Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 11, n. 1, 2025.

Trabalho apresentado no XLIII CNMAC, Centro de Convenções do Armação Resort - Porto de Galinhas - PE, 2024

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Exploring Spatial Continuity with Wendland's Covariance Functions: A Geostatistical Approach

Marcelo dos Santos[1]
IFBA, Paulo Afonso, BA
Fernanda de Bastiani[2]
UFPE, Recife, PE
Miguel A. Uribe-Opazo[3]
UNIOESTE, Cascavel, PR
Manuel Galea[4]
UC, Santiago, CHL

Astract. This study explores spatial continuity using Wendland's covariance functions in geostatistical modeling. Wendland's covariance family, defined within a compact support, offers flexibility with a smoothing parameter, competing with the well-known Matérn family. Expressions for covariance functions and a spatial dependency index are provided, along with sensitivity analysis using local influential diagnostics. Jackknife-after-Bootstrap resamples establish reference levels for potential influence detection. An application to soybean yield data validates the methodology.

Keywords. Compact Support, Geostatistics, Sparse Matrix, Spatial Variability.

1 Introduction

Several works have contributed to the development of positive definite radial functions with compact support, laying the foundation for spatial modeling. [13] introduced a class of such functions and established criteria for their positivity. [12] demonstrated the suitability of certain piecewise polynomial functions as positive definite radial functions with compact support. Building upon these concepts, [7] investigated correlation functions governing the smoothness of associated random fields, offering a computationally efficient approach for spatial prediction. [3] showed the uniform convergence of Wendland's functions to Gaussian covariance functions under appropriate variable rescheduling. Generalizations of radial-based functions, termed generalized Wendland functions, were studied by [2], with [1] focusing on their application in Gaussian random field estimation and prediction. The Wendland covariance family, characterized by its compact support within the interval $(0, 1]$, offers flexibility through a varying smoothing parameter, making it a competitor to the Matérn family. Despite Matérn's limitations in handling large distances, Wendland-type functions demonstrate advantages in generating sparse covariance matrices, particularly beneficial in high-dimensional datasets. Like other statistical models, spatial models require validation, prompting the focus of this study on influential diagnostic analysis within Gaussian spatial linear models employing the Wendland covariance structure. Specifically, local influence methodology proposed by [4] is explored, with the Jackknife-after-Bootstrap technique used to determine influential levels. The paper aims to assess the performance of Wendland family covariance

[1]marcelo.santos@ifba.edu.br
[2]fernanda.bastiani@ufpe.br
[3]mopazo@uol.com.br
[4]mgalea@mat.uc.cl

functions in diagnosing local influence within linear spatial Gaussian models, supplemented by a novel spatial index.

2 Wendland's Covariance Family Functions

The Wendland covariance family makes up a class of positive defined functions with the domain in a compact support, which were introduced by [13], and subsequently unfolded by [12]. Its domain is restricted to the interval $(0, 1]$, defined from the distances between two points. For the present study, the Euclidean distance between two distinct spatial locations was considered, i.e., $\mathbf{s}_i$, and $\mathbf{s}_i + \mathbf{h}$, where $i = 1, \cdots, n$, with $h = \|(\mathbf{s}_i + \mathbf{h}) - \mathbf{s}_i\|$, belonging to the $d-$dimensional space $\mathbb{R}^2 (d = 2)$.

Let us consider a stationary covariance function $C : \mathbb{R}^2 \longrightarrow \mathbb{R}$ [7], [1] for an $\mathbf{\Phi}_2$ family of continuous mappings $f : [0, \infty) \longrightarrow \mathbb{R}$, with (1):

$$cov(Z(\mathbf{s}), Z(\mathbf{s} + \mathbf{h})) = C\left((\mathbf{s} + \mathbf{h}) - \mathbf{s}\right) = f \left\|(\mathbf{s} + \mathbf{h}) - \mathbf{s}\right\|, \tag{1}$$

where $\|\cdot\|$ is the Euclidean norm and $Z = \{Z(\mathbf{s}) : \mathbf{s} \in \mathbf{S} \subset \mathbb{R}^2\}$ is a stochastic process. That said, us let consider the covariance functions of the Generalized type of Wendland (GW) similar to the one defined in [1] by

$$C_{\mu,\kappa}(h) = \begin{cases} \dfrac{1}{B(2\kappa, \mu + 1)} \int_h^1 x\left(x^2 - h^2\right)^{\kappa-1} (1 - x)^{\mu} dx, & if \quad 0 \leq h \leq 1 \\ 0, \quad if \quad h > 1, \end{cases} \tag{2}$$

where $B(\cdot)$ denotes the Beta function, h is the Euclidean distance, and the smoothing parameter is $\kappa \in \mathbb{Z}_+$. The parameter μ characterizes the concavity of the correlation curve, and in Equation (3), it is possible to define the positivity of the generated covariance matrix.

$$\mu \geq \frac{d + 1}{2} + \kappa, \quad \text{with} \quad d = 2. \tag{3}$$

[7] and [14] presented these arguments to show that (2) belongs to the class of positive defined functions, in a d-dimensional space if, and only if (3) is satisfied.

2.1 Practical Range and Spatial Dependency Index for Generalized Wendland Covariance Family

As the literature suggests, if $\rho(h) = 0$ for a distance h greater than a finite real number, $a > 0$, then the interval $(0, a]$ is called the spatial correlation interval. In general, the correlation function, $\rho(h)$, assumes a null value only asymptotically, and in this case, the spatial correlation interval is undefined. According to [6], it is common to adopt the conversion $\rho(a) = 0.05$. In other words, $\gamma(a) = \phi_1 + 0.95\phi_2$ to determine the practical range.

The relationship between the practical range and the ϕ_3 parameter for the GW covariance function with smoothing parameters $\kappa = 0, 1, 2$, and 3 is $a = 0.776\phi_3$, $a = 0.657\phi_3$, $a = 0.573\phi_3$, and $a = 0.514\phi_3$, with $\mu = 2$, 3, 4 and 5, respectively.

Known spatial dependency indices in the literature, use only the parameters ϕ_1 and ϕ_2 of the semivariance function. According to [11], this is a weakness of the method, given that in such case the aspects of the semivariogram geometry are not considered, which can lead the researcher to erroneous conclusions. Considering this, [11] proposed a new measure of spatial dependence,

DOI: 10.5540/03.2025.011.01.0399

called the Spatial Dependency Index (SDI), which takes into account all information inherent to the semivariance function. This new index is calculated by the following:

$$SDI(\%) = MF \times \left(\frac{\hat{\phi}_2}{\hat{\phi}_1 + \hat{\phi}_2}\right) \times \left(\frac{a}{q.MD}\right) \times 100, \tag{4}$$

where MF is a factor that measures the strength of the model's spatial dependence. The higher its value is, greater the spatial dependence of the model. The value of MF is obtained by calculating the spatial continuity area of the model. [11] defined the term $q.MD$ as the fraction (q) resulting from the maximum distance (MD) between the sampled points. The fraction $\frac{a}{q.MD}$ must be between zero and 1 and in the event of obtaining a value of greater than 1, its value is fixed at 1.

Given the relation in Equation (5), [11] calculated the "Spatial Continuity Area (SCA)," given in Equation (6) for the Spherical, Exponential and Gaussian models, obtaining of values 0.375, 0.317, and 0.50 for model factor (MF), respectively . Subsequently, based on the median and third quartile, [10] proposed interval categories of spatial dependence based on the values obtained by SDI, defining *categories of spatial variability.*

$$\rho(h) = 1 - \frac{\gamma(h)}{C(0)}, \tag{5}$$

where $C(0) > 0$ is equivalent to the total sill $\phi_1 + \phi_2$, $\rho(h)$ indicates the spatial correlation function, and $\gamma(h)$ indicates the theoretical model that defines the semivariance function. Then, SCA_κ is calculated by the following:

$$SCA_\kappa = \int_0^a \left(1 - \frac{\gamma(h)}{\phi_1 + \phi_2}\right) dh. \tag{6}$$

This study obtained the SCA for the GW covariance family, $SCA_0 = \dfrac{\phi_2}{\phi_1 + \phi_2} \times a \times 0.425$, $SCA_1 = \dfrac{\phi_2}{\phi_1 + \phi_2} \times a \times 0.502$, $SCA_2 = \dfrac{\phi_2}{\phi_1 + \phi_2} \times a \times 0.511$, and $SCA_3 = \dfrac{\phi_2}{\phi_1 + \phi_2} \times a \times 0.512$ for the smoothing parameters $\kappa = 0, 1, 2,$ and 3, respectively. The factors of the model (MF) are equal to $MF_0 = 0.425$, $MF_1 = 0.502$, $MF_2 = 0.511$, and $MF_3 = 0.512$. Details are given in [9].

2.2 Gaussian Spatial Linear Model

Let us consider an intrinsically stationary stochastic process $Z = \{Z(\mathbf{s}_i) : \mathbf{s}_i \in \mathbf{S} \subset \mathbb{R}^d, d = 2\}$. Let $\mathbf{Z}$ be an $n \times 1$ random vector, $\mathbf{Z} = (Z(\mathbf{s}_1), \cdots, Z(\mathbf{s}_n))^\top$, with $Z(\mathbf{s}_i)$ defining the observations of the stochastic process in different locations $\mathbf{s}_i, i = 1, \cdots, n$. According to [5] and [8], the observations of the process are generated by a linear spatial model defined as follows:

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{7}$$

where $\mu(\mathbf{s}_i)$ denotes the component of the fixed effects of the process, that is, the a deterministic term. The term $\epsilon(\mathbf{s}_i)$ indicates the error of the process, that is, the stochastic part in which we assume to have zero mean, $E(\epsilon(\mathbf{s}_i)) = 0$, and finite variance determined by a covariance function $Cov(\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_i + \mathbf{h})) = C(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h}), i, \cdots, n$, with $\mathbf{h}$ indicating the Euclidean distance between two locations. It was assumed that the covariance function $C(\mathbf{s}_i, \mathbf{s}_i + \mathbf{h})$ is specified by a parameter vector $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_3)^\top$.

Furthermore, according to [5], given a set $\{x_j(\mathbf{s}_i), j = 1, \cdots p, i = 1, \cdots n\}$ of explanatory variables, the mean of the stochastic process is defined by (8).

$$\mu(\mathbf{s}_i) = \sum_{j=1}^{p} x_j(\mathbf{s}_i)\beta_j, \tag{8}$$

where $\beta_j' s$ indicates the jth unknown parameter to be estimated. In view of this, we define a vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$, one design matrix, $\mathbf{X}$, on the order $n \times p$ with (i, j) entries defined by $x_j(\mathbf{s}_i), j = 1, \cdots, p, i = 1, \cdots, n$, that is, the ith line is composed by the vector $\mathbf{x}_i^\top = (x_1(\mathbf{s}_i), \cdots, x_p(\mathbf{s}_i))$, and $\boldsymbol{\epsilon} = (\epsilon(\mathbf{s}_1), \cdots, \epsilon(\mathbf{s}_n))$ is the stochastic error vector. Then, the spatial linear model defined in (7) can be rewritten in the following matrix form (9):

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{9}$$

where $\mathbf{X}\boldsymbol{\beta}$ is the mean process defined in (8). Then, $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and the covariance matrix for $\boldsymbol{\epsilon}$ is denoted by $\boldsymbol{\Sigma} = [\sigma_{ij}]$, where $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$. It is assumed that $\boldsymbol{\Sigma}$ is non-singular and that $\mathbf{X}$ is a full-rank column.

3 Application to Soybean Yield Data

In this section, the methodology studied in this paper will be illustrated using a set of 74 observations of soybean yield in $[t \times ha^{-1}]$ (ton per hectare), harvested in 2016/2017, collected from an experimental field of the Space Statistics Laboratory of Universidade Estadual do Oeste do Paraná, *campus* Cascavel, in a commercial area of 127.18 ha of grain production, located near the municipality of Cascavel, in the Western region of Paraná, with geographic coordinates are approximately 24.95 S e 53.57 W, and with a mean altitude of 650 m.

The Table 1 shows the parameters estimated using the ML for the model selected by cross-validation at $\kappa = 3$. The standard asymptotic errors are in parenthesis. The ϕ_3 parameter was fixed at 1.730, obtained by the least squares estimation method. In this case, as mentioned in Subsection 2.1, the practical range is $a = 0.514 \times \phi_3 = 0.889\ km$. The method of ordinary least squares was used to obtain the value of parameter ϕ_3 to be considered as fixed, and the initial values of ϕ_1 and ϕ_2, which helped to overcome the problem found with the singularity of the covariance matrix generated from these data. That said, as mentioned in Section 2.2, we consider the parametric form $\boldsymbol{\Sigma} = \tau_1\mathbf{I} + \tau_2\mathbf{R}(\phi_3)$, with $\tau_1 = \phi_1$, and $\tau_2 = \dfrac{\phi_2}{\phi_3^{2\kappa+1}}$.

Table 1: Parameters estimated by ML method for covariance family of Generalized Wendland for variable response of soybean yield in $(t \times ha^{-1})$ in the agricultural year of 2016/2017, with smoothing parameter $\kappa = 3$, and fixed ϕ_3.

Family	Intercept $\hat{\beta}_0$	P $\hat{\beta}_1$	K $\hat{\beta}_2$	Ca $\hat{\beta}_3$	Mg $\hat{\beta}_4$	Nugget $\hat{\tau}_1$	Sill $\hat{\tau}_2$	$SDI(\%)$
GW	2.759	0.004	0.331	-0.001	0.107	0.246	0.049	8.534
	(0.364)	(0.006)	(0.584)	(0.010)	(0.099)	(0.055)	(0.043)	$----$

P: phophorus, K: potassium, Ca: calcium, Mg: magnesium, SDI: Spatial Dependency Index. With $\kappa = 3$ and $FM = 0.512$ (Model factor). Standard deviations are presented in parentheses.

To identify the cutoff point, the jackknife-after-bootstrap method was used, from 1000 resamples of the variable soybean yield. We obtained a percentile range of 95% equal to $(0.062, 0.436)$, with the quantile 0.436 being used as the cutoff point of the dataset. The potentially influential points are #39, #52, #61, and #64, corresponding to the values of 2.927, 3.511, 3.543, and 3.400 $(t \times ha^{-1})$, respectively; see Figure 1.

DOI: 10.5540/03.2025.011.01.0399

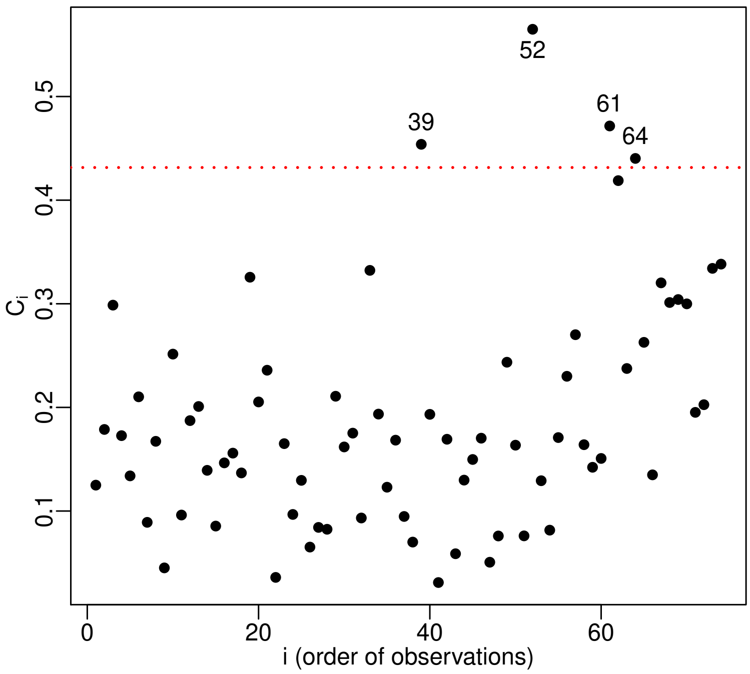


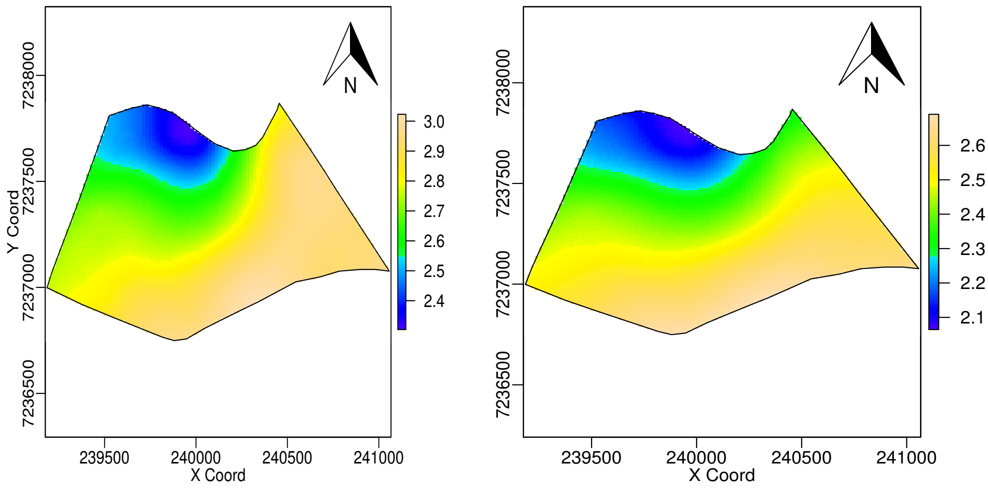
Figure 1: Local influence of maximum normal curvature. The cutoff line (horizontal dotted line) corresponds to the 95% quantile of the bootstrap distribution generated by the nBe^{-1} values, with $B = 1000$ Jackknife-after-Bootstrap technique samples. Source: Of the authors.

Regarding the spatial dependence structure, note that the changes were more decisive when considering that parameter τ_2 declined by almost half. This component seems to cause serious problems in terms of measuring the spatial dependence of the data.

When observations #52, #61, and #64 were removed together or individually, it only changes the sign of the estimated parameter β_3, and the influence of the explanatory variable Ca on the variable response. However, the spatial structure estimates an increase for the parameter τ_1 and a decreased to τ_2, highlighting the strictly inverse behavior in the individual removal of the #52 observation. Thus, in fact, this observation influences the structure that models the spatial dependence. Note that the asymptotic standard errors for the component that defines the spatial structure were increased only when excluding observation #52.

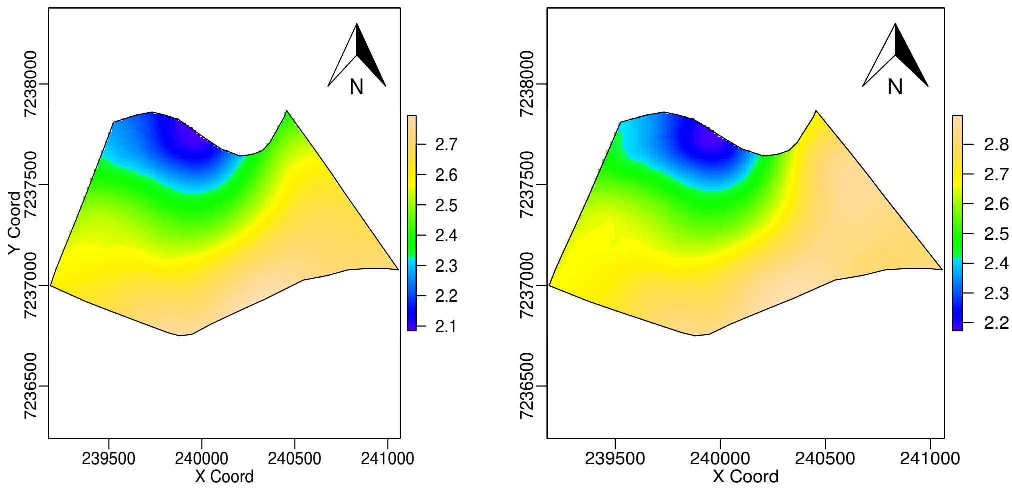
A positive relative variation ($\frac{9.147 - 8.534}{8.534} \times 100 = 7.183$) can be observed for the spatial dependency index, when observation #52 is removed. The influence of these observations on the construction of the maps was also verified; see Figures 2 and 3.

The predicted values were obtained using a universal kriging. Note that changes in the maps obtaining when the observations were removed in all scenarios, where it is more noticeable when observation #52 is excluded, as shown in Figure 3(b).



(a) (b)

Figure 2: Maps for the predicted values of soybean yield by $(t \times ha^{-1})$ in the 2016/2017 crop for Wendland's covariance function with $\kappa = 3$: (a) with all observations and (b) without observations #39, #52, #61, and #64. Source: Of the authors.



(a) (b)

Figure 3: Maps for the predicted values of soybean yield by $(t \times ha^{-1})$ in the crop of 2016/2017 for Wendland's covariance function with $\kappa = 3$: (a) without the presence of observations #52, #61 and #64, and (b) individually without the observation #52. Source: Of the authors.

4 Final Considerations

The results of this study were extended to measures of local influence with a spatial covariance matrix for covariance matrices belonging to the Wendland family. Its flexibility through a smoothing parameter placeds it as a strong competitor to the Matérn covariance family. As a consequence, the results of the data sets with one soybean yield showed that it is possible to detect influential observations. These conclusions are based on the analysis of the asymptotic standard errors, considering the individually influential observations or their sets as an alternative to considering their joint effect.

References

[1] M. Bevilacqua, T. Faouzi, R. Furrer, and E. Porcu. "Estimation and prediction using generalized wendland covariance functions under fixed domain asymptotics". In: **The Annals of Statistics** 47.2 (2019), pp. 828–856.

[2] A. Chernih and S. Hubbert. "Closed form representations and properties of the generalised Wendland functions". In: **Journal of Approximation Theory** 177 (2014), pp. 17–33.

[3] A. Chernih, I. H. Sloan, and R. S. Womersley. "Wendland functions with increasing smoothness converge to a Gaussian". In: **Advances in Computational Mathematics** 40.1 (2014), pp. 185–200.

[4] R. D. Cook. "Assessment of local influence". In: **Journal of the Royal Statistical Society. Series B (Methodological), 133–169** (1986).

[5] N. A. C. Cressie. **Statistics for spatial data**. Wiley series in probability, mathematical statistics. Applied probability, and statistics section, 1993.

[6] P. J. Diggle and P. J. Ribeiro. **Model-based Geostatistics**. 2007.

[7] T. Gneiting. "Compactly supported correlation functions". In: **Journal of Multivariate Analysis** 83.2 (2002), pp. 493–508.

[8] M. A. U. Opazo, J. A. Borssoi, and M. Galea. "Influence diagnostics in Gaussian spatial linear models". In: **Journal of Applied Statistics** 39.3 (2012), pp. 615–630.

[9] M. Santos. "Uma releitura inferencial e análise de diagnósticos em modelos geoestatísticos". PhD thesis. Universidade Federal de Pernambuco, 2022.

[10] E. Seidel and M. S. Oliveira. "A classification for a geostatistical index of spatial dependence". In: **Revista Brasileira de Ciência do Solo** 40 (2016).

[11] E. J. Seidel and M. S. Oliveira. "Novo índice geoestatístico para a mensuração da dependência espacial". In: **Revista Brasileira de Ciência do Solo** 38.3 (2014), pp. 699–705.

[12] H. Wendland. "Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree". In: **Advances in Computational Mathematics** 4.1 (1995), pp. 389–396.

[13] Z. Wu. "Compactly supported positive definite radial functions". In: **Advances in Computational Mathematics** 4.1 (1995), p. 283.

[14] V. P. Zastavnyi. "On some properties of Buhmann functions". In: **Ukrainian Mathematical Journal** 58.8 (2006), pp. 1184–1208.

DOI: 10.5540/03.2025.011.01.0399