Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Gaussian Approximation and Spatially Dependent Wild Bootstrap for High-Dimensional Spatial Data

Daisuke Kurisu, Kengo Kato & Xiaofeng Shao

To cite this article: Daisuke Kurisu, Kengo Kato & Xiaofeng Shao (03 Jul 2023): Gaussian Approximation and Spatially Dependent Wild Bootstrap for High-Dimensional Spatial Data, Journal of the American Statistical Association, DOI: 10.1080/01621459.2023.2218578

To link to this article: https://doi.org/10.1080/01621459.2023.2218578

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2023, VOL. 00, NO. 0, 1–13: Theory and Methods
https://doi.org/10.1080/01621459.2023.2218578

Taylor & Francis
Taylor & Francis Group

Check for updates

Gaussian Approximation and Spatially Dependent Wild Bootstrap for High-Dimensional Spatial Data

Daisuke Kurisu[a], Kengo Kato[b], and Xiaofeng Shao[c]

[a]The University of Tokyo, Tokyo, Japan; [b]Cornell University, Ithaca, NY; [c]University of Illinois at Urbana-Champaign, Urbana, IL

ABSTRACT
In this article, we establish a high-dimensional CLT for the sample mean of p-dimensional spatial data observed over irregularly spaced sampling sites in $\mathbb{R}^d$, allowing the dimension p to be much larger than the sample size n. We adopt a stochastic sampling scheme that can generate irregularly spaced sampling sites in a flexible manner and include both pure increasing domain and mixed increasing domain frameworks. To facilitate statistical inference, we develop the spatially dependent wild bootstrap (SDWB) and justify its asymptotic validity in high dimensions by deriving error bounds that hold almost surely conditionally on the stochastic sampling sites. Our dependence conditions on the underlying random field cover a wide class of random fields such as Gaussian random fields and continuous autoregressive moving average random fields. Through numerical simulations and a real data analysis, we demonstrate the usefulness of our bootstrap-based inference in several applications, including joint confidence interval construction for high-dimensional spatial data and change-point detection for spatio-temporal data. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received September 2021
Accepted May 2023

KEYWORDS
Change-point analysis; High-dimensional central limit theorem; Irregularly spaced spatial data; Spatio-temporal data

1. Introduction

Spatial data analysis plays an important role in many fields, such as atmospheric science, climate studies, ecology, hydrology and seismology. There are many classic textbooks and monographs devoted to modeling and inference of spatial data, see, for example, Cressie (1993), Stein (1999), Moller and Waagepetersen (2004), Gaetan and Guyan (2010), and Banerjee, Carlin, and Gelfand (2015), among others. This article aims to advance high-dimensional Gaussian approximation theory and bootstrap-based methodology related to the analysis of multivariate (and possibly high-dimensional) spatial data. Specifically, we assume that our data are from a multivariate random field $\boldsymbol{Y} = \{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with $\boldsymbol{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), \dots, Y_p(\boldsymbol{s}))'$, where $d \geqslant 2$ is the dimension of the spatial domain and $p \geqslant 2$ stands for the dimension of multivariate measurements at any location $\boldsymbol{s} \in \mathbb{R}^d$.

With recent technological advances and remote sensing technology, multivariate spatial data are becoming more prevalent. For example, levels of multiple air pollutants (e.g., ozone, $PM_{2.5}$, PM_{10}, nitric oxide, carbon monoxide) are monitored at many stations in many countries. To understand spatial distributions of carbon intake and emissions as well as their seasonal and annual evolutions, the total-column carbon dioxide (CO2) mole fractions (in units of parts per million) are measured using remote sensing instruments, which produce estimates of CO2 concentration, called profiles, at 20 different pressure levels; see Nguyen, Cressie, and Braverman (2017). The latter authors treated 20 measurements at different profiles as a 20-dimensional vector and proposed a modified spatial random effect model to capture spatial dependence and multivariate dependence across profiles. For an early literature on the modeling and inference of multivariate spatial data, we refer to Gelfand and Vounatsou (2003), Gelfand et al. (2004), and Gelfand and Banerjee (2010).

Motivated by the increasing availability of multivariate spatial data with increasing dimensions, we shall study a fundamental problem at the intersection of spatial statistics and high-dimensional statistics: central limit theorem (CLT) for the sample mean of high-dimensional spatial data observed at irregularly spaced sampling sites. When the dimension p is low and fixed, CLTs for weighted sums of spatial data have been derived when the sampling sites lie on the d-dimensional integer lattice, see Bulinskii and Zhurbenko (1976), Bolthausen (1982), and Guyon and Richardson (1984). To accommodate irregularly spaced sampling sites, which is the norm rather than the exception in spatial statistics, Lahiri (2003a) introduced a novel stochastic sampling design and derived CLTs under both pure increasing domain and mixed increasing domain settings. However, so far, all these results are restricted to the case when the dimension p is fixed, and there seem no CLT results that allow for the growing dimension in the literature.

To address the high-dimensional CLT for spatial data, we face the following challenges: (a) when the dimension p exceeds the sample size n, even for iid data, it is usually not known whether the distribution of normalized sample mean (or its norm, such as the ℓ^∞-norm) converges to a fixed limit, unless under very stringent assumptions; (b) spatial data have no

CONTACT Daisuke Kurisu daisukekurisu@csis.u-tokyo.ac.jp The University of Tokyo, Tokyo, Japan.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

natural ordering and sampling sites are often irregularly spaced. In the low-dimensional setting, Lahiri (2003a) showed that the asymptotic variance depends on the sampling density, and the convergence rate for the sample mean depends on which asymptotic regime we adopt (pure increasing-domain vs. mixed increasing domain). To meet the first challenge, we shall build on the celebrated high-dimensional Gaussian approximation techniques that have undergone recent rapid development (see a literature review below) and establish the asymptotic equivalence between the distribution of the normalized sample mean and that of its Gaussian counterpart in high dimensions. To tackle the challenge from the irregular spatial spacing, we shall adopt the stochastic sampling scheme of Lahiri (2003a), which allows the sampling sites to have a nonuniform density across the sampling region and enables the number of sampling sites n to grow at a different rate compared with the volume of the sampling region λ_n^d. This scheme accommodates both *the pure increasing domain* case ($\lim_{n\to\infty} n\lambda_n^{-d} = \kappa \in (0,\infty)$) and *the mixed increasing domain* case ($\lim_{n\to\infty} n\lambda_n^{-d} = \infty$). From a theoretical viewpoint, this scheme covers all possible asymptotic regimes since it is well-known that the sample mean is not consistent under the infill asymptotics (Lahiri 1996). See Lahiri (2003b), Lahiri and Zhu (2006), and Bandyopadhyay, Lahiri, and Nordman (2015) for some detailed discussions on the stochastic spatial sampling design.

Specifically, we establish a CLT for the sample mean of high-dimensional spatial data over the rectangles when $p = p_n \to \infty$ as $n \to \infty$ and possibly $p \gg n$ under a weak dependence condition, where the random field is observed at a finite number of discrete locations $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$ in a sampling region R_n whose volume scales as λ_n^d, where $\lambda_n \to \infty$ as $n \to \infty$. To facilitate statistical inference, we propose and develop the spatially dependent wild bootstrap (SDWB, hereafter), which is an extension of the dependent wild bootstrap of Shao (2010) to the spatial setting, and justify its asymptotic validity in high dimensions. Notably, we will show that the SDWB works for a wide class of random fields on $\mathbb{R}^d$ that includes multivariate Lévy-driven moving average (MA) random fields (see Kurisu (2022) for a detailed discussion of such random fields). Lévy-driven MA random fields constitute a rich class of models for spatial data and include both Gaussian and non-Gaussian random fields such as continuous autoregressive moving-average (CARMA) random fields (Brockwell and Matsuda 2017). To illustrate the usefulness of our theory and SDWB, we describe several applications, including (i) simultaneous inference for the mean vector of high-dimensional spatial data; (ii) construction of confidence bands for the mean function of spatio-temporal data, and (iii) multiple change-point detection for spatio-temporal data. In particular, we demonstrate in Appendix J that the change-point detection methodology we develop on the basis of Gaussian approximation and SDWB can be applied to detect the mean change in time for a spatio-temporal dataset. Our method differs from most of the existing change-point detection approaches developed for spatio-temporal data and high-dimensional data in terms of applicability and underlying assumptions; see Remark C.1 in Appendix C for a detailed comparison.

Contributions and Connections to the literature. To put our contributions in perspective, we shall review two lines of research that have inspired our work. The first line is related to *Gaussian approximation for both high-dimensional independent data and high-dimensional time series*. There is now a large and still rapidly growing literature on high-dimensional CLTs over the rectangles and related bootstrap theory when the dimension of the data is possibly much larger than the sample size. For the sample mean of independent random vectors, we refer to Chernozhukov, Chetverikov, and Kato (2013), Chernozhukov et al. (2014), Chernozhukov, Chetverikov, and Kato (2015), Chernozhukov, Chetverikov, and Kato (2016), Chernozhukov, Chetverikov, and Kato (2017), Chernozhukov et al. (2022), Deng and Zhang (2020), Kuchibhotla, Mukherjee, and Banerjee (2021), Fang and Koike (2020), and Chernozhukov, Chetverikov, and Koike (2020). For high-dimensional U-statistics and U-processes, see Chen (2018) and Chen and Kato (2019). In the time series setting, Zhang and Wu (2017) developed Gaussian approximation for the maximum of the sample mean of high-dimensional stationary time series with equidistant observations under the physical dependence measures developed by Wu (2005). Based on a nonparametric estimator for the long-run covariance matrix of the sample mean, they used a simulation-based approach to constructing simultaneous confidence intervals for the mean vector. Zhang and Cheng (2018) also developed high-dimensional CLTs for the maximum of the sample mean of high-dimensional time series under the physical/functional dependence measures and used nonoverlapping block bootstrap to perform inference. Chernozhukov, Chetverikov, and Kato (2019) studied high-dimensional CLTs for the maximum of the sum of β-mixing and possibly nonstationary time series and showed the asymptotic validity of a block multiplier bootstrap method. Also see Chang, Yao, and Zhou (2017), Koike (2019), and Yu and Chen (2021) among others for the use of Gaussian approximation or variants in high-dimensional testing problems.

To the best of our knowledge, our work is the first paper that establishes a high-dimensional CLT for spatial data and rigorously justifies the asymptotic validity of a bootstrap method for high-dimensional data in the spatial setting. From a technical point of view, the present paper builds on Chernozhukov, Chetverikov, and Kato (2017), Chernozhukov, Chetverikov, and Kato (2019), Zhang and Wu (2017), and Zhang and Cheng (2018), but our theoretical analysis differs substantially from those references in several important aspects. Specifically, (i) we establish a high-dimensional CLT and the asymptotic validity of SDWB that hold almost surely conditionally on the stochastic sampling sites. Precisely, we show that the conditional distribution of the sample mean (or its SDWB counterpart) given the sampling sites can be approximated by a (conditionally) Gaussian distribution. The randomness of the sampling sites yields additional technical complications in high dimensions; see for example, Lemma D.2 in Appendix D. (ii) We extend the coupling technique used in Yu (1994) for irregularly spatial data to prove the high-dimensional CLT. This extension is nontrivial since there is no natural ordering for spatial data and the number of observations in each block constructed is random. Our approach to the blocking construction is also quite different from those in Lahiri (2003b) and Lahiri and Zhu (2006) whose proofs essentially rely on approximating the characteristic function of the weighted sample mean by that of independent blocks; see also Remark 2.1. (iii) We explore in detail concrete random fields that satisfy our weak dependence condition and other regularity conditions. Indeed, we show that our regularity conditions can be satisfied for a wide class of multivariate Lévy-driven MA random fields that constitute a rich class of models for spatial data (see Brockwell and Matsuda 2017) but whose mixing properties have not been investigated so far. Verification of our regularity conditions to Lévy-driven MA fields is indeed nontrivial and relies on several probabilistic techniques from Lévy process theory and theory of infinitely divisible random measures (Sato 2006).

Our work also builds on the literature of *bootstrap methods for time series and spatial data*. For both time series and spatial data, the block-based bootstrap (BBB) has been fairly well studied since the introduction of moving block bootstrap (MBB) by Künsch (1989) and Liu and Singh (1992). Among many variants of the MBB, we mention Carlstein (1986) for the nonoverlapping block bootstrap, Politis and Romano (1992) for the circular bootstrap, Politis and Romano (1994) for the stationary bootstrap, and Paparoditis and Politis (2001, 2002) for the tapered block bootstrap. See Lahiri (2003b) for a book-length treatment. The BBB methods have been extended to spatial framework for both regular lattice and irregularly spaced non-lattice data. See for example, Politis and Romano (1993), Politis, Paparoditis, and Romano (1999), Lahiri and Zhu (2006), and Zhu and Lahiri (2007).

As we mentioned before, the proposed SDWB is an extension of the dependent wild bootstrap in Shao (2010), which was developed for time series data. The main difference between SDWB and DWB is that the SDWB observations in $\mathbb{R}^d$ are generated by simulating an auxiliary random field with suitable covariance function on $\mathbb{R}^d$ to mimic the spatial dependence. In contrast, the DWB in Shao (2010) aims to capture temporal dependence when $d = 1$. Thus, the multipliers (or external variables) in SDWB are spatially dependent, hence, the name SDWB. As argued in Shao (2010), the DWB/SDWB is much easier to implement for irregularly spaced data than BBB, as the latter requires partitioning the sampling region into blocks and can be less convenient to implement due to incomplete blocks. The SDWB is also different from the block multiplier bootstrap (BMB) proposed in Chernozhukov, Chetverikov, and Kato (2019) since the multipliers of the BMB are iid Gaussian random variables, while the multipliers of SDWB are dependent Gaussian random variables generated from a stationary Gaussian random field on the irregular spaced sampling sites.

Compared to other bootstrap methods and associated theory developed for spatial data (Lahiri and Zhu 2006; Zhu and Lahiri 2007), our bootstrap-based inference targets at a high-dimensional parameter and our theoretical argument is substantially different. Inference for a high-dimensional parameter related to spatial data is expected to grow due to the increasing need for the analysis of multivariate spatial data and spatio-temporal data that takes into account the effect of dimension and spatial dependence. We anticipate that the SDWB will be useful to inference for other parameters of interest, such as smooth function model, in the high-dimensional spatial setting, although the focus of this article is on the mean.

The rest of the article is organized as follows. In Section 2, we introduce the asymptotic framework for the sampling region, stochastic design of sampling locations, and dependence structure of the random field. In Section 3, we introduce the spatially dependent wild bootstrap and describe its implementation. In Section 4, we present a high-dimensional CLT for the sample mean of high-dimensional spatial data and derive the asymptotic validity of SDWB. In Section 5, we investigate finite sample properties of the SDWB via numerical simulations. All the proofs, some additional simulation results and a real data illustration are included in the supplement. The supplement contains the high-dimensional CLT under polynomial moment condition (Appendix A), discussion on examples of random fields that satisfy our regularity conditions (Appendix B), some applications of SDWB for spatial and spatio-temporal data (Appendix C), proofs of Theorems 4.1 and A.1, and Corollary 4.1 (Appendix D), proof of Theorem 4.2 (Appendix E), proof of Proposition B.1 (Appendix F), proof of Proposition C.1 (Appendix G), technical tools (Appendix H), additional simulation results (Appendix I), and real data analysis (Appendix J).

1.1. Notation

For any vector $\boldsymbol{x} = (x_1, \dots, x_q)' \in \mathbb{R}^q$, let $|\boldsymbol{x}| = \sum_{j=1}^q |x_j|$ and $\|\boldsymbol{x}\| = \sqrt{\sum_{j=1}^q x_j^2}$ denote the ℓ^1 and ℓ^2-norms of $\boldsymbol{x}$, respectively. For two vectors $\boldsymbol{x} = (x_1, \dots, x_q)'$ and $\boldsymbol{y} = (y_1, \dots, y_q)' \in \mathbb{R}^q$, the notation $\boldsymbol{x} \le \boldsymbol{y}$ means that $x_j \le y_j$ for all $j = 1, \dots, q$. For any set $A \subset \mathbb{R}^q$, let $|A|$ denote the Lebesgue measure of A, and let $[\![A]\!]$ denote the number of elements in A. For any positive sequences a_n, b_n, we write $a_n \lesssim b_n$ if there is a constant $C > 0$ independent of n such that $a_n \le C b_n$ for all n, $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$, and $a_n \ll b_n$ if $a_n/b_n \to 0$ as $n \to \infty$. For any $a, b \in \mathbb{R}$, let $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. For $a \in \mathbb{R}$ and $b > 0$, we use the shorthand notation $[a \pm b] = [a-b, a+b]$. Let $\|X\|_{\psi_1} = \inf\left\{c > 0 : E\left[\exp(|X|/c) - 1\right] \le 1\right\}$ denote the ψ_1-Orlicz norm for a real-valued random variable X. For random variables X and Y, we write $X \stackrel{d}{=} Y$ if they have the same distribution.

2. Settings

In this section, we discuss mathematical settings of our sampling design and spatial dependence structure. We observe discrete samples $\boldsymbol{Y}(\boldsymbol{s}_1), \dots, \boldsymbol{Y}(\boldsymbol{s}_n)$ from a random field $\boldsymbol{Y} = \{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with $\boldsymbol{Y}(\boldsymbol{s}) = (Y_1(\boldsymbol{s}), \dots, Y_p(\boldsymbol{s}))' \in \mathbb{R}^p$ and are interested in approximating the distribution of the sample mean $\overline{\boldsymbol{Y}}_n = n^{-1} \sum_{i=1}^n \boldsymbol{Y}(\boldsymbol{s}_i)$ when $p = p_n \to \infty$ as $n \to \infty$ and possibly $p \gg n$. The sampling sites $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n \in \mathbb{R}^d$ are stochastic and obtained by rescaling iid random vectors $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$; see below for details. Let $(\Omega^{(j)}, \mathcal{F}^{(j)}, P^{(j)})$, $j = 1, 2, 3$ be probability spaces on which the random field $\boldsymbol{Y}$, a sequence of iid random vectors $\{\boldsymbol{Z}_i\}_{i \ge 1}$ with values in $\mathbb{R}^d$, and an auxiliary real-valued Gaussian random field $W = \{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ are defined, respectively. The auxiliary Gaussian random field W will be used in the construction of SDWB. Consider the product probability space $(\Omega, \mathcal{F}, P)$ where $\Omega = \Omega^{(1)} \times \Omega^{(2)} \times \Omega^{(3)}$, $\mathcal{F} = \mathcal{F}^{(1)} \otimes \mathcal{F}^{(2)} \otimes \mathcal{F}^{(3)}$, and $P = P^{(1)} \times P^{(2)} \times P^{(3)}$. Then $\boldsymbol{Y}$, $\{\boldsymbol{Z}_i\}_{i\geq 1}$, and W are independent by construction. Let $P_{\boldsymbol{Z}}$ denote the joint distribution of the sequence of iid random vectors $\{\boldsymbol{Z}_i\}_{i\geq 1}$ and let $P_{\cdot|\boldsymbol{Z}}$ denote the conditional probability given $\sigma(\{\boldsymbol{Z}_i\}_{i\geq 1})$, the σ-field generated by $\{\boldsymbol{Z}_i\}_{i\geq 1}$. Let $E_{\boldsymbol{Z}}$ denote the expectation with respect to $\{\boldsymbol{Z}_i\}_{i\geq 1}$ and let $E_{\cdot|\boldsymbol{Z}}$ and $\mathrm{var}_{\cdot|\boldsymbol{Z}}$ denote the conditional expectation and variance given $\sigma(\{\boldsymbol{Z}_i\}_{i\geq 1})$, respectively. Finally, let $P_{\cdot|\boldsymbol{Y},\boldsymbol{Z}}$ and $\mathrm{var}_{\cdot|\boldsymbol{Y},\boldsymbol{Z}}$ denote the conditional probability and variance given $\sigma(\{\boldsymbol{Y}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\} \cup \{\boldsymbol{Z}_i\}_{i\geq 1})$, respectively.

2.1. Sampling Design

We follow the setting considered in Lahiri (2003a) and define the sampling region R_n as follows. Let R_0^* be an open connected subset of $(-1/2, 1/2]^d$ containing the origin and let R_0 be a Borel set satisfying $R_0^* \subset R_0 \subset \overline{R}_0^*$, where for any set $A \subset \mathbb{R}^d$, $\overline{A}$ denotes its closure. Let $\{\lambda_n\}_{n\geq 1}$ be a sequence of positive numbers such that $\lambda_n \to \infty$ as $n \to \infty$. We consider the following set as the sampling region. $R_n = \lambda_n R_0$. To avoid pathological cases, we also assume that for any sequence of positive numbers $\{a_n\}_{n\geq 1}$ with $a_n \to 0$ as $n \to \infty$, the number of cubes of the form $a_n(\boldsymbol{i} + [0, 1)^d)$, $\boldsymbol{i} \in \mathbb{Z}^d$ with their lower left corner $a_n\boldsymbol{i}$ on the lattice $a_n\mathbb{Z}^d$ that intersect both R_0 and R_0^c is $O(a_n^{-d+1})$ as $n \to \infty$.

Next we introduce our (stochastic) sampling designs. Let f be a continuous, everywhere positive probability density function on R_0, and let $\{\boldsymbol{Z}_i\}_{i\geq 1}$ be a sequence of iid random vectors with density f. Recall that $\{\boldsymbol{Z}_i\}_{i\geq 1}$ and $\boldsymbol{Y}$ are independent from the construction of the probability space $(\Omega, \mathcal{F}, P)$. We assume that the sampling sites $\boldsymbol{s}_1, \dots, \boldsymbol{s}_n$ are obtained from realizations $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n$ of the random vectors $\boldsymbol{Z}_1, \dots, \boldsymbol{Z}_n$ and the relation $\boldsymbol{s}_i = \lambda_n \boldsymbol{z}_i$, $i = 1, \dots, n$.

The boundary condition on the prototype set R_0 holds in many practical situations, including many convex subsets in $\mathbb{R}^d$ such as spheres, ellipsoids, polyhedrons, as well as many nonconvex sets in $\mathbb{R}^d$. See also Lahiri (2003a) and Chapter 12 in Lahiri (2003b) for more discussions.

2.2. Dependence Structure

In what follows, we assume that the random field $\boldsymbol{Y}$ can be decomposed as

$$\boldsymbol{Y}(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s}) + \boldsymbol{\Upsilon}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d, \tag{2.1}$$

where $\boldsymbol{X} = \{\boldsymbol{X}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with $\boldsymbol{X}(\boldsymbol{s}) = (X_1(\boldsymbol{s}), \dots, X_p(\boldsymbol{s}))'$ is a strictly stationary random field and $\boldsymbol{\Upsilon} = \{\boldsymbol{\Upsilon}(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ with $\boldsymbol{\Upsilon}(\boldsymbol{s}) = (\Upsilon_1(\boldsymbol{s}), \dots, \Upsilon_p(\boldsymbol{s}))'$ is a "residual" random field that is negligible in a certain sense. The decomposition (2.1) may (and in general does) depend on n, that is, $\boldsymbol{X} = \boldsymbol{X}^{(n)}$ and $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}^{(n)}$, but the dependence on n is suppressed for notational convenience. Throughout the article, we assume that $E[\boldsymbol{\Upsilon}(\boldsymbol{s})] = 0$ for any $\boldsymbol{s} \in \mathbb{R}^d$. Then $\boldsymbol{Y}$ is approximately stationary with constant mean.

We also assume that the random field $\boldsymbol{X}$ satisfies a certain mixing condition. Let $\sigma_{\boldsymbol{X}}(T) = \sigma(\{\boldsymbol{X}(\boldsymbol{s}) : \boldsymbol{s} \in T\})$ denote the σ-field generated by $\{\boldsymbol{X}(\boldsymbol{s}) : \boldsymbol{s} \in T\}$ for $T \subset \mathbb{R}^d$. For any subsets T_1 and T_2 of $\mathbb{R}^d$, the β-mixing coefficient between $\sigma_{\boldsymbol{X}}(T_1)$ and $\sigma_{\boldsymbol{X}}(T_2)$ is defined by $\check{\beta}(T_1, T_2) = \sup \sum_{j=1}^J \sum_{k=1}^K |P(A_j \cap B_k) - P(A_j)P(B_k)|/2$, where the supremum is taken over all partitions $\{A_j\}_{j=1}^J \subset \sigma_{\boldsymbol{X}}(T_1)$ and $\{B_k\}_{k=1}^K \subset \sigma_{\boldsymbol{X}}(T_2)$ of $\mathbb{R}^d$. Let $\mathcal{R}(b)$ denote the collection of all finite disjoint unions of cubes in $\mathbb{R}^d$ with total volume not exceeding b. Then, we define $\beta(a; b) = \sup\left\{\check{\beta}(T_1, T_2) : d(T_1, T_2) \geq a, T_1, T_2 \in \mathcal{R}(b)\right\}$, $a, b > 0$, where $d(T_1, T_2) = \inf\{|\boldsymbol{x} - \boldsymbol{y}| : \boldsymbol{x} \in T_1, \boldsymbol{y} \in T_2\}$. We assume that there exist a nonincreasing function β_1 with $\lim_{a\to\infty} \beta_1(a) = 0$ and a nondecreasing function g (that may be unbounded) such that

$$\beta(a; b) \leq \beta_1(a)g(b), \; a, b > 0. \tag{2.2}$$

Our mixing condition (2.2) is a β-mixing version of the α-mixing condition considered in Lahiri (2003a), Lahiri and Zhu (2006), and Bandyopadhyay, Lahiri, and Nordman (2015). In general, the function β_1 may depend on n since the random field $\boldsymbol{X}$ that appears in (2.1) depends on n. Here we assume that g does not depend on n for simplicity, but the extension to the general case that g changes with n is not difficult. The random field $\boldsymbol{Y}$ itself may not satisfy the mixing condition (2.2), since the mixing condition (2.2) is assumed on $\boldsymbol{X}$. With the decomposition (2.1), we allow $\boldsymbol{Y}$ to have a flexible dependence structure since the residual random field $\boldsymbol{\Upsilon}$ can accommodate a complex dependence structure. In particular, we will show in Appendix B that a wide class of Lévy-driven MA random fields admit the decomposition satisfying Condition (4.1).

Remark 2.1. Lahiri (2003b), Lahiri and Zhu (2006), and Bandyopadhyay, Lahiri, and Nordman (2015) assume the α-mixing version of Condition (2.2) to prove limit theorems for spatial data in the fixed dimensional case (i.e., p is fixed). Lahiri (2003b) established CLTs for weighted sample means of spatial data under an α-mixing condition in the univariate case. Lahiri's proof relies essentially on approximating the characteristic function of the weighted sample mean by that of independent blocks using the Volkonskii-Rozanov inequality (see Proposition 2.6 in Fan and Yao 2003) and then showing that the characteristic function corresponding to the independent blocks converges to the characteristic function of its Gaussian limit. However, in the high-dimensional case ($p_n \to \infty$ as $n \to \infty$), characteristic functions are difficult to capture the effect of dimensionality in approximation theorems, so we rely on a different argument than that of Lahiri (2003b). Indeed, we use a stronger blocking argument tailored to β-mixing sequences; see Lemma 4.1 in Yu (1994). It is not known that corresponding results hold for α-mixing sequences; see Remark (ii) right after the proof of Lemma 4.1 in Yu (1994). Hence, we assume Condition (2.2) in the present article.

Remark 2.2. It is important to restrict the size of index sets T_1 and T_2 in the definition of $\beta(a; b)$. Define the β-mixing coefficient of a random field $\boldsymbol{X}$ similarly to the time series as follows: Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be half-planes with boundaries L_1 and L_2, respectively. For each $a > 0$, define $\beta(a) = \sup\left\{\check{\beta}(\mathcal{O}_1, \mathcal{O}_2) : d(\mathcal{O}_1, \mathcal{O}_2) \geq a\right\}$. According to Theorem 1 in Bradley (1989), if $\{\boldsymbol{X}(s) : \boldsymbol{s} \in \mathbb{R}^2\}$ is strictly stationary, then $\beta(a) = 0$ or 1 for $a > 0$. This implies that if a random field $\boldsymbol{X}$ is β-mixing ($\lim_{a\to\infty} \beta(a) = 0$), then it is automatically m-dependent, that is, $\beta(a) = 0$ for some $a > m$, where m is a positive constant. To allow for certain flexibility, we restrict the size of T_1 and T_2 in the definition of $\beta(a; b)$. We refer to Bradley (1993) and Doukhan (1994) for more details on mixing coefficients for random fields.

3. Spatially Dependent Wild Bootstrap

In this section, we introduce the spatially dependent wild bootstrap (SDWB) method for the construction of joint confidence intervals for the mean vector $\boldsymbol{\mu} = E[\boldsymbol{Y}(\boldsymbol{s})] = (\mu_1, \dots, \mu_p)'$. Let $\overline{\boldsymbol{Y}}_n = n^{-1}\sum_{i=1}^n \boldsymbol{Y}(\boldsymbol{s}_i) = (\overline{Y}_{1,n}, \dots, \overline{Y}_{p,n})'$ denote the sample mean. In Section 4.1, we will show that as $n \to \infty$, $\sup_{A\in\mathcal{A}} \left| P_{\cdot|\boldsymbol{Z}}\left(\sqrt{\lambda_n^d}(\overline{\boldsymbol{Y}}_n - \boldsymbol{\mu}) \in A\right) - P_{\cdot|\boldsymbol{Z}}\left(\boldsymbol{V}_n \in A\right)\right| \to 0$, $P_{\boldsymbol{Z}}$-a.s., provided that $p = O(n^\alpha)$ for some $\alpha > 0$, where $\mathcal{A}$ is the class of closed rectangles in $\mathbb{R}^p$. Here $\boldsymbol{V}_n = (V_{1,n}, \dots, V_{p,n})'$ is a centered Gaussian random vector in $\mathbb{R}^p$ under $P_{\cdot|\boldsymbol{Z}}$ with (conditional) covariance matrix $\Sigma^{\boldsymbol{V}_n} = (\Sigma_{j,k}^{\boldsymbol{V}_n})_{1\le j,k\le p} = E_{\cdot|\boldsymbol{Z}}[\boldsymbol{V}_n\boldsymbol{V}_n']$, the form of which is specified later. This high-dimensional CLT implies that a joint $100(1-\tau)\%$ confidence interval for the mean vector $\boldsymbol{\mu}$ with $\tau \in (0,1)$ is given by $\widehat{C}_{1-\tau} = \prod_{j=1}^p \widehat{C}_{j,1-\tau}$, where $\widehat{C}_{j,1-\tau} = \left[\overline{Y}_{j,n} \pm \lambda_n^{-d/2}\sqrt{\Sigma_{j,j}^{\boldsymbol{V}_n}}\, q_n(1-\tau)\right]$, and $q_n(1-\tau)$ is the $(1-\tau)$-quantile of $\max_{1\le j\le p}\left|V_{j,n}/\sqrt{\Sigma_{j,j}^{\boldsymbol{V}_n}}\right|$. Indeed, we have $P_{\cdot|\boldsymbol{Z}}\left(\boldsymbol{\mu} \in \widehat{C}_{1-\tau}\right) = P_{\cdot|\boldsymbol{Z}}\left(\max_{1\le j\le p}\left|V_{j,n}/\sqrt{\Sigma_{j,j}^{\boldsymbol{V}_n}}\right| \le q_n(1-\tau)\right) + o(1) = 1 - \tau + o(1)$ with $P_{\boldsymbol{Z}}$-probability one, so that $\widehat{C}_{1-\tau}$ is a valid joint confidence interval for $\boldsymbol{\mu}$ with level approximately $1-\tau$.

In practice, we have to estimate the quantile $q_n(1-\tau)$, in addition to the coordinatewise variances $\Sigma_{j,j}^{\boldsymbol{V}_n}$. To this end, we develop the spatially dependent wild bootstrap (SDWB), as an extension of DWB proposed by Shao (2010) to the spatial setting. Given the observations $\{\boldsymbol{Y}(\boldsymbol{s}_i)\}_{i=1}^n$, we define the SDWB pseudo-observations as $\boldsymbol{Y}^*(\boldsymbol{s}_i) = \overline{\boldsymbol{Y}}_n + (\boldsymbol{Y}(\boldsymbol{s}_i) - \overline{\boldsymbol{Y}}_n)W(\boldsymbol{s}_i)$, $i = 1, \dots, n$, where $\{W(\boldsymbol{s}_i)\}_{i=1}^n$ are n discrete samples from a real-valued stationary Gaussian random field $W = \{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ such that $E[W(\boldsymbol{s})] = 0$, $\mathrm{var}(W(\boldsymbol{s})) = 1$ and $\mathrm{cov}(W(\boldsymbol{s}_1), W(\boldsymbol{s}_2)) = a(\|\boldsymbol{s}_2 - \boldsymbol{s}_1\|/b_n)$. Here $a(\cdot) : \mathbb{R} \to [0,1]$ is a continuous kernel function supported in $[-1,1]$ and b_n is a bandwidth such that $b_n \to \infty$ as $n \to \infty$.

To estimate the covariance matrix $\Sigma^{\boldsymbol{V}_n}$, we use the classical lag-window type estimator defined as $\widehat{\Sigma}^{\boldsymbol{V}_n} = \frac{1}{n^2\lambda_n^{-d}}\sum_{\ell_1,\ell_2=1}^n (\boldsymbol{Y}(\boldsymbol{s}_{\ell_1}) - \overline{\boldsymbol{Y}}_n)(\boldsymbol{Y}(\boldsymbol{s}_{\ell_2}) - \overline{\boldsymbol{Y}}_n)' a(\|\boldsymbol{s}_{\ell_1} - \boldsymbol{s}_{\ell_2}\|/b_n)$. Denote by $\widehat{\Sigma}_{j,k}^{\boldsymbol{V}_n}$ the (j,k)th component of $\widehat{\Sigma}^{\boldsymbol{V}_n}$. Let $\overline{\boldsymbol{Y}}_n^* = n^{-1}\sum_{i=1}^n \boldsymbol{Y}^*(\boldsymbol{s}_i) = (\overline{Y}_{1,n}^*, \dots, \overline{Y}_{p,n}^*)'$. It is not difficult to see that $\widehat{\Sigma}^{\boldsymbol{V}_n} = \lambda_n^d \mathrm{var}_{\cdot|\boldsymbol{Y},\boldsymbol{Z}}(\overline{\boldsymbol{Y}}_n^*)$. That is, the SDWB variance estimator coincides with the lag window estimator provided that the covariance function and bandwidth used in SDWB match the kernel and bandwidth in the above expression.

Then we can estimate the quantile $q_n(1-\tau)$ by the empirical quantile of SDWB sample means $\widehat{q}_n(1-\tau) = \inf\left\{t \in \mathbb{R} : P_{\cdot|\boldsymbol{Y},\boldsymbol{Z}}\left(\sqrt{\lambda_n^d}\max_{1\le j\le p}\left|\frac{\overline{Y}_{j,n}^* - \overline{Y}_{j,n}}{\sqrt{\widehat{\Sigma}_{j,j}^{\boldsymbol{V}_n}}}\right| \le t\right) \ge 1-\tau\right\}$. We will show in Sections 4.2 and C.1 in Appendix C that the plug-in joint confidence interval $\widehat{C}_{1-\tau}$ with $\Sigma_{j,j}^{\boldsymbol{V}_n}$ and $q_n(1-\tau)$ replaced by $\widehat{\Sigma}_{j,j}^{\boldsymbol{V}_n}$ and $\widehat{q}_n(1-\tau)$ will have asymptotically correct coverage probability under regularity conditions.

Remark 3.1 (Comparison with DWB and other bootstrap methods). Since the introduction of DWB for time series inference, there have been quite a bit further extensions in the time series literature. For example, its validity has been justified for degenerate U- and V-statistics by Leucht and Neumann (2013) and Chwialkowski, Sejdinovic, and Gretton (2014), and for empirical processes by Doukhan et al. (2015). It has also been used in several testing problems to cope with weak temporal dependence; see Bucchiaa and Wendler (2017), Rho and Shao (2019), and Hill and Motegi (2020) among others. Although conceptually simple, the theory associated with SDWB is considerably more involved than DWB and our proof techniques are substantially different from the above-mentioned papers due to our focus on its validity in the high-dimensional setting.

Recently, Hounyo (2022) introduced a novel wild bootstrap for dependent data (WBDD), which can be viewed as an extension of DWB in the time series setting. The WBDD is a kind of DWB but does not require the sequence of external variables to be stationary, and with suitable choice of the distribution of external variables, it can be shown to possess the second-order correctness property for stationary time series. However, the formulation of WBDD is limited to regularly spaced time series and hinges on a pre-blocking processing of the data, and there is not yet any extension to irregularly spaced time series or spatial data. We expect that such an extension is possible in view of grid-based block bootstrap of Lahiri and Zhu (2006) for irregularly spaced spatial data, but nevertheless some nontrivial implementation issues, such as the partition of the sampling region into complete and incomplete blocks, will come up. By contrast, the implementation of SDWB is agnostic to the spatial configuration and only requires the knowledge of pairwise distance between sampling sites, for a given kernel and bandwidth. As in the time series case, we do not expect the second order accuracy for SDWB to hold in general.

In the literature of bootstrapping time series, Jentsch and Politis (2015) showed the asymptotic validity of a multivariate version of the linear process bootstrap (McMurry and Politis 2010) for inference about the mean when the dimension of a time series is allowed to increase with the sample size. However, the growth rate of p has to be slower than that of the sample size. Since our method is mainly for spatial data, which does not have natural ordering, and our theory allows $p \gg n$, the applicability of linear process bootstrap and ours are fairly different, and a direct comparison seems unwarranted.

For spatial square-shaped data sample, Meyer, Jentsch, and Kreiss (2017) extended the autoregressive (AR) sieve bootstrap from time series to spatial process in $\mathbb{Z}^2$ by fitting AR models of increasing order to the given data, resampling of the residuals, and generating bootstrap replicates of the sample. They showed the validity of AR sieve bootstrap for an interesting class of statistics including sample autocorrelations and standardized sample variograms. As an important theoretical contribution, they obtained a weighted Baxter-inequality for spatial processes, which yields a rate of convergence for the finite predictor coefficients, that is, the coefficients of finite-order AR model fits, toward the autoregressive coefficients. Also see Jentsch and Meyer (2021) for related results on Akaike identity (Akaike 1969) for spatial lattice data. The applicability of spatial AR sieve bootstrap (SARB, hereafter) and SDWB are quite different. On one hand, our method is applicable to irregularly spaced spatial data, whereas SARB is only developed for spatial lattice data and an extension to allow irregular spacing seems highly nontrivial. On the other hand, the theoretical justification of SDWB is on the mean here, whereas the class of statistic to which SARB applies is broader. In both Meyer, Jentsch, and Kreiss (2017) and Jentsch and Meyer (2021), their focus was on univariate spatial process, whereas we mainly target multivariate spatial process with the dimension p being moderate or large.

4. Main Results

In this section, we first derive a high-dimensional CLT for the sample mean over the rectangles in Section 4.1. Building on the high-dimensional CLT, we establish the asymptotic validity of the SDWB over the rectangles in high dimensions in Section 4.2. In what follows, we maintain the baseline assumption discussed in Section 2.

4.1. High-Dimensional CLT

To state the high-dimensional CLT, we shall consider the two cases separately: (i) the coordinates of $\boldsymbol{X}$ are sub-exponential and (ii) have finite polynomial moments. The results for the second case are provided in Appendix A due to the page limit. Some applications of our main results are also discussed in Appendix C.

We make the following assumption.

Assumption 4.1. Suppose that $p = O(n^{\alpha})$ for some $\alpha > 0$. Let $\{\lambda_{1,n}\}_{n\geq 1}$ and $\{\lambda_{2,n}\}_{n\geq 1}$ be two sequences of positive numbers such that $\lambda_{1,n}, \lambda_{2,n} \to \infty$, $\lambda_{2,n} = o(\lambda_{1,n})$, and $\lambda_{1,n} = o(\lambda_n)$.

(i) The random field $\boldsymbol{X}$ has zero mean, that is, $E[\boldsymbol{X}(\boldsymbol{s})] = \boldsymbol{0}$ and the residual random field $\boldsymbol{\Upsilon}$ satisfies that for some $\zeta > 0$,

$$P_{\cdot|\mathbf{Z}}\left(\max_{1\leq j\leq p}\left|\frac{1}{\sqrt{n^2\lambda_n^{-d}}}\sum_{i=1}^{n}\Upsilon_j(\boldsymbol{s}_i)\right| > n^{-\zeta}\log^{-1/2}p\right) = O(n^{-\zeta}) \qquad \text{with } P_{\mathbf{Z}}\text{-a.s.} \tag{4.1}$$

There exist two sequences of positive constants $\{D_n\}_{n\geq 1}$ with $D_n \geq 1$ and $\{\delta_{n,\Upsilon}\}_{n\geq 1}$ with $\delta_{n,\Upsilon} \to 0$ such that

$$\max_{1\leq j\leq p}\|X_j(\boldsymbol{s})\|_{\psi_1} \leq D_n \qquad \text{and}$$
$$E_{\cdot|\mathbf{Z}}\left[\max_{1\leq i\leq n}\max_{1\leq j\leq p}\left|\Upsilon_j(\boldsymbol{s}_i)\right|^q\right] \leq \delta_{n,\Upsilon}^q \tag{4.2}$$
$$\text{with } P_{\mathbf{Z}}\text{-a.s. for some } q \in [8,\infty).$$

(ii) The probability density function f is continuous, everywhere positive with support $\overline{R}_0$.

(iii) We have $\lim_{n\to\infty} n\lambda_n^{-d} = \kappa \in (0,\infty]$ with $\lambda_n \geq n^{\bar{\kappa}}$ for some $\bar{\kappa} > 0$.

(iv) There exists a constant $0 < c < 1/2$ such that

$$\max\left\{\frac{1}{\sqrt{\lambda_{2,n}}},\ D_n\left(\frac{\log n}{n\lambda_n^{-d}}+1\right)\sqrt{\frac{\overline{\beta}_q\lambda_{2,n}}{\lambda_{1,n}}},\ \left(\lambda_{1,n}^{d/2}\lambda_{2,n}^{d}D_n^2 + \lambda_{1,n}^{d}D_n\right)\lambda_n^{-d/2},\ \frac{D_n^6\lambda_{1,n}^{3d}}{n^2\lambda_n^{-d}}\right\} n^{c} = O(1)$$

as $n\to\infty$, where $\overline{\beta}_q = \overline{\beta}_q(n) := 1+\sum_{k=1}^{\lambda_{1,n}} k^{d-1}\beta_1^{1-2/q}(k)$. Further, there exists a constant $0 < c' < c$ such that

$$\lambda_n^{d/2}n^{c'}\delta_{n,\Upsilon} = O(1). \tag{4.3}$$

(v) We have $\lim_{n\to\infty}\lambda_n^d\lambda_{1,n}^{-d}\beta(\lambda_{2,n};\lambda_n^d) = 0$, and there exist some constants $0 < \underline{c} < \overline{C} < \infty$ such that

$$\Sigma_{j,j}(\boldsymbol{0}) \leq \overline{C}, \qquad \int_{\mathbb{R}^d}\Sigma_{j,j}(\boldsymbol{s})d\boldsymbol{s} \geqslant \underline{c} \quad \text{and}$$
$$\int_{\mathbb{R}^d}|\Sigma_{j,j}(\boldsymbol{s})|d\boldsymbol{s} \leq \overline{C} \quad \text{for all } 1\leq j\leq p, \tag{4.4}$$

where $\Sigma(\boldsymbol{s}) = (\Sigma_{j,k}(\boldsymbol{s}))_{1\leq j,k\leq p} = \mathrm{cov}(\boldsymbol{X}(\boldsymbol{s}),\boldsymbol{X}(\boldsymbol{0}))$.

A discussion about the above assumptions is warranted. The sequences $\{\lambda_{1,n}\}$ and $\{\lambda_{2,n}\}$ will be used in the large-block-small-block argument, which is commonly used in proving CLTs for sums of mixing random variables; see Lahiri (2003b). Specifically, $\lambda_{1,n}$ corresponds to the side length of large blocks, while $\lambda_{2,n}$ corresponds to the side length of small blocks. The first part of Condition (4.2) requires the coordinates of $\boldsymbol{X}(\boldsymbol{s})$ to be (uniformly) sub-exponential, while the second part of Condition (4.2) partially ensures the asymptotic negligibility of the residual random field, along with the condition (4.1). Condition (ii) is concerned with the distribution of irregularly spaced sampling sites and allows a nonuniform density across the sampling region. Condition (iii) implies that our sampling design allows the pure increasing domain case ($\lim_{n\to\infty} n\lambda_n^{-d} = \kappa \in (0,\infty)$) and the mixed increasing domain case ($\lim_{n\to\infty} n\lambda_n^{-d} = \infty$). Condition (4.3) is used to guarantee the asymptotic negligibility of $\boldsymbol{\Upsilon}(\boldsymbol{s})$ for the asymptotic validity of the SDWB. For random fields to be discussed in Appendix B, $\delta_{n,\Upsilon}$ decays exponentially fast as $n\to\infty$, so that Condition (4.3) is satisfied. Condition (v) is a technical condition on the covariance function of $\boldsymbol{X}$. Condition (4.4) is used to guarantee that the (conditional) coordinatewise variances of the normalized sample mean $\sqrt{\lambda_n^d}\overline{\boldsymbol{X}}_n$ are bounded away from zero almost surely.

Let us briefly compare our conditions with Condition (S.5) in Lahiri (2003a), who established CLTs for weighted sums of spatial data in the univariate case (i.e., $p = 1$). Condition (iv) corresponds to Lahiri's Condition (S.5) Part (i), and the condition $\lim_{n\to\infty}\lambda_n^d\lambda_{1,n}^{-d}\beta(\lambda_{2,n};\lambda_n^d) = 0$ corresponds to the β-mixing version of Lahiri's Condition (S.5) Part (iii). In particular, $\sqrt{\overline{\beta}_q\lambda_{2,n}\lambda_{1,n}^{-1}} = O(n^{-c})$ and $\lambda_{1,n}^{d}D_n/\lambda_n^{-d/2} = O(n^{-c})$ imply Lahiri's Condition (S.5) Part (i). Although our conditions are slightly more restrictive than his, they enable us to obtain error bounds for the high-dimensional CLT for the sum of large blocks of high-dimensional spatial data with the dimension growing polynomially fast in the sample size.

We are ready to state the main results. Let $\mathcal{A} = \{\prod_{j=1}^{p}[a_j, b_j] : -\infty \le a_j \le b_j \le \infty, 1 \le j \le p\}$ denote the collection of closed rectangles in $\mathbb{R}^p$. For $\boldsymbol{\ell} = (\ell_1, \dots, \ell_d)' \in \mathbb{Z}^d$, we let $\Gamma_n(\boldsymbol{\ell}; \mathbf{0}) = (\boldsymbol{\ell} + (0,1]^d)\lambda_{3,n}$ with $\lambda_{3,n} = \lambda_{1,n} + \lambda_{2,n}$, and define the following hypercubes, $\Gamma_n(\boldsymbol{\ell}; \mathbf{1}) = \prod_{j=1}^{d}(\ell_j\lambda_{3,n}, \ell_j\lambda_{3,n} + \lambda_{1,n}]$. Intuitively, $\Gamma_n(\boldsymbol{\ell}; \mathbf{0})$ is a complete block of indices in $\mathbb{R}^d$ that contains a large block $\Gamma_n(\boldsymbol{\ell}; \mathbf{1})$ and many small blocks $\Gamma_n(\boldsymbol{\ell}; \mathbf{0}) \setminus \Gamma_n(\boldsymbol{\ell}; \mathbf{1})$. Let $L_n = \{\boldsymbol{\ell} \in \mathbb{Z}^d : \Gamma_n(\boldsymbol{\ell}, \mathbf{0}) \cap R_n \ne \emptyset\}$ denote the index set of all hypercubes $\Gamma_n(\boldsymbol{\ell}, \mathbf{0})$ that are contained in or intersects with the boundary of R_n. Define $S_n(\boldsymbol{\ell}; \mathbf{1}) = \sum_{i:\boldsymbol{s}_i \in \Gamma_n(\boldsymbol{\ell};\mathbf{1}) \cap R_n} \boldsymbol{X}(\boldsymbol{s}_i)$. If $[\![\{i : \boldsymbol{s}_i \in \Gamma_n(\boldsymbol{\ell}; \mathbf{1}) \cap R_n\}]\!] = 0$, we set $S_n(\boldsymbol{\ell}; \mathbf{1}) = \mathbf{0}$.

Theorem 4.1 (High-dimensional CLT). Under Assumption 4.1, the following result holds $P_{\boldsymbol{Z}}$-almost surely:

$$\sup_{A \in \mathcal{A}} \left| P_{\cdot|\boldsymbol{Z}}\left(\sqrt{\lambda_n^d}\,\overline{\boldsymbol{Y}}_n \in A\right) - P_{\cdot|\boldsymbol{Z}}\left(\boldsymbol{V}_n \in A\right) \right| \le C\left(\frac{\lambda_n}{\lambda_{1,n}}\right)^d \beta(\lambda_{2,n}; \lambda_n^d) + O(n^{-\left(\frac{c'}{6} \wedge \zeta\right)}), \tag{4.5}$$

where C is a positive constant that does not depend on n, and $\boldsymbol{V}_n = (V_{1,n}, \dots, V_{p,n})'$ is a centered Gaussian random vector under $P_{\cdot|\boldsymbol{Z}}$ with (conditional) covariance $E_{\cdot|\boldsymbol{Z}}\left[\boldsymbol{V}_n\boldsymbol{V}_n'\right] = \frac{1}{n^2\lambda_n^{-d}}\sum_{\boldsymbol{\ell} \in L_n} E_{\cdot|\boldsymbol{Z}}\left[S_n(\boldsymbol{\ell}; \mathbf{1})S_n(\boldsymbol{\ell}; \mathbf{1})'\right]$.

The proof of Theorem 4.1 relies on an extension of the coupling technique in Yu (1994) to irregularly spatial data. The proof proceeds by first approximating the sample mean by the sum of independent large blocks and then showing the high-dimensional CLT for the sum of independent large blocks. The terms $S_n(\boldsymbol{\ell}; \mathbf{1})$ that appear in the representation of $E_{\cdot|\boldsymbol{Z}}[\boldsymbol{V}_n\boldsymbol{V}_n']$ is the (conditional) covariance matrix of independent couplings for the large blocks. The first term $\left(\lambda_n/\lambda_{1,n}\right)^d \beta(\lambda_{2,n}; \lambda_n^d)$ in the error bound (4.5) comes from the blocking argument and reflects a bound on the contribution from small blocks, while the second term corresponds to the error bound of the high-dimensional CLT for the sum of independent large blocks.

The covariance matrix of the (conditionally) Gaussian vector $\boldsymbol{V}_n$ depends on the block construction. While the result of Theorem 4.1 is sufficient to establish the asymptotic validity of the SDWB, it is possible to replace the approximating Gaussian vector by that with covariance matrix independent of the block construction, as shown in the following corollary.

Corollary 4.1. If, in addition to Assumption 4.1, (i) $(n\lambda_n^{-d})^{-1} = \kappa^{-1} + o((\log n)^{-2})$, where $\kappa^{-1} = 0$ if $\kappa = \infty$; (ii) the density function f is Lipschitz continuous inside R_0; and (iii) $\int_{\|\boldsymbol{s}\| \ge \lambda_{2,n}} |\Sigma_{j,j}(\boldsymbol{s})| d\boldsymbol{s} = O(n^{-c'/2})$ uniformly over $1 \leqslant j \leqslant p$, then we have with $P_{\boldsymbol{Z}}$-probability one, $\sup_{A \in \mathcal{A}} \left| P_{\cdot|\boldsymbol{Z}}\left(\sqrt{\lambda_n^d}\,\overline{\boldsymbol{Y}}_n \in A\right) - P\left(\check{\boldsymbol{V}}_n \in A\right) \right| = o(1)$, where $\check{\boldsymbol{V}}_n$ is a centered Gaussian random vector with covariance $E[\check{\boldsymbol{V}}_n\check{\boldsymbol{V}}_n'] = \left(\int_{\mathbb{R}^d} \Sigma_{j,k}(\boldsymbol{x})d\boldsymbol{x} \int_{R_0} f^2(\boldsymbol{z})d\boldsymbol{z} + \kappa^{-1}\Sigma_{j,k}(\mathbf{0})\right)_{1 \leqslant j,k \leqslant p}$.

Corollary 4.1 is a high-dimensional extension of Theorem 3.1 in Lahiri (2003a) when $\omega_n(s) = 1$ in his notation. If there is no residual random field $\boldsymbol{\Upsilon}$, that is, $\boldsymbol{Y} = \boldsymbol{X}$ then each component of the covariance matrix $E[\check{\boldsymbol{V}}_n\check{\boldsymbol{V}}_n']$ in Corollary 4.1 corresponds to $\lim_{n\to\infty} \lambda_n^d \operatorname{cov}(\bar{Y}_{j,n}, \bar{Y}_{k,n})$, $1 \le j, k \le p$. The conclusion of Corollary 4.1 follows from Theorem 4.1 combined with the Gaussian comparison inequality. Indeed, under the assumption of Corollary 4.1, we will show that $\max_{1 \le j,k \le p} |E_{\cdot|\boldsymbol{Z}}[V_{j,n}V_{k,n}] - E[\check{V}_{j,n}\check{V}_{k,n}]| = o((\log n)^{-2})$ with $P_{\boldsymbol{Z}}$-probability one, which implies the conclusion of Corollary 4.1 via the Gaussian comparison.

4.2. Asymptotic Validity of the SDWB

In this section, we establish the asymptotic validity of SDWB in high dimensions. Recall that, given the observations $\{\boldsymbol{Y}(\boldsymbol{s}_i)\}_{i=1}^n$, the SDWB pseudo-observations are given by $\boldsymbol{Y}^*(\boldsymbol{s}_i) = \overline{\boldsymbol{Y}}_n + (\boldsymbol{Y}(\boldsymbol{s}_i) - \overline{\boldsymbol{Y}}_n)W(\boldsymbol{s}_i)$, $i = 1, \dots, n$, where $\{W(\boldsymbol{s}_i)\}_{i=1}^n$ are n discrete samples from a real-valued stationary Gaussian random field $W = \{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^d\}$ independent of $\boldsymbol{Y}$ and $\{\boldsymbol{Z}_i\}_{i \geqslant 1}$. We make the following assumption on W.

Assumption 4.2. The random field W is a stationary Gaussian random field with mean zero and covariance function $\operatorname{cov}(W(\boldsymbol{s}_1), W(\boldsymbol{s}_2)) = a(\|\boldsymbol{s}_2 - \boldsymbol{s}_1\|/b_n)$, where $a(\cdot) : \mathbb{R} \to [0,1]$ is a continuous kernel function and b_n is a bandwidth parameter. The kernel function satisfies that $a(0) = 1$ and $a(x) = 0$ for $|x| \ge 1$. There exist positive constants c_W and L_W such that $|1 - a(x)| \le L_W|x|$ for $|x| \le c_W$. Further, with $\mathrm{i} = \sqrt{-1}$, $K_a(x) = \int_{\mathbb{R}} a(u)e^{-\mathrm{i}ux}du \ge 0$ for all $x \in \mathbb{R}$.

The condition on K_a guarantees the positive semi-definiteness of the covariance matrix of $\{W(\boldsymbol{s}_i)\}_{i=1}^n$. Assumption 4.2 is satisfied by many commonly used kernel functions in the literature of spectral density estimation, in particular, Bartlett and Parzen kernels. See Andrews (1991) for details.

Remark 4.1 (Comments on the auxiliary random field W). The covariance function of the Gaussian random field W defined in Assumption 4.2 implies that the random field W is isotropic. We assume this condition for technical convenience, and it is not difficult to see from the proof that the conclusion of the following theorem holds for the following class of (possibly) non-isotropic covariance functions. Consider a function $\breve{a} : \mathbb{R}^d \to [0,1]$ with $\breve{a}(\mathbf{0}) = 1$, $\breve{a}(\boldsymbol{x}) = 0$ for $\|\boldsymbol{x}\| \ge 1$, and assume that there exist positive constants c_W and L_W such that $|1 - \breve{a}(\boldsymbol{x})| \le L_W\|\boldsymbol{x}\|$ for $\|\boldsymbol{x}\| \le c_W$. Further, assume that the function $\mathfrak{a} : \mathbb{R}^d \times \mathbb{R}^d \to [0,1]$ defined by $\mathfrak{a}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \breve{a}(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ is positive semidefinite. For example, these conditions are satisfied for product kernels of the form $\breve{a}(\boldsymbol{x}) = \prod_{j=1}^d a_j(\sqrt{d}|x_j|)$ where a_j are one-dimensional kernel functions that satisfy Assumption 4.2. In addition, the Gaussian random field assumption can also be relaxed but at the expense of additional technical complications; see Example 4.1 of Shao (2010) for an example of non-Gaussian distribution for external random variables $\{W(\boldsymbol{s}_i)\}_{i=1}^n$.

Theorem 4.2 (Asymptotic validity of SDWB in high dimensions). Suppose that Assumptions 4.1 (or A.1 in Appendix A) and 4.2 hold with $b_n \sim \lambda_{2,n}$. In addition, assume that

$$\sum_{n=1}^{\infty} n^{c'}(\log n)^2\lambda_n^d\lambda_{1,n}^{-d} \max_{1 \le j,k \le p} \int_{\|\boldsymbol{s}\| > \sqrt{\lambda_{2,n}}} |\Sigma_{j,k}(\boldsymbol{s})| d\boldsymbol{s} < \infty. \tag{4.6}$$

Then the following result holds $P_{\mathbf{Z}}$-almost surely: with $P_{\cdot|\mathbf{Z}}$-probability at least $1 - O(n^{-\left(\frac{c'}{6} \wedge \zeta\right)}) - C(\lambda_n/\lambda_{1,n})^d \beta(\lambda_{2,n}; \lambda_n^d)$, $\sup_{A \in \mathcal{A}} \left| P_{\cdot|\mathbf{Y},\mathbf{Z}} \left(\sqrt{\lambda_n^d}(\overline{\mathbf{Y}}_n^* - \overline{\mathbf{Y}}_n) \in A \right) - P_{\cdot|\mathbf{Z}}(\mathbf{V}_n \in A) \right| = O(n^{-c'/6})$ where C is a positive constant that does not depend on n and p.

From the definition of $\overline{\mathbf{Y}}_n^*$, $\sqrt{\lambda_n^d}(\overline{\mathbf{Y}}_n^* - \overline{\mathbf{Y}}_n)$ can be decomposed into the sum of three terms: $n^{-1}\lambda_n^{d/2} \left(\sum_{i=1}^n W(\mathbf{s}_i)\mathbf{X}(\mathbf{s}_i) - \sum_{i=1}^n W(\mathbf{s}_i)\overline{\mathbf{Y}}_n + \sum_{i=1}^n W(\mathbf{s}_i)\boldsymbol{\Upsilon}(\mathbf{s}_i)\right)$.

The proof of Theorem 4.2 proceeds with (i) showing asymptotic negligibility of second and third terms and (ii) approximating the first term by $\mathbf{V}_n$. Let $E_{\cdot|\mathbf{X},\mathbf{Z}}$ denote the conditional expectation given $\sigma(\{\mathbf{X}(\mathbf{s}) : \mathbf{s} \in \mathbb{R}^d\} \cup \{\mathbf{Z}_i\}_{i\geq 1})$ and let $U_n = \overline{\mathbf{Y}}_n^* - \overline{\mathbf{Y}}_n = (U_{1,n}, \ldots, U_{p,n})'$. We can show that $E_{\cdot|\mathbf{Z}}\left[\max_{1\leq j,k\leq p} \left|\lambda_n^d E_{\cdot|\mathbf{X},\mathbf{Z}}[U_{j,n}U_{k,n}] - E_{\cdot|\mathbf{Z}}[V_{j,n}V_{k,n}]\right|\right] = o(1)$ with $P_{\mathbf{Z}}$-probability one using the results in the proof of Theorem 4.2. This implies the consistency of SDWB for variance estimation in the high-dimensional setting.

Remark 4.2 (Comparisons with block-based subsampling and resampling methods). There have been substantial efforts in extending subsampling (Politis, Romano, and Wolf 1999b) and block-based bootstrap (BBB) methods (Lahiri 2003b) from time series (i.e., $d = 1$) to random fields (i.e., $d \geqslant 2$). For example, Politis, Paparoditis, and Romano (1998) proposed a subsampling method for irregularly spaced spatial data generated by a homogeneous Poisson process. Politis, Paparoditis, and Romano (1999) proposed a version of the spatial block bootstrap under the same framework. Lahiri and Zhu (2006) developed a grid-based block bootstrap for irregularly spaced spatial data with nonuniform stochastic sampling designs. While subsampling and BBB methods are able to capture spatial dependence nonparametrically, their implementation can be inconvenient when applied to irregularly spaced spatial data, as both require partitioning the sampling region into complete and incomplete blocks, and the implementation details can be highly dependent on spatial configuration of sampling region. By contrast, the implementation of SDWB only requires the generation of an auxiliary random field $W(\cdot)$ and irregularity of sampling sites brings no additional difficulty.

On the theory front, Shao (2010) showed that DWB and BBB (especially TBB) are often comparable in terms of theoretical properties in the time series setting with a proper choice of kernel function and bandwidth, but all theoretical results developed so far for BBB seem exclusively for low-dimensional time series/random fields. In particular, the analysis of Shao (2010) is focused on the case where the dimension p is fixed and relies on the explicit limit distribution of the normalized sample mean, while in the high-dimensional case, there are no explicit limit distributions, and the asymptotic analysis of the SDWB is substantially more involved than that of Shao (2010). Overall, the technical assumptions and probabilistic tools we use are considerably different due to our focus on high-dimensional Gaussian approximation for random fields. To the best of our knowledge, our work is the first attempt in the literature to show the validity of a bootstrap method for high-dimensional spatial data.

5. Simulation Results

In this section, we present some simulation results to evaluate the finite sample properties of the SDWB in constructing simultaneous confidence intervals for the mean vector of high-dimensional spatial data. Let the sampling region be $R_n = \lambda_n(-1/2, 1/2]^2 \subset \mathbb{R}^2$ with $\lambda_n \in \{15, 25\}$. We consider four data generating processes (DGPs).

The first DGP (DGP1) is the compound Poisson-driven CAR(1) (CP-CAR(1))-type random field $\mathbf{Y}(\mathbf{s}) = \sum_{i=1}^{\infty} \mathbf{g}(\|\mathbf{s} - \mathbf{x}_i\|)\mathbf{J}_i$, where $\mathbf{x}_i$ denotes the location of the ith unit point mass of a Poisson random measure on $\mathbb{R}^2$ with intensity $\lambda = 1$ and $\{\mathbf{J}_i\}_{i\geq 1}$ is a sequence of iid random variables in $\mathbb{R}^p$. In our simulation study, we set $\mathbf{g}(\|\mathbf{x}\|) = e^{-3\|\mathbf{x}\|}I_p$ and $\mathbf{J}_1 \sim N(0, I_p)$, where I_p denotes the $p \times p$ identity matrix. To simulate CP-CAR(1) random field, we follow the algorithm described in Brockwell and Matsuda (2017):

(i) Take R_n' to be a sufficiently large set containing R_n. In this simulation study, we take $R_n' = 35 \cdot (-1/2, 1/2]^2$.
(ii) Simulate a Poisson random variable $n(R_n')$ with mean $\lambda|R_n'|$ and set it as the number of knots contained in R_n'.
(iii) Simulate $n(R_n')$ independent and uniformly distributed points $\mathbf{x}_1, \ldots, \mathbf{x}_{n(R_n')}$ in R_n'.
(iv) Compute the truncated version of CP-CAR(1): $\mathbf{Y}(\mathbf{s}) = \sum_{i=1}^{n(R_n')} \mathbf{g}(\|\mathbf{s} - \mathbf{x}_i\|)\mathbf{J}_i$.

The second DGP (DGP2) is a p-variate Gaussian random field with mean zero and independent components, each of which admits the Matérn covariance function $\mathrm{cov}(Y_j(\mathbf{s}), Y_j(\mathbf{0})) = \frac{2^{1-\nu}}{\Gamma(\nu)}\left(\sqrt{2\nu}\|\mathbf{s}\|/a\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\|\mathbf{s}\|/a\right)$, $\nu > 0, a > 0, 1 \leq j \leq p$, where $\Gamma(\cdot)$ is the gamma function. In our simulation study, we set $\nu = 3/2$ and $a = 1/\sqrt{3}$.

The third DGP (DPG3) is the following factor model: $\mathbf{Y}(\mathbf{s}_i) = \mathbf{A}\mathbf{F}(\mathbf{s}_i) + \mathbf{R}(\mathbf{s}_i)$, $i = 1, \ldots, n$, where $\mathbf{A}$ is a $p \times 5$ matrix, $\mathbf{F}(\mathbf{s})$ is a 5-variate mean zero Gaussian random field with independent components that have the Matérn covariance function with $\nu = 3/2$, $a = 1/\sqrt{3}$, and $\mathbf{R}(\mathbf{s}_i)$ are p-variate iid standard Gaussian random vectors. For each combination of (n, p, λ_n, b_n), we generate all the $p \times 5$ elements of $\mathbf{A}$ independently from the uniform distribution on the interval $[-1, 1]$ and fix it for all Monte Carlo replications. Compared with DGP2, the Gaussian random field $\mathbf{Y}$ from DGP3 adds strong componentwise dependence through a factor model structure.

The fourth DGP (DGP4) is a p-variate non-Gaussian random field with independent components such that $\widetilde{\mathbf{Y}}(\mathbf{s}) = (Y_1^2(\mathbf{s}), \ldots, Y_p^2(\mathbf{s}))'$, where $Y_j(\mathbf{s})$ admits the Matérn covariance function with $\nu = 3/2$ and $a = 1/\sqrt{3}$. In this case, we have $E[Y_j^2(\mathbf{s})] = \mathrm{var}(Y_1(\mathbf{0})) = 1, j = 1, \ldots, p$.

In our simulation, we compare the finite sample properties of the SDWB with that of the grid-based block bootstrap (GBBB) proposed in Lahiri and Zhu (2006). Let $\overline{\mathbf{Y}}_n^{*,BB} = (\overline{Y}_{1,n}^{*,BB}, \ldots, \overline{Y}_{p,n}^{*,BB})'$ be the GBBB version of the sample mean $\overline{\mathbf{Y}}_n$ (see Section 5.1.3 of Lahiri and Zhu (2006) for details). A GBBB joint $100(1 - \tau)$% confidence interval for

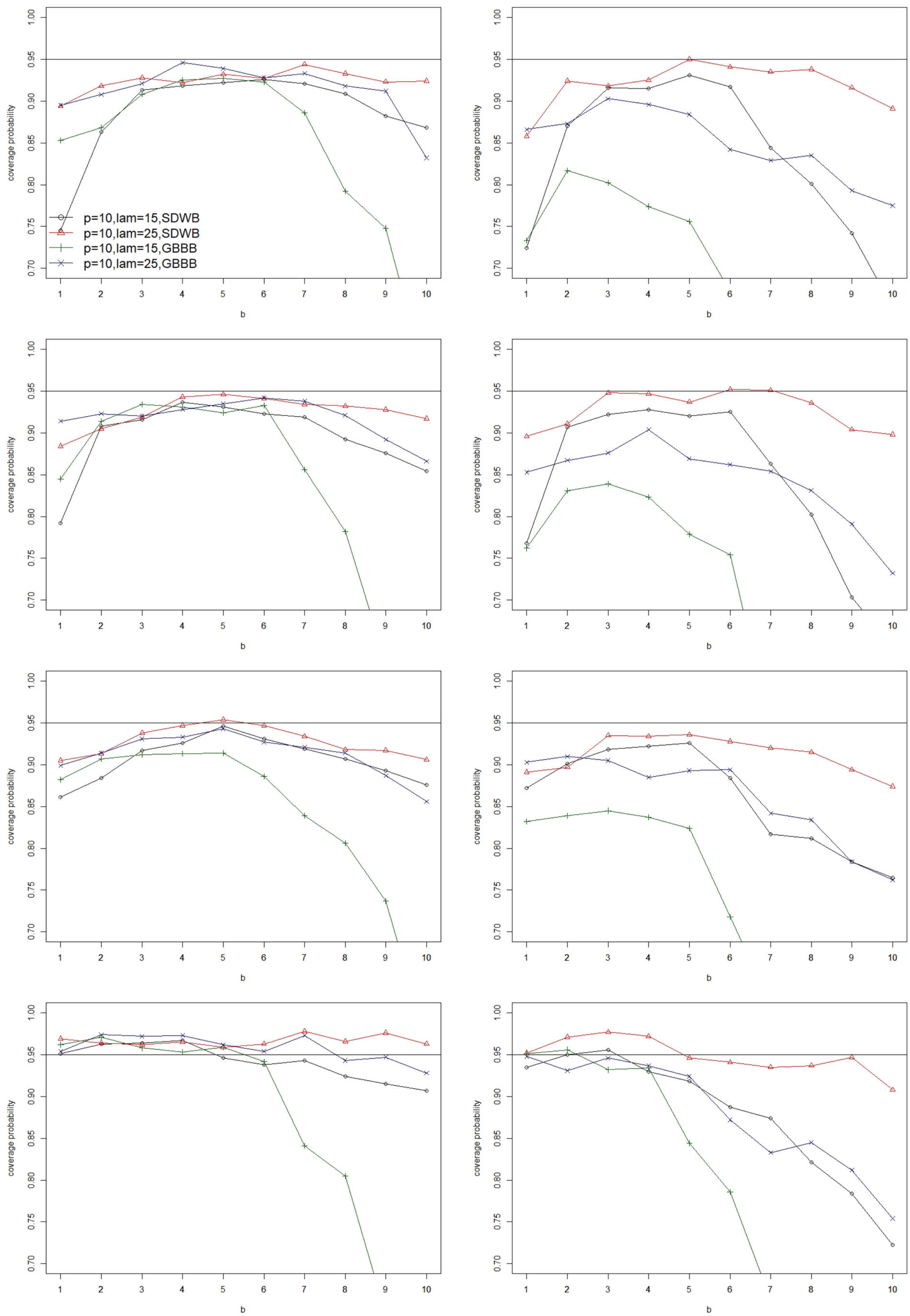

Figure 1. Coverage probabilities of SDWB and GBBB joint 95% confidence intervals for DGP1 (first row), DGP2 (second row), DGP3 (third row), and DGP4 (fourth row) with $n = 100$ and $p = 10$ under uniform (left-panel) and Gaussian (right-panel) sampling design.

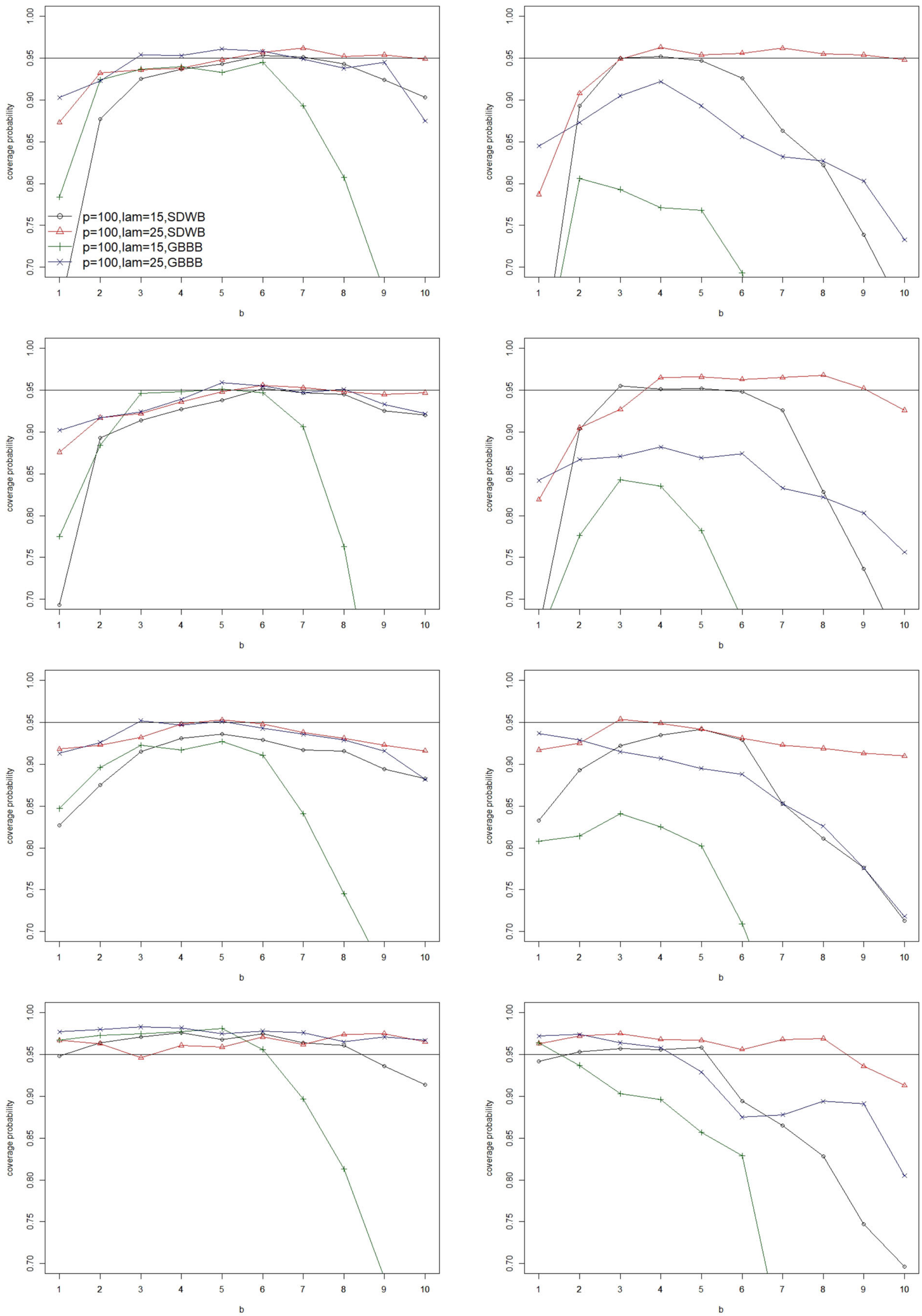

Figure 2. Coverage probabilities of SDWB and GBBB joint 95% confidence intervals for DGP1 (first row), DGP2 (second row), DGP3 (third row), DGP4 (fourth row) with $n = 100$ and $p = 100$ under uniform (left-panel) and Gaussian (right-panel) sampling design.

the mean vector $\boldsymbol{\mu}$ with $\tau \in (0,1)$ can be given by $\widehat{C}^{(BB)}_{1-\tau} = \prod_{j=1}^{p} \widehat{C}^{(BB)}_{j,1-\tau}$, where $\widehat{C}^{(BB)}_{j,1-\tau} = \left[\overline{Y}_{j,n} \pm \lambda_n^{-d/2}\widehat{q}_n^{(BB)}(1-\tau)\right]$, and $\widehat{q}_n^{(BB)}(1-\tau)$ is a quantile function defined as $\widehat{q}_n^{(BB)}(1-\tau) = \inf\left\{t \in \mathbb{R} : P_{\cdot|\mathbf{Y},\mathbf{Z}}\left(\sqrt{\lambda_n^d}\max_{1\le j\le p}\left|\overline{Y}^{*,BB}_{j,n} - \overline{Y}_{j,n}\right| \le t\right) \ge 1-\tau\right\}$. To compute the critical values $\widehat{q}_n(1-\tau)$ and $\widehat{q}_n^{(BB)}(1-\tau)$, we generate 1500 bootstrap samples for each run of the simulations. The number of Monte Carlo iterations is 1000.

We consider two types of sampling designs: uniform distribution over R_0 and a normal distribution $N((-1/4,-1/4)', 5/4I_2)$, truncated outside R_0. We set $p \in \{10, 100\}$ and $n \in \{100, 250\}$ (results for $n = 250$ are shown in the supplement). We use the Bartlett kernel for the covariance function of the Gaussian random field $\{W(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$ and examine the coverage accuracy for the bandwidth $b_n \in \{1, \ldots, 10\}$. Here b_n denotes the side length of the resampled blocks for GBBB and also the bandwidth parameter in SDWB. Figures 1 and 2 show coverage probabilities of SDWB and GBBB joint 95% confidence intervals for DGP1 (first row), DGP2 (second row), DGP3 (third row), and DGP4 (fourth row) with $n = 100$ and $p \in \{10, 100\}$ under uniform (left-panel) and Gaussian (right-panel) sampling design. We also performed simulations for $p = 400$, but the results were not much different from those for $p \in \{10, 100\}$.

A few remarks are in order. (a), Comparing the cases $\lambda_n = 15$ and $\lambda_n = 25$ for both SDWB and GBBB, we observe that the larger λ_n generally corresponds to more accurate coverages for the same combination of (n, p, b_n). This can be explained by the fact that the convergence rate of the sample mean is $\lambda_n^{d/2}$, and λ_n^d plays the role of effective sample size here. (b), for all settings, there is a broad range of b_ns that yield empirical coverage levels that are closest to the nominal one for SDWB. This suggests that in practice it is not necessary to find the optimal b_n that corresponds to the optimal coverage, but instead we only need to locate the range of b_n for which the coverage accuracy is almost optimal. By contrast, the performance of GBBB is more sensitive to b_n. When b_n is large, there is serious under-coverage, and thus the optimal block size selection becomes more critical for GBBB. (c), SDWB works for both low-dimensional (i.e., $p = 10$) and high-dimensional cases (i.e., $p = 100$), and seems to be robust to strong componentwise dependence and non-Gaussianness in view of results for DGP2, DGP3, and DGP4. (d) When we change the sampling design from uniform to truncated Gaussian, the performance for GBBB gets noticeably worse, but SDWB is less affected. (e) A direct comparison between SDWB and GBBB shows that the coverage for SDWB is closer to the nominal level than that for GBBB almost uniformly. In addition, the implementation of SDWB is much less involved than that of GBBB. Overall, the results are very encouraging as they demonstrate the advantage of SDWB over GBBB, the applicability of SDWB to low, medium and high-dimensional spatial data, and the robustness with respect to componentwise dependence, non-Gaussianness and sampling designs.

6. Conclusion

In this article, we have advanced Gaussian approximation to high-dimensional spatial data observed at irregularly spaced sampling sites and proposed the spatially dependent wild bootstrap (SDWB) to allow feasible inference. We provide a rigorous theory for Gaussian approximation and bootstrap consistency under the stochastic sampling design in Lahiri (2003a), which includes both pure increasing domain and mixed increasing domain asymptotic frameworks. SDWB is shown to be valid for a wide class of random fields that includes Lévy-driven MA random fields and the popular Gaussian random field as special cases. We demonstrate the usefulness of SDWB by constructing joint confidence intervals of the mean of random field over time, and performing change-point testing/estimation in the mean of spatio-temporal data. The validity of our Gaussian approximation and associated bootstrap theory hinges on the approximate spatial stationarity, suitable mixing and moment assumptions. Both irregularly temporal spacing and temporal nonstationarity can be accommodated in the application to inference for spatio-temporal data.

To conclude, we shall mention several important future research topics. First, an obvious one is to come up with a good data driven formula for the bandwidth parameter b_n, which plays an important role in the approximation accuracy of SDWB. For Gaussian approximation of time series and subsequent inference, a bandwidth parameter is often necessary; see Zhang and Wu (2017), Zhang and Cheng (2018), Chang, Yao, and Zhou (2017), among others, and it seems difficult to extend their data-driven formula (see e.g., Chang, Yao, and Zhou 2017) to the spatial setting. One way out is to adopt the minimal volatility approach, as advocated by Politis, Romano, and Wolf (1999b) for subsampling and block bootstrap of low-dimensional time series, and it remains to see whether it works in our setting. Second, the inference problem we study is limited to the mean of random field since our Gaussian approximation result is stated for the mean of p-dimensional spatial data. We are hopeful that our theory can be extended to cover inference for the parameter related to second order properties of a random field, such as variogram at a particular lag, given the recent work by Chang, Yao, and Zhou (2017) on testing white noise hypothesis for high-dimensional time series. Also the extension to the inference of the possibly high-dimensional parameter in spatial regression models is worth pursuing. See Zhu and Lahiri (2007). The stochastic sampling design in this article is inspired by that of Lahiri (2003a). However, one may consider other sampling designs based on point processes as considered in Garner and Politis (2013). We believe that establishing high-dimensional CLTs under different sampling design requires additional substantial works. We leave these topics for future investigation.

Supplementary Materials

The supplement contains the high-dimensional CLT under polynomial moment condition (Appendix A), discussion on examples of random fields that satisfy our regularity conditions (Appendix B), some applications of SDWB for spatial and spatio-temporal data (Appendix C), proofs of Theorems 4.1 and A.1, and Corollary 4.1 (Appendix D), proof of Theorem 4.2 (Appendix E), proof of Proposition B.1 (Appendix F), proof of Proposition C.1 (Appendix G), technical tools (Appendix H), additional simulation results (Appendix I), and real data analysis (Appendix J).

Acknowledgments

Daisuke Kurisu is Associate Professor at The University of Tokyo, Japan. Kengo Kato is Professor of Statistics at Cornell University, and Xiaofeng

Shao is Professor of Statistics at University of Illinois at Urbana-Champaign. We would like to thank three anonymous referees and an Associate Editor for constructive comments, which led to substantial improvements. We also would like to thank Adam Kashlak, Yuta Koike, Yasumasa Matsuda, Fabian Mies, Nan Zou, Taisuke Otsu and Yoshihiro Yajima for their helpful comments and suggestions.

Funding

D. Kurisu is partially supported by JSPS KAKENHI Grant Number 20K13468. K. Kato is partially supported by NSF grants DMS-1952306, DMS-2014636, and DMS-2210368. X. Shao is partially supported by NSF grants DMS-1807032 and DMS-2014018.

ORCID

Daisuke Kurisu https://orcid.org/0000-0001-8944-9098

References

Akaike, H. (1969), "Power Spectrum Estimation through Autoregressive Model Fitting," *Annals of the Institute of Statistical Mathematics*, 21, 401–419. [6]

Andrews, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858. [7]

Bandyopadhyay, S., Lahiri, S. N., and Nordman, D. J. (2015), "A Frequency Domain Empirical Likelihood Method for Irregularly Spaced Spatial Data," *Annals of Statistics*, 43, 519–545. [2,4]

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2015), *Hierarchical Modeling and Analysis for Spatial Data* (2nd ed.), Boca Raton, FL: CRC Press. [1]

Bolthausen, E. (1982), "On the Central Limit Theorem for Stationary Mixing Random Fields," *Annals of Probability*, 10, 1047–1050. [1]

Bradley, R. C. (1989), "A Caution on Mixing Conditions for Random Fields," *Statistics & Probability Letters*, 8, 489–491. [4]

——— (1993), "Some Examples of Mixing Random Fields," *Rocky Mountain Journal of Mathematics*, 23, 495–519. [5]

Brockwell, P. J., and Matsuda, Y. (2017), "Continuous Auto-Regressive Moving Average Random Fields on $\mathbb{R}^n$," *Journal of the Royal Statistical Society*, Series B, 79, 833–857. [2,3,8]

Bucchiaa, B., and Wendler, M. (2017), "Change-Point Detection and Bootstrap for Hilbert Space Valued Random Fields," *Journal of Multivariate Analysis*, 155, 344–368. [5]

Bulinskii, A. V., and Zhurbenko, I. G. (1976), "A Central Limit Theorem for Random Fields," *Doklady Akademii Nauk SSSR*, 226, 23–25. [1]

Carlstein, E. (1986), "The Use of Subseries Values for Estimating the Variance of a General Statistic from a Stationary Sequence," *Annals of Statistics*, 14, 1171–1179. [3]

Chang, J., Yao, Q., and Zhou, W. (2017), "Testing for High-Dimensional White Noise Using Maximum Cross Correlations," *Biometrika*, 104, 111–127. [2,11]

Chen, X. (2018), "Gaussian and Bootstrap Approximations for High-Dimensional u-statistics and their Applications," *Annals of Statistics*, 46, 642–678. [2]

Chen, X., and Kato, K. (2019), "Randomized Incomplete u-statistics in High Dimensions," *Annals of Statistics*, 47, 3127–3156. [2]

Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), "Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors," *Annals of Statistics*, 41, 2786–2819. [2]

——— (2014), "Gaussian Approximation of Suprema of Empirical Processes," *Annals of Statistics*, 42, 1564–1597. [2]

——— (2015), "Comparison and Anti-concentration Bounds for Maxima of Gaussian Random Vectors," *Probability Theory and Related Fields*, 162, 47–70. [2]

——— (2016), "Empirical and Multiplier Bootstraps for Suprema of Empirical Processes of Increasing Complexity, and Related Gaussian Couplings," *Stochastic Processes and their Applications*, 126, 3632–3651. [2]

——— (2017), "Central Limit Theorems and Bootstrap in High Dimensions," *Annals of Probability*, 45, 2309–2352. [2]

——— (2019), "Inference on Causal and Structural Parameters Using Many Moment Inequalities," *Review of Economic Studies*, 86, 1867–1900. [2,3]

Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2022), "Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions," *Annals of Statistics*, forthcoming. [2]

Chernozhukov, V., Chetverikov, D., and Koike, Y. (2020), "Nearly Optimal Central Limit Theorem and Bootstrap Approximations in High Dimensions," arXiv:2012.09513. [2]

Chwialkowski, K., Sejdinovic, D., and Gretton, A. (2014), "A Wild Bootstrap for Degenerate Kernel Tests," *Advances in Neural Information Processing Systems* (Vol. 27). [5]

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley. [1]

Deng, H., and Zhang, C.-H. (2020), "Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors," *Annals of Statistics*, 48, 3643–3671. [2]

Doukhan, P. (1994), *Mixing: Properties and Examples*, New York: Springer. [5]

Doukhan, P., Lang, G., Leucht, A., and Neumann, Michael, H. (2015), "Dependent Wild Bootstrap for the Empirical Process," *Journal of Time Series Analysis*, 36, 290–314. [5]

Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer. [4]

Fang, X., and Koike, Y. (2020), "High-Dimensional Central Limit Theorems by Stein's Method," *Annals of Applied Probability*, forthcoming. [2]

Gaetan, C., and Guyan, X. (2010), *Spatial Statistics and Modeling*, New York: Springer. [1]

Garner, W., and Politis, D. (2013), "The Correct Asymptotic Variance for the Sample Mean of a Homogeneous Poisson Marked Point Process," *Journal of Applied Probability*, 50, 889–892. [11]

Gelfand, A. E., and Banerjee, S. (2010), "Multivariate Spatial Process Models," in *Handbook of Spatial Statistics*, eds. P. Diggle, M. Fuentes, A. E. Gelfand, and P. Guttorp, pp. 1985–2016, Boca Raton, FL: Taylor and Francis. [1]

Gelfand, A. E., and Vounatsou, P. (2003), "Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis," *Biostatistics*, 4, 11–25. [1]

Gelfand, A. E., Vounatsou, P., Schimdt, Alexandra, M., Banerjee, S., and Sirmans, C. F. (2004), "Nonstationary Multivariate Process Modeling through Spatially Varying Coregionalization," *Test*, 13, 263–312. [1]

Guyon, X., and Richardson, S. (1984), "Vitesse de convergence du theoreme de la limite centrale pour des champs faiblement dependants," *Z. Wahrsch. Verw. Gebiete*, 66, 297–314. [1]

Hill, J. B., and Motegi, K. (2020), "A Max-Correlation White Noise Test for Weakly Dependent Time Series," *Econometric Theory*, 36, 907–960. [5]

Hounyo, U. (2022), "A Wild Bootstrap for Dependent Data," *Econometric Theory*, 39, 264–289. [5]

Jentsch, C., and Meyer, M. (2021), "On the Validity of Akaike's Identity for Random Fields," *Journal of Econometrics*, 222, 676–687. [6]

Jentsch, C., and Politis, D. (2015), "Covariance Matrix Estimation and Linear Process Bootstrap for Multivariate Time Series of Possibly Increasing Dimension," *Annals of Statistics*, 43, 1117–1140. [5]

Koike, Y. (2019), "Gaussian Approximation of Maxima of Wiener Functionals and its Application to High-Frequency Data," *Annals of Statistics*, 47, 1663–1687. [2]

Kuchibhotla, A. K., Mukherjee, S., and Banerjee, D. (2021), "High-dimensional clt: Improvements, Non-uniform Extensions and Large Deviations," *Bernoulli*, 27, 192–217. [2]

Künsch, H. R. (1989), "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241. [3]

Kurisu, D. (2022), "Nonparametric Regression for Locally Stationary Random Fields under Stochastic Sampling Design," *Bernoulli*, 28, 1250–1275. [2]

Lahiri, S. N. (1996), "On Inconsistency of Estimators Under Infill Asymptotics for Spatial Data," *Sankhya Series A*, 58, 403–417. [2]

——— (2003a), "Central Limit Theorems for Weighted Sum of a Spatial Process under a Class of Stochastic and Fixed Design," *Sankhya*, 65, 356–388. [1,2,4,6,7,11]

——— (2003b), *Resampling Methods for Dependent Data*, New York: Springer. [2,3,4,6,8]

Lahiri, S. N., and Zhu, J. (2006), "Resampling Methods for Spatial Regression Models Under a Class of Stochastic Designs," *Annals of Statistics*, 34, 1774–1813. [2,3,4,5,8]

Leucht, A., and Neumann, M. H. (2013), "Dependent Wild Bootstrap for Degenerate u and v-statistics," *Journal of Multivariate Analysis*, 117, 257–280. [5]

Liu, R., and Singh, K. (1992), "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence," in *Exploring the Limits of Bootstrap*, eds. R. LePage and L. Billard, pp. 225–248, New York: Wiley. [3]

McMurry, T., and Politis, D. (2010), "Banded and Tapered Estimates for Autocovariance Matrices and the Linear Process Bootstrap," *Journal of Time Series Analysis*, 31, 471–482. [5]

Meyer, M., Jentsch, C., and Kreiss, J.-P. (2017), "Baxter's Inequality and Sieve Bootstrap for Random Fields," *Bernoulli*, 23, 2988–3020. [5,6]

Moller, J., and Waagepetersen, R.-m. P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Boca Raton: Chapman and Hall/CRC. [1]

Nguyen, H., Cressie, N., and Braverman, A. (2017), "Multivariate Spatial Data Fusion for Very Large Remote Sensing Data," *Remote Sensing*, 9, 1–19. [1]

Paparoditis, E., and Politis, D. N. (2001), "Tapered Block Bootstrap," *Biometrika*, 88, 1105–1119. [3]

——— (2002), "The Tapered Block Bootstrap for General Statistics from Stationary Sequences," *Econometrics Journal*, 5, 131–148. [3]

Politis, D. N., Paparoditis, E., and Romano, J. P. (1998), "Large Sample Inference for Irregularlyspaced Dependent Observations based on Subsampling," *Sankya Series A*, 60, 274–292. [8]

——— (1999), *Resampling Marked Point Processes* (Vol. 159), New York: Dekker. [3,8]

Politis, D. N., and Romano, J. P. (1992), *A Circular Block-Resampling Procedure for Stationary Data,* New York: Wiley. [3]

——— (1993), "Nonparametric Resampling for Homogeneous Strong Mixing Random Fields," *Journal of Multivariate Analysis*, 47, 301–328. [3]

——— (1994), "The Stationary Bootstrap," *Journal of the American Statistical Association*, 89, 1303–1313. [3]

Politis, D. N., Romano, J. P., and Wolf, P. (1999b), *Subsampling*, New York: Springer. [8,11]

Rho, Y., and Shao, X. (2019), "Bootstrap-Assisted Unit Root Testing with Piecewise Locally Stationary Errors," *Econometric Theory*, 35, 142–166. [5]

Sato, K. (2006), "Additive Processes and Stochastic Integrals," *Illinois Journal of Mathematics*, 50, 825–851. [3]

Shao, X. (2010), "The Dependent Wild Bootstrap," *Journal of American Statistical Association*, 105, 218–235. [2,3,5,7,8]

Stein, M. L. (1999), *Interpolation of Spatial Data: Some Theory for Kriging*, New York: Springer. [1]

Wu, W.-B. (2005), "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Sciences USA*, 102, 14150–14154. [2]

Yu, B. (1994), "Rates of Convergence for Empirical Processes of Stationary Mixing Sequences," *Annals of Probability*, 22, 94–116. [2,4,7]

Yu, M., and Chen, X. (2021), "Finite Sample Change Point Inference and Identification for High-Dimensional Mean Vectors," *Journal of the Royal Statistical Society*, Series B, 83, 247–270. [2]

Zhang, D., and Wu, W. B. (2017), "Gaussian Approximation for High Dimensional Time Series," *Annals of Statistics*, 45, 1895–1919. [2,11]

Zhang, X., and Cheng, G. (2018), "Gaussian Approximation for High Dimensional Vector Under Physical Dependence," *Bernoulli*, 24, 2640–2675. [2,11]

Zhu, J., and Lahiri, S. N. (2007), "Bootstrapping the Empirical Distribution Function of a Spatial Process," *Statistical Inference for Stochastic Processes*, 10, 107–145. [3,11]