Spatial Variability Analysis of CBERS CCD Images in Forest Regions

Carlos Alberto Felgueiras [1]
Eduardo Celso Gerbi Camargo [1]
Camilo Daleles Rennó [1]
Jussara de Oliveira Ortiz [1]

[1] Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12245-970 - São José dos Campos - SP, Brasil
{carlos, eduardo, camilo, jussara}@dpi.inpe.br

Abstract. This paper analyses the spatial variability of remote sensing images obtained by the Charge Coupled Device (CCD) sensor presented in the China Brazil Earth Remote Satellite (CBERS) at spatial regions of forests. Semivariograms are mostly used to model the spatial variability of environmental data as elevation, temperature, health risks, geology, hydrology and mining information, etc... In this work the semivariograms were used to model the variability of the spectral information represented in remote sensing images. A limited number of sample points of images were considered, instead of full images, due to the size constraints of the input data to calculate the semivariances. A very large amount of input points requires a very high processing time to evaluate the semivariograms. Thus, this work aims to explore the possibility of representing the spatial variability of an entire image using a limited amount of samples draw from it. Semivariogram analysis were done with random sample sets of different sizes. Two spatial regions with different patterns were considered, one with typical forested area and other with partially deforested area. It was found that is feasible to get representative semivariograms of the whole image from small sample sets. In addition it was obtained different semivariograms for the two regions showing that stationary hypothesis cannot be assumed for forest images with different patterns.

Keywords: **R**emote Sensing, Image Processing, Spatial Analysis, Semivariogram, Geostatistics.

1. Introduction

The semivariogram is a powerful statistical tool to represent spatial variations of phenomena in geographic regions. Semivariogram measures the average dissimilarity between data separated by a distance vector **h** (Goovaerts, 1997). The semivariogram is used for geostatistical procedures for spatial predictions and simulations (Isaaks and Srivastava, 1989; Deutsch and Journel, 1998; Goovaerts, 1997). Typically the semivariogram is obtained from a set of sample points representing environmental phenomena, as temperature, elevation, soil texture and others (Felgueiras, 1999; Burgess and Webster, 1980). The semivariogram has been explored in applications of remote sensing images in classifications (Florenzano, 2002; Chica-Olme and Alberca-Hernández, 2004; Almeida-Filho et all, 2005), and predictions for missing data regions (Zhang, Li and Travis, 2009). Zhang et all, 2009 perform analyses of variograms with various sample sizes from a multispectral image.

The main drawback to assess semivariograms of remote sensing images is their size that uses to be very large, mainly for high resolution images. Large amounts of input points are computationally costly requiring very high processing time to plot the semivariograms.

The objective of this work is to analyze the variability of remote sensing images using a random sample set of points obtained directly from their pixel information. This work intends to find out whether is possible or not to represent the spatial variability of the entire image using a limited set of samples of the image of interest. Two different spatial regions are considered, one with typical forest area and other with partially deforested area. The semivariograms are computed and compared in terms of their whole structure (nugget effect, sill, range and mathematical model parameters) in order to understand their behavior in relation to different spatial regions and different sizes of sample sets.

The structure of this article begins with the introduction in Section 1. Section 2 addresses some basic important concepts and the methodology used in this work. Section 3 presents the results and several analyzes and discussions related to them. Conclusion remarks and some suggestions to continue this work is considered in the Section 4.

2. Concepts and Methodology

The semivariogram

The semivariance is defined as half the average of squared difference in z values between pair of sample points. *The semivariogram* is a plot of semivariances as a function of vector distances **h**. Semivariograms model the spatial variability of phenomena, represented by a *z* variable, in function of distances h. An *experimental,* also known as *empirical, semivariogram* can be estimated from a sample set by the following relation (Equation 1):

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2\mathrm{N}(\mathbf{h})} \sum_{\mathrm{i}=1}^{\mathrm{N}(\mathbf{h})} [z(\mathbf{u}_{\mathrm{i}}) - z(\mathbf{u}_{\mathrm{i}} + \mathbf{h})]^2 \quad (1)$$

where N(**h**) is the number of pairs stating approximately to a |**h**| distance between spatial positions $\mathbf{u}_i$ and positions $\mathbf{u}_i + \mathbf{h}$. Directional semivariogram consider also the direction of the vector **h** defined by each sample pair. Omnidirectional, or Isotropic, semivariogram does not take into account the directions of vector **h** and are used to represent isotropic phenomena.

Usually the experimental semivariograms are fitted by mathematical functions, know as theoretical models. The theoretical semivariogram is used to assess predictions and simulations values in geostatistical procedures. Figure 1 shows a typical semivariogram with its characteristic parameters, the nugget effect C_0, the sill $C(\mathbf{0})$ and the range *a*.

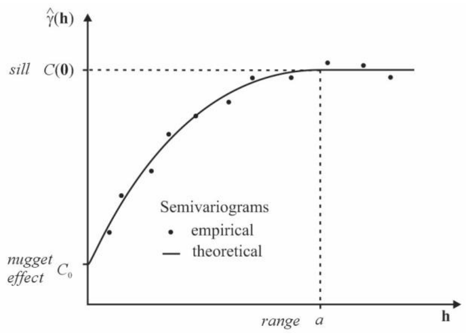


Figure 1. Parameters of experimental and theoretical semivariograms.

Frequently the mathematical model fitted to theoretical semivariogram is Spherical, Exponential, Power or Gaussian. For example, the Exponential semivariogram is defined by (Equation 2):

$$\gamma(\mathbf{h}) = \begin{cases} 0, & |\mathbf{h}| = 0 \\ C_0 + C_1\left[1 - \exp\left(-3\frac{|\mathbf{h}|}{a}\right)\right], & |\mathbf{h}| \neq 0 \end{cases} \quad (2)$$

where C_1, the model contribution, is the difference between $C(\mathbf{0})$ and C_0.

Methodology

Given a band of a CCD CBERS image of a region of interest, the methodology used in this work is based in the following steps:

- Draw randomly sample sets from the image for different number of samples.
- Obtain experimental semivariograms representing the variability of the sample set.
- Fit mathematical models to the experimental semivariograms.
- Compare and analyze the semivariograms for different sample size and different spatial regions.

3. Results and Discussions

3.1 The CCD CBERS images

In this work it was used two small regions of band 4, of a CCD CBERS image, shown in the Figures 2 and 3. These images are patches of 171/114 path/row scene of CCD CBERS sensor acquired in September, 01, 2008. This scene was imported to the SPRING software (Camara et all, 1996) where the patches were cut and save as new infolayers. The SPRING was also used to create the contrasted color composition of Figure 2 and to assess some deterministic and statistics properties of the images.

Figure 2 shows the region 1 representing a typical natural forest area while the Figure 3 depicts the region 2 representing a partially deforested area. The bounding box coordinates, in decimal degrees, of region 1 are: longitude -60.2146, latitude -12.3786 and longitude -60.1218, latitude -12.2873. The bounding box coordinates, in decimal degrees, of region 2 are: longitude -60.4195, latitude -13.0131 and longitude -60.3262, latitude -12.9218. Each image has 20 m x 20m, x and y resolutions respectively, and size of 500 rows x 500 columns. The statistical parameters, mean and the variance, of the pixel values for region 1 are 132.61 and 33.6793 while these values for region 2 are 115.235 and 115.188 respectively.

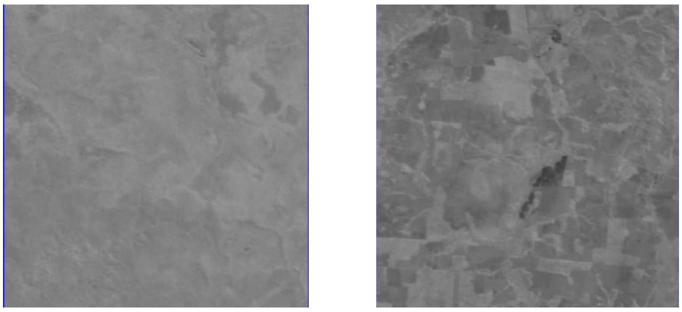


Figure 2. CBERS CCD Band 4 images of region1 (left) and region2 (right).

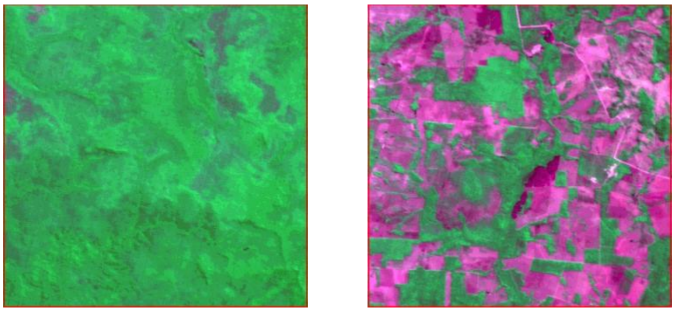


Figure 3. Color composition of CBERS CCD Bands 3, 4 and 2 for Red, Green and Blue color levels respectively of region1 (left) and region2 (right)

In Figure 3 greener regions represent forested spatial areas while pinker regions represent deforested spatial.

3.2 Spatial Variability Analysis

For each region presented in Figures 2 it was performed the spatial variability analysis with experimental semivariograms for random sample sets of 6 different sizes, 1000, 2500, 5000, 10000, 25000 and 50000 samples . All experimental semivariograms were fitted by exponential mathematical semivariograms. The semivariograms presented in this section were obtained using the GS+ software (http://www.gammadesign.com/). The Residual Sums of Squares (RSS) equation provides an exact measure of how well the theoretical model fits the experimental semivariogram information; the lower the reduced sums of squares, the better the model fits. GS+ uses the RSS criterion to assess the best parameters for each of the semivariogram models. This is done by determining the combination of parameter values that minimizes the RSS value for any given model.

3.2.1 Semivariograms for region 1

Figure 4 shows the surface semivariogram where one can conclude that the variability of the band4 reflectance in the region 2 can be considered isotropic.

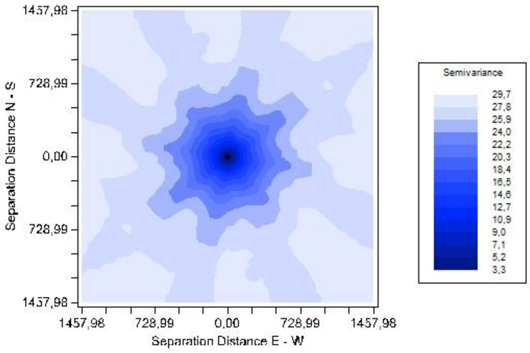
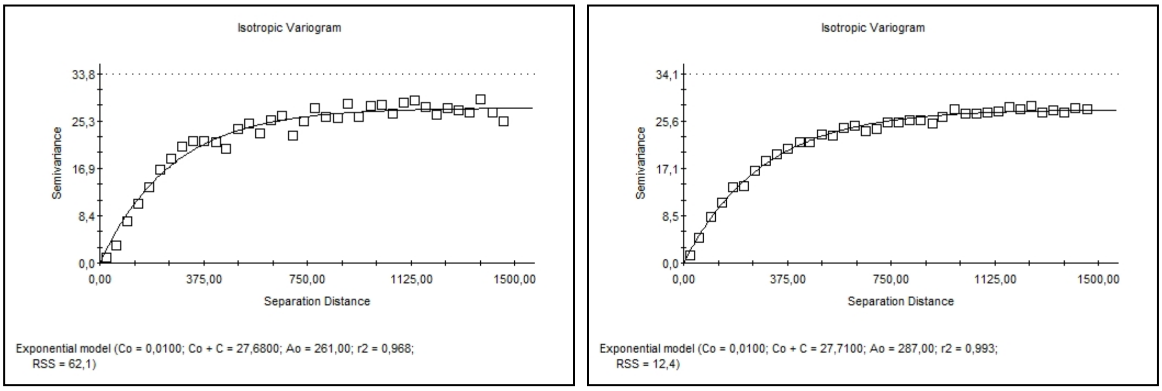


Figure 4. Surface semivariogram (region 1).

Figure 5 shows semivariograms fitted to exponential model for 1000, 2500, 5000 and 10000 samples set of random points taken from the region 1. For separation distances it was used lags of 40 m with a maximum distance equal to 1500 m.



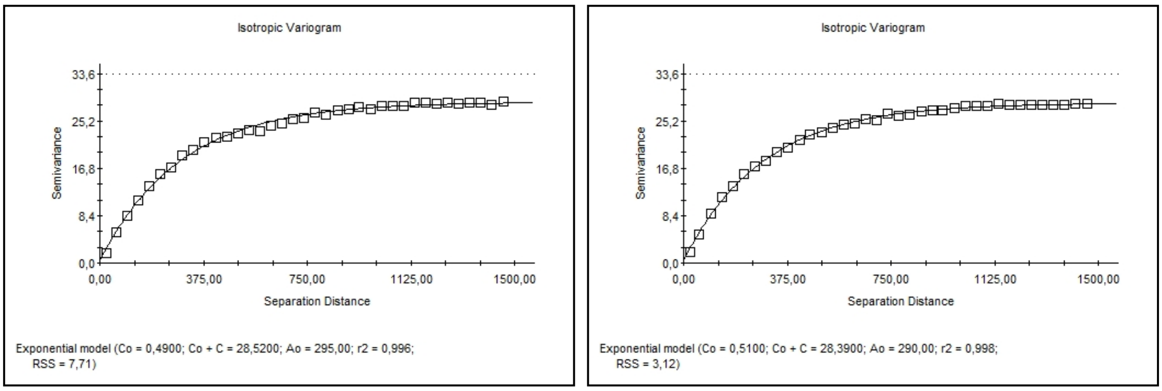


Figure 5. Isotropic semivariograms (region 1) from 1000 samples (upper left), 2500 samples (upper right), 5000 samples (lower left), 10000 samples (lower right)

It was also estimated semivariograms for 25000 and 50000 samples and their plot seemed very similar to that presented for 10000 samples in Figure 5. Table1 present the parameters of isotropic semivariograms for 6 different sizes of sample set of the region 1.

Table 1. Parameters of isotropic semivariograms for different sample sets of the region 1

# Samples	Nugget Effect	Contribution	Range (m)	RSS
1000	0.01	27.67	783	62.10
2500	0.01	27.70	861	12.40
5000	0.49	28.03	885	7.71
10000	0.51	27.88	870	3.12
25000	0.43	27.69	852	2.25
50000	0.29	28.61	864	2.00

3.2.2 Semivariograms for region 2

Figure 6 shows the surface semivariogram where one can conclude that the variability of the band4 reflectance in the region 2 can be considered isotropic.

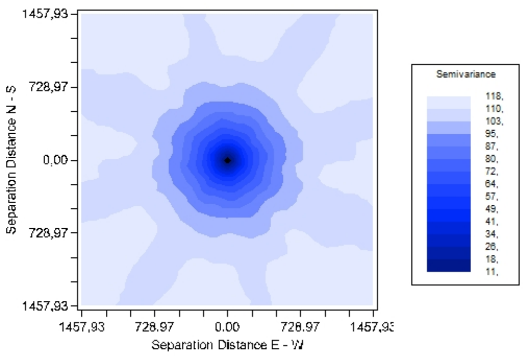


Figure 6. Surface semivariogram (region 1).

Figure 7 depicts semivariograms fitted for 1000, 2500, 5000 and 10000 samples set of random points taken from the region 2. For separation distances it was used lags of 40 m with a maximum distance equal to 1500 m.

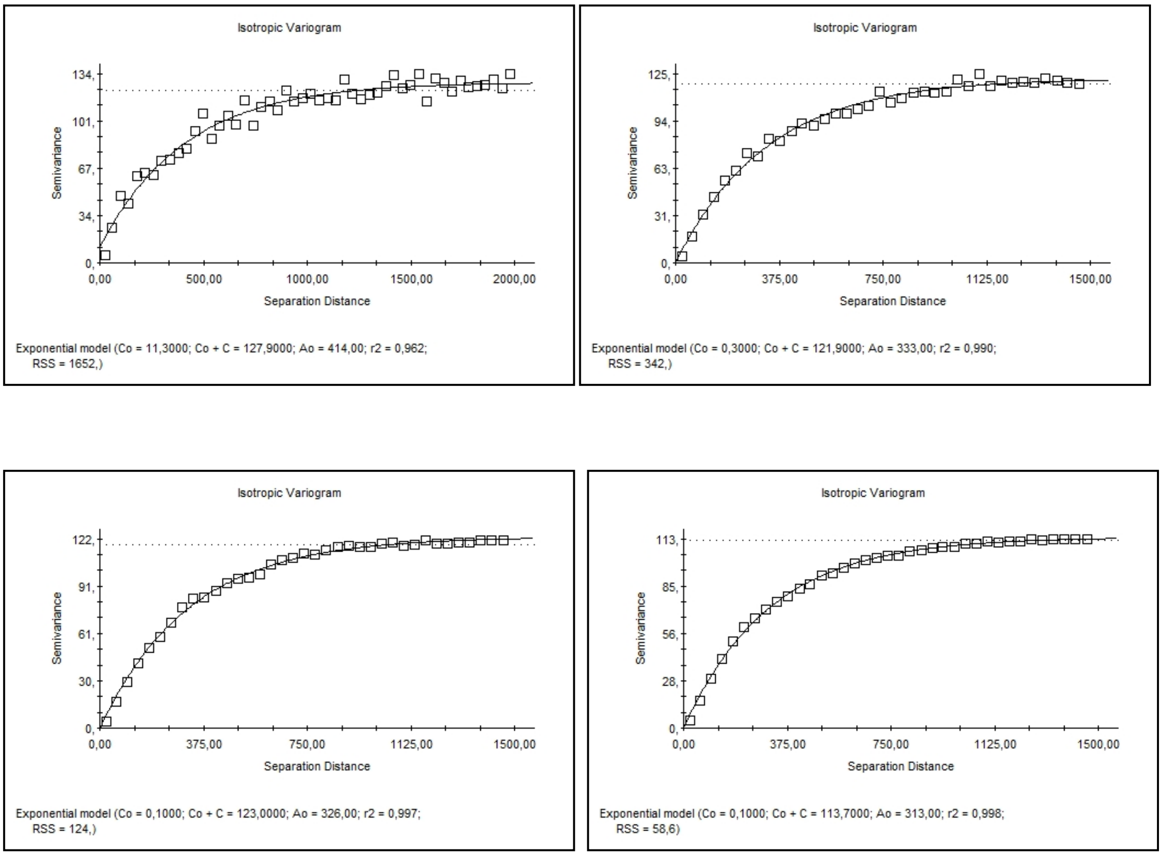


Figure 7. Isotropic semivariograms (region 2) from 1000 samples (upper left), 2500 samples (upper right), 5000 samples (lower left), 10000 samples (lower right)

It was also estimated semivariograms for 25000 and 50000 samples and their plot seemed very similar to that presented for 10000 samples in Figure 7. Table2 present the parameters of isotropic semivariograms for 6 different sizes of sample set of the region 2.

Table 2. Parameters of isotropic semivariograms for sample sets of the region 2

# Samples	Nugget Effect	Contribution	Range	RSS
1000	11.3	116.6	1242	1652
2500	0.3	121.6	999	342
5000	0.1	122.9	978	124
10000	0.1	113.6	939	58.6
25000	0.2	118.9	1020	67.8
50000	0.1	119.2	993	57.9

3.2.3 Time for processing the semivariograms

The processing time to calculate and plot the semivariograms varied from 9s, for 5000 samples, to 33 seconds, for 10000 samples, up to 13 minutes, for 50000. To obtain the semivariograms was used an ordinary personal computer with intel i5 processor and 4GBytes of RAM memory. It was detected an exponential growth in processing time relative to the number of samples showing that this processing demand vast computation and is time consuming for large amount of samples.

3.2.4 Analysis and discussions

Considering the results presented in the Figures 5 and 7 and in the Tables 1 and 2 it can be concluded:

- Semivariograms can be assessed and used to represent the spatial variability of image data of forest regions as the two images of the Figure 2.
- The fitted semivariograms are good representations of spatial variability of both regions because the nugget effects are very small, compared to their sills that are close to the total variance of the images. Also, as expected, the sills are larger for region 2 that has more heterogeneous texture than region 1. Also the range is larger to region 1 showing more spatial continuity for this information.
- In Tables 1 and 2, larger RSS values, for 1000 samples for example, mean that the sampling size was not enough to assess the semivariances with good accuracy for the lag distances considered.
- The RSS values for sets larger than, or equal to, 10000 samples are very low showing that the fitting for those experimental semivariograms were very accurate. For these sample sets the semivariogram parameters (nugget effect, contribution and range) are very similar, i.e., present small variance among them.

Therefore, for this experiment, where each image has 250000 pixels, a random sample set of size 10000, 4% of the total of pixels, seems to be enough to get high-quality spatial variability representation for each image. This shows that, in this case, it is not necessary to work in the total data of the image to get good approximations for semivariograms plots. Besides this avoid problems with time and computational costs to get semivariograms from samples. The semivariograms for 10000 samples were evaluated in about only 33 seconds using an ordinary personal computer.

The semivariograms, for the two regions considered, present significant differences in their parameters, mainly in the contribution and range. This is explained by the fact that those two images depicts different forest patterns. So the stationary hypothesis is not valid for the entire band 4 CCD CBERS scene from where the two images were originated. This indicates that it is better work with stratified areas, with similar textures, to get representative semivariograms in forest areas.

.

4. Conclusions

In this work it was found that is possible to get representative semivariograms of the whole image from small sample sets. Besides, it was obtained different semivariograms for the two regions showing that stationary hypothesis cannot be assumed for forest images with different patterns or textures.

Although this article analyses only with two specifics examples of the band4 CCD images of forest regions the results show that is feasible to work with random samples to get representative semivariogram models for remote sensing images. More experiments should be done with different sensors, image sizes, patterns and different wavelength or satellite bands to get more generalized conclusions.

In the future we intend to repeat this methodology with different type of images representing other land uses as urban, agriculture and others. For forest regions we could work also with images derived from combination of original bands as, for example, the normalized difference vegetation index and the principal component. Other types of sample set can be explored, as for example stratified sampling or very important points. Also other type of spatial variability tool as copulas, for example, could be considered.

5. Citations and References

Almeida-Filho, R.; Miranda, F. P.; Lorenzzetti, J. A.; Pedroso, E. C.; Beisl C. H.; Landau L.; Baptista M. C.; and Camargo, E. G. RADARSAT-1 images in support of petroleum exploration: the offshore Amazon River mouth example. **Canadian Journal Remote Sensing**, v. 31, n. 4, p. 289–303, 2005.

Burgess, T. M.; Webster, R. "Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging". **Journal of Soil Science**, v. 31, p. 315-331, 1980.

Camara G., Souza, RCM, Freitas UM and Garrido J. SPRING, Integrating Remote Sensing and GIS by object-oriented data modeling. **Computer & Graphics**, v. 20, n. 17, p.395-403, 1996.

Chica-Olme, M.; Alberca-Hernández, F. Chapter 6 **Derived Image Texture for Classifying Remotely Sensed Images** Book: S.M. de Jong and F.D. van der Meer. **Remote Sensing Image Analysis: Including the Spatial Domain**, Kluwer Academic Publishers. Printed in the Netherlands. 2004. p 93-111.

Deutsch, C.V.; Journel, A.G. **GSLIB: geostatistical software library and user's guide**. New York: Oxford University Press, 1998. 369p.

Felgueiras, C.A. **Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geoestatístico por indicação.** 1999. 165p. PhD Thesis, Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 1999.

Florenzano, T. G. **Imagens de satélites para estudos ambientais**. São Paulo: Oficina de Textos, 2002. 97 p.

Goovaerts, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997. 483p.

Isaaks, E. H; Srivastava, R.M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 561p.

Zhang, H., Lan, Y., Lacey, R. E., Huang, Y., Hoffmann, W. C., Martin, D., Bora, G. C. Analysis of variograms with various sample sizes from a multispectral image. **International Journal of Agriculture & Biological Engeneering.** Open Access at http://www.ijabe.org , v. 2, n. 4, p. 62-69 , 2009

Zhang, C., Li, W. and Travis, D., Restoration of clouded pixels in multispectral remotely sensed imagery with cokriging. **International Journal of Remote Sensing**, v. 30, p. 2173–2195, 2009.