

Transregio
391

TRR 391 Working Paper #14

November 2026

Nonparametric Covariance Estimation with Bias Correction for Spatial Lattice Processes

Antonio Fioravanti[a], Carsten Jentsch[a]

Suggested citation:
Fioravanti, A. and Jentsch, C. (2026). "Nonparametric Covariance Estimation with Bias Correction for Spatial Lattice Processes". *TRR 391 Working Paper #14*. DOI: 10.17877/DE290R-27046.

Version 1.0, November 2026

[a] Department of Statistics, TU Dortmund University
Corresponding author: antonio.fioravanti@tu-dortmund.de

Nonparametric covariance estimation with bias correction for spatial lattice processes

Antonio Fioravanti*, Carsten Jentsch

Department of Statistics, TU Dortmund University, 44221 Dortmund, Germany

Abstract

Accurate covariance estimation is crucial for spatial data analysis. While parametric methods can suffer from model misspecification leading to wrong conclusions, nonparametric approaches are often neglected in practice as they rely on the estimation of a large number of covariance parameters and often face finite-sample bias issues.

In this paper, we study the bias properties of sample covariance estimators for stationary lattice processes on $\mathbb{Z}^2$, when the true mean parameter is unknown and has to be estimated. We derive exact formulas for the finite-sample biases of sample covariance estimators and show that their expectations are linear combinations of population covariances determined by spatial lag and sample size. Based on this characterization, we propose jointly bias-corrected covariance estimators that are nearly unbiased. Additionally, we derive formulas for the mean-squared error and prove asymptotic normality results that show asymptotic equivalence for the estimators with and without bias correction. Simulations demonstrate substantial reductions of bias and often also in terms of MSE, particularly under strong spatial dependence. A kriging application to topography data further illustrates the practical benefits of the proposed approach.

Keywords: Spatial data, random fields, spatial dependence, covariance estimation, lattice processes, finite-sample bias
2020 MSC: 62M30, 62M40, 62G05

1. Introduction

The study of spatially dependent processes, that is, of random variables that are indexed by a set of discrete spatial coordinates of dimension larger than one, has become increasingly important across different scientific disciplines. Among others, these include geostatistics, environmental science, epidemiology, image analysis and econometrics. When studying spatial data, a key aspect is the analysis of the spatial dependence, that is, how strong the random variables are correlated with others in nearby locations and how fast this (linear) stochastic dependence actually decays when the distance between two locations becomes larger. The accurate estimation of the dependence a central problem in the analysis of any kind of dependent processes. Compared with time series, where dependence evolves along a single ordered dimension, spatial processes exhibit dependence across several directions. This makes the dependence structure more complex and accurate covariance estimation becomes even more crucial to guarantee meaningful inference on the underlying process

*Corresponding author
Email addresses: fioravanti@statistik.tu-dortmund.de (Antonio Fioravanti), jentsch@statistik.tu-dortmund.de (Carsten Jentsch)

and reliable prediction at unobserved locations. For a broad review of spatial statistics and spatial dependence concepts, see Cressie (1993); Sherman (2010); Schabenberger and Gotway (2017).

Two major approaches are commonly used to model spatial dependence. On the one hand, parametric approaches are used to specify models for the covariance function based on a (low-dimensional) parameter vector, such as, e.g., the Matérn, exponential or spherical model, among others (Cressie, 1993; Guttorp and Gneiting, 2006) and estimates its parameters from the data, typically by maximum likelihood (Mardia and Marshall, 1984) or by fitting an empirical variogram using weighted least squares (Cressie, 1985). Parametric models are attractive, because valid covariance families guarantee positive semidefiniteness of the estimated covariance function and boil down the problem to estimating some (low-dimensional) parameter vector that determines the covariance function. However, parametric approaches impose potentially restrictive assumptions on the covariance structure, and the effect of covariance misspecification, e.g., on spatial prediction depends critically on the appropriateness of the assumed covariance models (Stein and Handcock, 1989). Sometimes even under correct specification, e.g. of a Matérn covariance model, individual covariance parameters may fail to be consistently estimable under fixed-domain asymptotics (Zhang, 2004).

On the other hand, to address the possible restrictiveness of parametric assumptions, nonparametric approaches estimate the covariances or (semi-)variograms directly from the data without imposing any specific parametric form. Classical geostatistics predominately favors empirical variogram estimation (Matheron, 1962); while more flexible methods include kernel-methods (Hall et al., 1994), bootstrap methods (BjØrnstad and Falck, 2001), sieve maximum likelihood (Wang and Ghosh, 2023), among others. A recent review of nonparametric methods for the estimation of the covariance for spatial data is provided by Bilchouris and Olenko (2025), see also the references therein. Relying on the common method-of-moment principle, the most direct nonparametric technique is based on the sample covariance. This estimator, however, is well known to be biased in finite samples, when the true mean is unknown (Sherman, 2010).

For lattice covariance estimation, the focus has been on the bias caused by edge or boundary effects, which are typical for lattices and higher-dimensional data arrays. Guyon (1982) showed that such a bias is not asymptotically negligible in a lattice of dimension larger or equal to 2, and proposed to normalize the sample covariance by the number of observed pairs at its respective lag rather than using the total sample size, which in this paper is referred to as the *standard estimator*. This solved the issue of the bias induced by the edge effect (Ripley, 1984). Later, Dahlhaus and Künsch (1987) introduced data tapering to reduce the large variance caused by estimating covariances at high lags. More recently, Guillaumin et al. (2022) propose a debiasing of the Whittle quasi-likelihood estimator by replacing the spectral density by the expected periodogram. However, all these methods operate in settings where the process mean is known and often set to zero. In practice, this is not the case and the unknown mean needs to replaced by a suitable estimator (typically the sample mean) in the estimation process. In comparison to the edge effect bias, this introduces a different kind of bias Sherman (2010), which is the subject of the present paper.

The same bias problem arises also in time series analysis, where for univariate time series, Vogelsang and Yang (2016) proposed a matrix inversion framework that yields nearly unbiased sample covariance estimation, subsequently extended also to the multivariate time series case by Li (2023). To the best of our knowledge, no analogous bias correction exists for spatial lattice processes, where the two dimensional structure introduces additional complexity and the number of covariances to be estimated becomes large already for small sample sizes. The present paper aims to fill this gap.

The remainder of the paper is organized as follows. In Section 2, we introduce the spatial lattice setup and provide the notation used throughout the paper. Then, we present the two versions of sample covariance estimators studied in this paper, which differ only in their prefactors, and

derive formulas for their exact finite sample bias and also jointly for the whole matrix of sample covariances at all observed lags. Based on these findings, in Section 3, we propose a joint bias-corrected sample covariance (matrix) estimator that is demonstrated to be nearly unbiased. We quantify the remaining bias, calculate the mean squared error and suggest two data-driven rules for the selection of a required cutoff parameter. Moreover, we establish that the correction approach applied to both versions of sample covariance estimators studied in this paper lead exactly to the same bias-corrected estimator and we present asymptotic normality results. In Section 4, we report the results of a Monte Carlo simulation study for separable and nonseparable Gaussian random fields. A real-data application is shown in Section 5, where we apply bias-corrected covariance estimation to conduct ordinary kriging on a topographic dataset. Section 6 concludes and states future research directions. All proofs are collected in the Appendix.

2. The exact bias of sample covariances for lattice processes

In this section, after presenting the notation and assumptions employed in this paper, we introduce the two sample covariance estimators (which differ only in the prefactor used), we derive the finite sample bias of each of them, first for a single scalar estimator, and then for a joint matrix estimator of covariances. Finally, we describe the second-order structure of the estimators assuming Gaussianity.

2.1. Spatial lattice processes: setup and notation

We consider a real-valued univariate spatial process $\mathbf{X} = \{X_{\underline{t}},\ \underline{t} = (t_1, t_2) \in \mathbb{Z}^2\}$, indexed on the two-dimensional integer lattice[1]. A vector $\underline{h} = (h_1, h_2) \in \mathbb{Z}^2$ is called a spatial *lag* and represents the shift separating a location $\underline{t}$ from the location $\underline{t} + \underline{h}$. We state two commonly imposed assumptions in the following.

Assumption 2.1 (Second-order stationarity)**.** $\mathbf{X}$ *is second-order (or weakly) stationary with constant mean* $\mathbb{E}(X_{\underline{t}}) = \mu$ *for all* $\underline{t} \in \mathbb{Z}^2$ *and with spatial covariance function* $C : \mathbb{Z}^2 \to \mathbb{R}$*, such that*

$$C(\underline{h}) := \mathrm{Cov}(X_{\underline{t}+\underline{h}}, X_{\underline{t}}) = \mathbb{E}\big(\big(X_{\underline{t}+\underline{h}} - \mu\big)\big(X_{\underline{t}} - \mu\big)\big), \quad \forall \underline{t}, \underline{h} \in \mathbb{Z}^2. \tag{2.1}$$

In particular, it holds that $C(\underline{h}) = C(-\underline{h})$*.*

Assumption 2.2 (Short-range dependence)**.** *The covariance function* C *from* (2.1) *is absolutely summable over* $\mathbb{Z}^2$*. That is, we have*

$$\|C\|_1 = \sum_{\underline{h} \in \mathbb{Z}^2} |C(\underline{h})| < \infty. \tag{2.2}$$

In particular, under Assumption 2.1, a Gaussian random field is fully characterized by its first and second moments, while Assumption 2.2 imposes the common short-range dependence condition that excludes long-memory lattices fields.

Next, we define how data from the process $\mathbf{X}$ is observed. Suppose that for $\underline{n} = (n_1, n_2) \in \mathbb{N}^2$, we obtain a lattice realization and collect the observations in the data matrix $\mathbf{X}_{\underline{n}} := \big(X_{\underline{t}},\ \underline{t} \in G_{\underline{n}}\big)$ of the spatial process $\mathbf{X}$ on a $(n_1 \times n_2)$ rectangular grid, whose unique coordinate pairs are collected in the index set

$$G_{\underline{n}} = \big\{\underline{t} = (t_1, t_2) : 1 \le t_i \le n_i,\ i = 1, 2\big\} \subset \mathbb{Z}^2. \tag{2.3}$$

[1]Here and throughout the manuscript, vector-valued quantities are denoted by underlined lower-case letters $\underline{v}$ and matrices by bold upper-case letters $\mathbf{M}$.

In the following, we denote its cardinality $|G_{\underline{n}}|$, i.e., the total amount of unique coordinate pairs in the grid $G_{\underline{n}}$, also as the *total* sample size and we define

$$N := |G_{\underline{n}}| = n_1 n_2. \tag{2.4}$$

Accordingly, all lags realized by $G_{\underline{n}}$ are collected in

$$H_{\underline{n}} := \{\underline{h} = (h_1, h_2) : |h_i| \leq n_i - 1,\ i \in \{1, 2\}\} \subset \mathbb{Z}^2. \tag{2.5}$$

Note that not every location in $G_{\underline{n}}$ admits every lag in $H_{\underline{n}}$: a location $\underline{t}$ and lag $\underline{h}$ has an observation in $\mathbf{X}_{\underline{n}}$ if and only if the location $\underline{t} + \underline{h}$ is also contained in the observed grid $G_{\underline{n}}$. Thus, we define also the subset of all valid locations for the lag $\underline{h}$ as

$$G_{\underline{n}}(\underline{h}) := \left\{\underline{t} \in G_{\underline{n}} : \underline{t} + \underline{h} \in G_{\underline{n}}\right\} \subseteq G_{\underline{n}}, \qquad \underline{h} \in H_{\underline{n}}. \tag{2.6}$$

Its cardinality $|G_{\underline{n}}(\underline{h})|$, is the *effective* sample size corresponding to lag $\underline{h}$, i.e., the number of pairs $(X_{\underline{t}}, X_{\underline{t}+\underline{h}})$ separated by the lag $\underline{h}$ such that both $X_{\underline{t}}$ and $X_{\underline{t}+\underline{h}}$ are contained in the grid $G_{\underline{n}}$. This leads to

$$N_h := |G_{\underline{n}}(\underline{h})| = (n_1 - |h_1|)(n_2 - |h_2|). \tag{2.7}$$

Finally, we introduce some set notation. First, for arbitrary integers $a, b \in \mathbb{Z}$ with $a \leq b$, we use the short notation

$$I(a : b) := \{a,\ a + 1,\ a + 2,\ \ldots,\ b - 2,\ b - 1,\ b\} \tag{2.8}$$

for the set of all integers from a to b. Hence, for $i \in \{1, 2\}$, the lags of $\underline{h} \in H_{\underline{n}}$ run along the interval $I\big((-n_i + 1) : (n_i - 1)\big)$ and the locations $\underline{t} \in G_{\underline{n}}(\underline{h})$ run along the interval $I\big(\max(1, 1 - h_i) : \min(n_i, n_i - h_i)\big)$.

Throughout the paper, we will deal with the vectorization operator $\mathrm{vec}(\cdot)$ that stacks a rectangular lattice column-wise. Accordingly, the mapping that allows to recover the correct double indices of the lattice from the single index of the vectorized lattice can be defined as follows: for a lattice $F = I(a_1 : b_1) \times I(a_2 : b_2)$ with set cardinalities $c_i := |I(a_i : b_i)| = b_i - a_i + 1$, $i \in \{1, 2\}$ and total cardinality $|F| = c_1 c_2$, we use the mapping

$$\underline{\ell}_F : \{1, \ldots, |F|\} \to F, \quad \underline{\ell}_F(k) := \left(\big((k - 1) \bmod c_1\big) + a_1,\ \left\lfloor \frac{k - 1}{c_1} \right\rfloor + a_2\right), \tag{2.9}$$

where $\lfloor x \rfloor$ denotes the largest integer smaller or equal x, and $a \bmod b$ denotes the remainder of a upon division by b.

We illustrate this spatial lattice setup with a Gaussian example, which also serves as the data-generating process in our simulation study in Section 4.

Example 2.1 (Gaussian spatial lattice process). Suppose the spatial lattice process $\mathbf{X}$ is Gaussian. Further, suppose that the covariance $C(\underline{h})$ between arbitrary points $X_{\underline{t}+\underline{h}}$ and $X_{\underline{t}}$ having spatial lag $\underline{h} = (h_1, h_2)$ is given by the exponential covariance function

$$C_\lambda(h_1, h_2) = \sigma^2 \exp\left(-\frac{|h_1|}{\lambda} - \frac{|h_2|}{\lambda} - \beta\frac{|h_1 - h_2|}{\lambda}\right), \quad \underline{h} \in H_{\underline{n}}, \tag{2.10}$$

where $\sigma^2 = \mathrm{Var}(X_{\underline{t}}) = C_\lambda(0, 0)$ is the variance of the lattice process $\mathbf{X}$, the parameter λ controls the persistence of the spatial dependence and β the degree of (non)-separability. In this paper, we consider $\lambda \in \{2, 4, 6, 8\}$, where $\lambda = 2$ gives the weakest and $\lambda = 8$ the strongest spatial dependence. Further, $\beta = 0$ leads to a separable covariance function, while $\beta \in (0, 1]$ implies non-separability.

In Figure 1, we show realizations $\mathbf{X}_{\underline{n}}$ for $n_1 = n_2 = 30$ of two (Gaussian) spatial processes $\mathbf{X}$ for the exponential covariance function defined in (2.10) for $\lambda = 4$ (moderate dependence) and $\lambda = 8$ (large dependence), respectively, and $\beta = 0$ (separable case).

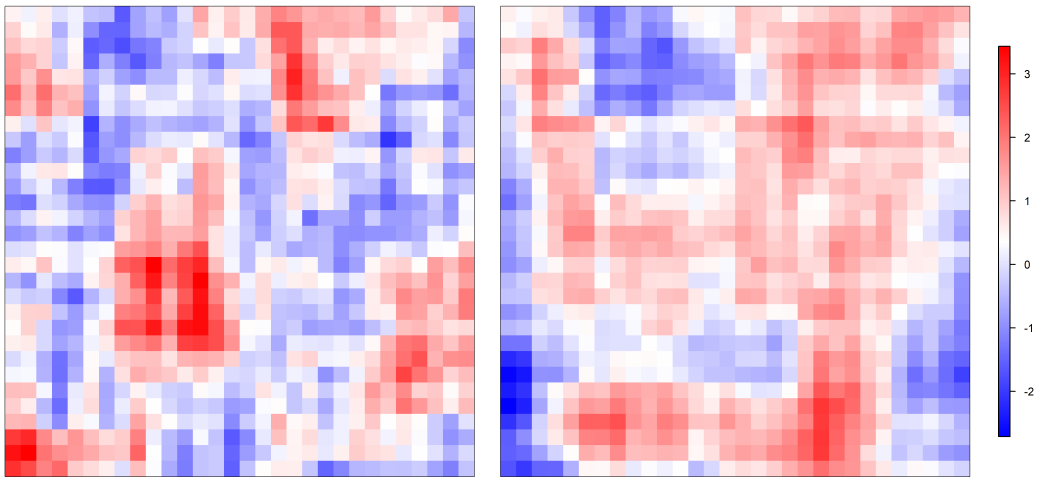


Figure 1: Realizations $\mathbf{X}_{\underline{n}}$ with $n_1 = n_2 = 30$ of spatial (Gaussian) processes $\mathbf{X}$ with exponential covariance function C_λ defined in (2.10) with $\lambda = 4$ (left panel) and $\lambda = 8$ (right panel) and $\beta = 0$ (separable case) in both panels.

2.2. Sample covariance estimators for spatial lattice processes

We consider two versions of spatial sample covariance estimators that have been proposed in the literature. As defined in the following, they differ only in their prefactors. On the one hand, using the prefactor $1/N_{\underline{h}}$ that scales by the *effective* sample size $N_{\underline{h}}$ from (2.7) for lag $\underline{h}$, we define

$$\widehat{C}(\underline{h}) = \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \big(X_{\underline{t}+\underline{h}} - \overline{X}\big)\big(X_{\underline{t}} - \overline{X}\big), \quad \underline{h} \in H_{\underline{n}}, \tag{2.11}$$

and call $\widehat{C}(\underline{h})$ a *standard covariance estimator*. On the other hand, using the prefactor $1/N$ that scales by the *total* sample size N from (2.4) for all lags $\underline{h}$, we define

$$\widetilde{C}(\underline{h}) = \frac{1}{N} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \big(X_{\underline{t}+\underline{h}} - \overline{X}\big)\big(X_{\underline{t}} - \overline{X}\big), \quad \underline{h} \in H_{\underline{n}}. \tag{2.12}$$

In the following, $\widetilde{C}(\underline{h})$ is called a *constant prefactor covariance estimator*. In the above, $G_{\underline{n}}(\underline{h})$ is the set of coordinates that admit the lag $\underline{h}$ from (2.6), $\overline{X}$ denotes the sample mean of the observed data on the whole grid $G_{\underline{n}}$, defined as

$$\overline{X} = \frac{1}{N} \sum_{\underline{u} \in G_{\underline{n}}} X_{\underline{u}}, \tag{2.13}$$

and, as usual, we set $\widehat{C}(\underline{h}) = \widetilde{C}(\underline{h}) := 0$ if $\underline{h} \notin H_{\underline{n}}$.

The standard estimator (2.11) is normalized by the effective sample size N_h from (2.7) and is the version most commonly used in geostatistics (Cressie, 1993). The constant prefactor estimator (2.12) instead divides every lag by the total sample size N. Although the latter approach is quite uncommon in the spatial statistics literature, it can be motivated and is largely inspired by analogous practices in time series analysis. In particular, the choice of the reciprocal sample size as a universal prefactor for all sample covariances (irrespective of the lag $\underline{h}$) guarantees positive definiteness of the resulting sample covariance function (and all sample covariance matrices); see Brockwell and Davis (1991); Kreiß and Neuhaus (2006).

The standard estimator $\widehat{C}(\underline{h})$ (2.11) is mainly motivated by the fact that its *infeasible* version

$$\widehat{C}_\mu(\underline{h}) = \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \big(X_{\underline{t}+\underline{h}} - \mu\big)\big(X_{\underline{t}} - \mu\big), \quad \underline{h} \in H_{\underline{n}}, \tag{2.14}$$

where $\overline{X}$ is replaced by μ, is an unbiased estimator of the covariance function $C(\underline{h})$, i.e., $\mathbb{E}\big(\widehat{C}_\mu(\underline{h})\big) = C(\underline{h})$; see Guyon (1982). In contrast, the analogously defined *infeasible* version $\widetilde{C}_\mu(\underline{h})$ of the constant prefactor estimator $\widetilde{C}(\underline{h})$ from (2.12) does still exhibit a bias (unless $\underline{h} = (0,0)$) due to the edge or boundary effect, i.e., $\mathbb{E}\big(\widetilde{C}_\mu(\underline{h})\big) \neq C(\underline{h})$ (Guyon, 1982). Still when comparing $\widehat{C}(\underline{h})$ and $\widetilde{C}(\underline{h})$ under the unknown mean setting, $\widehat{C}(\underline{h})$ exhibits a smaller bias at small lags, thus the standard estimator is often preferred over the constant prefactor estimator.

The issue on which we focus in this paper arises when the true mean μ is *unknown* and the sample mean estimator $\overline{X}$ from (2.13) (which is itself an unbiased estimator for μ) is used in (2.11)-(2.12). In this case, also $\widehat{C}(\underline{h})$ is biased, i.e., $\mathbb{E}(\widehat{C}(\underline{h})) \neq C(\underline{h})$, see Sherman (2010).

2.3. Finite sample bias of a single covariance estimator

In this section, we quantify the finite-sample biases of both versions of sample covariance estimators $\widehat{C}(\underline{h})$ and $\widetilde{C}(\underline{h})$ defined in (2.11) and (2.12), respectively, that arise when the true mean μ of the process $\mathbf{X}$ is unknown and the sample mean $\overline{X}$ from (2.13) is plugged-in for the estimation. Consequently, both estimators are computed from the deviations of the observations from the sample mean. Under Assumption 2.1, this implies that their moments are governed by the second-order structure of the demeaned process, rather than that of the process $\mathbf{X}$. By taking a closer look at the covariance between $X_{\underline{t}} - \overline{X}$ and $X_{\underline{s}} - \overline{X}$ for some $\underline{t}, \underline{s} \in G_{\underline{n}}$, we get

$$C_Z(\underline{t}, \underline{s}) := \mathrm{Cov}(X_{\underline{t}} - \overline{X}, X_{\underline{s}} - \overline{X}) = C(\underline{t} - \underline{s}) - \overline{C}(\underline{t}) - \overline{C}(\underline{s}) + \mathrm{Var}(\overline{X}), \tag{2.15}$$

where, for $\underline{t} \in G_{\underline{n}}$,

$$\overline{C}(\underline{t}) := \mathrm{Cov}(X_{\underline{t}}, \overline{X}) = \frac{1}{N} \sum_{\underline{u} \in G_{\underline{n}}} C(\underline{t} - \underline{u}) \tag{2.16}$$

denotes the covariance between a single observation $X_{\underline{t}}$ and the sample mean $\overline{X}$. Further, for the variance of the sample mean, we get

$$D := \mathrm{Var}(\overline{X}) = \frac{1}{N} \sum_{\underline{v} \in G_{\underline{n}}} \overline{C}(\underline{v}) = \frac{1}{N^2} \sum_{\underline{v} \in G_{\underline{n}}} \sum_{\underline{u} \in G_{\underline{n}}} C(\underline{v} - \underline{u}). \tag{2.17}$$

We refer to C_Z as the *centered covariance function* of the process $\mathbf{X}$. In contrast to C, which takes as arguments the lags $\underline{h}$, the centered covariance function C_Z is a function of the two locations. That is, $C_Z(\underline{t}, \underline{s})$ does depend on the positions $\underline{t}$ and $\underline{s}$ and not only on their difference $\underline{t} - \underline{s}$.

Now, with the notation introduced above, the expectation of the standard covariance estimator $\widehat{C}(\underline{h})$ becomes

$$\mathbb{E}\big(\widehat{C}(\underline{h})\big) = \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} C_Z(\underline{t} + \underline{h}, \underline{t}) = C(\underline{h}) - A_{\underline{h}} - B_{\underline{h}} + D, \tag{2.18}$$

where D is the variance of the sample mean (2.17) and where

$$A_{\underline{h}} := \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \overline{C}(\underline{t} + \underline{h}) = \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \mathrm{Cov}\big(X_{\underline{t}+\underline{h}}, \overline{X}\big) = \frac{1}{N_{\underline{h}} N} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{u} \in G_{\underline{n}}} C(\underline{t} + \underline{h} - \underline{u}), \tag{2.19}$$

$$B_{\underline{h}} := \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \overline{C}(\underline{t}) = \frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \mathrm{Cov}\big(X_{\underline{t}}, \overline{X}\big) = \frac{1}{N_{\underline{h}} N} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{u} \in G_{\underline{n}}} C(\underline{t} - \underline{u}). \tag{2.20}$$

The representation of $\mathbb{E}(\widehat{C}(\underline{h}))$ in (2.18) makes the source of the bias explicit. The terms A_h and B_h from (2.19) and (2.20) are averages of the covariances between the observations entering the sample covariance estimator, namely $X_{\underline{t}+\underline{h}}$ and $X_{\underline{t}}$, respectively, and the sample mean, while D is the variance of the sample mean itself. Note that each of the two terms (2.19) and (2.20) are double sums over the set $G_{\underline{n}}(\underline{h})$ and $G_{\underline{n}}$. Collecting the terms accordingly reduces each double sum to a single sum over $\underline{r} \in H_{\underline{n}}$, weighted by counts that factorize across the two lattice dimensions. Details are given in the Appendix, and the result is summarized in the following theorem.

Theorem 2.1 (Exact bias of a single $\widehat{C}(\underline{h})$)**.** *Suppose Assumption 2.1 holds and let $\underline{h} \in H_{\underline{n}}$. Then, the bias of the standard covariance estimator $\widehat{C}(\underline{h})$ defined in* (2.11) *is*

$$b(\underline{h}) := \mathbb{E}\big(\widehat{C}(\underline{h})\big) - C(\underline{h}) = -\frac{1}{N_{\underline{h}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \big[\mathrm{Cov}(X_{\underline{t}+\underline{h}}, \overline{X}) + \mathrm{Cov}(X_{\underline{t}}, \overline{X})\big] + \mathrm{Var}(\overline{X}) \tag{2.21}$$

or, equivalently represented, we have

$$b(\underline{h}) = -\frac{S(\underline{h}) + S(-\underline{h})}{N N_{\underline{h}}} + \frac{S(\underline{0})}{N^2}, \tag{2.22}$$

where $S(\underline{h})$ is a weighted sum running over the full lag set $H_{\underline{n}}$ defined by

$$S(\underline{h}) := \sum_{\underline{r} \in H_{\underline{n}}} C(\underline{r}) \prod_{i=1}^{2} \nu_{n_i}(r_i, h_i), \tag{2.23}$$

and, with $x^+ := \max(x, 0)$, the counting factors $\nu_{n_i}(r_i, h_i)$ are given by

$$\nu_{n_i}(r_i, h_i) := \big(n_i - \max(|r_i|, |h_i|, |r_i + h_i|)\big)^+, \quad i \in \{1, 2\}. \tag{2.24}$$

A similar bias formula holds for the sample covariance estimator with constant prefactor $\widetilde{C}(\underline{h})$ from (2.12), which we state as a simple corollary.

Corollary 2.1 (Exact bias of a single $\widetilde{C}(\underline{h})$)**.** *Under the conditions of Theorem 2.1, the bias of the constant prefactor estimator $\widetilde{C}(\underline{h})$ defined in* (2.12) *becomes*

$$\begin{aligned} \widetilde{b}(\underline{h}) := \mathbb{E}\big(\widetilde{C}(\underline{h})\big) - C(\underline{h}) &= \left(\frac{N_h}{N} - 1\right) C(\underline{h}) + \frac{N_h}{N} b(\underline{h}) \\ &= \left(\frac{N_{\underline{h}}}{N} - 1\right) C(\underline{h}) - \frac{S(\underline{h}) + S(-\underline{h})}{N^2} + \frac{N_{\underline{h}}}{N^3} S(\underline{0}), \end{aligned} \tag{2.25}$$

with $S(\underline{h})$ as defined in (2.23)*.*

Some remarks on the biases presented in Theorem 2.1 and Corollary 2.1 are in order.

Remark 2.1 (On the bias formulas)*.*

(i) Both biases $b(\underline{h})$, $\widetilde{b}(\underline{h})$ from (2.22) and (2.25) are weighted sums of all covariances between any two random variables in $\mathbf{X}_{\underline{n}}$, where the weights depend only on the sample sizes (n_1, n_2) and the spatial lags (h_1, h_2). In particular, the bias does *not* depend on the unknown mean $\mu = \mathbb{E}(X_{\underline{t}})$.

(ii) If the true mean μ is known and replaces the sample mean $\overline{X}$ in the standard estimator $\widehat{C}(\underline{h})$ (leading to $\widehat{C}_\mu(\underline{h})$ defined in (2.14)), we get an unbiased estimator as all the terms on the right-hand side of (2.21) become zero in that case.

(iii) Conversely, the constant prefactor estimator $\widetilde{C}(\underline{h})$ remains biased even when the true mean μ is known and replaces the sample mean $\overline{X}$ (leading to $\widetilde{C}_\mu(\underline{h})$), as that information does not remove the term $(N_h/N-1)C(\underline{h})$ from the bias in (2.25). This leftover bias in the $\widetilde{C}(\underline{h})$ estimator is caused by the *edge effect*, as it represents the fraction of lattice pairs lost to the edge or boundary of the lattice, as studied by Guyon (1982).

(iv) Under positive spatial dependence, i.e., whenever $C(\underline{h}) \geq 0$ for all $\underline{h} \in H_{\underline{n}}$, the bias is negative for both estimators, meaning that they systematically *underestimate* the true covariance.

(v) For the sample covariance at lag $\underline{h} = (0,0)$, i.e., when estimating the variance $\mathrm{Var}(X_{\underline{t}})$, we have $N_h = N$, and in this case $b(\underline{h})$ and $\widetilde{b}(\underline{h})$ from (2.22) and(2.25) coincide.

According to Remark 2.1 (i), for arbitrary spatial lags $\underline{h}$, the bias $b(\underline{h})$ of $\widehat{C}(\underline{h})$ (also for bias $\widetilde{b}(\underline{h})$ of $\widetilde{C}(\underline{h})$) does depend on *all* covariances between any two random variables in $\mathbf{X}_{\underline{n}}$. Hence, with the goal of bias corrected estimation in view, it appears to be advisable to consider sample covariances and their biases *jointly*, which is done in the following.

2.4. Joint bias of a matrix of the standard sample covariances

In this section, we adopt a matrix-based approach that allows for a simultaneous treatment of *all* spatial lags contained in $\underline{h} \in H_{\underline{n}}$ at once. We focus first on the standard estimator $\widehat{C}(\underline{h})$ from (2.11). For this purpose, we organize the population covariances in a comprehensive $(L_1 \times L_2)$ matrix $\mathbf{\Gamma}$, where $L_i = 2n_i - 1$ for $i \in \{1,2\}$ and $L_{\underline{n}} = L_1 L_2$, encompassing all lag combinations between any two random variables in $\mathbf{X}_{\underline{n}}$. That is, we define

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_{p,q}, \\ p = 1,\dots,L_1 \\ q = 1,\dots,L_2 \end{pmatrix} = \begin{pmatrix} C(h_1,h_2), \\ h_1 \in I(1-n_1 : n_1-1) \\ h_2 \in I(1-n_2 : n_2-1) \end{pmatrix}, \quad \big[\mathbf{\Gamma}\big]_{p,q} = C(p-n_1,\, q-n_2), \tag{2.26}$$

such that the matrix $\mathbf{\Gamma}$ contains all covariances $C(\underline{h})$, $\underline{h} \in H_{\underline{n}}$. In the same way, we define the corresponding matrix of sample covariances $\widehat{\mathbf{\Gamma}}$ by replacing all $C(h_1,h_2)$'s by $\widehat{C}(h_1,h_2)$'s in (2.26). Moreover, for $i \in \{1,2\}$, we arrange the counting factors $\nu_{n_i}(r_i,h_i)$ from (2.24) into three $(L_i \times L_i)$ matrices, whose rows are indexed by the lag $h_i = p - n_i$ and columns by the lag $r_i = q - n_i$. In the bias representation (2.22), the lag $\underline{h}$ enters as $S(\underline{h})$, $S(-\underline{h})$, $S(\underline{0})$, and thus affects only the second argument of the counting factor $\nu_{n_i}(r_i,h_i)$ from (2.24). Accordingly, for $i \in \{1,2\}$ and $p,q = 1,\dots,L_i$, we define the *raw counting matrices* $\mathbf{A}_i$, $\mathbf{B}_i$, and $\mathbf{D}_i$ having entries

$$\big[\mathbf{A}_i\big]_{p,q} := \nu_{n_i}\big(q-n_i,\ p-n_i\big), \quad \big[\mathbf{B}_i\big]_{p,q} := \nu_{n_i}\big(q-n_i,\ n_i-p\big), \quad \big[\mathbf{D}_i\big]_{p,q} := \nu_{n_i}\big(q-n_i,\ 0\big), \tag{2.27}$$

which are displayed in Figure 2 for $n_1 = n_2 = 4$. These raw counting matrices have to be normalized in order to incorporate the prefactors $(1/NN_h)$ and $1/N^2$ appearing in (2.22). Since $NN_{\underline{h}} = \prod_{i\in\{1,2\}} n_i\big(n_i - |h_i|\big)$, the relevant normalization in direction i is obtained by using the $(L_i \times L_i)$ diagonal matrix

$$\mathbf{N}_i := \mathrm{diag}\big(n_i(n_i - |p-n_i|)\big), \quad p = 1,\dots,L_i$$

and the *normalized count matrices* are defined by

$$\mathbf{A}_i^* := \mathbf{N}_i^{-1}\mathbf{A}_i, \quad \mathbf{B}_i^* := \mathbf{N}_i^{-1}\mathbf{B}_i, \quad \mathbf{D}_i^* := n_i^{-2}\,\mathbf{D}_i. \tag{2.28}$$

Matrices $A_1 = A_2$

$h_i \backslash r_i$	-3	-2	-1	0	1	2	3
-3	0	0	0	1	1	1	1
-2	0	0	1	2	2	2	1
-1	0	1	2	3	3	2	1
0	1	2	3	4	3	2	1
1	1	2	3	3	2	1	0
2	1	2	2	2	1	0	0
3	1	1	1	1	0	0	0

Matrices $B_1 = B_2$

$h_i \backslash r_i$	-3	-2	-1	0	1	2	3
-3	1	1	1	1	0	0	0
-2	1	2	2	2	1	0	0
-1	1	2	3	3	2	1	0
0	1	2	3	4	3	2	1
1	0	1	2	3	3	2	1
2	0	0	1	2	2	2	1
3	0	0	0	1	1	1	1

Matrices $D_1 = D_2$

$h_i \backslash r_i$	-3	-2	-1	0	1	2	3
-3	1	2	3	4	3	2	1
-2	1	2	3	4	3	2	1
-1	1	2	3	4	3	2	1
0	1	2	3	4	3	2	1
1	1	2	3	4	3	2	1
2	1	2	3	4	3	2	1
3	1	2	3	4	3	2	1

Figure 2: The matrices $A_1 = A_2$ (left panel), $B_1 = B_2$ (center panel) and $D_1 = D_2$ (right panel) for $n_1 = n_2 = 4$. As $L_i = 2n_i - 1$ for $i \in \{1, 2\}$, all matrices are of dimension 7×7.

Accordingly, their entries can be shown by using the counting factors from (2.24) and by indexing the matrices with $h_i, r_i \in I(1 - n_i : n_i - 1)$. That is, we have

$$[\mathbf{A}_i^*]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, h_i)}{n_i(n_i - |h_i|)}, \qquad [\mathbf{B}_i^*]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, -h_i)}{n_i(n_i - |h_i|)}, \qquad [\mathbf{D}_i^*]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, 0)}{n_i^2}. \tag{2.29}$$

Now, making use of (2.18), we can express the expectation of the whole matrix of sample covariances $\widehat{\mathbf{\Gamma}}$ compactly as

$$\mathbb{E}\big(\widehat{\mathbf{\Gamma}}\big) = \mathbf{\Gamma} - \mathbf{A}_1^* \mathbf{\Gamma} \mathbf{A}_2^{*\top} - \mathbf{B}_1^* \mathbf{\Gamma} \mathbf{B}_2^{*\top} + \mathbf{D}_1^* \mathbf{\Gamma} \mathbf{D}_2^{*\top}. \tag{2.30}$$

Finally, by applying the vectorization operator $\mathrm{vec}(\cdot)$ on both sides of Equation (2.30) and using the Kronecker product properties[2], we can represent the expectation of the vectorized sample covariance matrix $\widehat{\mathbf{\Gamma}}$ as a linear combination of all population covariances in $\mathbf{\Gamma}$ and provide its joint bias formula, which is summarized in the following theorem.

Theorem 2.2 (Joint mean and bias of $\mathrm{vec}(\widehat{\mathbf{\Gamma}})$)**.** *Under Assumption 2.1, using the vectorization operator* $\mathrm{vec}(\cdot)$*, Equation* (2.30) *can equivalently expressed as*

$$\mathbb{E}\big(\mathrm{vec}(\widehat{\mathbf{\Gamma}})\big) = \mathbf{W}_{\underline{n}}^* \, \mathrm{vec}(\mathbf{\Gamma}), \tag{2.31}$$

where the $(L_{\underline{n}} \times L_{\underline{n}})$ *weight matrix*

$$\mathbf{W}_{\underline{n}}^* = \mathbf{I}_{L_{\underline{n}}} - \mathbf{A}_2^* \otimes \mathbf{A}_1^* - \mathbf{B}_2^* \otimes \mathbf{B}_1^* + \mathbf{D}_2^* \otimes \mathbf{D}_1^* \tag{2.32}$$

is a rank-deficient deterministic matrix with $\mathrm{rank}(\mathbf{W}_{\underline{n}}^*) \leq L_{\underline{n}} - 1$ *that only depends on the grid sizes* (n_1, n_2)*. Further, the joint bias vector* $\underline{b}_{\mathrm{vec}(\widehat{\mathbf{\Gamma}})}$ *of* $\mathrm{vec}(\widehat{\mathbf{\Gamma}})$ *becomes*

$$\underline{b}_{\mathrm{vec}(\widehat{\mathbf{\Gamma}})} := \mathbb{E}\big(\mathrm{vec}(\widehat{\mathbf{\Gamma}})\big) - \mathrm{vec}(\mathbf{\Gamma}) = \big(\mathbf{W}_{\underline{n}}^* - \mathbf{I}_{L_{\underline{n}}}\big)\,\mathrm{vec}(\mathbf{\Gamma}). \tag{2.33}$$

Moreover, under Assumption 2.2, the bias (2.33) *is uniformly at most of order* $1/N$*. That is, we have*

$$\big\|\underline{b}_{\mathrm{vec}(\widehat{\mathbf{\Gamma}})}\big\|_\infty \leq \frac{1}{N}\big\|\mathrm{vec}(\mathbf{\Gamma})\big\|_1, \tag{2.34}$$

where, for a vector $\underline{v} \in \mathbb{R}^d$*,* $\|\underline{v}\|_\infty := \max_{1 \leq i \leq d} |v_i|$ *denotes its largest absolute component, whereas* $\|\underline{v}\|_1 := \sum_{i=1}^d |v_i|$ *denotes the sum of the absolute values of its components.*

[2]For matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times q}$, $\mathbf{A} \otimes \mathbf{B}$ indicates the Kronecker product, which is the $(mp) \times (nq)$ matrix obtained by multiplying each element a_{ij} of $\mathbf{A}$ by the matrix $\mathbf{B}$. In particular, for a fitting matrix $\mathbf{M} \in \mathbb{R}^{q \times n}$ it holds that $\mathrm{vec}(\mathbf{B}\mathbf{M}\mathbf{A}^\top) = (\mathbf{A} \otimes \mathbf{B})\mathrm{vec}(\mathbf{M})$. Moreover, for matrices $\mathbf{C} \in \mathbb{R}^{n \times r}$ and $\mathbf{D} \in \mathbb{R}^{q \times s}$, it holds that $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A}\mathbf{C}) \otimes (\mathbf{B}\mathbf{D})$.

Some remarks on the matrices $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{D}_i$, $i \in \{1,2\}$ from (2.27) and the weight matrix $\mathbf{W}^*_{\underline{n}}$ (2.32) and their properties are in order.

Remark 2.2 (On $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{D}_i$, and $\mathbf{W}^*_{\underline{n}}$).

(i) All matrices $\mathbf{A}_i$, $\mathbf{B}_i$ and $\mathbf{D}_i$ are "counting matrices" that have non-negative and integer-valued entries. All entries of $\mathbf{D}_i$ are strictly positive.

(ii) Despite $\mathbf{W}^*_{\underline{n}}$ may appear as a diagonally dominant matrix, for arbitrary sample sizes (n_1, n_2), all its row sums equal zero. For the case of $n_1 = n_2 = 4$, a heatmap of the resulting (49×49) weight matrix $\mathbf{W}^*_{\underline{n}}$ is shown in Figure 6 (left panel).

(iii) By construction, the entries of $\mathbf{N}_i$ are exactly the row sums of $\mathbf{A}_i$ and of $\mathbf{B}_i$. That is, for each fixed h_i, summing $\nu_{n_i}(r_i, h_i)$ over r_i recovers $n_i(n_i - |h_i|)$ admissible index pairs for direction i, and correspondingly n_i^2 when $h_i = 0$, which is the constant row sum of $\mathbf{D}_i$. Hence, the normalized count matrices $\mathbf{A}_i^*$, $\mathbf{B}_i^*$ and $\mathbf{D}_i^*$ from (2.29) are row-stochastic. That is, we have

$$\mathbf{A}_i^* \underline{\mathbb{1}}_i = \mathbf{B}_i^* \underline{\mathbb{1}}_i = \mathbf{D}_i^* \underline{\mathbb{1}}_i = \underline{\mathbb{1}}_i, \qquad \text{for} \quad i \in \{1,2\},$$

where $\underline{\mathbb{1}}_i = (1, 1, \ldots, 1)^\top$ denotes the $(L_i \times 1)$ vector of ones.

We illustrate further the introduced quantities and the derived formulas with a toy example for the case of a Gaussian spatial lattice process that was used already in Example 2.1.

Example 2.2 (Joint bias formulas for spatial lattice process). Let $\mathbf{X}$ be the Gaussian spatial lattice process from Example 2.1 with exponential covariance function $C_\lambda(\underline{h})$ defined in (2.10) for some $\lambda \in \{2, 4, 6, 8\}$ and with $\beta \in [0, 1]$. Further, suppose $n_1 = n_2 = 4$, that is, $\mathbf{X}_{\underline{n}}$ is a (4×4) data matrix and $\mathbf{\Gamma}$ defined in (2.26) becomes a (7×7) matrix of covariances. Figure 3 illustrates different instances of $\mathbf{\Gamma}$, where the panels show varying levels of spatial dependence, controlled by the parameter λ, for both separable ($\beta = 0$) and non-separable ($\beta = 1$) settings. For this setup, the (7×7) raw counting matrices $\mathbf{A}_1 = \mathbf{A}_2$, $\mathbf{B}_1 = \mathbf{B}_2$ and $\mathbf{D}_1 = \mathbf{D}_2$ are shown in Figure 2. Moreover, in Figure 6 (left panel), we show a heatmap of the corresponding (49×49) weight matrix $\mathbf{W}^*_{\underline{n}}$.

2.5. Joint bias of a matrix of the constant prefactor sample covariances

The characterization of the joint mean and bias for the standard covariance estimator $\widehat{C}(\underline{h})$ in Theorem 2.2 can be easily transferred also to the case of the constant prefactor estimator from $\widetilde{C}(\underline{h})$. This is because the expectations of the two estimators are linked by a lag dependent scaling $\mathbb{E}(\widetilde{C}(\underline{h})) = N_h/N\mathbb{E}(\widehat{C}(\underline{h}))$, see (A.24). This property can be used to define diagonal $(L_i \times L_i)$ matrices $\mathbf{\Delta}_i$, $i \in \{1,2\}$ that encode the relationship jointly. We define

$$\mathbf{\Delta}_i := \operatorname{diag}\left(1 - \frac{|p - n_i|}{n_i},\ p = 1, \ldots, L_i\right) = \operatorname{diag}\left(1 - \frac{|h_i|}{n_i},\ h_i \in I(1 - n_i : n_i - 1)\right). \tag{2.35}$$

That is, these matrices essentially collect the Bartlett windows (Bartlett, 1950) on their main diagonal. Hence, we refer to $\mathbf{\Delta}_i$, $i \in \{1,2\}$ as *Bartlett window matrices*. Since the *total* and *effective* sample size factorizes across the two lattice dimensions, i.e., $N = n_1 n_2$ and $N_h = (n_1 - |h_1|)(n_2 - |h_2|)$, their ratio factorizes as well. That is, for $(h_1, h_2) = (p - n_1, q - n_2)$, we have

$$\frac{N_h}{N} = \frac{n_1 - |h_1|}{n_1} \cdot \frac{n_2 - |h_2|}{n_2} = [\mathbf{\Delta}_1]_{p,p} \cdot [\mathbf{\Delta}_2]_{q,q},$$

such that we can exploit the above together with $\mathbb{E}(\widetilde{C}(\underline{h})) = N_h/N\mathbb{E}(\widehat{C}(\underline{h}))$ to connect the two estimators (for the whole covariance matrix) in the following corollary.

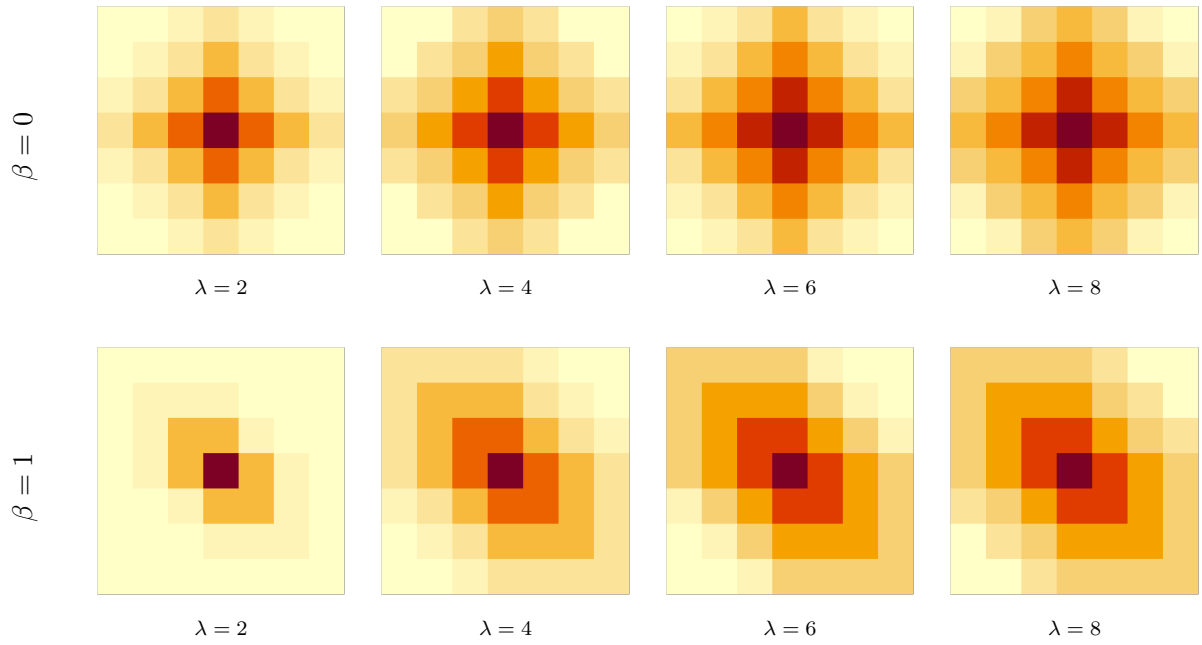


Figure 3: Covariance matrices $\mathbf{\Gamma}$ from (2.26) of size (7×7) for $\mathbf{X}_{\underline{n}}$ with $n_1 = n_2 = 4$ from a spatial (Gaussian) process $\mathbf{X}$ with exponential covariance function C_λ defined in (2.10). $\mathbf{\Gamma}$ is shown for separable ($\beta = 0$; top row) and non-separable ($\beta = 1$; bottom row) dependence with increasing spatial dependence $\lambda \in \{2, 4, 6, 8\}$.

Corollary 2.2 (Joint mean and bias of $\mathrm{vec}(\widetilde{\mathbf{\Gamma}})$)**.** *Let $\widetilde{\mathbf{\Gamma}}$ be the matrix of constant prefactor sample covariances that is defined as $\widehat{\mathbf{\Gamma}}$ below* (2.26)*, but uses $\widetilde{C}$ from* (2.12) *instead of $\widehat{C}$. Then, we have*

$$\mathbb{E}\big(\widetilde{\mathbf{\Gamma}}\big) = \mathbf{\Delta}_1 \mathbb{E}\big(\widehat{\mathbf{\Gamma}}\big) \mathbf{\Delta}_2, \qquad \mathbb{E}\big(\mathrm{vec}(\widetilde{\mathbf{\Gamma}})\big) = \widetilde{\mathbf{W}}^*_{\underline{n}} \mathrm{vec}(\mathbf{\Gamma}). \tag{2.36}$$

with joint bias vector

$$b_{\mathrm{vec}(\widetilde{\mathbf{\Gamma}})} := \mathbb{E}\big(\mathrm{vec}(\widetilde{\mathbf{\Gamma}})\big) - \mathrm{vec}(\mathbf{\Gamma}) = \big(\widetilde{\mathbf{W}}^*_{\underline{n}} - \mathbf{I}_{L_1 L_2}\big)\mathrm{vec}(\mathbf{\Gamma}). \tag{2.37}$$

Here, the $(L_{\underline{n}} \times L_{\underline{n}})$ constant-prefactor weight matrix $\widetilde{\mathbf{W}}^*_{\underline{n}}$ can be directly linked to $\widehat{\mathbf{W}}^*_{\underline{n}}$ in (2.32) *by using the Bartlett window matrices $\mathbf{\Delta}_i$ defined in* (2.35)*. Precisely, we have*

$$\begin{aligned} \widetilde{\mathbf{W}}^*_{\underline{n}} &:= (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1)\, \mathbf{W}^*_{\underline{n}} \\ &= (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1) - (\mathbf{\Delta}_2 \mathbf{A}^*_2) \otimes (\mathbf{\Delta}_1 \mathbf{A}^*_1) - (\mathbf{\Delta}_2 \mathbf{B}^*_2) \otimes (\mathbf{\Delta}_1 \mathbf{B}^*_1) + (\mathbf{\Delta}_2 \mathbf{D}^*_2) \otimes (\mathbf{\Delta}_1 \mathbf{D}^*_1), \end{aligned} \tag{2.38}$$

where $\mathbf{\Delta}_i \mathbf{A}^*_i = n_i^{-2} \mathbf{A}_i$ and $\mathbf{\Delta}_i \mathbf{B}^*_i = n_i^{-2} \mathbf{B}_i$, while $\mathbf{\Delta}_i \mathbf{D}^*_i = \mathrm{diag}\big((n_i - |h_i|)/n_i^3\big)\mathbf{D}_i$ and $\widetilde{\mathbf{W}}^*_{\underline{n}}$ inherits the same rank deficiency as $\mathbf{W}^*_{\underline{n}}$, i.e., $\mathrm{rank}(\widetilde{\mathbf{W}}^*_{\underline{n}}) \leq L_{\underline{n}} - 1$.

2.6. Exact finite sample second-order structure of the sample covariances

The bias analysis in the previous section was distribution free as it relied only on the weak stationarity Assumption 2.1. In this section, we present also the second-order structure of single covariance estimators $\widehat{C}(\underline{h})$ and $\widetilde{C}(\underline{h})$. For this purpose, some information on the (finite) fourth-order moments is required to describe the variability of the estimators. In the following, we impose Gaussianity to facilitate the derivation and the presentation of the resulting exact formulas for the covariances of sample covariance estimators.

Assumption 2.3 (Gaussianity)**.** *The spatial lattice process* $\mathbf{X}$ *is Gaussian.*

Focusing first on the standard estimator $\widehat{C}(\underline{h})$, Isserlis' theorem (Isserlis, 1918) applied to the centered random variables of $\mathbf{X}$ leads to the exact formula presented in the following theorem.

Theorem 2.3 (Exact finite-sample covariances of the standard covariance estimator)**.** *Let Assumptions 2.1 and 2.3 hold and let* $\underline{h} = (h_1, h_2)$ *and* $\underline{g} = (g_1, g_2)$ *be two lag vectors in* $H_{\underline{n}}$ (2.5) *with effective sample sizes* $N_h = (n_1 - |h_1|)(n_2 - |h_2|)$ *and* $N_g = (n_1 - |g_1|)(n_2 - |g_2|)$*, respectively. Then, using the centered form representation based on* C_Z *defined in* (2.15)*, the covariances of two standard sample covariance estimators* $\widehat{C}(\underline{h})$ *and* $\widehat{C}(\underline{g})$ *defined in* (2.11) *can be represented as*

$$\operatorname{Cov}\big(\widehat{C}(\underline{h}), \widehat{C}(\underline{g})\big) = \frac{1}{N_h N_g} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{s} \in G_{\underline{n}}(\underline{g})} \big[C_Z(\underline{t} + \underline{h}, \underline{s} + \underline{g})\, C_Z(\underline{t}, \underline{s}) + C_Z(\underline{t} + \underline{h}, \underline{s})\, C_Z(\underline{t}, \underline{s} + \underline{g})\big]\,. \tag{2.39}$$

Alternatively, the lag form representation with $\overline{C}$ *and* D *from* (2.16) *and* (2.17) *and the shift* $\underline{\delta} := \underline{s} - \underline{t}$ *becomes*

$$\begin{aligned} &\operatorname{Cov}\big(\widehat{C}(\underline{h}), \widehat{C}(\underline{g})\big) \qquad (2.40)\\ =&\frac{1}{N_{\underline{h}} N_{\underline{g}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{s} \in G_{\underline{n}}(\underline{g})} \Big\{ \big[C(\underline{\delta} + \underline{g} - \underline{h}) - \overline{C}(\underline{t} + \underline{h}) - \overline{C}(\underline{s} + \underline{g}) + D\big] \big[C(\underline{\delta}) - \overline{C}(\underline{t}) - \overline{C}(\underline{s}) + D\big] \\ &\quad + \big[C(\underline{\delta} - \underline{h}) - \overline{C}(\underline{t} + \underline{h}) - \overline{C}(\underline{s}) + D\big] \big[C(\underline{\delta} + \underline{g}) - \overline{C}(\underline{t}) - \overline{C}(\underline{s} + \underline{g}) + D\big] \Big\}. \end{aligned}$$

Notably, the right-hand sides of (2.39) *and* (2.40) *does not depend on the mean* μ*.*

A remark on Theorem 2.3 and the interpretation of the exact representation of the (finite sample) covariance of two standard sample covariance estimators is in order.

Remark 2.3 (On the representation of $\operatorname{Cov}(\widehat{C}(\underline{h}), \widehat{C}(\underline{g}))$ in (2.39)).

(i) If the true mean μ is known and replaces the sample mean $\overline{X}$ in the standard estimators $\widehat{C}(\underline{h})$ (leading to $\widehat{C}_\mu(\underline{h})$) in (2.39), $C_Z(\underline{t}, \underline{s})$ collapses to $C(\underline{t} - \underline{s})$ and (2.40) becomes

$$\frac{1}{N_{\underline{h}} N_{\underline{g}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{s} \in G_{\underline{n}}(\underline{g})} \big[C(\underline{\delta} + \underline{g} - \underline{h}) C(\underline{\delta}) + C(\underline{\delta} - \underline{h}) C(\underline{\delta} + \underline{g})\big]\,, \tag{2.41}$$

which is the two-dimensional lattice version of the Bartlett formula (Bartlett, 1946), which gives the covariance of sample covariances in the univariate time series setting. That is, all other terms in (2.39) arise from estimating μ by $\overline{X}$ in $\widehat{C}(\underline{h})$ as defined in (2.11).

Again, using the simple link between the two sample covariances $\widetilde{C}(\underline{h}) = N_{\underline{h}}/N\, \widehat{C}(\underline{h})$, we can state the second-order structure of the constant prefactor estimator in the following corollary.

Corollary 2.3 (Exact finite sample covariance of the constant estimator)**.** *Under the conditions of Theorem 2.3, and with* $\widetilde{C}$ *from* (2.12)*, we have*

$$\begin{aligned} \operatorname{Cov}\big(\widetilde{C}(\underline{h}), \widetilde{C}(\underline{g})\big) &= \frac{N_{\underline{h}} N_{\underline{g}}}{N^2} \operatorname{Cov}\big(\widehat{C}(\underline{h}), \widehat{C}(\underline{g})\big) \\ &= \frac{1}{N^2} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{s} \in G_{\underline{n}}(\underline{g})} \Big[C_Z(\underline{t} + \underline{h}, \underline{s} + \underline{g}) C_Z(\underline{t}, \underline{s}) + C_Z(\underline{t} + \underline{h}, \underline{s}) C_Z(\underline{t}, \underline{s} + \underline{g})\Big]. \end{aligned} \tag{2.42}$$

3. Bias-corrected covariance estimation

In this section, we use the linear bias representation derived in the previous Section 2 to construct a framework for bias-corrected estimation of covariances from stationary spatial lattice data. Indeed, Equation (2.31) suggests a natural route for bias correction. If the deterministic weight matrix $\mathbf{W}^*_{\underline{n}}$ *were* invertible, we could obtain an unbiased estimator for the sample covariances by left-multiplying $\mathrm{vec}(\widehat{\mathbf{\Gamma}})$ with $(\mathbf{W}^*_{\underline{n}})^{-1}$.

However, according to Proposition 2.2, the deterministic weight matrix $\mathbf{W}^*_{\underline{n}}$ from (2.32) is singular. Hence, since the inverse of the full weight matrix does not exist, we cannot recover such a correction directly. To remedy this issue, we propose an adaptation of the methods proposed by Vogelsang and Yang (2016) for univariate time series and extended by Li (2023) to multivariate time series to the spatial lattice case. That is, instead of using the whole sample covariance matrix $\widehat{\mathbf{\Gamma}}$, we consider suitable lower-dimensional (central) submatrices, denoted by $\widehat{\mathbf{\Gamma}}_{\underline{m}}$, of $\widehat{\mathbf{\Gamma}}$. As this also reduces the dimension of the corresponding weight matrix, it becomes invertible.

3.1. Bias correction for submatrices of the matrix of sample covariances

Recall that $L_i = 2n_i - 1$ for $i \in \{1, 2\}$, and $L_{\underline{n}} = L_1 L_2$. We define $\underline{m} = (m_1, m_2)$ to be a vector of *cutoff parameters* with $m_i \leq n_i - 2$ and set $M_i = (2m_i + 1)$, for $i \in \{1, 2\}$, and $M_{\underline{m}} := M_1 M_2$. In the following, we split the full lag set $H_{\underline{n}}$ from (2.5) into the set of *selected* (or *within-window*) lags

$$W_{\underline{m}} := \{\underline{h} \in \mathbb{Z}^2 : |h_i| \leq m_i,\ i \in \{1, 2\}\} \subset H_{\underline{n}}, \quad |W_{\underline{m}}| = M_{\underline{m}}$$

and the set of the *non-selected* (or *out-of-window*) lags

$$W^c_{\underline{m}} := H_{\underline{n}} \setminus W_{\underline{m}}, \qquad |W^c_{\underline{m}}| = L_{\underline{n}} - M_{\underline{m}}.$$

For $i \in \{1, 2\}$, let the $\mathbf{P}_{i,m_i} \in \mathbb{R}^{M_i \times L_i}$ be the matrix that retains the M_i central rows of $\mathbf{I}_{L_i}$. We call them *partition matrices* and their entries become

$$[\mathbf{P}_{i,m_i}]_{p,q} = \mathbb{1}\{q = n_i - m_i + p - 1\}, \qquad p = 1, \ldots, M_i, \quad q = 1, \ldots, L_i, \tag{3.1}$$

where $\mathbb{1}(\cdot)$ is the indicator function. In this way, the p-th row of $\mathbf{P}_{i,m_i}$ selects the lag $h_i = p - m_i - 1$ and the partition matrices have a single one in each row and no repeated column. Analogously, let $\mathbf{P}^c_{i,m_i} \in \mathbb{R}^{(L_i - M_i) \times L_i}$ be the complementary matrices that retain the remaining $L_i - M_i$ rows of $\mathbf{I}_{L_i}$ in increasing order, that is,

$$[\mathbf{P}^c_{i,m_i}]_{p,q} = \begin{cases} 1, & q = p, \qquad\qquad p = 1, \ldots, n_i - m_i - 1, \\ 1, & q = p + M_i, \qquad p = n_i - m_i, \ldots, K_i, \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

By construction, $\mathbf{P}_{i,m_i}$ and $\mathbf{P}^c_{i,m_i}$ together consist of all rows of $\mathbf{I}_{L_i}$. We apply (3.1) on both sides of $\mathbf{\Gamma}$ and $\widehat{\mathbf{\Gamma}}$ defined in (2.26) to extract an $(M_1 \times M_2)$ principal submatrix centered around the lag $(0, 0)$ that contains all the covariances with lags $\underline{h} \in W_{\underline{m}}$. That is, we define

$$\widehat{\mathbf{\Gamma}}_{\underline{m}} := \mathbf{P}_{1,m_1}\, \widehat{\mathbf{\Gamma}}\, \mathbf{P}^\top_{2,m_2}, \quad \mathbf{\Gamma}_{\underline{m}} := \mathbf{P}_{1,m_1}\, \mathbf{\Gamma}\, \mathbf{P}^\top_{2,m_2}. \tag{3.3}$$

In Figure 4, we give examples of $\widehat{\mathbf{\Gamma}}_{\underline{m}}$ for $n_1, n_2 = 4$ and three choices of cutoff parameters $m_1 = m_2 = 0$, $m_1 = m_2 = 1$ and $m_1 = m_2 = 2$.

For the vectorized versions of (3.3), we exploit the vectorization property of the Kronecker product and combine the partition matrices into a single *selection matrix* $\mathbf{S}_{\underline{m}}$ and its complement $\mathbf{S}^c_{\underline{m}}$. Precisely, we define

$$\mathbf{S}_{\underline{m}} := (\mathbf{P}_{2,m_2} \otimes \mathbf{P}_{1,m_1}) \in \mathbb{R}^{M_{\underline{m}} \times L_{\underline{n}}}, \quad \mathbf{S}^c_{\underline{m}} := \begin{pmatrix} \mathbf{I}_{L_2} \otimes \mathbf{P}^c_{1,m_1} \\ \mathbf{P}^c_{2,m_2} \otimes \mathbf{P}_{1,m_1} \end{pmatrix} \in \mathbb{R}^{(L_{\underline{n}} - M_{\underline{m}}) \times L_{\underline{n}}}. \tag{3.4}$$

where the two blocks of $\mathbf{S}^c_{\underline{m}}$ in (3.4) correspond to the disjoint decomposition of $W^c_{\underline{m}}$ into the lags with $|h_1| > m_1$, for which h_2 is unrestricted, and those with $|h_1| \leq m_1$ and $|h_2| > m_2$. Since, by construction, $\mathbf{S}_{\underline{m}}$ and $\mathbf{S}^c_{\underline{m}}$ consist of distinct rows of $\mathbf{I}_{L_i}$, exhausting them together, they have orthonormal rows and mutually orthogonal row spaces. That is, we have

$$\mathbf{S}_{\underline{m}}\mathbf{S}_{\underline{m}}^\top = \mathbf{I}_{M_{\underline{m}}}, \qquad \mathbf{S}_{\underline{m}}\mathbf{S}_{\underline{m}}^{c\top} = \mathbf{0}, \qquad \mathbf{S}_{\underline{m}}^\top\mathbf{S}_{\underline{m}} + \mathbf{S}_{\underline{m}}^{c\top}\mathbf{S}^c_{\underline{m}} = \mathbf{I}_{L_{\underline{n}}}. \tag{3.5}$$

Hence, pre-multiplying the selection matrix $\mathbf{S}_{\underline{m}}$ and its complementary $\mathbf{S}^c_{\underline{m}}$ from (3.4) to the $\text{vec}(\mathbf{\Gamma})$ and $\text{vec}(\widehat{\mathbf{\Gamma}})$ yields

$$\begin{pmatrix} \underline{C}_{\underline{m}} \\ \underline{C}^c_{\underline{m}} \end{pmatrix} := \begin{pmatrix} \mathbf{S}_{\underline{m}} \\ \mathbf{S}^c_{\underline{m}} \end{pmatrix} \text{vec}(\mathbf{\Gamma}), \quad \begin{pmatrix} \widehat{\underline{C}}_{\underline{m}} \\ \widehat{\underline{C}}^c_{\underline{m}} \end{pmatrix} := \begin{pmatrix} \mathbf{S}_{\underline{m}} \\ \mathbf{S}^c_{\underline{m}} \end{pmatrix} \text{vec}(\widehat{\mathbf{\Gamma}}), \tag{3.6}$$

where $\underline{C}_{\underline{m}} \in \mathbb{R}^{M_{\underline{m}}}$ contains the population covariances with lags in $W_{\underline{m}}$ and $\underline{C}^c_{\underline{m}} \in \mathbb{R}^{(L_{\underline{n}} - M_{\underline{m}})}$ contains those with lags in $W^c_{\underline{m}}$. By the last identity in (3.5), we can finally decompose the $L_{\underline{n}}$-dimensional vector of population covariances into the vector of selected and non-selected covariances leading to

$$\text{vec}(\mathbf{\Gamma}) = \mathbf{S}_{\underline{m}}^\top\mathbf{S}_{\underline{m}}\text{vec}(\mathbf{\Gamma}) + \mathbf{S}_{\underline{m}}^{c\top}\mathbf{S}^c_{\underline{m}}\text{vec}(\mathbf{\Gamma}). \tag{3.7}$$

Now, using (2.31), we examine the expectation of the vectorized vector of partitioned sample covariances $\text{vec}(\widehat{\mathbf{\Gamma}})$. By plugging in the decomposition of $\text{vec}(\mathbf{\Gamma})$ (3.7) and using the notation in (3.6), we get

$$\begin{aligned} \mathbb{E}[\text{vec}(\widehat{\mathbf{\Gamma}}_{\underline{m}})] = \mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\,\text{vec}(\mathbf{\Gamma}) &= \mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top\mathbf{S}_{\underline{m}}\text{vec}(\mathbf{\Gamma}) + \mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^{c\top}\mathbf{S}^c_{\underline{m}}\text{vec}(\mathbf{\Gamma}) \\ &= \mathbf{W}^*_{\underline{n},\underline{m}}\,\underline{C}_{\underline{m}} + \mathbf{W}^{*,c}_{\underline{n},\underline{m}}\,\underline{C}^c_{\underline{m}}, \end{aligned} \tag{3.8}$$

where the two factors in (3.8) are the *reduced weight matrix*, namely a central submatrix of the weight matrix in (2.32), and the rectangular *leakage matrix*

$$\mathbf{W}^*_{\underline{n},\underline{m}} := \mathbf{S}_{\underline{m}}\,\mathbf{W}^*_{\underline{n}}\,\mathbf{S}_{\underline{m}}^\top \in \mathbb{R}^{M_{\underline{m}} \times M_{\underline{m}}}, \quad \mathbf{W}^{*,c}_{\underline{n},\underline{m}} := \mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^{c\top} \in \mathbb{R}^{M_{\underline{m}} \times (L_{\underline{n}} - M_{\underline{m}})} \tag{3.9}$$

The reduced weight matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ inherits the Kronecker structure of $\mathbf{W}^*_{\underline{n}}$, such that, for $i \in \{1, 2\}$, we have

$$\mathbf{A}^*_{i,m_i} := \mathbf{P}_{i,m_i}\mathbf{A}^*_i\mathbf{P}_{i,m_i}^\top, \quad \mathbf{B}^*_{i,m_i} := \mathbf{P}_{i,m_i}\mathbf{B}^*_i\mathbf{P}_{i,m_i}^\top, \quad \mathbf{D}^*_{i,m_i} := \mathbf{P}_{i,m_i}\mathbf{D}^*_i\mathbf{P}_{i,m_i}^\top,$$

which are the central $(M_{\underline{m}} \times M_{\underline{m}})$ blocks of the normalized counting matrices in (2.28) and the reduced weight matrix is given by

$$\mathbf{W}^*_{\underline{n},\underline{m}} = \mathbf{I}_{M_{\underline{m}}} - \mathbf{A}^*_{2,m_2} \otimes \mathbf{A}^*_{1,m_1} - \mathbf{B}^*_{2,m_2} \otimes \mathbf{B}^*_{1,m_1} + \mathbf{D}^*_{2,m_2} \otimes \mathbf{D}^*_{1,m_1}, \tag{3.10}$$

which becomes invertible for $m_i \leq n_i - 2$ small enough. Finally, for estimating $\underline{C}_{\underline{m}}$, we define the *bias-corrected vector of (central) sample covariance estimators* $\widehat{\underline{C}}^*_{\underline{m}}$ by solving the reduced system in the first term of the last right-hand side of (3.8). That is, we define

$$\widehat{\underline{C}}^*_{\underline{m}} = \left(\mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top\right)^{-1}\widehat{\underline{C}}_{\underline{m}} = \left(\mathbf{W}^*_{\underline{n},\underline{m}}\right)^{-1}\widehat{\underline{C}}_{\underline{m}}. \tag{3.11}$$

Since (3.11) inverts only the first of the two terms in (3.8), the estimator $\widehat{\underline{C}}^*_{\underline{m}}$ is not exactly unbiased unless the second term vanishes. We formalize this in the following theorem.

$h_1 \backslash h_2$	-3	-2	-1	0	1	2	3
-3	$\widehat{C}(-3,-3)$	$\widehat{C}(-3,-2)$	$\widehat{C}(-3,-1)$	$\widehat{C}(-3,0)$	$\widehat{C}(-3,1)$	$\widehat{C}(-3,2)$	$\widehat{C}(-3,3)$
-2	$\widehat{C}(-2,-3)$	$\widehat{C}(-2,-2)$	$\widehat{C}(-2,-1)$	$\widehat{C}(-2,0)$	$\widehat{C}(-2,1)$	$\widehat{C}(-2,2)$	$\widehat{C}(-2,3)$
-1	$\widehat{C}(-1,-3)$	$\widehat{C}(-1,-2)$	$\widehat{C}(-1,-1)$	$\widehat{C}(-1,0)$	$\widehat{C}(-1,1)$	$\widehat{C}(-1,2)$	$\widehat{C}(-1,3)$
0	$\widehat{C}(0,-3)$	$\widehat{C}(0,-2)$	$\widehat{C}(0,-1)$	$\widehat{C}(0,0)$	$\widehat{C}(0,1)$	$\widehat{C}(0,2)$	$\widehat{C}(0,3)$
1	$\widehat{C}(1,-3)$	$\widehat{C}(1,-2)$	$\widehat{C}(1,-1)$	$\widehat{C}(1,0)$	$\widehat{C}(1,1)$	$\widehat{C}(1,2)$	$\widehat{C}(1,3)$
2	$\widehat{C}(2,-3)$	$\widehat{C}(2,-2)$	$\widehat{C}(2,-1)$	$\widehat{C}(2,0)$	$\widehat{C}(2,1)$	$\widehat{C}(2,2)$	$\widehat{C}(2,3)$
3	$\widehat{C}(3,-3)$	$\widehat{C}(3,-2)$	$\widehat{C}(3,-1)$	$\widehat{C}(3,0)$	$\widehat{C}(3,1)$	$\widehat{C}(3,2)$	$\widehat{C}(3,3)$
			$m=0$	$m=1$	$m=2$		

Figure 4: For $n_1 = n_2 = 4$, the $(L_1 \times L_2) = (7 \times 7)$ sample covariance matrix $\widehat{\mathbf{\Gamma}}$ is shown, where extracted submatrices $\widehat{\mathbf{\Gamma}}_{\underline{m}}$ are indicated for $m_1 = m_2 = 0$ (green) $m_1 = m_2 = 1$ (green and red) and $m_1 = m_2 = 2$ (green, red and blue), respectively.

Theorem 3.1 (Finite sample bias of $\widehat{\underline{C}}^*_{\underline{m}}$)**.** *Let* $\underline{m} = (m_1, m_2)$ *with* $m_i \leq n_i - 2$, $i \in \{1, 2\}$, *and suppose* $\mathbf{W}_{\underline{n},\underline{m}}$ *is invertible. Then, the bias vector of the bias-corrected estimator* $\widehat{\underline{C}}^*_{\underline{m}}$ *defined in* (3.11) *becomes*

$$\underline{b}^*_{\underline{m}} := \mathbb{E}\big(\widehat{\underline{C}}^*_{\underline{m}}\big) - \underline{C}_{\underline{m}} = \big(\mathbf{W}^*_{\underline{n},\underline{m}}\big)^{-1} \mathbf{W}^{*,c}_{\underline{n},\underline{m}}\, \underline{C}^c_{\underline{m}}. \tag{3.12}$$

Moreover, if $M_{\underline{m}}/N < 1$, *then*

$$\big\|\underline{b}^*_{\underline{m}}\big\|_\infty \leq \frac{1}{N - M_{\underline{m}}} \big\|\underline{C}^c_{\underline{m}}\big\|_1. \tag{3.13}$$

A few remarks on the result of Theorem 3.1 are in order.

Remark 3.1 (On the bias correction).

(i) The bias correction thus removes the contribution to the bias of the selected covariances exactly, whatever the covariance function is, while the leakage bias is pre-multiplied by the inverse of the reduced weight matrix $(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}$.

(ii) The bias-corrected estimator is exactly unbiased if and only if $\mathbf{W}^{*,c}_{\underline{n},\underline{m}} \underline{C}^c_{\underline{m}} = \underline{0}$. This holds whenever $\underline{C}^c_{\underline{m}} = \underline{0}$, that is, whenever the field $\mathbf{X}$ is $\underline{m}$-dependent, so that $C(\underline{h}) = 0$ for every $\underline{h} \notin W_{\underline{m}}$.

(iii) Both factors in (3.8) are governed by the deviation of $\mathbf{W}^*_{\underline{n}}$ from the identity. Indeed, by the first two identities in (3.5), the identity part of $\mathbf{W}^*_{\underline{n}}$ contributes exactly $\mathbf{I}_{M_{\underline{m}}}$ to $\mathbf{W}^*_{\underline{n},\underline{m}}$ and nothing at all to $\mathbf{W}^{*,c}_{\underline{n},\underline{m}}$, so that

$$\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}} = \mathbf{S}_{\underline{m}} \big(\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\big) \mathbf{S}^\top_{\underline{m}}, \qquad \mathbf{W}^{*,c}_{\underline{n},\underline{m}} = \mathbf{S}_{\underline{m}} \big(\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\big) \mathbf{S}^{c\top}_{\underline{m}}. \tag{3.14}$$

Both are therefore built from the counting matrices generated by the estimation of the mean. In particular, if μ were known and used in place of $\overline{X}$ (leading to $\widehat{C}_\mu(\underline{h})$), we would have $\mathbf{W}^*_{\underline{n}} = \mathbf{I}_{L_{\underline{n}}}$, both matrices in (3.14) would vanish, and $\widehat{\underline{C}}^*_{\underline{m}}$ would coincide with $\widehat{\underline{C}}_{\underline{m}}$ and be unbiased.

To illustrate the bias correction approach and the quantities introduced above, we conclude this section with a continuation of the toy example setup considered already in Examples 2.1 and 2.2.

$\mathbf{P}_{1,m} = \mathbf{P}_{2,m}$ **for** $m = 0$

$p\backslash h$	-3	-2	-1	0	1	2	3
1	0	0	0	1	0	0	0

$\mathbf{P}_{1,m} = \mathbf{P}_{2,m}$ **for** $m = 1$

$p\backslash h$	-3	-2	-1	0	1	2	3
1	0	0	1	0	0	0	0
2	0	0	0	1	0	0	0
3	0	0	0	0	1	0	0

$\mathbf{P}_{1,m} = \mathbf{P}_{2,m}$ **for** $m = 2$

$p\backslash h$	-3	-2	-1	0	1	2	3
1	0	1	0	0	0	0	0
2	0	0	1	0	0	0	0
3	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0
5	0	0	0	0	0	1	0

Figure 5: The selection matrices $\mathbf{P}_{1,m} = \mathbf{P}_{2,m}$ for a grid of size $n_1 = n_2 = 4$ ($L_1 = L_2 = 7$). Left: $m_1 = m_2 = 0$, selecting only lag $h_i = 0$. Middle: $m_1 = m_2 = 1$, selecting lags $|h_i| \leq 1$. Right: $m_1 = m_2 = 2$, selecting lags $|h_i| \leq 2$.

Example 3.1 (Bias correction for spatial lattice process). As in the setup of Examples 2.1 and 2.2, suppose again $n_1 = n_2 = 4$. That is, we observe the (4×4) data matrix $\mathbf{X}_{\underline{n}}$ such that $\mathbf{\Gamma}$ defined becomes a (7×7) matrix of covariances. For this setup, Figure 4 shows the corresponding (7×7) sample covariance matrix $\widehat{\mathbf{\Gamma}}$. Additionally, Figure 4 indicates the submatrices $\widehat{\mathbf{\Gamma}}_{\underline{m}}$ for the choices $m_1 = m_2 = 0$ (green), $m_1 = m_2 = 1$ (green and red) and $m_1 = m_2 = 2$ (green, red and blue) that are obtained by applying the selection matrices shown in Figure 5. In these cases, the green-and-red-shaded entries ($m_1 = m_2 = 1$) determine a (3×3) submatrix centered around the zero lag $(0, 0)$, capturing all lag combinations (h_1, h_2) where $|h_1| \leq m_1 = 1$ and $|h_2| \leq m_2 = 1$, while the green-and-red-and-blue-shaded region ($m_1 = m_2 = 2$) represents the corresponding (5×5) submatrix, which includes the lag combinations where $|h_1| \leq m_1 = 2$ and $|h_2| \leq m_2 = 2$. Finally, Figure 6 (right panel) shows a heatmap of the (25×25) inverted weight matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ for the case $m_1 = m_2 = 2$.

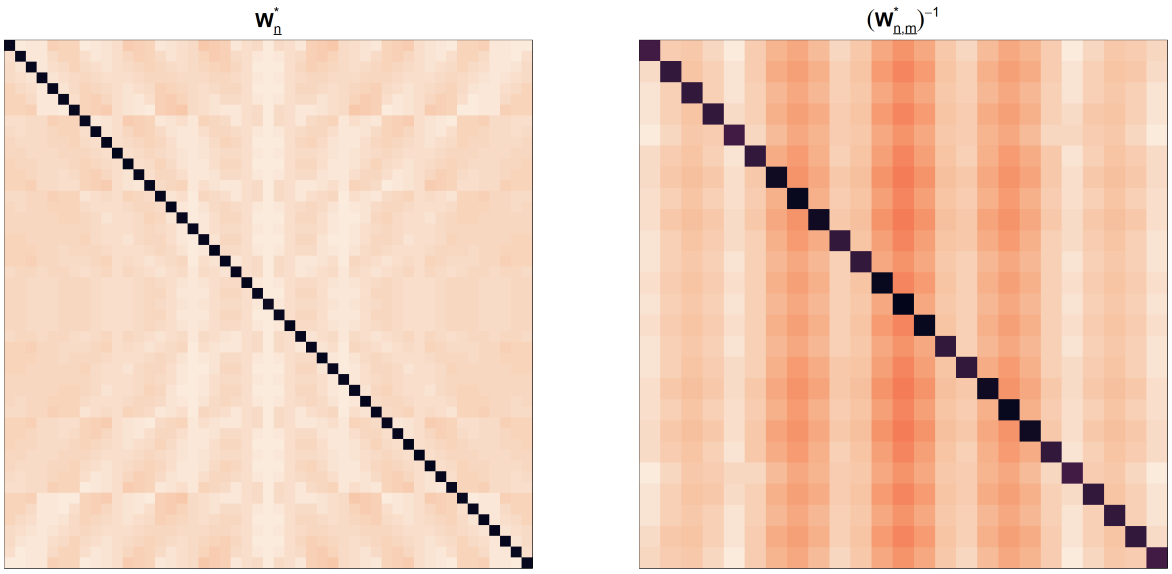


Figure 6: Heatmaps of the (singular) (49×49) weight matrix $\mathbf{W}^*_{\underline{n}}$ (left) and the inverse of its central (25×25) submatrix $\mathbf{W}^*_{\underline{n},m}$, i.e., $(\mathbf{W}^*_{\underline{n},m})^{-1}$, (right) for a grid of size $n_1 = n_2 = 4$ and cutoff parameters $m_1 = m_2 = 2$.

3.2. On the invertibility of the weight matrix

In the following lemma, we give sufficient conditions on the cutoff parameters $\underline{m} = (m_1, m_2)$ that guarantee invertibility of the resulting weight matrix $\mathbf{W}^*_{\underline{n},m}$.

Lemma 3.1 (Invertibility and bound of the inverse weight matrix $\mathbf{W}_{\underline{n},\underline{m}}$)**.** *Let $\underline{n} = (n_1, n_2)$ be the sizes of the observed lattice $\mathbf{X}_{\underline{n}}$ and $\underline{m} = (m_1, m_2)$ are the cutoff parameters leading to the weight matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ from* (3.10)*, with $L_{\underline{n}} = \prod_{i=1}^2 (2n_i - 1)$ and $M_{\underline{m}} = \prod_{i=1}^2 (2m_i + 1)$. Then, we have*

$$\|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\|_\infty \leq \frac{M_{\underline{m}}}{N} \quad \textit{and} \quad \|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\|_1 \leq \frac{M_{\underline{m}}}{N}. \tag{3.15}$$

Consequently, whenever $M_{\underline{m}}/N < 1$, the matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ is nonsingular with

$$\left\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\right\|_\infty \leq \frac{N}{N - M_{\underline{m}}} \quad \textit{and} \quad \left\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1} - \mathbf{I}_{M_{\underline{m}}}\right\|_\infty \leq \frac{M_{\underline{m}}}{N - M_{\underline{m}}}. \tag{3.16}$$

A few remarks on the condition $M_{\underline{m}}/N < 1$ of Lemma 3.1 are in order.

Remark 3.2 (On the invertibility condition).

(i) The condition is stricter than $m_i \leq n_i - 2$ for $i \in \{1, 2\}$. In fact, since it is based on a product of the two cutoff parameters, the two axes trade against each other. For axis 1, we have that that

$$m_1 < \frac{1}{2}\left(\frac{n_1 n_2}{2m_2 + 1} - 1\right),$$

so that a larger cutoff on one axis must be compensated by a smaller one on the other axis. For a square lattice with $n_1 = n_2 = n$ and $m_1 = m_2 = m$ it reduces to $2m + 1 < n$, that is $m \leq \lfloor (n-1)/2 \rfloor$, roughly half of the admissible range $m \leq n - 2$. For example for a 20×20 grid, we would only accept cutoff values $m \leq 9$.

(ii) The condition is sufficient but not necessary. We have evaluated numerically lattices with dimensions $3 \leq n_i \leq 40$ and their respective cutoffs with the more relaxed condition $m_i \leq n_i - 2$ and we see that the reduced weight matrix turned out to be nonsingular in all these cases. The spectral radius of $\mathbf{I}_{M_{\underline{m}}} - \mathbf{W}^*_{\underline{n},\underline{m}}$ increases with the lattice size but stays below 1, so that the Neumann series still converges and remains invertible throughout the admissible range, by Lemma 2.3.3 from Golub and Van Loan (2013). Hence, we conjecture that $\mathbf{W}^*_{\underline{n},\underline{m}}$ is nonsingular, whenever $m_i \leq n_i - 2$, $i \in \{1, 2\}$ holds.

3.3. Equivalence of the bias correction for $\widetilde{C}(\underline{h})$

The same construction discussed above applies also to the joint estimation using the constant-prefactor estimator $\widetilde{C}(\underline{h})$ from (2.12). We apply the joint characterization from Corollary 2.2 with the same partitioning matrices from (3.1) to get

$$\widetilde{\mathbf{\Gamma}}_{\underline{m}} = \mathbf{P}_{1,m_1} \widetilde{\mathbf{\Gamma}} \mathbf{P}_{2,m_2}^\top \tag{3.17}$$

which denotes the central submatrix of $\widetilde{\mathbf{\Gamma}}$, whose entries are obtained from $\widetilde{C}(\underline{h})$ from (2.12) instead of $\widehat{C}(\underline{h})$ from (2.11). Recall from (2.38) that the full constant prefactor weight matrix satisfies $\widetilde{\mathbf{W}}^*_{\underline{n}} = (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1)\mathbf{W}^*_{\underline{n}}$. We now apply the selection matrix $\mathbf{S}_{\underline{m}}$ to obtain the corresponding reduced weight matrix. For $i \in \{1, 2\}$, we define the central Bartlett window matrices by

$$\mathbf{\Delta}_{i,m_i} := \mathbf{P}_{i,m_i} \mathbf{\Delta}_i \mathbf{P}_{i,m_i}^\top.$$

Since $\mathbf{P}_{i,m_i}$ from (3.1) is a coordinate selection matrix, $\mathbf{P}_{i,m_i}^\top \mathbf{P}_{i,m_i}$ is the diagonal projection onto the selected coordinates, but also since $\mathbf{\Delta}_i$ is already diagonal we have that $\mathbf{P}_{i,m_i}\mathbf{\Delta}_i = \mathbf{\Delta}_{i,m_i}\mathbf{P}_{i,m_i}$. This implies that

$$\mathbf{S}_{\underline{m}}(\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1) = (\mathbf{\Delta}_{2,m_2} \otimes \mathbf{\Delta}_{1,m_1})\mathbf{S}_{\underline{m}}. \tag{3.18}$$

Therefore, the weight matrix for the constant prefactor bias correction is obtained by

$$\begin{aligned}\widetilde{\mathbf{W}}^*_{\underline{n},\underline{m}} &:= \mathbf{S}_{\underline{m}}\widetilde{\mathbf{W}}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top = \mathbf{S}_{\underline{m}}(\boldsymbol{\Delta}_2 \otimes \boldsymbol{\Delta}_1)\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top \\ &= (\boldsymbol{\Delta}_{2,m_2} \otimes \boldsymbol{\Delta}_{1,m_1})\mathbf{S}_{\underline{m}}\mathbf{W}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top = (\boldsymbol{\Delta}_{2,m_2} \otimes \boldsymbol{\Delta}_{1,m_1})\mathbf{W}^*_{\underline{n},\underline{m}}. \end{aligned} \tag{3.19}$$

Whenever the reduced weight matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ is invertible, the matrix $\widetilde{\mathbf{W}}^*_{\underline{n},\underline{m}}$ is invertible as well, since it is the product of $\mathbf{W}^*_{\underline{n},\underline{m}}$ and the invertible diagonal matrix $\boldsymbol{\Delta}_{2,m_2} \otimes \boldsymbol{\Delta}_{1,m_1}$. Analogously to (3.11), we define the bias-corrected vector of central sample covariance estimators with constant prefactor by

$$\widetilde{\underline{C}}^*_{\underline{m}} = \big(\mathbf{S}_{\underline{m}}\widetilde{\mathbf{W}}^*_{\underline{n}}\mathbf{S}_{\underline{m}}^\top\big)^{-1}\mathbf{S}_{\underline{m}}\mathrm{vec}\big(\widetilde{\boldsymbol{\Gamma}}\big) = \big(\widetilde{\mathbf{W}}^*_{\underline{n},\underline{m}}\big)^{-1}\mathrm{vec}\big(\widetilde{\boldsymbol{\Gamma}}_{\underline{m}}\big). \tag{3.20}$$

Interestingly, although the two bias corrections start from estimators with different prefactors and different (biased) expectations, the correction removes this discrepancy entirely: it turns out that both bias-corrected estimators coincide.

Corollary 3.1 (Both bias-corrected estimators coincide)**.** *For every cutoff $\underline{m} = (m_1, m_2)$ that implies invertibility of $\mathbf{W}^*_{\underline{n},\underline{m}}$, the bias-corrected estimators $\widehat{\underline{C}}^*_{\underline{m}}$ (standard) from* (3.11) *and $\widetilde{\underline{C}}^*_{\underline{m}}$ (constant prefactor) from* (3.20) *coincide. That is, we have*

$$\widetilde{\underline{C}}^*_{\underline{m}} = \widehat{\underline{C}}^*_{\underline{m}}. \tag{3.21}$$

The choice of the prefactor $1/N_h$ or $1/N$, which matters considerably for the standard and constant prefactor estimators (2.11) and(2.12), as discussed in Section 2.2, is thus irrelevant after the bias correction has been applied and the Bartlett taper is removed as a by-product. However, the main advantage of the constant prefactor estimator is also lost: the bias-corrected estimator does not necessarily lead to a positive semidefinite covariance. We address this issue in Section 3.5.

3.4. Mean-squared error analysis

For assessing the overall estimation performance of the bias-corrected covariance estimator $\widehat{\underline{C}}^*_{\underline{m}}$ beyond its bias properties, its mean-squared error (MSE) is typically considered. For this purpose, first, we define the covariance matrix of all sample covariances

$$\boldsymbol{\Omega} := \mathrm{Cov}\big(\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\big) \in \mathbb{R}^{L_{\underline{n}} \times L_{\underline{n}}} \tag{3.22}$$

and using $\widehat{\underline{C}}_{\underline{m}} = \mathbf{S}_{\underline{m}}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})$, we have

$$\boldsymbol{\Omega}^*_{\underline{m}} := \mathrm{Cov}\left(\widehat{\underline{C}}^*_{\underline{m}}\right) = (\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\mathbf{S}_{\underline{m}}\,\boldsymbol{\Omega}\,\mathbf{S}_{\underline{m}}^\top(\mathbf{W}^{*\top}_{\underline{n},\underline{m}})^{-1} \in \mathbb{R}^{M_{\underline{m}} \times M_{\underline{m}}}, \tag{3.23}$$

where each entry is given by

$$\big[\boldsymbol{\Omega}^*_{\underline{m}}\big]_{k,l} = \mathrm{Cov}\big(\widehat{C}^*(\underline{\ell}_{W_{\underline{m}}}(k)),\, \widehat{C}^*(\underline{\ell}_{W_{\underline{m}}}(l))\big), \qquad k,l = 1,\ldots,M_{\underline{m}},$$

with $\underline{\ell}_{W_{\underline{m}}}$ the enumeration (2.9) of the within-window lag set $W_{\underline{m}} = I(-m_1 : m_1) \times I(-m_2 : m_2)$, for which $a_i = -m_i$, $c_i = 2m_i + 1$ and $|W_{\underline{m}}| = M_{\underline{m}}$. That is, we have

$$\underline{\ell}_{W_{\underline{m}}}(k) = \big(((k-1) \bmod (2m_1+1)) - m_1,\ \lfloor (k-1)/(2m_1+1) \rfloor - m_2\big).$$

Hence, in the setup considered here, for estimating $\underline{C}_{\underline{m}}$, the MSE of the bias-corrected vector of (central) sample covariance estimators $\widehat{\underline{C}}^*_{\underline{m}}$ becomes

$$\mathrm{MSE}(\widehat{\underline{C}}^*_{\underline{m}}) := \mathbb{E}\left(\left\|\widehat{\underline{C}}^*_{\underline{m}} - \underline{C}_{\underline{m}}\right\|_2^2\right) = \mathbb{E}\left(\left(\widehat{\underline{C}}^*_{\underline{m}} - \underline{C}_{\underline{m}}\right)^\top\left(\widehat{\underline{C}}^*_{\underline{m}} - \underline{C}_{\underline{m}}\right)\right) = \mathrm{tr}\left(\boldsymbol{\Omega}^*_{\underline{m}}\right) + \|\underline{b}^*_{\underline{m}}\|_2^2, \tag{3.24}$$

where $\|\cdot\|_2$ denotes the Euclidean (vector) norm, $\mathrm{tr}(\cdot)$ the trace of a (square) matrix. That is, in the above, $\underline{b}^*_{\underline{m}}$ is the $M_{\underline{m}}$-dimensional bias vector of $\widehat{\underline{C}}^*_{\underline{m}}$ from Theorem 3.1, and

$$\mathrm{tr}(\boldsymbol{\Omega}^*_{\underline{m}}) = \mathrm{tr}\left((\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\mathbf{S}_{\underline{m}}\,\boldsymbol{\Omega}\,\mathbf{S}^\top_{\underline{m}}(\mathbf{W}^{*\top}_{\underline{n},\underline{m}})^{-1}\right) = \sum_{i=1}^{M_{\underline{m}}}\mathrm{Var}(\widehat{C}^*_{\underline{m},i}).$$

3.5. Plug-in estimator of the variance covariance matrix and its mean squared error

The MSE in (3.24) holds for the vector of bias-corrected estimators $\widehat{C}^*_{\underline{m}}$ only. In practical application, one is often interested in working with the proper variance covariance matrix of the vectorized spatial lattice

$$\boldsymbol{\Sigma} := \mathrm{Cov}\big(\mathrm{vec}(\mathbf{X}_{\underline{n}})\big) \in \mathbb{R}^{N\times N}. \tag{3.25}$$

We use the column-wise enumeration $G_{\underline{n}} = \{\underline{\ell}_{G_{\underline{n}}}(1),\ldots,\underline{\ell}_{G_{\underline{n}}}(N)\}$ given by (2.9) applied to the grid $G_{\underline{n}} = I(1:n_1)\times I(1:n_2)$, for which $a_i = 1$, $c_i = n_i$ and $|G_{\underline{n}}| = N$,

$$\underline{\ell}_{G_{\underline{n}}}(k) = \left(\big((k-1) \bmod n_1\big)+1,\ \left\lfloor\frac{k-1}{n_1}\right\rfloor+1\right), \qquad k=1,\ldots,N,$$

to define the entries as

$$[\boldsymbol{\Sigma}]_{k,l} = C\big(\underline{\ell}_{G_{\underline{n}}}(k) - \underline{\ell}_{G_{\underline{n}}}(l)\big), \qquad k,l=1,\ldots,N.$$

Then, we build a simple plug-in estimator by replacing C by $\widehat{C}^*_{\underline{m}}$ and tapering to zero the entries where the lag lies outside the within-window lag

$$\big[\widehat{\boldsymbol{\Sigma}}^*_{\underline{m}}\big]_{k,l} = \begin{cases}\widehat{C}^*_{\underline{m}}\big(\underline{\ell}_{G_{\underline{n}}}(k) - \underline{\ell}_{G_{\underline{n}}}(l)\big), & \big(\underline{\ell}_{G_{\underline{n}}}(k) - \underline{\ell}_{G_{\underline{n}}}(l)\big)\in W_{\underline{m}},\\ 0, & \text{otherwise.}\end{cases} \qquad k,l=1,\ldots,N. \tag{3.26}$$

We define the full matrix containing the effective sample sizes for all the lags $\underline{h}\in H_{\underline{n}}$ using the Bartlett weight matrices from (2.35)

$$\mathbf{N}_{\underline{n}} = N(\boldsymbol{\Delta}_2\otimes\boldsymbol{\Delta}_1) = \mathrm{diag}(N_h) \in \mathbb{R}^{L_{\underline{n}}\times L_{\underline{n}}}$$

Then, under Assumption (2.1), the exact mean squared error of (3.26) is given by the $\mathrm{MSE}(\widehat{\underline{C}}^*_{\underline{m}}(\underline{h}))$ plus the bias produced by the sum of the squared tapered covariances

$$\mathrm{MSE}\left(\widehat{\boldsymbol{\Sigma}}^*_{\underline{m}}\right) := \mathbb{E}\left(\left\|\widehat{\boldsymbol{\Sigma}}^*_{\underline{m}} - \boldsymbol{\Sigma}\right\|^2_F\right) = \mathrm{tr}\left(\mathbf{S}_{\underline{m}}\mathbf{N}_{\underline{n}}\mathbf{S}^\top_{\underline{m}}\left(\boldsymbol{\Omega}^*_{\underline{m}} + \underline{b}^*_{\underline{m}}\underline{b}^{*\top}_{\underline{m}}\right)\right) + \mathrm{tr}\left(\mathbf{S}^c_{\underline{m}}\mathbf{N}_{\underline{n}}\mathbf{S}^{c\top}_{\underline{m}}\mathbf{S}^c_{\underline{m}}\underline{C}\underline{C}^\top\mathbf{S}^{c\top}_{\underline{m}}\right), \tag{3.27}$$

where $\|\cdot\|_F$ denotes the Frobenius (matrix) norm.

Remark 3.3 (On the positive semidefiniteness of $\widehat{\boldsymbol{\Sigma}}^*_{\underline{m}}$).

(i) The tapering does not necessarily produce a positive semidefinite matrix, thus not a valid covariance matrix. However, by construction, it is a real symmetric matrix, which always admits the spectral decomposition $\widehat{\boldsymbol{\Sigma}}^*_{\underline{m}} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\top$ with orthogonal $\mathbf{Q}$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1,\ldots,\lambda_N)$. We apply the spectral projections used, e.g., in Jentsch and Politis (2015), and set all negative eigenvalues to zero and define the projected estimator

$$\widehat{\boldsymbol{\Sigma}}^{*+}_{\underline{m}} = \mathbf{Q}\boldsymbol{\Lambda}^+\mathbf{Q}^\top, \quad \text{where} \quad \boldsymbol{\Lambda}^+ = \mathrm{diag}\big(\max(\lambda_1,0),\ldots,\max(\lambda_N,0)\big), \tag{3.28}$$

which is assured to be positive semidefinite. Similarly, by setting

$$\widetilde{\mathbf{\Lambda}}^{+} = \operatorname{diag}\big(\max(\lambda_1, \epsilon), \ldots, \max(\lambda_N, \epsilon)\big), \quad \text{where} \quad \epsilon > 0$$

and using $\widetilde{\mathbf{\Lambda}}^{+}$ instead of $\mathbf{\Lambda}^{+}$, we get a positive definite estimator.

(ii) For fixed $\underline{m}$, When $n_1, n_2 \to \infty$, both estimators $\widehat{\mathbf{\Sigma}}^{*}_{\underline{m}}$ and $\widehat{\mathbf{\Sigma}}^{*+}_{\underline{m}}$ converge to the same limit that is always positive semidefinite.

3.6. Asymptotic theory

In this section, we provide asymptotic theory for the sample covariance estimator with and without bias correction. For this purpose, following Vogelsang and Yang (2016) and Li (2023) for the time series case, we suppose that the spatial process $\mathbf{X}$ is linear.

Assumption 3.1 (Linear spatial process)**.** *The spatial lattice process* $\mathbf{X} = \{X_{\underline{t}}, \underline{t} \in \mathbb{Z}^2\}$ *is linear,*

$$X_{\underline{t}} = \mu + \sum_{\underline{j} \in \mathbb{Z}^2} \psi_{\underline{j}} \epsilon_{\underline{t}-\underline{j}}, \quad \underline{t} \in \mathbb{Z}^2, \tag{3.29}$$

where $\{\epsilon_{\underline{t}}, \underline{t} \in \mathbb{Z}^2\}$ *is an i.i.d. white noise process with* $\mathbb{E}(\epsilon_{\underline{t}}) = 0$, $\operatorname{Var}(\epsilon_{\underline{t}}) = \sigma^2_{\epsilon} \in (0, \infty)$ *and with finite fourth moments* $\mathbb{E}(\epsilon^4_{\underline{t}}) = \eta\sigma^4_{\epsilon} < \infty$*. Further, suppose that the sequence* $\{\psi_{\underline{j}}, \underline{j} \in \mathbb{Z}^2\}$ *is absolutely summable,* $\sum_{\underline{j} \in \mathbb{Z}^2} |\psi_{\underline{j}}| < \infty$*.*

However, other types of dependence assumptions may be employed such as mixing-type conditions (Bolthausen, 1982; Politis et al., 1999; Bradley, 2007), weak dependence concepts (Dedecker et al., 2007), or physical/functional dependence concepts (El Machkouri et al., 2013). Irrespective of the dependency concept used, it turns out that the bias-corrected sample covariance estimator (for both considered prefactors) is asymptotically equivalent to the uncorrected standard one.

Proposition 3.1 (Asymptotics for $\widehat{\underline{C}}_{\underline{m}}$ and $\widehat{\underline{C}}^{*}_{\underline{m}}$)**.** *Let* $\underline{m} \in \mathbb{N}^2$ *be fixed and let Assumptions 2.1 and 3.1 hold. Then, as* $n_1, n_2 \to \infty$*, we have*

$$\sqrt{N}\left(\widehat{\underline{C}}_{\underline{m}} - \underline{C}_{\underline{m}}\right) \xrightarrow{d} \mathcal{N}\left(\underline{0}, \mathbf{V}_{\underline{m}}\right) \quad \textit{and} \quad \sqrt{N}\left(\widehat{\underline{C}}^{*}_{\underline{m}} - \underline{C}_{\underline{m}}\right) \xrightarrow{d} \mathcal{N}\left(\underline{0}, \mathbf{V}_{\underline{m}}\right),$$

with

$$\mathbf{V}_{\underline{m}} := \lim_{N \to \infty} N\, \mathbf{\Omega}^{*}_{\underline{m}}, \qquad \big[\mathbf{V}_{\underline{m}}\big]_{k,l} = V\big(\underline{\ell}_{W_{\underline{m}}}(k), \underline{\ell}_{W_{\underline{m}}}(l)\big), \quad k, l = 1, \ldots, M_{\underline{m}}, \tag{3.30}$$

with $\underline{\ell}_{W_{\underline{m}}}$ *the enumeration* (2.9) *of* $W_{\underline{m}}$ *and where the entries of* $\mathbf{V}_{\underline{m}}$ *are of the form*

$$V(\underline{h}, \underline{g}) = (\eta - 3)C(\underline{h})C(\underline{g}) + \sum_{\underline{\delta} \in \mathbb{Z}^2} \Big[C(\underline{\delta})C(\underline{\delta} + \underline{g} - \underline{h}) + C(\underline{\delta} + \underline{g})C(\underline{\delta} - \underline{h})\Big], \quad \underline{h}, \underline{g} \in W_{\underline{m}}. \tag{3.31}$$

Some remarks on the limiting results in Proposition 3.1 are in order.

Remark 3.4 (On Proposition 3.1 and possible extensions).

(i) The result in Proposition 3.1 extends Theorem 1 in Vogelsang and Yang (2016) to linear lattice processes in the plane.

(ii) The same limiting distribution does also hold for $\widetilde{\underline{C}}_m = \operatorname{vec}\big(\widetilde{\mathbf{\Gamma}}_m\big)$, which relies on the constant prefactor covariance estimator. However, according to Corollary 3.1, its bias corrected version $\widetilde{\underline{C}}_m = \operatorname{vec}\big(\widetilde{\mathbf{\Gamma}}_m\big)$ coincides with $\widehat{\underline{C}}_m = \operatorname{vec}\big(\widehat{\mathbf{\Gamma}}_m\big)$ and is trivially also covered by Proposition 3.1.

(iii) Without relying on the Assumption 3.1 of linearity of the underlying process $\mathbf{X}$, under suitable mixing conditions (that guarantee sufficient summability conditions for the cumulants), asymptotic normality of $\sqrt{N}(\widehat{\underline{C}}_m - \underline{C}_m)$ can still be obtained. However, by dropping Assumption 3.1 and using summability conditions for cumulants, the entries of the limiting variance change to become

$$V_{general}(\widehat{\underline{C}}_m)(\underline{h},\underline{g}) = \sum_{\underline{\delta}\in\mathbb{Z}^2} \operatorname{cum}\Big(X_{\underline{h}} - \mu, X_{\underline{0}} - \mu, X_{\underline{\delta}+\underline{g}} - \mu, X_{\underline{\delta}} - \mu\Big) + \sum_{\underline{\delta}\in\mathbb{Z}^2} \Big[C(\underline{\delta})C(\underline{\delta}+\underline{g}-\underline{h}) + C(\underline{\delta}+\underline{g})C(\underline{\delta}-\underline{h})\Big],$$

where $\operatorname{cum}(W,X,Y,Z)$ denotes the joint cumulant of random variables W, X, Y and Z.

3.7. Selection of the cutoff parameter

The cutoff $\underline{m}$ plays here the role that commonly is taken by the bandwidth in classical nonparametric estimation methods. In particular, it is not identified by the model, it governs a bias-variance trade-off and it has to be fixed before the estimator can be computed. Hyperparameter selection is a vast topic of interest in nonparametric methods, which have a need of defining the optimal hyperparameter to minimize the bias-variance trade-off. In our method, on the one hand, we have that a small $\underline{m}$ corrects the bias minimally at the cost of setting to 0 all non selected covariances with lags $\underline{h} \notin W_{\underline{m}}$, on the other hand, a higher $\underline{m}$ results in a larger bias correction, but it inflates the variance of the estimator, resulting in poorer results according to MSE criteria. As usual for hyperparameter selection of this kind, practitioners often rely on data-driven methods to find parameters that perform well given the task at hand. We propose here two methods for the selection of the cutoff parameters $\underline{m}$.

First, we suggest a *plug-in rule* from a fitted correlogram.

1. For $i = 1, 2$, with $\underline{e}_1 = (1,0)$ and $\underline{e}_2 = (0,1)$ and evaluate the correlogram by $\widehat{\rho}_i(h) := \widehat{C}(h\,\underline{e}_i)/\widehat{C}(\underline{0})$ for $h = 1, \ldots, K_i$, where one can choose $K_i = \lfloor \min(n_1, n_2)/2 \rfloor$.

2. For a correlogram model given by $\rho(h) = \exp(-h/\lambda)$, we fit a weighted least squares model through the origin on the log scale, with weights given by the lags similar to those in Cressie (1985), to obtain an estimate of the inverse of the dependence parameter,

$$\widehat{\lambda}_i^{-1} = -\frac{\sum_{h=1}^{K_i} w_{i,h} h \log \widehat{\rho}_i(h)}{\sum_{h=1}^{K_i} w_{i,h} h^2}, \quad w_{i,h} = n_i - h. \tag{3.32}$$

3. We scale the fitted dependence parameter by the axis length and we select the cutoff parameter by $\widehat{m}_i^{\text{rule}} := \lfloor \kappa \left(n_i \widehat{\lambda}_i\right)^{\eta} \rceil$, with $\lfloor \cdot \rceil$ denoting rounding to the nearest integer and with $\kappa = 0.445$ and $\eta = 0.5$, which are not implied by the model and were calibrated by numerical simulations.

Second, we suggest applying a *block cross-validation* method.

1. Fix a grid of cutoffs, where we set $1 \le m \le \lfloor \min(n_1, n_2)/2 \rfloor$, for $m_1 = m_2 = m$.

2. For each candidate $\underline{m}$, compute $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ on the whole lattice. Then, compute its inverse $(\mathbf{\Sigma}^{*+}_{\underline{m}})^{-1}$ with the spectral projection method from (3.28).

3. Determine B square blocks $P_1, \ldots, P_B$ of size $f \times f$, all kept inside the lattice, so that the criterion does not depend on one particular position of the block.

4. For each block, predict the held-out block from the remaining sites O_b by ordinary kriging,

$$\widehat{y}_{P_b} = \widehat{\mu}\underline{\mathbb{1}} + \widehat{\mathbf{\Sigma}}_{P_bO_b}(\widehat{\mathbf{\Sigma}}^{*+}_{O_bO_b})^{-1}\big(y_{O_b} - \widehat{\mu}\underline{\mathbb{1}}\big), \quad \widehat{\mu} = (\underline{\mathbb{1}}^\top(\widehat{\mathbf{\Sigma}}^{*+}_{OO})^{-1}\underline{\mathbb{1}})^{-1}\underline{\mathbb{1}}^\top(\widehat{\mathbf{\Sigma}}^{*+}_{OO})^{-1}\underline{y}_O, \tag{3.33}$$

5. Average the root mean squared prediction error over the blocks and select the $\underline{m}$, which minimizes the criterion,

$$\mathrm{CV}(\underline{m}) = \frac{1}{B}\sum_{b=1}^{B}\Big(\frac{1}{|P_b|}\sum_{s\in P_b}\big(\widehat{y}_s - y_s\big)^2\Big)^{1/2}, \qquad \widehat{\underline{m}}^{\mathrm{cv}} = \arg\min_{\underline{m}} \mathrm{CV}(\underline{m}). \tag{3.34}$$

Unlike the plug-in rule, the block cross-validation does not assume a shape of the correlogram, however it is evidently more expensive computationally.

4. Simulation Study

In this section, we perform a Monte Carlo simulation study to assess the finite-sample performances of the naive (uncorrected) and the bias-corrected covariance estimators under varying degrees of spatial dependence. We consider spatial Gaussian lattice processes as introduced in Example 2.1 with covariance function defined in (2.10). That is, we resume the same simulation setup as described in Example 2.1 and generate a true variance covariance matrix $\mathbf{\Sigma}_{\lambda,N} \in \mathbb{R}^{N\times N}$ by plugging the entries

$$\big[\mathbf{\Sigma}_{\lambda,N}\big]_{k,l} = C_{\lambda,N}\big(\underline{\ell}_{G_{\underline{n}}}(k) - \underline{\ell}_{G_{\underline{n}}}(l)\big), \quad k,l = 1,\dots,N. \tag{4.1}$$

with $\underline{\ell}_{G_{\underline{n}}}(k) = \big(((k-1) \bmod n_1) + 1,\ \lceil k/n_1 \rceil\big)$.

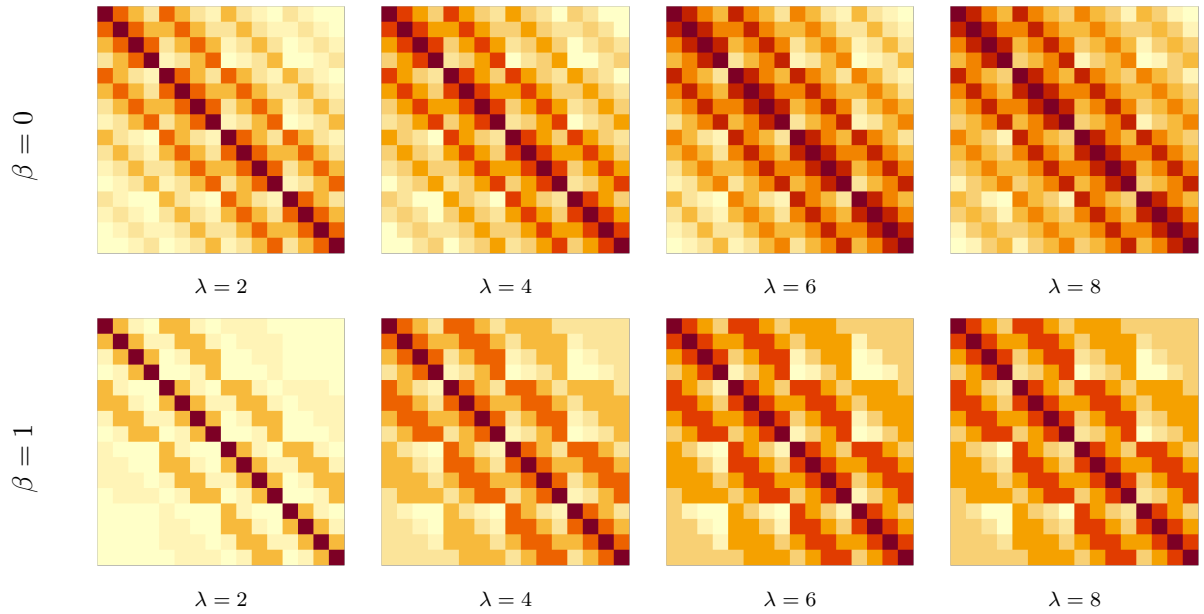


Figure 7: True (16×16)) covariance matrices $\mathbf{\Sigma}_{\lambda,16}$ for separable ($\beta = 0$; first row) and non-separable ($\beta = 1$; second row) dependence across increasing spatial dependence λ. Grid size is (4×4).

We then draw the realizations of the lattice process from a multivariate normal distribution $\text{vec}(\mathbf{X}_{\underline{n}}) \sim \mathcal{N}\big(0,\, \mathbf{\Sigma}_{\lambda,N}\big)$ and we use these realizations to estimate the covariances. We draw $10,000$ Monte Carlo replications for each setting that we examine, in order to evaluate the bias, variance and MSE of the estimators under each combinations of $\lambda \in \{2, 4, 6, 8\}$ and sample sizes $N = n_1 n_2 \in \{16, 100, 400, 900\}$, for convenience we have restricted the simulations to squared grids with $n_1 = n_2$. In Figures 7, we show the true covariance matrices for $N = 16$ and for $\beta = 0$ and $\beta = 1$, respectively.

Additionally, we compare the naive (uncorrected) and the bias-corrected estimates with a competitor variogram-based covariance estimator for the covariance function that is given by

$$\widehat{C}_{CH}(\underline{h}) = \widehat{C}(\underline{0}) - \widehat{\gamma}(\underline{h}),$$

where we use the popular variogram estimator proposed by Cressie and Hawkins (1980).

$$2\widehat{\gamma}(\underline{h}) := \left(\frac{1}{N_h} \sum_{i=1}^{N_h} \left| X\left(\underline{t}_{i+\underline{h}}\right) - X\left(\underline{t}_i\right) \right|^{1/2} \right)^4 \Big/ (0.457 + 0.494/N_h) \tag{4.2}$$

4.1. Separable case

We analyse the results of the case in which the true covariance $C_{\lambda,N}(\underline{h})$ from Equation (2.10) has a separable parametrization ($\beta = 0$). In order to aggregate the results of two dimensional lattices, we often use the Euclidean norm applied to the lag vector $\|\underline{h}\|_2$. We compare the performance of the vector of standard (or as we call below, naive) estimator $\widehat{\underline{C}}$, with entries given by (2.11), and the vector of the Cressie-Hawkins $\widehat{\underline{C}}_{CH}$, with entries given by (4.2), with the proposed bias-corrected estimator $\underline{C}^*_{\underline{m}}$ from (3.11). In Figure 8, in the top panel, we show the bias of the three estimators, where the bias-corrected estimator is evaluated with different cutoff parameters $\underline{m}$, where (for simplicity) we take the same cutoff value for each axis dimension, i.e., $m := m_1 = m_2$. The plot shows Monte Carlo averages of the bias as a function of the norm of the lag vector, for the grid of size $n_1 = n_2 = 10$ and for the four levels of spatial dependence determined by $\lambda \in \{2, 4, 6, 8\}$.

Across all settings, a reduction in the bias level is already achieved for $m = 1$, with increasing improvements as the cutoff parameter m increases. The naive estimator, shown as the light red solid curve, exhibits larger bias than all bias-corrected estimators over the entire range of lag norms. The Cressie–Hawkins estimator, shown as the dashed dark red line, performs slightly better than the naive estimator at small lags, but its bias increases rapidly at higher lags. Similar results are obtained for the other not reported grid sizes.

However, due to the common bias-variance trade-off effect, correcting the bias in the estimation may come at the cost of an increased variance. Hence, a more comprehensive measure for the quality of the estimator is given by the Mean Squared Error (MSE) studied in Section 3.4, which sums the squared biases and the variances of the estimator. When looking at this criterion, the benefits of the bias correction are less immediate. In Figure 8, in the bottom panel, we report the MSE results of the same setting. We see that at short range lags, all estimators exhibit similar MSE levels, but (as expected) higher-order bias corrections (corresponding to large values of m) lead to worse performances than simply using the naive estimation. At moderate to high lags, the MSE of the Cressie-Hawkins estimator blows up, reflecting the increasing inappropriateness of the bias correction it introduces. Nevertheless, for strong spatial dependence (λ large), the bias-corrected estimator shows to be not only competitive, but superior in comparison to the naive and the Cressie-Hawkins benchmarks also in terms of MSE for suitable choices of $\underline{m}$, reporting lower MSE-levels in the simulations.

We also examine the MSE of each lag pair $\underline{h} = (h_1, h_2)$ separately. For this purpose, Table 1 reports the ratios of the MSE for the bias-corrected estimator $\underline{C}^*_{\underline{m}}$ and the Cressie-Hawkins $\underline{C}_{CH}$

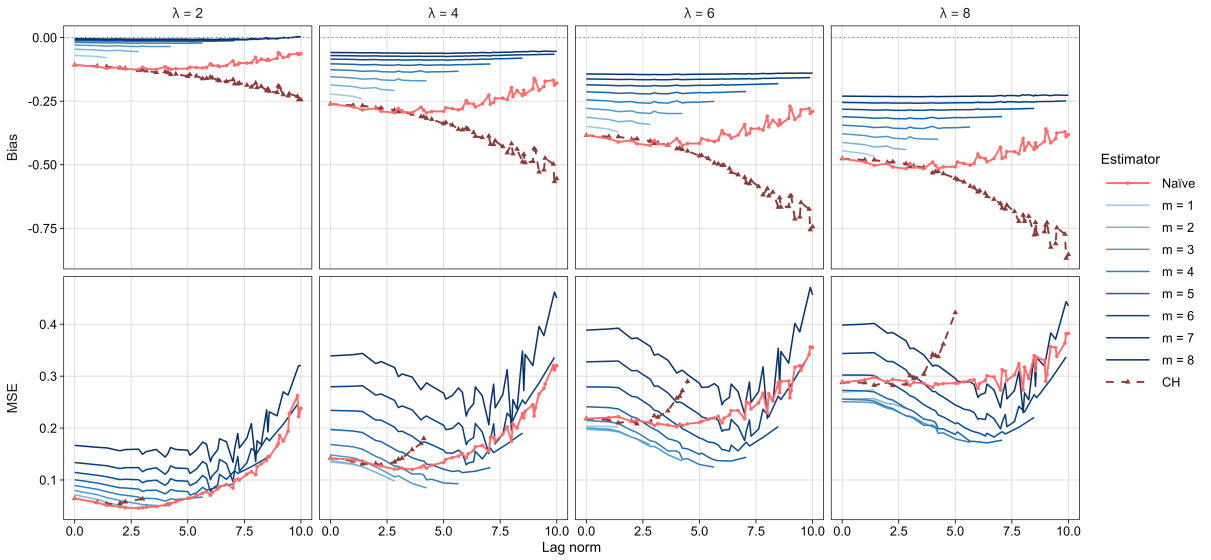


Figure 8: Bias and MSE performance with increasing lag norm $\|\underline{h}\|_2$ for the naive estimator $\widehat{\underline{C}}$ (solid light red line), Cressie-Hawkins $\widehat{\underline{C}}_{CH}$ estimator (dashed dark red line) and the bias-corrected estimator $\widehat{\underline{C}}^*_m$ for different m values (several blue shades) for four spatial dependence decay parameters $\lambda \in \{2, 4, 6, 8\}$ and for grid size $n_1 = n_2 = 10$. Separable model $\beta = 0$.

estimator over the standard (bias-uncorrected) one $\widehat{\underline{C}}$. Values that are smaller than 1 indicate those instances, where the competitor estimator works better than the standard one. We show that indeed the gain in MSE becomes quite substantial in the strong dependency case, while still some valuable gains are already achieved for short lags with cutoff value $m \in \{1, 2\}$.

Additionally, to visualize the MSE results from Table 1, we show heatmap plots in Figures F.15 (for $n_1 = n_2 = 10$) and F.16 (for $n_1 = n_2 = 20$) in the appendix. There, for each lag pair (h_1, h_2) shown in each plot axis, it is indicated whether the bias-corrected estimator $\underline{C}^*_m$ can achieve a better MSE than the uncorrected estimator $\widehat{\underline{C}}$ and whether simple tapering the covariance estimator to zero leads to the best MSE. We refer to Appendix F for details.

We now investigate the maximal MSE reduction attainable by selecting, for each lag pair, the optimal cutoff parameter m, namely we report the oracle choice for the cutoff parameter that minimizes the MSE of the bias-corrected estimator and compare this with the MSE of the naive estimator. This is indicated in Figure 9, which shows the relative increase or decrease in MSE compared to the naive estimator, when each lag pair (h_1, h_2), appearing in the plots' axes, is evaluated at its optimal m for the (10×10) grid and across the four dependence levels λ. Negative percentages indicate a reduction in MSE while positive percentages indicate an increase in MSE. We notice that there are some slight gains in MSE estimation at short lags like $\underline{h} = (0, 1)$ for the weak dependence case (λ small), while the gains in MSE also become more substantial as the dependency structure increases (λ large). In these plots, a crossed-out value indicates that tapering the covariance estimator to zero results in a better MSE than using the naive or the bias-corrected estimator. That is, in these cases, one should not (try to) estimate them at all from the data.

Table 1: MSE Ratio $\widehat{C}^*_m/\widehat{C}$ and $\widehat{C}_{CH}/\widehat{C}$ for selected (h_1, h_2) and cutoff parameter m.

λ	N	(h_1,h_2)	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$	$m=10$	CH
2	4	(0,0)	1.041	1.605	–	–	–	–	–	–	–	–	1.000
		(1,0)	0.956	1.619	–	–	–	–	–	–	–	–	1.130
		(1,1)	0.821	1.538	–	–	–	–	–	–	–	–	1.285
		(2,0)	–	1.140	–	–	–	–	–	–	–	–	1.663
		(2,2)	–	0.900	–	–	–	–	–	–	–	–	3.541
	10	(0,0)	1.011	1.105	1.235	1.384	1.558	1.775	2.072	2.583	–	–	1.000
		(1,0)	0.994	1.105	1.264	1.448	1.659	1.923	2.280	2.875	–	–	1.007
		(1,1)	0.976	1.104	1.295	1.517	1.772	2.083	2.499	3.174	–	–	1.043
		(2,0)	–	1.058	1.249	1.479	1.744	2.071	2.511	3.223	–	–	1.130
		(2,2)	–	0.966	1.185	1.470	1.807	2.203	2.718	3.498	–	–	1.526
	20	(0,0)	1.005	1.034	1.070	1.105	1.136	1.163	1.188	1.213	1.240	1.270	1.000
		(1,0)	1.003	1.037	1.081	1.124	1.163	1.197	1.228	1.259	1.290	1.327	0.996
		(1,1)	0.999	1.035	1.087	1.138	1.184	1.224	1.261	1.297	1.336	1.380	1.009
		(2,0)	–	1.029	1.084	1.140	1.192	1.238	1.280	1.320	1.360	1.407	1.058
		(2,2)	–	1.008	1.071	1.142	1.210	1.270	1.325	1.377	1.431	1.492	1.208
	30	(0,0)	1.003	1.016	1.033	1.049	1.063	1.073	1.081	1.088	1.094	1.100	1.000
		(1,0)	1.001	1.016	1.037	1.056	1.072	1.085	1.096	1.104	1.111	1.118	1.000
		(1,1)	1.000	1.017	1.040	1.063	1.082	1.097	1.110	1.120	1.129	1.137	1.014
		(2,0)	–	1.014	1.039	1.064	1.086	1.103	1.117	1.129	1.139	1.148	1.042
		(2,2)	–	1.007	1.036	1.067	1.096	1.120	1.139	1.156	1.170	1.182	1.165
4	4	(0,0)	0.862	0.993	–	–	–	–	–	–	–	–	1.000
		(1,0)	0.807	0.921	–	–	–	–	–	–	–	–	1.034
		(1,1)	0.743	0.843	–	–	–	–	–	–	–	–	1.031
		(2,0)	–	0.766	–	–	–	–	–	–	–	–	1.474
		(2,2)	–	0.658	–	–	–	–	–	–	–	–	2.583
	10	(0,0)	0.947	0.964	1.044	1.185	1.387	1.647	1.967	2.385	–	–	1.000
		(1,0)	0.932	0.943	1.027	1.179	1.398	1.679	2.021	2.459	–	–	0.983
		(1,1)	0.916	0.920	1.006	1.168	1.403	1.701	2.064	2.520	–	–	0.959
		(2,0)	–	0.896	0.975	1.136	1.376	1.685	2.063	2.537	–	–	1.034
		(2,2)	–	0.812	0.879	1.052	1.323	1.675	2.103	2.626	–	–	1.105
	20	(0,0)	0.987	0.995	1.026	1.074	1.134	1.200	1.271	1.347	1.429	1.516	1.000
		(1,0)	0.983	0.991	1.023	1.074	1.138	1.210	1.286	1.368	1.455	1.548	0.976
		(1,1)	0.980	0.986	1.019	1.075	1.144	1.222	1.305	1.394	1.487	1.587	0.958
		(2,0)	–	0.979	1.011	1.066	1.137	1.217	1.304	1.396	1.494	1.598	0.975
		(2,2)	–	0.960	0.991	1.052	1.135	1.231	1.335	1.446	1.564	1.689	0.978
	30	(0,0)	0.995	1.000	1.016	1.040	1.070	1.103	1.136	1.169	1.201	1.232	1.000
		(1,0)	0.994	0.998	1.015	1.041	1.073	1.108	1.144	1.179	1.214	1.247	0.982
		(1,1)	0.992	0.996	1.013	1.041	1.075	1.113	1.151	1.189	1.226	1.262	0.965
		(2,0)	–	0.994	1.011	1.040	1.076	1.115	1.155	1.196	1.235	1.272	0.977
		(2,2)	–	0.987	1.004	1.035	1.076	1.123	1.171	1.219	1.266	1.312	0.974
6	4	(0,0)	0.855	0.852	–	–	–	–	–	–	–	–	1.000
		(1,0)	0.822	0.803	–	–	–	–	–	–	–	–	1.021
		(1,1)	0.784	0.750	–	–	–	–	–	–	–	–	0.997
		(2,0)	–	0.734	–	–	–	–	–	–	–	–	1.337
		(2,2)	–	0.670	–	–	–	–	–	–	–	–	1.982
	10	(0,0)	0.933	0.904	0.919	0.984	1.105	1.280	1.500	1.781	–	–	1.000
		(1,0)	0.923	0.887	0.898	0.964	1.088	1.268	1.494	1.776	–	–	0.983
		(1,1)	0.912	0.868	0.875	0.940	1.067	1.252	1.484	1.769	–	–	0.946
		(2,0)	–	0.858	0.860	0.922	1.048	1.234	1.468	1.761	–	–	1.008
		(2,2)	–	0.802	0.785	0.838	0.965	1.160	1.407	1.710	–	–	1.006
	20	(0,0)	0.978	0.967	0.975	1.001	1.046	1.105	1.177	1.262	1.358	1.466	1.000
		(1,0)	0.975	0.962	0.969	0.995	1.041	1.103	1.178	1.267	1.367	1.480	0.979
		(1,1)	0.972	0.957	0.962	0.989	1.037	1.101	1.180	1.272	1.377	1.495	0.953
		(2,0)	–	0.953	0.956	0.982	1.030	1.095	1.175	1.269	1.377	1.497	0.972
		(2,2)	–	0.936	0.934	0.959	1.008	1.079	1.168	1.272	1.392	1.526	0.945
	30	(0,0)	0.990	0.987	0.992	1.007	1.032	1.064	1.102	1.144	1.189	1.236	1.000
		(1,0)	0.989	0.985	0.990	1.005	1.031	1.064	1.103	1.147	1.194	1.243	0.983
		(1,1)	0.988	0.983	0.987	1.003	1.030	1.064	1.105	1.151	1.200	1.251	0.961
		(2,0)	–	0.981	0.985	1.001	1.027	1.063	1.105	1.151	1.202	1.255	0.974
		(2,2)	–	0.975	0.976	0.992	1.020	1.058	1.105	1.157	1.213	1.272	0.949
8	4	(0,0)	0.870	0.826	–	–	–	–	–	–	–	–	1.000
		(1,0)	0.847	0.790	–	–	–	–	–	–	–	–	1.019
		(1,1)	0.821	0.749	–	–	–	–	–	–	–	–	0.984
		(2,0)	–	0.755	–	–	–	–	–	–	–	–	1.255
		(2,2)	–	0.712	–	–	–	–	–	–	–	–	1.649
	10	(0,0)	0.935	0.890	0.871	0.888	0.947	1.050	1.194	1.384	–	–	1.000
		(1,0)	0.927	0.876	0.853	0.868	0.927	1.032	1.179	1.370	–	–	0.983
		(1,1)	0.920	0.861	0.834	0.846	0.905	1.011	1.158	1.347	–	–	0.948
		(2,0)	–	0.856	0.825	0.833	0.889	0.993	1.140	1.331	–	–	0.999
		(2,2)	–	0.818	0.770	0.767	0.819	0.923	1.072	1.264	–	–	0.980
	20	(0,0)	0.974	0.954	0.945	0.951	0.974	1.012	1.065	1.133	1.217	1.315	1.000
		(1,0)	0.972	0.949	0.939	0.944	0.966	1.005	1.060	1.130	1.216	1.317	0.980
		(1,1)	0.970	0.945	0.932	0.937	0.959	0.999	1.055	1.127	1.215	1.319	0.956
		(2,0)	–	0.943	0.928	0.932	0.953	0.993	1.049	1.121	1.210	1.316	0.975
		(2,2)	–	0.930	0.909	0.908	0.928	0.968	1.027	1.104	1.199	1.311	0.945
	30	(0,0)	0.988	0.978	0.975	0.980	0.994	1.016	1.047	1.085	1.128	1.177	1.000
		(1,0)	0.987	0.976	0.972	0.976	0.990	1.014	1.045	1.084	1.128	1.178	0.982
		(1,1)	0.986	0.975	0.969	0.973	0.987	1.011	1.043	1.083	1.129	1.180	0.963
		(2,0)	–	0.973	0.967	0.971	0.984	1.008	1.041	1.081	1.127	1.180	0.974
		(2,2)	–	0.968	0.959	0.961	0.975	0.999	1.033	1.076	1.126	1.182	0.944

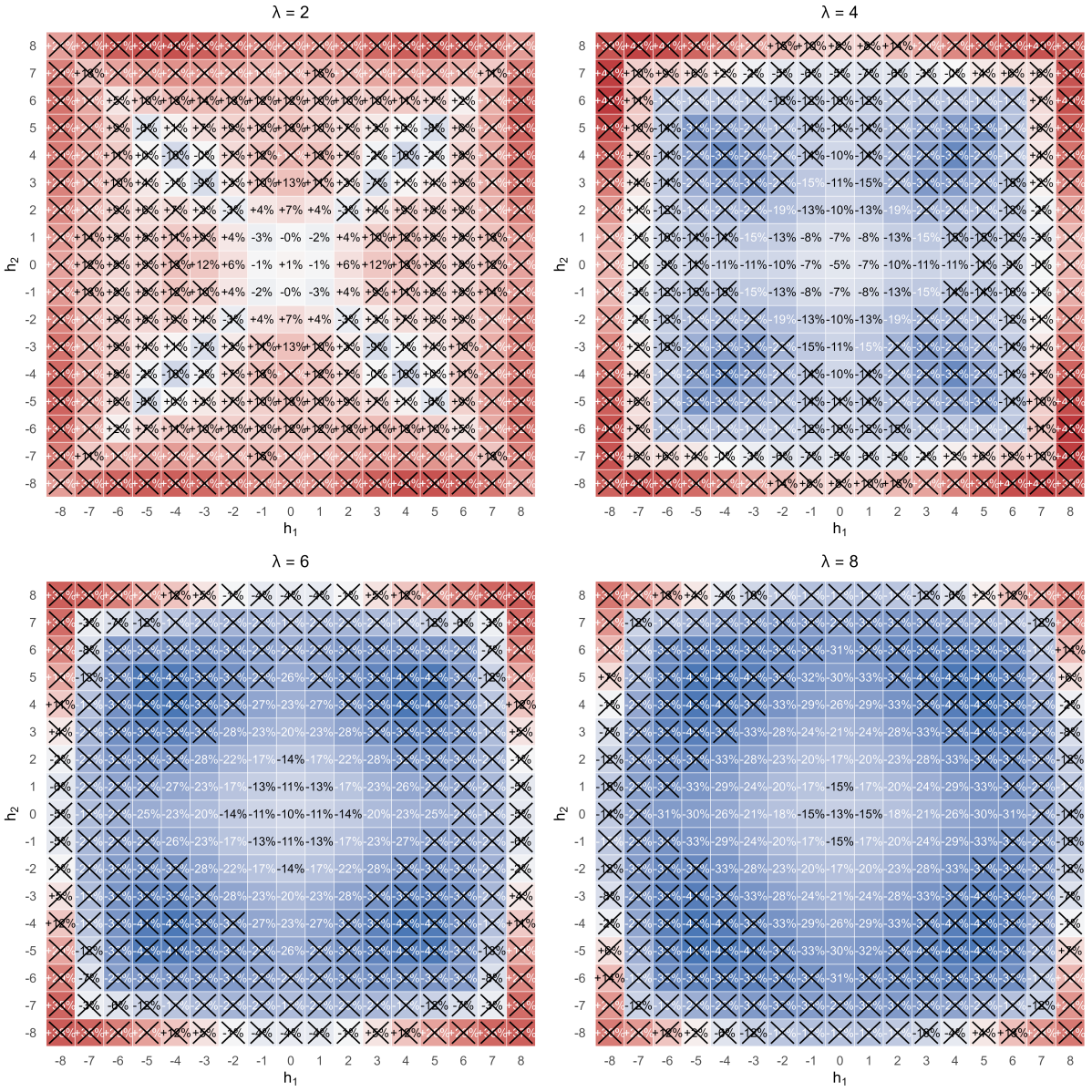


Figure 9: MSE ratio of the estimator with the optimal cutoff m for each lag pair (h_1, h_2) and naive estimator. Negative percentages indicate a reduction in MSE while positive percentages indicate an increase in MSE for the specific lag pair. Equivalently, blue shades indicate a reduction in MSE while red shades indicate a worsening of the MSE when compared to the naive estimator. Grid size (10×10). Separable case $\beta = 0$.

Finally, we are interested in reporting the properties of the estimations of the variance covariance matrix $\mathbf{\Sigma}$. We consider the naive estimator $\widehat{\mathbf{\Sigma}}$, whose entries are given by the standard sample estimator $\widehat{C}(\underline{h})$ from (2.11), the bias-corrected estimator $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ defined in (3.26) and a tapered estimator $\widehat{\mathbf{\Sigma}}_{\underline{m}}$, where the entries not selected by the m parameter are tapered by zero, just like in (3.26), but the selected entries are given by with the standard estimator $\widehat{C}(\underline{h})$ from (2.11), such that only the tapering is applied but not the bias correction. Following Cai et al. (2010), we evaluate in Figure 10 the normalized squared Frobenius norm

$$\frac{1}{N}\mathbb{E}\left(\left\|\widehat{\mathbf{\Sigma}}^*_{\underline{m}} - \mathbf{\Sigma}\right\|_F^2\right).$$

We observe that for suitably selected cutoff values both the bias-corrected and the tapered estimator obtain substantially lower errors than the simple naive estimator. This behavior is well-known and is due to the regularization induced by the tapering present in both $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ and $\widehat{\mathbf{\Sigma}}_{\underline{m}}$, see among others (Ledoit and Wolf, 2004; Furrer et al., 2006; Bickel and Levina, 2008b,a). The additional benefit of the bias-correction occurs for small grids and strong dependence, where $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ attains lower values than tapering alone done by $\widehat{\mathbf{\Sigma}}_{\underline{m}}$. In the other settings, we see that they attain similar levels.

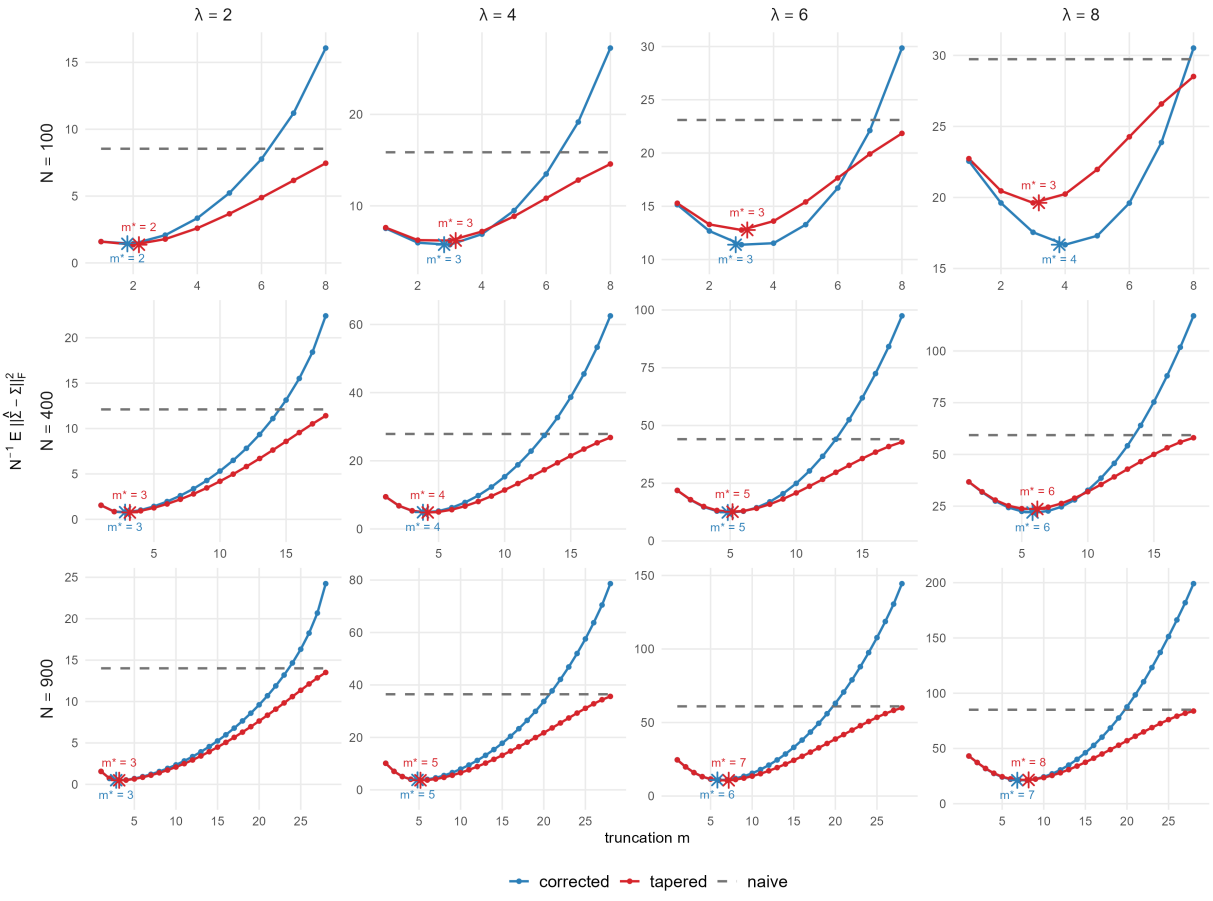


Figure 10: Average Normalized squared Frobenius norm difference of 10,000 Monte Carlo simulations vs different cutoff parameter m comparing the naive estimator $\widehat{\mathbf{\Sigma}}$ (gray), the bias-corrected estimator $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ (blue) and the tapered estimator $\widehat{\mathbf{\Sigma}}_{\underline{m}}$ (red)

To complement the aggregate error measure given by the Frobenius criterion, we also consider the spectral or operator norm, which for symmetric matrices, it corresponds to the absolute largest eigenvalue. Spectral norm consistency is often investigated in the literature, see e.g. Bickel and Levina (2008b), thus we report in Figure 11 the spectral norm

$$\mathbb{E}\left(\left\|\widehat{\mathbf{\Sigma}}^*_{\underline{m}} - \mathbf{\Sigma}\right\|_2\right).$$

The results of the simulations broadly confirm that the evidence presented in Figure 10, where the bias-correction estimator achieves visible improvements in settings with small grid sizes and strong dependence. However, we also see improvements from simple tapering also in the moderate and large grid sizes and when the dependence is moderate or strong.

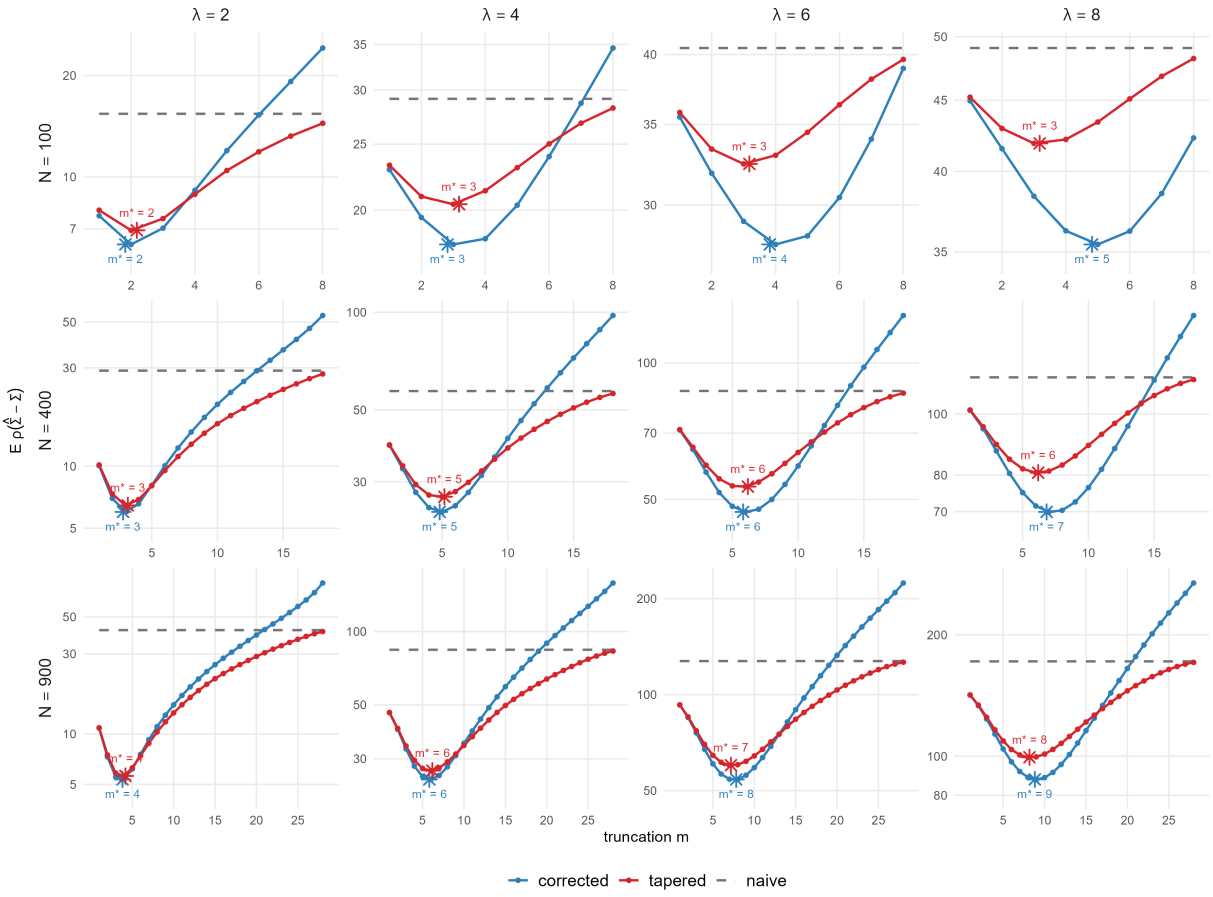


Figure 11: Average spectral norm difference of 10,000 Monte Carlo simulations vs different cutoff parameter m comparing the naive estimator $\widehat{\mathbf{\Sigma}}$ (gray), the bias-corrected estimator $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ (blue) and the tapered estimator $\widehat{\mathbf{\Sigma}}_{\underline{m}}$ (red)

4.2. Nonseparable case

Now, we examine the case where our true covariance from Equation 2.10, where we choose the parameter $\beta = 1$ leading to nonseparability of the covariance.

Figure 12 depicts the bias and the MSE for the bias-corrected estimator with standard prefactor with the competitor Cressie-Hawkins estimator and the naive estimator. Also here the bias correction is confirmed for almost all m cutoff values. The MSE level has a similar behavior to the separable

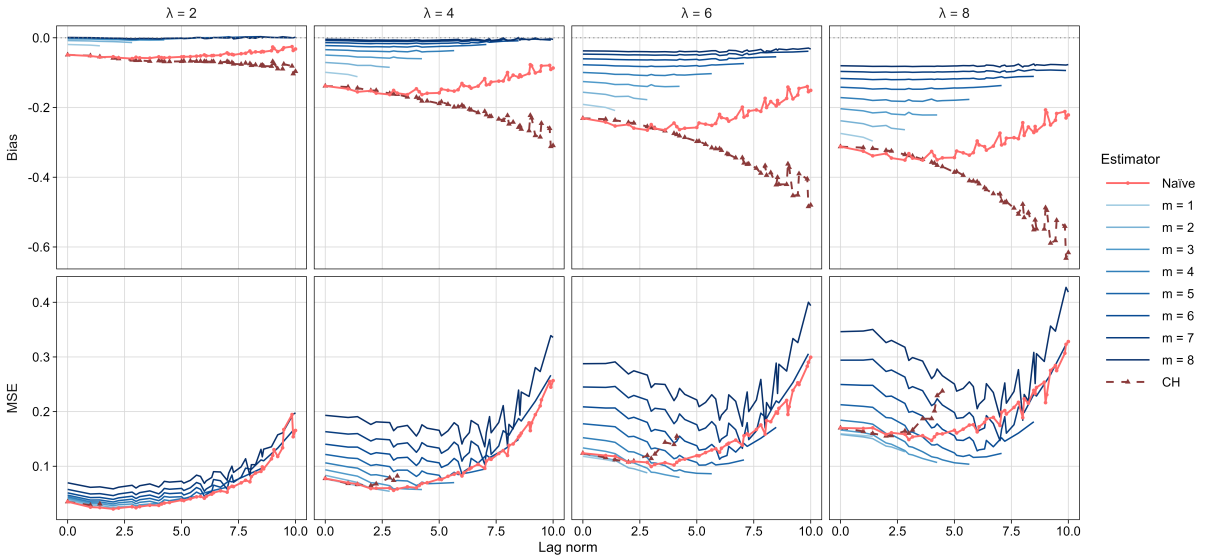


Figure 12: Bias and MSE performance with increasing lag norm $\|\underline{h}\|_2$ for the naive estimator $\widehat{\underline{C}}$ (solid light red line), Cressie-Hawkins $\widehat{\underline{C}}_{CH}$ estimator (dashed dark red line) and the bias-corrected estimator $\widehat{\underline{C}}_m^*$ for different m values (several blue shades) for four spatial dependence decay parameters $\lambda \in \{2, 4, 6, 8\}$ and for grid size $n_1 = n_2 = 10$. Nonseparable model $\beta = 1$.

case: at small levels of spatial dependence, already small values of m cause an inflation in MSE, which increases together with m, while as the spatial dependence increases low values of m obtain both a reduction of the bias and in the MSE. The effect of nonseparability becomes visible in Figure 13, where we lose the symmetry from the respective results in Figure 9, and a selection of lag pairs achieves a better MSE performance if we taper the estimation for that lag to zero. The bias correction method leads to better results with increasing strength of the spatial dependence. Note that, by the nature of the true covariance model (2.10), the nonseparable case generates overall a weaker dependence structure, because, for $\beta > 0$, we introduce a term that additionally supports the decay of the exponential function for increasing lags. This is also seen in Figure 7, where the nonseparable covariances (second row) appear in lighter shades compared to the respective separable ones (first row). Therefore, bearing this in mind, the results from the separable and non-separable case should be compared carefully. Nonetheless, we can see that the bias-correction achieves valuable results in the MSE sense, as reported in Figure 13. There, for the weakest dependence ($\lambda = 2$), the naive estimator is actually better than the bias-corrected one, but in the remaining three settings we see that the bias correction achieves MSE improvements for short lags, while covariances at large lags are indeed better tapered to zero. The non-separable nature of the covariance is manifested by the non-symmetric diagonal slant in the plot, where lag pairs with both positive or both negative lags can be estimated well with either the naive or bias corrected methods, while the cross-lags are better tapered early to zero as the lags increase.

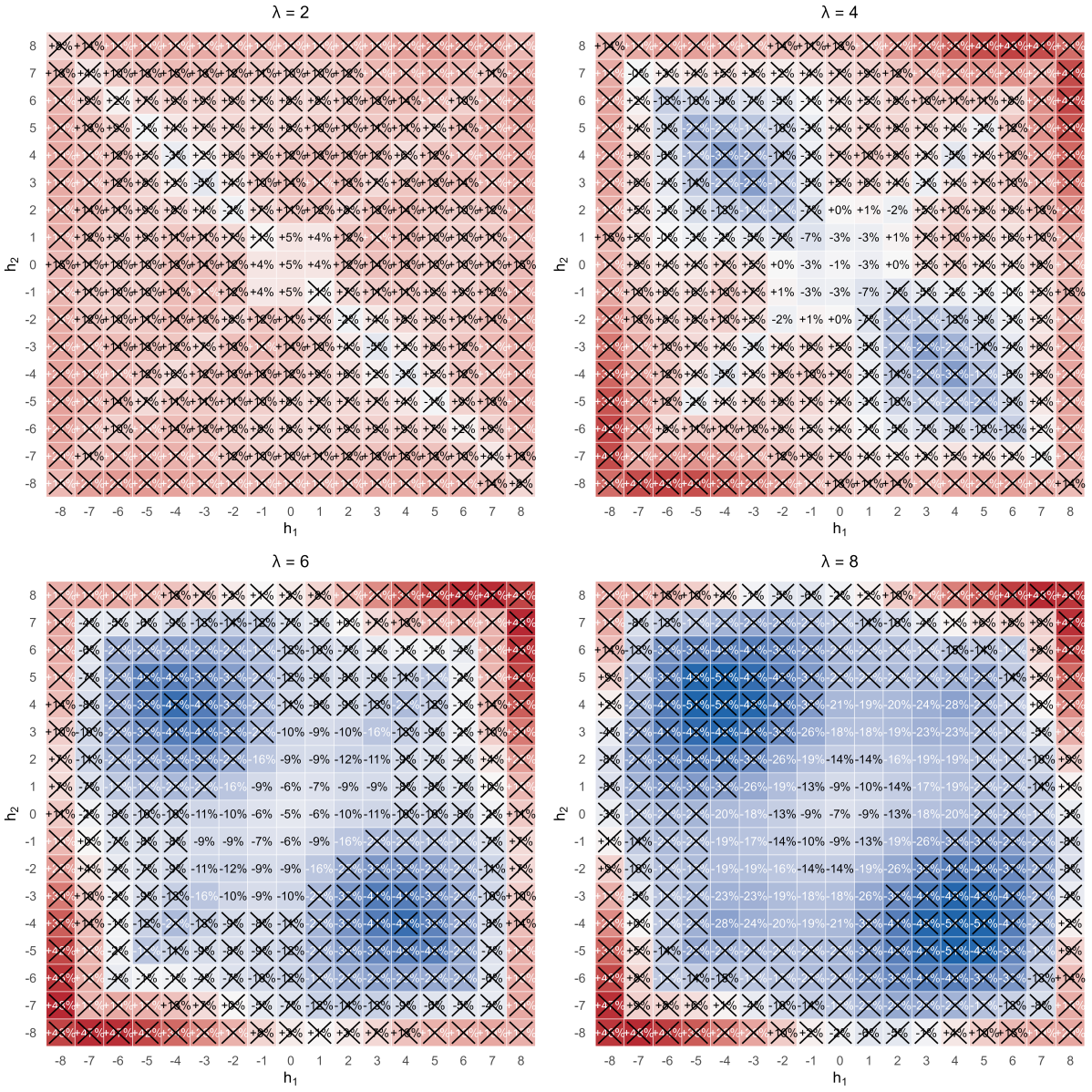


Figure 13: MSE ratio of the estimator with the optimal cutoff m for each lag pair (h_1, h_2) and naive estimator. Negative percentages indicate a reduction in MSE while positive percentages indicate an increase in MSE for the specific lag pair. Equivalently, blue shades indicate a reduction in MSE while red shades indicate a worsening of the MSE when compared to the naive estimator. Grid size (10×10). Nonseparable case $\beta = 1$.

5. Real data illustration: kriging on the topography of the Maungawhau volcano

We illustrate the proposed bias corrected covariance estimation from Equation (3.11) on a real data example, using it to drive a kriging-based prediction task on image data, a common type of lattice realization. The point of the exercise is not to claim a state-of-the-art interpolation method, but to show that correcting the finite-sample bias of the sample covariance on a lattice translates into a measurable improvement in a downstream task that depends entirely on the estimated covariance structure.

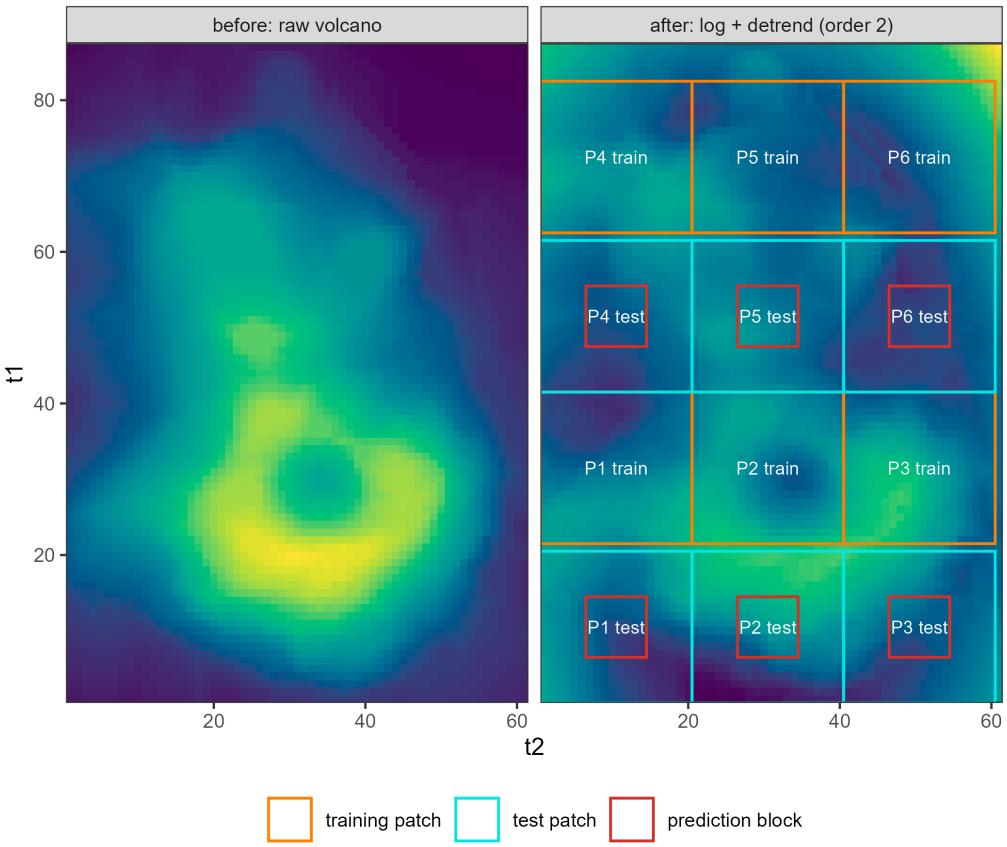


Figure 14: The `volcano` lattice before (left panel) and after (right panel) the log transform and the subtraction of the fitted order two polynomial surface.

We consider the `volcano` dataset, distributed with the `datasets` package in R (R Core Team, 2024). It records the topography of Maungawhau (Mount Eden), in Auckland, on a 87×61 lattice grid. As can be seen in the left panel of Figure 14, the elevation of the volcano is a cone, such that it does not meet the weak stationarity imposed by Assumption 2.1. Hence, we apply two transformations on the field, namely a log-transform and the detrending described in the following.

With the coordinates $\underline{t} = (t_1, t_2)$, $t_i = 1, \ldots, n_i$ of the lattice $G_{\underline{n}}$, being the process a record of the elevation, it is always positive $X_{\underline{t}} > 0$, thus it admits a log-transform $Y_{\underline{t}} = \log(X_{\underline{t}})$. This transformation is needed because the variability of the elevations grows with their level. Then, an order two polynomial with the coordinates as variables, that is,

$$\mu_{\underline{\beta}}(\underline{t}) := \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \beta_3 t_1^2 + \beta_4 t_2^2 + \beta_5 t_1 t_2,$$

is used and fitted by ordinary least squares for detrending the log-transformed data, such that we work with the residuals $\{y_{\underline{t}}\}$ obtained by

$$\widehat{\underline{\beta}} := \arg\min_{\underline{\beta}} \sum_{\underline{t}\in G_{\underline{n}}} \big(Y_{\underline{t}} - \mu_{\underline{\beta}}(\underline{t})\big)^2, \qquad y_{\underline{t}} := Y_{\underline{t}} - \mu_{\widehat{\underline{\beta}}}(\underline{t})$$

in the following.

Figure 14 shows the raw lattice data before (left panel) and the residuals after the two transformations (right panel). On the right panel, the lattice has been split into 6 training and 6 respective testing patches. We perform ordinary kriging for each patch pair. First, we estimate the variance covariance matrix $\mathbf{\Sigma}$ on each of the (20×20) training patches. We build four different estimators: two bias-corrected estimators $\widehat{\mathbf{\Sigma}}_{\underline{m}}$ from (3.26), one with cutoff value $\widehat{m}^{\text{rule}}$ obtained from the plug-in rule procedure and one with $\widehat{m}^{\text{cv}}$ obtained from the cross-validation procedure, both described in Section 3.7. Further, a naive estimator $\widehat{\mathbf{\Sigma}}_N$ obtained by plugging-in the standard estimator $\widehat{C}(\underline{h})$ from (2.11) in its entries, and $\widehat{\mathbf{\Sigma}}_{CH}$, which is obtained by plugging-in the Cressie-Hawkins estimator (4.2). All four estimators above undergo the spectral projection technique from (3.28) to secure a positive semidefinite covariance matrix, defined generally with $\widehat{\mathbf{\Sigma}}^+$ for the description of the ordinary kriging below.

In each testing patch, we split two disjoint sets: the set O of observed locations and the set P of locations to be predicted, indicated by the red square in Figure 14. The observed values are collected in the vector $\underline{y}_O$, while the central block values are collected in vector $\underline{y}_P$, and used only for the evaluation of the prediction. We reconfigure each estimator by the index sets O and P. Concretely, for all four estimators, $\widehat{\mathbf{\Sigma}}^+_{OO}$ represents the submatrix of rows and columns indexed by O and $\widehat{\mathbf{\Sigma}}^+_{PO}$ the submatrix of rows indexed by P and columns indexed by O. Ordinary kriging treats the process mean as an unknown constant, estimated by generalized least squares

$$\widehat{\mu} = (\underline{\mathbb{1}}^\top(\widehat{\mathbf{\Sigma}}^+_{OO})^{-1}\underline{\mathbb{1}})^{-1}\underline{\mathbb{1}}^\top(\widehat{\mathbf{\Sigma}}^+_{OO})^{-1}\underline{y}_O \tag{5.1}$$

and then used in the prediction of the held-out central block by ordinary kriging

$$\widehat{\underline{y}}_P = \widehat{\mu}\underline{\mathbb{1}} + \widehat{\mathbf{\Sigma}}^+_{PO}(\widehat{\mathbf{\Sigma}}^+_{OO})^{-1}\big(\underline{y}_O - \widehat{\mu}\underline{\mathbb{1}}\big), \tag{5.2}$$

for further details on the ordinary kriging methodology, see for example Sacks et al. (1989); Cressie (1993); Sherman (2010). The prediction accuracy is measured for each estimator $\widehat{\mathbf{\Sigma}}^+$ by the root mean squared prediction error (RMSPE)

$$\text{RMSPE}(\widehat{\mathbf{\Sigma}}^+) = \sqrt{\frac{1}{|P|}\sum_{t\in P}(\hat{y}_t - y_t)^2}. \tag{5.3}$$

The results are shown in Table 2. It appears that, for most patches, the bias-corrected estimator $\widehat{\mathbf{\Sigma}}^*_{\underline{m}}$ with cutoff value selected by the plug-in rule obtains the lowest prediction error, while in general both bias-corrected estimators compete well against the naive benchmark.

	full lag grid		cross-validated		plug-in rule	
patch	$\widehat{\mathbf{\Sigma}}_N^+$	$\widehat{\mathbf{\Sigma}}_{\text{CH}}^+$	$\widehat{m}^{\text{cv}}$	$\widehat{\mathbf{\Sigma}}_{\underline{m}}^{*+}$	$\widehat{m}^{\text{rule}}$	$\widehat{\mathbf{\Sigma}}_{\underline{m}}^{*+}$
P_1	0.031	0.030	8	0.020	5	0.019
P_2	0.044	0.064	8	0.025	4	0.030
P_3	0.044	0.011	9	0.028	5	0.018
P_4	0.031	0.045	5	0.023	4	0.022
P_5	0.072	0.034	9	0.031	4	0.020
P_6	0.125	0.081	8	0.037	5	0.017
avg	0.058	0.044	—	0.027	—	0.021

Table 2: RMSPE results of the ordinary kriging for each estimator considered. $\widehat{\mathbf{\Sigma}}_N^+$ standard estimator in its entries; $\widehat{\mathbf{\Sigma}}_{\text{CH}}^+$ Cressie-Hawkins estimator in its entries; $\widehat{\mathbf{\Sigma}}_{\underline{m}}^{*,+}$ bias-corrected estimator with cutoff selected by either plug-in rule or cross-validation.

The bias-corrected estimator is implemented in the `bclc` package that is available on Github (Fioravanti, 2026).

6. Conclusion

In this paper, we developed a framework for bias correction of sample covariance estimators for stationary lattice processes indexed in $\mathbb{Z}^2$. The starting point was the derivation of the exact finite-sample bias for the covariance estimator $\widehat{C}(h_1, h_2)$ at any single spatial lag (h_1, h_2) computed on a rectangular $(n_1 \times n_2)$ grid. When considering a subset of spatial sample covariances jointly, we derived how to construct a deterministic matrix-correction factor $\mathbf{W}_{\underline{n}}^*$ that is known and depends only on the grid dimensions. This weight matrix $\mathbf{W}_{\underline{n}}^*$ is shown to be singular, but the reduced weight matrix $\mathbf{W}_{\underline{n},m}^*$ obtained by extracting the central submatrix is argued to be invertible and the resulting bias-corrected estimator $\underline{\widehat{C}}_m^* = (\mathbf{W}_{\underline{n},m}^*)^{-1}\text{vec}(\widehat{\mathbf{\Gamma}}_m)$ becomes nearly unbiased for jointly estimating the corresponding covariances. We show that the correction applied to the standard estimator with prefactor $1/N_h$ and to the constant prefactor estimator with $1/N$ leads to exactly the same bias-corrected estimator. We also derive the finite sample formulas for the covariances of the estimators and establish asymptotic normality of our bias-corrected estimator, which is asymptotically equivalent to its uncorrected version. Also, we propose two data-driven methods to select the cutoff parameter m.

In Monte Carlo simulations, we study the bias-corrected estimator and compare it to the naive one and a benchmark variogram-based estimator proposed by Cressie and Hawkins (1980), which also attempts to mitigate the bias introduced by the sample covariance estimation. Our simulation focuses on the study of Gaussian lattice processes with an exponential covariance function with varying degrees of spatial dependence, covering both separable and non-separable structures. The bias-corrected estimator effectively manages to reduce the bias in many setups, especially when the spatial dependence is strong, which may be the practically more important case. In such setups, we also see that the reduction of the bias results also in a reduction of the MSE, when a conservative choice of the parameter m made. Finally, we applied the bias-corrected estimator in a ordinary kriging task, which showed good results in prediction compared to the naive counterpart.

Acknowledgments

This research was funded in the course of TRR 391 Spatio-temporal Statistics for the Transition of Energy and Transport (520388526) by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). In addition, the authors gratefully acknowledge the computing time provided on the Linux HPC cluster at TU Dortmund University (LiDO3), partially funded in the course of the Large-Scale Equipment Initiative by the German Research Foundation (DFG) as project 271512359.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude (Anthropic) and ChatGPT (OpenAI) in order to improve language and readability, check consistency of notation, assist with the review and debugging of the simulation code. After using these tools, the authors reviewed, verified and edited all content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bartlett, M.S., 1946. On the theoretical specification and sampling properties of autocorrelated time-series. Journal of the Royal Statistical Society Series B: Statistical Methodology 8, 27–41. URL: http://dx.doi.org/10.2307/2983611, doi:10.2307/2983611.

Bartlett, M.S., 1950. Periodogram analysis and continuous spectra. Biometrika 37, 1–16. URL: http://dx.doi.org/10.1093/biomet/37.1-2.1, doi:10.1093/biomet/37.1-2.1.

Bickel, P.J., Levina, E., 2008a. Covariance regularization by thresholding. The Annals of Statistics 36. URL: http://dx.doi.org/10.1214/08-AOS600, doi:10.1214/08-aos600.

Bickel, P.J., Levina, E., 2008b. Regularized estimation of large covariance matrices. The Annals of Statistics 36. URL: http://dx.doi.org/10.1214/009053607000000758, doi:10.1214/009053607000000758.

Bilchouris, A., Olenko, A., 2025. On nonparametric estimation of covariogram. Austrian Journal of Statistics 54, 112–137. URL: http://dx.doi.org/10.17713/ajs.v54i1.1975, doi:10.17713/ajs.v54i1.1975.

BjØrnstad, O.N., Falck, W., 2001. Nonparametric spatial covariance functions: Estimation and testing. Environmental and Ecological Statistics 8, 53–70. URL: http://dx.doi.org/10.1023/A:1009601932481, doi:10.1023/a:1009601932481.

Bolthausen, E., 1982. On the central limit theorem for stationary mixing random fields. The Annals of Probability 10. URL: http://dx.doi.org/10.1214/aop/1176993726, doi:10.1214/aop/1176993726.

Bradley, R.C., 2007. Introduction to Strong Mixing Conditions. Kendrick Press, Heber City, UT. Volumes 1–3.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods. Springer New York. URL: http://dx.doi.org/10.1007/978-1-4419-0320-4, doi:10.1007/978-1-4419-0320-4.

Cai, T.T., Zhang, C.H., Zhou, H.H., 2010. Optimal rates of convergence for covariance matrix estimation. The Annals of Statistics 38. URL: https://www.jstor.org/stable/20744485, doi:10.1214/09-aos752.

Cressie, N., 1985. Fitting variogram models by weighted least squares. Journal of the International Association for Mathematical Geology 17, 563–586. URL: http://dx.doi.org/10.1007/BF01032109, doi:10.1007/bf01032109.

Cressie, N., Hawkins, D.M., 1980. Robust estimation of the variogram: I. Journal of the International Association for Mathematical Geology 12, 115–125. URL: http://dx.doi.org/10.1007/BF01035243, doi:10.1007/bf01035243.

Cressie, N.A.C., 1993. Statistics for Spatial Data. Wiley. URL: http://dx.doi.org/10.1002/9781119115151, doi:10.1002/9781119115151.

Dahlhaus, R., Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random fields. Biometrika 74, 877–882. URL: http://dx.doi.org/10.1093/biomet/74.4.877, doi:10.1093/biomet/74.4.877.

Dedecker, J., Doukhan, P., Lang, G., Louhichi, S., Prieur, C., 2007. Weak Dependence: With Examples and Applications. volume 190 of *Lecture Notes in Statistics*. Springer, New York. URL: http://dx.doi.org/10.1007/978-0-387-69952-3, doi:10.1007/978-0-387-69952-3.

El Machkouri, M., Volný, D., Wu, W.B., 2013. A central limit theorem for stationary random fields. Stochastic Processes and their Applications 123, 1–14. URL: http://dx.doi.org/10.1016/j.spa.2012.08.014, doi:10.1016/j.spa.2012.08.014.

Fioravanti, A., 2026. bclc: Bias-corrected lattice covariance estimation. URL: https://github.com/antfioravanti/bclc. R code.

Furrer, R., Genton, M.G., Nychka, D., 2006. Covariance tapering for interpolation of large spatial datasets. Journal of Computational and Graphical Statistics 15, 502–523. URL: http://dx.doi.org/10.1198/106186006X132178, doi:10.1198/106186006x132178.

Golub, G.H., Van Loan, C.F., 2013. Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences. 4 ed., Johns Hopkins University Press, Baltimore, MD. URL: https://doi.org/10.56021/9781421407944, doi:10.56021/9781421407944.

Guillaumin, A.P., Sykulski, A.M., Olhede, S.C., Simons, F.J., 2022. The debiased spatial whittle likelihood. Journal of the Royal Statistical Society Series B: Statistical Methodology 84, 1526–1557. URL: http://dx.doi.org/10.1111/rssb.12539, doi:10.1111/rssb.12539.

Guttorp, P., Gneiting, T., 2006. Studies in the history of probability and statistics xlix on the matérn correlation family. Biometrika 93, 989–995. URL: http://dx.doi.org/10.1093/biomet/93.4.989, doi:10.1093/biomet/93.4.989.

Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69, 95–105. URL: http://dx.doi.org/10.1093/biomet/69.1.95, doi:10.1093/biomet/69.1.95.

Hall, P., Fisher, N.I., Hoffmann, B., 1994. On the nonparametric estimation of covariance functions. The Annals of Statistics 22, 2115–2134. URL: http://www.jstor.org/stable/2242501.

Isserlis, L., 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika 12, 134. URL: http://dx.doi.org/10.2307/2331932, doi:10.2307/2331932.

Jentsch, C., Politis, D.N., 2015. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. The Annals of Statistics 43. URL: http://dx.doi.org/10.1214/14-AOS1301, doi:10.1214/14-aos1301.

Kreiß, J.P., Neuhaus, G., 2006. Einführung in die Zeitreihenanalyse. Springer-Verlag. URL: http://dx.doi.org/10.1007/3-540-33571-4, doi:10.1007/3-540-33571-4.

Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management 30, 110–119. URL: http://dx.doi.org/10.3905/jpm.2004.110, doi:10.3905/jpm.2004.110.

Li, Y., 2023. Correcting the bias of the sample cross-covariance estimator. Journal of Time Series Analysis 45, 214–247. URL: http://dx.doi.org/10.1111/jtsa.12701, doi:10.1111/jtsa.12701.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146. URL: http://dx.doi.org/10.1093/biomet/71.1.135, doi:10.1093/biomet/71.1.135.

Matheron, G., 1962. Traité de géostatistique appliquée. volume 14. Editions Technip.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer New York. URL: http://dx.doi.org/10.1007/978-1-4612-1554-7, doi:10.1007/978-1-4612-1554-7.

R Core Team, 2024. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Ripley, B.D., 1984. Spatial statistics: Developments 1980-3, correspondent paper. International Statistical Review / Revue Internationale de Statistique 52, 141–150. URL: http://www.jstor.org/stable/1403097.

Sacks, J., Schiller, S.B., Welch, W.J., 1989. Designs for computer experiments. Technometrics 31, 41–47. URL: http://dx.doi.org/10.1080/00401706.1989.10488474, doi:10.1080/00401706.1989.10488474.

Schabenberger, O., Gotway, C.A., 2017. Statistical Methods for Spatial Data Analysis: Texts in Statistical Science. Chapman and Hall/CRC. URL: http://dx.doi.org/10.1201/9781315275086, doi:10.1201/9781315275086.

Sherman, M., 2010. Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties. Wiley. URL: http://dx.doi.org/10.1002/9780470974391, doi:10.1002/9780470974391.

Stein, M.L., Handcock, M.S., 1989. Some asymptotic properties of kriging when the covariance function is misspecified. Mathematical Geology 21, 171–190. URL: http://dx.doi.org/10.1007/BF00893213, doi:10.1007/bf00893213.

Vogelsang, T.J., Yang, J., 2016. Exactly/nearly unbiased estimation of autocovariances of a univariate time series with unknown mean. Journal of Time Series Analysis 37, 723–740. URL: http://dx.doi.org/10.1111/jtsa.12184, doi:10.1111/jtsa.12184.

Wang, Y., Ghosh, S.K., 2023. Nonparametric estimation of isotropic covariance function. Journal of Nonparametric Statistics 35, 198–237. URL: https://doi.org/10.1080/10485252.2022.2146111, doi:10.1080/10485252.2022.2146111.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association 99, 250–261. URL: http://dx.doi.org/10.1198/016214504000000241, doi:10.1198/016214504000000241.

Appendix A. Proof of Theorem 2.1

We adopt the following set notation for arbitrary integers $a, b \in \mathbb{Z}$ with $a \leq b$, we write

$$I(a:b) := \{a,\, a+1,\, a+2,\, \ldots,\, b-2,\, b-1,\, b\} \tag{A.1}$$

for the set of all integers from a to b. We let $I(a:b) = \emptyset$ when $a > b$. Hence, for all $a, b \in \mathbb{Z}$, the cardinality $|I(a:b)|$ of $I(a:b)$ becomes

$$\big|I(a:b)\big| = (b-a+1)^+, \quad \text{with} \quad x^+ := \max(x, 0), \quad \forall x \in \mathbb{Z}.$$

For integers $a_j \leq b_j$, $j = 1, \ldots J$ and for $x \in \mathbb{Z}$, we have that

$$x \in \bigcap_{j=1}^{J} I(a_j : b_j) \iff \max_{1\leq j\leq J} a_j \leq x \leq \min_{1\leq j\leq J} b_j \iff x \in I\Big(\max_j a_j : \min_j b_j\Big). \tag{A.2}$$

Then, the cardinality of the intersection in (A.2) is given by

$$\Big|\bigcap_{j=1}^{J} I(a_j : b_j)\Big| = \begin{cases} \min_j b_j - \max_j a_j + 1, & \max_j a_j \leq \min_j b_j, \\ 0, & \max_j a_j > \min_j b_j, \end{cases} = \Big(\min_j b_j - \max_j a_j + 1\Big)^+. \tag{A.3}$$

Lemma Appendix A.1 (Window count)**.** *Let $n \geq 1$, let $J \geq 1$ and let $d_1, \ldots, d_J \in \mathbb{Z}$ be arbitrary shifts. Then, we have*

$$\nu_n\{d_1, \ldots, d_J\} := \Big|\bigcap_{j=1}^{J} \big\{t \in \mathbb{Z} :\ t + d_j \in I(1:n)\big\}\Big| = \Big(n - \max_{j,j'\leq J} \big|d_j - d_{j'}\big|\Big)^+. \tag{A.4}$$

Proof. For the fixed j-th constraint given by the jth shift d_j, we have that

$$t + d_j \in I(1:n) \iff 1 \leq t + d_j \leq n \iff 1 - d_j \leq t \leq n - d_j \iff t \in I(1 - d_j : n - d_j)$$

so that the j-th constraint set, is itself a set of the form (A.1). The admissible t are therefore the elements that satisfy all J constraints simultaneously, thus we have an intersection of sets such as in (A.2) with $a_j = 1 - d_j$ and $b_j = n - d_j$,

$$\begin{aligned} \bigcap_{j=1}^{J} \big\{t \in \mathbb{Z} : t + d_j \in I(1:n)\big\} &= \bigcap_{j=1}^{J} I(1 - d_j : n - d_j) = I\Big(\max_{j\leq J}(1 - d_j) : \min_{j\leq J}(n - d_j)\Big) \\ &= I\Big(1 - \min_{j\leq J} d_j : n - \max_{j\leq J} d_j\Big). \end{aligned} \tag{A.5}$$

Then, by (A.3) the cardinality of (A.5) is

$$\begin{aligned} \Big|\bigcap_{j=1}^{J} \big\{t \in \mathbb{Z} : t + d_j \in I(1:n)\big\}\Big| &= \Big(\big(n - \max_{j\leq J} d_j\big) - \big(1 - \min_{j\leq J} d_j\big) + 1\Big)^+ \\ &= \Big(n - \big(\max_{j\leq J} d_j - \min_{j\leq J} d_j\big)\Big)^+ = \Big(n - \max_{j,j'\leq J} \big|d_j - d_{j'}\big|\Big)^+ \end{aligned} \tag{A.6}$$

□

Remark Appendix A.1 (Window count on a 2-dimensional lattice). For shifts $\underline{d}_1, \ldots, \underline{d}_J \in \mathbb{Z}^2$, with $\underline{d}_j = (d_{j,1}, d_{j,2})$, let

$$\nu_n\{\underline{d}_1, \ldots, \underline{d}_J\} := \Big|\bigcap_{j=1}^{J} \{\underline{t} \in \mathbb{Z}^2 : \ \underline{t} + \underline{d}_j \in G_{\underline{n}}\}\Big|. \tag{A.7}$$

where $G_{\underline{n}} = I(1:n_1) \times I(1:n_2)$. Since the membership of $\underline{t} + \underline{d}_j \in G_{\underline{n}}$ means that we impose one constraint per axis, and the constraints of the two axes involve disjoint coordinates $\{1, \ldots, n_1\} \times \{1, \ldots, n_2\}$. The intersection in (A.7) therefore factorizes into a Cartesian product,

$$\nu_n\{\underline{d}_1, \ldots, \underline{d}_J\} = \prod_{i=1}^{2} \nu_{n_i}\{d_{1,i}, \ldots, d_{J,i}\} = \prod_{i=1}^{2} \Big(n_i - \max_{j,j' \le J} |d_{j,i} - d_{j',i}|\Big)^{+}. \tag{A.8}$$

Proof of Theorem 2.1. From Assumption 2.1, $\{X_{\underline{t}} : \underline{t} \in \mathbb{Z}^2\}$ is a second-order stationary random field with mean $\mu = \mathbb{E}[X_{\underline{t}}]$ and covariance function $C(\underline{h}) = \mathrm{Cov}(X_{\underline{t}+\underline{h}}, X_{\underline{t}})$, which only depends on the lag/shift and not on the coordinates on the $(n_1 \times n_2)$ grid $G_{\underline{n}}$, where

$$G_{\underline{n}} = \{\underline{t} \in \mathbb{Z}^2 : 1 \le t_i \le n_i, \ i = \{1,2\}\} = I(1:n_1) \times I(1:n_2), \quad \text{with} |G_{\underline{n}}| = N = n_1 n_2,$$

while the set of all admissible lags is $H_{\underline{n}}$ defined by

$$H_{\underline{n}} = \{\underline{h} \in \mathbb{Z}^2 : |h_i| \le n_i - 1, \ i = \{1,2\}\}.$$

The subset $G_{\underline{n}}(\underline{h})$ represents the set of coordinates $\underline{t}$ that admit the shift by a lag $\underline{h} \in H_{\underline{n}}$, that is,

$$G_{\underline{n}}(\underline{h}) := \big\{\underline{t} \in G_{\underline{n}} : \underline{t} + \underline{h} \in G_{\underline{n}}\big\} \subseteq G_{\underline{n}}, \quad |G_{\underline{n}}(\underline{h})| = N_h = \prod_{i=1}^{2} (n_i - |h_i|).$$

More generally, for an arbitrary shift $\underline{k}$, we write $G_{\underline{n}}(\underline{k}) := \{\underline{t} \in G_{\underline{n}} : \underline{t} + \underline{k} \in G_{\underline{n}}\}$, with the case that $G_{\underline{n}}(\underline{0}) = G_{\underline{n}}$ and by Remark Appendix A.1, $|G_{\underline{n}}(\underline{k})| = \nu_{\underline{n}}\{\underline{0}, \underline{k}\}$.

The standard sample covariance estimator at lag $\underline{h} \in H_{\underline{n}}$ by

$$\widehat{C}(\underline{h}) = \frac{1}{N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} (X_{\underline{t}+\underline{h}} - \overline{X})(X_{\underline{t}} - \overline{X})$$

with the sample mean defined

$$\overline{X} = \frac{1}{N} \sum_{\underline{u} \in G_{\underline{n}}} X_{\underline{u}} = \frac{1}{N} \sum_{u_1=1}^{n_1} \sum_{u_2=1}^{n_2} X_{u_1,u_2}$$

Now we expand the product and take expectations

$$\begin{aligned} \mathbb{E}\big[\widehat{C}(\underline{h})\big] \ = \ & \underbrace{\frac{1}{N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \mathbb{E}\big[X_{\underline{t}+\underline{h}} X_{\underline{t}}\big]}_{=:\mathrm{C}} - \underbrace{\frac{1}{N N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{u} \in G_{\underline{n}}} \mathbb{E}\big[X_{\underline{t}+\underline{h}} X_{\underline{u}}\big]}_{=:\mathrm{A}} \\ & \underbrace{- \frac{1}{N_h N} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{u} \in G_{\underline{n}}} \mathbb{E}\big[X_{\underline{t}} X_{\underline{u}}\big]}_{=:\mathrm{B}} + \underbrace{\frac{1}{N^2} \sum_{\underline{u},\underline{v} \in G_{\underline{n}}} \mathbb{E}\big[X_{\underline{u}} X_{\underline{v}}\big]}_{=:\mathrm{D}}. \end{aligned} \tag{A.9}$$

Term C is a double summation and is immediate: since the summand $\mathbb{E}[X_{\underline{t}+\underline{h}}\, X_{\underline{t}}]$ is constant in $\underline{t}$, there are exactly N_h terms, which cancel with the prefactor $1/N_h$ of the summation and we obtain

$$\mathrm{C} = C(\underline{h}) + \mu^2.$$

The inner sum over $\underline{u}$ is $N\overline{X}$, so

$$\mathrm{A} = \frac{1}{N_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \mathbb{E}\big[X_{\underline{t}+\underline{h}}\overline{X}\big], \quad \mathrm{B} = \frac{1}{N_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \mathbb{E}\big[X_{\underline{t}}\overline{X}\big], \quad \mathrm{D} = \mathbb{E}\big[\overline{X}^{\,2}\big]. \tag{A.10}$$

Using $\mathbb{E}[X_{\underline{p}}\overline{X}] = \mathrm{Cov}(X_{\underline{p}}, \overline{X}) + \mu^2$ and $\mathbb{E}[\overline{X}^2] = \mathrm{Var}(\overline{X}) + \mu^2$, we have that each of the four terms in (A.9) has exactly one μ^2 such that they cancel. Hence, $\mathbb{E}[\widehat{C}(\underline{h})]$ does not depend on μ. Then, by subtracting $C(\underline{h})$, term C cancels and we obtain

$$b(\underline{h}) = \mathbb{E}\big[\widehat{C}(\underline{h})\big] - C(\underline{h}) = -\frac{1}{N_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \big[\mathrm{Cov}(X_{\underline{t}+\underline{h}}, \overline{X}) + \mathrm{Cov}(X_{\underline{t}}, \overline{X})\big] + \mathrm{Var}(\overline{X}), \tag{A.11}$$

which is the bias representation (2.21).

To get also the reduced form, note that terms A, B and D are fourfold summations. We want to exploit the stationarity property to reduce them to a double summation over a new lag index $\underline{r} \in H_{\underline{n}}$. However, in practice, we need to count how many times the new lag $\underline{r}$ appears in the new double summation. We proceed term by term.

Term A. By stationarity,

$$\mathrm{A} - \mu^2 = \frac{1}{NN_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \sum_{\underline{u}\in G_{\underline{n}}} \mathbb{E}\big[X_{\underline{t}+\underline{h}} X_{\underline{u}}\big] = \frac{1}{NN_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \sum_{\underline{u}\in G_{\underline{n}}} C\big(\underline{u} - \underline{t} - \underline{h}\big). \tag{A.12}$$

We apply a change of argument by fixing $\underline{t}$ and setting

$$\underline{r} := \underline{u} - \underline{t} - \underline{h} \quad \Longleftrightarrow \quad \underline{u} = \underline{t} + \underline{h} + \underline{r} \quad \Longrightarrow \quad \big(\underline{u} \in G_{\underline{n}} \iff \underline{t} + \underline{h} + \underline{r} \in G_{\underline{n}}\big).$$

The site $\underline{t} + \underline{h} + \underline{r}$ is $\underline{u}$ rewritten, namely $\underline{u} \mapsto \underline{r}$ is a bijection of $G_{\underline{n}}$ onto $\{\underline{r} : \underline{t} + \underline{h} + \underline{r} \in G_{\underline{n}}\}$, so the inner sum of (A.12) is

$$\sum_{\underline{u}\in G_{\underline{n}}} C\big(\underline{u} - \underline{t} - \underline{h}\big) = \sum_{\underline{r}\in\mathbb{Z}^2} C(\underline{r}) \mathbb{1}\{\underline{t} + \underline{h} + \underline{r} \in G_{\underline{n}}\}.$$

Exchanging the two summations in (A.12), for one with $\underline{r}$ as index,

$$\sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \sum_{\underline{u}\in G_{\underline{n}}} C\big(\underline{u} - \underline{t} - \underline{h}\big) = \sum_{\underline{r}\in\mathbb{Z}^2} C(\underline{r}) \big|\big\{\underline{t} : \underbrace{\underline{t} \in G_{\underline{n}}, \underline{t} + \underline{h} \in G_{\underline{n}}}_{\underline{t} \in G_{\underline{n}}(\underline{h})}, \underbrace{\underline{t} + \underline{h} + \underline{r} \in G_{\underline{n}}}_{\underline{u}\in G_{\underline{n}}}\big\}\big|. \tag{A.13}$$

The three constraints in (A.13) are the shifts $\underline{0}, \underline{h}, \underline{h} + \underline{r}$ of $\underline{t}$ and we apply the count function from Lemma Appendix A.1. By Remark Appendix A.1, the cardinality factorizes on the axis dimensions and on axis i the shifts $0, h_i, h_i + r_i$ have pairwise absolute distance $|h_i|, |h_i + r_i|, |r_i|$, such that Lemma Appendix A.1 gives

$$\nu_{\underline{n}}\{\underline{0}, \underline{h}, \underline{h} + \underline{r}\} = \prod_{i=1}^{2} \big(n_i - \max(|r_i|, |h_i|, |r_i + h_i|)\big)^+ =: \prod_{i=1}^{2} \nu_{n_i}(r_i, h_i), \tag{A.14}$$

which vanishes as soon as $|r_i| \geq n_i$; the sum over $\underline{r}$ is therefore a sum over $H_{\underline{n}}$. Combining (A.12)–(A.14)

$$\frac{1}{N_h} \sum_{\underline{t}\in G_{\underline{n}}(\underline{h})} \mathrm{Cov}\big(X_{\underline{t}+\underline{h}}, \overline{X}\big) = \frac{1}{NN_h} \sum_{\underline{r}\in H_{\underline{n}}} C(\underline{r}) \prod_{i=1}^{2} \nu_{n_i}(r_i, h_i) = \frac{S(\underline{h})}{NN_h}. \tag{A.15}$$

Term B. By stationarity,

$$\mathrm{B} - \mu^2 = \frac{1}{N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \mathrm{Cov}\big(X_{\underline{t}}, \overline{X}\big) = \frac{1}{N N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{u} \in G_{\underline{n}}} C\big(\underline{u} - \underline{t}\big). \tag{A.16}$$

The term differs from (A.12) only in being indexed by $\underline{t}$ instead of $\underline{t} + \underline{h}$. Thus, with $\underline{r} = \underline{u} - \underline{t}$, it follows in the same steps as before, that

$$\nu_{\underline{n}}\{\underline{0}, \underline{h}, \underline{r}\} = \prod_{i=1}^{2} \big(n_i - \max(|r_i|, |h_i|, |r_i - h_i|)\big)^+ = \prod_{i=1}^{2} \nu_{n_i}(r_i, -h_i), \tag{A.17}$$

where the last equality is the definition (A.14) of ν_{n_i} evaluated at $-h_i$. The weight again vanishes as soon as $|r_i| \geq n_i$, so the sum over $\underline{r}$ is a sum over $H_{\underline{n}}$, thus we get

$$\frac{1}{N_h} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \mathrm{Cov}\big(X_{\underline{t}}, \overline{X}\big) = \frac{1}{N N_h} \sum_{\underline{r} \in H_{\underline{n}}} C(\underline{r}) \prod_{i=1}^{2} \nu_{n_i}(r_i, -h_i) = \frac{S(-\underline{h})}{N N_h}. \tag{A.18}$$

Term D. By stationarity,

$$\mathrm{D} - \mu^2 = \mathrm{Var}(\overline{X}) = \frac{1}{N^2} \sum_{\underline{v} \in G_{\underline{n}}} \sum_{\underline{u} \in G_{\underline{n}}} C\big(\underline{u} - \underline{v}\big). \tag{A.19}$$

Here both summation sites are free in $G_{\underline{n}}$ and there is no lag $\underline{h}$ to respect. In a similar fashion, with $\underline{r} = \underline{u} - \underline{v}$, we obtain

$$\nu_{\underline{n}}\{\underline{0}, \underline{r}\} = \prod_{i=1}^{2} \big(n_i - |r_i|\big)^+ = \prod_{i=1}^{2} \nu_{n_i}(r_i, 0), \tag{A.20}$$

which is $N_{\underline{r}}$ for $\underline{r} \in H_{\underline{n}}$, the middle equality being the definition (A.14) at $h_i = 0$. Hence, combining (A.19)–(A.20), leads to

$$\mathrm{Var}(\overline{X}) = \frac{1}{N^2} \sum_{\underline{r} \in H_{\underline{n}}} C(\underline{r}) \prod_{i=1}^{2} \nu_{n_i}(r_i, 0) = \frac{S(\underline{0})}{N^2}. \tag{A.21}$$

The three weights obtained in (A.14), (A.17) and (A.20) are the same counting factor evaluated at h_i, at $-h_i$ and at 0, so we define the counting factor with the arguments the new index r_i and the lag h_i

$$\nu_{n_i}(r_i, h_i) := \big(n_i - \max(|r_i|, |h_i|, |r_i + h_i|)\big)^+, \quad i \in \{1, 2\}, \tag{A.22}$$

which is (2.24). (A.15), (A.18) and (A.21) are $S(\underline{h})$, $S(-\underline{h})$ and $S(\underline{0})$ as defined in (2.23). Inserting them into the bias representation (A.11) gives

$$b(\underline{h}) = -\left[\frac{S(\underline{h})}{N N_h} + \frac{S(-\underline{h})}{N N_h}\right] + \frac{S(\underline{0})}{N^2} = -\frac{S(\underline{h}) + S(-\underline{h})}{N N_h} + \frac{S(\underline{0})}{N^2}, \tag{A.23}$$

which is (2.22) and completes the proof. □

Proof of Corollary 2.1. The proof follows by applying a simple prefactor change N_h/N to connect the expectations of the two estimators and the definition of the bias of the standard estimator in Proposition 2.1, such that

$$\frac{N_h}{N} \mathbb{E}\left(\widehat{C}(\underline{h})\right) = \mathbb{E}\left(\widetilde{C}(\underline{h})\right) = \frac{N_h}{N} \left(C(\underline{h}) + b(\underline{h})\right) \tag{A.24}$$

Subtracting $C(\underline{h})$ on both sides of the equation gives (2.25). □

Appendix B. Proof of Theorem 2.2

Before proving Theorem 2.2, we need the following lemma.

Lemma Appendix B.1 (Entrywise bounds on the bias weights)**.** *For all $\underline{h}, \underline{r} \in H_{\underline{n}}$,*

$$\left|\left[\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right]_{\underline{h},\underline{r}}\right| \le \frac{1}{N}. \tag{B.1}$$

The bound is attained at $\underline{r} = \underline{0}$ where $\left[\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right]_{\underline{h},\underline{0}} = -1/N$ for every $\underline{h} \in H_{\underline{n}}$, so that

$$\left\|\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right\|_{\max} = \frac{1}{N}. \tag{B.2}$$

Proof. Fix $\underline{h}, \underline{r} \in H_{\underline{n}}$ so that $|h_i|, |r_i| \le n_i - 1$ for $i \in \{1, 2\}$. In particular, as $n_i - |h_i| \ge 1$, all denominators below are positive, and $\big(n_i - |r_i|\big)^+ = n_i - |r_i|$. For $i \in \{1, 2\}$, we define the entries of the normalized counting matrices as

$$a_i := [\mathbf{A}^*_i]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, h_i)}{n_i(n_i - |h_i|)}, \quad b_i := [\mathbf{B}^*_i]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, -h_i)}{n_i(n_i - |h_i|)}, \quad d_i := [\mathbf{D}^*_i]_{h_i,r_i} = \frac{\nu_{n_i}(r_i, 0)}{n_i^2}, \tag{B.3}$$

where $\nu_{n_i}(r_i, h_i) = \big(n_i - \max(|r_i|, |h_i|, |r_i + h_i|)\big)^+$ by (2.24). By the definition of the weight matrix $\mathbf{W}^*_{\underline{n}}$ from (2.32), the entries of interest are

$$\left[\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right]_{\underline{h},\underline{r}} = -a_1 a_2 - b_1 b_2 + d_1 d_2. \tag{B.4}$$

We use two elementary properties of a_i, b_i, d_i, namely

$$0 \le a_i,\, b_i,\, d_i \le \frac{1}{n_i} \quad \text{and} \quad \min(a_i, b_i) \le d_i \tag{B.5}$$

that are argued in the following. By construction, each element is a normalized count thus it is nonnegative and also $\nu_{n_i}(r_i, \pm h_i) \le n_i - |h_i|$, which gives $a_i, b_i \le 1/n_i$. While, $\nu_{n_i}(r_i, 0) = n_i - |r_i| \le n_i$ gives $d_i \le 1/n_i$. This shows the first part of (B.5).

For the second part note that for any $r, h \in \mathbb{Z}$,

$$\max\big(|r + h|, |r - h|\big) = |r| + |h| \implies \max\big(|r|, |h|, |r + h|, |r - h|\big) = |r| + |h|. \tag{B.6}$$

Since $\min(x^+, y^+) = (\min(x, y))^+$, the minimum between the two counting factors is

$$\begin{aligned}
\min\big(\nu_{n_i}(r_i, h_i), \nu_{n_i}(r_i, -h_i)\big) &= \min\Big(\big(n_i - \max\{|r_i|, |h_i|, |r_i + h_i|\}\big)^+, \big(n_i - \max\{|r_i|, |h_i|, |r_i - h_i|\}\big)^+\Big) \\
&= \big(n_i - \max\{|r_i|, |h_i|, |r_i + h_i|, |r_i - h_i|\}\big)^+ \\
&= \big(n_i - |r_i| - |h_i|\big)^+.
\end{aligned}$$

using (B.6) in the last step. We want to show that

$$\min(a_i, b_i) = \frac{(n_i - |r_i| - |h_i|)^+}{n_i(n_i - |h_i|)} \le \frac{n_i - |r_i|}{n_i^2} = d_i. \tag{B.7}$$

If $n_i - |r_i| - |h_i| < 0$ the left-hand side vanishes and the inequality is trivial. Otherwise, multiplying both sides by $n_i^2(n_i - |h_i|) > 0$, we have $n_i\big(n_i - |r_i| - |h_i|\big) \le (n_i - |r_i|)(n_i - |h_i|)$, and taking their difference

$$(n_i - |r_i|)(n_i - |h_i|) - n_i\big(n_i - |r_i| - |h_i|\big) = |r_i||h_i| \ \ge\ 0,$$

which proves (B.7) and thus the second part of (B.5).
Upper Bound. As $a_1a_2,\ b_1b_2 \geq 0$, we obtain the upper bound

$$\left[\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right]_{\underline{h},\underline{r}} \leq d_1d_2 \leq \frac{1}{n_1n_2} = \frac{1}{N}. \tag{B.8}$$

Lower Bound. From (B.4), the lower bound $-a_1a_2 - b_1b_2 + d_1d_2 \geq -\frac{1}{N}$ is equivalent to

$$a_1a_2 + b_1b_2 \leq \frac{1}{N} + d_1d_2, \tag{B.9}$$

By the second part of (B.5), for each i at least one of a_i, b_i is bounded by d_i. Since (B.9) is invariant under interchanging (a_1, a_2) and (b_1, b_2) this leaves the following two cases.

(i) Suppose that $a_1 \leq d_1$ and $a_2 \leq d_2$. Then $a_1a_2 \leq d_1d_2$, whereas $b_i \leq 1/n_i$, hence $b_1b_2 \leq 1/(n_1n_2) = 1/N$, which gives (B.9) The case $b_1 \leq d_1$ and $b_2 \leq d_2$ follows in the same way.

(ii) Suppose that $a_1 \leq d_1$ and $b_2 \leq d_2$. Since $a_2 \leq 1/n_2$ and $b_1 \leq 1/n_1$, respectively,

$$a_1a_2 \leq d_1a_2 \leq \frac{d_1}{n_2}, \quad b_1b_2 \leq b_1d_2 \leq \frac{d_2}{n_1},$$

so that it suffices to show

$$a_1a_2 + b_1b_2 \leq \frac{d_1}{n_2} + \frac{d_2}{n_1} \leq d_1d_2 + \frac{1}{N}.$$

Since $d_i = (n_i - |r_i|)/n_i^2$ and $N = n_1n_2$, multiplication by the positive quantity $n_1^2n_2^2$ transforms this inequality into

$$n_2(n_1 - |r_1|) + n_1(n_2 - |r_2|) \leq n_1n_2 + (n_1 - |r_1|)(n_2 - |r_2|).$$

The difference between the right-hand and left-hand sides is $|r_1||r_2| \geq 0$. The case $b_1 \leq d_1$ and $a_2 \leq d_2$ is identical.

Hence (B.9) holds in both cases, which together with (B.8) proves (B.1)..
For the attainment, let $\underline{r} = \underline{0}$. Then, for $i \in \{1, 2\}$,

$$\nu_{n_i}(0, h_i) = \nu_{n_i}(0, -h_i) = n_i - |h_i|, \qquad \nu_{n_i}(0, 0) = n_i,$$

so that $a_i = b_i = d_i = 1/n_i$ and

$$\left[\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\right]_{\underline{h},\underline{0}} = -\frac{1}{N} - \frac{1}{N} + \frac{1}{N} = -\frac{1}{N}, \quad \underline{h} \in H_{\underline{n}}$$

so the bound (B.1) is attained. This proves (B.2). □

Proof of Theorem2.2. While the first part of Theorem 2.2 are direct consequences of the formulas derived before stating the results, it remains to show the rank deficiency of $\mathbf{W}^*_{\underline{n}}$ and the bound in the following. For this purpose, we begin by noting a few properties of the counting matrices

$$\mathbf{A}_i, \mathbf{B}_i, \mathbf{D}_i \in \mathbb{R}^{L_i \times L_i}, \quad \text{with} \quad L_i = 2n_i - 1$$

and their normalized counterparts

$$\mathbf{A}^*_i = \mathbf{N}^{-1}_{h,i}\mathbf{A}_i, \qquad \mathbf{B}^*_i = \mathbf{N}^{-1}_{h,i}\mathbf{B}_i, \qquad \mathbf{D}^*_i = n_i^{-2}\mathbf{D}_i, \quad \text{for } i = 1, 2$$

where

$$\mathbf{N}_{h,i} := \operatorname{diag}\big(N_{h,1,i}, \ldots, N_{h,L_i,i}\big), \qquad N_{h,p,i} = n_i\big(n_i - |p - n_i|\big).$$

First, as they are counting matrices, $\mathbf{A}_i, \mathbf{B}_i$ (respectively $\mathbf{A}_i^*, \mathbf{B}_i^*$) have non-negative entries by construction. Additionally, $\mathbf{D}_i$ (respectively $\mathbf{D}_i^*$) has only strictly positive entries.

By construction, the p-th row–sum of $\mathbf{A}_i$ and $\mathbf{B}_i$ equals $n_i(n_i - |p - n_i|$, while the p-th row sum of $\mathbf{D}_i$ is n_i^2. This means that the normalization matrices $\mathbf{A}_i^*, \mathbf{B}_i^*, \mathbf{D}_i^*$ have row sums equal to 1 and thus are row–stochastic. Hence

$$\mathbf{A}_i^* \underline{\mathbb{1}}_i = \underline{\mathbb{1}}_i, \qquad \mathbf{B}_i^* \underline{\mathbb{1}}_i = \underline{\mathbb{1}}_i, \qquad \mathbf{D}_i^* \underline{\mathbb{1}}_i = \underline{\mathbb{1}}_i,$$

where $\underline{\mathbb{1}}_i = (1, 1, \ldots, 1)$ is the $L_i \times 1$ vector of ones. We define the weight matrix $\mathbf{W}^*$ as a linear mapping

$$W^* : \mathbb{R}^{L_1 \times L_2} \to \mathbb{R}^{L_1 \times L_2}, \quad W^*(\mathbf{X}) = \mathbf{X} - \mathbf{A}_1^* \mathbf{X} \mathbf{A}_2^{*\top} - \mathbf{B}_1^* \mathbf{X} \mathbf{B}_2^{*\top} + \mathbf{D}_1^* \mathbf{X} \mathbf{D}_2^{*\top}.$$

In particular, for the all-ones matrix $\underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top$ of size $L_1 \times L_2$,

$$\begin{aligned}
\mathbf{A}^* \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top \mathbf{A}^{*\top} &= \big(\mathbf{A}^* \underline{\mathbb{1}}_1\big)\big(\mathbf{A}^* \underline{\mathbb{1}}_2\big)^\top = \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top \\
\mathbf{B}^* \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top \mathbf{B}^{*\top} &= \big(\mathbf{B}^* \underline{\mathbb{1}}_1\big)\big(\mathbf{B}^* \underline{\mathbb{1}}_2\big)^\top = \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top \\
\mathbf{D}^* \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top \mathbf{D}^{*\top} &= \big(\mathbf{D}^* \underline{\mathbb{1}}_1\big)\big(\mathbf{D}^* \underline{\mathbb{1}}_2\big)^\top = \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top
\end{aligned}$$

so we have

$$W^*(\underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top) = \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top - \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top - \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top + \underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top = \mathbf{0}.$$

Thus, $\operatorname{span}\big(\underline{\mathbb{1}}_1 \underline{\mathbb{1}}_2^\top\big)$ is in the nullspace of the linear operator W^* and we have $\dim(\ker\big(W^*\big)) \geq 1$. Hence,

$$\operatorname{rank}\big(\mathbf{W}^*\big) = L_{\underline{n}} - \dim \ker\,(W^*) \implies \operatorname{rank}\big(\mathbf{W}^*\big) \leq L_{\underline{n}} - 1.$$

Finally, by

$$b_{\operatorname{vec}(\widehat{\mathbf{\Gamma}})} := \mathbb{E}\big(\operatorname{vec}(\widehat{\mathbf{\Gamma}})\big) - \operatorname{vec}(\mathbf{\Gamma}) = \big(\mathbf{W}_{\underline{n}}^* - \mathbf{I}_{L_{\underline{n}}}\big)\operatorname{vec}(\mathbf{\Gamma}),$$

and Lemma Appendix B.1, we apply the Hölder inequality[3] such that

$$\big\|b_{\operatorname{vec}(\widehat{\mathbf{\Gamma}})}\big\|_\infty = \big\|\big(\mathbf{W}_{\underline{n}}^* - \mathbf{I}_{L_{\underline{n}}}\big)\operatorname{vec}(\mathbf{\Gamma})\big\|_\infty \leq \big\|\mathbf{W}_{\underline{n}}^* - \mathbf{I}_{L_{\underline{n}}}\big\|_{\max} \big\|\operatorname{vec}(\mathbf{\Gamma})\big\|_1 = \frac{1}{N}\big\|\operatorname{vec}(\mathbf{\Gamma})\big\|_1,$$

which is the bound (2.34). Under Assumption 2.2 it is bounded uniformly.

□

Proof of Corollary 2.2. The scaling (A.24) $N_h/N\mathbb{E}\left(\widehat{C}(\underline{h})\right) = \mathbb{E}\left(\widetilde{C}(\underline{h})\right)$ holds entrywise, i.e., $\mathbb{E}(\widetilde{\mathbf{\Gamma}})_{p,q} = N_h/N\mathbb{E}(\widehat{\mathbf{\Gamma}})_{p,q}$. By the definition of the Bartlett window matrices

$$\mathbf{\Delta}_i := \operatorname{diag}\left(1 - \frac{|p - n_i|}{n_i},\ p = 1, \ldots, L_i\right) = \operatorname{diag}\left(1 - \frac{|h_i|}{n_i},\ h_i \in I(1 - n_i : n_i - 1)\right), \quad i \in \{1, 2\},$$

where we have the factorization

$$\frac{N_h}{N} = \frac{n_1 - |h_1|}{n_1} \cdot \frac{n_2 - |h_2|}{n_2} = [\mathbf{\Delta}_1]_{p,p}\,[\mathbf{\Delta}_2]_{q,q}\,, \quad (h_1, h_2) = (p - n_1, q - n_2), \tag{B.10}$$

[3] Row-wise Hölder inequality $\|\mathbf{M}v\|_\infty \leq \|\mathbf{M}\|_{\max}\|v\|_1$

we obtain

$$\mathbb{E}(\widetilde{\mathbf{\Gamma}}) = \mathbf{\Delta}_1 \mathbb{E}(\widehat{\mathbf{\Gamma}}) \mathbf{\Delta}_2.$$

By vectorizing the above and using the identity $\mathrm{vec}(\mathbf{\Delta}_1 \mathbf{X} \mathbf{\Delta}_2) = (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1)\mathrm{vec}(\mathbf{X})$ as well as $\mathbb{E}(\mathrm{vec}(\widehat{\mathbf{\Gamma}})) = \mathbf{W}_{\underline{n}}^* \mathrm{vec}(\mathbf{\Gamma})$, we get

$$\mathbb{E}(\mathrm{vec}(\widetilde{\mathbf{\Gamma}})) = (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1) \mathbf{W}_{\underline{n}}^* \mathrm{vec}(\mathbf{\Gamma}),$$

which is (2.38).

Finally, since $\mathbf{\Delta}_i$ are diagonal with positive entries $(n_i - |h_i|)/n_i > 0$, for $|h_i| < n_i$, so is their Kronecker product $\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1$, and since $\mathbf{W}_{\underline{n}}^* \mathbb{1} = \mathbf{0}$ this implies $\widetilde{\mathbf{W}}_{\underline{n}}^* \mathbb{1} = (\mathbf{\Delta}_2 \otimes \mathbf{\Delta}_1) \mathbf{W}_{\underline{n}}^* \mathbb{1} = \mathbf{0}$, so the null space of $\widetilde{\mathbf{W}}_{\underline{n}}^*$ is contained in that of $\mathbf{W}_{\underline{n}}^*$.

□

Appendix C. Proof of Theorem 2.3

Proof of Theorem 2.3. Each summand of the standard estimator (2.11) is a product of two deviations from the sample mean, so that

$$\mathrm{Cov}\big(\widehat{C}(\underline{h}), \widehat{C}(\underline{g})\big) = \frac{1}{N_{\underline{h}} N_{\underline{g}}} \sum_{\underline{t} \in G_{\underline{n}}(\underline{h})} \sum_{\underline{s} \in G_{\underline{n}}(\underline{g})} \mathrm{Cov}\Big(\big(X_{\underline{t}+\underline{h}} - \overline{X}\big)\big(X_{\underline{t}} - \overline{X}\big),\ \big(X_{\underline{s}+\underline{g}} - \overline{X}\big)\big(X_{\underline{s}} - \overline{X}\big) \Big), \tag{C.1}$$

by the bilinearity of the covariance. Fix $\underline{t} \in G_{\underline{n}}(\underline{h})$ and $\underline{s} \in G_{\underline{n}}(\underline{g})$. Under Assumption 2.3, the four centered variables in (C.1) are jointly Gaussian. They have constant mean, because $\mathbb{E}(X_{\underline{t}}) = \mathbb{E}(\overline{X}) = \mu$ under Assumption 2.1; and their pairwise covariances are, by the definition (2.15) of the centered covariance function, the values C_Z evaluated at the corresponding pairs of locations. Isserlis' theorem (Isserlis, 1918) applied to their fourth moment therefore gives three pairings, of which the first cancels against the product of expectations in the covariance, leaving the two cross pairings

$$\begin{aligned}
&\mathrm{Cov}\Big(\big(X_{\underline{t}+\underline{h}} - \overline{X}\big)\big(X_{\underline{t}} - \overline{X}\big),\ \big(X_{\underline{s}+\underline{g}} - \overline{X}\big)\big(X_{\underline{s}} - \overline{X}\big) \Big) \\
&\quad = \mathrm{Cov}\big(X_{\underline{t}+\underline{h}} - \overline{X},\, X_{\underline{s}+\underline{g}} - \overline{X}\big)\, \mathrm{Cov}\big(X_{\underline{t}} - \overline{X},\, X_{\underline{s}} - \overline{X}\big) \\
&\qquad + \mathrm{Cov}\big(X_{\underline{t}+\underline{h}} - \overline{X},\, X_{\underline{s}} - \overline{X}\big)\, \mathrm{Cov}\big(X_{\underline{t}} - \overline{X},\, X_{\underline{s}+\underline{g}} - \overline{X}\big) \\
&\quad = C_Z(\underline{t}+\underline{h}, \underline{s}+\underline{g})\, C_Z(\underline{t}, \underline{s}) + C_Z(\underline{t}+\underline{h}, \underline{s})\, C_Z(\underline{t}, \underline{s}+\underline{g}).
\end{aligned} \tag{C.2}$$

Inserting (C.2) into (C.1) proves the centered form (2.39).

For the lag form (2.40), we apply the form given by (2.15), that is by $\mathrm{Cov}(X_{\underline{t}}, X_{\underline{s}}) = C(\underline{s} - \underline{t})$,, as in (2.1). Then, with the shift $\underline{\delta} := \underline{s} - \underline{t}$, the four lags are

$$\begin{aligned}
(\underline{s}+\underline{g}) - (\underline{t}+\underline{h}) &= \underline{\delta} + \underline{g} - \underline{h}, & \underline{s} - \underline{t} &= \underline{\delta}, \\
\underline{s} - (\underline{t}+\underline{h}) &= \underline{\delta} - \underline{h}, & (\underline{s}+\underline{g}) - \underline{t} &= \underline{\delta} + \underline{g},
\end{aligned}$$

so that

$$C_Z(\underline{t}+\underline{h}, \underline{s}+\underline{g}) = C(\underline{\delta} + \underline{g} - \underline{h}) - \overline{C}(\underline{t}+\underline{h}) - \overline{C}(\underline{s}+\underline{g}) + D, \tag{C.3}$$

$$C_Z(\underline{t}, \underline{s}) = C(\underline{\delta}) - \overline{C}(\underline{t}) - \overline{C}(\underline{s}) + D, \tag{C.4}$$

$$C_Z(\underline{t}+\underline{h}, \underline{s}) = C(\underline{\delta} - \underline{h}) - \overline{C}(\underline{t}+\underline{h}) - \overline{C}(\underline{s}) + D, \tag{C.5}$$

$$C_Z(\underline{t}, \underline{s}+\underline{g}) = C(\underline{\delta} + \underline{g}) - \overline{C}(\underline{t}) - \overline{C}(\underline{s}+\underline{g}) + D. \tag{C.6}$$

Substituting (C.3)–(C.6) into (2.39) yields (2.40). Finally, neither C, nor $\overline{C}$, nor D depends on μ, so neither does the right-hand side of (2.39)–(2.40). □

Proof of Corollary 2.3. By (A.24), $\widetilde{C}(\underline{h}) = N_{\underline{h}}/N\,\widehat{C}(\underline{h})$ with $N_{\underline{h}}/N$ deterministic, so the bilinearity of the covariance yields

$$\mathrm{Cov}\big(\widetilde{C}(\underline{h}), \widetilde{C}(\underline{g})\big) = \frac{N_{\underline{h}} N_{\underline{g}}}{N^2}\,\mathrm{Cov}\big(\widehat{C}(\underline{h}), \widehat{C}(\underline{g})\big).$$

Substituting the representation (2.39) of Proposition 2.3, whose prefactor is $1/(N_{\underline{h}}N_{\underline{g}})$, the factors $N_{\underline{h}}N_{\underline{g}}$ cancel and leave $1/N^2$, which is (2.42). □

Appendix D. Proof of Theorem 3.1

In this Appendix, we prove Theorem 3.1 regarding the finite-sample bias of the bias-corrected estimator and its bound. For this purpose, first, we prove in Lemma 3.1 the invertibility of the weight matrix $\mathbf{W}^*_{\underline{n},\underline{m}}$ when the sufficient condition $M_{\underline{n}}/N < 1$ is satisfied.

Proof of Lemma 3.1. Let $\mathbf{S}_{\underline{m}}$ be the selection matrix as in (3.4) with $\mathbf{S}_{\underline{m}}\mathbf{S}_{\underline{m}}^\top$ and the reduced weight matrix $\mathbf{S}_{\underline{m}}\mathbf{W}^{*\top}_{\underline{m},\underline{n}}\mathbf{S}_{\underline{m}}$ as in (3.9), hence we have

$$\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}} = \mathbf{S}_{\underline{m}}\big(\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}\big)\mathbf{S}_{\underline{m}}^\top, \tag{D.1}$$

so that every entry of $\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}$ is an entry of $\mathbf{W}^*_{\underline{n}} - \mathbf{I}_{L_{\underline{n}}}$and is bounded in absolute value by $1/N$ by Lemma Appendix B.1. Since each row and each column in $\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}$ has $M_{\underline{m}}$ entries, we have the bounds in (3.15).

Suppose that $M_{\underline{m}}/N < 1$. By Lemma 2.3.3 of Golub and Van Loan (2013) (4th edition), since $\|\mathbf{I}_{M_{\underline{m}}} - \big(\mathbf{I}_{M_{\underline{m}}} - \mathbf{W}^*_{\underline{n},\underline{m}}\big)\|_\infty = \|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\|_\infty \le M_{\underline{m}}/N < 1$ and $\|\cdot\|_\infty$ is submultiplicative, the Neumann series applied to $\mathbf{W}^*_{\underline{n},\underline{m}} = \mathbf{I}_{M_{\underline{m}}} - \big(\mathbf{I}_{M_{\underline{m}}} - \mathbf{W}^*_{\underline{n},\underline{m}}\big)$ shows that $\mathbf{W}^*_{\underline{n},\underline{m}}$ is nonsingular with

$$\big\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\big\|_\infty \le \frac{1}{1 - \big\|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\big\|_\infty} \le \frac{1}{1 - (M_{\underline{m}}/N)} = \frac{N}{N - M_{\underline{m}}} \tag{D.2}$$

$$\big\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1} - \mathbf{I}_{M_{\underline{m}}}\big\|_\infty \le \frac{\big\|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\big\|_\infty}{1 - \big\|\mathbf{W}^*_{\underline{n},\underline{m}} - \mathbf{I}_{M_{\underline{m}}}\big\|_\infty} \le \frac{M_{\underline{m}}/N}{1 - (M_{\underline{m}}/N)} = \frac{M_{\underline{m}}}{N - M_{\underline{m}}}. \tag{D.3}$$

which completes the proof. □

Proof of Theorem 3.1. By applying the expectation to (3.11) then,

$$\mathbb{E}\big(\widehat{\underline{C}}^*_{\underline{m}}\big) = \big(\mathbf{W}^*_{\underline{n},\underline{m}}\big)^{-1}\Big(\mathbf{W}^*_{\underline{n},\underline{m}}\,\underline{C}_{\underline{m}} + \mathbf{W}^{*,c}_{\underline{n},\underline{m}}\,\underline{C}^c_{\underline{m}}\Big) = \underline{C}_{\underline{m}} + \big(\mathbf{W}^*_{\underline{n},\underline{m}}\big)^{-1}\mathbf{W}^{*,c}_{\underline{n},\underline{m}}\,\underline{C}^c_{\underline{m}},$$

Then subtracting $\widehat{\underline{C}}_{\underline{m}}$ on both sides gives (3.12).
Bound. By Lemma 3.1 and applying the Hölder inequality we have

$$\begin{aligned}\big\|\underline{b}^*_{\underline{m}}\big\|_\infty &= \big\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\mathbf{W}^{*,c}_{\underline{n},\underline{m}}\underline{C}^c_{\underline{m}}\big\|_\infty \le \big\|(\mathbf{W}^*_{\underline{n},\underline{m}})^{-1}\big\|_\infty \big\|\mathbf{W}^{*,c}_{\underline{n},\underline{m}}\underline{C}^c_{\underline{m}}\big\|_\infty \le \frac{N}{N - M_{\underline{m}}}\big\|\mathbf{W}^{*,c}_{\underline{n},\underline{m}}\big\|_{\max}\big\|\underline{C}^c_{\underline{m}}\big\|_1 \\ &\le \Big(\frac{N}{N - M_{\underline{m}}}\Big)\frac{1}{N}\sum_{\underline{h}\in W^c_{\underline{m}}} |C(\underline{h})| = \frac{1}{N - M_{\underline{m}}}\big\|\underline{C}^c_{\underline{m}}\big\|_1\end{aligned}$$

□

Proof of Corollary 3.1. Given that the deterministic scaling $\widetilde{C}(h_1,h_2) = (N_h/N)\cdot\widehat{C}(h_1,h_2)$ from (A.24) holds entrywise in $\mathrm{vec}(\widetilde{\mathbf{\Gamma}})$, as we have that $\mathrm{vec}(\widetilde{\mathbf{\Gamma}}) = (\mathbf{\Delta}_2\otimes\mathbf{\Delta}_1)\mathrm{vec}(\widehat{\mathbf{\Gamma}})$ and by (3.4) and (3.18) we have

$$\mathrm{vec}(\widetilde{\mathbf{\Gamma}}_{\underline{m}}) = \mathbf{S}_{\underline{m}}(\mathbf{\Delta}_2\otimes\mathbf{\Delta}_1)\mathrm{vec}(\widehat{\mathbf{\Gamma}}) = (\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})\mathbf{S}_{\underline{m}}\mathrm{vec}(\widehat{\mathbf{\Gamma}}) = (\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})\,\mathrm{vec}(\widehat{\mathbf{\Gamma}}_{\underline{m}}).$$

Substituting this together with (3.19) into (3.20), and using that the Kronecker product of the diagonal Bartlett window matrices $\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1}$ is invertible, since its diagonal entries are $(n_i - |h_i|)/n_i > 0$ for $|h_i|\le m_i < n_i$), we have

$$\begin{aligned}\widetilde{C}^*_{\underline{m}} &= \left[(\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})\,\mathbf{W}^*_{\underline{n},\underline{m}}\right]^{-1}(\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})\,\mathrm{vec}(\widehat{\mathbf{\Gamma}}_{\underline{m}})\\ &= \left(\mathbf{W}^*_{\underline{n},\underline{m}}\right)^{-1}(\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})^{-1}(\mathbf{\Delta}_{2,m_2}\otimes\mathbf{\Delta}_{1,m_1})\,\mathrm{vec}(\widehat{\mathbf{\Gamma}}_{\underline{m}}) = \left(\mathbf{W}^*_{\underline{n},\underline{m}}\right)^{-1}\mathrm{vec}(\widehat{\mathbf{\Gamma}}_{\underline{m}}) = \widehat{C}^*_{\underline{m}}.\end{aligned}$$

□

Appendix E. Proof of Proposition 3.1

By exploiting a common truncation argument (see, e.g., Brockwell and Davis (1991), Proposition 7.3.4), for $\underline{q} := (q_1,q_2)\in\mathbb{N}^2$, we define the truncated process $\mathbf{X}^{(\underline{q})} = \{X^{(\underline{q})}_{\underline{t}}, \underline{t}\in\mathbb{Z}^2\}$ by

$$X_{\underline{t}} = \mu + \sum_{j_1=-q_1}^{q_1}\sum_{j_2=-q_2}^{q_2}\psi_{\underline{j}}\epsilon_{\underline{t}-\underline{j}}, \quad (\underline{t})\in\mathbb{Z}^2.$$

Then, by exploiting the m-dependence property (in two-dimensions) of $\mathbf{X}^{(\underline{q})}$, it is straightforward to show that

$$\sqrt{N}\left(\widehat{\underline{C}}^{(\underline{q})}_m - \underline{C}^{(\underline{q})}_m\right)\xrightarrow{d}\mathcal{N}\left(\underline{0}, V(\widehat{\underline{C}}^{(\underline{q})}_m)\right),$$

where the entries of $V(\widehat{\underline{C}}^{(\underline{q})}_m)$ are of the form

$$\begin{aligned}V(\widehat{\underline{C}}^{(\underline{q})}_m)(\underline{h},\underline{g}) =&(\eta-3)C^{(\underline{q})}(\underline{h})C^{(\underline{q})}(\underline{g}) + \sum_{\underline{\delta}\in\mathbb{Z}^2}\left\{C^{(\underline{q})}(\underline{\delta})C^{(\underline{q})}(\underline{\delta}+\underline{g}-\underline{h})\right\}\\ &+\sum_{\underline{\delta}\in\mathbb{Z}^2}\left\{C^{(\underline{q})}(\underline{\delta}+\underline{g})C^{(\underline{q})}(\underline{\delta}-\underline{h})\right\},\end{aligned}$$

where $C^{(\underline{q})}(\underline{\delta})$ denotes the corresponding covariances of lag $\underline{\delta}$ of the truncated process $\mathbf{X}^{(\underline{q})}$. Hence, as $\mathbf{X}^{(\underline{q})}$ is m-dependent (in two-dimensions with $\underline{m} = (m_1,m_2)$ and $m_i = 2q_i+1$, in contrast to $C(\underline{\delta})$ and $V(\widehat{\underline{C}}_m)$, where all sums are infinite, all sums for $C^{(\underline{q})}(\underline{\delta})$ and $V(\widehat{\underline{C}}^{(\underline{q})}_m)$ become finite. Further, by letting $\underline{q}\to\infty$, we get $V(\widehat{\underline{C}}^{(\underline{q})}_m)\to V(\widehat{\underline{C}}_m)$ and, finally, by using standard arguments (i.e., the absolute summability of the sequence $\{\psi_{\underline{j}}, \underline{j}\in\mathbb{Z}^2\}$) and showing that for all $\delta>0$, we have

$$\lim_{\underline{q}\to\infty}\limsup_{n_1,n_2\to\infty} P\left(\left|\sqrt{N}\left(\widehat{\underline{C}}_m - \underline{C}_m\right) - \sqrt{N}\left(\widehat{\underline{C}}^{(\underline{q})}_m - \underline{C}^{(\underline{q})}_m\right)\right| > \delta\right) = 0,$$

the proof of $\sqrt{N}\left(\widehat{\underline{C}}_m - \underline{C}_m\right)\xrightarrow{d}\mathcal{N}\left(\underline{0}, V(\widehat{\underline{C}}_m)\right)$ is complete. Further, for showing $\sqrt{N}\left(\widehat{\underline{C}}^*_m - \underline{C}_m\right)\xrightarrow{d}\mathcal{N}\left(\underline{0}, V(\widehat{\underline{C}}_m)\right)$, using (3.11), we get

$$\begin{aligned}\sqrt{N}\left(\widehat{\underline{C}}^*_m - \underline{C}_m\right) =&\sqrt{N}\left(\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1}\widehat{\underline{C}}_m - \widehat{\underline{C}}_m + \widehat{\underline{C}}_m - \underline{C}_m\right)\\ =&\sqrt{N}\left(\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1} - I_{M_{\underline{m}}}\right)\widehat{\underline{C}}_m + \sqrt{N}\left(\widehat{\underline{C}}_m - \underline{C}_m\right).\end{aligned}$$

Hence, in view of the asymptotic normality result derived above and the consistency of $\widehat{\underline{C}}_m$, that is, $\widehat{\underline{C}}_m \to \underline{C}_m$, it remains to show

$$\sqrt{N}\left(\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1} - I_{M_{\underline{m}}}\right) = o(1).$$

From

$$I_{M_{\underline{m}}} = \mathbf{W}^*_{\underline{n},m}\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1}$$

and

$$\mathbf{W}^*_{\underline{n},m} = \mathbf{I}_{M_{\underline{m}}} - \mathbf{A}^*_{2,m} \otimes \mathbf{A}^*_{1,m} - \mathbf{B}^*_{2,m} \otimes \mathbf{B}^*_{1,m} + \mathbf{D}^*_{2,m} \otimes \mathbf{D}^*_{1,m},$$

we get

$$\begin{aligned}
&\sqrt{N}\left(\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1} - I_{M_{\underline{m}}}\right) \\
&= -\sqrt{N}\left(-\mathbf{A}^*_{2,m} \otimes \mathbf{A}^*_{1,m} - \mathbf{B}^*_{2,m} \otimes \mathbf{B}^*_{1,m} + \mathbf{D}^*_{2,m} \otimes \mathbf{D}^*_{1,m}\right)\left(\mathbf{W}^*_{\underline{n},m}\right)^{-1}. \qquad \text{(E.1)}
\end{aligned}$$

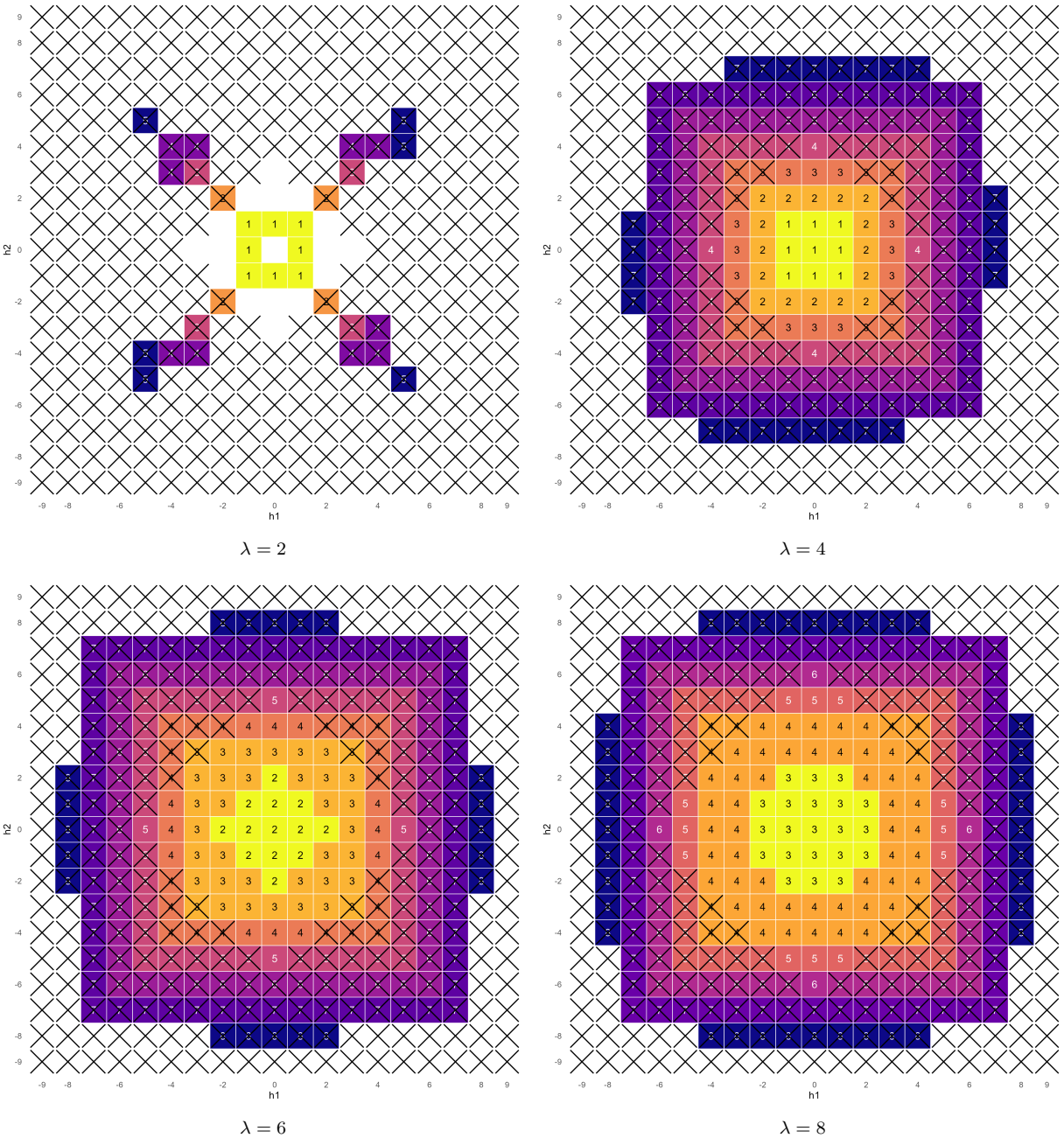
Finally, for all $i \in \{1,2\}$, note that all entries of $\mathbf{A}^*_{i,m}$, $\mathbf{B}^*_{i,m}$ and $\mathbf{D}^*_{i,m}$ are of order $O(1/n_i)$, respectively, due to the application of the selection matrices $\mathbf{P}_{i,m}$ to $\mathbf{A}^*_i$, $\mathbf{B}^*_i$ and $\mathbf{D}^*_i$. Hence, the Kronecker products $\mathbf{A}^*_{2,m} \otimes \mathbf{A}^*_{1,m}$, $\mathbf{B}^*_{2,m} \otimes \mathbf{B}^*_{1,m}$ and $\mathbf{D}^*_{2,m} \otimes \mathbf{D}^*_{1,m}$ are of order $O(1/N)$. Hence, (E.1) becomes

$$-\sqrt{N}O\left(\frac{1}{N}\right)O(1) = O\left(\frac{1}{\sqrt{N}}\right),$$

which completes the proof. □

Appendix F. Supplementary figures

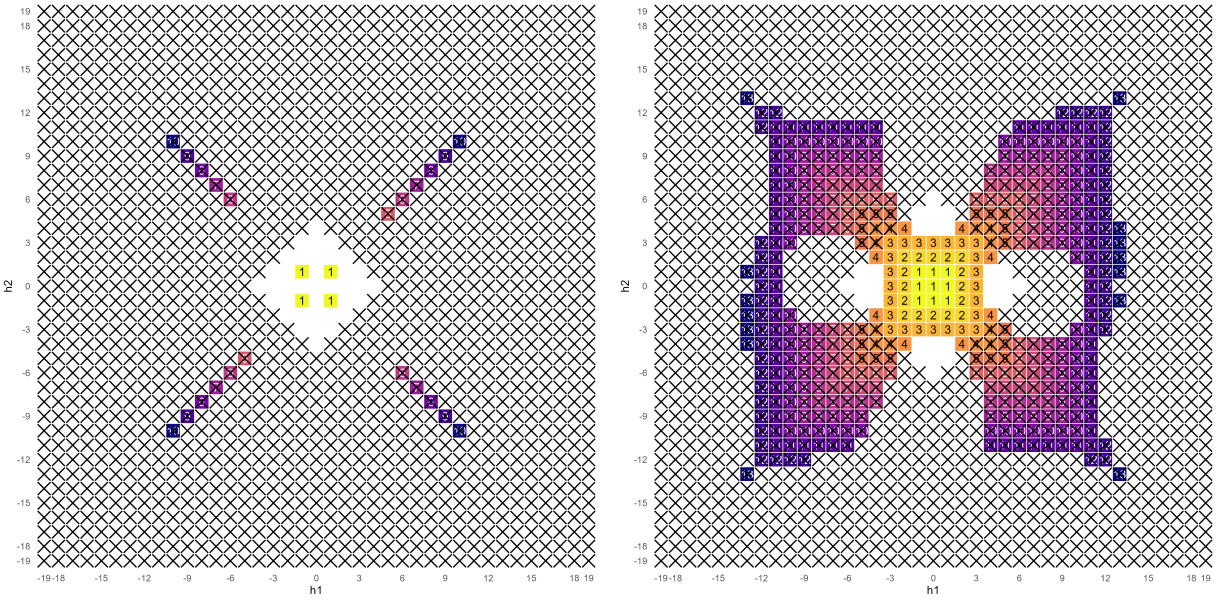
In Figure F.15, a heatmap for each spatial dependence level is shown, each related to observations in a grid of size (10×10). In the axis, the (h_1, h_2)-lag pairs are represented. Colored cells indicates if the bias-correction estimator $\widehat{C}^*_{\underline{m}}$ has achieved a better MSE than the standard estimator $\widehat{C}$, and, if so, it also shows which m values achieve the best possible MSE reduction for each specific lag pair. Recall that the higher m, the more covariances are used in the weighting of the bias-corrected estimator and therefore the more bias it may reduce. We see that especially for high levels of spatial dependence, i.e., for $\lambda \in \{6, 8\}$, the bias corrected method can afford to use higher values of m and thus is able to correct more of the bias. The crossed out, colored cells, corresponding to larger lag pairs, indicate cases where a better MSE is achieved by simply tapering the estimator, i.e., by setting $\widehat{C}(\underline{h}) := 0$. This behavior is typical especially for processes with short-range spatial dependence, where the true covariance function decays rapidly to zero, such that the estimation noise dominates any small signal. Consequently, tapering to zero these entries reduces the variance more than the bias it introduces, leading to an overall improvement in the MSE. Finally the cells that are left in white (without crosses) have a better MSE either with the standard estimation or with the tapered to zero estimation, but in those cases, bias correction performs worse than either of those two options. Figure F.16 shows the same MSE heatmap, but for a grid of size (20×20). On this larger lattice, the symmetry implied by the chosen (separable) covariance model becomes more clearly visible. Also here the optimal choice of m for the same lag-pair increases together with the strength of the spatial dependence.



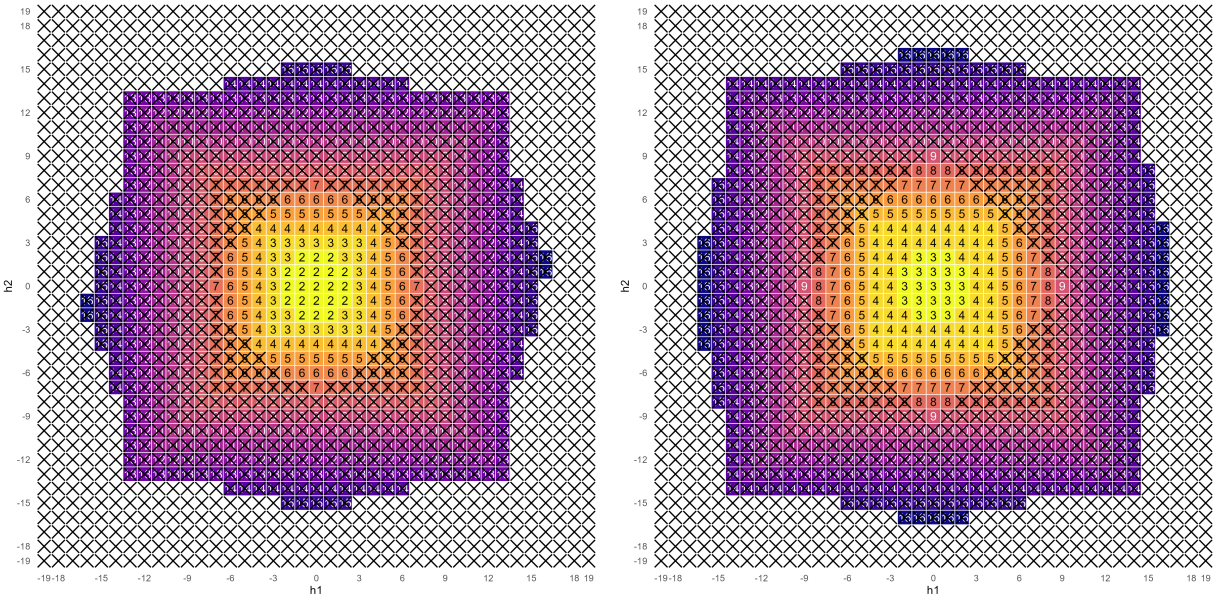
$\lambda = 2$ $\lambda = 4$

$\lambda = 6$ $\lambda = 8$

Figure F.15: Values of m for which $\widehat{\underline{C}}^*_m$ achieves the best MSE reduction compared to the naive estimator $\widehat{C}$. Crossed-out cells indicate that tapering to zero ($\widehat{C}(h_1, h_2) := 0$) yields the best MSE. Grid size 10×10.



$\lambda = 2$ $\lambda = 4$



$\lambda = 6$ $\lambda = 8$

Figure F.16: Values of m for which $\widehat{C}^*_m$ achieves the best MSE reduction compared to the naive estimator $\widehat{C}$. Crossed-out cells indicate that tapering to zero ($\widehat{C}(h_1, h_2) := 0$) yields the best MSE. Grid size 20×20.