Constructing a Spatial Concordance Correlation Coefficient

Ronny Vallejos[1,*], Javier Pérez[2], Aaron Ellison[3] and Andrew D. Richardson[4]

[1,2] *Departamento de Matemática, Universidad Técnica Federico Santa María, Avenida España 1680, Valparaíso, Chile; ronny.vallejos@usm.cl, javier.perez@alumnos.usm.cl,*
[3] *Harvard Forest, Harvard University, Petersham, Massachusetts, USA; aellison@fas.harvard.edu*
[4] *School of Informatics, Computing and Cyber Systems, Northern Arizona University, USA and Center for Ecosystem Science and Society, Northern Arizona University, USA; Andrew.Richardson@nau.edu*
[*] *Corresponding author*

Abstract. *In this work we define a spatial concordance coefficient for second-order stationary processes. This problem has been widely addressed in a non-spatial context, but here we consider a coefficient that for a fixed spatial lag allows one to compare two spatial sequences along a 45° line. The proposed coefficient is explored for the bivariate Matérn and Wendland covariance functions. The asymptotic normality of a sample version of the spatial concordance coefficient for an increasing domain sampling framework is established for the Wendland covariance function. Monte Carlo simulations are used to gain additional insights into the asymptotic properties for finite sample sizes. The results will be illustrated by real data examples to see how our method works in practice.*

Keywords. *Concordance; Correlation; Spatial correlation function; Lin's coefficient; Bivariate Wendland covariance function.*

1 A Concordance Correlation Coefficient

In recent decades, concordance correlation coefficients have been developed in a variety of different contexts. For instance, in assay or instrument validation processes, the reproducibility of the measurements from trial to trial is of interest. Also, when a new instrument is developed, it is relevant to evaluate whether its performance is concordant with other, existing ones. In the literature, this concordance has been tackled from different perspectives [1]. One way to approach this problem for continuous measurements is constructing a scaled summary index that can take on values between -1 and 1. Using this approach Lin [6] suggested a concordance correlation coefficient that evaluates the agreement between two continuous variables by measuring the variation from a 45° line through the origin.

More precisely, assume that X and Y are two continuous random variables such that the joint distribution of X and Y has finite second moments with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and covariance σ_{YX}. The mean squared deviation of $D = Y - X$ is

$$\text{MSD} = \varepsilon^2 = \mathbb{E}[D^2] = \mathbb{E}[(Y-X)^2].$$

It is straightforward to see that $\varepsilon^2 = (\mu_X - \mu_Y)^2 + \sigma_Y^2 + \sigma_X^2 - 2\sigma_{YX}$ and the sample counterpart satisfies $e^2 = (\bar{y} - \bar{x})^2 + s_Y^2 + s_X^2 - 2s_{XY}$. Under the above hypothesis, Lin [6] proposed a concordance correlation coefficient defined as

$$\rho_c = 1 - \frac{\varepsilon^2}{\varepsilon^2|\rho = 0} = \frac{2\sigma_{YX}}{\sigma_Y^2 + \sigma_X^2 + (\mu_Y^2 - \mu_X^2)^2}. \tag{1}$$

This coefficient satisfies the following properties:

1. $\rho_c = \alpha \cdot \rho$, where $\alpha = \frac{2}{w+1/w+v^2}$ and $w = \frac{\sigma_Y}{\sigma_X}$.
2. $|\rho_c| \leq 1$.
3. $\rho_c = 0$ if and only if $\rho = 0$.
4. $\rho_c = \rho$ if and only if $\sigma_Y = \sigma_X$ and $\mu_Y = \mu_X$.

The sample estimate of ρ_c is given as

$$\widehat{\rho}_c = \frac{2s_{YX}}{s_Y^2 + s_X^2 + (\bar{y} - \bar{x})^2}.$$

The inference for this coefficient was addressed via Fisher's transformation. Lin [6] proved that

$$Z = \frac{1}{2}\left(\frac{1+\widehat{\rho}_c}{1-\widehat{\rho}_c}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\psi, \sigma_Z^2), \text{ as } n \to \infty,$$

where $\psi = \tanh^{-1}(\rho_c) = \frac{1}{2}\left(\frac{1+\rho_c}{1-\rho_c}\right)$, $\sigma_Z^2 = \frac{1}{n-2}\left[\frac{(1-\rho^2)\rho_c^2}{(1-\rho_c^2)\rho^2} + \frac{2v^2(1-\rho_c)\rho_c^3}{(1-\rho_c^2)^2\rho} + \frac{v^4\rho_c^4}{2(1-\rho_c^2)^2\rho^2}\right]$, and $v^2 = \frac{(\mu_Y-\mu_X)^2}{\sigma_Y\sigma_X}$. As a consequence of the asymptotic normality of the sample concordance index, an approximate hypothesis testing problem of the form $\mathrm{H}_0 : \rho_c = \rho_0$ versus $\mathrm{H}_1 : \rho_c \neq \rho_0$ for a fixed ρ_0 can be constructed.

Applications and extensions of Lin's coefficient can be found in [5], [7], and [8], among others.

2 A Spatial Concordance Coefficient and its Properties

Here, we extend Lin's coefficient for bivariate second-order spatial processes for a fixed lag in space.

Definition 1. Let $(X(\boldsymbol{s}), Y(\boldsymbol{s}))^\top$ be a bivariate second-order stationary random field with $\boldsymbol{s} \in \mathbb{R}^2$, mean $(\mu_1, \mu_2)^\top$, and covariance function

$$C(\boldsymbol{h}) = \begin{pmatrix} C_X(\boldsymbol{h}) & C_{XY}(\boldsymbol{h}) \\ C_{YX}(\boldsymbol{h}) & C_Y(\boldsymbol{h}) \end{pmatrix}.$$

Then the spatial concordance coefficient is defined as

$$\rho^c(\boldsymbol{h}) = \frac{\mathbb{E}[(Y(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s}))^2]}{\mathbb{E}[(Y(\boldsymbol{s}+\boldsymbol{h}) - X(\boldsymbol{s}))^2|C_{XY}(\boldsymbol{0}) = 0]} = \frac{2C_{YX}(\boldsymbol{h})}{C_X(\boldsymbol{0}) + C_Y(\boldsymbol{0}) + (\mu_1 - \mu_2)^2}. \tag{2}$$

Some straightforward features of this coefficient are the following:

1. $\rho^c(\boldsymbol{h}) = \eta \cdot \rho_{YX}(\boldsymbol{h})$, where $\eta = \frac{2\sqrt{C_X(\boldsymbol{0})C_Y(\boldsymbol{0})}}{C_X(\boldsymbol{0})+C_Y(\boldsymbol{0})+(\mu_1-\mu_2)^2}$.
2. $|\rho^c(\boldsymbol{h})| \leq 1$.
3. $\rho^c(\boldsymbol{h}) = 0$ iff $\rho_{YX}(\boldsymbol{h}) = 0$.
4. $\rho^c(\boldsymbol{h}) = \rho_{YX}(\boldsymbol{h})$ iff $\mu_1 = \mu_2$ and $C_X(\boldsymbol{0}) = C_Y(\boldsymbol{0})$.

5. For a bivarite Matérn covariance function defined as $C_X(\boldsymbol{h}) = \sigma_1^2 M(\boldsymbol{h},\nu_1,a_1)$, $C_Y(\boldsymbol{h}) = \sigma_2^2 M(\boldsymbol{h},\nu_2,a_2)$, $\mu_1 = \mu_2$, $C_{YX}(\boldsymbol{h},\nu_{12},a_{12}) = \rho_{12}\sigma_1\sigma_2 M(\boldsymbol{h},\nu_{12},a_{12})$, where $M(\boldsymbol{h},\nu,a) = (a||\boldsymbol{h}||)^\nu K_\nu(a||\boldsymbol{h}||)$, and $K_\nu(\cdot)$ is a modified Bessel function of the second type and $\rho_{12} = \text{cor}[X(\boldsymbol{s}_i),Y(\boldsymbol{s}_j)]$ we have that

$$\rho^c(\boldsymbol{h}) = \frac{2\sigma_1\sigma_2 M(\boldsymbol{h},\nu_{12},a_{12})}{\sigma_1^2+\sigma_2^2} = \eta\cdot\rho_{12},$$

where $\eta = \frac{2\sigma_1\sigma_2 M(\boldsymbol{h},\nu_{12},a_{12})}{\sigma_1^2+\sigma_2^2}$.

6. For a bivariate Wendland-Gneiting covariance function [3] of the form

$$\boldsymbol{C}(\boldsymbol{h}) = [\rho_{ij}\sigma_{ii}\sigma_{jj}R_{ij}(\boldsymbol{h})]_{i,j=1}^2,$$

where $R(\boldsymbol{h},\psi_{12}) = c_{ij}b_{ij}^{\nu+2k+1}B(\nu+2k+1,\gamma_{ij}+1)\tilde{\psi}_{\nu+\gamma_{ij}+1,k}\left(\frac{\|\boldsymbol{h}\|}{b_{ij}}\right)$, $B(\cdot,\cdot)$ is the beta function, and $\tilde{\psi}_{\nu,k}$ is defined in [4], the spatial concordance coefficient is

$$\rho^c(\boldsymbol{h}) = \frac{2\rho_{12}\sigma_1\sigma_2 R(\boldsymbol{h},\psi_{12})}{\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)^2}, \quad \boldsymbol{h}\in\mathbb{R}^2, \tag{3}$$

In particular, considering $R_{ij}(\boldsymbol{h}) = p_k(\|\boldsymbol{h}\|)(1-\|\boldsymbol{h}\|/b_{ij})_+^l$, where $k=2$, $l=v+1$, and $b_{ij}>0$,

$$\rho^c(\boldsymbol{h}) = \frac{2\rho_{12}\sigma_1\sigma_2\,(1+l\|\boldsymbol{h}\|/b_{12})\,(1-\|\boldsymbol{h}\|b_{12})_+^l}{\sigma_1^2+\sigma_2^2+(\mu_1-\mu_2)}.$$

3 Inference

In the previous section we proved that for several covariance structures, the spatial concordance coefficient defined in (2) can be written as a product of the correlation coefficient and a constant. Thus, we can consider a plug-in estimator for these two quantities.

Let $(Z_1(\boldsymbol{s}),Z_2(\boldsymbol{s}))^\top$ $s\in D$ be a Gaussian process with mean $\boldsymbol{\mu} = (\mu_1,\mu_2)^\top$ and covariance function $\boldsymbol{C}(\boldsymbol{h})$, $\boldsymbol{s},\boldsymbol{h}\in\mathbb{R}^2$. Then a sample estimate of the concordance index (2) is

$$\widehat{\rho}_c(\boldsymbol{h}) = \widehat{\rho}_{12}(\boldsymbol{h})\widehat{C}_{ab}, \tag{4}$$

where $\widehat{C}_{ab} = ((\widehat{a}+1/\widehat{a}+\widehat{b}^2)/2)^{-1}$, $\quad \widehat{a} = \left(\frac{\widehat{C}_{11}(\boldsymbol{0})}{\widehat{C}_{22}(\boldsymbol{0})}\right)^{1/2}$, $\widehat{b} = \frac{\widehat{\mu_1}-\widehat{\mu}_2}{(\widehat{C}_{11}(\boldsymbol{0})\widehat{C}_{22}(\boldsymbol{0}))^{1/4}}$ and $\widehat{\mu}_1$, $\widehat{\mu}_2$, $\widehat{C}_{11}(\boldsymbol{0})$ and $\widehat{C}_{22}(\boldsymbol{0})$, are the maximum likelihood (ML) estimates of μ_1, μ_2, $C_{11}(\boldsymbol{0})$ and $C_{22}(\boldsymbol{0})$, respectively.

The asymptotic properties of an estimator as in (4) have been studied in the literature for specific cases. Moreno et al. [2] studied the asymptotic properties of the ML estimator for a separable Matérn covariance model. They used a result provided by Mardia and Marshall [9] in an increasing domain sampling framework. Using this result and the delta method we can establish the following result for the Wendland-Gneiting model.

Theorem 1. Let $(Z_1(\boldsymbol{s}),Z_2(\boldsymbol{s}))^\top$, $s\in D$ be a bivariate Gaussian spatial process with mean $\boldsymbol{0}$ and covariance function given by

$$\boldsymbol{C}(\boldsymbol{h}) = \left[\rho_{ij}\sigma_{ii}\sigma_{jj}\left(1+(\nu+1)\frac{\|\boldsymbol{h}\|}{b_{12}}\right)\left(1-\frac{\|\boldsymbol{h}\|}{b_{12}}\right)_+^{\nu+1}\right]_{i,j=1}^2,$$

for $\nu > 0$ fixed. Define $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \rho_{12}, b_{12})^\top$ and denote $\widehat{\boldsymbol{\theta}}_n$ the ML estimator of $\boldsymbol{\theta}$. Then

$$\left(\nabla g(\boldsymbol{\theta})^\top \boldsymbol{F}_n(\boldsymbol{\theta})^{-1} \nabla g(\boldsymbol{\theta})\right)^{-1/2} (g(\widehat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta})) \xrightarrow{D} \mathcal{N}(0,1), \text{ as } n \to \infty,$$

where $g(\boldsymbol{\theta}) = \dfrac{2\rho_{12}\sigma_1\sigma_2 \left(1 + (\nu+1)\dfrac{\|\boldsymbol{h}\|}{b_{12}}\right)\left(1 - \dfrac{\|\boldsymbol{h}\|}{b_{12}}\right)_+^{\nu+1}}{\sigma_1^2 + \sigma_2^2}$, and $\boldsymbol{F}_n(\boldsymbol{\theta})$ is the Fisher information matrix associated with $\widehat{\boldsymbol{\theta}}$.

In addition, we also provide an expression for the asymptotic variance of the spatial concordance coefficient.

4 Applications

We explore the properties of the spatial concordance coefficient ρ^c for finite samples sizes through Monte Carlo simulations. Misspecification of the covariance function will also be addressed to inspect the impact on estimations when assuming a misspecified covariance function for the processes.

An application with real data also will be analyzed. Two images taken from the same location will be compared using the spatial concordance coefficient for different spatial lags. This measure of concordance is applied to digital images of forest-tree spring leaf-out obtained by different cameras. The result measures the agreement between two different acquisition processes. A concordance map similar to a codispersion map can be built to explore (an)isotropy in spatial concordance.

5 Acknowledgments

This work has been partially supported by the AC3E, UTFSM, under grant FB-0008, and by grants from the US National Science Foundation and NASA.

References

[1] Barnhart, H. X., Haber, M. J., Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics* **17**, 529–569.

[2] Bevilacqua, M., Vallejos, R., Velandia, D. (2015). Assessing the significance of the correlation between the components of a bivariate Gaussian random field. *Environmetrics* **26**, 545–556.

[3] Daley, D. J., Porcu, E., Bevilacqua, M. (2015). Classes of compactly supported covariance functions for multivariate random fields. *Stochastic Environmental Research and Risk Assessment* **29**, 1249–1263.

[4] Gneiting, T., 2002. Compactly supported correlation functions. *J*ournal of Multivariate Analysis **83**, 493–508.

[5] Hiriote, S., Chinchilli, V. M. (2011). Matrix-based Concordance Correlation Coefficient for Repeated Measures. *Biometrics* **67**, 1007–1016.

[6] Lin, L. I-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **45**, 255–268.

[7] Lin, L., Hedayat, A. S., Sinha, B., Yang, M. (2002). Statistical methods in assessing agreement: Models, issues, and tools. *J*ournal of the American Statistical association **97**, 257–270.

[8] Lin, L., Hedayat, A. S., Wu, W. (2012). *Statistical Tools for Measuring Agreement*, Springer. New York.

[9] Mardia, K. V., Marshall, T. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.