The Codispersion Coefficient: An Application in the Evaluation of the Performance of Different Spatial Interpolators

Ronny Vallejos[1], Marcelo Miranda[2]

[1]Departamento de Estadística-CIMFAV /Universidad de Valparaíso, Chile
ronny.vallejos@deuv.cl

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

Abstract. We examine the quantitative assessment of the association between two spatial sequences to evaluate the performance between different spatial interpolators. In the context of natural resources this problem is relevant in several different fields. For example, in precision farming and forest productivity it is of interest to study the performance of interpolators in the generation of digital elevation models. In this work the codispersion coefficient is considered for intrinsic stationary spatial processes. Recent asymptotic results allow the construction of hypothesis testing and confidence intervals for the coefficient. To illustrate the applicability of the coefficient, a real data example was considered in the context of digital elevation models, in a study area located in a micro-basin in the south of Chile. Three surfaces generated by kriging, spline and inverse distance weighted (IDW) were evaluated using the conventional approach based on the root mean square error. Since this approach does not consider the spatial dependence present in the data, the codispersion coefficient was computed to compare all possible pairs of interpolations. The results show that the codispersion coefficient captures spatial association that is not possible to obtain using the conventional methods.

1 Notation and Definitions

The association between two spatial processes has been extensively studied in spatial statistics. Several coefficients of association have been proposed to summarized in a single number the association characteristics of two spatial or temporal sequences. In a nonparametric context Tjøtheim [15] introduced a coefficient for spatial variables. Subsequently Clifford et. al, [5], [7] proposed a coefficient that is a corrected version of the correlation coefficient. That measure was generalized later by Richardson and Clifford [12]. As a result, tests of spatial association based on the correlation coefficient have been implemented (see also [10]).

The codispersion coefficient was first introduced by Matheron [11] to study the spatial association between two spatial sequences. This coefficient is defined as a normalized version of the cross-variogram between two spatial variables at distance lag h. Thus it can be interpreted as a measure of the dependence of the spatial structure of different variables. The codispersion coefficient can also be interpreted as a linear correlation coefficient between spatial increments of both atributes (Goovaerts, [8]). This coefficient has been applied in several different fields such as soil sciences [8], time series [4], hydrology [1], biology [2] among others.

Definition 1. Consider two weakly stationary processes, X and Y, defined on $D \subset \mathbb{Z}^d$.

The cross-variogram between X and Y is defined as

$$\gamma(h) = \mathbb{E}[X(s+h) - X(s)][Y(s+h) - Y(s)], \tag{1}$$

such that $s, s+h \in D$.

Definition 2. The codispersion coefficient is a normalized version of (1) given by

$$\rho(h) = \frac{\gamma(h)}{\sqrt{V_X(h)V_Y(h)}}, \tag{2}$$

where $V_X(h) = \mathbb{E}[(X(s+h) - X(s))^2]$.

This coefficient has been derived for several different types of parametric models, for example, for MA(∞), AR(1) and first order spatial AR models explicit expressions for $\rho(h)$ can be found in Rukhin and Vallejos [13].

In a one-dimensional case, this coefficient can be interpreted as a comovement coefficient for two time series [16]. It is related to the mean squared successive difference $\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2/(n-1)$, that was studied by Von Neumann in the forties. Its exact distribution has been derived [9] and its variance in the i.i.d. normal case is $4\sigma^4(3n - 4)/(n-1)^2$. The ratio of the mean squared successive difference statistic to the sample variance $\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2/\sum_{i=1}^{n}(x_i - \overline{x})^2$ is known variously as the Von Neumann ratio, or the Durbin-Watson statistic in the context of regression. It is used as a test statistic for the independence of Gaussian observations, or as a test for independence versus the alternative of nonzero first order autocorrelation.

2 Asymptotic Results

For $Z(s) = (X(s), Y(s))^T$, assume that

$$Z(s_1 + h_1, s_2 + h_2) - Z(s_1, s_2) = \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} A(k,l)\epsilon(s_1 - k, s_2 - l), \tag{3}$$

where $A(k,l) = A_h(k,l)$ are 2×2 matrices defined for all integer k and l, such that

$$\sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} ||A(k,l)||^2 < \infty.$$

Here $||\cdot||$ denotes any matrix norm, and two-dimensional random vectors $\epsilon(t)$ are independent with mean 0 and the covariance matrix Σ. Condition (3) is desirable to guarantee consistency and asymptotic normality of the sample codispersion coefficient.

We define $K = E[Z(s+h) - Z(s)][Z(s+h) - Z(s)]^T$ as the covariance matrix of the vector process for which (3) holds. Then the codispersion coefficient can be computed using the elements of K as

$$\rho(h) = \frac{\kappa_{12}}{\sqrt{\kappa_{11}\kappa_{22}}},$$

where for $i, j = 1, 2$, κ_{ij} denote elements of K.

Assume now that both processes can be observed on the increasing part of the positive lattice $\{0 \le s_1 < M, 0 \le s_2 < M\}$, and suppose normality of the error vectors $\epsilon(t) = (\epsilon_1(t), \epsilon_2(t))^T$ in (3), although the main results hold for other distributions with four first moments. For the sample version of the codispersion coefficient given by

$$\widehat{\rho}(h) = \frac{\sum_{0\le s_i < M-h_i}(X(s_1+h_1, s_2+h_2) - X(s_1,s_2))(Y(s_1+h_1,s_2+h_2) - Y(s_1,s_2))}{\sqrt{\sum(X(s_1+h_1,s_2+h_2)-X(s_1,s_2))^2 \sum(Y(s_1+h_1,s_2+h_2)-Y(s_1,s_2))^2}},$$

asymptotically we have $M\,[\widehat{\rho}\,(h) - \rho(h)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2)$, where

$$v_h^2 = \frac{\varphi_{1122}}{\kappa_{11}\kappa_{22}} + \frac{\kappa_{12}^2\varphi_{1111}}{4\kappa_{11}^3\kappa_{22}} + \frac{\kappa_{12}^2\varphi_{2222}}{4\kappa_{11}\kappa_{22}^3} - \frac{\kappa_{12}\varphi_{1112}}{\kappa_{11}^2\kappa_{22}} - \frac{\kappa_{12}\varphi_{1222}}{\kappa_{11}\kappa_{22}^2} + \frac{\kappa_{12}^2\varphi_{1212}}{2\kappa_{11}^2\kappa_{22}^2}, \qquad (4)$$

are the entries of 4×4 matrix Φ defined as

$$\begin{aligned}\Phi = \mathbb{E}&\left[(Z(s+h)-Z(s))(Z(s+h)-Z(s))^T - K\right] \\ &\otimes \left[(Z(s+h)-Z(s))(Z(s+h)-Z(s))^T - K\right].\end{aligned}$$

where $\otimes$ represent the Kronecker product between two matrices.

3 An Application: Evaluation of the Performance of Spatial Interpolators

3.1 Data Description

The study area, covering an area of 3011.8 ha, is located in the "escuadrón" sector, south of Concepción in the southern portion of Chile ($36^o54'$, $73^o54'$) and belongs to the Forestal Mininco S.A. company. The general geomorphology of the site corresponds to a range of coastal mountains (Cordillera de la Costa) that has mainly an abrupt topography and an elevation reaching 500 m.a.s.l. The landscape is dominated by steep short-length topographic profiles [17].

Since in practice it is difficult to have a continuous representation of the terrain in the study area a Digital Elevation Model (DEM) was produced, applying kriging, spline and IDW interpolation methods [3], [6] to points located on 10-meter topographic contour lines, thereby producing a root mean square error (RMSE) for an elevation of 2.86 m, 3.08 m and 3.94 m respectively. The 271×155 images produced by these interpolation methods are shown in Figure 1 (a), (b) and (c).

3.2 Results

First notice that kriging was the best predictor, spline the second and IDW the worse one in terms of RMSE. However, the approach that considers the RMSE as a measure of discrepancy, does not consider the spatial dependence present in the data. Second, the codispersion coefficient was computed to compare all possible pairs of interpolations e.g., kriging versus spline. In all cases $\rho(h)$ was evaluated for $h_1, h_2 = 1, \ldots, 20$ (see Figure

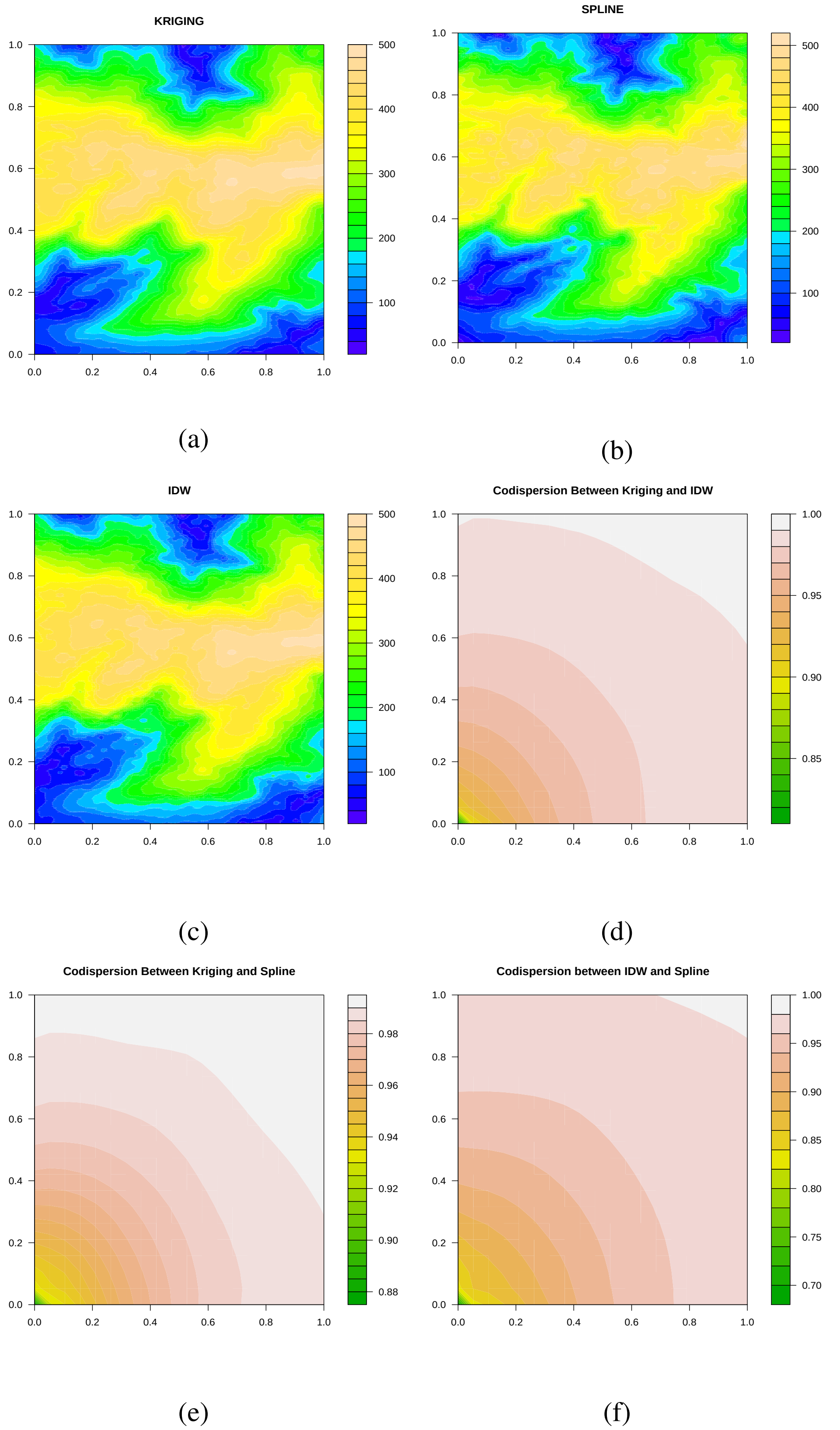

Figure 1: (a)Surface generated using kriging; ; (b) Surface generated using spline; (c) Surface generated using the IDW method;(d) $\rho(h)$ between kriging and IDW; (e) $\rho(h)$ between kriging and spline; (f) $\rho(h)$ between IDW and spline.

1 (d), (e) and (f)). The highest values for the codispersion coefficient were observed for kriging versus spline (0.88-1.00). A circular pattern of the codispersion coefficient for a fixed lag distance was also observed in all comparisons. In all cases the association at a large scale was bigger than spatial association at a small scale. The coefficient turns out to be constant for lag distances bigger than 400 meters. This distance is related with the mean distance between the highest and lowest point in the micro-basine. The codispersion coefficient captures spatial association that is not possible to obtain using the conventional methods.

4 CONCLUSIONS AND FUTURE WORK

In this paper we have discussed an application of the codispersion coefficient in the context of evaluation of spatial interpolators. The coefficient provides useful information about the spatial association between two processes. The construction of confidence intervals for $\rho(h)$ requires to fit suitable models to each image shown in Figure 1 (a), (b) and (c). Clearly, in this case one single spatial AR model will not be appropriate for any of these images since there are several textures that will not be captured by a single spatial AR model. One way to deal with these kind of image is to preprocess each image to produce K regions in which the process is homogeneous. Then a spatial AR model can be fitted to each category. Finally, an hypothesis test can be implemented for the resulting vector of codispersion coefficients between the same category for processes X and Y. The development of such a test is a matter of current research.

When the models involve a large number of parameters, the derivation of the asymptotic variance of the sample codispersion coefficient is difficult. Resampling techniques for spatial dependent variables seem adequate to obtain estimations for the variance of the coefficient. For instance, Sherman [14] developed a method to deal with the variance estimation for statistics computed from spatial lattice data. We are in the process of implementing Sherman's technique for the codispersion coefficient. We view the work described in this paper as only the beginning of a large project with several open problems to be tackle in the future.

5 ACKNOWLEDGEMENTS

The authors thank two anonymous referees and the editor for their valuable suggestions. The first author was partially supported by Fondecyt grant 11075095, Chile.

REFERENCES

[1] S. Barnes, J. Baxter and R. M. Lark. Spatial covariation of azotobacter abundance and soil properties: A case study using the wavalet transform. *Soil Biology & Biochemistry*, 39:295–310, 2007.

[2] L. Blanco-Moreno, J.M. Chamorro and F. X. Sans. Spatial and temporal patterns of lolium rigidum-avena sterilis mixed populations in a cereal field. *Weed Research*, 46:207–218, 2006.

[3] P.A. Burrough and R.A. McDonnell. Principles of geographical information systems. *Oxford University Pres Inc.*, page 333, 1998.

[4] A. Chouakria and P. Nagabhushan. Adaptive dissimilarity index for measuring time series proximity. *Advances in Data Analysis Classification*, 1:5–21, 2007.

[5] P. Clifford, S. Richardson and D. Hémon. Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45:123–134, 1989.

[6] P. Dionne. Comparaison de modeÌles numeìriques d'altitude de zones forestieÌres produits par altimeìtrie laser et interpolation de courbes de niveau. *MSc, Universite Laval, Canada*, 2003.

[7] P. Dutilleul. Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, 49:305–314, 1993.

[8] P. Goovaerts. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemica soil properties. *Biology and Fertility of soils*, 50:868–875, 1998.

[9] A. R. Kamat. Modified mean square successive difference with an exact distribution. *Sankya*, 15:295–302, 2008.

[10] S. Lee. Developing a bivariate spatial association measure: An integration of pearson's r and moran's i. *Journal of Geographical Systems*, 4:369–385, 2001.

[11] G. Matheron. *Les variables régionalisées et leur estimation.* Masson, 1965.

[12] S. Richardson and P. Clifford. Testing association between spatial processes. in: Possolo, a., (eds.), spatial statistics and imaging. *IMS Lecture Notes*, 20:295–308, 1991.

[13] A. Rukhin and R. Vallejos. Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters*, 78:1290–1300, 2008.

[14] M. Sherman. Variance estimation for statistics computed from spatial lattice data. *Journal of the Royal Statistical Society B*, 58:509–523, 1996.

[15] D. Tjøstheim. A measure of association for spatial variables. *Biometrika*, 65:109–114, 1978.

[16] R. Vallejos. Assessing the association between two spatial or temporal sequences. *Journal of Applied Statistics*, 35:1323–1343, 2008.

[17] C. Villagrán. El cuaternario de la región de los lagos del sur de chile. guia de excursión. *Taller Internacional El Cuaternario de Chile, Santiago*, 1993.