

Estimation of the asymptotic variance of univariate and multivariate random fields and statistical inference

Annabel Prause and Ansgar Steland

*RWTH Aachen University
Institute of Statistics*

Wüllnerstr. 3, D-52056 Aachen, Germany

e-mail: prause@stochastik.rwth-aachen.de; steland@stochastik.rwth-aachen.de

Abstract: Correlated random fields are a common way to model dependence structures in high-dimensional data, especially for data collected in imaging. One important parameter characterizing the degree of dependence is the asymptotic variance which adds up all autocovariances in the temporal and spatial domain. Especially, it arises in the standardization of test statistics based on partial sums of random fields and thus the construction of tests requires its estimation. In this paper we propose consistent estimators for this parameter for strictly stationary φ -mixing random fields with arbitrary dimension of the domain and taking values in a Euclidean space of arbitrary dimension, thus allowing for multivariate random fields. We establish consistency, provide central limit theorems and show that distributional approximations of related test statistics based on sample autocovariances of random fields can be obtained by the subsampling approach.

As in applications the spatial-temporal correlations are often quite local, such that a large number of autocovariances vanish or are negligible, we also investigate a thresholding approach where sample autocovariances of small magnitude are omitted. Extensive simulation studies show that the proposed estimators work well in practice and, when used to standardize image test statistics, can provide highly accurate image testing procedures. Having in mind automatized applications on a big data scale as arising in data science problems, these examinations also cover the proposed data-adaptive procedures to select method parameters.

MSC 2010 subject classifications: Primary 62H86, 62E20; secondary 60G60, 62H12.

Keywords and phrases: Data science, high-dimensional data, imaging, long-run variance estimation, threshold estimator, subsampling, spatial statistics, testing.

Received March 2017.

1. Introduction

Random fields are a natural approach to model high-dimensional data. They arise in a natural way when analyzing digitized image data as arising in medical imaging, e.g. MRI or CT images, or in industrial quality control, e.g. images of materials obtained from cameras capturing the visual or infrared spectrum or electroluminescence images of solar cells, in order to analyze the structure of the

material of interest and to detect defects. Although the main results go beyond that scope and even allow to treat, for example, voxel-by-voxel multivariate brain data collected at a sample of individuals, let us first stick to the following setting, in order to motivate the approach, outline basic ideas and fix notations: Consider a sequence of n_1 images represented by $n_2 \times n_3$ dimensional matrices

$$(Y_{(i_1, i_2, i_3)})_{(i_2, i_3) \in (1:n_2) \times (1:n_3)}, \quad i_1 = 1, \dots, n_1, \tag{1}$$

of real-valued random variables, where for any $q \in \mathbb{N}$ we put $\mathbf{a} : \mathbf{b} = \times_{i=1}^q \{a_i, \dots, b_i\}$ for $\mathbf{a} = (a_1, \dots, a_q), \mathbf{b} = (b_1, \dots, b_q) \in \mathbb{N}^q$ and $\mathbf{1} = (1, \dots, 1) \in \mathbb{N}^q$. The sequence (1) of matrices can be seen as the corresponding subset $\{Y_{\mathbf{i}}\}_{\mathbf{i} \in \mathbf{1} : \mathbf{n}}$ of a three-dimensional random field $\{Y_{\mathbf{i}} : \mathbf{i} \in \mathbb{Z}^3\}$ of random variables defined on a common probability space. A statistical test to check whether or not one observes a reference image $\mathbf{m}^{(0)} = (m_{(i_2, i_3)}^{(0)})_{(i_2, i_3) \in \mathbf{1} : (n_2, n_3)}$ can be based on the sum

$$S_{\mathbf{n}} = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} (Y_{\mathbf{i}} - m_{(i_2, i_3)}^{(0)}), \quad \mathbf{n} = (n_1, n_2, n_3) \in \mathbb{N}^3.$$

In order to test for the presence of a known reference series of images $m_{\mathbf{i}}^{(0)}, \mathbf{i} \in \mathbf{1} : \mathbf{n}$, one simply replaces $m_{(i_2, i_3)}^{(0)}$ by $m_{\mathbf{i}}^{(0)}$. If the null hypothesis $H_0 : \mathbb{E}(Y_{\mathbf{i}}) = m_{\mathbf{i}}^{(0)} (\forall \mathbf{i} \in \mathbf{1} : \mathbf{n})$ holds true, under fairly weak conditions on the error random field $\xi_{\mathbf{i}} = Y_{\mathbf{i}} - \mathbb{E}(Y_{\mathbf{i}}), \mathbf{i} \in \mathbb{Z}^q$, the partial sum scaled by the squared root of $|\mathbf{n}| = \prod_i n_i$ converges weakly (in distribution) to a Gaussian law,

$$\frac{1}{\sqrt{|\mathbf{n}|}} S_{\mathbf{n}} \Rightarrow \sigma B(\mathbf{1}), \tag{2}$$

as $\mathbf{n} \rightarrow \infty$, where from now on $\mathbf{n} \rightarrow \infty$ is understood as $\min_j n_j \rightarrow \infty$. Here $B(x, y, z), x, y, z \geq 0$, is a standard Brownian motion in dimension 3. For example, (2) has been shown for weakly stationary linear processes, see [19] and [20], strictly stationary φ -mixing fields on which we shall focus in the present paper, see [11], or, using other notions of weak dependence, in [7], [3], [27], [14] and [8].

The asymptotic variance σ^2 is given by $\sigma^2 = \lim_{\mathbf{n} \rightarrow \infty} \sigma_{\mathbf{n}}^2$, where

$$\sigma_{\mathbf{n}}^2 = \text{Var} \left(\frac{1}{\sqrt{|\mathbf{n}|}} \sum_{\mathbf{i} \in \mathbf{1} : \mathbf{n}} \xi_{\mathbf{i}} \right)$$

For a weakly stationary random field it holds

$$\sigma_{\mathbf{n}}^2 = \sum_{-(n-1) \leq \ell \leq n-1} \prod_{j=1,2,3} \frac{n_j - |\ell_j|}{n_j} \mathbb{E}(\xi_0 \xi_{\ell}),$$

and for an isotropic field where $\mathbb{E}(\xi_0 \xi_{\ell})$ is a function of $\|\ell\|$

$$\sigma_{\mathbf{n}}^2 = \mathbb{E}(\xi_0^2) + \sum_{\mathbf{0} \leq \ell \leq n-1} \prod_{j=1,2,3} \frac{n_j - \ell_j}{n_j} \mathbb{E}(\xi_0 \xi_{\ell}).$$

Here $\mathbf{a} \leq \mathbf{b}$ for vectors \mathbf{a}, \mathbf{b} is understood component-wise and $\mathbf{a} \leq^* \mathbf{b}$ means $\mathbf{a} \leq \mathbf{b}$ with $\mathbf{a} \neq \mathbf{b}$, $\|\ell\|$ is the Euclidean vector norm and $\|\ell\|_0$ denotes the number of non-zero coordinates of ℓ with components ℓ_j .

In order to use the central limit theorem (2), it is crucial to be in a position to estimate σ^2 consistently, and the lack of such estimators, contrary to the time series literature, motivated this paper. For example, for the above mentioned image testing problem an asymptotic level α test, $\alpha \in (0, 1)$, for H_0 can be devised by rejecting H_0 if

$$|\mathbf{n}|^{-1/2}|S_{\mathbf{n}}| > \hat{\sigma}_{\mathbf{n}}\Phi^{-1}(1 - \alpha/2),$$

where Φ denotes the distribution function of the standard normal law, as long as $\hat{\sigma}_{\mathbf{n}}^2$ is a consistent estimator of σ^2 .

In the same vein, one can construct a statistical test for a single image, in order to test for departures from a reference model $m^{(0)}$ for the expectation of the two-dimensional random field $\{\xi_{\mathbf{i}} : \mathbf{0} < \mathbf{i} \leq \mathbf{n}\}$ representing the image of pixel resolution $n_1 \times n_2$. In our simulations we do not only investigate the proposed estimators in their own right, but also examine the accuracy of such an image test in terms of the type I error rate, see Section 4.3. It turns out that our approach allows for highly accurate test procedures for images as required by present day image analysis.

The aims and contributions of the present paper are therefore as follows: Going beyond the above scope of sequences of images, we consider general random fields of arbitrary dimension q of the domain and taking values in the p -dimensional Euclidean space \mathbb{R}^p . We propose and study a nonparametric estimator of the asymptotic variance, which directly generalizes the class of Bartlett-type long run variance estimators studied in time series analysis. In addition, we provide a central limit theorem for that estimator and establish the consistency of subsampling under weak conditions. A concise model when capturing a sample of images of constant scenery, e.g. in order to estimate the true underlying object by the noisy images and to estimate the spatial-temporal correlations, is to assume that the data is given by a superposition of a time series and a spatial random field. Then one may center the observed images at their temporal average. We show that the corresponding estimator of σ^2 is still consistent under fairly weak conditions. Lastly, as spatial (or time) correlations in image data resp. sequences of images are often of a local nature, so that autocovariances corresponding to larger lags are negligible or even vanish, we introduce and study a class of cut-off or thresholding estimators, which resemble to some extent thresholding estimators studied in high-dimensional statistics. Those cut-off estimators aim at reducing the estimation variability by neglecting autocovariances of small order. We present extensive simulation results, in order to shed some light onto the accuracy of the proposed estimators, to identify situations where the thresholding estimator improves upon the classical lag-truncation approach, and to investigate how the estimators perform when used in statistical image testing.

Consider a general q -dimensional real-valued random field $\{\xi_{\mathbf{n}} : \mathbf{n} \in \mathbb{Z}^q\}$ with $\mathbb{E}(\xi_{\mathbf{n}}) = 0$ for all $\mathbf{n} \in \mathbb{Z}^q$, where, as above, the first dimension typically

represents time. Then the asymptotic variance of the random field is defined as

$$\sigma^2 = \sum_{\mathbf{k} \in \mathbb{Z}^q} \mathbb{E}(\xi_0 \xi_{\mathbf{k}}).$$

When the random field attains values in \mathbb{R}^p for some $p \in \mathbb{N}$, i.e. $\xi_{\mathbf{n}} : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^p, \mathcal{B}^p)$, for $\mathbf{n} \in \mathbb{Z}^q$, where $(\Omega, \mathcal{F}, \mathbb{P})$ denotes the underlying probability space and \mathcal{B}^p the usual Borel σ -field on \mathbb{R}^p , the asymptotic variance is the matrix

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{|\mathbf{n}|} \mathbb{E}(S_n S_n'),$$

since now $\sigma_{\mathbf{n}}^2 = \text{Var}(|\mathbf{n}|^{-1/2} S_n) = |\mathbf{n}|^{-1} \mathbb{E}(S_n S_n')$. Observe that σ^2 adds up all (cross-) autocovariances in the temporal and spatial domain. Obviously, for $q = 1$ we are given a univariate time series and σ^2 is the well known long run variance, for which an extensive literature on its estimation exists, e.g. smoothing window type estimators, cf. [17], or estimators based on batched means, cf. [1] and [28]. But for a random field of dimension $q > 1$ there are only a few results which address certain special cases. Indeed, most of the existing results are motivated from an economic point of view and concern the spatial heteroscedasticity and autocorrelation consistent estimation of covariance matrices with applications in two dimensions as in [13], [9], [15] and [16]. [13] obtain consistent estimates of the matrix of cross-sectional correlations by averaging over the time dimension, i.e. they request that the time dimension grows while the size of the cross-sectional dimension stays fixed. They construct an estimator that relies on the standard Newey and West estimator of the time series literature, see [18], and show that it is robust to very general forms of spatial and temporal dependence as the time dimension becomes large.

[15] as well as [16] study spatial heteroscedasticity and autocorrelation consistent estimators of covariance matrices of parameter estimators, where the spatial dependence is measured by a so-called economic distance. If this economic distance $d_{ij,n}$ between two units i and j is small these units are highly dependent, if it is large, however, the units are nearly independent. Examples for an economic distance include geographic distances as well as transportation costs. Both papers also allow for errors in the measurement of the distance. The main disadvantage is, however, that both papers focus on linear processes with i.i.d. innovations, which include certain non-stationary models, but exclude non-linear models.

A slightly different approach is the one of [9] who considers the estimation of the asymptotic variance for strictly stationary and α -mixing random fields in two dimensions. He constructs an estimator as the weighted average of products of the observations and finally shows its consistency.

However, to the best of our knowledge, no further estimators of the asymptotic variance for arbitrary stationary mixing random fields in q dimensions exist in the literature so far. Thus, in Section 2 we propose an estimator defined as a weighted sum of the sample autocovariances of the random field and show its consistency for mixing random fields and also provide results about the

asymptotic distribution. In Section 3 we show how one can improve the estimator and propose a data-adaptive procedure to select remaining unknowns which make use of subsampling. Section 4 is devoted to an extensive simulation study of both estimators. The simulations study for several models the behavior of the estimators and demonstrate that the thresholding estimator is preferable. We also investigate the statistical properties of the image test and show that the thresholding estimator leads to very accurate image tests. For its application and accuracy in change detection, see [23] and [25]. Rigorous proofs of all results are provided in Section 5.

2. Nonparametric estimation of the asymptotic variance

In this section, we introduce the proposed nonparametric estimator for the asymptotic variance of a random field. The estimator is based on the formula (4) and belongs to the class of lag-truncation estimators. We provide an interesting extension to the case of multiplicative random fields which are well suited to analyze image samples of a constant scenery.

2.1. Estimation for general random fields

Let $\{\xi_i\}$ be a random field defined on $\Gamma_n := \{1, \dots, n_1\} \times \dots \times \{1, \dots, n_q\}$ and taking values in \mathbb{R}^p with $\max_{1 \leq j \leq p} \mathbb{E}(\xi_i^{(j)})^2 < \infty$ for $i \in \Gamma_n$, where $\xi_i = (\xi_i^{(1)}, \dots, \xi_i^{(p)})'$. For a fixed $\mathbf{j} \in \mathbb{Z}^q$ denote by

$$\gamma(\mathbf{j}) = \mathbb{E}(\xi_0 \xi_{\mathbf{j}}'), \quad \mathbf{j} \in \mathbb{Z}^q$$

the autocovariances of $\{\xi_i\}$. Define the sample autocovariances

$$\hat{\gamma}_n(\mathbf{j}) := \frac{1}{|\tilde{\Gamma}_n(\mathbf{j})|} \sum_{i \in \tilde{\Gamma}_n(\mathbf{j})} \xi_i \xi_{i+\mathbf{j}}', \quad \tilde{\Gamma}_n(\mathbf{j}) := \{i \in \Gamma_n : i + \mathbf{j} \in \Gamma_n\}, \quad (3)$$

and set

$$\hat{\sigma}_n^2 := \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) \hat{\gamma}_n(\mathbf{j}). \quad (4)$$

Recall that $|\mathbf{j}| \leq \mathbf{m}$ is understood component-by-component, i.e. $|j_i| \leq m_i$ for all $1 \leq i \leq q$. The weights $w_{\mathbf{m}}(\mathbf{j})$, for fixed $\mathbf{j} \in \mathbb{Z}^q$ and $\mathbf{m} = \mathbf{m}_n \in \mathbb{N}^q$, arising in (4) are assumed to satisfy the following assumptions.

(W1) $w_{\mathbf{m}}(\mathbf{j}) \rightarrow 1$ as $\mathbf{m} \rightarrow \infty$ for all $\mathbf{j} \in \mathbb{Z}^q$.

(W2) $|w_{\mathbf{m}}(\mathbf{j})| \leq C_w < \infty$ independently of \mathbf{j} and \mathbf{m} .

The weights $w_{\mathbf{m}}(\mathbf{j})$ can be chosen as the product of one-dimensional weights, i.e. we can put $w_{\mathbf{m}}(\mathbf{j}) := \prod_{i=1}^q w_{m_i}(j_i)$, where the $w_{m_i}(j_i)$, $1 \leq i \leq q$, are weights satisfying (W1) and (W2). Examples for such weights are the Bartlett weights, $w_{m_i}(j_i) = 1 - \frac{|j_i|}{m_i}$ for $1 \leq i \leq q$, or the Tukey-Hanning weight sequence, $w_{m_i}(j_i) = \frac{1 + \cos(\pi j_i / m_i)}{2}$, $1 \leq i \leq q$.

For multivariate random fields, i.e. for $p > 1$, σ^2 and its estimator are matrix-valued, such that the evaluation of an estimator's accuracy requires to select a matrix norm. For simplicity of presentation, in the sequel the matrix maximum norm on $\mathbb{R}^{p \times p}$ defined by

$$\|\mathbf{A}\|_\infty = \max_{\substack{1 \leq i, j \leq p}} |a_{ij}|, \quad \text{for a matrix } \mathbf{A} = (a_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}},$$

will be used. One could also employ the frequently used Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j=1}^p |a_{ij}|^2}$, which, however, satisfies

$$\|\mathbf{A}\|_\infty = \sqrt{\max_{i,j} |a_{ij}|^2} \leq \|\mathbf{A}\|_F \leq p \|\mathbf{A}\|_\infty,$$

such that all results can be easily reformulated in terms of $\|\cdot\|_F$ instead of $\|\cdot\|_\infty$, or any other matrix norm, since all norms on $\mathbb{R}^{p \times p}$ are equivalent.

Our main results require the random field to be φ -mixing. For related and other notions of weak dependence for random fields we refer to [12] and [6] and the discussion at the end of this section. Let us briefly recall the definition of φ -mixing.

For each j with $1 \leq j \leq q$ and for each $r \geq 0$, define

$$\mathcal{A}^+(j; r) := \sigma \left(\{ \xi_{n_1, \dots, n_q} : n_j \geq r, n_i \text{ unrestricted for } i \neq j \} \right), \quad (5)$$

$$\mathcal{A}^-(j; r) := \sigma \left(\{ \xi_{n_1, \dots, n_q} : n_j \leq r, n_i \text{ unrestricted for } i \neq j \} \right). \quad (6)$$

Further, for each $r \geq 1$ introduce

$$\varphi(j; r) := \sup \{ |\mathbb{P}(B|A) - \mathbb{P}(B)| : A \in \mathcal{A}^-(j; 0), B \in \mathcal{A}^+(j; r), P(A) > 0 \} \quad (7)$$

and define the (half-space) φ -mixing coefficients

$$\varphi(r) := \max_{1 \leq j \leq q} \varphi(j; r). \quad (8)$$

Putting $\varphi(0) = 1$ we can observe that the set $\{\varphi(r)\}$ is a decreasing sequence of real numbers. The random field $\{\xi_{\mathbf{n}} : \mathbf{n} \in \mathbb{Z}^q\}$ is now called φ -mixing, if $\varphi(r) \rightarrow 0$ as $r \rightarrow \infty$.

Assumption 1. Let $\{\xi_{\mathbf{i}}, \mathbf{i} \in \Gamma_{\mathbf{n}}\}$ be a strictly stationary, φ -mixing random field with $\mathbb{E}(\xi_{\mathbf{0}}) = \mathbf{0}$, $\max_{1 \leq j \leq p} \mathbb{E} \left(\xi_{\mathbf{0}}^{(j)} \right)^4 < \infty$ and mixing coefficients satisfying

$$\sum_{r=1}^{\infty} r^{q-1} \varphi^{\frac{1}{2}}(r) < \infty. \quad (9)$$

We are now in a position to formulate the following theorem on the consistency of the estimator $\hat{\sigma}_{\mathbf{n}}^2$ under mild regularity conditions for a general multivariate random field.

Theorem 2.1. *Suppose that Assumption 1 holds and that the weights $w_{\mathbf{m}}(\mathbf{j})$ fulfill (W1) and (W2). Furthermore, assume that $\mathbf{m} \rightarrow \infty$, as $\mathbf{n} \rightarrow \infty$, with*

$$\frac{(m^*)^3}{n_*} = o(1), \quad (10)$$

where $m^* = \max_{1 \leq i \leq q} m_i$ and $n_* = \min_{1 \leq i \leq q} n_i$. Then the estimator $\hat{\sigma}_{\mathbf{n}}^2$ is weakly consistent, i.e.

$$\|\hat{\sigma}_{\mathbf{n}}^2 - \sigma^2\|_{\infty} \xrightarrow{\mathbb{P}} 0$$

as $\mathbf{n} \rightarrow \infty$.

It is interesting to note that the consistency can be strengthened to L_2 -consistency resp. L_1 -consistency, as shown by the following theorem.

Theorem 2.2. *Let the conditions of Theorem 2.1 be satisfied. Then for $p = 1$*

$$\mathbb{E} \left[(\hat{\sigma}_{\mathbf{n}}^2 - \sigma^2)^2 \right] \rightarrow 0$$

as $\mathbf{n} \rightarrow \infty$, and for $p > 1$

$$\mathbb{E} \|\hat{\sigma}_{\mathbf{n}}^2 - \sigma^2\|_{\infty} \rightarrow 0,$$

as $\mathbf{n} \rightarrow \infty$.

A direct consequence of Theorem 2.1 is the consistency of the autocovariance estimators, $\hat{\gamma}_{\mathbf{n}}(\mathbf{j})$, used to define the estimator of σ^2 , which are, of course, of independent interest when analyzing random field data.

Corollary 2.1. *The estimator $\hat{\gamma}_{\mathbf{n}}(\mathbf{j})$ of the lag \mathbf{j} -autocovariance of random field $\{\xi_i\}$,*

$$\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) := \frac{1}{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|} \sum_{i \in \tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})} \xi_i \xi'_{i+\mathbf{j}}$$

for $\mathbf{j} \in \mathbb{Z}^q$, is L_r - and thus also weakly consistent for the lag \mathbf{j} autocovariance $\gamma(\mathbf{j})$, where $r = 2$ for $p = 1$ and $r = 1$ otherwise.

2.2. Asymptotic distribution theory

Let us now turn to the asymptotic distribution theory, which we study for univariate as well as multivariate random fields. We show that, under weak regularity conditions, the sample autocovariances are asymptotically Gaussian and provide an associated limit theorem for the estimator $\hat{\sigma}_{\mathbf{n}}^2$. The latter is interesting in its own right, but is also needed to establish the subsampling central limit theorem to be discussed and applied in the next section.

We need the following approximation result, which shows that the sample autocovariances can be scaled with $\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}$ or $\sqrt{|\mathbf{n}|}$.

Lemma 2.3. *Under Assumption 1 it holds*

$$\max_{|\mathbf{j}| \leq \mathbf{m}} \left\| \sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|} (\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma(\mathbf{j})) - \frac{1}{\sqrt{|\mathbf{n}|}} \sum_{\mathbf{i} \in 1:\mathbf{n}} (\xi_{\mathbf{i}} \xi_{\mathbf{i}+\mathbf{j}} - \gamma(\mathbf{j})) \right\|_{\infty} = o_{\mathbb{P}}(1),$$

as $\mathbf{n} \rightarrow \infty$.

We have the following result providing the weak convergence of the sample autocovariances and the estimator $\hat{\sigma}_{\mathbf{n}}^2$ when appropriately centered and scaled. Let us denote by \Rightarrow the weak convergence in the Euclidean space \mathbb{R}^l , $l \in \mathbb{N}$. Recall that a matrix $\mathbf{A} = (a_{\nu\mu})_{\nu\mu}$ of dimension $p \times p$ defines the quadratic form $(\lambda_{\nu})_{\nu=1}^p \mapsto \sum_{\nu,\mu=1}^p \lambda_{\nu} \lambda_{\mu} a_{\nu\mu}$, $(\lambda_{\nu})_{\nu=1}^p \in \mathbb{R}^p$, which is positive definite if it is positive for all $(\lambda_{\nu})_{\nu=1}^p \neq \mathbf{0}$. More generally, a scheme $(a_{\mathbf{j}}^{(\nu,\mu)} : |\mathbf{j}| \leq \mathbf{m}, |\mathbf{k}| \leq \mathbf{m}, 1 \leq \nu, \mu \leq p)$ of real numbers defines a quadratic form via $(\lambda_{\mathbf{j}}^{(\nu)} : |\mathbf{j}| \leq \mathbf{m}, 1 \leq \nu \leq p) \mapsto \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{|\mathbf{k}| \leq \mathbf{m}} \sum_{\nu,\mu=1}^p \lambda_{\mathbf{j}}^{(\nu)} \lambda_{\mathbf{k}}^{(\mu)} a_{\mathbf{j}\mathbf{k}}^{(\nu,\mu)}$, which is positive definite if it is positive for all $(\lambda_{\mathbf{j}}^{(\nu)} : |\mathbf{j}| \leq \mathbf{m}, 1 \leq \nu \leq p) \neq \mathbf{0}$.

Theorem 2.4. *Suppose Assumption 1 holds true with $p = 1$ and fix $\mathbf{m} \geq 1$.*

(i) $\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|} (\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma(\mathbf{j})) \Rightarrow B_{\mathbf{j}}$, as $\mathbf{n} \rightarrow \infty$, where $(B_{\mathbf{j}} : |\mathbf{j}| \leq \mathbf{m})$ is Gaussian with mean zero and covariances given by

$$\text{Cov}(B_{\mathbf{j}}, B_{\mathbf{j}'}) = \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(\xi_{\mathbf{0}} \xi_{\mathbf{j}} - \gamma(\mathbf{j})) (\xi_{\mathbf{i}} \xi_{\mathbf{i}+\mathbf{j}'} - \gamma(\mathbf{j}')), \quad (11)$$

for $|\mathbf{j}| \leq \mathbf{m}$ and $|\mathbf{j}'| \leq \mathbf{m}$, provided (11) induces a positive definite quadratic form.

(ii) For $\mathbf{m} \geq 1$ we have

$$\sqrt{|\mathbf{n}|} (\hat{\sigma}_{\mathbf{n}}^2 - \mathbb{E}(\hat{\sigma}_{\mathbf{n}}^2)) \Rightarrow \mathcal{S},$$

as $\mathbf{n} \rightarrow \infty$, where $\mathcal{S} \stackrel{d}{=} \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) B_{\mathbf{j}}$.

For a multivariate random field the autocovariances $\gamma(\mathbf{j})$ are $p \times p$ matrices with elements $\gamma(\mathbf{j})_{\nu\mu}$, $1 \leq \nu, \mu \leq p$. We have the following multivariate extension of Theorem 2.4.

Theorem 2.5. *Suppose that Assumption 1 holds true.*

(i) We have $\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|} (\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma(\mathbf{j})) \Rightarrow B_{\mathbf{j}}$, as $\mathbf{n} \rightarrow \infty$, with $B_{\mathbf{j}} = (B_{\mathbf{j}}^{(\nu,\mu)})_{\substack{1 \leq \nu \leq p \\ 1 \leq \mu \leq p}}$, for $|\mathbf{j}| \leq \mathbf{m}$, where $(B_{\mathbf{j}}^{(\nu,\mu)})$ is Gaussian with mean zero and covariances given by

$$\text{Cov}(B_{\mathbf{j}}^{(\nu,\mu)}, B_{\mathbf{j}'}^{(\nu',\mu')}) = \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(\xi_{\mathbf{0}}^{(\nu)} \xi_{\mathbf{j}}^{(\mu)} - (\gamma(\mathbf{j}))_{\nu\mu}) (\xi_{\mathbf{i}}^{(\nu')} \xi_{\mathbf{i}+\mathbf{j}'}^{(\mu')} - (\gamma(\mathbf{j}'))_{\nu'\mu'}), \quad (12)$$

for $1 \leq \nu, \mu, \nu', \mu' \leq p$, $|\mathbf{j}| \leq \mathbf{m}$ and $|\mathbf{j}'| \leq \mathbf{m}$, provided (12) defines a positive definite quadratic form.

(ii) For $\mathbf{m} \geq \mathbf{1}$ we have under the assumption of (i)

$$\sqrt{|\mathbf{n}|}(\hat{\sigma}_{\mathbf{n}}^2 - \mathbb{E}(\hat{\sigma}_{\mathbf{n}}^2)) \Rightarrow \mathcal{S},$$

as $\mathbf{n} \rightarrow \infty$, where $\mathcal{S} \stackrel{d}{=} \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) B_{\mathbf{j}}$.

It is worth mentioning that, by Lemma 5.1 and Lemma 5.4 (a), the asymptotic variance of the random field $\xi_{\mathbf{i}} \xi_{\mathbf{i}+\mathbf{j}} - \gamma(\mathbf{j})$, $\mathbf{i} \in \mathbb{Z}^q$,

$$\zeta_{\mathbf{j}}^2 = \text{Var}(B_{\mathbf{j}}) = \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(\xi_{\mathbf{0}} \xi_{\mathbf{j}} - \gamma(\mathbf{j}))(\xi_{\mathbf{i}} \xi_{\mathbf{i}+\mathbf{j}} - \gamma(\mathbf{j})),$$

exists for each $\mathbf{j} \in \mathbb{Z}^q$.

Next we aim at studying the centered and scaled sample autocovariances as a process indexed by the lag \mathbf{j} , thus extending the above results. Observe that for fixed \mathbf{n} the estimator $\hat{\gamma}_{\mathbf{n}}(\mathbf{j})$ is only well defined for $\mathbf{j} \leq \mathbf{n}$, and it is common to put $\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) = \mathbf{0}$ if $j_i > n_i$ for some $i \in \{1, \dots, q\}$. This motivates to consider the multivariate random field

$$G_{\mathbf{n}}(\mathbf{j}) = \sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}(\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma(\mathbf{j})), \quad \mathbf{j} \in \mathbb{Z}^q,$$

indexed by \mathbb{Z}^q and thus taking values in the space $S = (\mathbb{R}^{p \times p})^{\mathbb{Z}^q}$ of mappings $\mathbb{Z}^q \rightarrow \mathbb{R}^{p \times p}$. The space $S = (\mathbb{R}^{p \times p})^{\mathbb{Z}^q}$ is a separable and complete metric space when equipped with the metric

$$\rho(x, y) = \sum_{\mathbf{k} \in \mathbb{Z}^q} 2^{-\mathbf{k}} d_0(x_{\mathbf{k}}, y_{\mathbf{k}}),$$

for $x = \{x_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^q\}, y = \{y_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^q\} \in (\mathbb{R}^{p \times p})^{\infty}$, where $d_0(\mathbf{a}, \mathbf{b}) = \min(1, \|\mathbf{a} - \mathbf{b}\|_{\infty})$ for matrices $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{p \times p}$ and $2^{-|\mathbf{k}|} = \prod_{j=1}^k 2^{-|k_j|}$ for $\mathbf{k} = (k_1, \dots, k_q) \in \mathbb{Z}^q$.

The question arises whether $G_{\mathbf{n}}$ converges weakly in the space (S, ρ) , as Theorem 2.5 already provides the convergence of the finite-dimensional distributions. Since weak convergence in (S, ρ) is not elaborated in the literature, let us briefly discuss some details. First note that convergence with respect to ρ is pointwise convergence. For $x \in S$ and $t_1, \dots, t_k \in \mathbb{Z}^q, k \in \mathbb{N}$, the projection $\pi_{t_1, \dots, t_k} = (x_{t_1}, \dots, x_{t_k}) \in \mathbb{R}^{p \times pk}$, is continuous. The finite-dimensional distributions of a random element $X = (X_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^q)$ taking values in $(\mathbb{R}^{p \times p})^{\mathbb{Z}^q}$ are given by the laws of the random matrices $(X_{t_1}, \dots, X_{t_n})$ of dimension $p \times (pn)$, or, equivalently, by the laws of the $p^2 n$ -dimensional random vectors $((\text{vec} X_{t_1})', \dots, (\text{vec} X_{t_n})')'$, for $t_1, \dots, t_n \in \mathbb{N}$, where $\text{vec} A$ denotes the vector obtained by stacking the columns of a matrix A . The finite-dimensional sets, $\pi_{t_1, \dots, t_k}^{-1}(H) = \{z \in S : (z_{t_1}, \dots, z_{t_k}) \in H, H \subset \mathbb{R}^{p \times pk}$ measurable, are a π -system. For $x \in S$ and $\varepsilon > 0$ let $\mathcal{A}_{x, \varepsilon}$ be the system of sets A with $\dot{A} \subset A \subset B(x, \varepsilon)$, where $B(x, \varepsilon)$ is the open ball around x with radius ε . Choose $\mathbf{k} \in \mathbb{Z}^q$ such that $\sum_{\mathbf{j} \in \mathbb{Z}^q} 2^{-|\mathbf{j}|} - \sum_{|\mathbf{j}| \leq \mathbf{k}} 2^{-|\mathbf{j}|} < \varepsilon/2$. Consider now the uncountable many disjoint sets

$$A_{\eta} = \{y \in S : \|y_{\mathbf{i}} - x_{\mathbf{i}}\|_{\infty} \leq \eta, |\mathbf{i}| \leq |\mathbf{k}|, \|y_{\mathbf{i}_0} - x_{\mathbf{i}_0}\|_{\infty} < \eta, \exists \mathbf{i}_0 \text{ with } |\mathbf{i}_0| \leq |\mathbf{k}|\}$$

for $0 < \eta < \varepsilon/(2 \cdot 3^q)$. Since $\sum_{j \in \mathbb{Z}^q} 2^{-|j|} = 3^q$ and d_0 is bounded by 1, any $y \in A_\eta$ satisfies

$$\rho(x, y) \leq \sum_{|j| \leq k} 2^{-|k|} d_0(x_k, y_k) + \varepsilon/2 \leq \eta 3^q + \varepsilon/2 < \varepsilon.$$

Hence, $A_\eta \subset B(x, \varepsilon)$. This shows that the system of boundaries of the sets $\mathcal{A}_{x, \varepsilon}$ contains uncountably many disjoint sets, such that by [5] the weak convergence in $(\mathbb{R}^{p \times pn})^{\mathbb{Z}^q}$ coincides with the convergence of the finite-dimensional distributions.

Therefore, Theorem 2.5 implies the following result.

Theorem 2.6. *Under the conditions of Theorem 2.5*

$$G_n \Rightarrow B,$$

as $n \rightarrow \infty$, where $B = \{B_j : j \in \mathbb{Z}^q\}$.

2.3. Subsampling

Let us briefly review how subsampling for random fields works. For more details we refer to [22], especially Chapter 5.3 therein. Define

$$\mathbf{E}_u := \{\mathbf{t} \in \mathbb{Z}^q : 0 < t_k \leq u_k, k = 1, 2, \dots, q\}$$

and suppose that given the random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_n\}$, we have a consistent estimator $\hat{\theta}_n = \hat{\theta}_n(X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_n)$ of an unknown real-valued parameter $\theta(\mathbb{P})$. If we define $J_n(\mathbb{P})$ as the sampling probability law of $\tau_n(\hat{\theta}_n - \theta(\mathbb{P}))$ under \mathbb{P} , where τ_n is a normalizing constant, and write $J_n(x, \mathbb{P})$ for the corresponding sampling probability distribution function, i.e.

$$J_n(x, \mathbb{P}) := \mathbb{P}\{\tau_n(\hat{\theta}_n - \theta(\mathbb{P})) \leq x\}, \quad x \in \mathbb{R},$$

the aim of subsampling is to find an approximation of $J_n(x, \mathbb{P})$ by recomputing the statistic $\hat{\theta}_n$ over random fields of smaller size than n and by considering the empirical distribution function of these subsampled values. Thus, for $\mathbf{b}, \mathbf{h} \in \mathbb{Z}^q$ we define these smaller random fields by

$$Y_j := \{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}_{j, \mathbf{b}, \mathbf{h}}\},$$

where $\mathbf{E}_{j, \mathbf{b}, \mathbf{h}}$ stands for the rectangle containing the points $\mathbf{i} \in \mathbb{Z}^q$ with $(j_k - 1)h_k < i_k \leq (j_k - 1)h_k + b_k$ for $1 \leq k \leq q$. The point \mathbf{b} represents the size of the smaller random field defined on $\mathbf{E}_{j, \mathbf{b}, \mathbf{h}}$, the point \mathbf{h} determines how many of these smaller random fields are taken into account, as Y_j is only defined if $0 < j_k \leq N_k$, where $N_k = \left\lfloor \frac{n_k - b_k}{h_k} \right\rfloor + 1$. This means, the vector \mathbf{h} defines a grid and at each point of that grid a rectangle (block) defined by the block size \mathbf{b} is located. Depending on the choice of \mathbf{h} and \mathbf{b} , those rectangles may

overlap or not. Now, one can easily see that the number of considered subrandom fields decreases when \mathbf{h} increases and that it increases when \mathbf{h} decreases with a maximum for $\mathbf{h} = (1, 1, \dots, 1)$.

We further define the subsample value $\widehat{\theta}_{\mathbf{n}, \mathbf{b}, i}$ as the statistic $\widehat{\theta}_{\mathbf{b}}$ evaluated at the smaller random field Y_i , i.e. $\widehat{\theta}_{\mathbf{n}, \mathbf{b}, i} := \widehat{\theta}_{\mathbf{b}}(Y_i)$. The desired subsampling approximation of the sampling probability distribution function $J_{\mathbf{n}}(x, \mathbb{P})$ is then defined as

$$L_{\mathbf{n}, \mathbf{b}}(x) := |\mathbf{N}|^{-1} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_q=1}^{N_q} \mathbb{1}_{\{\tau_{\mathbf{b}}(\widehat{\theta}_{\mathbf{n}, \mathbf{b}, i} - \widehat{\theta}_{\mathbf{n}}) \leq x\}},$$

where $|\mathbf{N}| = \prod_{i=1}^q N_i$.

The question arises under which conditions this approximation is consistent. The main assumption for this to hold is that $J_{\mathbf{n}}(\mathbb{P})$ converges weakly to a limit law $J(\mathbb{P})$ with corresponding distribution function $J(x, \mathbb{P})$, as $\mathbf{n} \rightarrow \infty$; the precise assumptions are as in [22, Theorem 5.3.1] and are listed in the theorem below, which is adapted to our setting. The assumptions mainly control the growth rate of \mathbf{b} with respect to \mathbf{n} and the mixing coefficients of the underlying random field. A common choice studied in greater detail in Section 4.2 is $\mathbf{b} = \lfloor \mathbf{n}^\gamma \rfloor$ for $\gamma \in (0, 1)$.

The following theorem establishes the consistency of subsampling for statistics calculated from the underlying random field or, going beyond that case, which depend on (collections) of the terms arising in the lag \mathbf{j} sample autocovariances. Define $Y_{\mathbf{i}}(\mathbf{j}) = \xi_{\mathbf{i}} \xi_{\mathbf{i}+\mathbf{j}} - \gamma(\mathbf{j})$, $\mathbf{i} \in \mathbb{Z}^q$, for any $\mathbf{j} \in \mathbb{Z}^q$.

Theorem 2.7. *Let $\widehat{\theta}_{\mathbf{n}}$ be a real-valued statistic depending on $\{\xi_{\mathbf{i}} : \mathbf{0} < \mathbf{i} \leq \mathbf{n}\}$, $\{Y_{\mathbf{i}}(\mathbf{j}) : \mathbf{i} \in \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j})\}$ or $\{Y_{\mathbf{i}}(\mathbf{j}) : \mathbf{i} \in \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j}), |\mathbf{j}| \leq \mathbf{m}\}$, used for estimating the unknown real-valued parameter $\theta(\mathbb{P})$, such that*

$$J_{\mathbf{n}}(x, \mathbb{P}) = \mathbb{P}(|\mathbf{n}|^{-1/2}(\widehat{\theta}_{\mathbf{n}} - \theta(\mathbb{P})) \leq x) \Rightarrow J(x, \mathbb{P}),$$

as $\mathbf{n} \rightarrow \infty$. Assume that \mathbf{h} is a non-zero constant and assume that

$$\prod_{j=1}^q b_j / (n_j - b_j) = o(1), \tag{13}$$

as $\mathbf{n} \rightarrow \infty$, and

$$\frac{1}{|\mathbf{N}|} \sum_{k=1}^{N^*} k^{q-1} \varphi(k - 2m^*) = o(1), \tag{14}$$

as $\mathbf{n} \rightarrow \infty$. Then, in continuity points x of $J(x, \mathbb{P})$,

$$L_{\mathbf{n}, \mathbf{b}} \xrightarrow{\mathbb{P}} J(x, \mathbb{P}),$$

as $\mathbf{n} \rightarrow \infty$. If $J(\cdot, \mathbb{P})$ is continuous, then

$$q_{\mathbf{n}, \mathbf{b}}(\gamma) = \inf\{x : L_{\mathbf{n}, \mathbf{b}}(x) \geq \gamma\}$$

converges in probability to $q(\gamma) = \inf\{x : J(x, \mathbb{P}) \geq \gamma\}$ and

$$\mathbb{P}(|\mathbf{n}|^{-1/2}(\widehat{\theta}_{\mathbf{n}} - \theta(\mathbb{P})) \leq q_{\mathbf{n},b}(\gamma)) \rightarrow \gamma,$$

as $\mathbf{n} \rightarrow \infty$.

2.4. Extensions to multiplicative models

An important subclass of spatial-temporal random fields arises as the additive superposition of a time series introducing the serial dependencies and a spatial univariate random field, i.e.

$$Y_{\mathbf{i}} = \eta_{i_1}^{(T)} + \eta_{\bar{\mathbf{t}}}^{(S)}, \quad \mathbf{i} = (i_1, \bar{\mathbf{t}}) \in \mathbb{Z}^q.$$

As discussed in the introduction, this is a reasonable model when capturing, say, n_1 images of a fixed experimental situation, in order to estimate the image, its spatial dependence structure and the serial correlations of the data acquisition process. For a related detection procedures we refer to [24] and [25], Section 4.2. Obviously, the above model can be also formulated as a multiplicative model, and thus we assume from now on that $\xi_{\mathbf{i}} = (\xi_{\mathbf{i}}^{(1)}, \dots, \xi_{\mathbf{i}}^{(p)})'$ with

$$\xi_{\mathbf{i}}^{(j)} := \varepsilon_{i_1}^{(T)} \varepsilon_{\bar{\mathbf{t}}}^{(S,j)}, \quad \mathbf{i} = (i_1, \bar{\mathbf{t}}) \in \mathbb{Z}^q, \quad j = 1, \dots, p, \tag{15}$$

where $\varepsilon^{(T)} = \{\varepsilon_i^{(T)} : i \in \mathbb{Z}\}$ is a strictly stationary second order φ -mixing univariate time series and $\varepsilon^{(S)} = \{\varepsilon_{\bar{\mathbf{t}}}^{(S)} : \bar{\mathbf{t}} \in \mathbb{Z}^{q-1}\}$ is a strictly stationary φ -mixing multivariate random field taking values in \mathbb{R}^p with $\max_{1 \leq j \leq p} \mathbb{E}((\varepsilon_{\mathbf{0}}^{(S,j)})^4) < \infty$. We further assume that $\varepsilon^{(T)}$ and $\varepsilon^{(S)}$ have φ -mixing coefficients satisfying (9) for dimension 1 and $q - 1$, respectively, and are independent from each other.

In the present setting, one may center the $\xi_{\mathbf{i}}$ at their temporal average and therefore we define

$$\check{\gamma}_{\mathbf{n}}(\mathbf{j}) := \frac{1}{|\check{\Gamma}_{\mathbf{n}}(\mathbf{j})|} \sum_{(i_1, \bar{\mathbf{t}}) \in \check{\Gamma}_{\mathbf{n}}(\mathbf{j})} (\xi_{\mathbf{i}} - \bar{\xi}_{\cdot, \bar{\mathbf{t}}}) (\xi_{\mathbf{i}+\mathbf{j}} - \bar{\xi}_{\cdot, \bar{\mathbf{t}}})',$$

where $\bar{\xi}_{\cdot, \bar{\mathbf{t}}} := \frac{1}{n_1} \sum_{l=1}^{n_1} \xi_{l, \bar{\mathbf{t}}}$, and introduce the estimator

$$\check{\sigma}_{\mathbf{n}}^2 := \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) \check{\gamma}_{\mathbf{n}}(\mathbf{j})$$

for σ^2 . We then obtain the following theorem.

Theorem 2.8. *Suppose the noise process satisfies in addition to the assumptions of Theorem 2.1, the multiplicative model (15). Then $\check{\sigma}_{\mathbf{n}}^2 \xrightarrow{\mathbb{P}} \sigma^2$, as $\mathbf{n} \rightarrow \infty$, and $\mathbb{E}\|\check{\sigma}_{\mathbf{n}}^2 - \sigma^2\|_{\infty} \rightarrow 0$, as $\mathbf{n} \rightarrow \infty$.*

2.5. Discussion of the mixing assumptions

The above results assume that the random field is φ -mixing. Inspecting the proof shows that this condition can be replaced by other weak dependence conditions ensuring that

$$|\mathbf{n}|^{-1} \mathbb{E} \left| \sum_{\mathbf{k} \in \mathbf{1}:\mathbf{n}} \xi_{\mathbf{0}} \xi_{\mathbf{k}+\mathbf{j}} - \gamma(\mathbf{j}) \right|^2 = O(1), \quad (16)$$

as $\mathbf{n} \rightarrow \infty$, cf. Lemma 5.4. For example, a natural condition is to assume that $\xi_{\mathbf{k}}$ is ρ^* -mixing, which implies that the strictly stationary random field $Y_{\mathbf{k}}(\mathbf{j}) = \xi_{\mathbf{0}} \xi_{\mathbf{k}+\mathbf{j}} - \gamma(\mathbf{j})$ has a continuous spectral density function, cf. Lemma 5.2, which is then given by

$$f(t) = \lim_{n \rightarrow \infty} n^{-q} \mathbb{E} \left| \sum_{\mathbf{k} \in \mathbf{1}:n\mathbf{1}} e^{-i\mathbf{k}'\lambda_t} Y_{\mathbf{k}}(\mathbf{j}) \right|^2, \quad t \in S^q,$$

where from now on we put $n\mathbf{1} = (n, \dots, n)$ for $n \in \mathbb{N}$. Further, $S^q = \{\mathbf{x} \in \mathbb{R}^q : \|\mathbf{x}\|_2 = 1\}$ and λ_t is the unique number $\lambda_t = (\lambda_{t_1}, \dots, \lambda_{t_q}) \in (-\pi, \pi]^q$ such that $e^{-i\lambda_{t_j}} = t_j$ for $j = 1, \dots, q$, see [6, Theorem 28.21]. Here $\{\xi_{\mathbf{k}}\}$ is called ρ^* -mixing, if $\rho^*(r) \rightarrow 0$, as $r \rightarrow \infty$, where

$$\rho^*(r) = \sup \rho(\sigma(\xi_{\mathbf{k}} : \mathbf{k} \in A_1), \sigma(\xi_{\mathbf{k}} : \mathbf{k} \in A_2)).$$

The supremum is taken over all nonempty disjoint sets $A_1, A_2 \subseteq \mathbb{Z}^q$ with distance $\text{dist}(A_1, A_2) \geq r$, and the ρ -mixing coefficient $\rho(\mathcal{G}, \mathcal{H})$ for two sub- σ -fields \mathcal{G} and \mathcal{H} of \mathcal{F} is defined as

$$\rho(\mathcal{G}, \mathcal{H}) = \sup |\text{Cor}(U, V)|,$$

where the supremum is taken over all \mathcal{G} -measurable random variables U and all \mathcal{H} -measurable random variables V , both with finite second moment. The distance between two sets $A_1, A_2 \subseteq \mathbb{Z}^q$ is defined by $\text{dist}(A_1, A_2) = \inf\{\|a_1 - a_2\|_2 : a_1 \in A_1, a_2 \in A_2\}$, where $\|\cdot\|$ is the usual Euclidean vector norm.

It is also worth mentioning that [3] establish a strong invariance principle for partial sums of α -mixing random fields taking values in \mathbb{R}^p and (16), if the mixing coefficients, defined as

$$\alpha(\mathbf{E}_1, \mathbf{E}_2) = \sup_{\substack{A \in \sigma(\xi_{\mathbf{i}} : \mathbf{i} \in \mathbf{E}_1) \\ B \in \sigma(\xi_{\mathbf{i}} : \mathbf{i} \in \mathbf{E}_2)}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$$

for any two disjoint non-empty sets $\mathbf{E}_1, \mathbf{E}_2 \subseteq \mathbb{N}^q$, can be bounded in terms of the distance, namely by

$$\alpha(\mathbf{E}_1, \mathbf{E}_2) \leq C (\text{dist}(\mathbf{E}_1, \mathbf{E}_2))^{-q(1+\varepsilon)(1+2/\delta)}$$

for some $0 < \varepsilon < 1/2$ and $\delta > 0$ is such that $E|\xi_{\mathbf{0}}^{(j)}|^{2+\delta} < \infty$ holds. The result of [3] is limited to upper summation limits \mathbf{n} satisfying a constraint. That

constraint is, however, satisfied for the important special case of proportional sample sizes, i.e. $n_j = c_j n_1$, $j = 2, \dots, q$, holds for constants c_2, \dots, c_q . The strong invariance principle of [3] and (16) have been also established by [8] for random fields, which are weakly dependent in the sense that the covariance between $f(\xi_i : i \in \mathbf{I})$ and $g(\xi_i : i \in \mathbf{J})$, for any pair of disjoint finite sets \mathbf{I}, \mathbf{J} with sup-distance r , see (46) for a definition of the latter distance, and any pair of bounded Lipschitz functions f and g , is of the order θ_r , for a sequence θ_r decaying exponentially fast.

3. Threshold cut-off estimation of the asymptotic variance

As the simulation studies in Section 4.1 will show, the RMSE of the variance estimator $\hat{\sigma}_n^2$ increases for larger values of m . Moreover, the smallest possible RMSE in most situations is still quite high. This problem often arises when the spatial autocovariance matrix is sparse, i.e. the number of non-vanishing entries is much smaller than the number of null entries, or when it is close to sparsity. The latter typically occurs when the autocovariances decrease fast. This estimation of null entries increases the variance but does not reduce the bias. By thresholding sample autocovariances which are small in magnitude, we may reduce the variance.

We allow for parameterized thresholding functions and propose to determine remaining unknown parameters by a subsampling procedure, which implicitly determines an approximating m -dependent random field. For that purpose, we adopt a known result about the consistency of subsampling for random fields, see [22], to the setting of this paper.

3.1. Cut-off estimation

These observations motivate to propose a thresholding estimator for σ^2 , where the idea is to multiply $\hat{\gamma}_n(\mathbf{j})$ with some weight function g . This leads to the following estimator $\hat{\sigma}_{n,th}^2$, defined as

$$\hat{\sigma}_{n,th}^2 := \sum_{|\mathbf{j}| \leq m} w_m(\mathbf{j}) \hat{\gamma}_n(\mathbf{j}) g(\hat{\gamma}_n(\mathbf{j}), c_n(\mathbf{j})), \quad (17)$$

where g is a function depending on $\hat{\gamma}_n(\mathbf{j})$ and some sequence $c_n(\mathbf{j})$ and which satisfies the following assumption.

Assumption 2. Let $g: \mathbb{R} \times [0, \infty) \rightarrow [0, 1]$ be a bounded function such that

$$\mathbb{E}(g(\hat{\gamma}_n(\mathbf{j}), c_n(\mathbf{j}))) \rightarrow 1,$$

as $n \rightarrow \infty$.

The next theorem states some sufficient conditions for $\hat{\sigma}_{n,th}^2$ to be weakly consistent.

Theorem 3.1. *Under the conditions of Theorem 2.1 and Assumption 2 the thresholding estimator $\hat{\sigma}_{\mathbf{n},th}^2$ defined as in (17) is a weakly consistent estimator for σ^2 .*

For the special choice of g as

$$g(\hat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j})) = \mathbb{1}_{\{|\hat{\gamma}_{\mathbf{n}}(\mathbf{j})| > c_{\mathbf{n}}(\mathbf{j})\}}$$

we obtain the so-called cut-off estimator

$$\hat{\sigma}_{\mathbf{n},c}^2 := \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) \hat{\gamma}_{\mathbf{n}}(\mathbf{j}) \mathbb{1}_{\{|\hat{\gamma}_{\mathbf{n}}(\mathbf{j})| > c_{\mathbf{n}}(\mathbf{j})\}}, \tag{18}$$

which belongs to the subclass of hard-thresholding estimators, where only auto-covariances are taken into account whose absolute value exceeds the threshold $c_{\mathbf{n}}(\mathbf{j})$. Theorem 3.1 then simplifies to the following Corollary.

Corollary 3.1. *Under the conditions of Theorem 17 and if for each $\mathbf{j} \in \mathbb{Z}^q$, $c_{\mathbf{n}}(\mathbf{j}) \rightarrow c(\mathbf{j})$ for $\mathbf{n} \rightarrow \infty$ with $c(\mathbf{j}) < |\gamma(\mathbf{j})|$, then $\hat{\sigma}_{\mathbf{n},c}^2$ defined as in (18) is a weakly consistent estimator for σ^2 .*

The extension to multivariate random fields is as follows. Define the matrix-valued threshold estimator

$$\hat{\sigma}_{\mathbf{n},th}^2 := \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) \hat{\gamma}_{\mathbf{n}}(\mathbf{j}) \circ g(\hat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j})), \tag{19}$$

where for a $p \times p$ matrix $\mathbf{A} = (a_{\nu\mu})_{\substack{1 \leq \nu \leq p \\ 1 \leq \mu \leq p}}$ and a real number $c \geq 0$ we extend the function g to the domain $\mathbb{R}^{p \times p} \times [0, \infty)$ by setting

$$g(\mathbf{A}, c) = \left(g(a_{\nu\mu}, c) \right)_{\substack{1 \leq \nu \leq p \\ 1 \leq \mu \leq p}}.$$

In (19) the symbol \circ denotes the Hadamard product, i.e. the element-wise multiplication of two matrices of the same dimension. This means, we threshold all elements of the variance-covariance matrix using the same threshold $c_{\mathbf{n}}(\mathbf{j})$ depending on the lag \mathbf{j} .

Theorem 3.2. *Under the conditions of Theorem 2.1 and Assumption 2 the thresholding estimator $\hat{\sigma}_{\mathbf{n},th}^2$ for a multivariate random field, defined in (19), is a weakly consistent estimator for σ^2 , i.e.*

$$\|\hat{\sigma}_{\mathbf{n},th}^2 - \sigma^2\|_{\infty} \xrightarrow{\mathbb{P}} 0,$$

as $\mathbf{n} \rightarrow \infty$.

Motivated by our empirical studies presented in the next section, we propose to use rules of the form

$$c_{\mathbf{n}}(\mathbf{j}) = \frac{\left(\sqrt{j_1^2 + \dots + j_q^2} \right)^{\alpha}}{n_1 \cdots n_q} - \delta. \tag{20}$$

with $\alpha \geq 0$ and $\delta = 0.0001$. For this rule we have $c_{\mathbf{n}}(\mathbf{j}) \rightarrow -\delta =: c(\mathbf{j})$ for all $\mathbf{j} \in \mathbb{Z}^q$ and $\mathbf{n} \rightarrow \infty$, such that the condition $c(\mathbf{j}) < |\gamma(\mathbf{j})|$ of Corollary 3.1 is fulfilled in any case.

Let us consider a special choice of the lag truncation constants \mathbf{m} for the case $q = 2$ corresponding to image data, in order to clarify, for a valid choice of the lag truncation constants \mathbf{m} , possible values for α and the relationship of the lag-dependent truncation rule $c_{\mathbf{n}}(\mathbf{j})$ to a constant-in-lag truncation constant typically used in time series analysis. Hence, let us assume that $n_2 = fn_1$ for some known $f > 0$, the aspect ratio of the image, and let

$$m_i = c_i n_1^\gamma, \quad i = 1, 2,$$

for constants c_1, c_2 and some $0 < \gamma < 1/3$, such that the sufficient condition (10) of Theorem 2.1 is satisfied. Clearly, for larger γ a larger number of spatial autocovariances is taken into account. Then for all $|\mathbf{j}| \leq \mathbf{m}$ we have

$$c_{\mathbf{n}}(\mathbf{j}) \leq \frac{\left(\sqrt{m_1^2 + m_2^2}\right)^\alpha}{fn_1^2} - \delta = \frac{(c_1^2 + c_2^2)^{\alpha/2}}{f} n_1^{\alpha\gamma-2} - \delta.$$

If $\alpha = 2/\gamma$, then at the boundary ($\mathbf{j} = \mathbf{m}$) the constant (in sample size) threshold

$$c_{th} = \frac{(c_1^2 + c_2^2)^\alpha}{f} - \delta$$

applies, whereas for $\alpha < 2/\gamma$ the boundary threshold converges to $-\delta$, as $n_1 \rightarrow \infty$. For γ slightly smaller than $1/3$, in order to take into account a large number of autocovariances, the parameter α can be selected slightly larger than 6. This is in nice agreement with our empirical findings concerning the choice of α when aiming at precise estimation of the asymptotic variance σ^2 . However, when the estimated asymptotic variance is used to standardize an image test statistic, our simulations indicate that smaller values of α (around 3.6) are preferable, which are admissible for all admissible values of γ .

3.2. Data-adaptive estimation of lag truncation constants and optimal thresholds

The proposed estimators require to select the lag truncation constants \mathbf{m} and, for the cut-off threshold estimator, the thresholds $c_{\mathbf{n}}(\mathbf{j})$. Let us assume that the thresholds are given by a parametric family of functions with parameter α , i.e. $c_{\mathbf{n}}(\mathbf{j}) = c_{\mathbf{n}}(\mathbf{j}; \alpha)$, e.g. as in (20). To determine \mathbf{m} and α we propose to proceed as follows: To determine the optimal value \mathbf{m}_{opt} for the discrete-valued parameter \mathbf{m} from the set $\{k\mathbf{1} : k \in \mathbb{N}\}$, we apply a sequential testing procedure which analyzes the sample autocovariances on the unit spheres with respect to the maximum norm and stops when they are no longer statistically different from zero. The statistical tests are based on subsampling in order to determine critical values. This procedure has the nice property that it is consistent for m -dependent random fields. For other fields, applying this procedure may serve

as a guide, in order to approximate the random field under investigation by an m -dependent one, e.g. a spatial moving average model.

A reasonable measure to evaluate $\hat{\sigma}_{\mathbf{n},c}^2$ is, of course, the root mean squared error (RMSE), which is a function of \mathbf{m} and α . We propose to plug-in \mathbf{m}_{opt} and then to determine α by minimizing an estimate of the RMSE obtained by a subsampling procedure, too.

The subsampling-based proposal to deal with the estimators (4) and (18) is as follows. In accordance with the above notation introduced in Section 2.3, $Y_{\mathbf{j}}$ now stands for the subrandom field $\{\xi_{\mathbf{t}}, \mathbf{t} \in \mathbf{E}_{\mathbf{j},\mathbf{b},\mathbf{h}}\}$, and we write $\hat{\sigma}_{\mathbf{n},\mathbf{b},i}^2$ for the subsample value that equals the statistic $\hat{\sigma}_{\mathbf{b}}^2$ with constant weights equal to one evaluated at the random field Y_i , i.e. $\hat{\sigma}_{\mathbf{n},\mathbf{b},i}^2 = \hat{\sigma}_{\mathbf{b}}^2(Y_i)$ with $w_{\mathbf{m}}(\mathbf{j}) \equiv 1$. We choose the constant weights here as these do not depend on \mathbf{m} (and thus not on \mathbf{n}) to avoid additional difficulties when working with the smaller random fields of size \mathbf{b} .

In order to find a subsampling approximation for the RMSE we need to find a reasonable centering term. A natural choice is $\hat{\sigma}_{\mathbf{n}}^2$ itself. However, as the summation in (4) still depends on \mathbf{m} , we need to choose \mathbf{m} first. For random fields with $\gamma(\mathbf{j}) \geq 0$ we propose the following sequential testing procedure. For each subrandom field of size \mathbf{b} we start by considering the test statistic

$$R_{\mathbf{b}}(\mathbf{m}) := \sum_{\mathbf{j}:|\mathbf{j}_k|=m_k} \hat{\gamma}_{\mathbf{b}}(\mathbf{j}) \quad (21)$$

which estimates $r(\mathbf{m}) = \sum_{\mathbf{j}:|\mathbf{j}_k|=m_k} \gamma(\mathbf{j})$. We reject $H_0 : r(\mathbf{m}) = 0$ if the 90%-confidence interval for $r(\mathbf{m})$ obtained by subsampling does not contain 0. We propose to determine \mathbf{m}_{opt} by the following sequential testing procedure: Starting with $\mathbf{m} = \mathbf{1}$ and proceeding with $k\mathbf{1}$, $k = 1, 2, \dots$, we apply the above test until it rejects the associated null hypothesis $H_0 : r(\mathbf{m}') = 0$ for the first time. Then we put $\mathbf{m}_{opt} = \mathbf{m}' - \mathbf{1}$.

The consistency of the proposed subsampling testing procedure follows from Theorem 2.7 and Theorem 2.4, since the latter ensures that the root

$$K_{\mathbf{n}}(\mathbb{P}) = \mathcal{L} \left(|\mathbf{n}|^{-1/2} (R_{\mathbf{n}}(\mathbf{m}) - r(\mathbf{m})) \right)$$

converges weakly to a Gaussian random vector:

Corollary 3.2. *Under the conditions of Theorem 2.4 and if (13) and (14) hold,*

$$K_{\mathbf{n}}(\mathbb{P}) \Rightarrow K(\mathbb{P}) = \mathcal{L} \left(\sum_{|\mathbf{j}| \leq \mathbf{m}} B_{\mathbf{j}} \right),$$

as $\mathbf{n} \rightarrow \infty$, where $(B_{\mathbf{j}} : |\mathbf{j}| \leq \mathbf{m})$ is as in Theorem 2.4. If $q(\gamma)$ denotes the γ -quantile of $K(\cdot, \mathbb{P})$ and $q_{\mathbf{n},\mathbf{b}}(\gamma)$ the γ -quantile of the empirical subsampling distribution for the statistic $R(\mathbf{m})$, then the asymptotic coverage of the confidence interval

$$[R_{\mathbf{n}}(\mathbf{m}) - |\mathbf{n}|^{-1/2} q_{\mathbf{n},\mathbf{b}}(\gamma), R_{\mathbf{n}}(\mathbf{m}) + |\mathbf{n}|^{-1/2} q_{\mathbf{n},\mathbf{b}}(\gamma)]$$

is the nominal confidence level 2γ , $\gamma \in (0, 1/2)$.

Once we have determined \mathbf{m}_{opt} , we can evaluate $\hat{\sigma}_{\mathbf{n}}^2$ for \mathbf{m}_{opt} and use $\hat{\sigma}_{\mathbf{n},\mathbf{m}_{opt}}^2$ as centering term for the subsampling approximation for the RMSE given a block size \mathbf{b} , whose selection is discussed and studied in Section 4.2. This approach finally leads to the approximation

$$\widehat{\text{RMSE}}(\hat{\sigma}_{\mathbf{n}}^2) := \sqrt{N^{-1} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_q=1}^{N_q} \left(\hat{\sigma}_{\mathbf{n},\mathbf{b},i}^2 - \hat{\sigma}_{\mathbf{n},\mathbf{m}_{opt}}^2 \right)^2}, \quad (22)$$

where $\hat{\sigma}_{\mathbf{n},\mathbf{b},i}^2$, $i \in \mathbf{1} : \mathbf{N}$, are the subsampling replicates and $\mathbf{N} = (N_1, \dots, N_q)$. That approximation is now minimized to determine the tuning parameter α .

4. Simulation studies

In this section we present simulations about the proposed methods. Sections 4.1 to 4.2 provide a thorough analysis of the variance estimators of Sections 2 and 3, where for the latter we will focus our attention on the cut-off estimator. The results show that thresholding can improve in terms of the RMSE and, especially, is more robust with respect to the choice of the lag truncation constants \mathbf{m} . Then we investigate the proposed data-adaptive procedure to estimate the remaining unknown tuning parameters of these estimators via subsampling.

Section 4.3 provides results about the accuracy of significance tests for image data based on partial sums standardized with the proposed cut-off estimator. The simulations show that highly accurate testing is achieved by our proposal, even in the presence of substantial image correlations.

4.1. Simulation results for the variance estimators

Having in mind applications in imaging, where quite often spatial dependencies are local, we investigate the statistical behavior of the variance estimators $\hat{\sigma}_{\mathbf{n}}^2$ and $\hat{\sigma}_{\mathbf{n},c}^2$ for selected stationary two- and three-dimensional random fields, including an autoregressive model to take into account serial correlations present in sequences of images.

Further, we investigate both estimators for different weighting functions and different values of \mathbf{m} and \mathbf{n} and analyse the resulting behaviour of the root mean square error and the bias. Finally, we will also analyse the second estimator dependent on the sequence $c_{\mathbf{n}}(\mathbf{j})$ and make a concrete proposal for its selection.

4.1.1. Spatial moving average models of order one in two dimensions

The first model for the random field for which we want to analyse the behaviour of the estimators is a spatial moving average model of order one in two dimensions. For that purpose, take i.i.d. innovations $\eta_{i,j}$ with $\eta_{i,j} \sim \mathcal{N}(0, 1)$ for all i

and j and define

$$(M1) \quad \varepsilon_{i,j} := a_1\eta_{i-1,j-1} + a_2\eta_{i-1,j} + a_3\eta_{i-1,j+1} + a_4\eta_{i,j-1} \\ + a_5\eta_{i,j} + a_6\eta_{i,j+1} + a_7\eta_{i+1,j-1} + a_8\eta_{i+1,j} + a_9\eta_{i+1,j+1}$$

with real weights a_k , $k = 1, \dots, 9$. We first take $a_5 = 1$ and $a_k = 0.3$ for $k \neq 5$. We can now calculate the theoretical asymptotic variance as follows. If we write $\mathbf{a} = (a_1, \dots, a_9)'$ for the column vector with the weights a_k , $k = 1, \dots, 9$, and

$$\boldsymbol{\eta}_{i,j} = (\eta_{i-1,j-1}, \eta_{i-1,j}, \eta_{i-1,j+1}, \eta_{i,j-1}, \eta_{i,j}, \eta_{i,j+1}, \eta_{i+1,j-1}, \eta_{i+1,j}, \eta_{i+1,j+1})'$$

for the column vector with the innovation $\eta_{i,j}$ and all the innovations that are located around it, we can rewrite (M1) as scalar product of the vectors \mathbf{a} and $\boldsymbol{\eta}_{i,j}$, i.e. as $\varepsilon_{i,j} = \mathbf{a}'\boldsymbol{\eta}_{i,j}$. Thus, we obtain

$$\sigma^2 = \sum_{(h_1, h_2) \in \mathbb{Z}^2} \mathbb{E}(\varepsilon_{0,0}\varepsilon_{h_1, h_2}) = \sum_{(h_1, h_2) \in \mathbb{Z}^2} \mathbb{E}((\mathbf{a}'\boldsymbol{\eta}_{0,0})(\mathbf{a}'\boldsymbol{\eta}_{h_1, h_2})) = \sum_{i,j=1}^9 a_i a_j, \tag{23}$$

as only $9^2 = 81$ summands in the above summation over $(h_1, h_2) \in \mathbb{Z}^2$ are non-zero, namely one for each combination of a_i and a_j , $i, j = 1, \dots, 9$. For our concrete vector \mathbf{a} one can thus calculate that the theoretical variance is $\sigma^2 = 11.56$.

Throughout this study we choose the weights $w_{\mathbf{m}}(\mathbf{j})$, that are needed for the calculation of $\hat{\sigma}_{\mathbf{n}}^2$ in (4), as the product of one-dimensional weights, where we take the $w_{m_i}(j_i)$, $1 \leq i \leq 2$, either as constant weights (CW) with $w_{m_i}(j_i) = 1$, $1 \leq i \leq 2$, or as the Quadratic Spectral weight sequence (QS), i.e.

$$w_{m_i}(j_i) = \frac{25}{12\pi^2 \left(j_m^{(i)}\right)^2} \left(\frac{\sin(6\pi j_m^{(i)}/5)}{6\pi j_m^{(i)}/5} - \cos(6\pi j_m^{(i)}/5) \right)$$

for $1 \leq i \leq 2$ and $j_m^{(i)} = j_i/(m_i + b_w)$, where b_w is a bandwidth parameter that still has to be chosen. In dimension one [2] could show that the QS kernel with a properly chosen bandwidth b_w is the optimal choice with respect to the asymptotic MSE, see Theorem 2 in [2]. In his simulation studies, however, the author could also show that in a lot of situations the constant weights lead to more efficient estimates than the quadratic spectral weights. The disadvantage of the constant weights, however, is that in contrast to the QS weights they do not necessarily generate nonnegative variance estimates.

Now, we first consider a random field of size (30,40) and investigate our estimator in (4) for several values of $\mathbf{m} = (m_1, m_2) \in \mathcal{M}_2$ with

$$\mathcal{M}_2 = \{(m_1, m_2) \in \{0, \dots, 29\}^2 : m_1 = m_2\} \\ \cup \{(1, 2), (2, 3), (3, 4), (4, 5), (5, 6), (6, 7), (10, 13), (15, 20)\}. \tag{24}$$

\mathbf{m}	(0,0)	(1,1)	(2,2)	(3,3)	(4,4)
Mean($\hat{\sigma}_{\mathbf{n}}^2$)	1.7217	8.6957	11.5924	11.5879	11.5588
RMSE($\hat{\sigma}_{\mathbf{n}}^2$)	9.8391	2.9996	1.8478	2.8328	3.8371

\mathbf{m}	(5,5)	(6,6)	(7,7)	(8,8)	(9,9)
Mean($\hat{\sigma}_{\mathbf{n}}^2$)	11.5290	11.5233	11.5297	11.5304	11.5165
RMSE($\hat{\sigma}_{\mathbf{n}}^2$)	4.8846	5.9867	7.1408	8.3481	9.5736

TABLE 1

Simulation results for (M1) for different values of \mathbf{m} with constant kernel.

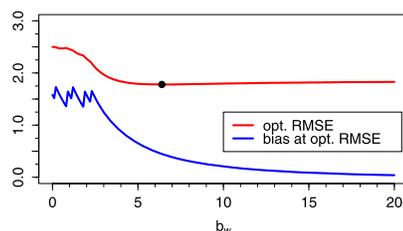


FIG 1. Optimal RMSE with respect to \mathbf{m} and corresponding bias of (M1) for bandwidths b_w .

Note, that \mathcal{M}_2 mainly consists of pairs with equal components. These pairs are quite natural choices for \mathbf{m} , as (M1) is a symmetric model where the influence of the innovations around $\eta_{i,j}$ is the same in both directions. Nevertheless, we also consider some pairs where the components differ, including the cases (10,13) and (15,20) which correspond to a third and a half of the random field of size (30,40), respectively.

The first table shows the values of the estimator and its RMSE for selected choices of $\mathbf{m} \in \mathcal{M}_2$ for the constant kernel based on 10000 repetitions.

Table 1 shows that the RMSE, seen as a function of \mathbf{m} , is convex with a minimum in $\mathbf{m} = (2,2)$, which corresponds to the largest lag for which the theoretical autocovariance in (M1) is non-zero. We also see that for values of \mathbf{m} larger than (2,2) the RMSE increases quite fast while the bias does not vary much. This implies that the variance increases.

Next, we also examine our estimator for the QS kernel, for which we still have to choose the bandwidth parameter b_w which we choose the same in all dimensions. We want to choose this parameter in such a way that the RMSE gets as small as possible. Figure 1 depicts for different values of b_w the optimal RMSE with respect to \mathbf{m} (i.e. when \mathbf{m} ranges over the values of \mathcal{M}_2) and the corresponding bias.

To abbreviate the notation we write $\mathbf{m} + b_w$ for $(m_1, m_2) + (b_w, b_w)$. We see that the smallest RMSE is achieved for $b_w = 6.4$ which corresponds to $\mathbf{m} = (2,2)$. Thus, we will fix the bandwidth at this value. Furthermore, we can also see that for $b_w \geq 4$ the RMSE stays nearly fixed such that different choices of b_w near the optimal one would lead to nearly as good estimates as for the optimal bandwidth $b_w = 6.4$. The jags in the bias for small choices of the

\mathbf{m}	(0,0)	(1,1)	(2,2)	(3,3)	(4,4)
Mean($\hat{\sigma}_n^2$)	1.7217	8.4380	11.1120	11.2030	11.2502
RMSE($\hat{\sigma}_n^2$)	9.8391	3.2388	1.7793	2.5630	3.3187

\mathbf{m}	(5,5)	(6,6)	(7,7)	(8,8)	(9,9)
Mean($\hat{\sigma}_n^2$)	11.2799	11.3132	11.3471	11.3741	11.3899
RMSE($\hat{\sigma}_n^2$)	4.0534	4.7795	5.4992	6.2172	6.9214

TABLE 2

Simulation results for (M1) for different values of \mathbf{m} with QS kernel and bandwidth $\mathbf{m} + 6.4$.

bandwidth parameter arise from the fact, that we allow \mathbf{m} to range over all the values of \mathcal{M}_2 . A jag occurs, when the value of \mathbf{m} , for which the smallest RMSE is attained, changes. For example, for $b_w = 0$ the optimal \mathbf{m} is (4,5) whereas for $b_w \geq 2.3$ it is (2,2), which is again the largest lag for which the theoretical autocovariance in (M1) is non-zero. For $0 < b_w \leq 2.3$ the optimal value of \mathbf{m} varies.

Table 2 shows the values of the estimator and its RMSE in the above scenario of (M1) for selected choices of \mathbf{m} for the QS kernel with $b_w = 6.4$ for 10000 repetitions.

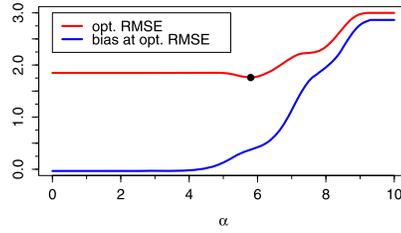
The smallest RMSE is now 1.7793 and again achieved for $\mathbf{m} = (2,2)$. Similar as in Table 1 the RMSE increases for values of \mathbf{m} larger than (2,2), but not as rapid as in Table 1. Thus, we see that in this scenario the QS kernel with a reasonably chosen bandwidth is nearly 4% more efficient than the constant kernel which is in accordance with the simulation results in [2].

As this simulation study shows, the smallest RMSE is still quite high and what is even more problematic, the RMSE depends quite heavily on the proper choice of \mathbf{m} . As already explained in the motivation for the improved variance estimator, this problem often appears when a lot of ‘zeros’ have to be estimated. This is the case here, as the dependencies of the random field are only weak, as the autocovariances vanish for $|j| > 2$. The estimation of many such null entries worsens the RMSE a lot. Thus we shall now investigate the improved variance estimator $\hat{\sigma}_{n,c}^2$ in (18) which is designed to circumvent this issue. For that we need to choose an appropriate cutting rule $c_n(\mathbf{j})$.

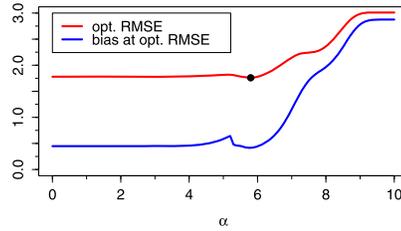
We focus on the rule $c_n(\mathbf{j})$ introduced in the previous section, which is parameterized by $\alpha > 0$. The parameter α is a further tuning parameter that we want to choose in such a way that the RMSE of $\hat{\sigma}_{n,c}^2$ attains the smallest possible value when \mathbf{m} ranges again over the values of \mathcal{M}_2 .

In the following, we investigate the same situation as before, the spatial moving average model (M1) with weights $a_5 = 1$ and $a_i = 0.3$ for $i \neq 5$. Figure 2 shows for both weighting schemes the curves of the optimal RMSE and the corresponding bias as a function of α .

We can see that the curves of the RMSE and the bias are very similar to each other. If α is too small the thresholds $c_n(\mathbf{j})$ do not lead to an improvement of the optimal RMSE. However, for $\alpha \in [5.0, 6.2]$ and constant weights, and for $\alpha \in [5.6, 6.0]$ and QS weights, one has slight improvements of the optimized



(a) Constant kernel.



(b) QS kernel.

FIG 2. Optimal RMSE and corresponding bias of (M1).

RMSE. Here the optimal choice for α is $\alpha = 5.8$ for both kernels. If $\alpha > 6.2$ and $\alpha > 6.0$ respectively the optimized RMSE gets worse than without cutting. The jags in the bias arise again due to the fact that we do not fix \mathbf{m} but allow it to vary over \mathcal{M}_2 . A jag occurs when the value of \mathbf{m} for which the optimal RMSE is attained changes.

We can, however, see more easily the improvement of the estimator $\hat{\sigma}_{\mathbf{n},c}^2$ upon $\hat{\sigma}_{\mathbf{n}}^2$ when we inspect tables similar to 1 and 2, but with $\hat{\sigma}_{\mathbf{n}}^2$ replaced by $\hat{\sigma}_{\mathbf{n},c}^2$ and a cutting rule of the form (20) with $\alpha = 5.8$. Tables 3 and 4 show, that the RMSE for both kernels levels off at approximately 1.76, which is in both cases an improvement of the RMSE of the estimator without cutting rule, though this improvement is higher for the estimator with constant weights. This shows that the proper choice of \mathbf{m} is now of minor importance than before, since now each choice of \mathbf{m} , that is larger or equal than the largest lag \mathbf{j} for which the theoretical autocovariances are non-zero, leads to nearly the same RMSE. If we had chosen a different α than the optimal one, but one close to it, we would not have improved the optimal RMSE, but nevertheless the effect for \mathbf{m} larger than (2,2) would have been the same.

We can observe a further interesting fact by inspecting Table 5. We have already obtained the results of the last two columns of Table 5 before. What is new now are the results of the first column. If we take $b_w = 0$ and thus have $j_m^{(i)} = j_i/m_i, i = 1, 2$, as argument in the QS kernel, the smallest RMSE, that is attained for a random field of size (30,40) in model (M1), is 2.4996 for

\mathbf{m}	(0,0)	(1,1)	(2,2)	(3,3)	(4,4)
Mean($\hat{\sigma}_{\mathbf{n},c}^2$)	1.7217	8.6957	11.1815	11.1815	11.1815
RMSE($\hat{\sigma}_{\mathbf{n},c}^2$)	9.8391	2.9996	1.7597	1.7597	1.7597

TABLE 3

Simulation results for (M1) for different values of \mathbf{m} with constant weights, $\alpha = 5.8$.

\mathbf{m}	(0,0)	(1,1)	(2,2)	(3,3)	(4,4)	(29,29)
Mean($\hat{\sigma}_{\mathbf{n},c}^2$)	1.7217	8.4380	10.7599	10.8431	10.9041	11.1572
RMSE($\hat{\sigma}_{\mathbf{n},c}^2$)	9.8391	3.2388	1.8038	1.7867	1.7768	1.7594

TABLE 4

Simulation results for (M1) for different values of \mathbf{m} with QS kernel, $\alpha = 5.8$.

$\mathbf{m} = (4,5)$. If we now optimize the cutting rule (20) for this situation as above, we obtain again $\alpha = 5.8$ as the optimal value for α . This leads to a smallest possible RMSE of 1.7593 for $\mathbf{m} = (28,28)$. Thus we see that the estimator $\hat{\sigma}_{\mathbf{n},c}^2$ with $\alpha = 5.8$ and a QS kernel with bandwidth $b_w = 0$ leads to a smaller RMSE than the estimator $\hat{\sigma}_{\mathbf{n}}^2$ with a QS kernel and bandwidth $b_w = 6.4$ or than the estimator $\hat{\sigma}_{\mathbf{n}}^2$ with constant weights and to an equally good RMSE if we further use the cutting rule. Similar observations also hold true for larger sample sizes.

Finally, we also want to investigate the behaviour of both estimators for both kernels for different sample sizes when we choose different weights for a_k . So, in the following let $a_5 = 1$ and $a_k = a$ for $k \neq 5$. We again fix $\alpha = 5.8$ and $b_w = 6.4$. This time we do not only calculate the optimal RMSE (denoted by RMSE_{opt}), the corresponding bias and the value of \mathbf{m} for which RMSE_{opt} is attained, but also the quotient $\text{RMSE}_{29}/\text{RMSE}_{opt}$, where RMSE_{29} is the RMSE of the estimator for $\mathbf{m} = (29,29)$. This is to get an impression of how much the RMSE can vary when we choose \mathbf{m} too large in the case that we do not use a cutting rule. Table 6 shows the corresponding simulation results.

The table shows that both kernels compete reasonably well with each other. The estimator $\hat{\sigma}_{\mathbf{n}}^2$ with a QS kernel, for example, ranges from being 6% less efficient to 4% more efficient than the estimator $\hat{\sigma}_{\mathbf{n}}^2$ with a constant kernel in the case of a random field of size (30,40) with different weights a . Similar observations hold true for the other cases. In some scenarios the optimized RMSE of $\hat{\sigma}_{\mathbf{n},c}^2$ is worse than the one for $\hat{\sigma}_{\mathbf{n}}^2$ as we do not vary the value of α ,

kernel	QS, $b_w = 0$	QS, $b_w = 6.4$	CW
RMSE($\hat{\sigma}_{\mathbf{n}}^2$)	2.4996	1.7793	1.8478
bias($\hat{\sigma}_{\mathbf{n}}^2$)	-1.5802	-0.4480	0.0324
\mathbf{m}	(4,5)	(2,2)	(2,2)
RMSE($\hat{\sigma}_{\mathbf{n},c}^2$)	1.7593	1.7593	1.7597
bias($\hat{\sigma}_{\mathbf{n},c}^2$)	-0.4173	-0.4162	-0.3785
\mathbf{m}	(28,28)	(22,22)	(2,2)
impr.(%)	29.62	1.12	4.76

TABLE 5

Simulation results of both estimators for (M1) for both kernels and different bandwidths and $\alpha = 5.8$.

weights a	0.01	0.1	0.3	0.5	0.7	0.9
$\hat{\sigma}_{\mathbf{n},c}^2$	1.1664	3.24	11.56	25	43.56	67.24
RMSE _{opt} (CW)	0.1367	0.4272	1.8478	3.9033	6.7235	10.3083
bias (CW)	-0.0015	-0.3154	0.0324	0.0681	0.1171	0.1792
\mathbf{m} (CW)	(1,1)	(1,1)	(2,2)	(2,2)	(2,2)	(2,2)
RMSE ₂₉ /RMSE _{opt}	54.5744	42.6213	33.1515	33.4509	33.6324	33.7522
RMSE _{opt} (CW, cut)	0.1367	0.4272	1.7597	3.9650	6.8246	10.3694
bias (CW, cut)	-0.0015	-0.3154	-0.3785	-0.3899	-0.0901	0.0933
\mathbf{m} (CW, cut)	(1,1)	(1,1)	(2,2)	(2,2)	(2,2)	(2,2)
RMSE ₂₉ /RMSE _{opt}	1.1185	1.0494	1.0000	1.0127	1.1006	1.2141
RMSE _{opt} (QS)	0.1324	0.4543	1.7793	3.8008	6.5890	10.1438
bias (QS)	-0.0076	-0.2483	-0.4480	-1.1059	-2.0494	-3.2787
\mathbf{m} (QS)	(1,1)	(1,2)	(2,2)	(2,2)	(2,2)	(2,2)
RMSE ₂₉ /RMSE _{opt}	21.6538	16.5137	14.6394	14.7177	14.7478	14.7626
RMSE _{opt} (QS, cut)	0.1324	0.4490	1.7593	3.9650	6.7261	10.2171
bias (QS, cut)	-0.0078	-0.1991	-0.4162	-1.4951	-2.2255	-3.3517
\mathbf{m} (QS, cut)	(1,1)	(29,29)	(22,22)	(2,2)	(2,2)	(2,2)
RMSE ₂₉ /RMSE _{opt}	1.1525	1.0000	1.0000	1.0102	1.1115	1.2253

TABLE 6

Simulation results for (M1) for constant / QS kernel, $\mathbf{n} = (30,40)$, $\alpha = 5.8$, and different weights a .

but keep it fix as $\alpha = 5.8$ which was the value of α adapted to the situation $\mathbf{n} = (30,40)$ with $a_k = 0.3$ for $k \neq 5$. However, the main advantage of the estimator $\hat{\sigma}_{\mathbf{n},c}^2$ has again to be seen in the fact that it stabilizes the RMSE for values of \mathbf{m} larger than (2,2). This can be seen by the fact that for $\hat{\sigma}_{\mathbf{n},c}^2$ the quotient RMSE₂₉/RMSE_{opt} is close to one in all cases. For example, for $\mathbf{n} = (30,40)$, $a_k = 0.7$ for $k \neq 5$, and the constant kernel the optimized RMSE worsens from 6.7235 for $\hat{\sigma}_{\mathbf{n}}^2$ to 6.8246 for $\hat{\sigma}_{\mathbf{n},c}^2$, while the quotient RMSE₂₉/RMSE_{opt} improves from 33.6324 to 1.1006. So regarding the RMSE we see that for $\hat{\sigma}_{\mathbf{n},c}^2$ it does not make a big difference if we choose \mathbf{m} as (2,2) or (29,29) or as some value in between, whereas for $\hat{\sigma}_{\mathbf{n}}^2$ it does. The same is also true in the case that we use a QS kernel, with the only difference that for $\hat{\sigma}_{\mathbf{n}}^2$ the quotient RMSE₂₉/RMSE_{opt} equals 14.7478 which is a lot smaller than 33.6324 in the case of a constant kernel. This observation also holds true for all other scenarios, i.e. for $\hat{\sigma}_{\mathbf{n}}^2$ this quotient is larger if we use the constant kernel instead of the QS kernel. Naturally, we could further improve the simulation results of the estimators with QS weights if we chose the bandwidth b_w differently in each scenario. Similarly, we could also improve the simulation results of $\hat{\sigma}_{\mathbf{n},c}^2$ with an α chosen differently in each scenario.

4.1.2. Spatial moving average models of order three in two dimensions

We also investigate the behaviour of the estimators for stronger dependence structures such as a spatial moving average model of order three. The model, (M2), used here is similar to (M1), but now with weights 1, a_1, a_2 and a_3 for the center pixel and the pixels in the first, second and third ring around it instead of only giving weights to the center pixel and the pixels in the first

ring around it. If $H_{i,j}$ is the (7×7) -matrix with the column by column entries $\eta_{i-3,j-3}, \eta_{i-3,j-2}, \eta_{i-3,j}, \dots, \eta_{i,j}, \dots, \eta_{i+3,j+1}, \eta_{i+3,j+2}, \eta_{i+3,j+3}$ and write $vec(\cdot)$ for the vectorization of a matrix, which transforms a matrix into a vector by writing the columns on top of one another, we can formulate model (M2) as

$$(M2) \quad \varepsilon_{i,j} = (vec(W))' vec(H_{i,j}).$$

Our findings are in agreement with model (M1), in particular the threshold estimator again stabilizes the RMSE for large values of \mathbf{m} , and are therefore deferred to the supplement.

4.1.3. *Autoregressive spatial moving average mixture models in three dimensions*

The last model that we consider is a mixture between an autoregressive model of order one in the time domain and a moving average model of order two in the spatial domain. More specifically we put for $\rho \in (-1, 1)$

$$(M4) \quad \varepsilon_{t,i,j} = X_{t,i,j} + v_{t,i,j}, \quad X_{t,i,j} = \rho X_{t-1,i,j} + u_{t,i,j}$$

where $u_{t,i,j} \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ for all $t, i, j \in \mathbb{Z}$ and the $v_{t,i,j}$ follow for each fixed t model (M1) and are uncorrelated for different values of t . Moreover, we suppose that u_{t_1, i_1, j_1} and v_{t_2, i_2, j_2} are uncorrelated for all $t_1, t_2, i_1, i_2, j_1, j_2 \in \mathbb{Z}$.

We now calculate the asymptotic variance as follows. Note, that by the one-dimensional theory for AR(1)-models in the time series literature we get for the $X_{t,i,j}$ the representation as a linear process by

$$X_{t,i,j} = \sum_{k=0}^{\infty} \rho^k u_{t-k,i,j}.$$

For $(h_t, h_i, h_j) \in \mathbb{Z}^3$ we then obtain

$$\begin{aligned} \mathbb{E}(\varepsilon_{t,i,j} \varepsilon_{t+h_t, i+h_i, j+h_j}) &= \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \rho^{k+l} \mathbb{E}(u_{t-k,i,j} u_{t+h_t-l, i+h_i, j+h_j}) \\ &+ \mathbb{E}(v_{t,i,j} v_{t+h_t, i+h_i, j+h_j}) \\ &+ \sum_{k=0}^{\infty} \rho^k \mathbb{E}(u_{t-k,i,j} v_{t+h_t, i+h_i, j+h_j}) \\ &+ \sum_{l=0}^{\infty} \rho^l \mathbb{E}(u_{t+h_t-l, i+h_i, j+h_j} v_{t,i,j}) \\ &=: E_1 + E_2 + E_3 + E_4. \end{aligned}$$

E_3 and E_4 equal zero as u_{t_1, i_1, j_1} and v_{t_2, i_2, j_2} are uncorrelated by assumption. E_1 is only non-zero if $h_i = h_j = 0$. In this case we get

$$E_1 = \sum_{k=0}^{\infty} \rho^{2k+h_t} = \frac{\rho^{h_t}}{1 - \rho^2}.$$

E_2 , on the contrary, is only non-zero if $h_t = 0$ and it then equals the autocovariances of model (M1). We thus obtain

$$\sigma^2 = \sum_{(h_t, h_i, h_j) \in \mathbb{Z}^3} \mathbb{E}(\varepsilon_{t,i,j} \varepsilon_{t+h_t, i+h_i, j+h_j}) = \frac{1}{(1-\rho)^2} + \sum_{i,j=1}^9 a_i a_j,$$

which corresponds to the sum of the variance caused by the AR(1)-model in the time domain and the asymptotic variance of (M1). In our first simulation setting for (M4) we take $\rho = 0.2$, $a_5 = 1$ and $a_i = 0.3$ for $i \neq 5$ and consequently have $\sigma^2 = 1.5625 + 11.56 = 13.1225$. Now, we first consider a random field of size (20,30,40) where the first component corresponds to the time domain, and investigate the estimator in (4) for several values of $\mathbf{m} = (m_1, m_2, m_3) \in \mathcal{M}_3 \cup \{(1, 2, 2)\}$ with \mathcal{M}_3 defined as

$$\mathcal{M}_3 = \{(m_1, m_2, m_3) \in \{0, \dots, 15\}^3 : m_1 = m_2 = m_3\}. \quad (25)$$

Note that this time we include the triple (1,2,2), as this is the largest lag with a significant contribution to the asymptotic variance of model (M4).

Table 7 shows the corresponding simulation results for the constant kernel using 10000 repetitions.

\mathbf{m}	(0,0,0)	(1,1,1)	(1,2,2)	(2,2,2)	(3,3,3)	(4,4,4)
Mean($\hat{\sigma}_n^2$)	2.7079	10.0659	12.9475	13.0239	13.0259	12.9901
RMSE($\hat{\sigma}_n^2$)	10.4146	3.0834	0.8432	1.0995	2.0106	3.1359

\mathbf{m}	(5,5,5)	(8,8,8)	(10,10,10)	(12,12,12)	(15,15,15)
Mean($\hat{\sigma}_n^2$)	12.9906	12.9402	12.8049	12.7995	12.7426
RMSE($\hat{\sigma}_n^2$)	4.4694	9.9540	15.0064	21.3540	34.7365

TABLE 7

Simulation results for (M4) for different values of \mathbf{m} with constant kernel.

We can see that the smallest RMSE is attained for $\mathbf{m} = (1,2,2)$ which corresponds in the spatial domain to the largest lag with non-zero autocovariances and in the time domain to the largest lag with a significant contribution to the asymptotic variance as the autoregressive model is of order one. Greater values of \mathbf{m} lead to a rapid increase of the RMSE while the bias stays stable.

In the case of QS weights we first have to determine the bandwidth parameter b_w . This time we allowed b_w to vary from 0 to 40, but for these values no minimum is attained (except on the boundary for $b_w = 40$). On the contrary, the RMSE decreases slowly when b_w increases. As, however, for $b_w \geq 20$ the RMSE decreases so slowly that visually it is constant, we take $b_w = 20$ from now on. Note that in the previous three simulation models a minimal RMSE with respect to b_w was always attained, but in all these cases the RMSE was also nearly constant around the minimum. Table 8 shows the corresponding simulation results.

We see that also for the QS kernel the smallest RMSE is achieved for $\mathbf{m} = (1,2,2)$ which is now slightly larger than for the constant kernel. However, if we

\mathbf{m}	(0,0,0)	(1,1,1)	(1,2,2)	(2,2,2)	(3,3,3)	(4,4,4)
Mean($\hat{\sigma}_{\mathbf{n}}^2$)	2.7079	10.0322	12.8747	12.9503	12.9588	12.9309
RMSE($\hat{\sigma}_{\mathbf{n}}^2$)	10.4146	3.1165	0.8515	1.0910	1.9561	2.9977
\mathbf{m}	(5,5,5)	(8,8,8)	(10,10,10)	(12,12,12)	(15,15,15)	
Mean($\hat{\sigma}_{\mathbf{n}}^2$)	12.9362	12.9115	12.8108	12.8094	12.7996	
RMSE($\hat{\sigma}_{\mathbf{n}}^2$)	4.1905	8.7241	12.5226	16.9393	25.3493	

TABLE 8
Simulation results for (M_4) for different values of \mathbf{m} with QS kernel.

chose a larger value for b_w we could probably get a smaller RMSE as in the previous simulation settings.

Next, we want to investigate how the improved variance estimator $\hat{\sigma}_{\mathbf{n},c}^2$ with the cutting rule

$$c_{\mathbf{n}}(\mathbf{j}) = \frac{\left(\sqrt{j_1^2 + j_2^2 + j_3^2}\right)^\alpha}{n_1 n_2 n_3} - \delta \quad (26)$$

with $\alpha \geq 0$ and $\delta = 0.0001$, behaves in this model. If we optimize again the RMSE with respect to α , we obtain for the constant as well as for the QS weights an optimal value of α as $\alpha = 9.4$. Due to very high computational costs we allowed \mathbf{m} only to vary up to values of \mathcal{M}_3 smaller than $\mathbf{m} = (7,7,7)$ which does not effect the results when using constant weights, but does not lead to the smallest possible RMSE when using QS weights. The corresponding results are gathered in Tables 9 and 10.

\mathbf{m}	(0,0,0)	(1,1,1)	(1,2,2)	(2,2,2)	(3,3,3)
Mean($\hat{\sigma}_{\mathbf{n},c}^2$)	2.7079	10.0659	12.5863	12.6569	12.6569
RMSE($\hat{\sigma}_{\mathbf{n},c}^2$)	10.4146	3.0832	0.7600	0.7169	0.7169

TABLE 9
Simulation results for (M_4) for different values of \mathbf{m} with constant kernel, $\alpha = 9.4$.

\mathbf{m}	(0,0,0)	(1,1,1)	(1,2,2)	(2,2,2)	(3,3,3)	(15,15,15)
Mean($\hat{\sigma}_{\mathbf{n},c}^2$)	2.7079	10.0322	12.5219	12.5917	12.5972	12.6310
RMSE($\hat{\sigma}_{\mathbf{n},c}^2$)	10.4146	3.1163	0.8040	0.7578	0.7542	0.7327

TABLE 10
Simulation results for (M_4) for different values of \mathbf{m} with QS kernel, $\alpha = 9.4$.

We see that for both weighting schemes we again have the stabilization property of the RMSE for values of \mathbf{m} larger than $\mathbf{m} = (2,2,2)$. For the constant weights the RMSE levels off at approximately 0.71 and for the QS weights between 0.73 and 0.76. But what is even more appealing now is that we can actually improve the optimized RMSE by almost 15% in the case of constant weights and by nearly 14% in the case of QS weights. The reason for this is that in this model the autocovariances for \mathbf{m} larger than $(1,2,2)$ are close to zero, but not identically zero as in the previous simulation settings.

Finally, we want to investigate if we still have an improvement of the optimized RMSE if we take the cutting rule with the optimal $\alpha = 9.4$ in cases where we change the weights of the autoregressive parameter ρ in model (M4). All other parameters are chosen as before.

Table 11 shows that this is indeed the case. For $\rho = 0.4$, for example, we have an improvement of around 17% for both kernels. If we optimized α to this situation we would probably obtain an even higher improvement. Also the other values for ρ lead to an improvement of the optimized RMSE of nearly or even more than 10%.

weights ρ	0.01	0.1	0.2	0.3	0.4	0.5
σ^2	12.5803	12.7946	13.1225	13.6008	14.3378	15.56
RMSE _{opt} (CW)	0.8019	0.8143	0.8432	0.9233	1.1546	1.5319
bias (CW)	-0.0449	-0.0779	-0.1750	-0.3749	-0.7595	-0.8961
\mathbf{m} (CW)	(1,2,2)	(1,2,2)	(1,2,2)	(1,2,2)	(1,2,2)	(2,2,2)
RMSE _{opt} (CW, cut)	0.6625	0.6821	0.7169	0.7837	0.9573	1.3909
bias (CW, cut)	-0.4065	-0.4254	-0.4656	-0.5506	-0.7644	-1.2529
\mathbf{m} (CW, cut)	(1,2,2)	(2,2,2)	(2,2,2)	(2,2,2)	(2,2,2)	(2,2,2)
impr. (%)	17.38	16.24	14.98	15.12	17.09	9.21
RMSE _{opt} (QS)	0.7993	0.8145	0.8515	0.9463	1.1971	1.5658
bias (QS)	-0.1166	-0.1501	-0.2478	-0.4485	-0.8339	-0.9784
\mathbf{m} (QS)	(1,2,2)	(1,2,2)	(1,2,2)	(1,2,2)	(1,2,2)	(2,2,2)
RMSE _{opt} (QS, cut)	0.6894	0.7073	0.7437	0.8137	0.9931	1.4341
bias (QS, cut)	-0.4489	-0.4679	-0.5090	-0.5953	-0.8109	-1.3021
\mathbf{m} (QS, cut)	(7,7,7)	(7,7,7)	(7,7,7)	(7,7,7)	(7,7,7)	(7,7,7)
impr. (%)	13.75	13.16	12.66	14.01	17.04	8.41

TABLE 11

Simulation results for (M4) for constant / QS kernel, $\alpha = 9.4$ and different weights ρ .

4.2. Data-adaptive variance estimation using subsampling

In the previous section we have performed an extensive simulation study about the behaviour of the RMSE of the variance estimators (4) and (18). We have seen that dependent on the dependence structure of the underlying random fields it is quite useful to consider the improved variance estimator in (18) instead of the one in (4), as the former is very robust with respect to the proper choice of \mathbf{m} and can also lead to improvements of the optimized RMSE up to more than 15% in some situations.

In the following simulation study we want to check how well the data-adaptive subsampling procedure proposed in Section 3.2 works. For this purpose, we consider model (M1) and choose $\mathbf{b} = \lfloor \mathbf{n}^\gamma \rfloor = (\lfloor n_1^\gamma \rfloor, \lfloor n_2^\gamma \rfloor)$ with $\gamma \in \{0.7, 0.8, 0.9\}$ as well as $\mathbf{h} = \mathbf{1}$, so that we consider all possible subrandom fields of size \mathbf{b} . Moreover, we do not only investigate the subsampling approximation of the RMSE, but also the one for the mean which is defined as

$$\widehat{\text{Mean}}(\widehat{\sigma}_{\mathbf{n}}^2) := |\mathbf{N}|^{-1} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_q=1}^{N_q} \widehat{\sigma}_{\mathbf{n}, \mathbf{b}, i}^2. \quad (27)$$

The general simulation setting of model (M1) stays the same as in Subsection 4.1.1; in particular we consider a random field of size (30,40) leading to subrandom fields of size (10,13), (15,19) and (21,27) for $\gamma = 0.7, 0.8$ and 0.9 , respectively. Table 12 shows the simulation results for different values of \mathbf{m} and 10000 repetitions. We can see that the subsampling approximation of the mean works very well for all values of γ and \mathbf{m} . Regarding the subsampling approximation of the RMSE, we see that this is most adequate if we choose $\gamma = 0.9$. Therefore, we fix this value from now on.

As our main goal is, however, to find a good estimate for the parameter α used in the cutting rules of the improved variance estimator (18), we now want to investigate how well we can imitate the graph of the RMSE in Figure 2 by the subsampling approximation. We compute the (optimal) subsampling approximation $\widehat{\text{RMSE}}(\widehat{\sigma}_{\mathbf{n},c}^2)$ of the RMSE of $\widehat{\sigma}_{\mathbf{n},c}^2$ with respect to \mathbf{m} (i.e. when \mathbf{m} ranges over the values of \mathcal{M}_2 in (24)) analogously to the one of $\widehat{\sigma}_{\mathbf{n}}^2$ in (22), whereby we now have to replace $c_{\mathbf{n}}(\mathbf{j})$ by $c_{\mathbf{b}}(\mathbf{j})$.

\mathbf{m}	(0,0)	(1,1)	(3,3)	(4,4)	(6,6)	(7,7)
Mean($\widehat{\sigma}_{\mathbf{n}}^2$)	1.7217	8.6957	11.5879	11.5588	11.5233	11.5297
RMSE($\widehat{\sigma}_{\mathbf{n}}^2$)	9.8391	2.9996	2.8328	3.8371	5.9867	7.1408
Mean _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.7$	1.7232	8.7022	11.5825	11.5530	11.5377	11.5527
RMSE _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.7$	7.2646	2.6152	8.7550	12.1514	20.6803	26.2114
Mean _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.8$	1.7231	8.7012	11.5781	11.5554	11.5585	11.5674
RMSE _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.8$	9.0152	2.7599	4.9774	7.0740	11.8211	14.4589
Mean _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.9$	1.7232	8.7021	11.5810	11.5564	11.5538	11.5565
RMSE _{Sub} ($\widehat{\sigma}_{\mathbf{n}}^2$), $\gamma = 0.9$	9.9136	3.5601	2.9771	3.8902	6.7117	8.3712

TABLE 12
Subsampling approximation of the mean and the RMSE in model (M1)

Figure 3 shows the optimal and subsampled RMSE in model (M1) for random fields of size (30,40) and (45,60) using constant weights.

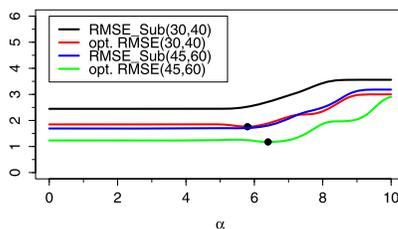


FIG 3. Optimal and subsampled RMSE in model (M1) for random fields of size (30,40) and (45,60) using constant weights.

We can see that the shape of the curves of the subsampling approximation of the RMSE coincides quite well with the shape of the optimal curves of the RMSE. Moreover, the difference between the approximated and the optimal curves becomes smaller for larger sample sizes. The only problem is that the

curves of the approximation attain their minimum for $\alpha = 0$; however, this is not very surprising as the minimum of the curves of the optimal RMSE is not very marked. To deal with such situations, we choose α as the largest value for which the corresponding subsampled RMSE differs less than 1%, say, from the subsampled RMSE for $\alpha = 0$. In the considered situation this leads to $\alpha = 5.4$ compared to the optimal value of $\alpha = 5.8$ for a random field of size (30,40) and to $\alpha = 5.7$ compared to the optimal value of $\alpha = 6.4$ for a random field of size (45,60). For a tolerance of 3% one would obtain $\alpha = 5.7$ for the smaller and $\alpha = 6.1$ for the larger random field. This shows that the proposed method leads to reasonable estimates for α if one chooses an appropriate tolerance level. Even if the optimal and the estimated values for α do not match exactly, they are still close enough to each other to guarantee good results and to profit from the advantages that one achieves when using $\hat{\sigma}_{\mathbf{n},c}^2$ instead of $\hat{\sigma}_{\mathbf{n}}^2$.

4.3. Accuracy for testing image data

Lastly, we analyzed the accuracy of the proposed estimator when used to standardize the image test statistic based on the partial sums $S_{\mathbf{n}}$, in order to test for the presence of deviations from a null or reference model for the image, as introduced and discussed in the introduction. Concretely, we investigated the test which rejects the null hypothesis H_0 , if

$$|\mathbf{n}|^{-1/2}|S_{\mathbf{n}}| > \hat{\sigma}_{\mathbf{n},c}\Phi^{-1}(1 - \alpha/2). \quad (28)$$

Observe that we use the threshold cut-off estimator with constant weights to standardize the partial sum. We used the threshold rule $c_{\mathbf{n}}(\alpha)$ investigated above in detail and investigated how one should select α to obtain accurate tests in terms of the type I error rate.

We simulated image data from spatial moving average models of order d given by

$$\xi_{i_1,i_2} = \sum_{j_1=-d}^d \sum_{j_2=-d}^d \theta_{j_1,j_2} \epsilon_{i_1+j_1,i_2+j_2} + \epsilon_{i_1,i_2}, \quad 1 \leq i_1 \leq n_1, 1 \leq i_2 \leq n_2,$$

where ϵ_{ij} are i.i.d. $N(0,1)$. The first model is model (M2), i.e. a spatial moving average model of order $d = 3$ with weights $a_1 = 0.5$, $a_2 = 0.3$ and $a_3 = 0.1$, cf. Section 4.1.2. The second model allows for much stronger dependencies and is of order $d = 40$ with coefficients given by

$$(M5) \quad \theta_{j_1,j_2} = \rho \sqrt{j_1^2 + j_2^2},$$

for $|\rho| < 1$. The degree was chosen as 40, in order to approximate a spatial linear process with $d = \infty$ and those coefficients, and nevertheless keeping the computational costs at a reasonable level. Depending on the parameter ρ the correlations can be substantial.

Table 13 provides the results for random fields following those models with sizes ranging between $n_1 = n_2 = 50$ and $n_1 = n_2 = 250$. For the second model the parameter ρ was chosen as 0.1, 0.3 and 0.5, in order to evaluate the procedure for weak, intermediate and stronger dependence structures. Indeed, visual inspection of random fields with $\rho = 0.5$ shows that the covariances induce an interesting texture-like random pattern which comes close to those observed in certain inhomogenous materials. For each case the type I error rate was simulated based on 5,000 simulation runs.

Compared to the results addressing the estimation accuracy, where values of α around 5.8 led to optimal RMSE values, it can be seen that smaller values around 3.6 are preferable, in order to obtain accurate type I error rates for the settings studied in this simulation, whatever the degree of correlation of the random field. This is, however, in good agreement with the findings of the previous simulations, especially the RMSE and bias curves shown in Figure 2. Indeed, according to those results for values around 3.6 the variance is slightly larger (leading to the slightly higher RMSE values), but the bias is smaller. Our simulations reveal the interesting observation that for the accuracy in terms of the type I error rate it seems to be better to have a smaller bias.

4.4. Concluding remarks

In all the settings of Subsection 4.1 we presented results for positive values of the MA- and AR-parameters. However, in our simulation studies we also considered negative values and mixtures of positive and negative values for these parameters. As the results are qualitatively the same as for only positive weights we do not present them here.

Moreover, the question arises whether cutting rules different from those in (20) and (26) respectively, would lead to similar or even better results regarding the RMSE of the improved variance estimator. One could for example also think of a cutting rule of the form

$$c_{\mathbf{n}}(\mathbf{j}) = \frac{(\max\{|j_1|, |j_2|, |j_3|\})^\alpha}{n_1 n_2 n_3} - \delta \quad (29)$$

or even of using a constant cutting rule, i.e. a rule that does not depend on the lag \mathbf{j} and the sample size \mathbf{n} . However, the simulation studies showed that a constant cutting rule in general cannot improve the RMSE as much as a cutting rule depending on \mathbf{j} and \mathbf{n} . Moreover, rules as in (29) or of similar forms lead to comparable but not better results as the cutting rules in (20) and (26) concerning the optimized RMSE. That is why we only focused our attention for the analysis of $\hat{\sigma}_{\mathbf{n},c}^2$ on sequences $c_{\mathbf{n}}(\mathbf{j})$ as in (20) and (26).

Lastly, we may conclude that the proposed estimators work well in various models of practical importance and lead to highly accurate results, both in terms of estimation accuracy and in terms of the type I error rates when considering significance testing for image data, when the parameters are selected appropriately. The latter can be achieved in a data-adaptive way by the proposed subsampling-based selection procedure.

n	ρ	α									
		3	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9
Model (M2)											
50	—	0.0622	0.06	0.0592	0.0518	0.0558	0.0558	0.0478	0.0506	0.0518	0.0474
100	—	0.0618	0.0598	0.0564	0.0504	0.0574	0.0538	0.0518	0.0456	0.0502	0.0458
250	—	0.0594	0.0524	0.0496	0.0542	0.0466	0.053	0.0574	0.051	0.05	0.0498
Model (M5)											
50	0.1	0.0574	0.0586	0.0588	0.061	0.055	0.0528	0.046	0.0556	0.0558	0.0558
	0.3	0.0622	0.0632	0.0566	0.0548	0.062	0.0578	0.0624	0.0562	0.0602	0.0632
	0.5	0.0914	0.085	0.078	0.0784	0.0812	0.0782	0.0768	0.0768	0.0802	0.081
100	0.1	0.0498	0.0534	0.0502	0.0492	0.0504	0.0468	0.0466	0.0502	0.0468	0.048
	0.3	0.054	0.0502	0.0572	0.0496	0.0542	0.0452	0.0496	0.0532	0.0508	0.0582
	0.5	0.0624	0.06	0.0642	0.0628	0.062	0.0538	0.0574	0.061	0.0652	0.0654
250	0.1	0.0516	0.0496	0.0524	0.0412	0.0502	0.0522	0.0554	0.056	0.0526	0.0516
	0.3	0.046	0.054	0.0462	0.0496	0.046	0.05	0.0572	0.0472	0.049	0.048
	0.5	0.0562	0.0538	0.0508	0.052	0.0534	0.0546	0.053	0.0566	0.0536	0.0524

TABLE 13. Simulated type I error rates of the test (28).

5. Proofs

This section is devoted to rigorous proofs of the presented theorems. The consistency proofs for the case $p = 1$, except Theorem 2.8, are part of the first author’s Ph.D. thesis, see [23]. The extensions to multivariate fields ($p > 1$) as well as Theorems 2.4, 2.5, 2.6 and 2.7 are newer and due to the second author.

5.1. Proofs of Section 2

Frequently we shall use the following result, see [4]: Fix $1 \leq j \leq q$ and put $\mathcal{F}_a^b = \sigma(\xi_i : a \leq i_j \leq b)$. Then ξ_0 is $\mathcal{F}_{-\infty}^0$ -measurable and ξ_ℓ is $\mathcal{F}_{\ell_*}^\infty$ -measurable, where $\ell_* = \min_j \ell_j$. It follows that

$$|\mathbb{E}(\xi_0 \xi_\ell) - \mathbb{E}(\xi_0)\mathbb{E}(\xi_\ell)| \leq 2\varphi_j^{1/2}(\ell_*)\sqrt{\mathbb{E}\xi_0^2}\sqrt{\mathbb{E}\xi_\ell^2} \leq 2\varphi^{1/2}(\ell_*)\sqrt{\mathbb{E}\xi_0^2}\sqrt{\mathbb{E}\xi_\ell^2},$$

by definition of φ . For the proof of Theorem 2.1 we need the following lemma.

Lemma 5.1. *Let $\{\xi_i, i \in \Gamma_n\}$ with $\xi_i = (\xi_i^{(1)}, \dots, \xi_i^{(p)})$, be a strictly stationary, φ -mixing (ρ^* -mixing) random field taking values in \mathbb{R}^p with $\max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^4 < \infty$. Fix $1 \leq \nu, \mu \leq p$. Then the random field $\{Y_i^{(\nu, \mu)}(\mathbf{j}), i \in \tilde{\Gamma}_n(\mathbf{j})\}$, defined by*

$$Y_i^{(\nu, \mu)}(\mathbf{j}) := \xi_i^{(\nu)}\xi_{i+\mathbf{j}}^{(\mu)} - \mathbb{E}(\xi_0^{(\nu)}\xi_{\mathbf{j}}^{(\mu)}), \tag{30}$$

is as well a strictly stationary, φ -mixing (ρ^* -mixing) random field with $\mathbb{E}(Y_0^{(\nu, \mu)}(\mathbf{j})) = 0$ and $\mathbb{E}(Y_0^{(\nu, \mu)}(\mathbf{j}))^2 < \infty$. The mixing coefficients satisfy

$$\varphi_{Y(j; \nu, \mu)}(i; r) \leq \varphi_\xi(i; r - 2j^*) \quad \text{for } r > 2j^* \tag{31}$$

Fix $\mathbf{j} \in \mathbb{Z}^q$ and $\nu, \mu \in \{1, \dots, p\}$. If (9) holds, i.e. $\sum_{r=1}^\infty r^{q-1}\varphi^{\frac{1}{2}}(r) < \infty$, then we have

$$\sum_{r=1}^\infty r^{q-1}\varphi_{Y(j; \nu, \mu)}^{\frac{1}{2}}(r) < \infty, \tag{32}$$

where $\varphi_{Y(j; \nu, \mu)}$ denote the φ -mixing coefficients of the random field $\{Y_i^{(\nu, \mu)}(\mathbf{j}), i \in \tilde{\Gamma}_n(\mathbf{j})\}$.

Proof. The stationarity of the random field $\{\xi_i, i \in \Gamma_n\}$ directly implies the stationarity of $\{Y_i^{(\nu, \mu)}(\mathbf{j}), i \in \tilde{\Gamma}_n(\mathbf{j})\}$. Moreover, we have

$$\mathbb{E}(Y_0^{(\nu, \mu)}(\mathbf{j}))^2 \leq \mathbb{E}(\xi_0^{(\nu)}\xi_{\mathbf{j}}^{(\mu)})^2 \leq \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^4,$$

by the Cauchy-Schwarz inequality and the stationarity of $\{\xi_i\}$. It remains to show that $\{Y_i^{(\nu,\mu)}(\mathbf{j})\}$ is φ -mixing and that the sum in (32) is finite. For this purpose set $j^* = \max_{1 \leq i \leq q} |j_i|$. On the one hand we have

$$\begin{aligned} \mathcal{A}_{Y(\mathbf{j})}^-(i; 0) &= \sigma \left(\left\{ Y_{n_1, \dots, n_q}^{(\nu,\mu)}(\mathbf{j}) : n_i \leq 0, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &= \sigma \left(\left\{ \xi_{n_1, \dots, n_q}^{(\nu)} \xi_{n_1+j_1, \dots, n_q+j_q}^{(\mu)} : n_i \leq 0, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &\subseteq \sigma \left(\left\{ \xi_{n_1, \dots, n_q} \xi_{n_1+j_1, \dots, n_q+j_q} : n_i \leq 0, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &\subseteq \sigma \left(\left\{ \xi_{n_1, \dots, n_q} : n_i \leq j^*, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &= \mathcal{A}_{\xi}^-(i; j^*). \end{aligned}$$

On the other hand we have for all $r \geq 0$ that

$$\begin{aligned} \mathcal{A}_{Y(\mathbf{j})}^+(i; r) &= \sigma \left(\left\{ Y_{n_1, \dots, n_q}^{(\nu,\mu)}(\mathbf{j}) : n_i \geq r, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &= \sigma \left(\left\{ \xi_{n_1, \dots, n_q}^{(\nu)} \xi_{n_1+j_1, \dots, n_q+j_q}^{(\mu)} : n_i \geq r, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &\subseteq \sigma \left(\left\{ \xi_{n_1, \dots, n_q} \xi_{n_1+j_1, \dots, n_q+j_q} : n_i \geq r, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &\subseteq \sigma \left(\left\{ \xi_{n_1, \dots, n_q} : n_i \geq r - j^*, n_l \text{ unrestricted for } l \neq i \right\} \right) \\ &= \mathcal{A}_{\xi}^+(i; r - j^*). \end{aligned}$$

Observe that the above inclusions also show that $\mathcal{A}_{Y(\mathbf{j})}^{\pm}(i; r) \subseteq \mathcal{A}_{Y(\mathbf{j})}^{\pm}(i; r)$. Thus, denoting the φ -mixing coefficients of $\{\xi_i, i \in \Gamma_{\mathbf{n}}\}$ by φ_{ξ} and those of $\{Y_i(\mathbf{j}), i \in \tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})\}$, $\mathbf{j} \in \mathbb{Z}^q$, by $\varphi_{Y(\mathbf{j})}$ we get

$$\varphi_{Y(\mathbf{j}; \nu, \mu)}(i; r) \leq \varphi_{Y(\mathbf{j})}(i; r)$$

as well as

$$\varphi_{Y(\mathbf{j}; \nu, \mu)}(i; r) \leq \varphi_{\xi}(i; r - 2j^*) \quad \text{for } r > 2j^*$$

leading to

$$\begin{aligned} \varphi_{Y(\mathbf{j}; \nu, \mu)}(r) &= \max_{1 \leq i \leq q} \varphi_{Y(\mathbf{j}; \nu, \mu)}(i; r) \\ &\leq \max_{1 \leq i \leq q} \varphi_{\xi}(i; r - 2j^*) = \varphi_{\xi}(r - 2j^*) \quad \text{for } r > 2j^*. \end{aligned} \tag{33}$$

Hence $\{Y_i^{(\nu,\mu)}(\mathbf{j})\}$ is φ -mixing. Lastly, to show that the sum in (32) is finite observe that

$$\sum_{r=1}^{\infty} r^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) = \sum_{r=1}^{2j^*} r^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) + \sum_{r=2j^*+1}^{\infty} r^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r).$$

The second sum is finite by inequality (33). We obtain

$$\sum_{r=2j^*+1}^{\infty} r^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) \leq \sum_{r=2j^*+1}^{\infty} r^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r - 2j^*) = \sum_{r=1}^{\infty} (r + 2j^*)^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r),$$

where the last sum is finite by assumption (9). □

Lemma 5.2. *Let $p = 1$. If $\{\xi_i, i \in \Gamma_n\}$ is ρ^* -mixing, then $\{Y_i(\mathbf{j}), \mathbf{i} \in \tilde{\Gamma}_n(\mathbf{j})\}$ is ρ^* -mixing and posses a continuous spectral density function.*

Proof. Observe that

$$\sup_{A_1, A_2: d(A_1, A_2) \geq r} \sup_{f, g \in L_2(P)} |\text{Corr}(f(\{\xi_i \xi_{i+j} : \mathbf{i} \in A_1\}), g(\{\xi_i \xi_{i+j} : \mathbf{i} \in A_2\}))|$$

is less or equal to

$$\sup_{A_1, A_2: d(A_1, A_2) \geq r} \sup_{f, g \in L_2(P)} \text{Corr}(f(\{\xi_i : \mathbf{i} \in A_1^{j^*}\}), g(\{\xi_i : \mathbf{i} \in A_2^{j^*}\})),$$

where $A_i^{j^*} = \{\mathbf{k} \in \mathbb{Z}^q : d(\mathbf{k}, A_i) \leq j^*\}$ is the j^* - enlargement of A_i , $i = 1, 2$, which is bounded by

$$\begin{aligned} &\sup_{A_1, A_2: d(A_1, A_2) \geq r-2j^*} \sup_{f, g \in L_2(P)} \text{Corr}(f(\{\xi_i : \mathbf{i} \in A_1\}), g(\{\xi_i : \mathbf{i} \in A_2\})) \\ &= \rho^*(r - 2j^*). \end{aligned}$$

It follows now from Theorem 28.21 [6] that $Y_i(\mathbf{j})$ possesses a continuous spectral density function, since for any ρ^* -mixing random field the linear dependence measure $\kappa(r)$ defined by

$$\kappa(r) = \sup \frac{|E(\sum_{i \in Q} a_i \xi_i)(\sum_{i \in S} b_i \xi_i)|}{\sqrt{\sum_{i \in Q} |a_i|^2} \sqrt{\sum_{i \in S} |b_i|^2}}$$

where the supremum is taken over all pairs of nonempty, finite and disjoint sets Q, S such that $d(Q, S) \geq r$, see [6, p. 148], satisfies the necessary and sufficient condition $\kappa(r) = o(1)$, as $r \rightarrow \infty$. □

For $p = 1$ the following basic result for φ -mixing random fields can be found in [11] without a proof.

Lemma 5.3. *Suppose $p = 1$. Let condition (9) be satisfied. Set $S_n := \sum_{1 \leq j \leq n} \xi_j$ for all $n \geq 1$. Then the following three assertions hold true.*

- (a) $\sum_{j \in \mathbb{Z}^q} |\gamma(\mathbf{j})| < \infty$.
- (b) $|n|^{-1} \mathbb{E}(S_n^2) \rightarrow \sigma^2$ as $n \rightarrow \infty$.
- (c) $|n|^{-1} \mathbb{E}(S_n^2) \leq A(q, \varphi) \mathbb{E}(\xi_0^2)$ for all $n \geq 1$, where

$$A(q, \varphi) = 1 + 2q \sum_{r=1}^{\infty} (2r + 1)^{q-1} \varphi^{\frac{1}{2}}(r).$$

The following result generalizes Lemma 5.3 to a general random field taking values in \mathbb{R}^p , $p \in \mathbb{N}$.

Lemma 5.4. *Assume that condition (9) is satisfied. Let $S_n := \sum_{1 \leq j \leq n} \xi_j$, $n \geq 1$.*

- (a) $\sum_{j \in \mathbb{Z}^q} \|\gamma(j)\|_\infty < \infty$.
- (b) It holds

$$\text{Var}(|\mathbf{n}|^{-1/2} S_{\mathbf{n}}) = \mathbb{E}(\xi_0 \xi'_0) + \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \mathbb{E}(\xi_0 \xi'_\ell) + o(1),$$

as $\mathbf{n} \rightarrow \infty$, and

$$\text{Var}(|\mathbf{n}|^{-1/2} S_{\mathbf{n}}) = |\mathbf{n}|^{-1} \mathbb{E}(S_{\mathbf{n}} S'_{\mathbf{n}}) \rightarrow \sigma^2,$$

as $\mathbf{n} \rightarrow \infty$, in $(\mathbb{R}^{p \times p}, \|\cdot\|_\infty)$.

- (c) If $\mathbf{n} \geq \mathbf{1}$, then

$$\| |\mathbf{n}|^{-1} \mathbb{E}(S_{\mathbf{n}} S'_{\mathbf{n}}) \|_\infty \leq A(q, \varphi) \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2,$$

where $A(q, \varphi)$ is as in Lemma 5.3.

- (d) For all $\ell \in \mathbb{Z}^q$ we have

$$\|\mathbb{E}(\xi_0 \xi'_\ell)\|_\infty = \max_{1 \leq i, j \leq p} |\mathbb{E}(\xi_0^{(i)} \xi_\ell^{(j)})| \leq 2\varphi^{1/2}(l_\star) \max_{1 \leq j \leq p} \mathbb{E}|\xi_0^{(j)}|^2,$$

where $l_\star = \min_j l_j$.

Proof. By [4, Lemma 1]

$$\max_{1 \leq i, j \leq p} |\mathbb{E}(\xi_0^{(i)} \xi_\ell^{(j)})| \leq 2\varphi^{1/2}(l_\star) \sqrt{\mathbb{E}|\xi_0^{(i)}|^2} \sqrt{\mathbb{E}|\xi_0^{(j)}|^2} \leq 2\varphi^{1/2}(l_\star) \max_{1 \leq j \leq p} \mathbb{E}|\xi_0^{(j)}|^2, \tag{34}$$

for any $\ell \in \mathbb{Z}^q$. Hence, by summing over all maximum-norm q -dimensional unit spheres,

$$\begin{aligned} \sum_{\ell \in \mathbb{Z}^q} \|\gamma(\ell)\|_\infty &= \sum_{\ell} \max_{1 \leq i, j \leq p} |\mathbb{E}(\xi_0^{(i)} \xi_\ell^{(j)})| \\ &\leq 2 \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2 \sum_{\ell \in \mathbb{Z}^q} \varphi^{1/2}(l_\star) \\ &= 2 \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2 \sum_{r=0}^{\infty} \varphi^{1/2}(r) d(r), \end{aligned}$$

where $d(r) = \#\{\ell \in \mathbb{Z}^q : l_\star = r\} \leq 2q(2r + 1)^{q-1}$. Hence (a) and (d) are shown. To verify (b), we have to show that

$$\text{Var}(|\mathbf{n}|^{-1/2} S_{\mathbf{n}}) = \mathbb{E}(\xi_0 \xi'_0) + \sum_{l_1 = -n_1 + 1}^{n_1 - 1} \frac{n_1 - l_1}{n_1} \dots \sum_{\ell_q = -n_q + 1}^{n_q - 1} \frac{n_q - |\ell_q|}{n_q} I_\ell \mathbb{E}(\xi_0 \xi'_\ell),$$

where $I_\ell = 1(\exists 1 \leq j \leq q : \ell_j \neq 0)$, satisfies

$$\text{Var}(|\mathbf{n}|^{-1/2} S_{\mathbf{n}}) = \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \mathbb{E}(\xi_0 \xi'_\ell) + o(1),$$

as $\mathbf{n} \rightarrow \infty$. Clearly, this follows from

$$\sum_{\ell_1=-n_1+1}^{n_1-1} \frac{n_1 - |\ell_1|}{n_1} \dots \sum_{\ell_q=-n_q+1}^{n_q-1} \frac{n_q - |\ell_q|}{n_q} I_{\ell} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) = \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) + o(1),$$

as $\mathbf{n} \rightarrow \infty$. In what follows, $\ell = (\ell_1, \dots, \ell_q)' \in \mathbb{N}_0^q$, $\ell_{-k} = (\ell_j)_{j=1, \dots, q, j \neq k}$ and (ℓ_{-k}, i_k) is obtained from ℓ by replacing the k th entry, ℓ_k , by i_k . Note that

$$\sum_{\ell_q=-n_q+1}^{n_q-1} \frac{n_q - |\ell_q|}{n_q} I_{\ell} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) = \sum_{\ell_q=-n_q+1}^{n_q-1} I_{\ell} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) - \frac{1}{n_q} \sum_{\ell_q=-n_q+1}^{n_q-1} I_{\ell} |\ell_q| \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}).$$

We shall show that the second term is $o(1)$ and remains $o(1)$ when propagated through the outer summations. To show that

$$\tilde{r}_q(\ell_{-q}) := \frac{1}{n_q} \sum_{\ell_q=-n_q+1}^{n_q-1} |\ell_q| I_{\ell} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) = o(1),$$

it suffices to consider the case $I_{\ell} = 1$, i.e. we may neglect this indicator. Decompose $\tilde{r}_q(\ell_{-q}) = \tilde{r}_q^+(\ell_{-q}) + \tilde{r}_q^-(\ell_{-q})$, where $r_q^+(\ell_{-q}) := \frac{1}{n_q} \sum_{\ell_q=1}^{n_q-1} \ell_q \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell})$ and $r_q^-(\ell_{-q}) := \frac{1}{n_q} \sum_{\ell_q=1}^{n_q-1} \ell_q \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell_{-q}, -\ell_q})$. We treat $r_q(\ell_{-q}) := r_q^+(\ell_{-q})$, since $r_q^-(\ell_{-q})$ can be handled analogously. Indeed, we will show that for all $k \in \{1, \dots, q\}$

$$r_k(\ell_{-k}) := \frac{1}{n_k} \sum_{\ell_k=1}^{n_k-1} \ell_k \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell}) = \frac{1}{n_k} \sum_{\ell_k=1}^{n_k-1} \sum_{i_k=\ell_k}^{n_k-1} \mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell_{-k}, i_k}) = o(1).$$

We have $\|r_k(\ell_{-k})\|_{\infty} \leq \frac{1}{n_k} \sum_{\ell_k=1}^{n_k-1} \sum_{i_k \geq \ell_k} \|\mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell_{-k}, i_k})\|_{\infty}$. Observe that for any $j \in \{1, \dots, q\} \setminus \{k\}$

$$\left\| \sum_{\ell_j=0}^{n_j-1} r(\ell_{-k}) \right\|_{\infty} \leq \frac{1}{n_k} \sum_{\ell_k=1}^{n_k-1} \sum_{i_k \geq \ell_k} \sum_{\ell_j=0}^{\infty} \|\mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell_{-k}, i_k})\|_{\infty} = o(1),$$

as $\mathbf{n} \rightarrow \infty$, by a Cesaro sum argument, since, of course, $\sum_{i_k=1}^{\infty} \sum_{\ell_j=0}^{\infty} \|\mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell_{-k}, i_k})\|_{\infty} \leq \sum_{\ell \in \mathbb{N}^q} \|\mathbb{E}(\xi_{\mathbf{o}} \xi'_{\ell})\|_{\infty} < \infty$. Similarly, using again a Cesaro sum argument,

$$\left\| \sum_{\ell_j=0}^{n_j-1} \frac{n_j - \ell_j}{n_j} r(\ell_{-k}) \right\|_{\infty} = \left\| \sum_{\ell_j=0}^{n_j-1} r(\ell_{-k}) - \frac{1}{n_j} \sum_{\ell_j=0}^{n_j-1} \sum_{\ell_k=\ell_j}^{n_k-1} r(\ell_{-k}) \right\|_{\infty} = o(1),$$

as $\mathbf{n} \rightarrow \infty$. By combining the results for r_k^+ and r_k^- we now also obtain

$$\left\| \sum_{\ell_j=-n_j+1}^{n_j-1} \frac{n_j - |\ell_j|}{n_j} r(\ell_{-k}) \right\|_{\infty} = o(1),$$

as $\mathbf{n} \rightarrow \infty$. More generally, by iterating the above argument, we have, firstly, for any pairwise different $j_1, \dots, j_r \in \{1, \dots, q\} \setminus \{k\}$, in the $\|\cdot\|_\infty$ -norm,

$$\sum_{\ell_{j_1}=-n_{j_1}+1}^{n_{j_1}-1} \dots \sum_{\ell_{j_q}=-n_{j_q}+1}^{n_{j_q}-1} r(\ell_{-k}) = o(1)$$

and, secondly,

$$\sum_{\ell_{j_1}=-n_{j_1}+1}^{n_{j_1}-1} \frac{n_{j_1}-|\ell_{j_1}|}{n_{j_1}} \dots \sum_{\ell_{j_q}=-n_{j_q}+1}^{n_{j_q}-1} \frac{n_{j_q}-|\ell_{j_q}|}{n_{j_q}} r(\ell_{-k}) = o(1), \quad (35)$$

as $\mathbf{n} \rightarrow \infty$. Now we may conclude that (35) yields

$$\begin{aligned} & \sum_{\ell_1=-n_1+1}^{n_1-1} \frac{n_1-|\ell_1|}{n_1} \dots \sum_{\ell_q=-n_q+1}^{n_q-1} \frac{n_q-|\ell_q|}{n_q} I_\ell \mathbb{E}(\xi_0 \xi'_\ell) \\ &= \sum_{\ell_1=-n_1+1}^{n_1-1} \frac{n_1-|\ell_1|}{n_1} \dots \sum_{\ell_{q-1}=-n_{q-1}+1}^{n_{q-1}-1} \frac{n_{q-1}-|\ell_{q-1}|}{n_{q-1}} \sum_{\ell_q=-n_q+1}^{n_q-1} I_\ell \mathbb{E}(\xi_0 \xi'_\ell) + o(1) \\ &= \dots = \sum_{\ell_1=-n_1+1}^{n_1-1} \dots \sum_{\ell_q=-n_q+1}^{n_q-1} I_\ell \mathbb{E}(\xi_0 \xi'_\ell) + o(1) \\ &= \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \mathbb{E}(\xi_0 \xi'_\ell) + o(1), \end{aligned}$$

as $\mathbf{n} \rightarrow \infty$. This shows (b). Assertion (c) can be shown by summing over all q -dimensional unit squares w.r.t to the maximum norm,

$$\begin{aligned} \|\text{Var}(|\mathbf{n}|^{-1/2} S_\mathbf{n})\|_\infty &= \left\| \frac{1}{|\mathbf{n}|} \mathbb{E}(S_\mathbf{n} S'_\mathbf{n}) \right\|_\infty \leq \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \left| \prod_{j=1}^q \frac{n_j-|\ell_j|}{n_j} \right| \|E(\xi_0 \xi'_\ell)\|_\infty \\ &\leq \sum_{-\mathbf{n} \leq^* \ell \leq^* \mathbf{n}} \|\mathbb{E}(\xi_0 \xi'_\ell)\|_\infty = \sum_{r=1}^\infty c(r) \|\mathbb{E}(\xi_0 \xi'_{r\mathbf{1}})\|_\infty, \end{aligned}$$

where $c(r) = \#\{-\mathbf{n} \leq^* \mathbf{i} \leq^* \mathbf{n} : \max_j |i_j| = r\} \leq 2q(2r+1)^{q-1}$. Using again (34), we obtain

$$\|\mathbb{E}(\xi_0 \xi'_\ell)\|_\infty \leq 2\varphi^{1/2}(r) \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2$$

for any $\ell \in \mathbf{1} : \mathbf{n}$ with $l_* = r$. Hence

$$\left\| \text{Var}(|\mathbf{n}|^{-1/2} S_\mathbf{n}) \right\|_\infty \leq \|\mathbb{E}(\xi_0 \xi'_\ell)\|_\infty + \sum_{r=1}^{n^*} 2q(2r+1)^{q-1} 4\varphi^{1/2}(r) \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2,$$

and we arrive at

$$\left\| \text{Var}(|\mathbf{n}|^{-1/2} S_\mathbf{n}) \right\|_\infty \leq A(\varphi, q) \max_{1 \leq j \leq p} \mathbb{E}(\xi_0^{(j)})^2. \quad \square$$

We are now in a position to prove the consistency of the estimator $\widehat{\sigma}_\mathbf{n}^2$ for multivariate random fields.

Proof of Theorem 2.1. Define the $p \times p$ matrix

$$\tilde{\sigma}_n^2 := \mathbb{E}(\hat{\sigma}_n^2) = \sum_{|\mathbf{j}| \leq m} w_m(\mathbf{j}) \gamma(\mathbf{j}).$$

We show that $\|\hat{\sigma}_n^2 - \tilde{\sigma}_n^2\|_\infty \xrightarrow{\mathbb{P}} 0$ and $\|\tilde{\sigma}_n^2 - \sigma^2\|_\infty \rightarrow 0$ as $n \rightarrow \infty$. Notice that

$$\begin{aligned} \|\tilde{\sigma}_n^2 - \sigma^2\|_\infty &= \left\| \sum_{|\mathbf{j}| \leq m} w_m(\mathbf{j}) \gamma(\mathbf{j}) - \sum_{\mathbf{j} \in \mathbb{Z}^q} \gamma(\mathbf{j}) \right\|_\infty \leq \left\| \sum_{|\mathbf{j}| \leq m} (w_m(\mathbf{j}) - 1) \gamma(\mathbf{j}) \right\|_\infty \\ &+ o(1) \leq \sum_{|\mathbf{j}| \leq m} |w_m(\mathbf{j}) - 1| \|\gamma(\mathbf{j})\|_\infty \mathbb{1}_{\{|\mathbf{j}| \leq m\}} + o(1), \end{aligned}$$

as $n \rightarrow \infty$, which implies $m \rightarrow \infty$, by Lemma 5.3(a). For the first summand this follows by dominated convergence: For $\mathbf{j} \in \mathbb{Z}^q$ let

$$f_m(\mathbf{j}) = |w_m(\mathbf{j}) - 1| \|\gamma(\mathbf{j})\|_\infty \mathbb{1}_{\{|\mathbf{j}| \leq m\}}.$$

Clearly, $f_m(\mathbf{j}) \rightarrow 0$ for each fixed $\mathbf{j} \in \mathbb{Z}^q$, as $m \rightarrow \infty$. Moreover,

$$|f_m(\mathbf{j})| \leq (C_w + 1) \|\gamma(\mathbf{j})\|_\infty =: g(\mathbf{j})$$

with $\sum_{\mathbf{j} \in \mathbb{Z}^q} |g(\mathbf{j})| < \infty$ by Lemma 5.3(a). Thus, the first summand also converges to zero and hence we obtain $\|\tilde{\sigma}_n^2 - \sigma^2\|_\infty \rightarrow 0$.

We now show that $\|\hat{\sigma}_n^2 - \tilde{\sigma}_n^2\|_\infty \xrightarrow{\mathbb{P}} 0$. By the Markov inequality we have

$$\mathbb{P}(\|\hat{\sigma}_n^2 - \tilde{\sigma}_n^2\|_\infty > \epsilon) \leq \frac{\mathbb{E}\|\hat{\sigma}_n^2 - \tilde{\sigma}_n^2\|_\infty}{\epsilon}.$$

We shall bound the right-hand side by an expression which is $o(1)$. Combined with $\|\tilde{\sigma}_n^2 - \sigma^2\|_\infty = o(1)$, as $n \rightarrow \infty$, this then shows the second assertion of Theorem 2.2. By boundedness of the weights $w_m(\mathbf{j})$ we obtain

$$\begin{aligned} \mathbb{E}\|\hat{\sigma}_n^2 - \tilde{\sigma}_n^2\|_\infty &\leq \sum_{|\mathbf{j}| \leq m} |w_m(\mathbf{j})| \mathbb{E}\|\hat{\gamma}_n(\mathbf{j}) - \gamma(\mathbf{j})\|_\infty \\ &\leq C_w \sum_{\nu, \mu=1}^p \sum_{|\mathbf{j}| \leq m} \mathbb{E} \left| \hat{\gamma}_n^{(\nu, \mu)}(\mathbf{j}) - \gamma^{(\nu, \mu)}(\mathbf{j}) \right| \\ &\leq C_w \sum_{\nu, \mu=1}^p \sum_{|\mathbf{j}| \leq m} \left(\mathbb{E} \left| \hat{\gamma}_n^{(\nu, \mu)}(\mathbf{j}) - \gamma^{(\nu, \mu)}(\mathbf{j}) \right|^2 \right)^{1/2}. \end{aligned} \tag{36}$$

Observe that

$$\begin{aligned} \hat{\gamma}_n^{(\nu, \mu)}(\mathbf{j}) - \gamma^{(\nu, \mu)}(\mathbf{j}) &= \frac{1}{|\tilde{\Gamma}_n(\mathbf{j})|} \sum_{i \in \tilde{\Gamma}_n(\mathbf{j})} \xi_i^{(\nu)} \xi_{i+\mathbf{j}}^{(\mu)} - \mathbb{E} \left(\xi_0^{(\nu)} \xi_{\mathbf{j}}^{(\mu)} \right) \\ &= \frac{1}{|\tilde{\Gamma}_n(\mathbf{j})|} \sum_{i \in \tilde{\Gamma}_n(\mathbf{j})} Y_i^{(\nu, \mu)}(\mathbf{j}) \end{aligned}$$

with $Y_i^{(\nu,\mu)}(\mathbf{j})$ defined in (30). Thus we obtain

$$\mathbb{E} \left| \widehat{\gamma}_n^{(\nu,\mu)}(\mathbf{j}) - \gamma^{(\nu,\mu)}(\mathbf{j}) \right|^2 = \frac{1}{|\widetilde{\Gamma}_n(\mathbf{j})|^2} \mathbb{E} \left(\sum_{i \in \widetilde{\Gamma}_n(\mathbf{j})} Y_i^{(\nu,\mu)}(\mathbf{j}) \right)^2.$$

For each $\mathbf{j} \in \mathbb{Z}^q$ we can now apply Lemma 5.3(c) to the expectation of the squared sum of the random field $\{Y_i^{(\nu,\mu)}(\mathbf{j})\}$ as condition (32) is fulfilled by Lemma 5.1. This leads to

$$\begin{aligned} & \mathbb{E} \left| \widehat{\gamma}_n^{(\nu,\mu)}(\mathbf{j}) - \gamma^{(\nu,\mu)}(\mathbf{j}) \right|^2 \\ & \leq \frac{1}{|\widetilde{\Gamma}_n(\mathbf{j})|} A(q, \varphi_{Y(\mathbf{j})}) \mathbb{E} \left(Y_{\mathbf{0}}^{(\nu,\mu)}(\mathbf{j}) \right)^2 \\ & \leq \frac{\left(1 + 2q \sum_{r=1}^{\infty} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) \right) \max_{1 \leq j \leq p} \mathbb{E}(\xi_{\mathbf{0}}^{(j)})^4}{|\widetilde{\Gamma}_n(\mathbf{j})|}, \end{aligned} \tag{37}$$

where the last inequality is fulfilled by Lemma 5.1 and the definition of $A(q, \varphi)$ in Lemma 5.3(c). Combining (36) with (37) we obtain

$$\mathbb{E} \|\widehat{\sigma}_n^2 - \widetilde{\sigma}_n^2\|_{\infty} \leq \max_{1 \leq j \leq p} \mathbb{E}^{1/2}(\xi_{\mathbf{0}}^{(j)})^4 \sum_{|\mathbf{j}| \leq m} \left(\frac{1 + 2q \sum_{r=1}^{\infty} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r)}{|\widetilde{\Gamma}_n(\mathbf{j})|} \right)^{1/2}$$

and therefore

$$\begin{aligned} & \mathbb{P} \left(\|\widehat{\sigma}_n^2 - \widetilde{\sigma}_n^2\|_{\infty} > \varepsilon \right) \\ & \leq \frac{p^2 C_w}{\varepsilon} \max_{1 \leq j \leq p} \mathbb{E}^{1/2}(\xi_{\mathbf{0}}^{(j)})^4 \sum_{|\mathbf{j}| \leq m} \left(\frac{1 + 2q \sum_{r=1}^{\infty} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r)}{|\widetilde{\Gamma}_n(\mathbf{j})|} \right)^{1/2}. \end{aligned} \tag{38}$$

To estimate the right-hand side consider the decomposition

$$\begin{aligned} \sum_{r=1}^{\infty} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) &= \sum_{r=1}^{2j^*} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) \\ &+ \sum_{r=2j^*+1}^{\infty} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) =: S_1 + S_2. \end{aligned}$$

Since the φ -mixing coefficients are decreasing with $\varphi(0) = 1$, we obtain

$$S_1 \leq \sum_{r=1}^{2j^*} (2r+1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(0) = \sum_{r=1}^{2j^*} (2r+1)^{q-1} \leq 2j^*(4j^*+1)^{q-1}.$$

As the sum in (38) only ranges over $\mathbf{j} \in \mathbb{Z}^q$ with $|\mathbf{j}| \leq \mathbf{m}$ we know that $j^* \leq m^*$. This implies

$$S_1 \leq 2m^*(4m^* + 1)^{q-1} \leq c_1(m^*)^q$$

for some finite constant $c_1 \in \mathbb{R}$ depending only on the dimension q of the random field. For the estimation of S_2 we use again (33) and get

$$\begin{aligned} S_2 &\leq \sum_{r=2j^*+1}^{\infty} (2r + 1)^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r - 2j^*) \\ &= \sum_{r=1}^{\infty} (2(r + 2j^*) + 1)^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r) \leq (4j^*)^{q-1} \sum_{r=1}^{\infty} (2r + 2)^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r). \end{aligned}$$

Since $\sum_{r=1}^{\infty} (2r + 2)^{q-1} \varphi_{\xi}^{\frac{1}{2}}(r)$ is finite by assumption, we can find a constant $c_2 \in \mathbb{R}$ such that for all $|\mathbf{j}| \leq \mathbf{m}$ we have

$$S_2 \leq c_2(m^*)^{q-1}.$$

Putting things together we obtain

$$\sum_{r=1}^{\infty} (2r + 1)^{q-1} \varphi_{Y(\mathbf{j})}^{\frac{1}{2}}(r) \leq c_3(m^*)^q \tag{39}$$

for $c_3 \in \mathbb{R}$.

Finally, observe that

$$|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})| \geq \prod_{i=1}^q (n_i - j^*) \geq \prod_{i=1}^q (n_{\star} - m^*) = (n_{\star} - m^*)^q \tag{40}$$

for all $\mathbf{j} \in \mathbb{Z}^q$ with $|\mathbf{j}| \leq \mathbf{m}$. Combining (39) and (40) with (38) we obtain

$$\begin{aligned} \mathbb{P}(\|\hat{\sigma}_{\mathbf{n}}^2 - \tilde{\sigma}_{\mathbf{n}}^2\|_{\infty} > \varepsilon) &\leq \frac{p^2 C_w}{\varepsilon} \max_{1 \leq j \leq p} \mathbb{E}^{1/2}(\xi_{\mathbf{0}}^{(j)})^4 \sum_{|\mathbf{j}| \leq \mathbf{m}} \left(\frac{1 + 2qc_3(m^*)^q}{(n_{\star} - m^*)^q} \right)^{1/2} \\ &\leq \frac{C_w}{\varepsilon} \max_{1 \leq j \leq p} \mathbb{E}^{1/2}(\xi_{\mathbf{0}}^{(j)})^4 \left(\prod_{i=1}^q (2m_i + 1) \right) \\ &\quad \times \left(\frac{1 + 2qc_3(m^*)^q}{(n_{\star} - m^*)^q} \right)^{1/2}. \end{aligned}$$

As $\prod_{i=1}^q (2m_i + 1) \leq (2m^* + 1)^q \leq c_5(m^*)^q$ for some $c_5 \in \mathbb{R}$, we can find a constant $C \in \mathbb{R}$ such that

$$\mathbb{P}(\|\hat{\sigma}_{\mathbf{n}}^2 - \tilde{\sigma}_{\mathbf{n}}^2\|_{\infty} > \varepsilon) \leq C \left(\frac{(m^*)^3}{n_{\star} - m^*} \right)^{q/2} = C \left(\frac{n_{\star}}{(m^*)^3} - \frac{1}{(m^*)^2} \right)^{-q/2}.$$

The last expression tends to zero for $\mathbf{n} \rightarrow \infty$ since by assumption $(m^*)^3/n_\star = o(1)$. We obtain $\|\hat{\sigma}_\mathbf{n}^2 - \tilde{\sigma}_\mathbf{n}^2\|_\infty \xrightarrow{\mathbb{P}} 0$ which completes the proof. \square

Proof of Theorem 2.2. We have $\mathbb{E} \left[(\hat{\sigma}_\mathbf{n}^2 - \sigma^2)^2 \right] = \text{Var}(\hat{\sigma}_\mathbf{n}^2) + \text{Bias}^2(\hat{\sigma}_\mathbf{n}^2)$, where $\text{Bias}^2(\hat{\sigma}_\mathbf{n}^2) = (\tilde{\sigma}_\mathbf{n}^2 - \sigma^2)^2 \rightarrow 0$ as $\mathbf{n} \rightarrow \infty$ by the proof of Theorem 2.1. For the variance we obtain by the Cauchy-Schwarz inequality and the uniform boundedness of the weights $w_m(\mathbf{j})$ that

$$\begin{aligned} \text{Var}(\hat{\sigma}_\mathbf{n}^2) &\leq \mathbb{E} \left[\sum_{|\mathbf{j}| \leq m} w_m^2(\mathbf{j}) (\hat{\gamma}_\mathbf{n}(\mathbf{j}) - \gamma(\mathbf{j}))^2 \sum_{|\mathbf{j}| \leq m} 1 \right] \\ &\leq c_1(m^*)^q \sum_{|\mathbf{j}| \leq m} \mathbb{E} \left[(\hat{\gamma}_\mathbf{n}(\mathbf{j}) - \gamma(\mathbf{j}))^2 \right] \end{aligned} \tag{41}$$

for some constant $c_1 \in \mathbb{R}$. By the proof of Theorem 2.1 we moreover have

$$\sum_{|\mathbf{j}| \leq m} \mathbb{E} \left[(\hat{\gamma}_\mathbf{n}(\mathbf{j}) - \gamma(\mathbf{j}))^2 \right] \leq c_2(m^*)^q \frac{(m^*)^q}{(n_\star - m^*)^q} \tag{42}$$

for some constant $c_2 \in \mathbb{R}$. By combining (41) with (42) we finally obtain $\text{Var}(\hat{\sigma}_\mathbf{n}^2) \leq C \left(\frac{(m^*)^3}{n_\star - m^*} \right)^q = o(1)$, by assumption, for some $C \in \mathbb{R}$. This completes the proof. \square

Proof of Corollary 2.1. This is a direct consequence of the proof of Theorem 2.1. \square

Let us now show the consistency of the estimator, when centering the terms at the temporal average, under the multiplicate model.

Proof of Theorem 2.8. $\tilde{\gamma}_\mathbf{n}(\mathbf{j})$ satisfies the decomposition

$$\tilde{\gamma}_\mathbf{n}(\mathbf{j}) = \tilde{\gamma}_\mathbf{n}(\mathbf{j}) + \tilde{R}_\mathbf{n}(\mathbf{j}), \quad \tilde{R}_\mathbf{n}(\mathbf{j}) = R_\mathbf{n}(\mathbf{j}) + R_\mathbf{n}(\mathbf{0}) + \bar{R}_\mathbf{n}, \tag{43}$$

where

$$\tilde{\gamma}_\mathbf{n}(\mathbf{j}) := \frac{1}{|\tilde{\Gamma}_\mathbf{n}(\mathbf{j})|} \sum_{i \in \tilde{\Gamma}_\mathbf{n}} \xi_i \xi'_{i+\mathbf{j}},$$

and

$$R_\mathbf{n}(\mathbf{j}) := -\frac{1}{|\tilde{\Gamma}_\mathbf{n}(\mathbf{j})|} \sum_{(i_1, \bar{i}) \in \tilde{\Gamma}_\mathbf{n}(\mathbf{j})} \xi_{i_1+\mathbf{j}} \bar{\xi}'_{i_1, \bar{i}}, \quad \bar{R}_\mathbf{n} := \frac{1}{|\tilde{\Gamma}_\mathbf{n}(\mathbf{j})|} \sum_{(i_1, \bar{i}) \in \tilde{\Gamma}_\mathbf{n}(\mathbf{j})} \bar{\xi}_{i_1, \bar{i}} \bar{\xi}'_{i_1, \bar{i}}.$$

(43) induces the decomposition

$$\check{\sigma}_\mathbf{n}^2 = \sum_{|\mathbf{j}| \leq m} w_m(\mathbf{j}) \tilde{\gamma}_\mathbf{n}(\mathbf{j}) + R_\mathbf{n}^\sigma, \quad R_\mathbf{n}^\sigma := R'_\mathbf{n} + R''_\mathbf{n},$$

with

$$R'_n := \sum_{|j| \leq m} w_m(j) R_n(j), \quad R''_n := \sum_{|j| \leq m} w_m(j) (R_n(\mathbf{0}) + \bar{R}_n).$$

Noting that Theorems 2.1 and 2.2 apply to the random field $\{\xi_i : i \in \mathbb{Z}^q\}$, the assertion follows if $R_n^\sigma = o_{\mathbb{P}}(1)$, as $n \rightarrow \infty$. We treat R'_n as R''_n can be dealt with in a similar way. Observe that with $\mathbf{j} = (j_1, \bar{\mathbf{j}})$, $\bar{\mathbf{j}} = (j_2, \dots, j_q)$, arranging terms leads to

$$\begin{aligned} R_n(j) &= - \frac{1}{|\bar{\Gamma}_n(\mathbf{j})|} \sum_{(i_1, \bar{i}) \in \bar{\Gamma}_n(\mathbf{j})} \varepsilon_{i_1+j_1}^{(T)} \varepsilon_{\bar{i}+\bar{j}}^{(S)} \frac{1}{n_1} \sum_{l=1}^{n_1} \varepsilon_l^{(T)} (\varepsilon_{\bar{i}}^{(S)})' \\ &= - \left(\frac{1}{n_1} \sum_{l=1}^{n_1} \varepsilon_l^{(T)} \right) \left(\frac{1}{n_1 - |j_1|} \sum_{i_1=1}^{n_1 - |j_1|} \varepsilon_{i_1+j_1}^{(T)} \right) \tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}}), \end{aligned}$$

where

$$\tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}}) := \frac{1}{\prod_{l=2}^q (n_l - |j_l|)} \sum_{\bar{i} \in \bar{\Gamma}_n(\bar{\mathbf{j}})} \varepsilon_{\bar{i}}^{(S)} (\varepsilon_{\bar{i}+\bar{j}}^{(S)})'.$$

Define $\tilde{\gamma}^{(S)}(\bar{\mathbf{j}}) := \mathbb{E}(\varepsilon_0^{(S)} (\varepsilon_{\bar{j}}^{(S)})')$, the average $\bar{S}_n := \frac{1}{n_1} \sum_{l=1}^{n_1} \varepsilon_l^{(T)}$, and $\bar{S}_n(j_1) := \frac{1}{n_1 - |j_1|} \sum_{l=1}^{n_1 - |j_1|} \varepsilon_l^{(T)}$. Let us denote by $R_n^{(\nu, \mu)}(\mathbf{j})$, $\tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}})_{\nu\mu}$ and $\tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu}$ the (ν, μ) th element of the corresponding (random) $p \times p$ matrices, $1 \leq \nu, \mu \leq p$. By Lemma 5.3(c) and independence, we obtain

$$\begin{aligned} &\max_{1 \leq \nu, \mu \leq p} \mathbb{E} \left| R_n^{(\nu, \mu)}(\mathbf{j}) \right| \\ &\leq \max_{1 \leq \nu, \mu \leq p} \mathbb{E} (|\bar{S}_n| |\bar{S}_n(j_1)|) \left(\mathbb{E} \left| \tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}})_{\nu\mu} - \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu} \right| + \left| \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu} \right| \right) \\ &\leq \max_{1 \leq \nu, \mu \leq p} \sqrt{\mathbb{E}(\bar{S}_n^2) \mathbb{E}(\bar{S}_n(j_1)^2)} \left(\sqrt{\mathbb{E}(\tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}})_{\nu\mu} - \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu})^2} + \left| \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu} \right| \right) \\ &\leq \max_{1 \leq \nu, \mu \leq p} \frac{K}{\sqrt{n_1(n_1 - |j_1|)}} \left(\sqrt{\mathbb{E}(\tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}})_{\nu\mu} - \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu})^2} + \left| \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu} \right| \right) \end{aligned}$$

with $K := A(1, \varphi_{\varepsilon^{(T)}}) \mathbb{E}((\varepsilon_0^{(T)})^2)$. Hence, for any $\varepsilon > 0$

$$\begin{aligned} \mathbb{E} \|R'_n\|_\infty &\leq \frac{KC_w}{n_1^{1/2}} \sum_{\nu, \mu=1}^p \sum_{|j| \leq m} \sqrt{\mathbb{E}(\tilde{\gamma}_n^{(S)}(\bar{\mathbf{j}})_{\nu\mu} - \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu})^2} \\ &\quad + \frac{KC_w}{n_1^{1/2}} \sum_{\nu, \mu=1}^p \sum_{|j| \leq m} \left| \tilde{\gamma}^{(S)}(\bar{\mathbf{j}})_{\nu\mu} \right|. \end{aligned}$$

Since $\sum_{\bar{j} \in \mathbb{Z}^{q-1}} |\tilde{\gamma}^{(S)}(\bar{j})_{\nu\mu}| < \infty$ and $(m^*)^2/n_1 = o(1)$, one may now argue as in the proof of Theorem 2.1, cf. (36), to show that $\mathbb{E}\|\tilde{\sigma}_{\mathbf{n}}^2 - \sigma^2\|_\infty = o(1)$, which completes the proof. \square

Next we establish the results on the asymptotic distribution of the estimators.

Proof of Lemma 2.3, Theorem 2.4 and Theorem 2.5. Recall that $Y_i(\mathbf{j}) = \xi_i \xi_{i+\mathbf{j}} - \gamma(\mathbf{j})$. We have

$$\begin{aligned} \sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}(\hat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma(\mathbf{j})) &= \frac{1}{\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}} \sum_{i \in \tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})} Y_i(\mathbf{j}) \\ &= \frac{1}{\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}} \sum_{i \in \mathbf{1:n}} Y_i(\mathbf{j}) - R_{\mathbf{n}}^{(1)}(\mathbf{j}) \\ &= \frac{1}{\sqrt{|\mathbf{n}|}} \sum_{i \in \mathbf{1:n}} Y_i(\mathbf{j}) - R_{\mathbf{n}}^{(1)}(\mathbf{j}) + R_{\mathbf{n}}^{(2)}(\mathbf{j}), \end{aligned}$$

where

$$R_{\mathbf{n}}^{(1)}(\mathbf{j}) = \frac{1}{\sqrt{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}} \sum_{i \in \mathbf{1:n} \setminus \tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})} Y_i(\mathbf{j})$$

and

$$R_{\mathbf{n}}^{(2)}(\mathbf{j}) = \left(\sqrt{\frac{|\mathbf{n}|}{|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|}} - 1 \right) \frac{1}{\sqrt{|\mathbf{n}|}} \sum_{i \in \mathbf{1:n}} Y_i(\mathbf{j}).$$

Because $Y_i(\mathbf{j})$ is a strictly stationary φ -mixing random field taking values in \mathbb{R}^p with asymptotic variance $\zeta_{\mathbf{j}}^2 > 0$, since (11) defines a positive definite quadratic form, and satisfying (9), the Cramér-Wold device and [11] show that

$$\frac{1}{\sqrt{|\mathbf{n}|}} \sum_{i \in \mathbf{1:n}} Y_i(\mathbf{j}) \Rightarrow B_{\mathbf{j}},$$

as $\mathbf{n} \rightarrow \infty$, for each $|\mathbf{j}| \leq \mathbf{m}$. This immediately implies $\max_{|\mathbf{j}| \leq \mathbf{m}} \|R_{\mathbf{n}}^{(2)}(\mathbf{j})\|_\infty = o_{\mathbb{P}}(1)$. Observe that $\mathbf{1} : \mathbf{n} \setminus \tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})$ is a finite union of index rectangles $I_{\mathbf{n}}^{(1,\ell)}$, $\ell = 1, \dots, L$, each of which has at least one index, denoted by k_ℓ , ranging over $n_{k_\ell} - j_{k_\ell} + 1, \dots, n_{k_\ell}$ for some $k_\ell \in \{1, \dots, q\}$. Therefore, $|I_{\mathbf{n}}^{(1,\ell)}| \leq j_{k_\ell} \prod_{\nu=1, \nu \neq k_\ell}^q (n_\nu - j_\nu) = o(|\mathbf{n}|)$, as $\mathbf{n} \rightarrow \infty$. It follows that for each $\nu, \mu \in \{1, \dots, p\}$,

$$|\tilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|^{-1} E |R_{\mathbf{n}}^{(1,\ell)}(\mathbf{j})_{\nu\mu}|^2 = O \left(j_{k_\ell} / n_{k_\ell} \sqrt{\mathbb{E}(Y_{\mathbf{0}}(\mathbf{j})_{\nu\mu})^2} \right),$$

where $R_n^{(1,\ell)}(\mathbf{j})_{\nu\mu} = \sum_{\mathbf{i} \in I_n^{(1,\ell)}} Y_{\mathbf{i}}(\mathbf{j})_{\nu\mu}$ for $\ell = 1, \dots, L$. Consequently, we obtain the estimate

$$\begin{aligned} \mathbb{P}\left(\max_{|\mathbf{j}| \leq \mathbf{m}} \|R_n^{(1)}(\mathbf{j})\|_\infty > \varepsilon\right) &\leq \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \sum_{\ell=1}^L \frac{\mathbb{E} |R_n^{(1,\ell)}(\mathbf{j})_{\nu\mu}|^2}{|\tilde{\Gamma}_n(\mathbf{j})| \varepsilon^2} \\ &= O\left(\sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \sum_{\ell=1}^L \frac{j_{k_\ell}}{n_{k_\ell}} \sqrt{\mathbb{E}(Y_{\mathbf{0}}(\mathbf{j})_{\nu\mu})^2}\right) \\ &= o(1), \end{aligned}$$

as $n \rightarrow \infty$, for all $\varepsilon > 0$, such that

$$\max_{|\mathbf{j}| \leq \mathbf{m}} \|R_n^{(1)}(\mathbf{j})\|_\infty = o_{\mathbb{P}}(1),$$

as $n \rightarrow \infty$. Hence, we may change the scaling factor to $\sqrt{|\mathbf{n}|}$ and shall now establish the weak convergence

$$\left(\sqrt{|\mathbf{n}|}(\hat{\gamma}_n(\mathbf{j}) - \gamma(\mathbf{j}))\right)_{|\mathbf{j}| \leq \mathbf{m}} \Rightarrow \mathcal{S}', \tag{44}$$

as $n \rightarrow \infty$, where \mathcal{S}' denotes the process \mathcal{S} for constant weights $w_{\mathbf{m}}(\mathbf{j}) = 1$, $|\mathbf{j}| \leq \mathbf{m}$. This follows, if we prove

$$\sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \lambda_j^{(\nu, \mu)} |\mathbf{n}|^{-1/2} \sum_{\mathbf{i} \in 1:\mathbf{n}} Y_{\mathbf{i}}(\mathbf{j})_{\nu\mu} \xrightarrow{d} \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \lambda_j^{(\nu, \mu)} B_j^{(\nu, \mu)} \tag{45}$$

for all arrays of coefficients $\lambda = \{\lambda_j^{(\nu, \mu)} : 1 \leq \nu, \mu \leq p, |\mathbf{j}| \leq \mathbf{m}\}$ of real numbers such that the variance of the limiting random variable is positive, by virtue of the Cramér-Wold device, see e.g. [10, Chapter 29.5]. Observe that the real-valued random field

$$Z_{\mathbf{i}}(\lambda) = \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \lambda_j^{(\nu, \mu)} Y_{\mathbf{i}}(\mathbf{j})_{\nu\mu}, \quad \mathbf{i} \in \mathbb{Z}^q,$$

is strictly stationary and φ -mixing with mixing coefficients $\varphi_\lambda(r)$ satisfying $\varphi_\lambda(r) \leq \varphi(r - m^*)$, which implies $\sum_{r=1}^\infty r^{q-1} \varphi_\lambda^{1/2}(r) < \infty$. Its asymptotic variance is positive, since the quadratic form induced by (12) is positive definite. Hence we may apply [11] and can conclude that the CLT holds true with asymptotic variance

$$\eta_\lambda^2 = \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(Z_{\mathbf{0}}(\lambda) Z_{\mathbf{i}}(\lambda)),$$

where

$$\mathbb{E}(Z_{\mathbf{0}}(\lambda) Z_{\mathbf{i}}(\lambda)) = \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{|\mathbf{k}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \sum_{\nu', \mu'=1}^p \lambda_j^{(\nu, \mu)} \lambda_{\mathbf{k}}^{(\nu', \mu')} \mathbb{E}(Y_{\mathbf{0}}(\mathbf{j})_{\nu\mu} Y_{\mathbf{i}}(\mathbf{k})_{\nu'\mu'}).$$

Hence,

$$\begin{aligned} & \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(Z_{\mathbf{0}}(\boldsymbol{\lambda})Z_{\mathbf{i}}(\boldsymbol{\lambda})) \\ &= \sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{|\mathbf{k}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \sum_{\nu', \mu'=1}^p \lambda_{\mathbf{j}}^{(\nu, \mu)} \lambda_{\mathbf{k}}^{(\nu', \mu')} \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(Y_{\mathbf{0}}(\mathbf{j})_{\nu\mu} Y_{\mathbf{i}}(\mathbf{k})_{\nu'\mu'}). \end{aligned}$$

Substituting

$$\mathbb{E}(B_{\mathbf{j}}^{(\nu\mu)} B_{\mathbf{k}}^{(\nu'\mu')}) = \sum_{\mathbf{i} \in \mathbb{Z}^q} \mathbb{E}(Y_{\mathbf{0}}(\mathbf{j})_{\nu\mu} Y_{\mathbf{i}}(\mathbf{k})_{\nu'\mu'}),$$

cf. (12), it follows that

$$\eta_{\boldsymbol{\lambda}}^2 = \text{Var} \left(\sum_{|\mathbf{j}| \leq \mathbf{m}} \sum_{\nu, \mu=1}^p \lambda_{\mathbf{j}}^{(\nu, \mu)} B_{\mathbf{j}}^{(\nu\mu)} \right)$$

such that (44) is shown. Now the first assertion, i.e. the case of general weights $w_{\mathbf{m}}(\mathbf{i})$, follows by an application of the continuous mapping theorem. The second assertion follows by noting that

$$\sqrt{|\mathbf{n}|}(\widehat{\sigma}_{\mathbf{n}}^2 - \mathbb{E}(\widehat{\sigma}_{\mathbf{n}}^2)) = \sum_{|\mathbf{j}| \leq \mathbf{m}} w_{\mathbf{m}}(\mathbf{j}) \sqrt{\frac{|\mathbf{n}|}{\widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j})}} \sqrt{|\widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j})|} (\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}) - \gamma_{\mathbf{n}}(\mathbf{j}))$$

and the fact that $|\mathbf{n}|/|\widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j})| \rightarrow 1$, as $\mathbf{n} \rightarrow \infty$. This completes the proof. \square

Lastly, we prove the results about subsampling.

Proof of Theorem 2.7. We show the result for $\widehat{\theta}_{\mathbf{n}} = \widehat{\theta}_{\mathbf{n}}(Y_{\mathbf{i}}(\mathbf{j}) : \mathbf{0} < \mathbf{i} \leq \mathbf{n})$, \mathbf{j} fixed, since then the corresponding results for $\widehat{\theta}_{\mathbf{n}} = \widehat{\theta}_{\mathbf{n}}(Y_{\mathbf{i}}(\mathbf{j}), |\mathbf{j}| \leq \mathbf{m} : \mathbf{0} < \mathbf{i} \leq \mathbf{n})$ and $\widehat{\theta}_{\mathbf{n}} = \widehat{\theta}_{\mathbf{n}}(\xi_{\mathbf{i}} : \mathbf{0} < \mathbf{i} \leq \mathbf{n})$ follow along the same lines. In [22] the authors use the strong mixing coefficient $\widehat{\alpha}_{Y(\mathbf{j})}(k; l_1)$ defined by

$$\sup_{\substack{\mathbf{E}_2 = \mathbf{E}_1 + \mathbf{t} \\ \mathbf{E}_1 \subseteq \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j}), \mathbf{t} \in \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j})}} \left\{ \left| \mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2) \right| : \begin{array}{l} A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2, \\ |\mathbf{E}_1| \leq l_1, d_{\infty}(\mathbf{E}_1, \mathbf{E}_2) \geq r \end{array} \right\}.$$

Here $\mathcal{E}_i = \sigma(Y_{\mathbf{i}}(\mathbf{j}) : \mathbf{i} \in \mathbf{E}_i)$, $i = 1, 2$, and the distance between two (index) sets $\mathbf{A}, \mathbf{B} \subset \mathbb{Z}^q$ is defined as

$$d_{\infty}(\mathbf{A}, \mathbf{B}) = \inf\{d_{\infty}(\mathbf{a}, \mathbf{b}) : \mathbf{a} \in \mathbf{A}, \mathbf{b} \in \mathbf{B}\}, \tag{46}$$

using the sup distance $d_{\infty}(\mathbf{a}, \mathbf{b}) = \max_{1 \leq j \leq q} |a_j - b_j|$ between two points $\mathbf{a} = (a_1, \dots, a_q)$ and $\mathbf{b} = (b_1, \dots, b_q)$.

If \mathbf{E}_1 and \mathbf{E}_2 have sup distance $\geq r$, then there exists some coordinate $k \in \{1, \dots, q\}$, such that one set is an element of $\mathcal{A}^-(k; l)$ and the other set is an element of $\mathcal{A}^+(k; l+r)$, where the σ -fields \mathcal{A}^{\pm} are defined in (6) and (5). By shifting both sets, we can assume that $l = 0$. Denote the shifted sets by \mathbf{E}'_1 and

\mathcal{E}'_2 and the associated σ -fields by \mathcal{E}'_1 and \mathcal{E}'_2 . If $A \in \mathcal{E}_1$ and $B \in \mathcal{E}_2$, then there is some Borel function h such that

$$1_A 1_B = h(Y_i(j) : i \in \mathbf{E}_1, i \in \mathbf{E}_2) \stackrel{d}{=} h(Y_i(j) : i \in \mathbf{E}'_1, i \in \mathbf{E}'_2).$$

Hence

$$\sup_{A \in \mathcal{E}_1, B \in \mathcal{E}_2} |\mathbb{P}(B|A) - \mathbb{P}(B)| = \sup_{A \in \mathcal{E}'_1, B \in \mathcal{E}'_2} |\mathbb{P}(B|A) - \mathbb{P}(B)|,$$

and we may conclude that

$$\varphi(\mathcal{E}_1, \mathcal{E}_2) = \sup_{A \in \mathcal{E}_1, B \in \mathcal{E}_2} |\mathbb{P}(B|A) - \mathbb{P}(B)| \leq \varphi(k, r - 2j_*) \leq \varphi(r - 2j_*).$$

Using the inequality $\alpha(\mathcal{E}_1, \mathcal{E}_2) \leq (1/2)\varphi(\mathcal{E}_1, \mathcal{E}_2)$, see e.g. [6, Theorem 25.16], we therefore obtain the estimate

$$\widehat{\alpha}_{Y(j)}(\cdot; |\mathbf{b}|) \leq (1/2)\varphi(\cdot - 2j_*). \tag{47}$$

Put $N_j = n_j - b_j + 1$, $j = 1, \dots, q$, and $\mathbf{N} = (N_1, \dots, N_q)$. In [22] it is shown that $L_{\mathbf{n}, \mathbf{b}}(x)$ converges in probability to $J(x, P)$, if (13) holds and the mixing condition

$$\frac{1}{|\mathbf{N}|} \sum_{k=1}^{N^*} k^{q-1} \widehat{\alpha}_\xi(k; |\mathbf{b}|) \rightarrow 0,$$

as $\mathbf{n} \rightarrow \infty$, is satisfied. By virtue of this estimate, the assertion can be shown as follows: Denote by $\widetilde{L}_{\mathbf{n}, \mathbf{b}}$ the quantity $L_{\mathbf{n}, \mathbf{b}}$ with the centering $\widehat{\theta}_{\mathbf{n}}$ replaced by $\theta(\mathbb{P})$. Since $\mathbb{E}(\widetilde{L}_{\mathbf{n}, \mathbf{b}}(x)) = J_{\mathbf{b}}(x, \mathbb{P}) \xrightarrow{\mathbb{P}} J(x, \mathbb{P})$, as $\mathbf{b} \rightarrow \infty$, it suffices to show that $\text{Var}(\widetilde{L}_{\mathbf{n}, \mathbf{b}}(x)) = o(1)$, as $\mathbf{n} \rightarrow \infty$. Letting

$$I_N = \{\mathbf{i} \in \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j}) : |\mathbf{i}| \leq N\} \quad \text{and} \quad I^* = \{\mathbf{i} \in \widetilde{\Gamma}_{\mathbf{n}}(\mathbf{j}) : |\mathbf{i}| \leq \mathbf{b}\}$$

and decomposing the variance as $\text{Var}(\widetilde{L}_{\mathbf{n}, \mathbf{b}}(x)) = A^* + A$ with

$$A^* = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{i} \in I^*} \prod_{j=1}^q \frac{N_j - |i_j|}{N_j} c(\mathbf{i}), \quad A = \frac{1}{|\mathbf{N}|} \sum_{\mathbf{i} \in I_N \setminus I^*} \prod_{j=1}^q \frac{N_j - |i_j|}{N_j} c(\mathbf{i}),$$

where $c(\mathbf{i}) = \mathbb{Cov}(1(|\mathbf{b}|^{-1/2}(\widehat{\theta}_{\mathbf{n}, \mathbf{b}, \mathbf{0}} - \theta(\mathbb{P})) \leq x), 1(|\mathbf{b}|^{-1/2}(\widehat{\theta}_{\mathbf{n}, \mathbf{b}, \mathbf{i}} - \theta(\mathbb{P})) \leq x))$, [22] establish the estimates

$$|A^*| = O\left(\prod_{j=1}^q b_j / (n_j - b_j)\right) = o(1),$$

as $\mathbf{n} \rightarrow \infty$, and

$$|A| \leq c_2 \frac{1}{|\mathbf{N}|} \sum_{k=\lfloor b^*/h_* \rfloor + 1}^{N^*} k^{q-1} \widehat{\alpha}_{Y(j)}(kh_* - b^*; |\mathbf{b}|),$$

for some constant c_2 , under the condition (13). The proof of the first assertion can now be completed by estimating the last expression using (47).

Recall that a sequence of distribution functions converges in all continuity points of the limit function, if and only if the associated sequence of quantile functions converges in all continuity points of the limiting quantile function, see [26, Lemma 21.2]. Consequently, since by continuity of $J(x, \mathbb{P})$, $x \in \mathbb{R}$,

$$\sup_{x \in \mathbb{R}} |J_n(x, \mathbb{P}) - J(x, \mathbb{P})| \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$, we obtain

$$\sup_{\gamma \in (0,1)} |q_{n,b}(\gamma) - q(\gamma)| \xrightarrow{\mathbb{P}} 0,$$

as $n \rightarrow \infty$. Therefore,

$$|n|^{1/2}(\widehat{\theta}_n - \theta(\mathbb{P})) - q_{n,b}(\gamma) \Rightarrow \mathcal{R} - q(\gamma),$$

as $n \rightarrow \infty$, where $\mathcal{R} \sim J(x, \mathbb{P})$, which implies

$$\mathbb{P}(|n|^{1/2}(\widehat{\theta}_n - \theta(\mathbb{P})) \leq q_{n,b}(\gamma)) \rightarrow \mathbb{P}(\mathcal{R} \leq q(\gamma)) = \gamma,$$

as $n \rightarrow \infty$, which proves the remaining assertions. □

5.2. Proofs of Section 3

Let us now establish the consistency of the thresholding estimators of the asymptotic variance.

Proof of Theorem 3.1 and Theorem 3.2. Let us first consider the univariate case $p = 1$. We show that $|\widehat{\sigma}_n^2 - \widehat{\sigma}_{n,th}^2| \rightarrow 0$ in probability and then the assertion follows with Theorem 2.1. First, we have

$$\begin{aligned} |\widehat{\sigma}_n^2 - \widehat{\sigma}_{n,th}^2| &\leq \left| \sum_{|j| \leq m} w_m(j) (\widehat{\gamma}_n(j) - \gamma(j)) \right| \\ &+ \left| \sum_{|j| \leq m} w_m(j) (\gamma(j) - \gamma(j)g(\widehat{\gamma}_n(j), c_n(j))) \right| \\ &+ \left| \sum_{|j| \leq m} w_m(j) (\gamma(j) - \widehat{\gamma}_n(j)) g(\widehat{\gamma}_n(j), c_n(j)) \right| =: I_1 + I_2 + I_3. \end{aligned}$$

Analogously to the proof of Theorem 2.1 we directly obtain $I_1 \xrightarrow{\mathbb{P}} 0$ and $I_3 \xrightarrow{\mathbb{P}} 0$ for $n \rightarrow \infty$, as g is bounded, so we only have to consider I_2 . With the Markov inequality we obtain

$$\begin{aligned} \mathbb{P}(|I_2| > \varepsilon) &= \mathbb{P}\left(\left| \sum_{|j| \leq m} w_m(j) \gamma(j) (1 - g(\widehat{\gamma}_n(j), c_n(j))) \right| > \varepsilon \right) \\ &\leq \frac{1}{\varepsilon} \sum_{|j| \leq m} |w_m(j)| |\gamma(j)| \mathbb{E}(1 - g(\widehat{\gamma}_n(j), c_n(j))) \end{aligned}$$

$$\begin{aligned} &\leq \frac{C_w}{\varepsilon} \sum_{\mathbf{j} \in \mathbb{Z}^q} |\gamma(\mathbf{j})| \mathbb{E}(1 - g(\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j}))) \mathbb{1}_{\{|\mathbf{j}| \leq m\}} \\ &= \frac{C_w}{\varepsilon} \sum_{\mathbf{j} \in \mathbb{Z}^q} f_{\mathbf{n}}(\mathbf{j}) \end{aligned}$$

with

$$f_{\mathbf{n}}(\mathbf{j}) = |\gamma(\mathbf{j})| \mathbb{E}(1 - g(\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j}))) \mathbb{1}_{\{|\mathbf{j}| \leq m\}}.$$

We now want to apply the dominated convergence theorem. As $|f_{\mathbf{n}}(\mathbf{j})| \leq |\gamma(\mathbf{j})|$ with $\sum_{\mathbf{j} \in \mathbb{Z}^q} |\gamma(\mathbf{j})| < \infty$ by Lemma 1 (a) in [11] we already have a convergent majorant and since

$$\mathbb{E}(g(\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j}))) \rightarrow 1,$$

as $\mathbf{n} \rightarrow \infty$, by Assumption 2 the assertion follows. For the multivariate case denote the (ν, μ) th element of $\widehat{\sigma}_{\mathbf{n}}^2$ by $(\widehat{\sigma}_{\mathbf{n}}^2)_{\nu\mu}$, $(\widehat{\sigma}_{\mathbf{n},th}^2)_{\nu\mu}$ denotes the corresponding entry of $\widehat{\sigma}_{\mathbf{n},th}^2$, $1 \leq \nu, \mu \leq p$. Now observing that

$$\|\widehat{\sigma}_{\mathbf{n}}^2 - \widehat{\sigma}_{\mathbf{n},th}^2\|_{\infty} \leq \sum_{\nu, \mu=1}^p |(\widehat{\sigma}_{\mathbf{n}}^2)_{\nu\mu} - (\widehat{\sigma}_{\mathbf{n},th}^2)_{\nu\mu}|$$

the consistency follows from the above estimates elaborated for the univariate case. \square

Proof of Corollary 3.1. We only have to show that the special choice of g in Corollary 3.1 fulfills Assumption 2 and then the assertion directly follows by Theorem 3.1. This means that we have to show that

$$\mathbb{E}(1 - g(\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}), c_{\mathbf{n}}(\mathbf{j}))) = P(|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| \leq c_{\mathbf{n}}(\mathbf{j})) \rightarrow 0$$

for all $\mathbf{j} \in \mathbb{Z}^q$. For that observe that by Lemma 5.3 in [25] and the assumptions of the corollary we have

$$|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| - c_{\mathbf{n}}(\mathbf{j}) \rightarrow |\gamma(\mathbf{j})| - c(\mathbf{j}) =: X$$

in probability, as $\mathbf{n} \rightarrow \infty$. This implies the convergence in distribution, i.e. if $F_{|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| - c_{\mathbf{n}}(\mathbf{j})}$ denotes the distribution function of $\widehat{\gamma}_{\mathbf{n}}(\mathbf{j}) - c_{\mathbf{n}}(\mathbf{j})$ and F_X the one of X , we have

$$\mathbb{P}(|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| \leq c_{\mathbf{n}}(\mathbf{j})) = F_{|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| - c_{\mathbf{n}}(\mathbf{j})}(0) \rightarrow F_X(0),$$

as $\mathbf{n} \rightarrow \infty$, if zero is a continuity point of F_X . But since

$$F_X(x) = \begin{cases} 0, & x < |\gamma(\mathbf{j})| - c(\mathbf{j}) \\ 1, & x \geq |\gamma(\mathbf{j})| - c(\mathbf{j}) \end{cases}$$

and $c(\mathbf{j}) < |\gamma(\mathbf{j})|$ by assumption this is fulfilled. Moreover, the last inequality also implies that $F_X(0) = 0$. Thus we have $\mathbb{P}(|\widehat{\gamma}_{\mathbf{n}}(\mathbf{j})| \leq c_{\mathbf{n}}(\mathbf{j})) \rightarrow 0$ for all $\mathbf{j} \in \mathbb{Z}^q$ as $\mathbf{n} \rightarrow \infty$ which completes the proof. \square

Acknowledgments

The authors thank an anonymous associate editor and an anonymous referee for their helpful comments. Part of the work of the second author has been supported by a grant from Deutsche Forschungsgemeinschaft (DFG), grant No. STE 1034/11-1, which he gratefully acknowledges.

References

- [1] Alexopoulos, C. and Goldsman, D. (2004). To batch or not to batch? *ACM Trans. Model. Comput. Simul.*, 14(1), 76–114.
- [2] Andrews, D. W. K. (1991). *Heteroskedasticity and autocorrelation consistent covariance matrix estimation*. *Econometrica*, 69(3), 817–858. [MR1106513](#)
- [3] Berkes, I. and Morrow, G. (1981). Strong invariance principles for mixing random fields. *Probability theory and related fields*, 57, 15–37. [MR0623453](#)
- [4] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Statistics, Probability and Statistics. John Wiley & Sons Inc., New York. [MR1700749](#)
- [5] Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd ed. Wiley Series in Probability and Statistics, Probability and Statistics. John Wiley & Sons Inc., New York. [MR1700749](#)
- [6] Bradley, R. C. (2005). *Introduction to Strong Mixing Conditions*, Volume 3, Kendrick Press, Heber City, U.S.A. [MR2325296](#)
- [7] Bulinski, A. V. and Kaene, M. S. (1996). Invariance principle for associated random fields. *Journal of Mathematical Sciences*, 81(5), 2905–2911. [MR1420893](#)
- [8] Bulinski, A. and Shaskin, A. (2006). *Strong invariance principle for dependent random fields*. *Dynamics & Stochastics, IMS Lecture Notes - Monograph Series*, 48, 128–143. [MR2306195](#)
- [9] Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, 92(1), 1–45. [MR1707000](#)
- [10] Davidson, J. (1994). *Stochastic Limit Theory*. Advanced texts in econometrics. Oxford Univ. Press, Oxford. [MR1430804](#)
- [11] Deo, C. M. (1975). *A functional central limit theorem for stationary random fields*. *The Annals of Probability*, 3(4), 708–715. [MR1106280](#)
- [12] Doukhan, P. (1994). *Mixing: Properties and Examples*. Lecture Notes in Statistics. Springer-Verlag, New York. [MR1312160](#)
- [13] Driscoll, J. C. and Kraay, A. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics*, 80(4), 549–560.
- [14] El Machkouri, M., Volný, D. and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, 123(1), 1–14. [MR2988107](#)
- [15] Kelejian, H. H. and Prucha, I. R. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140(1), 131–154. [MR2395919](#)

- [16] Kim, M. S. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160(2), 349–371. [MR2748557](#)
- [17] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory*, 26, 1218–1245. [MR2660298](#)
- [18] Newey, W. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–08. [MR0890864](#)
- [19] Marinucci, D. and Poghosyan, S. (2001). Asymptotics for linear random fields. *Statistics and Probability Letters*, 51(2), 131–141. [MR1808035](#)
- [20] Paulauskas, V. (2010). On Beveridge-Nelson decomposition and limit theorems for linear random fields. *Journal of Multivariate Analysis*, 101(3), 621–639.
- [21] Pawlak, M. and Steland, A. (2013). *Nonparametric sequential signal change detection under dependent noise*. IEEE Transactions on Information Theory, 59(6), 3514–3531. [MR3061262](#)
- [22] Politis, D. N., Romano, J. P. and Wolf, M. (1999). *Subsampling*. Springer series in statistics. Springer. [MR1707286](#)
- [23] Prause, A. (2015). *Sequential Nonparametric Detection of High-Dimensional Signals Under Dependent Noise*. PhD Thesis, RWTH Aachen University.
- [24] Prause, A. and Steland, A. (2015). *Detecting changes in spatial-temporal image data based on quadratic forms*. In: Ansgar Steland, Krzysztof Szajowski and Ewaryst Rafajłowicz (Eds.), *Stochastic Models, Statistics and Their Applications*, pp. 139–147. Heidelberg: Springer [MR3336433](#)
- [25] Prause, A. and Steland, A. (2017). *Sequential detection of three-dimensional signals under dependent noise*. *Sequential Analysis*, 36, 151–178. [MR3665833](#)
- [26] van der Vaart, A. W. (2000). *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge. [MR1652247](#)
- [27] Wang, Y. and Woodroffe, M. (2013). A new condition for the invariance principle for stationary random fields. *Statistica Sinica*, 23, 1673–1696. [MR3222815](#)
- [28] Wu, W. B. (2009). Recursive estimation of time-average variance constants. *The Annals of Applied Probability*, 19(4), 1529–1552.