A nonparametric study of the spatial association between forest variables

Francisco Cuevas[1], Emilio Porcu[1], Ronny Vallejos [1]

[1] Universidad Técnica Federico Santa María, Chile

E-mail for correspondence: ronny.vallejos@usm.cl

Abstract: We propose a nonparametric approach to assess the spatial association between two spatial variables. Our proposal is based on a Nadaraya-Watson version of the codispersion coefficient through a suitable kernel. The proposed method is useful for quantifying spatial associations between two variables measured at the same locations. We study forest data concerning the relationship among the tree height, basal area, elevation and slope of Pinus radiata plantations. A two-dimensional codispersion map is constructed to provide insight into the spatial association between these variables.

Keywords: Kernel; Spatial association; Basal area and height.

1 Introduction

In the analysis of spatial data, the quantification of spatial associations between two variables is an important issue, and considerable effort has been devoted to the construction of appropriate coefficients and tests for the association between two correlated variables.

The codispersion coefficient (Matheron, 1965) is a measure of association between two spatial variables and has been used in several applications (Chilés and Delfiner, 1999; Vallejos, 2012). Such a measure is a normalized version of the cross-variogram, being a crucial instrument for multivariate spatial prediction (Ver Hoef and Barry, 1998). Rukhin and Vallejos (2008) studied the codispersion coefficient from both theoretical and applied viewpoints, and established, for arbitrary lags, the consistency and limiting distribution of the sample coefficient. Recently, Vallejos (2012) studied some extensions of the codispersion in a time series context, while Ojeda et al., (2012) used the codispersion coefficient to assess the similarity between two digital images.

The goal of the work is to use a nonparametric version of the codispersion coefficient to assess the spatial association between several pairs of forest variables. Such extensions of this nature have previously been considered in the spatial statistics literature (García-Soidán et al., 2004).

The study of the forest variables is based on a data set of Pinus radiata plantations in the south of Chile. Through the use of codispersion maps, we explore the spatial association of these variables.

2 Methods

Throughout the paper we shall consider intrinsically stationary random fields $\{X(\boldsymbol{s}),\ \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ with semi-variogram defined as

$$\gamma_X(\boldsymbol{k}) = \frac{1}{2}\mathrm{var}\{X(\boldsymbol{s}+\boldsymbol{k}) - X(\boldsymbol{s})\}, \tag{1}$$

where $\boldsymbol{k} \in \mathbb{R}^d$ denotes the spatial lag. For n sampling sites $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n$, a natural and unbiased estimator based on the method of moments is the empirical semi-variogram given by

$$\widehat{\gamma}_X(\boldsymbol{k}) = \frac{1}{2|N(\boldsymbol{k})|}\sum_{N(\boldsymbol{k})}(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2, \tag{2}$$

where $N(\boldsymbol{k}) = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : ||\boldsymbol{s}_i - \boldsymbol{s}_j|| \in T(\boldsymbol{k}), 1 \leq i, j \leq n\}$, $T(\boldsymbol{k})$ is a tolerance region around $\boldsymbol{k}$, and where $|\cdot|$ denotes cardinality of a set. García- Soidán (2007) proposed a Nadaraya-Watson type estimator for the semi-variogram defined as

$$\breve{\gamma}_{X_h}(\boldsymbol{k}) = \frac{\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)(X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2}{2\sum_{i=1}^n \sum_{j=1}^n K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)}, \tag{3}$$

where h represents a bandwidth parameter and $K : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a symmetric and strictly positive density function. For such an estimator, García-Soidán (2007) establishes, under regularity conditions, consistency and asymptotic normality, and addresses the inadequate behavior of estimator (3) near the endpoints.
Let $\{(X(\boldsymbol{s}), Y(\boldsymbol{s})) : \boldsymbol{s} \in D \subset \mathbb{R}^d\}$ be a bivariate intrinsically stationary random field on D with cross-variogram $2\gamma_{XY}(\cdot) : \mathbb{R}^d \to \mathbb{R}$ defined through

$$2\gamma_{XY}(\boldsymbol{k}) = \mathbb{E}[(X(\boldsymbol{s}+\boldsymbol{k}) - X(\boldsymbol{s}))(Y(\boldsymbol{s}+\boldsymbol{k}) - Y(\boldsymbol{s}))], \tag{4}$$

for all $\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{k} \in D$, and with marginal variograms $2\gamma_X$, $2\gamma_Y$ as defined through Equation (1). The codispersion coefficient (Matheron, 1965) is a normalized version of (4) and defined through

$$\rho_{XY}(\boldsymbol{k}) = \frac{\gamma_{XY(\boldsymbol{k})}}{\sqrt{\gamma_X(\boldsymbol{k})\gamma_Y(\boldsymbol{k})}}.$$

Rukhin and Vallejos (2008) and Vallejos (2008) found a closed form for the coefficient $\rho_{XY}(\cdot)$ for spatial autoregressive processes under particular

assumptions on the correlation structure of the errors and when considering a rectangular lattice.
The analogue of Matheron's estimator for the cross-variogram is obtained through

$$\widehat{\gamma}_{XY}(\boldsymbol{k}) = \frac{1}{2|N(\boldsymbol{k})|}\sum_{N(\boldsymbol{k})}(X(\boldsymbol{s}_i)-X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i)-Y(\boldsymbol{s}_j)), \tag{5}$$

where $N(\boldsymbol{k})$ is defined as in Equation (2). The corresponding empirical estimator of the codispersion based on (5) is given by

$$\widehat{\rho}_{XY}(\boldsymbol{k}) = \frac{\widehat{\gamma}_{XY}(\boldsymbol{k})}{\sqrt{\widehat{\gamma}_X(\boldsymbol{k})\widehat{\gamma}_Y(\boldsymbol{k})}}. \tag{6}$$

The analogue of the Nadaraya-Watson type estimator for the cross-variogram is instead given by

$$\breve{\gamma}_{XY_h}(\boldsymbol{k}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)(X(\boldsymbol{s}_i)-X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i)-Y(\boldsymbol{s}_j))}{2\sum_{i=1}^{n}\sum_{j=1}^{n}K\left(\frac{\boldsymbol{k}-(\boldsymbol{s}_i-\boldsymbol{s}_j)}{h}\right)}, \tag{7}$$

where $K(\cdot)$ is a kernel function as in Equation (3).
A kernel type estimator of the codispersion coefficient is

$$\breve{\rho}_{XY_{\boldsymbol{h}}}(\boldsymbol{k}) = \frac{\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})}{\sqrt{\breve{\gamma}_{X_{h_2}}(\boldsymbol{k})\breve{\gamma}_{Y_{h_3}}(\boldsymbol{k})}}, \tag{8}$$

where $\boldsymbol{h} = (h_1, h_2, h_3)$, $\breve{\gamma}_{XY_{h_1}}(\boldsymbol{k})$ is as in (7) and $\breve{\gamma}_{X_{h_2}}(\boldsymbol{k})$ is as in (3).
Cuevas et al., (2013) established the consistency of estimator (8). In addition, asymptotic expressions for the bias and mean square error were derived for estimator (7). A bandwidth selection rule for the variogram and the cross-variogram was also provided.

3 An Application

Here, we present an example of an issue that motivated the present work. Pinus radiata is one of the most widely planted species in Chile; it is planted on a wide array of soil types and in a variety of regional climates. Two important measures of plantation development are the dominant tree height and the basal area. Snowdon argues convincingly that both measures are correlated with regional climate and local growing conditions (Snowdon, 2001). The variogram was used to characterize the spatial dependence of each variable. However, the assessment of the spatial association between tree height, tree basal area and other regional climate variables is of great interest for the quantification of spatial dependence and the detection of those directions in which there is either high or low degree of spatial association.

In the present article, we consider the relationship among the tree height, basal area, elevation and slope of Pinus radiata plantations. The study site is located in the sector *Escuadrón*, south of Concepción in the southern portion of Chile (36° 54' S, 73° 54' O) and has an area of 1244.43 hectare. In addition to more mature stands, we were also interested in the area containing young (i.e., four year old) stands of Pinus radiate, with an average density of 1600 trees per hectare. The basal area and dominant tree height at the year of plantation establishment (1993, 1994, 1995, and 1996) were used to represent the stand attributes. The three variables were obtained from 200 m^2 circular sample plots and point-plant sample plots. For the latter type of sample, four quadrants are established around the sample point; the four closest trees in each quadrant (16 trees in total) are then selected and measured in a clockwise direction. The samples were located systematically using a mean distance of 150 meters between samples. The total number of plots available for this study was 468. In addition to the tree height and basal area, the coordinates, elevation and slope were recorded for each site.

In this talk, we will discuss the construction of a codispersion map based on Equation (8) to provide better insight into the spatial associations between all pairs of variables in several different directions on a two-dimensional space. Figure 1 shows the codispersion maps that were created from the variables of interest using a rectangular grid. We provide a full description of the findings. In particular, we find those directions for which the codispersion coefficient is maximum or minimum for the forest data.

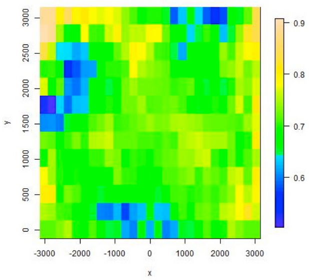
Acknowledgments: This work was supported in part by Fondecyt, Grant 1120048, Chile.

References

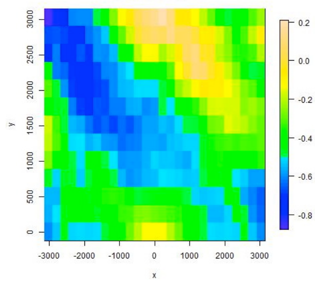
Chilés, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty.* New York: Wiley.

Cuevas, F., Porcu, E., Vallejos, R. (2013). Study of spatial relationships between two sets of variables: A nonparametric approach. Submitted.

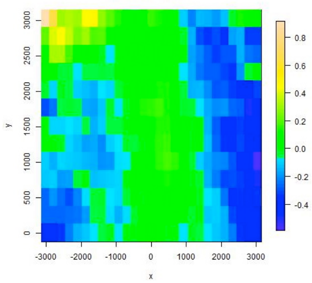
García-Soidán, P., Febrero, M. and González, W. (2004). *Journal of Statistical Planning and Inference*, **121**, 65 – 92.



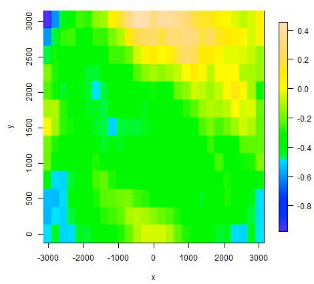
(a)



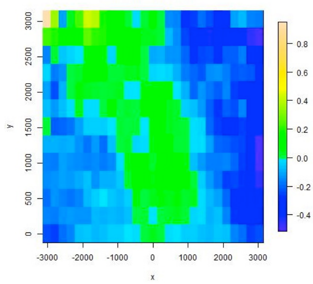
(b)



(c)



(d)



(e)

FIGURE 1. Codispersion map between all pairs of variables of interest.

García-Soidán, P. (2007). Asymptotic normality of the Nadaraya-Watson semivariogram estimators. *Test*, **16**, 479 – 503.

Matheron, P.J. (1965). *Les Variables Régionalisées et leur Estimation.* Paris: Masson.

Ojeda, S., Vallejos, R., Lamberti, P. (2012). Measure of Similarity Between Images Based on the Codispersion Coeofficient. *Journal of Electronic Imaging*, **21**, 023019.

Rukhin, A. and Vallejos, R. (2008). Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters*, **78**, 1290 – 1300.

Snowdon, P. (2001). Short-term predictions of growth of Pinus radiata with models incorporating indices of annual climatic variation. *Forest Ecology and Management*, **152** , 1-11.

Vallejos, R. (2008). Assessing the association between two spatial or temporal se quences. *Journal of Applied Statistics*, **35**, 1323 – 1343.

Vallejos, R. (2012). Testing for the absence of correlation between two spatial or temporal sequences. *Pattern Recognition Letters*, **33**, 1741 – 1748.

Ver Hoef, J. M. and Barry, R. P. (1998). Constructing and fitting models for cokriging and multivariate spatial prediction. *Journal of Statistical Planning and Inference*, **69**, 275 – 294.