg) of a tree species was estimated following Chave et al. (2005):

$$B_t = 0.112 \times \left( \rho \times H \times d^2 \right)^{0.916}, \quad (2)$$

10 with,  $\rho$  ( $\text{g cm}^{-3}$ ) the wood specific density,  $H$  (m) the species height, and  $d$  (cm) the species diameter at breast height. Information on species densities was obtained from Schöngart et al. (2008). Canopy height (CH) was considered the average height of the eight highest individuals in a plot.

## 5 Identifying plant communities

15 Veloso et al. (1991) developed a classification system of Brazilian vegetation, which was adapted by Nunes da Cunha et al. (2006), providing a detailed description of the plant communities in the Pantanal. Here, we aim at mapping the communities described by Nunes da Cunha et al. (2006), considering that these can be identified by means of quantification of structure and composition (only dominant species) of  
20 different vegetation layers. Communities are represented at a broad level as vegetation formation types rather than plant associations. We used factor analysis (Bray and Curtis, 1957) where the factor scores summarize the structural and compositional characteristics of different vegetation samples. These factors were found in a principal

---

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



component analysis of the correlation matrix, generating a small number of orthogonal factors explaining the correlation among the vegetation variables (Legendre and Legendre, 1998). The different factor scores are plotted against each other in Fig. 4, and the proximity among point-samples and our field background about the vegetation classes found in these points were used to classify them in vegetation classes/clusters. Finally, cluster centers were calculated by averaging factor scores corresponding to the community/cluster and were used in the final part of the mapping procedure below.

## 5.1 Ecological interpretation of ordination space

The first four factor axes explain 46% of the total variance (Table 1). We assumed that the strongest correlations with each axis reflect the main vegetation gradients captured by it. Factor 1 explains a relatively large proportion (22%) of the total variance. It mainly distinguishes communities dominated by a tall and rich tree layer (negative loadings) and those dominated by vine, shrub or herbaceous life forms (positive loadings). Although explaining considerably smaller proportions of the total variance, the remaining factors are still useful for identifying the vegetation classes. Factor 2 separates plant communities by their degree of coverage and richness of herbaceous species. Factor 3 mainly represents variation in biomass of two trees, *Brosimum latescens* and *Mouriri guianensis*, and one shrub, *Psychotria capitata*. Factor 4 mainly represents variation in the biomass of shrubs and of two species, the medium-sized tree *Sapium obovatum* and the shrub *Ruprechtia brachycephala*.

## 5.2 Defining plant communities through ecological interpretation of clusters

The seven clusters are indicated in a scatter plot of the different factors (Fig. 4) and are identified as: *Monodominant forest*, *Shrubland*, *Alluvial seasonal semideciduous forest (Alluvial forest)*, *Alluvial seasonal semideciduous low forest (Alluvial low forest)*, *Seasonally flooded grass-woody savanna (Grassland)*, *Low open tree and shrub savanna (Open savanna)* and *Low dense tree and shrub savanna (Dense savanna)*. The

**BGD**

7, 6889–6934, 2010

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

number of samples in each cluster and their distribution over the ordination space express the structural and floristic variability found within the community and which communities have dominated the floodplain landscape. Table 2 provides a statistical summary of structural and floristic characteristics of communities.

5 A number of communities show overlapping ranges of scores on some of the factor axes, while other factor axes provide clear boundaries between these communities. For instance, the transitions between *Alluvial forest* and *Monodominant forest* are smooth (Fig. 4a–c). These two communities are mainly separated through the tree biomass and coverage of herbaceous species in *Monodominant forest* (Fig. 4a) and the dominance of *Brosimum latescens* and *Mouriri guianensis* in *Alluvial fores* (Fig. 4b). *Dense savanna* lies between *Open savanna* and *Monodominant forest*. Richness and coverage of herbaceous life form distinguish these communities (Fig. 4a). *Dense savanna*, *Grassland*, *Open savanna* and *Monodominant forest* have similar correlation values with Factor 2, indicating that there may be a small variation in coverage of herbaceous species between these communities. The low tree biomass in *Shrubland* is responsible for its positive scores on the first factor. *Alluvial low forest* is distinguished from other forests based on Factor 4. Its high shrub coverage compared to *Shrubland* and the dominance of *Sapium obovatum* and *Ruprechtia brachycephala* generates scores on Factor 4 in intermediate position between *Shrubland* and *Alluvial forest*.

## 20 6 Mapping plant communities

### 6.1 Remote sensing and ancillary data

Remotely sensed imagery and ancillary data are frequently used in spatial vegetation modeling due to their capability of providing accurate environmental information related to vegetation patterns (Guisan and Zimmerman, 2000; Pfeffer et al., 2003; Miller et al., 25 2007). An IKONOS-2 image and a Digital Elevation Model (DEM) (Fig. 1c and d) were used in this study to derive variables related to vegetation patterns. The acquisition

**BGD**

7, 6889–6934, 2010

---

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

date of the IKONOS-2 image is 1 October 2003 corresponding to the dry season in the Pantanal and representing an optimal time for detecting spectral signatures of terrestrial vegetation on the floodplain, due to the availability of cloud free images and non-flooded soil conditions. The IKONOS-2 image consists of four spectral bands: three bands in the visible part of the spectrum located at blue (450–520 nm), green (520–600 nm) and red (630–690 nm) and one band in Near Infrared (760–900 nm). The pixel size is approximately 4 by 4 m. The registered radiance values by the IKONOS-2 sensor were converted to reflectance values using the calibration information provided by Bowen (2002). A Normalized Difference Vegetation Index (NDVI) was computed from the spectral bands by taking a ratio of the difference of the near infrared and red spectral bands and the sum of the near infrared and red band (Tucker, 1979). Such an NDVI image shows stronger contrast between vegetation and soil and water surfaces while reducing noise in the image. Furthermore, we applied a principal component (PC) transformation to the IKONOS-2 image to reduce inter-band correlation and extract new spectral information that arises from this transformation. The four original bands, the NDVI image and the four principal component images were used for further data analysis as described in the next section. Ancillary data, such as soil and topography maps, in combination with multi-spectral bands have been used to improve classification of wetlands (Ozesmi and Bauer, 2002). The 90-m resolution DEM of the study area was obtained from the SRTM (NASA Shuttle Radar Topography Mapping Mission) and used to provide continuous information of canopy height rather than soil surface (Jacobsen, 2006) (Fig. 1d).

### 6.2 Re-scaling and extracting image derived data

The original geodata with cell sizes of 4 m (IKONOS-2) and 90 m (DEM) were re-sampled to the support of the field data, i.e., to the plot size used to take measurements of large tree species (50 × 40 m). The resampling technique applied consists of: (1) delineating the irregular plot boundaries in the IKONOS-2 image using ARC/INFO GIS software (version 9.0; ESRI, 2006); (2) calculating average remote sensing and

elevation values for exactly these digitized plots; and (3) extracting variables from the IKONOS-2 derived images and SRTM DEM for the 115 plots to be used in the analysis. The two last steps were done in the PCRaster interactive raster GIS environment (PCRaster, 2002; Wesseling et al., 1996).

### 6.3 Correlating field data and image derived data

We first examined the relationship between image and DEM derived variables and the vegetation patterns captured in the four factorial axes. The functional relationships between each factor axis and image and DEM derived variable were found by Pearson's correlation analysis, to facilitate the ecological interpretation of the variables (James and McCulloch 1990). Next, the relationship between the image derived variables and the factor axes was found using the following multiple linear regression model:

$$Y_i = a_0 + a_1x_{1i} + a_2x_{2i} + \dots + a_px_{pi} + \varepsilon_i \quad (3)$$

where  $Y_i$  is the score value,  $a_0, a_1, \dots, a_p$  are the model parameters,  $x_{1i}, x_{2i}, \dots, x_{pi}$  are the values of the image derived variables and  $\varepsilon_i$  are uncorrelated residuals. The analyses were done with log transformed reflectance values to ensure that the statistical distribution of the data is close to Gaussian (Draper and Smith, 1998).

Before performing the multiple regression analysis, image derived variables were selected to be included in the multiple regression models using the best-subset regression method (Hofmann et al., 2007). In this method, all combinations of explanatory variables in regressions are tested, and Mallow's C-p statistic (Mallows, 1973) is used as eliminatory criterion of variables (Draper and Smith, 1998). We consider the best regression equation for each factor that one combining lowest C-p value and lowest number of explanatory variables.

### 6.4 Vegetation patterns captured by digital images

Table 3 shows correlations between explanatory variables and the factor axes. Except for NDVI, all image derived data present significant correlation with Factor 1. The

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

strongest correlations with this first axis are found with blue, green and red bands, PC1 and canopy topography. Lower reflectance values in the three spectral bands and lower score values in the PC1-3 images are linked to areas occupied by communities with high stored tree biomass such as *Monodominant forest* and *Alluvial forest* (Table 3).

5 Lower score values in the PC4 reflects communities with lower tree biomass values even though this axis explains the noise from the spectral band transformation. In spite of its weak correlation with Factor 1, NDVI shows an expected spectral behaviour: the values decrease toward areas with lower tree biomass, such as those areas covered by *Grassland*, *Open savanna*, *Shrubland* and *Dense savanna*. The strong negative correlation between canopy topography and Factor 1 shows that the boundaries between communities dominated by trees and those dominated by shrubs, lianas and herbs are detected by differences in canopy height.

The variability in cover degree and richness of herbaceous life forms expressed by the second axis is best described by the PC2 image (61%; Table 3). Communities with higher and richer coverage of herbaceous species such as *Grassland*, *Open savanna* and *Dense savanna* are associated with higher reflectance values in blue, green and red bands and higher score values in the PC2 image. The negative correlations between Factor 2 and infra-red band and NDVI show that communities dominated by herbaceous species present weaker spectral response to these two images.

20 As observed earlier, Factor 3 mostly justifies the spatial distribution pattern of three tree species that dominate in *Alluvial forest*. Relatively to *Monodominant forest*, the lower biomass content and canopy height of *Alluvial forest* might be the cause for the negative correlations between Factor 3 and NDVI and Factor 3 and canopy topography.

The spatial variability of Factor 4 represents vegetation patterns that are mainly explained by canopy topography (i.e. DEM) showing that areas with higher biomass of shrubs are associated with lower canopy height.

25 The equations found in the multiple regression analysis are shown in Table 4. The regression models significantly explain 70.4%, 66.3%, 31.3% and 25.6% of the variance in Factors 1 to 4, respectively.

## 6.5 Variogram analysis

We applied variogram analysis on the residuals of the multiple linear regression to derive information on their spatial structure (Wagner and Fortin, 2005). This information was used for two reasons: (1) to investigate the spatial autocorrelation associated with the observed vegetation patterns; (2) to use this information when making spatial predictions (Miller et al., 2007). Sample variograms were estimated and variogram models fit using the function *autofitVariogram* from the library *automap* (Hiemstra et al., 2008) in the statistical environment R (R Development Core Team, 2009).

The results indicate that the vegetation gradients represented by the residuals of each factor (Factor 1–4) vary on different spatial scales (Fig. 5). Variograms of the Matérn family, a family of semivariogram models where the degree of smoothness of the random field is controlled through a shape parameter ( $\kappa$ ) (Pardo-Iguzquiza and Chica-Olmo, 2008), were fit for Factors 1, 2 and 4, Fig. 5a,b,d), whereas an exponential variogram (special case of the Matérn family) was fit for Factor 3 (Fig. 5c). The first and third factors show large-scale patterns as revealed by their ranges of spatial dependence. The variogram of Factor 1 has a range of 3380 m, whereas the variogram of Factor 3 is monotonically increasing within the extent of the sample variogram and consequently has a larger range. The variograms of Factors 2 and 4 show short ranges of spatial dependence (close to a pure nugget effect) suggesting that processes governing their spatial patterns show small scale variability.

## 6.6 Universal kriging

Universal kriging is a spatial interpolation technique that can incorporate environmental data and spatial dependence in the modeled error to predict at locations without observations and generate accurate vegetation distribution maps (Pfeffer et al., 2003; Pebesma and Wesseling, 1998). Universal kriging was done on the regression residuals and the interpolated residuals were added to a trend surface to predict factor scores at unobserved locations. This trend surface was based on the regression equation in

**BGD**

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Eq. (3) (Pfeffer et al., 2003). The predicted scores were used to create four factor score maps. In addition, the universal kriging approach was used to estimate the prediction error (standard deviation), which is typically increasing as a function of the distance to observation locations (Stein and Corsten, 1991).

## 5 Continuum representation of vegetation spatial patterns

The score maps in Fig. 5 show the vegetation spatial patterns predicted by universal kriging. The score maps of the first and third factor axes (Fig. 5Aa and c) show mainly large-scale variability. These axes, as mentioned earlier, mostly represent spatial variation of tree life forms. Contrarily, the score maps of the second and fourth axes (Fig. 5b and d) representing the occurrence of herbaceous and shrub layers, respectively, show small-scale spatial variability.

Examining the pattern of the prediction errors of the scores for each factor axis (Fig. 6), one can infer to which extent sample data and image data contribute to predictions. When the range of the semivariogram is large, as seen in the semivariograms of Factor 1 and 3 (Fig. 5a,c), the prediction errors increase slowly with the distance away from samples. On the other hand, a short range in the variogram results in prediction errors increasing rapidly with distance away from samples, as is the case with Factor 2 and 4. Image data will in this case have greater impact on predictions. Nevertheless, the quality of the factor score maps is not only related to differences between small-scale and large-scale spatial variation but rather reflects the explanatory strength of the relationship between factor axes and image derived variables as shown by the mean error in the score maps. According to these averages, Factor 1 represents the most accurate map (mean SD = 0.51) followed by Factor 2 (mean SD = 0.64), Factor 3 (mean SD = 0.69) and Factor 4 (mean SD = 0.87).

**BGD**

7, 6889–6934, 2010

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





## 6.7 Spatial distribution of plant communities across the floodplain

In the final part of this procedure, we combined results from spatial and non-spatial analyses generated as described in the former sections to create the final map of plant communities. The clusters/communities centers, calculated in the section “Identifying plant communities” were used to assign each location on the map to a community class. This was done by calculating Euclidean distances between centers and predicted scores values. Each location was then assigned to the community whose center was nearest to the predicted score values at that location.

The map of plant communities (Fig. 7a) resulting from this classification method shows the predicted spatial distribution of the seven identified plant communities on the floodplain. *Grassland* (16% of coverage), *Shrubland* (30% of coverage) and *Monodominant forest* (32% of coverage) sum up to 78% of the coverage at the studied site. These communities mostly appear as large and contiguous patches across the site. *Alluvial forest* and *Alluvial low forest*, as expected, appear as strips covering exclusively places close to water bodies: along rivers, channels and surrounding *baías*, i.e. temporary or permanent lakes seasonally connected to the river. These two communities cover just 4% (2% each) of the studied floodplain. The greatest portion of the 10% of *Open savanna* that covers the study area is located towards the Northern boundary. The 8% of *Dense savanna* is found as small patches generally surrounded by *Open savanna* and as a big patch beside *Monodominant forest*.

## 6.8 Evaluating uncertainty

Vegetation mapping using statistical approaches carries different sources of uncertainties related to sampling scheme, interpolation errors, sampling support, data quality, lack of data and others, which may compromise the model’s capability of accurately predicting vegetation patterns (Guisan and Zimmerman, 2000; Pfeffer et al., 2003; Miller et al., 2007). The predictive success of our mapping approach was evaluated

**BGD**

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



using cross-validation (Efron and Tibshirani, 1986) and random-simulations (Bourenane et al., 2007), both performed in R (R Development Core Team, 2009).

### 6.8.1 Cross-validation

We have used cross-validation to investigate the sensitivity of vegetation predictions performed by universal kriging as a result of sampling variability (Pfeffer et al., 2003). Two resampling techniques were applied: *leave-one-out cross-validation (LOOCV)* and *leave-five-out cross-validation (LFOCV)*. The first technique is the standard procedure (Efron and Tibshirani, 1986) which consists of omitting one sample at a time from the data set and based on the remaining observed values make predictions at this location using the interpolation technique, i.e., universal kriging. Because samples/plots within the same trail are considerably closer to other observations than the typical distance between prediction locations and observations locations (Miller et al. 2007), LFOCV was used to test the prediction quality of the model when the whole trail, that is, five plots, is left out to make predictions. Vegetation classes were assigned from the predicted scores and compared with the observed vegetation classes at the 115 sample plots.

Overall agreement between predicted and observed classes does not differ substantially between the two resampling techniques: *leave-one-out* results in 52.2% agreement and *leave-five-out* in 48.7% agreement. Both techniques show that accuracy in classification varies according to the community type (Fig. 7b and c). Communities which have been observed on a large number of plots and occupy large portions of the vegetation map, such as *Monodominant forest* and *Shrubland*, are less sensitive to sampling density than those communities which occur in smaller and few patches, such as *Alluvial forest* and *Alluvial low forest*. Consequently, communities observed in few of the plots are wrongly classified also for LOOCV (Fig. 7b). Other possible causes of uncertainty in classification from our mapping approach derives from the similarity between community types having a small distance between cluster centers in the ordination space (Fig. 4). Communities such as *Alluvial forest* and *Dense savanna*

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



are frequently predicted to be their neighboring communities, namely, *Monodominant forest*; and *Alluvial low forest* are frequently predicted to be *Shrubland* (Fig. 7b and c).

## 6.8.2 Simulation

A Monte Carlo approach was applied to examine the uncertainty of our method (Legendre and Legendre, 1998). In this approach, we performed the same universal kriging, however creating random realizations of score maps conditioned to the observations instead of predicted values as was done in the original procedure. This was done by simulating 1000 realizations of score maps for each factor with Gstat (Pebesma, 2004), based on the scores at the observation locations and the fit variograms. These realizations reflect the prediction uncertainty at the prediction locations; all realizations are equally probable. For each realization, we calculated the vegetation pattern, using the same Euclidean distance algorithm applied in the original mapping procedure. This was repeated for all 1000 realizations, resulting in 1000 realizations of vegetation community maps. Two realizations are shown in Fig. 8b and c. From these 1000 realizations, we created a map showing the probability, from 0 to 1, that a certain community is found in a 40 m grid cell (Fig. 8). On this map, a value 1 indicates zero prediction uncertainty.

The result shows that the quality of classification varies spatially, even though the proportion and arrangement of communities observed in the original map is preserved to a great extent. The central zone of a community patch is more likely to be classified correctly than border areas, as shown by the increasing probabilities towards the center of patches of communities (Fig. 8a). This might reflect intrinsic uncertainties in classification of natural ecotones reflected in the overlapping of score values of very close communities in the factor space. The quality of classification also varied between communities. Classification of *Dense savanna* and *Open savanna*, for instance, exhibit lower probabilities of being in the correct class as indicated by their more random distribution across the landscape (Fig. 8b and c). Here, sampling configuration and distance between clusters in factor space are an important source of errors.

# Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 7 Flood duration-vegetation relationship

The relationship between vegetation distribution and flooding was assessed by comparing the plant community map with a flood duration map as in direct gradient analysis. The flood duration map (Fig. 9) was created from a digital elevation map and 38 years of daily recordings of the water level in the River Cuiabá (Fig. 1b) provided by the Brazilian National Water Agency (ANA; <http://hidroweb.ana.gov.br>). The 40-m resolution digital elevation map was created with universal kriging from 81 GPS elevation measurements at the site and using SRTM DEM as an auxiliary variable (Valeriano and Abdon, 2007). A base station was installed for increased precision of the GPS measurements. Flood duration and flood depth data were also monitored by direct reading of staff gauges for two years (2007–2008) at the 23 sampling trails. The relationship between flooding and elevation data was tested with Pearson's correlation coefficient. Statistically significant and strong correlations were found among them ( $r > 70\%$ ;  $P \leq 0.05$ ) indicating the possibility of calculating flooding duration values over the floodplain through the indirect relationship between river water depth and elevation. Flood duration of a cell was calculated by comparing the water level in the river and the topographical elevation of the cell for each day as follows: if the elevation value at a cell was lower than the water level in the river on a certain day, the cell was considered flooded that day. This approach ignores spatial variation in water level associated with downstream gradients in water level, local depressions containing water that is only partially connected with the main river, and surface water fed by groundwater. However, the effect of these processes is relatively small as indicated by additional field sampling with the staff gauges.

The flood duration map (Fig. 9) shows the number of flooded days per year in the study area. Flood duration data extracted from this map were classified into monthly intervals and the distribution of the plant communities found in the vegetation map along this flooding gradient was plotted in Fig. 10. Figure 10 shows that the zonation of plant communities along the floodplain is clearly related to the duration of inundation. *Alluvial forest* and *Dense savanna* occur in areas with a flooding duration of less than

**BGD**

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



two months. *Monodominant forest*, although occupying a high proportion of the highest areas, has the highest occurrence at intermediary flooding conditions, with a flooding duration between two and four months. *Open savanna* is mostly found where flooding lasts for four to six months per year. *Grassland* is found under almost the whole range of flooding durations, however with peaks of occurrence in areas with a flooding duration below two months and between four and six months of inundation. *Alluvial low forest* is mostly situated at locations with a flooding duration between 6–8 months. *Shrubland* dominates the areas with the highest flooding duration. Above eight months of flood duration, there is no suitable condition for tree species establishment and the landscape is occupied mostly by *Shrubland*, *Open savanna* and *Grassland*. The occurrence of monodominant forest in the last flood duration class might be associated with the coarse representation of spatial variation in flood duration, illustrated in Fig. 9.

## 8 Discussion

We showed that it is possible to classify vegetation at locations in the studied floodplain by measuring structural and floristic attributes of different vegetation layers (herbaceous, tree, shrub and vines), and combining these data with remote-sensing imagery and DEM data. The plant communities described in an existing classification could be clearly identified as clusters in the ordination space, thanks to the floristic properties included in the analyses that differentiated structurally similar but floristically different plant communities. However, sometimes clusters showed overlap on a number of factor axes and boundaries between clusters were not always accurate. Such overlap probably indicates the existence of gradual changes in vegetation (Brzeziecki et al., 1993), which is not represented in our model with sharp boundaries between vegetation communities. Thus, the vegetation community studied deviates slightly from our crisp plant community model. This had two implications for our analysis. One is the subjective determination of cluster boundaries in the ordination space, particularly in cases where boundaries were not crisp. The other is related to the interpretation of

**BGD**

7, 6889–6934, 2010

### Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the uncertainty analysis. One of the causes of uncertainty of the mapped vegetation is the uncertainty in the assignment of an interpolated point to a cluster in the ordination space. Overlap of clusters in the ordination space may actually represent transition zones between plant communities, and are related to intrinsic uncertainty in classification (see also, Fortin et al., 2000; Hernandez-Stefanoni and Dupuy, 2007). Potential misclassification in these zones appears as uncertainty on the interpolated crisp map, particular in areas close to mapped boundaries between plant communities. However, the assumption of crisp plant communities as a spatial concept is in most cases sufficient to interpret vegetation patterns, as in our study (see also Austin and Smith, 1989; Brzeziecki et al., 1993).

The statistical approach described here shows the value of integrating field observations and high resolution remote sensing. The field observations provide enough information to identify vegetation communities, while remote sensing reduces the interpolation error because derivatives of the remote sensing image explain a significant part of the spatial variation in vegetation. The use of universal kriging is valuable here because parts of the remaining variability can be modeled as a spatially correlated residual. These findings confirm that the use of remote sensing and a spatial interpolation method reduce the uncertainty in mapped vegetation, as was also stressed in other studies (Guisan and Zimmerman, 2000; Ferrier et al., 2002; Pfeffer et al., 2003; Miller et al., 2007).

The different techniques used to evaluate the behavior of the statistical model used for mapping vegetation, e.g. cross-validation and Monte Carlo simulation, allowed us to identify possible causes of misclassification and determine spatial prediction uncertainty (Congalton and Green, 1999; Guisan and Zimmerman, 2000; Pfeffer et al., 2003). The accuracy levels of the vegetation map derived from the mapping procedure described here and assessed by cross-validation (e.g. 49 of 52%) were of the same magnitude as to those found by Pfeffer et al. (2003) in their maps of Alpine vegetation (e.g. 50 to 65%). The uncertainties in vegetation classification that resulted from the sampling density and configuration suggest that the map quality may be improved



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

vegetation distribution of these communities (Fig. 10) (Austin, 2002). Based on these findings, we conclude that vegetation zonation in the studied floodplain might be influenced not just by physiographic limits from flood duration, as stressed in most of the studies in the Pantanal (Junk, 1989; Nunes da Cunha and Junk, 1999, 2000; Zeilhofer and Schessl, 2000), but also by biological constraints related to competition between neighbors (Tilman, 1994).

The significant advantage of the mapping approach described in this paper is that detailed biological information present in field observations can be integrated with high spatial resolution remotely sensed data producing accurate vegetation maps. Different from “classical” approaches to vegetation class mapping, our modeling carries quantitative information on vegetation variability allowing future application in modeling concerned with the effects of environmental shifts on biological patterns and processes (Arieira et al., 2010; Brzezicki et al., 1993). We believe that mapping of plant communities by integrating field observations and high-resolution imagery is a promising approach for conservation assessment and long-term ecological monitoring in extensive wetland areas.

*Acknowledgements.* The authors are grateful to the Brazilian governmental agencies, CAPES and CNPq, for the financial support. Helpful comments and assistance were provided by P. Girard, Peter Zeilhofer, Arnildo Pott, Vali J. Pott, Sandra Santos and José F. M. Valls. We also thanks to the Social Service of the Commerce (SESC) and technicians and students of the Federal University of Mato Grosso, for the technical support in field work.

## References

- Ab'Saber, A. N.: O Pantanal Matogrossense e a teoria dos refúgios, *Rev. Bras. Geog.*, 50, 9–57, 1988.
- Alvarenga, S. M., Brasil, A. E., Pinheiro, R., and Kux, H. J. H.: Estudo geomorfológico aplicado à Bacia do alto Rio Paraguai e Pantanaís Matogrossenses, *Boletim Técnico Projeto RADAM/BRASIL, Série Geomorfologia*, Salvador, 187, 89–183, 1984.



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Arieira, J., Karssenbergh, D., Couto, E. G., and Nunes da Cunha, C.: Modeling wetland vegetation dynamic based on spatio-temporal interaction and flooding tolerance of plant communities in the Pantanal Mato-Grossense (Brazil), Federal University of Mato Grosso, Cuiabá, Brazil/ Utrecht University, Utrecht, The Netherlands, in preparation, 2010.
- 5 Assine, M. L. and Soares, P. C.: Quaternary of the Pantanal, west-central Brazil, *Quatern. Int.*, 114, 23–34, 2004.
- Austin, M. P.: Spatial prediction of species distribution: an interface between ecological theory and statistical modelling, *Ecol. Model.*, 157, 101–118, 2002.
- Austin, M. P. and Smith, T. M.: A new model for the continuum concept, *Vegetatio*, 83, 35–47, 10 1989.
- Barbosa, R. I. and Ferreira, C. A. C.: Biomassa acima do solo de um ecossistema de “campina” em Roraima, norte da Amazônia Brasileira, *Acta Amazon.*, 34, 577–586, 2004.
- Bascompte, J. and Solé, R. V.: Habitat fragmentation and extinction thresholds in spatially explicit models, *J. Anim. Ecol.*, 65, 465–473, 1996.
- 15 Bourennane, H., King, D., Couturier, A., Nicoullaud, B., Mary, B., and Richard, G.: Uncertainty assessment of soil water content spatial patterns using geostatistical simulations: an empirical comparison of a simulation accounting for single attribute and a simulation accounting for secondary information, *Ecol. Model.*, 205, 323–335, 2007.
- Bray, J. R. and Curtis, J. T.: An ordination of the upland forest communities of southern Wisconsin, *Ecol. Monogr.*, 27, 325–349, 1957.
- 20 Brzeziecki, B., Kienast, F., and Wildi, O.: A simulated map of the potential natural forest vegetation of Switzerland, *J. Veg. Sci.*, 4, 499–508, 1993.
- Bullock, J.: Plants, in: *Ecological census techniques: a handbook*, edited by: Sutherland, W. J., Cambridge Univ. Press, New York, 111–138, 1996.
- 25 Chambers, J. Q., Asner, G. P., Morton, D. C., Morton, D. C., Anderson, L. O., Saatchi, S. S., Espírito-Santo, F. D. B., and Souza Jr., M. P. C.: Regional ecosystem structure and function: ecological insights from remote sensing of tropical forests, *Trends Ecol. Evol.*, 22, 414–423, 2007.
- Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., Folster, H., Fromard, F., Higuchi, N., Kira, T., Lescure, J. P., Nelson, B. W., Ogawa, H., Puig, H., Riera, B., and Yamakura, T.: Tree allometry and improved estimation of carbon stocks and balance in tropical forests, *Oecologia*, 145, 87–99, 2005.
- 30





## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

last access: 15 May 2009, Biota Neotropica, 5, 2005.

McIntosh, R. P.: Plant Communities, Science, 128, 115–120, 1958.

Minchin, P. R.: Montane vegetation of the Mt. Field Massif, Tasmania: a test of some hypotheses about properties of community patterns, Plant Ecol., 83, 97–110, 1989.

5 Miller, J., Franklin, J., and Aspinall, R.: Incorporating spatial dependence in predictive vegetation models, Ecol. Model., 202, 225–242, 2007.

Mueller-Dombois, D. and Ellenberg, H.: Aims and Methods of Vegetation Ecology, John Wiley & Sons, New York, 1974.

10 Nunes da Cunha, C. and Junk, W. J.: Composição florística de capões e cordilheiras: localização das espécies lenhosas quanto ao gradiente de inundação no Pantanal de Poconé, MT, Brasil, in: Anais do II Simpósio sobre Recursos Naturais e Sócio-Econômicos do Pantanal – Manejo e Conservação, Embrapa – CPAP, Corumbá, Brasil, 387–406, 1999.

15 Nunes da Cunha, C. and Junk, W. J.: The impact of flooding on distribution of woody plant communities in the Pantanal of Poconé, Mato Grosso, Brazil, in: German-Brazilian Workshop on Neotropical Ecosystems – Achievements and Prospects of Cooperative Research, Session 4: Living Resource Management: Approaches, Techniques, Variability, Hamburg, Germany, 3–8, 2000.

20 Nunes da Cunha, C., Rawiel, P., Wantzen, K. M., Junk, W. J., and Prado, A. L.: Mapping and characterization of vegetation units by means of Landsat imagery and management recommendations for the Pantanal of Mato Grosso (Brazil), north of Poconé, Amazoniana, XIX, 1–32, 2006.

Ozesmi, S. L. and Bauer, M. E.: Satellite remote sensing of wetlands, Wetl. Ecol. Manag., 10, 381–402, 2002.

25 Pardo-Iguzquiza, E. and Chica-Olmo, M.: Geostatistics with the Matérn semivariogram model: a library of computer programs for inference, kriging and simulation, Comput. Geosci., 34, 1073–1079, 2008.

Parolin, P.: Submerged in darkness: adaptations to prolonged submergence by woody species of the Amazonian floodplains, Ann. Bot., 103, 359–376, 2009.

30 PCRaster: PCRaster Software, Info at <http://www.geog.uu.nl/pcraster.nl>, last access: 8 February 2009, 2002.

Pebesma, E. J.: Multivariable geostatistics in S: the gstat package, Comput. Geosci., 30, 683–691, 2004.

**Integrating field  
sampling, spatial  
statistics and remote  
sensing in Brazil**

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Pebesma, E. J. and Wesseling, C. G.: GSTAT: a program for geostatistical modelling, prediction and simulation, *Computat. Geosci.*, 24, 17–31, 1998.
- Pfeffer, K., Pebesma, E. J., and Burrough, P. A.: Mapping alpine vegetation using vegetation observations and topographic attributes, *Landscape Ecol.*, 18, 759–776, 2003.
- 5 Pillar, V. D. and Sosinski Jr., E. E.: An improved method for searching plant functional types by numerical analysis, *J. Veg. Sci.*, 14, 323–332, 2003.
- R Development Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, <http://www.R-project.org>, last access: 8 February, 2009.
- 10 Scheller, R. and Mladenoff, D.: An ecological classification of forest landscape simulation models: tools and strategies for understanding broad-scale forested ecosystems, *Landscape Ecol.*, 22, 491–505, 2007.
- Schöngart, J., Arieira, J., Felfili Fortes, C., Cezarine de Arruda, E., and Nunes da Cunha, C.: Carbon dynamics in aboveground coarse wood biomass of wetland forests in the northern Pantanal, Brazil, *Biogeosciences Discuss.*, 5, 2103–2130, doi:10.5194/bgd-5-2103-2008, 2008.
- 15 Silva, M. P., Mauro, R., Mourão, G. E., and Coutinho, M.: Distribuição e quantificação de classes de vegetação do Pantanal através de levantamento aéreo, *Rev. Bras. Bot.*, 23, 143–152, 2000.
- 20 Stein, A. and Corsten, L. C. A.: Universal kriging and cokriging as a regression procedure, *Biometrics*, 47, 575–587, 1991.
- Svenning, J. C., Kinner, D. A., Stallard, R. F., Engelbrecht, B. M. J., and Wright, S. J.: Ecological determinism in plant community structure across a tropical forest landscape, *Ecology*, 85, 2526–2538, 2004.
- 25 Tilman, D.: *Plant Strategies and the Dynamic and Structure of Plant Communities*, Princeton University Press, Princeton, New Jersey, 1988.
- Tilman, D.: Competition and biodiversity in spatially structured habitats, *Ecology*, 75, 2–16, 1994.
- Tucker, C. J.: Red and photographic infrared linear combinations for monitoring vegetation, *Remote Sens. Environ.*, 8, 127–150, 1979.
- 30 Turner, M. G.: Landscape ecology: the effect of pattern on process, *Annu. Rev. Ecol. Syst.*, 20, 171–197, 1989.

**Integrating field sampling, spatial statistics and remote sensing in Brazil**

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Turner, M. G., Gardner, R. H., and O'Neill, R. V.: Landscape Ecology in Theory and Practice – Pattern and Process, Springer-Verlag, New York, Inc., USA, 2001.
- Velloso, H. P., Rangel Filho, A. L. R., and Lima, J. C. A.: Classificação da Vegetação Brasileira Adaptada a um Sistema Universal, IBGE, RJ, BR, 1991.
- 5 Wagner, H. H. and Fortin, M. J.: Spatial analysis of landscapes: concepts and statistics, Ecology, 86, 1975–1987, 2005.
- Wantzen, K. M., Drago, E., and da Silva, C. J.: Aquatic habitats of the Upper Paraguay River-Floodplain-System and parts of the Pantanal (Brazil), Ecohydrol. Hydrobiol., 5, 107–126, 2005.
- 10 Wesseling, C. G., Deursen, W. P. A., and Burrough, P. A.: A spatial modelling language that unifies dynamic environmental models and GIS, in: Proceeding, Third International Conference/Workshop of Integrating GIS and Environmental Modeling, Santa Fé, NM, <http://www.ncgia.ucsb.edu/conf/SANTAFECD-ROM/main.html>, last access: 15 May 2009, Santa Barbara, CA, National Center for Geographic Information and Analysis, 1996.
- 15 Whittaker, R. W.: Gradient analysis of vegetation, Biol. Rev., 49, 207–264, 1967.
- Zeilhofer, P. and Schesl, M.: Relationship between vegetation and environmental conditions in the northern Pantanal of Mato Grosso, J. Biogeogr., 27, 159–168, 2000.

**Table 1.** Summary statistics for the factor analysis. Numbers in bold highlight the highest correlation with the factor axis.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Richness tree	-0.76	-0.25	0.17	-0.28
Richness shrub	0.28	0.07	0.20	-0.35
Richness herb	0.31	0.58	0.43	-0.11
Canopy height	-0.86	-0.05	-0.33	0.09
Cover % herb	0.04	<b>0.79</b>	0.16	0.06
Cover % vine	0.82	-0.15	-0.16	0.16
Richness vine	0.73	-0.09	-0.14	0.16
Biomass tree (total)	<b>-0.88</b>	0.00	-0.26	0.18
Biomass shrub	0.54	-0.42	0.07	<b>-0.58</b>
Biomass (DBH 10 cm > 30 cm)				
<i>Vochysia divergens</i>	-0.31	-0.05	0.05	0.03
<i>Sapium obovatum</i>	-0.03	-0.23	-0.09	-0.56
<i>Licania parvifolia</i>	-0.43	-0.03	-0.17	0.11
<i>Brosimum latescens</i>	-0.29	-0.34	0.51	0.22
<i>Trichilia catigua</i>	-0.30	-0.27	0.45	0.06
<i>Duroia duckei</i>	-0.59	0.09	-0.35	0.19
<i>Cecropia pachystachya</i>	-0.31	0.23	-0.20	-0.02
<i>Mouriri guianensis</i>	-0.51	-0.35	0.36	0.22
Biomass (DHB > 30 cm)				
<i>Vochysia divergens</i>	-0.72	0.17	-0.47	0.21
<i>Mouriri guianensis</i>	-0.37	-0.42	<b>0.55</b>	0.23
Biomass (shrub)				
<i>Albizia polycephala</i>	0.62	-0.11	0.02	0.12
<i>Ruprechtia brachycephala</i>	-0.02	-0.25	0.05	-0.50
<i>Peritassa dulcis</i>	0.18	-0.52	-0.46	-0.06
<i>Melochia villosa</i>	0.41	0.16	0.18	0.24
<i>Byrsonima cydoniifolia</i>	-0.23	-0.40	0.43	0.15

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Table 1.** Continued.

Variable	Factor 1	Factor 2	Factor 3	Factor 4
<i>Psychotria capitata</i>	−0.48	−0.49	0.51	0.20
<i>Bauhinia rufa</i>	0.17	0.04	0.09	−0.32
<i>Mimosa pellita</i>	0.57	0.05	0.02	0.35
<i>Laetia americana</i>	0.74	−0.20	−0.09	0.14
<i>Solanum pseudoauriculatum</i>	0.26	0.19	0.15	0.22
<i>Eugenia florida</i>	0.03	−0.29	−0.32	0.07
<i>Alchornia discolor</i>	−0.22	0.37	0.03	−0.37
<i>Mabea paniculata</i>	−0.30	0.12	0.17	−0.25
<i>Byrsonima orbygniana</i>	0.00	0.19	0.25	−0.49
Cover % herbaceous species				
<i>Paspalum hydrophilum</i>	0.34	0.37	0.24	0.05
<i>Panicum guianense</i>	−0.17	0.06	−0.18	−0.35
<i>Scleria bracteata</i>	−0.57	0.24	−0.27	0.16
Cover % vine				
<i>Cissus spinosa</i>	0.67	−0.31	−0.20	−0.001
<i>Aniseia cernua</i>	0.54	0.14	0.09	0.34
<i>Paullinia pinata</i>	0.52	−0.31	−0.27	0.09
<i>Dollicarpus dentatus</i>	0.23	−0.57	−0.27	−0.01
<i>Ipomea rubens</i>	0.36	0.16	0.17	0.22
% Variance	22	9	8	7



**Integrating field sampling, spatial statistics and remote sensing in Brazil**

J. Arieira et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

**Table 2.** Structural and floristic characteristics of plant communities, given as mean and standard deviation.

	Monodominant forest	Shrubland	Alluvial forest	Alluvial Low Forest	Grassland	Open savanna	Dense
	<i>Vochysia divergens</i> Pohl.	<i>Laetia americana</i> L.	<i>Byrsonima cydoniifolia</i> A. Juss.	<i>Ruprechtia brachysepala</i> Meisn.	<i>Paspalum hydrophilum</i> Henrard	<i>Paspalum hydrophilum</i> Henrard	<i>Byrsonima orbignyana</i> A. Juss.
	<i>Duroia duckei</i> Huber	<i>Mimosa pellita</i> Humb. & Bonpl. ex Willd.	<i>Psychotria capitata</i> Ruiz & Pav.	<i>Crataeva tapia</i> L.	<i>Panicum guianense</i> Hitchc.	<i>Hybiscus furcellatus</i> Desr.	<i>Bauhinia rufa</i> (Bong.) Steud.
	<i>Licania parvifolia</i> Huber	<i>Peritassa dulcis</i> (Benth.) Miers	<i>Trichilia catigua</i> A. Juss.	<i>Banara arguta</i> Briq.		<i>Laetia americana</i> L.	<i>Alchornea discolor</i> Poepp.
Characteristic species	<i>Scleria bracteata</i> Cav.	<i>Albizia polycephala</i> (Benth.) Killip <i>Cissus spinosa</i> Cambess.  <i>Aniseia cernua</i> Moric. <i>Paullinia pinnata</i> L. <i>Ipomea rubens</i> Chousy	<i>Mouriri guianensis</i> Aubl.  <i>Brosimum lactescens</i> (S. Moore) C. C. Berg	<i>Sapium obovatum</i> Klotzsch ex Müll. Arg. <i>Cecropia pachystachya</i> Trécul			



## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 2. Continued.

	Monodominant forest	Shrubland	Alluvial forest	Alluvial Low Forest	Grassland	Open savanna	Dense
Richness of herbs (no. of sp. per sample)	1.98 ± 1.92	2.54 ± 1.75	2 ± 0.82	2.4 ± 1.52	3.67 ± 1.56	4.6 ± 1.34	4.2 ± 1.92
Richness of vines (no. of sp. per sample)	2.96 ± 1.52	7.71 ± 1.72	1.71 ± 1.25	4.6 ± 1.82	4.42 ± 1.31	3.8 ± 1.1	1.4 ± 1.52
Richness of shrubs (no. of sp. per sample)	7.85 ± 3.58	9.89 ± 2.36	6.71 ± 4.61	6.4 ± 1.95	8.58 ± 2.91	8.6 ± 2.3	14.6 ± 3.6
Richness of medium sized trees (no. of sp. per sample)	4.91 ± 1.64	0.32 ± 0.61	8 ± 2.65	6.2 ± 2.17	0.17 ± 0.39	1 ± 0.45	5.8 ± 1.64
Richness of large trees (no. of sp. per sample)	2.77 ± 0.71	0.14 ± 0.45	4.71 ± 1.11	1.6 ± 1.34	0.17 ± 0.39	0.6 ± 0.55	2 ± 0.7
Biomass of shrubs (Mg ha <sup>-1</sup> )	3.25 ± 6.71	10.11 ± 3.79	3.82 ± 3.01	13.54 ± 8.18	4.03 ± 2.08	2.63 ± 0.78	9.96 ± 6.71
Biomass of medium sized trees (Mg ha <sup>-1</sup> )	109.58 ± 17.29	2.38 ± 7.58	91.54 ± 29.87	28.50 ± 21.78	4.57 ± 9.56	29.5 ± 31.01	26.75 ± 17.29
Biomass of large trees (Mg ha <sup>-1</sup> )	84.94 ± 11.36	1.87 ± 6.80	57.15 ± 18.55	9.47 ± 8.41	1.81 ± 4.29	24.8 ± 28.06	9.40 ± 11.36
Canopy height (m)	20.19 ± 1.30	2.41 ± 0.84	15.2 ± 3.19	5.72 ± 1.74	1.76 ± 0.21	2 ± 0.1	3.14 ± 1.3
Cover % herbs	37.06 ± 9.55	27 ± 24.26	16 ± 7.66	18.4 ± 11.52	61.33 ± 21.46	97.6 ± 3.58	56.8 ± 9.55
Cover % vines	21.74 ± 6.07	83 ± 11.71	8.57 ± 6.70	47.2 ± 21.05	43.33 ± 19.66	27.2 ± 7.16	5.6 ± 6.07

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.

**Table 3.** Pearson's correlation coefficients between factor axes and image variables: four spectral bands, Normalized Difference Vegetation Index (NDVI), Principal Component transformation to the IKONOS-2 image (PC), and canopy topography derived from DEM-SRTM (DEM). \* $P \leq 0.05$

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Blue band	0.71*	0.30*	0.38*	0.008
Green band	0.71*	0.23*	0.38*	-0.028
Red band	0.67*	0.36*	0.37*	-0.009
Infra-red band	0.43*	-0.34*	0.023	-0.001
NDVI	-0.124	-0.47*	-0.30*	0.012
PC1	0.69*	-0.126	0.25*	-0.028
PC2	0.29*	0.61*	0.30*	0.035
PC3	0.26*	-0.076	0.025	0.095
PC4	-0.33*	0.129	-0.097	-0.1
Canopy topography (DEM)	-0.72*	0.153	-0.39*	-0.163

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

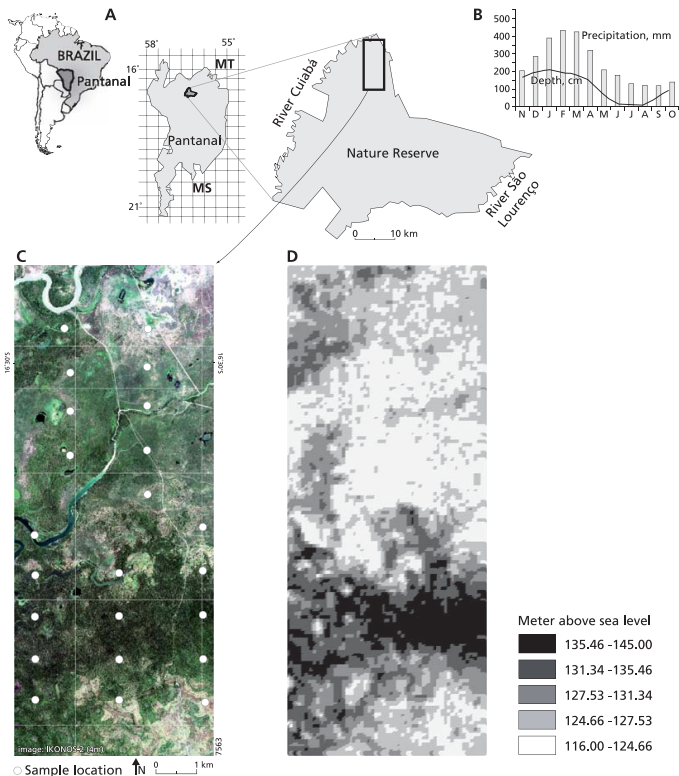
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Fig. 1.** Study site. **(A)** Natural Reserve SESC Pantanal located at the Pantanal Mato-grossense, Mato Grosso; Brazil; **(B)** Mean annual water depth fluctuation of the River Cuiabá (1963–2000) and mean precipitation near Cuiabá, northern Pantanal. Rainfall data from INMET (National Institute of Meteorology of Brazil), river level data from DNAEE (National Department of Waters and Electric Energy of Brazil); **(C)** Four meter resolution multispectral IKONOS-2 image, of the study site acquired in October 2003, true color. White circles are the sampling locations; **(D)** 90-m Resolution SRTM (NASA Shuttle Radar Topographic Mission, <http://ww2.jpl.nasa.gov/srtm/>) Digital Elevation Model of the study area.

6925

**BGD**

7, 6889–6934, 2010

**Integrating field sampling, spatial statistics and remote sensing in Brazil**

J. Arieira et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

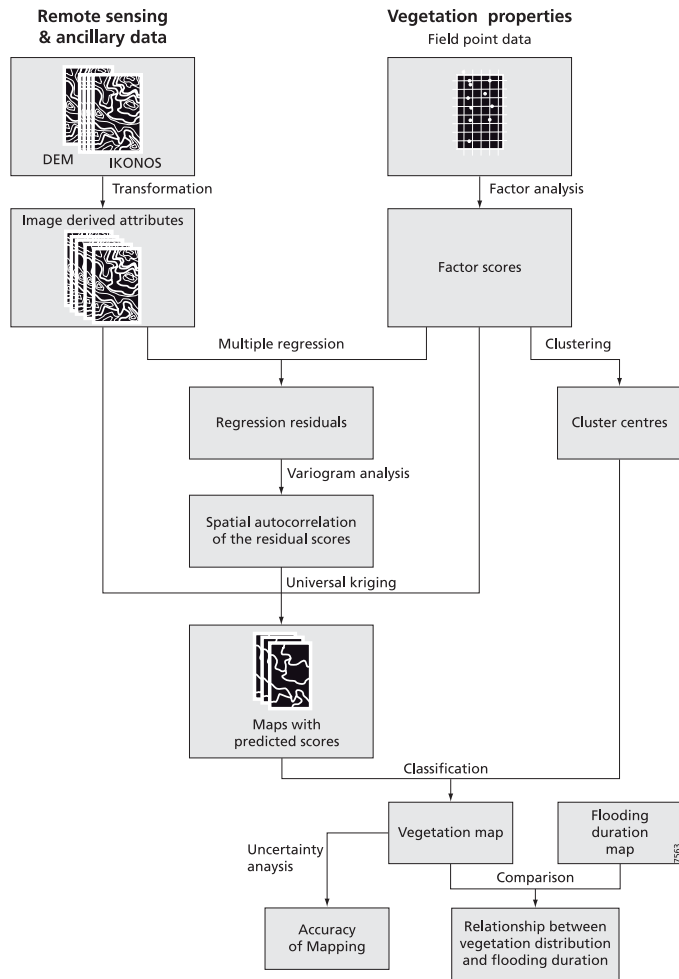
Close

Full Screen / Esc

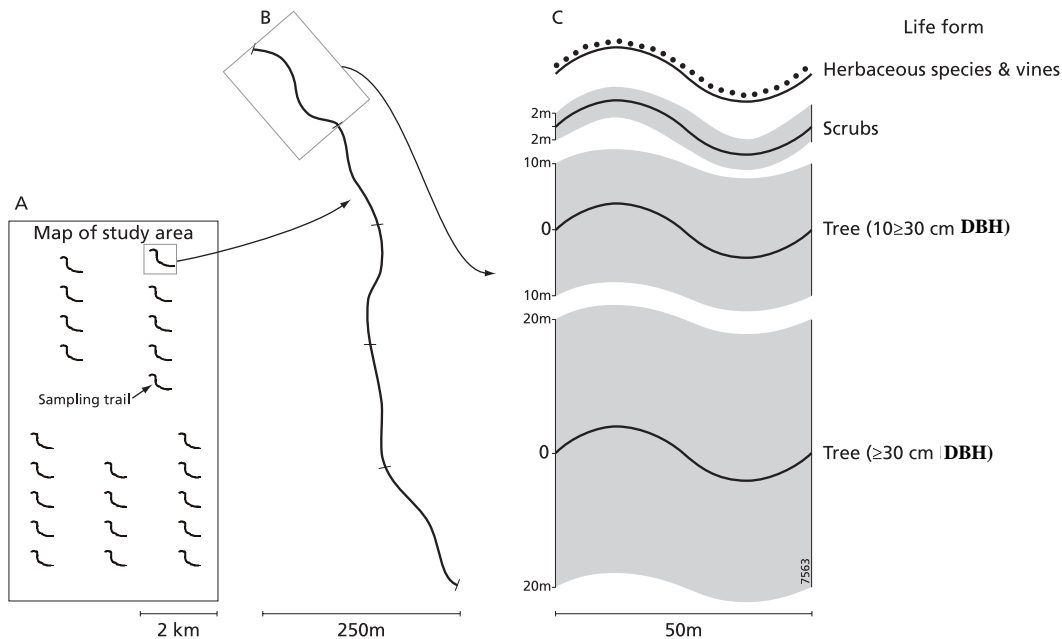
Printer-friendly Version

Interactive Discussion





**Fig. 2.** Flow diagram describing the procedural steps in the analysis of the data.



**Fig. 3.** Sampling scheme modified from the RAPELD method (see Magnusson et al., 2005). **(A)** 23 trails are regularly spaced over the site. **(B)** Each trail is placed at the same elevation level and divided in five sampling trails each of 50 m length. **(C)** Sampling along the trail. Herbaceous species: point samples at regular interval along the trail centre line. Shrubs, large-sized trees and medium-sized trees: exhaustive sampling in the indicated zone. The tree size category is based on diameter of the trunk at breast height (DBH).

**Integrating field sampling, spatial statistics and remote sensing in Brazil**

J. Arieira et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

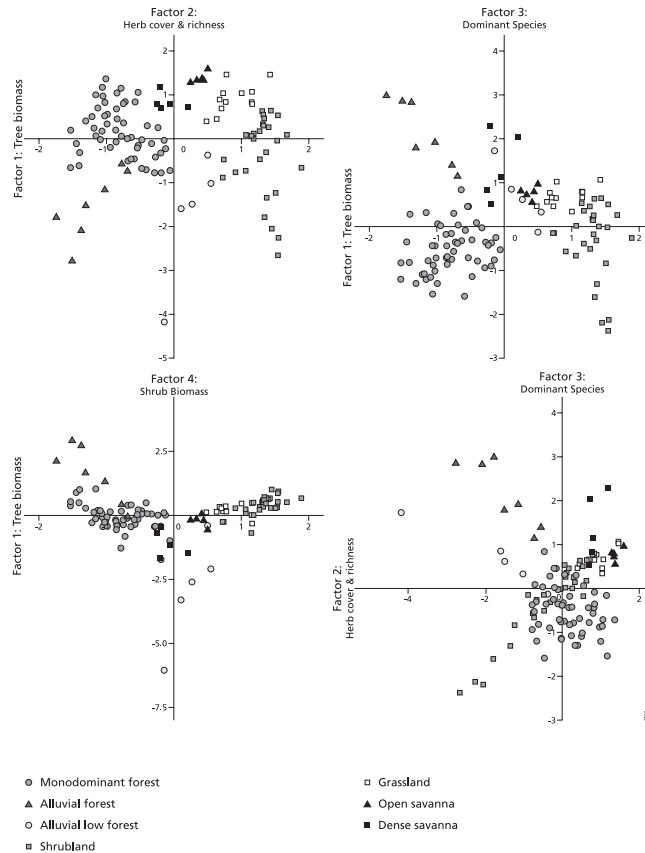
Full Screen / Esc

Printer-friendly Version

Interactive Discussion

## Integrating field sampling, spatial statistics and remote sensing in Brazil

J. Arieira et al.



**Fig. 4.** Factor analysis biplots of the axes 1 to 4 on vegetation variables obtained from 115 sampling locations. Seven clusters (symbols) represent the vegetation communities found in the studied floodplain. Factor 1 describes the gradient of tree biomass found in the study site; Factor 2 of herb cover and richness; Factor 3 of tree dominant species; and Factor 4 of shrub biomass.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

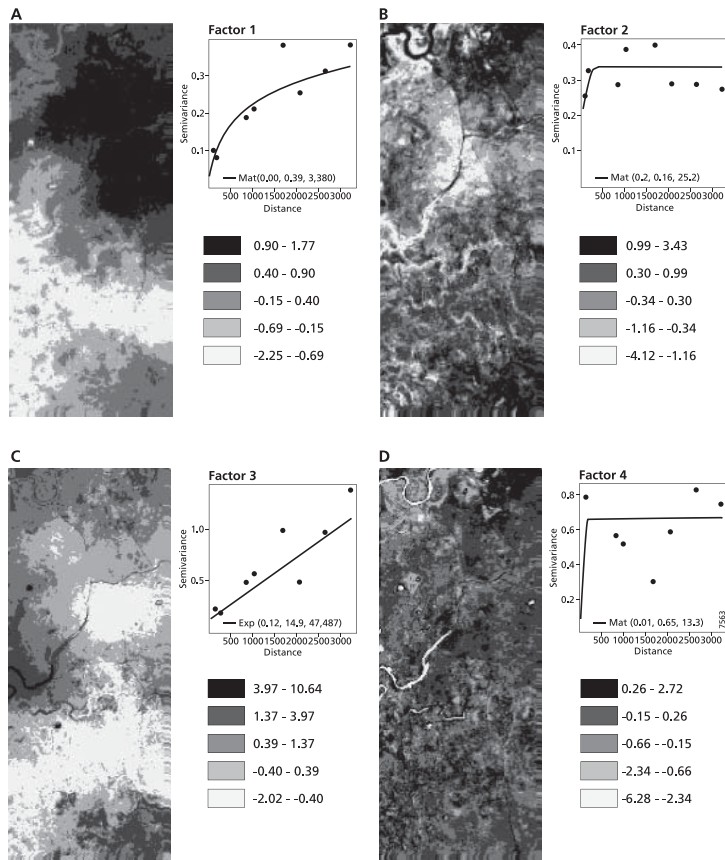
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Fig. 5.** Maps of the kriged estimates of factor scores and the semi-variograms of the residuals of the regression between factor axes and remotely sensed and ancillary data; Fitted variogram models: Mat: Matheron family, Exp: Exponential. Values between brackets are nugget effect, structured variance and variogram range, respectively. **(A)** Factor 1; **(B)** Factor 2; **(C)** Factor 3; **(D)** Factor 4.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

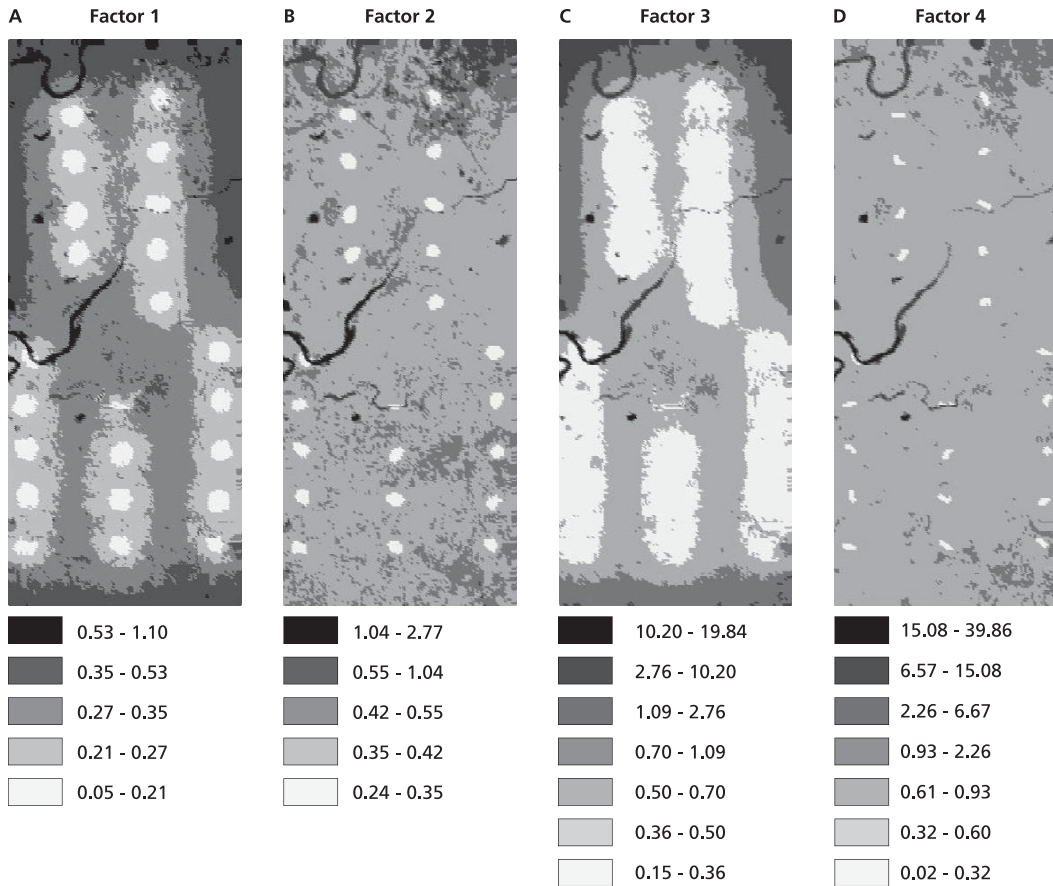
◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



**Fig. 6.** Maps of the standard deviation of the predicted error resulting from universal kriging; **(A)** Factor 1; **(B)** Factor 2; **(C)** Factor 3; **(D)** Factor 4.

**Integrating field sampling, spatial statistics and remote sensing in Brazil**

J. Arieira et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

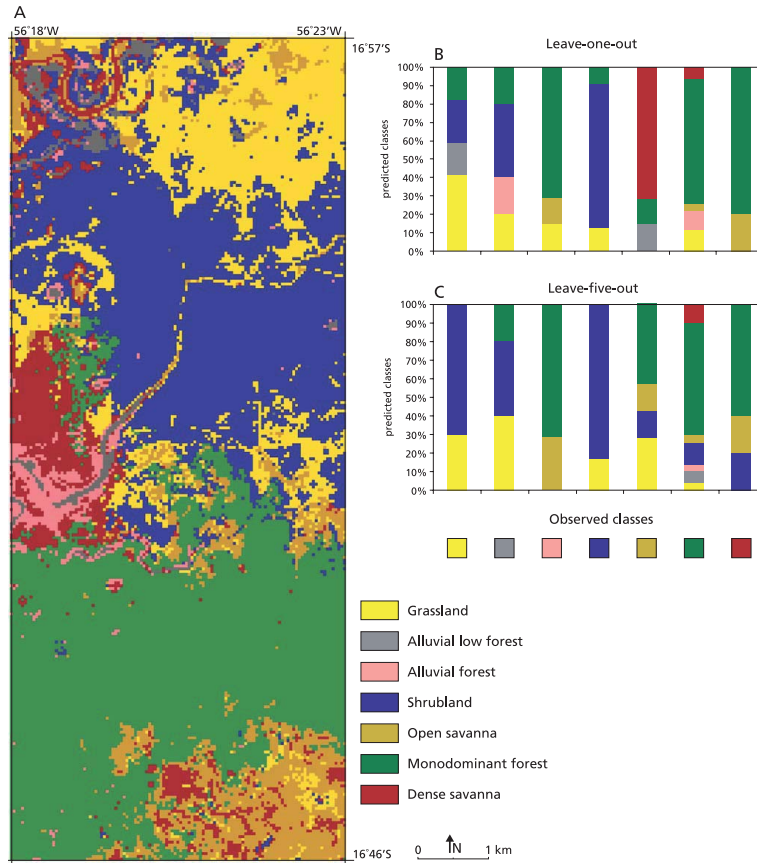
Back Close

Full Screen / Esc

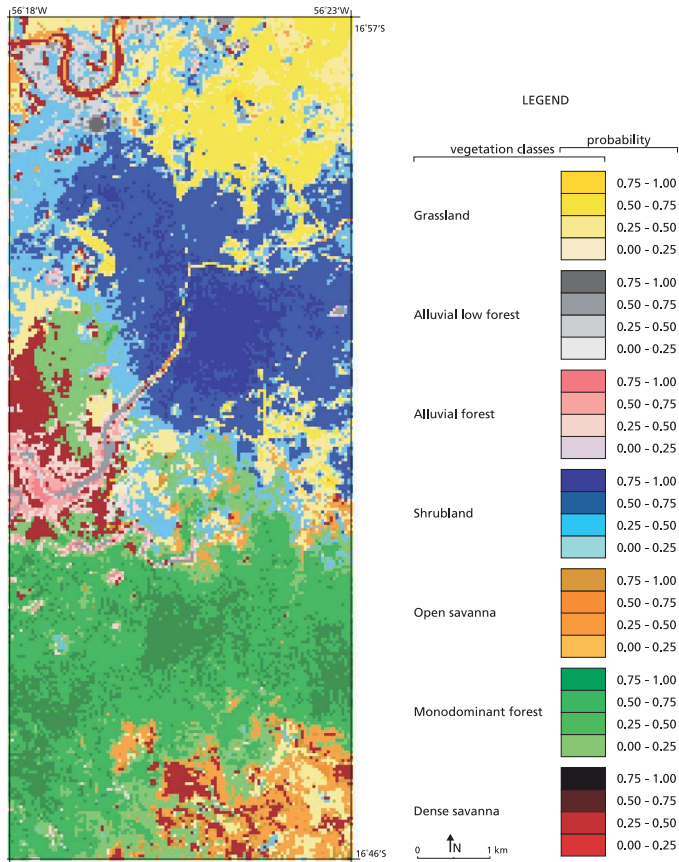
Printer-friendly Version

Interactive Discussion

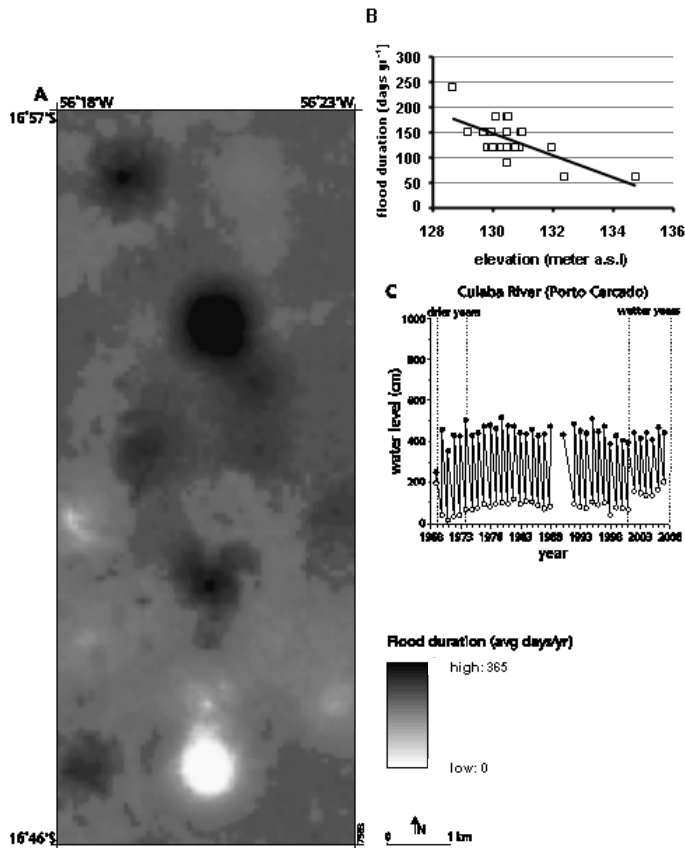




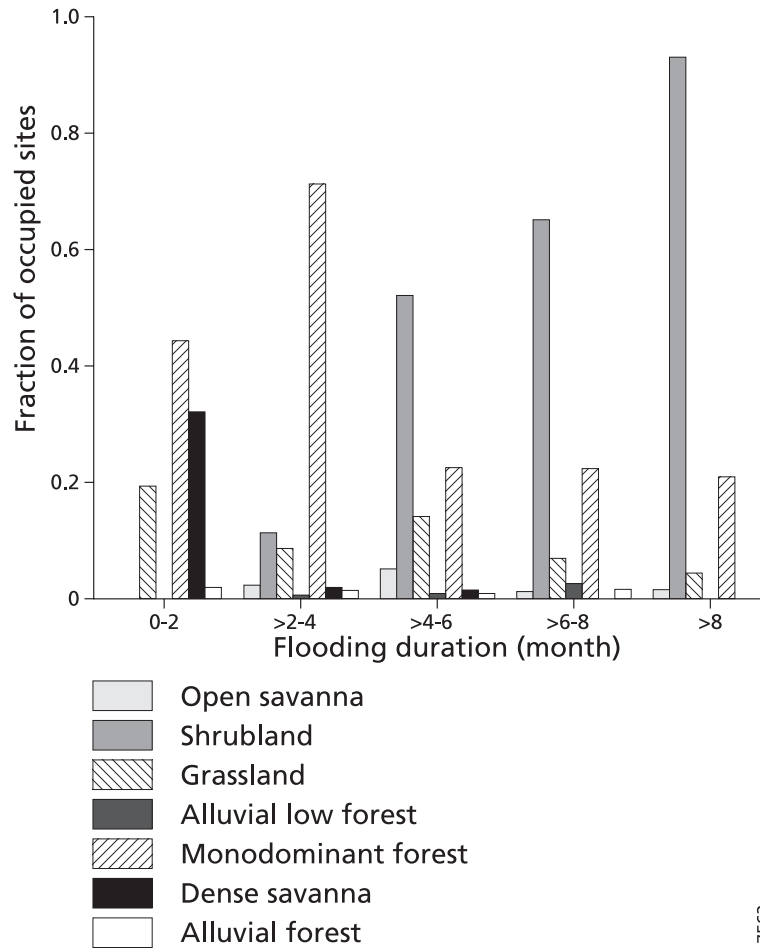
**Fig. 7.** (A) Predicted distribution of the plant communities identified at the study site. (B) Results of leave-one-sample out cross-validation. Percentage of predicted classes at sampling locations. Each bar shows the results for sampling locations with a certain observed class (indicated by the colors at the bottom of the C panel). (C) Idem, leave-five-out cross-validation.  $N = 115$ .



**Fig. 8.** Results of Monte Carlo simulation, using 1000 random simulations. **(A)** The colors on the map indicate the vegetation class with the highest probability of occurrence at a cell. A color gradient is used to show the value of this highest probability; **(B)** Maps of two single random realizations, color scale is identical to Fig. 8.



**Fig. 9. (A)** Map with average number of days flooded per year at the study site, calculated over the period 1969–2007; **(B)** Relationship between flood duration (days yr<sup>-1</sup>) and elevation of the soil surface (m.a.s.l.) observed at the 23 study trails; **(C)** water level fluctuation in the River Cuiabá between 1969 and 2007. Vertical dotted lines indicate the occurrence of drier and wetter years.



**Fig. 10.** Fraction of occupied sites by the seven identified communities along the flood duration gradient. Flooding gradient is divided in five flood classes representing number of flooded months.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

