


# On Nonparametric Estimation of Covariograms

Adam Bilchouris   
La Trobe University,  
Melbourne, Australia

Andriy Olenko   
La Trobe University,  
Melbourne, Australia

---

## Abstract

The paper overviews and investigates several nonparametric methods of estimating covariograms. It provides a unified approach and notation to compare the main approaches used in applied research. The primary focus is on methods that utilise the actual values of observations, rather than their ranks. We concentrate on such desirable properties of covariograms as bias, positive-definiteness and behaviour at large distances. The paper discusses several theoretical properties and demonstrates some surprising drawbacks of well-known estimators. Numerical studies provide a comparison of representatives from different methods using various metrics. The results provide important insight and guidance for practitioners who use estimated covariograms in various applications, including kriging, monitoring network optimisation, cross-validation, and other related tasks.

*Keywords:* nonparametric estimation, covariance, semivariogram, covariogram, auto-correlation, positive-definiteness, spatial, R.

---

## 1. Introduction

Estimating covariograms and semivariograms of stochastic processes and random fields is an important problem in time series and spatial statistics analysis. Most applied statisticians rely on the classical estimator, as evidenced by its implementation in most R and Python statistical packages used for the analysis of time series and spatial data, see Appendix A. The standard estimator has seen routine use in statistical work due to its intuitiveness (Jenkins and Watts 1969, p. 174). In fact, the correlation of a times series has been studied as early as 1905, see Cave-Browne-Cave and Pearson (1905). Issues with the standard estimator were also known for some time, but its use persisted, see for instance Jenkins and Watts (1969, p. 185), Яглом (1981, pp. 71-77), and Yaglom (1987, pp. 239-245). The study of variograms is often contributed to Matheron (1962), however, Cressie (1993, p. 58) refers to papers as early as 1941 discussing the same idea, albeit under different names. Matheron's variograms were originally used in geostatistics but have now seen use in spatial statistics as a whole, see the references in the following sections. The literature now presents several alternative approaches that are less known, but are often more robust and accurate estimates with desirable properties, as will be demonstrated in the paper. To the best of our knowledge, there are no comprehensive overview publications thoroughly discussing and comparing these methods. Thus, the initial aim of this paper was to collect and compare all known approaches, however, the literature search revealed

that it was an overly ambitious task for a single paper. Consequently, we restricted the scope of this manuscript to nonparametric estimators based on actual observations. Other significant areas, such as methods based on ranks and parametric approaches, are only briefly mentioned.

The paper discusses several different estimators of the covariance function known in the literature or variations upon well-known estimators. We tried to avoid the discussion of various technical modifications but rather presented nonequivalent methods or scenarios. Throughout the paper, we will use the terms *covariance*, *autocovariance*, and *covariogram* interchangeably.

First, when presenting these methods, the main emphasis was on several important properties of these estimators, including bias and positive-definiteness. The primary concern was homogeneous and isotropic random fields and stationary time series, see Ma (2007), characterised by covariance functions dependent only on distances. Nonetheless, most of the materials presented are applied or can be extended to the estimation of homogeneous and directional dependencies. This paper primarily discusses issues important for applications. For proofs and theoretical results concerning various models of positive-definite functions and theoretical inference, we recommend, for example, Yadrenko (1983), Ivanov and Leonenko (1989), and Porcu, Furrer, and Nychka (2021b).

Secondly, simulation studies were conducted for the mentioned methods, wherein their performance was compared using several covariance models with different properties. This included Gaussian, Bessel, and Cauchy covariance models, which exhibit short-, cyclic, and long-range dependencies. When it was possible, in this analysis, we utilised some available R packages. However, as most non-standard methods were only available in publications without code, we developed the corresponding R code to illustrate them and compare their performance. The comparison methods included metrics that compared distances between the theoretical covariogram, from which random processes were simulated, with the aforementioned nonparametric estimators. Also, a cross-validation approach based on kriging that utilised the estimated functions was used.

Most theoretical papers primarily focus on the asymptotic properties of estimators. It is well-known and rigorously proved in the literature that, within a fixed distance range, many estimators exhibit desirable asymptotic properties. However, the analysis presented in this paper reveals that if the distance range in the estimated covariogram is more than 10-20% of the total available distance, the performance of some estimators, including classical ones, yields unreliable results. This supports the idea presented by Yaglom (1987, p. 237). Also, it was demonstrated that accurate estimation of covariograms of long-range dependent data necessitates much longer distances than those typically used for estimation intervals. Consequently, most approaches that reliably handle short-range dependent data are inadequate for long-range dependent scenarios. Furthermore, this study demonstrates that for cases involving cyclic behaviours of the covariogram, certain foundational principles of well-established approaches are invalid, potentially leading to unreliable results.

The paper is structured as follows. Section 2 presents two key functions for estimating dependencies in the literature, covariograms and semivariograms. Section 3 overviews some principles of constructing and properties of the most frequently used classical estimator. Sections 4 through 9 introduce and analyse the main nonparametric estimation approaches examined within this study. Section 10 provides a brief overview of alternative approaches for estimating temporal and spatial dependencies in the literature. Section 11 presents numerical and simulation studies that compare the considered methods. Finally, the concluding section offers conclusions and outlines some problems for future research. The references include key publications that can serve as starting resources for exploring the considered methods. Some technical proofs and complimentary results are included in the Appendices.

Simulations, numerical computations and plotting in the paper were performed using the software R (version 4.1.0). The code is freely available in the folder ‘Research materials’ from the website <https://sites.google.com/site/olenkoandriy/>.

## 2. Semivariogram or covariogram?

In spatial statistics, there are two similar functions used in the analysis of the dependency structure of random fields, the semivariogram, and the covariance function.

The *semivariogram* characterises the second-moment structure of the increments of the random field, i.e.

$$\gamma(\mathbf{h}) := \frac{1}{2} \text{Var}(Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})), \quad \mathbf{s}, \mathbf{h} \in \mathbb{R}^d,$$

whereas the *covariance function* is a measure of spatial covariance,

$$C(\mathbf{h}) := E((Z(\mathbf{s}) - \mu)(Z(\mathbf{s} + \mathbf{h}) - \mu)), \quad \text{where } \mu := E(Z(\mathbf{s})).$$

There are several reasons why someone may want to use the semivariogram over the covariance function when performing estimation and analysis. For example, the semivariogram estimator does not require the mean in its estimation. The mean, if not known, is estimated from the data, and if the data is spatially correlated, a bias may be present in the standard estimate of the mean, and thus the estimated covariance function (Hristopulos 2020, p. 103). Under the assumption of weak stationarity, a Gaussian random field can be defined entirely by its first and second moments, i.e. its mean and covariance (Chilès and Delfiner 2012, p. 17). Under weak stationarity, the semivariogram and covariance functions are related through

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}). \quad (1)$$

However, under intrinsic stationarity, this does not hold, unless the semivariogram is bounded by a finite value (Chilès and Delfiner 2012, p. 31).

The following sections employ both functions, where it is more convenient or depending on how they were introduced in the literature, with some bias towards the covariance function to discuss positive-definiteness.

## 3. Standard estimators

The estimation of the spatial covariance function generalises the construction for the time series autocovariance function, aiming to utilise all points  $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$  with the separation vector  $\mathbf{h} \in \mathbb{R}^d$ , i.e.  $\mathbf{s} - \mathbf{t} = \mathbf{h}$ .

**Estimator 1.** (*standard or classical estimated covariance function*) (Cressie 1993, pp.69-70) The empirical spatial covariance function is given by:

$$\hat{C}(\mathbf{h}) := \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X}), \quad \mathbf{t}_i \in \mathbb{R}^d, i = 1, \dots, n,$$

where  $N(\mathbf{h})$  denotes all pairs of points  $(\mathbf{t}_i, \mathbf{t}_j)$  whose difference is equal to the vector  $\mathbf{h}$ ,

$$N(\mathbf{h}) := \{(\mathbf{t}_i, \mathbf{t}_j) : \mathbf{t}_i - \mathbf{t}_j = \mathbf{h}, i, j = 1, \dots, n\},$$

and  $\bar{X}$  denote the sample mean  $\bar{X} := n^{-1} \sum_{i=1}^n X(\mathbf{t}_i)$ .

This estimate  $\hat{C}(\mathbf{h})$  is also called the *covariogram* and is a method of moments estimator for the covariance function.

**Estimator 2.** (*standard semivariogram estimator*) The corresponding estimator of the semivariogram proposed by Matheron (1962) is

$$\hat{\gamma}(\mathbf{h}) := \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (X(\mathbf{t}_i) - X(\mathbf{t}_j))^2.$$

Often there might not be enough pairs of points whose difference is  $\mathbf{h}$ , so one may consider adding some neighbourhood around  $\mathbf{h}$ , controlled by  $\varepsilon$ , which can increase the number of pairs of points for the lag  $\mathbf{h}$  (Schabenberger and Gotway 2005, p. 153). A method to select this neighbourhood, for the isotropic case, will be presented in the next subsection. Computationally, this estimator is not ideal as one must determine which pairs of points have a separation vector equal to  $\mathbf{h}$ , which requires searching all possible pairs of points for each  $\mathbf{h}$ .

For an isotropic covariance function, this estimator is modified by  $\hat{C}(\tau)$  that uses points  $\mathbf{t}_i$  and  $\mathbf{t}_j$ , whose separation vector has the length  $\tau$ . This is equivalent to finding all points which fall on a circle of radius  $\tau$  around each point  $\mathbf{t}_i$ .

Appendix C provides a pseudocode for computing the empirical isotropic covariance function through the above method. For simplicity, it is assumed that the mean is zero, and the covariance function is estimated for some lag distance  $\tau$ . For each distance  $\tau$ , this algorithm has a time complexity of  $\mathcal{O}(N^2)$ .

For a sufficiently sampled random field on some bounded domain there are usually a high number of points which can be considered when computing the empirical covariance function at short distances, however, if the distance is too small, there may be no pairs of points. For larger distances, the number of points decreases, as points near the edges of the domain can contribute only in certain directions. As a result, one must carefully consider the distances they wish to use when estimating the covariance function.

When the domain size of a random field remains the same but the number of points increases, which is called *infill sampling*, the statistical properties of Estimator 2 change. For example, it is not consistent, see Lahiri (1996). Lahiri, Lee, and Cressie (2002) and Lee and Lahiri (2002) consider estimators using spatial subsampling to fit a variogram model. The proposed regions can cover a wide variety of shapes such as spheres, ellipsoids, and star-shaped sets (Lee and Lahiri 2002, p. 839). Furthermore, Lahiri, Lee, and Cressie (2002, p. 75) state when infill sampling is present in a region, it leads to strong dependence amongst the observations, similar to that of long-range dependence. This can cause issues when performing estimation, such as slower convergence to regression parameters.

**Estimator 3.** (*directional estimated covariance function*) A variation of the standard estimator exists, called the directional estimated covariance function, which considers a specific direction only. In this case, to get enough points along the selected azimuthal angle, a conic region (called a search cone) is used (Ma 2019, p. 310). The width of the cone depends on angle tolerance and bandwidth.

In the case of an isotropic random field, the directional estimated covariance functions should be similar regardless of the angles. For anisotropy, there should be a difference between some directions, assuming the anisotropy is strong enough.

### 3.1. On some properties of standard estimators

Now, let us discuss some desirable properties of covariance function estimators.

A covariance function must be *positive-definite*, i.e.

**Definition 3.1.** A function  $f : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{C}$  is called positive-definite if for any  $n \in \mathbb{N}$ ,  $\mathbf{t}_1, \dots, \mathbf{t}_n \in \mathbb{R}^d$  and  $a_1, \dots, a_n \in \mathbb{C}$ , it holds

$$\sum_{i=1}^n \sum_{j=1}^n c_i \bar{c}_j f(\mathbf{t}_i, \mathbf{t}_j) \geq 0.$$

For a weakly stationary random field it is satisfied for  $f(\mathbf{t}_i, \mathbf{t}_j) = C(\mathbf{t}_i, \mathbf{t}_j) = C(\mathbf{t}_i - \mathbf{t}_j)$ .

If an estimator is not positive-definite, then the result is not a valid covariance function, just a function that in some sense is close to a covariance function. In many cases, such estimates can be problematic, for example, kriging requires a valid covariance function (Cressie 1993); the

spectral density estimates based on non-positive-definite functions may take negative values, among other potential issues.

The following properties of the classical estimator illustrate several potential issues that require careful consideration when employing it.

- (i) The standard estimator does not always produce a positive-definite result, when the number of pairs of points  $|N(\tau)|$  varies as  $\tau$  varies, which can be shown by a simple example, see Appendix B.
- (ii) In fact, the classical estimator, is always positive-definite if  $|N(\tau)|$  is the same constant for all  $\tau$ , see Appendix D. A well-known variant of this estimator in the one-dimensional case, which is biased and positive-definite, is

$$\hat{C}^{**}(h) := \frac{1}{N} \sum_{t=1}^{N-h} (X(t) - \bar{X})(X(t+h) - \bar{X}),$$

and the associated autocorrelation estimator is

$$\hat{\rho}(h) := \frac{\sum_{t=1}^{N-h} (X(t) - \bar{X})(X(t+h) - \bar{X})}{\sum_{t=1}^N (X(t) - \bar{X})^2}. \quad (2)$$

- (iii) Similar to other statistics, *unbiasedness and consistency* are typically desirable properties of covariance function estimators. There is a substantial theoretical literature on this topic. However, as will be demonstrated, despite the presence of pointwise convergence of the estimators for each fixed lag, the classical and many other estimators do not exhibit uniform convergence and may display undesirable behaviour in instances involving finite spatial domains or large distances.

Hassani (2009) and Hassani, Leonenko, and Patterson (2012) provide a surprising result about the one-dimensional autocorrelation estimator, (2), when summing over all lags  $h \geq 1$ , when  $N \geq 2$ ,

$$\sum_{h=1}^{N-1} \hat{\rho}(h) = -\frac{1}{2},$$

which holds regardless of observed values. This has implications when considering long-memory, as the sum of the estimates converges to  $-1/2$ , whilst it needs to be divergent in the long-memory case.

Appendix E proves that a similar property holds for the spatial case too. Namely,

$$\sum_{\mathbf{h} \in H} \hat{\rho}(\mathbf{h}) = -1 \quad \text{and} \quad \sum_{\mathbf{h} \in H} \frac{|N(\mathbf{h})|}{N} \tilde{\rho}(\mathbf{h}) = -1,$$

where  $H := \{\mathbf{h} : \mathbf{h} = \mathbf{t}_i - \mathbf{t}_j, i, j = 1, \dots, N, i \neq j\}$  is the set of all separation vectors amongst different observation locations, and  $\hat{\rho}(\mathbf{h})$  and  $\tilde{\rho}(\mathbf{h})$  denote the empirical correlations that correspond to the standard covariogram estimators with constant and varying  $|N(\mathbf{h})|$ .

The classical empirical covariograms will have more pronounced waves for the estimated values  $\tilde{\rho}(\mathbf{h})$  as the multiplicative weights are within  $[0, 1]$ . These waves, as will be shown in the simulation results, are artefacts in the estimation process caused by fewer samples being available to compute the empirical covariance function as the estimation lag increases. Furthermore, analogous identities can be obtained for other commonly used weighted estimators where the weights depend only on the separation lag (or distance) between the locations of samples, rather than the sampled values at the locations.

- (iv) Furthermore, the classical estimator does not hold the aforementioned relation (1) between the variogram and covariance function, see Appendix F.

### 3.2. Selection scheme for the standard method

When estimating the covariance function, the data one is dealing with, whether it is simulated, or real, will only be sampled finitely. As a result, the empirical covariance function cannot be computed at all distances, except when there is a sufficient number of points for each distance (such as in the case of functional data).

**Estimator 4.** (*bin-type estimators*) There exist several variations, even within the standard method, which specify the sampled values to be used for each particular distance  $\tau$  (or the vector  $\mathbf{h}$ ) and choose the extrapolation technique between distinct  $\tau_i$  (or  $\mathbf{h}_i$ ).

Krajewski and Duffy (1988) provide two schemes to select the distances considered, and the error region,  $\Delta$ , when estimating an isotropic covariance function. Adding the error region is common in practice so a sufficient number of points are included for each distance  $\tau$ . We will only consider selection scheme 1 in Krajewski and Duffy (1988), as it is simpler when working with a square lattice. The error region is constructed around a circle of radius  $\tau$ , resulting in a region bounded by two circles with radii  $\tau + \Delta$  and  $\tau - \Delta$ . The scheme is specified as follows

1.  $\tau_{k+1} - \tau_k = \text{constant}$ ,  $k = 1, \dots, K - 1$ .
2.  $\tau_1 := \frac{1}{n} \sum_{i=1}^n \min [d_{ij} : j = 1, \dots, n : j \neq i]$ , where  $d_{ij}$  is the distance from point  $\mathbf{t}_i$  to  $\mathbf{t}_j$ .
3.  $\Delta := \tau_1/2$ .
4.  $K$  is fixed (usually  $K$  is between 10 and 15).

This does not allow any control for the number of pairs of points  $n_k$  for some  $\tau_k$ . If there are insufficient samples used when computing the empirical covariance function, even in the one-dimensional case, there will be waves present, see Yaglom (1987, p. 239) and Яглом (1981, p. 74). In Journal and Huijbregts (1978, p. 194), they recommend the number of pairs for each lag to be at least 30, and the maximum lag to be less than half the maximum separation lag, which were supported by the simulation studies in Cressie (1988). The choice of the maximum separation lag will determine how prominent the waves will be.

### 3.3. Estimators on the ball

It is difficult to mention numerous generalisations of the classical estimator in the literature. We only provide an example, although the case of sampled functional data is not considered here, of the following estimator by Дыховичный (1983) for the cases of functional data with observation windows that include a ball. It is the continuous equivalent of Estimator 3.

**Estimator 5.** (*classical estimated covariance function for functional data*) An estimator of a covariance function of a zero-mean isotropic random field on an  $n$ -ball of radius  $R$  centred at the origin,  $V_R(0)$ , is given by

$$\hat{C}(r) := \frac{1}{U_n(R-r)} \int_{V_{R-r}(0)} X(\mathbf{s}) \left( \frac{1}{\omega_n(r)} \int_{S_r(\mathbf{s})} X(\mathbf{t}) m_n^{(r)}(d\mathbf{t}) \right) d\mathbf{s},$$

where  $r$  is the distance lag,  $S_r(\mathbf{s})$  is a sphere of radius  $r$  centred at  $\mathbf{s}$ ,  $U_n$  and  $\omega_n$  are the volume of the ball and surface area of the sphere, respectively, and  $m_n^{(r)}$  is a Lebesgue measure on the sphere.

## 4. Variogram-based estimators

Although we are mainly concerned with covariance function estimation, it is important to note that in the cases when it is appropriate (see the discussion in (iv) of Section 3), one can employ the following approach.



**Estimator 6.** (*variogram-based estimated covariance function*)

$$\hat{C}(\mathbf{h}) = \hat{C}(\mathbf{0}) - \hat{\gamma}(\mathbf{h}) = \widehat{\text{Var}}(X) - \hat{\gamma}(\mathbf{h})$$

where  $\hat{\gamma}(\mathbf{h})$  is a variogram estimator known in the literature.

For example, one can use the popular estimator proposed by [Cressie and Hawkins \(1980\)](#)

$$2\hat{\gamma}(\mathbf{h}) := \left( \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} |X(\mathbf{t}_{i+\mathbf{h}}) - X(\mathbf{t}_i)|^{1/2} \right)^4 / (0.457 + 0.494/N(\mathbf{h})),$$

or its robust generalisation in [Genton \(1998, p. 214\)](#).

## 5. Kernel regression estimators

Below is an estimator similar to that of Nadaraya-Watson kernel regression, proposed by [Hall, Fisher, and Hoffmann \(1994\)](#) and [Hall and Patil \(1994\)](#).

**Estimator 7.** (*kernel-based estimated covariance function*)

$$\hat{C}_H(\mathbf{t}) := \frac{\sum_i \sum_j \check{X}_{ij} K((\mathbf{t} - (\mathbf{t}_i - \mathbf{t}_j))/b)}{\sum_i \sum_j K((\mathbf{t} - (\mathbf{t}_i - \mathbf{t}_j))/b)},$$

where  $\check{X}_{ij} := (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})$ .

One has a choice for the kernel function  $K(\cdot)$  used, however, it must have the properties of symmetric probability density, resulting in  $\hat{C}_H(-\mathbf{t}) = \hat{C}_H(\mathbf{t})$  ([Hall and Patil 1994, p. 403](#)).

A method in the paper by [Hall and Patil \(1994\)](#) was proposed to ensure the positive-definiteness of the estimator  $\hat{C}_H$ , by manipulating its Fourier transform,  $\mathcal{F}(\hat{C}_H(\mathbf{t}))$ , which forces negative values of this Fourier transform to zero before performing a Fourier inversion.

The approach can be utilised in more general settings, where an estimate of the covariance function is a non-positive-definite function. Namely,

**Estimator 8.** (*kernel-based positive-definite estimated covariance function*)

1. Compute  $\hat{C}_H(\mathbf{t})$ .
2. Compute  $\mathcal{F}(\hat{C}_H(\mathbf{t}))$ .
3. Set  $\hat{\mathcal{F}}(\hat{C}_H(\mathbf{t})) = \mathcal{F}(\hat{C}_H(\mathbf{t}))$  if  $\mathcal{F}(\hat{C}_H(\mathbf{t})) > 0$ , and  $\hat{\mathcal{F}}(\hat{C}_H(\mathbf{t})) = 0$  otherwise, for every frequency  $\boldsymbol{\theta}$ .
4. Compute the inverse Fourier transform of  $\hat{\mathcal{F}}(\hat{C}_H(\mathbf{t}))$  and use it as a new estimator  $\tilde{C}(\mathbf{t})$ .

A related approach was provided in the paper by [Hall, Fisher, and Hoffmann \(1994, p. 2118\)](#), which considered forcing the estimated covariance function down to zero linearly starting at the point  $T_1 > 0$  and ending at  $T_2 > T_1$ . For example, for Estimator 7,  $\hat{C}_H(\mathbf{t})$ , in the one-dimensional case, it gives:

**Estimator 9.** (*kernel-based truncated estimated covariance function*)

1. Compute the empirical covariance function as

$$\hat{C}_1(t) := \begin{cases} \hat{C}_H(t), & 0 \leq t \leq T_1 \\ \hat{C}_H(T_1)(T_2 - t)(T_2 - T_1)^{-1}, & T_1 < t \leq T_2, \\ 0 & t > T_2. \end{cases}$$

2. Perform a Fourier transform (which is a Fourier-cosine transform due to the real symmetric nature of  $\hat{C}_1$ )

$$\mathcal{F}^c(\hat{C}_1(t)) = 2 \int_0^\infty \cos(\theta t) \hat{C}_1(t) dt.$$

3. Determine  $\hat{\theta} = \inf \{ \theta > 0 : \mathcal{F}^c(\hat{C}_1(t)) < 0 \}$ . Then, set all values of  $\mathcal{F}^c(\hat{C}_1(t))$ , when  $\theta > \hat{\theta}$ , to zero.

4. Perform an inversion to obtain

$$\tilde{C}(t) := \int_{-\hat{\theta}}^{\hat{\theta}} (2\pi)^{-1} \mathcal{F}^c(\hat{C}_1(t)) \cos(\theta t) d\theta.$$

This method and its resulting estimator will be referred to as *Hall's estimator*.

One issue with bringing the tail of the estimator down to zero is it can no longer be used for long-range dependent random fields (or long-memory processes), as the absolute values of their covariance functions are non-integrable. This is the same issue as Estimator 20, which will be introduced later, although this can help to remove the waves mentioned earlier. Also, the last method discussed in this section, making the estimated function positive-definite, may fail if  $\hat{\theta}$  is very close to zero.

The kernel-based estimated covariance function has been adapted in the literature for semivariogram estimation. For example, García-Soidán, Febrero-Bande, and González-Manteiga (2004) considers a variant to estimate an isotropic semivariogram, García-Soidán and Menezes (2012) considers it for an indicator variogram estimator, Cuevas, Porcu, and Vallejos (2013) construct a cross-variogram estimator, and Yu, Mateu, and Porcu (2007) construct a variable nearest neighbour estimator for the semivariogram.

## 6. $Q_n$ estimators

The papers by Genton (1998) and Ma and Genton (2000) provide a robust estimator for the variogram and covariance function based on the  $Q_n$  estimator of Rousseeuw and Croux (1993).  $Q_n$  is an estimator of scale, and has a breakdown point of 50%, meaning up to 50% of observations can be outliers before an incorrect (arbitrarily large) result is given (Rousseeuw and Croux 1993, p. 1277). Unfortunately, the covariance version of this estimator only has a breakdown point of 25%, which is on par with the interquartile range (Rousseeuw and Croux 1992, p. 78; Ma and Genton 2000; p. 655).  $Q_n$  is defined as follows:

$$Q_n(\mathbf{X}) := c \{ |X_i - X_j|; i < j \}_{(m)},$$

where  $\mathbf{X}$  is a vector of observations  $(X_1, \dots, X_n)'$ ,  $c$  is a constant for consistency, and  $m = \lceil ((\binom{n}{2} + 2) / 4) + 1 \rceil$ , where  $\lceil \cdot \rceil$  denotes the integer part. Thus, it computes the  $m^{\text{th}}$ -order statistic of the absolute differences of the observations. For large  $n$ , this is approximately the first quartile, and in the case of a Gaussian distribution, the consistency constant can be selected as  $c = 2.2191$  (Ma and Genton 2000, p. 655). For  $Q_n$ , the outliers refer to the absolute pairwise differences, not the observed values in  $\mathbf{X}$ .

In the variogram case, Genton (1998, pp. 214-216) considers constructing a random field of differences at the spatial lag  $\mathbf{h}$ ,  $V(\mathbf{h}) = X(\mathbf{t} + \mathbf{h}) - X(\mathbf{t})$ , which has zero mean and variance  $2\gamma(\mathbf{h})$ . From this, a sample  $\{V_1(\mathbf{h}), \dots, V_{|N(\mathbf{h})|}\}$  of  $V(\mathbf{h})$  which corresponds to the sample  $\{X(\mathbf{t}_1), \dots, X(\mathbf{t}_N)\}$  of  $X$ , gives

$$Q_{|N(\mathbf{h})|} := 2.2191 \{ |V_i(\mathbf{h}) - V_j(\mathbf{h})| \}_{(m_S)},$$

where  $m_S = \lceil (|N(\mathbf{h})|/2)^{+1} \rceil$ , and finally,  $2\hat{\gamma}(\mathbf{h}) = (Q_{|N(\mathbf{h})|})^2$  (Genton 1998, pp. 214-216).



The estimator about to be introduced, and the estimator to follow, are motivated by the following identity, which assumes finite second-moment (i.e.  $X, Y \in L^2$ ) (Huber 1981, p. 202):

$$\text{Cov}(X, Y) = \frac{1}{4ab} (\text{Var}(aX + bY) - \text{Var}(aX - bY)),$$

where  $a, b$  are arbitrary constants, although they should be chosen on a similar scale to avoid strange results. In the following two estimators, as we are considering processes with unit variance, we have  $a = b = 1$ .

**Estimator 10.** (*quantile-based estimated covariance function*) (Ma and Genton 2000, p. 665) The  $Q_n$  estimator for the covariance function is given by

$$\hat{C}_Q(h, \mathbf{X}) := \frac{1}{4} (Q_{n-h}^2(\mathbf{X}_{1:(n-h)} + \mathbf{X}_{(h+1):n}) - Q_{n-h}^2(\mathbf{X}_{1:(n-h)} - \mathbf{X}_{(h+1):n})),$$

where the vector  $\mathbf{X}_{1:(n-h)}$  consists of first  $n - h$  observations of the vector  $\mathbf{X}$  and the vector  $\mathbf{X}_{(h+1):n}$  includes last  $n - h$  observations.

For this estimator, the autocorrelation function cannot be obtained from the autocovariance in the usual way,  $\rho(h) = C(h)/C(0)$ , as this is not bounded between  $-1$  and  $1$ . Instead, Ma and Genton (2000, p. 666) provide the following robust estimate.

**Estimator 11.** (*quantile-based estimated correlation function*)

$$\hat{\rho}_Q(h, \mathbf{X}) := \frac{Q_{n-h}^2(\mathbf{X}_{1:(n-h)} + \mathbf{X}_{(h+1):n}) - Q_{n-h}^2(\mathbf{X}_{1:(n-h)} - \mathbf{X}_{(h+1):n})}{Q_{n-h}^2(\mathbf{X}_{1:(n-h)} + \mathbf{X}_{(h+1):n}) + Q_{n-h}^2(\mathbf{X}_{1:(n-h)} - \mathbf{X}_{(h+1):n})}.$$

Unlike other estimators, this estimator is location-free, as it does not depend upon any knowledge of the location of the points. For example, the standard estimator, Estimator 1, depends on the locations between points, whereas this one considers pairwise differences. This estimator has a breakdown point of 25%, which is the highest possible value when considering autocovariance (Ma and Genton 2000, p. 665).

### 6.1. $P_n$ estimator

Another robust estimator was proposed by Tarr, Müller, and Weber (2013), based on the  $P_n$  estimator of pairwise means in Tarr, Müller, and Weber (2012), and is related to U-statistics.

**Estimator 12.** (*pairwise mean-based estimated covariance function*)

$$\hat{C}_P(h) := \frac{1}{4} (P_{n-h}^2(X_{1:(n-h)} + X_{(h+1):n}) - P_{n-h}^2(X_{1:(n-h)} - X_{(h+1):n})),$$

where  $P_n = c(M_n^{-1}(0.75) - M_n^{-1}(0.25))$ , with  $c \approx 1.048$  being a correction factor ensuring  $P_n$  is consistent for the standard deviation when the observations are Gaussian.  $M_n(\cdot)$  denotes the empirical distribution of pairwise means,

$$M_n(t) := \frac{2}{n(n-1)} \sum_{i < j} \mathbf{1} \left( \frac{X_i + X_j}{2} \leq t \right), \quad t \in \mathbb{R},$$

and  $\mathbf{1}(\cdot)$  denotes the indicator function.

$P_n$  has a breakdown point of 13.4%, whilst  $Q_n$  has a breakdown point of 50%, although for Gaussian asymptotic efficiency,  $P_n$ 's is 0.86 whereas  $Q_n$ 's is 0.82 (Tarr, Müller, and Weber 2012, p. 193; Tarr, Müller, and Weber 2013).

## 7. Tapered estimates

### 7.1. Edge effect

An important factor to consider when working with stochastic processes, especially as the dimension increases, is the *edge effect*. Let  $D$  be some study domain, such as a finite rectangular lattice  $P_n := \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_d\}$  in  $d$ -dimensional integer space,  $\mathbb{Z}^d$ . The points on the boundary of  $D$  will have their nearest neighbours lying outside of  $D$ , meaning observations are unavailable at those locations. As  $d$  increases, so does the number of boundary points. For example, for  $N = 100$  points, a one-dimensional interval has two points, or 2%, on the edge, but for a square with side lengths 10, there are 36 points on the boundary or 36%. This number quickly increases as  $d$  increases (Cressie 1993, p. 478).

As a result, the edge effect also introduces bias, so one must be aware of its presence (Cressie 1993, p. 607). There are many techniques to deal with edge effects in practice, such as treating a rectangular study region as a torus (Griffith and Amrhein 1983). In practice, edge effects can arise in unexpected scenarios, such as when patients receive healthcare outside of a study area (Fortney, Rost, and Warren 2000).

### 7.2. Tapered covariance estimators

To deal with the edge effect several corrections of standard statistical estimators (tapered estimators) were proposed. For example, Guyon (1982) suggested to use the following unbiased estimator

**Estimator 13.** (*tapered unbiased estimated covariance function*)

$$\hat{C}_N(\mathbf{h}) := \frac{1}{C_{N-|\mathbf{h}|}} \sum_{\mathbf{t}, \mathbf{t}+\mathbf{h} \in P_n} X(\mathbf{t})X(\mathbf{t}+\mathbf{h}),$$

where  $C_{N-|\mathbf{h}|}$  is the cardinality of the set  $\{\mathbf{t} : \mathbf{t}, \mathbf{t}+\mathbf{h} \in P_n\}$  and  $N = \prod_{i=1}^d n_i$ .

However, it was shown that this estimator is not positive-definite and has some other potential issues. One of the corrections, which properly takes the edge effect into account was proposed by Dahlhaus and Künsch (1987):

**Estimator 14.** (*tapered estimated covariance function*)

$$\hat{C}_N^a(\mathbf{h}) := \left( \prod_{i=1}^d H_{2,n_i}(0) \right)^{-1} \sum_{\mathbf{t}, \mathbf{t}+\mathbf{h} \in P_n} X(\mathbf{t})X(\mathbf{t}+\mathbf{h}) \left( \prod_{i=1}^d a((t_i - 1/2)/n_i; \rho) a((t_i + h_i - 1/2)/n_i; \rho) \right).$$

Here  $a(\cdot; \cdot)$  is a taper (window) function with smoothness parameter  $\rho$  and  $u \in [0, 1]$ ,

$$a(u; \rho) := \begin{cases} w(2u/\rho), & 0 \leq u < \frac{1}{2}\rho, \\ 1, & \frac{1}{2}\rho \leq u \leq \frac{1}{2}, \\ a(1-u; \rho), & \frac{1}{2} < u \leq 1, \end{cases}$$

where  $w(\cdot)$  is a continuous increasing function with  $w(0) = 0$  and  $w(1) = 1$ . Additionally,

$$H_{2,n}(0) := \sum_{s=1}^n a((s-1/2)/n; \rho)^2.$$

A possible choice for the function  $w(\cdot)$ , as suggested by Dahlhaus and Künsch (1987, p. 878), is the Tukey window, although one is not limited to this choice.

This estimator is biased, although it becomes asymptotically negligible due to the use of windowing. As  $H_{2,n}(0)$  is a constant, a similar idea can be applied as in the standard estimator

to show the tapered estimator is positive-definite. Additionally, it also can be applied in non-Gaussian scenarios (Yao and Brockwell 2006, p. 403).

Much like Hall's estimator, this incorporates windowing on the standard estimate of the covariance, under the assumption of a zero-mean process. The incorporation of a taper does not inhibit this estimator's ability to consider long-range dependent processes to the same extent as Hall's estimator and Estimator 20, which will be introduced soon.

## 8. Estimates via linear combinations of basis functions

An estimator for the isotropic autocovariance function can be constructed closely matching an empirical estimator with a linear combination of properly selected basis functions.

The first approach uses the fact that an isotropic function  $C(h)$  is positive-definite for all  $d \geq 1$ , if and only if  $\phi(h) = C(h^{1/2})$  is completely monotone (Cressie 1993, p. 86).

Let  $\{B_1^{(p)}, \dots, B_{m+p+1}^{(p)}\}$  denote  $m + p + 1$  B-splines of order  $p$ , with  $m$  equally spaced knots, with end knots  $k_0 = 0$  and  $k_m = 1$ . Also, let  $\{b_1, \dots, b_{m+p+1}\}$  denote the B-spline coefficients. Choi, Li, and Wang (2013, p. 615) provided the following representation for  $\phi(\cdot)$  using the B-splines:

$$\phi(x) = \sum_{j=1}^{m+p} \beta_j f_j^{(p-1)}(x),$$

with

$$f_j^{(l)}(x) = \int_0^1 (m+1)t^x B_{j+1}^{(l)}(t) dt,$$

where  $\beta_j = b_{j+1} - b_j \geq 0$  for all  $j = 1, \dots, m+p$ . The functions  $f_j^{(l)}$  are completely monotone and serve as basis functions.

Therefore, the covariance estimator can be given as

**Estimator 15.** (*linear combination of completely monotone functions*) (Choi, Li, and Wang 2013, p. 615)

$$\hat{C}^B(\tau) = \sum_{j=1}^{m+p} \beta_j f_j^{(p-1)}(\tau^2).$$

The coefficients  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{m+p})'$  are estimated using the weighted least squares (WLS) approach,

$$\hat{\boldsymbol{\beta}}_{\text{WLS}} = \arg \min_{\beta_j \geq 0} \sum_{i=1}^L w_i \left( \hat{C}(\tau_i) - \sum_{j=1}^{m+p} \beta_j f_j^{(p-1)}(\tau_i^2) \right)^2,$$

where  $\hat{C}(\cdot)$  is an estimator of a covariance function,  $\{\tau_1, \dots, \tau_L\}$  is a set of distance lags, and  $\{w_1, \dots, w_L\}$  is a set of weights. The standard estimator, Estimator 1, was used in Choi, Li, and Wang (2013, p. 617).

As mentioned earlier, some estimators suffer from oscillations, meaning the estimator for the isotropic autocovariance function can become unpredictable as the distance lag changes, particularly depending on how many points are separated by a distance  $\tau$ . To address this issue, either a reasonably short range of  $\tau$  is selected, or the set of weights is chosen based on the number of available points when computing the empirical autocovariance function. For example, Choi, Li, and Wang (2013, p. 617) recommended the following weights:  $w_i = |N(\tau_i)| / (1 - \hat{C}(\tau_i))^2$ . Furthermore, one is not limited to the choice of the standard estimator of the autocovariance function when selecting the weights and the fitting process.

An isotropic covariance function in  $\mathbb{R}^n$  has the following form, see Yadrenko (1983), and Ivanov and Leonenko (1989),

$$C(\tau) = \int_0^\infty \Lambda_n(k\tau) dF(k),$$

where  $F(\cdot)$  is a cumulative distribution function,

$$\Lambda_n(k\tau) = 2^{\frac{n-2}{2}} \Gamma\left(\frac{n}{2}\right) \frac{J_{\frac{n-2}{2}}(k\tau)}{(k\tau)^{\frac{n-2}{2}}}$$

and  $J_\nu(\cdot)$  denotes the Bessel function of the first kind of the order  $\nu$ .

Li (2023, p. 12) considered an approximation of  $F(\cdot)$  by a step function, with jump points at  $k_i$ , resulting in the following nonparametric estimator of the isotropic covariance function.

**Estimator 16.** (*linear combination of isotropic basis functions*)

$$\hat{C}(\tau) := \sum_{i=1}^m w_i \Lambda_n(k_i \tau),$$

where  $\Lambda_n(\cdot)$  is acting as a basis function and  $\{w_i\}$  are weights.

The locations of the jumps  $\{k_i\}$  are determined in an ad hoc manner, although some methods exist such as selecting them to be roots of the Bessel function of the first kind (Li 2023, p. 13). The above estimator is related to the semivariogram estimator in Shapiro and Botha (1991), although considered for the autocovariance case instead. The method in Shapiro and Botha (1991) has some limitations, such as the number of jumps having no upper limit, or how to nonparametrically estimate the nugget, sill, and range (Cherry 1994, p. 25).

Wang and Ghosh (2023) considered the next estimator using a different class of basis functions.

**Estimator 17.** (*linear combination of Bernstein polynomials*) For any dimension  $d \geq 1$ , a linear combination of the basis functions

$$A_{i,m}(\tau) = \prod_{j=i}^m \left(1 + \frac{\tau^2}{j}\right)^{-1}, \quad 1 \leq i \leq m, \quad i, m \in \mathbb{N},$$

provides a valid empirical isotropic covariance function

$$\hat{C}_m(\tau) := \sum_{i=1}^m w_i A_{i,m}(\tau).$$

## 9. Correction of established methods

When established or new methods for estimating covariograms lack the required properties, such as positive-definiteness, several general transformation strategies can be employed to achieve them. Two popular approaches are discussed in this section.

### 9.1. Shrunken covariance

Numerous papers estimate a covariance matrix and subsequently transform the estimate into a positive-definite matrix if it is not inherently so. As the sample covariogram can be derived from the estimated covariance matrix, and conversely, the estimated covariance matrix can be constructed from the sample covariogram, these transformation methods are also applicable to covariograms.

One of the popular transformations, shrinking, was introduced in Devlin, Gnanadesikan, and Kettenring (1975).

**Estimator 18.** (*linear shrinking*) A  $p \times p$  pseudo-correlation matrix  $\mathbf{R}$  is shrunk towards the  $p$ -dimensional identity matrix  $\mathbf{I}_p$ , as

$$\tilde{\mathbf{R}} := \lambda \mathbf{R} + (1 - \lambda) \mathbf{I}_p,$$

where  $\lambda \in [0, 1]$  is the largest value, which makes  $\tilde{\mathbf{R}}$  positive-definite.

An alternative to linear shrinking is

**Estimator 19.** (*nonlinear shrinking*) Each off-diagonal element  $r$  of  $\mathbf{R}$  is replaced by  $\tilde{r}$  through the following transformation procedure

$$\tilde{r} := \begin{cases} f^{-1}(f(r) + \Delta), & \text{if } r < -f^{-1}(\Delta), \\ 0, & \text{if } |r| \leq f^{-1}(\Delta), \\ f^{-1}(f(r) - \Delta), & \text{if } r > f^{-1}(\Delta), \end{cases}$$

where  $\Delta$  is a small positive constant (such as 0.05) and  $f : \mathbb{R} \rightarrow [-1, 1]$  is a monotone increasing continuous function. This process is repeated until a positive-definite matrix is given.

Two popular choices for  $f$  are  $\tanh(x)$  and  $\frac{2}{\pi} \arctan(x)$  (Rousseeuw and Molenberghs 1993). Shrinking does not consider the entire correlation structure of the matrix, although it is easy to apply practically, as both linear and nonlinear shrinking operate on each correlation separately.

## 9.2. Kernel correction

One can use the following modification of the approach presented in Яглом (1981, eq. (1.76)) for the case of  $d = 1$  to correct the estimate  $\hat{C}(\mathbf{h})$ .

**Estimator 20.** (*kernel-corrected estimated covariance function*)

$$\hat{C}^{(a)}(\mathbf{h}) = a(\mathbf{h})\hat{C}(\mathbf{h}),$$

where  $a(\cdot)$  is a kernel function approaching to 0, when  $\mathbf{h}$  increases.

We will now extend some of the comments in the discussion in Яглом (1981, eq. (1.76)) which are also relevant to the spatial case.

As mentioned earlier, covariance function estimators suffer from waves as the number of samples used to estimate the covariance function decreases. Estimator 20 introduces a kernel function to remove such waves, by gradually bringing down the estimator to zero. A common choice of  $a(\mathbf{h})$  is a function that vanishes after a certain distance, for example, approximately 0.1 of the observation period in the case of time series. When working under the assumption of short-range dependence, such choices are acceptable, but in the case of long-range dependence, this may not be an appropriate estimator. Applying the vanished estimator to a long-range dependent process results in a short-short-range dependent covariance function, which does not accurately reflect the process' dependencies. A possible strategy is to increase the support (range) of  $a(\mathbf{h})$  with the observation area or the number of observations.

One can choose  $a(\mathbf{h})$  as a positive-definite kernel to keep the estimator positive-definite, given  $\hat{C}(\mathbf{h})$  is a positive-definite estimator. However, it restricts the class of eligible functions one can use. Genton (2002) provides a list of some valid kernels, see Table 1, where  $\tau = \|\mathbf{h}\|$  and  $\theta > 0$  in all cases. These kernels are not only positive-definite but also isotropic, meaning they can also be applied to the estimators of isotropic covariance functions.

## 10. Other estimators

In this paper, our primary focus has been on the estimation of Pearson-type correlations, which are based on the actual values of observations. We assumed that one realisation of the stationary spatial process is available and mainly considered nonparametric estimators given their significance in addressing numerous spatial issues. Even parametric estimators are frequently derived by fitting a parametric covariance function family to a nonparametric empirical covariance function.

Table 1: List of isotropic positive-definite kernels

Name	Equation	Validity
Circular	$a(\tau; \theta) = \begin{cases} \frac{2}{\pi} \arccos(\tau/\theta) - \frac{2}{\pi} \frac{\tau}{\theta} \sqrt{1 - (\tau/\theta)^2}, & \text{for } \tau < \theta \\ 0, & \text{otherwise} \end{cases}$	$\mathbb{R}^2$
Spherical	$a(\tau; \theta) = \begin{cases} 1 - \frac{3}{2} \frac{\tau}{\theta} + \frac{1}{2} \left(\frac{\tau}{\theta}\right)^3, & \text{for } \tau < \theta \\ 0, & \text{otherwise} \end{cases}$	$\mathbb{R}^3$
Rational Quadratic	$a(\tau; \theta) = 1 - \frac{\tau^2}{\tau^2 + \theta}$	$\mathbb{R}^d$
Exponential	$a(\tau; \theta) = \exp(-\tau/\theta)$	$\mathbb{R}^d$
Gaussian	$a(\tau; \theta) = \exp(-\tau^2/\theta)$	$\mathbb{R}^d$
Wave	$a(\tau; \theta) = \frac{\theta}{\tau} \sin(\tau/\theta)$	$\mathbb{R}^3$

Several alternative or closely related nonparametric approaches exist, though beyond the scope of this discussion due to space constraints and will be only briefly mentioned.

Firstly, there are methods based not on actual values but on the signs or ranks of the data. For additional information, one can refer to [Dürre, Vogel, and Fried \(2015\)](#) and [Sang, Dang, and Sang \(2016\)](#), and the comprehensive citations included within. [Sang, Dang, and Sang \(2016\)](#) argues that Pearson-type estimates perform poorly when dealing with heavy-tailed or asymmetric distributions, as well as being impacted greatly by outliers. There were also several modifications of the approach considered in Section 6. The majority of these modifications were based on the methods utilising a robust scale measure as proposed in the highly cited publication [Gnanadesikan and Kettenring \(1972\)](#).

There also exist several approaches that used weighted, affine equivariant estimators in combination with M-estimators, see, for example, the publications [Dürre, Vogel, and Fried \(2015\)](#), [Tyler \(1987\)](#), and [Lumley and Heagerty \(1999\)](#) and the discussions in them. It is worth mentioning that, similar to the discussion in Appendix E, most of these models satisfy certain summability constraints that may impact their performance.

An overview of nonparametric Bayesian approaches can be found in [Kidd and Katzfuss \(2022\)](#), which has an emphasis on methods which have nice properties computationally, namely parallelisation and scalability. A nonparametric Bayesian approach was also considered in [Porcu, Bissiri, Tagle, Soza, and Quintana \(2021a\)](#), where they studied the estimation of isotropic covariance functions for random fields on the sphere.

Finally, there are other options that use variogram- or spectrum-based ideas, for example, [Huang, Hsing, and Cressie \(2011\)](#) used spline polynomials to estimate the isotropic spectral density, and then it was employed to compute the variogram. Such methods that produce negative-definite variogram estimates can be easily adapted to the covariance case.

## 11. Simulation studies

This section will demonstrate some of the considered properties and empirically compare the estimators via realisations of Gaussian random fields with different covariance functions. For the sake of simplicity and visualisations, we only present results for two-dimensional cases, but the corresponding one-dimensional results were also obtained.

For illustration, we will consider the three isotropic covariance functions, namely

- *Gaussian*:  $C(\tau; \sigma) = \exp(-\tau^2/\sigma^2)$ , where  $\sigma > 0$ ,
- *Bessel*:  $C(\tau; \nu) = 2^\nu \tau^{-\nu} \Gamma(\nu + 1) J_\nu(\tau)$ , where  $\nu \geq (d - 2)/2$ ,
- *Cauchy*:  $C(\tau; \gamma) = (1 + \tau^2)^{-\gamma}$ , where  $\gamma > 0$ .

They are positive-definite in all dimensions  $d$ .



For each covariance function, a single spatial realisation will be used, as it can be deemed more realistic in many applications. To obtain reliably estimated values, these realisations have been generated either in a large observation window or a dense sampling grid. The analysis can also be extended to the methods which use multiple realisations (often modelled as independent realisations of random fields with the same statistical properties), for example, see [Ivanov and Leonenko \(1989, Section 4.5\)](#).

For the convenience of the readers, we reiterate the previous notations for the estimators to be discussed in this section:

- $C^*$ , the standard estimator, Estimator 1 with the normalising factor varying with lags,
- $C^{**}$ , the standard estimator, Estimator 1 with the constant normalising factor and the total number of sampled locations  $N$ ,
- $C^{(a)}$ , corrected version of  $C^{**}$ , using the correction method of Estimator 20,
- $\tilde{C}$ , Hall's estimator, Estimator 9 ,
- $\hat{C}_Q$ , the quantile-based estimator, Estimator 10,
- $\hat{C}_N^a$ , the tapered estimator, Estimator 14,
- $\hat{C}^B$ , the splines estimator, Estimator 15. In this case  $C^*$  was used as the estimated covariance function during the fitting process with  $m = 2$  and  $p = 3$ .

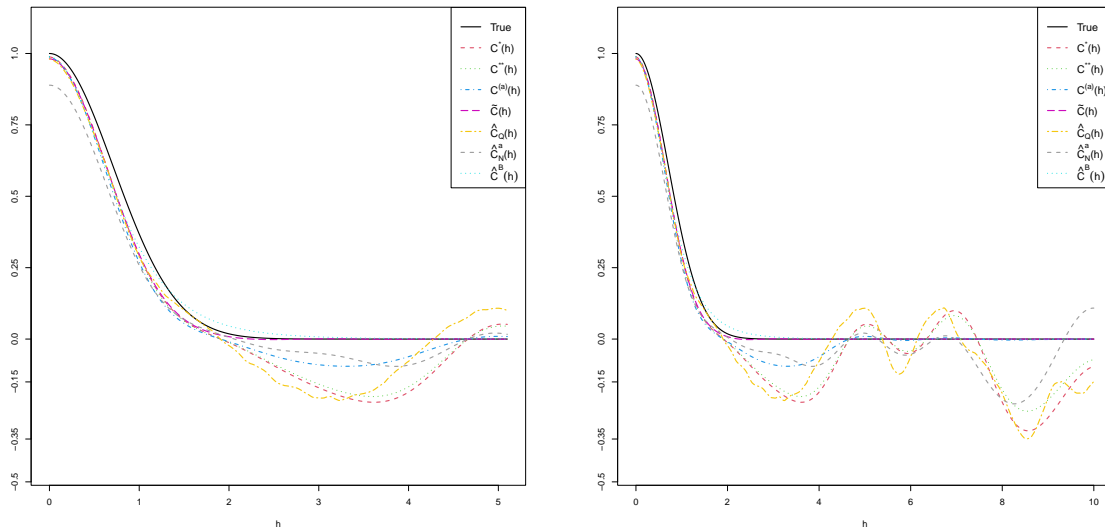
As the spatial case is the focus of this paper, we will briefly demonstrate only one example from the one-dimensional case that highlights the worsening of waves with increasing estimation lag. Let  $d = 1$  and consider a Gaussian process with the Gaussian covariance and with  $\sigma = 1$ . Realisations of the process were simulated on the grid  $\{0, 0.02, 0.04, \dots, 40\}$  and used to estimate its covariance function. It is clear from Figure 1a that all methods estimate the variance of the process well, all around 1, with the lowest being  $\hat{C}_N^a$  (the grey line) with a value of 0.89, and the highest being  $C^*$ ,  $C^{**}$ , and  $C^{(a)}$  at 0.99. The truncation points for Hall's estimator were chosen as  $T_1 = 1.5$ ,  $T_2 = 2$ , and the kernel was the Gaussian kernel, which  $C^{(a)}$  also used. For larger distances, as Hall's estimator brings down the estimate linearly, it removes waves, as one may expect, although other estimators still possess these waves. The waves of estimator  $C^{(a)}$  have small amplitude due to the introduction of a multiplicative kernel transformation. When the estimation range increases, the amplitudes of the waves become more significant as seen in Figure 1b. From this, one can conclude that most estimators suffer from the issue with waves, except Hall's estimator,  $C^{(a)}$ , and  $\hat{C}^B$ . It is also worth noting the waves occur in similar locations.

In practice, estimating over long distances is not advisable. For example, for the one-dimensional case the range of 10-20% of the total observation range  $N$  is recommended ([Yaglom 1987, p. 237](#)). In Figure 1b, 25% of  $N$  was used, whereas Figure 1a considers only 12.8%. Another reason why one should avoid high lag values when estimating the covariance function when using estimator  $C^*$ , the averaging interval,  $N - h$ , becomes smaller ([Yaglom 1987, p. 237](#)). One may also want to avoid high lag values as the sum of the estimated values of  $C^{**}$  is always constant, see the discussion in (iii) from Section 3.1.

Now, we compare the performance of the considered estimators of the three mentioned isotropic covariance functions using simulated realisations in  $\mathbb{R}^2$ . For a fixed  $\tau_0 \in \mathbb{R}$ , which represents the maximum estimation lag, we use the following five metrics:

- the *area between* the theoretical covariance function and an estimated covariance function, i.e.

$$A(C(\tau), \hat{C}(\tau)) = \int_0^{\tau_0} |C(\tau) - \hat{C}(\tau)| d\tau,$$



(a) Estimated autocovariance functions over a short range

(b) Estimated autocovariance functions over a longer range with prominent waves

Figure 1: Waves worsening in autocovariance estimators

- the *maximum vertical distance* between the theoretical covariance function and the estimated covariance function,

$$D(C(\tau), \hat{C}(\tau)) = \max_{\tau \in [0, \tau_0]} |C(\tau) - \hat{C}(\tau)|,$$

- the *mean-square prediction error (MSPE)* is considered by taking the differences between the actual values and those computed by kriging. There are two versions, one using the theoretical and estimated covariance function, and one just the estimated covariance function, called *MSPE gstat* and *MSPE*, respectively.
- the *MSPE gstat*, requires using both the theoretical and estimated covariance functions, whose steps are outlined below.

- (1) Generate a Gaussian random field  $X$  on  $[-15, 15]^2$  using a grid covering this square with a 0.1 step in both the  $x$  and  $y$  directions.
- (2) Estimate the covariance functions using the generated values from the subregion  $[-10, 10]^2$ .
- (3) Estimate the parameter of the theoretical covariance model by using nonlinear least squares. For Gaussian covariance, this is  $\sigma$ , for Bessel  $\nu$ , and for Cauchy  $\gamma$ .
- (4) Sample 50 locations  $\mathbf{t}_{s1}, \dots, \mathbf{t}_{s50}$  outside the subregion  $[-10, 10]^2$ .
- (5) Perform kriging on the set of sample locations using the observations from the subregion  $[-10, 10]^2$  and obtain  $\hat{X}_S$ .
- (6) Compute

$$\text{MSPE}_g = \frac{1}{50} \sum_{i=1}^{50} \left( X(\mathbf{t}_{si}) - \hat{X}_S(\mathbf{t}_{si}) \right)^2.$$

To implement the *MSPE gstat*, the R package *gstat* (Pebesma 2004; Gräler, Pebesma, and Heuvelink 2016) was used. As *gstat* does not support the Bessel and Cauchy covariance functions, the corresponding package functions were modified.

- the *MSPE*, which is the second version of the *MSPE* metric, uses only the estimated covariance function  $\hat{C}$ . For this metric, one uses steps (1), (2) and (4) as above and then proceeds as follows.

- (5) For the  $M$  different points  $\mathbf{t}_j$ , construct a  $(M + 1) \times (M + 1)$  matrix with the estimated covariances

$$\mathbf{\Gamma} = \begin{bmatrix} \hat{C}(\|\mathbf{t}_1 - \mathbf{t}_1\|) & \hat{C}(\|\mathbf{t}_1 - \mathbf{t}_2\|) & \dots & \hat{C}(\|\mathbf{t}_1 - \mathbf{t}_M\|) & 1 \\ \hat{C}(\|\mathbf{t}_2 - \mathbf{t}_1\|) & \hat{C}(\|\mathbf{t}_2 - \mathbf{t}_2\|) & \dots & \hat{C}(\|\mathbf{t}_2 - \mathbf{t}_M\|) & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \hat{C}(\|\mathbf{t}_M - \mathbf{t}_1\|) & \hat{C}(\|\mathbf{t}_M - \mathbf{t}_2\|) & \dots & \hat{C}(\|\mathbf{t}_M - \mathbf{t}_M\|) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}.$$

To speed computations, we considered the  $M = 512$  closest locations to each  $\mathbf{t}_{si}$ . There may not always be 512 different closest locations available, so the largest possible  $M$  was used in such cases.

It is worth noting that not all distances between some data locations equate to a sample distance lag, so interpolation can be done to estimate the covariance function at a specific distance. For example, in the simulation studies, the mean covariance value of points within a small neighbourhood can be taken for lags which do not have an empirical covariance value.

- (6) For each randomly sampled location,  $\mathbf{t}_{si}$ , compute the distances between it and the  $M$  closest sample locations  $\mathbf{t}_j$  from the above step.
- (7) Construct the column vector  $\mathbf{c} = (\hat{C}(\|\mathbf{t}_{si} - \mathbf{t}_1\|), \hat{C}(\|\mathbf{t}_{si} - \mathbf{t}_2\|), \dots, \hat{C}(\|\mathbf{t}_{si} - \mathbf{t}_M\|), 1)'$ .
- (8) Compute  $\hat{X}(\mathbf{t}_{si}) = \left( (\mathbf{\Gamma}^+ \mathbf{c})_{-(M+1)} \right)' (X(\mathbf{t}_1), \dots, X(\mathbf{t}_M))'$ , where  $\mathbf{\Gamma}^+$  is the Moore-Penrose inverse, and the notation  $\mathbf{a}_{-j}$  is the vector  $\mathbf{a}$  with the  $j^{\text{th}}$  element removed. The Moore-Penrose inverse was chosen as in some cases, due to a computational issue, the matrix  $\mathbf{\Gamma}$  may be singular.
- (9) Compute the MSPE metric

$$\text{MSPE} = \frac{1}{50} \sum_{i=1}^{50} \left( X(\mathbf{t}_{si}) - \hat{X}(\mathbf{t}_{si}) \right)^2.$$

- the spectral norm of a symmetric matrix constructed from the difference between the true covariance function and the empirical covariance function,  $D(\tau) = C(\tau) - \hat{C}(\tau)$  for all lags  $\{0, \tau_1, \dots, \tau_N\}$  used in the estimation process. The matrix  $\mathbf{D}$  is given by

$$\mathbf{D} = \begin{bmatrix} D(0) & D(\tau_1) & \dots & D(\tau_{N-1}) & D(\tau_N) \\ D(\tau_1) & D(0) & \dots & D(\tau_{N-2}) & D(\tau_{N-1}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ D(\tau_{N-1}) & D(\tau_{N-2}) & \dots & D(0) & D(\tau_1) \\ D(\tau_N) & D(\tau_{N-1}) & \dots & D(\tau_1) & D(0) \end{bmatrix}.$$

For the two-dimensional case, the spectral norm of the matrix  $\mathbf{D}$  equals its largest singular value. In the following text, this metric's name will be shortened to  $SN$ .

We computed metrics for 20 realisations of 2D random fields, the average and the average rank of the metrics. The rank was computed for each metric per realisation, resulting in the ranks from 1 to 7 per realisation, where a lower rank indicates better performance.

First, we considered the isotropic Gaussian covariance case, with  $\sigma = 1$ . The truncation points for Hall's estimator were  $T_1 = 1.5$ ,  $T_2 = 2$ , and the kernel was the Gaussian kernel, which was also chosen for  $C^{(a)}$ . The estimated Gaussian covariance on the interval  $[0, 19.9]$  (i.e.  $\tau_0 = 19.9$ ) was used for all metrics. The same interval was used for the other covariance model cases too.

Hall's estimator performed the best across area, distance, and SN, whilst the second-best estimator varied across these. For area and SN,  $C^{(a)}$  was the second best estimator, and for

distance, it was  $C^{**}$ , which is reflected partially in the average rank Table 2b. The large areas of  $\hat{C}_Q$  and  $C^*$  can be seen, especially at the end of the estimation range (see Figure 2a). These estimators suffered from waves, although Hall's,  $C^{(a)}$ , and  $\hat{C}^B$  reduced the significance of the waves due to their properties. In this case,  $C^{(a)}$  had the lowest MSPE. The MSPE gstat values were quite close, except for  $\hat{C}^B$ . The ranks in Table 2b show that Hall's estimator had the lowest average rank for area, distance, and SN, whilst  $\hat{C}_Q$  had the highest average rank for those three. For MSPE,  $C^{**}$  performs poorly, as do Hall's estimator and  $\hat{C}_N^a$ .

Table 2: Average performance of the estimators for the isotropic Gaussian case

(a) Sampled average errors for various estimators of isotropic Gaussian covariogram

Metric \ Method	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	0.9984164	0.1868498	5.125285	3.810552	0.4981920
$C^{**}$	0.4078536	0.0714691	3.186527	27.400889	0.4765476
$C^{(a)}$	0.1410914	0.1183420	2.671193	2.045802	0.4719200
$\tilde{C}$	0.0689988	0.0481300	1.199598	10.457498	0.4640216
$\hat{C}_Q$	1.2938920	0.2064251	11.070654	9.062758	0.4952457
$\hat{C}_N^a$	0.2839754	0.0793276	3.172531	9.695051	0.4607385
$\hat{C}^B$	0.1748706	0.0864897	2.928384	2.587638	0.6607006

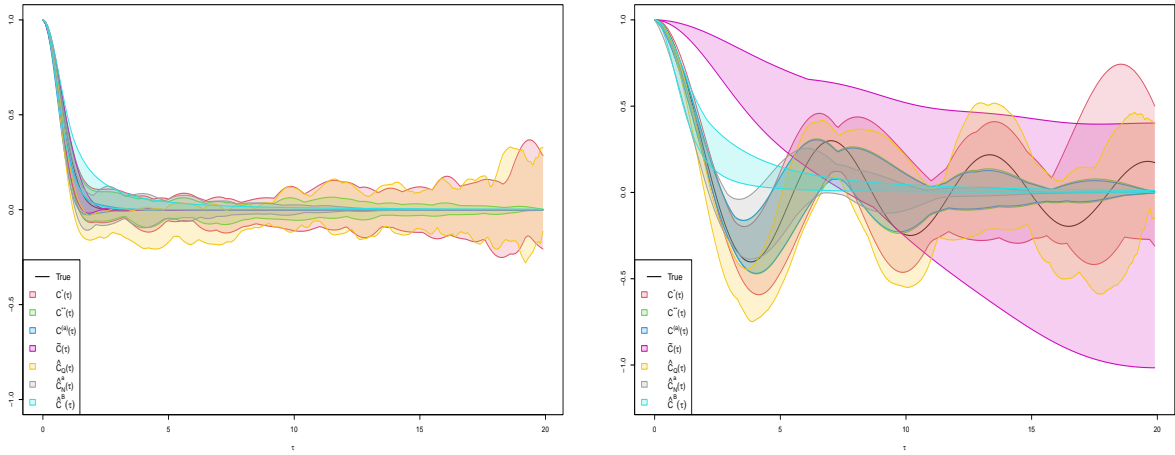
(b) Average performance rank for various estimators of isotropic Gaussian covariogram

Metric's Rank \ Method	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	6.05	6.25	5.90	4.35	3.00
$C^{**}$	4.95	2.80	3.70	4.10	3.40
$C^{(a)}$	2.50	4.50	3.25	2.85	4.85
$\tilde{C}$	1.15	1.35	1.30	3.30	3.35
$\hat{C}_Q$	6.95	6.70	6.85	4.25	6.35
$\hat{C}_N^a$	3.90	3.05	4.00	4.80	2.80
$\hat{C}^B$	2.50	3.35	3.00	4.35	4.25

For simulations from the model with the Bessel covariance function, the parameter  $\nu = 0$  was selected, which is valid for the considered case  $d = 2$ . The kernel for Hall's estimator was Gaussian, whilst the wave kernel was chosen for  $C^{(a)}$ . The truncation points  $T_1$  and  $T_2$  for Hall's estimator were beyond the used estimation range, making the truncation of the estimator unnecessary.

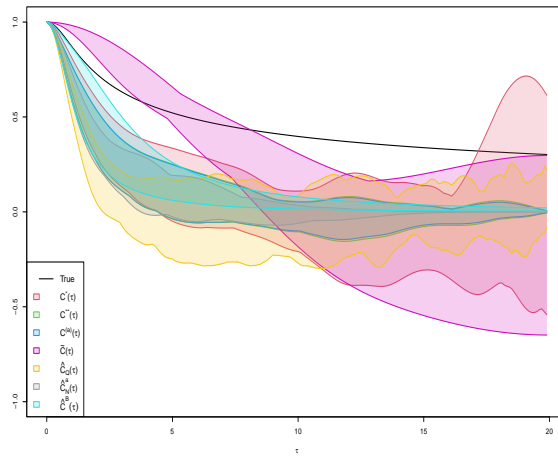
Hall's estimator performed the most poorly in this case. It is clear from Figure 2b that Hall's estimator may suffer if there is an early sample spectral frequency with a corresponding negative value.  $C^{**}$  had the lowest error for area and SN,  $C^{(a)}$  for distance,  $\hat{C}^B$  for MSPE, and for MSPE gstat, all were similar with the exception of  $\hat{C}^B$  (see Table 3a).  $\hat{C}^B$  having the lowest MSPE was quite surprising, given it did not follow the shape of the true covariance function, and had the second largest area. Despite  $C^{(a)}$  and  $C^{**}$  having the lowest errors, it is visually clear that neither followed the true function as nice as  $C^*$  or  $\hat{C}_Q$  did.  $C^{**}$  and  $\hat{C}^B$  were going to zero, whilst  $C^{(a)}$  decayed to zero due to the use of a kernel function, and  $C^*$  and  $\hat{C}_Q$  roughly followed the true function, despite having a wide prediction range. If ranks are considered (see Table 3b),  $C^{**}$  was still the best for area and SN, whilst  $C^{(a)}$  for distance,  $\hat{C}_N^a$  for MSPE, as discussed above. This is surprising given they did not follow the shape of the function nicely.

Finally, for the Cauchy covariance function case,  $\gamma = 0.2$  was selected. For Hall's estimator and  $C^{(a)}$ , the rational quadratic kernel was chosen, and the truncation points were once again beyond the estimation range, making truncation unnecessary.



(a) Isotropic Gaussian covariance case

(b) Isotropic Bessel covariance case



(c) Isotropic Cauchy covariance case

Figure 2: Upper and lower bounds for each estimator for 20 realisations

In this case, for lower values of  $\tau$ , the estimators performed well, but as the estimation range increased, the estimators performed worse, which can be seen in Figure 2c. In terms of error, Hall's estimator performed the best for area and SN, whilst  $\hat{C}^B$  for distance,  $C^*$  for MSPE, and for MSPE gstat,  $\hat{C}_Q$  was the best, and all others were similar. If we considered the average rank instead (see Table 4b), it is clear that  $\hat{C}^B$  had the lowest average rank for area and distance, and Hall's for SN, but it had the worst MSPE rank, and  $\hat{C}_Q$  had the best MSPE gstat rank. Unlike Hall's estimator,  $\hat{C}^B$  did not flare out after  $\tau = 10$ , and instead, the minimum and maximum values of the band approached one another. As a result,  $\hat{C}^B$  can be seen as a more stable estimator when dealing with noncyclic nonnegative covariance functions.

## 12. Discussion and conclusion

Several nonparametric methods for estimating spatial covariance functions were reviewed, studied, and compared based on unified principles. The main focus was on the case of a single realisation and the methods utilising the actual values of observations. Some surprising drawbacks of several well-known estimators were identified. Numerical studies compared the accuracy of the considered estimators using various metrics and three theoretical covariogram models. When dealing with random fields whose covariance functions are not oscillating and decay sufficiently fast, such as the Gaussian covariance, estimators like Hall's estimator and  $C^{(a)}$  outperform other estimators. However, in the case of cyclic or long-range dependence, the vanishing estimators often fail to capture the behaviour of the theoretical covariogram model.

Table 3: Average performance of the estimators for the isotropic Bessel case

(a) Sampled average errors for various estimators of isotropic Bessel covariogram

Method \ Metric	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	2.489791	0.3507251	15.61262	125.475762	0.0056134
$C^{**}$	2.098724	0.2503224	12.70419	6.760378	0.0084915
$C^{(a)}$	2.109226	0.2496318	12.78919	106.180658	0.0086290
$\tilde{C}$	7.513908	1.0682663	82.24297	75.854967	0.0764184
$\hat{C}_Q$	2.748823	0.3275784	18.99515	443.374390	0.0049442
$\hat{C}_N^a$	2.492002	0.2714209	16.79564	2.626784	0.0140113
$\hat{C}^B$	3.486087	0.5350350	29.29765	1.950848	0.1618823

(b) Average performance rank for various estimators of isotropic Bessel covariogram

Method \ Metric's Rank	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	3.20	4.00	2.80	5.20	1.875
$C^{**}$	1.90	2.20	1.75	3.95	3.200
$C^{(a)}$	2.40	2.10	2.25	4.00	3.800
$\tilde{C}$	7.00	7.00	7.00	5.35	6.100
$\hat{C}_Q$	4.15	3.85	4.35	5.30	1.225
$\hat{C}_N^a$	3.75	3.00	4.00	1.95	4.900
$\hat{C}^B$	5.60	5.85	5.85	2.25	6.900

Table 4: Average performance of the estimators for the isotropic Cauchy case

(a) Sampled average errors for various estimators of isotropic Cauchy covariogram

Method \ Metric	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	7.486252	0.6165021	70.18825	2.582530	7.051711
$C^{**}$	7.191006	0.4875699	72.18749	3.678911	6.974136
$C^{(a)}$	7.187045	0.4864138	72.23088	5.487214	6.970320
$\tilde{C}$	5.671946	0.5083464	36.96571	786.580629	7.069838
$\hat{C}_Q$	8.393729	0.6678209	91.26490	10.102567	5.850791
$\hat{C}_N^a$	7.212647	0.4913294	73.58152	17.153026	7.003175
$\hat{C}^B$	6.280143	0.3943084	61.03689	5.872001	6.990817

(b) Sampled average errors for various estimators of isotropic Cauchy covariogram

Method \ Metric's Rank	Area	Distance	SN	MSPE	MSPE gstat
$C^*$	4.80	5.65	3.70	3.50	4.3
$C^{**}$	4.35	3.60	4.40	3.20	4.2
$C^{(a)}$	3.95	3.65	4.75	3.25	4.0
$\tilde{C}$	2.25	3.55	1.10	6.95	4.2
$\hat{C}_Q$	6.40	6.30	6.80	3.70	2.6
$\hat{C}_N^a$	4.35	3.80	5.10	3.25	4.4
$\hat{C}^B$	1.90	1.45	2.15	4.15	4.3

In such instances, other estimators, such as  $C^*$  in the Bessel case, may be better suited, even if they have some constant summability properties. The results are useful for understanding



the limitations and proper usage of covariograms in various applications, including kriging, monitoring network optimisation, and cross-validation.

Future work includes an overview and comparative studies of other estimation methods for covariograms, as well as similar studies for estimation methods of spectral densities. Other potential extensions could involve analogous investigations of estimates of dependency and the spectrum of multivariate and spherical spatial data, as well as spatio-temporal data.

## Acknowledgments

This research was partially supported under the Australian Research Council's Discovery Projects funding scheme (project number DP220101680). Andriy Olenko was also partially supported by La Trobe University's SCEMS CaRE and Beyond grant. We would like to thank Professors N.Cressie, N.Leonenko, and E.Porcu for their constructive feedback on the early draft of this paper.

## References

- Cave-Browne-Cave FE, Pearson K (1905). "On the Influence of the Time Factor on the Correlation between the Barometric Heights at Stations More than 1000 Miles Apart." *Proceedings of the Royal Society of London*, **74**, 403–413. doi:10.1098/rspl.1904.0136.
- Cherry JS (1994). *Nonparametric Estimation of Semivariogram Functions*. Ph.D. thesis, Montana State University.
- Chilès J, Delfiner P (2012). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons. doi:10.1002/9781118136188.
- Choi I, Li B, Wang X (2013). "Nonparametric Estimation of Spatial and Space-Time Covariance Function." *Journal of Agricultural, Biological, and Environmental Statistics*, **18**(4), 611–630. doi:10.1007/s13253-013-0152-z.
- Cressie N (1988). "A Graphical Procedure for Determining Nonstationarity in Time Series." *Journal of the American Statistical Association*, **83**(404), 1108–1116. doi:10.1080/01621459.1988.10478708.
- Cressie N (1993). *Statistics for Spatial Data*. John Wiley & Sons. doi:10.1002/9781119115151.
- Cressie N, Hawkins DM (1980). "Robust Estimation of the Variogram: I." *Journal of the International Association for Mathematical Geology*, **12**(2), 115–125. doi:10.1007/BF01035243.
- Cuevas F, Porcu E, Vallejos R (2013). "Study of Spatial Relationships between Two Sets of Variables: A Nonparametric Approach." *Journal of Nonparametric Statistics*, **25**(3), 695–714.
- Dahlhaus R, Künsch H (1987). "Edge Effects and Efficient Parameter Estimation for Stationary Random Fields." *Biometrika*, **74**(4), 877–882. doi:10.1093/biomet/74.4.877.
- Devlin SJ, Gnanadesikan R, Kettenring JR (1975). "Robust Estimation and Outlier Detection with Correlation Coefficients." *Biometrika*, **62**(3), 531–545. doi:10.1093/biomet/62.3.531.
- Dürre A, Vogel D, Fried R (2015). "Spatial Sign Correlation." *Journal of Multivariate Analysis*, **135**, 89–105. doi:10.1016/j.jmva.2014.12.002.
- Дыховичный АА (1983). "Об Оценке Корреляционной Функции Однородного и Изотропного Гауссовского Случайного Поля." *Теория Вероятностей и Математическая Статистика*, **29**, 37–40.

- Яглом АМ (1981). *Корреляционная Теория Стационарных Случайных Функций: С Примерами из Метеорологии*. Гидрометеиздат.
- Fortney J, Rost K, Warren J (2000). “Comparing Alternative Methods of Measuring Geographic Access to Health Services.” *Health Services and Outcomes Research Methodology*, **1**, 173–184. ISSN 1572-9400. doi:10.1023/A:1012545106828.
- García-Soidán P, Menezes R (2012). “Estimation of the Spatial Distribution through the Kernel Indicator Variogram.” *Environmetrics*, **23**(6), 535–548. doi:https://doi.org/10.1002/env.2151.
- García-Soidán PH, Febrero-Bande M, González-Manteiga W (2004). “Nonparametric Kernel Estimation of an Isotropic Variogram.” *Journal of Statistical Planning and Inference*, **121**(1), 65–92. ISSN 0378-3758. doi:https://doi.org/10.1016/S0378-3758(02)00507-4.
- Genton MG (1998). “Highly Robust Variogram Estimation.” *Mathematical Geology*, **30**(2), 213–221. doi:10.1023/A:1021728614555.
- Genton MG (2002). “Classes of Kernels for Machine Learning: A Statistics Perspective.” *The Journal of Machine Learning Research*, **2**, 299–312.
- Gnanadesikan R, Kettenring JR (1972). “Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data.” *Biometrics*, **28**(1), 81. doi:10.2307/2528963.
- Griffith DA, Amrhein CG (1983). “An Evaluation of Correction Techniques for Boundary Effects in Spatial Statistical Analysis: Traditional Methods.” *Geographical Analysis*, **15**, 352–360. doi:10.1111/j.1538-4632.1983.tb00794.x.
- Gräler B, Pebesma E, Heuvelink G (2016). “Spatio-Temporal Interpolation Using gstat.” *The R Journal*, **8**, 204–218. doi:10.32614/RJ-2016-014. URL https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.
- Guyon X (1982). “Parameter Estimation for a Stationary Process on a  $d$ -Dimensional Lattice.” *Biometrika*, **69**(1), 95–105. doi:10.1093/biomet/69.1.95.
- Hall P, Fisher NI, Hoffmann B (1994). “On the Nonparametric Estimation of Covariance Functions.” *The Annals of Statistics*, **22**(4), 2115 – 2134. doi:10.1214/aos/1176325774.
- Hall P, Patil P (1994). “Properties of Nonparametric Estimators of Autocovariance for Stationary Random Fields.” *Probability Theory and Related Fields*, **99**(3), 399–424. doi:10.1007/bf01199899.
- Hassani H (2009). “Sum of the Sample Autocorrelation Function.” *Random Operators and Stochastic Equations*, **17**(2), 125–130. doi:10.1016/j.physa.2009.12.050.
- Hassani H, Leonenko N, Patterson K (2012). “The Sample Autocorrelation Function and the Detection of Long-Memory Processes.” *Physica A: Statistical Mechanics and Its Applications*, **391**(24), 6367–6379. doi:10.1016/j.physa.2012.07.062.
- Hristopulos DT (2020). *Random Fields for Spatial Data Modeling: A Primer for Scientists and Engineers*. Springer Nature B.V. doi:10.1007/978-94-024-1918-4.
- Huang C, Hsing T, Cressie N (2011). “Nonparametric Estimation of the Variogram and Its Spectrum.” *Biometrika*, **98**(4), 775–789. doi:10.1093/biomet/asr056.
- Huber PJ (1981). *Robust Statistics*. John Wiley & Sons. doi:10.1002/0471725250.
- Ivanov AV, Leonenko NN (1989). *Statistical Analysis of Random Fields*. Springer Dordrecht. doi:10.1007/978-94-009-1183-3.

- Jenkins GM, Watts DG (1969). *Spectral Analysis and Its Applications*. Holden-Day.
- Journel AG, Huijbregts CJ (1978). *Mining Geostatistics*. Academic Press.
- Kidd B, Katzfuss M (2022). “Bayesian Nonstationary and Nonparametric Covariance Estimation for Large Spatial Data (with Discussion).” *Bayesian Analysis*, **17**(1). doi:10.1214/21-BA1273.
- Krajewski WF, Duffy CJ (1988). “Estimation of Correlation Structure for a Homogeneous Isotropic Random Field: A Simulation Study.” *Computers & Geosciences*, **14**(1), 113–122. doi:10.1016/0098-3004(88)90055-6.
- Lahiri SN (1996). “On Inconsistency of Estimators Based on Spatial Data under Infill Asymptotics.” *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, **58**(3), 403–417.
- Lahiri SN, Lee Y, Cressie N (2002). “On Asymptotic Distribution and Asymptotic Efficiency of Least Squares Estimators of Spatial Variogram Parameters.” *Journal of Statistical Planning and Inference*, **103**, 65–85. doi:10.1016/s0378-3758(01)00198-7.
- Lee YD, Lahiri SN (2002). “Least Squares Variogram Fitting by Spatial Subsampling.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **64**(4), 837–854. doi:10.1111/1467-9868.00364.
- Li Y (2023). *Non-Parametric and Semi-Parametric Estimation of Spatial Covariance Function*. Ph.D. thesis, Iowa State University. doi:10.31274/etd-180810-3216. URL <https://dr.lib.iastate.edu/entities/publication/9e132aec-1625-4a7c-9365-dc20fb7782dc>.
- Lumley T, Heagerty P (1999). “Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **61**(2), 459–477. doi:10.1111/1467-9868.00187.
- Ma C (2007). “Why is Isotropy so Prevalent in Spatial Statistics?” *Proceedings of the American Mathematical Society*, **135**(3), 865–871. doi:10.1090/S0002-9939-06-08592-3.
- Ma Y, Genton MG (2000). “Highly Robust Estimation of the Autocovariance Function.” *Journal of Time Series Analysis*, **21**(6), 663–684. doi:10.1111/1467-9892.00203.
- Ma YZ (2019). *Quantitative Geosciences: Data Analytics, Geostatistics, Reservoir Characterization and Modeling*. Springer Nature Switzerland. doi:10.1007/978-3-030-17860-4.
- Matheron G (1962). *Traité de Géostatistique Appliquée, Tome I*. 14. Editions Technip.
- Pebesma EJ (2004). “Multivariable Geostatistics in S: The gstat Package.” *Computers & Geosciences*, **30**, 683–691. doi:10.1016/j.cageo.2004.03.012.
- Porcu E, Bissiri PG, Tagle F, Soza R, Quintana FA (2021a). “Nonparametric Bayesian Modeling and Estimation of Spatial Correlation Functions for Global Data.” *Bayesian Analysis*, **16**(3), 845–873.
- Porcu E, Furrer R, Nychka D (2021b). “30 Years of Space–Time Covariance Functions.” *WIREs Computational Statistics*, **13**(2), e1512. doi:https://doi.org/10.1002/wics.1512.
- Rousseeuw PJ, Croux C (1992). “Explicit Scale Estimators with High Breakdown Point.” *L1-Statistical Analysis and Related Methods*, **1**, 77–92.
- Rousseeuw PJ, Croux C (1993). “Alternatives to the Median Absolute Deviation.” *Journal of the American Statistical Association*, **88**(424), 1273–1283. doi:10.1080/01621459.1993.10476408.

- Rousseeuw PJ, Molenberghs G (1993). “Transformation of Non Positive Semidefinite Correlation Matrices.” *Communications in Statistics - Theory and Methods*, **22**(4), 965–984. doi:10.1080/03610928308831068.
- Sang Y, Dang X, Sang H (2016). “Symmetric Gini Covariance and Correlation.” *Canadian Journal of Statistics*, **44**(3), 323–342. doi:10.1002/cjs.11287.
- Schabenberger O, Gotway CA (2005). *Statistical Methods for Spatial Data Analysis*. Chapman and Hall/CRC.
- Shapiro A, Botha JD (1991). “Variogram Fitting with a General Class of Conditionally Non-negative Definite Functions.” *Computational Statistics & Data Analysis*, **11**(1), 87–96. doi:10.1016/0167-9473(91)90055-7.
- Tarr G, Müller S, Weber N (2012). “A Robust Scale Estimator Based on Pairwise Means.” *Journal of Nonparametric Statistics*, **24**(1), 187–199. doi:10.1080/10485252.2011.621424.
- Tarr G, Müller S, Weber N (2013). “Robust Covariance Estimation with  $P_n$ .” URL <https://www.maths.usyd.edu.au/u/gartht/GarthTarrASC.pdf>.
- Tyler DE (1987). “A Distribution-Free M-Estimator of Multivariate Scatter.” *The Annals of Statistics*, **15**(1), 234–251. doi:10.1214/aos/1176350263.
- Wang Y, Ghosh SK (2023). “Nonparametric Estimation of Isotropic Covariance Function.” *Journal of Nonparametric Statistics*, **35**(1), 198–237. doi:10.1080/10485252.2022.2146111.
- Yadrenko MI (1983). *Spectral Theory of Random Fields*. Optimization Software, New York.
- Yaglom AM (1987). *Correlation Theory of Stationary and Related Random Functions*. Volume I: Basic Results. Springer New York. doi:10.1007/978-1-4612-4628-2.
- Yao Q, Brockwell PJ (2006). “Gaussian Maximum Likelihood Estimation for ARMA Models II: Spatial Processes.” *Bernoulli*, **12**(3), 403–429. doi:10.3150/bj/1151525128.
- Yu K, Mateu J, Porcu E (2007). “A Kernel-Based Method for Nonparametric Estimation of Variograms.” *Statistica Neerlandica*, **61**(2), 173–197.

## A. R and Python packages with estimators of covariance functions

R packages: stats, RandomFields, spatcov, fields, TSA, astda, forecast, Stat2Data

Python packages: Numpy, statsmodels, pandas

Many of these packages simply use wrapper functions for the acf function. It is important to note that this list is not exhaustive.

## B. Non-positive-definiteness of classical estimator

Consider the points  $\mathbf{s}_1 = (1, 0)$ ,  $\mathbf{s}_2 = (2, 0)$  and  $\mathbf{s}_3 = (3, 0)$ , with values  $X(\mathbf{s}_1) = X(\mathbf{s}_3) = 1$ ,  $X(\mathbf{s}_2) = 0$ , and coefficients  $a_1 = i$ ,  $a_2 = 1$ , and  $a_3 = -i$ . We will have three pairs of points for the distance  $\tau = 0$ ,  $\{(\mathbf{s}_1, \mathbf{s}_1), (\mathbf{s}_2, \mathbf{s}_2), (\mathbf{s}_3, \mathbf{s}_3)\}$ , at  $\tau = 1$  there are four pairs,  $\{(\mathbf{s}_1, \mathbf{s}_2), (\mathbf{s}_2, \mathbf{s}_1), (\mathbf{s}_2, \mathbf{s}_3), (\mathbf{s}_3, \mathbf{s}_2)\}$ , and for  $\tau = 2$  there are two pairs of points,  $\{(\mathbf{s}_1, \mathbf{s}_3), (\mathbf{s}_3, \mathbf{s}_1)\}$ . Expanding the sum, we obtain a contradiction with the positive-definite condition:

$$\sum_{i=1}^3 \sum_{j=1}^3 a_j X(s_i) \bar{X}(s_j) \bar{a}_j = \frac{1}{3} \cdot 2 + \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot -2 = -\frac{1}{3} < 0.$$

**Data:**  $N$  sample values of a random field  $X$ , desired distance  $\tau$

sum  $\leftarrow 0$ ;

$N(\tau) \leftarrow \emptyset$ ;

**for** Each coordinate  $\mathbf{t}_i$  in  $X$  **do**

**for** Each coordinate  $\mathbf{t}_j$  in  $X$  **do**

**if**  $\|\mathbf{t}_i - \mathbf{t}_j\| = \tau$  **then**

            Append  $(\mathbf{t}_i, \mathbf{t}_j)$  to  $N(\tau)$ ;

**end**

**end**

**end**

**for** Each pair  $(\mathbf{t}_i, \mathbf{t}_j)$  in  $N(\tau)$  **do**

    sum  $\leftarrow$  sum  $+ X(\mathbf{t}_i)X(\mathbf{t}_j)$ ;

**end**

**return** (sum/ $|N(\tau)|$ )

**Result:** An estimated isotropic covariance function for  $X$  at a desired radius  $\tau$

### C. Pseudocode for computing isotropic covariance function

### D. Positive-definiteness of classical estimator with constant $N(h)$

Without loss of generality, assume that  $X$  is a zero-mean random field. Let  $K = \{\tilde{s}_k\}$  denote the set of observations, and let us set  $X(s) = 0$  if  $s \notin K$ . As in the covariance function estimator, we use only the values of  $X$  over the set  $K$ , this assumption does not change the estimate but proves to be useful to avoid considering different subcases. For  $I = \{s_i, i = 1, \dots, n\}$ , consider a set  $S_{K,I} = \{\tilde{s}_k - s_i, \tilde{s}_k \in K, s_i \in I\}$ , and let  $a_1, \dots, a_n \in \mathbb{C}$  be arbitrary constants. Then, for  $N(s_i - s_j) = N$ , where  $N$  is constant, we obtain

$$\begin{aligned} \sum_{i,j=1}^n a_i \hat{C}(s_i - s_j) \bar{a}_j &= \sum_{i,j=1}^n a_i \left( \frac{1}{|N(s_i - s_j)|} \sum_{k \in N(s_i - s_j)} X(\tilde{s}_k) \overline{X(\tilde{s}_k + (s_i - s_j))} \right) \bar{a}_j \\ &= \frac{1}{N} \sum_{i,j=1}^n a_i \left( \sum_{s \in S_{K,I}} X(s + s_i) \overline{X(s + s_j)} \right) \bar{a}_j = \frac{1}{N} \sum_{s \in S_{K,I}} \sum_{i,j=1}^n a_i X(s + s_i) \overline{X(s + s_j)} \bar{a}_j \\ &= \frac{1}{N} \sum_{s \in S_{K,I}} \left| \sum_{i=1}^n a_i X(s + s_i) \right|^2 \geq 0. \end{aligned}$$

### E. Summability identity for classical spatial covariogram estimator

By Definition (1), the standard estimate of spatial autocorrelation function is

$$\hat{\rho}(\mathbf{h}) = \frac{\sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X}) / N(\mathbf{h})}{\sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X})^2 / N},$$

Assume that  $N(\mathbf{h}) \equiv N$ . Then,

$$\begin{aligned} \sum_{\mathbf{h} \in H} \hat{\rho}(\mathbf{h}) &= \sum_{\mathbf{h} \in H} \frac{\sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})}{\sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X})^2} \\ &= \frac{\sum_{\mathbf{h} \in H} \sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})}{\left( \sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X}) \right)^2 - \sum_{i=1}^N \sum_{\substack{j=1 \\ j \neq i}}^N (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})} \end{aligned}$$

$$= \frac{\sum_{\mathbf{h} \in H} \sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})}{-\sum_{\mathbf{h} \in H} \sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})} = -1.$$

Note, that a similar identity also holds for the case of nonconstant  $N(\mathbf{h})$ . Namely, from the above identity, it follows that

$$\begin{aligned} -1 &= \sum_{\mathbf{h} \in H} \hat{\rho}(\mathbf{h}) = \frac{\sum_{\mathbf{h} \in H} \hat{C}^{**}(\mathbf{h})}{\widehat{\text{Var}}(X)} = \frac{\sum_{\mathbf{h} \in H} \frac{1}{N} \sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})}{\frac{1}{N} \sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X})^2} \\ &= \sum_{\mathbf{h} \in H} \frac{|N(\mathbf{h})|}{N} \frac{1}{|N(\mathbf{h})|} \frac{\sum_{\mathbf{t}_i, \mathbf{t}_j \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X})}{\frac{1}{N} \sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X})^2} = \sum_{\mathbf{h} \in H} \frac{|N(\mathbf{h})|}{N} \tilde{\rho}(\mathbf{h}), \end{aligned}$$

where  $\tilde{\rho}(\mathbf{h}) = \hat{C}(\mathbf{h})/\hat{C}(\mathbf{0})$ . Thus, the weighted sum of  $\tilde{\rho}(\mathbf{h})$  maintains a constant value regardless of the observations  $X(\mathbf{t}_i)$ , and that the weights  $|N(\mathbf{h})|/N$  depend only the locations of the observations, not the values at the observations.

## F. Relation between classical covariogram and semivariogram estimators

The relation  $\hat{\gamma}(\mathbf{h}) = \hat{C}(\mathbf{0}) - \hat{C}(\mathbf{h})$  does not hold for the classical estimators as follows from

$$\begin{aligned} \hat{\gamma}(\mathbf{h}) &= \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} (X(\mathbf{t}_i) - X(\mathbf{t}_j))^2 = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} ((X(\mathbf{t}_i) - \bar{X}) - (X(\mathbf{t}_j) - \bar{X}))^2 \\ &= \frac{1}{2|N(\mathbf{h})|} \sum_{i \in N(\mathbf{h})} 2(X(\mathbf{t}_i) - \bar{X})^2 - \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} 2(X(\mathbf{t}_i) - \bar{X})(X(\mathbf{t}_j) - \bar{X}) \\ &= \frac{1}{|N(\mathbf{h})|} \sum_{i \in N(\mathbf{h})} (X(\mathbf{t}_i) - \bar{X})^2 - \hat{C}(\mathbf{h}). \end{aligned}$$

As the estimate  $\hat{C}(\mathbf{0}) = N^{-1} \sum_{i=1}^N (X(\mathbf{t}_i) - \bar{X})^2$  incorporates all spatial locations, not only those separated by a vector  $\mathbf{h}$ , the relation  $\hat{\gamma}(\mathbf{h}) = \hat{C}(\mathbf{0}) - \hat{C}(\mathbf{h})$  is not valid in general, except the case when  $\mathbf{h} = \mathbf{0}$ .

### Affiliation:

Adam Bilchouris  
Department of Mathematical and Physical Sciences  
La Trobe University  
Melbourne, 3086, Victoria, Australia  
E-mail: [A.Bilchouris@latrobe.edu.au](mailto:A.Bilchouris@latrobe.edu.au)  
URL: <https://scholars.latrobe.edu.au/abilchouris>

Andriy Olenko  
Department of Mathematical and Physical Sciences  
La Trobe University  
Melbourne, 3086, Victoria, Australia  
E-mail: [A.Olenko@latrobe.edu.au](mailto:A.Olenko@latrobe.edu.au)  
URL: <https://sites.google.com/site/olenkoandriy/>