Ann. Inst. Statist. Math.
Vol. 50, No. 1, 165–186 (1998)

ON MAXIMUM LIKELIHOOD ESTIMATION FOR GAUSSIAN SPATIAL AUTOREGRESSION MODELS

JAROSLAV MOHAPL

Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

(Received September 25, 1996; revised February 17, 1997)

Abstract. The article presents a central limit theorem for the maximum likelihood estimator of a vector-valued parameter in a linear spatial stochastic difference equation with Gaussian white noise right side. The result is compared to the known limit theorems derived for the approximate likelihood e.g. by Whittle (1954, *Biometrika*, **41**, 434–439), Guyon (1982, *Biometrika*, **69**, 95–105) and Rosenblatt (1985, *Stationary Sequences and Random Fields*, Birkhäuser, Boston) and to the asymptotic properties of the quasi-likelihood studied by Heyde and Gay (1989, *Stochastic Process. Appl.*, **31**, 223–236; 1993, *Stochastic Process. Appl.*, **45**, 169–182). Application of the theory is demonstrated on several classes of models including the one considered by Niu (1995, *J. Multivariate Anal.*, **55**, 82–104).

Key words and phrases: Spatial process, asymptotic normality, consistency, lattice sampling, stochastic difference equation.

1. Introduction

Parameter estimation in spatial autoregression models (*SPAR* models) concerns stationary temporal-spatial processes satisfying the equation

$$\mathcal{P}_\theta \xi(n) = \epsilon(n), \qquad n \in \boldsymbol{Z}^d, \tag{1.1}$$

where $\mathcal{P}_\theta$ is a linear difference operator and ϵ is a Gaussian white noise. The maximum likelihood procedure has been studied already since early fifties. Whittle (1954) noticed, that ordinary least squares applied to a general *SPAR* process may lead to "nonsensical results". He observed, that failure of the least squares may occur even in case of a one dimensional processes, but he did not specify the reason. Ord (1975) pointed out, that the bias of least squares (LS) estimating equations one may encounter in some cases, is responsible for inconsistency of the LS estimators. The equation $\theta\xi(n) - \xi(n-1) = \epsilon(n)$, where $n \in \boldsymbol{Z}$, $|\theta| > 1$ and ϵ consists of $N(0, \sigma^2)$ i.i.d. r.v.'s, provides the simplest example. If we estimate θ directly, without reparametrizing to the commonly used $AR(1)$ model, then the

LS estimator is inconsistent. Instead of investigating the likelihood itself, Whittle modified the residual sum of squares and replaced the Gaussian likelihood with an approximate likelihood function. The logarithm of approximate likelihood yields the well known measure Q of discrepancy between the spectral density f_θ and the periodogram $\hat{f}$ of the process ξ defined by the relation

$$Q(\hat{f}, f_\theta) = \int \ln f_\theta(\lambda) d\lambda + \int \hat{f}(\lambda)/f_\theta(\lambda) d\lambda. \tag{1.2}$$

The periodogram of the process is fairly easy to compute and provides an asymptotically unbiased estimator of the spectral density function. However, the estimating equations obtained by minimizing (1.2) subject to θ are in general biased. Inconsistency of estimators generated by (1.2) led Guyon (1982) to a thorough revision of the Whittle (1954) central limit theorem for approximate likelihood estimators and suitable modifications like bias reduction and smoothing of the periodogram. Recall that sometimes variance of the periodogram fails to converge to zero as the number of observations grows to infinity. Guyon (1982) and Rosenblatt (1985) formulated interesting theorems on asymptotic behavior of estimators computed by means of smoothed periodograms. In order to remove the bias, Heyde and Gay (1989, 1993) suggested replacing Whittle's approximate likelihood estimating equations by the quasi-likelihood estimating equations of the form

$$\int (\hat{f} - E_\theta \hat{f}) \frac{\partial_{\theta_k} f_\theta}{f_\theta^2} = 0, \qquad k = 1, \ldots, r. \tag{1.3}$$

They claim that equation (1.3) yields consistent, asymptotically normal estimators under certain assumptions, referring to Rosenblatt's (1985) proof.

The aim of this paper is to study the conditional likelihood function of a general stationary Gaussian $SPAR$ process and related maximum likelihood (ML) estimators. Mardia and Marshall (1984) studied asymptotic properties of the unconditional ML for an arbitrary lattice process. The general result concerning consistency and asymptotic normality of the ML estimator is obtained under regularity assumptions that are difficult to verify, because asymptotic behavior of the covariance matrix's eigenvalues must be known. In Section 6 of their paper, they compare the ML with the approximate likelihood estimators for a data set. One may use the relationship between the approximate, conditional and unconditional likelihood to compare the properties of the corresponding estimators in the case of unilateral processes. For such processes the approximate and conditional estimators have asymptotically the same normal distribution. The optimality of the unconditional maximum likelihood estimating function implies that the difference between the asymptotic covariance matrix of conditional and unconditional ML estimators is non-negative definite, see Godambe and Heyde (1987).

The number of results concerning exact conditional and unconditional likelihoods is continuously increasing in recent years. They focus on specific models and the $SPAR$ processes are investigated rather as a part of weighted linear regression problems. It is worthwhile to mention papers by Niu (1995), Basu and

Reinsel (1994), Martin (1990) and references there in. There is also a series of papers dealing with unilateral processes, c.f. Tjøstheim (1978, 1983). It is to note that the possibility to give a $SPAR$ process a unilateral representation does not mean the problem of parameter estimation for $SPAR$ processes can be reduced to estimation in processes with unilateral representation, c.f. Whittle (1954).

Here we give the exact conditional likelihood function of a general $SPAR$ process and the limit distribution of its consistent ML estimator. We compare the result to those by Whittle (1954) and Heyde and Gay (1989, 1993). We also demonstrate how our central limit theorem applies to certain classes of $SPAR$ processes.

2. The main results

In this section we introduce the notation used throughout the paper, derive the likelihood function, show its relation to the approximate likelihood and the quasi-likelihood and finally formulate the central limit theorem.

A $SPAR$ process ξ is formed by random variables $\xi(n)$ indexed by multi-indices $n \in \boldsymbol{Z}^d$, where $\boldsymbol{Z}^d = \{n : n = (n_1, \ldots, n_d)$ and n_k is an integer$\}$. The Gaussian white noise is a spatial process $\epsilon = (\epsilon(n))_{n \in \boldsymbol{Z}^d}$ such that $\epsilon(n)$ are i.i.d. r.v.'s, each with distribution $N(0, \sigma^2)$. The linear difference operator $\mathcal{P}_\theta$ from (1.1) is defined by the relation

$$\mathcal{P}_\theta \xi(n) = \sum_{k \in \mathcal{K}} a_k(\theta) \xi(n-k), \qquad n \in \boldsymbol{Z}^d, \tag{2.1}$$

where $\mathcal{K}$ is a fixed subset of $\boldsymbol{Z}^d$ and θ belongs to a parameter set $\Theta \subset (-\infty, \infty)^r$. The power r is a natural number. Notice that components of any $k \in \mathcal{K}$ may be negative or positive. Hence, $\mathcal{P}_\theta$ contains generally both forward and backward shifts. Let us denote by $\langle n, \lambda \rangle$ the Euclidean inner product of two elements $n, \lambda \in (-\infty, \infty)^d$: $\langle n, \lambda \rangle = n_1\lambda_1 + \cdots + n_d\lambda_d$. The difference operator $\mathcal{P}_\theta$ determines the so called characteristic polynomial P_θ defined by the relation

$$\mathcal{P}_\theta e^{i\langle n, \lambda \rangle} - P_\theta(\lambda) e^{i\langle n, \lambda \rangle}. \tag{2.2}$$

It is easy to see that P_θ is a polynomial in variables $e^{i\lambda_k}$ and $e^{-i\lambda_k}$, $k = 1, \ldots, d$, respectively. Due to the periodicity of the functions $e^{i\lambda_k}$ we may restrict our considerations to $\lambda \in [-\pi, \pi]^d$. According to Cramér (1940), if ϵ is a spatial Gaussian white noise, then there is an orthogonal Gaussian random measure Z on the Borel σ-algebra $\mathcal{B}(-\infty, \infty)^d$ such that

$$\epsilon(n) = \int e^{i\langle n, \lambda \rangle} Z(d\lambda) \tag{2.3}$$

almost surely (a.s.), $EZ(d\lambda) = 0$, $EZ^2(d\lambda) = \sigma^2 d\lambda/(2\pi)^d$ and $\lambda \in [-\pi, \pi]^d$. If $P_\theta(\lambda)$ has no zeros in $[-\pi, \pi]^d$ for every θ and the function $R(n) = \int e^{i\langle n, \lambda \rangle} |P_\theta(\lambda)|^{-2} d\lambda$ is non-negative definite, then the stationary solution to (1.1) may be described by the equation

$$\xi(n) = \int e^{i\langle n, \lambda \rangle} Z_\xi(d\lambda), \tag{2.4}$$

where $E_\theta Z_\xi(d\lambda) = 0$, $E_\theta Z_\xi^2(d\lambda) = \sigma^2(2\pi)^{-d}|P_\theta(\lambda)|^{-2}d\lambda$ and $\lambda \in [-\pi,\pi]^d$. Conversely, the residuals of the process described by (1.1) admit the representation

$$\epsilon_\theta(n) = \int P_\theta(\lambda)e^{i(n,\lambda)}Z_\xi(d\lambda), \tag{2.5}$$

where the subscript emphasizes the dependence of the white noise representation on θ under which ξ was generated.

THEOREM 2.1. *Let ϵ be a zero mean Gaussian white noise process with variance σ^2, let ξ satisfy the difference equation* (1.1) *and let $P_\theta(\lambda)$ have coefficients differentiable with respect to θ and no zeros in λ, $\lambda \in [-\pi,\pi]^d$. Suppose we compute from a set $(\xi(n))_{n\in\mathcal{N}}$ of observations N residuals $(\epsilon(n))_{n\in\mathcal{N}_0}$, $\mathcal{N}_0 \subset \mathcal{N}$, and the model* (1.1) *provides a one to one mapping between the sets $(\xi(n))_{n\in\mathcal{N}_0}$ and $(\epsilon(n))_{n\in\mathcal{N}_0}$ when conditioning on the observations in $\mathcal{N}\backslash\mathcal{N}_0$. Then the conditional likelihood of the observations $(\xi(n))_{n\in\mathcal{N}_0}$ given $(\xi(n))_{n\in\mathcal{N}\backslash\mathcal{N}_0}$ can be described by the equation*

$$\begin{aligned} &l_\theta(\xi(n), n \in \mathcal{N}_0 \mid \xi(n), n \in \mathcal{N}\backslash\mathcal{N}_0) \\ &\quad = \frac{v_N}{(2\pi v(\theta)\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n\in\mathcal{N}_0}(\mathcal{P}_\theta\xi(n))^2\right\}, \end{aligned} \tag{2.6}$$

where v_N is a constant that does not depend on θ and $v(\theta)$ is described by the relation

$$v(\theta) = \exp\left\{(2\pi)^{-d}\int \ln|P_\theta(\lambda)|^{-2}d\lambda\right\}, \tag{2.7}$$

where we integrate over $[-\pi,\pi]^d$.

PROOF. The likelihood for N i.i.d. zero mean Gaussian random variables $\epsilon_\theta(n)$ is given by the relation

$$l_\theta(\epsilon_\theta(n), n \in \mathcal{N}_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n\in\mathcal{N}_0}\epsilon_\theta^2(n)\right\}. \tag{2.8}$$

The likelihood of the observations $(\xi(n))_{n\in\mathcal{N}_0}$ conditioned on $(\xi(n))_{n\in\mathcal{N}\backslash\mathcal{N}_0}$ is obtained by substitution for ϵ_θ in (2.8) from (2.5). The new likelihood is of the form

$$\begin{aligned} &l_\theta(\xi(n), n \in \mathcal{N}_0 \mid \xi(n), n \in \mathcal{N}\backslash\mathcal{N}_0) \\ &\quad = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n\in\mathcal{N}_0}\left[\int P_\theta(\lambda)e^{i(n,\lambda)}Z_\xi(d\lambda)\right]^2\right\} J_N(\theta), \end{aligned} \tag{2.9}$$

where $J_N(\theta)$ is the Jacobian of the transformation. The Jacobian is well defined because of the one to one assumption about $(\xi(n))_{n\in\mathcal{N}_0}$ and $(\epsilon(n))_{n\in\mathcal{N}_0}$. We determine $J_N(\theta)$ from the well known condition

$$E_\theta \partial_{\theta_k} \ln l_\theta(\xi) = \partial_{\theta_k} \int l_\theta(\xi) d\xi = 0. \tag{2.10}$$

The likelihood l_θ must satisfy (2.10) for every $k = 1, \ldots, r$ and this happens only if

$$\partial_{\theta_k} \ln J_N(\theta) = \frac{1}{2\sigma^2} \sum_{n\in\mathcal{N}_0} E_\theta \partial_{\theta_k} \left[\int P_\theta(\lambda) e^{i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right]^2. \tag{2.11}$$

The expectation in (2.11) may be calculated as follows:

$$\begin{aligned}
&E_\theta \partial_{\theta_k} \left| \int P_\theta(\lambda) e^{i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right|^2 \\
&= E_\theta \left\{ \left(\int \partial_{\theta_k} P_\theta(\lambda) e^{i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right) \left(\int P_\theta^*(\lambda) e^{-i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right) \right. \\
&\qquad \left. + \left(\int P_\theta(\lambda) e^{i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right) \left(\int \partial_{\theta_k} P_\theta^*(\lambda) e^{-i\langle n,\lambda\rangle} Z_\xi(d\lambda) \right) \right\} \\
&= \frac{\sigma^2}{(2\pi)^d} \int \{ \partial_{\theta_k} P_\theta(\lambda) P_\theta^*(\lambda) + P_\theta(\lambda) \partial_{\theta_k} P_\theta^*(\lambda) \} \frac{1}{|P_\theta(\lambda)|^2} d\lambda \\
&= \frac{\sigma^2}{(2\pi)^d} \int \partial_{\theta_k} |P_\theta(\lambda)|^2 |P_\theta(\lambda)|^{-2} d\lambda = \partial_{\theta_k} \frac{\sigma^2}{(2\pi)^d} \int \ln |P_\theta(\lambda)|^2 d\lambda.
\end{aligned} \tag{2.12}$$

The asterisk denotes the complex conjugate number. The proof is now straight forward.

The representation (2.6), (2.7) goes back to Kolmogorov (1941), who derived it for one-sided stationary time series (c.f. also Whittle (1953, 1954)). The following assertion due to Whittle (1954) states that for large samples the likelihood (2.6) may be approximated by the function (2.13) below.

COROLLARY 2.1. *Consider the function*

$$\begin{aligned}
&\hat{l}_\theta(\xi(n), n \in \mathcal{N}) \\
&= \frac{v_N}{(2\pi\sigma^2)^{N/2}} \exp\left\{ -\frac{N}{2} \left\{ \ln \frac{(2\pi)^d}{\sigma^2} + \frac{1}{(2\pi)^d} \left[\int \ln f_\theta(\lambda) d\lambda \right.\right.\right. \\
&\qquad \left.\left.\left. + \int \frac{\hat{f}_N(\lambda)}{f_\theta(\lambda)} d\lambda \right] \right\} \right\},
\end{aligned} \tag{2.13}$$

where we define

$$\hat{f}_N(\lambda) = \frac{1}{(2\pi)^d} \sum_{m\in\mathcal{N}} \hat{C}_m e^{i\langle m,\lambda\rangle}, \qquad \hat{C}_m = \frac{1}{N} \sum_{n\in\mathcal{M}(m)} \xi(n)\xi(n+m), \tag{2.14}$$

$$f_\theta(\lambda) = \frac{\sigma^2}{(2\pi)^d} |P_\theta(\lambda)|^{-2}, \tag{2.15}$$

$\mathcal{M}(m) = \{n : n \in \mathcal{N}, n + m \in \mathcal{N}\}$ and N is the number of observations in $\mathcal{N}$. Otherwise we preserve the notation from Theorem 2.1. If the sample size grows to infinity then the function (2.13) converges to the likelihood (2.6) almost surely.

PROOF. To simplify the notation in the proof we suppress writing of the parameter θ. In sake of precision we assume that $\mathcal{N} \subset \boldsymbol{Z}^d$ contains a d-dimensional rectangle. By sample size growing to infinity we understand that all sides of the rectangle grow to infinity. The function P is a polynomial in variables $e^{i\lambda_k}$ and $e^{-i\lambda_k}$, $k = 1, \dots, d$, with coefficients a_n, $n \in \boldsymbol{Z}^d$. All but finitely many of the coefficients are zero. Hence

$$|P(\lambda)|^2 = \sum_{k \in \boldsymbol{Z}^d} \sum_{l \in \boldsymbol{Z}^d} a_k a_l^* e^{i\langle k-l, \lambda\rangle} = \sum_{m \in \boldsymbol{Z}^d} \sum_{k \in \boldsymbol{Z}^d} a_k a_{k+m}^* e^{-i\langle m, \lambda\rangle}. \tag{2.16}$$

The function $|P|^2$ itself may be expanded into a series with respect to the complete orthonormal system $((2\pi)^{-d/2} e^{-i\langle n,\lambda\rangle})_{n \in \boldsymbol{Z}^d}$:

$$|P(\lambda)|^2 = (2\pi)^{-d/2} \sum_{m \in \boldsymbol{Z}^d} c_m e^{-i\langle m,\lambda\rangle}. \tag{2.17}$$

Comparison of (2.16) and (2.17) yields the equality $(2\pi)^{-d/2} c_m = \sum_{k \in \boldsymbol{Z}^d} a_k a_{k+m}^*$ for all $m \in \boldsymbol{Z}^d$. According to the strong law of large numbers in Section 4, the averages

$$\hat{C}_{k,m} = \frac{1}{N} \sum_{n \in \mathcal{N}_0} \xi(k+n)\xi(k+n+m), \qquad \hat{C}_m = \frac{1}{N} \sum_{n \in \mathcal{M}(m)} \xi(n)\xi(n+m), \tag{2.18}$$

converge to $R(m)$ almost surely as N increases to infinity. Consequently for large samples $\hat{C}_{k,m}$ may be replaced by $\hat{C}_m$ and

$$\begin{aligned} \frac{1}{N} \sum_{n \in \mathcal{N}_0} (\mathcal{P}\xi(n))^2 &= \sum_{k \in \boldsymbol{Z}^d} \sum_{l \in \boldsymbol{Z}^d} a_k a_l^* \frac{1}{N} \sum_{n \in \mathcal{N}_0} \xi(k+n)\xi(l+n) \\ &\approx (2\pi)^{-d/2} \sum_{m \in \boldsymbol{Z}^d} c_m \hat{C}_m. \end{aligned} \tag{2.19}$$

Expansion of $\hat{f}_N$ with respect to the same system of orthonormal functions has coefficients $(2\pi)^{-d/2}\hat{C}_m$ for $m \in \mathcal{N}$ and zero for other m. The Parseval equality combined with (2.17) thus says that $\int \hat{f}_N(\lambda)|P(\lambda)|^2 d\lambda = (2\pi)^{-d/2} \sum_{m \in \boldsymbol{Z}^d} \hat{C}_m c_m$ and the result follows.

Minimization of $Q(\hat{f}, f_\theta)$ in (1.2) with respect to θ leads to the approximate likelihood estimating equations

$$\int (1 - \hat{f}_N(\lambda)) \frac{\partial_{\theta_k} f_\theta(\lambda)}{f_\theta^2(\lambda)} d\lambda = 0, \qquad k = 1, \dots, r. \tag{2.20}$$

Their solution, if exists, determines the so called approximate likelihood estimator. The estimating equations (2.20) are generally biased. That means the expectation of the left side of (2.20) is not zero for all $\theta \in \Theta$.

Remark 2.1. If we replace $\hat{f}_N$ in (2.20) by $\tilde{f}_N$ defined by

$$\tilde{f}_N(\lambda) = \frac{1}{(2\pi)^d}\frac{N_0}{N}\sum_{m\in\mathcal{M}}\tilde{C}_m e^{i\langle m,\lambda\rangle}, \qquad \tilde{C}_m = \frac{1}{N_m}\sum_{n\in\mathcal{M}(m)}\xi(n)\xi(n+m), \tag{2.21}$$

where $\mathcal{M} = \{m : c_m \neq 0, m \in \boldsymbol{Z}^d\}$ and N_m is the power of $\mathcal{M}(m)$, then the estimating equations (2.20) are unbiased.

This means the risk of inconsistency of the approximate likelihood estimator caused by the bias of estimating equations can be reduced by using in (1.2) the periodogram (2.21) instead of $\hat{f}_N$. In order to prove it, consider in the previous proof the periodogram (2.21) instead of (2.14). Define $\tilde{C}_m = 0$ in case when $\mathcal{M}(m)$ is an empty set and follow the proof to (2.19). Then take derivatives on both sides of the approximate equality (2.19) and calculate the expected values. Rest of the verification is trivial. Replacement of the biased periodogram estimator by the unbiased one was advocated already by Guyon (1982). Heyde and Gay (1989) suggested to remove the bias by replacement of one in (2.20) by the expectation of $\hat{f}_N$. This leads to equations (1.3).

In order to formulate the central limit theorem for the conditional maximum likelihood estimator computed from (2.6) we introduce the following regularity assumptions.

i) The function $P_\theta(\lambda)$ has no zeros for $\lambda \in [-\pi,\pi]^d$, the coefficients of $P_\theta(\lambda)$ are three times continuously differentiable and their derivatives, up to third order, are bounded on Θ.

ii) For every $\theta \in \Theta$, the matrices $Q(\theta)$ and $W(\theta)$ with elements

$$\begin{aligned} Q_{j,k}(\theta) &= \frac{1}{2}(2\pi)^{-d}\int \partial_{\theta_j}\ln|P_\theta(\lambda)|^2\partial_{\theta_k}\ln|P_\theta(\lambda)|^2 d\lambda, \\ W_{j,k}(\theta) &= (2\pi)^{-d}\Big(\int \frac{\partial_{\theta_j}P_\theta(\lambda)}{P_\theta(\lambda)}\frac{\partial_{\theta_k}P^*_\theta(\lambda)}{P^*_\theta(\lambda)}d\lambda \\ &\qquad + \frac{1}{4}(2\pi)^{-d}\int \partial_{\theta_j}\ln|P_\theta(\lambda)|^2 d\lambda \int \partial_{\theta_k}\ln|P_\theta(\lambda)|^2 d\lambda\Big), \end{aligned} \tag{2.22}$$

where we integrate over $[-\pi,\pi]^d$, are invertible.

THEOREM 2.2. *Let ξ be a stationary Gaussian process on a rectangular lattice $\mathcal{N}$ satisfying equation* (1.1) *and regularity conditions* (i) *and* (ii). *Suppose we compute from a set $(\zeta(n))_{n\in\mathcal{N}}$ of observations N residuals $(\epsilon(n))_{n\in\mathcal{N}_0}$, $\mathcal{N}_0 \subset \mathcal{N}$. If there exists a sequence $(\hat{\theta}_N)_{N=1}^\infty$ of ML estimates such that in probability $\lim_{N\to\infty}\hat{\theta}_N = \theta_0$ as the sides of $\mathcal{N}$ grow to infinity and θ_0 is the true parameter*, *then*

$$\lim_{N\to\infty}\sqrt{N}(\hat{\theta}_N - \theta_0) = \mathcal{N}(0, W^{-1}(\theta_0)Q(\theta_0)W^{-1}(\theta_0)) \tag{2.23}$$

and

$$\lim_{N\to\infty} \sqrt{N}(\hat{\sigma}_N^2 - \sigma_0^2) = \mathcal{N}(0, 2\sigma_0^4) \tag{2.24}$$

in distribution.

PROOF. See Section 4.

The random variable $\hat{\sigma}_N^2$ is the common maximum likelihood estimator of the true parameter σ_0^2 computed from (2.6).

Remark 2.2. For stationary unilateral processes described by the equation

$$\sum_{k_1=0}^{N_1} \cdots \sum_{k_d=0}^{N_d} a_{k_1,\ldots,k_d}(\theta)\xi(n_1 - k_1, \ldots, n_d - k_d) = \epsilon(n_1, \ldots, n_d), \tag{2.25}$$

where ϵ is a Gaussian white noise, the random variables $\sqrt{N}(\hat{\theta}_N - \theta_0)$ and $\sqrt{N}(\hat{\sigma}_N^2 - \sigma_0^2)$ are asymptotically independent. This is a consequence of the arguments by Rosenblatt (1985), Chapter IV.

A vector of estimating equations is called unbiased if its expectation is zero for each $\theta \in \Theta$. Under fairly broad conditions the solution of an unbiased estimating equation is a consistent, asymptotically normal estimator of the true parameter. In particular, under the assumptions in Heyde and Gay (1989), the matrix $Q^{-1}(\theta_0)$ agrees with the asymptotic covariance matrix of the quasi-likelihood estimator. The maximum likelihood estimating equations are in some sense optimal in the class of unbiased estimating equations. This allows comparison of the asymptotic covariance matrices of the *ML* and quasi-likelihood estimators.

Remark 2.3. If $Q^{-1}(\theta_0)$ is the asymptotic covariance matrix of the quasi-likelihood estimators determined by (1.3) or of the Whittle's approximate likelihood modified using the periodogram (2.21) then $Q^{-1}(\theta_0) - W^{-1}(\theta_0) \cdot Q(\theta_0)W^{-1}(\theta_0)$ is a non-negative definite matrix. See Example 3 in Heyde and Gay (1989), Remark 2.1 and the paper by Godambe and Heyde (1987).

The central limit theorem for estimators obtained from the approximate likelihood function (2.13) was studied in several papers, often for non-Gaussian processes, c.f. Giraitis and Surgailis (1990), Heyde and Gay (1989), Hosoya and Taniguchi (1982), Guyon (1982), Whittle (1953, 1954). It is believed, that under the Gaussianity assumption, a consistent sequence of approximate likelihood estimators substituted in the left side of the equation (2.23), has asymptotically a zero mean normal distribution with covariance matrix $Q^{-1}(\theta_0)$ defined in (2.22), i.e. the same asymptotic distribution as the quasi-likelihood estimators have. This result was first formulated by Whittle (1954). Its criticism can be found in Guyon (1982), who verified this assertion for approximate likelihood estimators modified according to (2.21). However, Guyon seems to work only with the unilateral representation of the studied processes. In order to extend the limit theorems to quasi-likelihood estimators, Heyde and Gay (1989) used an argument by Rosenblatt (1985), which is also valid only for unilateral processes.

Remark 2.4. A simple calculation shows, that for unilateral processes matrices $Q(\theta_0)$ and $W(\theta_0)$ in (2.22) agree. This means, for unilateral processes the exact, approximate and quasi-likelihood estimators have the same asymptotic normal distribution.

Comparison of matrices $Q^{-1}(\theta_0)$ and $W^{-1}(\theta_0)Q(\theta_0)W^{-1}(\theta_0)$ in (2.22) for a general spatial process is not easy.

Remark 2.5. Let $\Theta \subset (-\infty, \infty)$ and suppose that ξ satisfies assumptions of Theorem 2.2. Denote by $q(\theta_0)$ and $w(\theta_0)$ the 1×1 matrices $Q(\theta_0)$ and $W(\theta_0)$, respectively. If the characteristic polynomial of ξ is a real-valued function then $w(\theta_0) = q(\theta_0)$.

In consequence of the Schwartz inequality $q^{-1}(\theta_0) \leq w^{-1}(\theta_0)q(\theta_0)w^{-1}(\theta_0)$. The converse inequality can be derived from Remark 2.3.

Example 2.1. The model $\theta\xi(n) - \xi(n-1) = \sigma\epsilon(n)$ with $|\theta| > 1$ and $\epsilon(n) \sim N(0,1)$ i.i.d. is commonly fitted using reparametrization to the standard $AR(1)$ model by means of the relations $\alpha = 1/\theta$, $\delta = \sigma/\theta$ and $\xi(n) - \alpha\xi(n-1) = \delta\epsilon(n)$. The ML estimating equation utilizing observations $\xi(0), \ldots, \xi(N)$ and obtained by conditioning on $\xi(0)$ provides the consistent estimate of θ, $\hat{\theta}_N = \sum_{n=1}^N \xi^2(n) / \sum_{n=1}^N \xi(n)\xi(n-1)$, with asymptotic distribution $\lim_{N\to\infty} \sqrt{N}(\hat{\theta}_N - \theta) = N(0, (1-\theta^{-2})\theta^4)$.

Direct application of Theorem 2.2 yields the estimating equation

$$\theta^2 \sum_{n=1}^N \xi^2(n) - \theta \sum_{n=1}^N \xi(n)\xi(n-1) - N\sigma^2 = 0. \tag{2.26}$$

If $\theta > 1$ then the consistent estimator of θ is

$$\hat{\theta}_N(\sigma) = \frac{1}{2}(a_N + \sqrt{a_N^2 + 4b_N(\sigma)}), \tag{2.27}$$

where the random variables a and b are defined by the relations

$$a_N = \frac{\sum_{n=1}^N \xi(n)\xi(n-1)}{\sum_{n=1}^N \xi^2(n)}, \qquad b_N(\sigma) = \frac{N\sigma^2}{\sum_{n=1}^N \xi^2(n)}.$$

Under the notation of Remark 2.5 we may calculate $q(\theta) = w(\theta) = \sigma^2(\theta^{-2} + (\theta^2 - 1)^{-1})$. The asymptotic distribution of $\hat{\theta}_N(\sigma)$ is therefore, according to Theorem 2.2, $\lim_{N\to\infty} \sqrt{N}(\hat{\theta}_N(\sigma) - \theta) = N(0, (1-\theta^{-2})\theta^4(2\theta^2-1)^{-1})$. As the reader may notice, if we know the true value of σ then direct utilization of the model $\theta\xi(n) - \xi(n-1) = \sigma\epsilon(n)$ for estimation of $\theta > 1$ provides an estimator with substantially lower variance than the reparametrized model. But if σ is not known and is estimated using the ML estimator $\hat{\sigma}_N^2 = N^{-1}\sum_{n=1}^N (\theta\xi(n) - \xi(n-1))^2$ then we receive the same estimator for θ as by the reparametrization.

3. An unconditional likelihood

Theorem 2.1 can be used in case of unilateral multiplicative processes for computation of the unconditional likelihood function. We derive it for processes on the plane. Generalization to domains with higher dimension is straight forward. We recall that the model described by the equation (1.1) with $d = 2$ is called multiplicative if its characteristic polynomial $P_{(\theta,\vartheta)}(\lambda,\omega) = P_\theta(\lambda)P_\vartheta(\omega)$, where $P_\theta(\lambda) = 1 - \theta_1 e^{-i\lambda} - \cdots - \theta_p e^{-ip\lambda}$ and $P_\vartheta(\omega) = 1 - \vartheta_1 e^{-i\omega} - \cdots - \vartheta_q e^{-iq\omega}$. We take $\lambda, \omega \in \boldsymbol{C}$, p and q are natural numbers. This means the model is described by the equation

$$(3.1)\quad \xi(n,m) - \phi_{1,0}\xi(n-1,m) - \phi_{0,1}\xi(n,m-1) - \cdots - \phi_{p,q}\xi(n-p,m-q) = \epsilon(n,m),$$

where $\phi_{i,j}$ are the coefficients of the characteristic polynomial and $\epsilon(n,m)$ are i.i.d. Gaussian r.v.'s with variance σ^2. The covariance function of this model satisfies the equation $R_{\theta,\vartheta}(n,m) = \sigma^2 R_{1,\theta}(n)R_{2,\vartheta}(m)$, where $R_{1,\theta}(n)$ and $R_{2,\vartheta}(m)$ are covariance functions of ordinary autoregression models of degree p and q, respectively. Suppose the observations are collected from a rectangular lattice of points $\mathcal{N} = \{(n,m) : n = 0,1,\dots,N, m = 0,1,\dots,M\}$, $N \geq p$ and $M \geq q$. Then $\mathcal{N}_0 = \{(n,m) : n = p, p+1, \dots, N, m = q, q+1, \dots, M\}$ and we condition on $(\xi(n))_{n\in\mathcal{N}\setminus\mathcal{N}_0}$. If we fix $m \in \{0,1,\dots,q-1\}$ and set $\zeta(n) = \xi(n,m)$, $n = 0,1,\dots,N$, then conditioning on $\zeta(0),\dots,\zeta(p-1)$, we can derive the likelihood for observations $\zeta(n) = \xi(n,m)$, $n = p, p+1, \dots, N$. Such a likelihood may be obtained for each $m = 0,1,\dots,q-1$ and its form is

$$(3.2)\quad \begin{aligned} & l_{\theta.m}(\xi(p,m),\dots,\xi(N,m) \mid \xi(0,m),\dots,\xi(p-1,m)) \\ &= \frac{1}{(2\pi\sigma^2 R_{2,\vartheta}(0)v_1(\theta))^{(N+1-p)/2}} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=p}^{N}(\xi(n,m) - \theta_1\xi(n-1,m) \right. \\ &\qquad\qquad \left. - \cdots - \theta_p\xi(n-p,m))^2\right\}. \end{aligned}$$

The constant $v_1(\theta) = (2\pi)^{-1}\int_{-\pi}^{\pi} \ln|P_\theta(\lambda)|^{-2}d\lambda$. Now we can fix any $n = 0,1,\dots,p$ and derive a similar likelihood for any row of observations $(\xi(n,m))_{m=q}^{M}$ conditioned on $(\xi(n,m))_{m=0}^{q-1}$. This way we can obtain the conditional likelihood for observations $\{\xi(n,m) : n = p,\dots,N, m = q,\dots,M\}$ conditioned on observations in $\mathcal{N}_c = \{(n,m) : n = 0,\dots,p-1, m = 0,\dots,q-1\}$:

$$(3.3)\quad \begin{aligned} & l_\theta(\xi(n,m),(n,m) \in \mathcal{N}\setminus\mathcal{N}_c \mid \xi(n,m),(n,m) \in \mathcal{N}_c) \\ &= \frac{1}{(2\pi\sigma^2)^{(N+1-p)(M+1-q)/2}} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=p}^{N}\sum_{m=q}^{M}(\xi(n,m) - \phi_{1,0}\xi(n-1,m)\right. \end{aligned}$$

$$
\begin{aligned}
&\qquad - \phi_{0,1}\xi(n, m-1) - \cdots - \phi_{p,q}\xi(n-p, m-q))^2 \Bigg\} \\
&\times \frac{1}{(2\pi\sigma^2 R_{2,\vartheta}(0) v_1(\theta))^{q(N+1-p)/2}} \\
&\times \exp\Bigg\{ -\frac{1}{2\sigma^2 R_{2,\vartheta}(0)} \sum_{n=p}^{N} \sum_{m=0}^{q-1} (\xi(n,m) - \theta_1\xi(n-1,m) \\
&\qquad\qquad - \cdots - \theta_p \xi(n-p,m))^2 \Bigg\} \\
&\times \frac{1}{(2\pi\sigma^2 R_{1,\theta}(0) v_2(\vartheta))^{p(M+1-q)/2}} \\
&\times \exp\Bigg\{ -\frac{1}{2\sigma^2 R_{1,\theta}(0)} \sum_{n=0}^{p-1} \sum_{m=q}^{M} (\xi(n,m) - \vartheta_1\xi(n,m-1) \\
&\qquad\qquad - \cdots - \vartheta_q \xi(n,m-q))^2 \Bigg\},
\end{aligned}
$$

where $v_2(\vartheta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \ln |P_\vartheta(\omega)|^{-2} d\omega$ and the coefficients $\phi_{j,k}$ depend on θ and ϑ. The unconditional likelihood for the observations with indices $(n,m) \in \mathcal{N}_c$ must be computed directly from the definition. Combined with (3.3) it yields the unconditional likelihood.

Example 3.1. For $p = 1$ and $q = 1$, relation (3.3) leads to the likelihood function of the doubly geometric series:

$$
\begin{aligned}
(3.4)\quad l(\sigma, \theta, \vartheta \mid \xi) &= (2\pi\sigma^2)^{-(N+1)(M+1)/2} (1-\theta^2)^{(M+1)/2} (1-\vartheta^2)^{(N+1)/2} \\
&\times \exp\Bigg\{ -\frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{m=1}^{M} (\xi(n,m) - \theta\xi(n-1,m) \\
&\qquad - \vartheta\xi(n,m-1) + \theta\vartheta\xi(n-1,m-1))^2 \Bigg\} \\
&\times \exp\left\{ -\frac{1-\vartheta^2}{2\sigma^2} \sum_{n=1}^{N} (\xi(n,0) - \theta\xi(n-1,0))^2 \right\} \\
&\times \exp\left\{ -\frac{1-\theta^2}{2\sigma^2} \sum_{m=1}^{M} (\xi(0,m) - \vartheta\xi(0,m-1))^2 \right\} \\
&\times \exp\left\{ -\frac{1}{2\sigma^2} (1-\theta^2)(1-\vartheta^2)\xi^2(0,0) \right\}.
\end{aligned}
$$

Properties of the estimators were studied e.g. by Basu and Reinsel (1994) and Ying (1993).

THEOREM 3.1. *Let us consider a stochastic process ξ that satisfies the spatial autoregression equation*

$$\xi(n) = \sum_{k\in\mathcal{K}} a_k(\theta)\xi(n-k) + \epsilon(n), \tag{3.5}$$

where $\mathcal{K} = \times_{\alpha=1}^{p}\{0,1,\ldots,n_\alpha\}$, ϵ is a Gaussian white noise with variance σ^2 and the vector of real-valued parameters $\{\theta_k : k \in \mathcal{L}\} \subset \Theta$, $\mathcal{L} \subset \times_{\alpha=1}^{q}\{0,1,\ldots,n_\alpha\}$, is unknown. Suppose that there exists a set $\mathcal{L}' \subset \mathcal{K}$ such that $\{a_k(\theta) : k \in \mathcal{L}'\}$ define a homeomorphism of Θ on itself. If the true parameter θ_0 is in Θ and the observations are sampled on a rectangular lattice then the conditional and unconditional ML estimators of θ_0 and σ_0^2 exist and are strongly consistent. If $(\hat\theta_N)_{N=1}^{\infty}$ and $(\hat\sigma_N^2)_{N=1}^{\infty}$ are corresponding sequences of consistent estimators then the random vectors $\sqrt{N}(\hat\theta_N - \theta_0)$ and $\sqrt{N}(\hat\sigma_N^2 - \sigma_0^2)$ are asymptotically independent with normal distribution specified by the relations (2.23) *and* (2.24).

PROOF. Maximization of the conditional Gaussian likelihood subject to θ in the relation (3.5) is equivalent to finding the minimum of the quadratic form

$$Q_N(a(\theta)) = \frac{1}{N}\sum_{n\in\mathcal{N}_0}\left(\xi(n) - \sum_{k\in\mathcal{K}} a_k(\theta)\xi(n-k)\right)^2. \tag{3.6}$$

i) If $p = d$, $a_k = \theta_k$ for all $k \in \mathcal{K}\setminus\{(0,\ldots,0)\}$ and $a_{(0,\ldots,0)} = 0$, then the form has a.s. a unique minimum and the sequence of these minimums converges to the true parameter according to the law of large numbers.

ii) Next we consider the general case. Let $\hat\alpha = \{\hat\alpha_k(N) : k \in \mathcal{K}\}$ be the unique minimum of Q considered as a function of a and $\hat\theta = \{\hat\theta_k(N) : k \in \mathcal{L}\}$ be the ML of θ. Both are considered for a fixed sample from ξ of size N. According to our assumption the values $\{\hat\alpha_k(N) : k \in \mathcal{L}'\}$ determine a unique element $\tilde\theta(N)$ such that $a_k(\tilde\theta(N)) = \hat{\alpha_k}(N)$ for all $k \in \mathcal{L}'$. The values θ depend continuously on α. Hence, according to i) we have $\lim_{N\to\infty}\tilde\theta(N) = \theta_0$ a.s. and by the law of large numbers

$$\lim_{N\to\infty} Q_N(a(\tilde\theta(N))) = Q(a(\theta_0)) \tag{3.7}$$

a.s., where Q arises from Q_N by fixing a and passing with N to infinity. The last relation and the inequalities

$$Q_N(a(\tilde\theta(N))) \geq Q_N(a(\hat\theta(N))) \geq Q_N(\hat\alpha(N)) \tag{3.8}$$

have in consequence that

$$\lim_{N\to\infty} Q_N(a(\hat\theta(N))) = Q(a(\theta_0)). \tag{3.9}$$

The function $Q_N(a)$ is conclave in each component a_k, $k \in \mathcal{K}$. Therefore, relation (3.7) has in consequence the inequalities $a_k(\tilde\theta(N)) \geq a_k(\hat\theta(N)) \geq a_k(\theta_0(N))$. Using the relation $\lim_{N\to\infty}\tilde{\theta}(N)=\theta_0$ a.s. and continuity of $a_k(\theta)$, $k\in\mathcal{L}'$, with respect to θ we can easily prove the result for the conditional likelihood. Assumptions of the theorem guarantee that for large samples the difference between the conditional and unconditional maximum likelihood estimating equations may be neglected. Hence, proof of the first part of the theorem is complete. A similar argument allows to consider the second part of the theorem as a consequence of Theorem 2.2 and Remark 2.2.

Example 3.2. Niu (1995) investigated a stationary temporal-spatial process described by the equation

$$\sum_{k=-q}^{q}\theta_k\xi(t,n-k)+\sum_{l=1}^{p}\phi_l\xi(t-l,n)=\epsilon(t,n), \tag{3.10}$$

where θ_k and ϕ_l are parameters to be estimated and $\epsilon(t,n)$ are $N(0,1)$ i.i.d. r.v.'s. Compared to Niu we chose the parameter θ_q equal to one, not θ_0. This has little effect on application of the model as long as we do not want to draw inferences about θ_q. The choice $\theta_q=1$ provides a noticeable simplification in resulting estimating equations for θ. Niu considers the variance of ϵ time dependent but with a twelve month period which substantially simplifies the problem. We will suppose the variance σ^2 constant over time. The aim is to obtain the conditional maximum-likelihood estimate of the unknown parameters assuming a sufficient number of observations over a rectangular lattice of points is available. The characteristic polynomial of the model (3.10) is

$$P_{(\theta,\phi)}(\lambda,\omega)=\sum_{k=-q}^{q}\theta_k e^{-ik\omega}+\sum_{l=1}^{p}\phi_l e^{-il\lambda}. \tag{3.11}$$

In order to determine the Jacobian of the conditional distribution we assume the function $\sum_{k=0}^{2q}\theta_{q-k}e^{-ik\omega}$ has no zeros for $\omega\in[-\pi,\pi]$ and $P_{(\theta,\phi)}(\lambda,\omega)/\sum_{k=0}^{2q}\theta_{q-k}e^{-ik\omega}$ as a polynomial in $e^{-i\lambda}$ has all roots in the unit circle for every $\omega\in[-\pi,\pi]$. We take any $k=-q,\dots,q-1$ and calculate (see (2.11) and (2.12))

$$\begin{aligned}
&\partial_{\theta_k}\frac{\sigma^2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\ln|P_{(\theta,\phi)}(\lambda,\omega)|^2 d\lambda d\omega \\
&\quad=\frac{\sigma^2}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\left(\frac{e^{-ik\omega}}{P_{(\theta,\phi)}(\lambda,\omega)}+\frac{e^{ik\omega}}{P^*_{(\theta,\phi)}(\lambda,\omega)}\right)d\lambda d\omega \\
&\quad=\frac{\sigma^2}{2\pi}\int_{-\pi}^{\pi}\left(\frac{e^{-i(q-k)\omega}}{\sum_{l=0}^{2q}\theta_{q-l}e^{-il\omega}}+\frac{e^{i(q-k)\omega}}{\sum_{l=0}^{2q}\theta_{q-l}e^{il\omega}}\right)d\omega=0.
\end{aligned} \tag{3.12}$$

This and a similar consideration for ϕ instead of θ shows that the Jacobian is a constant independent of both θ and ϕ. Consequently, if we observe the process on the rectangle $\mathcal{N} = \{(t,n) : t = 0,\dots,T, n = 0,\dots,N\}$ then the conditional *ML* estimators of θ and ϕ are computed from the linear equations

$$\sum_{k=-q}^{q} \theta_k \sum_{t=1}^{T} \sum_{n=q-1}^{N-q} \xi(t,n-k)\xi(t,n-k') + \sum_{l=1}^{p} \phi_l \sum_{t=1}^{T} \sum_{n=q-1}^{N-q} \xi(t-l,n)\xi(t,n-k') = 0,$$

$$\sum_{k=-q}^{q} \theta_k \sum_{t=1}^{T} \sum_{n=q-1}^{N-q} \xi(t,n-k)\xi(t-l',n) + \sum_{l=1}^{p} \phi_l \sum_{t=1}^{T} \sum_{n=q-1}^{N-q} \xi(t-l,n)\xi(t-l',n) = 0,$$

$k' = -q,\dots,-1,1,\dots,q$ and $l' = 1,\dots,T$. Niu studied the asymptotic properties of the estimators for T increasing to infinity. Theorem 2.2 is thus not applicable. But it is easy to see that after appropriate scaling and passing with T to infinity we obtain

$$\begin{aligned} &\sum_{k=-q}^{q} \lim_{T\to\infty} \sqrt{T}(\theta_k - \hat\theta_{k.T})R(0,k-k') + \sum_{l=1}^{p} \lim_{T\to\infty} \sqrt{T}(\phi_l - \hat\phi_{l,T})R(l,k') = \eta(k'), \\ &\sum_{k=-q}^{q} \lim_{T\to\infty} \sqrt{T}(\theta_k - \hat\theta_{k.T})R(l',k) + \sum_{l=1}^{p} \lim_{T\to\infty} \sqrt{T}(\phi_l - \hat\phi_{l,T})R(l-l',0) = \eta(l'), \end{aligned} \tag{3.13}$$

where η is formed by zero mean normally distributed random variables and the convergence is in distribution. If we set $\varphi = (\theta,\phi)'$ and $\hat\varphi_T = (\hat\theta_T,\hat\phi_T)'$ then the *ML* estimators satisfy equations $\Gamma\hat\varphi_T = 0$, where Γ is made up by values of the covariance function R, c.f. equations (3.13). Analogy with the ordinary autoregression and unilateral processes suggests

$$\lim_{T\to\infty} \sqrt{T}(\varphi - \hat\varphi_T) = \mathcal{N}(0,\Gamma^{-1}). \tag{3.14}$$

We leave the reader to verify if this relation is true. Arguments by Niu (1995) show that the unconditional *ML* estimators have the same asymptotic distribution as the conditional ones. The paper also includes further details on the *ML* estimation procedure.

4. Supplementary results

A wide sense stationary process $(\xi(n))_{n\in Z^d}$ is called ergodic if it satisfies a version of the law of large numbers:

$$\lim_{n\to\infty}\frac{1}{n_1\cdots n_d}\sum_{k_1=1}^{n_1}\cdots\sum_{k_d=1}^{n_d}\xi(k_1,\ldots,k_d)=E\xi(0,\ldots,0) \tag{4.1}$$

in the mean or almost surely. The symbol $n\to\infty$ means that all components of $n=(n_1,\ldots,n_d)$ tend to infinity.

THEOREM 4.1. *If $(\xi(n))_{n\in Z^d}$ is a strictly stationary sequence then there is a random variable ξ such that*

$$\lim_{n\to\infty}\frac{1}{n_1\cdots n_d}\sum_{k_1=1}^{n_1}\cdots\sum_{k_d=1}^{n_d}\xi(k_1,\ldots,k_d)=\xi \tag{4.2}$$

in the mean and almost surely.

PROOF. The terminology used throughout the proof is introduced e.g. in Krengel (1985). Denote $C(\boldsymbol{Z}^d)$ the space of all real-valued functions upon $\boldsymbol{Z}^d$ and define on $C(\boldsymbol{Z}^d)$ a set of bijections τ_k, $k=1,\ldots,d$. The mapping τ_k is given by the relation $\tau_k(x)(n_1,\ldots,n_k,\ldots,n_d)=x(n_1,\ldots,n_k+1,\ldots,n_d)$ for each $x\in C(\boldsymbol{Z}^d)$ and $n=(n_1,\ldots,n_d)\subset\boldsymbol{Z}^d$. Briefly, τ_k is the shift on the k-th coordinate. If we denote the product σ-algebra of $C(\boldsymbol{Z}^d)$ by $\mathcal{B}$ then the strictly stationary sequence $(\xi(n))_{n\in Z^d}$ induces on $\mathcal{B}$ a probability measure ν, τ_k are endomorphisms under ν and the mappings $T_kf(x)=f\circ\tau_k(x)$ operating from $L_1(C(\boldsymbol{Z}^d),\mathcal{B},\nu)$ into $L_1(C(\boldsymbol{Z}^d),\mathcal{B},\nu)$ are positive contractions. The assertion is therefore a consequence of Theorem 1.2 in Chapter 6 by Krengel (1985).

THEOREM 4.2. (Law of Large Numbers) *Let $(\xi(n))_{n\in Z^d}$ be a strictly stationary process with finite second moments. Suppose the spectral measure of the process is absolutely continuous with respect to the Lebesgue measure. Then the process is ergodic.*

PROOF. In consequence of the previous theorem it remains to prove that $\xi=E\xi(0,\ldots,0)$ almost surely. We can and will assume $E\xi(0,\ldots,0)=0$. Set $\chi_0(\lambda)=1$ if $\lambda=0$ and $\chi_0(\lambda)=0$ otherwise. Representation (2.4), (4.2) and the relation

$$\lim_{n\to\infty}\frac{1}{n_1\cdots n_d}\sum_{k_1=1}^{n_1}\cdots\sum_{k_d=1}^{n_d}e^{i(n_1\lambda_1+\cdots+n_d\lambda_d)}=\chi_0(\lambda) \tag{4.3}$$

for every $\lambda\in(-\pi,\pi]^d$ have in consequence that $\xi=\int_{\{0\}}Z(d\lambda)=Z(\{0\})$ almost surely. With regard to (4.2)

$$\lim_{n\to\infty}\frac{1}{n_1\cdots n_d}\sum_{k_1=1}^{n_1}\cdots\sum_{k_d=1}^{n_d}R(k_1,\ldots,k_d)=\int_{\{0\}}dF(\lambda), \tag{4.4}$$

where F is the spectral function of the process. Due to the absolute continuity of the spectral function, the left side of (4.4) is zero. Hence, $E\xi^2 = E|Z(\{0\})|^2 = \int_{\{0\}} dF(\lambda)d\lambda = 0$. This may happen only if ξ is zero almost surely.

Consider on $\boldsymbol{Z}^d$ a metric ρ which assigns to every pair of points in $\boldsymbol{Z}^d$ their Euclidean distance. Using this metric we can define the distance between any two subsets of $\boldsymbol{Z}^d$. We denote this distance by ρ again. A strictly stationary random field $(\xi(n))_{n\in\boldsymbol{Z}^d}$ is called α-mixing (with function α) or strongly mixing if there is a function α of ρ such that for every pair of disjoint sets $S \subset \boldsymbol{Z}^d$ and $S' \subset \boldsymbol{Z}^d$

$$\sup\{|P(AB) - P(A)P(B)| : A \in \mathcal{A}, B \in \mathcal{B}\} \leq \alpha(\rho(S, S')) \tag{4.5}$$

and $\alpha(\rho)$ tends to zero as ρ goes to infinity. The σ-algebras $\mathcal{A}$ and $\mathcal{B}$ are generated by the random variables $\{\xi(n) : n \in S\}$ and $\{\xi(n) : n \in S'\}$, respectively. The α-mixing property thus means that correlation between the random variables in the field decreases with increasing distance between them.

PROPOSITION 4.1. *If a Gaussian stationary random field has a spectral density function which is positive and continuous on $(-\pi, \pi]^d$, then this field is strongly mixing.*

PROOF. See Rosenblatt (1985).

In order to formulate the main assertion of this section we accept the notation $\sum_{n=0}^{N}$ instead of $\sum_{n_1=0}^{N_1} \cdots \sum_{n_d=0}^{N_d}$ and set $N = (N_1 \cdots N_d)$.

THEOREM 4.3. *Let $(\xi(n))_{n\in\boldsymbol{Z}^d}$ be a strictly stationary strongly mixing random field with $E\xi(n) = 0$ and finite second moments. Denote $s_N = (E|\sum_{n=1}^{N} \xi(n)|^2)^{1/2}$. If the normed sums $|\sum_{n=1}^{N} \xi(n)|^2/s_N^2$ are uniformly integrable and $lim_{N\to\infty} s_N = \infty$, then*

$$\lim_{N\to\infty} \frac{1}{s_N} \sum_{n=1}^{N} \xi(n) = \mathcal{N}(0, 1) \tag{4.6}$$

in distribution.

PROOF. It can be obtained by minor changes from the proof of Theorem 2 in Volný (1986).

COROLLARY 4.1. *Let $(\xi(n))_{n\in\boldsymbol{Z}^d}$ be a Gaussian stationary zero mean random field which has a continuous spectral density f positive on $(-\pi, \pi]^d$. Define for a pair of arbitrarily chosen but fixed indices $k, l \in \boldsymbol{Z}^d$ the random variables $\eta(n) = \xi(n + k)\xi(n + l) - E\xi(n + k)\xi(n + l)$. Then in distribution*

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \eta(n) = \sqrt{2}(2\pi)^d \int e^{i\langle k-l,\lambda\rangle} f(\lambda) Z(d\lambda), \tag{4.7}$$

where Z is a Gaussian orthogonal measure with spectral density $g(\lambda) = (2\pi)^{-d}$ on $(-\pi, \pi]^d$.

PROOF. By Proposition 4.1 the field $(\xi(n))_{n\in Z^d}$ is α-mixing, thus $\eta(n)$ form again an α-mixing sequence. Due to the Gaussianity, the process $(\xi(n))_{n\in Z^d}$ has finite mixed moments of all orders. This draws with it uniform integrability of the sums $\sum_{n=1}^{N} \eta(n)/(E|\sum_{n=1}^{N} \eta(n)|^2)^{-1/2}$. The ratio has therefore an asymptotic normal distribution given by Theorem 4.3.

Let us investigate in more detail the expression $E|\sum_{n=1}^{N} \eta(n)|^2$. Using the Gauss property and stationarity we obtain by a direct computation:

$$
\begin{aligned}
(4.8)\qquad E\left|\sum_{n=1}^{N}\eta(n)\right|^2 &= \sum_{n=1}^{N}\sum_{m=1}^{N} E\eta(n)\eta(m)\\
&= \sum_{n=1}^{N}\sum_{m=1}^{N}\{E\xi(n+k)\xi(n+l)\xi(m+k)\xi(m+l)\\
&\qquad\qquad - E\xi(n+k)\xi(n+l)E\xi(m+k)\xi(m+l)\}\\
&= \sum_{n=1}^{N}\sum_{m=1}^{N}\{R^2(k-l)+R^2(n-m)\\
&\qquad\qquad + R(n+k-m-l)R(n+l-m-k)\\
&\qquad\qquad - R^2(k-l)\}\\
&= \sum_{n\in B_N}(N-|n|)R^2(n)\\
&\quad + \sum_{n\in B_N}(N-|n|)R(n+l-k)R(n+k-l),
\end{aligned}
$$

where $B_N = \{n : n = (n_1, \ldots, n_d), |n_k| < N_k, k = 1, \ldots, d\} \subset Z^d$ is a cube with sides of length $2N_k$, $k = 1, \ldots, d$, and $(N - |n|) = (N_1 - |n_1|)\cdots(N_d - |n_d|)$. The functions $((2\pi)^{-d/2}e^{-i\langle n,\lambda\rangle})_{n\in Z^d}$ form a complete orthonormal system in $L_2(-\pi,\pi]^d$. Thus using the Parseval equality and the relation $f(\lambda) = (2\pi)^{-d}\sum_{t\in Z^d} e^{-i\langle t,\lambda\rangle}R(t)$ we have

$$
(4.9)\qquad
\begin{aligned}
&\sum_{n\in Z^d} R^2(n) = (2\pi)^d\int f^2(\lambda)d\lambda,\\
&\sum_{n\in Z^d} R(n+l-k)R(n+k-l) = (2\pi)^d\int f^2(\lambda)e^{2i\langle l-k,\lambda\rangle}d\lambda
\end{aligned}
$$

and the sum of integrals in (4.9) equals to $2(2\pi)^d\int f^2(\lambda)\cos^2(\langle l-k,\lambda\rangle)d\lambda$. If we divide both sides of (4.8) by N and take the limit, the term $(N-|n|)/N$ converges to one. Since $(N-|n|)/N \le 1$, the expectation in (4.8) divided by N is dominated by $2(2\pi)^d\int f^2(\lambda)\cos^2(\langle l-k,\lambda\rangle)d\lambda$. We can thus apply the dominated convergence theorem and show that

$$
(4.10)\qquad \lim_{N\to\infty}\frac{1}{N}E\left|\sum_{n=1}^{N}\eta(n)\right|^2 = 2(2\pi)^d\int f^2(\lambda)\cos^2(\langle l-k,\lambda\rangle)d\lambda.
$$

From the previous considerations it follows that the left side of (4.7) has asymptotic normal distribution with zero mean and variance $2(2\pi)^d \int f^2(\lambda) \cdot \cos^2(\langle l-k,\lambda\rangle)d\lambda$. In order to verify that let us write

$$(2\pi)^d \int e^{i\langle k-l,\lambda\rangle} f(\lambda)Z(d\lambda) = (2\pi)^d \int \cos(\langle k-l,\lambda\rangle)f(\lambda)Z(d\lambda) + i(2\pi)^d \int \sin(\langle k-l,\lambda\rangle)f(\lambda)Z(d\lambda). \tag{4.11}$$

The random variable on the left is real-valued, thus the imaginary part in the right must be zero almost surely. Variance of the real part equals to the right side of (4.10). The normal random variables in (4.7) have thus the same first and second moments and determine therefore the same normal distribution.

PROOF OF THEOREM 2.2. Define, for the given sets $\mathcal{N}_0 \subset \mathcal{N}$ and observations $(\xi(n))_{n\in\mathcal{N}}$, the functions

$$f_n(\theta,\xi) = \frac{v_N^{1/N}}{(2\pi\sigma^2)^{1/2}} \exp\left\{\frac{1}{2(2\pi)^d}\int \ln|P_\theta(\lambda)|^2 d\lambda - \frac{1}{2\sigma^2}(\mathcal{P}_\theta\xi(n))^2\right\}, \tag{4.12}$$

where $n \in \mathcal{N}_0$. The function $f_n(\theta,\xi)$ is the conditional likelihood of $\xi(n)$ conditioned on the neighboring elements entering $\mathcal{P}_\theta\xi(n)$. The ML estimate minimizes the negative log-likelihood $L_N(\theta,\xi) = -\sum_{n\in\mathcal{N}_0} \ln f_n(\theta,\xi)$. This means it satisfies the equation $\partial_\theta L_N(\hat\theta_N,\xi) = 0$, where $\partial_\theta L_N = (\partial_{\theta_1}L_N,\ldots,\partial_{\theta_r}L_N)^T$. According to the multivariate Taylor theorem there is a point θ^* on the line between $\hat\theta$ and θ_0 such that

$$\partial_\theta L_N(\hat\theta_N,\xi) = \partial_\theta L_N(\theta_0,\xi) + \partial_\theta^2 L_N(\theta_0,\xi)(\hat\theta_N-\theta_0)^T + Z(\theta_N^*,\xi,\theta_0). \tag{4.13}$$

In our notation $\partial_\theta^2 L_N(\theta_0,\xi)$ is the Jacobi matrix of the vector $\partial_\theta L_N(\theta,\xi)$ at $\theta = \theta_0$. The last expression in (4.13), $Z(\theta_N^*,\xi,\theta_0)$, is a column vector with components made up of sums of mixed third order partial derivatives $\partial_{\theta_u\theta_v\theta_j}L_N(\theta,\xi)$ at $\theta=\theta^*$ multiplied by $(\hat\theta_{u,N}-\theta_{0,u})(\hat\theta_{v,N}-\theta_{0,v})$, u, $v = 1,\ldots,r$. The left side of (4.13) equals to zero by the definition of the ML estimator. If we denote by $a_k(\theta)$ the coefficients of $P_\theta(\lambda)$, then the components of $\partial_\theta L_N(\theta,\xi)$ are

$$\begin{aligned} & -\sum_{n\in\mathcal{N}_0} \partial_{\theta_j} \ln f_n(\theta,\xi) \\ &= \frac{1}{2\sigma^2}\sum_{n\in\mathcal{N}_0}\{\partial_{\theta_j}(\mathcal{P}_\theta\xi(n))^2 - E_{\theta_0}\partial_{\theta_j}(\mathcal{P}_\theta\xi(n))^2\} \\ &= \frac{1}{\sigma^2}\sum_{n\in\mathcal{N}_0}\{\mathcal{P}_\theta\xi(n)\partial_{\theta_j}\mathcal{P}_\theta\xi(n) - E_{\theta_0}\mathcal{P}_\theta\xi(n)\partial_{\theta_j}\mathcal{P}_\theta\xi(n)\} \\ &= \frac{1}{\sigma^2}\sum_{k\in Z^d}\sum_{l\in Z^d} a_k(\theta)a_l^*(\theta)\partial_{\theta_j}(a_k(\theta)a_l^*(\theta)) \\ &\quad \cdot \sum_{n\in\mathcal{N}_0}\{\xi(k+n)\xi(l+n) - E_{\theta_0}\xi(k+n)\xi(l+n)\}. \end{aligned} \tag{4.14}$$

We recall that all but finitely many of the a_k are zero. If $\theta = \theta_0$, then by the strong law of large numbers, the last sum divided by N has limit zero a.s. for N growing to infinity. Thus

$$\lim_{N\to\infty} \frac{1}{N}\partial_\theta L_N(\theta_0,\xi) = 0 \quad \text{a.s.} \tag{4.15}$$

If we divide the sum on the left of (4.14) by $\sqrt{N}$ rather than by N, then we can apply Corollary 4.1 and obtain that, in distribution,

$$\begin{aligned} -\lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{n\in\mathcal{N}_0} \partial_{\theta_j} \ln f_n(\theta,\xi) \\ &= \sqrt{2}\sum_{k\in Z^d}\sum_{l\in Z^d} a_k(\theta)a_l^*(\theta)\partial_{\theta_j}(a_k(\theta)a_l^*(\theta)) \int \frac{e^{i\langle k-l,\lambda\rangle}}{|P_\theta(\lambda)|^2} Z(d\lambda) \\ &= \frac{1}{\sqrt{2}}\int \frac{\partial_{\theta_j}|P_\theta(\lambda)|^2}{|P_\theta(\lambda)|^2} Z(d\lambda) = \frac{1}{\sqrt{2}}\int \partial_{\theta_j} \ln|P_\theta(\lambda)|^2 Z(d\lambda), \end{aligned} \tag{4.16}$$

where Z is a Gaussian white noise measure with spectral function $g(\lambda) = (2\pi)^{-d}$ on $[-\pi,\pi]^d$. Therefore,

$$\lim_{N\to\infty} \frac{1}{\sqrt{N}}\partial_{\theta_j} L_N(\theta,\xi) = \mathcal{N}(0, Q_{j,j}(\theta)). \tag{4.17}$$

This proves that the vector $\partial_\theta L_N(\theta,\xi)$ has asymptotically normal distribution with a covariance matrix whose diagonal elements equal to $Q_{j,j}(\theta)$, $j = 1,\dots,r$. Form of the off diagonal elements may be found accordingly. The covariance matrix Q is thus described by (2.22). Next we examine the limit properties of $\partial^2_\theta L_N(\theta_0,\xi)$. By differentiation of $\partial_{\theta_j} L_N(\theta,\xi)$ with respect to θ_k we obtain

$$\partial_{\theta_k\theta_j} L_N(\theta,\xi) = \sum_{n\in\mathcal{N}_0} \frac{\partial_{\theta_j} f_n(\theta,\xi)\partial_{\theta_k} f_n(\theta,\xi) - f_n(\theta,\xi)\partial_{\theta_j\theta_k} f_n(\theta,\xi)}{f_n^2(\theta,\xi)}. \tag{4.18}$$

According to the strong law of large numbers

$$\begin{aligned} \lim_{N\to\infty} \frac{1}{N}\partial_{\theta_k\theta_j} L_N(\theta_0,\xi) \\ &= E_{\theta_0}\partial_{\theta_j} \ln f_n(\theta_0,\xi)\partial_{\theta_k} \ln f_n(\theta_0,\xi) - E_{\theta_0}\frac{\partial_{\theta_j\theta_k} f_n(\theta_0,\xi)}{f_n(\theta_0,\xi)}. \end{aligned} \tag{4.19}$$

As we mentioned, f_n may be viewed as a conditional likelihood function of $\xi(n)$, conditioned on the neighboring elements that form a vector $\xi_{0,b}$. With the obvious notation we obtain

$$E_{\theta_0}\frac{f_n(\theta,\xi(n)\mid\xi_{0,b})}{f_n(\theta_0,\xi(n)\mid\xi_{0,b})} = E_{\theta_0}\left[E_{\theta_0}\left[\frac{f_n(\theta,\xi(n)\mid\xi_{0,b})}{f_n(\theta_0,\xi(n)\mid\xi_{0,b})}\,\middle|\,\xi_{0,b}\right]\right] = E_{\theta_0}1 = 1. \tag{4.20}$$

Taking derivatives with respect to θ_j and θ_k on both sides we see that

$$E_{\theta_0} \frac{\partial_{\theta_j \theta_k} f_n(\theta, \xi)}{f_n(\theta_0, \xi)} = \partial_{\theta_j \theta_k} E_{\theta_0} \frac{f_n(\theta, \xi)}{f_n(\theta_0, \xi)} = 0 \tag{4.21}$$

for every $\theta \in \Theta$. In particular if $\theta = \theta_0$ then

$$\begin{aligned}
(4.22)\quad & E_{\theta_0} \partial_{\theta_j} \ln f_n(\theta_0, \xi) \partial_{\theta_k} \ln f_n(\theta_0, \xi) \\
&= \frac{1}{4\sigma^4} E_{\theta_0} \{\partial_{\theta_j} \epsilon_\theta^2(n) - E_{\theta_0} \partial_{\theta_j} \epsilon_\theta^2(n)\}\{\partial_{\theta_k} \epsilon_\theta^2(n) - E_{\theta_0} \partial_{\theta_k} \epsilon_\theta^2(n)\} \\
&= \frac{1}{\sigma^4} \{E_{\theta_0} \epsilon_\theta(n) \partial_{\theta_j} \epsilon_\theta(n) \epsilon_\theta(n) \partial_{\theta_k} \epsilon_\theta(n) \\
&\qquad - E_{\theta_0} \epsilon_\theta(n) \partial_{\theta_j} \epsilon_\theta(n) E_{\theta_0} \epsilon_\theta(n) \partial_{\theta_k} \epsilon_\theta(n)\} \\
&= \frac{1}{\sigma^4} (\sigma^2 E_{\theta_0} \partial_{\theta_j} \epsilon_\theta(n) \partial_{\theta_k} \epsilon_\theta(n) \\
&\qquad + E_{\theta_0} \epsilon_\theta(n) \partial_{\theta_j} \epsilon_\theta(n) E_{\theta_0} \epsilon_\theta(n) \partial_{\theta_k} \epsilon_\theta(n)) \\
&= (2\pi)^{-d} \left\{ \int \frac{\partial_{\theta_j} P_\theta(\lambda) \partial_{\theta_k} P_\theta^*(\lambda)}{|P_\theta(\lambda)|^2} d\lambda \right. \\
&\qquad \left. + \frac{1}{4}(2\pi)^{-d} \int \partial_{\theta_j} \ln |P_\theta(\lambda)|^2 d\lambda \int \partial_{\theta_k} \ln |P_\theta(\lambda)|^2 d\lambda \right\}.
\end{aligned}$$

This is the component of the matrix W defined in (2.22). We used the relation $E\eta_1\eta_2\eta_3\eta_4 = E\eta_1\eta_2 E\eta_3\eta_4 + E\eta_1\eta_3 E\eta_2\eta_4 + E\eta_1\eta_4 E\eta_2\eta_3$ valid for any quadruple $\eta_1, \eta_2, \eta_3, \eta_4$ of normal random variables. The second summand in the last line of (4.22) comes from the relation (2.12). The expectation $E_{\theta_0} \partial_{\theta_j} \epsilon_\theta(n) \partial_{\theta_k} \epsilon_\theta(n)$ may be evaluated in a similar way.

Finally we examine the limit of the last term in (4.13). According to (4.14)

$$\begin{aligned}
(4.23)\quad \partial_{\theta_u \theta_v \theta_j} L_N(\theta, \xi) = \frac{1}{\sigma^2} \sum_{k \in \mathbf{Z}^d} \sum_{l \in \mathbf{Z}^d} \partial_{\theta_u \theta_v} \Bigg[& a_k(\theta) a_l^*(\theta) \partial_{\theta_j} (a_k(\theta) a_l^*(\theta)) \\
& \times \sum_{n \in \mathcal{N}_0} \{\xi(k+n)\xi(l+n) \\
& - E_\theta \xi(k+n)\xi(l+n)\} \Bigg]
\end{aligned}$$

considered for $\theta = \theta_0$. Note that $k+n$ and $l+n$ belong to $\mathcal{N}$ whenever the coefficients $a_k(\theta)$ and $a_l(\theta)$ are non-zero. The expected value of (4.23) equals to $R_\theta(k-l) = \sigma^2 (2\pi)^d \int |P_\theta(\lambda)|^{-2} e^{i\langle k-l, \lambda\rangle} d\lambda$. The coefficients of $P_\theta(\lambda)$ are assumed three times differentiable and therefore $R_\theta(k-l)$ is three times differentiable as well. By the strong law of large numbers

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n \in \mathcal{N}_0} \xi(k+n)\xi(l+n) = R_\theta(k-l). \tag{4.24}$$

The coefficients of $P_\theta(\lambda)$ and their mixed partial derivatives up to third order are bounded on Θ. Hence, the last term in (4.13) converges to a finite constant for every θ and it is bounded. If $(\hat\theta_N)_{N=1}^\infty$ is a sequence of ML estimators that converges to θ_0 in probability then the third term in (4.13) converges to zero in probability as well with N growing to infinity. The relation (2.23) is now a consequence of (4.13), (4.17) and (4.22).

Next we compute the limit (2.24). For convenience we replace in our notation $\hat\sigma^2$ with the averaged sum of squared residuals $s_N^2(\hat\theta_N)$. Using continuity of the coefficients $a_k(\theta)$ and the strong law of large numbers it is easy to verify that if $(\hat\theta_N)_{N=1}^\infty$ is a consistent sequence of estimators then $(s_N^2(\hat\theta_N))_{N=1}^\infty$, where N is the number of elements in $\mathcal{N}_0$, is also a consistent sequence of estimators. That follows from the relations

$$(4.25)\qquad \begin{aligned} s_N^2(\hat\theta_N) &- s_N^2(\theta_0) - 2\sum_{k\in\mathcal{K}}(a_k(\hat\theta_N) - a_k(\theta_0))\frac{1}{N}\sum_{n\in\mathcal{N}_0}\epsilon(n)\xi(n-k) \\ &+ \frac{1}{N}\sum_{n\in\mathcal{N}_0}\left(\sum_{k\in\mathcal{K}}(a_k(\hat\theta_N) - a_k(\theta_0))\xi(n-k)\right)^2 \\ &= s_N^2(\theta_0) + O\left(\frac{1}{N}\right). \end{aligned}$$

Relation (4.21) has in consequence asymptotic unbiasedness of the estimator $s_N^2(\hat\theta_N)$. In order to determine the limit variance we can calculate

$$(4.26)\qquad \begin{aligned} E(s_N^2(\theta_0) - \sigma^2)^2 &= \frac{1}{N}\sum_{n\in\mathcal{N}_0}\sum_{m\in\mathcal{N}_0} E(\epsilon(n)-\sigma)(\epsilon(n)+\sigma) \\ &\quad \times (\epsilon(m)-\sigma)(\epsilon(m)+\sigma) \\ &= \frac{1}{N}\sum_{n\in\mathcal{N}_0}\sum_{m\in\mathcal{N}_0} \{E(\epsilon^2(n)-\sigma^2)(\epsilon^2(m)-\sigma^2) \\ &\qquad + E(\epsilon(n)-\sigma)(\epsilon(m)-\sigma) \\ &\qquad \times E(\epsilon(n)-\sigma)(\epsilon(m)-\sigma) \\ &\qquad + E(\epsilon(n)-\sigma)(\epsilon(m)+\sigma) \\ &\qquad \times E(\epsilon(n)+\sigma)(\epsilon(m)-\sigma)\} = 2\sigma^4. \end{aligned}$$

This is a direct consequence of normality and independence of the noise components $\epsilon(n)$. Rest of the proof follows from Theorem 4.3.

References

Basu, S. and Reinsel, G. C. (1994). Regression models with spatially correlated errors, *J. Amer. Statist. Assoc.*, **89**, 88–99.

Cramér, H. (1940). On the theory of stationary random processes, *Ann. of Math.*, **41**, 215–230.

Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittles estimate, *Probab. Theory Related Fields*, **86**, 87–104.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation, *International Statistical Review*, **55**, 231–244.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice, *Biometrika*, **69**, 95–105.

Heyde, C. C. and Gay, R. (1989). On asymptotic quasi-likelihood, *Stochastic Process. Appl.*, **31**, 223–236.

Heyde, C. C. and Gay, R. (1993). Smoothed periodogram asymptotics and estimation for processes and fields with possible long-range dependence., *Stochastic Process. Appl.*, **45**, 169–182.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes, *Ann. Statist.*, **10**, 132–153.

Kolmogorov, A. N. (1941). Interpolation and extrapolation of stationary random series, *Bulletin of the Academy of Sciences of USSR*, **5**, 3–14 (in Russian).

Krengel, V. (1985). *Ergodic Theorems*, de Gruyter, Berlin.

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

Martin, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials, *Comm. Statist. Theory Methods*, **19**, 55–81.

Niu, X. (1995). Asymptotic properties of maximum likelihood estimates in a class of space-time regression models, *J. Multivariate Anal.*, **55**, 82–104.

Ord, K. (1975). Estimation methods for models of spatial interaction, *J. Amer. Statist. Assoc.*, **70**, 120–126.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*, Birkhäuser, Boston.

Tjøstheim, D. (1978). Statistical spatial series modeling, *Advances in Applied Probability*, **10**, 130–154.

Tjøstheim, D. (1983). Statistical spatial series modeling II. Some further results on unilateral lattice processes, *Advances in Applied Probability*, **15**, 562–584.

Volný, D. (1986). Approximation of stationary processes and the central limit problem, Probability Theory and Mathematical Statistics, 5th Japan-USSR Symposium Proceedings, *Lecture Notes Math.*, **1299**, Springer, Berlin.

Whittle, P. (1953). The analysis of multiple stationary time series, *J. Roy. Statist. Soc. Ser. B*, **15**, 125–139.

Whittle, P. (1954). On stationary processes in the plane, *Biometrika*, **41**, 434–439.

Ying, Z. (1993). Maximum likelihood of parameters under spatial sampling scheme, *Ann. Statist.*, **21** (3), 1567–1590.