We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter 2

Image Segmentation and Time Series Clustering Based on Spatial and Temporal ARMA Processes

Ronny Vallejos and Silvia Ojeda

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/50513

1. Introduction

During the past decades, image segmentation and edge detection have been two important and challenging topics. The main idea is to produce a partition of an image such that each category or region is homogeneous with respect to some measures. The processed image can be useful for posterior image processing treatments.

Spatial autoregressive moving average (ARMA) processes have been extensively used in several applications in image/signal processing. In particular, these models have been used for image segmentation, edge detection and image filtering. Image restoration algorithms based on robust estimation of a two-dimensional process have been developed (Kashyap & Eom 1988). Also the two-dimensional autoregressive model has been used to perform unsupervised texture segmentation (Cariou & Chehdi, 2008). Generalizations of the previous algorithms using the generalized M estimators to deal with the effect caused by additive contamination was also addressed (Allende et al., 2001). Later on, robust autocovariance (RA) estimators for two dimensional autoregresive (AR-2D) processes were introduced (Ojeda, 2002). Several theoretical contributions have been suggested in the literature, including the asymptotic properties of a nearly unstable sequence of stationary spatial autoregressive processes (Baran et al., 2004). Other contributions and applications of spatial ARMA processes have been considered in many publications (Basu & Reinsel, 1993, Bustos 2009a, Francos & Friendlaner1998, Guyon 1982, Ho 2011, Illig & Truong-Van 2006, Martin1996, Vallejos & Mardesic 2004).

A new approach to perform image segmentation based on the estimation of AR-2D processes has been recently suggested (Ojeda 2010). First an image is locally modeled using a spatial autoregressive model for the image intensity. Then the residual autoregressive image is computed. This resulting image possesses interesting texture features. The borders and edges

INTECH
open science | open minds

© 2012 Vallejos and Ojeda; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

are highlighted, suggesting that the algorithm can be used for border detection. Experimental results with real images clarify how the algorithm works in practice. A robust version of the algorithm was also proposed, to be used when the original image is contaminated with additive outliers. Applications in the context of image inpainting were also offered.

Another concern that has been pointed out in the context of spatial statistics is the development of coefficients to compare two spatial processes. Coefficients that take into account the spatial association between two processes have been proposed in the literature. (Tjostheim, 1978) suggested a nonparametric coefficient to assess the spatial association between two spatial variables. Later on, (Clifford et al. 1989) proposed an hypothesis testing procedure to study the spatial dependence between two spatial sequences. Rukhin & Vallejos (2008) studied asymptotic properties of the codispersion coefficient first introduced by Matheron(1965). The performance and impact of this coefficient to quantify the spatial association between two images is currently under study Ojeda et al. (2012). An adaptation of this coefficient to time series analysis was studied in Vallejos (2008).

In the context of clustering time series Chouakria & Nagabhushan (2007) proposed a distance measure that is a function of the codispersion coefficient. This measure includes the correlation behavior and the proximity of two time series. They proposed to combine these distances in a multiplicative way, introducing a tuning constant controlling the weight of each quantity in the final product. This makes the measure flexible to model sequences with different behaviors, comparing them in terms of both correlation and dissimilarity between the values of the series.

The structure of this chapter consist in two parts. In the first part we review some theoretical aspects of the spatial ARMA processes. Then the algorithm suggested by Ojeda(2010), its limitations and advantages are briefly described. In order to propose a more efficient algorithm new variants of this algorithm are suggested specially to address the problem of determining the most convenient (in terms of the quality of the segmentation) prediction window of unilateral AR-2D processes. The computation of the distance between the filtered images and the original one will be done by using the codispersion coefficient and other image quality measures (Wang and Bovik 2002). Examples with real images will highlight the features of the modified algorithm. In the second part, the codispersion coefficient previously used to measure the closeness between images is utilized in a distance measure to perform cluster analysis of time series. The distance measure introduced in Chouakria & Nagabhushan (2007) is generalized in the sense that considers an arbitrary lag h that allows us to capture a higher serial correlation of two temporal or spatial sequences. Examples and numerical studies are presented to explore our proposal in several different scenarios. We explore the performance of hierarchical methods to classify correlated sequences when the proposed proximity measure is used, employing the Monte Carlo simulation. An application is discussed for time series measuring the Normalized Difference Vegetation Index (NDVI) in four locations of Argentina. The clusters formed using hierarchical classification techniques with the proposed distance measure preserve the geographical location where the series were obtained, providing information that is unavailable when using hierarchical methods with conventional distance measures.

http://dx.doi.org/10.5772/50513

2. Image Segmentation Through Estimation of Spatial ARMA Processes

2.1. The Spatial ARMA Processes

Spatial ARMA processes have been studied in the context of random fields indexed over $\mathbb{Z}^d$, $d \geq 2$, where $\mathbb{Z}^d$ is endowed with the usual partial order. That is, for $s=(s_1, s_2, \ldots, s_d)$, $u=(u_1, u_2, \ldots, u_d)$ in $\mathbb{Z}^d$, $s \leq u$ if for $i=1, 2, \ldots, d$, $s_i \leq u_i$ For $a, b \in \mathbb{Z}^d$, such that $a \leq b$ and $a \neq b$, we define $S[a, b]=\{x \in \mathbb{Z}^d \mid a \leq x \leq b\}$ and $S\langle a, b]=S[a, b] \setminus \{a\}$.

A random field $(X_s)_{s \in \mathbb{Z}^d}$ is said to be a spatial ARMA(p, q) with parameters $p, q \in \mathbb{Z}^d$ if it is weakly stationary and satisfies the equation

$$X_s - \sum_{j \in S\langle 0, p]} \phi_j X_{s-j} = \varepsilon_t + \sum_{k \in S\langle 0, q]} \theta_j \varepsilon_{s-j}, \tag{1}$$

where $(\phi_j)_{j \in S\langle 0, p]}$ and $(\varepsilon_j)_{k \in S\langle 0, q]}$, respectively, denote the autoregressive and moving average parameters with $\phi_0=\theta_0=1$, and $(\varepsilon_s)_{s \in \mathbb{Z}^d}$ denotes a sequence of independent and identically distributed centered random variables with variance σ^2. Notice that if $p=0$, the sum over $S\langle 0, p]$ is supposed to be zero, and the process is called a spatial moving average MA(q) random field. Similarly, if $q=0$, the process is called a spatial autoregressive AR(p) random field. The ARMA random field is labeled as causal if it has the following unilateral representation.

$$X_s = \sum_{[j \in S\ 0, \infty]} \psi_j \varepsilon_{s-j}$$

with $\sum_j \left| \phi_j \right| < \infty$. Similar to the time series case, there are conditions on the (AR or MA) polynomials that ensure stationarity and invertibility, respectively. Let $\Phi(z)=1-\sum_{j \in S\langle 0, p]} \phi_j z^j$ and $\Theta(z)=1-\sum_{j \in S\langle 0, q]} \theta_j z^j$, where $z=(z_1, z_2, \ldots, z_d)$ and $z^j = z_1^{j_1} z_2^{j_2} \ldots z_d^{j_d}$. A sufficient condition for the random field to be causal is that the AR polynomial $\Phi(z)$ has no zeros in the closure of the open disc D^d in $\mathbb{C}^d$. For example, if $d=2$, the process is causal if $\Phi(z_1, z_2)$ is not zero for any z_1 and z_2 that simultaneously satisfy $|z_1|<1$ and $|z_2|<1$ (Jain et al., 1999).

Applications of spatial ARMA processes have been developed, including analysis of yield trials in the context of incomplete block designs (Cullis & Glesson 1991, Grondona et al. 1996) and the study of spatial unilateral first-order ARMA model (Basu & Reinsel, 1993). Other theoretical extensions of time series and spatial ARMA models can be found in (Baran et al., 2004, Bustos et al., 2009b, Gaetan & Guyon 2010, Choi 2000, Genton & Koul 2008, Guo 1998, Vallejos and Garccía-Donato 2006).

2.2. An Image Segmentation Algorithm

In this section, we describe an image segmentation algorithm that is based on a previous fitting of spatial autoregressive models to an image. This fitted image is constructed by dividing the original image into squared sub-images (e.g.,8×8) and then fitting a spatial autoregressive model to each sub-image (i.e., block). Then, we generate a sub-image from each local fitted model, preserving intensities on the boundary to smooth the edges between blocks. The final fitted image is yielded by putting together all generated sub-images.

Let$Z=Z_{m,n}, 0\le m\le M-1$,$0\le n\le N-1$, be the original image, and let$X=X_{m,n}$, $0\le m\le M-1$, $0\le n\le N-1$, where for all$0\le m\le M-1, 0\le n\le N-1$,$X_{m,n}=Z_{m,n}-\overline{Z}$, and $\overline{Z}$is the mean of$Z$. Let $4\le k\le \min(M, N)$and consider the rearrange images

$Z=Z_{m,n}$,

$X=X_{m,n}$,

where$0\le m\le M'-1, 0\le n\le N'-1$, $M'=\left[\frac{M-1}{k-1}\right](k-1)+1, N'=\left[\frac{N-1}{k-1}\right](k-1)+1.$ For all $i_b=1, \cdots, \left[\frac{M-1}{k-1}\right]$and for all $j_b=1, \cdots, \left[\frac{N-1}{k-1}\right]$the $(k-1)\times(k-1)$block $(i_b, j_b)$of the image $X$is defined as

$B_X(i_b, j_b)=X_{r,s}$,

where $(k-1)(i_b-1)+1\le r\le (k-1)i_band(k-1)(j_b-1)+1\le s\le (k-1)j_b$. Then, the approximated image $\hat{X}$of $X$is provided by Algorithm 1.

Algorithm 1.

For each block $B_X(i_b, j_b)$

1. Compute estimators$\hat{\phi}_1(i_b, j_b)$, $\hat{\phi}_2(i_b, j_b)$of ϕ_1 and $\phi_2$corresponding to the block $B_X(i_b, j_b)$extended to:

$B_X'(i_b, j_b)=X_{r,s}$,

where$(k-1)(i_b-1)\le r\le (k-1)i_b, (k-1)(j_b-1)\le s\le (k-1)j_b$.

2. Let $\hat{X}$ be defined on the block $B_X(i_b, j_b)$by

$\hat{X}_{r,s}=\hat{\phi}_1(i_b, j_b)X_{r-1,s}+\hat{\phi}_2(i_b, j_b)X_{r,s-1}$

where $(k-1)(i_b-1)+1\le r\le (k-1)i_b$and $(k-1)(j_b-1)+1\le s\le (k-1)j_b$.

Then the approximated image $\hat{Z}$of the original image $Z$is:

$\hat{Z}_{m,n}=\hat{X}_{m,n}+\overline{Z}, \ 0\le m\le M'-1, 0\le n\le N'-1.$

The image segmentation algorithm we describe below is supported by a widely known notion in regression analysis. If a fitted image very well represents the patterns on the original image, then the residual image (i.e., the fitted image minus the observed image) will not contain useful information about the original patterns because the model already explains the features that are present in the original image. On the contrary, if the model does not

http://dx.doi.org/10.5772/50513

well represent the patterns that are present in the original image, then the residual image will contain useful information that has not been explained by the model. Thus, to implement an algorithm based on these notions, we must characterize which patterns are present in the residual image when the fitted image is not a good representation of the original one, and we must develop a technique to produce a fitting that is satisfactory in terms of segmentation but not a very good estimation in that the residual image still contains valuable information. (Ojeda et al. 2010) investigated these concerns and, based on several numerical experiments with images, determined that the residual image associated with a good local fitting is in fact poor in terms of structure (i.e., it is very similar to a white noise). However, when the fitted image is poor in terms of estimation, the residual image is useful for highlighting the boundaries and edges of the original image. Moreover, a bad fitting is related to the size of the block (or window) used in Algorithm 1. The best performance is attained for the maximum block size, which would be the size of the original image. The image segmentation algorithm introduced by (Ojeda et al. 2010) can be summarized as follows.

Algorithm 2.

1. Use Algorithm 1 to generate an approximated image $\hat{Z}$ of Z.

2. Compute the residual autoregressive image given by $Z - \hat{Z}$

Example 1. We present examples with real images to illustrate the performance of Algorithms 1 and 2. These images were taken from the database http://sipi.usc.edu/database. Figure 1(a) shows an original image of size512×512 (aerial), and Figure 1(b) shows the image generated by Algorithm 1 when a moving window of size 512×512is used to define an AR-2D process on the plane. It is not possible to visualize the differences between the original and fitted images. However, the residual image (Fig 1(c)) shows patterns that the model is not able to capture. Basically, the AR-2D model does not capture the changes in the texture produced by lines, borders and object boundaries. These features are contained in the residual image produced by Algorithm 2 such that the good performance of Algorithm 2 is associated with a moderate fitting of the AR-2D model. Another image (peepers) was processed by Algorithm 2 to show the effect of the size of the moving window. Figure 2(b) shows the segmentation produced by Algorithm 2 using a moving window of size 128×128. Another segmentation with a moving of size 512×512is shown in Figure 2. In both cases, the segmentations highlight the borders and boundaries present in the original image.

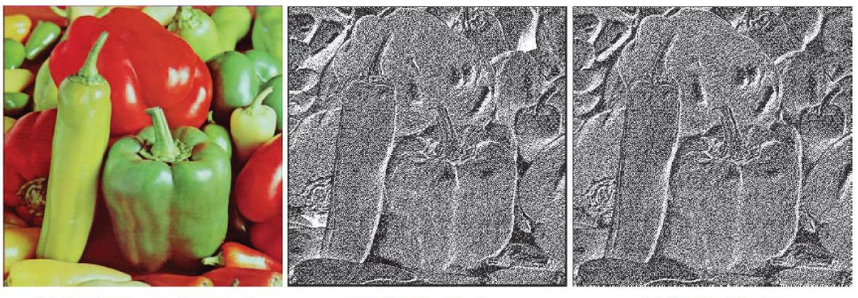
2.3. Improving the Segmentation Algorithm

In all experiments carried out in (Ojeda et al., 2010) and (Quintana et al., 2011), Algorithm 1 was implemented using the same prediction window for the AR-2D process, which contains only two elements belonging to a strongly causal region on the plane. Here, we consider other prediction windows to observe the effect on the performance of Algorithm 2. A description



(a) Original image (aerial) (b) Image generated by Algorithm 1 (c) Residual image generated by Algorithm 2

Figure 1. Images generated by Algorithms 1 and 2.



(a) Original image (peppers) (b) Residual image (c) Residual image

Figure 2. (b)-(c) Images generated by Algorithm 2 with prediction windows of 128×128and 512×512respectively.

of the most commonly used prediction windows in statistical image processing is in Bustos et al., (2009a). A brief description of the strongly causal prediction windows is given below.

For all $(m, n) \in \mathbb{Z}^2$, a strongly causal region at (m, n) is defined as

$$S(m, n) = \{(k, l) \in \mathbb{Z}^2 : k \leq m, l \leq n)\} - \{(m, n)\} \tag{2}$$

For a given $M \in \mathbb{N}$, a strongly causal prediction window is

$$W = \{(k, l) \in s(m, n) : m - M \leq k \leq m, n - M \leq l \leq n\}. \tag{3}$$

In particular, if$M$ =1, then a strongly causal prediction window containing three elements is

$$W_1 = \{(k, l) \in s(m, n) : m - 1 \leq k \leq m, n - 1 \leq l \leq n\} \tag{4}$$

The set W_1 is shown in Figure 3 (b). Similarly, strongly causal prediction windows can be defined by considering not only the top left quadrant on the plane $\mathbb{Z}^2$.The definition of such sets generates the prediction windows W_2, W_3, andW_4, as shown in Figure3(b). Algorithms

http://dx.doi.org/10.5772/50513

1 and 2 were implemented using the prediction windows W_1, W_2, W_3, and W_4, with two elements each (Figure 3(a)).

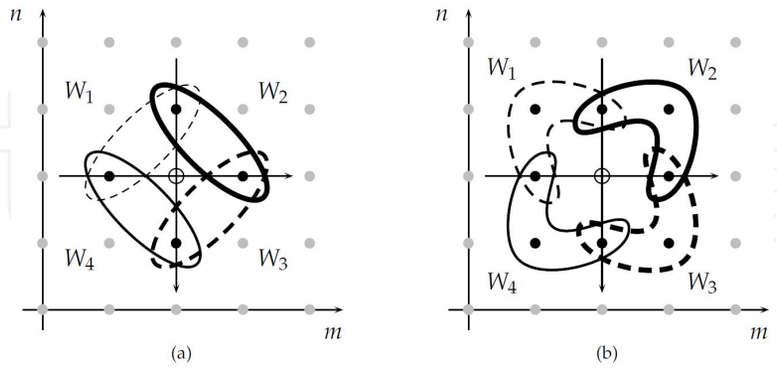


Figure 3. Strongly causal prediction windows.

Visually, the best segmentation for the aerial image is yielded by the prediction window W_2. The lines and edges are better highlighted in this segmentation (Figure 4(b)) than in the other segmentations. The dark regions are also stressed, which provides a more intense and brighter partition of the original features.

To gain insight on image quality measures, the fitted images produced by Algorithm 1 associated with the images shown in Figure 4(a) -(d) were compared aerially with the original image using three coefficients described in (Ojeda et al., 2012). These coefficients are briefly described below.

Consider two weakly stationary processes, $(X_s)_{s\in D}$ *and* $(Y_s)_{s\in D}$, $D\subset\mathbb{Z}^d$.For a given $h\in D$, the codispersion coefficient is defined as

$$\rho(h)=\frac{\gamma(h)}{\sqrt{V_X(h)V_Y(h)}}, \tag{5}$$

where $s,\ s+h\in D$, $\gamma(h)=\mathsf{E}[X(s+h)-X(s)][Y(s+h)-Y(s)]$, $V_X(h)=\mathsf{E}[X(s+h)-X(s)]^2$(and similarly for$V_Y(h)$).

For $d=2,$ the sample codispersion coefficient is defined by

$$\hat{\rho}(h)=\frac{\sum_{s,s+h\in D'}a_s b_s}{\sqrt{\hat{V}_X(h)\hat{V}_Y(h)}} \tag{6}$$

with$s=(s_1,\ s_2)$, $h=(h_1,\ h_2)$, $D'\subseteq D$, $\#(D')<\infty$, $a_s=\mathrm{X}(s_1+h_1,\ s_2+h_2)-X(s_1,\ s_2)$, $b_s=\mathrm{Y}(s_1+h_1,\ s_2+h_2)-Y(s_1,\ s_2)$, $\hat{V}_X(h)=\sum_{s,s+h\in D'}a_s^2$, and $\hat{V}_Y(h)=\sum_{s,s+h\in D'}b_s^2$.

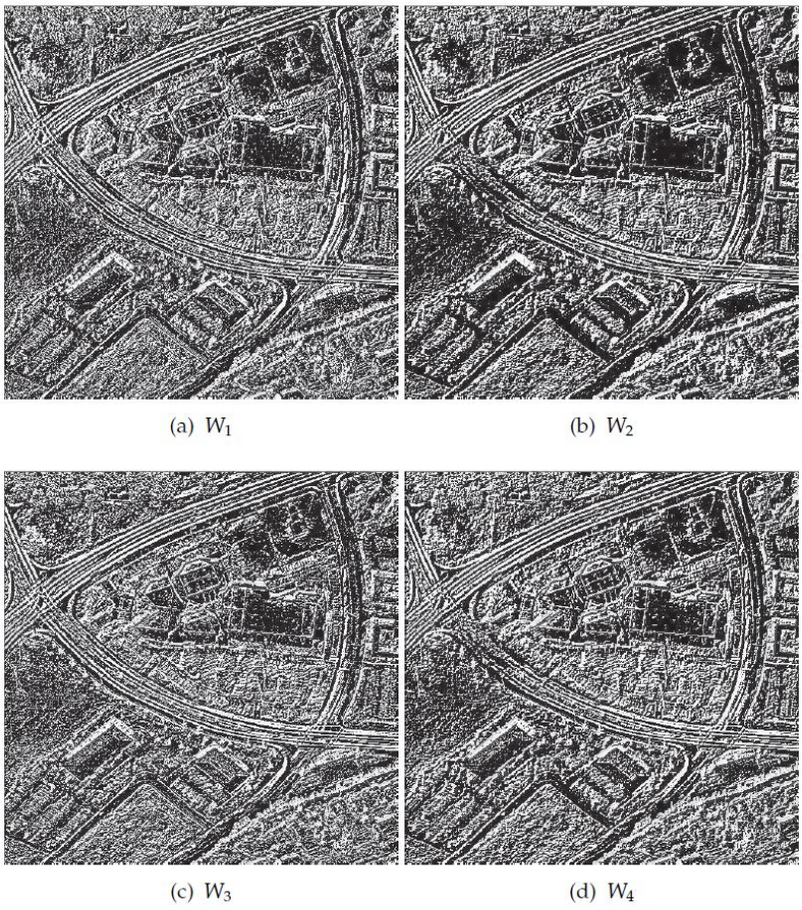


Figure 4. a)-(d) Images generated by Algorithm 2 with prediction windows $W_1 - W_4$ respectively.

The index Q (Wang and Bovik, 2002) is

$$Q = \frac{4S_{XY}\overline{X}\,\overline{Y}}{(S_X^2 + S_Y^2)[\overline{X}^2 + \overline{Y}^2]} = \frac{S_{XY}}{S_X S_Y} \cdot \frac{2\overline{X}\,\overline{Y}}{\overline{X}^2 + \overline{Y}^2} \cdot \frac{2S_X S_Y}{S_X^2 + S_Y^2} = C \cdot M \cdot V, \tag{7}$$

where $\overline{X}$ is the mean of $(X_s)_{s\in D}$, S_X is the standard deviation of $(X_s)_{s\in D}$, and S_{XY} is the covariance between $(X_s)_{s\in D}$ and $(Y_s)_{s\in D}$ (and similarly for $\overline{Y}$ and S_Y). The quantity $C = S_{XY} / S_X S_Y$ models the linear correlation between $(X_s)_{s\in D}$ and $(Y_s)_{s\in D}$, $M = 2\overline{X}\,\overline{Y} / (\overline{X}^2 + \overline{Y}^2)$ measures the similarity between the sample means (luminance) of $(X_s)_{s\in D}$ and $(Y_s)_{s\in D}$, and $V = 2S_X S_Y / (S_X^2 + S_Y^2)$ measures the similarity related to the contrast between the images. Coefficient Q is defined as a function of the correlation coefficient; hence, it is able to capture only the linear association between $(X_s)_{s\in D}$ and $(Y_s)_{s\in D}$ It is unable to account for other types of relationships between these sequences, including the spatial association in a specific direction h. Ojeda et al. (2012) suggested by the CQ index, which is defined as:

$$CQ(h)=\hat{\rho}(h)\cdot M\cdot V, \tag{8}$$

where M and V are defined as in (7).

The correlation coefficient and the coefficients defined in (6), (7) and (8) were computed to compare the fitted images, which were generated with a prediction window with two elements and associated with the images shown in Figure 4(a) -(f), and the original images. The results are shown in Table 1. In all cases, the highest values of the image quality measures are attained for the image fitted using the prediction window W_2.This means that the residual image shown in Figure 4 (b) is the best segmentation yielded by Algorithm 2. The same

Coefficient	Original-Fitted(W_1)	Original-Fitted(W_2)	Original- Fitted(W_3)	Original-Fitted(W_4)
Correlation	0.8985	0.9400	0.8899	0.8953
Q	0.8968	0.9384	0.8882	0.8935
		$h=(0,1)$		
$\hat{\rho}(h)$	0.3341	0.6130	0.3954	0.4779
$CQ(h)$	0.3336	0.6120	0.3947	0.4770
		$h=(1,1)$		
$\hat{\rho}(h)$	0.5457	0.7072	0.5040	0.4645
$CQ(h)$	0.5447	0.7060	0.5030	0.4036
		$h=(1,0)$		
$\hat{\rho}(h)$	0.4740	0.6655	0.3681	0.4980
$CQ(h)$	0.4731	0.6644	0.3675	0.4970

Table 1. Image quality measures between the fitted and original (aerial) images related to the residual images shown in Figure 4.

experiment was carried out for the image shown in Figure 2(a). Table 2 summarizes the values of the image quality coefficients for the fitted images generated by Algorithm 2 with prediction windows W_1, W_2, W_3, and W_4. In this case, the best performance is for the fitted image

Coefficient	Original-Fitted(W_1)	Original-Fitted(W_2)	Original- Fitted(W_3)	Original-Fitted(W_4)
Correlation	0.9801	0.9848	0.9800	0.9855
Q	0.9801	0.9847	0.9800	0.9854
		$h=(0,1)$		
$\hat{\rho}(h)$	0.2294	0.5191	0.2366	0.5321
$CQ(h)$	0.2294	0.5191	0.5303	0.2366
		$h=(1,1)$		
$\hat{\rho}(h)$	0.5158	0.5862	0.5134	0.6030
$CQ(h)$	0.5158	0.5862	0.5134	0.6029
		$h=(1,0)$		
$\hat{\rho}(h)$	0.2537	0.5267	0.2420	0.5386
$CQ(h)$	0.2537	0.5267	0.2420	0.5386

Table 2. Image quality measures between the fitted and original (peppers) images.

generated with prediction window W_4.In general, the performance of Algorithm 2 depends on the choice of the prediction window. One way to choose the prediction window that yields the best segmentation is to maximize the association between the fitted and original images. Indeed, if we denote the original image by $Z$and the fitted image generated by Algo-

rithm 1 with the prediction window W_i by $\hat{Z}_{W_i}$, then the prediction window that produces the best segmentation can be obtained by finding the maximum value of one of the quality measures (6), (7) or (8) between Z and Z_{W_i}. This criterion is summarized in the following algorithm.

Algorithm 3.

1. Use Algorithm 1 to generate the approximated images $\hat{Z}_{W_i}$ of Z, $i=1, 2, 3, 4$.

2. Compute an image quality index between Z and $\hat{Z}_{W_i}$ for all $i=1, 2, 3, 4$. Suppose that the maximum value for the image quality index is attained for $\hat{Z}_{W_j}$, $1 \le j \le 4$. Then, the best fitted image is $\hat{Z}_{W_j}$.

3. Compute the residual autoregressive image $Z - \hat{Z}_{W_j}$.

3. Clustering Time series

3.1. Measuring Closeness and Association Between Time Series

Let $x=(x_1, x_2, \ldots, x_p)$ and $y=(y_1, y_2, \ldots, y_q)$ be two time series. There are several conventional distance measures between time series. For example, if $p=q=n$, then the Euclidean distance between x and y is defined as $d_E(x, y)=\left(\sum_{i=1}^{n}(x_i-y_i)^2\right)^{1/2}$. As is evident, d_E ignores information about the dependence between x and y. The Minkowski distance is a generalization of the Euclidean distance, which is defined as

$$d_M(x, y)=\left(\sum_{i=1}^{n}(x_i-y_i)^q\right)^{1/q}, \tag{9}$$

where q is a positive integer. The Fréchet distance was introduced to measure the proximity between continuous curves. Let m be a natural number such that $m \le \min(p, q)$. Let M be the set of all mappings r between x and y such that r is a sequence of m pairs preserving the order

$r=((x_{a_1}, y_{b_1}), (x_{a_2}, y_{b_2}), \ldots, (x_{a_m}, y_{b_m}))$,

where $a_i \in \{1, 2, ..., p\}$, $b_j \in \{1, 2, ...q\}$ with $a_1=1$, $b_1=1$, $a_m=p$, $b_m=q$ and for $i \in \{1, 2, \ldots, m-1\}$, $a_{i+1}=(a_i \text{ or } a_i+1)$ and $b_{i+1}=(b_i \text{ or } b_i+1)$. Note that $|r|=\max_{i=1,2,\ldots,m}|x_{a_i}-y_{b_i}|$ is the mapping length representing the maximum span between two coupled observations. Thus, the Fréchet distance between the series x and y is given by

$$d_F(x, y)=\min_{r \in M}|r|=\min_{r \in M}\left(\max_{i=1,2,\ldots,m}|x_{a_i}-y_{b_i}|\right). \tag{10}$$

http://dx.doi.org/10.5772/50513

Dynamic time warping (DTW) is a variant of the Fréchet distance that considers mapping length as the sum of the spans of all coupled observations. That is,
$\left| r \right| = \sum_{i=1,2,\ldots,m} | x_{a_i} - y_{b_i} |$.

Dynamic time warping is then defined as

$$d_{DTW}(x, y) = \min_{r \in M} \left| r \right| = \min_{r \in M} \sum_{i=1,2,\ldots,m} | x_{a_i} - y_{b_i} | . \tag{11}$$

The distances defined above are based on the proximity of the values $| x_{a_i} - y_{b_i} |$. However, these distances disregard both the temporal dependence between the sequences x and y and the correlation structure of each sequence.

Several distance measures that are functions of the correlation between two sequences ($Cor(x, y)$) have been suggested. For example, (Golay et al.,1998) proposed
$d_{cc}(x, y) = \left(\frac{1 - Cor(x, y)}{1 + Cor(x, y)} \right)^{\beta}$ and $d_{cc}^2(x, y) = 2(1 - Cor(x, y))$,

where β is a parameter related to the fuzzy c-means classification algorithm (Macqueen, 1967). However, none of these measures takes into account the serial association between the sequences because the correlation coefficient is a crude measure of association. This approach requires the study of coefficients that are capable of capturing the spatial or serial correlation between two sequences.

3.2. The Codispersion Coefficient for Time Series

Consider two weakly stationary processes, $X = \{X_s : s \in D \subset \mathbb{Z}\}$ and $Y = \{Y_s : s \in D \subset \mathbb{Z}\}$, and let x and y be realizations of these processes as in Section 3.1. For $d = 1$, the estimator (6) becomes

$$\hat{\rho}(h) = \frac{\sum_{t \in N(h)} (x_{t+h} - x_t)(y_{t+h} - y_t)}{\sqrt{\sum_{t \in N(h)} (x_{t+h} - x_t)^2 \sum_{t \in N(h)} (y_{t+h} - y_t)^2}} \tag{12}$$

where $N(h) = \{t : t + h \in D\}$, $N = | N(h) |$ is the cardinality of $N(h)$, and sequences x and y are realizations of processes x and y, respectively. The coefficient $\hat{\rho}(h)$ is called the comovement coefficient when $h = 1$. Although $\hat{\rho}(h)$ is not the correlation coefficient, the codispersion coefficient shares a number of its standard properties. For example, $\hat{\rho}(h)$ is translation invariant, positive homogeneous, symmetric in its arguments, positive definite for a sequence and lagged versions of itself, and interpretable as the cosine of the angle between the vectors formed by the first difference of the sampled series. As in the case of classic correlation, a codispersion coefficient of +1 indicates that the compared functions or processes are rescaled or retranslated versions of one another. Similarly, a profile matched with its reflection across the time axis yields a codispersion of −1. The value $\hat{\rho}(h) = 0$ expresses that there is no monoto-

nicity between x and y and that their growth rates are stochastically linearly independent. More details can be found in (Rukhin & Vallejos, 2008 , Vallejos, 2008).

3.3. Dissimilarity Index for Time Series

This coefficient involves a distance measure and a correlation-type measure that addresses both the correlation behavior and the proximity of two time series. The dissimilarity index depends on similarity behaviors, which should be specified in advance. The suggested dissimilarity index $D(x, y, h)$ for the realizations x and y and $d=1$ is given by

$$D(x, y, h)=f(\hat{\rho}(h))\cdot d(x, y), \tag{13}$$

where f is an adaptive tuning function, and $d(x, y)$ is one of the conventional distances described in Section 3.1 that summarizes the closeness of sequences x and y. There are many possible ways to choose a function f. Here, we follow the guidelines given in (Chouakria & Nagabhushan, 2007), according to which f is considered an exponential adaptive tuning function given by

$$f_k(t)=\frac{2}{1+exp(kt)}, \tag{14}$$

where $k\geq 0$ modulates the contributions of the proximity with respect to values and behavior. For example, when $|\hat{\rho}(1)|$ is large and $k=2$, the proximity with respect to behavior contributes 76.2% to D. The flexibility of D allows us to choose k such that for highly dependent sequences, the correlation structure can have a large weight in (13).

Note that (13) is a generalization of the dissimilarity index introduced in Chouakria & Nagabhushan, (2007). The dissimilarity index (13) can capture high-order serial correlations between the sequences because the distance lag h is arbitrary, while Chouakria's index only captures the first-order correlation.

The dependence of (13) on h is crucial, and in some specific cases, h can be chosen using an optimal criterion. For example, for two AR(1) processes with parameters ϕ_1 and ϕ_2, respectively, and a correlation structure between the errors (Rukhin & Vallejos,2008), it is possible to find ϕ_1, ϕ_2 such that $Var(\hat{\rho}(1))Var(\hat{\rho}(2))$. In other words, for those processes in which the asymptotic variance of the codispersion coefficient is known, we suggest setting the value of h to produce the minimum variance. That is,

$$\hat{h}=\arg\min_h[Var(\hat{\rho}(h))]. \tag{15}$$

When the variance of the codispersion coefficient is difficult to compute, resampling methods can be use to estimate the variance of the sample codispersion coefficient (Politis & Romano, 1994, Vallejos, 2008).

http://dx.doi.org/10.5772/50513

In the next section, we present two simulation examples to illustrate the capabilities of the hierarchical methods using the distance measure (13) under the tuning function given by (14). All else being constant, the clusters produced using traditional distances are usually different from those yielded using the distance measure (13).

3.4. Simulations

In this example, we simulate observations from six first-order autoregressive models to illustrate the clustering produced by hierarchical methods when the sequences exhibit serial correlation. To generate the series, we consider the following models.

$$X_t^i = \phi_i X_{t-1}^i + \varepsilon_t^i, \qquad i = 1, 2, \ldots 6,$$

where $\forall\, i = 1, 2, \ldots, 6$, $X^i = \{X_t^i\}_{t \in \mathbb{Z}}$ define the $i-$model, and the sequence $\varepsilon^i = \{\varepsilon_t^i\}_{t \in \mathbb{Z}}$ is zero-mean white noise. Note that the sequences ε^1 and ε^2 have the covariance structure

$$Cov(\varepsilon_t^1, \varepsilon_s^2) = \begin{cases} \rho\sigma\tau, & s = t, \\ 0, & \text{otherwise}, \end{cases}$$

with $\sigma^2 = \tau^2 = 1$, and $\rho = 0.9$. The same covariance structure is assumed for ε_t^1 and ε_t^3, with $\sigma^2 = \tau^2 = 1$ and $\rho = 0.7$. ε^i, $i = 4, 5, 6$, are assumed to be white noise processes with variance 1 and are uncorrelated with all other error sequences. If $i,\ j \leq 3$, the correlation structure between processes X^i and X^j, $i \neq j$, is not null due to the correlation structure between ε^i and ε^j. Instead, if $i,\ j \geq 4$, $i \neq j$, the correlation structure between processes X^i and X^j vanishes.

Two hundred observations were generated from each model for $\phi_1 = -0.5$, $\phi_2 = -0.3$, $\phi_3 = -0.7$, $\phi_4 = 0.1$, $\phi_5 = -0.9$, and $\phi_6 = -0.2$. The goal was to perform time series clustering with the Euclidean distance and (13). Hierarchical methods with complete linkage using both measures were implemented to evaluate whether the methods are capable of capturing the correlation structure between the sequences described above. We used the distance measure (13) under the tuning function (14) for $h = 1$, and $k = 3$. That is, the correlation structure contributes 90.5% to D, whereas the proximity with respect to values contributes 9.5%.

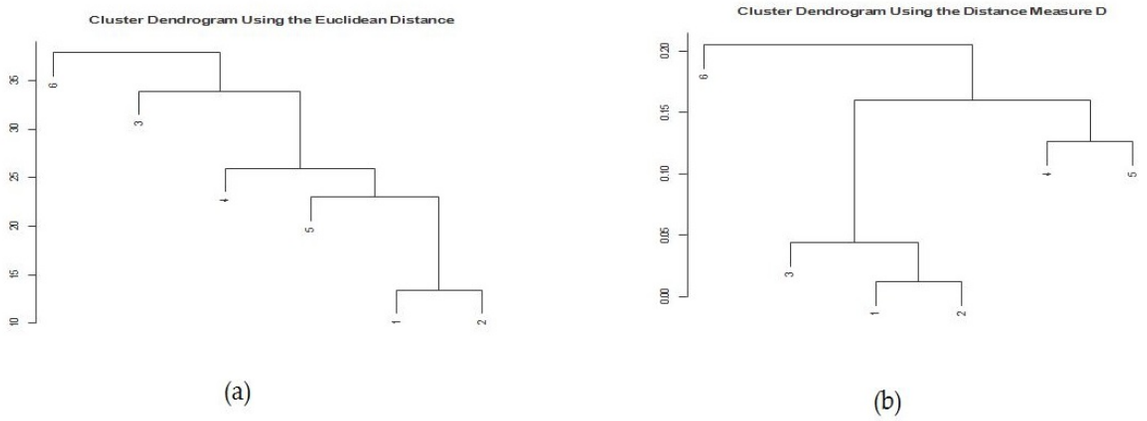


Figure 5. a) Time series clustering using the Euclidean distance, (b) Time series clustering using*D*.

In Figure 5, we see that the dendrogram obtained using hierarchical methods with the Euclidean distance does not recognize the correlation structure between $X^1$and X^3.In this case, sequences X^1, X^2, X^4, and $X^5$are pulled together before sequence X^3.However, hierarchical methods using (13) yield the expected results, combining sequences X^1, X^2, and X^3before the rest of the series.

To obtain better insight into the classification process using the proposed distance measure (13), we carried out a second simulation study that involves clustering measures based on other distances (but using the same setup). Observations from models 1-6 were generated using Gaussian white noise sequences for the errors, thereby preserving the same correlation structure used in the first study. The goal was to explore the ability of the distance measure (13) to group strongly correlated series first. A total of 1000 runs were considered for this

	d_E	d_{DTW}	d_M	d_F
Yes	65%	3%	99%	11%
No	35%	97%	1%	89%

Table 3. Percentage of correct clustering of the correlated series 1,2 and 3.

experiment, and 200 observations were generated in each run. We used measure (13) under the tuning function (14) for h =1, 2and k =1, 2, 3, 4.We counted the number of times that the hierarchical algorithm with complete linkage was able to pull together series 1, 2 and 3 before connecting them with other sequences. The traditional distances described in Section 2 were also implemented. After the 1000 simulation runs were finished, the percentage of times that the algorithm was able to recognize the correlated series was recorded. The results of the experiment are summarized in Table 3 and 4.

http://dx.doi.org/10.5772/50513

	$D_E(h,k)$	$D_{DTW}(h,k)$	$D_M(h,k)$	$D_F(h,k)$
	$h=1,k=1$	$h=1,k=1$	$h=1,k=1$	$h=1,k=1$
Yes	82%	60%	100%	13%
No	18%	40%	0	87%
	$h=1,k=2$	$h=1,k=2$	$h=1,k=2$	$h=1,k=2$
Yes	87%	68%	99%	22%
No	13%	32%	1%	78%
	$h=1,k=3$	$h=1,k=3$	$h=1,k=3$	$h=1,k=3$
Yes	90%	64%	99%	26%
No	10%	36%	1%	74%
	$h=1,k=4$	$h=1,k=4$	$h=1,k=4$	$h=1,k=4$
Yes	90%	65%	98%	29%
No	10%	35%	2%	71%
	$h=2,k=1$	$h=2,k=1$	$h=2,k=1$	$h=2,k=1$
Yes	83%	58%	100%	16%
No	17%	42%	0%	84%
	$h=2,k=2$	$h=2,k=2$	$h=2,k=2$	$h=2,k=2$
Yes	87%	59%	100%	26%
No	13%	41%	0%	74%
	$h=2,k=3$	$h=2,k=3$	$h=2,k=3$	$h=2,k=3$
Yes	90%	50%	100%	21%
No	10%	50%	0%	79%
	$h=2,k=4$	$h=2,k=4$	$h=2,k=4$	$h=2,k=4$
Yes	89%	56%	100%	26%
No	11%	44%	0%	74%

Table 4. Percentage of correct clustering of the correlated series (1,2, and 3). $D_{DTW}(h,k)$is distance measure (13) with the DTW distance. $D_M(h,k)$and $D_F(h,k)$denote distance measure (10) with Minkowski and Frechet distances respectively.

Note from Table 3 that the traditional distance measures failed to group the correlated sequences, with the exception of the Minkowski distance, which correctly grouped the correlated series 99% of the time. The hierarchical algorithm that uses the distance measure (13) has a higher percentage of well-clustered correlated sequences than the same algorithm using the traditional distance measures described in Section 2 (see Table 4). The percentage of correct clusters increased in all cases with the distance measure (13), suggesting that hierarchical algorithms can be improved by including coefficients of association that consider high-order cross-correlation.

3.5. The NDVI Data Set

In this section, we consider time series from four different locations in Argentina. The data set consists of 15 monthly NDVI series measured during a period of 19 years (i.e., January 1982-December 2000). The observed values correspond to a transformation to the interval

[0, 255]of the original NDVI series, which commonly resides in the interval [−1, 1].The data were collected by a NOAA sensor at 1 km resolution and provided by the Comisión Nacional de Actividades Espaciales (CONAE) in Córdoba, Argentina. Fifteen time series were collected from the following: the Amazon region in the northeast of Argentina (1, 2, 3), the Patagonia region in the south of Argentina (4, 5, 6, 7), the Pampean region in the center of Argentina (8, 9, 10, 11) and the Pre-Andean zone of Argentina (12, 13, 14, 15). The time series are shown in Figure 6.

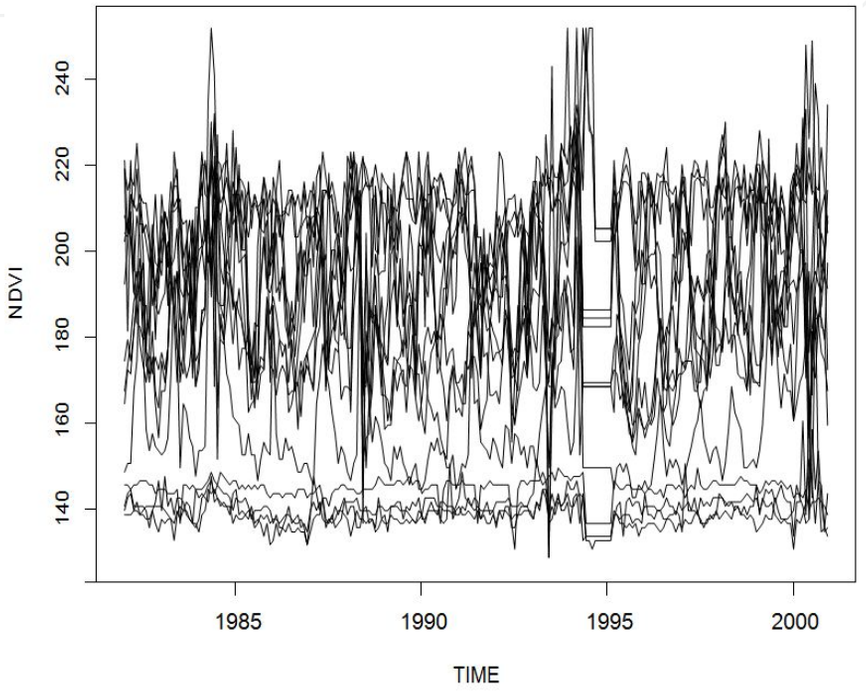


Figure 6. Fifteen NDVI series collected from four different regions in South America.

We can observe a variety of different patterns in Figure 6. In particular, the data collected during the period 1994-1995 show irregular behavior. Additionally, the original data lack some information (less than one percent) for all series over the period 1999-2000. An imputation technique based on moving averages, which takes into account past and future values of the series, was used to replace missing values. The series were grouped by geographical region and then plotted (Figure 7). Similar patterns are observed for the series across each group.

An exploratory data analysis was carried out for each of the 15 series. There exists significant autocorrelation of order of at least one in all series. Seasonal components are present in most of partial autocorrelations. Because there is no large departure from the weakly stationary assumptions (i.e., constant means and variances), all series can be modeled using the Box-Jenkins approach. Specifically, seasonal ARIMA models can be fitted to each single series with a small number of parameters (i.e., $p \leq 5$, $q \leq 5$and$d \leq 2$).

http://dx.doi.org/10.5772/50513

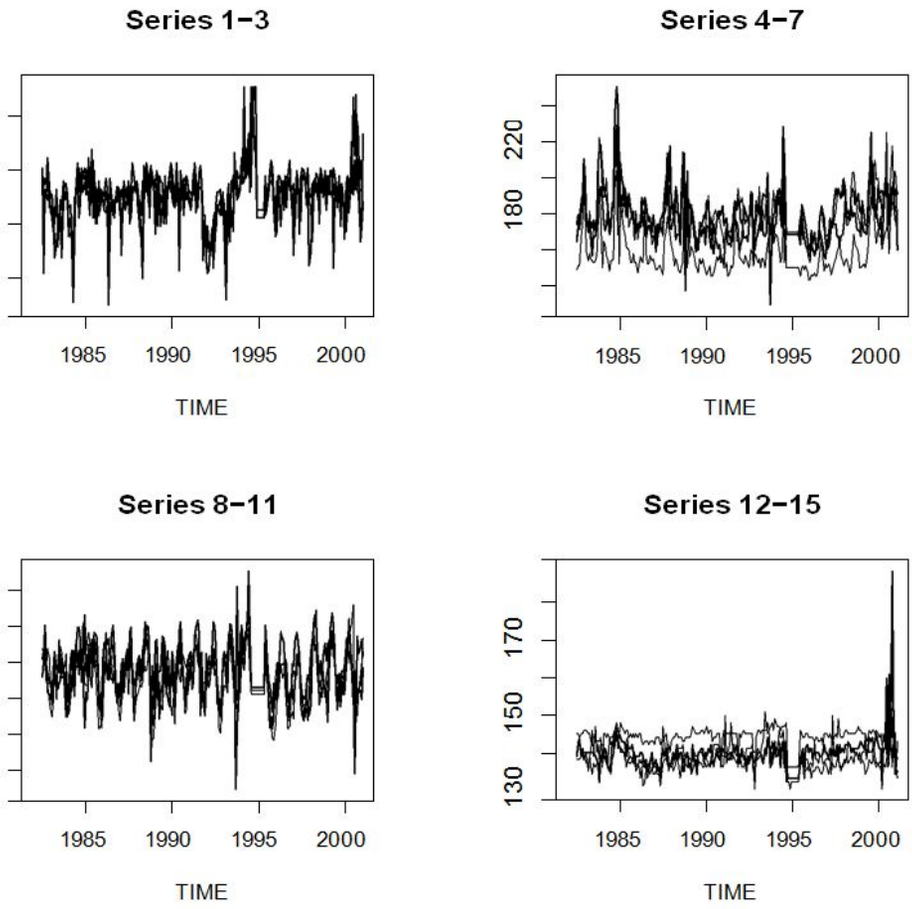


Figure 7. The fifteen NDVI series grouped by area.

3.6. Clustering

Using the NVDI data set described in Section 3.5, the distance measure Din (13) was computed for all possible pairs. Then, dendrogram plots were constructed using a hierarchical procedure (i.e., complete linkage) to compare the mergers and clusters produced using $D$with those produced using the conventional distances described in Section 2.2 and (13). In Figure 8, we observe the agglomeration produced by using the Euclidean distance (top left); the other five plots show the results produced by distance (13) for different values of $k$and h.The agglomeration algorithm using the Euclidean distance merges series 5 together with series 12-15 and thus does not preserve location when grouping series. However, with $k=1$and $h=1$ in (13), the location dependence of the 15 series is captured. Higher values of $k$and h do not modify the original clusters formed using$D$. In Figure 9, we see the clusters and merges yielded by using the Minkowski distance in (13). Note that series 4 is classified together with series 8, 9, 10 and 11 in the top left plot, but with $k=1$and $h=1$, the algorithm handles the series differently (shown in the top right plot) by merging together those series that are in the same

location. The effects of higher values of k and h .are again negligible. The Fréchet distance produced unsatisfactory results. In this case, the hierarchical algorithm does not take into account the geographical location when using both the conventional distances measures and (13). For example, series 1, 2 and 3 were merged in different stages. Nevertheless, when k and h are increased, the algorithm using (13) still clusters the series by geographical location. Indeed, if our goal is to produce four clusters as before, the hierarchical algorithm with $h = 2$ and $k = 3$ produces a geographically consistent agglomeration (dendrograms not shown here). The same analysis was performed using the hierarchical algorithm with the dynamic time warping distance measure. In this case, this distance measure produced an agglomeration by geographical location and thus did not need to be modified to capture serial correlation. The result yielded with (13) produced the same outcome for all values of k and h .

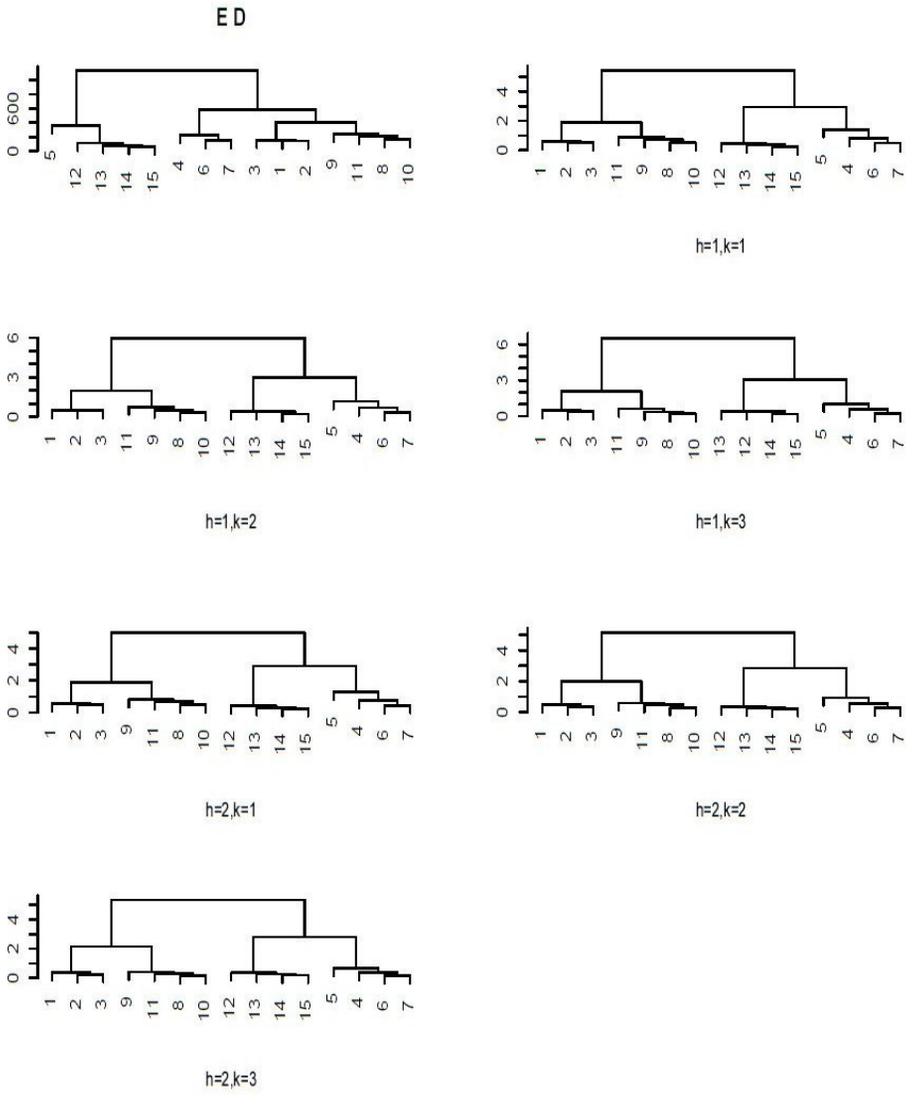


Figure 8. Clusters produced by using a hierarchical method with the Minkowski distance and (13).

http://dx.doi.org/10.5772/50513

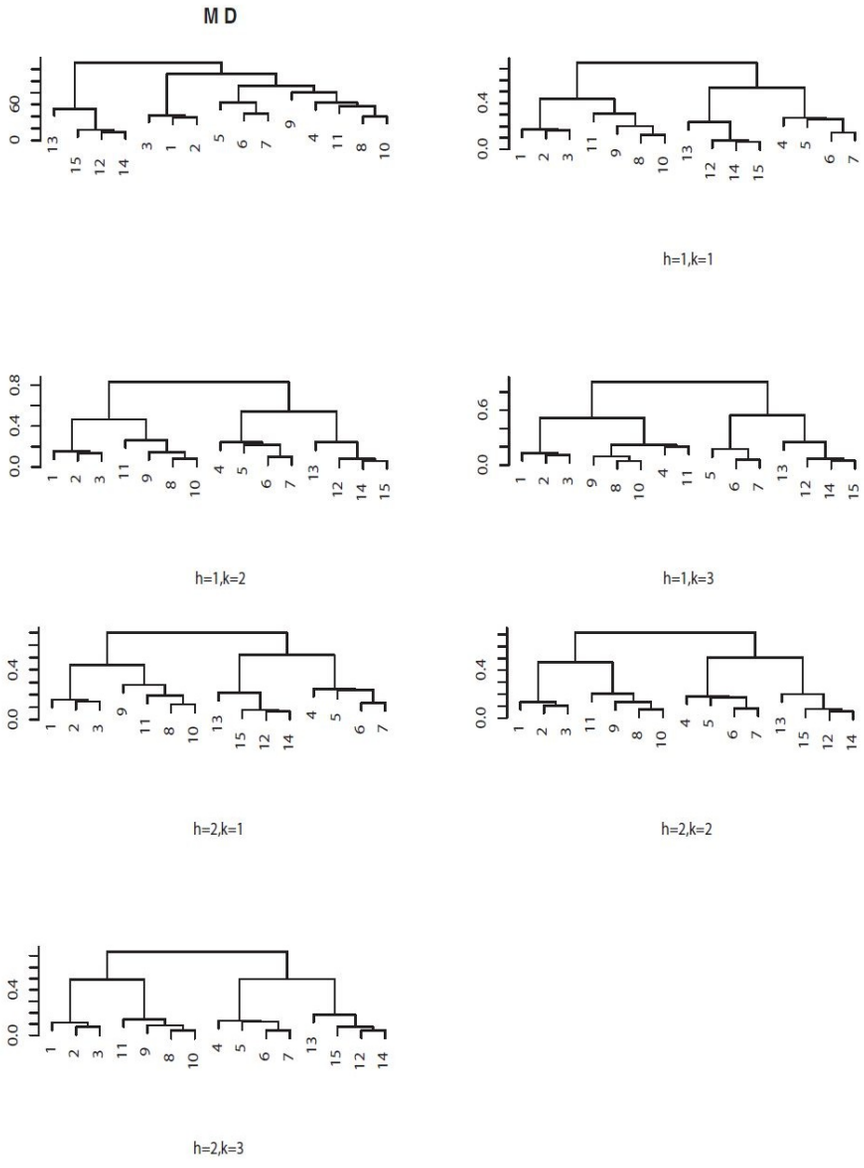


Figure 9. Clusters produced by using a hierarchical method with the Fréchet distance and (13)

4. Concluding Remarks and Future Work

This chapter described two problems. The first problem involved image segmentation, while the second problem involved clustering time series. For the first problem, a new algorithm was proposed that enhances the segmentation yielded by a previous algorithm (Ojeda et al., 2010). Identifying the best prediction window improves segmentation based on the estimation of AR-2D processes and generalizes the previous algorithm to different prediction windows associated with unilateral processes on the plane. An analysis of the association between the original and fitted images relies on the selection of a suitable image quality

measure. Using three image quality coefficients that are commonly used in image segmentation, we carried out experiments that support our algorithm. Specifically, a set of images belonging to the image database (http://sipi.usc.edu/database/) were processed and provided satisfactory results (not shown here) in terms of image segmentation.

This chapter also proposed an extension of the dissimilarity measure first introduced in (Chouakria & Nagabhushan,2007). The simulation experiments performed and the data analysis carried out for relevant ecological series show that the distance lag h plays an important role in capturing the higher-order correlation of each series. Cluster analysis performed using the proposed distance measure produced different merges and dendrograms. Furthermore, the percentage of times that the hierarchical algorithms correctly classified the highly correlated sequences increased in all cases in which the distance measure (13) was used. For the NDVI series discussed in Section 3.5, the distance measure Dimproved the performance of the Euclidean, Fréchet and Minkowski distances in the presence of high-order autocorrelation in the series. The dynamic time warping distance measure showed the best performance in capturing the serial correlation between the NDVI series, and thus, it was not necessary to introduce modified distance measures such as (13) to ensure agglomeration by geographical location.

Now, further research for the topics presented in this chapter is outlined.

Following the notation used in the Algorithm 3, consider the following residual image. $R_{W_i} = Z - \hat{Z}_{W_i}$.

One interesting open problem involves the characterization of the types of images and distributions associated with the segmentation produced by Algorithms 2 and 3. In addition, the definition and study of linear combinations of residual images produced by distinct prediction windows is also of interests. For example,

$$I = \sum_{j=1}^{4} a_j R_{W_i},$$

where $a_j$is a weight associated with the residual image R_{W_i}.

Regarding the clustering technique problem, the distribution of Dcan be studied from a parametric point of view. This is an open problem that we expect to address in future research.

Acknowledgements

The first author was partially supported by Fondecyt grant 1120048, UTFSM under grant 12.12.05, and Proyecto Basal CMM, Universidad de Chile. The second author was supported in part by CIEM-FAMAF, UNC, Argentina.

http://dx.doi.org/10.5772/50513

Author details

Ronny Vallejos[1*] and Silvia Ojeda[2]

*Address all correspondence to: ronny.vallejos@usm.cl

1 Department of Mathematics, Universidad Técnica Federico Santa María,, Chile

2 FAMAF, Universidad Nacional de Córdoba,, Argentina

References

[1] Allende, H., Galbiati, J., & Vallejos, R. (2001). Robust image modeling on image processing. *Pattern Recognition Letters*, 22(11), 1219 -1231.

[2] Baran, S., Pap, G. , & Zuijlen, M. C. A. (2004). Asymptotic inference for a nearly unstable sequence of stationary spatial AR models. *Statistics & Probability Letters*, 69(1), 53 -61.

[3] Basu, S., & Reinsel, G. (1993). Properties of the spatial unilateral first-order ARMA model. *Advances in Applied Probbability*, 25(3), 631 -648.

[4] Bustos, O., Ojeda, S., & Vallejos, R. (2009a). Spatial ARMA models and its applications to image filtering. *Brazilian Journal of Probability and Statistics*, 23(2), 141 -165 .

[5] Bustos, O., Ojeda, S., Ruiz, M., Vallejos, R., & Frery, A. (2009b). Asymptotic Behavior of RA-estimates in Autoregressive 2D Gaussian Processes. *Journal of Statistical Planning and Inference*, 139(10), 3649-3664.

[6] Cariou, C., & Chehdi, K. (2008). Unsupervised texture segmentation/classification usind 2-D autoregressive modeling and the stochastic expectation-maximization algorithm. *Pattern Recognition Letters*, 29(7), 905-917.

[7] Choi, B. (2000). On the asymptotic distribution of mean, autocovariance, autocorrelation, crosscovariance and impulse response estimators of a stationary multidimensional random field. *Communications in Statistics-Theory and Methods*, 29(8), 1703-1724.

[8] Chouakria, A. D., & Nagabhushan, P. N. (2007). Adaptive dissimilarity for measuring time series proximity. *Advances in Data Analysis and Classification*, 1(1), 5 -21 .

[9] Clifford, P., Richardson, S., & Hémon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, 45(1), 123 -134.

[10] Cullis, B. R., & Glesson, A. C. (1991). Spatial analysis of field experiments-an extension to two dimensions. *Biometrics*, 47(4), 1449-1460.

[11] Francos, J., & Friedlander, B. (1998). Parameter estimation of two-dimensional moving average random fields. *IEEE Transaction Signal Processing*, 46(8), 2157-2165.

[12] Gaetan, C., & Guyon, X. (2010). *Spatial Statistics and Modelling*, Springer, New York.

[13] Genton, M. G., & Koul, H. L. (2008). Minimum distance inference in unilateral autoregressive laticce processes. *Statistica Sinica*, 18, 617-631.

[14] Golay, X., Kollias, S., Stoll, G., Meier, D., & Valavanis, A. (1998). A new correlation-based fuzzy logic clustering algorithm of FMRI. *Magnetic Resonance in Medicine*, 40(2), 249-260.

[15] Grondona, M. R., Crossa, J., Fox, P. N., & Pfeiffer, W. H. (1996). Analysis of variety yield trials using two-dimensional separable ARIMA processes. *Biometrics*, 52(2), 763-770.

[16] Guo, J., & Billard, L. (1998). Some inference results for causal autoregressive processes on a plane. *Journal of Time Series Analysis*, 19(6), 681-691.

[17] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika*, 69(1), 95-105.

[18] Ho, P. G. P. (2011). Image segmentation by autoregressive time series model. In Image Segmentation edited by Pei-Gee Ho. InTech.

[19] Illig, A., & Truong Van, B. (2006). Asymptotic results for spatial ARMA Models. *Communications in Statistics-Theory and Methods*, 35(4), 671 -688 .

[20] Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Comput. Surveys,*, 31(3), 264-323.

[21] Kashyap, R., & Eom, K. (1988). Robust images techniques with an image restoration application. *IEEE Trans. Acoust. Speech Signal Process*, 36(8), 1313 -1325 .

[22] Mac, Queen. J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5 -th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1, 281-297.

[23] Martin, R. J. (1996). Some results on unilateral ARMA laticce processes. *Journal of Statistical Planning and Inference*, 50(3), 395-411.

[24] Matheron, G. (1965). Les Variables Régionalisées et leur Estimation Masson Paris.

[25] Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303 -1313.

[26] Ojeda, S. M., Vallejos, R. O., & Lucini, M. (2002). Performance of RA Estimator for Bidimensional Autoregressive Models. *Journal of Statistical Simulation and Computation*, 72(1), 47 -62 .

[27] Ojeda, S. M., Vallejos, R., & Bustos, O. (2010). A New Image Segmentation Algorithm with Applications to Image Inpainting. .Computational Statistics & Data Analysis , 54(9), 2082 -2093 .

[28] Ojeda, S. M., Vallejos, R., & Lamberti, W. P. (2012). A Measure of Similarity Between Images. , in press Journal of Electronic Imaging.

[29] Quintana, C., Ojeda, S., Tirao, G., & Valente, M. (2011). Mammography image detection processing for automatic micro-calcification recognition. *Chilean Journal of Statistics*, 2(2), 69-79.

[30] Rukhin, A., & Vallejos, R. (2008). Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters*, 78(11), 1290 -1300 .

[31] Vallejos, R., & Mardesic, T. (2004). A recursive algorithm to restore images based on robust estimation of NSHP autoregressive models. *Journal of Computational and Graphical Statistics*, 13(3), 674 -682 .

[32] Vallejos, R., & Garcia-Donato, G. (2006). Bayesian analysis of contaminated quarter plane moving average models. *Journal of Statistical Computation and Simulation*, 76(2), 131-147.

[33] Vallejos, R. (2008). Assessing the association between two spatial or temporal sequences. *Journal of Applied Statistics*, 35(12), 1323 -1343 .

[34] Tjostheim, D. (1978). A measure of association for spatial variables. *Biometrika*, 65(1), 109 -114.

[35] Wang, Z., & Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3), 81 -84 .

IntechOpen

IntechOpen