

## Estimation of Linear Correlation Coefficient of two Correlated Spatial Processes

F. Pascual

*Washington State University, Pullman, USA*

H. Zhang

*Washington State University, Pullman, USA*

---

### Abstract

Standard inferential methods for the correlation coefficient for two normally distributed variables are based on the assumption of independent sampling. In general, this assumption is not appropriate for spatial data where contiguous locations exhibit some significant level of spatial-correlation or autocorrelation. Past works have demonstrated the dangers, in the form of underestimated standard errors and inflated Type I errors, of applying independent-sample inference methods on spatially correlated data. Here, we study the distribution of maximum likelihood estimators in a bivariate spatial model and give the finite sample distribution and asymptotic distribution of estimators. We show that if the estimator appropriately accounts for spatial correlation, some existing distributional results derived under the assumption of independence are still valid.

*AMS (2000) subject classification.* Primary 62M30, 60G15, 86A32.

*Keywords and phrases.* Cross-correlation, exponential covariogram, geostatistics, infill asymptotics, multivariate covariogram.

---

### 1 Introduction

*1.1. Motivation.* In many disciplines such as environmental and agricultural sciences, it is common practice to collect data on multiple spatial variables. These data are often spatially correlated and this spatial correlation is one of the major research areas in spatial statistics. While most works in spatial statistics deal with spatial correlation of one variable, researchers are often interested in how one variable correlates with another at the same (linear correlation) or two different spatial locations (spatial correlation). For example, if a state county experiences inclement weather, the neighboring counties are expected to experience relatively similar conditions. Thus,

measurements on, say, crop yields and rainfall amounts from these counties will exhibit some significant degree of positive correlation.

In this paper, we study the estimation of the correlation coefficient of two variables in the presence of spatial correlation. Our data consist of samples of the bivariate Gaussian stationary process  $[Y_1(\mathbf{s}), Y_2(\mathbf{s})]'$  measured at locations  $\mathbf{s}_i, i = 1, \dots, n$ . The linear correlation coefficient between the two spatial process, denoted by  $r$ , is the correlation coefficient between  $Y_1(\mathbf{s})$  and  $Y_2(\mathbf{s})$  at any location  $\mathbf{s}$ . Standard inference methods for  $r$  have been derived under the assumption that  $[Y_1(\mathbf{s}_i), Y_2(\mathbf{s}_i)]', i = 1, \dots, n$  are independent pairs, i.e., there are no spatial correlations. Clearly, when cross-correlations do exist as in many applications, standard inferential methods may be inappropriate and could result in incorrect and possibly dangerous conclusions. However, our results show that some of these inferential methods may still hold if a cross-correlation structure is incorporated in the model.

*1.2. Related work.* The sample correlation is a natural estimator of  $r$  and has been extensively studied. It is defined as

$$\hat{r}_0 = \hat{\sigma}_{12} / \sqrt{\hat{\sigma}_{11}\hat{\sigma}_{22}}, \quad (1.1)$$

where

$$\hat{\sigma}_{ij} = (1/n) \sum_{k=1}^n [Y_i(\mathbf{s}_k) - \bar{Y}_i][Y_j(\mathbf{s}_k) - \bar{Y}_j], \quad \bar{Y}_i = (1/n) \sum_{k=1}^n Y_i(\mathbf{s}_k) \quad (1.2)$$

for  $i, j = 1, 2$ .

Graybill (1976, Section 3.8 and Chapter 11) discussed the properties of  $\hat{r}_0$  for the normal case when data consist of independent pairs. In this particular case, a useful result has been established by Fisher (1915, 1921) who used a geometrical argument to derive the finite sample and asymptotic distributions of  $\hat{r}_0$ . He suggested a hyperbolic arctangent transformation of  $\hat{r}_0$  to achieve a faster convergence to normality. Kendall and Stuart (1979) provided more details and also discussed the joint distribution of all the ML estimators through the geometrical approach.

When the pairs are dependent, the properties of the sample correlation also have received attention from past research work. One of the earliest works is by Student (1914) who showed how to eliminate additional variability due to the correlation between positions in space or time. Bivand (1980) provided a review of previous works about the dangers of standard inference

methods (based on independent pairs) for the correlation coefficient when measurements were actually correlated. He conducted a Monte Carlo study with the bivariate normal distribution to show that the standard errors of estimators were underestimated and that the bias depended on the regularity of the lattice over which data were collected. This underestimation resulted in lower actual coverage probabilities for confidence intervals or, equivalently, inflated Type I errors in hypothesis tests. The author did not suggest ways to address these issues. Clifford et al. (1989), Dutilleul (1993) and Clifford et al. (1993) provided references on more recent works on the estimation of the variance of the sample covariance when data are spatially correlated. Clifford et al. (1989) discussed a procedure for testing the hypothesis of no correlation (i.e.  $H_0 : r = 0$ ) between two spatially correlated normal processes. To describe the cross-correlation between locations, they divided the locations into strata so that the covariances between the same property measurements were constant within the same stratum. They concluded that using the usual linear correlation to test associations between two normal processes when spatial correlations are present would lead to underestimation of the variance of the sample covariance and inflated Type I errors. Dutilleul (1993) provided the theoretical arguments for the empirical observations of Clifford et al. (1989).

*1.3. Outline.* In this article, we focus on the estimation of the linear correlation  $r$  between two measured variables under the existence of spatial correlations by employing an explicit model for bivariate spatial variables. We introduce this model in Section 2. In Section 3, we consider fitting the model to data using ML methods and provide explicit expressions for ML estimation, which lead to an efficient algorithm. In Section 4, we discuss distributional properties of ML estimators and in particular, that of the ML estimator of  $r$ . We then present simulation studies in Section 5 that show how well the large sample properties work for finite samples. These studies also show that ignoring spatial correlations when they actually exist can cause serious problems in inferences for  $r$ .

## 2 A Statistical Model for Bivariate Spatial Processes

Suppose that there are two spatial variables of interest, and let the random variable  $Y_i(\mathbf{s})$  represent the  $i^{\text{th}}$  variate at location  $\mathbf{s} \in \mathbb{R}^d$  for  $i = 1, 2$ . Let  $\mathbf{Y}(\mathbf{s}) = [Y_1(\mathbf{s}), Y_2(\mathbf{s})]'$  be a bivariate Gaussian stationary process so that

$$E[Y_i(\mathbf{s})] = \mu_i, \quad \text{Cov}[Y_i(\mathbf{s}), Y_j(\mathbf{s} + \mathbf{h})] = c_{ij}(\mathbf{h})$$

for all  $\mathbf{s}, \mathbf{h} \in \mathbb{R}^d$  and  $i, j = 1, 2$ . In geostatistical terms, the scalar  $c_{ij}(\cdot)$  is referred to as the direct covariogram when  $i = j$  and the cross covariogram when  $i \neq j$ . The matrix-valued function  $\mathbf{C}(\mathbf{h}) = [c_{ij}(\mathbf{h})]_{i,j=1,2}^2$  is called the multivariate covariogram. If  $\mathbf{C}(\mathbf{h})$  depends only on the Euclidean distance  $\|\mathbf{h}\|$  between  $\mathbf{s}$  and  $\mathbf{s} + \mathbf{h}$ , then the process  $\mathbf{Y}(\cdot)$  is said to be isotropic. The distribution of the bivariate Gaussian process  $\mathbf{Y}(\mathbf{s})$  is then fully specified by the means  $\mu_i$  and the multivariate covariogram  $\mathbf{C}(\mathbf{h})$ .

In order for a matrix-valued function  $\mathbf{C}(\cdot)$  to be a valid multivariate covariogram, it must satisfy some necessary conditions. For instance, it must be positive definite in the sense that for any spatial locations  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$  and vectors  $\mathbf{a}_i \in \mathbb{R}^2$ ,  $i = 1, \dots, n$ ,

$$\text{Var} \left[ \sum_{i=1}^n \mathbf{a}_i' \mathbf{Y}(\mathbf{s}_i) \right] = \sum_{i,j=1}^n \mathbf{a}_i' \mathbf{C}(\mathbf{s}_i - \mathbf{s}_j) \mathbf{a}_j \geq 0.$$

It is not a simple task to identify a valid multivariate covariogram that is relatively easy to estimate and, at the same time, provide an adequate description of the spatial correlations. Only a handful of multivariate covariograms have been proposed and used to analyze multivariate spatial data.

The simplest form of a covariogram is the intrinsic correlation model given by

$$\mathbf{C}(\mathbf{h}) = \rho(\mathbf{h}) \mathbf{V}$$

where  $\mathbf{V}$  is a  $(2 \times 2)$  positive definite matrix and the scalar  $\rho(\cdot)$  is a correlation function, also known as a correlogram, in  $\mathbb{R}^d$ . See Wackernagel (1998, Chapter 23). Chilés and Delfiner (1999) also refer to this as the proportional covariogram. The correlation function  $\rho(\mathbf{h})$  usually depends on a parameter  $\psi$ , i.e.,  $\rho(\mathbf{h}) = \rho(\mathbf{h}; \psi)$ . For example, the exponential isotropic correlation function takes the form

$$\rho(\mathbf{h}; \psi) = \exp(-\|\mathbf{h}\|/\psi),$$

where  $\psi > 0$ . The proportional covariogram assumes that the two direct covariograms differ only by a constant multiplicative factor and hence appears to be restrictive to model bivariate or multivariate spatial processes. Nonetheless, this model has been applied to real problems successfully. For example, Banerjee and Gelfand (2002) applied this model to an ecological dataset. They developed a fully Bayesian methodology to obtain spatial association and prediction. However, they did not explicitly consider estimation of the correlation coefficient.

The proportional model can also be used as building blocks for more complex models. We refer readers to Wackernagel (1998), Chilés and Delfiner (1999) and Banerjee, Carlin and Gelfand (2004) for more information.

In this present work, we assume that the two Gaussian processes have the bivariate proportional covariogram and are both observed at  $n$  locations given by  $\mathbf{s}_1, \dots, \mathbf{s}_n$ . Group together variable measurements by defining

$$\mathbf{U}_i = [Y_i(\mathbf{s}_1), \dots, Y_i(\mathbf{s}_n)]' \text{ for } i = 1, 2, \quad \mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2)'. \quad (2.1)$$

Let  $\otimes$  stand for the Kronecker product. Then

$$\mathbf{U} \sim MVN(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{V} \otimes \boldsymbol{\Xi}), \quad (2.2)$$

where  $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ ,  $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\psi) = [\rho(\mathbf{s}_i - \mathbf{s}_j; \psi)]_{i,j=1}^n$  is the common correlation matrix of  $\mathbf{U}_i$ ,  $i = 1, 2$ , and  $\mathbf{V} = (\sigma_{ij})_{i,j=1}^2$  is the  $2 \times 2$  covariance matrix of  $\mathbf{Y}(\mathbf{s}_i)$  for any  $\mathbf{s}_i$ . Hereafter, we refer to this model on  $\mathbf{U}$  as the cross-correlation model. The model parameter vector is thus  $\boldsymbol{\theta} = (\psi, \mu_1, \mu_2, \sigma_{11}, \sigma_{22}, \sigma_{12})'$ .

### 3 Maximum Likelihood Estimation

In this section, we discuss the ML estimation of the cross-correlation model defined in the previous section. The log likelihood function is given by, apart from an additive constant,

$$\log L(\boldsymbol{\theta}, \mathbf{U}) = -(1/2) \log(|\boldsymbol{\Sigma}|) - (1/2)(\mathbf{U} - \boldsymbol{\mu} \otimes \mathbf{1}_n)' \boldsymbol{\Sigma}^{-1} (\mathbf{U} - \boldsymbol{\mu} \otimes \mathbf{1}_n), \quad (3.1)$$

where  $\boldsymbol{\Sigma} = \mathbf{V} \otimes \boldsymbol{\Xi}(\psi)$ . Denote the ML estimator by  $\hat{\boldsymbol{\theta}} = (\hat{\psi}, \hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_{11}, \hat{\sigma}_{22}, \hat{\sigma}_{12})$ , which maximizes the likelihood. We first give some explicit functional relationship between the ML estimators of different parameters. Let  $v_{ij}$  be the  $(i, j)$ <sup>th</sup> element of  $\mathbf{V}^{-1}$ . Then we can write (3.1) as

$$\begin{aligned} \log L(\boldsymbol{\theta}, \mathbf{U}) &= (n/2) \log(|V^{-1}|) - (p/2) \log(|\boldsymbol{\Xi}|) \\ &\quad - (1/2) \sum_{i,j=1}^2 v_{ij} (\mathbf{U}_i - \mu_i \mathbf{1}_n)' [\boldsymbol{\Xi}(\psi)]^{-1} (\mathbf{U}_j - \mu_j \mathbf{1}_n). \end{aligned}$$

Since

$$\frac{\partial \log |V^{-1}|}{\partial v_{ij}} = \text{tr} \left[ V \frac{\partial V^{-1}}{\partial v_{ij}} \right] = 2d_{ij} \sigma_{ij},$$

then

$$\frac{\partial \log L}{\partial v_{ij}} = d_{ij} \{ n\sigma_{ij} - (\mathbf{U}_i - \mu_i \mathbf{1}_n)' [\Xi(\psi)]^{-1} (\mathbf{U}_j - \mu_j \mathbf{1}_n) \}, \quad (3.2)$$

where  $d_{ij} = 1$  if  $i \neq j$  and  $d_{ij} = 1/2$  if  $i = j$ . Letting (3.2) be equal to 0, we see that for any fixed  $\psi$ ,  $L$  as a function of  $\mu_i$  and  $\sigma_{ij}$  is maximized at

$$\begin{aligned} \hat{\mu}_i(\psi) &= \frac{\mathbf{1}'_n [\Xi(\psi)]^{-1} \mathbf{U}_i}{\mathbf{1}'_n [\Xi(\psi)]^{-1} \mathbf{1}_n}, \\ \hat{\sigma}_{ij}(\psi) &= (1/n) (\mathbf{U}_i - \hat{\mu}_i(\psi) \mathbf{1}_n)' [\Xi(\psi)]^{-1} (\mathbf{U}_j - \hat{\mu}_j(\psi) \mathbf{1}_n). \end{aligned} \quad (3.3)$$

If  $\psi$  is known, the ML estimator of  $r$  is then

$$\hat{r}(\psi) = \frac{\hat{\sigma}_{12}(\psi)}{\sqrt{\hat{\sigma}_{11}(\psi) \hat{\sigma}_{22}(\psi)}}. \quad (3.4)$$

Let  $\hat{\boldsymbol{\mu}}(\psi) = [\hat{\mu}_1(\psi), \hat{\mu}_2(\psi)]'$  and  $\hat{\mathbf{V}}(\psi) = [\hat{\sigma}_{ij}(\psi)]_{i,j=1}^2$ . It is clear from (3.3) that for any  $\psi$ , the estimate matrix  $\hat{\mathbf{V}}$  is non-negative definite. If in (3.3) and (3.4),  $\psi$  is replaced by its ML estimator  $\hat{\psi}$ , then the resultant estimators  $\hat{\mu}_i(\hat{\psi})$ ,  $\hat{\sigma}_{ij}(\hat{\psi})$  and  $\hat{r}(\hat{\psi})$  are the ML estimators of  $\mu_i$ ,  $\sigma_{ij}$  and  $r$ , respectively. To obtain  $\hat{\psi}$ , we can maximize the profile log likelihood for  $\psi$  given by

$$PLL(\psi) = \sup_{\mathbf{V}, \boldsymbol{\mu}} \log L(\boldsymbol{\theta}, \mathbf{U}) \quad (3.5)$$

over  $\psi > 0$ . Clearly,  $PLL(\psi)$  is maximum when  $\mu_i = \hat{\mu}_i(\psi)$  and  $\sigma_{ij} = \hat{\sigma}_{ij}(\psi)$  for  $i, j = 1, 2$ , and the closed form of  $PLL(\psi)$  is obtained by replacing  $\boldsymbol{\Sigma}$  with  $\hat{\mathbf{V}} \otimes \Xi$  and  $\boldsymbol{\mu}$  with  $\hat{\boldsymbol{\mu}}$  in (3.1). Simple calculation yields

$$PPL(\psi) = -(n/2) \log [\hat{\mathbf{V}}(\psi)] - (p/2) \log [|\Xi(\psi)|] - (n/2).$$

We apply the Newton-Raphson algorithm on  $PPL(\psi)$  to derive the ML estimate  $\hat{\psi}$ , and then calculate  $\hat{\sigma}_{ij}$  and  $\hat{\mu}_i$  using (3.3). This strategy is an efficient way of computing ML estimates because optimization is performed over  $\psi > 0$  instead of the full six-dimensional parameter space.

#### 4 Properties of Maximum Likelihood Estimators

In this section, we provide finite-sample properties as well as asymptotic properties of ML estimators for the cross-correlation model assuming that

the correlation matrix  $\Xi(\psi)$  is known. In particular, we establish the asymptotic distribution of the ML estimator of correlation coefficient  $r$ . While the correlation matrix has to be estimated in practice, the assumption that this correlation matrix is known avoids some difficulties in the asymptotic theory, namely, possible inconsistency of estimators when the correlation function is estimated. In the next subsection, we add more details on this.

*4.1. Inconsistency of estimators.* There are two different asymptotic frameworks in spatial statistics: infill asymptotic framework where the spatial domain is fixed and more data are collected by sampling more densely, and increasing domain asymptotic framework where the minimum distance between two sampling points is bounded away from 0. Asymptotic properties of maximum likelihood estimators are quite different under the two asymptotic frameworks. It is known that all parameters are consistently estimable and the ML estimators are asymptotically normal under the increasing domain asymptotic framework (Mardia and Marshall, 1984). However, not all parameters in a covariance function are consistently estimable under the infill asymptotics (Ying, 1991, Zhang, 2004). For example, Zhang (2004) considered a stationary Gaussian process having a Matérn covariogram and showed that the parameters in the covariogram are not consistently estimable. In particular, for the isotropic exponential covariogram  $\sigma^2 \times \exp(-\|\mathbf{h}\|/\psi)$ , where  $\mathbf{h} \in R^d$ ,  $d = 1, 2$  or  $3$ , which is a special case of the Matérn class, neither  $\sigma^2$  nor  $\psi$  is consistently estimable under the infill asymptotic framework. The simulation results of Zhang (2004) show that infill asymptotics can appropriately explain some finite sample properties that are hard to explain by the increasing domain asymptotics. Zhang and Zimmerman (2005) further compared the two asymptotics frameworks.

Ying (1991) showed that the ratio  $\sigma^2/\psi$  is consistently estimable for a univariate Gaussian process having an isotropic covariogram  $\sigma^2 \exp(-h/\psi)$  under the infill asymptotic framework. In addition, consistent and asymptotically normally distributed estimators of this ratio can be constructed by fixing the parameter  $\psi$  at an arbitrary value, and the choice of the fixed value does not affect the asymptotic efficiency of the estimator of the ratio. Zhang (2004) showed that this ratio is more important to interpolation than the individual parameters  $\sigma^2$  and  $\psi$ .

On the other hand, for univariate Gaussian process with a known mean, if the correlation function is known, there is no difference between the two asymptotic frameworks and the ML estimator of the variance has the same asymptotic normal distribution under both asymptotic frameworks (Zhang

and Zimmerman, 2005). Although analogous infill asymptotic results are unavailable for multivariate spatial process, it seems to us that inconsistency of estimators for the variances still occurs for multivariate spatial process if the correlation parameter  $\psi$  is to be estimated. This is because two different values for the parameter vector  $(\sigma_1^2, \sigma_2^2, \psi)$  can define two equivalent Gaussian measures on the bivariate process, which can be proved by extending the techniques for scalar Gaussian processes (e.g., Ibragimov and Rozanov, 1978 and Stein, 1999). Considering the potential problem that the estimation of  $\psi$  can bring about, we assume that the parameter  $\psi$  is known and establish the asymptotic distribution of the ML estimator of  $r$  in the following subsection. We then conduct simulation study to see how well this asymptotic distribution applies when  $\psi$  is also estimated by ML methods.

*4.2. Properties of the maximum likelihood estimators.* Assume that the observed vector of variables  $\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2)'$  follows the cross-correlation model, i.e. ,

$$\mathbf{U} \sim N(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{V} \otimes \boldsymbol{\Xi}(\psi)),$$

where  $\mathbf{V} = (\sigma_{ij})_{i,j=1}^2$ ,  $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ , and  $\boldsymbol{\Xi} = [\rho(-\|\mathbf{s}_i - \mathbf{s}_j\|/\psi)]_{i,j=1}^n$  is the correlation matrix of  $\mathbf{U}_1$  or  $\mathbf{U}_2$  with a known correlation function  $\rho(\cdot)$  and a known  $\psi$ . Parameters to be estimated are the components of  $\boldsymbol{\mu}$  and  $\mathbf{V}$ .

We transform each  $\mathbf{U}_i$  into a vector of independent components. Let  $\mathbf{A}$  be a nonsingular matrix from a spectral decomposition  $\mathbf{A}\boldsymbol{\Xi}(\psi)\mathbf{A}' = \mathbf{I}_n$ , and define  $\mathbf{X}_i = \mathbf{A}\mathbf{U}_i$  for  $i = 1, 2$  and  $\mathbf{a} = \mathbf{A}\mathbf{1}_n$ . Then  $\mathbf{X}_i \sim N(\mu_i\mathbf{a}, \sigma_{ii}\mathbf{I}_n)$ ,  $i = 1, 2$ , and jointly

$$\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2)' \sim N(\boldsymbol{\mu} \otimes \mathbf{a}, \mathbf{V} \otimes \mathbf{I}_n). \quad (4.1)$$

The ML estimators in (3.3) can be alternatively written as

$$\hat{\mu}_i(\psi) = \frac{\mathbf{a}'\mathbf{X}_i}{\|\mathbf{a}\|^2}, \quad \hat{\sigma}_{ij}(\psi) = \frac{1}{n} [\mathbf{X}_i - \hat{\mu}_i(\psi)\mathbf{a}]' [\mathbf{X}_j - \hat{\mu}_j(\psi)\mathbf{a}] \quad (4.2)$$

for  $i, j = 1, 2$ .

Hereafter in this section, we suppress  $\psi$  in the estimators. Write  $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \hat{\mu}_2)'$  and  $\hat{\mathbf{V}} = (\hat{\sigma}_{ij})_{i,j=1}^2$ . Following the standard procedures, we can show that

$$\begin{aligned} \hat{\boldsymbol{\mu}} &\sim N(\boldsymbol{\mu}, \|\mathbf{a}\|\mathbf{V}) \\ \frac{n\mathbf{c}'\hat{\mathbf{V}}\mathbf{c}}{\mathbf{c}'\mathbf{V}\mathbf{c}} &\sim \chi_{n-1}^2 \text{ for any vector } \mathbf{c} \in R^2 \text{ such that } \mathbf{c} \neq \mathbf{0}. \end{aligned}$$

In addition,  $\hat{\boldsymbol{\mu}}$  and  $\hat{\mathbf{V}}$  are statistically independent and completely sufficient.

Next we consider the finite-sample distribution of the estimators which will lead to the asymptotic distribution of the ML estimator of the correlation  $r = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$  between  $Y_1(\mathbf{s})$  and  $Y_2(\mathbf{s})$ . To simplify notation, write  $\sigma_1 = \sqrt{\sigma_{11}}$  and  $\sigma_2 = \sqrt{\sigma_{22}}$ . The ML estimator of  $r$  is then  $\hat{r} = \hat{\sigma}_{12}/(\hat{\sigma}_1\hat{\sigma}_2)$  where  $\hat{\sigma}_i$  is the ML estimator of  $\sigma_i$ ,  $i = 1, 2$ . For the discussions below, we reparameterize by using  $\sigma_1, \sigma_2$  and  $r$  in place of  $\sigma_{11}, \sigma_{22}$  and  $\sigma_{12}$ , respectively. Thus, when  $\Xi(\psi)$  is known, the parameter vector to be estimated is  $\boldsymbol{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, r)'$ . The ML estimators are obtained from (3.3) or (4.2).

**THEOREM 4.1.** *Assume that  $\mathbf{U} = (\mathbf{U}'_1, \mathbf{U}'_2)'$  satisfies*

$$\mathbf{U} \sim N(\boldsymbol{\mu} \otimes \mathbf{1}_n, \mathbf{V} \otimes \Xi(\psi)),$$

where  $\mathbf{V} = (\sigma_{ij})_{i,j=1}^2$ ,  $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ ,  $\Xi = [\rho(-\|\mathbf{s}_i - \mathbf{s}_j\|/\psi)]_{i,j=1}^n$  is the correlation matrix of  $\mathbf{U}_1$  or  $\mathbf{U}_2$ ,  $\mathbf{s}_1, \dots, \mathbf{s}_n$  are  $n$  known spatial locations, and  $\rho(\cdot)$  and  $\psi$  are known. Let the ML estimators be given by (3.3) and  $\hat{r} = \hat{\sigma}_{12}/(\hat{\sigma}_1\hat{\sigma}_2)$ . Then the following holds.

1. For any sample size  $n$ , the marginal probability density function of  $\hat{r}|\Xi$  is given by

$$f_{\hat{r}}(x) = \frac{(1-r)^{(n-1)/2}}{\pi\Gamma(n-2)} (1-x^2)^{(n-4)/2} \frac{d^{n-2}}{d(xr)^{n-2}} \left\{ \frac{\cos^{-1}(-rx)}{\sqrt{1-r^2x^2}} \right\}.$$

2. If  $Z = \text{arctanh}(\hat{r})$ , where  $\text{arctanh}(x) = \log[(1+x)/(1-x)]/2$ , then for a sufficiently large  $n$ ,  $Z|\Xi$  has an approximate  $N[\text{arctanh}(r), 1/(n-3)]$  distribution.

**PROOF.** To derive the sampling distribution of  $\hat{\boldsymbol{\theta}}$ , we first rewrite the joint probability density function (pdf) of  $\mathbf{X}$  in (4.1) in terms of the ML estimators  $\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_1, \hat{\sigma}_2$ , and  $\hat{r}$ . The joint pdf can be written as

$$f_{\mathbf{X}}(\mathbf{x}|\Xi) = \frac{1}{(2\pi\sigma_1\sigma_2\sqrt{1-r^2})^n} \times \exp \left\{ -\frac{\|\mathbf{a}\|^2}{2(1-r^2)} \left[ \left( \frac{\hat{\mu}_1 - \mu_1}{\sigma_1} \right)^2 - 2r \left( \frac{\hat{\mu}_1 - \mu_1}{\sigma_1} \right) \left( \frac{\hat{\mu}_2 - \mu_2}{\sigma_2} \right) + \left( \frac{\hat{\mu}_2 - \mu_2}{\sigma_2} \right)^2 \right] + n \left( \frac{\hat{\sigma}_1^2}{\sigma_1^2} - \frac{2r\hat{r}\hat{\sigma}_1\hat{\sigma}_2}{\sigma_1\sigma_2} + \frac{\hat{\sigma}_2^2}{\sigma_2^2} \right) \right\}. \quad (4.3)$$

Next, we follow the geometric arguments in Fisher (1915, 1921) and Kendall and Stuart (1979, pp 411-413), who derived the joint pdf of  $\hat{\boldsymbol{\theta}}$  for the independent-samples case, which corresponds to our model with  $\mathbf{a} = \mathbf{1}_n$  or  $\Xi = \mathbf{I}_n$ . For  $\mathbf{a} = \mathbf{1}_n$ , Fisher (1915) made the following geometric arguments, which is extendable in a straightforward way to any vector  $\mathbf{a}$ .

If  $\mathbf{X}_i$  is a point in  $R^n$  for  $i = 1$  or  $2$ , then  $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2)'$  is a point in  $R^{2n}$ . The geometric argument of Fisher uses the fact that  $\hat{\mu}_i \mathbf{a}$  is the projection of  $\mathbf{X}_i$  onto the one-dimensional space spanned by  $\mathbf{a}$ , for  $i = 1, 2$ . Hence,  $\hat{\mu}_i \mathbf{a}$  and  $\mathbf{X}_i - \hat{\mu}_i \mathbf{a}$  are orthogonal vectors. It is geometrically clear that when  $\hat{\mu}_1$  and  $\hat{\sigma}_1 = \|\mathbf{X}_1 - \hat{\mu}_1 \mathbf{a}\|/\sqrt{n}$  are fixed, the point  $\mathbf{X}_1$  varies on an  $(n-2)$ -dimensional hypersphere with volume proportional to  $(\sqrt{n}\hat{\sigma}_1)^{n-2}$ . Therefore, the probability mass of the set of points where  $\hat{\mu}_1$  lies in  $d\hat{\mu}_1$  and  $\hat{\sigma}_1$  lies in  $d\hat{\sigma}_1$  is proportional to  $\hat{\sigma}_1^{n-2} d\hat{\mu}_1 d\hat{\sigma}_1$ .

Consider  $P_i = \mathbf{X}_i - \hat{\mu}_i \mathbf{a}$  for  $i = 1, 2$  on the same vector space with origin  $O$ . The angle  $\theta$  between  $\overleftrightarrow{OP_1}$  and  $\overleftrightarrow{OP_2}$  satisfies

$$\cos(\theta) = \frac{P_1 \bullet P_2}{(OP_1)(OP_2)}$$

where " $\bullet$ " denotes inner product and  $OP$  is the distance between  $O$  and  $P$ . We have

$$\cos(\theta) = \frac{\sum_{i=1}^n (x_{i1} - a_i \hat{\mu}_1)(x_{i2} - a_i \hat{\mu}_2)}{\sqrt{n}\hat{\sigma}_1 \sqrt{n}\hat{\sigma}_2} = \frac{n\hat{\sigma}_{12}}{\sqrt{n}\hat{\sigma}_1 \sqrt{n}\hat{\sigma}_2} = \hat{r}.$$

Now, fix  $\theta$  and fix the projection  $\mathbf{X}_2 - \hat{\mu}_2 \mathbf{a}$  at any point on the sphere of radius  $\sqrt{n}\hat{\sigma}_2$ . Then the  $(n-3)$ -dimensional region for which  $\hat{r}$  falls in the range  $d\hat{r}$  is a zone with radius  $\sqrt{n}\hat{\sigma}_2 \sin(\theta) = \sqrt{n}\hat{\sigma}_2 \sqrt{1 - \hat{r}^2}$  and width  $\sqrt{n}\hat{\sigma}_2 d\theta = \sqrt{n}\hat{\sigma}_2 d\hat{r} / \sqrt{1 - \hat{r}^2}$ . Thus, this region has volume proportional to  $[\hat{\sigma}_2 \sqrt{n(1 - \hat{r}^2)}]^{n-3} \hat{\sigma}_2 \sqrt{n} d\hat{r} / \sqrt{1 - \hat{r}^2} = \hat{\sigma}_2^{n-2} (1 - \hat{r}^2)^{(n-4)/2} d\hat{r}$ . Thus, we can write

$$d\mathbf{X}_1 d\mathbf{X}_2 \propto (\hat{\sigma}_1 \hat{\sigma}_2)^{n-2} (1 - \hat{r}^2)^{(n-4)/2} d\hat{\mu}_1 d\hat{\mu}_2 d\hat{\sigma}_1 d\hat{\sigma}_2 d\hat{r}.$$

It follows from (4.3) that  $\hat{\boldsymbol{\theta}}$  has the following probability density given  $\Xi = \Xi(\psi)$

$$\begin{aligned} & f(\hat{\boldsymbol{\theta}}|\Xi) \\ & \propto \exp \left\{ -\frac{\|\mathbf{a}\|^2}{2(1-r^2)} \left[ \left( \frac{\hat{\mu}_1 - \mu_1}{\sigma_1} \right)^2 - 2r \left( \frac{\hat{\mu}_1 - \mu_1}{\sigma_1} \right) \left( \frac{\hat{\mu}_2 - \mu_2}{\sigma_2} \right) + \left( \frac{\hat{\mu}_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \\ & \quad \times \hat{\sigma}_1^{n-2} \hat{\sigma}_2^{n-2} (1 - \hat{r}^2)^{(n-4)/2} \times \exp \left[ n \left( \frac{\hat{\sigma}_1^2}{\sigma_1^2} - \frac{2r\hat{r}\hat{\sigma}_1\hat{\sigma}_2}{\sigma_1\sigma_2} + \frac{\hat{\sigma}_2^2}{\sigma_2^2} \right) \right]. \end{aligned} \quad (4.4)$$

Therefore, the marginal distribution of  $\hat{\boldsymbol{\theta}}_1 = (\hat{\sigma}_1, \hat{\sigma}_2, \hat{r})'$  given  $\boldsymbol{\Xi} = \boldsymbol{\Xi}(\psi)$  is

$$f_1(\hat{\boldsymbol{\theta}}_1|\boldsymbol{\Xi}) \propto \hat{\sigma}_1^{n-2}\hat{\sigma}_2^{n-2} (1 - \hat{r}^2)^{(n-4)/2} \times \exp \left[ n \left( \frac{\hat{\sigma}_1^2}{\sigma_1^2} - \frac{2r\hat{r}\hat{\sigma}_1\hat{\sigma}_2}{\sigma_1\sigma_2} + \frac{\hat{\sigma}_2^2}{\sigma_2^2} \right) \right]. \tag{4.5}$$

Note that this joint pdf of  $\hat{\boldsymbol{\theta}}_1$  does not depend on the correlation matrix  $\boldsymbol{\Xi}$  although the estimator itself depends on  $\boldsymbol{\Xi}$ . Hence for any known  $\boldsymbol{\Xi}$ , this distribution of  $\hat{\boldsymbol{\theta}}_1$  is the same as that of  $\hat{\boldsymbol{\theta}}_1$  when  $\boldsymbol{\Xi} = \mathbf{I}_n$ , i.e. , there is no spatial correlation. For the case  $\boldsymbol{\Xi} = \mathbf{I}_n$ , Fisher (1915, p. 516) derived the marginal pdf  $f_{\hat{r}}(x)$  from (4.5) (see also, Kendall and Stuart (1979, p. 415), who expressed the pdf slightly differently). The arctanh transformation was first suggested by Fisher (1915) and then formally addressed in Fisher (1921).  $\square$

Note that in Theorem 4.1, the variables  $\mathbf{U}, \hat{r}, Z$  all depend on  $n$  which is suppressed to simplify the notation. Alternative expressions for  $f_{\hat{r}}(x)$  are given in Kendall and Stuart (1979, pp 416-417). For  $n > 4$ ,  $f_{\hat{r}}(x)$  is unimodal and is more skewed for larger values of  $|r|$ . The distribution tends to normality as  $n \rightarrow \infty$  but at a slow rate. Fisher (1921) observed that  $Z = \text{arctanh}(\hat{r})$  approaches normality much more quickly than  $\hat{r}$  and has a variance almost independent of  $r$ . Generally, for a small  $r$  or  $n > 50$ , the normal approximation to the distribution of  $Z$  is quite close. See also Kendall and Stuart (1979) or Graybill (1976) for details. Hotelling (1953) studied the transformation  $Z^* = Z - (3Z + \hat{r})/(4n)$  which is even closer to normality than  $Z$  with  $\text{Var}(Z^*) \doteq 1/(n - 1)$ .

### 5 Simulation Studies on the Estimation of $r$

In this section, we present the results of simulations that we conducted to study the estimation of the correlation  $r$ . We will compare three estimators of  $r$ : (a)  $\hat{r}(\psi)$  defined in (3.4), where  $\psi$  is the true value of the parameter; (b)  $\hat{r}(\hat{\psi})$ , where  $\hat{\psi}$  is the ML estimate of  $\psi$ ; and (c)  $\hat{r}_0$  defined by (1.1).

We consider measuring two properties at  $n = 121$  equally-spaced locations  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_{121}$  on the unit square  $[0, 1] \times [0, 1]$  and assume that the variables follow the cross-correlation model. We also assume an exponential correlation function. In this case, we have

$$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \quad \boldsymbol{\Xi} = [\exp(-\|\mathbf{s}_i - \mathbf{s}_j\|/\psi)]_{i,j=1}^n,$$

and a dataset consisting of  $n = 121$  pairs of observations. Because  $\Xi$  depends on the scalar  $\psi$ , deriving the ML estimates is a univariate search over  $\psi > 0$  as described in Section 3.

Without loss of generality, we assign  $\mu_1 = \mu_2 = 0$  and fix  $\sigma_1 = \sigma_2 = 1$ . Other values for  $\sigma_i$ 's are not necessary because rescaling each variable by a multiplicative constant does not change the estimators defined in (3.3). We then have 9 combinations of  $r \in \{0, .3, .6\}$  and  $\psi \in \{.1, .2, .3\}$  to be used in the simulation study. For similar reasons as stated previously, it is not necessary to include negative values of  $r$ .  $\psi = .1$  represents weak cross-correlations between locations, while  $\psi = .2, .3$  represent stronger cross-correlations. We use the `rmvnorm` function of S-Plus to generate data based on the cross-correlation model with the chosen set of parameter values.

Our simulation study consists of two parts. The first part investigates and compares  $\hat{r}(\psi)$ , the ML estimator of  $r$  when  $\psi$  is specified correctly, and  $\hat{r}(\hat{\psi})$ , the ML estimator of  $r$  when  $\psi$  is also estimated. This will provide insight on how the ML estimation of  $r$  is affected by replacing  $\psi$  with the estimator  $\hat{\psi}$ . The second part explores the consequence of ignoring the spatial cross-correlation when it actually exists by comparing  $\hat{r}(\hat{\psi})$  with  $\hat{r}_0$ . It also investigates how well the asymptotic approximation by Theorem 4.1 applies to the finite-sample cases.

*5.1. Estimation of  $r$  when  $\psi$  is known and when  $\psi$  is estimated.* In this section, we study the difference between  $\hat{r}(\psi)$  and  $\hat{r}(\hat{\psi})$ . Consider the situation when  $\sigma_1 = 1$ ,  $\sigma_2 = 1$ ,  $r = 0$ , and  $\psi = 0.1$ . For each of the 1000 simulated datasets, we obtain ML estimates and compute  $\hat{r}$  for the two cases, respectively. A scatterplot of the 1000 pairs of  $\text{arctanh}(\hat{r})$  values is given in Figure 1 (a). The  $45^\circ$  line is drawn in the plot to assess the similarity between the pairs of estimates. The points lie very close to this line, and the linear correlation of .995 agrees with this observation. We observe similar trends in the other 8 simulation scenarios for which the linear correlation between  $\hat{r}$  values ranged between .995 and .997.

Figure 1 (b) is a normal probability plot of the 1,000  $\text{arctanh}[\hat{r}(\hat{\psi})]$  values when  $\sigma_1 = 1$ ,  $\sigma_2 = 1$ ,  $r = 0$ , and  $\psi = 0.1$ . The Shapiro-Wilk test for normality yields a test statistic of  $w = 0.9988$  and p-value=0.7833. Normality of  $\text{arctanh}[\hat{r}(\hat{\psi})]$  is not rejected at the 0.05 level. The test and the linearity of the probability plot do not provide enough evidence against the normality of  $\text{arctanh}[\hat{r}(\hat{\psi})]$ .

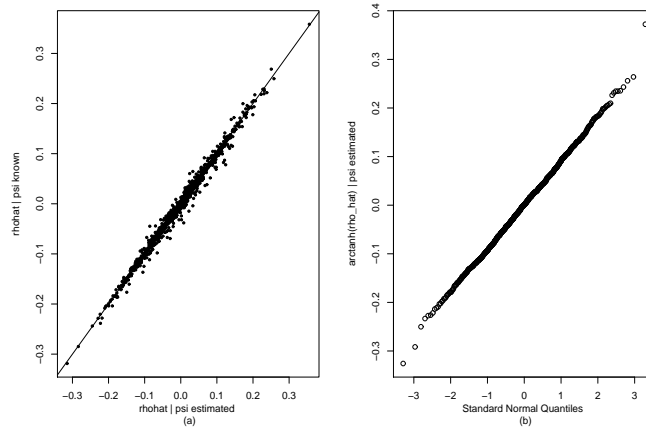


Figure 1: (a) Scatterplot of 1000 pairs of  $\{\text{arctanh}[\hat{\psi}], \text{arctanh}[\psi]\}$  values (b) Normal probability plot of 1000  $\text{arctanh}[\hat{\psi}]$  values (see text for details).

The analogous normal probability plots for the other 8 simulations sets exhibit similar patterns. Table 1 gives the Shapiro-Wilk test statistics and p-values (in parentheses) for all of the 9 simulations sets.

TABLE 1. SHAPIRO-WILK TEST STATISTICS AND P-VALUES FOR TESTING THE NORMALITY OF SIMULATED  $\text{arctanh}[\hat{\psi}]$  VALUES

| Shapiro-Wilk Test<br>for Normality | $\psi$        |               |               |
|------------------------------------|---------------|---------------|---------------|
|                                    | .1            | .2            | .3            |
| $r$                                | .0            | .0            | .0            |
|                                    | .9988 (.7833) | .9984 (.4713) | .9978 (.1982) |
|                                    | .3            | .3            | .3            |
|                                    | .9989 (.8029) | .9979 (.2274) | .9978 (.2034) |
|                                    | .6            | .6            | .6            |
|                                    | .9985 (.5225) | .9959 (.0098) | .9962 (.0159) |

All the tests are insignificant at the 0.05 level, except for the scenarios when  $r = 0.6$  and  $\psi = 0.2, 0.3$ . These suggest that perhaps larger samples sizes are necessary in order to achieve a better normal approximation. For example, for  $r = 0.6$ ,  $\psi = 0.2$ , and a  $14 \times 14$  regular lattice (75 observations more), the test for normality is not significant with test statistic 0.9985 and p-value 0.5796.

Our simulation results suggest that, if the cross-correlation model is correct, then substituting  $\psi$  with  $\hat{\psi}$  will not significantly affect  $\hat{r}$ . Also, the normal probability plots and tests for normality suggest that, for a large

enough sample, a normal distribution can be used to describe the distribution of  $\text{arctanh}[\widehat{r}(\widehat{\psi})]$  for the situations considered here.

5.2. *Estimating  $r$  when spatial correlation is modeled or ignored.* Our goal here is to compare the sampling properties of  $\widehat{r}_0$ , the sample correlation coefficient when spatial correlation is ignored, and  $\widehat{r}(\widehat{\psi})$  when spatial correlation is modeled with  $\psi$  estimated. Let  $Z_0 = \text{arctanh}(\widehat{r}_0)$  and  $Z = \text{arctanh}[\widehat{r}(\widehat{\psi})]$ . We study the

- variance of  $Z_0$  and  $Z$ : by Theorem 4.1, the variance of  $Z$  is approximately  $1/(n-3)$ ;
- coverage of confidence intervals for  $r$ : The 95% confidence intervals for  $r$  based on  $Z$  and  $Z_0$  are, respectively,

$$\left( \tanh \left[ Z - 1.96/\sqrt{n-3} \right], \quad \tanh \left[ Z + 1.96/\sqrt{n-3} \right] \right), \quad (5.1)$$

$$\left( \tanh \left[ Z_0 - 1.96/\sqrt{n-3} \right], \quad \tanh \left[ Z_0 + 1.96/\sqrt{n-3} \right] \right). \quad (5.2)$$

We use the same 9 combinations of the parameter values as in the previous section. For each parameter combination, we generated 1000 datasets each from which we computed 95% CIs for  $r$  according to (5.1) and (5.2). Results are summarized in the following two subsections.

5.2.1. *Results for  $\widehat{r}_0$ .* We have the following observations.

- **Variance.** The  $\text{arctanh}$  transformation method estimates the variance of  $\text{arctanh}(\widehat{r}_0)$  to be  $1/(n-3) = 8.474 \times 10^{-3}$ . The sample variances of the ML estimates for  $\text{arctanh}(\widehat{r}_0)$  are given in Table 2. These sample variances range between 0.0135 and 0.0505, and these are from 59% to 496% above the projected variance. These suggest that ignoring spatial correlations when they actually exist could inflate estimator variance even for weak cross-correlations. Moreover, because the sample variance gets larger with  $\psi$  the underestimation of the true variance is worse with stronger cross-correlations.

TABLE 2. ESTIMATED VARIANCES OF  $\text{arctanh}(\widehat{r}_0)$   
BASED ON 1,000 SIMULATIONS FROM THE  
CROSS-CORRELATION MODEL

| Sample Variance<br>of $\text{arctanh}(\widehat{r}_0)$ | $\psi$ |       |       |
|---|--------|-------|-------|
|   | .1     | .2    | .3    |
| .0  | .0153  | .0313 | .0503 |
| $r$   | .0143  | .0318 | .0505 |
| .6  | .0135  | .0326 | .0474 |

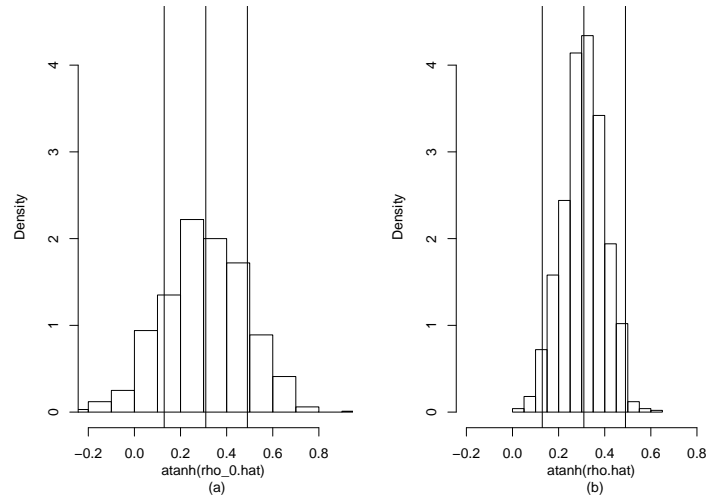


Figure 2: Histograms of estimates (a)  $\arctanh(\hat{r}_0)$  and (b)  $\arctanh(\hat{r}(\hat{\psi}))$  based on 1000 simulated samples (see text for details).

Figure 2 (a) is a histogram of the 1,000 estimates  $\arctanh(\hat{r}_0)$  for  $\sigma_1 = 1$ ,  $\sigma_2 = 1$ ,  $r = 0.3$  and  $\psi = 0.3$ . The center line indicates the true value  $\arctanh(0.3)$ . The left and right lines indicate points that are 1.96 standard deviations (i.e.,  $1.96 \times \sqrt{8.474 \times 10^{-3}}$ ) away from the mean  $\arctanh(0.3)$ . If the confidence interval (5.1) is appropriate, we would expect roughly 95% of all  $\arctanh(\hat{r}_0)$  values to lie between these 2 lines. The histogram shows that the fraction of  $\arctanh(\hat{r}_0)$  values that fall within these bounds is well below 95% (actually 73.2%). This is expected due to the underestimation of the variance of  $\hat{r}_0$  as already pointed out previously. This is also true for the other 9 sets of simulations.

- **Confidence Interval Coverage.** We compute 95% confidence intervals for  $r$  using (5.1) and determine the percentages of CIs that include the true value of  $r$ . Table 3 gives the coverage percentages for the 9 sets of simulations. These percentages range between 57.2% and 87.5% with lower percentages for larger values of  $\psi$  or stronger cross-correlations. Clearly, CI coverage probabilities fall significantly below the nominal 95% when cross-correlations are ignored. Furthermore, the CI coverage is extremely poor when cross-correlations are stronger.

TABLE 3. COVERAGE PERCENTAGES FOR 95%  
CONFIDENCE INTERVALS FOR  $r$  BASED ON  $\hat{r}_0$ 

|     | Coverage<br>Percentages | $\psi$ |      |      |
|-----|-------------------------|--------|------|------|
|     |                         | .1     | .2   | .3   |
| $r$ | .0                      | 84.7   | 68.6 | 57.2 |
|     | .3                      | 87.5   | 67.9 | 59.9 |
|     | .6                      | 86.8   | 68.0 | 58.6 |

5.2.2. *Results for  $\hat{r}(\hat{\psi})$ .* Consider fitting the cross-correlation model to the same simulated datasets in the previous section. We have the following observations about  $\hat{r}(\hat{\psi})$ :

- **Variance.** Variances of  $\arctanh[\hat{r}(\hat{\psi})]$  are given in Table 4, each obtained based on 1,000 simulations. The sample variances here are quite close to the projected value of 0.0085.

Figure 2 (b) is a histogram of the 1000 simulated values of  $\arctanh(\hat{r})$  for  $\sigma_1 = 1$ ,  $\sigma_2 = 1$ ,  $r = 0.3$  and  $\psi = 0.2$ . The histogram shows that close to 95% (actually 95.8%) of  $\arctanh(\hat{r})$  values are within 1.96 standard deviations of the true mean  $\arctanh(0.3)$ . The same holds for the other 8 sets of simulations.

- **Confidence Interval Coverage.** For the 9 simulation sets, Table 5 gives estimates of the percent of the time that 95% confidence intervals based on the  $\arctanh$  transformation covers the true value of  $r$ . These percentages ranging between 94.2% and 95.8% are certainly close to the nominal 95%.

TABLE 4. ESTIMATED VARIANCES OF  $\arctanh[\hat{r}(\hat{\psi})]$   
BASED ON 1,000 SIMULATIONS

| Sample Variance<br>of $\arctanh(\hat{r})$ | $\psi$ |       |       |
|---|--------|-------|-------|
|   | .1     | .2    | .3    |
| .0  | .0085  | .0089 | .0082 |
| $r$                                       | .3     | .0086 | .0083 |
|   | .6     | .0084 | .0080 |

TABLE 5. ACTUAL COVERAGE PERCENTAGES FOR 95%  
CONFIDENCE INTERVALS FOR  $r$  USING (16)

|     | Coverage<br>Percentages | $\psi$ |      |      |
|-----|-------------------------|--------|------|------|
|     |                         | .1     | .2   | .3   |
| $r$ | .0                      | 95.3   | 94.3 | 95.2 |
|     | .3                      | 94.2   | 95.8 | 95.2 |
|     | .6                      | 94.8   | 95.3 | 95.7 |

*5.2.3. Summary of Simulation Results.* It is clear that ignoring cross-correlations results in an actual variance for  $\text{arctanh}(\hat{r}_0)$ , and subsequently for  $\hat{r}_0$ , that is much larger than what is projected by inferential methods based on the assumption that samples or spatial locations are independent. Also, because the variance of  $\hat{r}_0$  is larger than expected, confidence interval coverage of the true value of  $r$  could be poor particularly in situations of strong cross-correlations. A remedy for this problem is to fit a model that describes the cross-correlations structure. Our simulations show that doing so results in accurate projections of the variance of  $\text{arctanh}(\hat{r})$  and confidence interval coverage percentages close to nominal.

## 6 Discussion and Areas for Further Research

Previous works have shown that when cross-correlations exist but is ignored, inference methods for the sample correlation underestimate standard errors. This in turn results in lower than nominal coverage probabilities by confidence intervals and, equivalently, larger than nominal Type I errors for hypothesis tests. Our investigations suggest a remedy for correcting these problems by considering a statistical model which explicitly models cross-correlations.

We established the asymptotic distribution of  $\text{arctanh}(\hat{r})$  under the assumption that the correlation parameter  $\psi$  is known. However, our numerical results show that  $\hat{r}$  does not change significantly when  $\psi$  is also estimated. Existing infill asymptotic results lead us to believe that this insensitivity may be justified by asymptotic results and can be a problem for further research. To relate this to existing asymptotic results, let us assume that the means  $\mu_i$  are known to be 0 so that  $\hat{\mu}_i$  in (3.3) disappears. In this case, for any fixed value  $\tilde{\psi}$ , by Theorem 3.1 in Zhang (2004),  $\hat{\sigma}_{ii}(\tilde{\psi})/\tilde{\psi}$  is a consistent estimator of  $\sigma_{ii}/\psi$ . If one can show that  $\hat{\sigma}_{12}(\tilde{\psi})/\tilde{\psi}$  consistently estimates  $\sigma_{12}/\psi$ , then  $\hat{r}(\tilde{\psi})$  is a consistent estimator of  $r$ . When the observational domain is one-dimensional, more in-depth asymptotic results exist and further consolidate the above belief. Ying (1991) has shown that if  $\tilde{\psi}$  is fixed or is the ML estimator, then  $\hat{\sigma}_{ii}(\tilde{\psi})/\tilde{\psi}$  has the same asymptotic normal distribution as  $\hat{\sigma}_{ii}(\psi)/\psi$ . Hence, if analogous result can be established for  $\hat{\sigma}_{12}(\tilde{\psi})/\tilde{\psi}$ , then it is likely that the asymptotic distribution of  $\hat{r}(\tilde{\psi})$  does not depend on  $\tilde{\psi}$ .

Bivand (1980) noted that the bias in estimating the standard errors varied with the lattice type. Recent advances in geostatistics show that having some locations closer to each other improves the efficiency of estimators

(Stein, 1999). More simulation studies could be done to study how the sampling design affects the estimation of  $r$ . A more rigorous approach is to derive sampling designs that optimize specific design criteria (e.g.,  $D$ -optimal, minimum asymptotic variance of predictor or estimator, etc.).

There are few models for multivariate covariogram which are less restrictive than the proportional covariogram model (PCM). For example, the covariogram of a linear coregionalization model is the sum of several PCMs. It is an interesting and hard problem to study the estimation of  $r$  for those models.

*Acknowledgement.* The authors would like to thank the referee for valuable comments and suggestions. The research of H. Zhang is partially supported by a grant from National Science Foundation of the U.S.A. (DMS-0405782).

### References

- BANERJEE, S. and GELFAND, A.E. (2002). Prediction, interpolation and regression for spatially misaligned data. *Sankhyā Ser. A*, **64**, 227–245.
- BIVAND, R. (1980). A Monte Carlo study of correlation coefficient estimation with spatially autocorrelated observations. *Quaestiones Geographicae*, **6**, 5–9.
- CHILÉS, J.P. and DELFINER, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York.
- CLIFFORD, P., RICHARDSON, S. and HEMON, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics*, **45**, 123–134.
- CLIFFORD, P., RICHARDSON, S. and HEMON, D. (1993). Response to “(Reader Reaction) Modifying the t test for assessing the correlation between two spatial processes.” *Biometrics*, **45**, 312–314.
- DUTILLEUL, P. (1993). (Reader Reaction) Modifying the t test for assessing the correlation between two spatial processes. *Biometrics*, **49**, 305–312.
- FISHER, R.A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521.
- FISHER, R.A. (1921). On the “probable” error of a coefficient of correlation from a small sample. *Metron*, **1**, 3–32.
- GRAYBILL, F.A. (1976). *Theory and Application of the Linear Model*. Wadsworth Publishing Co., Belmont, California.
- HOTELLING, H. (1953). New light on the correlation coefficient and its transformations. *J. Roy. Statist. Soc. Ser. B*, **15**, 193–232.

- IBRAGIMOV, I.A. and ROZANOV, Y.A. (1978). *Gaussian Random Processes*. Springer-Verlag, New York.
- KENDALL, M.G. and STUART, A. (1979). *The Advanced Theory of Statistics, Vol. 1: Distribution Theory* (Fourth edition). Charles Griffin, London.
- MARDIA, K.V. and GOODALL, C.R. (1993). Spatial-temporal Analysis of multivariate environmental monitoring data. In *Multivariate Environmental Statistics* (G.P. Patil and C.R. Rao, eds.), Elsevier, Amsterdam, 347–386.
- SEARLE, S.R. (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.
- STEIN, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.
- STUDENT (GOSSET, W.W.) (1914). The elimination of spurious correlation due to position in time or space. *Biometrika*, **65**, 109–114.
- WACKERNAGEL, H. (1998). *Multivariate Geostatistics: An Introduction with Applications*. Springer-Verlag, New York.
- YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.*, **36**, 280–296.
- ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.*, **99**, 250–261.
- ZHANG, H. and ZIMMERMAN, D. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, **92**, 921–936.

F. PASCUAL  
DEPARTMENT OF MATHEMATICS  
WASHINGTON STATE UNIVERSITY  
PULLMAN, WA 99164-3113, USA  
E-mail: jpascual@math.wsu.edu

H. ZHANG  
DEPARTMENT OF STATISTICS  
WASHINGTON STATE UNIVERSITY  
PULLMAN, WA 99164-3144, USA  
E-mail: zhanghao@wsu.edu