ASSOCIATION CHARACTERISTICS IN SPATIAL STATISTICS

A. L. Rukhin

Dept. of Mathematics and Statistics, Univ. of Maryland at Baltimore County,
Baltimore, MD 21250 USA, and
Statistical Engineering Division, National Inst. of Standards and Technology
Gaithersburg, MD 20899 USA
E-mail: rukhin@math.umbc.edu

A coefficient of association for two spatial sequences is suggested. Some properties of this characteristic are discussed. By using the Central Limit Theorem for stationary random fields its limiting asymptotic normality is established for any error distribution admitting finite fourth moment. The mean and the variance of this limiting law are found. Several examples of moving average and autoregression models are presented.

Key words: autoregression, confidence interval, correlation, cross-variogram, moving average, tests of independence, time series.

2000 Mathematics Subject Classification: ??????.

1. Introduction and Summary

There has been a considerable interest in spatial processes in diverse fields such as economics, environmental science, climatology, ecology, oceanography, soil science, real estate markets, image modelling, etc. Many developments have been summarized in the books by Ripley (1981), Cliff and Ord (1981), and Cressie (1993).

In this paper we examine the quantitative assessment of the similarity between two spatial or two temporal sequences. In time series it is of interest to study how well two time sequences move together. Nonparametric tests for the association or comovement of time series have appeared in the literature (e.g., Yang and Schreck-engost, 1981). Such tests have typically been oriented to establish independence of two sequences.

There are several proposed measures of association between two spatial processes. A nonparametric association coefficient for two spatial variables has been suggested by Tjostheim (1978a). Later this characteristic was generalized by Hubert and

Golledge (1982). Hubert *et al.* (1985) criticized Tjostheim's coefficient for being "aspatial in design" and pointed "to the need for renewed interest in the concept of spatial association and the way in which it can be evaluated."

The correlation coefficient also has been used as a measure of spatial association. A modified test of association based on this coefficient was suggested by Clifford, Richardson and Hémon (1989) with later adjustments by Richardson and Clifford (1991). The problem of testing independence between two series of auto-correlated variables can also be studied by regression techniques (Ord, 1975) when a relationship between a dependent variable and a set of independent variables is postulated.

In this paper in Section 2 we suggest a coefficient to measure the association between two spatial processes. This measure is based on the cross-variogram, a characteristic which has been extensively used in multivariate spatial prediction (see Ver Hoef and Cressie, 1993). Banerjee, Carlin and Gelfand (2004), p. 219, refer to a related coefficient as generalized correlation or coherence.

Explicit expressions for the proposed coefficient are derived for spatial moving average models in Section 3. The asymptotic behavior of the sample coefficient is studied and its limiting normality is obtained for any error distribution admitting finite fourth moment in Section 4. Section 5 gives approximate confidence intervals and a test of independence, while Section 6 contains several examples of moving average models and of autoregression models in which the optimal lag is determined. The paper is concluded with a practical application of the suggested coefficient to the similarity of two polymer images.

2. Definition and Model

Consider two spatial processes, X and Y, defined on part of a rectangular lattice $D \subset \mathbb{Z}^d$. The cross-variogram provides spatial information on the relationship between X and Y. For processes with stationary increments the cross-variogram is defined as

$$C_{XY}(h) = E[X(s + h) - X(s)]\,[Y(s + h) - Y(s)],$$

for a "lag" h such that $s, s+h \in D$ (Cressie, 1993, p. 67 and p. 140). Implementation of the important kriging procedure (i.e., spatial prediction) in terms of the cross-variogram is discussed by Ver Hoef and Barry (1998). When the covariance function is symmetric and data of both variable types are measured at the same locations, using the same units, the cross-variogram provides consistent equations to perform cokriging (Myers, 1982). Cressie (1993), pp. 70–73, discusses in detail numerous advantages of working with the variogram, $C_{XX}(h) = E[X(s + h) - X(s)]^2$, rather than with the covariogram, $\mathrm{Cov}(X(s + h), X(s))$, and the following definition (1) is very much in spirit of variogram concept.

Define the similarity coefficient as the normalized version of the cross-variogram,

$$(1) \qquad\qquad \gamma = \gamma_{XY}(h) = \frac{C_{XY}(h)}{\sqrt{C_{XX}(h)C_{YY}(h)}}.$$

It is obvious that $\gamma_{XY}(h)$ is bounded in absolute value by 1. As defined, the similarity coefficient is different from the correlation coefficient between the corresponding coordinates, although they share some common properties. When $d = 1$, this coefficient was suggested as a comovement coefficient for two time series (Leigh, 1996).

We consider processes with stationary increments of the form

$$X(s+h) - X(s) = \sum c_t \varepsilon_{s-t}$$

with squarely summable coefficients $c_t = c_{t_1 \ldots t_d}$, and ε_i being zero mean independent random variables (innovations) with the same variance. These processes are characterized by the property of absolute continuity of their increments spectral measure. More generally, we assume that the vector process $Z(s) = (X(s), Y(s))^T$ formed by two spatial processes $X(s)$ and $Y(s)$ admits the following structure,

$$\tag{2} Z(s+h) - Z(s) = \sum_t C_t \varepsilon(s-t),$$

where $C_t = C_t(h)$ are 2×2 matrices such that $\sum_s ||C_s||^2 < \infty$ and $\varepsilon(t)$ are independent random vectors with mean 0 and covariance matrix Σ. In other words, we assume that stationary increments of the random field $Z(s)$ have absolutely continuous spectrum. We denote $\mathbb{S} = \{t : t \in \mathbb{Z}^d, C_t \neq 0\}$.

The idea of representation of a (rough) surface as a realization of a stationary field has been suggested by Linnik and Husu (1974). Whitenton and Deckman (1986) describe the process of matching wear scar traces of the surfaces of two balls to assess lubricant and material performance. Processes of the form (2) also appear in the simultaneous autoregressive models for a spatial process $X(s)$ which have been introduced by Whittle (1954); they are reviewed in Chapter 6 of Cressie (1993) (see also Tjostheim, 1978b.)

3. Distribution Characteristics of the Coefficient of Association

For two processes forming a two-dimensional random field of the form (2) the formula for (1) is immediate. Indeed,

$$\tag{3} E[Z(s+h) - Z(s)][Z(s+h) - Z(s)]^T = \sum_{t,u} C_t E\varepsilon(s-t)\varepsilon(s-u)^T C_u^T$$
$$= \sum_t C_t \Sigma C_t^T = K.$$

For $i, j = 1, 2$, denote by κ_{ij} the elements of symmetric matrix K above. Then clearly,

$$\gamma = \frac{\kappa_{12}}{\sqrt{\kappa_{11}\kappa_{22}}}.$$

Let $A \otimes B$ be the tensor (Kronecker) product of matrices A and B. We will need the form for a 4×4 matrix,

$$\tag{4} \Phi = E\big[(Z(s+h) - Z(s))(Z(s+h) - Z(s))^T - K\big]$$
$$\otimes \big[(Z(s+h) - Z(s))(Z(s+h) - Z(s))^T - K\big]$$

for a two-dimensional random field of the form (2).

Properties of the tensor product imply that

$$
(5) \quad \Phi = \sum_{t_1,t_2,t_3,t_4} E\big[C_{t_1}\big(\varepsilon(s-t_1)\varepsilon(s-t_2)^T - E\varepsilon(s-t_1)\varepsilon(s-t_2)^T\big)C_{t_2}^T\big]
$$

$$
\otimes \big[C_{t_3}\big(\varepsilon(s-t_3)\varepsilon(s-t_4)^T - E\varepsilon(s-t_3)\varepsilon(s-t_4)^T\big)C_{t_4}^T\big]
$$

$$
= \sum_{t_1,t_2,t_3,t_4} (C_{t_1}\otimes C_{t_3})E\big(\varepsilon(s-t_1)\varepsilon(s-t_2)^T - E\varepsilon(s-t_1)\varepsilon(s-t_2)^T\big)
$$

$$
\otimes \big(\varepsilon(s-t_3)\varepsilon(s-t_4)^T - E\varepsilon(s-t_3)\varepsilon(s-t_4)^T\big)(C_{t_2}^T\otimes C_{t_4}^T).
$$

Since the ε's are independent, the expected value in the sum (5) is a nonzero matrix if and only if either $t_1 = t_3$, $t_2 = t_4$, or $t_1 = t_4$, $t_2 = t_3$, or both of these conditions hold. Let for $s \neq t$,

$$
U = E\big(\varepsilon(t)\varepsilon(s)^T - E\varepsilon(t)\varepsilon(s)^T\big) \otimes \big(\varepsilon(t)\varepsilon(s)^T - E\varepsilon(t)\varepsilon(s)^T\big)
$$

$$
= E\varepsilon(t)\varepsilon(s)^T \otimes \varepsilon(t)\varepsilon(s)^T
$$

and

$$
V = E\big(\varepsilon(t)\varepsilon(s)^T - E\varepsilon(t)\varepsilon(s)^T\big) \otimes \big(\varepsilon(s)\varepsilon(t)^T - E\varepsilon(s)\varepsilon(t)^T\big)
$$

$$
= E\varepsilon(t)\varepsilon(s)^T \otimes \varepsilon(s)\varepsilon(t)^T.
$$

Put $W = E\big(\varepsilon(t)\varepsilon(t)^T - \Sigma\big) \otimes \big(\varepsilon(t)\varepsilon(t)^T - \Sigma\big)$. The matrix W, which determines the covariance matrix of the limiting normal distribution of $\sum_t \varepsilon(t)\varepsilon(t)^T$, has the form

$$
W = \begin{pmatrix}
\mu_{40}-\mu_{20}^2 & \mu_{31}-\mu_{20}\mu_{11} & \mu_{31}-\mu_{20}\mu_{11} & \mu_{22}-\mu_{11}^2 \\
\mu_{31}-\mu_{20}\mu_{11} & \mu_{22}-\mu_{20}\mu_{02} & \mu_{22}-\mu_{11}^2 & \mu_{13}-\mu_{11}\mu_{02} \\
\mu_{31}-\mu_{20}\mu_{11} & \mu_{22}-\mu_{11}^2 & \mu_{22}-\mu_{20}\mu_{02} & \mu_{13}-\mu_{11}\mu_{02} \\
\mu_{22}-\mu_{11}^2 & \mu_{13}-\mu_{11}\mu_{02} & \mu_{13}-\mu_{11}\mu_{02} & \mu_{04}-\mu_{02}^2
\end{pmatrix},
$$

while

$$
U = \begin{pmatrix}
\mu_{20}^2 & \mu_{20}\mu_{11} & \mu_{20}\mu_{11} & \mu_{20}\mu_{02} \\
\mu_{20}\mu_{11} & \mu_{11}^2 & \mu_{11}^2 & \mu_{11}\mu_{02} \\
\mu_{20}\mu_{11} & \mu_{11}^2 & \mu_{11}^2 & \mu_{11}\mu_{02} \\
\mu_{20}\mu_{02} & \mu_{11}\mu_{02} & \mu_{11}\mu_{02} & \mu_{02}^2
\end{pmatrix}
$$

and

$$
V = \begin{pmatrix}
\mu_{20}^2 & \mu_{20}\mu_{11} & \mu_{20}\mu_{11} & \mu_{11}^2 \\
\mu_{20}\mu_{11} & \mu_{11}^2 & \mu_{20}\mu_{02} & \mu_{11}\mu_{02} \\
\mu_{20}\mu_{11} & \mu_{20}\mu_{02} & \mu_{11}^2 & \mu_{11}\mu_{02} \\
\mu_{11}^2 & \mu_{11}\mu_{02} & \mu_{11}\mu_{02} & \mu_{02}^2
\end{pmatrix}.
$$

It follows that

$$
(6) \quad \Phi = \sum_{t\neq u}(C_t\otimes C_t)U(C_u^T\otimes C_u^T)
$$

$$
+ \sum_{t\neq u}(C_t\otimes C_u)V\big(C_u^T\otimes C_t^T\big) + \sum_t (C_t\otimes C_t)W(C_t^T\otimes C_t^T)
$$

$$
= \sum_{t,u}(C_t\otimes C_t)U(C_u^T\otimes C_u^T) + \sum_{t,u}(C_t\otimes C_u)V(C_u^T\otimes C_t^T)
$$

$$
+ \sum_t (C_t\otimes C_t)[W - U - V](C_t^T\otimes C_t^T) = \Phi_1 + \Phi_2 + \Phi_3.
$$

4. Limiting Distribution of the Sample Coefficient of Association

In this section we look for the asymptotic behavior of the sample association coefficient for two processes of the form (2),

$$\hat\gamma = \frac{\sum_{s\in N(h)}(X(s+h)-X(s))(Y(s+h)-Y(s))}{\sqrt{\sum_{s\in N(h)}(X(s+h)-X(s))^2\sum_{s\in N(h)}(Y(s+h)-Y(s))^2}}.$$

Here $N(h) = N_M(h) = \{s\colon s, s+h \in D_M\}$, where the observation domain D_M is increasing as $M \to \infty$. A typical example is that of a d-dimensional rectangle $D_M = [0, n_1] \times \cdots \times [0, n_d]$ with $M = \min(n_i)$ (Guyon, 1982). The cardinality $N = |N(h)|$ of $N(h)$ in this case is $\prod_1^d (n_i - |h_i|)_+$. The common distribution of possibly non-normal independent error vectors $\varepsilon(t) = (\varepsilon_1(t), \varepsilon_2(t))^T$ in (2) is supposed to have finite fourth moment. Denote by $\mu_{ij} = E\varepsilon_1^i(t)\varepsilon_1^j(t)$, $i, j = 0, 1, 2, 3, 4, i+j \le 4$, so that

$$\Sigma = \begin{pmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{pmatrix}.$$

Then $\rho^2 = \mu_{11}^2/(\mu_{20}\mu_{02})$ is the square of the correlation coefficient between components of ε.

The following (almost sure) convergence holds by the law of large numbers,

$$\hat K = \frac{1}{N} \sum_{s\in N(h)} [Z(s+h) - Z(s)][Z(s+h) - Z(s)]^T \to K.$$

In particular, $\hat\gamma \to \gamma$. To find the limiting distribution of $|N(h)|^{1/2}(\hat\gamma - \gamma)$, we use the Central Limit Theorem for stationary random fields (see Theorem 1 in Bolthausen, 1982, and application of this theorem to processes of the form (2) by Moore, 1988). This theorem implies that in distribution

$$\sqrt{N}(\hat K - K) = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} + o_P(1),$$

where z_{ij} are zero mean normal variables such that

$$E \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} \otimes \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} = \Phi,$$

with Φ defined by (5). The entries $\varphi_{ijk\ell}$ of Φ, which can be found from (6), give the covariances $\mathrm{Cov}(z_{ij}, z_{k\ell})$.

The limiting variance of the sample association coefficient is determined by the elements $\varphi_{1111} = \mathrm{Var}(z_{11})$, $\varphi_{2222} = \mathrm{Var}(z_{22})$, $\varphi_{1122} = \mathrm{Var}(z_{12})$, $\varphi_{1112} = \mathrm{Cov}(z_{11}, z_{12})$, $\varphi_{1212} = \mathrm{Cov}(z_{11}, z_{22})$, and $\varphi_{1222} = \mathrm{Cov}(z_{12}, z_{22})$. Indeed,

$$N^{1/2}[\hat\gamma - \gamma] = \frac{z_{12}}{\sqrt{\kappa_{11}\kappa_{22}}} - \frac{\kappa_{12}z_{11}}{2\sqrt{\kappa_{11}^3\kappa_{22}}} - \frac{\kappa_{12}z_{22}}{2\sqrt{\kappa_{11}\kappa_{22}^3}} + o_P(1),$$

converges in distribution to a zero mean normal random variable with the variance,

$$(7) \quad v^2 = \frac{\mathrm{Var}(z_{12})}{\kappa_{11}\kappa_{22}} + \frac{\kappa_{12}^2 \, \mathrm{Var}(z_{11})}{4\kappa_{11}^3\kappa_{22}} + \frac{\kappa_{12}^2 \, \mathrm{Var}(z_{22})}{4\kappa_{11}\kappa_{22}^3}$$

$$- \frac{\kappa_{12} \, \mathrm{Cov}(z_{11}, z_{12})}{\kappa_{11}^2\kappa_{22}} - \frac{\kappa_{12} \, \mathrm{Cov}(z_{12}, z_{22})}{\kappa_{11}\kappa_{22}^2} + \frac{\kappa_{12}^2 \, \mathrm{Cov}(z_{11}, z_{22})}{2\kappa_{11}^2\kappa_{22}^2}$$

$$= \frac{\varphi_{1122}}{\kappa_{11}\kappa_{22}} + \frac{\kappa_{12}^2\varphi_{1111}}{4\kappa_{11}^3\kappa_{22}} + \frac{\kappa_{12}^2\varphi_{2222}}{4\kappa_{11}\kappa_{22}^3} - \frac{\kappa_{12}\varphi_{1112}}{\kappa_{11}^2\kappa_{22}} - \frac{\kappa_{12}\varphi_{1222}}{\kappa_{11}\kappa_{22}^2} + \frac{\kappa_{12}^2\varphi_{1212}}{2\kappa_{11}^2\kappa_{22}^2}.$$

This establishes the following result.

Theorem. *If the observed two processes can be represented by* (2) *with independent, zero mean innovations* $\varepsilon(t)$ *possessing finite fourth moment, then for the matrix* K *in* (3) *the limiting distribution of matrices* $\sqrt{N}(\hat{K} - K)$ *is normal with zero mean and matrix* Φ *in* (4) *satisfying* (6). *In particular,* $N^{1/2}(\hat{\gamma} - \gamma)$ *is asymptotically normal with mean* 0 *and the variance specified in* (7).
   With

$$a = \left(\frac{\kappa_{12}}{2\kappa_{11}\sqrt{\kappa_{11}\kappa_{22}}}, -\frac{1}{2\sqrt{\kappa_{11}\kappa_{22}}}, -\frac{1}{2\sqrt{\kappa_{11}\kappa_{22}}}, \frac{\kappa_{12}}{2\kappa_{22}\sqrt{\kappa_{11}\kappa_{22}}} \right)^T,$$

one can write

$$v^2 = a^T \Phi a.$$

5. Confidence Intervals and Tests of Independence

The formulas derived in this section are more convenient to use in practical situations. Denote by $\mathrm{Vec}(\Sigma)$ the 4×1 vector formed by stacking the columns of the matrix Σ under each other, $\mathrm{Vec}(\Sigma) = \left(\mu_{20}, \mu_{11}, \mu_{11}, \mu_{02} \right)^T$. Then (3) can be rewritten in the form

$$\mathrm{Vec}(K) = \left(\sum_t C_t \otimes C_t \right) \mathrm{Vec}(\Sigma),$$

so that with $Q = \left(\sum_t C_t \otimes C_t \right)^{-1}$ (assuming that this inverse matrix exists),

$$(8) \quad \mathrm{Vec}(\Sigma) = Q\,\mathrm{Vec}(K) = (Q\,\mathrm{Vec}(K)_1, Q\,\mathrm{Vec}(K)_2, Q\,\mathrm{Vec}(K)_2, Q\,\mathrm{Vec}(K)_4)^T.$$

This formula leads to an estimate of the correlation coefficient ρ and to a test of the hypothesis $H_0 \colon \rho = 0$. Indeed, as K is estimable from the observed data by $\hat{K}$, for given matrices C_t, Σ can be estimated from (8). The proof of Theorem gives the asymptotic variance of the estimator $\hat{\rho}$ of the correlation coefficient ρ based on $\hat{K}$,

$$(9) \quad \hat{\rho} = \frac{Q\,\mathrm{Vec}(\hat{K})_2}{[Q\,\mathrm{Vec}(\hat{K})_1 Q\,\mathrm{Vec}(\hat{K})_4]^{1/2}},$$

with

$$\mathrm{Var}(\hat{\rho}) = a^T Q \Phi Q^T a.$$

Here the matrix Φ satisfies (6), which shows that

$$
\begin{aligned}
\mathrm{Vec}(\Phi) &= \mathrm{Vec}(\Phi_1) + \mathrm{Vec}(\Phi_2) + \mathrm{Vec}(\Phi_3) \\
&= \sum_{t,u} \big(C_t \otimes C_t \otimes C_u \otimes C_u\big) \mathrm{Vec}(U) + \sum_{t,u} \big(C_t \otimes C_u \otimes C_u \otimes C_t\big) \mathrm{Vec}(V) \\
&\quad + \sum_{t} \big(C_t \otimes C_t \otimes C_t \otimes C_t\big) \mathrm{Vec}(W - U - V).
\end{aligned}
$$

As Σ is symmetric, all these identities can be written in terms of matrices of size three by using the duplicating matrix and its left inverse (Harville, 1997, Section 16.4 or Magnus and Neudecker, 1999, Section 3.8). More precisely, let $\mathrm{Vech}(\Sigma) = \big(\mu_{20}, \mu_{11}, \mu_{02}\big)^T$. Then for every symmetric matrix A, $\mathrm{Vec}(A) = G\,\mathrm{Vech}(A)$, with the *duplicating* matrix $G = G_2$ of size 4×3. If H is a left inverse of G (its size is 3×4), then (8) means that

$$
\mathrm{Vech}(\Sigma) = \Big[H\Big(\sum_{t} C_t \otimes C_t\Big)G\Big]^{-1} \mathrm{Vech}(K),
$$

and $\mathrm{Vech}(\Phi)$ can be obtained similarly.

In particular, if all matrices C_s are diagonal, $C_s = \mathrm{diag}[\alpha_s, \beta_s]$, γ is proportional to the correlation coefficient ρ since

$$
(10) \qquad \mathrm{Vech}(K) = \Big(\sum_{s} \alpha_s^2 \mu_{20}, \sum_{s} \alpha_s \beta_s \mu_{11}, \sum_{s} \beta_s^2 \mu_{02}\Big)^T,
$$

so that

$$
(11) \qquad \gamma = \frac{\rho \sum_i \alpha_i \beta_i}{\sqrt{\sum_i \alpha_i^2 \sum_i \beta_i^2}}.
$$

Using formulas for U, V, and W in Section 3 one obtains,

$$
\varphi_{1111} = 2\Big(\sum_{t} \alpha_t^2\Big)^2 \mu_{20}^2 + \sum_{t} \alpha_t^4(\mu_{40} - 3\mu_{20}^2),
$$

$$
\varphi_{2222} = 2\Big(\sum_{t} \beta_t^2\Big)^2 \mu_{02}^2 + \sum_{t} \beta_t^4(\mu_{04} - 3\mu_{02}^2),
$$

$$
\varphi_{1112} = 2\sum_{t} \alpha_t^2 \sum_{t} \alpha_t \beta_t \mu_{20}\mu_{11} + \sum_{t} \alpha_t^3 \beta_t(\mu_{31} - 3\mu_{20}\mu_{11}),
$$

$$
\varphi_{1222} = 2\sum_{t} \beta_t^2 \sum_{t} \alpha_t \beta_t \mu_{11}\mu_{02} + \sum_{t} \alpha_t \beta_t^3(\mu_{13} - 3\mu_{11}\mu_{02}),
$$

$$
\varphi_{1212} = 2\Big(\sum_{t} \alpha_t \beta_t\Big)^2 \mu_{11}^2 + \sum_{t} \alpha_t^2 \beta_t^2(\mu_{22} - 2\mu_{11}^2 - \mu_{20}\mu_{02}),
$$

$$
\varphi_{1122} = \sum_{t} \alpha_t^2 \sum_{t} \beta_t^2 \mu_{20}\mu_{02} + \Big(\sum_{t} \alpha_t \beta_t\Big)^2 \mu_{11}^2 + \sum_{t} \alpha_t^2 \beta_t^2(\mu_{22} - 2\mu_{11}^2 - \mu_{20}\mu_{02}).
$$

It follows from (7) that

$$(12) \qquad v^2 = 1 - 2\rho^2 \frac{(\sum_t \alpha_t \beta_t)^2}{\sum_s \alpha_s^2 \sum_s \beta_s^2} + \frac{\rho^4 (\sum_s \alpha_s \beta_s)^4}{(\sum_s \alpha_s^2)^2 (\sum_s \beta_s^2)^2}$$
$$+ \frac{\sum_s \alpha_s^2 \beta_s^2 (\mu_{22} - 2\mu_{11}^2 - \mu_{20}\mu_{02})}{\sum_s \alpha_s^2 \sum_s \beta_s^2 \mu_{20}\mu_{02}}$$
$$- \frac{(\sum_s \alpha_s \beta_s) \sum_t \alpha_t^3 \beta_t (\mu_{31} - 3\mu_{20}\mu_{11})\mu_{11}}{(\sum_s \alpha_s^2)^2 \sum_s \beta_s^2 \mu_{20}^2 \mu_{02}}$$
$$- \frac{(\sum_s \alpha_s \beta_s) \sum_t \alpha_t \beta_t^3 (\mu_{13} - 3\mu_{11}\mu_{02})\mu_{11}}{\sum_s \alpha_s^2 (\sum_s \beta_s^2)^2 \mu_{20}\mu_{02}^2}$$
$$+ \frac{\sum_s \alpha_s^2 \beta_s^2 (\sum_s \alpha_s \beta_s)^2 (\mu_{22} - 2\mu_{11}^2 - \mu_{20}\mu_{02})\mu_{11}^2}{2(\sum_s \alpha_s^2)^2 (\sum_s \beta_s^2)^2 \mu_{20}^2 \mu_{02}^2}$$
$$+ \frac{\sum_s \alpha_s^4 (\sum_s \alpha_s \beta_s)^2 (\mu_{40} - 3\mu_{20}^2)\mu_{11}^2}{(\sum_s \alpha_s^2)^3 \sum_s \beta_s^2 \mu_{20}^2 \mu_{02}}$$
$$+ \frac{\sum_s \beta_s^4 (\sum_s \alpha_s \beta_s)^2 (\mu_{04} - 3\mu_{02}^2)\mu_{11}^2}{\sum_s \alpha_s^2 (\sum_s \beta_s^2)^3 \mu_{20}\mu_{02}^3}.$$

In the normal case

$$(13) \qquad\qquad U + V = W,$$

so that $\Phi_3 = 0$ in (6). In this situation (12) becomes a quadratic polynomial in ρ^2,

$$(14) \qquad\qquad v^2 = \left(1 - r\rho^2\right)^2,$$

where

$$r = \frac{(\sum \alpha_s \beta_s)^2}{\sum \alpha_s^2 \sum \beta_s^2} \le 1.$$

When α's are proportional to β's, $\alpha_i \propto \beta_i$, one gets $r = 1$, and formula (14) is merely

$$(15) \qquad\qquad v^2 = (1 - \rho^2)^2.$$

Provided that $\sum_s \alpha_s \beta_s \ne 0$, one can use (14) to test the null hypothesis $H_0: \rho = 0$ with critical region, $|\hat{\rho}|r^{1/2} > z_{\alpha/2}$, and $\hat{\rho}$ defined by (9). Under this condition,

$$\mathrm{Var}(\hat{\rho}) = r^{-1}\left(1 - r\rho^2\right)^2 = r^{-1} - 2\rho^2 + r\rho^4,$$

so that the quality of such an estimator is characterized by the ratio r (the larger r, the smaller its variance). The smallest possible value of this variance is given in (15).

An (approximate) confidence interval for ρ also can be obtained from (14) in the form of the inequality,

$$\frac{|\hat{\rho} - \rho|}{\sqrt{\mathrm{Var}(\hat{\rho})}} \le z_{\alpha/2}\sqrt{N} = a^{-1},$$

which gives

$$(16) \qquad \frac{a - \sqrt{a^2 + 4(1 - ra\hat\rho)}}{2r} \le \rho \le \frac{\sqrt{a^2 + 4(1 + ra\hat\rho)} - a}{2r}.$$

The right-hand side of (16) has to be replaced by 1 if it exceeds 1, and similarly for the lower bound. Simulations for the models considered further in this section and in Section 5 show that this interval has a shorter length especially for larger ρ than the plug-in confidence procedure $\hat\rho \pm z_{\alpha/2}\sqrt{N}(1 - r\hat\rho^2)/r$.

Proposition 1. *If the random field Z admits the moving average structure (2) with diagonal matrices C_s, $C_s = \mathrm{diag}[\alpha_s, \beta_s]$, $s \in \mathbb{S}$, the limiting distribution of $\sqrt{N}(\hat\gamma - \gamma)$ is normal with mean 0 and variance specified in (12). Under condition (13) the asymptotic variance of $\hat\gamma$ admits representation (14). A $(1 - \alpha)$-confidence interval for ρ can be obtained from (16).*

6. Examples

6.1. Moving Average Models.
Consider a moving average process

$$(17) \qquad Z(s) = \sum_{t\in\mathbb{T}} A_t \varepsilon(s - t),$$

where $\mathbb{T} = \{t\colon t \in \mathbb{Z}^d, A_t \ne 0\}$, $\varepsilon(t)$ are independent random vectors with mean 0 and the covariance matrix Σ. Then

$$Z(s + h) - Z(s) = \sum_{t\in(\mathbb{T}-h)\cap\mathbb{T}} (A_{t+h} - A_t)\varepsilon(s - t) + \sum_{t\in(\mathbb{T}-h)\cap\mathbb{T}^c} A_{t+h}\varepsilon(s - t)$$
$$- \sum_{t\in(\mathbb{T}-h)^c\cap\mathbb{T}} A_t\varepsilon(s - t).$$

It follows that the representation (2) holds with $\mathbb{S} = (\mathbb{T} - h) \cup \mathbb{T}$. If all matrices A_t, $t \in \mathbb{T}$, are diagonal, $A_t = \mathrm{diag}[\xi_t, \psi_t]$, the corresponding matrices C_s are also diagonal, and

$$\sum_s \alpha_s\beta_s = 2\sum_{t\in\mathbb{T}} \xi_t\psi_t - \sum_{t\in(\mathbb{T}-h)\cap\mathbb{T}} (\xi_{t+h}\psi_t + \xi_t\psi_{t+h}),$$
$$\sum_s \alpha_s^2 = 2\left(\sum_{t\in\mathbb{T}} \xi_t^2 - \sum_{t\in(\mathbb{T}-h)\cap\mathbb{T}} \xi_{t+h}\xi_t\right),$$

with a similar formula for $\sum_s \beta_s^2$. Notice that the condition $\xi_t \propto \psi_t$ implies that $\alpha_i \propto \beta_i$, and then the formula (15) is valid.

If $(\mathbb{T} - h) \cap \mathbb{T} = \emptyset$, then

$$Z(s + h) - Z(s) = \sum_{t\in(\mathbb{T}-h)^c\cap\mathbb{T}} A_t[\varepsilon(s + h - t) - \varepsilon(s - t)],$$

and the random variables $\varepsilon'(s) = \varepsilon(s + h) - \varepsilon(s)$ are independent. In this case one can use formulas (11) and (14) with α_s, β_s replaced by ξ_s, ψ_s. For example, if in a representation of a moving average process $Z(s)$ for $d \geq 2$, $\mathbb{T} = \{(0, 0, \cdots, 0), (-1, 0, \cdots, 0), \ldots, (0, 0, \cdots, -1)\}$, then with $h = (1, \ldots, 1)$,

$$
\begin{aligned}
Z(s+h) - Z(s) = \sum_{i=1}^{d} A_i \big[\varepsilon'(s_1 + 1, \ldots, s_i, \ldots, s_d + 1) - \varepsilon'(s_1, \ldots, s_i - 1, \ldots, s_d) \big] \\
+ A_0 \big[\varepsilon'(s_1 + 1, \ldots, s_i + 1, \ldots, s_d + 1) - \varepsilon'(s_1, \ldots, s_i, \ldots, s_d) \big],
\end{aligned}
$$

and (2) holds for the increments of this process. It follows that for diagonal matrices $A_i = \mathrm{diag}[\alpha_i, \beta_i], i = 1, \ldots, d$, γ has the form (11) and the variance of the limiting normal distribution of $\sqrt{N}[\hat{\gamma}(X, Y) - \gamma]$ is given by (14). Haining (1978) advocates the use of models (17) with some matrices A_s in geostatistics.

When $d = 1$, in the classical model of two time series,

$$
X(t) = \phi_1 v(t - 1) + v(t), \qquad Y(t) = \phi_2 \zeta(t - 1) + \zeta(t),
$$

where $|\phi_1|, |\phi_2| < 1$, the errors $\varepsilon(t) = (v(t), \zeta(t))^T, t = 1, \ldots, M$, form i.i.d. random vectors with mean 0 and the covariance matrix Σ. In this case, for $h = 1$,

$$
Z(s + 1) - Z(s) = \varepsilon(s + 1) + (A - I)\varepsilon(s) - A\varepsilon(s - 1)
$$

with $A = \mathrm{diag}[\phi_1, \phi_2]$, so that $\mathbb{T} = \{0, 1\}$ and $\mathbb{S} = \{-1, 0, 1\}$. One has

$$
\sum_s \alpha_s^2 = 1 + \phi_1^2 + (1 - \phi_1)^2,
$$

$$
\sum_s \beta_s^2 = 1 + \phi_2^2 + (1 - \phi_2)^2,
$$

$$
\sum_s \alpha_s \beta_s = 1 + \phi_1 \phi_2 + (1 - \phi_1)(1 - \phi_2),
$$

so that

$$
\gamma = \frac{\rho(2 - \phi_1 - \phi_2 + 2\phi_1\phi_2)}{2\sqrt{(1 - \phi_1 + \phi_1^2)(1 - \phi_2 + \phi_2^2)}}.
$$

Formula (14) shows that

$$
v^2 = \left(1 - \frac{\rho^2(2 - \phi_1 - \phi_2 + 2\phi_1\phi_2)^2}{4(1 - \phi_1 + \phi_1^2)(1 - \phi_2 + \phi_2^2)} \right)^2.
$$

When $h \geq 2$,

$$
Z(s + h) - Z(s) = \varepsilon(s + h) - \varepsilon(s) + A\big[\varepsilon(s + h - 1) - \varepsilon(s - 1)\big],
$$

so that one can take $\mathbb{T} = \mathbb{S} = \{0, 1\}$ with new independent errors, $\varepsilon(s + h) - \varepsilon(s)$, whose covariance matrix is 2Σ. Therefore, for such h,

$$
\gamma = \frac{\rho(1 + \phi_1\phi_2)}{\sqrt{(1 + \phi_1^2)(1 + \phi_2^2)}},
$$

and

$$v^2 = \left(1 - \frac{\rho^2(1 + \phi_1\phi_2)^2}{(1 + \phi_1^2)(1 + \phi_2^2)}\right)^2.$$

It is instructive to compare the related estimators of ρ. Let $\hat{\rho}_1$ be the estimator corresponding to $h = 1$ and $\hat{\rho}_2$ to $h = 2$ as defined by (9). Then

$$\text{Var}(\hat{\rho}_1) = \left(1 - \frac{\rho^2(2 - \phi_1 - \phi_2 + 2\phi_1\phi_2)^2}{4(1 - \phi_1 + \phi_1^2)(1 - \phi_2 + \phi_2^2)}\right)^2 \frac{4(1 - \phi_1 + \phi_1^2)(1 - \phi_2 + \phi_2^2)}{(2 - \phi_1 - \phi_2 + 2\phi_1\phi_2)^2 N},$$

and

$$\text{Var}(\hat{\rho}_2) = \left(1 - \frac{\rho^2(1 + \phi_1\phi_2)^2}{(1 + \phi_1^2)(1 + \phi_2^2)}\right)^2 \frac{(1 + \phi_1^2)(1 + \phi_2^2)}{(1 + \phi_1\phi_2)^2 N}.$$

It follows that $\text{Var}(\hat{\rho}_1) < \text{Var}(\hat{\rho}_2)$ for all ρ if and only if

$$\frac{4(1 - \phi_1 + \phi_1^2)(1 - \phi_2 + \phi_2^2)}{(2 - \phi_1 - \phi_2 + 2\phi_1\phi_2)^2} < \frac{(1 + \phi_1^2)(1 + \phi_2^2)}{(1 + \phi_1\phi_2)^2}.$$

Figure 1 shows that the area of values ϕ_1, ϕ_2, where $\text{Var}(\hat{\rho}_2) < \text{Var}(\hat{\rho}_1)$, almost coincides with the lower diagonal half of the square, $-1 < \phi_1, \phi_2 < 1$. Notice that $\text{Var}(\hat{\rho}_1) = \text{Var}(\hat{\rho}_2)$ when $\phi_1 = \phi_2$.
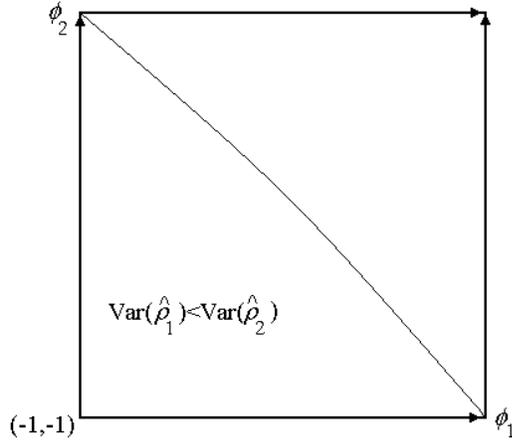

FIGURE 1. The region where $\text{Var}(\hat{\rho}_2) < \text{Var}(\hat{\rho}_1)$ for a moving average process.

6.2. FIRST ORDER AUTOREGRESSIVE AND RELATED MODELS. Here we consider a special case when $d = 2$. Basu and Reinsel (1993) investigated the correlation structure of a general first order autoregressive process of the form,

(18) $X(i, j) = \xi_1 X(i - 1, j) + \xi_2 X(i, j - 1) + \xi_3 X(i - 1, j - 1) + \varepsilon_1(i, j),$

by deriving its stationary representation as

$$X(i,j) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{r=0}^{\infty} \frac{(k+l+r)!}{k!l!r!}\xi_1^k\xi_2^l\xi_3^r\varepsilon_1(i-k-r,j-l-r)$$

(which is different from (2)). They showed that the conditions, $|\xi_i| < 1$, $1 - \xi_2^2 > |\xi_1 + \xi_2\xi_3|$, $(1 + \xi_1^2 - \xi_2^2 - \xi_3)^2 > 4(\xi_1 + \xi_2\xi_3)^2$, guarantee convergence of the triple series.

In the particular case $\xi_3 = -\xi_1\xi_2$, this multiplicative (separable) process simplifies to

$$X(i,j) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\xi_1^k\xi_2^l\varepsilon_1(i-k,j-l).$$

This model has been investigated by Martin (1979), (1990) who argued that it has many practical uses. If a two-dimensional process has the corresponding form,

$$Z(i,j) = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}A_1^k A_2^l \varepsilon(i-k,j-l)$$

with diagonal matrices $A_\ell = \text{diag}[\xi_\ell, \psi_\ell]$, $\ell = 1, 2$, and mean zero i.i.d. innovations $\varepsilon(i,j)$ whose covariance matrix is given by Σ, then with $h = (h_1, h_2)$, $h_1, h_2 \geq 0$,

$$\gamma = \frac{\rho[1 - (\xi_1^{h_1}\xi_2^{h_2} + \psi_1^{h_1}\psi_2^{h_2})/2]\sqrt{(1-\xi_1^2)(1-\xi_2^2)(1-\psi_1^2)(1-\psi_2^2)}}{(1-\xi_1\psi_1)(1-\xi_2\psi_2)\sqrt{(1-\xi_1^{h_1}\xi_2^{h_2})(1-\psi_1^{h_1}\psi_2^{h_2})}}.$$

This formula demonstrates that for positive ξ_ℓ, ψ_ℓ, $\ell = 1, 2$, the optimal value of h minimizing (14) is $h = (1, 0)$, if

$$\frac{1 - (\xi_1 + \psi_1)/2}{\sqrt{(1-\xi_1)(1-\psi_1)}} > \frac{1 - (\xi_2 + \psi_2)/2}{\sqrt{(1-\xi_2)(1-\psi_2)}}$$

or $h = (0, 1)$, otherwise.

We also look at the models (18), when $\xi_3 = 0$, which were studied earlier by Whittle (1954) and Besag (1972). Then

$$X(i,j) = \xi_1 X(i-1,j) + \xi_2 X(i,j-1) + \varepsilon_1(i,j),$$
$$Y(i,j) = \psi_1 Y(i-1,j) + \psi_2 Y(i,j-1) + \varepsilon_2(i,j),$$

where $\varepsilon(t) = (\varepsilon_1(t), \varepsilon_2(t))^T$ form i.i.d. random vectors with mean 0 and covariance matrix Σ. Under conditions mentioned above on the parameters, $Z(i,j) = (X(i,j), Y(i,j))^T$ has the convergent representation (17) with $\mathbb{T} = \mathbb{Z}_+^2$, and

$$A_t = A_{kl} = \binom{k+l}{k}\begin{pmatrix} \xi_1^k\xi_2^l & 0 \\ 0 & \psi_1^k\psi_2^l \end{pmatrix}.$$

By setting for $h = (h_1, h_2)$ and x, y such that $1 + x^4 + y^4 - 2x^2 - 2y^2 - 2x^2y^2 > 0$,

$$D_h(x, y) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \frac{(k+l)!(k+l+h_1+h_2)!}{k!l!(k+h_1)!(l+h_2)!} x^{2k+h_1} y^{2l+h_2},$$

we get from formulas in Section 5

$$\sum_s \alpha_s \beta_s = 2D_0\big(\sqrt{\xi_1 \xi_2},\, \sqrt{\psi_1 \psi_2}\big)$$

$$- D_h\big(\sqrt{\xi_1 \xi_2},\, \sqrt{\psi_1 \psi_2}\big) \left[\left(\frac{\xi_1}{\xi_2}\right)^{h_1/2} \left(\frac{\psi_1}{\psi_2}\right)^{h_2/2} + \left(\frac{\xi_2}{\xi_1}\right)^{h_1/2} \left(\frac{\psi_2}{\psi_1}\right)^{h_2/2} \right],$$

$$\sum_s \alpha_s^2 = 2\big[D_0(\xi_1, \psi_1) - D_h(\xi_1, \psi_1)\big], \qquad \sum_s \beta_s^2 = 2\big[D_0(\xi_2, \psi_2) - D_h(\xi_2, \psi_2)\big].$$

The form of the association coefficient γ follows. Its numerical evaluation is facilitated by noticing that

$$D_h(x, y) = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{e^{-i(\omega_1 h_1 + \omega_2 h_2)}\, d\omega_1\, d\omega_2}{|1 - e^{-i\omega_1} x - e^{-i\omega_2} y|^2}.$$

In particular,

$$D_0(x, y) = \frac{1}{(2\pi)^2} \int_0^{2\pi} \int_0^{2\pi} \frac{d\omega_1\, d\omega_2}{|1 - e^{-i\omega_1} x - e^{-i\omega_2} y|^2}$$

$$= \frac{1}{\sqrt{1 + x^4 + y^4 - 2x^2 - 2y^2 - 2x^2 y^2}},$$

$$D_{(h,0)}(x, y) = D_0(x, y) \left[\frac{1 + x^4 - y^4 - \sqrt{1 + x^4 + y^4 - 2x^2 - 2y^2 - 2x^2 y^2}}{2x} \right]^h,$$

and a similar formula holds for $D_{(0,h)}(x, y)$.

For example, when $\xi_1 = \psi_1$, $\xi_2 = \psi_2$, $\mathrm{Var}(\hat{\rho}_{(h,0)}) = \mathrm{Var}(\hat{\rho}_{(0,h)})$ for all h. The region, where $\mathrm{Var}(\hat{\rho}_{(1,0)}) < \mathrm{Var}(\hat{\rho}_{(2,0)})$, is depicted in Figure 2.

When $d = 1$,

$$X(t) = \phi_1 X(t-1) + \varepsilon_1(t), \qquad Y(t) = \phi_2 Y(t-1) + \varepsilon_2(t),$$

are two classical AR(1) processes. We assume that $|\phi_1|, |\phi_2| < 1$ and the error vectors $\varepsilon(t) = \big(\varepsilon_1(t), \varepsilon_2(t)\big)^T$, $t = 1, \ldots, N$, are i.i.d. random vectors with mean 0 and covariance matrix Σ, so that $Z(i, j) = (X(i, j), Y(i, j))^T$ has the form

$$Z(i, j) = \sum_{k=0}^{\infty} A^k \varepsilon(i - k)$$

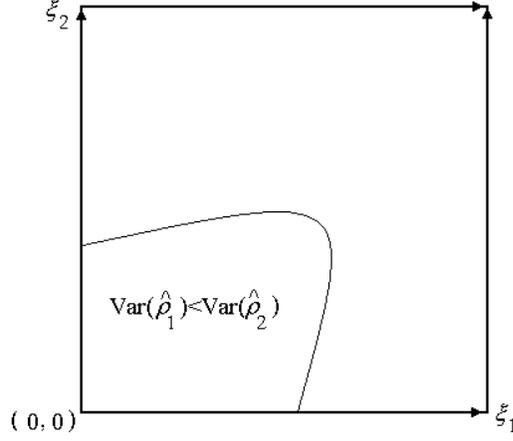with a diagonal matrix $A = \mathrm{diag}[\phi_1, \phi_2]$.

FIGURE 2. The region of $\xi_1, \xi_2 \geq 0$, where $\mathrm{Var}(\hat\rho_2) < \mathrm{Var}(\hat\rho_1)$ in the model (18) when $\xi_3 = 0$.

The asymptotic variance of the normalized sum of successive differences,

$$\sqrt{N}\left[N^{-1}\sum_{t=1}^{N-1}\left(X(t+1)-X(t)\right)^2 - 2\mu_{20}\right]$$

for an arbitrary error distribution is equal to

$$\varphi_{1111} = \frac{8\mu_{20}^2}{(1+\phi_1)^2} + \frac{(1-\phi_1)^3(\mu_{40}-3\mu_{20}^2)}{(1+\phi_1)(1+\phi_1^2)}.$$

The mean square successive difference, $\sum_{t=1}^{N-1}[X(t+1)-X(t)]^2/(N-1)$, was introduced by von Neumann in the forties. It is known that the variance of this statistic when both sequences $X(t)$ and $Y(t)$ are i.i.d. normal random variables, $\phi_1 = \phi_2 = 0$, is $8N\mu_{20}^2/(N-1)^2$. This agrees with our formula as then $\varphi_{1111} = 8\mu_{20}^2$.

The asymptotic distribution of $\sum_t(X(t+1)-X(t))(Y(t+1)-Y(t))/\sqrt{N}$ is normal with zero mean and variance

$$\varphi_{1122} = \frac{4\mu_{20}\mu_{02}}{(1+\phi_1)(1+\phi_2)} + \frac{(2-\phi_1-\phi_2+2\phi_1\phi_2)^2\mu_{11}^2}{(1-\phi_1\phi_2)^2}$$
$$+ \frac{(2-\phi_1^2-\phi_2^2+2\phi_1^2\phi_2^2)(\mu_{22}-\mu_{20}\mu_{02}-2\mu_{11}^2)}{1-\phi_1^2\phi_2^2}.$$

When the normal errors are independent, this variance becomes

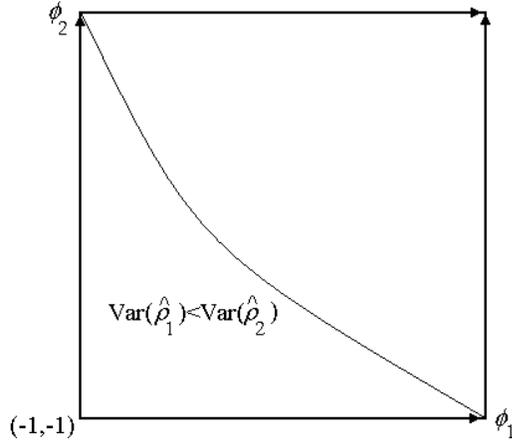$$\varphi_{1122} = \frac{4\mu_{20}\mu_{02}}{(1+\phi_1)(1+\phi_2)}.$$

FIGURE 3. The region where $\text{Var}(\hat{\rho}_2) < \text{Var}(\hat{\rho}_1)$ in the ARMA model.

For $h \geq 1$, in the representation (2) $C_k = A^{h+k}$, $-h \leq k \leq -1$, $C_k = (A^h - \mathbf{I})A^k$, $k \geq 0$, so that

$$\text{Vech}(K) = \left(\frac{2\mu_{20}(1-\phi_1^h)}{1-\phi_1^2}, \frac{\mu_{11}(2-\phi_1^h-\phi_2^h)}{1-\phi_1\phi_2}, \frac{2\mu_{02}(1-\phi_2^h)}{1-\phi_2^2} \right).$$

In particular, when $h = 1$,

$$\gamma(1) = \frac{\rho(2-\phi_1-\phi_2)\sqrt{(1+\phi_1)(1+\phi_2)}}{2(1-\phi_1\phi_2)}$$

and when $h = 2$,

$$\gamma(2) = \frac{\rho(2-\phi_1^2-\phi_2^2)}{2(1-\phi_1\phi_2)}.$$

The limiting distribution of $\sqrt{N}[\hat{\gamma} - \gamma(h)]$ is normal with mean 0. Its variance according to (14) has the form

$$v_h^2 = \left[1 - \frac{\rho^2(1-(\phi_1^h+\phi_2^h)/2)^2(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)^2(1-\phi_1^h)(1-\phi_2^h)} \right]^2.$$

The comparison of the corresponding estimators (9) of ρ, $\hat{\rho}_h$, can be performed as in the case $d = 2$ treated above. The optimal value of h corresponds to the largest ratio, $(2-\phi_1^h-\phi_2^h)^2/[(1-\phi_1^h)(1-\phi_2^h)]$. In particular, $\text{Var}(\hat{\rho}_1) < \text{Var}(\hat{\rho}_2)$ if and only if $\phi_1^2 + \phi_1\phi_2 + \phi_2^2 - \phi_1 - \phi_2 < 1$. Figure 3 shows the region where $\text{Var}(\hat{\rho}_2) < \text{Var}(\hat{\rho}_1)$. For larger values of h this region quickly shrinks to the corner $\phi_1 = \phi_2 = -1$.

7. Application: Similarity of Carbon Nanotubes Optical Images

To illustrate a practical application of the association coefficient an example related to flammability properties of polymers is presented here. It is known that one
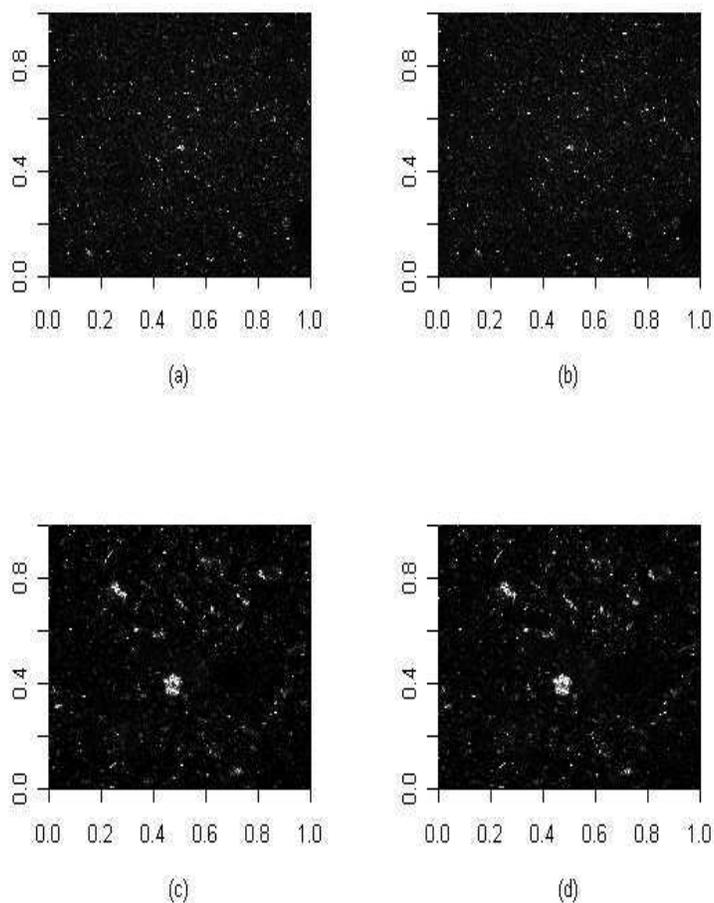
Figure 4. Four images of carbon nanotubes taken at different distances.

weak aspect of polymers is that they are combustible under certain conditions. The flame retardant of clay-polymer nanocomposites has been demonstrated to improve physical and flammability properties of polymers by Kashiwagi *et al.* (2005). To investigate the effects of the dispersion and concentration of one of these retardants, the so-called single-walled carbon nanotube, on the flammability of polymers, these authors examined by optical microscopy the distribution of this nanotube. This distribution is believed to depend mainly on the distance from the top surface to the location in the polymer matrix (polymethyl methacrylate). In the study it was important to characterize the dispersion and the concentration of the flame

retardant after an image of the polymer. In other words it was desirable to define a metric for closeness of two sample polymers in terms of their image characteristics. Figure 4 shows four images taken at the National Institute of Standards and Technology (NIST) kindly provided by T. Kashiwagi.

Samples (a) and (b) were taken at the same distance from the matrix with similar concentration of the flame retardant; samples (c) and (d) also with almost identical concentration were taken at the same distance from the matrix but different than images (a) and (b). In this case it is of interest to have an association characteristic to capture similarity between images that were taken at the same distance (and dissimilarity of images taken at different distances like (a) and (c)). The association coefficient ρ_{XY} suggested in this paper proved to be a useful tool to address this problem.

The available data set was represented as 512×512 images indexed by the greyness level taking 256 values shown in Figure 4. In this case $\hat{\rho}_{ab}(1,0) = 0.804$, $\hat{\rho}_{ab}(1,1) = 0.841$, $\hat{\rho}_{ab}(0,1) = 0.807$. The same patterns are observed for the association coefficient between images (c) and (d) $\hat{\rho}_{cd}(1,0) = 0.815$, $\hat{\rho}_{ab}(1,1) = 0.865$, $\hat{\rho}_{ab}(0,1) = 0.794$. However $\hat{\rho}_{ac}(1,0) = 0.002$, $\hat{\rho}_{ac}(1,1) = 0.003$, $\hat{\rho}_{ac}(1,0) = 0.001$. Moreover, when both lags were used in the computation of the coefficient, stronger evidence of similarity was found. For example, $\hat{\rho}_{ac}(2,2) = 0.875$. The association coefficient convincingly captures similarity between images (a) and (b), and dissimilarity between images (a) and (c).

8. Conclusions and Acknowledgments

The paper gives very explicit formulas for the mean and the variance of a sample association coefficient whose definition is based on that of a cross-variogram. Derived formulas allow for the optimal choice of the lag h in popular spatial and temporal models. These results lead to tests of independence and provide confidence intervals for the correlation coefficient in these models.

References

[1] S. Banerjee, B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Chapman&Hall CRC, London.

[2] S. Basu and G. C. Reinsel (1993), *Properties of spatial unilateral first-order ARMA model*, Adv. Appl. Probab., **25**, 631–648.

[3] J. E. Besag (1972), *On the correlation structure of some two-dimensional stationary process*, Biometrika, **59**, 43–48.

[4] R. Bolthausen (1982), *On the central limit theorem for stationary random fields*, Ann. Probab., **10**, 1047–1050.

[5] A. Cliff and J. Ord (1981), *Spatial Processes: Models and Applications*, Pion Ltd., London.

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

[7] N. Cressie (1993), *Statistics for Spatial Data*, Wiley, New York.

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

[9] R. P. Haining (1978), *The moving average model for spatial interaction*, Trans. Inst. British Geographers, **3**, 202–225.

[10] D. A. Harville (1997), *Matrix Algebra from a Statistician's Perspective*, Springer, New York.

[11] L. J. Hubert and R. G. Golledge (1982), *Measuring association between spatially defined variables: Tjostheim index and some extensions*, Geographical Analysis, **14**, 273–278.

[12] L. J. Hubert, R. G. Golledge, C. Constanzo, and N. Gale (1985), *Measuring association between spatially defined variables: an alternative approach*, Geographical Analysis, **17**, 36–46.

[13] T. Kashiwagi, F. Du, K. I. Winey, K. M. Groth, J. R. Shields, S. P. Bellayer, H. Kim, and J. F. Douglas (2005), *Flammability properties of polymer nanocomposites with single-walled carbon nanotubes: effects of nanotube dispersion and concentration*, Polymer, **46**, 471–481.

[14] S. Leigh (1996), personal communication.

[15] Yu. V. Linnik and A. P. Husu (1974), *Probabilistic methods of evaluation of surfaces roughness* (in Russian). In: *Probabilistic and Statistical Foundations of Processes of Grinding and Finishing*, Leningrad, pp. 7–12.

[16] J. R. Magnus and H. Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York.

[17] R. J. Martin (1979), *A subclass of lattice processes applied to a problem in planar sampling*, Biometrika, **66**, 209–217.

[18] R. J. Martin (1990), *The use of time-series models and methods in the analysis of agricultural trials*, Commun. Statist. — Theory Methods, **A19**, 55–81.

[19] M. Moore (1988), *Spatial linear processes*, Commun. Statist. — Stochastic Models, **4**, 45–75.

[20] D. Myers (1982), *Matrix formulation of co-kriging*, Mathematical Geology, **14**, 249–257.

[21] K. Ord (1975), *Estimation methods for models of spatial interaction*, J. Amer. Statist. Assoc., **70**, 120–126.

[22] S. Richardson and P. Clifford (1998), *Testing association between spatial processes*. In: *Spatial Statistics and Imaging*, A. Possolo, ed., IMS Lecture Notes, Vol. 20, pp. 295–308

[23] B. Ripley (1993), *Spatial Statistics*, Wiley, New York.

[24] D. Tjostheim (1978a), *A measure of association for spatial variables*, Biometrika, **65**, 109–114.

[25] D. Tjostheim (1978b), *Statistical spatial series modelling*, Adv. Appl. Probab., **10**, 130–154.

[26] J. Ver Hoef and R. Barry (1998), *Constructing and fitting models for cokriging and multivariate spatial prediction*, J. Statist. Plann. Inf., **69**, 275–294.

[27] J. Ver Hoef and N. Cressie (1993), *Multivariate spatial prediction*, Mathematical Geology, **25**, 219–240.

[28] E. P. Whitenton and D. E. Deckman (1998), *Measuring matching wear scars on balls and flats*, Surface Topography, **2**, 205–222.

[29] P. Whittle (1954), *On stationary processes in the plane*, Biometrika, **41**, 434–449.

[30] M. C. K. Yang and J. F. Schreckengost (1981), *Difference sign test for comovements between two time series*, Commun. Statist. — Theory Methods, **A10**, 355–369.