SURVEYING THE TOPOGRAPHIC HEIGHT FROM SRTM DATA FOR CANOPY MAPPING IN THE BRAZILIAN PANTANAL

Viviana Aguilar MUÑOZ¹,²
Márcio de Morisson VALERIANO¹,³
Polyanna da Conceição BISPO¹,³

Abstract

An algorithm was developed in Geographical Information Systems (GIS) for the extraction of topographic height from the Digital Elevation Models (DEM) of the Shuttle Radar Topography Mission (SRTM), C and X bands, applied to mapping vegetation canopy in the Pantanal Floodlands. According to previous studies, these bands are sensitive to surface vegetation and thus elevation values increase in relation to terrain proportional to the height of the canopy, known as canopy effect. The proposed algorithm identifies minimum elevations within a search radius, which are likely to represent bare earth values, to generate a reference surface. The topographic height results from the subtraction between the elevations of this surface and of the original DEM. It is expected that, in this region, the height values be related to the vegetation height. Whenever possible, the algorithm results were observed together with optical images and vegetation maps of RADAMBRASIL project, for the establishment of height slicing levels as related to vegetal groups. The visual examination and statistical analysis have provided three levels of slicing height, which would be related to herbaceous, shrub and tree (forest) vegetation communities. Although slicing levels could be related to general classes of vegetation canopy in this region, field data and, or, fine resolution optical data are required for more detailed mappings. Regardless of the classification approach, height estimates from SRTM DEM represent a subsidiary data for the remote characterization of the Pantanal Floodlands vegetation, which complements traditionally used optical data.

Key-words: Topography height. Canopy effect. Floodlands. Pantanal. SRTM.

¹ National Institute of Spatial Research – INPE - Av. dos Astronautas, 1758 - Caixa Postal 515 - 12201-970 - São José dos Campos - SP, Brasil - E-mails: {viaguila; valerian; polyanna}@dsr.inpe.br
² PhD CAPES/CNPq – IEL National - Brazil
³ CNPq Researcher

Resumo

Extração da altura topográfica a partir de dados SRTM para mapeamento da vegeação no Pantanal brasileiro

Foi desenvolvido em Sistemas de Informação Geográfica (SIG) um algoritmo para a extração da altura topográfica a partir de ModelosDigitais de Elevação (MDE) derivados da missão SRTM (Shuttle Radar Topography Mission), nas bandas C e X, para ser aplicado ao mapeamento da cobertura vegetal no Pantanal. Trabalhos precursorres indicam que estes comprimentos de onda são sensíveis à superfície da vegetação e assim ocorre um acréscimo no valor da elevação em relação ao terreno proporcional à altura da cobertura vegetal, conhecido como efeito dossel. O algoritmo aqui proposto identifica as cotas mínimas dentro de um raio de busca, possivelmente associadas ao terreno nu, para serem utilizadas na geração de uma superfície de referência. A altura topográfica resulta da subtração entre o modelo de elevação e a superfície de cotas mínimas. Espera-se que, na área de estudo, seus valores estejam relacionados à altura da vegetação. Os resultados do algoritmo, quando possíveis, foram observados junto a imagens óticas e a mapas de vegetação do projeto RADAMBRASIL, com o intuito de estabelecer níveis de fatiamento da altura relacionados a grupos vegetais. A interpretação visual e análises estatísticas permitiram propor o fatiamento em três classes de altura, que estariam relacionadas a formações vegetais de porte herbáceo, arbustivo e florestal. Embora estes níveis de fatiamento indiquem categorias gerais de cobertura vegetal no Pantanal, seu mapeamento em detalhe requer dados de campo e, ou, dados óticos de resolução fina. Independente da representação em classes, a altura calculada com dados SRTM representa um dado subsidiário na caracterização remota da vegetação pantaneira, que complementa os dados óticos tradicionalmente utilizados.


INTRODUCTION

In studies of distribution and characterization of plant species through Remote Sensing (RS), data from optical spectrum band are traditionally used, complemented by field data. In research involving radar images of Synthetic Aperture Radar (SAR) the use of full-polarimetric images have been highlighted as an interesting alternative (MARTINS, 2012). Besides forestry classification, vegetation height is critical to estimate biomass and thus raising inputs for carbon cycle modeling. Research has indicated the potential of Polarimetric SAR Interferometry data (PolInSAR) for vegetation height estimation, but this data has limitations in signal to noise ratio for this application. A promising alternative for that question is the dual-wavelength SAR Interferometry (InSAR). The surface of the canopy is obtained from X-band data and the surface of bare earth from L-band data (BALZTER et al, 2007).

Some authors (i.e. KIRKBY, et al. 1990; FLORINSKY; KURIAKOVA, 1996) investigated the inverse relationship, that is, relief was studied as an indicator of the physical elements (soil, climate, hydrology, etc.), since these metrics are related to characteristics of the terrain that influence the availability of resources for plant survival (GUISAN; ZIMMERMANN, 2000). Data from interferometric radar mission SRTM (Shuttle Radar Topography Mission, USGS EROS-2009) have been in this sense in Brazil (BISPO et al, 2009; MUÑOZ, 2009; BISPO, 2012). Interferometric data are still useful in estimating the vegetation cover height and biomass by considering the canopy effect (KELLNDORFER et al., 2004; WALKER et al., 2007; TIGUE et al, 2009).

The canopy effect is an error in elevation due to the ability of partially penetrate in plant canopies of C and X radar waves (like SRTM), which depends on the structural and dielectric characteristics of vegetation components (LE TOAN et al. 1992; BECEK, 2008). Once radar beam penetration is partial, the return signal originates at higher elevations (the terrain altitude) (VALERIANO et al, 2006), which constitutes the canopy effect. This approach
of exploring canopy effect for vegetation height estimation (KELNDORFER et al., 2004) requires knowledge of the land elevation, to be used as a reference (height zero). This condition is not possible in Brazil because of the very limited availability of good quality and detailed topographic maps and of InSAR-L data.

An attempt to extract the canopy effect from SRTM data in the Brazilian Pantanal was firstly conducted by Valeriano and Abdon (2007), whose methodology relied on land surface perception given by SRTM data itself. Their approach consisted in a point selection of places that meet metric criteria typical of low lands (low slope, convergent plan curvature, concave profile curvature and local thalwegs), followed by interpolation into the reference (terrain) DEM. Finally, the original DEM is subtracted by the estimated reference DEM. The results showed negative values, suggesting some error in the algorithm selection of minimum quotas. As a consequence, the result analyses were limited by post-processing attempts to correct this error. A reported alternative to this approach consisted in fitting a statistical trend surface to topography. This approach was used for the characterization of the Taquari megafan active and inactive channels (ZANI et al., 2012), but it results in positive and negative quotas, thus heights could only be observed on a relative basis.

In soil distribution studies, a specific methodology was developed for the digital estimation of topographic height (MUÑOZ, 2009) based on other fundamentals. The proposed GIS processing consisted basically in the identification of regional minima through the gathering of SRTM (C-band) data values in a given range at the same point location. This was accomplished by means of directional filters and logical operations, in combined displacements to gather regional quotas in different GIS layers. This set allows analyses of quota distribution around each point, such as the extraction of minima quotas of the land for the construction of the reference DEM to subtract from original SRTM data.

Due to the large variation of elevation over short distances in rugged relief conditions (mountainous regions) the topographic height value is not expected to represent the vegetation height under this condition. On the other hand, in flat regions of homogeneous elevation, such as Pantanal, the quota variations in short distances are due to the result of surface objects height variations, mostly vegetation. The relief of the Pantanal is characterized by extreme flatness, a characteristic expected to be a favorable condition for the extraction of height features, which include the height of the vegetation. The main purpose of this paper is to apply this methodology and assess its usefulness in mapping vegetation canopy height in the Brazilian Pantanal.

**MATERIALS AND METHOD**

Pantanal is located in the center of South America, Upper Paraguay River Basin, covering territories of Brazil, Paraguay and Bolivia. The Brazilian Pantanal occupies large areas of Mato Grosso and Mato Grosso do Sul States. The experimental area of this work (Figure 1) is bounded by 17°S and 17°30’S latitude and 57°W and 55°30’W longitude, north region of the Taquari river megafan, which gathers water from Cuiabá, San Lorenzo and Piquiri rivers. In that place, elevation ranges approximately between 90m and 198m and slopes range between 0.05% and 0.08%.

Pantanal is a region of seasonal flooding (seasonal wetland) composed of a complex mosaic of habitat (SARMIENTO; MONASTERIO, 1975; ZEILHOFER; SCHESSL, 1999). The seasonally flooded vegetation consists of grasses and different types of forests. Some of the periodically flooded forests comprise mono specific communities, or several co-dominant species (POTT; POTT, 1994). Riparian forests along the rivers are described in the literature as flood forests of immense species richness (WITTMANN et al., 2008).
Surveying the topographic height from SRTM DATA for canopy mapping in the Brazilian Pantanal

The phytophysiognomies of the study site include: Savanna (Cerrado and Chaco), with vegetation varying from dense to woody-grassy formations; Deciduous Seasonal Forest (Floresta Estacional Decidual); and Savanna-Seasonal Forest (Savanna – Floresta Estacional) hotspot contact, with semideciduous forest communities with emergent canopy (RADAMBRASIL, 1982), as shown in figure 2. Dense Arboreal Savanna - Sd (A), Open Arboreal Savanna without Gallery Forest - Sas (B and C); Park Savanna without Gallery Forest - Sps (B and C), Grassy-woody Savanna without Gallery Forest - Sgs (A and B), contact between Savanna - Alluvial Semideciduous Seasonal Forest with emergent canopy - Fae1 (a, B and C), and contact between Savanna - Lowland Semideciduous Seasonal Forest with emergent canopy - Fbe1 (B) are the phytophysiognomies among the highlights of figure 1.

The Savanna structure can be described as small trees with cork bark and large leaves, spaced over a field of graminoid and large-leaved plants, typically developed on plain terrains. The Deciduous Seasonal Forest is very heterogeneous, with a large number of deciduous arboreal species. These communities present general height around 20m, with many emergent species, which present high foliar seasonality in unfavorable periods. The Savanna-Seasonal Forest contact area is the largest class, notably important near drainage channels; the graminoid types prevail on floodplain areas, whereas forest fragments occur in relatively higher lands disposed on elongated patches, locally known as cordilheiras (ridges).

To calculate the topographic height, elevation models from both the topographic database Topodata and the German Aerospace Center DLR X-SAR SRTM Digital Elevation Model (DEM X-SAR) were taken as input. Topodata DEM (www.dsr.inpe.br/topodata) corresponds to a refinement of the SRTM data (C-band) by kriging (geostatistical interpolation technique), from ~90m to ~30m (VALERIANO; ROSSETTI, 2012).
The X-SAR SRTM data can be obtained free of charge on the website of Earth Observation Center (DLR EOWEB-NG, http://eoweb.dlr.de). These data have been available since December 2010, with 30m of resolution. Just as C band data, X band data covers the globe between latitude range of 60° north and 60° south. As the X-SAR instrument did not use the system Scan-SAR, the data had a slightly higher resolution and better signal to noise ratio (SNR) than that of the C-SAR instrument. However, it has a narrower spatial coverage (FARR et al., 2007). From the X-SAR data tracks available for experimental area (Figure 3), a new DEM (15m) was refined by kriging with the same geostatistical coefficients used in the refinement of Topodata.

Figure 2 – Vegetation map of the study site (RADAMBRASIL, 1982). Highlighted phytosociognomc regions: Savanna (Sd: Dense Arboreal, Sas: Dense Arboreal without Gallery Forest, Sps: Park without Gallery Forest, and Sgs: grass-woody without Gallery Forest); Lowland Semidecidual Seasonal Forest with emergent canopy (Fbe), Ap.: pasture and farming area, and Ecological Tension Area (hotspot) (Sd1: contact Savanna - Seasonal Forest and Dense Arboreal Savanna; Fae1: contact of Savanna - Seasonal Forest and Alluvial Semidecidual Forest with emergent canopy, and Fbe1: contact of Savanna-Seasonal Forest and Lowland Semidecidual Forest with emergent canopy)

Figure 3 – SRTM X-SAR data available for the experimental area. Box A and B correspond to detailed areas that will be shown throughout this paper for viewing.

The topographic height was obtained using the methodology developed in Muñoz (2009), under the three elevation models (Topodata-30m SRTMX-30m and SRTMX-15m). The first step is to identify the DEM minimum within areas defined within a 27x27 pixel
moving window by comparing the heights among the encompassed pixels of the model. From the selected values, a surface of minimum elevations is computed. This surface is expected to represent the minimum elevations of the terrain, or bare ground (bald-Earth). Accordingly, the last step consists in calculating the arithmetic difference between the original DEMs (affected by canopy effect) and the respective bare-ground surface (Figure 4).

The topographic heights from the three models were compared as a first result evaluation. Next, height levels were established for slicing into meaningful plant canopy height classes in the experimental area with guidance of visual assessments of results together with optical images from Google Earth. Further, Topodata DEM results were selected to statistical analysis (Figure 2), since this dataset covered the entire experimental area at the closest scale to RADAMBRASIL vegetation map. This vegetation map was scanned and superimposed on the model for sampling height in the different phytophysiognomic subsets. The whole dataset consisted of 140 height records for each of the nine vegetal phytophysiognomies. The height distribution in each vegetation class was observed and the hypothesis of equality between classes was verified by applying both Kruskal-Wallis (homoscedasticity) and Wilcox (paired comparison) nonparametric tests.

For surface viewing and kriging the Surfer Software was used (GOLDEN SOFTWARE, 1995). The topographic height was computed by an IML - Idrisi Macro Language algorithm (EASTMAN, 1995). Global Mapper (GLOBAL MAPPER SOFTWARE Llc, 2007) was used for geoprocessing tests and to make graphs. Google Earth images were used for visual interpretation in comparisons with topographic height results. Finally, R Statistic (The R Foundation for Statistical Computing, 2012) was used for statistical analysis.
RESULTS AND DISCUSSION

A topographic height detailed excerpt, corresponding to A-box highlights in Figure 3, can be observed in Figure 5. The map obtained from Topodata DEM presents the highest level of generalization of the enhanced features. Although at the same resolution, the map derived from SRTMX-30m showed more detailed features than the former, but also presented a greater number of artifacts, probably due to the fact that the model was not pre-processed for the topographic height derivation, but certainly originated from the prevalence of top canopy-surface backscattering in the X-band. The most detailed features were registered on the map from SRTMX-15m.

Figure 5 – Topographic height detail in the experimental area (A-box highlights in Figure 3) obtained from SRTM C and X data. C-SAR data correspond to Topodata DEM and X-SAR 15m data correspond to resampling of X-SAR 30m by geostatistics
The frequency distribution shape is equivalent for the three topographical height models (Figure 6). Values between 2m and 6m predominate; values between 6 and 10m are less frequent but may represent taller and more scattered vegetation; values greater than 10 are rare and related to an artifact in the three images, particularly expressive on SRTM-X-30m derived topographical height model.

![Figure 6 - Topographic height frequency distribution in Figure 5 area](image)

This artifact is a type of square feature, originated from landmarks in an isolated condition and below the surrounding elevation, particularly noticeable in the X-SAR 30m data (Figure 7). Solutions to remove it may focus on the moving window used to extract the topographic height, with changes in the area of analysis, so that bald-Earth surface results are free from those negative singularities. Alternatives may consider DEM itself, perhaps a pre-processing technique that eliminates these singularities or, less suitable, on post-processing operations to operate directly on the results of topographic height.

![Figure 7 - Topographic height detail in experimental area (B-box highlights in Figure 3) where artifacts of square shape were observed](image)

The topographic height distribution, restricted to positive values, indicates an important gain in relation to the geomorphometric classification approach experienced in a prior study (Valeriano; Abdon, 2007). In that experiment, the minimum quota selection criteria were flawed because the resulting terrain surface has elevations above those of DEM in 25% of the tested area. This caused computing negative heights up to -13.7m and, when applying an offset (sum) of 2m, the negative value area is reduced to less than 5%. However, the remainder topographic height had its potential extent reduced, with evident commitment of relative differences. In this paper, selecting minimum quotas ensured that the ground surface was always estimated below the DEM surface.

Another possible difference on behalf of this result quality refers to the DEM used in both studies. The DEM used in the first study was interpolated with geostatistical coefficients raised in Pantanal, whilst the current study uses data from the X-band SRTM mission, with greater surface detail, and the Topodata DEM whose interpolation coefficients were selected from a broader base. In a retrospective study, Valeriano and Rossetti (2012) compared...
manifold interpolations of regional projects and a greater power of detail interpolator was used to make the Topodata DEM, whilst the old Pantanal DEM stood among those that caused surface smoothing. In addition to the mentioned +2m offset needed in the previous studies, the higher sharpness of Topodata DEM led to the expectation that the topographic heights in this work would be larger and more approximate to the variations that in fact occurred.

The results of topographic height, when observed in general scales, highlight the DEM singularities in the experimental area, and promote the enhancement of high features, which contrast with large homogeneous areas of lower elevation prevailing in the surroundings. The former corresponds to contact of Savana - Seasonal Forest and Alluvial Semideciduial Forest with emergent canopy (Fae1) and contact of Savanna-Seasonal Forest and Lowland Semidecidual Seasonal Forest with emergent canopy (Fbe), which occurs in association with drainage channels and lagoons (Figure 8). The latter corresponds to Savanna areas, where Savanna grass-woody without Gallery Forest (Sgs in Figure 2) prevails in the west portion of the presented area.

In studies of plant species distribution through radar and optical Remote Sensing (RS), the two data sets bring different descriptive aspects, as the vegetation types and their biophysical conditions, according to the interaction mechanisms affecting the acquired signals from the canopies. The observation of topographic height with optical images (Figure 9) showed that the main features are due to the vegetation height along rivers. However, incremental differences in tree density resulting in intermediate heights to those of herbaceous cover were observed, as indicated by the remarkable height transitions in detail B (I) and to a lesser extent at C (I). The abrupt contact of forest and non-forest areas (A detail) results in higher topographic height values and in a higher contrast among the features. The different transitions reflect the canopy effect, integrated in the SRTM-90m resolution, as the tree density is expected to affect the average altimetry (KELLNDORFER et al. 2004).

Exploratory analysis indicated that in the forest classes there is a greater heterogeneity of height canopy than that of savanna classes. In the Forest, the height is distributed between 0 and 15m, with an average of 5m and a lower frequency of heights above 10m. In the Savanna, height ranges between 0 and 10m, with an average of 3m. Although the average savanna canopy heights is more restricted to values below 4m, occasionally, species over 6m may occur (Figures 10 and 11). Overlapping of the topographic height distributions among the different vegetation classes hampers the determination of thresholds for the separation between groups (forest savanna and forest) and between forest types within groups (forest type and savanna type).

An attempt to identify slicing thresholds for height was made by pairwise comparison of topographic height samples with vegetation classes. Firstly the Kruskal-Wallis test was applied, whereby the homogeneity hypothesis was rejected (H$_0$: samples are taken from the same population) at 5% significance. This result indicates that at least two classes are different. By Wilcox-test statistically equal or different classes were identified. The equality hypothesis (H$_0$: class X equals Y class) was rejected in 30 and accepted in six of 36 possible combinations ($C_{\text{combination}} = (9x8) / 2$). While this result may suggest the separation probability of 30 pairs, in practice this is not so simple, as between each pair of classes accepted as different there is 5% probability of confusion. This deprives the establishment of such thresholds without a post-processing that also includes analyzes of data collected in the field or population statistics.

The vegetation classes indicated as equals by the Wilcox-test according to topographic height were Fae1-Fbe1, Fae1-Sd1, Fbe1-Sd1, Sas-Sgs, Sas-Sps and Ap-Sgs. This suggests that the topographic height allow grouping the experimental area vegetation into at least two classes: savanna and forest. This result is in agreement with the prior exploratory analyses (Figure 11). It is worthy of note that the forest-savanna (Fbe1-Sd1) contact areas were not distinguishable by the test.
Figure 8 – Topographic height in the experimental area (A) against forest structures mapping: (B) contact of Savana - Seasonal Forest and Alluvial Semideciduous Forest with emergent canopy (Fae1), (C) contact of Savanna-Seasonal Forest and Lowland Semideciduous Forest with emergent canopy (Fbe1), e (D) Lowland Semideciduous Seasonal Forest with emergent canopy (Fbe)
Figure 9 – Comparison of optical images and topographic height results. A, B and C are in the box highlighted areas in Figure 1. Rows: I – topographic height obtained from Topodata DEM; III – optical image from GoogleEarth; and II – detail from the same image.
Surveying the topographic height from SRTM DATA for canopy mapping in the brazilian Pantanal

Figure 10 – Topographic height distribution within vegetation classes, extracted by overlapping of height model from Topodata DEM and RADANBRASIL vegetation mapping of experimental area

Figure 11 – Topographic height Boxplot within vegetation classes, extracted by overlapping of height model from Topodata DEM and RADANBRASIL vegetation mapping of the experimental area
The selection of threshold levels should take into account the fact that the topographic height values derived from SRTM DEM represent a variable fraction of the canopy height, according to the penetration in the vegetation mass of radar waves C and X bands, which depends on many factors (LE TOAN et al. 1992). In the absence of accurate and appropriate references to specific vegetation of the experimental area, we can consider the general relation obtained by Kellndorfer et al. (2004) which indicates an adjustment factor (ratio) of about 0.8 (80%) between height canopy values collected in the field and those topographic height values derived from SRTM DEM. It happens, however, that this ratio is highly variable depending on the factors mentioned above. Therefore, more suitable than using absolute values of topographic height to directly estimate canopy height, mapping the differences in geographic scales proved to be feasible for the separation of phyto physiognomies of structural contrast, as a pre-sampling technique for more detailed surveys, and as a complementary data input for image classification of vegetation in flat areas like Pantanal.

Based on the statistical analysis results and the aforementioned limitations, and considering the entire experimental area rather than the single classes, the topographical height values were classified into three general categories which should indicate herbaceous (0 to 2m), shrub (2 to 6m), and forest (larger 6m) vegetation; that is consistent with the structural ranges suggested by the IBGE (1992) and with the interpretation of optical images obtained from Google Earth (Figure 12).

In a visual comparison (Figure 13), the information gains from elevation map (A) over topographic height maps (numeric and classified, A and B) in relation to forest structures (RADAMBRASIL, 1982) are evident. Moreover, height maps present a clear improvement of the planimetric detail in relation to RADAMBRASIL reference map itself, as an indication of their potential for the refinement of existing vegetation maps. The results can be easily replicated both for studies in the Pantanal and for other similar areas in relief and vegetation structure. In another contribution, the presented approach can be adopted for the development of biomass estimates in similar environments, by providing vegetal structure details not provided by optical remote sensing techniques. The biomass modeling is critical to quantify forest production. In carbon flux modeling research (BROWN et al., 1993) forest biomass is one of the greatest uncertainty parameters to compute the global balance because of the difficulty in obtaining its structural data. The proposed methodology is a viable approach when compared with other more accurate yet more expensive techniques, such as PolInSAR and InSAR.
Figure 13 – Maps of elevation (A), topographic height in the numerical way (B), topographic height in classes (C) and vegetation cover (D) of the experimental area.
CONCLUSION

Calculation of the vegetation height in the Brazilian Pantanal from DEM was developed by adapting a computational algorithm based on a regionalized analysis of elevation data dispersion. Once developed for the estimation of topographic feature height, the algorithm was applied to different DEM obtained by remote sensing (radar interferometry). Due to its consistency with the concept of topographic height, this calculation was superior to previous attempts, and the efficiency in calculating the surface minimum was evident in the numerical distribution of the results. Besides overcoming this major task, the visual interpretation of optical data and statistical analysis allowed evaluating the potential and limitations of the method compared to observable differences in vegetation cover. The selection of appropriate levels of slicing allowed the rapid mapping of herbaceous, shrubs, and forest formations. This methodology proved to be of great potential for forestry study applications, both for the structural characterization of the forest and for research related to forest biomass estimations. The improvements of the DEM-X SAR, refined by the pre-processing geostatistic approach, could not be evaluated due to the lack of vegetation data in finer and compatible resolutions. Research to advance on this particular data can be exploited only experimentally, due to its limited spatial coverage. Projects aiming to include L-band InSAR into this approach are also suggested.

REFERENCES


