The codispersion map: a graphical tool to visualize the association between two spatial variables

*Ronny Vallejos, Felipe Osorio and Diego Mancilla

Departamento de Matemática, Universidad Técnica Federico Santa María
Avenida España 1680, Valparaíso, Chile

Abstract

The codispersion coefficient quantifies the association between two spatial processes for a particular direction (spatial lag) on the two-dimensional space. When this coefficient is computed for many directions, it is useful to display those values on a single graph. In this article, we suggest a graphical tool called a codispersion map to visualize the spatial correlation between two sequences on a plane. We describe how to construct a codispersion map for regular and non-regular lattices, providing algorithms in both cases. Three numerical examples are given to illustrate how useful this map can be to detect those directions for which the codispersion coefficient attains its maximum and minimum values. We also provide the R code to construct the codispersion map in practice.

Keywords and Phrases: spatial process, spatial association, codispersion coefficient, semi-variogram estimation.

1 Introduction

Computational and graphical tools recently developed in spatial statistics help to visualize spatial patterns, correlations and several other trends on a low dimensional space. For example Genton and Ruiz-Gasen (2010) introduce the hair-plot to visualize influential observations in dependent data, while Kahle and Wickham (2013) developed a tool called `ggmap` which combines the spatial information of static maps from Google Maps, OpenStreetMap, Stamen Maps or CloudMade Maps with the layered grammar of graphics implementation of `ggplot2`. On the other hand, the increasing number of packages and software (Anselin, 2006) that allow one to perform spatial analysis enhance the number of researchers getting involve in different aspects of the spatial modelling.

The assessment of the correlation between two spatial processes has been tackled from two different perspectives. First, the problem is assessed from an hypothesis testing approach, mainly transforming the *t* test in a suitable way to include the spatial information available for each process (Clifford et al, 1989; Dutilleul, 1993). Recently a computational method based on permutations and smoothing of the original variables has been suggested in the context of biodiversity (Viladomat et al, 2014). The second way to assess the association between two spatial processes is considering coefficients of spatial association which have been increasingly used in several applied

[*] Corresponding author: R. Vallejos, email: `ronny.vallejos@usm.cl`

areas, such as hydrology and soil sciences (Goovaerts, 1997; Pringe and Lark, 2006; Córdoba et al, 2013). In particular, the codispersion coefficient, first introduced by Matheron (1965), has been receiving attention in recent years from researchers because it allows the quantification of the existing spatial association between two processes in a particular direction (Ojeda et al, 2012). Besides the codispersion coefficient, other coefficients of association and hypothesis testing procedures have been implemented computationally (Osorio and Vallejos, 2014) which in practice facilitates the analysis of correlation between two real spatial sequences. From a theoretical perspective, some subjects have been studied. For example, under precise conditions, Rukhin and Vallejos (2008) found the limiting distribution of the sample codispersion coefficient for spatial autoregressive processes. Some extensions of these results were studied by Vallejos (2008) and Vallejos (2012) in a time series context. Cuevas et al. (2013) proposed a Nadaraya-Watson type estimator of the codispersion to assess the spatial association between two processes from a nonparametric point of view. Asymptotic expressions for the bias and variance of these estimations were established, as well as consistency. Bandwidth rules for the semi-variogram and cross-variogram were also provided. In image processing Vallejos (2012) and Ojeda et al (2012) explored the codispersion coefficient as a measure of similarity between digital images. Recently, Vallejos et al (2014) extended the codispersion coefficient for a multivariate process. Asymptotic properties were established and an image classification method for multispectral images was proposed.

In this article the problem of how to visualize the codispersion coefficient when several spatial lags are considered is addressed. Based on the good performance of the codispersion coefficient to capture the hidden correlation in specific directions we suggest a graphical display called codispersion map. Similar type of maps have been suggested for the semi-variogram and discussed extensively by Isaaks and Srivastava (1989). A cross-variogram map has been developed and implemented in the R package `gstat` (Pebesma, 2004). Our proposal is described for regular and non-regular lattices. In both cases the algorithms used to construct the codispersion map are outlined. The computation and visualization of the map was implemented in the R language. The R code associated with the suggested algorithms is also provided. Three examples with real data are described. The first one deals with the Murray smelter site dataset in an industrially contaminated area in Utah, USA. The second example deals with a forest dataset collected in the south of Chile. The third example illustrates the capability of the codispersion coefficient to assess the similarity between digital images. Finally, some remarks and an outline of topics to be addressed in future research are given.

2 The Codispersion Coefficient

Let $X(\boldsymbol{s})$ and $Y(\boldsymbol{s})$ be two spatial processes defined for $\boldsymbol{s} \in S \subset \mathbb{R}^d$, $d \geq 2 \in \mathbb{N}$. Here, we restrict our study to those processes that are intrinsically stationary on the plane ($d = 2$) and for which the semi-variogram of $X(\cdot)$ can be defined as

$$\gamma_X(\boldsymbol{h}) = \frac{1}{2}\mathbb{E}[(X(\boldsymbol{s}) - X(\boldsymbol{t}))^2], \qquad \boldsymbol{h} = \boldsymbol{t} - \boldsymbol{s}$$

and similarly for $Y(\cdot)$. In geostatistics, the cross-variogram between the processes $X(\cdot)$ and $Y(\cdot)$ defined as

$$\gamma_{X,Y}(\boldsymbol{h}) = \mathbb{E}[(X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))]$$

has been widely used in several applications (Chilès and Delfiner, 1999) In particular, the cross-variogram helps to define a coefficient of spatial association in the direction $\boldsymbol{h} \in \mathbb{R}^2$, called the codispersion coefficient, which is given by

$$\rho_{X,Y}(\boldsymbol{h}) := \frac{\gamma_{X,Y}(\boldsymbol{h})}{\sqrt{\mathbb{E}[(X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))^2]\mathbb{E}[(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2]}}, \qquad \boldsymbol{s} \in D_{\boldsymbol{h}},$$

where $D_{\boldsymbol{h}} = \{\boldsymbol{s} \in D : \boldsymbol{s}+\boldsymbol{h} \in D\}$. Now, consider the process $Z(\boldsymbol{s}) = X(\boldsymbol{s}) + Y(\boldsymbol{s}), \boldsymbol{s} \in S$. We emphasize the existing relationship between the semi-variograms of $X(\cdot)$, $Y(\cdot)$ and $X(\cdot) + Y(\cdot)$ and the cross-variogram $\gamma_{X,Y}(\boldsymbol{h})$, noting that

$$\begin{aligned}\gamma_Z(\boldsymbol{h}) &= \frac{1}{2}\mathbb{E}[(Z(\boldsymbol{s}) - Z(\boldsymbol{s}+\boldsymbol{h}))^2] \\ &= \frac{1}{2}\mathbb{E}[(X(\boldsymbol{s}) + Y(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2] \\ &= \frac{1}{2}\left(\mathbb{E}[(X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))^2] + 2\mathbb{E}[(X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))]\right. \\ &\quad \left. + \mathbb{E}[(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2]\right) \\ &= \gamma_X(\boldsymbol{h}) + \gamma_{X,Y}(\boldsymbol{h}) + \gamma_Y(\boldsymbol{h}),\end{aligned}$$

then

$$\gamma_{X,Y}(\boldsymbol{h}) = \gamma_{X+Y}(\boldsymbol{h}) - \gamma_X(\boldsymbol{h}) - \gamma_Y(\boldsymbol{h}). \tag{1}$$

There are several ways to estimate the semi-variogram, and the method used in this work assumes isotropy. In the following, we suppose that $S_{loc} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subset S \subset \mathbb{R}^2$ is the set of locations on the plane for which the realization of the process is available. In practice, it is difficult to obtain pairs of the form $(\boldsymbol{s}_i, \boldsymbol{s}_j)$, such that $\|\boldsymbol{s}_i - \boldsymbol{s}_j\| = \|\boldsymbol{h}\|$ for an arbitrary $\boldsymbol{h}$. This consideration is easily surpassed considering the set of all pairs $(\boldsymbol{s}_i, \boldsymbol{s}_j)$ located in a certain region with a Euclidean distance between the upper and lower bounds. i.e., a restriction of the form $h_k^{(l)} < \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \leq h_k^{(u)}$ is imposed for $k = 1, \ldots, B$. Thus, we can precisely define the set N_k as

$$N_k = \{(\boldsymbol{s}_i, \boldsymbol{s}_j) : \|\boldsymbol{s}_i - \boldsymbol{s}_j\| \in]h_k^{(l)}, h_k^{(u)}]\}, \qquad k = 1, \ldots, B. \tag{2}$$

In practice, these estimations of the semi-variogram are obtained as fixed values of B. For instance, the `geoR` package (Ribeiro and Diggle, 2001) considers by default $B = 13$. To obtain the values that determine the sets N_k, it is necessary to obtain

$$d_{\max} = \max_{(\boldsymbol{s}_i, \boldsymbol{s}_j) \in S_{loc}} \{\|\boldsymbol{s}_i - \boldsymbol{s}_j\|\}.$$

Given that $d_{\max}$ divides the interval $]0, d_{\max}]$ into B sub-intervals, we can define the center of each sub-interval, h_k, as

$$h_k = \frac{k\, d_{\max}}{B}, \qquad k = 1, \ldots, B. \tag{3}$$

Then, the symmetric intervals with respect to the center h_k are the following:

$$]h_k^{(l)}, h_k^{(u)}] = \left]h_k - \frac{d_{\max}}{B}, h_k + \frac{d_{\max}}{B}\right], \qquad k = 1, 2, \ldots, B.$$

This is the standard procedure to compute the empirical semi-variogram $\widehat{\gamma}_X(h_k)$, as defined through

$$\widehat{\gamma}_X(h_k) = \frac{1}{2|N_k|} \sum_{(i,j)\in N_k} (X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2, \tag{4}$$

where h_k is as in (3), and N_k is as defined in (2). Similarly, we can estimate the codispersion coefficient by emulating the estimation of the semi-variogram. If the processes $X(\boldsymbol{s})$ and $Y(\boldsymbol{s})$ are defined over a set $S \subset \mathbb{R}^2$, $S_{loc} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\} \subset S$ is the set of locations on the space where both processes are observed, $\boldsymbol{h} = (h_1, h_2) \in \mathbb{R}^2$ and $S_{\boldsymbol{h}} = \{\boldsymbol{s} \in S_{loc} : \boldsymbol{s} + \boldsymbol{h} \in S_{loc}\}$, then the empirical codispersion coefficient is

$$\widehat{\rho}_{X,Y}(\boldsymbol{h}) := \frac{\displaystyle\sum_{s\in S_{\boldsymbol{h}}} (X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))(Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))}{\sqrt{\displaystyle\sum_{\boldsymbol{s}\in S_{\boldsymbol{h}}} (X(\boldsymbol{s}) - X(\boldsymbol{s}+\boldsymbol{h}))^2 \sum_{\boldsymbol{s}\in S_{\boldsymbol{h}}} (Y(\boldsymbol{s}) - Y(\boldsymbol{s}+\boldsymbol{h}))^2}}. \tag{5}$$

The estimator of the codispersion given in (5) can be computed for a general fixed spatial lag $\boldsymbol{h}$. This computation can be difficult if the number of points in $S_{\boldsymbol{h}}$ is small or if $S_{\boldsymbol{h}}$ is an empty set. We emphasize that the empirical estimator of the codispersion makes real sense when the processes are defined on a finite rectangular grid in the two-dimensional space that corresponds to the assessment of the similarities between two digital images Ojeda et al (2012). The similarities between the images provided by the codispersion coefficient are not necessarily concerned with the similarities between the shapes or textures present in both images, but rather, by their (possibly hidden) correlations.

For a general estimation of the codispersion coefficient, the same considerations as used for the semi-variogram can be assumed. Then, the omnidirectional estimation of the codispersion becomes

$$\widehat{\rho}_{X,Y}(h_k) := \frac{\displaystyle\sum_{(\boldsymbol{s}_i,\boldsymbol{s}_j)\in N_k} (X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))(Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_j))}{\sqrt{\displaystyle\sum_{(\boldsymbol{s}_i,\boldsymbol{s}_j)\in N_k} (X(\boldsymbol{s}_i) - X(\boldsymbol{s}_j))^2 \sum_{(\boldsymbol{s}_i,\boldsymbol{s}_j)\in N_k} (Y(\boldsymbol{s}_i) - Y(\boldsymbol{s}_i))^2}},$$

where h_k and N_k are as in (4). A third estimation of the codispersion can be formulated by considering the expression for the cross-variogram, as given by (1). Then, a plug-in estimator of the codispersion that depends only on the semi-variogram estimations is given by

$$\widehat{\rho}_{X,Y}(h_k) = \frac{\widehat{\gamma}_{X+Y}(h_k) - \widehat{\gamma}_X(h_k) - \widehat{\gamma}_Y(h_k)}{\sqrt{\widehat{\gamma}_X(h_k)\widehat{\gamma}_Y(h_k)}}. \tag{6}$$

In order to illustrate how in some cases the information provided by the codispersion coefficient is related to that of the linear correlation coefficient, Table 1 lists several cases where the codispersion can be written as a product between the correlation coefficient associated with the errors and a constant depending on the parameters of models. In all cases the codispersion coefficient was derived assuming that the error processes $\epsilon_1(\boldsymbol{s})$ and $\epsilon_2(\boldsymbol{s})$ are white noise sequences showing variances σ^2 and τ^2 respectively and the covariance structure given by

Models	Equations	$\rho_{X,Y}(1)$	Constants
AR(1)	$X(t) = \phi_1 X(t-1) + \epsilon_1(t)$ $Y(t) = \phi_2 Y(t-1) + \epsilon_2(t)$	$c_1 \cdot \rho$	$c_1 = \frac{(2-\phi_1-\phi_2)\sqrt{(1+\phi_1)(1+\phi_2)}}{2(1-\phi_1\phi_2)}$
MA(1)	$X(t) = \epsilon_1(t) + \theta_1\epsilon_1(t-1)$ $Y(t) = \epsilon_2(t) + \theta_2\epsilon_2(t-1)$	$c_2 \cdot \rho$	$c_2 = \frac{(2-\theta_1-\theta_2+2\theta_1\theta_2)}{2\sqrt{(1-\theta_1+\theta_1^2)(1-\theta_2+\theta_2^2)}}$
AR(2)	$X(t) = 2\phi_1 X(t-1) - \phi_1^2 X(t-2) + \epsilon_1(t)$ $Y(t) = 2\phi_2 Y(t-1) - \phi_2^2 Y(t-2) + \epsilon_2(t)$	$c_3 \cdot \rho$	$c_3 = \frac{(1-\phi_1^2)(1-\phi_2^2)\sqrt{(1-\phi_1^2)(1-\phi_2^2)}}{(1-\phi_1\phi_2)^3}$
		$\rho_{X,Y}(\boldsymbol{h})$	
MA(∞)	$X(t) = \sum_{j=0}^{\infty} \phi_j \epsilon_1(t-j)$ $Y(t) = \sum_{k=0}^{\infty} \psi_k \epsilon_2(t-k)$	$c_4 \cdot \rho$	$c_4 = \frac{\rho\sum_{j=0}^{\infty}(2\phi_j\psi_j - \phi_{j+h}\psi_j - \phi_j\psi_{j+h})}{2\sqrt{\sum_{j=0}^{\infty}(\phi_j^2-\phi_j\phi_{j+h})\sum_{j=0}^{\infty}(\psi_j^2-\psi_j\psi_{j+h})}}$
2D AR	$X(i,j) = \phi_1 X(i-1,j) + \phi_2 X(i,j-1) + \epsilon_1(i,j)$ $Y(i,j) = \psi_1 Y(i-1,j) + \psi_2 Y(i,j-1) + \epsilon_2(i,j)$	$c_5 \cdot \rho$	$c_5 = \frac{\overset{\boldsymbol{h}}{} 2D(\phi_1,\phi_2,\psi_1,\psi_2,0,0) - D(\phi_1,\phi_2,\psi_1,\psi_2,h_1,h_2)(\psi_1^{h_1}\psi_2^{h_2} + \phi_1^{h_1}\phi_2^{h_2})\big]}{2\sqrt{R}}$
		where $D(\phi_1,\phi_2,\psi_1,\psi_2,h_1,h_2)$ $= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} \frac{(k+l)!(l+h_2+k+h_1)!}{k!l!(l+h_2)!(k+h_1)!}(\phi_1\psi_1)^k(\phi_2\psi_2)^l$ and $R = [D(\phi_1,\phi_2,\phi_1,\phi_2,0,0) - \phi_1^{h_1}\phi_2^{h_2} D(\phi_1,\phi_2,\phi_1,\phi_2,h_1,h_2)]$ $\times[D(\psi_1,\psi_2,\psi_1,\psi_2,0,0) - \psi_1^{h_1}\psi_2^{h_2} D(\psi_1,\psi_2,\psi_1,\psi_2,h_1,h_2)]$	
3D AR	$X(i,j,k) = \phi_1 X(i-1,j,k) + \phi_2 X(i,j-1,k) - \phi_3 X(i,j,k-1) - \phi_1\phi_2 X(i-1,j-1,k)$ $-\phi_2\phi_3 X(i,j-1,k-1) - \phi_1\phi_3 X(i-1,j,k-1) + \phi_1\phi_2\phi_3 X(i-1,j-1,k-1) + \epsilon_1(i,j,k)$ $Y(i,j,k) = \xi_1 Y(i-1,j,k) + \xi_2 Y(i,j-1,k) + \xi_3 Y(i,j,k-1) - \xi_1\xi_2 Y(i-1,j-1,k)$ $-\xi_2\xi_3 Y(i,j-1,k-1) - \xi_1\xi_3 Y(i-1,j,k-1) + \xi_1\xi_2\xi_3 Y(i-1,j-1,k-1) + \epsilon_2(i,j,k)$	$c_6 \cdot \rho$	$c_6 = \frac{\sqrt{\prod_{i=1}^{3}\left[(1-\phi_i)^2(1-\xi_i)^2\right]}(2-\phi_1^{h_1}\phi_2^{h_2}\phi_3^{h_3} - \xi_1^{h_1}\xi_2^{h_2}\xi_3^{h_3})}{2(1-\phi_1\xi_1)(1-\phi_2\xi_2)(1-\phi_3\xi_3)\sqrt{(1-\phi_1^{h_1}\phi_2^{h_2}\phi_3^{h_3})}}$

Table 1: Codispersion coefficient for spatial and time series models. $\boldsymbol{h} = (h_1, h_2)$.

$$\operatorname{cov}[\epsilon_1(\boldsymbol{t}), \epsilon_2(\boldsymbol{s})] = \begin{cases} \rho\sigma\tau, & \boldsymbol{t} = \boldsymbol{s}, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $|\rho| \leq 1$. As a result, the codispersion coefficient which can be interpreted as a linear correlation coefficient between spatial increments of both processes can also be seen as an adapted (corrected) version of the correlation coefficient between the white noise sequences as is shown in Table 1. The coefficients $c_1, \dots, c_6$ incorporate the spatial or temporal association between the processes that is contained in the parameters of the models.

3 The Codispersion Map

Similar to the variogram map described by Isaaks and Srivastava (1989), this codispersion map is a set of estimations for the codispersion coefficient in many different directions on the plane. The main goal is to obtain a summary of the codispersion values for angles belonging to the interval $[0, \pi]$ and radius varying on the range $]0, d_{\max}]$, where d_{max} depends on the type of spatial data (defined on rectangular or general grids). In other words, plotting the codispersion map summarizes the information about the spatial association between two sequences in a radial way on the plane circumscribing the map in a semisphere of radius $d_{\max}$.

3.1 The Map for Spatial Data Defined on a General Lattice

The construction of the codispersion map for spatial data, as defined on a general grid, can be summarized by three steps. Step 1 consists of selecting the maximum distance $d_{\max}$. The second step involves the definition of the points h_k where the estimations will be evaluated; these points are chosen uniformly on the interval $]0, d_{\max}]$. The third step consists of choosing angles from the interval $[0, \pi]$ to fix the directions that will be used in the estimation. In practice, for each of the selected directions, the estimation of the semi-variograms are carried out for the distances h_k, as previously selected. The Algorithm 3.1 reflects the computational procedure to estimate the values of the codispersion and the plotting procedure to yield the codispersion map, as based on Equation (6). It should be stressed that, in Algorithm 3.1, the notation used for the estimation of the semi-variogram is $\widehat{\gamma}_X(h[k], \alpha[i])$, highlighting the dependency of the estimated semi-variogram on the angle. This is crucial to restrict the search of the points belonging to each N_k. The computational implementation of the semi-variograms relies upon the numerical computations provided by the R package `geoR` (Ribeiro and Diggle, 2001). Once the estimations in all defined directions are available, the algorithm fits a rectangular grid of the estimations, and bilinear interpolations are generated to fill that rectangular grid. The R code to generate such a codispersion map can be found in the Appendix.

3.2 The Map for Spatial Data Defined on a Regular Grid

In the case of a rectangular grid, the estimation of the codispersion coefficient is much more simple. The implemented estimation is based on Equation (5). The domain is the set of locations $\{1, \dots, n\} \times \{1, \dots, m\}$, where n and m, respectively, denote the numbers of rows and columns of the grid. The simplicity arises from the fact that the distances between two contiguous points are always the same. The possible directions considered are those that satisfy $h = (h_1, h_2)$, with $h_1 \in \{-m+1, \dots, m+1\}$, $h_2 \in \{0, \dots, n-1\}$. To ensure a sufficient number of differences in a particular direction, we imposed the restriction $|h_1|, h_2 \leq \min(n, m)/3$. The observations

Algorithm 3.1: Algorithm to construct the codispersion map for spatial data defined on non-rectangular grids.

```
input  : Two spatial sequences $X$ and $Y$ and the corresponding set of coordinates $S_{loc}$
output: A codispersion map between the variables $X$ and $Y$
begin
    $d_{\max}/2 \longleftarrow$ Half of the maximum distance among the locations
    $B \longleftarrow 13$
    $h \longleftarrow \{d_{max}/B, 2d_{max}/B, \ldots, d_{max}\}$
    $\alpha \longleftarrow \{0, 0.01, 0.02, \ldots, \pi\}$
    $circ = (xcirc, ycirc) \longleftarrow$ Cartesian coordinates for the pairs $(h[k], \alpha[i])$,
    $k = 1, \ldots, B, i = 1, \ldots, |\alpha|$
    $z \longleftarrow$ Array of length $|\alpha| \cdot B$, where $|\alpha|$ is the length of $\alpha$
    for $i = 1, 2, \ldots, |\alpha|$ do
        for $k = 1, 2, \ldots, B$ do
            $z[k + (i-1)B] \longleftarrow \widehat{\gamma}_{X+Y}(h[k], \alpha[i]) - \widehat{\gamma}_Y(h[k], \alpha[i]) - \widehat{\gamma}_X(h[k], \alpha[i])$
        end
    end
    return $plot(interpolate(xcirc, ycirc, z, n))$, display the interpolated cells in an
    inscribed circumference on a semicircle of diameter $n$
end
```

$X(\boldsymbol{s}_k)$ and $Y(\boldsymbol{s}_k)$, $k = 1, \ldots, nm$, are described in the matrix notation used for the images. i.e., $X(j,i)$ and $Y(j,i)$, $(j,i) \in \{1,\ldots,n\} \times \{1,\ldots,m\}$. Contrary to the usual convention, we used the index j for the row and i for the column, and the position $(1,1)$ is located at the top left corner of the grid. Given the observations of the processes $X(\cdot)$ and $Y(\cdot)$ on a finite rectangular grid of size $n \times m$, the observations using the new notation are $X(j,i)$ and $Y(j,i)$. Considering the direction $\boldsymbol{h} = (h_1, h_2)$ with $|h_1| \leq m-1$ y $0 \leq h_2 \leq n-1$, the estimator of the codispersion (5) can be written as follows:

$$\widehat{\rho}_{X,Y}(\boldsymbol{h}) = \frac{\displaystyle\sum_{(j,i)\in S_{\boldsymbol{h}}} (X(j,i) - X(j+h_1, i+h_2))(Y(j,i) - Y(j+h_1, i+h_2))}{\sqrt{\displaystyle\sum_{(j,i)\in S_{\boldsymbol{h}}} (X(j,i) - X(j+h_1, i+h_2))^2 \sum_{(j,i)\in S_{\boldsymbol{h}}} (Y(j,i) - Y(j+h_1, i+h_2))^2}}, \tag{8}$$

where $S_{\boldsymbol{h}} = \{(j,i) : \ (j+h_1, i+h_2) \in S_{loc}\}$. We recall that we are assuming a positive axis in the vertical downward direction. The Algorithm 3.2 computes the codispersion map in the fashion described above. From a practical point of view, the algorithm was written in C and can be run from R using the interface .Call. Thus, the R code is simply a wrapper that allows us to run the routines developed in C internally. The R and C routines and directions for compiling these files can be found on the website `http://spatialpack.mat.utfsm.cl/codispmap/`.

4 Numerical Examples

In this section, three examples are presented to illustrate the use of the codispersion map in practical applications for both spatial data defined in a rectangular grid and spatial data defined on a

Algorithm 3.2: Algorithm to construct the codispersion map for spatial data defined for rectangular grids

```
input  : Two matrices $X$ and $Y$
output: A codispersion map between the variables $X$ and $Y$
begin
    $n \longleftarrow$ Number of rows
    $m \longleftarrow$ Number of columns
    $h_{max} \longleftarrow$ Integer less or equal than one third of the minimum between $n$ and $m$
    $h \longleftarrow$ Set the directions to be consider in the map
    for $i \in \{1, \ldots, nrows(h)\}$ do
        $z[h[i,1], h[i,2]] \longleftarrow \widehat{\rho}_{X,Y}(h[i,1], h[i,2])$
    end
    return $plot(z)$
end
```

general lattice. The first two datasets are defined on non-rectangular grids, and the third example illustrates the similarity between images defined on a finite rectangular grid in the plane.

4.1 The Murray Smelter Site Dataset

The dataset consists of soil samples collected in and around the vacant industrially contaminated Murray smelter site (Utah, USA). This area was polluted by airborne emissions and the placement of waste slag from the smelting process. A total of 253 locations were included in the study, and soil samples were taken from each location. Each georeferenced sampling quantity is a pooled composite of four closely adjacent soil samples, for which the levels of the heavy metals, arsenic (As) and lead (Pb), were measured. A complete description of the Murray smelter site dataset can be found in Griffith (2002) and Griffith and Paelinck (2011). The attributes As and Pb for each location are shown in Figure 1.

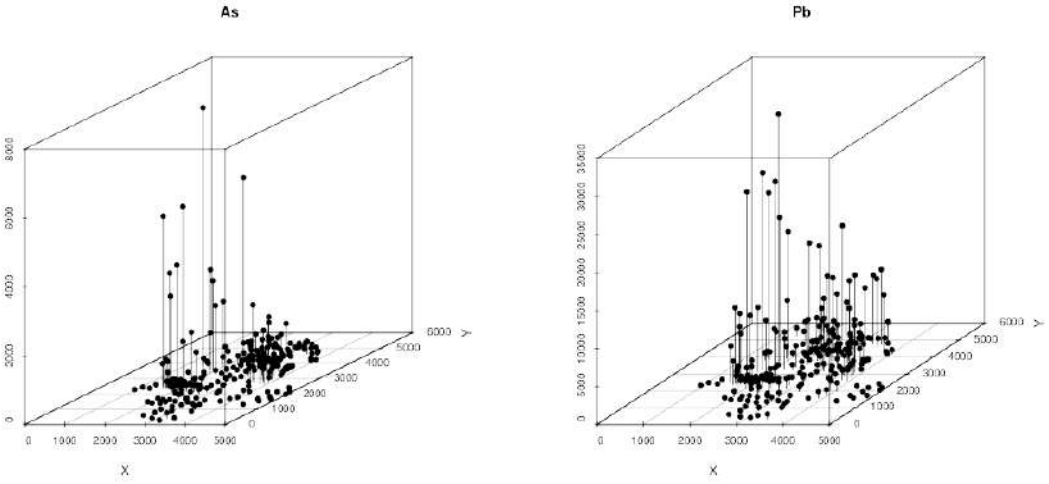


Figure 1: Locations and levels for As and Pb in the Murray smelter site dataset.

In Figure 2(a), we observe the values of the distances and directions for which the codispersion coefficient has been computed. Spatially, we can see the maximum and minimum values of the codispersion in Figure 2(b). The codispersion map for the circular grid considers distances up to 3000 meters. There is no clear pattern for any particular direction; however, the highest values of the codispersion are attained for the middle range distances. The correlation coefficient between As and Pb is $r = 0.5893$. Looking at the codispersion map, we observe a weak circular pattern of high values of the codispersion for distances ranging from 1000 to 2000 meters. This behavior can also be observed in the bubble charts displayed in Figure 3. Although we do not have an explanation for this, types of spatial associations we observe from a codispersion map are difficult to be obtained by the usual exploratory data analysis techniques.

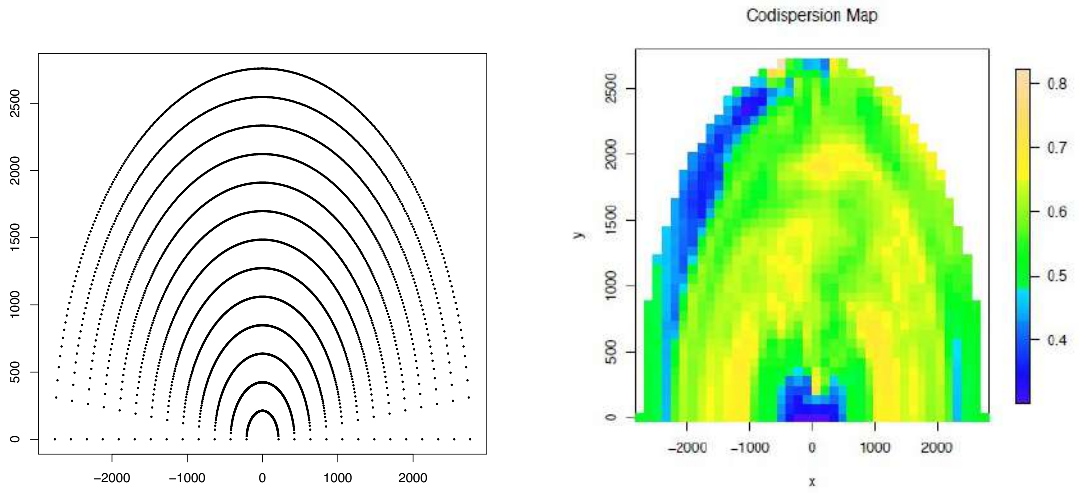


Figure 2: (a) Directions for which Algorithm 3.1 computes the values of the codispersion. (b) Codispersion map for the Murray smelter site dataset.

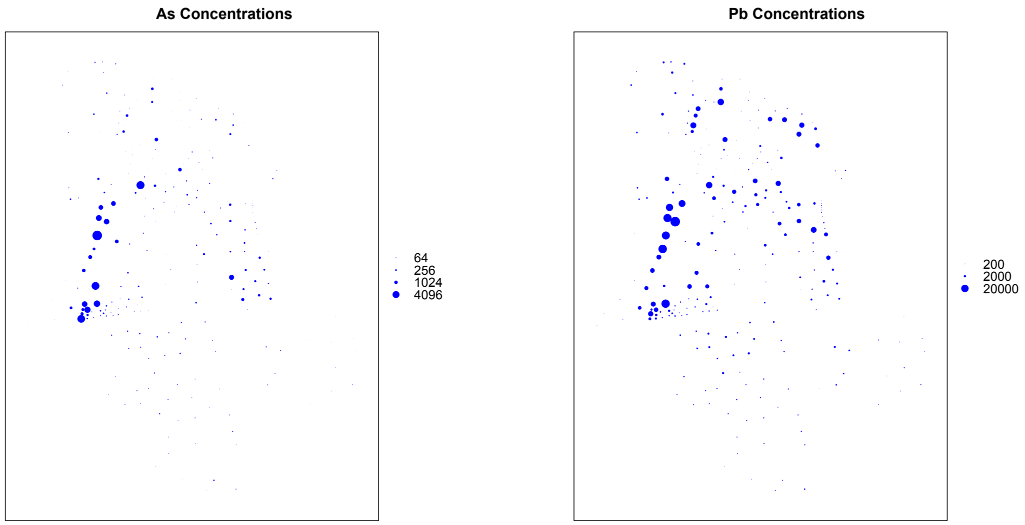
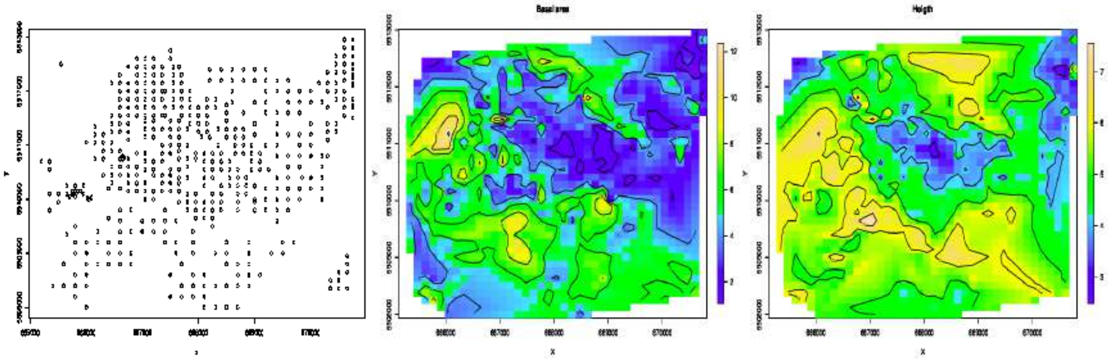


Figure 3: (a) Bubble chart for the levels of As (b) Bubble chart for the levels of Pb.

4.2 Forest Data

Pinus radiata is one of the most widely planted species in Chile and can thrive 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. The study site was 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 hectares. In addition to the more mature stands, we were also interested in the area containing younger (i.e., four years old) stands of Pinus radiata. This area had an average density of 1600 trees per hectare. The basal area and dominant tree height at the years of the plantation's establishment (1993, 1994, 1995, and 1996) were used to represent the stand attributes. The two variables we employed were obtained from 200 m^2 circular sample plots and point-plant sample plots. For the latter type of sample, four quadrants were established around the sample point; the four closest trees in each quadrant (16 trees in total) were 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 (Figure 4(a)). Figure 4 shows a simple bilinear interpolation and the corresponding contours for the two variables. Algorithm 3.1 was used to generate the codispersion map between the tree basal area and the tree height. The resulting plot is shown in Figure 5. In general it can be seen that the association is high in all directions. In particular we observe the highest values (0.8) of the codispersion on the vertical direction (90°) from 2000 meters up. We also observe a circular pattern on the top part of the codispersion map for angles between 45 and 135 degrees. These codispersion values are aproximately 0.75.



(a) (b) (c)

Figure 4: (a) Locations; (b) Bilinear interpolation of the tree basal area; (c) Bilinear interpolation of the tree height.

Unexpectedly, the smallest values of the codispersion are associated with the smallest distances (blue values in the center of the graph). This can happen often for forest variables due to the topography of the terrain which could significantly affect the spatial association of points that are very close one to each other. Another possible reason for this is a poor estimation of the codispersion coefficient for small distances, which can happen for processes defined on regular grids.

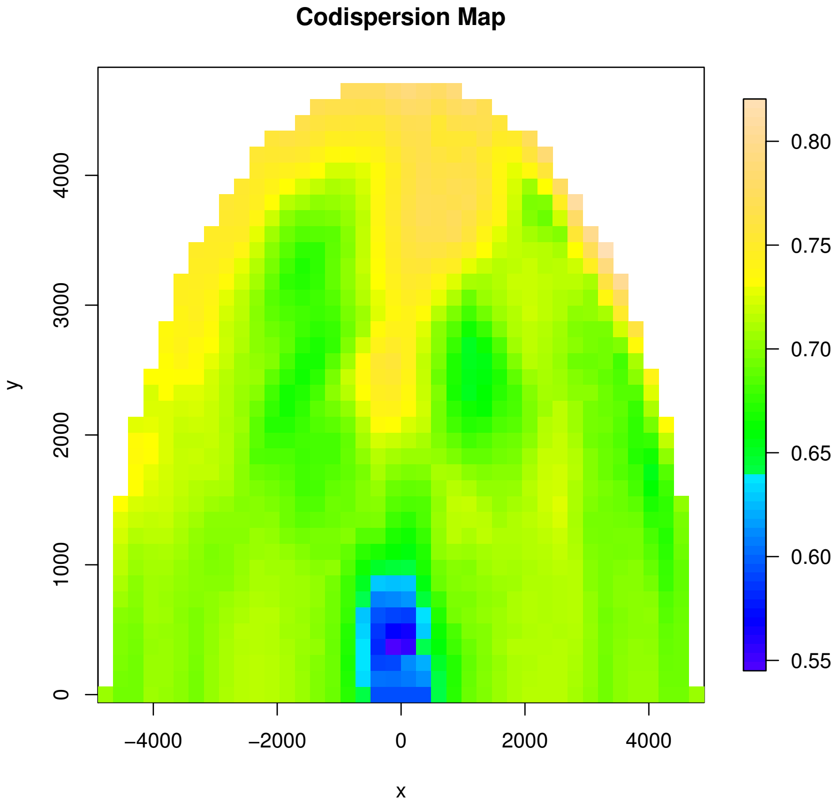


Figure 5: Codispersion map for the Pinus radiata dataset.

4.3 Image Similarity

In this example, we show the capacity of the codispersion coefficient to capture different levels of similarity between two images, considering different directions in the images. This subject was first studied by Ojeda et al (2012), who concluded that the codispersion coefficient was able to capture the hidden spatial correlations between two images in a particular direction, a feature not commonly assessed by other coefficients of similarity between images. This example also elucidates the difference between the classical association provided by the correlation coefficient and the spatial association provided by the codispersion coefficient. In order to accomplish this task, an original image (Lenna) of size 512×512 (Figure 6(a)) was taken from the USC-SIPI image database `http://sipi.usc.edu/database/`. This image was processed by the Algorithm 4.1 described below, which transforms the original image in one image with a clear pattern in the direction $(1, 1)$. The image produced by the Algorithm 4.1 is displayed in Figure 6(b). It should be stressed that the correlation coefficient between the images shown in Figure 6(a) and (b) is $r_{I,W} = 0.6909202$, whereas the codispersion coefficient for $\boldsymbol{h} = (1, 1)$ is $\widehat{\rho}_{I,W}(1, 1) = 1$. These results are not surprising because a detailed inspection of the pseudo-code displayed in Algorithm 4.1 shows that the differences needed for the computation of the codispersion coefficient take the form $W[i, j] - W[i + h_1, j + h_2]$, but $W[i + h_1, j + h_2] = W[i, j] - \mathcal{D}$, where $\mathcal{D}$ is the difference associated with pixel (i, j) in the original image. i.e., $\mathcal{D} = I[i, j] - I[i + h_1, j + h_2]$. Thus, in the original and transformed images, the sums of these differences are exactly the same, producing a codispersion coefficient equal to 1. The introduction of the variable W (line 5 in Algorithm 4.1) as a normal random variable is arbitrary and generates one type of transformed image into the direction $(1, 1)$ for the purposes of this illustration. Using different distributions for W will generate other types of transformed images. In addition, salt and pepper contamination noise was added to the original image shown in Figure 6(a) with percentages of contamination of 1%, 5%, 10%, and 25%, respectively. Then Algorithm 4.1 was applied to these blurred images producing the images shown in Figure 6(c)-(f). The transformation of the original image without contamination is displayed in Figure 6(b). Visually it is possible to observe a degradation of the original patterns as effect of the percentage of contamination. Moreover, the sample correlation coefficients between the original and contaminated images are 0.684555, 0.645544, 0.595579, and 0.517395, respectively, exhibiting a clear decreasing pattern as contamination increases.

Algorithm 4.1: Algorithm to transform an image into the direction $\boldsymbol{h} = (1, 1)$

```
   input : An image $I$ of size $n \times m$
   output: An image $W$ of size $n \times m$
1  for $i = 1, 2, \ldots, n$ do
2      for $j = 1, 2, \ldots, m$ do
3          if $(i == 1)$ or $(j == 1)$ then
4              $W[i, j] \longleftarrow$ Simulation of a standard normal random variable;
5          end
6          else
7              $W[i, j] \longleftarrow W[i-1, j-1] - I[i-1, j-1] + I[i, j]$
8          end
9      end
10     $W \longleftarrow \frac{W - \min W}{\max W - \min W}$;
11     return $W$;
12 end
```

In this case, the codispersion map yielded by the application of Algorithm 3.2 to the images shown in Figure 6 produces valuable information highlighting the directional features of the codispersion coefficient that cannot be accounted for by other coefficients. This occurrence can clearly be seen in the codispersion map displayed in Figure 7(a), where there is an evident pattern associated with the perfect spatial association in the 45° line, corresponding to the transformation of the original image (Figure 6(a)) in the direction $\boldsymbol{h} = (1, 1)$, as expected. In Figure 7(b)-(e) are displayed the codispersion maps between the original and contaminated images. It is evident that the existing spatial association in the direction (1,1) between the original and transformed images is corrupted by the effect of the contamination.

5 Conclusion

In this paper, we have introduced a new graphical tool called the codispersion map to visualize the spatial associations between two spatial processes on a plane. Algorithms for the construction of such a map were discussed for processes defined on rectangular and general grids. The R code in both cases has also been provided. The foundations of these computations are based on the evaluation of the codispersion coefficients at specific points on a suitable grid. In light of the real

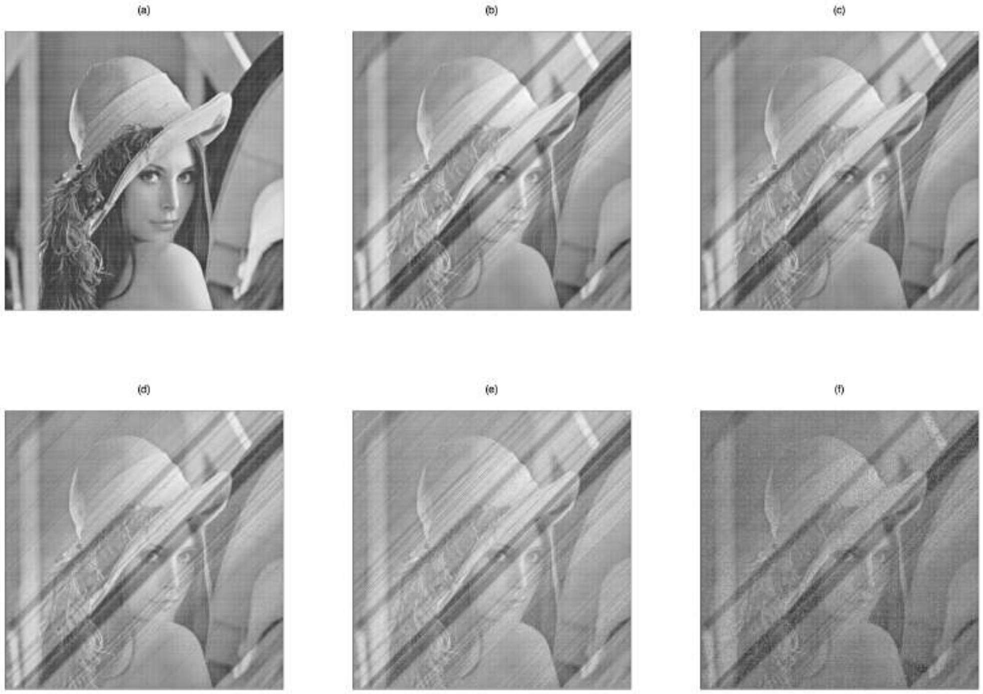


Figure 6: (a) Original image (Lenna); (b) Transformed image into the direction $\boldsymbol{h} = (1, 1)$; (c) Image (a) with 1% of salt and pepper noise transformed by Algorithm 4.1; (d) Image (a) with 5% of salt and pepper noise transformed by Algorithm 4.1; (e) Image (a) with 10% of salt and pepper noise transformed by Algorithm 4.1; (f) Image (a) with 25% of salt and pepper noise transformed by Algorithm 4.1.

data examples shown in Section 4, we conclude that the codispersion map is able to account for the hidden spatial associations between two processes for specific directions. This feature highlights the codispersion coefficient and its graphical developments as useful tools to be used in future practical applications.

We view the work described in this paper as only the beginning of a large project with several related problems to be tackled in the future. The study of a measure of similarity between images based on the codispersion coefficient seems to be an interesting problem to be explored in further research, drawing from Brunet et al (2012) and Wang et al (2004). The construction of a panel of codispersion maps comparing the associations among more than two spatial processes can be constructed based on Vallejos et al (2014). This codispersion map will be included in future versions of the R package SpatialPack.

Acknowledgements

Ronny Vallejos was partially supported by FONDECYT grant 1120048, Chile and AC3E grant FB-0008, UTFSM. Felipe Osorio was supported by CONICYT grant 791100007 and USM grant 12.14.14. The authors thank Jonathan Acosta for reading an early version of this manuscript.

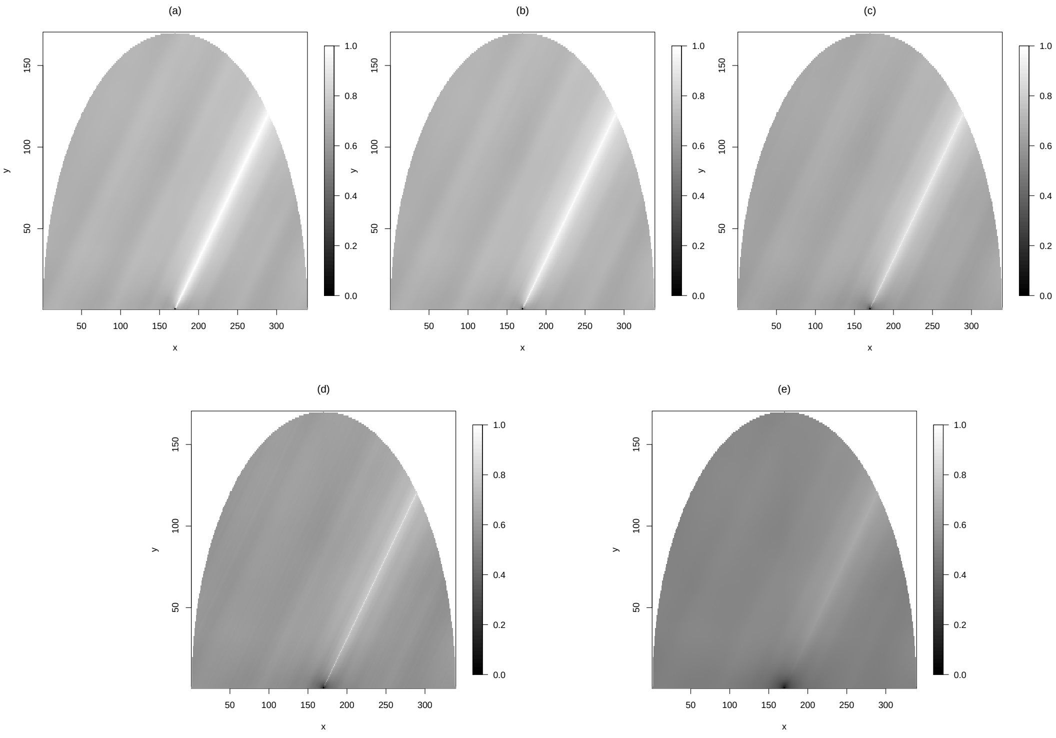


Figure 7: Codispersion maps between the pairs of images in Figure 6. (a) Codispersion map between images (a) and (b); (b) Codispersion map between images (a) and (c); (c) Codispersion map between images in (a) and (d); (d) Codispersion map between images (a) and (e); (e) Codispersion map between images (a) and (f).

A Appendix

R Code to Generate the Codispersion Map for Nonrectangular Grids

```
codisp.map <-
function(x, y, coords, nclass = 13, ncell = 40, plot.it = TRUE)
{
    require("akima")
    require("fields")
    require("geoR")
    rho <- function(x, y, uvec, max.dist, angle)
    {
       z <- as.geodata(cbind(x$coords, x$data + y$data))
       nz <- variog(z, uvec = uvec, max.dist = max.dist,
          direction = angle, messages = FALSE)
       dx <- variog(x, uvec = uvec, max.dist = max.dist,
          direction = angle, messages = FALSE)
       dy <- variog(y, uvec = uvec, max.dist = max.dist,
          direction = angle, messages = FALSE)
       rho <- .5 * (nz$v - dx$v - dy$v) / sqrt(dx$v * dy$v)
    }

    x <- as.geodata(cbind(coords, x))
    y <- as.geodata(cbind(coords, y))
    dmax <- .5 * max(dist(coords))
    angles <- seq(from = 0, to = pi, by = 0.01)
    nangles <- length(angles)
    uvec <- seq(from = 0, to = dmax, length = nclass + 1)[-1]

    xcirc <- 0
    ycirc <- 0

    for (i in seq_len(nclass)) {
       xcirc[(nangles*(i-1)+1):(nangles*i)] <- seq(-uvec[i],
          uvec[i], length = nangles)
       ycirc[(nangles*(i-1)+1):(nangles*i)] <- sqrt(uvec[i]^2 -
          xcirc[(nangles*(i-1)+1):(nangles*i)]^2)
    }
    z <- matrix(0, nrow = nangles, ncol = nclass)
    for (i in seq_len(nangles))
       z[i,] <- rho(x, y, uvec = uvec, max.dist = dmax, angle = angles[i])
    z <- as.vector(z)
    xl <- seq(min(xcirc), max(ycirc), length=ncell)
    yl <- seq(min(ycirc), max(ycirc), length=ncell)

    if (plot.it) {
       par(pty = "s")
       image.plot(interp(xcirc, ycirc, as.vector(z), xo = xl,yo = yl),
       col = topo.colors(256), xlab = "x", ylab = "y")
       title(main = "Codispersion Map")
    }

    invisible(list(xcirc = xcirc, ycirc = ycirc, z = z))
}
```

References

ANSELIN, L., I. SIABRI and Y. KHO (2006), GeoDa: An Introduction to Spatial Data Analysis, *Geographical Analysis* **38**, 5-22.

BRUNET, D., E. R. VRSCAY and Z. WANG, Z (2012), On the mathematical properties of the structural similarity index, *IEEE Transactions on Image Processing* **21**, 1488-1498.

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

CLIFFORD, P., S. RICHARDSON and D. HÉMON (1989), Assessing the significance of the correlation between two spatial processes, *Biometrics* **45**, 123-134.

CÓRDOBA, M., C. BRUNO., J. COSTA and M. BALZARINI (2013), Subfield management class delineation using cluster analysis from spatial principal components of soil variables, *Computers and Electronics Agriculture* **97**, 6-14.

CUEVAS, F., E. PORCU and R. VALLEJOS (2013), Study of Spatial Relationships Between Two Sets of Variables: A Nonparametric Approach, *Journal of Nonparametric Statistics* **25**, 695-714.

DUTILLEUL, P. (1993), Modifying the t test for assessing the correlation between two spatial processes, *Biometrics* **49**, 305-314.

GENTON, M and A. RUIZ-GAZEN (2010), Visualizing Influential Observations in Dependent Data, *Journal of Computational and Graphical Statistics* **19**, 808-825.

GOOVAERTS, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Oxford.

GRIFFITH, D. (2002), The geographic distribution of soil-lead concentration: description and concerns, *URISA Journal* **14**, 5-15.

GRIFFITH, D. and J. H. P. PAELINCK (2011), *Non-Stardard Spatial Statistics*, Springer, New York.

ISAAKS, E. H. and R. M. SRIVASTAVA (1989), *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

KHALE, D. and H. WICKHAM (2013), ggmap: Spatial visualization with ggplot2, *The R Journal* **5**, 144-161.

MATHERON, C. (1965), *Les Variables Régionalisées et leur Estimation*, Masson, Paris.

OJEDA, S., R. VALLEJOS and P. LAMBERTI (2012), Measure of similarity between images based on the codispersion coefficient, *Journal of Electronic Imaging* **21**, 023019.

OSORIO, F., R. and R. VALLEJOS (2014), SpatialPack: A package to assess the association between two spatial processes. R package version 0.2-3, URL:`http://cran.r-project.org/package=SpatialPack`.

PEBESMA, E. J., (2004), Multivariable geostatistics in S: the gstat package, *Computers & Geosciences* **30**, 683-691.

PRINGLE, M. J. and R. M. LARK (2006), Spatial analysis of model error, illustrated by soil carbon dioxide emissions, *Vadose Zone Journal* **5**:168-183.

RIBEIRO Jr. P. J. and P. J. DIGGLE (2001), geoR: a package for geostatistical analysis, *R-NEWS* **1**, 15-18.

RUKHIN, A. and R. VALLEJOS (2008), Codispersion coefficient for spatial and temporal series, *Statistics and Probability Letters* **78**, 1290-1300.

VALLEJOS, R. (2008), Assessing the association between two spatial or temporal sequences, *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.

VALLEJOS, R., A. MALLEA, M. HERRERA and S. M. OJEDA (2014), A multivariate geostatistical approach for landscape classification from remotely sensed image data, *Stochastic Environmental Research and Risk Assessment* (in press).

VILADOMAT, J. R. MAZUMDER, A. McINTURF, D. J. McCAULEY and T. HASTIE (2014) Assessing the significance of global and local correlations under spatial autocorrelation: A nonparametric approach, *Biometrics* (in press) DOI:10.1111/biom.12139.

WANG, Z., A. BOVIK, H. R. SHEIKH and E. P. SIMONCELLI (2004), Image quality assessment: from error visibility to structural similarity, *IEEE Transactions on Image Processing* **13**, 1-14.