Testing Spatial Isotropy by Using a Non Parametric Bootstrap Approach

Marcelo Miranda[1], Ronny Vallejos[2]

[1]Departamento de Ciencias Forestales / Pontificia Universidad Católica de Chile
mmirands@uc.cl

[2]Departamento de Estadística CIMFAV / Universidad de Valparaíso
ronny.vallejos@uv.cl

Abstract. A method that allows testing the presence of spatial isotropy is proposed. The procedure is based on a non-parametric bootstrap approach where the means of the semi-variance are compared for two fixed azimuths and an arbitrary lag distance. A simulation study and application in forest science real data were developed to investigate the performance of this method. In the case of an isotropy process (simulated data), the hypothesis of isotropy was accepted for all analyzed lag distances. For an anisotropic process (real data) and depending on the lag distances that were analyzed, the isotropy hypothesis was accepted or rejected. Bootstrap for dependent observations will be studied in future to address the possible dependency among the semi-variance pair values.

1 INTRODUCTION

Spatial dependency in terms of spatial isotropy or anisotropy is a fundamental feature in spatial modeling for natural resources [1], [2], [3]. A correlation structure is called isotropic if the correlation between observations at any two sites depends only on the distance between those sites and not on their relative orientation [4]. Otherwise, it is said to be anisotropic. On the other hand, two types of anisotropy are usually found in geostatistical studies: i) geometrical, which implies that the correlation iso-level contours are ellipses o ellipsoids and ii) zonal, which implies a direction-dependent variogram sill [5]. The focus of our work is on zonal anisotropy.

Conventional methods to evaluate the presence of spatial anisotropy are based on a visual analysis of the directional semi-variance, variogram maps and the anisotropy ratio [6], [7], [8], [9]. These analyses allow the observation of differences in the mean values of semi-variance between several directions or azimuths; however, it does not allow determining if these dissimilarities have a real statistical significance.

Few statistical approaches have been developed to evaluate the spatial anisotropy and the computational implementation is difficult in most of them [10], [11], [4].

MIRANDA, M. and VALLEJOS, R., 2009 Testing Spatial Isotropy by using a non parametric bootstrap approach. StatGIS 2009: Geoinformatics for Environmental Surveillance. Milos Greece June 17-19.

In the case of verification of spatial isotropy it is important to determine if, for a specific lag distance h, the mean semi-variance values for two azimuths θ and Φ have the probability to be equal.

A mean comparison test could be implemented to accomplish this objective, but it is necessary to know the basic parameters of the probability distribution of semi-variance for azimuths θ and Φ for the specific lag distance analyzed.

The Bootstrap Re-sampling Method [12] is an approach that let the estimation of parameters of a distribution (mean, standard error and confidence intervals) and allows to implement a test. Bootstrap is based on the idea that a sampling is a good representation of the distribution studied. Therefore, the distribution can be estimated generating new samples repeatedly from the original sample.

The advantages of bootstrap(s), compared to the classical statistical approach, is its better accurateness, flexibility in relation to the dimension of the samplings and their independence as regards a law of classic distribution of data. According to Jeong and Chung [13], if we have a set of observations $x_1, x_2..., x_n$ and a statistic $\hat{u}$, the basic Bootstrap procedure is the following:

1) To generate a «bootstrap sample» $B_1 = x_1^*, x_2^*, \ldots, x_n^*$ from the original sample. Every x_i^* is chosen randomly from $x_1, x_2 \ldots, x_n$ with replacement.

2) Evaluate the statistic $\hat{u}$ from B_1.

3) Repeat the steps 1) and 2) m time to get $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_m$.

4) Estimate the distribution of $\hat{u}$ from the Bootstrap distribution F assigning mass $1/m$ to every $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_m$. In this case, the mean and the *BCa* confidence intervals (Bootstraps bias–corrected accelerated intervals) are the most interesting parameters to be estimated from the distribution F [12].

The aim of this work is to present a test to determine the statistical significance of spatial isotropy based on the Bootstrap re-sampling method. The mean values of semi-variance for two azimuths θ and Φ, for a specific lag distance h, are compared based on Bootstrap distribution F. We introduce the test procedure using a simulation data set and a forest real data set.

2 DATA DESCRIPTION AND METHOD

2.1 Simulation Study

We considered an isotropic normal distributed random field X defined on $D \subset Z^2 \rightarrow N(\mu, \Sigma)$, with an exponential distance-based covariance. The corresponding variogram model is:

$$\gamma(h) = 1 - \exp\left(\frac{-3h}{a}\right)$$

A random field X is said to be isotropic when i) the mean vector function μ is constants on D and, ii) the covariance function does not depend on the azimuth angle. The simulated normal distributed random field was a 20 x 20 grid (400 sample points) and a =11.1.

2.2 Application Study

We applied the proposed test to topographic and forest data represented by elevation and tree height at the age of four years. The study zone, covering an area of 3011.8 ha, is located in the "Escuadrón" sector, south of Concepción in Southern Chile (36°54', 73°54'). The general geomorphology of the site corresponds to a range of Coastal Mountains having mainly an abrupt topography and an elevation reaching 500 m.a.s.l. The landscape is dominated by steep, short-length topographic profiles [14] and the site is representative of the *P. radiata* plantations existing in mountainous conditions near the sea in the south-central part of Chile.

Elevation was obtained from a Digital Elevation Model produced applying a kriging method [15], to points located in 10-meter topographic contour lines, thus producing a RMSE for elevation of 2.8 m. Tree height was obtained from a pre-pruning forest inventory, composed of 200 m^2 circular sample plots. The plots were located systematically using a mean distance of 150 m among them. The total number of sample plots available in the study site was 332.

2.3 The test procedure

If the means of semi-variance for two azimuths θ and Φ are given, it is of interest to find out if this process could be considered isotropic or anisotropic. The difference of means of semi-variance values $\gamma_\theta(h)$ and $\gamma_\Phi(h)$ at different lag distances h gives a natural estimate. Considering that statistical inference in this case is difficult [16], since the distribution of this estimate does not have a closed form, we propose the following resampling procedure:

For two fixed azimuths θ and Φ and for an arbitrary lag distance h:

Step 1. Calculate the list of pairs of semi-variance values for each azimuth θ and Φ.

Step 2. Select a random sample B_θ and B_Φ of pairs of semi-variance values $B_\theta = x_{\theta 1}, x_{\theta 2} \dots, x_{\theta n}$ and $B_\Phi = x_{\Phi 1}, x_{\Phi 2} \dots, x_{\Phi n}$ for each azimuth θ and Φ. With n > 50.

Step 3. To generate a «bootstrap sample» $B_{\theta 1} = x_{\theta 1}^*, x_{\theta 2}^*, \dots, x_{\theta n}^*$ and $B_{\Phi 1} = x_{\Phi 1}^*, x_{\Phi 2}^*, \dots, x_{\Phi n}^*$ from the original sample B_θ and B_Φ. , every element of $B_{\theta 1}$ and $B_{\Phi 1}$ is chosen randomly from B_θ and B_Φ with repetition.

Step 4. Evaluate the statistic $\hat{u}$; $\hat{u} = B_{\theta 1} - B_{\Phi 1}$.

Step 5. Repeat the steps 3) and 4) m times to get $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_m$.

Step 6. Estimate the distribution of $\hat{u}$ from the Bootstrap distribution F assigning mass $1/m$ to every $\hat{u}_1, \hat{u}_2, \dots, \hat{u}_m$. Calculate the mean and the *BCa* confidence intervals (Bootstrap bias–corrected accelerated intervals) from the distribution F ([14].

Step 7. Based on the *BCa* confidence intervals, formulate the following hypotheses:

H_0: The process is isotropic. There is no significant difference between the means of the semi-variance of azimuths θ and Φ for the lag distance h. The test is carried out using a 95% confidence interval based on the percentiles (2,5% - 97,5%) of the Bootstrap distribution F.

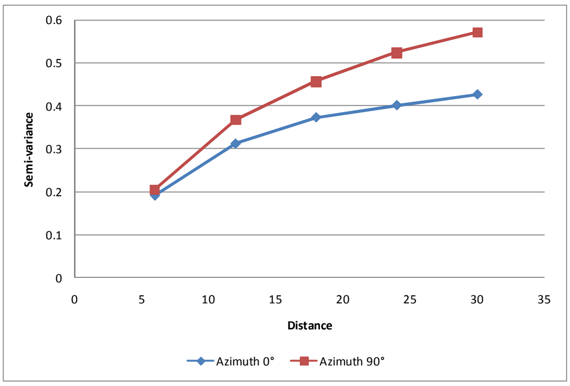
H_1:The process is anisotropic. The means of semi-variance of azimuths θ and Φ are significantly different for a lag distances h.

For the simulation and application studies, semi-variograms were generated considering: azimuths 0° and 90° for lag distances $h = 6$ units and $h=200$ m respectively (Figure 1a, b and c). The number of simulation runs in the bootstrap test were $m = 5000$.

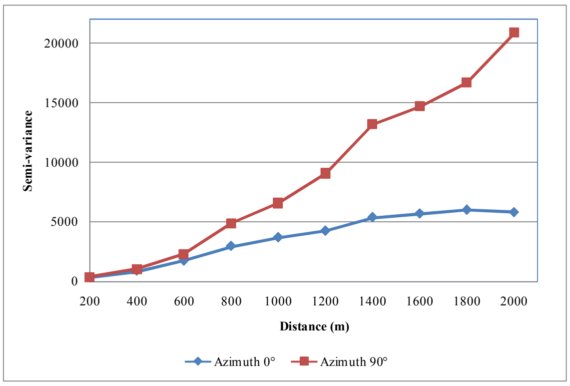
3 RESULTS

For the simulation study (Figure 1a), it is first noticed that for all lag distances analyzed, the isotropy hypothesis was not rejected with significance levels of 5% (Figure 1d). Next, it is also observed that this hypothesis is accepted with more power for large lag distances. These behaviors are the result of the semi-variance pair values distribution inside each lag distance analyzed. With an increasing distance, neighbors tend to be more dissimilar but always with a probability to have means of semi-variance similar.

For elevation and tree height, a visual analysis shows an apparent anisotropy at different spatial scales (Figure 1b and c). However, for both variables the test does not reject the hypothesis of isotropy for lag distances of less than 600 meters for elevation and 1,600 meters for tree heights (Figure 1e and f). Over these distances, the null hypothesis begins to be rejected more strongly.

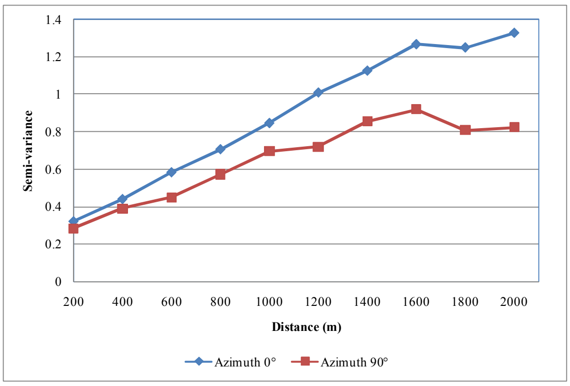


a) Semi-variance of simulated data.

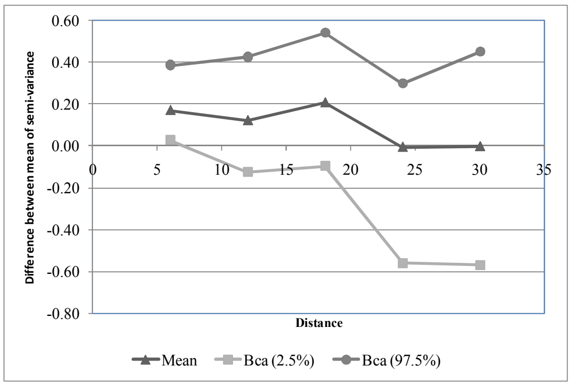


b) Semi-variance of Elevation.

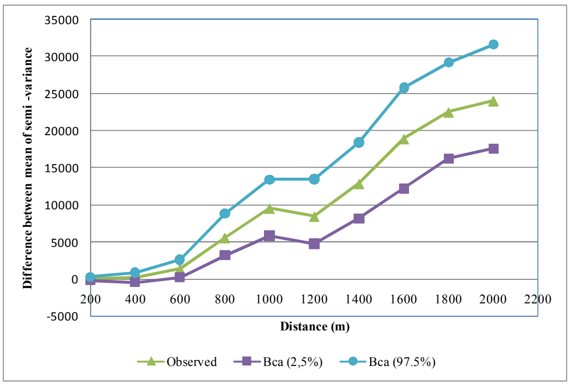
MIRANDA, M. and VALLEJOS, R., 2009 Testing Spatial Isotropy by using a non parametric bootstrap approach. StatGIS 2009: Geoinformatics for Environmental Surveillance. Milos Greece June 17-19.



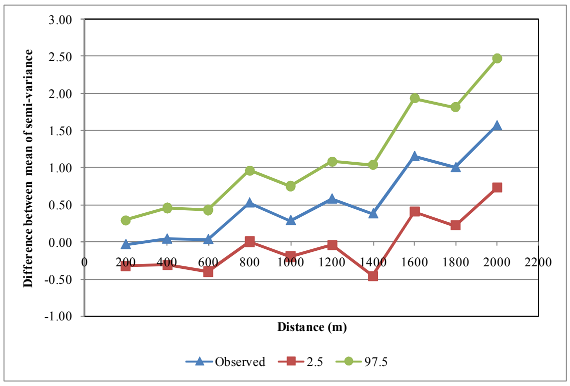
c) Tree height semi-variance .



d) Bootstrap test for simulated data.



e) Bootstrap test for elevation.



f) Bootstrap test for Tree height.

Figure 1. Semi-variance and results of Bootstrap test for the simulation and application studies.

4 Conclusion and Future work

In this work we have introduce a simple test to prove isotropy base on Bootstrap approach. As a result, with this approach it is possible to determine the lag distances values where a spatial process can be considered either isotropic or anisotropic. This test can be used as an alternative to other methods based on visual analysis. Bootstrap for dependent observations will be studied in the future to address the possible dependency among semi-variance pair values. The possible improvement of the proposed method when using bootstrap for spatial dependent observations (e.g. Sherman's proposal [17]), is an interesting open issue to be examined in future research from the theoretical and applied perspective.

5 References

[1] Lichstein, J., Simon, T.R., Shriner, S.A. and Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models. Ecological Monographs. 72: 445-463.

[2] Austerlitz, F., Dutech, C., Smouse, P., Davis, F. and Sork, V., 2007. Estimating anisotropic pollen dispersal: a case study in Quercus lobata. Heredity 99: 193–204.

MIRANDA, M. and VALLEJOS, R., 2009 Testing Spatial Isotropy by using a non parametric bootstrap approach. StatGIS 2009: Geoinformatics for Environmental Surveillance. Milos Greece June 17-19.

[3] Miranda, M., 2007. Modélisation de la productivité forestière de Pinus radiata D. Don à partir de patrons de variabilité spatio-temporelle de l'espèce et de l'environnement. Thèse Doctorale. Université Laval, Canadá. 150 p.

[4] Guan, Y., Sherman, M., and Calvin, J. 2004. A Nonparametric test for spatial isotropy using subsampling. Journal of the American Statistical Association, 99: 810-821.

[5] Chorti, A. and Hristopulos, D. T., 2008. Non-parametric Identification of Anisotropic (Elliptic) Correlations in Spatially Distributed Data Sets. IEEE Transactions on Signal Processing, 56: 4738-4751.

[6] Isaaks, E.H. and Srivastava, R.M., 1989. An introduction to applied geostatistics Oxford University Press, New York. 561 pp.

[7] Rossi, R., Mulla, D., Jurnel, A. and Franz, H. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecological Monograph. 62:277-314.

[8] Zimmerman, D. 1993. Another look at anisotropy in geostatistics. Mathematical Geology. 24: 453-471.

[9] Goovaerts, P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New-York. 483 pp.

[10] Lu, H., 1994. On the Distributions of the sample covariogram and semivariogram and their use in testing for isotropy. Unpublished doctoral dissertation. University of Iowa, Dept. of Statistics and Actuarial Science.

[11] Molina, A., and Feito, F., 2002. A Method for testing anisotropy and quantifying its direction in digital images. Computers & Graphics. 26:771–784.

[12] Efron, E., and Tibshirani, R. 1993. Bootstrap method for standard errors, confidence intervals and other measures of statistical accuracy. Statistical Science. 1: 54-75.

[13 Jeong, J., and Chung, S., 2001. Bootstrap tests for autocorrelation. Computational Statistics & Data Analysis. 38:49-69.

[14] Villagran, C. 1993. El cuaternario de la región de Los Lagos del sur de Chile. Guía de excursión. In: Taller Internacional El Cuaternario de Chile, Santiago.

[15] Burrough, P. and McDonnell, R. 1998. Principles of geographical information systems. Oxford University Press Inc., New York. 333 pp.

[16] Timmer, J. Lauk, M., Vach, W. and Lücking, C. 1999. A test for difference between spectral peak frequencies. Computational Statistics & Data Analysys. 30:45-55.

[17] Sherman, M., 1996. Variance estimation for statistics computed from spatial lattice data. Journal of the Royal Statistical Society, Series B, 58, 509-523.

MIRANDA, M. and VALLEJOS, R., 2009 Testing Spatial Isotropy by using a non parametric bootstrap approach. StatGIS 2009: Geoinformatics for Environmental Surveillance. Milos Greece June 17-19.