Bernoulli **23**(4B), 2017, 2988–3020
DOI: 10.3150/16-BEJ835

Baxter's inequality and sieve bootstrap for random fields

MARCO MEYER[1,*], CARSTEN JENTSCH[2] and JENS-PETER KREISS[1,**]

[1]*Inst. f. Math. Stochastik, Technische Universität Braunschweig, Pockelsstrasse 14, D-38106 Braunschweig, Germany. E-mail:* [*]*marco.meyer@tu-bs.de;* [**]*j.kreiss@tu-bs.de*
[2]*Department of Economics, University of Mannheim, L7, 3–5, D-68131 Mannheim, Germany. E-mail: cjentsch@mail.uni-mannheim.de*

The concept of the autoregressive (AR) sieve bootstrap is investigated for the case of spatial processes in $\mathbb{Z}^2$. This procedure fits AR models of increasing order to the given data and, via resampling of the residuals, generates bootstrap replicates of the sample. The paper explores the range of validity of this resampling procedure and provides a general check criterion which allows to decide whether the AR sieve bootstrap asymptotically works for a specific statistic of interest or not. The criterion may be applied to a large class of stationary spatial processes. As another major contribution of this paper, a weighted Baxter-inequality for spatial processes is provided. This result yields a rate of convergence for the finite predictor coefficients, i.e. the coefficients of finite-order AR model fits, towards the autoregressive coefficients which are inherent to the underlying process under mild conditions. The developed check criterion is applied to some particularly interesting statistics like sample autocorrelations and standardized sample variograms. A simulation study shows that the procedure performs very well compared to normal approximations as well as block bootstrap methods in finite samples.

Keywords: autoregression; bootstrap; random fields

1. Introduction

We consider stationary real-valued spatial processes in the plane $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with zero mean and finite second moments. By imposing only very mild regularity conditions on the processes the framework of this paper remains very general. Particularly, without making any parametric/linearity assumptions on the process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, we are interested in fitting spatial autoregressive models of the form

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p) X_{\underline{t}-\underline{k}} + e_{\underline{t}} \tag{1.1}$$

to data, where $\Theta(p)$ denotes some suitable finite index set and $(e_{\underline{t}})$ is some white noise process. In few words, this paper has two main purposes: First, we will show that models of the form (1.1) are well-suited to describe the behaviour of very general stationary spatial processes since a very large class of these processes possesses an inherent autoregressive structure. This structure can be approximated well by models such as (1.1), which will be shown by proving a generalization of Baxter's inequality, cf. [4], to the spatial setting. As a second major contribution of this paper,

1350-7265

the concept of the autoregressive sieve bootstrap scheme will be transferred to the case of spatial processes. In the following, the aforementioned purposes will be explained in more detail.

By classical results going back to the work of [40], general spatial processes in the plane $\mathbb{Z}^2$ always possess half-plane representations with respect to *each half-plane* of $\mathbb{Z}^2$ that might be chosen, as long as mild assumptions are fulfilled. More precisely, there exist one-sided autoregressive (AR) as well as moving-average (MA) representations

$$X_{\underline{t}} = \sum_{\underline{k} \in \Theta} a_{\underline{k}} X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \quad \text{and} \quad X_{\underline{t}} = \sum_{\underline{k} \in \Theta} b_{\underline{k}} \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}} \tag{1.2}$$

with respect to some (weak) white noise process $(\varepsilon_{\underline{t}})$, where Θ can be *any* half-plane in the sense of [17]. Throughout this paper, we stick to the so-called *lower half-plane* representation corresponding to lexicographical ordering of the plane $\mathbb{Z}^2$ as described by [21], among others. It is important to note that choosing this particular half-plane representation is *not* restrictive at all because any other choice of the half-plane would be fine as well. It has to be understood as a suitable vehicle to establish meaningful theory in this paper and the lower half-plane is just chosen for notational convenience. During the course of this paper, we will also clarify a common misunderstanding in the discussion of spatial and time series autoregressions, that should at least be mentioned briefly at this point: It is often criticized that, for spatial processes, one has to choose a concept of "past" values for one-sided autoregressions, that is, choose a direction from which the random variables $X_{\underline{t}}$ are influenced. This choice is of course arbitrary. Hence, one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true since our assumptions do not constrain the class of processes any further than demanding the spectral density to be positive and smooth.

In contrast to our framework, most of the existing literature on autoregressive modeling in the plane is heavily based on the assumption that the underlying spatial process actually fulfills some specific model structure. Autoregressive processes in the plane have been pioneered in [40], where unilateral and bilateral autoregressive models are studied. Correlation properties of these processes have been studied in [6] and for some special cases in [3]. Spatial autoregressive processes with a "quarter-plane past" form a popular sub-class of unilateral processes in the plane. These processes have been investigated in detail by [37,38] and [39]. However, although the class of spatial AR processes with a quarter-plane past appears to be appealing at first sight due to its simple structure, we still consider half-plane instead of quarter-plane representations in this paper. This is due to the fact that, under very mild assumptions, general spatial processes are *always* assured to have half-plane representations as in (1.2), which is in general not true for quarter-plane representations (at least not with uncorrelated innovations). Hence, imposing a quarter-plane past structure on the process $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ turns out to be very restrictive and is therefore omitted in this paper. The same is true for models considered in [12], who discuss the properties of models with several regions of support. Yule–Walker type estimation of spatial AR models has been investigated by [2,16] and [18], who particularly addressed an inaccuracy in [38].

The crucial property that spatial processes can always be represented as in (1.2) is also well-known for time series processes $(X_t)_{t \in \mathbb{Z}}$, cf. among others [35]. Here, the AR representation

corresponding to (1.2) reads

$$X_t = \sum_{k=1}^{\infty} a_k X_{t-k} + \varepsilon_t. \tag{1.3}$$

To deal with these infinite dimensional autoregressive representations in the time series case, the famous Baxter-inequality (cf. [4] for univariate processes and [20] or [11] for the multivariate case) plays a fundamental role and allows for meaningful asymptotic theory. When fitting AR models of finite order p to time series, for instance by Yule–Walker estimation, one typically estimates the so-called finite predictor coefficients $a_1(p), \ldots, a_p(p)$, which are simply the coefficients of the L^2-projection of X_t onto the finite past $\mathrm{sp}\{X_{t-1}, \ldots, X_{t-p}\}$. Here, and in the following, if A is an arbitrary subset of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $\mathrm{sp}(A)$ denotes the span of all vectors $a \in A$. Baxter's inequality provides a connection between these finite predictor coefficients and the AR coefficients from (1.3) and reads as follows: Under mild smoothness conditions on the spectral density of the process, there exist a constant $C < \infty$ and $p_0 \in \mathbb{N}$ such that

$$\sum_{k=1}^{p} \nu(k)\big|a_k(p) - a_k\big| \le C \cdot \sum_{k=p+1}^{\infty} \nu(k)|a_k|, \qquad \forall p \ge p_0. \tag{1.4}$$

Here, $\nu(\cdot)$ denotes a weight function which is connected to the smoothness condition on the spectral density. Notice that the right-hand side of (1.4) is finite and therefore converges to zero as $p \to \infty$. Hence, the left-hand side also vanishes for $p \to \infty$ which yields convergence for the predictors $a_k(p)$ towards the AR coefficients a_k. In fact, the weights $\nu(k)$ determine the rate of convergence. If this rate is fast enough, then even autoregressive fits of rather small order p are suitable to describe the process (X_t) properly. The goal in this paper is to derive a similar inequality for the AR fits of shape (1.1) in connection with representations (1.2).

The original proof of (1.4) for univariate time series is mainly based on the analytical result of [5]. One might think that the original proof of [5] transfers straightforwardly from time series to the spatial case, but this is not the case. Heuristically, this is due to the following observations. The proof of Baxter's inequality for time series is heavily based on the fact that by predicting X_t based on $\mathrm{sp}\{X_{t-1}, \ldots, X_{t-p}\}$, the two sets $\{s : s < t - p\}$ and $\{s : s \ge t\}$ can be separated arbitrarily far apart, for sufficiently large p, by the set $M(p) := \{s : t - p \le s \le t - 1\}$. Thus, $|\mathrm{Cov}(X_r, X_q)|$ becomes arbitrarily small for sufficiently large p and for $X_r \in \{X_s, s < t - p\}$ and $X_q \in \{X_s, s \ge t\}$. For the spatial case such a separation is no longer possible as no finite subset analogous to $M(p)$ exists that is capable to separate $\mathbb{Z}^2$ in this fashion. As one major contribution of this paper, we come up with a different approach to prove a version of Baxter's inequality that is suitable for spatial processes. This result allows to derive rigorous asymptotic theory for AR fits of increasing order for spatial processes.

For time series, Baxter's inequality is a key ingredient when establishing validity of the AR sieve bootstrap scheme. This procedure was introduced for stationary univariate linear time series by [22,23] and [9] who established validity for different statistics including autocovariances and autocorrelations. The main contribution of the AR sieve methodology is to allow the autoregressive order $p = p(n)$ to increase with the sample size n. Thus, the AR sieve bootstrap extends the model-based (parametric) AR bootstrap – first considered by [14] – to the much richer (nonparametric) class of AR(∞)-processes.

Paparoditis and Streitberg [33] established asymptotic validity of the AR sieve bootstrap to infer properties of high order autocorrelations, and [31] established its validity in a multivariate linear time series context. Furthermore, the AR sieve bootstrap is used for testing for unit roots in [10] and [32], and in econometrics literature for several purposes such as for example, forecasting in [1] or in the setup of time series panels in [36].

However, while all the aforementioned results were derived under the explicit assumption of an underlying AR(∞) process, [24] extended the range of applicability of the AR sieve significantly. Under very mild conditions and without having to assume any autoregressive structure of the underlying process, they were able to show that the AR sieve remains valid whenever the so-called companion process mimics the proper limiting distribution, which constitutes a simple and general check criterion. Recently, [28] extended the results of [24] to the multivariate case. To generalize their concept, as a second main contribution of this paper, we introduce a spatial AR sieve methodology in the spirit of [24] and provide rigorous theory.

The proposed AR sieve bootstrap performs favourably compared to block bootstrap techniques, as will be shown in a simulation study in this paper. Block bootstrap and subsampling for random fields were proposed by [19] and [25], whereas [34] addressed block resampling schemes for general statistics. Zhu and Lahiri [41] proved bootstrap consistency for the empirical process of a non-overlapping block bootstrap. Optimal block size and subsample size selection have been addressed in [30] and [29], respectively.

The remainder of this paper is organised as follows: In Section 2, we will introduce the basic notations and definitions and formulate the algorithm of the AR sieve bootstrap procedure precisely. In addition, we will show how the rate of decay of the autocovariances of a spatial process carries over to its cepstral coefficients – the Fourier coefficients of the logarithm of the spectral density – and then to the AR coefficients.

In Section 3, we will establish sufficiently fast convergence of the finite-order AR models that are fitted in the course of the sieve bootstrap procedure, to the aforementioned AR coefficients. Here, we will derive a generalisation of Baxter's inequality, cf. [4], to the case of random fields. Beyond its application in connection with the AR sieve bootstrap, this result may be of its own interest.

The conditions for AR sieve bootstrap validity are given in Section 4, and the result will be a check-criterion which allows to decide whether the procedure is asymptotically consistent or not; with the criterion being solely based on the asymptotics of the particular test statistic one is looking at. This result closely resembles the concept of the so-called companion process introduced by [24]. We will apply the derived check criterion in Section 5 to some particularly interesting statistics, including variogram estimators. It follows a simulation study in Section 6 which compares the performance of the AR sieve bootstrap to normal approximations and the block bootstrap. Section 7 contains the proofs of the two central theorems, Baxter's inequality and the result about bootstrap validity, while all other proofs of auxiliary results are deferred to the corresponding technical report [27].

2. Preliminaries

Consider a stationary real-valued spatial process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ with mean zero and finite second moments. In the following we will switch between the two equivalent notations $X_{\underline{t}} = X_{t_1,t_2}$. While

the vector index notation $X_{\underline{t}}$ allows for a more compact presentation of the results, the notation X_{t_1,t_2} is sometimes necessary if we want to describe operations on the components of the index vector. For convenience reasons, we will also sometimes use a mixed notation, for example, in expressions such as $\sum_{t_1\in A}\sum_{t_2\in B} X_{\underline{t}}$.

The autocovariance function of $(X_{\underline{t}})$ at lag $\underline{h} = (h_1, h_2)^T$ is denoted by $\gamma(\underline{h}) = E(X_{\underline{t}+\underline{h}}X_{\underline{t}})$. We assume to have a square-shaped data sample $\{X_{\underline{t}} : 1 \le t_1, t_2 \le n\}$ consisting of n^2 observations at hand. Define $\Pi := \{\underline{t} \in \mathbb{Z}^2 : 1 \le t_1, t_2 \le n\}$ and $\Pi_{\underline{h}} := \{\underline{t} \in \mathbb{Z}^2 : 1 \le t_1, t_2, t_1 + h_1, t_2 + h_2 \le n\}$; that is, $\Pi_{\underline{h}}$ describes the set of vectors $\underline{t} \in \mathbb{Z}^2$ such that both $\underline{t}$ and $\underline{t} + \underline{h}$ are elements of Π. The empirical autocovariance function can then be stated as

$$\widehat{\gamma}(\underline{h}) := \frac{1}{|\Pi_{\underline{h}}|}\sum_{\underline{t}\in\Pi_{\underline{h}}}(X_{\underline{t}+\underline{h}} - \overline{X})(X_{\underline{t}} - \overline{X}), \tag{2.1}$$

where $\overline{X} = n^{-2}\sum_{\underline{t}\in\Pi} X_{\underline{t}}$ denotes the sample mean.

We now turn to the algorithm of the autoregressive sieve bootstrap for random fields. Our proposal depends on fitting an autoregressive model of finite order $p \in \mathbb{N}$ to the data. Since it is not obvious how such an AR fit would look like in the spatial setting, we first define the following set of vectors in $\mathbb{Z}^2$ which characterises the collection of sites for the pth order AR fit:

$$\Theta(p) := \big\{\underline{k} \in \mathbb{Z}^2 : (1 \le k_1 \le p \text{ and } k_2 = 0) \text{ or } (-p \le k_1 \le p \text{ and } 1 \le k_2 \le p)\big\}. \tag{2.2}$$

An autoregressive model with sites given by $\Theta(p)$ could be stated as

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}} X_{\underline{t}-\underline{k}} + e_{\underline{t}} \tag{2.3}$$

for some white noise $(e_{\underline{t}})$. Figure 1 illustrates the shape of these types of AR models with an example of order $p = 3$; the index vectors $\underline{t} - \underline{k}$ from (2.3) are marked by the black dots while $\underline{t}$ can be found at the center. The AR model from (2.3) is one-sided in the sense of so-called lexicographical ordering of the plane $\mathbb{Z}^2$; we will discuss this property extensively further along the line in this section, but first formulate the AR sieve bootstrap algorithm.

Let $T_n = T_n(\{X_{\underline{t}} : \underline{t} \in \Pi\})$ be an estimator for some unknown parameter θ of the process, based on the given data sample. For an appropriately increasing sequence of real numbers $(c_n)_{n\in\mathbb{N}}$, we assume that the distributions $\mathcal{L}_n = \mathcal{L}(c_n(T_n - \theta))$ converge to a non-degenerated limiting distribution as $n \to \infty$. Our goal is to estimate the distribution $\mathcal{L}_n$ for some finite number $n \in \mathbb{N}$. We propose the following procedure:

The autoregressive sieve bootstrap algorithm for random fields:

(1) Select an order $p = p(n) \in \mathbb{N}$, $p \ll n$ and fit a pth order autoregressive model of shape (2.3) to the given observations, for example, by Yule–Walker estimation. Denote the estimated coefficients by $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

(2) Let $\Pi(n, p) := \{(t_1, t_2) \in \mathbb{Z}^2 : p + 1 \le t_1 \le n - p, p + 1 \le t_2 \le n\}$, that is, $\Pi(n, p)$ is the set of all vectors $\underline{t} \in \Pi$ such that $(\underline{t} - \underline{k}) \in \Pi$ for all $\underline{k} \in \Theta(p)$. Denote the residuals of the autoregressive fit by $\varepsilon'_{\underline{t}}(p) = X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p)X_{\underline{t}-\underline{k}}$ for all $\underline{t} \in \Pi(n, p)$, and

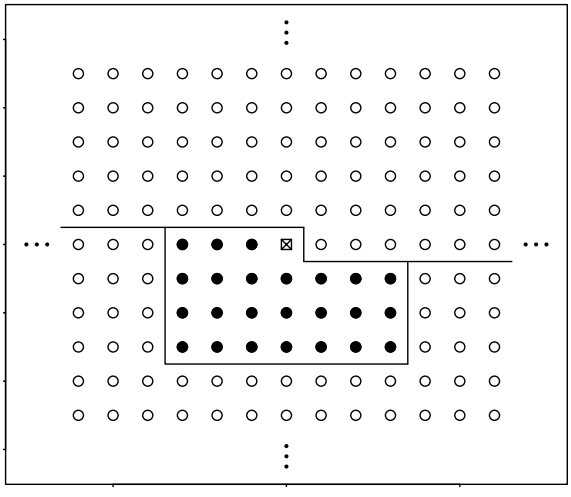

Figure 1. Illustration of the shape of an AR(3)-model with respect to $\Theta(3)$, cf. (2.3); locations of sites $\underline{t} - \underline{k}$ marked by the black dots.

let $\widehat{F}_n$ be the empirical distribution function of the centered residuals $\widehat{\varepsilon}_{\underline{t}}(p) = \varepsilon'_{\underline{t}}(p) - \overline{\varepsilon}$, where $\overline{\varepsilon} = (n-2p)^{-1}(n-p)^{-1} \sum_{\underline{t} \in \Pi(n,p)} \varepsilon'_{\underline{t}}(p)$. Generate independent random variables $\varepsilon^*_{\underline{j}}$ having identical distribution $\widehat{F}_n$, for example, by drawing with replacement from the set of centered residuals. Use these resampled residuals and the parameter estimators to calculate a bootstrap sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ according to the generating equation

$$X^*_{\underline{t}} = \sum_{\underline{k} \in \Theta(p)} \widehat{a}_{\underline{k}}(p) X^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}, \tag{2.4}$$

that is, the $X^*_{\underline{t}}$ are generated recursively, as $\underline{t}$ increases lexicographically through Π.

(3) Let $T^*_{n,(1)} := T_n(\{X^*_{\underline{t}} : \underline{t} \in \Pi\})$ be the same estimator as T_n based on the pseudo sample $\{X^*_{\underline{t}} : \underline{t} \in \Pi\}$ and θ^* the analogue of θ associated with the bootstrap process $(X^*_{\underline{t}})$.

(4) Repeat steps (1)–(3) M times, where M is sufficiently large, in order to obtain independent realisations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$ of the plug-in estimator.

(5) The estimator for $\mathcal{L}_n$ is then given by the empirical distribution of $\mathcal{L}^*_n = \mathcal{L}^*(c_n(T^*_n - \theta^*))$, based on the observations $T^*_{n,(1)}, \ldots, T^*_{n,(M)}$.

Here, $\mathcal{L}^*$ and E^* denote probability law and expectation, conditional on the given data sample.

In the following, we will investigate under which conditions the underlying process $(X_{\underline{t}})$ possesses one-sided autoregressive representations, since this property is crucial for showing asymptotic validity of the AR sieve bootstrap. For the remainder of this chapter, we will be working with spatial processes fulfilling the following assumptions. We use the notation $|\underline{k}|_\infty := \max\{|k_1|, |k_2|\}$ for the maximum vector norm of each $\underline{k} \in \mathbb{Z}^2$. For any arbitrary subset A of some vector space over $\mathbb{R}$ or $\mathbb{C}$, $\overline{\mathrm{sp}}(A)$ denotes the closed span of all vectors $a \in A$.

Assumption 1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a strictly stationary real-valued spatial process with mean zero and finite second moments. The autocovariance function $\gamma(\cdot)$ of $(X_{\underline{t}})$ fulfils $\sum_{\underline{k}\in\mathbb{Z}^2}(1+|\underline{k}|_\infty)^r|\gamma(\underline{k})|<\infty$ for some $r\in\mathbb{N}_0$ to be specified in the respective results later on. The spectral density of $(X_{\underline{t}})$,*

$$f(\underline{\lambda})=\frac{1}{4\pi^2}\sum_{\underline{k}\in\mathbb{Z}^2}\gamma(\underline{k})e^{-i\langle\underline{k},\underline{\lambda}\rangle},\qquad \underline{\lambda}\in(-\pi,\pi]^2,$$

fulfils the so-called boundedness condition: There exists a constant $c>0$ such that $f(\underline{\lambda})\geq c$ uniformly for all frequencies $\underline{\lambda}\in(-\pi,\pi]^2$.

It can be seen, that the representation of $\widehat{X}_{\underline{t}}$, the projection of $X_{\underline{t}}$ onto $\overline{\mathrm{sp}}\{X_{\underline{s}},\underline{s}\neq\underline{t}\}$, is closely related to the Fourier coefficients $d_{\underline{k}}$ of $1/f$ (compare to [35], Section 1.4). More precisely, one can show that

$$E|X_{\underline{t}}-\widehat{X}_{\underline{t}}|^2=\int_{(-\pi,\pi]^2}\left|1-\left(1-d_{\underline{0}}^{-1}f^{-1}(\underline{\lambda})\right)\right|^2 f(\underline{\lambda})\,d\underline{\lambda}=\int_{(-\pi,\pi]^2}d_{\underline{0}}^{-2}f^{-1}(\underline{\lambda})\,d\underline{\lambda}>0,$$

by the boundedness condition, which implies that $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is a so-called basic process, that is, $X_{\underline{t}}\notin\overline{\mathrm{sp}}\{X_{\underline{s}},\underline{s}\neq\underline{t}\}$.

Furthermore, Assumption 1 merely requires the spectral density to be positive and smooth, because the weighted summability condition on the autocovariances just implies that certain partial derivatives of f exist. For $u,v\in\mathbb{N}$ with $u+v\leq r$, we get from differentiating the Fourier series of f:

$$\frac{\partial^{u+v}f}{\partial\lambda_1^u\partial\lambda_2^v}(\underline{\lambda})=\frac{1}{4\pi^2}\sum_{\underline{k}\in\mathbb{Z}^2}(-ik_1)^u(-ik_2)^v\gamma(\underline{k})e^{-i\langle\underline{k},\underline{\lambda}\rangle}.$$

The derivative of the Fourier series of f on the right-hand side of the latter equation is absolutely summable because $|(-ik_1)^u(-ik_2)^v|\leq(1+|\underline{k}|_\infty)^r$ and because of Assumption 1. Therefore, the derivative of f itself, given by the left-hand side, exists and is equal to the derivative of the Fourier series.

We will now establish the aforementioned one-sided autoregressive and moving average representations for all processes that fulfil Assumption 1. Here, *one-sided* refers to the lexicographical ordering of the plane $\mathbb{Z}^2$, cf. [17]. Defining

$$\Theta:=\left\{(k_1,k_2)\in\mathbb{Z}^2:(k_1\geq1\text{ and }k_2=0)\text{ or }(k_1\text{ arbitrary and }k_2\geq1)\right\}$$

one can observe that $\mathbb{Z}^2$ can be partitioned as $\{\underline{0}\}\cup\Theta\cup(-\Theta)$. Θ is commonly referred to as the *upper half-plane* with respect to the origin while $-\Theta$ is the *lower half-plane*, cf. [21]. An illustration is given by Figure 2; the upper half-plane Θ is given by the white dots, the lower half-plane by the black dots. Obviously, it holds $\Theta(p)\to\Theta$, as $p\to\infty$.

We now get the following result on one-sided representations for spatial processes.

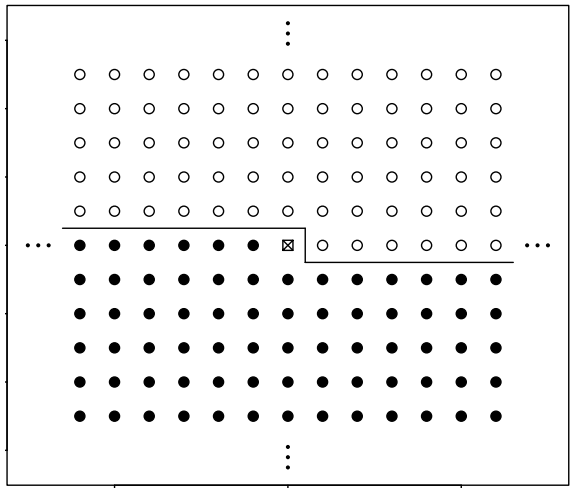

Figure 2. Illustration of the upper (white dots) and lower (black dots) half-plane of $\mathbb{Z}^2$.

Lemma 2.1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a spatial process that fulfils Assumption 1 with some $r \geq 1$. Then there exist uniquely determined autoregressive (AR) coefficients $(a_{\underline{k}})_{\underline{k}\in\Theta}$, uniquely determined moving average (MA) coefficients $(b_{\underline{k}})_{\underline{k}\in\Theta}$ and a uniquely determined uncorrelated white noise process $(\varepsilon_{\underline{t}})$, $\underline{t} \in \mathbb{Z}^2$, such that $(X_{\underline{t}})$ possesses the one-sided AR and MA representations*

$$X_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}} X_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \qquad X_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}} \varepsilon_{\underline{t}-\underline{k}} + \varepsilon_{\underline{t}}, \tag{2.5}$$

respectively, and $\sum_{\underline{k}\in\Theta} a_{\underline{k}} X_{\underline{t}-\underline{k}}$ represents the L^2-projection of $X_{\underline{t}}$ onto $\overline{\mathrm{sp}}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. The white noise process $(\varepsilon_{\underline{t}})$ is called the innovation process of $(X_{\underline{t}})$. The coefficients in (2.5) fulfil the summability conditions

$$\sum_{\underline{k}\in\Theta} \big(1 + |\underline{k}|_\infty\big)^{r-1} |a_{\underline{k}}| < \infty, \qquad \sum_{\underline{k}\in\Theta} \big(1 + |\underline{k}|_\infty\big)^{r-1} |b_{\underline{k}}| < \infty. \tag{2.6}$$

It should be noted that the existence of representations (2.5) has already been proven by [40]. However, we are especially interested in the summability conditions (2.6), which are not available in the literature. Hence, we derive these conditions in the proof of Lemma 2.1, which can be found [27].

Remark 2.2. At this point, we should clarify a common misunderstanding in the discussion of spatial and time series autoregressions: For time series, the "past" and the "future" of a time value $t \in \mathbb{Z}$ are naturally defined, and it is generally accepted that random variables X_t are influenced by its past values $X_{t-1}, X_{t-2}, \ldots$. Since this is not the case for spatial processes, it is often criticized that one has to choose a concept of "past" values, that is, choose a direction from

which the random variable $X_{\underline{t}}$ is influenced, such as the lower half-plane illustrated by Figure 2. This choice is of course arbitrary, which is why one might come to the conclusion that the whole concept of one-sided autoregressions implies a very specific model assumption which is not fulfilled for real-world data. However, the opposite is true: The AR sieve bootstrap, as an example, only uses the one-sided autoregressions as a vehicle in the proof of bootstrap validity. Under the mild conditions from Assumption 1, which only depend on the spectral density and which do not include any choice of direction whatsoever, the process $(X_{\underline{t}})$ possesses autoregressive representations with respect to *each half-plane* of $\mathbb{Z}^2$ that might be chosen. Therefore, the whole procedure is by no means arbitrary; and the concept of approximating a particular one-sided autogression does not constrain the class of processes any further than demanding the spectral density to be positive and smooth.

In order to prove the summability conditions from Lemma 2.1, we need the following auxiliary result. The AR and MA coefficients are strongly connected to the so-called *cepstral coefficients* of the process, that is the Fourier coefficients of the logarithm of the spectral density. The following lemma provides a result that carries over the summability condition from the Fourier coefficients of a function f to the Fourier coefficients of its logarithm. The result holds not only for spectral densities but for arbitrary integrable functions, and seems not to be available in the literature so far, at least not in this explicit form.

Lemma 2.3. *Denote for every integrable function* $f : (-\pi, \pi]^2 \to \mathbb{R}$ *its Fourier coefficients by* $\widetilde{f}_{\underline{k}} = (1/4\pi^2) \int_{(-\pi,\pi]^2} f(\underline{\lambda}) e^{-i\langle \underline{k}, \underline{\lambda} \rangle} \, d\underline{\lambda}$ *and by* $\sum_{\underline{k} \in \mathbb{Z}^2} \widetilde{f}_{\underline{k}} e^{i\langle \underline{k}, \underline{\lambda} \rangle}$ *its formal Fourier series. We define the following classes of functions*:

$$C_r := \left\{ f : (-\pi, \pi]^2 \to \mathbb{R}, \|f\|_r := \sum_{\underline{k} \in \mathbb{Z}^2} \left(1 + |\underline{k}|_\infty\right)^r |\widetilde{f}_{\underline{k}}| < \infty \right\},$$

$$D_{r_1,r_2} := \left\{ f : (-\pi, \pi]^2 \to \mathbb{R}, \|f\|_{r_1,r_2} := \sum_{\underline{k} \in \mathbb{Z}^2} \left(1 + |k_1|\right)^{r_1} \left(1 + |k_2|\right)^{r_2} |\widetilde{f}_{\underline{k}}| < \infty \right\}.$$

Assume that $f(\underline{\lambda}) \geq c > 0$ *for all* $\underline{\lambda} \in (-\pi, \pi]^2$. *Then it holds*:

(i) *If* $f \in C_r$ *for some* $r \geq 2$, *it follows* $\log f \in C_{r-1}$.
(ii) *If* $f \in D_{r_1,r_2}$ *for some* $r_1, r_2 \geq 1$, *it follows* $\log f \in D_{r_1,r_2}$.

Remark 2.4. In Assumption 1 and Lemma 2.3(i), we use the weight function $\nu(\underline{k}) = (1 + |\underline{k}|_\infty)^r$. This is due to the fact that we will later establish a weighted version of a Baxter-inequality for spatial processes, cf. Theorem 3.2. The proof of this Baxter-inequality requires the weights to be strictly non-decreasing in $|\underline{k}|_\infty$, that is, $\nu(\underline{k}) \geq \nu(\underline{j})$ whenever $|\underline{k}|_\infty \geq |\underline{j}|_\infty$ or, in other words, whenever $\underline{j} \in \Theta(p)$ and $\underline{k} \in \Theta \setminus \Theta(p)$. Other weights one might think of, like replacing the $|\cdot|_\infty$-norm in $\nu(\underline{k})$ by the Euclidean norm, the 1-norm or letting $\widetilde{\nu}(\underline{k}) = (1 + |k_1|)^{r_1}(1 + |k_2|)^{r_2}$, do not fulfil the property of being strictly non-decreasing in $|\underline{k}|_\infty$ and are, therefore, not suitable in order to establish a weighted Baxter-inequality. However, for Assumption 1 to be fulfilled, it suffices to check whether $\sum_{\underline{k} \in \mathbb{Z}^2} (1 + |\underline{k}|)^r |\gamma(\underline{k})| < \infty$ for any vector norm $|\underline{k}|$, since all vector norms are

equivalent. One could also switch to any other vector norm $|\cdot|_\alpha$, but in this case the projection set $\Theta(p)$ has to be modified such that $|\underline{k}|_\alpha \geq |\underline{j}|_\alpha$ whenever $\underline{j} \in \Theta(p)$ and $\underline{k} \in \Theta \setminus \Theta(p)$.

Remark 2.5. Classes of functions with weighted absolutely summable Fourier coefficients, such as C_r and D_{r_1,r_2} from Lemma 2.3, are commonly referred to as *Beurling algebras*; C_r represents the special case for the weight function $\nu(\underline{k}) = (1 + |\underline{k}|_\infty)^r$. Remark 2.4 explains why we are looking at these particular weights, although we get the somehow unsatisfactory result that $f \in C_r$ does *not imply* $\log f \in C_r$, but instead $\log f \in C_{r-1}$. While we will only work with assertion (i) from Lemma 2.3 for the remainder of this paper, it is still worthwile to consider the class D_{r_1,r_2} from (ii). Here, we get with analogous arguments as in (i) that $f \in D_{r_1,r_2}$ implies $\log f \in D_{r_1,r_2}$, i.e. the Fourier coefficients of $\log f$ fulfil the same summability condition as the ones of f. This result is strongly connected to the well-known Wiener–Lévy-theorem (cf. [42], Chapter VI, Theorem 5.2); and, for the special case of $\phi(f) = \log f$, our result even represents a slight generalisation of the latter, with respect to functions in several variables. We will shed some light on this situation:

Originally, Norbert Wiener proved for functions in one variable that if $f \neq 0$ has absolutely summable Fourier coefficients, then the same holds true for $1/f$. This assertion, also known as Wiener's lemma, can be transferred to functions in several variables; and, moreover, weighted summability versions in the spirit of Lemma 2.3 are available, cf. Theorem 6.2 in [15]. For functions in one variable, Paul Lévy generalised Wiener's result, concluding that if f has absolutely summable Fourier coefficients, the same holds true for $\phi(f)$, where ϕ is a smooth functional. This assertion became known as the Wiener–Lévy-theorem. In contrast to what happens for $\phi(f) = 1/f$, weighted versions in several variables are much harder to come by for general functions ϕ. Typically, one only gets that $\phi(f)$ is the element of a Beurling algebra with weights increasing at a slower rate than the ones of f, cf. [7].

Our proof of Lemma 2.3(ii) shows that a generalisation to functions in several variables for the special case of $\phi(f) = \log f$ is possible. However, the proof relies heavily on the structure of the logarithmic function and cannot be generalised to other functions.

3. Convergence of finite-order model fits

In this section, we will establish results that ensure convergence of the estimated parameters $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ from step (1) of the AR sieve bootstrap procedure, cf. Section 2, towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process given by Lemma 2.1. We will split up the results in two subsections: The first one will be concerned with convergence of the finite predictor coefficients of the process $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ towards $\{a_{\underline{k}} : \underline{k} \in \Theta\}$. The finite predictors are the L^2-projection coefficients of random variable $X_{\underline{t}}$ to the finite-dimensional space $\mathrm{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$. In this context, we will introduce a Baxter-inequality for spatial processes. Section 3.2 deals with conditions which ensure that the difference between the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and the finite predictor coefficients vanishes asymptotically in probability. The results from both subsections combined then yield the desired convergence of the finite-order AR model fits.

3.1. Convergence of finite predictor coefficients

The finite predictor coefficients with respect to the set $\Theta(p)$ are the coefficients of the L^2-projection of $X_{\underline{t}}$ onto $\mathrm{sp}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta(p)\}$, and will be denoted by $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. They can be obtained from solving the minimization problem

$$\big\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\big\} := \underset{\{c_{\underline{k}}(p):\underline{k}\in\Theta(p)\}}{\arg\min} \; E\bigg(X_{\underline{t}} - \sum_{\underline{k}\in\Theta(p)} c_{\underline{k}}(p) X_{\underline{t}-\underline{k}}\bigg)^2. \tag{3.1}$$

Solving (3.1) leads to the well-known Yule–Walker equations. We now want to introduce the notation which allows us to write the Yule–Walker equations in a convenient form: The number of elements in $\Theta(p)$ is $\bar{p} := 2p(p+1)$. Let $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ be an arbitrary enumeration of the vectors $\underline{k} \in \Theta(p)$. Define $\underline{a}(p) := (a_{\underline{k}_1}(p), \ldots, a_{\underline{k}_{\bar{p}}}(p))^T \in \mathbb{R}^{\bar{p}}$ and $\underline{Y}_{\underline{t}} := (X_{\underline{t}-\underline{k}_1}, \ldots, X_{\underline{t}-\underline{k}_{\bar{p}}})^T$. Note that the indices $\underline{k}_j$ appear in the same order in both vectors. Due to the projection property, it is easy to see that any solution of (3.1) fulfils

$$E\big(\big(X_{\underline{t}} - \underline{a}(p)^T \underline{Y}_{\underline{t}}\big) \cdot \underline{Y}_{\underline{t}}^T \underline{e}_j\big) = 0, \qquad j = 1, \ldots, \bar{p}, \tag{3.2}$$

where $\underline{e}_j$ denotes the jth unit vector. Using the notation $\Gamma(p) := E(\underline{Y}_{\underline{t}} \underline{Y}_{\underline{t}}^T)$ and $\underline{\gamma}(p) := E(X_{\underline{t}} \underline{Y}_{\underline{t}})$, system (3.2) is equivalent to

$$\Gamma(p)\underline{a}(p) = \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_1) & \cdots & \gamma(\underline{k}_1 - \underline{k}_{\bar{p}}) \\ \vdots & \ddots & \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_1) & \cdots & \gamma(\underline{k}_{\bar{p}} - \underline{k}_{\bar{p}}) \end{pmatrix} \cdot \begin{pmatrix} a_{\underline{k}_1}(p) \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix} = \underline{\gamma}(p). \tag{3.3}$$

System (3.3) is called the Yule–Walker equations. Note that the matrix $\Gamma(p)$ is symmetric, regardless of the order of indices in the vectors $\underline{Y}_{\underline{t}}$ and $\underline{a}(p)$. The following result ensures the existence of a unique solution of (3.3). Moreover, we establish a uniform bound for the spectral norms of the inverse matrices $\Gamma(p)^{-1}$, which will turn out to be crucial for proving the Baxter-inequality. The spectral norm of a real-valued quadratic matrix A is defined as the square root of the largest eigenvalue of $A^T A$, denoted by $\|A\|_{\mathrm{spec}} = \sqrt{\sigma_{\max}(A^T A)}$. For symmetric positive definite matrices, this formula can be simplified to $\|A\|_{\mathrm{spec}} = \sigma_{\max}(A)$.

Lemma 3.1. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1. Then the matrix $\Gamma(p)$ from the Yule–Walker equations (3.3) is invertible for all $p \in \mathbb{N}$. Furthermore, it holds $\|\Gamma(p)^{-1}\|_{\mathrm{spec}} \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$, where c is the lower bound of the spectral density from Assumption 1, and $\|\cdot\|_{\mathrm{spec}}$ denotes the spectral norm.*

The previous lemma justifies calling the unique solution $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ of (3.3) *the* finite predictor coefficients of the process for order p. As already mentioned, it is of critical importance for our sieve bootstrap scheme that the $a_{\underline{k}}(p)$ converge towards the autoregressive coefficients $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ of the underlying process from (2.5), as p tends to infinity. In particular, we have to ensure that this convergence is fast enough. Therefore, we introduce the following version of Baxter's inequality for random fields.

Theorem 3.2 (Baxter's inequality). *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils Assumption 1 with some $r \geq 2$ and $c > 0$. Let $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ be its finite predictor coefficients as defined above, and $\{a_{\underline{k}} : \underline{k} \in \Theta\}$ be its autoregressive coefficients given by (2.5). Denote by $K := \sum_{\underline{k}\in\mathbb{Z}^2} |\gamma(\underline{k})|$. Then it holds for all $s \in \mathbb{N}_0$ with $s + 1 < r$ and for all $p \in \mathbb{N}$:*

$$\sum_{\underline{k}\in\Theta(p)} \big(1 + |\underline{k}|_\infty\big)^s \big|a_{\underline{k}}(p) - a_{\underline{k}}\big| \leq \frac{K}{2\sqrt{2}\pi^2 c} \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \big(1 + |\underline{k}|_\infty\big)^{s+1} |a_{\underline{k}}|.$$

Due to Lemma 2.1 the right-hand side converges to zero as $p \to \infty$.

The established convergence of the autoregressive coefficients in Baxter's inequality is closely related to a similar convergence of moving average parameters, which shall be derived in the next step. To do this, we take a look at so-called *z-transforms*, also called *transfer functions*, cf. [8], Section 4.4. Based on the AR and MA representations from (2.5) with the coefficients $(a_{\underline{k}})$ and $(b_{\underline{k}})$, we define the z-transforms

$$A(\underline{z}) = 1 - \sum_{\underline{k}\in\Theta} a_{\underline{k}} z_1^{k_1} z_2^{k_2}, \qquad B(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}} z_1^{k_1} z_2^{k_2} \qquad \forall \underline{z} \in S, \tag{3.4}$$

where

$$S := \big\{\underline{z} \in \mathbb{C}^2 : |z_1| = 1, |z_2| \leq 1\big\}.$$

The series $A(\underline{z})$ and $B(\underline{z})$ converge absolutely on its domain S because of Lemma 2.1. It is worth noting that we have to make the distinction between z_1 and z_2 in S. Since z_2 shows up exclusively with exponents $k_2 \geq 0$ in (3.4), as can be seen from the definition of Θ in Section 2, we have $|z_2|^{k_2} \leq 1$ for the entire closed disk $|z_2| \leq 1$, while z_1 shows up with both positive and negative exponents k_1. Hence we get $|z_1|^{k_1} \leq 1$, and thus absolute convergence of the series $A(\underline{z})$ and $B(\underline{z})$, only for the circle $|z_1| = 1$.

In analogy to the definition of $A(\underline{z})$, we now define the z-transform of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ by

$$A_p(\underline{z}) = 1 - \sum_{\underline{k}\in\Theta(p)} a_{\underline{k}}(p) z_1^{k_1} z_2^{k_2} \qquad \forall \underline{z} \in S_p, \tag{3.5}$$

where

$$S_p := \left\{\underline{z} \in \mathbb{C}^2 : \frac{p}{p+1} \leq |z_1| \leq \frac{p+1}{p}, 0 \leq |z_2| \leq \frac{p+1}{p}\right\}.$$

Note that $A_p(\underline{z})$ is defined on an extended domain compared to $A(\underline{z})$, but for $p \to \infty$ the domains S_p converge to S.

From the proof of Lemma 2.1 we already have $B(\underline{z}) = 1/A(\underline{z})$ for all $\underline{z} \in S$. In particular, both $A(\underline{z})$ and $B(\underline{z})$ are non-zero on their domain S. The next lemma shows that, for p large enough, the inverse of $A_p(\underline{z})$ has a z-transform similar to the one of $B(\underline{z})$.

Lemma 3.3. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 with some $r \geq 2$. Then there exist $\delta > 0$ and $p_0 \in \mathbb{N}$ such that for all $p \geq p_0$ it holds $|A_p(\underline{z})| \geq \delta$ uniformly for all $\underline{z} \in S_p$. For those $p \geq p_0$, $B_p(\underline{z}) := 1/A_p(\underline{z})$ can be expressed as a convergent series of the form*

$$B_p(\underline{z}) = 1 + \sum_{\underline{k}\in\Theta} b_{\underline{k}}(p) z_1^{k_1} z_2^{k_2} \qquad \forall \underline{z} \in S_p, \tag{3.6}$$

for suitable coefficients $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$.

We conclude this section with a result which transfers the convergence of the autoregressive parameters from Baxter's inequality to the moving average parameters $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ and $\{b_{\underline{k}} : \underline{k} \in \Theta\}$:

Lemma 3.4. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 with some $r \geq 2$. For all p large enough such that $A_p(\underline{z}) \neq 0$ for all $\underline{z} \in S_p$, let $\{b_{\underline{k}}(p) : \underline{k} \in \Theta\}$ be the coefficients as defined in (3.6) and let $(a_{\underline{k}})_{\underline{k}\in\Theta}$ and $(b_{\underline{k}})_{\underline{k}\in\Theta}$ be the AR and MA coefficients of $(X_{\underline{t}})$ given by (2.5). Then there exists a constant $C < \infty$ such that it holds for all p large enough, and for all $s \in \mathbb{N}_0$ with $s + 1 < r$:*

$$\sum_{\underline{k}\in\Theta} \big(1 + |\underline{k}|_\infty\big)^s \big|b_{\underline{k}}(p) - b_{\underline{k}}\big| \leq C \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \big(1 + |\underline{k}|_\infty\big)^{s+1} |a_{\underline{k}}|.$$

Due to Lemma 2.1, the right-hand side converges to zero as $p \to \infty$.

The proof of Theorem 3.2 can be found in Section 7, while the proofs of the remaining results of this section are deferred to [27].

3.2. Conditions on the fitted-model order $p(n)$ and convergence of estimated coefficients

It is important for the validity of the AR sieve bootstrap scheme that the parameter estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ used in step 1 of the procedure converge towards the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ at a sufficient rate. At this point, one has to keep in mind that the order p of the autoregressive fits actually depends on the sample size n, which is suppressed in the notation for most parts of this paper due to convenience reasons. In order to use the results from the previous section, we need $p = p(n) \to \infty$ as $n \to \infty$. This implies that the dimension of the Yule–Walker matrices $\Gamma(p)$ given by (3.3) also increases for $n \to \infty$.

Probably the most popular form of fitting an AR model as in step (1) of the sieve bootstrap procedure, is Yule–Walker estimation: One replaces the autocovariances in $\Gamma(p)$ by its empirical versions, cf. (2.1), and solves the linear system. Informally speaking, we then have to make sure that $p(n)$ increases slowly enough such that for n large enough all autocovariances showing up in $\Gamma(p)$ can be estimated sufficiently well, in order to obtain a small difference between $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$ and $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

The following assumption formalizes this condition. Essentially it contains two assertions: First, the underlying process allows for consistent estimation of the finite predictor coefficients $\{a_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$. Second, by restricting the rate of increase of $p = p(n)$, we can achieve sufficiently fast uniform convergence of the estimators $\{\widehat{a}_{\underline{k}}(p) : \underline{k} \in \Theta(p)\}$.

Assumption 2. *For $p = p(n)$, with $p(n) \to \infty$ as $n \to \infty$, assume for the following sequence in n:*

$$p^4 \cdot \sum_{\underline{k} \in \Theta(p)} \left|\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)\right| = \mathcal{O}_P(1).$$

In the remainder of this section, we will investigate whether the fitted AR models can also be represented as moving averages of possibly infinite order, which will be crucial for asymptotic inference later on. Based on the parameter estimators $\widehat{a}_{\underline{k}}(p)$ we can define the z-transform $\widehat{A}_p(\underline{z})$ analogously to $A_p(\underline{z})$ in (3.5) as

$$\widehat{A}_p(\underline{z}) = 1 - \sum_{\underline{k} \in \Theta(p)} \widehat{a}_{\underline{k}}(p) z_1^{k_1} z_2^{k_2} \qquad \forall \underline{z} \in S_p.$$

The following calculations will make sure that $\widehat{A}_p(\underline{z})$ is bounded away from zero for n large enough. Assumption 2 implies

$$\begin{aligned} \sup_{\underline{z} \in S_p} \left|\widehat{A}_p(\underline{z}) - A_p(\underline{z})\right| &\leq \sum_{\underline{k} \in \Theta(p)} \left|\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)\right| \left(\frac{p+1}{p}\right)^{|k_1|+k_2} \\ &\leq \left(\frac{p+1}{p}\right)^{2p} \sum_{\underline{k} \in \Theta(p)} \left|\widehat{a}_{\underline{k}}(p) - a_{\underline{k}}(p)\right| \qquad (3.7) \\ &= \frac{1}{p^4} \mathcal{O}_P(1) = o_P(1), \end{aligned}$$

because $((p+1)/p)^{2p}$ is a bounded sequence (convergent with limit e^2), and because the definition of S_p yields

$$|z_1|^{k_1} \leq \left\{ \begin{array}{ll} \left(\frac{p+1}{p}\right)^{k_1}, & \text{for } k_1 \geq 0, \\ \left(\frac{p}{p+1}\right)^{k_1}, & \text{for } k_1 < 0 \end{array} \right\} = \left(\frac{p+1}{p}\right)^{|k_1|},$$

$$|z_2|^{k_2} \leq \left(\frac{p+1}{p}\right)^{k_2},$$

for all $\underline{z} \in S_p$. Assumption 2 ensures $p \to \infty$, as $n \to \infty$, which implies that $A_p(\underline{z})$ is bounded away from zero for all n large enough, cf. Lemma 3.3. It follows from (3.7) that $\widehat{A}_p(\underline{z})$ is uniformly bounded away from zero in probability for all $\underline{z} \in S_p$ and for all n large enough. For all

those n large enough, the inverse of $\widehat{A}_p(\underline{z})$ possesses the expansion

$$\widehat{B}_p(\underline{z}) = \frac{1}{\widehat{A}_p(\underline{z})} = 1 + \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p) z_1^{k_1} z_2^{k_2} \qquad \forall \underline{z} \in S_p, \tag{3.8}$$

in probability, following the same arguments as for (3.6). Hence, the bootstrap process given by (2.4), which can be described by the transfer function $\widehat{A}_p(\underline{z})$, has the moving average representation

$$X_{\underline{t}}^* = \sum_{\underline{k}\in\Theta} \widehat{b}_{\underline{k}}(p) \varepsilon_{\underline{t}-\underline{k}}^* + \varepsilon_{\underline{t}}^* \tag{3.9}$$

for all n large enough, in probability. The convergence of the parameter estimators $\widehat{a}_{\underline{k}}(p)$ towards $a_{\underline{k}}(p)$ in Assumption 2 carries over to the corresponding moving average parameters, as shows the following lemma.

Lemma 3.5. *Let $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a process that fulfils the conditions of Theorem 3.2 and Assumption 2. Then, for all n large enough (and thus p large enough) such that $A_p(\underline{z})$ and $\widehat{A}_p(\underline{z})$ are bounded away from zero (the latter in probability), it holds uniformly for all $\underline{k} \in \Theta$ and for some $C < \infty$:*

$$\left|\widehat{b}_{\underline{k}}(p) - b_{\underline{k}}(p)\right| \le C \cdot \left(1 + \frac{1}{p}\right)^{-|k_1|-k_2} \frac{1}{p^4} \qquad \textit{in probability.}$$

The proof can be found in [27].

4. Asymptotic validity of the bootstrap

In this section, we will derive asymptotic validity of the AR sieve bootstrap procedure under appropriate conditions for a class of statistics which will be specified in Assumption 3. Similar to what happens in the time series case, cf. [24], it turns out that the bootstrap procedure asymptotically mimics the behaviour of the so-called companion process, a modification of the underlying process $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$. This yields a check criterion which basically says that the bootstrap procedure works asymptotically for a test statistic T_n, whenever the asymptotic distributions of T_n applied to the underlying and the companion process coincide. We will elaborate this, and start with the definition of the companion process:

Based on representation (2.5) for the underlying process, we define the *companion process* of $(X_{\underline{t}})$ as the stationary spatial process $(\widetilde{X}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$, generated by

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} a_{\underline{k}} \widetilde{X}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{4.1}$$

where the coefficients $a_{\underline{k}}$ are exactly the ones from (2.5) and $(\widetilde{\varepsilon}_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ is an i.i.d. white noise process with identical marginal distribution as $(\varepsilon_{\underline{t}})$, that is, $\mathcal{L}(\widetilde{\varepsilon}_{\underline{t}}) = \mathcal{L}(\varepsilon_{\underline{t}})$. Therefore, the companion

process also possesses the moving average representation

$$\widetilde{X}_{\underline{t}} = \sum_{\underline{k}\in\Theta} b_{\underline{k}} \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}}, \tag{4.2}$$

with the exact same coefficients $b_{\underline{k}}$ as in (2.5). The only difference between $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ is the dependence structure of the respective noise processes $(\varepsilon_{\underline{t}})$ and $(\widetilde{\varepsilon}_{\underline{t}})$. While $(\widetilde{\varepsilon}_{\underline{t}})$ is i.i.d., $(\varepsilon_{\underline{t}})$ is strictly stationary but *not necessarily independent*, the random variables $\varepsilon_{\underline{s}}$ and $\varepsilon_{\underline{t}}$ in general are only uncorrelated for $\underline{s} \neq \underline{t}$. Nevertheless, it is easy to see from (4.2) that all second order properties of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ are identical, that is, the two processes possess identical autocovariances and spectral densities.

In our main theorem, we will establish bootstrap validity for a class of statistics which will be specified in the following Assumption 3. This class is a natural extension of the so-called *functions of generalized means*, introduced by [25], to the case of random fields. These statistics will be based on smooth functions g applied to rectangular-shaped subsamples of the available data sample $\{X_{\underline{t}} : \underline{t} \in \Pi\}$, with $\Pi := \{\underline{t} \in \mathbb{Z}^2 : 1 \leq t_1, t_2 \leq n\}$. We first specify the necessary notation: For $1 \leq m_1, m_2 \leq n$ let

$$\begin{aligned} S(m_1, m_2) :&= \left\{\underline{s} = (s_1, s_2)^T \in \mathbb{N}_0^2 : 0 \leq s_1 \leq m_1 - 1, 0 \leq s_2 \leq m_2 - 1\right\} \\ &= \left\{\underline{s}(1), \ldots, \underline{s}(m_1 m_2)\right\}, \end{aligned}$$

i.e., $\underline{s}(1), \ldots, \underline{s}(m_1 m_2)$ is any fixed enumeration of the $m_1 m_2$ vectors in $S(m_1, m_2)$. We define the $m_1 m_2$-dimensional random vector

$$\mathbf{Y}_{\underline{t}} := (X_{\underline{t}+\underline{s}(1)}, \ldots, X_{\underline{t}+\underline{s}(m_1 m_2)})^T.$$

Observe that for each $\underline{t}$ with $1 \leq t_1 \leq n - m_1 + 1$ and $1 \leq t_2 \leq n - m_2 + 1$, the components of $\mathbf{Y}_{\underline{t}}$ form a rectangular-shaped subsample of dimension $m_1 \times m_2$ of the original data sample. We can now specify the class of statistics we will be investigating.

Assumption 3. *Let $\bar{n}_1 := n - m_1 + 1$, $\bar{n}_2 := n - m_2 + 1$ for some $1 \leq m_1, m_2 \leq n$, and let $m := m_1 m_2$. Define the statistic T_n as*

$$T_n = f\left(\frac{1}{\bar{n}_1 \bar{n}_2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} g(\mathbf{Y}_{\underline{t}})\right),$$

where the functions $g : \mathbb{R}^m \to \mathbb{R}^k$ and $f : \mathbb{R}^k \to \mathbb{R}$, with $k \geq 1$, fulfil the following smoothness conditions: f is continuously differentiable in a neighborhood of $\underline{\theta} := Eg(\mathbf{Y}_{\underline{t}})$ and the gradient of f at $\underline{\theta}$ does not vanish, that is,

$$\nabla f(\theta) = \left(\frac{\partial f(\underline{x})}{\partial x_1}, \ldots, \frac{\partial f(\underline{x})}{\partial x_k}\right)\Big|_{\underline{x}=\underline{\theta}} \neq (0, \ldots, 0).$$

For some $h \geq 1$ all component functions $g_1, \ldots, g_k$ of g are h times continuously differentiable and all hth-order derivatives satisfy a Lipschitz condition, that is, for all $i = 1, \ldots, k$ and for all

$(h_1, \ldots, h_m) \in \mathbb{N}_0^m$ *with* $\sum_{u=1}^m h_u = h$ *the derivative*

$$\frac{\partial^h g_i(\underline{x})}{\partial^{h_1} x_1 \cdots \partial^{h_m} x_m}$$

is Lipschitz.

Remark 4.1. The conditions from the previous assumption should be explained at this point: The class of statistics from Assumption 3 contains, among other things, the sample mean and versions of the sample autocovariance and sample autocorrelation. To obtain the latter two statistics, one typically uses a function g which is not Lipschitz. For example, in the case of sample autocovariances at lag $\underline{h} = (h_1, h_2)^T$, one may choose $m_1 = h_1 + 1$, $m_2 = h_2 + 1$ and $g(x_1, \ldots, x_m) = x_1 x_m$. Then T_n from Assumption 3 translates to taking the empirical mean of observations $X_{\underline{t}+\underline{h}} X_{\underline{t}}$. Now observe that g itself is *not Lipschitz*, but all of its first order partial derivatives are. This is the why we allow for non-Lipschitz functions g in Assumption 3, and merely assume that there exists a number $1 \leq h < \infty$ such that all derivatives of order h (but *not* up to order h) are Lipschitz.

In order to state the main theorem, we define $\widetilde{T}_n$ and T_n^* as the statistic T_n applied to samples from the companion process $(\widetilde{X}_{\underline{t}})$ and the bootstrap process $(X_{\underline{t}}^*)$, respectively, that is,

$$\widetilde{T}_n := f\left(\frac{1}{\bar{n}_1 \bar{n}_2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} g(\widetilde{\mathbf{Y}}_{\underline{t}})\right), \qquad T_n^* := f\left(\frac{1}{\bar{n}_1 \bar{n}_2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} g\big(\mathbf{Y}_{\underline{t}}^*\big)\right),$$

where

$$\widetilde{\mathbf{Y}}_{\underline{t}} := (\widetilde{X}_{\underline{t}+\underline{s}(1)}, \ldots, \widetilde{X}_{\underline{t}+\underline{s}(m_1 m_2)})^T, \qquad \mathbf{Y}_{\underline{t}}^* := \big(X_{\underline{t}+\underline{s}(1)}^*, \ldots, X_{\underline{t}+\underline{s}(m_1 m_2)}^*\big)^T.$$

We can prove bootstrap validity under the following assumptions, which ensure convergence of empirical moments and the empirical distribution function to their theoretical counterparts for the innovations.

Assumption 4. *For all continuity points $x \in \mathbb{R}$ of the distribution function F of $\varepsilon_{\underline{0}}$ it holds*

$$F_n(x) \xrightarrow{P} F(x) \qquad \text{as } n \to \infty,$$

where $F_n(x)$ is the empirical distribution function

$$F_n(x) = \frac{1}{|\Pi(n, p)|} \sum_{\underline{t} \in \Pi(n,p)} \mathbb{1}\{\varepsilon_{\underline{t}} \leq x\},$$

and where $\Pi(n, p) := \{(t_1, t_2) \in \mathbb{Z}^2 : p + 1 \leq t_1 \leq n - p, p + 1 \leq t_2 \leq n\}$.

Furthermore, it holds $E(\varepsilon_{\underline{t}}^{2(h+2)}) < \infty$, *where* h *is the constant specified in Assumption 3, as well as the following convergence of empirical moments*:

$$\frac{1}{|\Pi(n,p)|}\sum_{\underline{t}\in\Pi(n,p)}(\varepsilon_{\underline{t}})^{2w} \xrightarrow{P} E\big((\varepsilon_{\underline{0}})^{2w}\big) \qquad \forall w \le h+2.$$

Theorem 4.2. *Let* $(X_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ *be a process fulfilling Assumptions 2–4, as well as Assumption 1 with* $r = 4$.

Then, for $\widetilde{T}_n$ *and* T_n^* *as defined above, it holds*

$$d_K\big(\mathcal{L}^*\big(n\big(T_n^* - f\big(\underline{\theta}^*\big)\big)\big), \mathcal{L}\big(n\big(\widetilde{T}_n - f\big(\widetilde{\underline{\theta}}\big)\big)\big)\big) = o_P(1)$$

as $n \to \infty$, *where* $\underline{\theta}^* = E^*(g(\mathbf{Y}_{\underline{t}}^*))$, $\widetilde{\underline{\theta}} = E(g(\widetilde{\mathbf{Y}}_{\underline{t}}))$ *and* d_K *denotes the Kolmogorov distance.*

This result shows for all statistics from Assumption 3 that the sieve bootstrap procedure asymptotically approximates the distribution $\widetilde{T}_n$ instead of the one of T_n. Therefore, the bootstrap procedure works asymptotically *if and only if* the limiting distributions of T_n and $\widetilde{T}_n$ coincide. We will give a few examples of the application of this check criterion in the following section. The proof of Theorem 4.2 can be found in Section 7.

5. Applications

In this section, we will give a few examples of prominent statistics to which the check criterion derived in the previous section can be applied. For a simulation study concerning sample autocorrelations, see Section 6.

Example 5.1 (Sample mean). We can use the AR sieve bootstrap procedure for the sample mean, even for processes which are not centered as required per Assumption 1. Let $(Z_{\underline{t}})_{\underline{t}\in\mathbb{Z}^2}$ be a strictly stationary process with mean μ which, other than being non-centered, fulfils the conditions stated in Assumption 1. Since all autocovariances of $(Z_{\underline{t}})$ and the centered process $(X_{\underline{t}}) := (Z_{\underline{t}} - \mu)$ coincide, $(X_{\underline{t}})$ obviously fulfils Assumption 1. Now let $\{Z_{\underline{t}}, \underline{t} \in \Pi\}$ be a data sample generated by $(Z_{\underline{t}})$. We apply the bootstrap procedure described in Section 2 to the data $\{Z_{\underline{t}}, \underline{t} \in \Pi\}$, which produces bootstrap samples $\{X_{\underline{t}}^*, \underline{t} \in \Pi\}$, generated by

$$X_{\underline{t}}^* = \sum_{\underline{k}\in\Theta(p)} \widehat{a}_{\underline{k}}(p) X_{\underline{t}-\underline{k}}^* + \varepsilon_{\underline{t}}^*.$$

Then, compute $Z_{\underline{t}}^* := \overline{Z} + X_{\underline{t}}^*$ for all $\underline{t} \in \Pi$, where $\overline{Z} := |\Pi|^{-1}\sum_{\underline{t}\in\Pi} Z_{\underline{t}}$ (for the bootstrap data, $\overline{Z}^*$ is analogously defined). We can approximate the distribution of $n(\overline{Z} - \mu)$ by the one of $n(\overline{Z}^* - \overline{Z})$. Asymptotic validity of this approach can be established via Theorem 4.2 in the following way:

The companion process associated with $(X_{\underline{t}})$ is denoted by $(\widetilde{X}_{\underline{t}})$ and we define $\widetilde{Z}_{\underline{t}} := \widetilde{X}_{\underline{t}} + \mu$. The functions f and g in Assumption 3 can be chosen appropriately such that T_n is the sample

mean of $\{X_{\underline{t}}, \underline{t} \in \Pi\}$, and $\widetilde{T}_n = \overline{\widetilde{X}}$ is the mean of $\{\widetilde{X}_{\underline{t}}, \underline{t} \in \Pi\}$. For the linear process $(\widetilde{Z}_{\underline{t}})$, with an obvious notation for $\overline{\widetilde{Z}}$, it is known that

$$n(\overline{\widetilde{Z}} - \mu) = n(\overline{\widetilde{X}}) = n\widetilde{T}_n \xrightarrow{d} \mathcal{N}\Big(0, \sum_{\underline{h} \in \mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big),$$

where $\gamma_{\widetilde{Z}}$ denotes the autocoavariance function of $(\widetilde{Z}_{\underline{t}})$. Noting that $\overline{Z}^* = \overline{Z} + \overline{X}^*$, it follows immediately from Theorem 4.2

$$n\big(\overline{Z}^* - \overline{Z}\big) = n\big(\overline{X}^*\big) = nT_n^* \xrightarrow{d^*} \mathcal{N}\Big(0, \sum_{\underline{h} \in \mathbb{Z}^2} \gamma_{\widetilde{Z}}(\underline{h})\Big) \qquad \text{in prob.} \tag{5.1}$$

For the sample mean $\overline{Z}$ of the actually observed data it holds under suitable regularity conditions that

$$n(\overline{Z} - \mu) \xrightarrow{d} \mathcal{N}\Big(0, \sum_{\underline{h} \in \mathbb{Z}^2} \gamma_Z(\underline{h})\Big). \tag{5.2}$$

Now observe that $(Z_{\underline{t}})$ and $(\widetilde{Z}_{\underline{t}})$ have identical second order properties per definition. In particular, $\gamma_Z(\underline{h}) = \gamma_{\widetilde{Z}}(\underline{h})$ for all lags $\underline{h} \in \mathbb{Z}^2$. Thus, the limiting distributions in (5.1) and (5.2) coincide and it follows

$$d_K\big(\mathcal{L}^*\big(n\big(\overline{Z}^* - \overline{Z}\big)\big), \mathcal{L}\big(n(\overline{Z} - \mu)\big)\big) = o_P(1).$$

Therefore, the AR sieve bootstrap proposal is asymptotically valid for the sample mean under the stated conditions.

In contrast to the preceeding example, the limiting distribution of sample autocovariances does not depend exclusively on second-order properties of the underlying process. This result is well known, particularly for the time-series case, that is, $d = 1$. Even if the data are generated by a linear spatial process, that is a process of the form

$$X_{\underline{t}} = \sum_{\underline{\nu} \in \mathbb{Z}^2} \alpha_{\underline{\nu}} u_{\underline{t} - \underline{\nu}}, \tag{5.3}$$

with absolutely summable coefficients $(\alpha_{\underline{\nu}})_{\underline{\nu} \in \mathbb{Z}^2}$ and an i.i.d. white noise process $(u_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ with finite fourth moments, the limiting variance depends on the fourth-order cumulants of $(u_{\underline{t}})$. This can be verified with analogous calculations as for the times series case, cf. [8], Proposition 7.3.4. However, the situation is different if one switches to sample autocorrelations of linear processes, instead of autocovariances. Then, the limiting distribution depends only on the autocorrelations of the underlying process, as shows the following theorem, which is a direct generalisation of the well-known Bartlett formula for time series, cf. [8], Proposition 7.2.1.

Lemma 5.2. *Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a linear spatial process as defined in (5.3), that is, with i.i.d. white noise and finite fourth moments, and with autocorrelation function ρ. For the sample autocor-*

relations $\widehat{\rho}(\underline{h}) = \widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$, *with* $\widehat{\gamma}(\cdot)$ *as defined in* (2.1), *we define the comparative quantity* $\check{\rho}(\underline{h}) := \check{\gamma}(\underline{h})/\check{\gamma}(\underline{0})$ *with*

$$\check{\gamma}(\underline{h}) := \frac{1}{|\Pi|} \sum_{\underline{t} \in \Pi} X_{\underline{t}+\underline{h}} X_{\underline{t}},$$

where $\Pi = \{\underline{t} \in \mathbb{Z}^2 : 1 \le t_1, t_2 \le n\}$. $\check{\rho}(\underline{h})$ *and* $\widehat{\rho}(\underline{h})$ *are asymptotically equivalent. Then it holds*

$$n^2 \operatorname{Cov}\big(\check{\rho}(\underline{h}), \check{\rho}(\underline{k})\big) \longrightarrow V(\underline{h}, \underline{k}), \qquad \text{as } n \to \infty,$$

where

$$\begin{aligned} V(\underline{h}, \underline{k}) = \sum_{\underline{r} \in \mathbb{Z}^2} \big\{ & 2\rho(\underline{r})^2 \rho(\underline{k})\rho(\underline{h}) - 2\rho(\underline{r}+\underline{k})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{k}) \\ & + \rho(\underline{r}-\underline{h}+\underline{k})\rho(\underline{r}) + \rho(\underline{r}+\underline{k})\rho(\underline{r}-\underline{h}) \big\}. \end{aligned}$$

The proof is analogous to the time-series case and can be found in [27].

Example 5.3 (Sample autocorrelations/correlogram). Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with corresponding companion process $(\widetilde{X}_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$. We consider the autocorrelation function $\rho(\underline{h}) = \gamma(\underline{h})/\gamma(\underline{0})$ at lag $\underline{h}$, together with the usual estimator $T_n := \widehat{\rho}(\underline{h}) = \widehat{\gamma}(\underline{h})/\widehat{\gamma}(\underline{0})$, where $\widehat{\gamma}(\cdot)$ is given by (2.1). For spatial processes, $\rho(\underline{h})$ (and accordingly $\widehat{\rho}(\underline{h})$) are often referred to as the (sample) correlogram, cf. [13], Section 2.3.2. Note that the autocorrelations of $(\widetilde{X}_{\underline{t}})$ are given by the function ρ as well. Under suitable assumptions on the dependence structure of the process, such as weak dependence or mixing conditions, it is known that

$$n\big(\widehat{\rho}(\underline{h}) - \rho(\underline{h})\big) \xrightarrow{d} \mathcal{N}\big(0, \tau_X^2\big), \qquad n\big(\widetilde{T}_n - \rho(\underline{h})\big) \xrightarrow{d} \mathcal{N}\big(0, \tau_{\widetilde{X}}^2\big),$$

where the limiting variances τ_X^2 and $\tau_{\widetilde{X}}^2$ in general depend on the fourth order cumulants of $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$, respectively. Hence, it follows $\tau_X^2 \neq \tau_{\widetilde{X}}^2$ in general, because $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$ share second order but not fourth order properties. For T_n^*, denoting the sample autocorrelation applied to the bootstrap sample $\{X_{\underline{t}}^*, \underline{t} \in \Pi\}$, Theorem 4.2 yields

$$n\big(T_n^* - f\big(\underline{\theta}^*\big)\big) \xrightarrow{d} \mathcal{N}\big(0, \tau_{\widetilde{X}}^2\big).$$

Therefore, $\tau_X^2 \neq \tau_{\widetilde{X}}^2$ implies that the AR sieve bootstrap in general is asymptotically not valid for sample autocorrelations.

However, if the data are generated by a linear process $(X_{\underline{t}})$ as given by (5.3), Lemma 5.2 shows that the limiting variance of $n(\check{\rho}(\underline{h}) - \rho(\underline{h}))$ is given by

$$\begin{aligned} \tau_X^2 = \sum_{\underline{r} \in \mathbb{Z}^2} \big\{ & 2\rho(\underline{r})^2 \rho(\underline{h})^2 - 2\rho(\underline{r}+\underline{h})\rho(\underline{r})\rho(\underline{h}) - 2\rho(\underline{r}-\underline{h})\rho(\underline{r})\rho(\underline{h}) \\ & + \rho(\underline{r})^2 + \rho(\underline{r}+\underline{h})\rho(\underline{r}-\underline{h}) \big\}. \end{aligned} \tag{5.4}$$

Since $\check{\rho}(\underline{h})$ and $\widehat{\rho}(\underline{h})$ are asymptotically equivalent, $n(\widehat{\rho}(\underline{h}) - \rho(\underline{h}))$ also has limiting variance τ_X^2. This expression depends only on the autocorrelations of the underlying process, which coincide for $(X_{\underline{t}})$ and $(\widetilde{X}_{\underline{t}})$. Thus, it follows for this case $\tau_X^2 = \tau_{\widetilde{X}}^2$, and the bootstrap procedure is asymptotically valid for sample autocorrelations of data generated from linear processes.

Remark 5.4. When checking for asymptotic validity of the AR sieve bootstrap procedure, it is of critical importance to ensure that the limiting distributions of T_n and $\widetilde{T}_n$ are identical, as has been done in the previous examples. In general, this will be the case whenever the limiting distribution depends only on second order entities such as autocovariances or the spectral density of the underlying process. For data generated by a linear process $X_{\underline{t}} = \sum_{\underline{\nu} \in \mathbb{Z}^2} \alpha_{\underline{\nu}} u_{\underline{t}-\underline{\nu}}$, one might be tempted to conclude that $(X_{\underline{t}})$ and its companion process $(\widetilde{X}_{\underline{t}})$ are identical since $(u_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ is already i.i.d. However, Example 3.2 from [24] shows for the special case of time series that this is not the case. To be precise, the companion process $(\widetilde{X}_{\underline{t}})$ is always derived from the AR representation (2.5), where $(\varepsilon_{\underline{t}})$ is the uniquely determined innovation process of $(X_{\underline{t}})$. Even if the process has linear representation $X_{\underline{t}} = \sum_{\underline{\nu} \in \mathbb{Z}^2} \alpha_{\underline{\nu}} u_{\underline{t}-\underline{\nu}}$ with i.i.d. noise $(u_{\underline{t}})$, its innovation process might differ from $(u_{\underline{t}})$, and might be only uncorrelated but not i.i.d. Remark 2.1 of [24] gives a specific example of this situation. Therefore, linear processes are in general not identical to their companion processes, which makes a careful inspection of the limiting distributions as in the previous examples a necessity.

Example 5.5 (Standardized sample variogram). Let $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ be a spatial process fulfilling Assumption 1 with autocovariance function γ. The variogram at lag $\underline{h} \in \mathbb{Z}^2$ is defined as

$$V(\underline{h}) = \mathrm{Var}(X_{\underline{t}} - X_{\underline{t}+\underline{h}}) = E\big((X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2\big) = 2\gamma(\underline{0}) - 2\gamma(\underline{h})$$

for centered fields, and $V^{(s)}(\underline{h}) := V(\underline{h})/\gamma(\underline{0})$ is called the standardized variogram. Using the notation from (2.1), two classical estimators for $V(\underline{h})$ are given by

$$\widehat{V}_1(\underline{h}) = 2\widehat{\gamma}(\underline{0}) - 2\widehat{\gamma}(\underline{h}), \qquad \widehat{V}_2(\underline{h}) = \frac{1}{|\Pi_{\underline{h}}|} \sum_{\underline{t} \in \Pi_{\underline{h}}} (X_{\underline{t}} - X_{\underline{t}+\underline{h}})^2,$$

which are asymptotically equivalent, cf. [13], Section 2.4. In particular, one can easily check that

$$n\big(\widehat{V}_1(\underline{h}) - \widehat{V}_2(\underline{h})\big) = o_P(1). \tag{5.5}$$

Versions of both of these estimators are included in the class of functions of generalized means, as given by Assumption 3. Furthermore, both $\widehat{V}_1(\underline{h})$ and $\widehat{V}_2(\underline{h})$ can be used to construct standardized sample variogram estimators via $\widehat{V}_j^{(s)}(\underline{h}) := \widehat{V}_j(\underline{h})/\widehat{\gamma}(\underline{0})$, $j = 1, 2$. It holds

$$\widehat{V}_1^{(s)}(\underline{h}) = 2 - 2\widehat{\rho}(\underline{h}).$$

Now assume the data are generated by a linear process. Then it follows from Example 5.3 asymptotic validity of the AR sieve bootstrap procedure for the standardized sample variogram, as long as the data are generated by a linear spatial process.

Remark 5.6. Our main result Theorem 4.2 provides a check criterion for asymptotic validity of the AR sieve bootstrap for all statistics from Assumption 3. This class of statistics contains, among other things, the statistics from Examples 5.1–5.5. However, we conjecture that analogous results can be proven, in the same spirit as in the proof of Theorem 4.2, for a much wider class of statistics beyond those covered by Assumption 3. If T_n denotes an estimator for some parameter θ, under the condition that $\mathcal{L}(c_n(T_n - \theta))$ has a non-degenerated limiting distribution for some sequence (c_n), we conjecture that the AR sieve bootstrap procedure is asymptotically valid, as long as the limiting distribution depends on second order properties of the underlying process, only.

For example, according to Section 4.5 in [17], one can prove central limit theorems for kernel-based nonparametric spectral density estimators for strictly stationary spatial processes under appropriate mixing conditions. The limiting distribution then depends exclusively on the spectral density of the underlying process, which is a second order quantity, and we conjecture that the AR sieve bootstrap is asymptotically valid in this situation.

6. A simulation study

In this section, we will present simulation results that compare the performance of the AR sieve bootstrap to classic normal approximations and block bootstrap methods. First, we generated square-shaped samples $\{X_{\underline{t}} = X_{t_1,t_2} : 1 \le t_1, t_2 \le n\}$ as defined in Section 2, where the sample size is set to be $n = 15$ which corresponds to $15 \times 15 = 225$ observations. The samples are generated by a moving average model given by

$$X_{t_1,t_2} = e_{t_1,t_2} + 0.5 \cdot e_{t_1+1,t_2} - 0.2 \cdot e_{t_1-1,t_2} + 0.3 \cdot e_{t_1,t_2+1} + 0.1 \cdot e_{t_1,t_2-1}, \tag{6.1}$$

where $(e_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ is an i.i.d. white noise process with marginal distribution $\mathcal{N}(0, 1)$. The process $(X_{\underline{t}})_{\underline{t} \in \mathbb{Z}^2}$ fulfils the conditions of Assumption 1. Furthermore, each realisation $X_{\underline{t}}$ depends on noise terms from four different directions, two from the lower and two from the upper half-plane, cf. Section 2. This means that the process is *not* "tailor-made" for an AR approximation in the direction of the lower half-plane as performed in the AR sieve algorithm. In fact, the data generating process from (6.1) does not "favor' any direction of one-sided autoregressive fits; one could as well fit models that are one-sided with respect to the upper, left or right half-plane.

The statistic that we investigated is the sample autocorrelation $\widehat{\rho}(\underline{h})$ as defined in Example 5.3, with $\underline{h} = (1, -1)^T$. For the process from (6.1), the true autocorrelation is given by $\rho(1, -1) = 0.13/1.39$. We approximated the distribution of

$$n\big(\widehat{\rho}(1, -1) - \rho(1, -1)\big) \tag{6.2}$$

for $n = 15$ with a normal approximation and with the AR sieve bootstrap, via the empirical distribution of $n(\widehat{\rho}^*(1, -1) - \widehat{\rho}(1, -1))$. To implement the normal approximation, we considered the limiting distribution of (6.2) given by $\mathcal{N}(0, \tau_X^2)$ with τ_X^2 from (5.4), cf. Example 5.3. For the process $(X_{\underline{t}})$ from (6.1) one can easily verify that τ_X^2 is given by

$$\tau_X^2 = \sum_{|\underline{r}|_\infty \le 2} \big\{2\rho(\underline{r})^2 \rho(1, -1)^2 - \cdots\big\}, \tag{6.3}$$

since all summands with $|\underline{r}|_\infty := \max\{|r_1|, |r_2|\} > 2$ vanish due to $\rho(\underline{r}) = 0$ for all $|\underline{r}|_\infty > 2$. Hence, we estimated τ_X^2 by replacing ρ with $\widehat{\rho}$ in (6.3). It should be noted that this approach represents a best-case scenario for the normal approximation because we used the additional information that for the present data τ_X^2 has the special form (6.3), that is, we chose the optimal point of cutting off the infinite sum in (5.4). For real-world data, this information would not be known, and one would have to estimate τ_X^2 based on equation (5.4) by cutting off the infinite sum at some non-optimal point which would generate an additional error in the estimation.

In Figure 3, the display in the top left corner shows the comparison of three different choices for the order p of the AR sieve bootstrap. We simulated the 95%-quantile of the distribution of (6.2) for $n = 15$. In each iteration, we generated $M = 500$ bootstrap samples to approximate this quantile, subsequently using the AR sieve bootstrap with orders $p = 1$, $p = 2$ and $p = 3$. We also calculated the normal approximation estimate of the quantile in each iteration as described previously. All of this was carried out for $N = 50$ independent iterations to generate boxplots of the locations of the estimates. The three AR sieve approximations with $p = 1, 2, 3$ are shown in the boxplots 1, 2, 3 in the top left display of Figure 3, while the normal approximation values are given boxplot 4. The target value, that is, the 95%-quantile of the distribution of (6.2), is determined from Monte-Carlo simulations with 500 000 repetitions and illustrated by the horizontal dashed line. One can see that the AR sieve bootstrap works very well compared to the normal approximation, even for small orders p and even though the normal approximation is already improved by additional information, as was explained earlier.

In the aforementioned setting, we also compared the performances of the AR sieve bootstrap and block bootstrap techniques (each based on $M = 500$ repetitions). The target was again the 95%-quantile of the distribution of (6.2) for $n = 15$. The order of the AR sieve bootstrap was fixed to $p = 2$ and we considered block sizes of $l = 2, \ldots, 8$. Here, the block size refers to square-shaped blocks, that is, a block size of l means drawing blocks of $l \times l$ observations from the original data sample and then sticking the blocks together to form a sample of size $n \times n$. The result can be seen in the top right corner in Figure 3. Boxplot 1 corresponds to the AR sieve bootstrap and the results for the block bootstrap are given in boxes $2, \ldots, 8$ with block length l depicted in box l. Arguably the best result for the block bootstrap is achieved for $l = 4$; however, the AR sieve bootstrap performs considerably stronger than all block bootstrap approaches implemented here.

In order to show that the results obtained so far are not only specific to the 95%-quantile but to the distribution of (6.2) as a whole, we will now look at an approximation of the variance of this distribution instead of a single quantile. The picture in the bottom left corner of Figure 3 shows these approximations of the variance with all parameters as before. The AR sieve bootstrap ($p = 2$) is depicted in box 1, the block bootstrap in boxes 2 and 3 (block lengths $l = 5, 6$) and the normal approximation in box 4. Similar to what happens for the 95%-quantile, the AR sieve bootstrap outperforms the other methods.

To conclude this section, we modified some of the parameters from the simulations performed so far. The data are still generated by a moving average model, but now following the model equation

$$X_{t_1,t_2} = e_{t_1,t_2} + 4 \cdot e_{t_1+1,t_2} - 5 \cdot e_{t_1-1,t_2} + 3 \cdot e_{t_1,t_2+1} - 2 \cdot e_{t_1,t_2-1}, \tag{6.4}$$

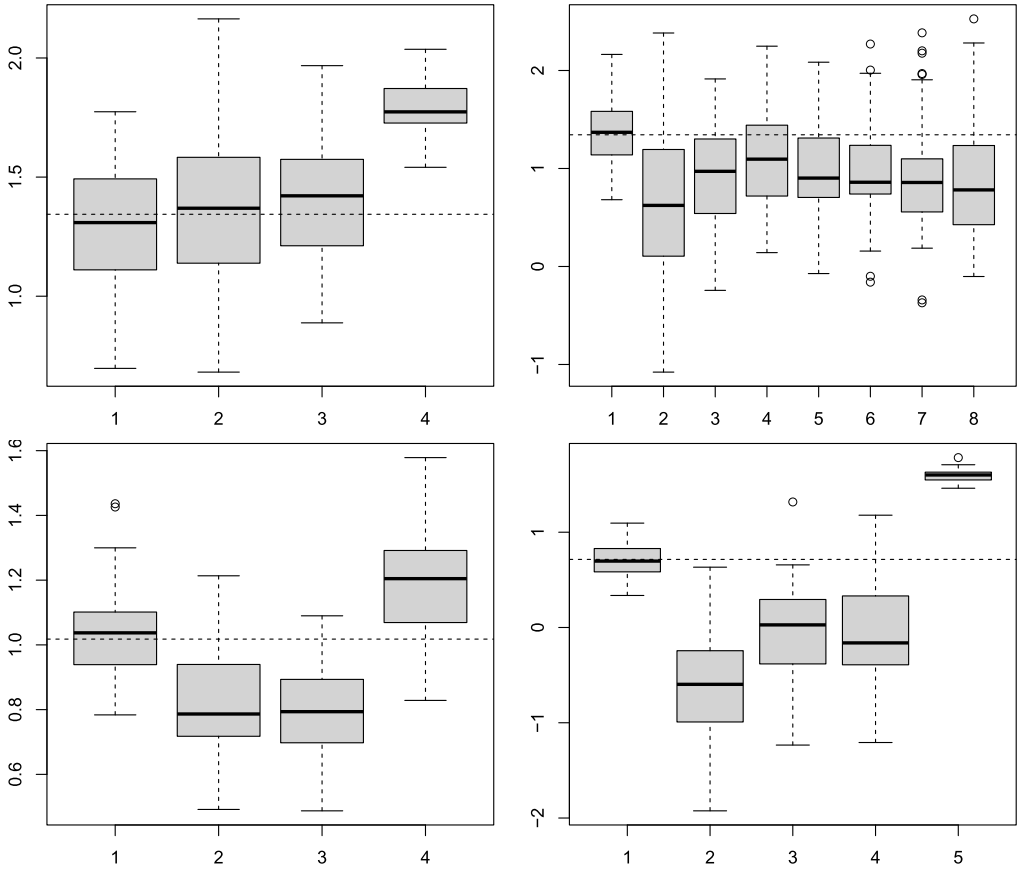

Figure 3. Top left: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n=15$, data generated by model (6.1); boxplots based on $N=50$ iterations. Boxes 1 to 3: AR sieve bootstrap (based on $M=500$ repetitions) with $p=1$, $p=2$ and $p=3$, followed by the normal approximation in box 4. Target value given by the horizontal dashed line. Top right: Same setting as in top left; approximation of the 95%-quantile with the AR sieve bootstrap in box 1, approximations with the block bootstrap and block sizes $l=2,\ldots,8$ in boxes $2,\ldots,8$ (each bootstrap with $M=500$ repetitions). Bottom left: Same setting as in top left; approximation of the variance of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$. Box 1: AR sieve bootstrap with $p=2$. Boxes 2 and 3: Block bootstrap with block sizes $l=5,6$. Box 4: Normal approximation. Bottom right: Approximations of the 95%-quantile of the distribution of $n(\widehat{\rho}(1,-1)-\rho(1,-1))$ for $n=25$, data generated by model (6.4); boxplots based on $N=50$ iterations and each bootstrap method based on $M=300$ repetitions. Box 1: AR sieve bootstrap with $p=4$. Boxes 2, 3, 4: Block bootstrap with block sizes $l=8,9,10$. Box 5: Normal approximation.

where the noise is no longer symmetrically distributed but has an i.i.d. centered exponential distribution. In this model, the dependence of neighbouring random variables is higher than in model (6.1). For example, the true autocorrelation at lag $\underline{h}=(1,-1)^T$ is here given by

$\rho(1,-1) = 0.4$ compared to $\rho(1,-1) \approx 0.094$ in model (6.1). We also increase the sample size to $n = 25$ – corresponding to $25 \times 25 = 625$ observations – and choose the order $p = 4$ for the AR sieve bootstrap. The picture in the bottom right corner of Figure 3 shows the results for the approximation of the 95%-quantile of the distribution of (6.2) for $n = 25$; box 1 shows the AR sieve bootstrap, boxes 2, 3, 4 the block bootstrap with $l = 8, 9, 10$ and box 5 the normal approximation. It can be seen that, for this increased sample size, the normal approximation is close to its limit which, however, differs considerably from the true quantile of the finite sample distribution. This is mainly due to a negative bias for the distribution of (6.2) which can be obtained from the Monte Carlo simulations that were performed to determine the 95%-quantile. The block bootstrap clearly does not show desirable results, this might stem from the increased dependence between neighbouring realisations in the present model compared to the model used previously. However, the AR sieve bootstrap performs very well for this choice of (increased values of) n and p. This emphasizes the fact that convergence of the AR sieve bootstrap can be achieved as long as $p = p(n) \to \infty$ at an appropriate rate.

7. Proofs of the main results

The proof of Theorem 4.2 depends in large parts on some auxiliary results that will be collected in the following lemmas. We will make use of a truncated version $(X^*_{\underline{t},M})$ of the bootstrap process, which is based on the moving average representation of $(X^*_{\underline{t}})$ from (3.9). For arbitrary $M \in \mathbb{N}$ we define

$$X^*_{\underline{t},M} = \sum_{\underline{k} \in \Theta(M)} \widehat{b}_{\underline{k}}(p) \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}}, \tag{7.1}$$

where the finite collection of sites $\Theta(M)$ is defined in (2.2), whereas the non-truncated version $(X^*_{\underline{t}})$ has the infinite collection of sites Θ. Analogously, a truncated version $(\widetilde{X}_{\underline{t},M})$ of the companion process can be defined by replacing Θ with $\Theta(M)$ in (4.2). As a natural extension of the definition of $\mathbf{Y}^*_{\underline{t}}$ and $\widetilde{\mathbf{Y}}_{\underline{t}}$, we denote by

$$\mathbf{Y}^*_{\underline{t},M} := \left(X^*_{\underline{t}+\underline{s}(1),M}, \ldots, X^*_{\underline{t}+\underline{s}(m_1 m_2),M}\right)^T,$$
$$\widetilde{\mathbf{Y}}_{\underline{t},M} := \left(\widetilde{X}_{\underline{t}+\underline{s}(1),M}, \ldots, \widetilde{X}_{\underline{t}+\underline{s}(m_1 m_2),M}\right)^T.$$

With the notations introduced so far we can state the following auxiliary results.

Lemma 7.1. *Let the Assumptions 1–4 be fulfilled with $r = 4$ and h as specified in Assumption 3. Let $\underline{c} \in \mathbb{R}^k$ be arbitrary. Then it holds*:

$$\bullet \sum_{\underline{k} \in \Theta} \left(1 + |\underline{k}|_\infty\right)^2 \left|\widehat{b}_{\underline{k}}(p)\right| = \mathcal{O}_P(1), \tag{7.2}$$

$$\bullet\ E^*\left(\left|\varepsilon^*_{\underline{t}}\right|^{2w}\right) \xrightarrow{P} E\left(|\varepsilon_{\underline{t}}|^{2w}\right) \qquad \forall w \le h+2, \tag{7.3}$$

$$\bullet \ \big(X^*_{\underline{t}_1}, \ldots, X^*_{\underline{t}_d}\big)^T \xrightarrow{d^*} (\widetilde{X}_{\underline{t}_1}, \ldots, \widetilde{X}_{\underline{t}_d})^T \qquad \text{in } P\text{-prob.} \tag{7.4}$$

for all $d \geq 1$ *and all* $\underline{t}_1, \ldots, \underline{t}_d \in \mathbb{Z}^2$,

$$\bullet \ E^*\big(\big|\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big)\big|^{2+2/(h+1)}\big) = \mathcal{O}_P(1), \qquad E\big(\big|\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{t},M})\big|^{2+2/(h+1)}\big) \leq C \tag{7.5}$$

uniformly for all $t \in \mathbb{Z}^2$,

$$\bullet \ \mathrm{Cov}^*\big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{h},M}\big), \underline{c}^T g\big(\mathbf{Y}^*_{\underline{0},M}\big)\big) \xrightarrow{P} \mathrm{Cov}\big(\underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{h},M}), \underline{c}^T g(\widetilde{\mathbf{Y}}_{\underline{0},M})\big) \tag{7.6}$$

for all $\underline{h} \in \mathbb{Z}^2$,

$$\bullet \ \textit{The series } \Sigma^{(u,v)} := \sum_{\underline{h} \in \mathbb{Z}^2} \mathrm{Cov}\big(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}})\big) \textit{ converges} \tag{7.7}$$

absolutely for all $1 \leq u, v \leq k$.

The following auxiliary result will also be used several times.

Lemma 7.2. *Let the Assumptions 1–4 be fulfilled with* $r = 4$. *Let* $W \subset \Theta \cup \{\underline{0}\}$ *be any subset of vectors in the upper half-plane* Θ *or in the origin. We define* $\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}}$ *and* $\mathbf{Y}^{*(W)}_{\underline{t}}$ *to be truncated versions of* $\widetilde{\mathbf{Y}}_{\underline{t}}$ *and* $\mathbf{Y}^*_{\underline{t}}$, *respectively, where*

$$\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}} := \big(\widetilde{X}^{(W)}_{\underline{t}+\underline{s}(1)}, \ldots, \widetilde{X}^{(W)}_{\underline{t}+\underline{s}(m_1 m_2)}\big)^T, \qquad \mathbf{Y}^{*(W)}_{\underline{t}} := \big(X^{*(W)}_{\underline{t}+\underline{s}(1)}, \ldots, X^{*(W)}_{\underline{t}+\underline{s}(m_1 m_2)}\big)^T,$$

and

$$\widetilde{X}^{(W)}_{\underline{t}} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} b_{\underline{k}} \widetilde{\varepsilon}_{\underline{t}-\underline{k}} + \widetilde{\varepsilon}_{\underline{t}} \mathbb{1}_{\{\underline{0} \in W\}}, \qquad X^{*(W)}_{\underline{t}} := \sum_{\underline{k} \in W \setminus \{\underline{0}\}} \widehat{b}_{\underline{k}}(p) \varepsilon^*_{\underline{t}-\underline{k}} + \varepsilon^*_{\underline{t}} \mathbb{1}_{\{\underline{0} \in W\}}.$$

Then there exists $C < \infty$, *such that it holds for any* $\underline{t} \in \mathbb{Z}^2$ *and any* $v = 1, \ldots, k$

$$\big\| g_v(\widetilde{\mathbf{Y}}_{\underline{t}}) - g_v\big(\widetilde{\mathbf{Y}}^{(W)}_{\underline{t}}\big) \big\|_2 \leq C \cdot \Bigg(\sum_{\underline{k} \in \Theta \setminus W} |b_{\underline{k}}| + \mathbb{1}_{\{\underline{0} \notin W\}} \Bigg),$$

$$\big\| g_v\big(\mathbf{Y}^*_{\underline{t}}\big) - g_v\big(\mathbf{Y}^{*(W)}_{\underline{t}}\big) \big\|_{*2} \leq \mathcal{O}_P(1) \cdot \Bigg(\sum_{\underline{k} \in \Theta \setminus W} \big|\widehat{b}_{\underline{k}}(p)\big| + \mathbb{1}_{\{\underline{0} \notin W\}} \Bigg),$$

where $\|z\|_2 := (E(z)^2)^{1/2}$ *and* $\|z\|_{*2} := (E^*(z)^2)^{1/2}$ *denote the usual* L^2*-norms.*

The previous lemma explicitly incorporates the two cases $\underline{0} \in W$ and $\underline{0} \notin W$, both of which will be needed in the proofs of the main results. The proofs of the lemmas from this section can be found in [27].

Proof of Lemma 3.1. As a preliminary consideration, we recall for the vectors $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ from (3.3) and arbitrary $r, s \in \{1, \ldots, \bar{p}\}$

$$\int_{(-\pi,\pi]^2} \exp\bigl(i\,\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle\bigr)\, d\underline{\lambda} = \begin{cases} 4\pi^2, & r = s, \\ 0, & r \neq s, \end{cases} \tag{7.8}$$

because $\underline{k}_r = \underline{k}_s$ if and only if $r = s$. Let $\underline{d} \in \mathbb{R}^{\bar{p}}$ be arbitrary with $\underline{d} \neq \underline{0}$ and denote by $\underline{w}(\underline{\lambda}) := (\exp(i\,\langle \underline{k}_1, \underline{\lambda}\rangle), \ldots, \exp(i\,\langle \underline{k}_{\bar{p}}, \underline{\lambda}\rangle))^T$. Observe that $|\underline{d}^T \underline{w}(\underline{\lambda})|^2 = \sum_{r,s=1}^{\bar{p}} d_r d_s \exp(i\,\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle)$. Using (7.8) as well as $\gamma(\underline{h}) = \int_{(-\pi,\pi]^2} f(\underline{\lambda}) e^{i\,\langle \underline{h}, \underline{\lambda}\rangle}\, d\underline{\lambda}$ and $f(\underline{\lambda}) \geq c > 0$, cf. Assumption 1, we can derive

$$\begin{aligned} \underline{d}^T \Gamma(p) \underline{d} &= \int_{(-\pi,\pi]^2} f(\lambda) \bigl|\underline{d}^T \underline{w}(\underline{\lambda})\bigr|^2 d\underline{\lambda} \\ &\geq c \cdot \int_{(-\pi,\pi]^2} \bigl|\underline{d}^T \underline{w}(\underline{\lambda})\bigr|^2 d\underline{\lambda} \\ &= c \cdot \sum_{r,s=1}^{\bar{p}} d_r d_s \int_{(-\pi,\pi]^2} \exp\bigl(i\,\langle \underline{k}_r - \underline{k}_s, \underline{\lambda}\rangle\bigr)\, d\underline{\lambda} \\ &= 4\pi^2 c \cdot \underline{d}^T \underline{d}. \end{aligned}$$

On the one hand this shows that $\Gamma(p)$ is positive definite and therefore invertible for each $p \in \mathbb{N}$. On the other hand it follows

$$\frac{\underline{d}^T \Gamma(p) \underline{d}}{\underline{d}^T \underline{d}} \geq 4\pi^2 c,$$

which implies for the smallest eigenvalue $\sigma_{\min}(\Gamma(p)) \geq 4\pi^2 c$, cf. [26], 5.2.2 (2). This yields for the largest eigenvalue of the inverse matrix $\sigma_{\max}(\Gamma(p)^{-1}) \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$. The spectral norm of the symmetric matrix $\Gamma(p)^{-1}$ is given by its largest eigenvalue, that is, $\|\Gamma(p)^{-1}\|_{\text{spec}} \leq (4\pi^2 c)^{-1}$ for all $p \in \mathbb{N}$, which yields the desired assertion. □

Proof of Theorem 3.2. In order to write the Yule–Walker equations (3.3) in compact form, we denoted $\bar{p} = 2p(p+1)$ and introduced the arbitrary but fixed enumeration $\underline{k}_1, \ldots, \underline{k}_{\bar{p}}$ of the vectors $\underline{k} \in \Theta(p)$. Now we extend this enumeration to the infinite but countable set Θ, by choosing an arbitrary enumeration $\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \ldots$ of the vectors $\underline{k} \in \Theta \setminus \Theta(p)$ such that

$$\Theta = \{\underline{k}_1, \ldots, \underline{k}_{\bar{p}}\} \cup \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \ldots\}.$$

While the finite predictor coefficients $(a_{\underline{k}}(p))_{\underline{k} \in \Theta(p)}$ are given by (3.3), Lemma 2.1 shows that the autoregressive coefficients $(a_{\underline{k}})_{\underline{k} \in \Theta}$ determine the L^2-projection of $X_{\underline{t}}$ onto $\overline{\text{sp}}\{X_{\underline{t}-\underline{k}} : \underline{k} \in \Theta\}$. Therefore, $X_{\underline{t}} - \sum_{\underline{k} \in \Theta} a_{\underline{k}} X_{\underline{t}-\underline{k}}$ is orthogonal to each $X_{\underline{t}-\underline{s}}$, $\underline{s} \in \Theta$. Equivalently, with the

introduced enumeration of Θ this means

$$\operatorname{Cov}\left(X_{\underline{t}} - \sum_{j=1}^{\infty} a_{\underline{k}_j} X_{\underline{t}-\underline{k}_j}, X_{\underline{t}-\underline{k}_m}\right) = \gamma(\underline{k}_m) - \sum_{j=1}^{\infty} a_{\underline{k}_j} \gamma(\underline{k}_m - \underline{k}_j) = 0 \qquad \forall m \in \mathbb{N}.$$

From this system of equations, we consider only those ones with $m = 1, \ldots, \bar{p}$, which is equivalent to

$$\Gamma(p) \cdot \begin{pmatrix} a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}} \end{pmatrix} + \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix} = \begin{pmatrix} \gamma(\underline{k}_1) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}}) \end{pmatrix}.$$

Since the right-hand sides of this system and (3.3) coincide, we can infer

$$\begin{pmatrix} a_{\underline{k}_1}(p) - a_{\underline{k}_1} \\ \vdots \\ a_{\underline{k}_{\bar{p}}}(p) - a_{\underline{k}_{\bar{p}}} \end{pmatrix} = \Gamma(p)^{-1} \cdot \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \begin{pmatrix} \gamma(\underline{k}_1 - \underline{k}_j) \\ \vdots \\ \gamma(\underline{k}_{\bar{p}} - \underline{k}_j) \end{pmatrix}. \tag{7.9}$$

In the following we will denote the (n, r)th entry of $\Gamma(p)^{-1}$ by $(\Gamma(p)^{-1})^{(n,r)}$. We are interested in a weighted sum of the absolute values of the entries on the left-hand side of (7.9). For $s \in \mathbb{N}_0$ such that $s + 1 < r$, we get

$$\begin{aligned} &\sum_{n=1}^{\bar{p}} \bigl(1 + |\underline{k}_n|_\infty\bigr)^s \bigl|a_{\underline{k}_n}(p) - a_{\underline{k}_n}\bigr| \\ &\quad = \sum_{n=1}^{\bar{p}} \bigl(1 + |\underline{k}_n|_\infty\bigr)^s \left| \sum_{j=\bar{p}+1}^{\infty} a_{\underline{k}_j} \sum_{r=1}^{\bar{p}} \bigl(\Gamma(p)^{-1}\bigr)^{(n,r)} \gamma(\underline{k}_r - \underline{k}_j) \right| \\ &\quad \le \sum_{j=\bar{p}+1}^{\infty} |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} \bigl|\gamma(\underline{k}_r - \underline{k}_j)\bigr| \max_{r=1,\ldots,\bar{p}} \sum_{n=1}^{\bar{p}} \bigl(1 + |\underline{k}_n|_\infty\bigr)^s \bigl|\bigl(\Gamma(p)^{-1}\bigr)^{(n,r)}\bigr|. \end{aligned} \tag{7.10}$$

We denote the max-column-sum norm of an arbitrary $n \times n$-matrix B by $\|B\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |B^{(i,j)}|$. It is well known that $\|\cdot\|_1$ is submultiplicative which allows us to derive

$$\begin{aligned} &\max_{r=1,\ldots,\bar{p}} \sum_{n=1}^{\bar{p}} \bigl(1 + |\underline{k}_n|_\infty\bigr)^s \bigl|\bigl(\Gamma(p)^{-1}\bigr)^{(n,r)}\bigr| \\ &\quad = \bigl\|\operatorname{diag}\bigl[\bigl(1 + |\underline{k}_1|_\infty\bigr)^s, \ldots, \bigl(1 + |\underline{k}_{\bar{p}}|_\infty\bigr)^s\bigr] \cdot \Gamma(p)^{-1}\bigr\|_1 \\ &\quad \le \max_{n=1,\ldots,\bar{p}} \bigl(1 + |\underline{k}_n|_\infty\bigr)^s \cdot \bigl\|\Gamma(p)^{-1}\bigr\|_1. \end{aligned}$$

Hence, (7.10) can be bounded from above by

$$\begin{aligned}&\big\|\Gamma(p)^{-1}\big\|_1 \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\dots,\bar{p}} \big(1+|\underline{k}_n|_\infty\big)^s |a_{\underline{k}_j}| \sum_{r=1}^{\bar{p}} \big|\gamma(\underline{k}_r - \underline{k}_j)\big| \\ &\leq \big\|\Gamma(p)^{-1}\big\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} \big|\gamma(\underline{k})\big| \cdot \sum_{j=\bar{p}+1}^{\infty} \max_{n=1,\dots,\bar{p}} \big(1+|\underline{k}_n|_\infty\big)^s |a_{\underline{k}_j}|.\end{aligned}$$

Since our numeration was chosen such that $\Theta(p) = \{\underline{k}_1, \dots, \underline{k}_{\bar{p}}\}$ and $\Theta \setminus \Theta(p) = \{\underline{k}_{\bar{p}+1}, \underline{k}_{\bar{p}+2}, \dots\}$, the inequality derived so far reads

$$\begin{aligned}&\sum_{\underline{k}\in\Theta(p)} \big(1+|\underline{k}|_\infty\big)^s \big|a_{\underline{k}}(p) - a_{\underline{k}}\big| \\ &\quad\leq \big\|\Gamma(p)^{-1}\big\|_1 \sum_{\underline{k}\in\mathbb{Z}^2} \big|\gamma(\underline{k})\big| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \max_{\underline{v}\in\Theta(p)} \big(1+|\underline{v}|_\infty\big)^s |a_{\underline{k}}|.\end{aligned} \tag{7.11}$$

Per definition of $\Theta(p)$, we have

$$\max_{\underline{v}\in\Theta(p)} \big(1+|\underline{v}|_\infty\big)^s = (1+p)^s \leq \big(1+|\underline{k}|_\infty\big)^s \qquad \forall \underline{k} \in \Theta \setminus \Theta(p), \tag{7.12}$$

as $|\underline{k}|_\infty \geq p+1$ for all $\underline{k} \in \Theta\setminus\Theta(p)$; this is why we need a weight function strictly nondecreasing in $|\underline{k}|_\infty$. Furthermore, it holds $\|A\|_1 \leq \sqrt{n}\|A\|_{\text{spec}}$ for all $n \times n$-matrices A, i.e. $\|\Gamma(p)^{-1}\|_1 \leq \sqrt{2p(p+1)}\|\Gamma(p)^{-1}\|_{\text{spec}}$ and

$$\sqrt{2p(p+1)} \leq \sqrt{2}(p+1) < \sqrt{2}\big(1+|\underline{k}|_\infty\big) \qquad \forall \underline{k} \in \Theta \setminus \Theta(p).$$

Therefore, and due to (7.12) and Lemma 3.1, (7.11) can be bounded by

$$\begin{aligned}&\sqrt{2p(p+1)}\big\|\Gamma(p)^{-1}\big\|_{\text{spec}} \sum_{\underline{k}\in\mathbb{Z}^2} \big|\gamma(\underline{k})\big| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \big(1+|\underline{k}|_\infty\big)^s |a_{\underline{k}}| \\ &\leq \frac{1}{2\sqrt{2}\pi^2 c} \sum_{\underline{k}\in\mathbb{Z}^2} \big|\gamma(\underline{k})\big| \cdot \sum_{\underline{k}\in\Theta\setminus\Theta(p)} \big(1+|\underline{k}|_\infty\big)^{s+1} |a_{\underline{k}}|,\end{aligned}$$

which completes the proof. □

Proof of Theorem 4.2. The basic structure of this proof resembles the one of Theorem 3.3 in [9]. Since the proof is lengthy and technical, we will restrict ourselves to give a sketch of the proof in this paper. The enhanced version can be found in [27]. At first, we will neglect the outer function f in T_n^* and show for the bootstrap quantities

$$(\bar{n}_1\bar{n}_2)^{-1/2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} \big(g\big(\mathbf{Y}_{\underline{t}}^*\big) - E^*\big(g\big(\mathbf{Y}_{\underline{t}}^*\big)\big)\big) \xrightarrow{d^*} \mathcal{N}(\underline{0}, \Sigma) \qquad \text{in prob.}, \tag{7.13}$$

where the entries of Σ are given by $\Sigma^{(u,v)} := \sum_{\underline{h}\in\mathbb{Z}^2} \operatorname{Cov}(g_u(\widetilde{\mathbf{Y}}_{\underline{h}}), g_v(\widetilde{\mathbf{Y}}_{\underline{0}}))$, for $u, v = 1, \ldots, k$. Since the companion process $(\widetilde{X}_{\underline{t}})$, just as the bootstrap process, is a linear spatial process (recall that the innovations $(\widetilde{\varepsilon}_{\underline{t}})$ are i.i.d.), one can follow the exact same arguments as in proving (7.13) to derive

$$(\bar{n}_1\bar{n}_2)^{-1/2} \sum_{t_1=1}^{\bar{n}_1} \sum_{t_2=1}^{\bar{n}_2} \big(g(\widetilde{\mathbf{Y}}_{\underline{t}}) - E\big(g(\widetilde{\mathbf{Y}}_{\underline{t}})\big)\big) \xrightarrow{d} \mathcal{N}(\underline{0}, \Sigma) \tag{7.14}$$

with the very same limiting distribution as above. In the end, we will incorporate the function f by applying the delta method to both CLT's which will complete the proof of Theorem 4.2 since $(\bar{n}_1\bar{n}_2)^{1/2}$ and n are asymptotically equivalent.

The strategy for proving (7.13) is the following: For arbitrary, but fixed $M \in \mathbb{N}$, we consider the truncated quantity $(\mathbf{Y}^*_{\underline{t},M})$ based on $(X^*_{\underline{t},M})$ from (7.1) and invoke the Cramér–Wold device by showing

$$\frac{1}{n} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big) - E^*\big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big)\big)\big) \xrightarrow{d^*} \mathcal{N}\big(0, \underline{c}^T \Sigma_M \underline{c}\big) \qquad \text{in prob.,} \tag{7.15}$$

for arbitrary $\underline{c} \in \mathbb{R}^k$, where

$$\Sigma_M^{(u,v)} := \sum_{h_1=-2M-m_1+1}^{2M+m_1-1} \sum_{h_2=-M-m_2+1}^{M+m_2-1} \operatorname{Cov}\big(g_u(\widetilde{\mathbf{Y}}_{\underline{h},M}), g_v(\widetilde{\mathbf{Y}}_{\underline{0},M})\big).$$

The conditional variance of the left-hand side of (7.15), subsequently abbreviated by v_n^*, converges to $\underline{c}^T \Sigma_M \underline{c}$ in probability due to straightforward (yet tedious) calculations, using strict stationarity of $(\underline{c}^T g(\mathbf{Y}^*_{\underline{t},M}))$ and the obvious fact that $g(\mathbf{Y}^*_{\underline{h},M})$ and $g(\mathbf{Y}^*_{\underline{0},M})$ are independent whenever $|h_1| \geq 2M + m_1$ or $|h_2| \geq M + m_2$. Hence, showing (7.15) reduces to showing

$$\frac{1}{n\sqrt{v_n^*}} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big) - E^*\big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big)\big)\big) \xrightarrow{d^*} \mathcal{N}(0, 1) \qquad \text{in prob.} \tag{7.16}$$

Next, we use a blocking technique: Define appropriate sequences of integers $a(n), b(n) \in \mathbb{N}$ with $a(n) \to \infty$, $b(n) \to \infty$ such that $b(n)/a(n) \to 0$ and $N(n) := n/(a(n)+b(n)) \to \infty$ as $n \to \infty$. The $n \times n$ summands in (7.16) are decomposed into dominating, square-shaped blocks A_{j_1,j_2} of size $a(n) \times a(n)$, and negligible, asymptotically vanishing remainder terms B_{j_1,j_2}. The left-hand side of (7.16) then equals

$$\frac{1}{n\sqrt{v_n^*}} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} (A_{j_1,j_2} + B_{j_1,j_2}), \tag{7.17}$$

where, with $a := a(n)$,

$$A_{j_1,j_2} := \sum_{t_1=(j_1-1)(a+b)+1}^{j_1a+(j_1-1)b} \sum_{t_2=(j_2-1)(a+b)+1}^{j_2a+(j_2-1)b} \big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big) - E^*\big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big)\big)\big).$$

One can easily check that all blocks A_{j_1,j_2} are independent as soon as $b(n) > M$. Hence, Lindeberg's CLT can be applied since the Lyapunov condition

$$\frac{1}{\tau_N^{2+\delta}} \sum_{j_1=1}^{N} \sum_{j_2=1}^{N} E^*\big(|A_{j_1,j_2}|^{2+\delta}\big) = \frac{N^2}{\tau_N^{2+\delta}} a(n)^{2+\delta} \mathcal{O}_P(1) = o_P(1)$$

can be verified for $\tau_N := (\sum_{j_1=1}^{N} \sum_{j_2=1}^{N} \mathrm{Var}^*(A_{j_1,j_2}))^{1/2}$. This yields (7.16) and therefore (7.15). The proof is then completed by using Lemma 7.2 to show

$$\mathrm{Var}^*\left(\frac{1}{n} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \big(\underline{c}^T g\big(\mathbf{Y}^*_{\underline{t}}\big) - \underline{c}^T g\big(\mathbf{Y}^*_{\underline{t},M}\big)\big)\right) \leq \frac{1}{M^2} \mathcal{O}_P(1), \tag{7.18}$$

since this condition is sufficient for Proposition 6.3.9 of [8] to hold, which states that (7.14) implies (7.13). □

Acknowledgements

We thank two anonymous referees for their thorough inspection of the present paper – their comments are much appreciated and helped to improve the quality of this work. The authors gratefully acknowledge financial support by the German Research Foundation (DFG). The research of Marco Meyer was partly supported by the Research Training Group RTG 1953 and the position of Carsten Jentsch was financed by the Research Center 884 "Political Economy of Reform" (Project B6). We also thank Efstathios Paparoditis, Mohsen Pourahmadi and Martin Schlather for interesting discussions and helpful comments.

References

[1] Alonso, A.M., Peña, D. and Romo, J. (2002). Forecasting time series with sieve bootstrap. *J. Statist. Plann. Inference* **100** 1–11. MR1869807

[2] Basu, S. and Reinsel, G.C. (1992). A note on properties of spatial Yule–Walker estimators. *J. Stat. Comput. Simul.* **41** 243–255. MR1276185

[3] Basu, S. and Reinsel, G.C. (1993). Properties of the spatial unilateral first-order ARMA model. *Adv. in Appl. Probab.* **25** 631–648. MR1234300

[4] Baxter, G. (1962). An asymptotic result for the finite predictor. *Math. Scand.* **10** 137–144. MR0149584

[5] Baxter, G. (1963). A norm inequality for a "finite-section" Wiener–Hopf equation. *Illinois J. Math.* **7** 97–103. MR0145285

[6] Besag, J.E. (1972). On the correlation structure of some two-dimensional stationary processes. *Biometrika* **59** 43–48. MR0315775
[7] Bhatt, S.J. and Dedania, H.V. (2003). Beurling algebra analogues of the classical theorems of Wiener and Lévy on absolutely convergent Fourier series. *Proc. Indian Acad. Sci. Math. Sci.* **113** 179–182. MR1983751
[8] Brockwell, P.J. and Davis, R.A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. New York: Springer. MR1093459
[9] Bühlmann, P. (1997). Sieve bootstrap for time series. *Bernoulli* **3** 123–148. MR1466304
[10] Chang, Y. and Park, J.Y. (2003). A sieve bootstrap for the test of a unit root. *J. Time Series Anal.* **24** 379–400. MR1997120
[11] Cheng, R. and Pourahmadi, M. (1993). Baxter's inequality and convergence of finite predictors of multivariate stochastic processes. *Probab. Theory Related Fields* **95** 115–124. MR1207310
[12] Choi, B. and Politis, D.N. (2007). Modeling 2-D AR processes with various regions of support. *IEEE Trans. Signal Process.* **55** 1696–1707. MR2472330
[13] Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York: Wiley. MR1239641
[14] Freedman, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842. MR0751275
[15] Gröchenig, K. (2007). Weight functions in time-frequency analysis. In *Pseudo-Differential Operators*: *Partial Differential Equations and Time-Frequency Analysis* (L. Rodino et al., eds.) *Fields Inst. Commun.* **52** 343–366. Providence, RI: Amer. Math. Soc. MR2385335
[16] Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika* **69** 95–105. MR0655674
[17] Guyon, X. (1995). *Random Fields on a Network – Modeling*, *Statistics*, *and Applications*. New York: Springer. MR1344683
[18] Ha, E. and Newton, H.J. (1993). The bias of estimators of causal spatial autoregressive processes. *Biometrika* **80** 242–245. MR1225229
[19] Hall, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR0808159
[20] Hannan, E.J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. New York: Wiley. MR0940698
[21] Helson, H. and Lowdenslager, D. (1958). Prediction theory and Fourier series in several variables. *Acta Math.* **99** 165–202. MR0097688
[22] Kreiss, J.-P. (1988). Asymptotical inference for a class of stochastic processes. Habilitationsschrift, Universität Hamburg.
[23] Kreiss, J.-P. (1992). Bootstrap procedures for AR(∞)-processes. In *Bootstrapping and Related Techniques* (*Trier*, 1990) (K.H. Jöckel, G. Rothe and W. Sender, eds.) *Lecture Notes in Econom. and Math. Systems* **376** 107–113. Berlin: Springer. MR1238505
[24] Kreiss, J.-P., Paparoditis, E. and Politis, D.N. (2011). On the range of validity of the autoregressive sieve bootstrap. *Ann. Statist.* **39** 2103–2130. MR2893863
[25] Künsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147
[26] Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley. MR1433592
[27] Meyer, M., Jentsch, C. and Kreiss, J.-P. (2015). Baxter's inequality and sieve bootstrap for random fields. Working Paper ECON 15–06, Department of Economics, University of Mannheim, Germany. Available at http://www2.vwl.uni-mannheim.de/3032.1.html.
[28] Meyer, M. and Kreiss, J.-P. (2015). On the vector autoregressive sieve bootstrap. *J. Time Series Anal.* **36** 377–397. MR3343007
[29] Nordman, D.J. and Lahiri, S.N. (2004). On optimal spatial subsample size for variance estimation. *Ann. Statist.* **32** 1981–2027. MR2102500

[30] Nordman, D.J., Lahiri, S.N. and Fridley, B.L. (2007). Optimal block size for variance estimation by a spatial block bootstrap method. *Sankhyā* **69** 468–493. MR2460005

[31] Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *J. Multivariate Anal*. **57** 277–296. MR1391173

[32] Paparoditis, E. and Politis, D.N. (2005). Bootstrapping unit root tests for autoregressive time series. *J. Amer. Statist. Assoc*. **100** 545–553. MR2160558

[33] Paparoditis, E. and Streitberg, B. (1992). Order identification statistics in stationary autoregressive moving-average models: Vector autocorrelations and the bootstrap. *J. Time Series Anal*. **13** 415–434. MR1183155

[34] Politis, D.N. and Romano, J.P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal*. **47** 301–328. MR1247380

[35] Pourahmadi, M. (2001). *Foundations of Time Series Analysis and Prediction Theory*. New York: Wiley. MR1849562

[36] Smeekes, S. and Urbain, J.-P. (2013). On the applicability of the sieve bootstrap in time series panels. *Oxford Bulletin of Economics and Statistics* **76** 139–151.

[37] Tjøstheim, D. (1978). Statistical spatial series modelling. *Adv. in Appl. Probab.* **10** 130–154. MR0471224

[38] Tjøstheim, D. (1981). Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Trans. Geosci. Remote Sens*. **19** 15–24.

[39] Tjøstheim, D. (1983). Statistical spatial series modelling. II. Some further results on unilateral lattice processes. *Adv. in Appl. Probab.* **15** 562–584. MR0706617

[40] Whittle, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. MR0067450

[41] Zhu, J. and Lahiri, S.N. (2007). Bootstrapping the empirical distribution function of a spatial process. *Stat. Inference Stoch. Process*. **10** 107–145. MR2273326

[42] Zygmund, A. (2002). *Trigonometric Series. Vol. I, II*, 3rd ed. *Cambridge Mathematical Library*. Cambridge: Cambridge Univ. Press. MR1963498

Received June 2015 and revised February 2016