Asymptotic Procedures for a Change-Point Analysis of Random Fields

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von
Béatrice Bucchia
aus Köln

Köln 2016

Berichterstatter: Prof. i. R. Dr. Josef G. Steinebach, Universität zu Köln
Prof. i. R. Dr. Wolfgang Wefelmeyer, Universität zu Köln
Prof. Dr. Marie Hušková, Karls-Universität, Prag

Tag der letzten mündlichen Prüfung: 3. Juni 2016

Abstract

The aim of this thesis is to extend some methods of change-point analysis, where classically, measurements in time are examined for structural breaks, to random field data which is observed over a grid of points in multidimensional space. The thesis is concerned with the a posteriori detection and estimation of changes in the marginal distribution of such random field data.

In particular, the focus lies on constructing nonparametric asymptotic procedures which take the possible stochastic dependence into account. In order to avoid having to restrict the results to specific distributional assumptions, the tests and estimators considered here use a nonparametric approach where the inference is justified by the asymptotic behavior of the considered procedures (i.e. their behavior as the sample size goes towards infinity). This behavior can often be derived from functional central limit theorems which make it possible to write the limit variables of the statistics as functionals of Wiener processes, independent of the distribution of the original data.

A large part of this thesis is concerned with constructing viable asymptotic tests for an epidemic change. For a change in the mean, an asymptotic test is derived under the assumption of a functional central limit theorem. The asymptotic critical values of the test are estimated using the special form of the limit of the statistic. Estimators for the long-run variance, which influences the asymptotic distribution, are discussed. These need to be consistent under the null hypothesis, while staying stochastically bounded under the alternative hypothesis. A special focus there lies on taking a possible change in the mean into account. For a general change in the marginal distribution of the random field under mixing assumptions, the dependent wild bootstrap is introduced to construct an asymptotic test. This is achieved by constructing a test for a change in the mean of Hilbert space valued random fields and translating the change in the marginal distribution of a vector-valued random field into this setting.

Under the assumption that a change has taken place, one is interested in determining the location of the change-set. For a change in the mean over a rectangular index set, consistent estimators for the edge points of the rectangle are presented and the rate of convergence is derived. Finally, for changes in the mean over more general sets, the consistency and rate of convergence of an argmax-type estimator of the change-set are obtained under the assumption of maximal inequalities. The latter general results are illustrated by examples for classes of sets which fulfill the assumptions.

Zusammenfassung

Ziel dieser Arbeit ist die Übertragung und Erweiterung von Methoden der Change-Point Analyse, bei der klassischerweise Beobachtungen in der Zeit auf Strukturbrüche untersucht werden, auf die Anwendung auf Zufallsfelder, bei denen Beobachtungen auf Gitterpunkten im Raum gemacht werden. Die Arbeit beschäftigt sich mit a posteriori Problemen, bei denen ein gegebener Datensatz auf Strukturbrüche in der Randverteilung des Zufallsfeldes getestet und die Change-Menge gegebenenfalls geschätzt wird.

Der besondere Fokus liegt dabei auf der Konstruktion nichtparametrischer asymptotischer Verfahren, die auf stochastisch abhängige Daten anwendbar sind. Um Verteilungsannahmen an die Daten zu vermeiden, werden dabei nichtparametrische Tests und Schätzer betrachtet, deren Funktionsweise auf ihrem asymptotischen Verhalten (für wachsende

Beobachtungszahlen) beruht. Diese Asymptotik kann oft anhand von funktionalen Zentralen Grenzwertsätzen hergeleitet werden, anhand derer die Grenzvariablen unabhängig von der ursprünglichen Verteilung der Daten als Funktionale von Wiener Prozessen geschrieben werden können.

Ein großer Teil dieser Arbeit dreht sich um die Konstruktion praktisch anwendbarer asymptotischer Tests für epidemische Strukturbrüche. Für einen Strukturbruch im Erwartungswert wird ein asymptotischer Test unter der Annahme eines funktionalen Zentralen Grenzwertsatzes hergeleitet, dessen kritische Werte anhand der speziellen Form der Grenzvariable hergeleitet werden. Des Weiteren werden Schätzer für die asymptotische Varianz, welche die asymptotische Verteilung der Teststatistik beeinflusst und daher unter der Nullhypothese konsistent geschätzt werden sollte, untersucht. Dabei liegt der Fokus auf der Berücksichtigung möglicher Strukturbrüche im Erwartungswert, unter denen der Schätzer weiterhin stochastisch beschränkt bleiben sollte. Für einen allgemeinen Strukturbruch in der Randverteilung unter Mixing-Annahmen wird ein Bootstrap-Verfahren vorgestellt, anhand dessen ein asymptotischer Test konstruiert wird. Letzteres wird erreicht, indem zunächst ein Test für einen Strukturbruch im Erwartungswert von Hilbertraum-wertigen Zufallsfeldern konstruiert und dann das Problem eines Strukturbruchs in der Randverteilung mehrdimensionaler Zufallsfelder in diese Art Fragestellung übersetzt wird.

Liegt ein Strukturbruch vor, so ist man daran interessiert, die Change-Menge zu schätzen. Für einen Strukturbruch im Erwartungswert über einer rechteckigen Indexmenge werden konsistente Schätzer für die Eckpunkte des Rechtecks vorgestellt und die Konvergenzrate bestimmt. Schließlich werden die Konsistenz und Konvergenzrate eines argmax-Schätzers für Strukturbrüche im Erwartungswert über allgemeineren Change-Mengen mit Hilfe von Maximalungleichungen bestimmt. Diese allgemeinen Resultate werden durch Beispiele für Klassen von Mengen, für die die Annahmen erfüllt sind, ergänzt.

Acknowledgement

I am grateful to a number of my professors, colleagues and friends for their help and support. First, I want to thank my thesis adviser Prof. i. R. Dr. Josef G. Steinebach who introduced me to the topic of change-point analysis and suggested this interesting research topic to me. Thank you for your unwavering support and patience, and for helping me when I had questions or needed guidance. Thank you also for giving me the opportunity and encouraging me to present and discuss my work at international conferences, which was not only beneficial to my research but also an exciting experience.

Thanks also to Prof. i. R. Dr. Wolfgang Wefelmeyer and Prof. Dr. Marie Hušková for agreeing to referee this thesis.

I am also especially grateful to my colleagues in the stochastic group who created a good working atmosphere and with whom I had many enjoyable lunches and coffee breaks and engaging discussions.

I want to thank my coauthors Christoph Heuser and Prof. Dr. Martin Wendler. Christoph, thank you for your contagious enthusiasm and creativity. I had a lot of fun working with you! Martin, it was a pleasure to work with you. Thank you for letting me benefit from your knowledge and experience.

Finally, I want to thank my friends and family for their support and endless patience which carried me through many ups and downs in these last three years. Thanks especially to my brother René whose proof reading and helpful comments greatly improved my work.

To my family

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1 Change-point problems for spatial data

In the following, we give a short introduction into the change-point problems discussed here in the context of spatial data. Since the further chapters contained in this thesis give detailed explanations and citations, we will keep the present introduction general.

Given a data set of observations, a common problem for statisticians is to determine whether all the observations have the same underlying stochastic structure or whether there is a subset of the data where the structure differs. In the latter case, one is additionally interested in knowing which subset presents a change. Since this type of problem was originally discussed for one-parameter processes where the subsets over which changes might take place can be characterized by their edge points (so-called change-points), we refer to this type of problem as change-point problems or change-set problems when we want to stress the spatial nature of the data considered here. Applications of this type of statistical problem arise in various scientific fields such as image analysis, medicine, meteorology, forestry or geology.

While in the classical change-point literature the observations are commonly assumed to be measurements in time, the current work focuses on data that are observed over a grid of points in space. Such data arise, for instance, when one measures the color at each pixel of an image or in neuroimaging applications when measurements are taken at different locations of the brain.

A lot of research has been done for spatio-temporal data, where spatial data are observed at certain moments in time and the question arises if there has been a change-point (in time) where the structure has changed (cf. e.g. Majumdar et al. (2005), Aston and Kirch (2012b) and Gromenko et al. (2016)). In this setting, one can compare measurements taken at different points in time, where each measurement is a spatial process, and detect changes for asymptotically infinitely many such time points. Such problems arise whenever one takes measurements of a spatial phenomenon evolving over time, such as fMRI (functional magnetic resonance imaging) data, which is simultaneously measured at different locations of the brain and could be used to detect changes in the activation of brain regions over time. Alternatively, such data is of interest in motion detection problems. In contrast, the data considered here is measured at points in multidimensional space where we do not differentiate between time and space and consider samples whose size growth in every direction. In this setting, there is no direction which is a priori distinguished as the direction of interest for a change and therefore the change-point problem becomes a change-set problem, where one is interested in testing for general subsets of the data with a changed stochastic structure. Furthermore, — unlike, for instance, in problems when two given images are to be tested for dissimilarities — we do not assume that we have a template for the spatial data with respect to which changes can be measured. Instead, changes within a single realization of a spatial process are discussed.

Applications for this kind of problem include statistical image processing and detection of edges in noisy images, where one is, for instance, interested in detecting specific objects in the image or in distinguishing between foreground and background of the image. For land use related data obtained by remote sensing, one might test for the presence of different land uses and subsequently estimate the location of the different sub-regions.

Since it has a straightforward generalization to multidimensional parameters, the epidemic change problem plays an important role in this thesis. For time series, an epidemic change occurs if there are two change-points such that the stochastic structure changes after the first and reverts back to its original state after the second change-point. An early application of these stochastic considerations in medicine motivates the terminology: the period of time between the change-points represents an epidemic, during which incidence of a disease is structurally more likely, and this epidemic is preceded and followed by periods of normalcy, i.e. (relative) health. For multiparameter processes, this translates into a changed structure on a rectangular subset of the data (often also called a block), and the term epidemic change is retained due to the similarity of this problem to its one-dimensional parameter counterpart. For a d-dimensional rectangle $(k_1, m_1] \times \cdots \times (k_d, m_d]$ $(d \in \mathbb{N})$, we call the points $(k_1, \ldots, k_d)$ and $(m_1, \ldots, m_d)$ the edge points of the rectangle and also refer to them as change-points. Changes of this form are of interest e.g. in the detection of vehicles or buildings in aerial images or generally in object detection problems whenever the object to be detected has a rectangular shape.

The extension of change-point methods to spatial data has received a lot of attention in the literature. Hahubia and Mnatsakanov (1996) discuss the asymptotic behavior of test statistics and estimators based on general set-indexed models which in particular include the partial sum processes for change-point problems in time or space. Khmaladze et al. (2006a) (cf. also Khmaladze et al. (2006b)) consider change-set estimators for changes in the conditional marginal distribution of a sequence of i.i.d. observations (X_i, Y_i) $(i = 1, 2, \ldots)$ that each consist of a location X_i in space and a mark Y_i, where the distribution of Y_i conditioned on X_i takes on different values depending on whether or not the location X_i is within a change-set. In contrast, we use a deterministic design and the data considered here are random fields, i.e. observations are assumed to lie on a discrete grid in space with rectangular mesh. This type of model was used e.g. by Carlstein and Krishnamoorthy (1992) and Ferger (2004) for the estimation of the change-set and Xie (1996) for corresponding testing problems.

A lot of the literature focuses on stochastically independent observations. However, this assumption is too restrictive for many applications where spatial autocorrelation is an unavoidable feature of the problem. For instance, as noted in Griffith and Layne (1999), spatial dependence is introduced into yield data by external influences such as rainfall or slope position. One main aim of the present work is therefore to derive a theory that is applicable to weakly dependent data. There are various ways to define weak dependence conditions (cf. e.g. the monographs by Dedecker et al. (2007) and Bulinski and Shashkin (2007)). In contrast to works such as Brodsky and Darkhovsky

(1993) and Puri and Ruymgaart (1994), we aim to avoid as much as possible restricting the analysis to a specific type of dependence. Instead, we use functional central limit theorems or moment assumptions for the partial sums, giving examples for types of dependence under which these assumptions are fulfilled. An exception to this approach is the paper Bucchia and Wendler (2015), where all the results were derived under mixing assumptions.

The change-point problem discussed here is an a posteriori change problem. Unlike in the sequential change problem where observations are continuously gathered and the new observations are tested for changes with respect to the old ones, in a posteriori analysis one considers a given finite data set and examines it for changes in the stochastic structure.

There are several approaches to modeling a change in the stochastic structure of the data. For a general change in the marginal distribution, one is usually interested in the distribution function of the observations. More specific changes in the stochastic structure are commonly modeled by using parameters of interest for the distribution (e.g. the mean or variance of the random variables or the correlation between them) and investigating these for changes — often under the additional assumption that the other parameters of the distribution remain constant. Under the assumption that the observations belong to a specific parametric family of distributions, specialized approaches for testing for changes in the parameters of the distribution can be used. This type of approach for spatial data was e.g. discussed in MacNeill and Jandhyala (1993), who used a Bayes-type derivation method for changes in the parameter of a one-parameter exponential family of distributions, and Ivanoff and Merzbach (2010) and Jarušková and Piterbarg (2011), who considered changes in the intensity of Poisson processes.

This thesis follows a nonparametric approach where the observations are not assumed to belong to a specific family of distributions or follow specific parametric data generation models such as (spatial) autoregression. The validity of the presented tests and estimators is based on their asymptotic behavior for sample sizes going to infinity.

The main focus of this work lies on changes in the mean where all other parameters of the distribution are unaffected by the change. As described in Brodsky and Darkhovsky (1993), changes in the mean are of particular interest to statisticians because oftentimes changes in other parameters of the marginal distributions can be translated into changes in the mean, thus making the methods developed for changes in the mean applicable to other types of change. Changes in the mean have received widespread attention in the change-point literature as can be seen e.g. from the monographs by Brodsky and Darkhovsky (1993) and Csörgő and Horváth (1997) or the more recent overview article by Aue and Horváth (2013). For the change in the mean problem, the observations are usually modeled as the sum of a deterministic mean function and a centered stochastic process. One advantage of this model is that the break from stationarity can thus be restricted to the deterministic part, making it possible to assume (weak) stationarity for the stochastic part of the process. Furthermore, having a single underlying stochastic process simplifies the specification of the stochastic dependence between the observations, since it is unaffected by the change. This type of model reflects the interpretation of the data as being the sum of a signal (the mean function) and additional error terms that introduce noise. Such noise terms might be used to account for measurement errors or inaccuracies. For example, one might be trying to reconstruct an image based on a grainy version of the image.

If additional information on the underlying structure of the signal is known, a popular model for the mean function is to view it as a linear combination of known functions (which might even be nondeterministic) with unknown coefficients. Then, a change in the mean can be viewed as a special case of a change in the coefficients of such a linear regression.

After having described the general topic of this thesis, we now discuss the challenges involved in the testing and the estimation problem separately.

Tests for changes in spatial data. In this thesis, we follow a global approach where we do not try do classify every single observation as changed or not changed but rather test the whole data for the presence of a change over an unknown subset of the data. For this, asymptotic tests are derived to distinguish between the null hypothesis of stationarity and epidemic change alternatives.

The basic idea behind the change-point tests considered is the same for multiparameter processes as for time series data. Only, instead of using points in time where there might be a change to divide the data into subsets, one has to specify a class of candidate sets that each divide the (multidimensional) index set into a possible change-set and its complement. For the epidemic change problem, these candidate sets are rectangles with sides parallel to the coordinate axes.

We consider the following testing problem for changes in the mean. Given n^d ($n \in \mathbb{N}$) observations on a d-dimensional regular grid of side length n, we want to test the null hypothesis that the observations correspond to a random field with constant (but unknown) mean μ against the epidemic alternative that there exists an unknown subrectangle of the grid where the mean is $\mu + \delta$ (for an unknown δ). For real-valued observations, one can then differentiate between tests for positive change heights δ or two-sided tests that make no restriction on the sign of the change. In this thesis, the latter type of tests are considered. We use CUSUM (cumulative sum) procedures that are based on the partial sums of the observations over rectangular subsets of the data. Statistics of CUSUM-type are widely popular in change-point tests for time series (cf. e.g Csörgő and Horváth (1997) or Aue and Horváth (2013)). For changes in the mean, they correspond to the following heuristic: A common approach for testing for changes in a parameter is to first divide the data into two subsets. Assuming for a moment that the parameter is constant on each of these subsets, one can then estimate it on each subset separately and use the difference of the estimators as a measure of the presence and size of the change. Since the true change-set is unknown, this procedure is used on all candidates for the change-set and the resulting test statistics are aggregated by taking the maximum of all the statistics for the corresponding two-sample problems. Additionally, weighting functions can be used to facilitate the detection of changes over particularly small or large subsets. However, depending on the weighting function, obtaining meaningful asymptotic results might make it necessary to trim the statistic, i.e. only consider change-sets of a certain size. The latter is the case for the statistic discussed in Bucchia (2014), whose specific form was motivated by a pseudo log-likelihood approach where the data are assumed to be normally distributed for the derivation of the statistic (cf. Bucchia (2012)).

Since the distribution of the statistic for finite sample sizes is unknown and depends on the specific distribution of the data, we use asymptotic tests which are based on the asymptotic distribution (for $n \to \infty$) of the statistic under the null hypothesis. For the CUSUM-type test statistics considered in this thesis, this can be achieved by writing the test statistic as a (continuous) functional of the partial sum process of the data. In analogy to the time series case, where the partial sum process $\{S_n(t)\}_{t\in[0,1]}$ is the process of (rescaled) sums $S_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i$ (for observations $X_1, \dots, X_n$), we consider (rescaled) sums over indices $\{1, \dots, \lfloor nt_1 \rfloor\} \times \cdots \times \{1, \dots, \lfloor nt_d \rfloor\}$ for $(t_1, \dots, t_d) \in [0,1]^d$. The resulting partial sum process has sample paths in the multiparameter Skorohod space $D[0,1]^d$. Under various dependence assumptions on the random field, a lot of research has been done on functional central limit theorems (cf. e.g. Bucchia (2014) and the references contained therein) which yield the weak convergence of this type of process to a Brownian sheet.

Using the continuous mapping theorem, we can then derive the weak limit of the statistic. Its distribution is independent of the (unknown) distribution of the original data (up to nuisance parameters that need to be estimated). An asymptotic test of level α is constructed by taking the limit of the test statistic under the null hypothesis and rejecting the null if the $(1-\alpha)$-quantile of the limit distribution is exceeded. If the test considered has asymptotic power one, we call it consistent. Consistency of the test can e.g. be proven by showing that the test statistic diverges to infinity under the alternative. Then, any choice of critical value will eventually be exceeded.

The general procedure described above can in principle be applied not only to real- or vector-valued observations, but more generally to observations with values in a Hilbert space. Such observations are, for instance, of interest whenever the observations made on the spatial grid are functions themselves, as is e.g. the case in some brain imaging or spatial physics problems.

Another application of the Hilbert space framework consists of changes in the marginal distribution of vector-valued observations: For this type of problem, the empirical distribution function replaces the partial sums as a natural indicator of a change. Using the fact that the empirical distribution function can be viewed as a random element of a suitable Hilbert space, the change in distribution problem for vector-valued random fields can be translated into a change in the mean problem in the chosen Hilbert space.

A lot of change-point tests were originally constructed for i.i.d. observations. When one introduces dependence, certain modifications are necessary to take the autocorrelations into account. In the approach described above, when one assumes that the dependence is such that the functional central limit theorem is still fulfilled, the main change is in the asymptotic variance of the partial sums (the so-called long-run variance), which needs to be estimated. Where in the independent case correcting for the variance of the partial sums boils down to estimating the variance of a single random variable, one now needs to estimate the sum of autocovariances. The idea is to use an estimator of the long-run variance as a rescaling factor in the statistic. If this estimator is consistent under the null, the limit variable will then be independent of the long-run variance. Many of the long-run variance estimators discussed in the literature focus on processes with constant (or at least continuous) mean functions. However, since by definition the mean has an abrupt jump under the alternative, these procedures might not be consistent (or at least stochastically bounded) under the alternative, which could in turn have a negative effect on the power of the resulting change-point test. One approach to deal with this problem is described in Bucchia and Heuser (2015).

Unfortunately, in most cases, the limit variable has a distribution whose quantiles are not theoretically known, even after eliminating the long-run variance as a nuisance parameter, and the problem of determining — or at least approximating — the $(1-\alpha)$-

quantile of the limit distribution remains. Most such limit distributions have not been tabulated yet, so one needs to develop new approaches for the approximation. One possibility would be to use a Monte-Carlo simulation of the limit variable to estimate its quantiles. However, due to the large number of sets over which the supremum is taken in the limit, this quickly becomes computationally intensive. Another possibility is to extend already existing approximation techniques for the time series case to multiparameter processes, as was done, for instance, in Xie (1996), who adapted a method by Eastwood (1993) to obtain approximations for the limiting distribution by considering the distribution of a suitable chi-square random variable.

In the present work, two approaches to derive critical values are presented. The method employed in Bucchia (2014) was introduced by Jarušková. Jarušková (2011) developed an approximation method for the limit distribution of statistics arising from change-point tests for multiple change-points. It is based on the fact that the limit variable she considered is a maximum of a normalized multiparameter Gaussian random field and thus the methodology by Piterbarg (1996) for the approximation of its tail probabilities can be used for this context. Jarušková and Piterbarg (2011) then applied this procedure to test for epidemic changes in i.i.d. random fields.

This method is, however, not easily applicable to Hilbert space valued processes. Additionally, the estimation of the long-run variance in this case is much more involved than for real- or vector-valued observations. Following the approach discussed in Sharipov et al. (2016), a nonparametric resampling method is introduced to solve both these problems. For this, a variant of the dependent wild bootstrap by Shao (2010) is introduced for random fields. The idea is to replicate the asymptotic behavior of the partial sum process by using a weighted version of the process where the weights are random variables which are independent of the original data but fulfill certain dependence assumptions. One can then consider a bootstrapped change-point statistic which is based on the bootstrapped partial sum process. For given observations, the (conditional) empirical quantiles of the bootstrapped statistic can be obtained from Monte-Carlo simulations. Then, if one can show that the original statistic and the bootstrapped statistic jointly converge to the same distribution, the (conditional) quantiles of the bootstrapped statistic converge to the quantiles of the limit process. Thus, the empirical quantiles can be used as critical values for the test.

While a lot of publications derive results for Hilbert space valued observations by projecting the statistics onto finite-dimensional subspaces, the approach described above has the added advantage that no such projection is necessary.

Change-set estimation. When a change in the stochastic structure of the data has been detected, a further question is the location of the change. As described above, for spatial data, this location is given as a subset of the index set, the change-set. In this thesis, estimation procedures both for epidemic changes — where the change-set can be characterized by its edge points and the change-set estimation thus becomes a change-point estimation — and more general change-sets are considered.

We discuss change-set estimators for abrupt changes in the mean function of either real- or vector-valued random fields. The mean function is modeled as a step function with one value inside the change-set and a different value outside of it. Using an analogous approach as for the detection of changes, we consider maxima of set-indexed CUSUM-type statistics that correspond to weighted differences of sample means over chosen candidate sets. Then, maximizers of the CUSUM-statistic are chosen as change-set estimators. For the special case of epidemic changes, we estimate the edge points directly by treating the CUSUM-type statistic not as a set-indexed functional but as a function of the edge points defining the rectangle. As a special case, this procedure reduces to an estimator of the form considered in Aston and Kirch (2012a) when one limits oneself to one-parameter processes.

Argmax estimators of this form were also considered in Müller and Song (1994), who derived convergence rates for the estimator for unions of rectangles as change-sets and independent observations, and Brodsky and Darkhovsky (1993), who proved the consistency of the estimator for parametric families of change-sets under mixing conditions.

Since choosing whether to label a subset or its complement the change-set is essentially arbitrary in most cases, a lot of research has focused on the so-called change-boundary problem, where one estimates the change-set's boundary, rather than the set itself. This is in direct analogy to the time series case, where the change-points, which characterize the boundaries of sets without structural breaks, are estimated. For spatial data, one needs to specify how the index set is to be segmented into subsets with different stochastic properties based on change-boundaries. In the present work, this is achieved by considering classes of subsets of the index set as candidate sets and defining the change-boundary to be the common topological boundary of a candidate set and its complement.

The estimation of change-sets or -boundaries has applications in image analysis. Assuming a signal plus noise model for the image, where the observed data is the sum of a noise process and a regression function, reconstructing the image amounts to estimating the regression function. The change-boundaries correspond to curves or edges in the image where the regression function has discontinuities. Since a lot of regression estimators (as described e.g. in El Machkouri (2007)) assume the continuity of the regression function, a first step in the reconstruction of the image is to segment the index set into subsets where the regression function is continuous.

By contrast, an application where there is a distinction between the change-set and its complement is the related problem of threshold estimation, where one assumes that the mean function $\mu(\cdot)$ has a constant value τ on a subset S of the index set and is strictly greater than τ on the complement of S. Mallik (2013) considered such problems for convex sets S in two dimensions and derived estimators based on minimizing the p-values of corresponding testing problems for $\mu(\mathbf{x}) = \tau$ against $\mu(\mathbf{x}) > \tau$ at each index point $\mathbf{x}$. He noted that this problem is closely related to level-set estimation, since the estimation of the complement $S^c = \{\mathbf{x} : \mu(\mathbf{x}) > \tau\}$ of S could also be viewed as the estimation of a level-set for the level τ.

In keeping with the global approach used for the testing problem, we do not investigate the probability of misclassification for a single grid point but measure the distance between the change-set and its estimator using the symmetric difference of sets, which gives a measure of the number of misclassified points. For the change-boundary estimation considered here, the minimum of the distances to a specified change-set and its complement is used. Alternatively, for the special case of the estimation of the edge points of rectangular change-sets, we use the distance between the true and the estimated change-points as a measure for the accuracy of the estimation procedure. Using such distance measures, the consistency (i.e. stochastic convergence to zero of the distance between the estimator and the true value) and rate of convergence of the estimation are discussed for sample sizes going to infinity.

In order to clarify how the present work fits into the general framework described above and give a more detailed account of the content of this thesis, we now give brief summaries of the material contained in each chapter of this thesis.

2 Chapter summaries

The thesis consists of three articles, two of which are already published while the third is under review, and two chapters of additional (unpublished) material, one containing additional results related to the ones from the articles, the other a discussion of the entire project. The first article gives both testing and estimation procedures for epidemic changes in real-valued random fields, the second discusses in detail the estimation of a nuisance parameter that arises in the testing problem and the third focuses on the testing problem for epidemic changes for Hilbert space valued observations, introducing a bootstrap method to derive asymptotic critical values. A further chapter contains additional material on the estimation of change-sets when the assumption that the change takes place over a rectangle is relaxed. Following this, a final chapter discusses the results obtained throughout these four chapters. As the chapters 2 to 5 are each essentially self-contained, the numbering of equations, theorems etc. starts anew in each chapter. The chapters 2, 3 and 4 correspond to Bucchia (2014), Bucchia and Heuser (2015) and Bucchia and Wendler (2015), respectively, and the bibliographic references and chapter names are used interchangeably.

Testing for epidemic changes in the mean of a multiparameter stochastic process
By Béatrice Bucchia

The article discusses the epidemic change in the mean problem for real-valued random fields and treats both the associated testing and estimation problems.

The asymptotic behavior of CUSUM-type statistics can be inferred from the behavior of the partial sum process. Therefore, the article employs a slightly more general model for the data, which encompasses both partial sums and set-indexed Lévy processes. All the asymptotic results are derived under the assumption of an invariance principle for which several examples are given. For the testing problem, a trimmed maximum type test statistic is presented. Given the assumption that an invariance principle is fulfilled, its limit distribution under the null hypothesis of no change is derived. For practical use of the statistic, an approach by Jarušková and Piterbarg (2011) is employed to approximate the tail behavior of the limit distribution. The methodology for this approach was introduced by Piterbarg (1996) and has since been used e.g. by Jarušková (2011), Jarušková and Piterbarg (2011), Jarušková (2015) for change-point tests in various settings. It takes advantage of the fact that the limit variable is the maximum of a homogeneous Gaussian field over a compact set.

A proof of consistency of the test under alternatives that do not vanish too fast is given.

In order to construct a test, the unknown long-run variance, which plays a role in the asymptotic distribution, needs to be estimated. After first describing the asymptotics under the assumption that a fitting estimator is available, a kernel-type estimator for the long-run variance is discussed.

As seen in the proof of consistency, under the epidemic alternative, the test statistic becomes asymptotically large at the change rectangle. Therefore, for constant change heights, a natural approach to construct a change-point estimator is to take the maximizer of the test statistic. For such an argmax-type estimator, the consistency is shown.

The theoretical results are complemented by a short simulation study that investigates the finite sample behavior of the test and estimators for both independent observations and weakly dependent moving average observations. The simulation results show that the estimator works well, with increasing accuracy for larger change-sets. The constructed test is (empirically) conservative under the null and has high power against the alternative hypothesis, even though the long-run variance is underestimated under the null and often overestimated under the alternative.

Although the article is based on results presented in Bucchia (2012), it extends the work presented therein in several ways. A different approach to modeling the process, which directly focuses on the partial sum process, was chosen. For this, several examples of processes that fulfill the main assumption were added. The derivation of asymptotic critical values using the method of Jarušková and Piterbarg (2011) and the discussion of change-point estimators, which where restricted to the one- and two-dimensional cases in Bucchia (2012), have been expanded to general dimensions d. The change-point estimation procedure for multiparameter processes was further extended from the case of positive changes to general change heights. Finally, the section on the estimation of the long-run variance and the simulation study are new.

Long-run variance estimation for spatial data under change-point alternatives

By Béatrice Bucchia and Christoph Heuser

As seen in Bucchia (2014), a common problem when constructing asymptotic change-point tests is the estimation of nuisance parameters which need to be determined in order to obtain the quantiles of the limiting random variable. Since the critical values are derived under the null hypothesis, most of the literature focuses on estimators which are consistent under the assumption of mean functions without discontinuities.

In contrast, this article is concerned with the estimation of the long-run variance (matrix) of a weakly dependent random field, with a special focus on the estimators' behavior under the alternative of mean functions with jumps. This is of particular interest because these estimators are often used as normalizing factors in change-point tests and therefore the overestimation of the long-run variance might lead to tests with less power against change in the mean alternatives.

The mean functions considered correspond to a single change in mean model: They take on two values, one inside a change-set and one for all indices outside of the change-set. The change-sets considered are finite unions of rectangles.

The classical kernel-type estimation of the long-run variance uses the weighted sum of estimators of the autocovariances of the process. In this setting, the arithmetic mean over the observations is used as an estimator of the mean function. Since, by assumption, the mean does not stay constant under the alternative, a natural idea is to replace the arithmetic mean — which is based on a constant mean assumption — by a mean estimator that reflects this. Therefore, this paper introduces a variation of the classical kernel-type long-run variance estimator with a different mean-function estimator, which uses a change-set estimator as an approximation of the unknown change-set.

Throughout the paper, the asymptotics for the classical long-run variance estimator as well as the modified estimator are developed in parallel. As anticipated, both estimators are consistent for constant mean functions, but while the classical estimator diverges for non-vanishing change heights, the modified estimator allows bandwidth choices that make it consistent under both constant means and changes in the mean. The error rate for the latter depends on the convergence rate of the change-set estimation.

The paper first discusses the long-run variance estimation under the assumption that an unspecified change-set estimator, which converges to the true change-set with a given rate, is available and then gives an example for such an estimator for rectangular change-sets. A convergence rate for the estimator is derived.

In the paper's final section, a simulation study gives a comparison of the empirical behavior of the long-run variance estimators with and without the modification of the mean function estimator. The simulations show that while the behavior of both estimators depends strongly on the choice of bandwidth, the newly introduced estimator exhibits the expected robustness with respect to changes. It leads to change-point tests with higher false rejection rates under the null hypothesis but also higher empirical power under epidemic alternatives. Finally, the methods are applied to a brain tumor detection problem.

Change-point detection and bootstrap for Hilbert space valued random fields

By Béatrice Bucchia and Martin Wendler

The aim of the article is twofold: The epidemic change in the mean problem for multiparameter processes with values in a Hilbert space is discussed, and a variant of the dependent wild bootstrap by Shao (2010) is introduced and its consistency is proven.

In contrast to the rest of this thesis, we do not assume an invariance principle and derive results under general weak dependence assumptions, but assume specific ρ- and α-mixing conditions and show that all the results hold under these. To this end, a functional central limit theorem for Hilbert space valued random fields under mixing conditions is proven.

Using the functional central limit theorem, the limit variable of a CUSUM-type test statistic for the change in the mean problem is derived. To make the test applicable without having to estimate the long-run variance operator, the dependent wild bootstrap is adapted to this context. Since the bootstrapped version of the test statistic should not be sensitive to changes in the mean but rather retain its behavior under the alternative, both the sample mean and an estimator for the mean function which takes possible epidemic changes into account are discussed for the bootstrapped process. The validity of the bootstrap is shown by deriving the joint weak convergence of the partial sum and the bootstrapped partial sum process to Wiener processes which are independent copies of each other. Thus, the continuous mapping theorem yields the joint weak convergence of the test statistic and the bootstrapped test statistic. Since the bootstrapped statistic therefore mimics the behavior of the original statistic, asymptotic critical values can be derived by Monte-Carlo approximation of the quantiles of the bootstrapped statistic.

While the problem of changes in the mean of functional data is of independent interest, the paper is also concerned with an application of the derived theory to testing for a change in the marginal distributions of a vector-valued random field. A test statistic based on the empirical distribution function is presented, whose convergence can be shown by using the Hilbert space theory previously described.

Finally, the results of a small simulation study are discussed, which shows that although the procedures considered show the typical over-rejection of bootstrap tests under the null hypothesis, they have good empirical power against epidemic changes in the mean or in the skewness of a random field.

Additional Material: Change-Set Estimation

For most of the thesis, the change-set estimation is restricted to rectangular sets. This chapter aims to relax that assumption.

Given a data set on a multidimensional grid, the problem of change-set estimation for $\mathbb{R}^p$-valued multiparameter processes is studied. Under the assumption that there is a change in the mean on a subset of the data, an estimator for this change-set is presented and results for its consistency and rate of convergence for general classes of sets and weakly dependent observations are obtained. As seen in Bucchia and Heuser (2015), the rate of convergence — as a measure of the number of misclassified data points — is of interest, in particular, for plug-in estimators of the mean function.

As a measure of stochastic dependence, moment inequalities for partial sums are assumed. The change-set estimator is a maximizer of a set-indexed process based on weighted differences of sample means over points inside and outside of candidate sets. Since the data is observed on a grid, the aim is to estimate the grid points contained in the change-set. Therefore, the number of misclassified grid points is used as a measure of the distance between the estimator and the true set. In parallel to the change-set estimation, the related problem of the estimation of the change-boundary is studied.

After introducing the model and main assumptions, the main results concerning the consistency and convergence rates for general classes of sets are presented. For these, maximal inequalities are assumed and, for the change-set estimation, identifiability assumptions are used. The latter are necessary for change-set estimation to ensure that the change-set and not its complement is estimated. The results are therefore supplemented by several remarks giving examples of classes of sets and stochastic processes that fulfill the assumptions. In an additional section, the results are applied to specific classes of sets. As a byproduct, some maximal inequalities, and in particular an exponential inequality under mixing assumptions, are derived.

Chapter 2

Testing for epidemic changes in the mean of a multiparameter stochastic process

Béatrice Bucchia
University of Cologne

Abstract

In this paper, multiparameter stochastic processes $\{Z_n(\mathbf{x})\}_{\mathbf{x}\in[0,n]^d}$, $n \in \mathbb{N}$, are tested for the occurrence of a block $(\mathbf{k}_0, \mathbf{m}_0] = (k_{0,1}, m_{0,1}] \times \cdots \times (k_{0,d}, m_{0,d}] \subset [0,n]^d$ where the mean of the process changes. This is modeled in the form

$$Z_n(\mathbf{x}) = \lambda((\underline{\mathbf{0}}, \mathbf{x}])\mu_n + \sigma Y(\mathbf{x}) + \lambda((\underline{\mathbf{0}}, \mathbf{x}] \cap (\mathbf{k}_0, \mathbf{m}_0])\delta_n,$$

where $\underline{\mathbf{0}} = (0, \ldots, 0)'$, $\lambda(A)$ denotes the Lebesgue measure of a set $A \subset \mathbb{R}^d$, and $\mu_n, \delta_n \in \mathbb{R}$ as well as $0 < \sigma < \infty$ are unknown parameters. The stochastic process $\{Y_n(\mathbf{t}) = Y(\lfloor n\mathbf{t} \rfloor) : \mathbf{t} \in [0,1]^d\}$ is assumed to fulfill a weak invariance principle. Under the null hypothesis, an approximation for the tail behavior of the limit variable of a trimmed maximum-type test statistic is given. Then, the (weak) consistency of the test under the alternative is proven. The corresponding estimation problem for the points $\mathbf{k}_0$ and $\mathbf{m}_0$ is also considered and consistent estimators are given for local alternatives $\delta_n = \delta\, n^{-d/2}$ with $\delta > 0$.

Keywords: change point detection, trimmed maximum-type test statistic, maxima of Gaussian fields, invariance principle, change point estimation

AMS subject classification: 62H15, 62E20, 62M99, 60G60, 62H12, 60F17

1 Introduction

This paper deals with the problem of detecting epidemic changes over a block. Assuming that we have observed values $\{X_{\mathbf{j}} : \mathbf{j} \in \{1, \ldots, n\}^d\}$ of a random field (where $d \in \mathbb{N}$ is fixed and small relative to n), we may ask whether these observations all have the same mean

μ_n, or whether there is a block $(\mathbf{k}_0, \mathbf{m}_0] = (k_{0,1}, m_{0,1}] \times \cdots \times (k_{0,d}, m_{0,d}]$ over which the mean has changed to a value $\mu_n + \delta_n$. Such a change point problem is the straightforward generalization to the multiparameter case of a one-dimensional change point problem with two change-points $0 < k_0 < m_0 < n$. Levin and Kline (1985) coined the term epidemic change for the latter in their paper about the connection between chromosomal abnormalities and the number of spontaneous abortions. In this medical context, the term epidemic change corresponds to a period of normal behavior, followed by a sudden change in patient numbers and finally by the return to normalcy. The epidemic change problem for processes with one-dimensional parameter space and independent observations has been the subject of several research papers, for example by Yao (1993), Antoch and Hušková (1996), Račkauskas and Suquet (2004) and Jarušková (2011), who studied several test statistics, and Hušková (1995), who considered estimators for the change points (cf. also Csörgő and Horváth (1997) and Brodsky and Darkhovsky (1993)). The change-point problem considered here, namely a change in the mean over a block in the index-space of a random field, was also studied by Jarušková and Piterbarg (2011) and Zemlys (2008). In both of these publications, the asymptotic distributions of the considered test statistics are determined by the fact that the random variables are independent and therefore the associated partial sum processes converge weakly to a Wiener process. Thus, the process of interest for the statistical analysis is not the original random field, but rather the associated partial sum process $Z_n(\mathbf{x}) = n^{-d/2} \sum_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \lfloor \mathbf{x} \rfloor} X_{\mathbf{j}}$. If we denote the Lebesgue measure by λ, a change over the block $(\mathbf{k}, \mathbf{m}]$ then corresponds to a change $\lambda((\mathbf{k}, \mathbf{m}] \cap (\underline{\mathbf{0}}, \lfloor \mathbf{x} \rfloor])\delta_n$. Due to this fact, we have chosen a model that includes the partial sum process as an example and replaced the independence assumption by the weaker assumption that an invariance principle be fulfilled. The statistic we use for change detection was inspired by the trimmed pseudo log-likelihood statistic employed by Jarušková and Piterbarg (2011) and adapted to our model. This approach differs from the one in Zemlys (2008), where a different weight function was used instead of a trimmed maximum. Examples of the problem of detecting inhomogeneity arise in image analysis and in textile fabric quality control (e.g. Zhang and Bresee, 1995). In particular, the search for an inhomogeneity in the shape of a rectangle might be of interest in the context of rectangular shape object detection problems. For instance, finding rectangular objects in an image is a step in the detection of buildings or vehicles from aerial imagery (Vinson et al., 2001; Vinson and Cohen, 2002; Moon et al., 2002), license plate detection (Kim et al., 2002; Huang et al., 2008) and in the detection of filaments in cryoelectron microscopy images (Zhu et al., 2001).

The structure of this paper is as follows: First, we introduce a few notations and describe the model we chose. Then, in the third section, we treat the change detection problem by studying the behavior of a test statistic under the null and the alternative hypotheses. Finally, using a similar approach to the one in Aston and Kirch (2012a), we give consistent estimators for the boundary points of the changed block under the alternative. A final section is devoted to a small simulation study in order to give some idea of the finite sample behavior of the suggested procedures.

2 The model

First, we introduce some notations that will be used throughout this paper. We consider the vector space $\mathbb{R}^d$ ($d \in \mathbb{N}$) equipped with the usual partial order. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we write $\mathbf{x}\vee\mathbf{y} = (\max\{x_1, y_1\}, \ldots, \max\{x_d, y_d\})'$ and $\mathbf{x}\wedge\mathbf{y} = (\min\{x_1, y_1\}, \ldots, \min\{x_d, y_d\})'$ as well as $\lfloor\mathbf{x}\rfloor = (\lfloor x_1\rfloor, \ldots, \lfloor x_d\rfloor)'$ for the integer part of $\mathbf{x}$, $|\mathbf{x}| = (|x_1|, \ldots, |x_d|)'$ and $[\mathbf{x}] = x_1 \cdots x_d$. Furthermore, for any integer $k \in \mathbb{N}_0$, we denote $(k, \ldots, k)' \in \mathbb{N}_0^d$ by $\underline{\mathbf{k}}$. For a set $A \subset \mathbb{R}^d$, a vector $\mathbf{x} \in \mathbb{R}^d$ and a number $y \in \mathbb{R}$, the sets $A+\mathbf{x}$ and yA are defined as

$$A + \mathbf{x} = \{\mathbf{a} + \mathbf{x} : \mathbf{a} \in A\}$$

and

$$yA = \{y\mathbf{a} : \mathbf{a} \in A\}.$$

A block in $\mathbb{R}^d$ is a set of the form

$$(\mathbf{x}, \mathbf{y}] = \{\mathbf{z} : x_i < z_i \leq y_i,\ i = 1, \ldots, d\}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ ($(\mathbf{x}, \mathbf{y}] = \emptyset$, if $x_i \geq y_i$ for some $i \in \{1, \ldots, d\}$). A block in $\mathbb{Z}^d$ is the intersection of a block in $\mathbb{R}^d$ and the set $\mathbb{Z}^d$. We denote the Lebesgue measure on $\mathbb{R}^d$ by λ. Note that for $\mathbf{x} \in \mathbb{R}_+^d$, $[\mathbf{x}]$ is the Lebesgue measure of the block $(\underline{\mathbf{0}}, \mathbf{x}] \subset \mathbb{R}^d$. For a function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^d$, we define the increment of f over a block $(\mathbf{s}, \mathbf{t}] \subset D$ as

$$f(\mathbf{s}, \mathbf{t}] = \begin{cases} \sum\limits_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \varepsilon_i} f(\mathbf{s} + \boldsymbol{\varepsilon}(\mathbf{t} - \mathbf{s})), & \mathbf{s} < \mathbf{t} \\ 0, & \mathbf{s} \not< \mathbf{t}. \end{cases}$$

For instance, in the case $d = 2$ and $\mathbf{s} < \mathbf{t}$, the increment is

$$f(\mathbf{s}, \mathbf{t}] = f(t_1, t_2) - f(t_1, s_2) - f(s_1, t_2) + f(s_1, s_2).$$

We write

$$\sum_{\mathbf{k} < \mathbf{j} \leq \mathbf{m}} x_{\mathbf{j}} = \begin{cases} \sum\limits_{\mathbf{j} \in (\mathbf{k}, \mathbf{m}] \cap \mathbb{Z}^d} x_{\mathbf{j}}, & \mathbf{k} < \mathbf{m} \\ \sum\limits_{\mathbf{j} \in \emptyset} x_{\mathbf{j}} = 0, & \mathbf{k} \not< \mathbf{m}. \end{cases}$$

We will use the notations $X(\mathbf{t})$ and $X_{\mathbf{t}}$ synonymously. For each $n \in \mathbb{N}$, we consider a stochastic process $\{Z_n(\mathbf{x})\}_{\mathbf{x} \in [0,n]^d}$ of the form

$$Z_n(\mathbf{x}) = \lambda((\underline{\mathbf{0}}, \mathbf{x}])\mu_n + \sigma Y(\mathbf{x}) + \lambda((\mathbf{k}_0, \mathbf{m}_0] \cap (\underline{\mathbf{0}}, \mathbf{x}])\delta_n,\ \mathbf{x} \in [0, n]^d, \tag{1}$$

where the constants $\mu_n, \delta_n \in \mathbb{R}$ and $0 < \sigma < \infty$ are unknown and there are (also unknown) points $\mathbf{k}_0, \mathbf{m}_0 \in [0, n]^d \cap \mathbb{Z}^d$, $\mathbf{k}_0 < \mathbf{m}_0$, such that the mean changes over the block $(\mathbf{k}_0, \mathbf{m}_0]$. $\{Y(\mathbf{x})\}_{\mathbf{x} \in \mathbb{R}_+^d}$ is a centered stochastic process that defines a process $\{Y_n(\mathbf{t}) = Y(\lfloor n\mathbf{t}\rfloor) : \mathbf{t} \in [0, 1]^d\}$ with sample paths in $D[0, 1]^d$. Furthermore, we assume that $Y(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}_+^d$ with $x_i = 0$ for some $i \in \{1, \ldots, d\}$ and $\{Y_n(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d}$ fulfills a weak invariance principle:

$$\{Y_n(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d} \stackrel{D[0,1]^d}{\longrightarrow} \{W(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d}, \quad n \to \infty, \tag{2}$$

where $\{W(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d}$ is a d-parameter standard Wiener process, i.e. a zero mean Gaussian process with covariance function $E[W(\mathbf{s})W(\mathbf{t})] = [\mathbf{s} \wedge \mathbf{t}]$, and $\stackrel{D[0,1]^d}{\longrightarrow}$ denotes weak convergence in the space $D[0, 1]^d$ (cf. Bickel and Wichura (1971)). In the following, we will always assume that (2) is satisfied.

Example 2.1. *Let* $\{X_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ *be a stochastic process that satisfies*

$$X_{\mathbf{k}} = e_{\mathbf{k}} + a_n + b_n I_{(\mathbf{k}_0,\mathbf{m}_0]}(\mathbf{k}),$$

where $a_n, b_n \in \mathbb{R}$ *and* $\{e_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ *is a centered, weakly stationary process with finite long-run variance*

$$0 < \sigma^2 = \sum_{\mathbf{k}\in\mathbb{Z}^d} \operatorname{Cov}(e_{\underline{\mathbf{0}}}, e_{\mathbf{k}}) < \infty$$

that satisfies the functional central limit theorem

$$\left\{ \frac{1}{n^{d/2}\sigma} \sum_{\underline{\mathbf{1}}\le\mathbf{k}\le\lfloor n\mathbf{t}\rfloor} e_{\mathbf{k}} \right\}_{\mathbf{t}\in[0,1]^d} \overset{D[0,1]^d}{\longrightarrow} \{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}, \quad n\to\infty. \tag{3}$$

This covers a large class of processes, e.g. i.i.d. (cf. Wichura (1969), Corollary 1), (positively and negatively) associated and (BL,θ)-dependent (cf. Bulinski and Shashkin (2007), Theorem 5.1.5), as well as martingale-difference (cf. Poghosyan and Rœlly (1998), Theorem 3) random fields fulfill this assumption under certain conditions. Define

$$Z_n(\mathbf{x}) = n^{-d/2} \sum_{\underline{\mathbf{1}}\le\mathbf{k}\le\lfloor\mathbf{x}\rfloor} X_{\mathbf{k}}$$

and

$$Y(\mathbf{x}) = \sigma^{-1}\left(Z_n(\mathbf{x}) - \lambda((\underline{\mathbf{0}},\mathbf{x}])\mu_n - \lambda((\mathbf{k}_0,\mathbf{m}_0]\cap(\underline{\mathbf{0}},\mathbf{x}])\delta_n\right),$$

where $\mu_n = a_n n^{-d/2}$ *and* $\delta_n = b_n n^{-d/2}$. *Then* Z_n *has the form* (1) *and* $Y_n(\mathbf{t}) = \sigma^{-1} n^{-d/2} \sum_{\underline{\mathbf{1}}\le\mathbf{k}\le\lfloor n\mathbf{t}\rfloor} \varepsilon_{\mathbf{k}}$, $\mathbf{t}\in[0,1]^d$, *fulfills* (2).

Example 2.2. *We now consider a special case of Example 2.1. Let* $\{\xi_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ *be a centered, stationary random field such that* $E[|\xi_{\mathbf{j}}|^q] < \infty$ *for some* $q > 2d$ *and*

$$0 < \rho^2 = \sum_{\mathbf{k}\in\mathbb{Z}^d} \operatorname{Cov}(\xi_{\underline{\mathbf{0}}}, \xi_{\mathbf{k}}) < \infty.$$

We assume further that the $\{\xi_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ *fulfill the functional central limit theorem* (3) *with* $\sigma = \rho$. *For* $\mathbf{k}\in\mathbb{Z}^d$ *and real numbers* $\{a(\mathbf{j})\}_{\mathbf{j}\in\mathbb{Z}^d}$ *that fulfill the assumption*

$$\sum_{i_1=0}^{\infty}\cdots\sum_{i_d=0}^{\infty}\sum_{k_1=i_1+1}^{\infty}\cdots\sum_{k_d=i_d+1}^{\infty} |a(k_1,\ldots,k_d)| < \infty,$$

we define

$$e_{\mathbf{k}} = \sum_{j_1=0}^{\infty}\cdots\sum_{j_d=0}^{\infty} a(j_1,\ldots,j_d)\xi(k_1-j_1,\ldots,k_d-j_d).$$

Then Ko et al. (2008) showed that $\{e_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ *satisfies* (3) *with*

$$\sigma = \rho\cdot\sum_{i_1=0}^{\infty}\cdots\sum_{i_d=0}^{\infty} a(i_1,\ldots,i_d).$$

In the case when the $\{\xi_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ *are i.i.d., this result was proven by Marinucci and Poghosyan (2001) without the assumption that the* $\{\xi_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ *fulfill the invariance principle themselves.*

Example 2.3. *(cf. Kabluchko and Spodarev (2009))*
Let $\{\xi(t)\}_{t\geq 0}$ *be a Lévy-process. Consider a set-indexed process* $\{Z(B) : B \in \mathcal{B}^d\}$ *that fulfills*

(i) $Z(B)$ *has the same distribution as* $\xi(\lambda(B))$ *for every* $B \in \mathcal{B}^d$.

(ii) $Z(B_1), \ldots, Z(B_k)$ *are independent with* $Z(\cup_{i=1}^k B_i) = \sum_{i=1}^k Z(B_i)$ *for disjoint sets* $B_1, \ldots, B_k \in \mathcal{B}^d$.

We call this a Lévy noise. The corresponding Lévy sheet is the random field $\{Z_{\mathbf{x}} : \mathbf{x} \in \mathbb{R}_+^d\}$ *with* $Z_{\mathbf{x}} = Z([\underline{\mathbf{0}}, \mathbf{x}])$.

Theorem 2.1. *Let* $\{Z(B) : B \in \mathcal{B}^d\}$ *be a Lévy noise with some* $0 < \sigma^2 < \infty$, *such that*

$$E[Z(B)] = 0 \quad \text{and} \quad \text{var} Z(B) = \sigma^2 \lambda(B),$$

for each $B \in \mathcal{B}^d$. *Then the corresponding Lévy sheet fulfills the following invariance principle:*

$$\left\{ \frac{1}{\sigma n^{d/2}} Z(\lfloor n\mathbf{t} \rfloor) \right\}_{\mathbf{t}\in[0,1]^d} \overset{D[0,1]^d}{\longrightarrow} \{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}, \quad n \to \infty$$

Proof. First, we observe that, in view of $\xi(0) = 0$ a.s., one has $Z(\mathbf{x}) = 0$ a.s. for $\mathbf{x} \in \mathbb{R}_+^d$ with $x_i = 0$ for some $i \in \{1, \ldots, d\}$. Then the assertion follows from the functional central limit theorem for i.i.d. random fields once we observe that due to (ii),

$$Z(\lfloor n\mathbf{t} \rfloor) = \sum_{\underline{\mathbf{1}} \leq \mathbf{k} \leq \lfloor n\mathbf{t} \rfloor} Z((\mathbf{k} - \underline{\mathbf{1}}, \mathbf{k}]) \quad a.s.$$

Obviously, assumptions (i) and (ii) together with the moment assumptions on Z yield that the random variables $Z((\mathbf{k} - \underline{\mathbf{1}}, \mathbf{k}])$ are i.i.d. and centered with variance σ^2. □

3 Testing for epidemic changes in the mean

We assume that we have n^d observations $\{Z_n(\mathbf{k}) : \mathbf{k} \in \{1, \ldots, n\}^d\}$ and we want to test the null hypothesis

$$H_0 : (\mathbf{k}_0, \mathbf{m}_0] = \emptyset \quad \text{(no change in the mean)}$$

against the alternative

$$\begin{aligned} H_{\alpha,\beta} :& \exists\, \mathbf{k}_0, \mathbf{m}_0 \in [0, n]^d \cap \mathbb{Z}^d,\ \mathbf{k}_0 < \mathbf{m}_0,\ \lfloor \alpha n^d \rfloor \leq [\mathbf{m}_0 - \mathbf{k}_0] \leq \lfloor (1-\beta) n^d \rfloor, \\ & \text{and } \delta_n \neq 0 \quad \text{(change over the block } (\mathbf{k}_0, \mathbf{m}_0]), \end{aligned}$$

for $0 < \alpha < 1 - \beta < 1$. To do that, we use the following test statistic:

$$\begin{aligned} T_n(\alpha, \beta) &= \hat{\sigma}_n^{-1} \max_{\substack{\mathbf{k},\mathbf{m} \in [\underline{\mathbf{0}}, \underline{\mathbf{n}}] \cap \mathbb{Z}^d,\ \mathbf{k} < \mathbf{m} \\ \lfloor \alpha n^d \rfloor \leq [\mathbf{m}-\mathbf{k}] \leq \lfloor (1-\beta) n^d \rfloor}} \frac{\left| Z_n(\mathbf{k}, \mathbf{m}] - \frac{[\mathbf{m}-\mathbf{k}]}{n^d} Z_n(\underline{\mathbf{n}}) \right|}{\sqrt{\frac{[\mathbf{m}-\mathbf{k}]}{n^d}\left(1 - \frac{[\mathbf{m}-\mathbf{k}]}{n^d}\right)}} \\ &= \hat{\sigma}_n^{-1} \sup_{\substack{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}} \\ \lfloor \alpha n^d \rfloor \leq [\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor] \leq \lfloor (1-\beta) n^d \rfloor}} \frac{\left| Z_n(\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor] - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} Z_n(\underline{\mathbf{n}}) \right|}{\sqrt{\frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d}\left(1 - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d}\right)}}, \end{aligned}$$

where we assume that $\hat{\sigma}_n$ is a consistent estimator for σ under the null hypothesis (i.e. $\hat{\sigma}_n - \sigma = o_P(1)$, $n \to \infty$) and bounded in probability under the alternative (cf. Section 3.3 for a discussion of possible estimators). This corresponds to the following heuristic: Since $Z_n(\mathbf{x}) = 0$ for $\mathbf{x} \in [0,n]^d$ with $x_i = 0$ for some $i \in \{1,\dots,d\}$, $Z_n(\underline{\mathbf{n}})$ is the increment of Z_n on $(\underline{\mathbf{0}}, \underline{\mathbf{n}}]$. Let us assume for a moment that the points $\mathbf{k}_0$ and $\mathbf{m}_0$ are known. Then the term

$$\frac{\left| Z_n\left(\mathbf{k}_0, \mathbf{m}_0\right] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Z_n(\underline{\mathbf{n}}) \right|}{\sqrt{\frac{[\mathbf{m}_0-\mathbf{k}_0]}{n^d}\left(1 - \frac{[\mathbf{m}_0-\mathbf{k}_0]}{n^d}\right)}}$$

is a weighted comparison of the increments of Z_n over the block $(\mathbf{k}_0, \mathbf{m}_0]$, on which the change takes place, and over $(\underline{\mathbf{0}}, \underline{\mathbf{n}}]$. Since we usually do not know the points $\mathbf{k}_0$ and $\mathbf{m}_0$, it is then natural to maximize over all possible blocks. However, due to the law of the iterated logarithm for the Wiener process at the origin, we have to restrict the sizes of the considered blocks. Since such a trimmed statistic cannot in general be expected to detect changes over blocks that do not fulfill the restriction $\lfloor \alpha n^d \rfloor \leq [\mathbf{m} - \mathbf{k}] \leq \lfloor (1-\beta) n^d \rfloor$, we have restricted the considered alternative accordingly. First, we observe that the test statistic is independent of μ_n, because it holds for $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^d$, $\mathbf{x} < \mathbf{y}$, that

$$\sum_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \varepsilon_i} \lambda((\underline{\mathbf{0}}, \mathbf{x} + \boldsymbol{\varepsilon}(\mathbf{y} - \mathbf{x})]) = [\mathbf{y} - \mathbf{x}].$$

Therefore, we will assume without loss of generality that $\mu_n = 0$.

3.1 Limit behavior under the null hypothesis

To define a test that has a given asymptotic level, we need to determine the asymptotic behavior of our test statistic under the null hypothesis. We do this in two steps, by first determining its limit variable and then finding an approximation for the tail behavior of the limit distribution.

Theorem 3.1. *Let $\hat{\sigma}_n$ be a (weakly) consistent estimator for σ under H_0. Then under H_0 it holds that for $n \to \infty$*

$$T_n(\alpha, \beta) \xrightarrow{\mathcal{D}} \sup_{\substack{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}} \\ \alpha \leq [\mathbf{t}-\mathbf{s}] \leq 1-\beta}} \frac{|W(\mathbf{s}, \mathbf{t}] - [\mathbf{t} - \mathbf{s}] W(\underline{\mathbf{1}})|}{\sqrt{[\mathbf{t} - \mathbf{s}](1 - [\mathbf{t} - \mathbf{s}])}}. \qquad (4)$$

Proof. The proof is based on the invariance principle (2) and the following fact: Let S be a metric space and let $f : D[0,1]^d \to S$ be a map which is continuous with respect to the uniform metric on $D[0,1]^d$. Then f is continuous with respect to the Skorohod metric (cf. Bickel and Wichura (1971)) at each point $x \in C[0,1]^d \subset D[0,1]^d$. Using $P(W \in C[0,1]^d) = 1$ and the fact that $f(Y_n)$ are random variables for all the considered maps, we can therefore use the continuous mapping theorem for functions that are continuous with respect to the uniform metric. First, we define the sets

$$A_n = \{(\mathbf{s}, \mathbf{t}) \in [0,1]^{2d} : \ \mathbf{s} < \mathbf{t}, \ \lfloor \alpha n^d \rfloor \leq [\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor] \leq \lfloor (1-\beta) n^d \rfloor\}$$

and

$$A = \{(\mathbf{s}, \mathbf{t}) \in [0,1]^{2d} : \ \mathbf{s} < \mathbf{t}, \ \alpha \leq [\mathbf{t} - \mathbf{s}] \leq 1 - \beta\}.$$

Then the test statistic has the form

$$T_n(\alpha,\beta) = \frac{\sigma}{\hat{\sigma}_n} \sup_{(\mathbf{s},\mathbf{t})\in A_n} \frac{\left|Y_n(\mathbf{s},\mathbf{t}] - \frac{[\lfloor n\mathbf{t}\rfloor - \lfloor n\mathbf{s}\rfloor]}{n^d} Y_n(\underline{\mathbf{1}})\right|}{\sqrt{\frac{[\lfloor n\mathbf{t}\rfloor - \lfloor n\mathbf{s}\rfloor]}{n^d}\left(1 - \frac{[\lfloor n\mathbf{t}\rfloor - \lfloor n\mathbf{s}\rfloor]}{n^d}\right)}}$$
$$= \frac{\sigma}{\hat{\sigma}_n} \sup_{(\mathbf{s},\mathbf{t})\in A_n} \frac{|Y_n(\mathbf{s},\mathbf{t}] - [\mathbf{t}-\mathbf{s}]Y_n(\underline{\mathbf{1}})|}{\sqrt{[\mathbf{t}-\mathbf{s}]\,(1-[\mathbf{t}-\mathbf{s}])}} + o_P(1), \quad n\to\infty,$$

where we have used the fact that $\sigma/\hat{\sigma}_n \xrightarrow{P} 1$ and the invariance principle (2) in combination with the continuous mapping theorem. For $\varepsilon < \min\{\alpha,\beta\}$ and

$$h_\varepsilon(\mathbf{s},\mathbf{t}) = \begin{cases} \varepsilon(1-\varepsilon), & |[\mathbf{t}-\mathbf{s}]| < \varepsilon \text{ or } |[\mathbf{t}-\mathbf{s}]| > 1-\varepsilon \\ |[\mathbf{t}-\mathbf{s}]|(1-|[\mathbf{t}-\mathbf{s}]|), & \varepsilon \le |[\mathbf{t}-\mathbf{s}]| \le 1-\varepsilon, \end{cases}$$

we define the random field

$$X_n(\mathbf{s},\mathbf{t}) = \frac{|Y_n(\mathbf{s},\mathbf{t}] - \lambda((\mathbf{s},\mathbf{t}])Y_n(\underline{\mathbf{1}})|}{\sqrt{h_\varepsilon(\mathbf{s},\mathbf{t})}}, \quad (\mathbf{s},\mathbf{t})\in[0,1]^{2d},$$

and obtain

$$\sup_{(\mathbf{s},\mathbf{t})\in A_n} \frac{|Y_n(\mathbf{s},\mathbf{t}] - [\mathbf{t}-\mathbf{s}]Y_n(\underline{\mathbf{1}})|}{\sqrt{[\mathbf{t}-\mathbf{s}]\,(1-[\mathbf{t}-\mathbf{s}])}} = \sup_{(\mathbf{s},\mathbf{t})\in A_n} X_n(\mathbf{s},\mathbf{t})$$

for large n. Using the invariance principle (2) and the continuous mapping theorem, we find that for $n\to\infty$:

$$\{X_n(\mathbf{s},\mathbf{t})\}_{(\mathbf{s},\mathbf{t})\in[0,1]^{2d}} \xrightarrow{D[0,1]^{2d}} \{X(\mathbf{s},\mathbf{t})\}_{(\mathbf{s},\mathbf{t})\in[0,1]^{2d}} = \left\{\frac{|W(\mathbf{s},\mathbf{t}]-\lambda((\mathbf{s},\mathbf{t}])W(\underline{\mathbf{1}})|}{\sqrt{h_\varepsilon(\mathbf{s},\mathbf{t})}}\right\}_{(\mathbf{s},\mathbf{t})\in[0,1]^{2d}}$$

Another application of the continuous mapping theorem results in

$$\sup_{(\mathbf{s},\mathbf{t})\in K} X_n(\mathbf{s},\mathbf{t}) \xrightarrow{\mathcal{D}} \sup_{(\mathbf{s},\mathbf{t})\in K} X(\mathbf{s},\mathbf{t}) \tag{5}$$

for any subset $K\subset[0,1]^{2d}$. Now, we can consider the sets

$$A_{\pm\vartheta} = \{(\mathbf{s},\mathbf{t})\in[0,1]^{2d}:\ \mathbf{s}<\mathbf{t},\ \alpha\pm\vartheta \le [\mathbf{t}-\mathbf{s}] \le 1-\beta\mp\vartheta\}$$

for sufficiently small $\vartheta>0$. The fact that

$$\sup_{(\mathbf{s},\mathbf{t})\in A_{+\vartheta}} X_n(\mathbf{s},\mathbf{t}) \le \sup_{(\mathbf{s},\mathbf{t})\in A_n} X_n(\mathbf{s},\mathbf{t}) \le \sup_{(\mathbf{s},\mathbf{t})\in A_{-\vartheta}} X_n(\mathbf{s},\mathbf{t})$$

for large $n\in\mathbb{N}$, together with (5) and

$$\sup_{(\mathbf{s},\mathbf{t})\in A_{\pm\vartheta}} X(\mathbf{s},\mathbf{t}) \xrightarrow{\mathcal{D}} \sup_{(\mathbf{s},\mathbf{t})\in A} X(\mathbf{s},\mathbf{t}), \quad \vartheta\to 0,$$

imply the proposition. □

In order to derive asymptotic critical values for the test, we approximate the tail behavior of the limit distribution. This is made easier by the fact that the limit variable is the supremum of a Gaussian field over a compact set. We define

$$C_d(\alpha,\beta) = \int_{\alpha}^{1-\beta} \frac{1}{4^d \xi_d^2 (1-\xi_d)^{2d}} \int_{\xi_d}^{1} \cdots \int_{\xi_2}^{1} \frac{(1-\xi_1)(\xi_1-\xi_2)\cdots(\xi_{d-1}-\xi_d)}{\xi_1^2 \cdots \xi_{d-1}^2} d\xi_1 \cdots d\xi_{d-1} d\xi_d$$

and consider a random field $\{X(\mathbf{s},\mathbf{t})\}_{(\mathbf{s},\mathbf{t})\in D}$ of the form

$$X(\mathbf{s},\mathbf{t}) = \frac{W(\mathbf{s},\mathbf{t}] - [\mathbf{t}-\mathbf{s}]W(\underline{\mathbf{1}})}{\sqrt{[\mathbf{t}-\mathbf{s}](1-[\mathbf{t}-\mathbf{s}])}},$$

where

$$D = \{(\mathbf{x},\mathbf{y}) \in [0,1]^{2d} : \ \mathbf{x} < \mathbf{y}, \ \alpha \le [\mathbf{y}-\mathbf{x}] \le 1-\beta\}.$$

In particular, this yields

$$C_d(\alpha,\beta) = \begin{cases} \frac{1}{4}\left(\log\frac{(1-\beta)(1-\alpha)}{\beta\alpha} + \frac{1}{\alpha} - \frac{1}{1-\beta}\right), & d=1 \\ \int_{\alpha}^{1-\beta} \frac{-2(1-\xi)-(1+\xi)\log\xi}{16\xi^2(1-\xi)^4} d\xi, & d=2 \\ \int_{\alpha}^{1-\beta} \frac{3(1+\xi)\log\xi - 6\xi - 1/2(\xi-1)\log^2\xi + 6}{64\xi^2(1-\xi)^6} d\xi, & d=3. \end{cases} \tag{6}$$

We write $a_k \sim b_k$ for two sequences $(a_k)_{k\in\mathbb{N}}$ and $(b_k)_{k\in\mathbb{N}}$ if $\lim_{k\to\infty} a_k/b_k = 1$. The following theorem is a direct consequence of Theorem 7.1 of Piterbarg (1996) (cf. also Jarušková (2011), Theorem A.1).

Theorem 3.2. *Let $\phi(u)$ be the density of the standard normal distribution. For $u \to \infty$ it holds that:*

$$P\left(\sup_{(\mathbf{s},\mathbf{t})\in D} X(\mathbf{s},\mathbf{t}) > u\right) \sim C_d(\alpha,\beta) u^{4d-1}\phi(u) \tag{7}$$

Proof. We give only an abbreviated version of the proof, for a fuller version in the case $d \in \{1,2\}$, readers are referred to Bucchia (2012). Since the proof is analogous, we treat the general case here. Following the proofs of Theorems 1 and 2 in Jarušková and Piterbarg (2011), we first note that simple calculations show that the correlation function of X has the following representation for $(\mathbf{s},\mathbf{t}), (\tilde{\mathbf{s}},\tilde{\mathbf{t}}) \in D$ with $\|(\mathbf{s},\mathbf{t}) - (\tilde{\mathbf{s}},\tilde{\mathbf{t}})\| \to 0$:

$$\begin{aligned} &\operatorname{Cov}(X(\mathbf{s},\mathbf{t}), X(\tilde{\mathbf{s}},\tilde{\mathbf{t}})) \\ =&1 - \sum_{i=1}^{d} c_i(\mathbf{t}-\mathbf{s})|\tilde{s}_i - s_i| - \sum_{i=d+1}^{2d} c_i(\mathbf{t}-\mathbf{s})|\tilde{t}_i - t_i| + o(\|\tilde{\mathbf{s}}-\mathbf{s}\|_1 + \|\tilde{\mathbf{t}}-\mathbf{t}\|_1), \end{aligned}$$

where $\|x\|_1 = \sum_{i=1}^d |x_i|$ for $x \in \mathbb{R}^d$ and $c_h(\mathbf{t}-\mathbf{s})$ are the functions

$$c_i(\mathbf{t}-\mathbf{s}) = c_{d+i}(\mathbf{t}-\mathbf{s}) = \frac{\prod_{j\neq i}(t_j - s_j)}{2[\mathbf{t}-\mathbf{s}](1-[\mathbf{t}-\mathbf{s}])}, \ i = 1,\ldots,d.$$

Furthermore, D is a compact set and the covariance $\mathrm{Cov}(X(\mathbf{s},\mathbf{t}), X(\tilde{\mathbf{s}},\tilde{\mathbf{t}}))$ of X is strictly less than 1 for $(\mathbf{s},\mathbf{t}),(\tilde{\mathbf{s}},\tilde{\mathbf{t}}) \in D$, $(\mathbf{s},\mathbf{t}) \neq (\tilde{\mathbf{s}},\tilde{\mathbf{t}})$ (cf. Bucchia (2012), Lemma 54).

Using Theorem 7.1 of Piterbarg (1996) (cf. also Theorem A.1 of Jarušková (2011)), we obtain

$$P\left(\sup_{(\mathbf{s},\mathbf{t})\in D} X(\mathbf{s},\mathbf{t}) > u\right) \sim \int_D H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t})\, u^{4d-1}\phi(u),$$

where H_d has the form

$$H_d(\mathbf{s},\mathbf{t}) = \prod_{i=1}^{2d} c_i(\mathbf{t}-\mathbf{s}) = \frac{1}{4^d[\mathbf{t}-\mathbf{s}]^2(1-[\mathbf{t}-\mathbf{s}])^{2d}}.$$

Now, the proposition follows from the fact that using the transformations $\mathbf{x} = \mathbf{t}-\mathbf{s}$ and $\boldsymbol{\xi} = (x_1, x_1x_2, x_1x_2x_3, \ldots, [\mathbf{x}])$ yields

$$\begin{aligned}
&\int_D H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}) \\
=& \int_{[0,1]^d} I_{\{\alpha\le[\mathbf{x}]\le1-\beta\}}(\mathbf{x}) \frac{[\underline{\mathbf{1}}-\mathbf{x}]}{4^d[\mathbf{x}]^2(1-[\mathbf{x}])^{2d}} d\mathbf{x} \\
=& \int_{[0,1]^d} I_{\{\alpha\le[\mathbf{x}]\le1-\beta\}}(\mathbf{x}) \frac{1}{4^d[\mathbf{x}]^2(1-[\mathbf{x}])^{2d}} \frac{(1-x_1)(x_1-x_1x_2)\cdots(\prod_{j\neq d} x_j - [\mathbf{x}])}{x_1\cdot x_1x_2\cdots\prod_{j\neq d} x_j} d\mathbf{x} \\
=& \int_{\alpha}^{1-\beta} \frac{1}{4^d\xi_d^2(1-\xi_d)^{2d}} \int_{\xi_d}^{1}\cdots\int_{\xi_2}^{1} \frac{(1-\xi_1)(\xi_1-\xi_2)\cdots(\xi_{d-1}-\xi_d)}{\xi_1^2\cdots\xi_{d-1}^2} d\xi_1\cdots d\xi_{d-1}d\xi_d \\
=& C_d(\alpha,\beta).
\end{aligned}$$

□

This result can be used to obtain an approximation for the tail behavior of the right hand side of (4):

Corollary 3.1. *With the same notations as in Theorem 3.2, it holds for $u\to\infty$ that*

$$P\left(\sup_{(\mathbf{s},\mathbf{t})\in D} |X(\mathbf{s},\mathbf{t})| > u\right) \sim 2\, C_d(\alpha,\beta) u^{4d-1}\phi(u).$$

Proof. The main idea of the proof (suggested by Z. Kabluchko in a private communication) is to consider a random field $X^\star$ of the form

$$X^\star(\mathbf{s},\mathbf{t}) = \begin{cases} X(\mathbf{s},\mathbf{t}), & (\mathbf{s},\mathbf{t}) \in D^{(1)} \\ -X(\mathbf{s}-\underline{\mathbf{2}},\mathbf{t}-\underline{\mathbf{2}}), & (\mathbf{s},\mathbf{t}) \in D^{(2)}, \end{cases}$$

where $D^{(1)} = D$ and $D^{(2)} = D + \underline{\mathbf{2}}$ are copies of D. Then

$$\sup_{(\mathbf{s},\mathbf{t})\in D} |X(\mathbf{s},\mathbf{t})| = \sup_{(\mathbf{s},\mathbf{t})\in D^{(1)}\cup D^{(2)}} X^\star(\mathbf{s},\mathbf{t})$$

and it stands to reason that the local behavior of the correlation function of $X^\star$ is the same as the local behavior of the correlation function of X. Unfortunately, Theorem 7.1 of Piterbarg (1996) can only be applied to random fields whose correlation function is strictly smaller than one on the domain over which the supremum is taken. This is not the case here, because the covariances of $X^\star$ are the covariances of X times plus or minus one and e.g. $\operatorname{Cov}(X(0,1/2), X(1/2,1)) = -1$ for $d = 1$. This difficulty can be avoided by considering the restricted random field $\{X^\star(\mathbf{s},\mathbf{t}) : (\mathbf{s},\mathbf{t}) \in D_\delta^{(1)} \cup D_\delta^{(2)}\}$, where

$$D_\delta^{(1)} = D_\delta = \{(\mathbf{x},\mathbf{y}) \in \mathbb{R}_+^{2d} : \underline{\delta} \leq \mathbf{x} < \mathbf{y} \leq \underline{\mathbf{1}} - \underline{\delta},\ \alpha \leq [\mathbf{y}-\mathbf{x}] \leq 1-\beta\},$$

and $D_\delta^{(2)} = D_\delta^{(1)} + \underline{\mathbf{2}}$ for $0 < \delta < 1/2$, and

$$X^\star(\mathbf{s},\mathbf{t}) = \begin{cases} X(\mathbf{s},\mathbf{t}), & (\mathbf{s},\mathbf{t}) \in D_\delta^{(1)} \\ -X(\mathbf{s}-\underline{\mathbf{2}}, \mathbf{t}-\underline{\mathbf{2}}), & (\mathbf{s},\mathbf{t}) \in D_\delta^{(2)}. \end{cases}$$

Then it can be shown (cf. Bucchia (2012), Lemma 54) that

$$\operatorname{Cov}(X^\star(\mathbf{s},\mathbf{t}), X^\star(\tilde{\mathbf{s}},\tilde{\mathbf{t}})) < 1$$

for all $(\mathbf{s},\mathbf{t}) \neq (\tilde{\mathbf{s}},\tilde{\mathbf{t}})$ in $D_\delta^{(1)} \cup D_\delta^{(2)}$. Because the correlation function of $X^\star$ behaves locally like the correlation function of X, Theorem 7.1 of Piterbarg (1996) can be used as in the proof of Theorem 3.2 to obtain

$$P\left(\sup_{(\mathbf{s},\mathbf{t})\in D_\delta^{(1)}\cup D_\delta^{(2)}} X^\star(\mathbf{s},\mathbf{t}) > u\right) \sim u^{4d-1}\phi(u) \int\limits_{D_\delta^{(1)}\cup D_\delta^{(2)}} H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}).$$

Since due to symmetry,

$$\int\limits_{D_\delta^{(2)}} H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}) = \int\limits_{D_\delta^{(1)}} H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}),$$

this implies

$$P\left(\sup_{(\mathbf{x},\mathbf{y})\in D_\delta} |X(\mathbf{s},\mathbf{t})| > u\right) \sim 2\, u^{4d-1}\phi(u) \int_{D_\delta} H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}). \tag{8}$$

For $\delta \to 0$, we obtain

$$\int_{D_\delta} H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}) \overset{\delta\to 0}{\longrightarrow} \int_D H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}) \tag{9}$$

as a simple consequence of Lebesgue's theorem. For $u \in \mathbb{R}$ define $C(u) = \phi(u)u^{4d-1}$,

$$a(u) = P\left(\sup_{(\mathbf{s},\mathbf{t})\in D} |X(\mathbf{s},\mathbf{t})| > u\right) \text{ and } b(u) = 2\, C(u) \int_D H_d(\mathbf{s},\mathbf{t})\, d(\mathbf{s},\mathbf{t}).$$

Then the proposition can be written in the form

$$\lim_{u\to\infty} \frac{a(u)}{b(u)} \overset{!}{=} 1.$$

Using (7), (8) and (9), it can be shown that

$$1 \leq \liminf_{u\to\infty} \frac{a(u)}{b(u)} \leq \limsup_{u\to\infty} \frac{a(u)}{b(u)} \leq 1$$

and therefore this proposition holds. □

3.2 Behavior under the alternative

Now, we focus on the behavior of the test statistic under the alternative.

Theorem 3.3. *If $|\delta_n| n^d \to \infty$ for $n \to \infty$ and $\hat{\sigma}_n = \mathcal{O}_P(1)$, $\hat{\sigma}_n > 0$, it holds under the alternative $H_{\alpha,\beta}$ that*

$$T_n(\alpha,\beta) \xrightarrow[n\to\infty]{P} \infty.$$

Proof. Let $(\mathbf{k}_0, \mathbf{m}_0]$ be the block over which the change takes place. Due to our restriction for the sizes of the blocks, we have

$$\alpha - 1 \leq \alpha - \frac{1}{n^d} \leq \frac{\lfloor \alpha n^d \rfloor}{n^d} \leq \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} \leq 1 - \beta.$$

This implies that there is a $C > 0$ such that

$$\begin{aligned} T_n(\alpha,\beta) &\geq \frac{\hat{\sigma}_n^{-1}}{\sqrt{(1-\beta)(2-\alpha)}} \left| Z_n(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Z_n(\underline{\mathbf{n}}) \right| \\ &\geq C \hat{\sigma}_n^{-1} \left| Z_n(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Z_n(\underline{\mathbf{n}}) \right| \end{aligned}$$

Moreover,

$$\begin{aligned} &Z_n(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Z_n(\underline{\mathbf{n}}) \\ =&\sigma \left(Y(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Y(\underline{\mathbf{n}}) \right) - \delta_n \frac{[\mathbf{m}_0 - \mathbf{k}_0]^2}{n^d} \\ &+ \delta_n \overbrace{\sum_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \varepsilon_i} \lambda((\mathbf{k}_0, \mathbf{m}_0] \cap (\underline{\mathbf{0}}, \mathbf{k}_0 + \boldsymbol{\varepsilon}(\mathbf{m}_0 - \mathbf{k}_0)])}^{=[\mathbf{m}_0 - \mathbf{k}_0]} \\ =& n^d \delta_n \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} \left(1 - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} \right) + \sigma \left(Y(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Y(\underline{\mathbf{n}}) \right), \end{aligned}$$

and thus,

$$T_n(\alpha,\beta) \geq C \left\{ n^d |\delta_n| \hat{\sigma}_n^{-1} \left(\alpha - \frac{1}{n^d} \right) \beta - \sigma \hat{\sigma}_n^{-1} \left| Y(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Y(\underline{\mathbf{n}}) \right| \right\}.$$

The assumed invariance principle for Y implies

$$\begin{aligned} &\left| Y(\mathbf{k}_0, \mathbf{m}_0] - \frac{[\mathbf{m}_0 - \mathbf{k}_0]}{n^d} Y(\underline{\mathbf{n}}) \right| \\ \leq& \sup_{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}} \left| Y(\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor] - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} Y(\lfloor n\underline{\mathbf{1}} \rfloor) \right| = \mathcal{O}_P(1), \ n \to \infty, \end{aligned}$$

and finally the assumptions $n^d |\delta_n| \to \infty$ and $\hat{\sigma}_n = \mathcal{O}_P(1)$ yield the weak consistency of the test. □

3.3 Long-run variance estimators

In the test statistics presented above, we have used an unspecified estimator for σ^2 in order to show that the main requirements for such an estimator are consistency under the null and stochastic boundedness under the alternative hypothesis. The problem of how to obtain such an estimator in the general model (1), possibly also taking into account the form of the specific alternative, is highly complex and requires further research which is beyond the scope of this paper. At present, in order to give some idea of possible estimators, we restrict ourselves to an example for an estimator that fulfills our requirements in the partial sum case (Example 2.1) with absolutely summable covariance function. In this case, we have observations of the form $X_{\mathbf{j}} = e_{\mathbf{j}} + a_n + b_n I_{(\mathbf{k}_0,\mathbf{m}_0]}(\mathbf{j})$ and the parameter σ^2 is the long-run variance $\sum\limits_{\mathbf{k}\in\mathbb{Z}^d} \operatorname{Cov}(e_{\underline{\mathbf{0}}}, e_{\mathbf{k}})$. Therefore, we can apply generalizations of well-known kernel-based variance estimators from the time series literature to our model. To match our general approach, we consider a nonparametric estimator. In order to shorten notation, we write $r(\mathbf{j}) = \operatorname{Cov}(e_{\underline{\mathbf{0}}}, e_{\mathbf{j}})$ and define

$$\hat{r}_X(\mathbf{j}) = \frac{1}{n^d} \sum_{\mathbf{k}\in N_{\mathbf{j}}} (X_{\mathbf{k}} - \bar{X}_n)(X_{\mathbf{k}+\mathbf{j}} - \bar{X}_n),$$

with $\bar{X}_n = n^{-d} \sum\limits_{\underline{\mathbf{1}}\leq\mathbf{k}\leq\underline{\mathbf{n}}} X_{\mathbf{k}}$ and $N_{\mathbf{j}} = \{\mathbf{k} \in \mathbb{Z}^d : \underline{\mathbf{1}} \leq \mathbf{k}, \mathbf{k}+\mathbf{j} \leq \underline{\mathbf{n}}\}$. We consider estimators of the form

$$\hat{\sigma}_n^2 = \sum_{\mathbf{j}\in B_{q-1}} \omega_{q,\mathbf{j}}\, \hat{r}_X(\mathbf{j}),$$

where $q = q(n) \in [1, n]$ is an integer with $q = q(n) \to \infty$ and $\lim\limits_{n\to\infty} q/n = 0$, $B_q = \{-q, \dots, q\}^d$ and $\omega_{q,\mathbf{j}}$ is a bounded weight function that fulfills $\omega_{q,\mathbf{j}} \to 1$ for $q \to \infty$. Analogously, we define

$$\tilde{\sigma}_n^2 = \sum_{\mathbf{j}\in B_{q-1}} \omega_{q,\mathbf{j}}\, \tilde{r}_e(\mathbf{j}) \quad \text{with} \quad \tilde{r}_e(\mathbf{j}) = \frac{1}{n^d} \sum_{\mathbf{k}\in N_{\mathbf{j}}} e_{\mathbf{k}} e_{\mathbf{k}+\mathbf{j}}.$$

If we assume additional moment and homogeneity conditions on $e_{\mathbf{k}}$ (cf. Lavancier (2008), hypothesis H0), a careful reading of the proof of Lemma 1 in Lavancier (2008) shows that his proof that $\hat{\sigma}^2$ converges stochastically to σ^2 remains valid if we replace $|\mathbf{j}|$ by $\mathbf{j}$ and consider different weight functions (e.g. flat-top kernels as suggested by Politis and Romano (1996)). This more general case is therefore discussed here. As in the time series case (cf. Berkes et al. (2006), Proposition D.1), in order to prove consistency under the null hypothesis and stochastic boundedness under the alternative it then suffices to show that the difference $\hat{\sigma}_n^2 - \tilde{\sigma}_n^2$ converges to 0 in probability under the null and remains bounded under the alternative:

Lemma 3.1. *For $q = q(n) \to \infty$ with $\lim\limits_{n\to\infty} q/n = 0$ and*

$$b_n^2 q^d = \mathcal{O}(1) \tag{10}$$

it holds that

$$\hat{\sigma}_n^2 - \tilde{\sigma}_n^2 = o_P(1), \quad n \to \infty,$$

if $b_n = 0$ *and*

$$\hat{\sigma}_n^2 - \tilde{\sigma}_n^2 = \mathcal{O}_P(1)$$

if b_n *satisfies* (10) *with* $b_n n^{d/2} \to \infty$.

Proof. Let $C > 0$ be a constant whose value may change from line to line and write $R = (\mathbf{k}_0, \mathbf{m}_0]$. It holds that

$$X_{\mathbf{k}} - \bar{X}_n = e_{\mathbf{k}} - \bar{e}_n + b_n \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right), \text{ with } \bar{e}_n = n^{-d} \sum_{\underline{\mathbf{1}} \le \mathbf{k} \le \underline{\mathbf{n}}} e_{\mathbf{k}},$$

so that we can assume without loss of generality that $a_n = 0$. Note that $\#N_{\mathbf{j}} = [\underline{\mathbf{n}} - |\mathbf{j}|]$ and $\#B_{q-1} = (\#\{-q+1, \dots, 0, 1, \dots, q-1\})^d \le 2^d q^d$. We find that

$$\begin{aligned}
&\hat{r}_X(\mathbf{j}) - \tilde{r}_e(\mathbf{j}) \\
=&\frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left\{ \left(e_{\mathbf{k}} - \bar{e}_n + b_n \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right) \right) \left(e_{\mathbf{k}+\mathbf{j}} - \bar{e}_n + b_n \left(I_R(\mathbf{k}+\mathbf{j}) - \frac{\lambda(R)}{n^d} \right) \right) \right\} \\
&- \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} e_{\mathbf{k}} e_{\mathbf{k}+\mathbf{j}} \\
=&\frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left\{ \bar{e}_n^2 - \bar{e}_n [e_{\mathbf{k}+\mathbf{j}} + e_{\mathbf{k}}] - \bar{e}_n b_n \left[I_R(\mathbf{k}) + I_R(\mathbf{k}+\mathbf{j}) - 2\frac{\lambda(R)}{n^d} \right] \right. \\
&\qquad + b_n \left[e_{\mathbf{k}} \left(I_R(\mathbf{k}+\mathbf{j}) - \frac{\lambda(R)}{n^d} \right) + e_{\mathbf{k}+\mathbf{j}} \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right) \right] \\
&\qquad \left. + b_n^2 \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right) \left(I_R(\mathbf{k}+\mathbf{j}) - \frac{\lambda(R)}{n^d} \right) \right\} \\
=&\frac{[\underline{\mathbf{n}} - |\mathbf{j}|]}{n^d} \bar{e}_n^2 - \frac{1}{n^d} \bar{e}_n \sum_{\mathbf{k} \in N_{\mathbf{j}}} (e_{\mathbf{k}} + e_{\mathbf{k}+\mathbf{j}}) \\
&+ \bar{e}_n b_n \left(-\frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} (I_R(\mathbf{k}) + I_R(\mathbf{k}+\mathbf{j})) + 2 \frac{[\underline{\mathbf{n}} - |\mathbf{j}|]}{n^d} \frac{\lambda(R)}{n^d} \right) \\
&+ b_n \frac{1}{n^d} \left(\sum_{\mathbf{k} \in N_{\mathbf{j}}} e_{\mathbf{k}} \left(I_R(\mathbf{k}+\mathbf{j}) - \frac{\lambda(R)}{n^d} \right) + \sum_{\mathbf{k} \in N_{\mathbf{j}}} e_{\mathbf{k}+\mathbf{j}} \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right) \right) \\
&+ b_n^2 \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left(I_R(\mathbf{k}) - \frac{\lambda(R)}{n^d} \right) \left(I_R(\mathbf{k}+\mathbf{j}) - \frac{\lambda(R)}{n^d} \right)
\end{aligned}$$

and thus

$$
\begin{aligned}
&\sum_{\mathbf{j}\in B_{q-1}} E|\hat{r}_X(\mathbf{j})-\tilde{r}_e(\mathbf{j})| \\
&\leq \sum_{\mathbf{j}\in B_{q-1}} \left\{ E\left|\frac{[\underline{\mathbf{n}}-|\mathbf{j}|]}{n^d}\bar{e}_n^2\right| + E\left|n^{-d}\bar{e}_n\sum_{\mathbf{k}\in N_{\mathbf{j}}}(e_{\mathbf{k}}+e_{\mathbf{k}+\mathbf{j}})\right|\right\} \\
&\quad + |b_n| \overbrace{\sum_{\mathbf{j}\in B_{q-1}} \underbrace{\left\{\frac{1}{n^d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}(I_R(\mathbf{k})+I_R(\mathbf{k}+\mathbf{j})) + 2\frac{[\underline{\mathbf{n}}-|\mathbf{j}|]}{n^d}\frac{\lambda(R)}{n^d}\right\}}_{\leq 4}}^{\leq 4\#B_{q-1}} E|\bar{e}_n| \\
&\quad + |b_n|n^{-d}\sum_{\mathbf{j}\in B_{q-1}}\left\{E\left|\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}}\left(I_R(\mathbf{k}+\mathbf{j})-\frac{\lambda(R)}{n^d}\right)\right| + E\left|\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}+\mathbf{j}}\left(I_R(\mathbf{k})-\frac{\lambda(R)}{n^d}\right)\right|\right\} \\
&\quad + b_n^2 \underbrace{\sum_{\mathbf{j}\in B_{q-1}} \underbrace{\frac{1}{n^d}\left\{\left|\sum_{\mathbf{k}\in N_{\mathbf{j}}}\left(I_R(\mathbf{k})-\frac{\lambda(R)}{n^d}\right)\left(I_R(\mathbf{k}+\mathbf{j})-\frac{\lambda(R)}{n^d}\right)\right|\right\}}_{\leq 4}}_{\leq 4\#B_{q-1}} \\
&\leq \sum_{\mathbf{j}\in B_{q-1}}\left\{E\left|\frac{[\underline{\mathbf{n}}-|\mathbf{j}|]}{n^d}\bar{e}_n^2\right| + E\left|\bar{e}_n n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}}\right| + E\left|\bar{e}_n n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}+\mathbf{j}}\right|\right\} \\
&\quad + C|b_n|q^d E|\bar{e}_n| + Cb_n^2 q^d \\
&\quad + |b_n|\sum_{\mathbf{j}\in B_{q-1}}\left\{E\left|n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}}\left(I_R(\mathbf{k}+\mathbf{j})-\frac{\lambda(R)}{n^d}\right)\right| + E\left|n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}+\mathbf{j}}\left(I_R(\mathbf{k})-\frac{\lambda(R)}{n^d}\right)\right|\right\}
\end{aligned}
$$

[1]The Cauchy-Schwarz inequality and the absolute summability of the covariance function imply

$$
\begin{aligned}
E\bar{e}_n^2 &= n^{-2d}\sum_{\underline{\mathbf{1}}\leq\mathbf{k},\mathbf{l}\leq\underline{\mathbf{n}}}\mathrm{Cov}(e_{\mathbf{k}},e_{\mathbf{l}}) \\
&\leq n^{-d}\sum_{-\underline{\mathbf{n}}\leq\mathbf{k}\leq\underline{\mathbf{n}}}|r(\mathbf{k})| \\
&\leq n^{-d}\sum_{\mathbf{k}\in\mathbb{Z}^d}|r(\mathbf{k})| \leq Cn^{-d}
\end{aligned}
$$

and analogously

$$
E\left|n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}}\left(I_R(\mathbf{k}+\mathbf{j})-\frac{\lambda(R)}{n^d}\right)\right| \leq \sqrt{2n^{-2d}\sum_{\mathbf{k},\mathbf{l}\in N_{\mathbf{j}}}|r(\mathbf{k}-\mathbf{l})|} \leq n^{-d/2}C
$$

as well as

$$
E\left|n^{-d}\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}+\mathbf{j}}\left(I_R(\mathbf{k})-\frac{\lambda(R)}{n^d}\right)\right| \leq Cn^{-d/2}.
$$

[1]In comparison to the published article, the present passage has been slightly modified.

Finally, an analogous argument yields

$$E\left|n^{-d}\bar{e}_n\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}}\right|+E\left|n^{-d}\bar{e}_n\sum_{\mathbf{k}\in N_{\mathbf{j}}}e_{\mathbf{k}+\mathbf{j}}\right|\le Cn^{-d}.$$

In conclusion, we have obtained

$$\sum_{\mathbf{j}\in B_{q-1}}E|\hat{r}_X(\mathbf{j})-\tilde{r}_e(\mathbf{j})|\le\mathcal{O}(q^dn^{-d})+\mathcal{O}(|b_n|q^dn^{-d/2}+q^db_n^2)=\begin{cases}o(1), & \text{under } H_0\\ \mathcal{O}(1), & \text{under } H_{\alpha,\beta}\end{cases}$$

since in conjunction with $b_nn^{d/2}\to\infty$, (10) implies $|b_n|q^dn^{-d/2}=o(1)$. □

Remark 3.1. *Using*

$$X_{\mathbf{k}}=n^{d/2}\sum_{\boldsymbol{\varepsilon}\in\{0,1\}^d}(-1)^{\sum_{j=1}^d\varepsilon_j}Z_n(\mathbf{k}-\boldsymbol{\varepsilon}),\ \mathbf{k}\ge\underline{\mathbf{1}},$$

we can view Z_n from (1) *as the partial sum process of the random field $X_{\mathbf{k}}=a_n+e_{\mathbf{k}}+I_{(\mathbf{k}_0,\mathbf{m}_0]}b_n$ with $a_n=n^{d/2}\mu_n$, $b_n=n^{d/2}\delta_n$ and*

$$e_{\mathbf{k}}=n^{d/2}\sum_{\boldsymbol{\varepsilon}\in\{0,1\}^d}(-1)^{\sum_{j=1}^d\varepsilon_j}\sigma Y(\mathbf{k}-\boldsymbol{\varepsilon}),\ \mathbf{k}\ge\underline{\mathbf{1}},$$

where the latter fulfills the functional central limit theorem.

4 Estimation of the change-points

In this section, we consider the alternative

$$H_A(\boldsymbol{\vartheta},\boldsymbol{\gamma}):\exists\ \underline{\mathbf{0}}<\boldsymbol{\vartheta}<\boldsymbol{\gamma}<\underline{\mathbf{1}}:\mathbf{k}_0=\lfloor n\boldsymbol{\vartheta}\rfloor,\ \mathbf{m}_0=\lfloor n\boldsymbol{\gamma}\rfloor,$$

and the "change" δ_n is assumed to be a constant multiple of $n^{-d/2}$, i.e.

$$\delta_n=\delta\,n^{-d/2},\quad \delta\ne0. \tag{11}$$

Our aim is to estimate the points $\boldsymbol{\vartheta}$ and $\boldsymbol{\gamma}$. Using a similar approach to the one employed by Aston and Kirch (2012a), the estimators we consider are points where the maximum of a slightly modified version of our test statistic is reached. To do so, we define

$$\arg\max_B Z=\{\mathbf{a}\in B:\ Z(\mathbf{a})=\max_{\mathbf{b}\in B}Z(\mathbf{b})\}$$

for functions $Z:A\to\mathbb{R}$ ($A\subseteq[0,1]^d$, $d\in\mathbb{N}$) in $D[0,1]^d$ and compact subsets $B\subseteq A$. Furthermore, let

$$K_d=\{(\mathbf{s},\mathbf{t})\in[0,1]^{2d}:\ \underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}\}$$

and

$$G_{n,d}(\mathbf{s},\mathbf{t})=\frac{1}{n^{d/2}}\left(Z_n(\lfloor n\mathbf{s}\rfloor,\lfloor n\mathbf{t}\rfloor]-\frac{[\lfloor n\mathbf{t}\rfloor-\lfloor n\mathbf{s}\rfloor]}{n^d}Z_n(\underline{\mathbf{n}})\right)I_{K_d}(\mathbf{s},\mathbf{t}).$$

Then $\arg\max_{K_d}|G_{n,d}|\ne\emptyset$, and arbitrary points $(\hat{\boldsymbol{\vartheta}}_n,\hat{\boldsymbol{\gamma}}_n)$ in $\arg\max_{K_d}|G_{n,d}|$ give consistent estimators for $(\boldsymbol{\vartheta},\boldsymbol{\gamma})$:

Theorem 4.1. *Under $H_A(\boldsymbol{\vartheta},\boldsymbol{\gamma})$, with δ_n as in* (11)*, it holds that*

$$(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta},\hat{\boldsymbol{\gamma}}_n-\boldsymbol{\gamma})=o_P(1),\quad n\to\infty.$$

Before we give the proof of Theorem 4.1, we introduce two useful lemmas.

Lemma 4.1. *(cf. Bucchia (2012), Lemma 66) Let K be a compact subset of $\mathbb{R}^d$ and $f: K\to\mathbb{R}$ a continuous function with a unique maximizer $\mathbf{x}_0\in K$ (i.e. $\{\mathbf{x}_0\}=\arg\max_K f$). Moreover, let $f_n: K\to\mathbb{R}$, $n=1,2,\dots,$ be functions with*

$$\max_{\mathbf{x}\in K}|f_n(\mathbf{x})-f(\mathbf{x})|\overset{n\to\infty}{\longrightarrow}0$$

and (not necessarily unique) maximizers $\hat{\mathbf{x}}_n$ (i.e. $f_n(\hat{\mathbf{x}}_n)=\max_{\mathbf{x}\in K}f_n(x)$). Then it holds that

$$\hat{\mathbf{x}}_n\overset{n\to\infty}{\longrightarrow}\mathbf{x}_0.$$

Lemma 4.2. *Define*

$$\mathcal{R}=\{R\subseteq[0,1]^d:\ R=(\mathbf{a},\mathbf{b}],\ \mathbf{a},\mathbf{b}\in[0,1]^d\}.$$

Then $\mathcal{R}$ is closed under intersection and $R_1\subsetneq R_2$ implies $\lambda(R_1)<\lambda(R_2)$ for all $R_1,R_2\in\mathcal{R}$. For $A\in\mathcal{R}$ with $0<\lambda(A)<1$, define a function $F:\mathcal{R}\to\mathbb{R}$ by setting

$$F(B)=\lambda(A\cap B)-\lambda(A)\lambda(B),\ B\in\mathcal{R}.$$

Then $F(B)$ is maximal for $B=A$, with $F(A)>0$ and $F(B)<F(A)$ for all $B\neq A$. Assume further that $A^c=[0,1]^d\setminus A\notin\mathcal{R}$. Then A uniquely maximizes $|F|$.

Proof. First, note that $F(A)=\lambda(A)(1-\lambda(A))$. Therefore, $F(A)>0$ and it suffices to show that

$$F(B)<\lambda(A)(1-\lambda(A))$$

for all $B\neq A$. Let $B\in\mathcal{R}$, $B\neq A$. If $\lambda(A)=\lambda(B)$, neither $A\subsetneq B$ nor $A\supsetneq B$ can hold, and therefore $\lambda(A\cap B)<\lambda(A)$. It follows that

$$F(B)=\lambda(A\cap B)-\lambda(A)^2<F(A).$$

If $\lambda(A)<\lambda(B)$, the fact that $\lambda(A\cap B)\le\lambda(A)$ implies

$$F(B)=\lambda(A\cap B)-\lambda(A)\lambda(B)<\lambda(A\cap B)-\lambda(A)^2\le F(A).$$

Finally, if $\lambda(B)<\lambda(A)$, we obtain

$$F(B)=\lambda(A\cap B)-\lambda(A)\lambda(B)\le\lambda(B)(1-\lambda(A))<F(A).$$

Now, we additionally assume that $A^c\notin\mathcal{R}$. It suffices to show that $F(B)>-F(A)$ for all $B\neq A$. Note that $F(A)=\lambda(A)\lambda(A^c)$. By our assumption, if $A\cap B=\emptyset$, we have $B\subset A^c$ and $\lambda(B)<\lambda(A^c)$. It follows that

$$F(B)=-\lambda(B)\lambda(A)>-\lambda(A^c)\lambda(A)=-F(A).$$

If $A \cap B \neq \emptyset$ and therefore $\lambda(A \cap B) > 0$, we obtain

$$\begin{aligned} F(B) &= \lambda(A \cap B) - \lambda(A)\lambda(B) \\ &= \lambda(A \cap B) - \lambda(A)\{\lambda(A \cap B) + \lambda(B \cap A^c)\} \\ &= \underbrace{\lambda(A \cap B)(1 - \lambda(A))}_{>0} - \lambda(A) \underbrace{\lambda(B \cap A^c)}_{\leq \lambda(A^c)} > -F(A), \end{aligned}$$

which completes the proof. □

We can now address the proof of Theorem 4.1.

Proof. We can again assume without loss of generality that $\mu_n = 0$. Using the same approach as in Aston and Kirch (2012a), we first show that

$$\sup_{\underline{\mathbf{0}} \leq \mathbf{s} \leq \mathbf{t} \leq \underline{\mathbf{1}}} |G_{n,d}(\mathbf{s}, \mathbf{t}) - \delta f(\mathbf{s}, \mathbf{t})| = o_P(1), \tag{12}$$

for some continuous function $f : [0,1]^{2d} \to \mathbb{R}$ such that $(\boldsymbol{\vartheta}, \boldsymbol{\gamma})$ is the unique maximizer of $|f|$. Define $f : [0,1]^{2d} \to \mathbb{R}$,

$$f(\mathbf{s}, \mathbf{t}) = \lambda((\boldsymbol{\vartheta}, \boldsymbol{\gamma}] \cap (\mathbf{s}, \mathbf{t}]) - \lambda((\boldsymbol{\vartheta}, \boldsymbol{\gamma}])\lambda((\mathbf{s}, \mathbf{t}]).$$

For $\underline{\mathbf{0}} \leq \mathbf{s} \leq \mathbf{t} \leq \underline{\mathbf{1}}$, it holds that

$$\begin{aligned} &\frac{1}{n^{d/2}} \left(Z_n(\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor] - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} Z_n(\underline{\mathbf{n}}) \right) \\ =&\sigma \frac{1}{n^{d/2}} \left(Y_n(\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor] - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} Y_n(\underline{\mathbf{n}}) \right) \\ &+ \delta \left(\frac{1}{n^d} \sum_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \varepsilon_i} \lambda((\lfloor n\boldsymbol{\vartheta} \rfloor, \lfloor n\boldsymbol{\gamma} \rfloor] \cap (\underline{\mathbf{0}}, \lfloor n\mathbf{s} \rfloor + \boldsymbol{\varepsilon}(\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor)]) \right. \\ &\qquad \left. - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} \frac{[\lfloor n\boldsymbol{\gamma} \rfloor - \lfloor n\boldsymbol{\vartheta} \rfloor]}{n^d} \right) \\ =&\sigma \frac{1}{n^{d/2}} \left(Y_n(\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor] - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} Y_n(\underline{\mathbf{n}}) \right) \\ &+ \delta \left(\frac{\lambda((\lfloor n\boldsymbol{\vartheta} \rfloor, \lfloor n\boldsymbol{\gamma} \rfloor] \cap (\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor])}{n^d} - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} \frac{[\lfloor n\boldsymbol{\gamma} \rfloor - \lfloor n\boldsymbol{\vartheta} \rfloor]}{n^d} \right). \end{aligned}$$

Define

$$\begin{aligned} f_n(\mathbf{s}, \mathbf{t}) &= \frac{\lambda((\lfloor n\boldsymbol{\vartheta} \rfloor, \lfloor n\boldsymbol{\gamma} \rfloor] \cap (\lfloor n\mathbf{s} \rfloor, \lfloor n\mathbf{t} \rfloor])}{n^d} - \frac{[\lfloor n\mathbf{t} \rfloor - \lfloor n\mathbf{s} \rfloor]}{n^d} \frac{[\lfloor n\boldsymbol{\gamma} \rfloor - \lfloor n\boldsymbol{\vartheta} \rfloor]}{n^d} \\ &= \lambda \left(\left(\tfrac{\lfloor n\boldsymbol{\vartheta} \rfloor}{n}, \tfrac{\lfloor n\boldsymbol{\gamma} \rfloor}{n} \right] \cap \left(\tfrac{\lfloor n\mathbf{s} \rfloor}{n}, \tfrac{\lfloor n\mathbf{t} \rfloor}{n} \right] \right) - \lambda \left(\left(\tfrac{\lfloor n\boldsymbol{\vartheta} \rfloor}{n}, \tfrac{\lfloor n\boldsymbol{\gamma} \rfloor}{n} \right] \right) \lambda \left(\left(\tfrac{\lfloor n\mathbf{s} \rfloor}{n}, \tfrac{\lfloor n\mathbf{t} \rfloor}{n} \right] \right) \end{aligned}$$

and $f_n(\mathbf{s},\mathbf{t}) = 0$ for $(\mathbf{s},\mathbf{t}) \notin K_d$. Then

$$\begin{aligned}
&\sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} \left| \frac{1}{n^{d/2}} \left(Z_n(\lfloor n\mathbf{s}\rfloor, \lfloor n\mathbf{t}\rfloor] - \frac{[\lfloor n\mathbf{t}\rfloor - \lfloor n\mathbf{s}\rfloor]}{n^d} Z_n(\underline{\mathbf{n}}) \right) - \delta f(\mathbf{s},\mathbf{t}) \right| \\
\le &\sigma \frac{1}{n^{d/2}} \overbrace{\sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} \left| Y_n(\lfloor n\mathbf{s}\rfloor, \lfloor n\mathbf{t}\rfloor] - \frac{[\lfloor n\mathbf{t}\rfloor - \lfloor n\mathbf{s}\rfloor]}{n^d} Y_n(\underline{\mathbf{n}}) \right|}^{\overset{(2)}{=}\mathcal{O}_P(1)} \\
&+ |\delta| \sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} |f_n(\mathbf{s},\mathbf{t}) - f(\mathbf{s},\mathbf{t})| \\
= &o_P(1) + |\delta| \sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} |f_n(\mathbf{s},\mathbf{t}) - f(\mathbf{s},\mathbf{t})|.
\end{aligned}$$

Now, (12) follows if $f(\mathbf{s},\mathbf{t}) = \lim_{n\to\infty} f_n(\mathbf{s},\mathbf{t})$ uniformly. To see that this is the case, we define a function $h : [0,1]^{4d} \to \mathbb{R}_+$ by setting

$$h(\mathbf{a},\mathbf{b},\mathbf{s},\mathbf{t}) = \lambda((\mathbf{a},\mathbf{b}] \cap (\mathbf{s},\mathbf{t}]) - \lambda((\mathbf{a},\mathbf{b}])\lambda((\mathbf{s},\mathbf{t}]).$$

Then $f(\mathbf{s},\mathbf{t}) = h(\boldsymbol{\vartheta},\boldsymbol{\gamma},\mathbf{s},\mathbf{t})$ and

$$f_n(\mathbf{s},\mathbf{t}) = h\left(\frac{\lfloor n\boldsymbol{\vartheta}\rfloor}{n}, \frac{\lfloor n\boldsymbol{\gamma}\rfloor}{n}, \frac{\lfloor n\mathbf{s}\rfloor}{n}, \frac{\lfloor n\mathbf{t}\rfloor}{n} \right).$$

Note that h has the form

$$\begin{aligned}
h(\mathbf{a},\mathbf{b},\mathbf{s},\mathbf{t}) &= \lambda(\times_{i=1}^d((a_i,b_i] \cap (s_i,t_i])) - \lambda(\times_{i=1}^d(a_i,b_i])\lambda(\times_{i=1}^d(s_i,t_i]) \\
&= \prod_{i=1}^d \lambda((a_i,b_i] \cap (s_i,t_i]) - \prod_{i=1}^d (b_i - a_i)^+ \prod_{i=1}^d (t_i - s_i)^+ \\
&= \prod_{i=1}^d (t_i \wedge b_i - a_i \vee s_i)^+ - [(\mathbf{b}-\mathbf{a})^+]\,[(\mathbf{t}-\mathbf{s})^+] \\
&= [(\mathbf{t}\wedge\mathbf{b} - \mathbf{a}\vee\mathbf{s})^+] - [(\mathbf{b}-\mathbf{a})^+]\,[(\mathbf{t}-\mathbf{s})^+],
\end{aligned}$$

where $(\mathbf{x})^+ = \mathbf{x}\vee\underline{\mathbf{0}}$, hence h is uniformly continuous on the compact set $[0,1]^{4d}$. Therefore, f_n converges to f uniformly and $|f(\cdot,\cdot)| = |h(\boldsymbol{\vartheta},\boldsymbol{\gamma},\cdot,\cdot)|$ is also continuous. Lemma 4.2 shows that $(\boldsymbol{\vartheta},\boldsymbol{\gamma})$ is the unique point at which $|f|$ attains its maximum (since $\underline{\mathbf{0}} < \boldsymbol{\vartheta} < \boldsymbol{\gamma} < \underline{\mathbf{1}}$, $(\boldsymbol{\vartheta},\boldsymbol{\gamma}]^c$ cannot be a rectangle in $[0,1]^d$). Since $\delta \neq 0$, this is also the case for $|\delta||f(\cdot,\cdot)|$. Now, the proposition follows if we can show that each subsequence of $(\hat{\boldsymbol{\vartheta}}_n,\hat{\boldsymbol{\gamma}}_n)_{n\in\mathbb{N}}$ has a further subsequence that converges to $(\boldsymbol{\vartheta},\boldsymbol{\gamma})$ almost surely. Let $(\hat{\boldsymbol{\vartheta}}_{n'},\hat{\boldsymbol{\gamma}}_{n'})$ be a subsequence of $(\hat{\boldsymbol{\vartheta}}_n,\hat{\boldsymbol{\gamma}}_n)_{n\in\mathbb{N}}$. Our previous arguments together with the triangle inequality show that

$$h_{n,d} = \sup_{(\mathbf{x},\mathbf{y})\in K_d} \{||G_{n,d}(\mathbf{x},\mathbf{y})| - |\delta||f(\mathbf{x},\mathbf{y})||\} = o_P(1), \quad n \to \infty.$$

Therefore, there is a subsequence $(n'') \subset (n')$, such that $h_{n'',d}$ converges almost surely to 0. Let Ω_0, $P(\Omega_0) = 1$, be the set on which $h_{n'',d}$ converges to 0. Then Lemma 4.1 implies

$$(\hat{\boldsymbol{\vartheta}}_{n''}(\omega),\hat{\boldsymbol{\gamma}}_{n''}(\omega)) \overset{n''\to\infty}{\longrightarrow} (\boldsymbol{\vartheta},\boldsymbol{\gamma})$$

for each $\omega \in \Omega_0$. Since $P(\Omega_0) = 1$, this concludes the proof. □

5 Some simulations

In this last section, we present some simulation results in order to illustrate the finite sample behavior of the presented procedures. For $d = 1, 2, 3$, we have generated n^d observations of two random fields $\{X^{(1)}_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ and $\{X^{(2)}_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ which correspond to Example 2.1 in the i.i.d. case and Example 2.2: $X^{(i)}_{\mathbf{k}} = I_{(\mathbf{k}_0,\mathbf{m}_0]}(\mathbf{k})\delta + e^{(i)}_{\mathbf{k}}$, where $\{e^{(1)}_{\mathbf{k}}\}$ are i.i.d. $N(0, 1)$-distributed and $\{e^{(2)}_{\mathbf{k}}\}$ is the MA-field $e^{(2)}_{\mathbf{k}} = \sum_{j_1=0}^{\infty}\cdots\sum_{j_d=0}^{\infty} 4^{-j_1}\cdots 4^{-j_d}\xi_{\mathbf{k}-\mathbf{j}}$ with i.i.d. $N(0, 1)$ distributed noise $\{\xi_{\mathbf{k}}\}$. The MA-field is simulated using the equivalent autoregressive presentation of $e^{(2)}_{\mathbf{k}}$ (cf. Tjøstheim (1978)). The corresponding true values for the parameter σ are

$$\sigma_1 = 1$$

and

$$\sigma_2 = \sum_{j_1=0}^{\infty}\cdots\sum_{j_d=0}^{\infty} 4^{-j_1}\cdots 4^{-j_d} = \left(\frac{1}{1-0.25}\right)^d = 1.333,\ 1.778,\ 2.370\ (d = 1, 2, 3).$$

We use the variance estimator of Subsection 3.3 with Bartlett weights

$$\omega_{q,\mathbf{j}} = \prod_{i=1}^{d}\left(1 - \frac{|j_i|}{q}\right)^+$$

and $q = \sqrt{n}$. We use $\alpha = 0.01$, $\beta = 0.01$ and a 5% significance level. The corresponding critical values 4.167 ($d = 1$), 5.971 ($d = 2$) and 7.095 ($d = 3$) were computed using Corollary 3.1 and (6). Under the alternative, we consider changes with different change height δ and over increasingly sized rectangles $(\mathbf{k}_0, \mathbf{m}_0]$, with $\mathbf{k}_0 = \lfloor n\boldsymbol{\theta}\rfloor$, $\mathbf{m}_0 = \lfloor n\boldsymbol{\gamma}\rfloor$, where the values for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ can be found in table 2.2. The accuracy of the estimated change-points $\hat{\mathbf{k}}_0$, $\hat{\mathbf{m}}_0$ is measured with the Jaccard similarity (Rajaraman and Ullman, 2012, Sec. 3.1.1, p. 54)

$$J((\mathbf{k}_0, \mathbf{m}_0], (\hat{\mathbf{k}}_0, \hat{\mathbf{m}}_0]) = \frac{\lambda((\mathbf{k}_0, \mathbf{m}_0] \cap (\hat{\mathbf{k}}_0, \hat{\mathbf{m}}_0])}{\lambda((\mathbf{k}_0, \mathbf{m}_0] \cup (\hat{\mathbf{k}}_0, \hat{\mathbf{m}}_0])}.$$

The Jaccard similarity is the ratio of the size of the intersection to the size of the union of the two rectangles and as such varies between 0 and 1, where 1 represents complete accuracy. Since the considered statistics only depend on $\lfloor n\boldsymbol{\theta}\rfloor$ and $\lfloor n\boldsymbol{\gamma}\rfloor$, we compare $\hat{\mathbf{k}}_0 = \lfloor n\hat{\boldsymbol{\theta}}\rfloor$ and $\hat{\mathbf{m}}_0 = \lfloor n\hat{\boldsymbol{\gamma}}\rfloor$ with $\mathbf{k}_0$ and $\mathbf{m}_0$ instead of considering $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\gamma}}$ directly.

Increasing sample sizes n^d with $n = 100, 500, 1000$ ($d = 1$), $n = 50, 100, 150$ ($d = 2$) and $n = 10, 20, 30$ ($d = 3$) were used. All simulated values were obtained from 1000 repetitions.

5.1 Discussion

Under H_0, the test is conservative for both known and unknown asymptotic variance, staying below the 5% nominal significance level even though the variance estimations with the chosen bandwidth $q = \sqrt{n}$ are smaller than the theoretical values. As expected, under the alternative the test and change-point estimator improve with increasing sample or change size. The test works better for large rectangles over which the change takes

Table 2.1: Empirical size of the test

		known σ		unknown σ			
n^d	d	i.i.d.	MA	$\hat{\sigma}_1$	i.i.d.	$\hat{\sigma}_2$	MA
100	1	0.003	0.001	0.9239	0.036	1.2141	0.007
500	1	0.015	0.012	0.9740	0.019	1.2700	0.012
1000	1	0.026	0.011	0.9783	0.018	1.2955	0.009
2500	2	0	0	0.9876	0	1.6034	0
10000	2	0.001	0	0.9935	0.001	1.6628	0
22500	2	0.006	0.001	0.9984	0.001	1.6822	0
1000	3	0	0	0.9890	0	1.6436	0
8000	3	0	0	0.9984	0	1.8367	0
27000	3	0	0	0.9983	0	1.9504	0

Table 2.2: Change points

Block parameters	$d=1$	$d=2$	$d=3$
$\boldsymbol{\theta}_1$	0.2	(0.2,0.2)	(0.2,0.2,0.2)
$\boldsymbol{\theta}_2$	0.4	(0.4,0.4)	(0.4,0.4,0.4)
$\boldsymbol{\theta}_3$	0.1	(0.1,0.1)	(0.1,0.1,0.1)
$\boldsymbol{\gamma}_1$	0.25	(0.4,0.6)	(0.3,0.5,0.7)
$\boldsymbol{\gamma}_2$	0.8	(0.8,0.8)	(0.8,0.8,0.8)
$\boldsymbol{\gamma}_3$	0.9	(0.9,0.9)	(0.9,0.9,0.9)

place. For sufficiently many observations and known asymptotic variance, almost all changes are detected even for small rectangles. The power of the test with unknown variance is slightly worse due to the overestimation of the variance, but nevertheless shows a strong improvement for increasing sample sizes. The estimator is more sensitive to the size of the changed rectangle than the test, with low accuracy for small rectangles and improving with increasing rectangle size. In general, the procedures work better for the i.i.d. than the moving average random field, but they work well in both cases.

Table 2.3: Empirical power of the test, $\boldsymbol{\theta}_1, \boldsymbol{\gamma}_1$, known σ

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.027	0.512	0.978	0.005	0.134	0.661
500	1	0.745	1	1	0.277	0.999	1
1000	1	0.996	1	1	0.834	1	1
2500	2	1	1	1	0.977	1	1
10000	2	1	1	1	1	1	1
22500	2	1	1	1	1	1	1
1000	3	0	0.725	1	0	0	0
8000	3	1	1	1	0.001	0.995	1
27000	3	1	1	1	0.926	1	1

Table 2.4: Empirical power of the test, $\boldsymbol{\theta}_2, \boldsymbol{\gamma}_2$, known σ

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.784	1	1	0.340	0.998	1
500	1	1	1	1	1	1	1
1000	1	1	1	1	1	1	1
2500	2	1	1	1	1	1	1
10000	2	1	1	1	1	1	1
22500	2	1	1	1	1	1	1
1000	3	0.768	1	1	0	0.228	1
8000	3	1	1	1	0.999	1	1
27000	3	1	1	1	1	1	1

Table 2.5: Empirical power of the test, $\boldsymbol{\theta}_3, \boldsymbol{\gamma}_3$, known σ

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.480	1	1	0.120	0.958	1
500	1	1	1	1	0.998	1	1
1000	1	1	1	1	1	1	1
2500	2	1	1	1	1	1	1
10000	2	1	1	1	1	1	1
22500	2	1	1	1	1	1	1
1000	3	1	1	1	0.274	1	1
8000	3	1	1	1	1	1	1
27000	3	1	1	1	1	1	1

Table 2.6: Empirical power of the test, unknown σ, i.i.d., $\boldsymbol{\theta}_1, \boldsymbol{\gamma}_1$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power
100	1	1.0284	0.018	1.2632	0.055	1.5828	0.203
500	1	1.2749	0.109	1.9360	0.94	2.7056	1
1000	1	1.4508	0.813	2.3584	1	3.3654	1
2500	2	1.8115	1	3.2043	1	4.6828	1
10000	2	2.5333	1	4.7694	1	7.0691	1
22500	2	3.0563	1	5.8664	1	8.7302	1
1000	3	1.0224	0	1.1192	0.385	1.2652	0.998
8000	3	1.1278	1	1.4474	1	1.8617	1
27000	3	1.2805	1	1.8913	1	2.6094	1

Table 2.7: Empirical power of the test, unknown σ, i.i.d., $\boldsymbol{\theta}_2, \boldsymbol{\gamma}_2$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power
100	1	1.6991	0	2.9922	0	4.3664	0
500	1	2.4216	0.508	4.5400	1	6.7212	1
1000	1	2.8297	1	5.4047	1	8.0346	1
2500	2	2.4033	1	4.4892	1	6.6408	1
10000	2	3.4209	1	6.6264	1	9.8794	1
22500	2	4.1499	1	8.1162	1	12.1222	1
1000	3	1.2869	0.028	1.9199	1	2.6588	1
8000	3	1.7715	1	3.0961	1	4.5098	1
27000	3	2.3683	1	4.4113	1	6.5227	1

Table 2.8: Empirical power of the test, unknown σ, i.i.d., $\boldsymbol{\theta}_3, \boldsymbol{\gamma}_3$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power	$\hat{\sigma}_1$	power
100	1	1.3953	0.001	2.2672	0	3.2344	0
500	1	1.9741	0.528	3.5838	1	5.2704	1
1000	1	2.3183	1	4.3215	1	6.3912	1
2500	2	2.6572	1	5.0345	1	7.4717	1
10000	2	4.0223	1	7.8611	1	11.7403	1
22500	2	5.0070	1	9.8662	1	14.7581	1
1000	3	1.7399	1	3.0311	1	4.4110	1
8000	3	2.7731	1	5.2755	1	7.8369	1
27000	3	3.9913	1	7.7956	1	11.6411	1

Table 2.9: Empirical power of the test, unknown σ, MA, $\boldsymbol{\theta}_1, \boldsymbol{\gamma}_1$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power
100	1	1.2786	0.005	1.4752	0.021	1.7573	0.085
500	1	1.5117	0.023	2.0975	0.591	2.8224	0.992
1000	1	1.6814	0.332	2.5062	1	3.4705	1
2500	2	2.2069	0.716	3.4420	1	4.8482	1
10000	2	2.8645	1	4.9547	1	7.1966	1
22500	2	3.3375	1	6.0152	1	8.8291	1
1000	3	1.6626	0	1.7224	0.001	1.8191	0.161
8000	3	1.9098	0.074	2.1137	1	2.4157	1
27000	3	2.1061	0.992	2.5220	1	3.0951	1

Table 2.10: Empirical power of the test, unknown σ, MA, $\boldsymbol{\theta}_2, \boldsymbol{\gamma}_2$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power
100	1	1.8521	0	3.0788	0	4.4257	0
500	1	2.5641	0.11	4.6238	0.99	6.78265	1
1000	1	2.9557	1	5.4733	1	8.0821	1
2500	2	2.7078	0.995	4.6558	1	6.7524	1
10000	2	3.6745	1	6.7630	1	9.9731	1
22500	2	4.3640	1	8.2275	1	12.1969	1
1000	3	1.8321	0	2.3144	0.165	2.9516	0.984
8000	3	2.3535	1	3.4656	1	4.7739	1
27000	3	2.8980	1	4.7152	1	6.7311	1

Table 2.11: Empirical power of the test, unknown σ, MA, $\boldsymbol{\theta}_3, \boldsymbol{\gamma}_3$

		$\delta = 1$		$\delta = 2$		$\delta = 3$	
n^d	d	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power	$\hat{\sigma}_2$	power
100	1	1.5923	0	2.3974	0	3.3304	0
500	1	2.1380	0.134	3.6791	0.993	5.3378	1
1000	1	2.4703	0.995	4.4064	1	6.4502	1
2500	2	2.9383	1	5.1865	1	7.5739	1
10000	2	4.2389	1	7.9751	1	11.8178	1
22500	2	5.1947	1	9.9662	1	14.8275	1
1000	3	2.1715	0.642	3.2921	1	4.5907	1
8000	3	3.1748	1	5.4988	1	7.9901	1
27000	3	4.3284	1	7.9740	1	11.7617	1

Table 2.12: Jaccard similarity for the change point estimator, $\boldsymbol{\theta}_1, \boldsymbol{\gamma}_1$

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.073	0.149	0.253	0.071	0.118	0.204
500	1	0.178	0.388	0.543	0.121	0.286	0.428
1000	1	0.275	0.509	0.655	0.192	0.401	0.557
2500	2	0.536	0.792	0.905	0.381	0.614	0.765
10000	2	0.769	0.936	0.979	0.589	0.835	0.933
22500	2	0.875	0.976	0.993	0.718	0.921	0.974
1000	3	0.028	0.043	0.065	0.022	0.026	0.034
8000	3	0.055	0.110	0.171	0.029	0.051	0.079
27000	3	0.093	0.203	0.327	0.045	0.088	0.136

Table 2.13: Jaccard similarity for the change point estimator, $\boldsymbol{\theta}_2, \boldsymbol{\gamma}_2$

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.867	0.968	0.989	0.811	0.949	0.984
500	1	0.969	0.993	0.998	0.946	0.988	0.997
1000	1	0.982	0.996	0.999	0.973	0.994	0.998
2500	2	0.880	0.980	0.998	0.735	0.933	0.982
10000	2	0.971	0.998	1.000	0.912	0.989	0.999
22500	2	0.988	1.000	1.000	0.960	0.996	1.000
1000	3	0.285	0.464	0.635	0.186	0.299	0.401
8000	3	0.548	0.839	0.952	0.320	0.546	0.739
27000	3	0.775	0.963	0.995	0.463	0.779	0.922

Table 2.14: Jaccard similarity for the change point estimator, $\boldsymbol{\theta}_3, \boldsymbol{\gamma}_3$

		i.i.d.			MA		
n^d	d	$\delta=1$	$\delta=2$	$\delta=3$	$\delta=1$	$\delta=2$	$\delta=3$
100	1	0.809	0.938	0.972	0.747	0.907	0.956
500	1	0.945	0.985	0.994	0.909	0.975	0.990
1000	1	0.971	0.992	0.997	0.952	0.988	0.995
2500	2	0.999	1.000	1.000	0.994	1.000	1.000
10000	2	1.000	1.000	1.000	1.000	1.000	1.000
22500	2	1.000	1.000	1.000	1.000	1.000	1.000
1000	3	1.000	1.000	1.000	0.991	1.000	1.000
8000	3	1.000	1.000	1.000	1.000	1.000	1.000
27000	3	1.000	1.000	1.000	1.000	1.000	1.000

Acknowledgements

I would like to thank my thesis adviser, Josef Steinebach, for suggesting the topic of this paper and for his continued and ongoing support of my academic pursuits. In addition, I would like to express my gratitude to Zakhar Kabluchko, who took time out of his busy schedule to consider a problem that arose in my work, ultimately proposing a helpful strategy for completing the proof of Corollary 3.1. I would also like to thank an anonymous referee, whose comments led to the inclusion of illustrative data (Section 5) and helped clarifying several results.

Chapter 3

Long-run variance estimation for spatial data under change-point alternatives

Béatrice Bucchia and Christoph Heuser
University of Cologne

Abstract

In this paper, we consider the problem of estimating the long-run variance (matrix) of an $\mathbb{R}^p$-valued multiparameter stochastic process $\{X_{\mathbf{k}}\}_{\mathbf{k}\in\{1,\ldots,n\}^d}$, $(n, p, d \in \mathbb{N}$, p, d fixed) whose mean-function has an abrupt jump. We consider processes of the form

$$X_{\mathbf{k}} = Y_{\mathbf{k}} + \mu + I_{C_n}(\mathbf{k})\Delta,$$

where I_C is the indicator function for a set C, the change-set $C_n \subset [1, n]^d$ is a finite union of rectangles and $\mu, \Delta \in \mathbb{R}^p$ are unknown parameters. The stochastic process $\{Y_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^d\}$ is assumed to fulfill a weak invariance principle. Due to the non-constant mean, kernel-type long-run variance estimators using the arithmetic mean of the observations as a mean estimator have an unbounded error for changes Δ that do not vanish for $n \to \infty$. To reduce this effect, we use a mean estimator which is based on an estimation of the set C_n. In the case where $C_n = (\lfloor n\boldsymbol{\theta}_1^0 \rfloor, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$ is a rectangle, we introduce an estimator $\hat{C}_n = (\lfloor n\hat{\boldsymbol{\theta}}_1 \rfloor, \lfloor n\hat{\boldsymbol{\theta}}_2 \rfloor]$ and study its convergence rate.

Keywords: long-run variance estimation, change-point estimation, change-point detection, random fields

AMS subject classification: 62H15, 62E20, 62M99, 60G60, 62H12

1 Introduction

In this paper, we present and analyze a kernel-type long-run variance matrix (LRV in the following) estimator for a multivariate random field under the assumption of a non-constant mean. Such an estimator is needed e.g. in change-point analysis when one is interested in testing whether a given data-set is stationary or whether there is a jump in the mean, dividing the data into two sets with (different) constant means. In this case, the magnitude of the difference between the arithmetic means over suitable subsets of the data can be used as an indicator of the likelihood of a non-constant mean. The resulting tests are often based on the asymptotic behavior of the test statistic under the null hypothesis. For tests based on the partial sums of observations under suitable weak dependence conditions, a functional central limit theorem can be used to determine the distributional limit of the test statistic as a function of a multiparameter Brownian motion, and appropriate normalization can be used to standardize the limit process, leaving the LRV Σ as the only nuisance parameter. In order to construct asymptotic tests it is therefore important to estimate Σ consistently under the null hypothesis, so that the unknown LRV Σ may be replaced by its estimator for sufficiently large sample sizes. This has already been widely investigated for processes with constant mean functions, amongst others by Newey and West (1987) and Andrews (1991) for multivariate time series and later by Politis and Romano (1996), Robinson (2007) and Lavancier (2008) for univariate random fields. Most of the publications on the subject focus on the (null hypothesis) case of constant means to derive consistency of the LRV estimators. However, since the estimator for Σ is often used as a scaling factor in change-point tests, it is also important to have an estimator which remains stable and bounded with respect to a change under the alternative. Otherwise, error in the estimation of Σ might lead to tests which display lower power for bigger changes. For example Vogelsang (1999) and Crainiceanu and Vogelsang (2001) investigate the problem of nonmonotonic power under data-dependent bandwidth choices for a test of mean shift in a univariate time series, noting that this might even lead to tests with no power against "obvious" changes, which could be detected with the naked eye. They conclude that this is due to the fact that the LRV estimator is constructed under the (misspecified) model of a stable mean. Indeed, under alternatives with abrupt changes in the mean, the arithmetic mean displays a bias which causes associated kernel-type LRV estimators to diverge for growing bandwidths. In order to avoid this effect — or at least attenuate it —, we consider LRV estimators that use a mean estimator which is more adapted to the change alternative. Depending on the accuracy of the change-set estimation, it is then possible to obtain a consistent estimator. This method has been well studied in the time series literature. For instance, Juhl and Xiao (2009) present an LRV estimator for a univariate time series which remains consistent and bounded under both the null and alternative hypotheses, where the mean function fulfills a Lipschitz condition under the alternative, and Antoch et al. (1997), Kejriwal (2009) and Hušková and Kirch (2010) investigate an At-Most-One-Change location model. The aim of this paper is to extend this methodology to the random field case.

This paper is organized as follows: In Section 2, we present notations, the model and the main assumptions on the considered process. In Section 3, we study the behavior of an LRV estimator constructed without taking the change into account and compare it to a modification which makes use of estimators for the magnitude and location of the change. Section 4 gives an example of a change-set estimator with the associated estimation rate. Finally, Section 5 contains a small simulation study in order to give an impression of the finite sample behavior of the estimators and associated change-point tests, both for simulated data and a real data-set. Technical proofs are relegated to the appendix.

2 Model and main assumptions

The following notations will be used throughout this paper. Let $\mathbb{R}^d$ ($d \in \mathbb{N}$) be the vector space of real vectors equipped with the usual partial order. For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we write $\mathbf{x} \vee \mathbf{y} = (\max\{x_1, y_1\}, \ldots, \max\{x_d, y_d\})^T$ and $\mathbf{x} \wedge \mathbf{y} = (\min\{x_1, y_1\}, \ldots, \min\{x_d, y_d\})^T$ as well as $\lfloor \mathbf{x} \rfloor = (\lfloor x_1 \rfloor, \ldots, \lfloor x_d \rfloor)^T$ for the integer part of $\mathbf{x}$, $|\mathbf{x}| = (|x_1|, \ldots, |x_d|)^T$ and $[\mathbf{x}] = x_1 \cdots x_d$. We use the notations $x^{(i)}$ or x_i for the i-th entry of a vector and analogously for matrices. The notation $\| \cdot \|$ is used to denote the maximum norm $\|\mathbf{x}\| = \max\limits_{i=1,\ldots,d} |x_i|$. Furthermore, for any integer $k \in \mathbb{N}_0$, we denote $(k, \ldots, k)' \in \mathbb{N}_0^d$ by $\underline{\mathbf{k}}$. A rectangle in $\mathbb{R}^d$ is a set of the form

$$(\mathbf{x}, \mathbf{y}] = \{\mathbf{z} = (z_1, \ldots, z_d)^T : x_i < z_i \leq y_i, \ i = 1, \ldots, d\}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ ($(\mathbf{x}, \mathbf{y}] = \emptyset$, if $x_i \geq y_i$ for some $i \in \{1, \ldots, d\}$). A rectangle in $\mathbb{Z}^d$ is the intersection of a rectangle in $\mathbb{R}^d$ and the set $\mathbb{Z}^d$. We denote the Lebesgue measure on $\mathbb{R}^d$ by λ. Note that for the union of two disjoint rectangles $(\mathbf{k}_1, \mathbf{m}_1]$ and $(\mathbf{k}_2, \mathbf{m}_2]$ with endpoints $\mathbf{k}_i, \mathbf{m}_i \in \mathbb{Z}^d$ it holds that

$$\lambda((\mathbf{k}_1, \mathbf{m}_1] \cup (\mathbf{k}_2, \mathbf{m}_2]) = \#((\mathbf{k}_1, \mathbf{m}_1] \cap \mathbb{Z}^d) + \#((\mathbf{k}_2, \mathbf{m}_2] \cap \mathbb{Z}^d),$$

where $\#A$ denotes the cardinality of a finite set A. Therefore, we do not always explicitly distinguish between the notations and take $\lambda(C)$ to mean either the Lebesgue measure of a set in $\mathbb{R}^d$ or (for finite sets) its cardinality. To simplify notation we write $\lambda(\mathbf{k}, \mathbf{m}] = \lambda((\mathbf{k}, \mathbf{m}])$ for any rectangle $(\mathbf{k}, \mathbf{m}]$. We denote the symmetric difference of two sets A and B by $A \triangle B$. For a function $f : D \to \mathbb{R}$, $D \subseteq \mathbb{R}^d$, the increment of f over a rectangle $(\mathbf{s}, \mathbf{t}] \subset D$ takes the form

$$f(\mathbf{s}, \mathbf{t}] = \begin{cases} \sum\limits_{\boldsymbol{\varepsilon} \in \{0,1\}^d} (-1)^{d - \sum_{i=1}^d \varepsilon_i} f(\mathbf{s} + \boldsymbol{\varepsilon}(\mathbf{t} - \mathbf{s})), & \mathbf{s} < \mathbf{t} \\ 0, & \mathbf{s} \not< \mathbf{t}. \end{cases}$$

Unless stated otherwise, we will always denote the complement of a set $R \subseteq (\underline{\mathbf{0}}, \underline{\mathbf{n}}]$ by $R^c = (\underline{\mathbf{0}}, \underline{\mathbf{n}}] \setminus R$ and take sums of the form $\sum_{\mathbf{k} \in R}$ to mean the summation over all $\mathbf{k} \in R \cap \mathbb{Z}^d$. The data-generating process considered here is an $\mathbb{R}^p$-valued random field $\{X_{\mathbf{k}}\}$ with

$$X_{\mathbf{k}} = Y_{\mathbf{k}} + \mu + I_{\mathbf{k} \in C_n} \Delta = Y_{\mathbf{k}} + \mu(\mathbf{k}), \quad \mathbf{k} \in [1, n]^d \cap \mathbb{Z}^d, \tag{1}$$

with a shift Δ that fulfills $\Delta^T \Delta > 0$, a subset $C_n \subset [1, n]^d$ and the mean function $\mu(\mathbf{k}) = EX_{\mathbf{k}} = \mu + I_{\mathbf{k} \in C_n} \Delta$. All the parameters are considered unknown. Since the mean deviates from its value μ on C_n, we call this the change-set. In particular, we have $C_n = (0, k_2^0]$ ($d = 1$) and $C_n = \left(\mathbf{k}_1^0, \mathbf{k}_2^0\right]$ ($d \geq 1$) in mind. For such rectangles C_n the resulting change-set problem is the straightforward generalization to the multiparameter case of a one-dimensional change-point problem with two change-points $0 < k_0 < m_0 < n$. This type of problem is known in the change-point literature as an epidemic change-point. A more detailed description of the epidemic change-point problem and its multiparameter version, as well as some references to further research, can be found in Bucchia (2014). In order to allow slightly more general change-sets for the LRV estimation, we consider the following case:

Assumption (C). *C_n is the finite union of disjoint rectangles, i.e.*

$$C_n \in \mathcal{A}_n = \{A \subseteq (\underline{\mathbf{0}}, \underline{\mathbf{n}}] : \exists N \leq m_n,\, i = 1, \ldots, N, \mathbf{k}_i, \mathbf{m}_i \in \mathbb{Z}^d : A = \sum_{i=1}^{N} (\mathbf{k}_i, \mathbf{m}_i]\},$$

where $m = m_n$ is known. Additionally, $n^{-d}\lambda(C_n) \xrightarrow{n\to\infty} a$ for some $a \in (0, 1)$ and $0 < \lambda(C_n) < n^d$ for all $n \in \mathbb{N}$.

Models of the form (1) are used in image segmentation and reconstruction problems. The observations fall into two segments, each with different statistical characteristics (in this case, different means), and the task is to find the segments and estimate the distributions on the different segments. This, as well as the related problem of edge detection, where the focus lies on detecting the boundary of the change-set, are well-known problems in image analysis (see e.g. Korostelev and Tsybakov (1993), Müller and Song (1994), Müller and Song (1996), Ferger (2004), Mallik (2013) and many more). However, while a lot of different models for the change-set C_n and the type of change have been considered, most of the literature deals with independent observations, whereas the model considered here allows (weak) dependence between the observations.

Remark 2.1. *Assumption* (C) *is fulfilled e.g. if*

$$C_n = \sum_{i=1}^{N} (\lfloor n\mathbf{s}_i \rfloor, \lfloor n\mathbf{t}_i \rfloor]$$

is created by scaling of a template set $C^0 = \sum_{i=1}^{N} (\mathbf{s}_i, \mathbf{t}_i] \subset (\underline{\mathbf{0}}, \underline{\mathbf{1}}]$ with $0 < \lambda(C^0) < 1$. For instance, in the particular cases mentioned above, this would be $C^0 = (\underline{\mathbf{0}}, \boldsymbol{\theta}_2^0]$ with $C_n = (\underline{\mathbf{0}}, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$ or $C^0 = (\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0]$ ($\underline{\mathbf{0}} < \boldsymbol{\theta}_1^0 < \boldsymbol{\theta}_2^0 < \underline{\mathbf{1}}$) with $C_n = (\lfloor n\boldsymbol{\theta}_1^0 \rfloor, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$. To simplify notations, we will write $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$ and assume $\underline{\mathbf{0}} \leq \boldsymbol{\theta}_1^0 < \boldsymbol{\theta}_2^0 < \underline{\mathbf{1}}$ instead of distinguishing between these two particular change-sets.

In order to derive asymptotic results, we make the following assumption about the process $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$, which we will always require in the remainder of this paper.

Assumption (Y1). *$\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ is an $\mathbb{R}^p$-valued, centered, weakly stationary random field with autocovariance function $\Gamma(\mathbf{k}) = \mathrm{Cov}(Y_{\underline{\mathbf{0}}}, Y_{\mathbf{k}})$, for which*

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} |\Gamma_{ij}(\mathbf{k})| < \infty \text{ for all } i, j \in \{1, \ldots, p\},$$

and

$$\Sigma := \sum_{\mathbf{k}\in\mathbb{Z}^d} \Gamma(\mathbf{k})$$

is positive-definite. Furthermore, we assume that Y fulfills a weak invariance principle, i.e.

$$\left\{ \Sigma^{-1/2} \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}} \leq \mathbf{k} \leq \lfloor n\mathbf{t} \rfloor} Y_{\mathbf{k}} \right\}_{\mathbf{t}\in[0,1]^d} \xrightarrow{D^p[0,1]^d} \{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}, \tag{2}$$

where W is a p-dimensional vector of independent Brownian sheets and $\xrightarrow{D^p[0,1]^d}$ denotes weak convergence in the multivariate Skorohod space $D^p[0,1]^d$.

Remark 2.2. 1. *Note that it follows from Assumption* (Y1) *that*

$$\left\| \sum_{\mathbf{k} \in M} Y_{\mathbf{k}} \right\| = \mathcal{O}_P(n^{d/2}) \quad \textit{for any } M \subseteq (\underline{\mathbf{0}}, \underline{\mathbf{n}}]$$

and

$$\max_{\underline{\mathbf{0}} \leq \mathbf{k} < \mathbf{m} \leq \underline{\mathbf{n}}} n^{-d/2} \left\| \sum_{\mathbf{k} < \mathbf{j} \leq \mathbf{m}} Y_{\mathbf{j}} \right\| = \sup_{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}} n^{-d/2} \left\| \sum_{\lfloor n\mathbf{s} \rfloor < \mathbf{j} \leq \lfloor n\mathbf{t} \rfloor} Y_{\mathbf{j}} \right\| = \mathcal{O}_P(1),$$

which implies

$$\max_{A \in \mathcal{A}_n} n^{-d/2} \left\| \sum_{\mathbf{j} \in A} Y_{\mathbf{j}} \right\| = \mathcal{O}_P(m),$$

with $\mathcal{A}_n$ as in Assumption (C).

2. *Instead of Assumption* (Y1), *one could assume a central limit theorem for set-indexed processes (cf. e.g. Bass and Pyke (1985), Alexander and Pyke (1986)) to obtain better rates. However, since our main focus will later be on change-sets with $m = 1$, it suffices for our purpose to assume the more classical version of the invariance principle. Examples of real-valued processes that fulfill this assumption can e.g. be found in Truquet (2008) and Bucchia (2014).*

3. *In order to avoid unnecessary complications in the notation, we only consider observations on an equal-sided index-set $\{1, \ldots, n\}^d$. This could easily be adapted to more general sets $\{1, \ldots, n_1\} \times \cdots \times \{1, \ldots, n_d\}$.*

3 Long-run variance estimators

A commonly used estimation method for the LRV consists of summing up estimators for the sample covariances, using a kernel-function to obtain lag-dependent weights. Denoting the arithmetic mean over all observations by $\overline{X}_n$, the classical LRV estimator applied to a random field has the form

$$\hat{\Sigma}_n = \sum_{\mathbf{j} \in B_{q-1}} \omega_{q,\mathbf{j}} \, \hat{\Gamma}_X(\mathbf{j}), \tag{3}$$

where $q = q(n) \in \{1, \ldots, n\}$ is a bandwidth parameter, $B_q = \{-q, \ldots, q\}^d$, $\omega_{q,\mathbf{j}}$ are weighting functions and

$$\hat{\Gamma}_X(\mathbf{j}) = \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} (X_{\mathbf{k}} - \overline{X}_n)(X_{\mathbf{k}+\mathbf{j}} - \overline{X}_n)^T$$

is an estimator of the covariance matrix with $N_{\mathbf{j}} = \{\mathbf{k} \in \mathbb{N}^d : \underline{\mathbf{1}} \leq \mathbf{k}, \mathbf{k} + \mathbf{j} \leq \underline{\mathbf{n}}\}$. This choice of covariance estimator is consistent under the assumption that there is no change in the mean, in which case $\overline{X}_n$ is a consistent estimator for the mean. However, it fails to take changes in the mean into account. To address this problem, we consider more general LRV estimators

$$\tilde{\Sigma}_n = \sum_{\mathbf{j} \in B_{q-1}} \omega_{q,\mathbf{j}} \, \tilde{\Gamma}_X(\mathbf{j}) \tag{4}$$

with

$$\tilde{\Gamma}_X(\mathbf{j}) = \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} (X_{\mathbf{k}} - \tilde{\mu}(\mathbf{k}))(X_{\mathbf{k}+\mathbf{j}} - \tilde{\mu}(\mathbf{k}+\mathbf{j}))^T,$$

where $\tilde{\mu}(\cdot)$ is an estimator of the mean function $\mu(\cdot) = \mu + I_{C_n}(\cdot)\Delta$. In order to study the effect of the mean estimation, we will compare the estimation results obtained using $\bar{\mu} \equiv \overline{X}_n$ with the result when using a more complex mean estimator which explicitly takes a possible change into account.

We do not use a specific weighting function but merely assume the following restrictions:

Assumption (W). *There is a constant $c > 0$ such that $0 \leq \omega_{q,\mathbf{j}} \leq c$ for all $q \in \mathbb{N}, \mathbf{j} \in \mathbb{Z}^d$ and $\omega_{q,\mathbf{j}} = 0$ for $\mathbf{j}$ with $\max_i |j_i| \geq q$. $\omega_{q,\mathbf{j}} = \omega(\mathbf{j}/q)$, where ω is a symmetric function that is continuous at zero with $\omega(\underline{\mathbf{0}}) = 1$.*

Remark 3.1. *Our results also hold for dimension-dependent bandwidths $\mathbf{q} = (q_1, \ldots, q_d)$ summed over sets $B_{\mathbf{q}} = \{-q_1 + 1, \ldots, q_1 - 1\} \times \cdots \times \{-q_d + 1, \ldots, q_d - 1\}$, but in order to avoid the associated technicalities, we limit the exposition to the simpler case of one-dimensional q.*

A natural approach to the estimation of the mean under a change alternative is to use an estimator of the change-set C_n (cf. e.g. Antoch et al. (1997), Kejriwal (2009) and Hušková and Kirch (2010)) and estimate the different mean levels as arithmetic means over fitting subsets. The estimation of C_n will be the subject of Section 4. For now, we derive results under the assumption that we have a change-set estimator $\hat{C}_n$ which fulfills the following assumption:

Assumption ($\hat{C}$). *$\hat{C}_n \in \mathcal{A}_n$ and there are constants $0 < \alpha < 1 - \beta < 1$ such that $\alpha n^d \leq \lambda(\hat{C}_n) \leq (1-\beta)n^d$ for all $n \in \mathbb{N}$.*

Assuming we have such an estimator $\hat{C}_n$, the following lemma quantifies the resulting estimation error for $\tilde{\mu}$.

Lemma 3.1. *Let Assumptions* (C) *and* (Y1) *be fulfilled and let $\hat{C}_n$ be an estimator of the change-set C_n fulfilling Assumption ($\hat{C}$) with*

$$\lambda(\hat{C}_n \triangle C_n) = \mathcal{O}_P(n^{d-\delta}) \tag{5}$$

for some $\delta > 0$. Then the estimator

$$\tilde{\mu}(\mathbf{k}) = \begin{cases} \frac{1}{\lambda(\hat{C}_n)} \sum\limits_{\mathbf{j} \in \hat{C}_n} X_{\mathbf{j}}, & \mathbf{k} \in \hat{C}_n \\ \frac{1}{\lambda(\hat{C}_n^c)} \sum\limits_{\mathbf{j} \notin \hat{C}_n} X_{\mathbf{j}}, & \mathbf{k} \notin \hat{C}_n \end{cases}$$

fulfills

$$\max_{\mathbf{k} \in (C_n \cap \hat{C}_n) \cup (C_n^c \cap \hat{C}_n^c)} \|\mu(\mathbf{k}) - \tilde{\mu}(\mathbf{k})\| = \mathcal{O}_P(mn^{-d/2}) + \mathcal{O}_P(n^{-\delta}\|\Delta\|)$$

and

$$\max_{\mathbf{k} \in \{1,\ldots,n\}^d} \|\tilde{\mu}(\mathbf{k}) - \mu(\mathbf{k})\| = \mathcal{O}_P(\|\Delta\|) + \mathcal{O}_P(mn^{-d/2}).$$

Proof. The assumption for the accuracy of the change-point estimator immediately implies $\lambda(\hat{C}_n \cap C_n^c) = \mathcal{O}_P(n^{d-\delta})$ and therefore it holds for $\mathbf{k} \in C_n \cap \hat{C}_n$ that

$$\begin{aligned}
\tilde{\mu}(\mathbf{k}) - \mu(\mathbf{k}) = &\frac{1}{\lambda(\hat{C}_n)} \sum_{\mathbf{j} \in \hat{C}_n} (X_{\mathbf{j}} - \mu - \Delta) \\
= &\frac{n^{d/2}}{\lambda(\hat{C}_n)} \left\{ n^{-d/2} \sum_{\mathbf{j} \in \hat{C}_n} Y_{\mathbf{j}} - n^{-d/2} \lambda(\hat{C}_n \cap C_n^c) \Delta \right\} \\
= &\mathcal{O}_P(mn^{-d/2}) + \mathcal{O}_P(n^{-\delta} \|\Delta\|),
\end{aligned}$$

where we have used

$$n^{-d/2} \left\| \sum_{\mathbf{j} \in \hat{C}_n} Y_{\mathbf{j}} \right\| \leq \max_{A \in \mathcal{A}_n} n^{-d/2} \left\| \sum_{\mathbf{j} \in A} Y_{\mathbf{j}} \right\| = \mathcal{O}_P(m)$$

and $n^d \lambda(\hat{C}_n)^{-1} = \mathcal{O}_P(1)$. For any $\mathbf{k} \in C_n^c \cap \hat{C}_n^c$, the relation

$$\tilde{\mu}(\mathbf{k}) - \mu(\mathbf{k}) = \frac{1}{\lambda(\hat{C}_n^c)} \sum_{\mathbf{j} \in \hat{C}_n^c} (X_{\mathbf{j}} - \mu) = \mathcal{O}_P(mn^{-d/2}) + \mathcal{O}_P(n^{-\delta} \|\Delta\|)$$

follows analogously. Finally, the above results imply

$$\max_{\mathbf{k} \in \{1,\ldots,n\}^d} \|\tilde{\mu}(\mathbf{k}) - \mu(\mathbf{k})\| \leq \max_{\mathbf{k} \in C_n \triangle \hat{C}_n} \|\mu(\mathbf{k}) - \tilde{\mu}(\mathbf{k})\| + \mathcal{O}_P(n^{-\delta} \|\Delta\|) + \mathcal{O}_P(mn^{-d/2}),$$

where

$$\begin{aligned}
&\max_{\mathbf{k} \in C_n \triangle \hat{C}_n} \|\mu(\mathbf{k}) - \tilde{\mu}(\mathbf{k})\| \\
\leq &\|\Delta\| + \left\| \frac{1}{\lambda(\hat{C}_n^c)} \sum_{\mathbf{j} \in \hat{C}_n^c} (X_{\mathbf{j}} - \mu) \right\| + \left\| \frac{1}{\lambda(\hat{C}_n)} \sum_{\mathbf{j} \in \hat{C}_n} (X_{\mathbf{j}} - \mu - \Delta) \right\| \\
= &\|\Delta\| + \mathcal{O}_P(n^{-\delta} \|\Delta\|) + \mathcal{O}_P(mn^{-d/2}).
\end{aligned}$$

□

Remark 3.2. *1. If there is no change in the mean, i.e.* $\Delta = 0$*, Assumptions* $(\hat{C})$ *and* (Y1) *imply* $\max_{\mathbf{k} \in \{1,\ldots,n\}^d} \|\tilde{\mu}(\mathbf{k}) - \mu(\mathbf{k})\| = \mathcal{O}_P(mn^{-d/2})$.

2. Under Assumption (C) *with* $m = 1$*, estimators* $\hat{C}_n = \left(\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2\right] = \left(\lfloor n\hat{\boldsymbol{\theta}}_1 \rfloor, \lfloor n\hat{\boldsymbol{\theta}}_2 \rfloor\right]$ *fulfill* (5) *if* $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ *satisfies* $n^\delta(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) = \mathcal{O}_P(1)$.

In the following, we consider the LRV estimator $\tilde{\Sigma}_n$ with mean estimator $\tilde{\mu}$ as defined in Lemma 3.1. In order to analyze the asymptotic behavior of the LRV estimators 3 and 4, we use the decompositions

$$\hat{\Sigma}_n = \Sigma_{Y,n} + \hat{R}_n$$

and

$$\tilde{\Sigma}_n = \Sigma_{Y,n} + \tilde{R}_n,$$

where

$$\Sigma_{Y,n} = \sum_{\mathbf{j} \in B_{q-1}} \omega_{q,\mathbf{j}} \, \tilde{\Gamma}_Y(\mathbf{j})$$

and

$$\Gamma_Y(\mathbf{j}) = \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} Y_{\mathbf{k}} Y_{\mathbf{k}+\mathbf{j}}^T.$$

Now, $\Sigma_{Y,n}$ is the well known consistent LRV estimator for centered random fields (cf. Lemma 3.2) and $\hat{R}_n$ and $\tilde{R}_n$ are noise terms. In order to prove consistency for $\Sigma_{Y,n}$, we need the following additional assumptions:

Assumption (Y2)**.** *$\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$ is fourth order stationary with summable cumulants, i.e. for all $a, b, c, d \in \{1, \ldots, p\}$*

$$E[Y_{\mathbf{k}}^{(a)} Y_{\mathbf{k}+\mathbf{l}}^{(b)} Y_{\mathbf{k}+\mathbf{m}}^{(c)} Y_{\mathbf{k}+\mathbf{n}}^{(d)}] = E[Y_{\underline{\mathbf{0}}}^{(a)} Y_{\mathbf{l}}^{(b)} Y_{\mathbf{m}}^{(c)} Y_{\mathbf{n}}^{(d)}]$$

and

$$\sup_{\mathbf{i} \in \mathbb{Z}^d} \sum_{(\mathbf{j},\mathbf{k}) \in \mathbb{Z}^{2d}} |c_{a,b}(\mathbf{i}, \mathbf{j}, \mathbf{k})| < \infty,$$

where

$$c_{a,b}(\mathbf{i}, \mathbf{j}, \mathbf{k}) = E[Y_{\underline{\mathbf{0}}}^{(a)} Y_{\mathbf{i}}^{(b)} Y_{\mathbf{j}}^{(a)} Y_{\mathbf{k}}^{(b)}] - \Gamma_{a,b}(\mathbf{i})\Gamma_{a,b}(\mathbf{k}-\mathbf{j}) - \Gamma_{a,a}(\mathbf{j})\Gamma_{b,b}(\mathbf{k}-\mathbf{i}) - \Gamma_{a,b}(\mathbf{k})\Gamma_{a,b}(\mathbf{j}-\mathbf{i})$$

represent the fourth order cumulants of the components of $\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$.

Remark 3.3. *Proofs of the consistency of the empirical LRV for a centered process $\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$ often involve the asymptotic variance of the estimator. It is therefore natural to require assumptions on the fourth moments of $\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$. Like similar conditions in this context (cf. e.g. Andrews (1991)), this assumption is a typical condition to prove the consistency of LRV estimators (cf. e.g. Giraitis et al. (2003), Lavancier (2008)). It is fulfilled e.g. for Gaussian random fields, linear fields with absolutely summable coefficients and some α-mixing random fields (cf. Guyon (1995), Lemmas 4.6.2 and 4.6.3).*

Lemma 3.2 (cf. Lavancier (2008))**.** *For $q = q(n) \to \infty$ with $\lim_{n \to \infty} q/n = 0$, it holds under Assumptions* (W)*,* (Y1) *and* (Y2) *that*

$$\Sigma_{Y,n} \xrightarrow{P} \Sigma \quad \textit{for } n \to \infty.$$

Proof. The convergence follows from componentwise convergence of the matrices, a proof of which can be found in Lavancier (2008). (Although it is stated for the Bartlett-kernel $\omega_{q,\mathbf{j}} = \prod_{i=1}^{d} \left(1 - \frac{|j_i|}{q}\right)$, it can easily be seen that Lavancier (2008)'s proof can be applied to any kernel satisfying Assumption (W).) □

Remark 3.4. *As remarked by Giraitis et al. (2003), Lemma 3.2 also holds if we replace the assumptions $q = o(n)$ and Assumption* (Y2) *by $q = o(n^{1-\alpha})$, where $0 \leq \alpha < 1$ with*

$$\sup_{\mathbf{h} \in \mathbb{Z}^d} \sum_{\mathbf{r},\mathbf{s} \in [-\underline{\mathbf{N}}, \underline{\mathbf{N}}]} |c_{a,b}(\mathbf{h}, \mathbf{r}, \mathbf{s})| \leq D N^{\alpha d}$$

for all $N \in \mathbb{N}$ and $a, b \in \{1, \ldots, p\}$ and some $D > 0$.

Theorem 3.1. *Suppose Assumption* (W) *as well as the assumptions of Lemma 3.1 are fulfilled and* $q = q(n) \to \infty$ *with* $\lim_{n\to\infty} q/n = 0$. *Then it holds that*

$$\hat{R}_n = \mathcal{O}_P(n^{-d/2}q^d\|\Delta\| + q^d\|\Delta\|^2)$$

and

$$\tilde{R}_n = \mathcal{O}_P(mn^{-d/2}q^d\|\Delta\| + n^{-\delta}q^d\|\Delta\|^2 + m^2(q/n)^d).$$

Remark 3.5. 1. *Since* $q \to \infty$ *is a necessary assumption for the consistency of* $\Sigma_{Y,n}$, *the rate for* $\hat{R}_n$ *only leads to a consistent LRV estimator for* $\Delta = \Delta_n \to 0$ $(n \to \infty)$. *The estimator* $\tilde{\Sigma}_n$ *has no such restriction and can therefore be used for constant change magnitudes* Δ *(as is e.g. assumed in Section 4).*

2. *If we additionally assume that* $C_n = (\lfloor n\boldsymbol{\theta}_1^0\rfloor, \lfloor n\boldsymbol{\theta}_2^0\rfloor]$ *and* $\hat{C}_n = (\lfloor n\hat{\boldsymbol{\theta}}_1\rfloor, \lfloor n\hat{\boldsymbol{\theta}}_2\rfloor]$ *are rectangles, the rate can be improved to*

$$\tilde{R}_n = \mathcal{O}_P(n^{-d/2}q^d\|\Delta\|) + \mathcal{O}_P(n^{-(\delta+1)}q^{d+1}\|\Delta\|^2).$$

3. *If* $m = \mathcal{O}(1)$ *and there is no change, i.e.* $\Delta = 0$, *straightforward calculations imply that* $T_i = \mathcal{O}_P(n^{-d})$, $i = 1, 2, 3$ *(see below), for both mean estimators* $\bar{\mu}$. *Therefore, the estimators* $\hat{\Sigma}_n$ *and* $\tilde{\Sigma}_n$ *are consistent under the assumptions of Lemma 3.2, as long as* $\alpha n^d \le \lambda(\hat{C}_n) \le (1-\beta)n^d$ *for some parameters* $0 < \alpha < 1-\beta < 1$.

Proof of Theorem 3.1. Let $\bar{\mu}$ be either $\overline{X}_n$ or the estimator $\tilde{\mu}$ of Lemma 3.1. It holds that

$$\begin{aligned}
&\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}}(X_\mathbf{k} - \bar{\mu}(\mathbf{k}))(X_{\mathbf{k}+\mathbf{j}} - \bar{\mu}(\mathbf{k}+\mathbf{j}))^T \\
=&\Gamma_Y(\mathbf{j}) + \frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}}(\mu(\mathbf{k}) - \bar{\mu}(\mathbf{k}))Y_{\mathbf{k}+\mathbf{j}}^T + \frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}} Y_\mathbf{k}(\mu(\mathbf{k}+\mathbf{j}) - \bar{\mu}(\mathbf{k}+\mathbf{j}))^T \\
&+\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}}(\mu(\mathbf{k}) - \bar{\mu}(\mathbf{k}))(\mu(\mathbf{k}+\mathbf{j}) - \bar{\mu}(\mathbf{k}+\mathbf{j}))^T \\
=&\Gamma_Y(\mathbf{j}) + T_1 + T_2 + T_3.
\end{aligned}$$

For $\bar{\mu} \equiv \overline{X}_n = \overline{Y}_n + \mu + n^{-d}\lambda(C_n)\Delta$, we obtain for any $a, b \in \{1, \ldots, p\}$, using Remark 2.2,

$$\begin{aligned}
|T_1^{(a,b)}| &= \left|\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}}(\mu^{(a)} + I_{\mathbf{k}\in C_n}\Delta^{(a)} - \overline{Y}_n^{(a)} - \mu^{(a)} - n^{-d}\lambda(C_n)\Delta^{(a)})Y_{\mathbf{k}+\mathbf{j}}^{(b)}\right| \\
&\le |\overline{Y}_n^{(a)}|\left|\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}} Y_{\mathbf{k}+\mathbf{j}}^{(b)}\right| + |\Delta^{(a)}|\left|\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}\cap C_n} Y_{\mathbf{k}+\mathbf{j}}^{(b)}\right| + |\Delta^{(a)}|\left|\frac{1}{n^d}\sum_{\mathbf{k}\in N_\mathbf{j}} Y_{\mathbf{k}+\mathbf{j}}^{(b)}\right|\frac{\lambda(C_n)}{n^d} \\
&= \mathcal{O}_P(|\Delta^{(a)}|n^{-d/2})
\end{aligned}$$

and $T_2 = \mathcal{O}_P(\|\Delta\| n^{-d/2})$ analogously. Furthermore,

$$\begin{aligned}
T_3 =& \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} (\mu(\mathbf{k}) - \overline{X}_n)(\mu(\mathbf{k}+\mathbf{j}) - \overline{X}_n)^T \\
=& \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left(I_{C_n}(\mathbf{k}) - n^{-d}\lambda(C_n)\right)\left(I_{C_n}(\mathbf{k}+\mathbf{j}) - n^{-d}\lambda(C_n)\right) \Delta\Delta^T + n^{-d}\lambda(N_{\mathbf{j}})\overline{Y}_n \overline{Y}_n^T \\
& - \overline{Y}_n \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left(I_{C_n}(\mathbf{k}+\mathbf{j}) - n^{-d}\lambda(C_n)\right) \Delta^T - n^{-d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} \left(I_{C_n}(\mathbf{k}) - n^{-d}\lambda(C_n)\right) \Delta \overline{Y}_n^T \\
=& \mathcal{O}(\|\Delta\|^2) + \mathcal{O}_P(n^{-d/2}\|\Delta\|)
\end{aligned}$$

implies

$$\hat{R}_n = \sum_{\mathbf{j} \in B_q} \omega_{q,\mathbf{j}}(T_1 + T_2 + T_3) = \mathcal{O}_P(n^{-d/2} q^d \|\Delta\|) + \mathcal{O}_P(q^d \|\Delta\|^2).$$

When $\bar{\mu} = \tilde{\mu}$ from Lemma 3.1, we write $Y(\hat{C}_n) = \lambda(\hat{C}_n)^{-1} \sum_{\mathbf{i} \in \hat{C}_n} Y_{\mathbf{i}}$ and $Y(\hat{C}_n^c) = \lambda(\hat{C}_n^c)^{-1} \sum_{\mathbf{i} \in \hat{C}_n^c} Y_{\mathbf{i}}$ and, using $Y(\hat{C}_n) = \mathcal{O}_P(mn^{-d/2})$, $Y(\hat{C}_n^c) = \mathcal{O}_P(mn^{-d/2})$, we obtain for any $a, b \in \{1, \ldots, p\}$

$$\begin{aligned}
|T_1^{(a,b)}| &= \left| \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}}} (\mu^{(a)} + I_{\mathbf{k} \in C_n}\Delta^{(a)} - \tilde{\mu}^{(a)}(\mathbf{k})) Y_{\mathbf{k}+\mathbf{j}}^{(b)} \right| \\
&= \left| -Y(\hat{C}_n)^{(a)} \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap \hat{C}_n} Y_{\mathbf{k}+\mathbf{j}}^{(b)} - Y(\hat{C}_n^c)^{(a)} \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap \hat{C}_n^c} Y_{\mathbf{k}+\mathbf{j}}^{(b)} \right. \\
&\quad + \Delta^{(a)} \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap C_n} Y_{\mathbf{k}+\mathbf{j}}^{(b)} - \frac{\lambda(C_n \cap \hat{C}_n)}{\lambda(\hat{C}_n)} \Delta^{(a)} \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap \hat{C}_n} Y_{\mathbf{k}+\mathbf{j}}^{(b)} \\
&\quad \left. - \frac{\lambda(C_n \cap \hat{C}_n^c)}{\lambda(\hat{C}_n^c)} \Delta^{(a)} \frac{1}{n^d} \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap \hat{C}_n^c} Y_{\mathbf{k}+\mathbf{j}}^{(b)} \right| \\
&= \mathcal{O}_P(|\Delta^{(a)}| mn^{-d/2} + m^2 n^{-d})
\end{aligned}$$

and $T_2 = \mathcal{O}_P(\|\Delta\| mn^{-d/2} + m^2 n^{-d})$ analogously. Next, Lemma 3.1 implies

$$\begin{aligned}
& \sum_{\mathbf{k} \in N_{\mathbf{j}}} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})^2 \\
=& \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap C_n \cap \hat{C}_n} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})^2 + \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap C_n^c \cap \hat{C}_n^c} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})^2 \\
& + \sum_{\mathbf{k} \in N_{\mathbf{j}} \cap (C_n \triangle \hat{C}_n)} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})^2 \\
\leq& (\lambda(C_n \cap \hat{C}_n) + \lambda(C_n^c \cap \hat{C}_n^c))(\mathcal{O}_P(m^2 n^{-d}) + \mathcal{O}_P(n^{-2\delta}\|\Delta\|^2)) \\
& + \lambda(N_{\mathbf{j}} \cap (C_n \triangle \hat{C}_n))(\mathcal{O}_P(\|\Delta\|^2) + \mathcal{O}_P(m^2 n^{-d})) \\
=& \mathcal{O}_P(m^2) + \mathcal{O}_P(n^{d-\delta}\|\Delta\|^2)
\end{aligned}$$

and thus

$$
\begin{aligned}
|T_3^{(a,b)}| &= \left| \frac{1}{n^d} \sum_{\mathbf{k}\in N_{\mathbf{j}}} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})(\mu(\mathbf{k}+\mathbf{j})^{(b)} - \tilde{\mu}(\mathbf{k}+\mathbf{j})^{(b)}) \right| \\
&\leq \left(\frac{1}{n^d} \sum_{\mathbf{k}\in N_{\mathbf{j}}} (\mu(\mathbf{k})^{(a)} - \tilde{\mu}(\mathbf{k})^{(a)})^2 \cdot \frac{1}{n^d} \sum_{\mathbf{k}\in N_{\mathbf{j}}} (\mu(\mathbf{k}+\mathbf{j})^{(b)} - \tilde{\mu}(\mathbf{k}+\mathbf{j})^{(b)})^2 \right)^{1/2} \\
&= \mathcal{O}_P(m^2 n^{-d}) + \mathcal{O}_P(n^{-\delta}\|\Delta\|^2).
\end{aligned}
$$

Finally, we obtain

$$
\tilde{R}_n = \sum_{\mathbf{j}\in B_q} \omega_{q,\mathbf{j}}(T_1 + T_2 + T_3) = \mathcal{O}_P(mn^{-d/2}q^d\|\Delta\| + n^{-\delta}q^d\|\Delta\|^2 + m^2(q/n)^d).
$$

□

Remark 3.6. *We have restricted our presentation to the case where the mean changes but the variance of the process stays constant. Note that if C_n is a rectangle (i.e. $m = 1$), similar arguments can be used to obtain an LRV estimator with the same rate for the model $X_{\mathbf{k}} = \sigma_{\mathbf{k}} Y_{\mathbf{k}} + \mu + \Delta I_{C_n}(\mathbf{k})$ with*

$$
EY_{\mathbf{k}}^2 = 1 \quad \textit{and} \quad \sigma_{\mathbf{k}} = \begin{cases} \sigma, & \mathbf{k} \notin C_n \\ \sigma*, & \mathbf{k} \in C_n \end{cases}.
$$

4 Change-point estimation

We now focus on the special case of change-sets $C_n = (\lfloor n\boldsymbol{\theta}_1^0 \rfloor, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$ (cf. Remark 2.1). In this case, the problem of estimating C_n can be reduced to finding estimators of $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$. In the following, we consider estimators of the form

$$
(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \in \arg\max\{Q_n(\mathbf{s},\mathbf{t}) : \underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}\}
$$

for $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$ with $\underline{\mathbf{0}} \leq \boldsymbol{\theta}_1^0 < \boldsymbol{\theta}_2^0 \leq \underline{\mathbf{1}}$ and

$$
Q_n(\mathbf{s},\mathbf{t}) = \left(\sum_{\lfloor n\mathbf{s}\rfloor < \mathbf{i} \leq \lfloor n\mathbf{t}\rfloor} (X_{\mathbf{i}} - \overline{X}_n) \right)^T \left(\sum_{\lfloor n\mathbf{s}\rfloor < \mathbf{i} \leq \lfloor n\mathbf{t}\rfloor} (X_{\mathbf{i}} - \overline{X}_n) \right).
$$

This corresponds to the change-point estimator proposed by Aston and Kirch (2012a) for $d = 1$. Since $Q_n(\mathbf{s},\mathbf{t})$ only depends on $(\mathbf{k}_1, \mathbf{k}_2) = (\lfloor n\mathbf{s}\rfloor, \lfloor n\mathbf{t}\rfloor)$, we can equivalently consider — writing $Q_n(\mathbf{s},\mathbf{t}) = Q_n(\lfloor n\mathbf{s}\rfloor, \lfloor n\mathbf{t}\rfloor)$ in a slight abuse of notation — $(\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2) \in \arg\max\{Q_n(\mathbf{k}_1, \mathbf{k}_2) : \underline{\mathbf{0}} \leq \mathbf{k}_1 < \mathbf{k}_2 \leq \underline{\mathbf{n}}\} = \arg\max\{Q_n(\mathbf{k}_1, \mathbf{k}_2) - Q_n(\mathbf{k}_1^0, \mathbf{k}_2^0)\}$ and $\hat{\boldsymbol{\theta}}_i = n^{-1} \cdot \hat{\mathbf{k}}_i$ $(i = 1, 2)$. For notational convenience, we will denote the vectors $(\mathbf{k}_1, \mathbf{k}_2)$ and $(\mathbf{k}_1^0, \mathbf{k}_2^0)$ by $\mathbf{k}$ and $\mathbf{k}^0$ respectively and analogously for $\boldsymbol{\theta}^0$ and $\hat{\boldsymbol{\theta}}$. In addition, we will denote the rectangles $(\mathbf{k}_1, \mathbf{k}_2]$ and $(\mathbf{k}_1^0, \mathbf{k}_2^0]$ by $R_{\mathbf{k}}$ and $R_{\mathbf{k}^0}$ respectively.

Remark 4.1. *If we additionally assume that for some $\alpha \in (0, 1)$ it holds that $\alpha \leq [\boldsymbol{\theta}_2^0 - \boldsymbol{\theta}_1^0] \leq 1 - \alpha$, we can restrict the estimators to*

$$
(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2) \in \arg\max\{Q_n(\mathbf{s},\mathbf{t}) : \underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}, \alpha \leq [\mathbf{t} - \mathbf{s}] \leq 1 - \alpha\}
$$

and

$$(\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2) \in \arg\max\{Q_n(\mathbf{k}_1, \mathbf{k}_2) : \underline{\mathbf{0}} \leq \mathbf{k}_1 < \mathbf{k}_2 \leq \underline{\mathbf{n}}, \lfloor \tilde{\alpha} n^d \rfloor \leq [\mathbf{k}_2 - \mathbf{k}_1] \leq \lfloor (1-\tilde{\alpha}) n^d \rfloor\}$$

for some $0 < \tilde{\alpha} \leq \alpha$. *Therefore, we can always assume w.l.o.g. that there are* $0 < \alpha < 1 - \beta < 1$ *with* $\alpha n^d \leq \lambda((\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2]) \leq (1-\beta) n^d$.

For this section we do not need to assume the existence of finite fourth moments. In fact, we can replace Assumption (Y2) by the weaker

Assumption $(Y2^\star)$**.** *There is an* $r > 2$ *such that for all* $l = 1, \ldots, p$

$$E \left| \sum_{\mathbf{j} \in (\mathbf{k}, \mathbf{m}]} Y_{\mathbf{j}}^{(l)} \right|^r \leq \tilde{c} \lambda(\mathbf{k}, \mathbf{m}]^{r/2}$$

for all $\mathbf{k} \leq \mathbf{m}$ *and a constant* $\tilde{c} > 0$ *that may depend on the dimension* d *and on* l *but not on* $\mathbf{k}$ *or* $\mathbf{m}$.

Remark 4.2. 1. *It can easily be seen that Assumption* (Y2) *implies* $(Y2^\star)$ *with* $r = 4$. *Assumption* $(Y2^\star)$ *implies tightness in* $D^p[0,1]^d$ *and is therefore often used together with the convergence of the finite-dimensional distributions to prove the functional central limit theorem* (2) *(cf. e.g. Bickel and Wichura (1971), Billingsley (1999), Truquet (2008)).*

2. *Since for any two rectangles* $(\mathbf{k}_1, \mathbf{m}_1]$, $(\mathbf{k}_2, \mathbf{m}_2]$ *the set* $(\mathbf{k}_1, \mathbf{m}_1] \setminus (\mathbf{k}_2, \mathbf{m}_2]$ *is the union of a finite number of rectangles, Assumption* $(Y2^\star)$ *implies*

$$E \left| \sum_{\mathbf{j} \in (\mathbf{k}_1, \mathbf{m}_1] \setminus (\mathbf{k}_2, \mathbf{m}_2]} Y_{\mathbf{j}}^{(l)} \right|^r \leq \tilde{c} \lambda((\mathbf{k}_1, \mathbf{m}_1] \setminus (\mathbf{k}_2, \mathbf{m}_2])^{r/2}.$$

Theorem 4.1. *Assume the change-set has the form* $C_n = (\lfloor n\boldsymbol{\theta}_1^0 \rfloor, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$ *with* $\underline{\mathbf{0}} < \boldsymbol{\theta}_1^0 < \boldsymbol{\theta}_2^0 < \underline{\mathbf{1}}$ *and* $0 < [\boldsymbol{\theta}_2^0 - \boldsymbol{\theta}_1^0] < 1$. *Under the Assumptions* (Y1) *and* $(Y2^\star)$, *it holds for a constant change size* $\Delta \neq 0$ *that*

$$n \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| = O_P(1).$$

Proof. The following proof is inspired by the proof by Aston and Kirch (2012a) and follows roughly the same lines. Nevertheless, we provide a fairly detailed proof below since the techniques required by our framework differ notably from those used by Aston and Kirch (2012a). W.l.o.g. let $\mu = 0$. Consider

$$\begin{aligned}
&Q_n(\mathbf{k}_1, \mathbf{k}_2) - Q_n(\mathbf{k}_1^0, \mathbf{k}_2^0) \\
&= \left(\sum_{\mathbf{k}_1 < \mathbf{i} \leq \mathbf{k}_2} (X_{\mathbf{i}} - \overline{X}_n) - \sum_{\mathbf{k}_1^0 < \mathbf{i} \leq \mathbf{k}_2^0} (X_{\mathbf{i}} - \overline{X}_n) \right)^T \\
&\quad \cdot \left(\sum_{\mathbf{k}_1 < \mathbf{i} \leq \mathbf{k}_2} (X_{\mathbf{i}} - \overline{X}_n) + \sum_{\mathbf{k}_1^0 < \mathbf{i} \leq \mathbf{k}_2^0} (X_{\mathbf{i}} - \overline{X}_n) \right) \\
&= \left(A_{\mathbf{k}_1, \mathbf{k}_2}^{(1)} + \Delta B_{\mathbf{k}_1, \mathbf{k}_2}^{(1)} \right)^T \left(A_{\mathbf{k}_1, \mathbf{k}_2}^{(2)} + \Delta B_{\mathbf{k}_1, \mathbf{k}_2}^{(2)} \right),
\end{aligned}$$

where

$$A^{(1)}_{\mathbf{k}_1,\mathbf{k}_2} = \sum_{\mathbf{i}\in R_{\mathbf{k}}\backslash R_{\mathbf{k}^0}} Y_{\mathbf{i}} - \sum_{\mathbf{j}\in R_{\mathbf{k}^0}\backslash R_{\mathbf{k}}} Y_{\mathbf{j}} - \frac{\lambda(R_{\mathbf{k}}\backslash R_{\mathbf{k}^0}) - \lambda(R_{\mathbf{k}^0}\backslash R_{\mathbf{k}})}{n^d} \sum_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}} Y_{\mathbf{i}},$$
$$B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2} = -\lambda(R_{\mathbf{k}^0}\backslash R_{\mathbf{k}}) - (\lambda(R_{\mathbf{k}}\backslash R_{\mathbf{k}^0}) - \lambda(R_{\mathbf{k}^0}\backslash R_{\mathbf{k}}))\frac{\lambda(R_{\mathbf{k}^0})}{n^d},$$
$$A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2} = \sum_{\mathbf{i}\in R_{\mathbf{k}}} Y_{\mathbf{i}} + \sum_{\mathbf{i}\in R_{\mathbf{k}^0}} Y_{\mathbf{i}} - \frac{\lambda(R_{\mathbf{k}}) + \lambda(R_{\mathbf{k}^0})}{n^d} \sum_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}} Y_{\mathbf{i}},$$
$$B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2} = \lambda(R_{\mathbf{k}}\cap R_{\mathbf{k}^0}) + \lambda(R_{\mathbf{k}^0}) - (\lambda(R_{\mathbf{k}}) + \lambda(R_{\mathbf{k}^0}))\frac{\lambda(R_{\mathbf{k}^0})}{n^d}.$$

We define $L_{n,\mathbf{k}_1,\mathbf{k}_2} = B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2} B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}$. Using

$$\alpha_n(N) = \begin{cases} \min\{N^{1/2-1/r}, \left(\sum_{i=N}^{\infty} \frac{1}{i^{r/2}}\right)^{-1/r}\}, & d=1 \\ n^{(d-1)(1/2-1/r)}, & d>1 \end{cases}$$

and combining the results from Lemma A2 yields

$$\begin{aligned}
&\mathbb{P}\left(n\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| \ge N+1\right) \le \mathbb{P}\left(\|\hat{\mathbf{k}} - \mathbf{k}^0\| \ge N\right) \\
&= \mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\ge N}} (Q_n(\mathbf{k}_1,\mathbf{k}_2) - Q_n(\mathbf{k}_1^0,\mathbf{k}_2^0)) \ge \max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|< N}} (Q_n(\mathbf{k}_1,\mathbf{k}_2) - Q_n(\mathbf{k}_1^0,\mathbf{k}_2^0)) \right) \\
&\le \mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\ge N}} (Q_n(\mathbf{k}_1,\mathbf{k}_2) - Q_n(\mathbf{k}_1^0,\mathbf{k}_2^0)) \ge 0 \right) \\
&\le \mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\ge N}} L_{n,\mathbf{k}_1,\mathbf{k}_2} \left(\frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2} A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{L_{n,\mathbf{k}_1,\mathbf{k}_2}} + \frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}}\Delta + \Delta^T \frac{A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}} + \Delta^T\Delta \right) \ge 0 \right) \\
&\le \mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\ge N}} L_{n,\mathbf{k}_1,\mathbf{k}_2} \left(\mathcal{O}_P(n^{-d/2}) + \alpha_n^{-1}(N)\mathcal{O}_P(1) + \Delta^T\Delta\right) \ge 0 \right) \\
&\le \mathbb{P}\left(\mathcal{O}_P(n^{-d/2}) + \alpha_n^{-1}(N)\mathcal{O}_P(1) + \Delta^T\Delta \le 0\right),
\end{aligned}$$

where the last inequality follows from $L_{n,\mathbf{k}_1,\mathbf{k}_2} \le -C < 0$. Since $\Delta^T\Delta > 0$ and $\alpha_n(N) \overset{n,N\to\infty}{\longrightarrow} \infty$, this probability becomes arbitrarily small for large N ($d=1$) and n ($d>1$). □

Although we do not explicitly mention this in the proof, our arguments can also be used to treat the case $C_n = (\underline{\mathbf{0}}, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$, and we have therefore implicitly proved the following corollary:

Corollary 4.1. *If $C_n = (\underline{\mathbf{0}}, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]$, $\underline{\mathbf{0}} < \boldsymbol{\theta}_2^0 < \underline{\mathbf{1}}$, it holds under the assumptions of Theorem 4.1 that*

$$n\|\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}_2^0\| = O_P(1),$$

with $\hat{\boldsymbol{\theta}}_2 = \hat{\mathbf{k}}_2/n$, $\hat{\mathbf{k}}_2 \in \arg\max\{Q_n(\mathbf{k}_2)\} = \arg\max\{Q_n(\mathbf{k}_2) - Q_n(\mathbf{k}_2^0)\}$, *where* $\mathbf{k}_2^0 = \lfloor n\boldsymbol{\theta}_2^0 \rfloor$ *and*

$$Q_n(\mathbf{k}_2) = \left(\sum_{\underline{\mathbf{0}} < \mathbf{i} \le \mathbf{k}_2} (X_{\mathbf{i}} - \overline{X}_n) \right)^T \left(\sum_{\underline{\mathbf{0}} < \mathbf{i} \le \mathbf{k}_2} (X_{\mathbf{i}} - \overline{X}_n) \right).$$

5 Finite sample results by simulations

5.1 Considered model

In this section we compare, by simulations, the finite sample behavior of the long-run variance estimator $\tilde{\Sigma}_n$ introduced in (4), where the estimator of the mean function is presented in Lemma 3.1, with the classical long-run variance estimator $\hat{\Sigma}_n$, where the mean function is estimated by the sample mean. To do so, we consider the real-valued random field

$$X_{\mathbf{k}} = Y_{\mathbf{k}} + \Delta I_{(\lfloor n\boldsymbol{\theta}_1^0 \rfloor, \lfloor n\boldsymbol{\theta}_2^0 \rfloor]}(\mathbf{k}), \quad \mathbf{k} \in \{1, \ldots, n\}^d,$$

for $d = 1, 2, 3$ and

$$Y_{\mathbf{k}} = \sum_{j_1=0}^{\infty} \cdots \sum_{j_d=0}^{\infty} a^{j_1} \ldots a^{j_d} \epsilon_{\mathbf{k}-\mathbf{j}} \tag{6}$$

with i.i.d. $N(0,1)$-distributed innovations $\epsilon_{\mathbf{k}}$ and $a \in \{-0.5, 0.5\}$. We consider $\Delta = 0, 0.5, 1, \ldots, 4$ as well as different sample sizes n_1, n_2, n_3 for each d (cf. Table 3.1).

Note that $\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$ fulfills Assumptions (Y1) and (Y2), since the coefficients are absolutely summable (cf. Marinucci and Poghosyan (2001) and Remark 3.3). The MA-field $\{Y_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^d}$ is simulated using its equivalent autoregressive representation (cf. Tjøstheim (1978)). We use the estimator presented in Section 4 to estimate the change-points $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0)$. Since the choice of kernel function is not the subject of this paper, we use the Bartlett-kernel $\omega_{q,\mathbf{j}} = \prod_{i=1}^d (1 - |j_i|/q)$ (or, where applicable, $\omega_{\mathbf{q},\mathbf{j}} = \prod_{i=1}^d (1 - |j_i|/q_i)$) for both LRV estimators as an example.

In order to investigate the effect of the volume (Vol) of the change-set on the estimation, we consider three different change-point settings, where C_n is small, medium sized and large (cf. Table 3.2).

All simulated values were obtained using 1000 repetitions.

5.2 Accuracy of the long-run variance estimation

To distinguish between the effect of the variance and the covariance in the estimation, we start by investigating the behavior of the variance estimators $\hat{\Gamma}_X(\underline{\mathbf{0}})$ and $\tilde{\Gamma}_X(\underline{\mathbf{0}})$. Since the behavior is similar for the different cases we only give a detailed analysis for Example 2, $d = 2$ and $a = 0.5$. First, we observe in Figure 3.1 that both $\hat{\Gamma}_X(\underline{\mathbf{0}})$ and $\tilde{\Gamma}_X(\underline{\mathbf{0}})$

Table 3.1: Sample sizes

	$d=1$	$d=2$	$d=3$
n_1	250	30	10
n_2	500	50	20
n_3	1000	70	30

	Example 1	Example 2	Example 3
$d=1$	$(0.2, 0.4]$ Vol= 0.2	$(0.3, 0.9]$ Vol= 0.6	$(0.1, 0.9]$ Vol= 0.8
$d=2$	$\left(\begin{pmatrix}0.2\\0.3\end{pmatrix}, \begin{pmatrix}0.6\\0.55\end{pmatrix}\right]$ Vol= 0.1	$\left(\begin{pmatrix}0.1\\0.1\end{pmatrix}, \begin{pmatrix}0.9\\0.85\end{pmatrix}\right]$ Vol= 0.6	$\left(\begin{pmatrix}0.05\\0.1\end{pmatrix}, \begin{pmatrix}0.95\\1.0\end{pmatrix}\right]$ Vol= 0.81
$d=3$	$\left(\begin{pmatrix}0.1\\0.2\\0.3\end{pmatrix}, \begin{pmatrix}0.7\\0.7\\0.7\end{pmatrix}\right]$ Vol= 0.12	$\left(\begin{pmatrix}0.1\\0.1\\0.0\end{pmatrix}, \begin{pmatrix}0.9\\0.85\\0.7\end{pmatrix}\right]$ Vol= 0.442	$\left(\begin{pmatrix}0.05\\0.05\\0.05\end{pmatrix}, \begin{pmatrix}0.95\\0.95\\0.95\end{pmatrix}\right]$ Vol= 0.729

Table 3.2: Values of $(\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0]$ and corresponding volumes for the different examples.

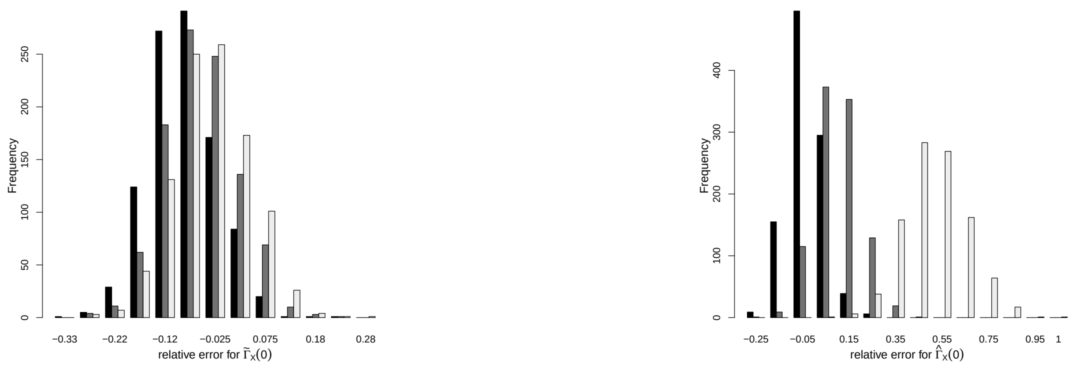
estimate the variance $(1-a^2)^{-d}$ of the random field $\{X_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ quite well for $\Delta = 0$. The classical estimator approximates the variance better than $\tilde{\Gamma}_X(\underline{\mathbf{0}})$ for $\Delta = 0$ but increases fast for growing Δ. By contrast, $\tilde{\Gamma}_X(\underline{\mathbf{0}})$ underestimates the variance for small n and Δ, but the estimation gains precision when either n or Δ increases.

Now, we investigate the behavior of the LRV estimators $\hat{\Sigma}_n$ and $\tilde{\Sigma}_n$. It is well known that the accuracy of kernel-type estimators often depends on the choice of bandwidth. In order to give an impression of the estimators' behavior independently of the chosen bandwidth, we considered bandwidths between 1 and 20 for $d = 1, 2$ and between 1 and 6 for $d = 3$ and plotted the minimal and maximal relative mean square error for each case. Since the estimates behave similarly in Examples 1 and 3, the figures presented here omit Example 1. Figures 3.2 and 3.3 show that under both choices of a the estimator $\tilde{\Sigma}_n$ stays stable for growing values of Δ. For both a, the worst approximation for $\tilde{\Sigma}_n$ is obtained with the bandwidth $q = 1$. Indeed, it is to be expected that estimating the LRV using an estimator of the variance — especially since, as seen in Figure 3.1, the variance already tends to be underestimated — leads to stronger underestimation and thus to greater errors. Since the relative difference between the LRV and the variance is $1-3^{-d}$ for $a = 0.5$ and $1-3^d$ for $a = -0.5$, the relative error when estimating the LRV using $q = 1$ is much greater for $a = -0.5$ than for $a = 0.5$ and both errors increase for higher dimensions d.

In comparison, the behavior of $\hat{\Sigma}_n$ depends on the choice of a and Δ. For $a = -0.5$, the relative mean square error strongly increases for larger Δ and q, as can be seen by the fact that the worst case in Figure 3.2 is often reached for the biggest possible bandwidth, while the optimal q decreases for growing Δ. An exception to this is for small Δ, where the worst error is attained when using small bandwidths. This is likely due to the tendency of the classical Bartlett estimator to overestimate the LRV, which is worsened by growing bandwidths. In general, the relative mean square errors are worse than the corresponding errors for $\tilde{\Sigma}_n$. For $a = 0.5$, the covariance part of the LRV is positive and Figure 3.3 shows that the overestimation of the variance by the classical Bartlett estimator (cf. Figure 3.1) can balance the underestimation of the LRV. Hence, even for some $\Delta > 0$, the estimator $\hat{\Sigma}_n$ is a little better than $\tilde{\Sigma}_n$ for decreasing q. However, a wrong choice of q has a considerable negative effect on the mean square error of $\hat{\Sigma}_n$, whereas the difference between the best and worst cases for $\tilde{\Sigma}_n$ stays small. The different examples illustrate that the choice of q has more influence on the error if the change-set C_n and C_n^c have similar volumes.

Values for $\Delta = 0, 1, 2$ in black, gray and light gray, respectively.

(a) $n_1 = 30$



(b) $n_2 = 50$

(c) $n_3 = 70$

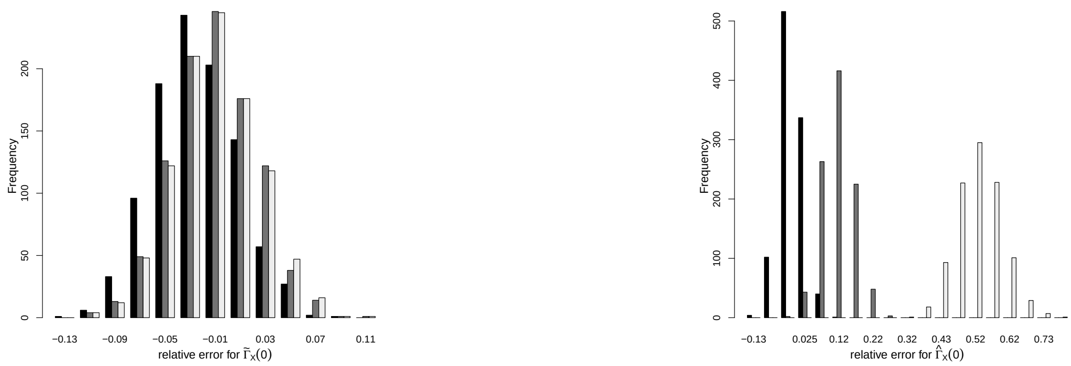


Figure 3.1: Multihistogram of the relative error $(\tilde{\Gamma}_X(\underline{\mathbf{0}}) - \Gamma(\underline{\mathbf{0}}))/\Gamma(\underline{\mathbf{0}})$ on the left and $(\hat{\Gamma}_X(\underline{\mathbf{0}}) - \Gamma(\underline{\mathbf{0}}))/\Gamma(\underline{\mathbf{0}})$ on the right. This shows Example 2 for $d = 2$ and $a = 0.5$.

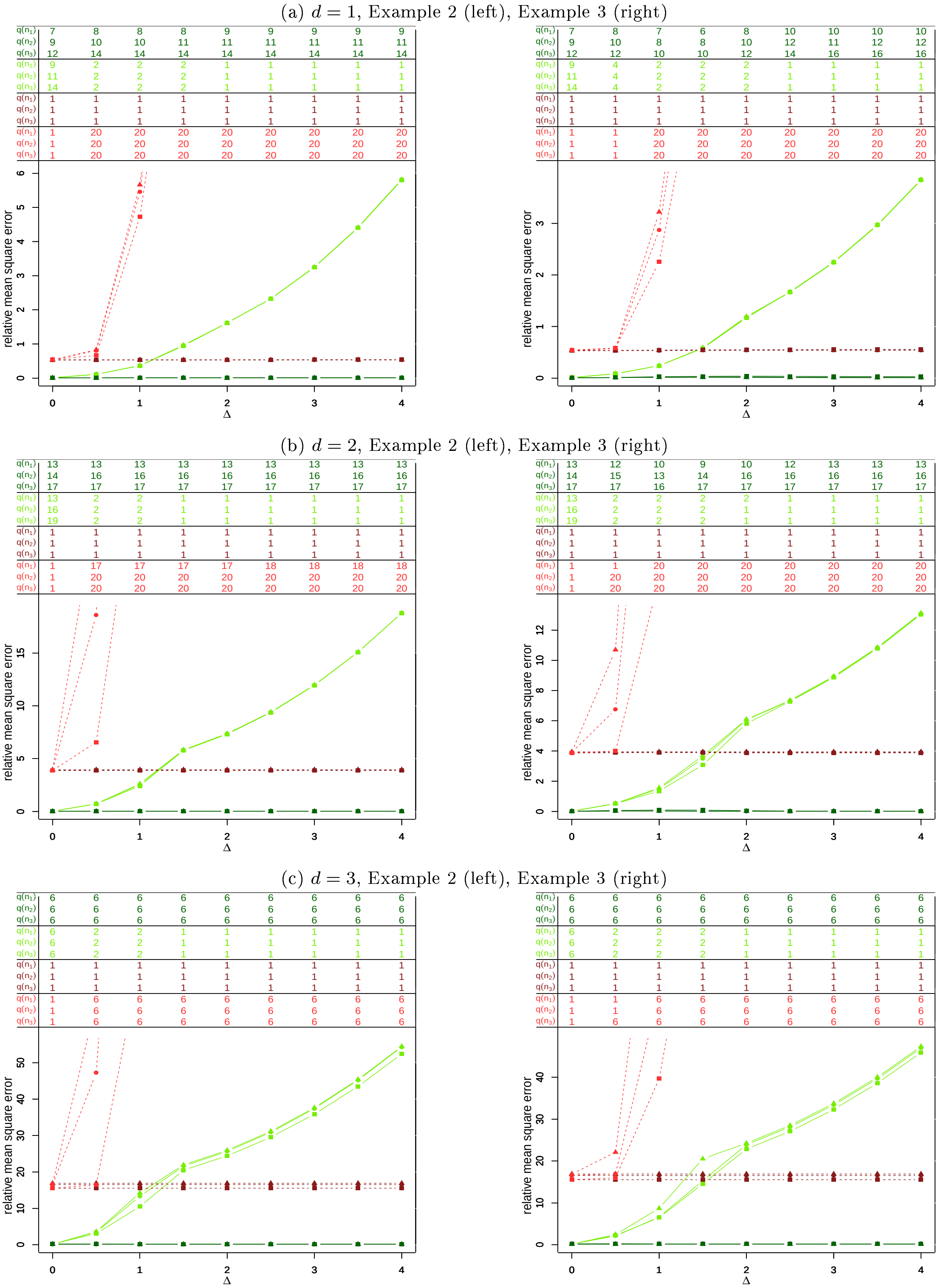


Figure 3.2: Best- and worst-case behavior of the LRV estimators for each Δ and n for $a = -0.5$ and the corresponding bandwidths q such that the relative mean square error is either minimal or maximal. Within each window, the values of q correspond to: the optimal case for $\tilde{\Sigma}_n$ (rows 1–3), the optimal case for $\hat{\Sigma}_n$ (rows 4–6), the worst case for $\tilde{\Sigma}_n$ (rows 7–9), the worst case for $\hat{\Sigma}_n$ (rows 10–12). The plots marked with $\square$ correspond to n_1, the $\circ$ to n_2 and $\triangle$ is for n_3. Solid lines correspond to the best and dashed lines to the worst case, the lighter shades are used for $\hat{\Sigma}_n$ and the darker shades for $\tilde{\Sigma}_n$.

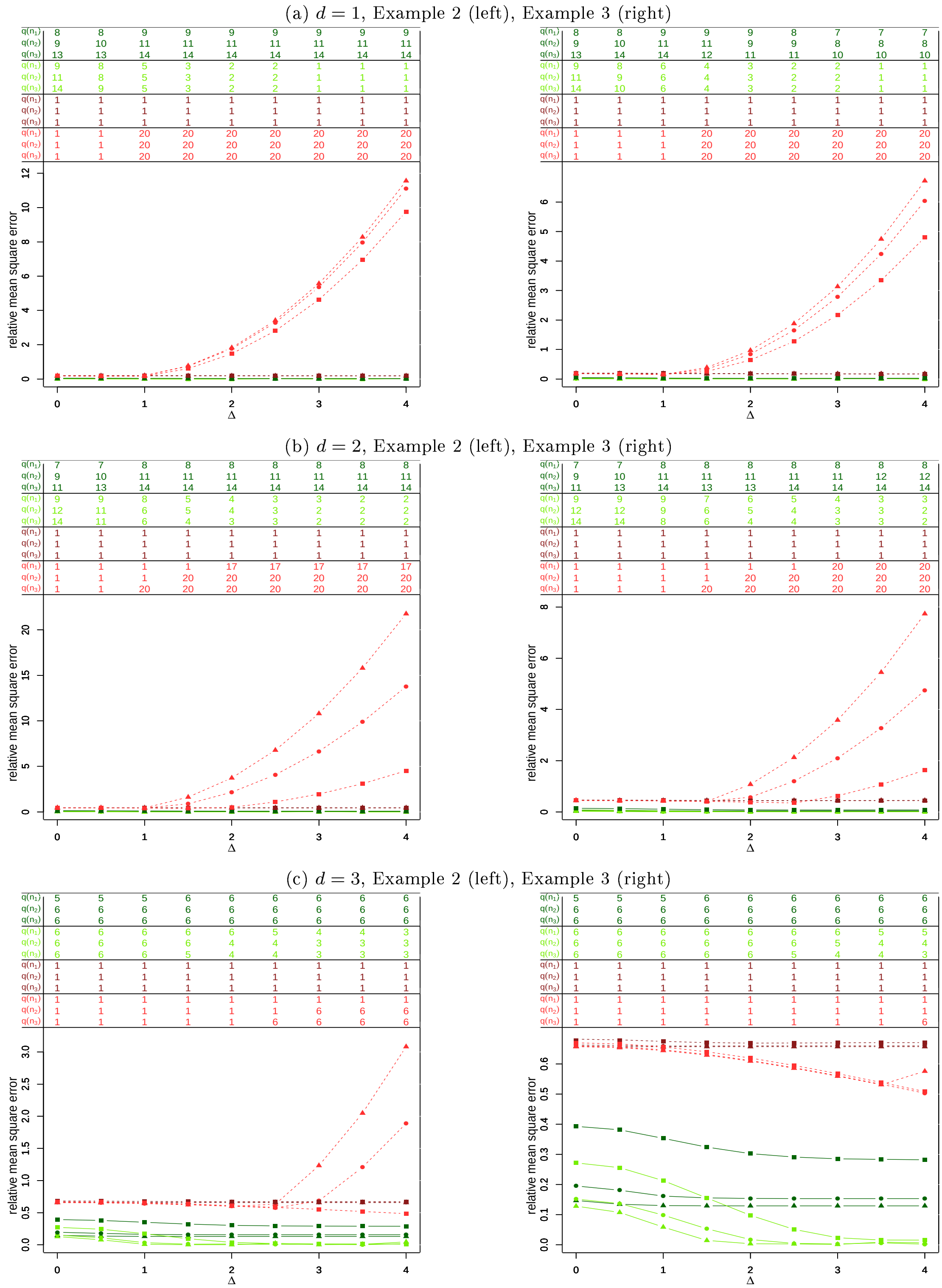


Figure 3.3: Best- and worst-case behavior of the LRV estimators for each Δ and n for $a = 0.5$ and the corresponding bandwidths q such that the relative mean square error is either minimal or maximal. Within each window, the values of q correspond to: the optimal case for $\tilde{\Sigma}_n$ (rows 1–3), the optimal case for $\hat{\Sigma}_n$ (rows 4–6), the worst case for $\tilde{\Sigma}_n$ (rows 7–9), the worst case for $\hat{\Sigma}_n$ (rows 10–12). The plots marked with □ correspond to n_1, the ∘ to n_2 and △ is for n_3. Solid lines correspond to the best and dashed lines to the worst case, the lighter shades are used for $\hat{\Sigma}_n$ and the darker shades for $\tilde{\Sigma}_n$.

5.3 Application to change-point tests

As an illustration of the effect of the LRV estimator on change-point detection, in this subsection, we study the behavior of the change-point test presented in Bucchia (2014) for a change in the mean over a rectangle C_n (where the null hypothesis is a constant unknown mean μ). Since this type of result is well-known in the one-dimensional case ($d = 1$), we focus on $d = 2, 3$. The tests are based on the two change-point test statistics $\hat{T}_n = \hat{\Sigma}_n^{-\frac{1}{2}} T_n$ and $\tilde{T}_n = \tilde{\Sigma}_n^{-\frac{1}{2}} T_n$, where

$$T_n := n^{-d/2} \max_{\substack{\mathbf{0} \leq \mathbf{k}_1 < \mathbf{k}_2 \leq \mathbf{n} \\ \lfloor 0.01 n^d \rfloor \leq [\mathbf{k}_2 - \mathbf{k}_1] \leq \lfloor 0.99 n^d \rfloor}} \frac{\left| \sum_{\mathbf{k}_1 < \mathbf{j} \leq \mathbf{k}_2} \left(X_{\mathbf{j}} - \overline{X}_n \right) \right|}{\sqrt{\frac{[\mathbf{k}_2 - \mathbf{k}_1]}{n^d} \left(1 - \frac{[\mathbf{k}_2 - \mathbf{k}_1]}{n^d} \right)}}.$$

The null hypothesis is rejected if the test statistic exceeds a critical value $c^\star(\alpha)$ for a given level α. We consider a 5% significance level and use the critical values

$$c^\star(0.05) = \begin{cases} 4.167, & d = 1 \\ 5.971, & d = 2 \\ 7.095, & d = 3 \end{cases}$$

obtained in Bucchia (2014). We denote the test based on $\hat{T}_n$ by $\hat{\Phi}$ and the test based on $\tilde{T}_n$ by $\tilde{\Phi}$. For both choices of a and the different examples, the empirical power of $\tilde{\Phi}$ is almost always higher than the empirical power of $\hat{\Phi}$ for a fixed bandwidth q. Unfortunately, the test using $\tilde{T}_n$ also often leads to a higher false rejection probability under the null hypothesis, and for some q, it seems that the given level α cannot be held (for the finite sample sizes considered here). For example, the probability of false rejection is highest for the largest q ($q = 20$) if $a = -0.5$, and for $q = 1$, if $a = 0.5$, with a rejection rate of over 60% in the latter case. To give an impression of the empirical power and size of the test without these extremes, Figure 3.4 shows results for the change-point setting 2 for $d = 2$ restricted to bandwidths which lead to a smaller empirical size than 0.2 for both tests. For each Δ and n, Figure 3.4 shows the best and worst possible empirical size and power of both tests and the corresponding q under the aforementioned restriction. We see that for $\Delta \geq 1$ and under the allowed bandwidths, the worst empirical power of $\tilde{\Phi}$ is only slightly worse than the best power of $\hat{\Phi}$. For $a = -0.5$, the best and the worst chosen q is the same for $\Delta \geq 1$. In summary, the simulations show that $\tilde{\Phi}$ has high empirical power under different q, whereas the empirical size of the test is sensitive to the choice of bandwidth. By contrast, the empirical size of $\hat{\Phi}$ is a little more stable under different bandwidths, but the empirical power of $\hat{\Phi}$ is more sensitive. For $a = -0.5$, the best empirical power is reached for $q = 1$, where the LRV is estimated by the variance. However, as mentioned before, most of the corresponding tests would not hold the significance level.

The problem of selecting an optimal bandwidth is beyond the scope of this paper, but in order to illustrate the behavior of the tests if we avoid the problem of underestimating the LRV under the null hypothesis for values of q which are too small or too large, we also consider a data-dependent bandwidth choice $\mathbf{q}_n$, which uses all the covariances above a certain (fixed) size. Since the decrease in the correlations may depend on the direction,

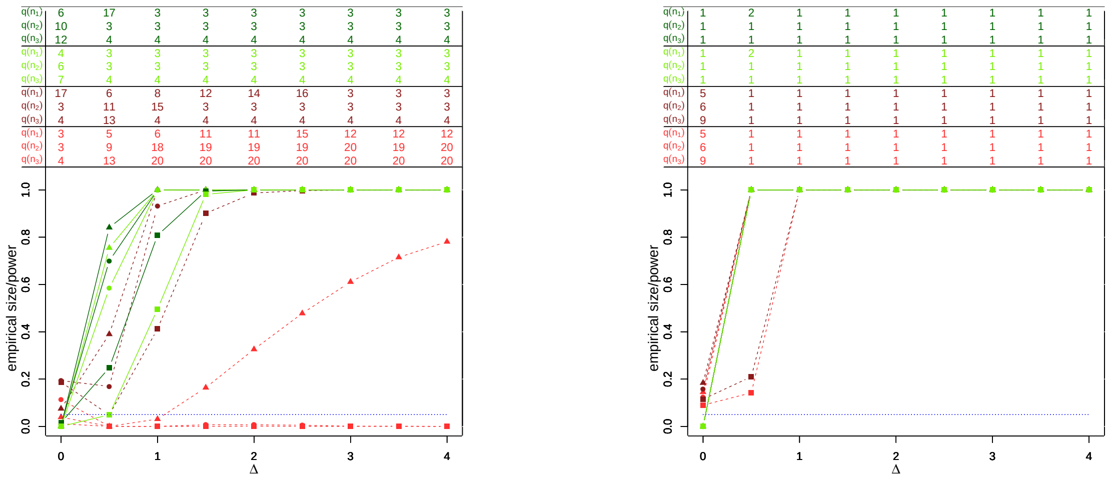


Figure 3.4: Best and worst empirical size and power of $\hat{\Phi}$ and $\tilde{\Phi}$ and the corresponding bandwidths q, for $a = 0.5$ on the left and $a = -0.5$ on the right side, both for Example 2 and $d = 2$. Within each window, the values of q correspond to: the optimal case for $\tilde{\Phi}$ (rows 1–3), the optimal case for $\hat{\Phi}$ (rows 4–6), the worst case for $\tilde{\Phi}$ (rows 7–9), the worst case for $\hat{\Phi}$ (rows 10–12). The plots marked with □ correspond to n_1, the ∘ to n_2 and △ is for n_3. Solid lines correspond to the best and dashed lines to the worst case, the lighter shades are used for $\hat{\Phi}$ and the darker shades for $\tilde{\Phi}$. The dotted line shows the theoretical significance level 0.05.

we now use a vector of bandwidths $q_{n,i}$ $(i = 1, \ldots, d)$ (cf. Remark 3.1) such that

$$q_{n,i} = \inf\{q \in \mathbb{N} : |\bar{\Gamma}_X(\mathbf{v}_q)/\bar{\Gamma}_X(\underline{\mathbf{0}})| < \delta\},$$

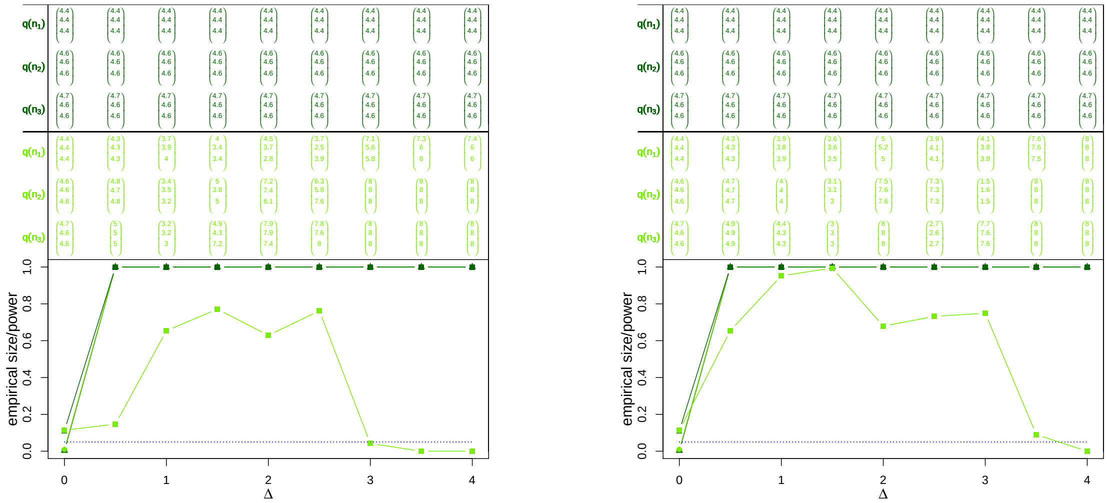
where $\delta = 0.05$, $\bar{\Gamma}_X$ is either the covariance estimator $\hat{\Gamma}_X$ or $\tilde{\Gamma}_X$ and $\mathbf{v}_q$ are vectors with i-th entry q and null otherwise. Figures 3.5a and 3.5b show the behavior of the above procedure and the average of the calculated bandwidths of each of the 1000 simulation runs. We again leave out Example 1 and, since the results were qualitatively the same, we also omit the cases $d = 1, 2$. For $a = -0.5$, the empirical size of $\tilde{\Phi}$ is a little over the given level of 0.05 for n_1, but for growing sample size n the test becomes more conservative, whereas for $a = 0.5$ the empirical size increases for growing sample size, leading to a rejection rate that is above the given level for $d = 1$. The power of $\tilde{\Phi}$ is quite good and monotonic in the magnitude of change Δ and the sample size n. By contrast, the empirical size and power of $\hat{\Phi}$ is always lower than the empirical size of $\tilde{\Phi}$, leading to a better adherence to the significance level but also nonmonotonic power for growing Δ. For $a = 0.5$ and n_1 the test $\hat{\Phi}$ nearly never detects the changes. This is also due to the bandwidth selection heuristic: For $\hat{\Phi}$, growing magnitudes of the mean change lead to higher estimates of the covariances and thus to the selection of bigger q, which worsens the nonmonotonic power problem. In contrast, the calculated bandwidths for $\tilde{\Phi}$ stay stable for different values of Δ.

5.4 An application to brain tumor detection

In this subsection, we apply the statistics from Section 5.3 to an MRT image of a brain with a possible tumor.[1] As can be seen in Figure 3.6, there is an obvious inhomogeneity in the upper left corner of the picture and therefore the aim of this section is to test if our

[1] The picture we used is an excerpt from a picture which was published on a website by the Neuroonkologische Arbeitsgemeinschaft (NOA) (2012).

(a) Results for $a = -0.5$



(b) Results for $a = 0.5$

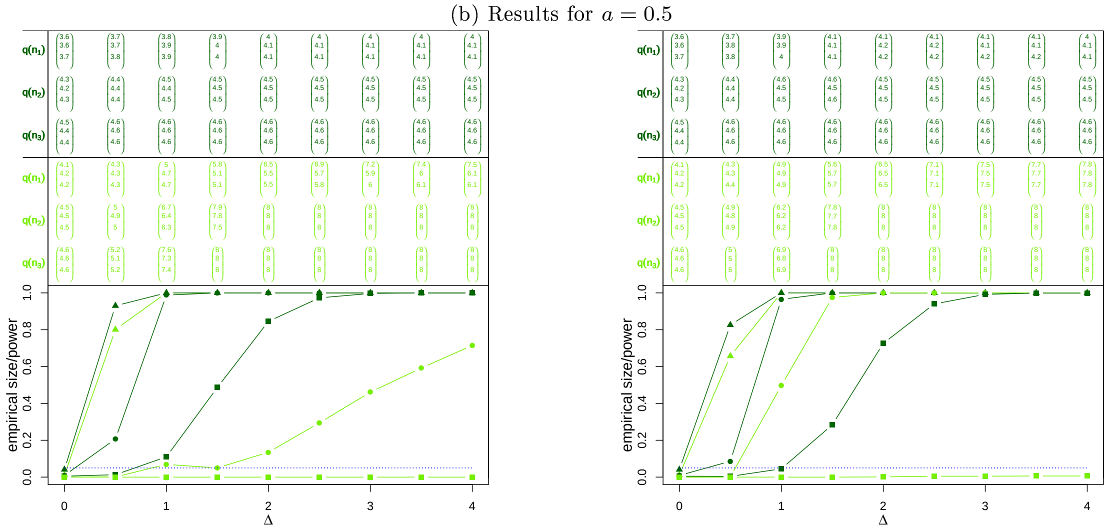


Figure 3.5: Empirical size and power of the tests ($d = 3$, Example 2 (left), Example 3 (right)) for each Δ and n and the data-driven bandwidth choice q (in each case three lines corresponding to the different n_i in ascending order). The plots marked with $\square$ correspond to n_1, the $\circ$ to n_2 and $\triangle$ is for n_3. The light shade is used for $\hat{\Phi}$ and the dark color for $\tilde{\Phi}$, the dotted line shows the theoretical significance level.

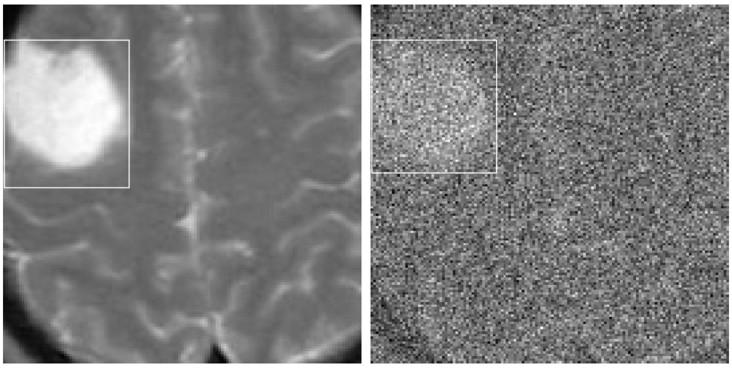


Figure 3.6: Original data (left) and an example of data with added noise (right). The estimated change-set is marked by a white rectangle.

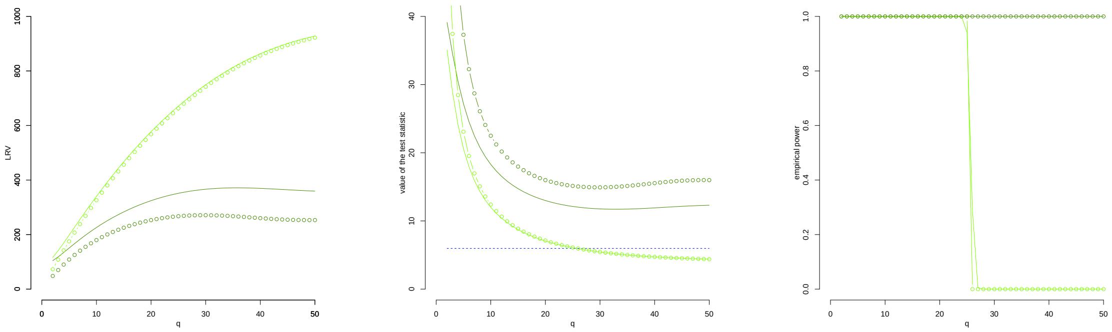


Figure 3.7: LRV-estimation and test results (for a 5% significance level) for 1000 realizations of the noisy data. From left to right, the figures show the average of the estimated values $\hat{\Sigma}_n$ and $\tilde{\Sigma}_n$ respectively, the calculated average values of the test statistics $\hat{T}_n$ and $\tilde{T}_n$, and the empirical power of the tests. Lighter shades are used for values corresponding to $\hat{\Sigma}_n$, values corresponding to $\tilde{\Sigma}_n$ are in darker tones. The plots marked with $\circ$ correspond to the results for the original data set, the solid lines correspond to the data with added noise. Additionally, the middle figure contains a dotted horizontal line corresponding to the critical value $c = 5.971$ for the asymptotic tests.

statistics confirm this observation. In order to further test the robustness of our procedure with respect to random errors in the data (as might result from imprecise measuring), we apply the statistics not only to the original picture but also to a version with added noise. To be precise, we consider observations $x_{\mathbf{k}}$ ($\mathbf{k} \in \{1, \ldots, 135\} \times \{1, \ldots, 146\}$) corresponding to an image and add noise $\{Y_{\mathbf{k}} + \tilde{\varepsilon}_{\mathbf{k}}\}$, where $\{Y_{\mathbf{k}}\}$ is an MA random field as defined in (6) with $a = 0.9$, and $\{\tilde{\varepsilon}_{\mathbf{k}}\}$ is an independent random field consisting of independent $N(0, 90)$-distributed random variables. We use the change-point estimator from Section 4 and the tests Φ and $\tilde{\Phi}$ described in Section 5.3, each with bandwidths q ranging from 2 to 50. The statistics are applied to the original data $\{x_{\mathbf{k}}\}$ and to 1000 realizations of $\{x_{\mathbf{k}} + Y_{\mathbf{k}} + \tilde{\varepsilon}_{\mathbf{k}}\}$, respectively. As can be seen in Figure 3.6, although the observations do not perfectly match the model (there are potentially several different segments and the change-set is not a rectangle), the statistics nevertheless produce an acceptable rectangular estimate of the change-set. Figure 3.7 confirms the empirical results from Section 5.3 for this real data example: Although both tests reject the homogeneity hypothesis, the LRV-estimator $\tilde{\Sigma}_n$ leads to smaller average values than the estimator $\hat{\Sigma}_n$, and therefore the corresponding test statistic has bigger values, resulting in better empirical power.

5.5 Conclusion

Our simulations confirm the theoretical qualities of $\tilde{\Sigma}_n$. The estimator is generally much more stable under alternatives than the corresponding classical Bartlett estimator, leading to change-point tests with monotonic power function. However, while the heuristic bandwidth choice considered here yields acceptable results for sufficiently large samples, the problem of finding an optimal bandwidth, which would guarantee adherence to the significance level, is still open.

Appendix: Some technical lemmas

Lemma A1. *For $\underline{\mathbf{0}} \le \mathbf{k}_1 \le \mathbf{k}_2 \le \underline{\mathbf{n}}$ and $\underline{\mathbf{0}} \le \mathbf{k}_1^0 < \mathbf{k}_2^0 \le \underline{\mathbf{n}}$ with $k_2^{0(i)} - k_1^{0(i)} \ge cn$ for all $i \in \{1, \dots, d\}$ and some $c > 0$, it holds that*

$$\lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0}) + \lambda(R_{\mathbf{k}^0} \setminus R_{\mathbf{k}}) \ge C n^{d-1} \|\mathbf{k} - \mathbf{k}^0\|$$

for some $C > 0$ which may depend on d but not on n.

Proof. W.l.o.g. we assume $\|\mathbf{k} - \mathbf{k}^0\| > 0$. Note that

$$\begin{aligned} &\lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0}) + \lambda(R_{\mathbf{k}^0} \setminus R_{\mathbf{k}}) \\ &= \lambda(R_{\mathbf{k}}) + \lambda(R_{\mathbf{k}^0}) - 2\lambda(R_{\mathbf{k}^0} \cap R_{\mathbf{k}}). \end{aligned}$$

Since $\lambda(R_{\mathbf{k}^0}) \ge cn^d \ge cn^{d-1}\|\mathbf{k} - \mathbf{k}^0\|$, we can assume w.l.o.g. that $R_{\mathbf{k}} \cap R_{\mathbf{k}^0} \ne \emptyset$. We prove the lemma by induction. For $d = 1$ it holds that

$$\begin{aligned} &\lambda(k_1, k_2] + \lambda(k_1^0, k_2^0] - 2\lambda((k_1, k_2] \cap (k_1^0, k_2^0]) \\ &= k_2 - k_2 \wedge k_2^0 + k_2^0 - k_2 \wedge k_2^0 - (k_1 - k_1 \vee k_1^0) - (k_1^0 - k_1 \vee k_1^0) \\ &= |k_2 - k_2^0| + |k_1 - k_1^0| \ge \|\mathbf{k} - \mathbf{k}^0\|. \end{aligned}$$

Assuming the assertion holds for d, we consider the case $d+1$. For any vector $\mathbf{x} \in \mathbb{Z}^{d+1}$, we denote by $\mathbf{x}'$ the vector $(x_1, \dots, x_d)$. W.l.o.g. we assume $\|\mathbf{k}' - \mathbf{k}^{0\prime}\| = \|\mathbf{k} - \mathbf{k}^0\|$. Writing $A = (k_2^{0(d+1)} - k_1^{0(d+1)})^{-1}(k_2^{(d+1)} - k_1^{(d+1)})$ and noting that

$$\frac{\left(k_2^{0(d+1)} \wedge k_2^{(d+1)} - k_1^{0(d+1)} \vee k_1^{(d+1)}\right)_+}{k_2^{0(d+1)} - k_1^{0(d+1)}} \le A \wedge 1,$$

we obtain

$$\begin{aligned} &\lambda(R_{\mathbf{k}}) + \lambda(R_{\mathbf{k}^0}) - 2\lambda(R_{\mathbf{k}^0} \cap R_{\mathbf{k}}) \\ &\ge \left(k_2^{0(d+1)} - k_1^{0(d+1)}\right) \left(A\lambda(\mathbf{k}_1', \mathbf{k}_2'] + \lambda(\mathbf{k}_1^{0\prime}, \mathbf{k}_2^{0\prime}] - 2(A \wedge 1)\lambda((\mathbf{k}_1^{0\prime}, \mathbf{k}_2^{0\prime}] \cap (\mathbf{k}_1', \mathbf{k}_2'])\right) \\ &\ge \left(k_2^{0(d+1)} - k_1^{0(d+1)}\right) \left((A \wedge 1) \left(\lambda(\mathbf{k}_1', \mathbf{k}_2'] + \lambda(\mathbf{k}_1^{0\prime}, \mathbf{k}_2^{0\prime}] - 2\lambda((\mathbf{k}_1^{0\prime}, \mathbf{k}_2^{0\prime}] \cap (\mathbf{k}_1', \mathbf{k}_2'])\right)\right) \\ &\quad + (1 - (A \wedge 1))\lambda(R_{\mathbf{k}^0}) \\ &\overset{Ind.hyp.}{\ge} \underbrace{(k_2^{0(d+1)} - k_1^{0(d+1)})}_{\ge cn} (A \wedge 1)\, C n^{d-1} \|\mathbf{k} - \mathbf{k}^0\| + (1 - (A \wedge 1)) \underbrace{\lambda(R_{\mathbf{k}^0})}_{\ge cn^{d+1}} \\ &\ge ((cC) \wedge c) \cdot n^d \left(n(1 - (A \wedge 1)) + (A \wedge 1)\|\mathbf{k} - \mathbf{k}^0\|\right) \\ &= ((cC) \wedge c) \cdot n^d \left(\|\mathbf{k} - \mathbf{k}^0\| + (n - \|\mathbf{k} - \mathbf{k}^0\|)(1 - (A \wedge 1))\right) \\ &\ge ((cC) \wedge c) \cdot n^d \|\mathbf{k} - \mathbf{k}^0\|. \end{aligned}$$

□

Lemma A2. *Under the assumptions of Theorem 4.1 and using the notations from its proof, it holds that*

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2} A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{L_{n,\mathbf{k}_1,\mathbf{k}_2}} \right| = \mathcal{O}_P(n^{-d/2})$$

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}} \Delta \right| = \mathcal{O}_P(n^{-d/2}) + \alpha_n(N)^{-1}\mathcal{O}_P(1)$$

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \Delta^T \frac{A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}} \right| = \mathcal{O}_P(n^{-d/2}).$$

Furthermore, there is a constant C such that

$$L_{n,\mathbf{k}_1,\mathbf{k}_2} \leq -C < 0$$

for all $\mathbf{k}_1 < \mathbf{k}_2$, $\|\mathbf{k}-\mathbf{k}^0\| \geq N$.

Proof. We use c or C to denote positive constants which are independent of n or N and whose values may change from line to line. First, we derive an upper bound for $B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}$:

$$\begin{aligned}\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \leq & \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left(\lambda\left(\left(\frac{\mathbf{k}_1}{n}, \frac{\mathbf{k}_2}{n}\right] \cap \left(\frac{\mathbf{k}_1^0}{n}, \frac{\mathbf{k}_2^0}{n}\right]\right) + \lambda\left(\frac{\mathbf{k}_1^0}{n}, \frac{\mathbf{k}_2^0}{n}\right]\right) \\ & + \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left\{\lambda\left(\frac{\mathbf{k}_1}{n}, \frac{\mathbf{k}_2}{n}\right] + \lambda\left(\frac{\mathbf{k}_1^0}{n}, \frac{\mathbf{k}_2^0}{n}\right]\right\} \lambda\left(\frac{\mathbf{k}_1^0}{n}, \frac{\mathbf{k}_2^0}{n}\right] \\ \leq & 4.\end{aligned}$$

Now we show that $\frac{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \geq c > 0$ for all $\mathbf{k}_1 < \mathbf{k}_2$, $\|\mathbf{k}-\mathbf{k}^0\| \geq N$.

$$\frac{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} = \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\left(1 - \frac{\lambda(R_\mathbf{k} \cup R_{\mathbf{k}^0})}{n^d}\right) + \frac{\lambda(R_\mathbf{k} \cap R_{\mathbf{k}^0})}{n^d}\left(1 - \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right)$$

1. Case: $\forall i = 1, \dots, d : (k_1^{0(i)}, k_2^{0(i)}] \subset (k_1^{(i)}, k_2^{(i)}]$
We obtain

$$\frac{\lambda(R_\mathbf{k} \cap R_{\mathbf{k}^0})}{n^d}\left(1 - \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right) = \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\left(1 - \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right) \geq c$$

and therefore $\frac{\lambda(R_\mathbf{k} \cup R_{\mathbf{k}^0})}{n^d} \leq 1$ implies $\frac{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \geq c > 0$.
2. Case: $\exists i \in \{1, \dots, d\} : (k_1^{0(i)}, k_2^{0(i)}] \not\subset (k_1^{(i)}, k_2^{(i)}]$
Then either $(0, k_1^{0(i)}] \cap (k_1^{(i)}, k_2^{(i)}] = \emptyset$ or $(k_2^{0(i)}, 1] \cap (k_1^{(i)}, k_2^{(i)}] = \emptyset$ and it follows that either $(k_1^{(i)}, k_2^{(i)}] \cup (k_1^{0(i)}, k_2^{0(i)}] \subset (0, k_2^{0(i)}]$ or $(k_1^{(i)}, k_2^{(i)}] \cup (k_1^{0(i)}, k_2^{0(i)}] \subset (k_1^{0(i)}, n]$. Since $\varepsilon_1 n < k_j^{0(i)} < \varepsilon_2 n$ for some $0 < \varepsilon_1, \varepsilon_2 < 1$ and $j = 1, 2$, there exists an $\epsilon < 1$ such that

$$\lambda(R_\mathbf{k} \cup R_{\mathbf{k}^0}) \leq \lambda((k_1^{(i)}, k_2^{(i)}] \cup (k_1^{0(i)}, k_2^{0(i)}]) \prod_{j\neq i} \lambda((k_1^{(j)}, k_2^{(j)}] \cup (k_1^{0(j)}, k_2^{0(j)}]) \leq \epsilon n n^{d-1}$$

and therefore

$$\frac{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \geq \frac{\lambda(R_{\mathbf{k}^0})}{n^d}(1-\epsilon) + \overbrace{\frac{\lambda(R_{\mathbf{k}} \cap R_{\mathbf{k}^0})}{n^d}\left(1 - \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right)}^{\geq 0} \geq c.$$

Consider now $B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}$. Similarly to $B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}$, we want to give lower and upper bounds. To do that, we start with a few preliminary considerations. In order to deal with more general cases, we define for any $\mathbf{k}_1 < \mathbf{k}_2$ corresponding vectors $\mathbf{a} \in \mathbb{Z}^d$ with $a^{(i)} \in \{k_1^{(i)}, k_2^{(i)}\}$ for $i = 1, \ldots, d$, as well as

$$\begin{aligned}
M_{\mathbf{a}} &:= (\underline{0}, \mathbf{a}] \setminus (\underline{0}, \mathbf{a}^0] \\
M^0_{\mathbf{a}} &:= (\underline{0}, \mathbf{a}^0] \setminus (\underline{0}, \mathbf{a}] \\
f(\mathbf{a}) &:= \max\{\lambda(M_{\mathbf{a}}), \lambda(M^0_{\mathbf{a}})\} \text{ and} \\
T_{\mathbf{k}_1,\mathbf{k}_2} &:= \max\{f(\mathbf{a}) : a^{(i)} \in \{k_1^{(i)}, k_2^{(i)}\},\ i = 1, \ldots, d\}.
\end{aligned}$$

Now, we show for $\|\mathbf{a} - \mathbf{a}^0\| \geq 1$ that $f(\mathbf{a}) \sim n^{d-1}\|\mathbf{a} - \mathbf{a}^0\|$ and therefore $T_{\mathbf{k}_1,\mathbf{k}_2} \sim n^{d-1}\|\mathbf{k} - \mathbf{k}^0\|$ for $\|\mathbf{k} - \mathbf{k}^0\| \geq N$. Here, the notation $x_n \sim y_n$ means that there are constants $0 < c, C$ such that $cy_n < x_n < y_n C$ for all $n \in \mathbb{N}$. We start with a few preliminary observations:
1. Note that $f(\mathbf{a})$ has the form

$$f(\mathbf{a}) = \max\left\{\prod_{i=1}^d a^{(i)}, \prod_{i=1}^d a^{0(i)}\right\} - \prod_{i=1}^d a^{(i)} \wedge a^{0(i)}.$$

First, we show by induction that there exists a $c > 0$ such that $f(\mathbf{a}) \leq cn^{d-1}\|\mathbf{a} - \mathbf{a}^0\|$.
$d = 1$: $f(\mathbf{a}) = a^{(1)} \vee a^{0(1)} - a^{(1)} \wedge a^{0(1)} = |a^{(1)} - a^{0(1)}| = \|\mathbf{a} - \mathbf{a}^0\| \leq cn^{d-1}\|\mathbf{a} - \mathbf{a}^0\|$.
$d - 1 \to d$: Choose $j \in \{1, \ldots, d\}$ with $|a^{(j)} - a^{0(j)}| = \|\mathbf{a} - \mathbf{a}^0\|$. Then it holds that

$$\begin{aligned}
f(\mathbf{a}) &= \max\left\{\prod_{i=1}^d a^{(i)}, \prod_{i=1}^d a^{0(i)}\right\} - \prod_{i=1}^d a^{(i)} \wedge a^{0(i)} \\
&\leq \left(\max\left\{\prod_{i\neq j} a^{(i)}, \prod_{i\neq j} a^{0(i)}\right\} - \prod_{i\neq j} a^{(i)} \wedge a^{0(i)}\right) \cdot \overbrace{a^{(j)} \vee a^{0(j)}}^{\leq n} \\
&+ \prod_{i\neq j} a^{(i)} \wedge a^{0(i)} \underbrace{\left(a^{(j)} \vee a^{0(j)} - a^{(j)} \wedge a^{0(j)}\right)}_{=\|\mathbf{a}-\mathbf{a}^0\|} \\
&\overset{Ind.hyp.}{\leq} cn^{d-1}\|\mathbf{a} - \mathbf{a}^0\|.
\end{aligned}$$

2. Since $f(\mathbf{a}) \geq 1/2\left(\lambda(M_{\mathbf{a}}) + \lambda(M^0_{\mathbf{a}})\right)$, Lemma A1 implies that there also exists a $c > 0$ such that $f(\mathbf{a}) \geq cn^{d-1}\|\mathbf{a} - \mathbf{a}^0\|$.
3. It holds that $\max\{\lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0}), \lambda(R_{\mathbf{k}^0} \setminus R_{\mathbf{k}})\} \leq cn^{d-1}\|\mathbf{k} - \mathbf{k}^0\|$:
$d = 1$: It holds that

$$\begin{aligned}
\lambda((k_1, k_2] \setminus (k_1^0, k_2^0]) &= k_2 - k_1 - (k_2 \wedge k_2^0 - k_1 \vee k_1^0)_+ \\
&\leq 2\|\mathbf{k} - \mathbf{k}^0\|.
\end{aligned}$$

and $\lambda((k_1^0, k_2^0] \setminus (k_1, k_2]) \leq 2\|\mathbf{k} - \mathbf{k}^0\|$ analogously.
$d-1 \to d$: Let $\mathbf{v}' = (v_1, \dots, v_{d-1})^T$ for a d-dimensional vector $\mathbf{v} = (v_1, \dots, v_d)^T$. Since $\lambda(\mathbf{k}_1', \mathbf{k}_2'] \leq n^{d-1}$ and $\lambda(\mathbf{k}_1^{(d)}, \mathbf{k}_2^{(d)}] \leq n$, we obtain

$$\begin{aligned}\lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0}) &\leq \lambda((\mathbf{k}_1', \mathbf{k}_2'] \setminus (\mathbf{k}_1^{0\prime}, \mathbf{k}_2^{0\prime}])\lambda(k_1^{(d)}, k_2^{(d)}] \\ &\quad + \lambda((k_1^{(d)}, k_2^{(d)}] \setminus (k_1^{0(d)}, k_2^{0(d)}])\lambda(\mathbf{k}_1', \mathbf{k}_2'] \\ &\overset{Ind.hyp.}{\leq} cn^{d-1}\|\mathbf{k} - \mathbf{k}^0\|\end{aligned}$$

and $\lambda((\mathbf{k}_1^0, \mathbf{k}_2^0] \setminus (\mathbf{k}_1, \mathbf{k}_2]) \leq cn^{d-1}\|\mathbf{k} - \mathbf{k}^0\|$ analogously. Combining our observations yields:

$$\begin{aligned}&\max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{-B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}}{T_{\mathbf{k}_1,\mathbf{k}_2}} \\ &= \max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{\lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0})\frac{\lambda(R_{\mathbf{k}^0})}{n^d} + \lambda(R_{\mathbf{k}^0} \setminus R_{\mathbf{k}})\left(1 - \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right)}{T_{\mathbf{k}_1,\mathbf{k}_2}} \\ &\overset{2.+3.}{\leq} 2 \max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{c_1 n^{d-1}\|\mathbf{k} - \mathbf{k}^0\|}{c_2 n^{d-1}\|\mathbf{k} - \mathbf{k}^0\|} \leq c\end{aligned}$$

for some $c_1, c_2 > 0$. For any $\mathbf{k}_1 < \mathbf{k}_2$, $\|\mathbf{k} - \mathbf{k}^0\| \geq N$, observation 1 and Lemma A1 imply

$$\frac{-B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}}{T_{\mathbf{k}_1,\mathbf{k}_2}} \geq c.$$

Finally, we obtain for $\mathbf{k}_1 < \mathbf{k}_2$ and $\|\mathbf{k} - \mathbf{k}^0\| \geq N$

$$\frac{|L_{n,\mathbf{k}_1,\mathbf{k}_2}|}{n^d T_{\mathbf{k}_1,\mathbf{k}_2}} \geq c > 0$$

and

$$\begin{aligned}\max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} L_{n,\mathbf{k}_1,\mathbf{k}_2} &= -\min_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} -B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2} B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2} \\ &\leq -\min_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} Cn^{d-1}\|\mathbf{k} - \mathbf{k}^0\| n^d \leq -c < 0.\end{aligned}$$

Now, we consider the $A^{(i)}$–terms:
First, we observe with $\max_{\mathbf{k}_1 < \mathbf{k}_2} |\sum_{\mathbf{i} \in R_{\mathbf{k}}} Y_{\mathbf{i}}^{(l)}| = \mathcal{O}_P(n^{d/2})$ that for all $l \in \{1, \dots, p\}$

$$\begin{aligned}\max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{|A^{(l)(2)}_{\mathbf{k}_1,\mathbf{k}_2}|}{n^d} &\leq \max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{1}{n^d}\left|\sum_{\mathbf{i} \in R_{\mathbf{k}}} Y_{\mathbf{i}}^{(l)}\right| + \max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \frac{1}{n^d}\left|\sum_{\mathbf{i} \in R_{\mathbf{k}^0}} Y_{\mathbf{i}}^{(l)}\right| \\ &\quad + \max_{\substack{\mathbf{k}_1 < \mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\| \geq N}} \left|\frac{\lambda(R_{\mathbf{k}})}{n^d} + \frac{\lambda(R_{\mathbf{k}^0})}{n^d}\right| \frac{1}{n^d}\left|\sum_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}}} Y_{\mathbf{i}}^{(l)}\right| = \mathcal{O}_P(n^{-d/2}).\end{aligned}$$

Now, we show

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{|A^{(l)(1)}_{\mathbf{k}_1,\mathbf{k}_2}|}{T_{\mathbf{k}_1,\mathbf{k}_2}} = \mathcal{O}_P(n^{-d/2}) + \alpha_n^{-1}(N)\mathcal{O}_P(1)$$

with $\alpha_n(N) = \min\{N^{1/2-1/r}, \left(\sum_{i=N}^{\infty}\frac{1}{i^{r/2}}\right)^{-1/r}\}I_{\{d=1\}} + n^{(d-1)(1/2-1/r)}I_{\{d>1\}}$. Using **2.** and **3.**, we obtain

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{\lambda(R_{\mathbf{k}}\backslash R_{\mathbf{k}^0}) - \lambda(R_{\mathbf{k}^0}\backslash R_{\mathbf{k}})}{T_{\mathbf{k}_1,\mathbf{k}_2}} = \mathcal{O}(1)$$

and therefore

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{|A^{(l)(1)}_{\mathbf{k}_1,\mathbf{k}_2}|}{T_{\mathbf{k}_1,\mathbf{k}_2}} \leq \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{\left|\sum_{\mathbf{i}\in R_{\mathbf{k}}} Y^{(l)}_{\mathbf{i}} - \sum_{\mathbf{i}\in R_{\mathbf{k}^0}} Y^{(l)}_{\mathbf{i}}\right|}{T_{\mathbf{k}_1,\mathbf{k}_2}} + \mathcal{O}_P(n^{-d/2})$$

and

$$\begin{aligned}
&\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{\left|\sum_{\mathbf{i}\in R_{\mathbf{k}}} Y^{(l)}_{\mathbf{i}} - \sum_{\mathbf{i}\in R_{\mathbf{k}^0}} Y^{(l)}_{\mathbf{i}}\right|}{T_{\mathbf{k}_1,\mathbf{k}_2}} \\
&= \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \frac{\left|\sum_{\boldsymbol{\epsilon}\in\{0,1\}^d}(-1)^{d-\sum_{i=1}^d \epsilon_i}\left(\sum\limits_{\mathbf{i}\leq \mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)} Y^{(l)}_{\mathbf{i}} - \sum\limits_{\mathbf{i}\leq \mathbf{k}^0_1+\boldsymbol{\epsilon}(\mathbf{k}^0_2-\mathbf{k}^0_1)} Y^{(l)}_{\mathbf{i}}\right)\right|}{T_{\mathbf{k}_1,\mathbf{k}_2}} \\
&\leq \sum_{\boldsymbol{\epsilon}\in\{0,1\}^d} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{1}{T_{\mathbf{k}_1,\mathbf{k}_2}} \sum_{\mathbf{i}\in(\underline{\mathbf{0}},\mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)]\backslash(\underline{\mathbf{0}},\mathbf{k}^0_1+\boldsymbol{\epsilon}(\mathbf{k}^0_2-\mathbf{k}^0_1)]} Y^{(l)}_{\mathbf{i}}\right| \\
&+ \sum_{\boldsymbol{\epsilon}\in\{0,1\}^d} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{1}{T_{\mathbf{k}_1,\mathbf{k}_2}} \sum_{\mathbf{i}\in(\underline{\mathbf{0}},\mathbf{k}^0_1+\boldsymbol{\epsilon}(\mathbf{k}^0_2-\mathbf{k}^0_1)]\backslash(\underline{\mathbf{0}},\mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)]} Y^{(l)}_{\mathbf{i}}\right| \\
&= T_1 + T_2.
\end{aligned}$$

Since both terms can be treated very similarly, we only show the estimation for T_1. We use the notation $M_{\mathbf{a}}$ defined above for the set over which the summation takes place, where $a^{(i)} \in \{k_1^{(i)}, k_2^{(i)}\}$ for all $i = 1, \ldots, d$. Per assumption, we have

$$\mathbb{E}\left(\left|\sum_{\mathbf{i}\in M_{\mathbf{a}}} Y^{(l)}_{\mathbf{i}}\right|^r\right) \leq \tilde{c}\lambda(M_{\mathbf{a}})^{r/2}.$$

For $d = 1$, $T_{k_1,k_2} = \|\mathbf{k}-\mathbf{k}^0\|$ and $\alpha_n = \alpha$ is independent of n. Therefore, the Markov inequality implies for $a \in \{k_1, k_2\}$ that

$$\begin{aligned}
&\mathbb{P}\left(\max_{\substack{k_1<k_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{1}{T_{k_1,k_2}} \sum_{\mathbf{i}\in(0,a]\setminus(0,a^0]} Y_i^{(l)} \right| \geq C\alpha^{-1}(N) \right) \\
&\leq \left(\frac{C}{2}\right)^{-r} \alpha^r(N)\mathbb{E}\left[\max_{1\leq a-a^0\leq N} \left| \frac{1}{N} \sum_{i=a^0+1}^{a} Y_i^{(l)} \right|^r \right] \\
&+ \left(\frac{C}{2}\right)^{-r} \alpha^r(N)\mathbb{E}\left[\max_{a-a^0\geq N} \left| \frac{1}{a-a^0} \sum_{i=a^0+1}^{a} Y_i^{(l)} \right|^r \right] \\
&\leq \left(\frac{C}{2}\right)^{-r} \alpha^r(N) \sum_{l=1}^{N} \frac{1}{N^{r/2}} + \left(\frac{C}{2}\right)^{-r} \alpha^r(N) \sum_{i=N}^{\infty} \frac{1}{i^{r/2}} \leq c.
\end{aligned}$$

And for $d \geq 2$:

$$\begin{aligned}
&\mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{1}{T_{\mathbf{k}_1,\mathbf{k}_2}} \sum_{\mathbf{i}\in M_\mathbf{a}} Y_\mathbf{i}^{(l)} \right| \geq C\alpha_n^{-1}(N) \right) \\
&\leq \mathbb{P}\left(\max_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\underline{\mathbf{n}} \\ \|\mathbf{a}-\mathbf{a}^0\|\geq 1}} \left| \frac{1}{f(\mathbf{a})} \sum_{\mathbf{i}\in M_\mathbf{a}} Y_\mathbf{i}^{(l)} \right| \geq C\alpha_n^{-1}(N) \right)
\end{aligned}$$

It follows with Markov's inequality and

$$\#\{\mathbf{a} : \underline{\mathbf{0}} \leq \mathbf{a} \leq \underline{\mathbf{n}},\ \|\mathbf{a}-\mathbf{a}^0\| = h\} \leq ch^{d-1} \leq cn^{d-1}, \quad h \leq n,$$

that

$$\begin{aligned}
&\mathbb{P}\left(\max_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\underline{\mathbf{n}} \\ \|\mathbf{a}-\mathbf{a}^0\|\geq 1}} \left| \frac{1}{f(\mathbf{a})} \sum_{\mathbf{i}\in M_\mathbf{a}} Y_i^{(l)} \right| \geq C\alpha_n^{-1}(N) \right) \\
&\leq C^{-r}\alpha_n^r(N) \sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\underline{\mathbf{n}} \\ \|\mathbf{a}-\mathbf{a}^0\|\geq 1}} \tilde{c}\frac{1}{f(\mathbf{a})^{r/2}} \leq C^{-r}\tilde{c}\alpha_n^r(N) \sum_{h=1}^{n} \sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\underline{\mathbf{n}} \\ \|\mathbf{a}-\mathbf{a}^0\|=h}} \frac{1}{f(\mathbf{a})^{r/2}} \\
&\leq C^{-r}c\alpha_n^r(N) \sum_{h=1}^{n} \sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\underline{\mathbf{n}} \\ \|\mathbf{a}-\mathbf{a}^0\|=h}} \frac{1}{(n^{d-1}\|\mathbf{a}-\mathbf{a}^0\|)^{r/2}} \\
&\leq C^{-r}c\alpha_n^r(N)n^{-(d-1)(r/2-1)} \sum_{h=1}^{n} \frac{1}{h^{r/2}} \leq C^{-r}c,
\end{aligned}$$

which implies the stated convergence order. This implies the lemma:

$$\begin{aligned}
&\max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2} A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{L_{n,\mathbf{k}_1,\mathbf{k}_2}} \right| \\
&\leq \max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{n^d T_{\mathbf{k}_1,\mathbf{k}_2}}{L_{n,\mathbf{k}_1,\mathbf{k}_2}} \right| \max_{l=1,\dots,p} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(l)(1)}_{\mathbf{k}_1,\mathbf{k}_2}}{T_{\mathbf{k}_1,\mathbf{k}_2}} \right| \max_{l=1,\dots,p} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2 \\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(l)(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \right| \\
&= \mathcal{O}_P(n^{-d/2})(\mathcal{O}_P(n^{-d/2}) + \alpha_n(N)^{-1}\mathcal{O}_P(1)) = \mathcal{O}_P(n^{-d/2}),
\end{aligned}$$

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(1)T}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}} \Delta \right| \leq \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{T_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(1)}_{\mathbf{k}_1,\mathbf{k}_2}} \right| \max_{l=1,\ldots,p} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(l)(1)}_{\mathbf{k}_1,\mathbf{k}_2}}{T_{\mathbf{k}_1,\mathbf{k}_2}} \right| \|\Delta\|$$
$$= \mathcal{O}_P(n^{-d/2}) + \alpha_n(N)^{-1}\mathcal{O}_P(1)$$

and

$$\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \Delta^T \frac{A^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}} \right| \leq \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{n^d}{B^{(2)}_{\mathbf{k}_1,\mathbf{k}_2}} \right| \|\Delta\| \max_{l=1,\ldots,p} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \|\mathbf{k}-\mathbf{k}^0\|\geq N}} \left| \frac{A^{(l)(2)}_{\mathbf{k}_1,\mathbf{k}_2}}{n^d} \right| = \mathcal{O}_P(n^{-d/2}).$$

□

Acknowledgements

We are thankful to two anonymous referees for the careful reading of this paper and the very helpful suggestions for improvements. We would like to thank our thesis adviser, Josef Steinebach, for his feedback on drafts of this paper and more generally for his continued support of our research.

Chapter 4

Change-point detection and bootstrap for Hilbert space valued random fields

Béatrice Bucchia* and Martin Wendler**
**University of Cologne*, ***Ernst Moritz Arndt University Greifswald*

Abstract

The problem of testing for the presence of epidemic changes in random fields is investigated. In order to be able to deal with general changes in the marginal distribution, a Cramér-von-Mises-type test is introduced which is based on Hilbert space theory. A functional central limit theorem for ρ-mixing Hilbert space valued random fields is proven. In order to avoid the estimation of the long-run variance and obtain critical values, Shao's dependent wild bootstrap method is adapted to this context. For this, a joint functional central limit theorem for the original and the bootstrap sample is shown. Finally, the theoretic results are supplemented by a short simulation study.

Keywords: change-point detection, dependent wild bootstrap, FCLT for Hilbert space valued r.v., random fields
AMS subject classification: 62H15, 62E20, 62M99, 60G60, 62H12

1 Introduction

1.1 Change-point tests for random fields

The focus of this paper lies on the problem of epidemic change in the mean for Hilbert space valued random fields. Given a data set of observations, a classical problem in change-point analysis consists of testing whether all the observations have the same stochastic structure (i.e. marginal distribution) or whether there is a subset (the change-set) of the data where the structure is different. For data corresponding to a time series, the split into different data subsets can be characterized by the points in time (the change-points) at which there is a structural break. In the epidemic change model, there are two possible change-points (the start and end of an "epidemic") and the structure of the data changes after the first change-point but reverts back to its original state after the second change-point. Extended to random fields, this becomes the problem of testing for rectangular change-sets. Epidemic changes are of interest not only in medicine (cf. e.g. Levin and Kline (1985)) but also e.g. in signal detection and textile fabric quality control (cf. e.g. Zhang and Bresee (1995)). The epidemic change-point problem was introduced by Levin and Kline (1985) and has since been the subject of numerous publications (see e.g. Csörgő and Horváth (1997), Račkauskas and Suquet (2004), Jarušková (2011), Aston and Kirch (2012a) and the publications listed therein). For random fields with a change in the mean, a nonparametric approach for this type of problem was considered in Jarušková and Piterbarg (2011) and Zemlys (2008) for i.i.d. observations and in Bucchia (2014) and Bucchia and Heuser (2015) for weakly dependent data. The test statistics considered in these publications are a special type of scan statistic, variants of which could - under the assumption that the distributions of the observations belong to a parametric family - also be used to test for changes in other parameters of a distribution (cf. e.g. Jarušková and Piterbarg (2011), Loader (1991), Siegmund and Yakir (2000)). For the nonparametric problem of a change in the distributions without any prior information on the family of distributions, however, a test based on the empirical distribution function F_n with

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}}$$

might be more useful. Equipped with the appropriate norm, one can regard these as sums of Hilbert space valued random variables, where the true distribution function of X_i is the expected value (in the Hilbert space) of $\mathbb{1}_{\{X_i \le \cdot\}}$. Therefore, the change in distribution problem can be translated into a change in mean problem for Hilbert space valued random variables.

The analysis of functional data over a spatial region is of independent interest. As a special case of spatio-temporal data, where measurements over time are taken at different locations in space, functional data may arise for instance in brain imaging or in space physics (cf. Gromenko and Kokoszka (2012)).

For weakly dependent time series of functional data, the epidemic change model was investigated by Aston and Kirch (2012a), who constructed test statistics based on projections on the principal components. By contrast, we aim to apply the approach used by Sharipov et al. (2016), who take the full functional structure into account. To the best of our knowledge, there are no results on asymptotic change-point tests for the specific setting considered here.

A popular approach for the construction of asymptotic tests for change in mean problems are so-called CUSUM-type tests, where the mean is estimated using cumulative sums of the observations. This leads to test statistics that can be written as functionals of the partial sum process of the data. Thus, under weak dependence, the main tool for the proof of the weak convergence of such CUSUM-type test statistics is a functional central limit theorem (FCLT). The continuous mapping theorem can then be applied to obtain the limit distribution. Therefore, one aim of this paper is to give an FCLT for Hilbert space valued random fields which can then be used for change-point tests.

Although the central limit theorem is known for multivariate and even Hilbert space valued weakly dependent random fields (cf. Bulinski (2004), Tone (2010, 2011)), most of the literature on FCLTs for random fields has focused on real-valued fields. For this setting, numerous results have been given not only for independent observations (cf.

Wichura (1969)) but also for weakly dependent fields. For instance, the monographs by Bulinski and Shashkin (2007) and Lin and Lu (1996) give examples of FCLTs under conditions related to association and mixing conditions respectively. For mixing random fields, Deo (1975, 1976) proved FCLTs under φ-mixing conditions and Kim and Seok (1995) extended the ideas of Deo's proofs to obtain FCLTs for ρ-mixing random fields. For i.i.d. Hilbert space valued random fields, Zemlys (2008) introduced a Hölderian FCLT. The FCLT presented here can be viewed as an extension of the approach by Deo (1975) first to vector-valued fields and then to Hilbert space valued fields.

After describing the bootstrap method considered here (section 1.2), we introduce the notations used throughout this article (section 1.3). We then present our main results in section 2. To illustrate our theoretical findings, our third section reports some simulation results. Proofs of our main results are relegated to section 4.

1.2 Bootstrap for Hilbert space valued processes

Nonparametric resampling methods like bootstrap are especially useful when dealing with stochastic processes, as the asymptotic distribution typically depends on a parameter function, which is hard to estimate. The bootstrap of the empirical distribution function has been well studied, starting with Bickel and Freedman (1981) in the independent case. This was extended to time series data by Naik-Nimbalkar and Rajarshi (1994), Peligrad (1998) and Radulović (2009) using block bootstrap methods adjusted for dependence. For an overview of the block bootstrap methods, see the book by Lahiri (2003). Shao (2010) introduced a different resampling method for time series: the dependent wild bootstrap, which generalizes Wu's (1986) wild bootstrap. Recently, Doukhan et al. (2015) extended the dependent wild bootstrap to empirical distribution functions and were able to show its validity. As seen above, the empirical distribution function can be interpreted as a function of Hilbert space valued random variables.

For more general Hilbert spaces, the bootstrap has been investigated by Politis and Romano (1994) and Dehling et al. (2015).

For the application to change-point detection, one needs a sequential bootstrap to mimic the behavior of the partial sum process. The consistency of the sequential multiplier bootstrap for the empirical distribution function under independence was shown by Gombay and Horváth (1999) and by Holmes et al. (2013) for the sequential empirical process indexed by functions. For dependent data, Inoue (2001) proposed a block multiplier bootstrap for the sequential empirical distribution function. Sharipov et al. (2016) studied block bootstrap for the partial sum process of Hilbert space valued random variables.

While there is a broad range of results for different bootstrap methods in the time series setting, much less work has been done for random fields, although ideas for this can be traced back thirty years to Hall (1985). Politis and Romano (1993) studied block bootstrap for partial sums, Zhu and Lahiri (2007) for the empirical distribution function. We are not aware of any bootstrap methods for Hilbert space valued random fields or of sequential bootstrap methods for the partial sums process of random fields (even in the real valued case).

The second aim of the paper is thus to give a sequential bootstrap method for Hilbert space valued random fields. We propose a generalization of the dependent wild bootstrap to random fields: Let $(X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^d}$ be a random field and $\bar{X}_n = \frac{1}{n^d}\sum_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}} X_i$. Furthermore, let $(V_n(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}$ be a real valued random field, independent of $(X_{\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^d}$, with

$E[V_n(\mathbf{i})] = 0$, $\mathrm{Var}[V_n(\mathbf{i})] = 1$ and a dependence structure to be specified later. The partial sum process $(S_n(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$ with

$$S_n(\mathbf{t}) = n^{-d/2} \sum_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} (X_\mathbf{i} - \mu)$$

will be bootstrapped by

$$S_n^\star(\mathbf{t}) = n^{-d/2} \sum_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} V_n(\mathbf{i})\,(X_\mathbf{i} - \hat{\mu}(\mathbf{i}))\,, \tag{1}$$

where $\hat{\mu}(\cdot)$ is an estimator for the mean function.

If the bootstrapped partial sum process mimics the behavior of the original partial sum process, by the continuous mapping theorem, the same holds for the bootstrap version of our test statistic. The classical choice proposed by Shao (2010) for the mean estimator is $\hat{\mu} \equiv \bar{X}_n$. However, under the alternative (presence of a change), the bootstrap with this choice of estimator might not be close to the distribution under the null hypothesis (no change). Therefore, we propose a different variant of our bootstrap. Let $\hat{C}_n$ be an estimator of the change-set such that $\varepsilon_1 n^d \leq \#\hat{C}_n \leq (1-\varepsilon_2)n^d$ for some $0 < \varepsilon_1 < 1-\varepsilon_2 < 1$ and all $n \in \mathbb{N}$. Define

$$\tilde{\mu}(\mathbf{k}) = \begin{cases} \frac{1}{\#\hat{C}_n}\sum_{\mathbf{i}\in\hat{C}_n} X_\mathbf{i} & \text{if } \mathbf{k}\in\hat{C}_n, \\ \frac{1}{\#\hat{C}_n^c}\sum_{\mathbf{i}\notin\hat{C}_n} X_\mathbf{i} & \text{if } \mathbf{k}\notin\hat{C}_n. \end{cases}$$

In the following, we will consider bootstrapped versions of $(S_n(\mathbf{t}))_{\mathbf{t}\in[0,1]^d}$ with either of these two mean estimators, i.e. $\hat{\mu}$ will denote either $\bar{X}_n$ or $\tilde{\mu}(\cdot)$. We will not specify the change-set estimator $\hat{C}_n$, but assume that it is a subblock of $(\underline{\mathbf{0}}, \underline{\mathbf{n}}]$ which fulfills the size restriction above (cf. Bucchia and Heuser (2015) for some example for $\mathbb{R}^p$-valued random fields).

1.3 Notations

Before introducing the main results, we will now cover some notations and conventions that will be used throughout this paper. $\mathbb{R}^d$ denotes the vector space of real vectors, equipped with the usual partial order, and $\mathbb{Z}^d$ and $\mathbb{N}^d$ denote the subsets of integer and positive integer vectors, respectively. For an integer $k \in \mathbb{Z}$, we denote $(k,\ldots,k)^t \in \mathbb{Z}^d$ by $\underline{\mathbf{k}}$, and write general vectors $(x_1,\ldots,x_d)^t \in \mathbb{R}^d$ as $\mathbf{x}$. For $\mathbf{x}\in\mathbb{R}^d$, we use the following notations: $\lfloor\mathbf{x}\rfloor = (\lfloor x_1\rfloor,\ldots,\lfloor x_d\rfloor)^t$ is the integer part of $\mathbf{x}$, $|\mathbf{x}| = (|x_1|,\ldots,|x_d|)$ and $[\mathbf{x}] = x_1\cdots x_d$. For a set $S\in\mathbb{R}^d$ and a number $n\in\mathbb{N}$, we write

$$S\ominus S = \{\mathbf{x}\in\mathbb{R}^d : \exists \mathbf{s},\mathbf{t}\in S,\, \mathbf{x} = \mathbf{s}-\mathbf{t}\},$$

$\#S = \mathrm{card}(S)$ if S is finite, and $nS := \{n\mathbf{x} : \mathbf{x}\in S\}$, where $n\mathbf{x} = (nx_1,\ldots,nx_d)^t$.

A block in $\mathbb{R}^d$ is a set of the form $(\mathbf{x},\mathbf{y}] = \{\mathbf{z} : x_i < z_i \leq y_i,\ i = 1,\ldots,d\}$ for $\mathbf{x},\mathbf{y}\in\mathbb{R}^d$ ($(\mathbf{x},\mathbf{y}] = \emptyset$, if $x_i \geq y_i$ for some $i\in\{1,\ldots,d\}$). A block in $\mathbb{Z}^d$ is the intersection of a block in $\mathbb{R}^d$ and the set $\mathbb{Z}^d$. In particular, for a block $B = (\mathbf{s},\mathbf{t}] \subseteq [0,1]^d$ and $n\in\mathbb{N}$, we denote the associated block $nB\cap\mathbb{Z}^d = (\lfloor n\mathbf{s}\rfloor, \lfloor n\mathbf{t}\rfloor]\cap\mathbb{Z}^d$ by B_n. Writing λ for the Lebesgue measure on $\mathbb{R}^d$, it then holds that $\lambda((\lfloor n\mathbf{s}\rfloor,\lfloor n\mathbf{t}\rfloor]) = \#B_n$.

Denoting the supremum norm on $\mathbb{R}^d$ by $\|\cdot\|_\infty$, we define the distance

$$\mathrm{dist}(S,Q) = \inf\{\|\mathbf{x}-\mathbf{y}\|_\infty : \mathbf{x}\in S, \mathbf{y}\in Q\}$$

between two sets S and Q.

Given observations $(X_{\mathbf{j}})_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}$ ($n\in\mathbb{N}$), a real-valued random field $(V_n(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}$ will be called a dependent multiplier field with bandwidth $q=q_n$ if it is a Gaussian random field, independent of $(X_{\mathbf{j}})_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}$, with $E[V_n(\mathbf{i})]=0$, $\mathrm{Var}[V_n(\mathbf{i})]=1$ and

$$\mathrm{Cov}\left(V_n(\mathbf{i}),V_n(\mathbf{j})\right)=\omega((\mathbf{i}-\mathbf{j})/q)$$

for a symmetric bounded function ω that is continuous at zero with $\omega(\underline{\mathbf{0}})=1$ and

$$\sum_{-\underline{\mathbf{n}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}|\omega(\mathbf{j}/q)|=\mathcal{O}(q^d).$$

We consider a separable (real) Hilbert space H with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|x\|=\sqrt{|\langle x,x\rangle|}$. (Since $\mathbb{R}^k$ with the inner product $\langle x,y\rangle=x^ty$ is also a Hilbert space, we will also denote the usual l_2-norm in $\mathbb{R}^k$ by $\|\cdot\|$.) Unless stated otherwise, the spaces considered are always seen as measurable spaces with their Borel σ-algebra. Let $L(H,H)$ be the space of bounded (with respect to the operator norm $\|S\|=\sup\{\|S(h)\| : h\in H, \|h\|\leq 1\}$) linear operators from H to H. $\mathcal{S}(H)$ denotes the set of all self-adjoint positive nuclear operators in $L(H,H)$. The notation $\{e_k\}_{k\in\mathbb{N}}$ is used for complete orthonormal systems in H. The trace of a nuclear operator $S\in\mathcal{S}(H)$ is $tr(S)=\sum_{i=1}^\infty\langle Se_i,e_i\rangle$, and $\|S-S'\|_{tr}=tr(S-S')$ defines a metric on $\mathcal{S}(H)$. Consider the span H_k of the first k e_i. Then the orthogonal projections on H_k are $P_k: H\to H_k$, $h\mapsto\sum_{i=1}^k\langle h,e_i\rangle e_i$, and the corresponding complementary operators are $A_k: H\to H$, $h\mapsto h-\sum_{i=1}^k\langle h,e_i\rangle e_i=\sum_{i=k+1}^\infty\langle h,e_i\rangle e_i$. For any H-valued random variable, we write $X^{(k)}=P_kX$ and $X^k=<X,e_k>$.

In analogy to the case $H=\mathbb{R}$, we will consider stochastic processes in the space

$$D_H([0,1]^d)=\{x:[0,1]^d\to H|\, x \text{ has quadrant limits and is cont. from above}\}$$

endowed with the metric

$$d_S(x,y)=\inf_{\lambda\in\Lambda}\{\max\{\sup_{\mathbf{t}\in[0,1]^d}\|x(\mathbf{t})-y(\lambda(\mathbf{t}))\|,\sup_{\mathbf{t}\in[0,1]^d}\|\mathbf{t}-\lambda(\mathbf{t})\|\}\},$$

where

$$\Lambda=\{\lambda:[0,1]^d\to[0,1]^d : \lambda(t_1,\ldots,t_d)=(\lambda_1(t_1),\ldots,\lambda_d(t_d)),\ \lambda_p:[0,1]\to[0,1]$$
$$\text{cont., strict. increasing and } \lambda_p(0)=0,\lambda_p(1)=1 \text{ for all } p=1,\ldots,d\}$$

(cf. e.g. Neuhaus (1969) for $D_{\mathbb{R}}([0,1]^d)$). Let $C_H([0,1]^d)$ be the subset of functions in $D_H([0,1]^d)$ that are continuous with respect to the supremum-norm $\|x\|_\infty=\sup\{\|x(\mathbf{t})\| : \mathbf{t}\in[0,1]^d\}$.

It can be seen that the proofs which Neuhaus (1969) provides for $D_{\mathbb{R}}([0,1]^d)$ can be extended to our present setting with only minor changes. In particular, $(D_H([0,1]^d),d_S)$ is separable and (topologically) complete and the Borel σ-algebra coincides with the σ-algebra generated by the coordinate mappings (for dense subsets of $[0,1]^d$). The relation between d_S and the supremum norm on $D_H([0,1]^d)$ is the same as in $D_{\mathbb{R}}([0,1]^d)$, and $(C_H([0,1]^d), \|\cdot\|_\infty)$ is a separable Banach space with $C_H([0,1]^d) \subseteq D_H([0,1]^d)$.

If $(X_{\mathbf{t}})_{\mathbf{t}\in[0,1]^d}$ is a stochastic process with values in $D_H([0,1]^d)$, then the increment $X(B)$ of X around a block $B = \prod_{i=1}^d (s_i, t_i]$ is given by

$$X(B) = \sum_{\varepsilon_1=0,1} \cdots \sum_{\varepsilon_d=0,1} (-1)^{d-\sum_{i=1}^d \varepsilon_i} X(s_1 + \varepsilon_1(t_1 - s_1), \ldots, s_d + \varepsilon_d(t_d - s_d)),$$

where we use the notations $X(\mathbf{t})$ and $X_{\mathbf{t}}$ synonymously. For ease of notation, we will often write this as

$$X(B) = \sum_{\boldsymbol{\varepsilon}\in\{0,1\}^d} (-1)^{d-\sum_{j=1}^d \varepsilon_j} X(\mathbf{s} + \boldsymbol{\varepsilon}(\mathbf{t} - \mathbf{s})).$$

Since $X_{\mathbf{t}} = X((\underline{\mathbf{0}}, \mathbf{t}])$ a.s. for a process which vanishes at zero (i.e. $X_{\mathbf{s}} = 0$ a.s. for any $\mathbf{s} \in [0,1]^d$ with $\min s_i = 0$), we often denote $X((\underline{\mathbf{0}}, \mathbf{t}])$ and $X(n(\underline{\mathbf{0}}, \mathbf{t}])$ by $X(\mathbf{t})$ and $X_n(\mathbf{t})$ respectively. For $\mathbf{k}, \mathbf{m} \in \mathbb{Z}^d$ and $\{x_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$, we write

$$\sum_{\mathbf{k}<\mathbf{j}\le\mathbf{m}} x_{\mathbf{j}} = \begin{cases} \sum\limits_{\mathbf{j}\in(\mathbf{k},\mathbf{m}]\cap\mathbb{Z}^d} x_{\mathbf{j}}, & \mathbf{k} < \mathbf{m} \\ \sum\limits_{\mathbf{j}\in\emptyset} x_{\mathbf{j}} = 0, & \mathbf{k} \not< \mathbf{m}. \end{cases}$$

We will now define the Hilbert space valued analogue of the Brownian sheet (or Chentsov process):

Definition 1.1. *An H-valued stochastic process $X = (X_{\mathbf{t}})_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in H with covariance operator $S \in \mathcal{S}(H)$ iff*

1. *$P(X \in C_H([0,1]^d)) = 1$,*
2. *$X_{\mathbf{t}} = 0$ a.s. if $t_i = 0$ for any $i \in \{1, \ldots, d\}$ and*
3. *for pairwise disjoint blocks $B_1, \ldots, B_m$ in $[0,1]^d$, the increments $X(B_1), \ldots, X(B_m)$ are independent Gaussian random elements in H with mean zero and covariance operators $\lambda(B_i)S$, where $S \in \mathcal{S}(H)$ does not depend on B_i.*

Remark 1.1.
- *In order to see that the independence and Gaussian distribution of the increments over pairwise disjoint blocks yields a Gaussian process, one can proceed analogously to the one-dimensional case and write any linear combination of $X_{\mathbf{t}_i} = X((\underline{\mathbf{0}}, \mathbf{t}_i])$ for points $\mathbf{t}_i \in [0,1]^d$ $(i = 1, \ldots, l)$ as a linear combination of increments over pairwise disjoint blocks whose union is $\cup_{i=1}^l (\underline{\mathbf{0}}, \mathbf{t}_i]$.*
- *If $X = (X_{\mathbf{t}})_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in H with covariance operator $S \in \mathcal{S}(H)$, then $(\langle X(\mathbf{t}), h\rangle)_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet with covariance $\langle Sh, h\rangle$ in $\mathbb{R}$ for any $h \in H$.*

For a σ-algebra $\mathcal{A}$, we define $L^p(\mathcal{A}, H)$ as the set of all $\mathcal{A}$-measurable H-valued random elements X with $\|X\|_p = (E[\|X\|^p])^{1/p} < \infty$.

As a measure of dependence, we use the following mixing conditions: For two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, we can define the usual strong mixing coefficients

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \mathcal{A}, B \in \mathcal{B}\}$$

as well as the ρ-mixing coefficients

$$\rho_{\mathbb{R}}(\mathcal{A},\mathcal{B}) = \sup\left\{\frac{|\mathrm{Cov}(X,Y)|}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}} : X \in L^2(\mathcal{A},\mathbb{R}), Y \in L^2(\mathcal{B},\mathbb{R}), \mathrm{var}(X), \mathrm{var}(Y) > 0\right\},$$

which lead to the following types of mixing coefficients for random fields. Let $\mathcal{A}_S = \sigma(X_{\mathbf{k}} : \mathbf{k} \in S)$ and define

$$\begin{aligned}\rho_{\mathbb{R}}(r) = \sup\{\rho_{\mathbb{R}}(\mathcal{A}_S,\mathcal{A}_Q) : S, Q \subseteq \mathbb{Z}^d, \exists i \in \{1,\dots,d\}\, \exists A, B \subset \mathbb{Z}, \mathrm{dist}(A,B) \geq r :\\ \forall \mathbf{j} \in S, \mathbf{k} \in Q : j_i \in A, k_i \in B\}\end{aligned} \tag{2}$$

and

$$\rho^*_{\mathbb{R}}(r) = \sup\{\rho_{\mathbb{R}}(\mathcal{A}_S,\mathcal{A}_Q) : S, Q \subseteq \mathbb{Z}^d, \mathrm{dist}(S,Q) \geq r\}.$$

As usual, we say that a random field is $\rho_{\mathbb{R}}$-mixing ($\rho^*_{\mathbb{R}}$-mixing), if $\lim_{r\to\infty} \rho_{\mathbb{R}}(r) = 0$ ($\lim_{r\to\infty} \rho^*_{\mathbb{R}}(r) = 0$).

Finally, we use an α-mixing coefficient where the cardinality of the index sets is restricted: For $k, m \in \mathbb{N}$, define

$$\alpha_{k,m}(r) = \sup\{\alpha(\mathcal{A}_S,\mathcal{A}_Q) : S, Q \subseteq \mathbb{Z}^d, \mathrm{dist}(S,Q) \geq r, \#S \leq k, \#Q \leq m\}.$$

2 Main results

2.1 Change-point problem for random fields

We now present our FCLT for Hilbert space valued $\rho_{\mathbb{R}}$-mixing random fields. For real-valued ρ-mixing random fields, Kim and Seok (1995) used an approach proposed by Ibragimov (1975) to prove the FCLT under an additional assumption on the growth of the variance of the partial sums. Here, we have used a ρ-mixing condition that is stronger than the one in Kim and Seok (1995) (we allow interlaced index sets in (2)) and, since it is unclear how the growth condition would translate to the Hilbert space context, we use assumption 2 (see below) on the α-mixing coefficients instead, which implies condition (2.6) in Corollary 2.3 of Kim and Seok (1995). However, although our assumptions are therefore stronger for real-valued fields, the following result is applicable not only to this special case but to general separable Hilbert spaces. As a byproduct of our proof, we extend a result from Deo (1975) to multivariate $\rho_{\mathbb{R}}$-mixing random fields.

Theorem 2.1. *Let $\{X_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ be a strictly stationary H-valued random field with $EX_{\underline{\mathbf{1}}} = \mu$. Assume that $\{X_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ is $\rho_{\mathbb{R}}$-mixing and that the following conditions hold for some $\delta > 0$:*

1. $E\|X_{\underline{\mathbf{1}}}\|^{2+\delta} < \infty$
2. $\sum_{m\geq 1} m^{d-1}\alpha_{1,1}(m)^{\delta/(2+\delta)} < \infty$

Then

$$\left\{\frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq \mathbf{j} \leq \lfloor n\mathbf{t}\rfloor} (X_{\mathbf{j}} - \mu)\right\}_{\mathbf{t}\in[0,1]^d} \Rightarrow \{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d},$$

where $\{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ *is a Brownian sheet in* H *and* $W(\underline{\mathbf{1}})$ *has the covariance operator* $S \in \mathcal{S}(H)$*, defined by*

$$\langle Sx, y\rangle = \sum_{\mathbf{k}\in\mathbb{Z}^d} E[\langle X_{\underline{\mathbf{0}}} - \mu, x\rangle \langle X_{\mathbf{k}} - \mu, y\rangle], \quad \textit{for } x, y \in H. \tag{3}$$

Furthermore, the series in (3) *converges absolutely.*

This can be used for the following change-point problem: Given observations $\{X_{\mathbf{j}}\}_{\mathbf{j}\in\{1,\ldots,n\}^d}$ with values in H, we want to test the null-hypothesis

$$H: \quad EX_{\mathbf{j}} = \mu \quad \forall\, \mathbf{k} \in \{1,\ldots,n\}^d$$

against the epidemic change alternative

$$H_A: \quad \exists\, \underline{\mathbf{1}} \le \mathbf{k}_0 < \mathbf{m}_0 \le \underline{\mathbf{n}} : EX_{\mathbf{k}} = \begin{cases} \mu, & \mathbf{k} \in \{1,\ldots,n\}^d \setminus (\mathbf{k}_0, \mathbf{m}_0] \\ \mu + \delta, & \mathbf{k} \in (\mathbf{k}_0, \mathbf{m}_0], \end{cases}$$

where $\mu, \delta \in H$ and $\mathbf{k}_0, \mathbf{m}_0$ are unknown. CUSUM-type asymptotic tests for the epidemic change in the mean problem have been investigated e.g. by Yao (1993), Csörgő and Horváth (1997), Račkauskas and Suquet (2004) and Jarušková (2011) for real-valued time series. These were extended to i.i.d. random fields by Zemlys (2008) - who used an approach similar to Račkauskas and Suquet (2004) - and Jarušková and Piterbarg (2011). For weakly dependent random fields, Bucchia (2014) gave an extension of some results from Jarušková and Piterbarg (2011). The epidemic change problem for weakly dependent time series of functional observations was treated by Aston and Kirch (2012a), who constructed asymptotic tests based on the principal components of the data.

Consider the test statistic

$$T_n = \max_{\underline{\mathbf{0}}\le\mathbf{k}<\mathbf{m}\le\underline{\mathbf{n}}} \frac{1}{n^{d/2}} \left\| \sum_{\mathbf{k}<\mathbf{j}\le\mathbf{m}} X_{\mathbf{j}} - \frac{[\mathbf{m}-\mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\underline{\mathbf{n}}} X_{\mathbf{j}} \right\|.$$

Analogously to the univariate case, since both the maximum function and the Hilbert space norm are continuous, Theorem 2.1 together with the continuous mapping theorem can be used to obtain the limit distribution of these statistics under H:

Corollary 2.1. *Under the assumptions of Theorem 2.1, it holds that*

$$T_n \Rightarrow \sup_{\underline{\mathbf{0}}\le\mathbf{s}<\mathbf{t}\le\underline{\mathbf{1}}} \|W(\mathbf{s},\mathbf{t}] - [\mathbf{t}-\mathbf{s}]W(\underline{\mathbf{1}})\| = T,$$

where $\{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ *is the* H*-valued Brownian sheet defined in Theorem 2.1.*

For $\mathbb{R}^p$-valued observations $\{X_{\mathbf{j}}\}_{\mathbf{j}\in\{1,\ldots,n\}^d}$, this result can be used to obtain a test for the change in distribution problem of testing

$$H: \quad F(\mathbf{t}) = P(X_{\mathbf{i}} \le \mathbf{t}) \quad \forall\, \mathbf{i} \in \{1,\ldots,n\}^d, \mathbf{t} \in \mathbb{R}^p$$

against the alternative

$$H_A: \quad \exists\, \underline{\mathbf{1}} \le \mathbf{k}_0 < \mathbf{m}_0 \le \underline{\mathbf{n}} : P(X_{\mathbf{k}} \le \mathbf{t}) = \begin{cases} F(\mathbf{t}), & \mathbf{k} \in \{1,\ldots,n\}^d \setminus (\mathbf{k}_0, \mathbf{m}_0] \\ G(\mathbf{t}), & \mathbf{k} \in (\mathbf{k}_0, \mathbf{m}_0], \end{cases}$$

where the distribution functions F and G, $F \neq G$, are unknown. Our goal is to write this as a change in mean problem for a suited Hilbert space. Common test statistics depend on the empirical distribution functions as estimators for the unknown parameters F and G. These are sums over the indicator functions $\mathbb{1}_{\{X_{\mathbf{j}} \leq \mathbf{t}\}}$, $\mathbf{t} \in \mathbb{R}^p$. For some nonnegative, bounded weight function $w : \mathbb{R}^p \to \mathbb{R}$ with $\int_{\mathbb{R}^p} w(\mathbf{t}) d\mathbf{t} < \infty$, the latter can be interpreted as random elements of the Hilbert space $L^2(\mathbb{R}^p, w)$ of measurable functions $f : \mathbb{R}^p \to \mathbb{R}$, with $\|f\| < \infty$ for the norm induced by the inner product

$$\langle f, g \rangle = \int_{\mathbb{R}^p} f(\mathbf{t}) g(\mathbf{t}) w(\mathbf{t}) d\mathbf{t}.$$

If F is the distribution function of $X_{\mathbf{j}}$, it can be seen that for any $h \in L^2(\mathbb{R}^p, w)$,

$$E\left[\left\langle \mathbb{1}_{\{X_{\mathbf{j}} \leq \cdot\}}, h \right\rangle\right] = E\left[\int_{\mathbb{R}^p} \mathbb{1}_{\{X_{\mathbf{j}} \leq \mathbf{t}\}} h(\mathbf{t}) w(\mathbf{t}) d\mathbf{t}\right] = \int_{\mathbb{R}^p} F(\mathbf{t}) h(\mathbf{t}) w(\mathbf{t}) d\mathbf{t} = \langle F, h \rangle$$

by Fubini's theorem. Therefore, F is the expected value of $\mathbb{1}_{\{X_{\mathbf{j}} \leq \cdot\}}$ in $L^2(\mathbb{R}^p, w)$ and we obtain a Cramér-von Mises type test for the change in distribution problem by translating Corollary 2.1 for this special case:

Corollary 2.2. *Let $\{X_{\mathbf{j}}\}_{\mathbf{j} \in \mathbb{Z}^d}$ be an $\mathbb{R}^p$-valued stationary random field with marginal distribution function F, which is $\rho_{\mathbb{R}}$-mixing with α-mixing coefficients that satisfy*

$$\sum_{m \geq 1} m^{d-1} \alpha_{1,1}(m)^{\delta/(2+\delta)} < \infty$$

for some $\delta > 0$. The change-point statistic

$$T_{n,w} = \max_{\underline{\mathbf{0}} \leq \mathbf{k} < \mathbf{m} \leq \underline{\mathbf{n}}} \frac{1}{n^d} \int_{\mathbb{R}^p} \left(\sum_{\mathbf{k} < \mathbf{j} \leq \mathbf{m}} \mathbb{1}_{\{X_{\mathbf{j}} \leq \mathbf{x}\}} - \frac{[\mathbf{m} - \mathbf{k}]}{n^d} \sum_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \underline{\mathbf{n}}} \mathbb{1}_{\{X_{\mathbf{j}} \leq \mathbf{x}\}} \right)^2 w(\mathbf{x}) d\mathbf{x}$$

then satisfies

$$T_{n,w} \Rightarrow \sup_{\underline{\mathbf{0}} \leq \mathbf{s} < \mathbf{t} \leq \underline{\mathbf{1}}} \|W(\mathbf{s}, \mathbf{t}] - [\mathbf{t} - \mathbf{s}] W(\underline{\mathbf{1}})\|^2 = T_w,$$

where $\{W(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d}$ is a Brownian sheet in $L^2(\mathbb{R}^p, w)$ and $W(\underline{\mathbf{1}})$ has the covariance operator $S \in \mathcal{S}(L^2(\mathbb{R}^p, w))$ defined by

$$\langle Sx, y \rangle = \sum_{\mathbf{k} \in \mathbb{Z}^d} E\left[\int_{\mathbb{R}^p} \left(\mathbb{1}_{\{X_{\underline{\mathbf{0}}} \leq \mathbf{t}\}} - F(\mathbf{t})\right) x(\mathbf{t}) w(\mathbf{t}) d\mathbf{t} \int_{\mathbb{R}^p} \left(\mathbb{1}_{\{X_{\mathbf{k}} \leq \mathbf{t}\}} - F(\mathbf{t})\right) y(\mathbf{t}) w(\mathbf{t}) d\mathbf{t}\right],$$

for $x, y \in L^2(\mathbb{R}^p, w)$.

Note that since $x \mapsto \mathbb{1}_{\{x \leq \cdot\}}$ is a measurable bijection, the mixing properties of $\{X_{\mathbf{j}}\}_{\mathbf{j} \in \mathbb{Z}^d}$ are preserved. Due to the non-negativity and integrability of w, the moment condition of Theorem 2.1 is satisfied.

2.2 Dependent wild bootstrap for change-point detection

We formulate our theorem on the consistency of the bootstrap version of the partial sum process for Hilbert space valued random fields.

Theorem 2.2. *Let the assumptions of Theorem 2.1 hold and assume additionally that*

$$\sum_{m\geq 1} m^{d-1}\alpha_{2,2}(m)^{\delta/(2+\delta)} < \infty \tag{4}$$

and $E\|X_{\underline{\mathbf{1}}}\|^{4+2\delta} < \infty$*. Furthermore, let* $(V_{n,1}(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}, \ldots, (V_{n,K}(\mathbf{i}))_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}$ *(*$K\in\mathbb{N}$*) be independent copies of the same dependent multiplier field.* [1] *Lastly, let the bandwidth* $q = q_n$ *fulfill* $q_n \to \infty$ *and* $q_n = o(\sqrt{n})$*. Then*

$$\left(S_n, S^\star_{n,1}, \ldots, S^\star_{n,K}\right) \Rightarrow (W, W^\star_1, \ldots, W^\star_K) \quad \text{in } D_H([0,1]^d)^{K+1}$$

where $S^\star_{n,1}, \ldots, S^\star_{n,K}$ *are bootstrapped partial sum processes defined as in* (1) *and* $W^\star_1, \ldots, W^\star_K$ *are independent copies of the Hilbert space valued Brownian sheet* W *from Theorem 2.1.*

Remark 2.1. *The additional assumption* (4) *is used to obtain the convergence of long-run variance estimators over sets* $B \subseteq (0,1]^d$ *which are either blocks or finite unions of disjoint blocks. Consider estimators of the form*

$$\hat{\Sigma}_n(B) = \sum_{\mathbf{h}\in B_n\ominus B_n} \omega(\mathbf{h}/q)\,\frac{1}{n^d}\sum_{\mathbf{a}\colon \mathbf{a},\mathbf{a}+\mathbf{h}\in B_n} (X^{(k)}_{\mathbf{a}} - \hat{\mu}^{(k)}(\mathbf{a}))(X^{(k)}_{\mathbf{a}+\mathbf{h}} - \hat{\mu}^{(k)}(\mathbf{a}+\mathbf{h}))^t.$$

As shown in Bucchia and Heuser (2015), these estimators can be written as kernel-type long-run variance estimators $\hat{\Sigma}_{Y,n}(B)$ *for the centered process* $Y^{(k)} = \{X^{(k)}_{\mathbf{j}} - \mu^{(k)}\}_{\mathbf{j}\in\mathbb{Z}^d}$*, plus an error term which converges to 0 in probability (cf. Bucchia and Heuser (2015), Theorem 3.1). The classical proof of* $\hat{\Sigma}_{Y,n}(B) \xrightarrow{P} \lambda(B)\Sigma$ *(cf. e.g. Lavancier (2008)), where* Σ *is the long-run variance matrix of* $X^{(k)}$*, works by showing* $E\left[\hat{\Sigma}_{Y,n}(B)\right] \longrightarrow \lambda(B)\Sigma$ *and* $E\left[\left(\hat{\Sigma}_{Y,n}(B) - E\left[\hat{\Sigma}_{Y,n}(B)\right]\right)^2\right] \to 0$*. Only a slight modification of the proof in Lavancier (2008) (who considered* $B = (0,1]^d$ *and slightly less general kernel-functions* ω*) is needed to obtain the convergence of the mean. For the second part, we concentrate on the case* $k = 1$ *to simplify notation, but the cases* $k \geq 2$ *work the same way. Note that by assumption* (4) *(cf. Guyon (1995), p. 110), there exists a* $C > 0$ *such that*

$$\begin{aligned}
&E\left[\left(\frac{1}{n^d}\sum_{\mathbf{a}\colon \mathbf{a},\mathbf{a}+\mathbf{h}\in B_n}\left(Y^{(1)}_{\mathbf{a}}Y^{(1)}_{\mathbf{a}+\mathbf{h}} - E\left[Y^{(1)}_{\mathbf{a}}Y^{(1)}_{\mathbf{a}+\mathbf{h}}\right]\right)\right)^2\right]\\
=&\frac{1}{n^{2d}}\sum_{\mathbf{a},\mathbf{a}'\colon \mathbf{a},\mathbf{a}',\mathbf{a}+\mathbf{h},\mathbf{a}'+\mathbf{h}\in B_n}\operatorname{Cov}(Y^{(1)}_{\mathbf{a}}Y^{(1)}_{\mathbf{a}+\mathbf{h}}, Y^{(1)}_{\mathbf{a}'}Y^{(1)}_{\mathbf{a}'+\mathbf{h}})\\
\leq&\frac{1}{n^d}\sum_{\mathbf{l}\in\mathbb{Z}^d}\left|\operatorname{Cov}(Y^{(1)}_{\underline{\mathbf{0}}}Y^{(1)}_{\mathbf{h}}, Y^{(1)}_{\mathbf{l}}Y^{(1)}_{\mathbf{l}+\mathbf{h}})\right| \leq C\frac{1}{n^d}
\end{aligned}$$

[1]Note that the restriction on ω is weaker than the restriction $\omega(\mathbf{j}/q) = 0$ for $\mathbf{j}$ with $\max_i |j_i| \geq q$ used in Bucchia and Heuser (2015), but since the proofs remain essentially unaffected, all results from that paper are still applicable.

and therefore

$$
\begin{aligned}
&E\left[\left(\hat{\Sigma}_{Y,n}(B)-E\left[\hat{\Sigma}_{Y,n}(B)\right]\right)^2\right]\\
\leq&\left(\sum_{\mathbf{h}\in B_n\ominus B_n}|\omega(\mathbf{h}/q)|\left\|\frac{1}{n^d}\sum_{\mathbf{a}:\,\mathbf{a},\mathbf{a}+\mathbf{h}\in B_n}\left(Y_{\mathbf{a}}^{(1)}Y_{\mathbf{a}+\mathbf{h}}^{(1)}-E\left[Y_{\mathbf{a}}^{(1)}Y_{\mathbf{a}+\mathbf{h}}^{(1)}\right]\right)\right\|_2\right)^2\\
\leq&C\frac{1}{n^d}\left(\sum_{-\underline{\mathbf{n}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}|\omega(\mathbf{j}/q)|\right)^2\leq C\frac{q^{2d}}{n^d}\longrightarrow 0.
\end{aligned}
$$

Alternatively, in order to use the proof presented e.g. in Lavancier (2008) for bandwidths $q=o(n)$, one could replace assumption (4) *by stronger mixing and integrability conditions (cf. e.g. Guyon (1995), Lemma 4.6.2) in order to obtain the summability of the fourth-order cumulants (cf. Assumption (Y2) in Bucchia and Heuser (2015)).*

Write $T_{n,1}^\star,\dots,T_{n,K}^\star$ and $T_{n,w,1}^\star,\dots,T_{n,w,K}^\star$ for the bootstrapped analogues of the above change-point statistics, where $X_{\mathbf{j}}$ and $\mathbb{1}_{\{X_{\mathbf{j}}\leq\cdot\}}$ are replaced by $V_{n,l}(\mathbf{j})\,(X_{\mathbf{j}}-\hat{\mu}(\mathbf{j}))$ and $V_{n,l}(\mathbf{j})\left(\mathbb{1}_{\{X_{\mathbf{j}}\leq\cdot\}}-\hat{\mu}(\mathbf{j})\right)$ respectively $(l=1,\dots,K)$. As a direct consequence of Theorem 2.2, we obtain the same limit distributions as for the original statistics:

Corollary 2.3. *(a) Let the assumptions of Theorem 2.2 hold. Then it holds that*

$$(T_n,T_{n,1}^\star,\dots,T_{n,K}^\star)\Rightarrow(T,T_1^\star,\dots,T_K^\star),$$

where $T_1^\star,\dots,T_K^\star$ are independent copies of T.

(b) Let $\{X_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ be an $\mathbb{R}^p$-valued stationary random field that fulfills the assumptions of Corollary 2.2 and (4). *Let $(V_{n,1}(\mathbf{j}))_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}},\dots,(V_{n,K}(\mathbf{j}))_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}$ be as in Theorem 2.2. Then it holds that*

$$(T_{n,w},T_{n,w,1}^\star,\dots,T_{n,w,K}^\star)\Rightarrow(T_w,T_{w,1}^\star,\dots,T_{w,K}^\star),$$

where $T_{w,1}^\star,\dots,T_{w,K}^\star$ are independent copies of T_w.

Using this corollary, we can obtain critical values for the test statistic T_n (and analogously for $T_{n,w}$) in the following way: Simulate the K conditionally independent copies $T_{n,1}^\star,\dots,T_{n,K}^\star$. For a given significance level $\alpha\in(0,1)$, calculate the $(1-\alpha)$ sample quantile $q_{n,K}^\star(1-\alpha)$ of $T_{n,1}^\star,\dots,T_{n,K}^\star$ and reject the hypothesis of stationarity if $T_n\geq q_{n,K}^\star(1-\alpha)$. Then Lemma F.1 in Bücher and Kojadinovic (2014) yields $\lim_{K\to\infty}\lim_{n\to\infty}P(T_n\geq q_{n,K}^\star(1-\alpha))=\alpha$.

3 Simulation study

To illustrate the finite sample behavior of the Cramér-von Mises type change-point test (using $T_{n,w}$) with dependent wild bootstrap, we present the results of a small simulation study. We use the distribution function of the $N(100,1000)$-distribution as a weight function w to define the Hilbert space $L^2(\mathbb{R},w)$. As a data generating process, we use an autoregressive process

$$Y_{\mathbf{k}}=aY_{k_1-1,k_2}+aY_{k_1,k_2-1}-a^2Y_{k_1-1,k_2-1}+\epsilon_{k_1,k_2},\quad \mathbf{k}\in\{1,\dots,n\}^2$$

Example 1	Example 2	Example 3
$\left(\begin{pmatrix}0.2\\0.3\end{pmatrix},\begin{pmatrix}0.6\\0.55\end{pmatrix}\right]$	$\left(\begin{pmatrix}0.1\\0.1\end{pmatrix},\begin{pmatrix}0.9\\0.85\end{pmatrix}\right]$	$\left(\begin{pmatrix}0.05\\0.1\end{pmatrix},\begin{pmatrix}0.95\\1.0\end{pmatrix}\right]$
Vol= 0.1	Vol= 0.6	Vol= 0.81

Table 4.1: Values of $(\boldsymbol{\theta}, \boldsymbol{\gamma}]$ and corresponding volumes for the different examples.

for dimension $d = 2$, where the parameter a, which reflects the dependence structure of the process, takes the values $a = 0.2, 0.5$ and the innovations $\{\epsilon_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ are i.i.d. $N(0,(1-a^2)^d)$-distributed. Applying the results in Doukhan (1994), Section 2.1.1, it can be seen that this process fulfills the mixing assumptions of Theorems 2.1 and 2.2. We use sample sizes $n = 30, 40, 50$. We consider two types of changes in distribution, changes in the mean and changes in the skewness of the process, each over a change-set of the form $C = (\boldsymbol{\theta}, \boldsymbol{\gamma}]$ ($\underline{\mathbf{0}} < \boldsymbol{\theta} < \boldsymbol{\gamma} < \underline{\mathbf{1}}$). For the change in mean, we consider

$$X^{(1)}_{\mathbf{k}} = Y_{\mathbf{k}} + \Delta \mathbb{1}_{C_n}(\mathbf{k}), \quad \mathbf{k} \in \{1, \dots, n\}^d,$$

with $\Delta = 0, 0.5, 1$. For the change in skewness, we use the same approach as in Sharipov et al. (2016) and simulate a second data generating process $\{Y'_{\mathbf{k}}\}_{\mathbf{k}\in\{1,\dots,n\}^d}$ which is independent of $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\{1,\dots,n\}^d}$, using the same scheme as for $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\{1,\dots,n\}^d}$. We define

$$X^{(2)}_{\mathbf{k}} = \begin{cases} Y^2_{\mathbf{k}} + Y'^2_{\mathbf{k}}, & \mathbf{k} \notin C_n \\ 4 - (Y^2_{\mathbf{k}} + Y'^2_{\mathbf{k}}), & \mathbf{k} \in C_n. \end{cases}$$

In order to investigate the effect of the volume (Vol) of the change block on the test, we consider three different change-point settings, where $C = (\boldsymbol{\theta}, \boldsymbol{\gamma}]$ is small, medium-sized and large (cf. Table 4.1). We compare two bootstrap methods:

- Discretely sampled Ornstein-Uhlenbeck sheets (autoregressive wild bootstrap (**AR**))

  $$V_n(\mathbf{k}) = aV_n(k_1 - 1, k_2) + aV_n(k_1, k_2 - 1) - a^2 V_n(k_1 - 1, k_2 - 1) + \varepsilon_{n,k_1,k_2},$$

  with $a = \exp(-1/q(n))$ and i.i.d. $N(0,(1-a^2)^d)$-distributed innovations $\varepsilon_{n,\mathbf{k}}$. This corresponds to the exponential weight function $\omega_{q(n),\mathbf{j}} = \prod_{i=1}^d \exp\left(-\frac{|j_i|}{q(n)}\right)$.

- Moving average random fields (**MA**): Let $\{\varepsilon_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ be a random field of i.i.d. $N(0,1)$-distributed r.v. For $a = (q(n)+1)^{-d/2}$ (i.e. $a = |B_{q(n)/2}|^{-1/2}$, with $B_{q(n)/2} := \{-\frac{q(n)}{2}, \dots, \frac{q(n)}{2}\}^d$), we consider the process defined by

  $$V_n(\mathbf{k}) = a \sum_{\mathbf{j}\in B_{q(n)/2}} \varepsilon_{\mathbf{k}-\mathbf{j}}.$$

  This corresponds to the Bartlett-type weight function $\omega_{q(n),\mathbf{j}} = \prod_{i=1}^d \left(1 - \frac{|j_i|}{q(n)+1}\right)^+$.

For both methods, we consider bandwidths $q = 2, 6, 10$ and use the mean estimators $\hat{\mu} = F_n$ and

$$\hat{\mu}(\mathbf{k}) = \tilde{F}_n(\mathbf{k}) = \begin{cases} \frac{1}{\#\hat{C}_n} \sum_{\mathbf{i}\in\hat{C}_n} \mathbb{1}_{\{X_{\mathbf{i}}\le\cdot\}} & \text{if } \mathbf{k} \in \hat{C}_n, \\ \frac{1}{\#\hat{C}_n^c} \sum_{\mathbf{i}\notin\hat{C}_n} \mathbb{1}_{\{X_{\mathbf{i}}\le\cdot\}} & \text{if } \mathbf{k} \notin \hat{C}_n \end{cases}$$

(cf. section 1.2). The change-set estimator $\hat{C}_n = (\hat{\mathbf{k}}, \hat{\mathbf{m}}]$ used for $\tilde{F}_n$ is obtained by taking the maximizing values for the test statistic $T_{n,w}$ as estimators $\hat{\mathbf{k}}$ and $\hat{\mathbf{m}}$. The empirical size and power of the tests are estimated using $N = 500$ repetitions, for each of which $J = 500$ wild bootstrap-iterations are used to derive the critical values. The nominal size was chosen as $\alpha = 0.05, 0.1$.
Table 4.2 shows the empirical size of the tests. Unsurprisingly, for both choices of a the empirical size depends strongly on the bandwidth q, which is a measure of the dependence of the bootstrap process. The greater q, the greater the dependence in the bootstrap sample and the smaller the empirical size of the test. For $\hat{\mu} = F_n$ and $a = 0.2$, the nominal size is always held for $q = 10$ and can be adequately held for $q = 6$, whereas the empirical size for $a = 0.5$ tends to be greater than the nominal one even for $q = 10$. For $\hat{\mu} = \tilde{F}_n$, the empirical size is much larger than the nominal one for all choices of a and q. The over-rejection under the null hypothesis seems to be typical for bootstrap methods (cf. Doukhan et al. (2015)). Conversely, under the alternative, the empirical power decreases with rising bandwidth q, but the effect is more pronounced for $\hat{\mu} = F_n$ than for $\hat{\mu} = \tilde{F}_n$ (cf. e.g. Tables 4.3 and 4.4). This effect is however less important than the choice of change-set for the power of the test: Where both the change in mean and the change in skewness are well detected for medium-sized and large change-sets (Examples 2 and 3), the empirical power for small change-sets (Example 1) can be very small for $q = 6, 10$ (cf. Tables 4.3, 4.5, 4.7, 4.9, 4.11 and 4.13). Again, the tests based on $\tilde{F}_n$ have a higher empirical power than the tests based on F_n and retain their good detection properties even for small change-sets (cf. Tables 4.4, 4.6, 4.8, 4.10, 4.12 and 4.14). The tests perform better under weaker dependence in the observations, but for medium-sized and large change-sets the empirical power is good for both choices of a and $\Delta = 0.5$ and excellent for $\Delta = 1$. Except for small change-sets and $\hat{\mu} = F_n$ (cf. e.g. Table 4.13), the change in skewness is well detected by all procedures (cf. Tables 4.11-4.14). Rising numbers n of observations improve the empirical power of the tests. The different choices of the random variables $(V_n(\mathbf{i}))_{\underline{\mathbf{1}} \leq \mathbf{i} \leq \underline{\mathbf{n}}}$ (**AR** or **MA**) do not seem to influence the power of the test strongly, with only slightly better empirical power under **MA** for $\hat{\mu} = F_n$ (cf. e.g. Tables 4.5 and 4.7).

3.1 Conclusion

In conclusion, the simulations show that the proposed tests display the typical over-rejection property of bootstrap tests but have good empirical power against changes in the distribution. The latter is strongly influenced by the size of the set on which there is a change. While the two considered bootstrap procedures (**MA** and **AR**) show comparable results, the choice of the bandwidth has a significant effect, with smaller bandwidths leading to higher rejection rates. In comparison to $\hat{\mu} = F_n$, the estimator $\hat{\mu} = \tilde{F}_n$ has worse adherence to the nominal level under the null hypothesis but also better power against changes in mean or in the skewness. This might be due to the fact that $\tilde{F}_n$ is a more accurate estimator for the mean under the alternative but performs slightly worse under the null hypothesis.

Table 4.2: Hypothesis (stationarity)

				$n=30$			$n=40$			$n=50$		
				q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
$\hat{\mu}=F_n$	**AR**	$\alpha=0.05$	$\mathbf{a}=\mathbf{0.2}$	0.18	0.01	0.00	0.19	0.04	0.00	0.17	0.04	0.01
			$\mathbf{a}=\mathbf{0.5}$	0.58	0.08	0.00	0.65	0.14	0.03	0.67	0.15	0.04
		$\alpha=0.1$	$\mathbf{a}=\mathbf{0.2}$	0.28	0.08	0.02	0.30	0.13	0.06	0.28	0.12	0.07
			$\mathbf{a}=\mathbf{0.5}$	0.71	0.24	0.07	0.76	0.30	0.13	0.76	0.31	0.16
	MA	$\alpha=0.05$	$\mathbf{a}=\mathbf{0.2}$	0.15	0.03	0.01	0.17	0.07	0.03	0.15	0.05	0.02
			$\mathbf{a}=\mathbf{0.5}$	0.58	0.15	0.03	0.63	0.20	0.07	0.66	0.19	0.06
		$\alpha=0.1$	$\mathbf{a}=\mathbf{0.2}$	0.28	0.12	0.03	0.28	0.15	0.09	0.26	0.13	0.09
			$\mathbf{a}=\mathbf{0.5}$	0.71	0.29	0.13	0.76	0.34	0.18	0.77	0.33	0.19
$\hat{\mu}=\tilde{F}_n$	**AR**	$\alpha=0.05$	$\mathbf{a}=\mathbf{0.2}$	0.26	0.18	0.13	0.24	0.17	0.14	0.20	0.13	0.13
			$\mathbf{a}=\mathbf{0.5}$	0.68	0.40	0.31	0.71	0.38	0.29	0.71	0.35	0.26
		$\alpha=0.1$	$\mathbf{a}=\mathbf{0.2}$	0.36	0.29	0.27	0.35	0.28	0.25	0.34	0.22	0.21
			$\mathbf{a}=\mathbf{0.5}$	0.80	0.54	0.47	0.82	0.53	0.46	0.81	0.49	0.42
	MA	$\alpha=0.05$	$\mathbf{a}=\mathbf{0.2}$	0.23	0.18	0.13	0.21	0.16	0.14	0.18	0.12	0.11
			$\mathbf{a}=\mathbf{0.5}$	0.66	0.40	0.30	0.71	0.38	0.26	0.70	0.32	0.24
		$\alpha=0.1$	$\mathbf{a}=\mathbf{0.2}$	0.32	0.26	0.25	0.33	0.24	0.23	0.29	0.19	0.18
			$\mathbf{a}=\mathbf{0.5}$	0.77	0.51	0.44	0.79	0.49	0.41	0.80	0.46	0.38

Table 4.3: Change in Mean, $\hat{\mu}=F_n$, a=0.2, $\Delta=0.5$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.43	0.03	0.00	0.66	0.20	0.03	0.83	0.34	0.05
		Ex. 2	1.00	0.92	0.37	1.00	1.00	0.97	1.00	1.00	1.00
		Ex. 3	0.81	0.38	0.07	0.97	0.85	0.57	1.00	0.99	0.93
	$\alpha=0.1$	**Ex. 1**	0.57	0.21	0.04	0.78	0.43	0.19	0.91	0.61	0.30
		Ex. 2	1.00	0.99	0.91	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.90	0.66	0.42	0.99	0.94	0.87	1.00	1.00	0.99
MA	$\alpha=0.05$	**Ex. 1**	0.43	0.12	0.01	0.65	0.33	0.10	0.85	0.50	0.18
		Ex. 2	0.99	0.97	0.82	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.81	0.54	0.27	0.96	0.88	0.80	1.00	1.00	0.98
	$\alpha=0.1$	**Ex. 1**	0.56	0.28	0.10	0.79	0.51	0.29	0.91	0.72	0.45
		Ex. 2	1.00	1.00	0.96	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.88	0.74	0.56	0.99	0.94	0.90	1.00	1.00	1.00

Table 4.4: Change in Mean, $\hat{\mu} = \tilde{F}_n$, a=0.2, $\Delta = 0.5$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.56	0.38	0.32	0.75	0.59	0.52	0.89	0.73	0.64
		Ex. 2	1.00	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.88	0.75	0.66	0.98	0.94	0.91	1.00	1.00	0.99
	$\alpha = 0.1$	**Ex. 1**	0.70	0.54	0.50	0.85	0.73	0.70	0.94	0.83	0.79
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.93	0.87	0.82	0.99	0.97	0.96	1.00	1.00	1.00
MA	$\alpha = 0.05$	**Ex. 1**	0.52	0.40	0.32	0.73	0.57	0.52	0.87	0.75	0.65
		Ex. 2	0.99	0.99	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.86	0.77	0.69	0.97	0.94	0.91	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.67	0.53	0.47	0.82	0.71	0.66	0.93	0.83	0.79
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.92	0.86	0.82	0.99	0.98	0.96	1.00	1.00	1.00

Table 4.5: Change in Mean, $\hat{\mu} = F_n$, a=0.5, $\Delta = 0.5$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.70	0.12	0.00	0.79	0.24	0.05	0.84	0.32	0.09
		Ex. 2	0.95	0.53	0.09	0.98	0.83	0.54	1.00	0.96	0.88
		Ex. 3	0.80	0.25	0.03	0.90	0.46	0.19	0.96	0.68	0.44
	$\alpha = 0.1$	**Ex. 1**	0.82	0.30	0.09	0.88	0.45	0.22	0.91	0.48	0.30
		Ex. 2	0.98	0.73	0.49	0.99	0.91	0.83	1.00	0.98	0.95
		Ex. 3	0.89	0.48	0.24	0.96	0.67	0.48	0.98	0.82	0.67
MA	$\alpha = 0.05$	**Ex. 1**	0.69	0.20	0.04	0.80	0.34	0.11	0.85	0.40	0.17
		Ex. 2	0.94	0.63	0.37	0.98	0.86	0.71	1.00	0.97	0.93
		Ex. 3	0.80	0.36	0.14	0.91	0.57	0.35	0.97	0.74	0.56
	$\alpha = 0.1$	**Ex. 1**	0.79	0.35	0.16	0.87	0.52	0.28	0.91	0.53	0.36
		Ex. 2	0.97	0.77	0.60	0.99	0.92	0.86	1.00	0.98	0.96
		Ex. 3	0.88	0.54	0.33	0.95	0.69	0.55	0.98	0.84	0.72

Table 4.6: Change in Mean, $\hat{\mu} = \tilde{F}_n$, a=0.5, $\Delta = 0.5$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.79	0.46	0.38	0.84	0.57	0.47	0.87	0.56	0.46
		Ex. 2	0.97	0.82	0.74	0.98	0.93	0.89	1.00	0.99	0.97
		Ex. 3	0.86	0.58	0.48	0.93	0.70	0.61	0.98	0.83	0.74
	$\alpha = 0.1$	**Ex. 1**	0.86	0.63	0.53	0.91	0.70	0.61	0.93	0.70	0.62
		Ex. 2	0.99	0.91	0.86	1.00	0.96	0.94	1.00	0.99	0.99
		Ex. 3	0.92	0.71	0.64	0.96	0.81	0.74	0.98	0.89	0.83
MA	$\alpha = 0.05$	**Ex. 1**	0.77	0.47	0.36	0.84	0.55	0.45	0.87	0.54	0.45
		Ex. 2	0.97	0.82	0.74	0.98	0.93	0.89	1.00	0.98	0.97
		Ex. 3	0.84	0.58	0.47	0.93	0.69	0.61	0.97	0.81	0.72
	$\alpha = 0.1$	**Ex. 1**	0.85	0.61	0.51	0.90	0.68	0.60	0.94	0.68	0.59
		Ex. 2	0.98	0.90	0.85	1.00	0.96	0.94	1.00	0.99	0.99
		Ex. 3	0.91	0.71	0.61	0.96	0.79	0.73	0.98	0.88	0.83

Table 4.7: Change in Mean, $\hat{\mu} = F_n$, a=0.2, $\Delta = 1$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.90	0.04	0.00	1.00	0.32	0.01	1.00	0.81	0.02
		Ex. 2	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	0.78	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.97	0.39	0.03	1.00	0.87	0.23	1.00	1.00	0.54
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha=0.05$	**Ex. 1**	0.92	0.20	0.01	1.00	0.80	0.10	1.00	1.00	0.39
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.97	0.58	0.13	1.00	0.98	0.59	1.00	1.00	0.97
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.8: Change in Mean, $\hat{\mu} = \tilde{F}_n$, a=0.2, $\Delta = 1$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.97	0.87	0.72	1.00	0.99	0.95	1.00	1.00	1.00
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.99	0.95	0.91	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha=0.05$	**Ex. 1**	0.97	0.87	0.73	1.00	1.00	0.97	1.00	1.00	1.00
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.99	0.94	0.90	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.9: Change in Mean, $\hat{\mu} = F_n$, a=0.5, $\Delta = 1$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.86	0.15	0.00	0.97	0.39	0.04	0.99	0.59	0.10
		Ex. 2	1.00	1.00	0.75	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.76	0.26	1.00	0.98	0.86	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.93	0.43	0.11	0.99	0.66	0.30	1.00	0.83	0.46
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.92	0.75	1.00	1.00	0.98	1.00	1.00	1.00
MA	$\alpha=0.05$	**Ex. 1**	0.86	0.30	0.04	0.97	0.59	0.16	0.99	0.77	0.33
		Ex. 2	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.88	0.59	1.00	0.99	0.95	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.93	0.53	0.23	0.99	0.75	0.48	0.99	0.90	0.65
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.95	0.85	1.00	1.00	0.99	1.00	1.00	1.00

Table 4.10: Change in Mean, $\hat{\mu} = \tilde{F}_n$, a=0.5, $\Delta = 1$

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.93	0.67	0.56	0.98	0.83	0.73	0.99	0.92	0.83
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.95	0.88	1.00	1.00	0.98	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.96	0.80	0.73	1.00	0.92	0.87	1.00	0.98	0.91
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.99	0.96	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha = 0.05$	**Ex. 1**	0.92	0.66	0.53	0.98	0.83	0.73	0.99	0.93	0.82
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.95	0.89	1.00	1.00	0.99	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.96	0.80	0.70	0.99	0.92	0.84	1.00	0.98	0.93
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.99	0.96	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.11: Change in Skewness, $\hat{\mu} = F_n$, a=0.2

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.28	0.01	0.00	0.80	0.15	0.00	0.99	0.56	0.02
		Ex. 2	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.91	0.37	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.49	0.15	0.02	0.93	0.55	0.11	1.00	0.94	0.36
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.99	0.92	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha = 0.05$	**Ex. 1**	0.31	0.07	0.00	0.82	0.41	0.04	0.99	0.89	0.23
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.98	0.83	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.51	0.23	0.07	0.93	0.73	0.31	1.00	0.98	0.80
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.12: Change in Skewness, $\hat{\mu} = \tilde{F}_n$, a=0.2

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha = 0.05$	**Ex. 1**	0.45	0.35	0.28	0.88	0.80	0.69	0.99	0.98	0.93
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	0.99	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.62	0.54	0.50	0.95	0.91	0.87	1.00	0.99	0.99
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha = 0.05$	**Ex. 1**	0.43	0.36	0.28	0.88	0.82	0.74	0.99	0.98	0.96
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha = 0.1$	**Ex. 1**	0.60	0.54	0.47	0.95	0.92	0.88	1.00	0.99	0.99
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4.13: Change in Skewness, $\hat{\mu} = F_n$, a=0.5

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.38	0.03	0.00	0.68	0.15	0.00	0.85	0.32	0.05
		Ex. 2	1.00	1.00	0.82	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.89	0.50	0.11	1.00	0.96	0.82	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.53	0.16	0.04	0.82	0.38	0.13	0.95	0.64	0.27
		Ex. 2	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.96	0.78	0.55	1.00	1.00	0.97	1.00	1.00	1.00
MA	$\alpha=0.05$	**Ex. 1**	0.36	0.09	0.01	0.67	0.26	0.06	0.83	0.52	0.18
		Ex. 2	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.88	0.63	0.37	1.00	0.98	0.93	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.51	0.22	0.07	0.81	0.49	0.22	0.93	0.72	0.44
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.96	0.83	0.67	1.00	1.00	0.99	1.00	1.00	1.00

Table 4.14: Change in Skewness, $\hat{\mu} = \tilde{F}_n$, a=0.5

			$n=30$			$n=40$			$n=50$		
			q=2	q=6	q=10	q=2	q=6	q=10	q=2	q=6	q=10
AR	$\alpha=0.05$	**Ex. 1**	0.50	0.30	0.23	0.76	0.54	0.41	0.89	0.72	0.64
		Ex. 2	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.92	0.78	0.67	1.00	0.99	0.96	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.65	0.46	0.39	0.87	0.73	0.65	0.96	0.84	0.79
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.97	0.89	0.83	1.00	1.00	1.00	1.00	1.00	1.00
MA	$\alpha=0.05$	**Ex. 1**	0.45	0.30	0.24	0.73	0.54	0.44	0.87	0.72	0.64
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.92	0.79	0.70	1.00	1.00	0.97	1.00	1.00	1.00
	$\alpha=0.1$	**Ex. 1**	0.61	0.45	0.36	0.85	0.70	0.65	0.95	0.84	0.77
		Ex. 2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
		Ex. 3	0.96	0.88	0.84	1.00	1.00	1.00	1.00	1.00	1.00

4 Proofs

4.1 Preliminary results

Lemma 4.1. *Let $\{X_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{N}^d}$ be an H-valued centered random field with $\lim_{\tau\to\infty} \rho_{\mathbb{R}}(\tau) < 1$. Then for any $r \geq 2$, there exists a positive constant $B_{d,r}$ depending only on r, d and $\rho_{\mathbb{R}}(\cdot)$ such that for any finite set $S \subset \mathbb{N}^d$,*

$$E\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r \leq B_{d,r}\left(\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^r + \left(\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^2\right)^{r/2}\right). \tag{5}$$

If $\sup_{\mathbf{k}\in\mathbb{N}^d} E\|X_{\mathbf{k}}\|^r < \infty$, this implies

$$E\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r \leq B_{d,r}\left(\sup_{\mathbf{k}\in\mathbb{N}^d} E\|X_{\mathbf{k}}\|^r + (\sup_{\mathbf{k}\in\mathbb{N}^d} E\|X_{\mathbf{k}}\|^2)^{r/2}\right)(\#S)^{r/2} =: C(d,r,X)(\#S)^{r/2}. \tag{6}$$

We say a block W in $\mathbb{Z}^d$ belongs standardly to a block U and denote this by $W \lhd U$ whenever $W \subset U$ and the minimal vertices of W and U (in the sense of the lexicographic order) coincide. If (6) holds for $r > 2$ and blocks S in $\mathbb{Z}^d$, U is any block in $\mathbb{Z}^d$, and

$$M(U) = \max_{W\lhd U}\left\|\sum_{\mathbf{j}\in W} X_{\mathbf{j}}\right\|,$$

then (6) implies

$$E\left(M(U)^r\right) \leq \tilde{C}C(d,r,X)(\#U)^{r/2} \tag{7}$$

with $\tilde{C} = \left(\frac{5}{2}\right)^d (1 - 2^{(1-\frac{r}{2})/r})^{-dr}$.

Remark 4.1. *For H-valued processes, an alternative definition of ρ-mixing is given by the coefficients*

$$\rho_H(\mathcal{A},\mathcal{B}) = \sup\left\{\frac{|E(\langle X,Y\rangle) - \langle EX,EY\rangle|}{\|X\|_2\|Y\|_2} : X \in L^2(\mathcal{A},H), Y \in L^2(\mathcal{B},H), \|X\|_2, \|Y\|_2 > 0\right\}.$$

Analogously to the real-valued case, one can then define $\rho_H(r)$ and $\rho_H^*(r)$ for random fields. As shown in Bradley and Bryc (1985), Theorem 4.2, the coefficients ρ_H and $\rho_{\mathbb{R}}$ coincide and therefore $\rho_H(\cdot) = \rho_{\mathbb{R}}(\cdot)$ and $\rho_H^*(\cdot) = \rho_{\mathbb{R}}^*(\cdot)$.

Proof. For $\rho_{\mathbb{R}}^* = \rho_H^*$ instead of $\rho_{\mathbb{R}} = \rho_H$, (5) is Theorem 2 of Zhang (1998). Since the two definitions of mixing coincide for $d = 1$, we can use Theorem 2 of Zhang (1998) for the one-dimensional case and obtain (5) by induction over d (cf. Bradley (2007), Volume III, p.234). For any $j \in \mathbb{N}$, define sets $S(j) = \{\mathbf{k} \in S : k_1 = j\}$, $T(j) = \{\mathbf{k} \in \mathbb{N}^d : k_1 = j\}$ and $Y_j = \sum_{\mathbf{k}\in S(j)} X_{\mathbf{k}}$ if $S(j) \neq \emptyset$ and $Y_j = 0$ otherwise. Then $\{Y_j\}_{j\in\mathbb{N}}$ satisfies $\rho_{\mathbb{R},Y}(\tau) \leq \rho_{\mathbb{R},X}(\tau)$. The random field $\zeta^{(j)} = \{X_{\mathbf{k}} : \mathbf{k} \in T(j)\}$ can be viewed as a $(d-1)$-parameter field with $\rho_{\mathbb{R},\zeta^{(j)}}(\tau) \leq \rho_{\mathbb{R},X}(\tau)$ since $T(j) \cong \mathbb{N}^{d-1}$. Now, with $N(S) = \{j \in \mathbb{N} : S(j) \neq \emptyset\}$, it holds that

$$E\left\|\sum_{\mathbf{k}\in S} X_{\mathbf{k}}\right\|^r = E\left\|\sum_{j\in N(S)} Y_j\right\|^r$$

$$\leq B_{1,r}\left(\sum_{j\in N(S)} E\|Y_j\|^r + \left(\sum_{j\in N(S)} E\|Y_j\|^2\right)^{r/2}\right)$$

$$= B_{1,r}\sum_{j\in N(S)} E\left\|\sum_{\mathbf{k}\in S(j)} \zeta_{\mathbf{k}}^{(j)}\right\|^r + B_{1,r}\left(\sum_{j\in N(S)} E\left\|\sum_{\mathbf{k}\in S(j)} \zeta_{\mathbf{k}}^{(j)}\right\|^2\right)^{r/2}$$

$$\overset{\text{I.H.}}{\leq} B_{1,r}\sum_{j\in N(S)} B_{d-1,r}\left[\sum_{\mathbf{k}\in S(j)} E\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^r + \left(\sum_{\mathbf{k}\in S(j)} E\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2\right)^{r/2}\right]$$

$$+ B_{1,r}\left(\sum_{j\in N(S)} B_{d-1,2}\left[\sum_{\mathbf{k}\in S(j)} E\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2 + \left(\sum_{\mathbf{k}\in S(j)} E\left\|\zeta_{\mathbf{k}}^{(j)}\right\|^2\right)^{2/2}\right]\right)^{r/2}$$

$$= B_{1,r}B_{d-1,r}\left[\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^r + \sum_{j\in N(S)}\left(\sum_{\mathbf{k}\in S(j)} E\|X_{\mathbf{k}}\|^2\right)^{r/2}\right]$$

$$+ 2^{r/2}B_{1,r}B_{d-1,2}^{r/2}\left(\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^2\right)^{r/2}$$

$$\leq (B_{1,r}B_{d-1,r} + 2^{r/2}B_{d-1,2}^{r/2}B_{1,r})\left[\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^r + \left(\sum_{\mathbf{k}\in S} E\|X_{\mathbf{k}}\|^2\right)^{r/2}\right]$$

where the inequality $(\sum_{k=1}^m a_k)^q \geq \sum_{k=1}^m a_k^q$ (A2902 in Bradley (2007), Volume III) is used to obtain the last inequality.

(6) is a trivial consequence of (5). If (6) holds for some $r > 2$, (7) follows from Corollary 1 in Móricz (1983) (cf. also Bulinski and Shashkin (2007), Theorem 2.1.2). (The Corollary can be applied in any normed space without changing the proof.) □

Following an approach that is similar in spirit to Davidson (2002) (cf. Theorems 29.6 and 29.18), we aim to reduce the multivariate functional central limit theorem to the corresponding results for the univariate case. For real-valued processes, Deo (1975) gave a version for random fields of Theorems 19.1 and 19.2 of Billingsley (1968), which use a characterization of Brownian motion to obtain a general functional central limit theorem (cf. Lemmas 2 and 3 in Deo (1975)). We extend this result to multivariate random fields by taking advantage of the fact that Gaussian random vectors can be characterized by their behavior under projections.

Lemma 4.2. *Let Σ be a symmetric positive semidefinite matrix and $S_n = \{S_n(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ a sequence of stochastic processes with sample paths in $D_{\mathbb{R}^k}([0,1]^d)$, such that*

(i) $ES_n(\mathbf{t}) \to \underline{0}$ *and* $\mathrm{Cov}S_n(\mathbf{t}) \to [\mathbf{t}]\Sigma$ *as* $n \to \infty$, *for each* $\mathbf{t} \in [0,1]^d$,

(ii) the set $\{\|S_n(\mathbf{t})\|^2\}_n$ *is uniformly integrable for each* $\mathbf{t}$,

(iii) if $B_1, B_2, \ldots, B_p$ *is a collection of strongly separated blocks, then the increments* $S_n(B_1), S_n(B_2), \ldots, S_n(B_p)$ *are asymptotically independent in the sense that if*

$H_1, H_2, \ldots, H_p$ are arbitrary Borel sets in $\mathbb{R}^k$, then the difference

$$P(S_n(B_1) \in H_1, \ldots, S_n(B_p) \in H_p) - \prod_{i=1}^{p} P(S_n(B_i) \in H_i)$$

goes to zero as $n \to \infty$ and,

(iv) for each $\varepsilon > 0$, $\eta > 0$, we can find a $\delta > 0$ such that

$$P(w^k(S_n, \delta) > \varepsilon) < \eta$$

for all sufficiently large n, where we define the modulus of continuity

$$w^k(x; \delta) := \sup\{\|x(\mathbf{t}) - x(\mathbf{s})\| : \|\mathbf{t} - \mathbf{s}\| \leq \delta\},$$

for $x \in D_{\mathbb{R}^k}([0,1]^d)$ and $0 < \delta < 1$.

Then S_n converges weakly in $D_{\mathbb{R}^k}([0,1]^d)$ to the k-dimensional Brownian sheet on $[0,1]^d$ with covariance matrix Σ.

Proof. Consider $\boldsymbol{\lambda} \in \mathbb{R}^k$ and define $\{S_n^{\boldsymbol{\lambda}}(\mathbf{t})\}_{\mathbf{t} \in [0,1]^d}$ by $S_n^{\boldsymbol{\lambda}}(\mathbf{t}) = \boldsymbol{\lambda}^t S_n(\mathbf{t})$. First, note that for any $x \in D_{\mathbb{R}^k}([0,1]^d)$, $\mathbf{t} \in [0,1]^d$ and $\boldsymbol{\lambda} \in \mathbb{R}^k$, it holds that if $\mathbf{t}_n \xrightarrow{\rho} \mathbf{t}$ (cf. notations in Neuhaus (1969)), then $x(\mathbf{t} + 0_\rho) = \lim_{n\to\infty} x(\mathbf{t}_n)$ exists, and since $\mathbf{y} \mapsto \boldsymbol{\lambda}^t \mathbf{y}$ is a continuous map, it follows that

$$\lim_{n\to\infty} \boldsymbol{\lambda}^t x(\mathbf{t}_n) = \boldsymbol{\lambda}^t \lim_{n\to\infty} x(\mathbf{t}_n) = \boldsymbol{\lambda}^t x(\mathbf{t} + 0_\rho)$$

also exists. Therefore, if $x \in D_{\mathbb{R}^k}([0,1]^d)$, then $\boldsymbol{\lambda}^t x \in D_{\mathbb{R}}([0,1]^d)$ and if $x \in C_{\mathbb{R}^k}([0,1]^d)$, then $\boldsymbol{\lambda}^t x \in C_{\mathbb{R}}([0,1]^d)$. Furthermore, since $D_{\mathbb{R}^k}([0,1]^d) \to D_{\mathbb{R}}([0,1]^d)$, $x \mapsto \boldsymbol{\lambda}^t x$, is a continuous map, $S_n^{\boldsymbol{\lambda}}$ are random elements in $D_{\mathbb{R}}([0,1]^d)$. Assumptions $(i) - (iii)$ imply:

(i) $ES_n^{\boldsymbol{\lambda}}(\mathbf{t}) = \boldsymbol{\lambda}^t ES_n(\mathbf{t}) \to 0$, $\mathrm{Cov}(\boldsymbol{\lambda}^t S_n(\mathbf{t})) = \boldsymbol{\lambda}^t \mathrm{Cov}(S_n(\mathbf{t}))\boldsymbol{\lambda} \to [\mathbf{t}]\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda}$ for any $\mathbf{t} \in [0,1]^d$

(ii) $\{|S_n^{\boldsymbol{\lambda}}(\mathbf{t})|^2\}_{n\geq 1}$ is uniformly integrable for each $\mathbf{t}$, since due to the Cauchy-Schwarz inequality $\left|\boldsymbol{\lambda}^t S_n(\mathbf{t})\right|^2 \leq \|\boldsymbol{\lambda}\|^2 \|S_n(\mathbf{t})\|^2$.

(iii) For arbitrary linear Borel sets $H_1, \ldots, H_p$, the sets $f_{\boldsymbol{\lambda}}^{-1}(H_1), \ldots, f_{\boldsymbol{\lambda}}^{-1}(H_p)$ where $f_{\boldsymbol{\lambda}}$ is the continuous map $f_{\boldsymbol{\lambda}} : \mathbb{R}^k \to \mathbb{R}$, $\mathbf{x} \mapsto \boldsymbol{\lambda}^t \mathbf{x}$, lie in $\mathcal{B}(\mathbb{R}^k)$. Therefore, for any collection of strongly separated blocks $B_1, \ldots, B_p$,

$$\begin{aligned}&P(S_n^{\boldsymbol{\lambda}}(B_1) \in H_1, \ldots, S_n^{\boldsymbol{\lambda}}(B_p) \in H_p) - \prod_{i=1}^{p} P(S_n^{\boldsymbol{\lambda}}(B_i) \in H_i)\\ =&P(S_n(B_1) \in f_{\boldsymbol{\lambda}}^{-1}(H_1), \ldots, S_n(B_p) \in f_{\boldsymbol{\lambda}}^{-1}(H_p)) - \prod_{i=1}^{p} P(S_n(B_i) \in f_{\boldsymbol{\lambda}}^{-1}(H_i))\end{aligned}$$

goes to zero as $n \to \infty$ by assumption (iii).

(iv) Since by the Cauchy-Schwarz inequality

$$|\boldsymbol{\lambda}^t(S_n(\mathbf{t}) - S_n(\mathbf{s}))| \leq \|\boldsymbol{\lambda}\| \|S_n(\mathbf{t}) - S_n(\mathbf{s})\|,$$

it trivially holds that:

$$\forall \varepsilon > 0, \eta > 0 \, \exists \delta > 0: \; P(\omega^1(S_n^{\boldsymbol{\lambda}}, \delta) > \varepsilon \|\boldsymbol{\lambda}\|) \leq P(\omega^k(S_n, \delta) > \varepsilon) < \eta$$

Therefore, if $\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda} > 0$, $(\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda})^{-1/2} S_n^{\boldsymbol{\lambda}}$ fulfills the conditions of Lemma 3 in Deo (1975) and thus converges to a standardized Brownian sheet $(\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda})^{-1/2} W^{\boldsymbol{\lambda}}$ in $D_{\mathbb{R}}([0,1]^d)$. By continuous mapping, this implies $S_n^{\boldsymbol{\lambda}} \Rightarrow W^{\boldsymbol{\lambda}}$ in $D_{\mathbb{R}}([0,1]^d)$. If $\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda} = 0$, the processes $S_n^{\boldsymbol{\lambda}} \equiv 0$ and $W^{\boldsymbol{\lambda}} \equiv 0$ are both degenerated and therefore $S_n^{\boldsymbol{\lambda}} \Rightarrow W^{\boldsymbol{\lambda}}$ holds trivially.
In particular, every coordinate process $S_n^i = S_n^{e_i}$ (where $e_i \in \mathbb{R}^k$ is the vector with one in position i and zero elsewhere ($i \in \{1, \ldots, k\}$)) is tight in $D_{\mathbb{R}}([0,1]^d)$ and thus for any $\varepsilon > 0$, we can find $M_\varepsilon \in (0, \infty)$ such that

$$P(\|S_n\|_\infty > M_\varepsilon) \leq \sum_{i=1}^k P(\|S_n^i\|_\infty > M_\varepsilon) \leq \varepsilon \quad \forall n \in \mathbb{N}.$$

Therefore, assumption (iv) implies that S_n is tight in $D_{\mathbb{R}^k}([0,1]^d)$. Now, consider a convergent subsequence $S_{n'}$, say $S_{n'} \Rightarrow W$. Then the continuity of the mappings $D_{\mathbb{R}^k}([0,1]^d) \to D_{\mathbb{R}}([0,1]^d)$, $x \mapsto \boldsymbol{\lambda}^t x$, for any $\boldsymbol{\lambda} \in \mathbb{R}^k$ implies $S_{n'}^{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^t S_{n'} \Rightarrow \boldsymbol{\lambda}^t W = W^{\boldsymbol{\lambda}}$, where $W^{\boldsymbol{\lambda}}$ is a Brownian sheet in $D_{\mathbb{R}}([0,1]^d)$ with covariance $\boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda} \geq 0$. In order to show that S_n converges in $D_{\mathbb{R}^k}([0,1]^d)$, it suffices to show that W (and therefore any limit of a convergent subsequence) is indeed the Brownian sheet in $H = \mathbb{R}^k$. Denote the coordinate processes by $W^i = W^{e_i}$. Since this holds for all the coordinate processes, W is a.s. continuous and $W(\mathbf{t}) = \underline{\mathbf{0}}$ a.s. for any $\mathbf{t} \in [0,1]^d$ with $[\mathbf{t}] = 0$.
The increments $W(B)$ of W have a Gaussian distribution with mean zero and covariance $\lambda(B)\Sigma$ in $\mathbb{R}^k$, since $W(B) = (W^1(B), \ldots, W^k(B))^t$ and $\sum_{i=1}^k \lambda_i W^i(B) = W^{\boldsymbol{\lambda}}(B)$ is a centered Gaussian random variable with variance $\lambda(B) \boldsymbol{\lambda}^t \Sigma \boldsymbol{\lambda}$ for any $\boldsymbol{\lambda} \in \mathbb{R}^k$. In particular, the distribution of $W(B)$ is absolutely continuous, so that for any collection of strongly separated blocks $B_1, \ldots, B_p$ and any $y_1, \ldots, y_p \in \mathbb{R}^k$, we have

$$P(S_{n'}(B_j) \leq y_j) \to P(W(B_j) \leq y_j) \quad (j = 1, \ldots, p)$$

and therefore

$$\begin{aligned}
&P(W(B_1) \leq y_1, \ldots, W(B_p) \leq y_p) \\
&= \lim_{n' \to \infty} P(S_{n'}(B_1) \leq y_1, \ldots, S_{n'}(B_p) \leq y_p) \\
&= \lim_{n' \to \infty} \left(P(S_{n'}(B_1) \leq y_1, \ldots, S_{n'}(B_p) \leq y_p) - \prod_{j=1}^p P(S_{n'}(B_j) \leq y_j) \right) \\
&\quad + \prod_{j=1}^p \lim_{n' \to \infty} P(S_{n'}(B_j) \leq y_j) \\
&\overset{(iii)}{=} \prod_{j=1}^p P(W(B_j) \leq y_j).
\end{aligned}$$

(We have used the fact that since $W \in C_{\mathbb{R}^k}([0,1]^d)$ a.s., the projection maps $\pi_{\mathbf{t}_1,\dots,\mathbf{t}_l}$ are P_W-a.s. continuous and therefore $S_{n'} \Rightarrow W$ implies the convergence of the finite dimensional distributions. Since for a block $B = (\mathbf{s},\mathbf{t}]$,

$$S_{n'}(B) = \sum_{\boldsymbol{\varepsilon}\in\{0,1\}^d} (-1)^{d-\sum_{j=1}^d \varepsilon_j} S_{n'}(\mathbf{s}+\boldsymbol{\varepsilon}(\mathbf{t}-\mathbf{s})),$$

this implies the weak convergence of the increments.)
Note that due to the a.s. continuity of W, this also yields the independence of the increments over any (not necessarily strongly separated) collection of pairwise disjoint blocks. □

Lemma 4.3. *Let $\{X_\mathbf{j}\}_{\mathbf{j}\in\mathbb{Z}^d}$ be an $\mathbb{R}^k$-valued $\rho_\mathbb{R}$-mixing, weakly stationary centered random field, $\{S_n(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ a process in $D_{\mathbb{R}^k}([0,1]^d)$ with*

$$S_n(\mathbf{t}) = n^{-d/2} \sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\lfloor n\mathbf{t}\rfloor} X_\mathbf{j}$$

and $\Sigma(n,\mathbf{t}) = \operatorname{Cov}(S_n(\underline{\mathbf{t}}))$. If

(i) $\sup_{\mathbf{j}\in\mathbb{Z}^d} E\|X_\mathbf{j}\|^{2+\delta} < \infty$ *for some* $\delta > 0$ *and*

(ii) $\sum_{m\ge1} m^{d-1}\alpha_{1,1}(m)^{\delta/(2+\delta)} < \infty$,

then $\Sigma(n,\mathbf{t}) \to [\mathbf{t}]\Sigma$ for any $\mathbf{t}\in[0,1]^d$ and a positive semidefinite matrix $\Sigma = (\sigma_{i,j})_{1\le i,j\le k}$ with $\sigma_{i,j} = \sum_{\mathbf{v}\in\mathbb{Z}^d}\gamma_{i,j}(\mathbf{v})$, where $\gamma_{i,j}(\mathbf{v}) = \operatorname{Cov}(X^i_{\underline{\mathbf{0}}}, X^j_\mathbf{v})$, and the series converges absolutely. Furthermore, $\{S_n(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ converges in $D_{\mathbb{R}^k}([0,1]^d)$ to a k-dimensional Brownian sheet with covariance matrix Σ.

Proof. As remarked by Guyon (1995) (p. 109 f.), for any $i,j\in\{1,\dots,k\}$ the covariance inequality (cf. Doukhan (1994), Theorem 3)

$$|\gamma_{i,j}(\mathbf{v})| = |\operatorname{Cov}(X^i_{\underline{\mathbf{0}}}, X^j_\mathbf{v})| \le 10\alpha_{1,1}(\|\mathbf{v}\|_\infty)^{\delta/(2+\delta)}\|X_{\underline{\mathbf{0}}}\|^2_{2+\delta}$$

together with assumptions (i) and (ii) implies $\sum_{\mathbf{v}\in\mathbb{Z}^d}|\gamma_{i,j}(\mathbf{v})| < \infty$. Using this and the dominated convergence theorem, we obtain

$$\sum_{-\lfloor n\mathbf{t}\rfloor\le\mathbf{v}\le\lfloor n\mathbf{t}\rfloor} \gamma_{i,j}(\mathbf{v}) = \sum_{\underline{\mathbf{n}}\le\mathbf{v}\le\underline{\mathbf{n}}} I_{\{|\mathbf{v}|\le\lfloor n\mathbf{t}\rfloor\}}\gamma_{i,j}(\mathbf{v}) \overset{n\to\infty}{\longrightarrow} \sigma_{i,j}$$

for any $\mathbf{t}\in[0,1]^d$. Furthermore,

$$\begin{aligned}
\Sigma(n,\mathbf{t})^{(i,j)} &= n^{-d}\operatorname{Cov}\left(\sum_{\underline{\mathbf{1}}\le\mathbf{m}\le\lfloor n\mathbf{t}\rfloor} X^i_\mathbf{m}, \sum_{\underline{\mathbf{1}}\le\mathbf{m}'\le\lfloor n\mathbf{t}\rfloor} X^j_{\mathbf{m}'}\right)\\
&= n^{-d}\sum_{-\lfloor n\mathbf{t}\rfloor<\mathbf{v}<\lfloor n\mathbf{t}\rfloor}\gamma_{i,j}(\mathbf{v})\prod_{l=1}^d(\lfloor nt_l\rfloor - |v_l|)\\
&= n^{-d}\prod_{i=1}^d\lfloor nt_i\rfloor\sum_{-\lfloor n\mathbf{t}\rfloor<\mathbf{v}<\lfloor n\mathbf{t}\rfloor}\gamma_{i,j}(\mathbf{v})\\
&\quad+\sum_{h=1}^d\binom{d}{h}(-1)^h\sum_{\substack{-\lfloor n\mathbf{t}\rfloor<\mathbf{v}<\lfloor n\mathbf{t}\rfloor,\\ \mathbf{v}\ne\underline{\mathbf{0}}}}\gamma_{i,j}(\mathbf{v})\sum_{\substack{I\subseteq\{1,\dots,d\},\\ |I|=h}}\prod_{l\in I^c}\frac{\lfloor nt_l\rfloor}{n}\prod_{l\in I}\frac{|v_l|}{n},
\end{aligned}$$

where analogous arguments to the proof of Lemma 3 in Berkes and Morrow (1981) can be used to show that the last sum goes to zero. Therefore, $\Sigma(n, \mathbf{t})^{(i,j)} \stackrel{n\to\infty}{\longrightarrow} [\mathbf{t}]\sigma_{i,j}$. Now, it remains to show that the matrix Σ is positive semidefinite. For any vector $\mathbf{u} \in \mathbb{R}^k$, applying the statement just proven to the real-valued random field $\{\mathbf{u}^t X_\mathbf{m}\}_{\mathbf{m}\in\mathbb{Z}^d}$ yields:

$$\mathbf{u}^t\Sigma\mathbf{u} = \lim_{n\to\infty} n^{-d} E\left(\sum_{\mathbf{1}\le\mathbf{m}\le\lfloor n\mathbf{t}\rfloor} \mathbf{u}^t X_\mathbf{m}\right)^2 \ge 0$$

Therefore, Σ is positive semidefinite and symmetric.
We show that Lemma 4.2 can be applied to obtain the stated convergence. First, note that condition (i) of Lemma 4.2 is fulfilled, since $\{X_\mathbf{j}\}_{\mathbf{j}\in\mathbb{Z}^d}$ is centered and $\Sigma(n,\mathbf{t}) \stackrel{n\to\infty}{\longrightarrow} [\mathbf{t}]\Sigma$.
The assumptions imply the moment inequality (6) from Lemma 4.1. Therefore, condition (ii) follows from

$$\sup_{n\ge1} E\|S_n(\mathbf{t})\|^{2+\delta} \le [\mathbf{t}]^{1+\delta/2} C(r,d,X) \le C(r,d,X) < \infty$$

for any $\mathbf{t}$.
For strongly separated blocks $B_1 = (\mathbf{s}_1, \mathbf{t}_1], \dots, B_q = (\mathbf{s}_q, \mathbf{t}_q]$, there is an $i \in \{1,\dots,d\}$ such that $0 \le s_1^i \le t_1^i < s_2^i \le t_2^i < \dots < s_q^i \le t_q^i \le 1$ (after reordering the blocks if necessary), i.e. $\min_{j=1,\dots,q-1}(s_{j+1}^i - t_j^i) > 0$, and therefore $\min_{j=1,\dots,q-1}(\lfloor ns_{j+1}^i\rfloor - \lfloor nt_j^i\rfloor) \to \infty$ for $n \to \infty$. Then

$$\begin{aligned}
&P\left(\bigcap_{j=1}^q \{S_n(B_j) \in H_j\}\right) - \prod_{j=1}^q P(S_n(B_j) \in H_j) \\
=&P\left(\left\{\bigcap_{j=1}^{q-1}\{S_n(B_j)\in H_j\}\right\} \cap \{S_n(B_q)\in H_q\}\right) \\
&- P\left(\bigcap_{j=1}^{q-1}\{S_n(B_j)\in H_j\}\right) P(S_n(B_q)\in H_q) \\
&+ P(S_n(B_q)\in H_q)\left[P\left(\left\{\bigcap_{j=1}^{q-2}\{S_n(B_j)\in H_j\}\right\}\cap\{S_n(B_{q-1})\in H_{q-1}\}\right)\right. \\
&\left.-P\left(\bigcap_{j=1}^{q-2}\{S_n(B_j)\in H_j\}\right)P(S_n(B_{q-1})\in H_{q-1})\right] \\
&+P(S_n(B_q)\in H_q)P(S_n(B_{q-1})\in H_{q-1})\left[P\left(\left\{\textstyle\bigcap_{j=1}^{q-3}\{S_n(B_j)\in H_j\}\right\}\cap\{S_n(B_{q-2})\in H_{q-2}\}\right)\right. \\
&\left.-P(S_n(B_1)\in H_1,\dots,S_n(B_{q-3})\in H_{q-3})P(S_n(B_{q-2})\in H_{q-2})\right] \\
&+\dots+\prod_{j=1}^q P(S_n(B_j)\in H_j) - \prod_{j=1}^q P(S_n(B_j)\in H_j) \\
\le& q\,\rho_\mathbb{R}(\min_{j=1,\dots,q-1}(\lfloor ns_{j+1}^i\rfloor - \lfloor nt_j^i\rfloor)) \stackrel{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

Finally, using (the proof of) Theorem 5.1.3 in Bulinski and Shashkin (2007), we will now show that condition (iv) of the Lemma is implied by (6). As noted in Lemma 4.1, (6) together with assumption (i) imply (7) for any block U. Analogously to the proof of condition (ii), this implies the uniform integrability of $\{(\#U_n)^{-1}M(U_n)^2\}_{n\geq 1}$ for any sequence of blocks U_n growing to infinity. The proof of Theorem 5.1.3 in Bulinski and Shashkin (2007) therefore shows (iv). □

The following corollary of Theorem 4.2 in Billingsley (1968) is an adaptation of Lemma 4.1 in Chen and White (1998) to multiparameter processes.

Lemma 4.4. *Let $\{X_n : n \geq 1\}$ be a sequence of $D_H([0,1]^d)$-random elements and for any $k \in \mathbb{N}$, let $X^{(k)}$ be a Brownian sheet in H_k with $EX^{(k)}(\underline{\mathbf{1}}) = 0$ and $\mathrm{Cov}X^{(k)}(\underline{\mathbf{1}}) = S^k$. Suppose the following conditions hold:*

(a) For each $k \geq 1$, $P_k X_n \Rightarrow X^{(k)}$ in $D_{H_k}([0,1]^d)$ as $n \to \infty$;

(b) $X^{(k)} \Rightarrow X$ in $D_H([0,1]^d)$ as $k \to \infty$;

(c) $\limsup_{n\to\infty} E\left(\sup_{\mathbf{t}\in[0,1]^d} \|X_n(\mathbf{t}) - P_k X_n(\mathbf{t})\|^r\right) \to 0$ as $k \to \infty$ for some $r \geq 2$.

Then $X_n \Rightarrow X$ in $D_H([0,1]^d)$, where X is a Brownian sheet in H with $EX(\underline{\mathbf{1}}) = 0$ and $\mathrm{Cov}X(\underline{\mathbf{1}}) = S$ for $S = \lim_{k\to\infty} S^k$.

Now, we give some preliminary results needed for the proof of Theorem 2.2. In the next two lemmas, we will establish a Rosenthal inequality for the bootstrapped partial sum process.

Lemma 4.5. *Let X, Y be random variables taking values in a Hilbert space H_1, X is $\mathcal{F}$-measurable and Y is $\mathcal{G}$-measurable. Let V be a random variable which is independent of $\sigma(\mathcal{F},\mathcal{G})$ and takes values in a Hilbert space H_2. Furthermore, let $g, h : H_1 \times H_2 \to H$ be measurable functions with*

$$E\left[g(X,V)\middle|V\right] = E\left[h(Y,V)\middle|V\right] = 0 \quad a.s.$$

If $\rho = \rho_{\mathbb{R}}(\mathcal{F},\mathcal{G}) < 1$, then for any $p > 1$ such that $E[\|g(X,V)\|^p] < \infty$ and $E[\|h(Y,V)\|^p] < \infty$, there exists a constant $C_{\rho,p}$ such that

$$E\left[\left\|g(X,V)\right\|^p\right] \leq C_{\rho,p} E\left[\left\|g(X,V) + h(Y,V)\right\|^p\right].$$

Proof. We will make use of the conditional expectations

$$E\left[\left\|g(X,V)\right\|^p\middle|V=v\right] = E[\|g(X,v)\|^p] \quad \text{and} \quad E\left[\left\|h(Y,V)\right\|^p\middle|V=v\right] = E[\|h(Y,v)\|^p].$$

$g(X,v)$ and $h(Y,v)$ are H-valued random variables which are $\mathcal{F}$- and $\mathcal{G}$-measurable, respectively. So we can apply Theorem 1 of Zhang (1998) to the conditional expectations and obtain

$$E\left[\left\|g(X,V)\right\|^p\middle|V=v\right] \leq C_{\rho,p} E\left[\left\|g(X,V)+h(Y,V)\right\|^p\middle|V=v\right]$$

and consequently

$$\begin{aligned} E\left[\left\|g(X,V)\right\|^p\right] &= E\left[E\left[\left\|g(X,V)\right\|^p\middle|V\right]\right] \\ &\leq E\left[C_{\rho,p} E\left[\left\|g(X,V)+h(Y,V)\right\|^p\middle|V\right]\right] \\ &= C_{\rho,p} E\left[\left\|g(X,V)+h(Y,V)\right\|^p\right]. \end{aligned}$$

□

Lemma 4.6. *Under the assumptions of Theorem 2.2, for any r there exists a constant $B_{d,r}$ such that for any finite subset $S \subset \mathbb{N}^d$ and $i \in \{1, \dots, K\}$*

$$
\begin{aligned}
&E\Big\| \sum_{\mathbf{k}\in S^{(n)}} (X_{\mathbf{k}} - \mu)V_{n,i}(\mathbf{k})\Big\|^r \\
\leq & B_{d,r}\left(\sum_{\mathbf{k}\in S^{(n)}} E\left\|(X_{\mathbf{k}} - \mu)\right\|^r E\left\|V_{n,i}(\mathbf{k})\right\|^r + \Big(\sum_{\mathbf{k}\in S^{(n)}} E\left\|(X_{\mathbf{k}} - \mu)\right\|^2 E\left\|V_{n,i}(\mathbf{k})\right\|^2 \Big)^{r/2} \right),
\end{aligned}
$$

where $S^{(n)} = S \cap \{1, \dots, n\}$. For any block $U \subseteq \{1, \dots, n\}^d$ and

$$M^\star(U) = \max_{W \lhd U} \left\| \sum_{\mathbf{j}\in W} (X_{\mathbf{j}} - \mu)V_{n,i}(\mathbf{j}) \right\|,$$

it then holds that

$$E\left(M^\star(U)^r\right) \leq C_r(\#U)^{r/2}$$

for $r \in (2, 2+\delta]$ and some $C_r > 0$ that may depend on r but not on U or n.

Proof. This inequality follows in the same way as Theorem 2 of Zhang (1998) and Lemma 4.1 above, using Lemma 4.5 instead of Theorem 1 of Zhang (1998). □

4.2 Proofs of the main results

Proof of Theorem 2.1. We assume without loss of generality that $\mu = 0$ and proceed as in the proof of Theorem 1 in Sharipov et al. (2016) by showing the three conditions of Lemma 4.4. First, note that for any $h \in H \setminus \{0\}$, the random field $\{Y_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ with $Y_{\mathbf{j}} = \langle X_{\mathbf{j}}, h\rangle$ is centered, stationary and $\rho_{\mathbb{R}}$-mixing with $\rho_{\mathbb{R},Y}(x) \leq \rho_{\mathbb{R},X}(x)$ and $\alpha_{1,1,Y}(x) \leq \alpha_{1,1,X}(x)$, since any $Y_{\mathbf{j}}$ is a measurable transform of $X_{\mathbf{j}}$. Furthermore,

$$E|Y_{\mathbf{j}}|^{2+\delta} \leq \|h\|^{2+\delta} E\|X_{\mathbf{j}}\|^{2+\delta}$$

ensures that $\{Y_{\mathbf{j}}\}_{\mathbf{j}\in\mathbb{Z}^d}$ has finite $(2+\delta)$-moments. Now, Lemma 4.3 implies

$$\left\{ \frac{1}{n^{d/2}} \sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\lfloor n\mathbf{t}\rfloor} Y_{\mathbf{j}} \right\}_{\mathbf{t}\in[0,1]^d} \Rightarrow \{W_h(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}, \quad \text{in } D_{\mathbb{R}}([0,1]^d),$$

where $\{W_h(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in $\mathbb{R}$ with covariance

$$\sigma^2(h) = \sum_{\mathbf{j}\in\mathbb{Z}^d} EY_{\underline{\mathbf{0}}}Y_{\mathbf{j}} = \sum_{\mathbf{j}\in\mathbb{Z}^d} E\left(\langle X_{\underline{\mathbf{0}}}, h\rangle \langle X_{\mathbf{j}}, h\rangle\right),$$

and the series converges absolutely. Define the covariance operator S as in (3), then $\langle Sh, h\rangle = \sigma^2(h)$ holds for all $h \in H \setminus \{0\}$, and S is positive, linear and self-adjoint. Then $S \in \mathcal{S}(H)$, because for any complete orthonormal system $\{e_i\}_{i\in\mathbb{N}}$ in H, we obtain

$$\sum_{i=1}^{\infty} |\langle Se_i, e_i\rangle| = \sum_{i=1}^{\infty} \langle Se_i, e_i\rangle = \sum_{i=1}^{\infty} \lim_{n\to\infty} n^{-d} E\left[\sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}} \langle X_{\mathbf{j}}, e_i\rangle \right]^2$$

and Theorem 28.10 of Bradley (2007) (Volume III, p. 154) implies

$$n^{-d}E\left[\sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\underline{\mathbf{n}}}\langle X_{\mathbf{j}},e_i\rangle\right]^2\le Cn^{-d}\sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\underline{\mathbf{n}}}E(\langle X_{\mathbf{j}},e_i\rangle^2)\le CE(\langle X_{\underline{\mathbf{0}}},e_i\rangle^2)$$

with a single constant C for all n and i. Therefore,

$$\sum_{i=1}^{\infty}|\langle Se_i,e_i\rangle|\le C\sum_{i=1}^{\infty}E(\langle X_{\underline{\mathbf{0}}},e_i\rangle^2)=CE\|X_{\underline{\mathbf{0}}}\|^2<\infty.$$

Define $S_n(\mathbf{t})=n^{-d/2}\sum_{\underline{\mathbf{1}}\le\mathbf{k}\le\lfloor n\mathbf{t}\rfloor}X_{\mathbf{k}}$ and consider a Brownian sheet $\{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ in H whose covariance operator is defined as in (3). Then $\{W^{(k)}(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}=\{P_kW(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in H_k with covariance operator $S_k=P_kSP_k$. In particular, the covariance operator can be identified with the $k\times k$ nonnegative definite covariance matrix $\Sigma=(\gamma_{i,j})_{1\le i,j\le k}$ with $\gamma_{i,j}=\sum_{\mathbf{v}\in\mathbb{Z}^d}E\left(\langle X_{\underline{\mathbf{0}}},e_i\rangle\langle X_{\mathbf{v}},e_j\rangle\right)$.
For each $k\ge1$, the convergence

$$\{P_kS_n(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}\Rightarrow\{W^{(k)}(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d},\quad\text{in }D_{H_k}([0,1]^d),$$

is equivalent to the functional central limit theorem for the k-dimensional random field $\tilde{X}_{\mathbf{j}}^{(k)}=(\langle X_{\mathbf{j}},e_1\rangle,\ldots,\langle X_{\mathbf{j}},e_k\rangle)^t$. Since $\{\tilde{X}_{\mathbf{j}}^{(k)}\}_{\mathbf{j}\in\mathbb{Z}^d}$ fulfills the assumptions of the Lemma, Lemma 4.3 yields

$$\left\{\frac{1}{n^{d/2}}\sum_{\underline{\mathbf{1}}\le\mathbf{j}\le\lfloor n\mathbf{t}\rfloor}\tilde{X}_{\mathbf{j}}^{(k)}\right\}_{\mathbf{t}\in[0,1]^d}\Rightarrow\{\tilde{W}^{(k)}(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d},\quad\text{in }D_{\mathbb{R}^k}([0,1]^d),$$

where $\{\tilde{W}^{(k)}(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in $\mathbb{R}^k$ with covariance matrix Σ, i.e. condition (a) of Lemma 4.4 is satisfied.
Let $\{W(\mathbf{t})\}_{\mathbf{t}\in[0,1]^d}$ be a Brownian sheet in H with $\mathrm{Cov}W(\mathbf{1})=S$, where S is as defined in (3). For every e_i, $\{\langle W(\mathbf{t}),e_i\rangle\}_{\mathbf{t}\in[0,1]^d}$ is a Brownian sheet in $\mathbb{R}$, and therefore Cairoli's strong inequality (Corollary 2.3.1 in Chapter 7 of Khoshnevisan (2002)) for submartingale random fields in $\mathbb{R}$ yields

$$\begin{aligned}E\left(\sup_{\mathbf{t}\in[0,1]^d}\left\|W(\mathbf{t})-W^{(k)}(\mathbf{t})\right\|^2\right)&=E\left(\sup_{\mathbf{t}\in[0,1]^d}\sum_{i=k+1}^{\infty}\langle W(\mathbf{t}),e_i\rangle^2\right)\\&\le\sum_{i=k+1}^{\infty}E\left(\sup_{\mathbf{t}\in[0,1]^d}\langle W(\mathbf{t}),e_i\rangle^2\right)\\&\le4^d\sum_{i=k+1}^{\infty}E\left(\langle W(\underline{\mathbf{1}}),e_i\rangle^2\right)\\&=4^d\sum_{i=k+1}^{\infty}\langle Se_i,e_i\rangle\overset{k\to\infty}{\longrightarrow}0,\end{aligned}$$

which implies $\sup_{\mathbf{t}\in[0,1]^d}\|W(\mathbf{t})-W^{(k)}(\mathbf{t})\|^2\to0$ in probability and therefore $W^{(k)}\Rightarrow W$ in $D_H([0,1]^d)$.

Finally, using (5) for $r = 2 + \delta > 2$, we show condition (c). We note that due to the Hilbert space property,

$$\|A_k(X_{\underline{\mathbf{1}}})\| = \left\| X_{\underline{\mathbf{1}}} - \sum_{i=1}^{k} \langle X_{\underline{\mathbf{1}}}, e_i \rangle e_i \right\| \overset{k\to\infty}{\longrightarrow} 0 \quad a.s.$$

Using $\|A_k(X_{\underline{\mathbf{1}}})\|^r \leq \|X_{\underline{\mathbf{1}}}\|^r$ and the dominated convergence theorem, this implies

$$\max\{E\|A_k(X_{\underline{\mathbf{1}}})\|^r, E\|A_k(X_{\underline{\mathbf{1}}})\|^2\} \overset{k\to\infty}{\longrightarrow} 0.$$

We can therefore apply Lemma 4.1 to $\{A_k(X_{\mathbf{j}})\}_{\mathbf{j}\in\mathbb{Z}^d}$ and obtain

$$\begin{aligned} &E\left(\sup_{\mathbf{t}\in[0,1]^d} \|S_n(\mathbf{t}) - P_k S_n(\mathbf{t})\|^r\right) \\ =& n^{-rd/2} E\left(\max_{\underline{\mathbf{1}}\leq\mathbf{l}\leq\underline{\mathbf{n}}} \left\|\sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\mathbf{l}} A_k(X_{\mathbf{j}})\right\|^r\right) \\ \leq& \tilde{C} B_{d,r}\left(E\|A_k(X_{\underline{\mathbf{1}}})\|^r + (E\|A_k(X_{\underline{\mathbf{1}}})\|^2)^{r/2}\right) \overset{k\to\infty}{\longrightarrow} 0, \end{aligned}$$

and therefore (c) in Lemma 4.4. □

Proof of Theorem 2.2. We will use Lemma 4.4. For $k \in \mathbb{N}$, we start by establishing the tightness of $S_{n,1}^{\star(k)}, \ldots, S_{n,K}^{\star(k)}$. Since $S_n^{(k)}$ is also tight (cf. the proof of Theorem 2.1), the tightness of $(S_n^{(k)}, S_{n,1}^{\star(k)}, \ldots, S_{n,K}^{\star(k)})$ will then follow immediately.

Note that for any $j \in \{1, \ldots, K\}$

$$S_{n,j}^{\star(k)}(\mathbf{t}) = \frac{1}{n^{d/2}} \sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} \left(X_{\mathbf{i}}^{(k)} - \mu^{(k)}\right) V_{n,j}(\mathbf{i}) - \frac{1}{n^{d/2}} \sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} \left(\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right) V_{n,j}(\mathbf{i}).$$

Using Lemma 4.6, we obtain that the first summand is stochastically bounded and fulfills the tightness condition (iv) of Lemma 4.2 (cf. the proof of Theorem 2.1). Since by assumption the change-set estimator $\hat{C}_n$ is a subblock of $(\underline{\mathbf{0}}, \underline{\mathbf{n}}]$, we can bound the second summand by

$$\begin{aligned} &\left\| \frac{1}{n^{d/2}} \sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} \left(\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right) V_{n,j}(\mathbf{i}) \right\| \\ \leq& n^{d/2} \max_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}} \left\|\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}\right\| \frac{1}{n^d} \sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} |V_{n,j}(\mathbf{i})| \\ \leq& C \max_{\underline{\mathbf{1}}\leq\mathbf{l}<\mathbf{m}\leq\underline{\mathbf{n}}} \left\| \frac{1}{n^{d/2}} \sum_{\mathbf{l}\leq\mathbf{i}\leq\mathbf{m}} \left(X_{\mathbf{i}}^{(k)} - \mu^{(k)}\right) \right\| \frac{1}{n^d} \sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} |V_{n,j}(\mathbf{i})| \end{aligned}$$

for some $C > 0$. By Lemma 4.1, the first factor is stochastically bounded. For the second factor, note that due to the Gaussian distribution of $V_{n,j}(\mathbf{i})$, for any block S and $r \geq 2$,

$$E\left|\sum_{\mathbf{k}\in S} |V_{n,j}(\mathbf{k})|\right|^r \leq C_r\, (\#S)^r \tag{8}$$

holds for some constant $C_r > 0$. Therefore, the second summand is stochastically bounded. Writing

$$Y_n(\cdot) = \frac{1}{n^{d/2}} \sum_{\mathbf{1} \leq \mathbf{i} \leq \lfloor n \cdot \rfloor} \left(\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)} \right) V_{n,j}(\mathbf{i}) \quad \text{and} \quad W_n(\cdot) = n^{-d} \sum_{\mathbf{1} \leq \mathbf{i} \leq \lfloor n \cdot \rfloor} |V_{n,j}(\mathbf{i})|,$$

the modulus of continuity of the second summand can be bounded in the following way:

$$\begin{aligned}
&P(\omega^k_{Y_n}(\delta) \geq \varepsilon) \\
\leq& \sum_{h=1}^d P\left(\sup_{\mathbf{t} \in [0,1]^d \colon t_h \leq 1-\delta, \gamma \in (0,\delta)} \| Y_n(t_1, \ldots, t_{h-1}, t_h + \gamma, t_{h+1}, \ldots, t_d) - Y_n(\mathbf{t}) \| \geq \varepsilon d^{-1} \right) \\
=& \sum_{h=1}^d P\left(\sup_{\mathbf{t} \in [0,1]^d \colon t_h \leq 1-\delta, \gamma \in (0,\delta)} \frac{1}{n^{d/2}} \left\| \sum_{\substack{\mathbf{1} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor \\ \lfloor n t_h \rfloor < i_h \leq \lfloor n(t_h+\gamma) \rfloor}} (\hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)}) V_{n,j}(\mathbf{i}) \right\| \geq \varepsilon d^{-1} \right) \\
\leq& \sum_{h=1}^d P\left(\max_{\mathbf{1} \leq \mathbf{i} \leq \mathbf{n}} n^{d/2} \| \hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)} \| \cdot \sup_{\mathbf{t} \in [0,1]^d \colon t_h \leq 1-\delta, \gamma \in (0,\delta)} \frac{1}{n^d} \sum_{\substack{\mathbf{1} \leq \mathbf{i} \leq \lfloor n\mathbf{t} \rfloor \\ \lfloor n t_h \rfloor < i_h \leq \lfloor n(t_h+\gamma) \rfloor}} |V_{n,j}(\mathbf{i})| \geq \varepsilon d^{-1} \right) \\
\leq& dP(\max_{\mathbf{1} \leq \mathbf{i} \leq \mathbf{n}} n^{d/2} \| \hat{\mu}^{(k)}(\mathbf{i}) - \mu^{(k)} \| > C) \\
&+ \sum_{h=1}^d P\left(\sup_{\mathbf{t} \in [0,1]^d \colon t_h \leq 1-\delta, \gamma \in (0,\delta)} | W_n(t_1, \ldots, t_{h-1}, t_h + \gamma, t_{h+1}, \ldots, t_d) - W_n(\mathbf{t}) | \geq \varepsilon d^{-1} C^{-1} \right).
\end{aligned}$$

The first summand goes to 0 uniformly in n for $C \to \infty$. For the second summand, define

$$A_m(h, \delta) = (0,1] \times \cdots \times ((m-1)\delta, m\delta \wedge 1] \times \cdots \times (0,1]$$

for $m = 1, \ldots, p$ with $p = p(\delta) = \lfloor \delta^{-1} \rfloor + 1$ and

$$U_{m,n} = \{ \lfloor n\mathbf{t} \rfloor \, : \, \mathbf{t} \in A_m(h,\delta) \}.$$

Then, $\# U_{m,n} \leq n^d \delta$, and therefore,

$$\begin{aligned}
&P\left(\sup_{\mathbf{t} \in [0,1]^d \colon t_h \leq 1-\delta, \gamma \in (0,\delta)} | W_n(t_1, \ldots, t_{h-1}, t_h + \gamma, t_{h+1}, \ldots, t_d) - W_n(\mathbf{t}) | \geq \varepsilon d^{-1} C^{-1} \right) \\
\leq& \sum_{m=1}^p P\left(\sup_{\mathbf{s}, \mathbf{t} \in A_m(h,\delta), \, s_r = t_r \ (r \neq h)} | W_n(\mathbf{t}) - W_n(\mathbf{s}) | \geq \frac{\varepsilon}{2} d^{-1} C^{-1} \right) \\
\leq& \sum_{m=1}^p P\left(\sup_{V \lhd U_{m,n}} n^{-d} \sum_{\mathbf{i} \in V} |V_{n,j}(\mathbf{i})| \geq \frac{\varepsilon}{4} d^{-1} C^{-1} \right) \\
\leq& \sum_{m=1}^p P\left(n^{-d} \sum_{\mathbf{i} \in U_{m,n}} |V_{n,j}(\mathbf{i})| \geq \frac{\varepsilon}{4} d^{-1} C^{-1} \right) \\
\leq& \sum_{m=1}^p n^{-dr} 4^r d^r C^r \varepsilon^{-r} C_r (\# U_{m,n})^r \\
\leq& 4^r d^r C^r \varepsilon^{-r} C_r (1 + \delta^{-1}) \delta \cdot \delta^{r-1} \overset{\delta \to 0}{\longrightarrow} 0.
\end{aligned}$$

Thus, condition (iv) of Lemma 4.2 is fulfilled for the second summand as well. Therefore, the sum $S^{\star(k)}_{n,j}$ is stochastically bounded with a modulus of continuity that fulfills the tightness condition, and therefore it is tight.

Next, we establish the finite dimensional convergence. Note that due to the tightness of the process, it suffices to show that for any subsequence, there exists a further subsequence such that the finite dimensional distributions converge to the right limit distribution. To do this, we first show the following result: For any subsequence $(n_m)_{m\in\mathbb{N}}$, there is another subsequence $(n_m)_{m\in M}$ with $M \subset \mathbb{N}$, such that for all $k, l \in \mathbb{N}$ and all disjoint blocks $B_1, \dots, B_l$ with corners in $([0,1]\cap\mathbb{Q})^d$, the weak convergence of the conditional (on $X_{\mathbf{i}}$, $\mathbf{i} \le \underline{\mathbf{n_m}}$) distribution of the random vectors

$$\mathbf{W}^\star_{m,j} := \left(S^{\star(k)}_{n_m,j}(B_1), S^{\star(k)}_{n_m,j}(B_2), \dots, S^{\star(k)}_{n_m,j}(B_l)\right)^t, \quad j = 1, \dots, K$$

to $\mathbf{W}^\star_j := (W^{\star(k)}_j(B_1), W^{\star(k)}_j(B_2), \dots, W^{\star(k)}_j(B_l))^t$, $j = 1, \dots, K$, holds almost surely for $(n_m)_{m\in M}$.

By the following argument, the almost sure weak convergence yields the weak convergence of the joint distribution. Define the random vectors

$$\mathbf{W}_m := \left(S^{(k)}_{n_m}(B_1), S^{(k)}_{n_m}(B_2), \dots, S^{(k)}_{n_m}(B_l)\right)^t$$

and $\mathbf{W} := (W^{(k)}(B_1), W^{(k)}(B_2), \dots, W^{(k)}(B_l))^t$. Note that by assumption, $\mathbf{W}, \mathbf{W}^\star_1, \dots, \mathbf{W}^\star_K$ are stochastically independent. It holds for any Borel sets $A_0, A_1, \dots, A_K \subset H^l_k$ that

$$\begin{aligned}
&\left|P\left(\mathbf{W}_m \in A_0, \mathbf{W}^\star_{m,1} \in A_1, \dots, \mathbf{W}^\star_{m,K} \in A_K\right) - P\left(\mathbf{W} \in A_0, \mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right)\right| \\
=&\left|E\left[P\left(\mathbf{W}_m \in A_0, \mathbf{W}^\star_{m,1} \in A_1, \dots, \mathbf{W}^\star_{m,K} \in A_K \middle| \mathbf{X_i}, \mathbf{i} \le \underline{\mathbf{n_m}}\right)\right]\right. \\
&\left. - P\left(\mathbf{W} \in A_0, \mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right)\right| \\
\le& E\left[\mathbb{1}_{\{\mathbf{W}_m \in A_0\}}\left|P\left(\mathbf{W}^\star_{m,1} \in A_1, \dots, \mathbf{W}^\star_{m,K} \in A_K \middle| \mathbf{X_i}, \mathbf{i} \le \underline{\mathbf{n_m}}\right) - P\left(\mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right)\right|\right] \\
&+ \left|P\left(\mathbf{W}_m \in A_0\right) P\left(\mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right) - P\left(\mathbf{W} \in A_0\right) P\left(\mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right)\right| \\
\le& E\left[\left|P\left(\mathbf{W}^\star_{m,1} \in A_1, \dots, \mathbf{W}^\star_{m,K} \in A_K \middle| \mathbf{X_i}, \mathbf{i} \le \underline{\mathbf{n_m}}\right) - P\left(\mathbf{W}^\star_1 \in A_1, \dots, \mathbf{W}^\star_K \in A_K\right)\right|\right] \\
&+ \left|P\left(\mathbf{W}_m \in A_0\right) - P\left(\mathbf{W} \in A_0\right)\right| \\
&\to 0,
\end{aligned}$$

as almost sure convergence implies convergence in L_1 for bounded random variables and as the last summand converges to 0 by Theorem 2.1. To show the conditional weak convergence of $(\mathbf{W}^\star_{m,1}, \dots, \mathbf{W}^\star_{m,K})$, note that conditional on $X_{\mathbf{i}}, \mathbf{i} \le \underline{\mathbf{n_m}}$, $\mathbf{W}^\star_{m,1}, \dots, \mathbf{W}^\star_{m,K}$ are stochastically independent and have a Gaussian distribution with mean 0, so it suffices to show the convergence of the conditional covariance operators. For $j \in \{1, \dots, K\}$ and $l_1, l_2 \in \{1, \dots, l\}$, the covariance operators are given by

$$\begin{aligned}
&\mathrm{Cov}^\star(S^{\star(k)}_{n,j}(B_{l_1}), S^{\star(k)}_{n,j}(B_{l_1})) \\
=& E\left[S^{\star(k)}_{n,j}(B_{l_1}) S^{\star(k)}_{n,j}(B_{l_2})^t | X_{\mathbf{i}}, \mathbf{i} \le \underline{\mathbf{n}}\right] \\
=& \frac{1}{n^d} \sum_{\mathbf{a}\in B_{l_1,n}} \sum_{\mathbf{b}\in B_{l_2,n}} (X^{(k)}_{\mathbf{a}} - \hat{\mu}^{(k)}(\mathbf{a}))(X^{(k)}_{\mathbf{b}} - \hat{\mu}^{(k)}(\mathbf{b}))^t E\left[V_{n,j}(\mathbf{a}) V_{n,j}(\mathbf{b})\right] \\
=& \sum_{\mathbf{h}\in B_{l_2,n}\ominus B_{l_1,n}} \omega\left(\frac{\mathbf{h}}{q(n)}\right) \frac{1}{n^d} \sum_{\mathbf{a}:\, \mathbf{a}\in B_{l_1,n}, \mathbf{a}+\mathbf{h}\in B_{l_2,n}} (X^{(k)}_{\mathbf{a}} - \hat{\mu}^{(k)}(\mathbf{a}))(X^{(k)}_{\mathbf{a}+\mathbf{h}} - \hat{\mu}^{(k)}(\mathbf{a}+\mathbf{h}))^t.
\end{aligned}$$

For $l_1 = l_2$, this is the covariance estimator proposed by Bucchia and Heuser (2015). As seen in Remark 2.1, under the assumptions of Theorem 2.2 this estimator converges in probability to $\lambda(B_{l_1})\Sigma$, where Σ is the long-run variance matrix. Write $\mathrm{var}^\star(S^{\star(k)}_{n,j}(B_{l_1})) = \mathrm{Cov}^\star(S^{\star(k)}_{n,j}(B_{l_1}), S^{\star(k)}_{n,j}(B_{l_1}))$. For $l_1 \neq l_2$, it holds that

$$\mathrm{var}^\star\left(S^{\star(k)}_{n,j}(B_{l_1} \cup B_{l_2})\right) \xrightarrow{P} \lambda(B_{l_1} \cup B_{l_2})\Sigma = (\lambda(B_{l_1}) + \lambda(B_{l_2}))\Sigma \quad \text{(cf. Remark 2.1)},$$

and thus

$$\begin{aligned}
&\mathrm{Cov}^\star(S^{\star(k)}_{n,j}(B_{l_1}), S^{\star(k)}_{n,j}(B_{l_2})) \\
&= \frac{1}{2}\left(\mathrm{var}^\star\left(S^{\star(k)}_{n,j}(B_{l_1}) + S^{\star(k)}_{n,j}(B_{l_2})\right) - \mathrm{var}^\star(S^{\star(k)}_{n,j}(B_{l_1})) - \mathrm{var}^\star(S^{\star(k)}_{n,j}(B_{l_2}))\right) \\
&= \frac{1}{2}\left(\mathrm{var}^\star\left(S^{\star(k)}_{n,j}(B_{l_1} \cup B_{l_2})\right) - \mathrm{var}^\star(S^{\star(k)}_{n,j}(B_{l_1})) - \mathrm{var}^\star(S^{\star(k)}_{n,j}(B_{l_2}))\right) \xrightarrow{P} 0.
\end{aligned}$$

Therefore, for any subsequence $(n_m)_{m\in\mathbb{N}}$, there exists a further subsequence $(n_m)_{m\in M}$ such that the estimator converges almost surely. Since we only consider countably many blocks B_i, by a diagonal sequence argument we can choose a single subsequence $(n_m)_{m\in M}$ so that the almost sure convergence holds for all $k \in \mathbb{N}$ and all blocks with edges in $(\mathbb{Q} \cap [0,1])^d$.

As the process $(S^{(k)}_{n_m}, S^{\star(k)}_{n_m,1}, \ldots, S^{\star(k)}_{n_m,K})$ is right-continuous, the convergence of all finite dimensional distributions follows from the convergence for all disjoint $B_1, \ldots, B_l$ with corners in $([0,1] \cap \mathbb{Q})^d$ (see the remark after Theorem 3 in Bickel and Wichura (1971)). Together with the tightness of $(S^{(k)}_n, S^{\star(k)}_{n,1}, \ldots, S^{\star(k)}_{n,K})$, condition (a) of Lemma 4.4 follows: for every k, the process $(S^{(k)}_n, S^{\star(k)}_{n,1}, \ldots, S^{\star(k)}_{n,K})$ converges to $(W^{(k)}, W^{\star(k)}_1, \ldots, W^{\star(k)}_K)$.

From the proof of Theorem 2.1, we already know that $W^{(k)} \Rightarrow W$ as $k \to \infty$. $W^{\star(k)}_1, \ldots, W^{\star(k)}_K$ and $W^\star_1, \ldots, W^{\star(k)}_K$ are independent copies of $W^{(k)}$ respectively W, so condition (b) is obvious.

For condition (c), note that for $r = 2 + \delta$

$$\begin{aligned}
&E\left[\sup_{\mathbf{s},\mathbf{t}\in[0,1]^d}\left\|(S_n(\mathbf{s}), S^\star_{n,1}(\mathbf{t}), \ldots, S^\star_{n,K}(\mathbf{t})) - (S^{(k)}_n(\mathbf{s}), S^{\star(k)}_{n,1}(\mathbf{t}), \ldots, S^{\star(k)}_{n,K}(\mathbf{t}))\right\|^r\right] \\
\leq& 2^{r-1}E\left[\sup_{\mathbf{s}\in[0,1]^d}\left\|S_n(\mathbf{s}) - S^{(k)}_n(\mathbf{s})\right\|^r\right] + 2^{K(r-1)}\sum_{j=1}^K E\left[\sup_{\mathbf{t}\in[0,1]^d}\left\|S^\star_{n,j}(\mathbf{t}) - S^{\star(k)}_{n,j}(\mathbf{t})\right\|^r\right] \\
\leq& 2^{r-1}E\left[\sup_{\mathbf{s}\in[0,1]^d}\left\|S_n(\mathbf{s}) - S^{(k)}_n(\mathbf{s})\right\|^r\right] \\
&+2^{(K+1)(r-1)}\sum_{j=1}^K E\left[\sup_{\mathbf{t}\in[0,1]^d}\left\|\frac{1}{n^{d/2}}\sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} V_{n,j}(\mathbf{i})\left(A_k(X_{\mathbf{i}}) - A_k(\mu)\right)\right\|^r\right] \\
&+2^{(K+1)(r-1)}\sum_{j=1}^K E\left[\sup_{\mathbf{t}\in[0,1]^d}\left\|\frac{1}{n^{d/2}}\sum_{\mathbf{1}\leq\mathbf{i}\leq\lfloor n\mathbf{t}\rfloor} V_{n,j}(\mathbf{i})\left(A_k(\hat{\mu}(\mathbf{i})) - A_k(\mu)\right)\right\|^r\right].
\end{aligned}$$

By the same reasoning as in the proof of Theorem 2.1, the first two terms converge to 0 for $k \to \infty$ by Lemma 4.1 and 4.6 respectively. For the third term, consider

$$
\begin{aligned}
&E\left[\sup_{\mathbf{t}\in[0,1]^d}\left\|\frac{1}{n^{d/2}}\sum_{\underline{\mathbf{1}}\le\mathbf{i}\le\lfloor n\mathbf{t}\rfloor} V_{n,j}(\mathbf{i})\left(A_k(\hat{\mu}(\mathbf{i}))-A_k(\mu)\right)\right\|^r\right]\\
\le &E\left[n^{rd/2}\max_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}}\|A_k(\hat{\mu}(\mathbf{i})-\mu)\|^r\right]\cdot E\left[\left|\frac{1}{n^d}\sum_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}}|V_{n,j}(\mathbf{i})|\right|^r\right]\\
\le &CE\left[\max_{\underline{\mathbf{1}}\le\mathbf{l}<\mathbf{m}\le\underline{\mathbf{n}}}\left\|\frac{1}{n^{d/2}}\sum_{\mathbf{l}\le\mathbf{i}\le\mathbf{m}} A_k\left(X_{\mathbf{i}}-\mu\right)\right\|^r\right]\cdot E\left[\left|\frac{1}{n^d}\sum_{\underline{\mathbf{1}}\le\mathbf{i}\le\underline{\mathbf{n}}}|V_{n,j}(\mathbf{i})|\right|^r\right].
\end{aligned}
$$

Since the first factor goes to 0 (cf. the proof of Theorem 2.1) and the second factor remains bounded (cf. (8)), the proof of condition (c) is finished. □

Chapter 5

Additional material: Change-set estimation

1 Introduction

1.1 Change-set estimation

Given observations at the nodes of a d-dimensional ($d \in \mathbb{N}$ fixed) grid in $[0,1]^d$ with rectangular mesh, we study the problem of dividing these observations into two subsets such that all the observations in one subset have the same distribution but the distributions differ for each subset. Applications of this type of model can be found in image analysis, where the observations are divided into a fore- and a background and the aim is to first find the corresponding segments of the grid and then estimate the distribution on each segment, thus reconstructing the image. Further applications lie in forestry, medicine, geology or meteorology, as discussed e.g. in Carlstein and Krishnamoorthy (1992). Here, we restrict ourselves to changes in the mean, where the mean in both segments is unknown and constant.

Since this type of problem can be interpreted as a multidimensional change-point problem — and indeed the nonparametric estimators employed in this chapter are generalizations of estimators used in classical change-point analysis — we call one of the two segments the change-set and focus on its estimation. A closely related problem is the boundary estimation problem, where instead of estimating the change-set, one aims to estimate the common boundary of the two segments. It stands to reason that the methods employed for one problem might also be used for the other, and we will therefore develop the theory for the boundary estimation problem alongside the theory for the change-set estimation.

While there are a lot of results for change-set and change-boundary estimation problems (see Korostelev and Tsybakov (1993), Müller and Song (1994), Carlstein et al. (1994), Khmaladze et al. (2006b), Mallik (2013) to name a few), most of the published works focus on independent observations. Articles like Carlstein and Krishnamoorthy (1992) and Ferger (2004) can thus draw on a more sophisticated theory (e.g. exponential inequalities) that is not available under the more general weak dependence assumptions considered here.

As a measure of the quality of the estimation, we count the total number of grid nodes that have been misclassified. This is in contrast to the local methods employed by e.g. Qiu (2005) and Garlipp and Müller (2007), which thrive to give a pixelwise reconstruction of the mean function and yield statements about the probability of misclassification of a single pixel but not of the total number of misclassified pixels. Letting the number of grid points go to infinity, we derive asymptotic results, namely conditions under which the estimated set converges in probability to the true change-set and the rate at which it does so.

From a technical point of view, as seen in Bucchia and Heuser (2015), the question of convergence rates for the total number of misclassified nodes is of interest when the change-set estimator is used to estimate the mean-function of the process: There, convergence for the mean-function estimator was derived by assuming that a change-set estimator with a given quality of approximation had been given. As an example for such an estimator, Bucchia and Heuser (2015) considered the case of rectangular change-sets. Here, we aim to obtain estimators for more general types of change-set, improving the result by Bucchia and Heuser (2015) along the way (cf. section 3.1). In contrast to Bucchia and Heuser (2015), in the current setting, we try to keep to the discrete framework as much as possible, i.e. we do not seek to estimate an abstract set $\theta \subseteq [0,1]^d$ but rather the grid points contained therein. This is reasonable, since the information given in the model pertains to the grid points rather than the whole set $[0,1]^d$, making it infeasible to expect a change-set estimator to have greater accuracy. In keeping with the rest of this thesis, we derive general results under relatively general assumptions and then give examples for the applications of these results.

The structure of the remainder of this chapter is as follows: First, we introduce the model considered and the assumptions which will be needed throughout the chapter. In section 2, the main results are presented, namely the consistency of the estimation and some general results on the rate of convergence. Since the latter reduces the problem of deriving convergence rates to obtaining a maximal inequality, we then give an example where such an inequality is fulfilled under mixing conditions. Finally, in section 3 we give examples for classes of change-sets where the main results can be applied to obtain the consistency and rates of convergence for the estimation. The proofs of both the main results and some auxiliary maximal inequalities are relegated to section 4.

Notations: Before we describe the model, we want to introduce some notations. We use the same notational conventions for vectors as in the rest of this thesis: $\mathbb{N}^d$ and $\mathbb{R}^d$ ($d \in \mathbb{N}$) are the spaces of d-dimensional integer and real vectors, respectively, equipped with the usual partial order. We write vectors in $\mathbb{R}^d$ as $\mathbf{x} = (x_1, \ldots, x_d)$. All operations on vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ are meant componentwise: $\lfloor \mathbf{x} \rfloor := (\lfloor x_1 \rfloor, \ldots, \lfloor x_d \rfloor)$, where $\lfloor x_i \rfloor$ is the integer part of x_i, $\frac{1}{\mathbf{x}} := \mathbf{x}^{-1} := (x_1^{-1}, \ldots, x_d^{-1})$ and $\mathbf{x} \cdot \mathbf{y} := (x_1 y_1, \ldots, x_d y_d)$. For $n \in \mathbb{N}$ and $\mathbf{x} \in \mathbb{R}^d$, we write $\underline{\mathbf{n}} = (n, \ldots, n)$ and $[\mathbf{x}] := \prod_{i=1}^d x_i$. For a set $T \subseteq \mathbb{R}^d$ and a vector $\mathbf{x} \in \mathbb{R}^d$, we write $\mathbf{x}T := \{\mathbf{x}\mathbf{t} : \mathbf{t} \in T\}$.

I_A is the indicator function and $\lambda(A)$ is the Lebesgue-measure of a set $A \subseteq \mathbb{R}^d$. In analogy to the one-dimensional case, we write

$$\{\underline{\mathbf{1}}, \ldots, \mathbf{N}\} := \{1, \ldots, N_1\} \times \cdots \times \{1, \ldots, N_d\}$$

for $\mathbf{N} = (N_1, \ldots, N_d) \in \mathbb{N}^d$ and

$$(\mathbf{x}, \mathbf{y}] := (x_1, y_1] \times \cdots \times (x_d, y_d]$$

for $\mathbf{x} = (x_1, \ldots, x_d), \mathbf{y} = (y_1, \ldots, y_d) \in \mathbb{R}^d$ with $\mathbf{x} \leq \mathbf{y}$.

$\|\cdot\|_2$ denotes the euclidean and $\|\cdot\|_\infty$ denotes the supremum norm. With this, we define the distance measure $\rho(\mathbf{x},\mathbf{y}) := \|\mathbf{x}-\mathbf{y}\|_\infty$ between two points and the distance $\rho(A,B) := \inf\{\rho(\mathbf{x},\mathbf{y}) : \mathbf{x}\in A, \mathbf{y}\in B\}$ between two sets A and B. Finally, $\rho(\mathbf{x},A) := \rho(\{\mathbf{x}\},A)$, and for $\delta > 0$, $A(\delta) := \{\mathbf{x} : \rho(\mathbf{x},\partial A) < \delta\}$ is the δ-annulus around the topological boundary ∂A of A. For a random variable X, $\|X\|_p$ denotes the usual L_p-norm ($p \in [1,\infty]$).

1.2 Model and main assumptions

We consider a sequence of grids $(\mathcal{I}_n)_{n\in\mathbb{N}}$ in $[0,1]^d$ ($d\in\mathbb{N}$ fixed) with

$$\mathcal{I} = \mathcal{I}_n = \left\{\kappa_{\mathbf{i},n} = \left(\frac{i_1}{N_1},\ldots,\frac{i_d}{N_d}\right) : 1 \le i_j \le N_j\right\} \subset [0,1]^d,$$

where $N_i = N_i(n) \in \mathbb{N}$ ($i = 1,\ldots,d$) and $\operatorname{card}(\mathcal{I}_n) = \prod_{j=1}^d N_j(n) \to \infty$ for $n\to\infty$. Assuming that $[0,1]^d = \theta \cup \theta^c$ for a measurable set θ, for each $n\in\mathbb{N}$ we consider $\mathbb{R}^p$-valued ($p\in\mathbb{N}$ fixed) observations $X_{\mathbf{i}} = X_{\mathbf{i},n}$ on the grid $\mathcal{I}_n$ with

$$X_{\mathbf{i},n} = aI_\theta(\kappa_{\mathbf{i},n}) + bI_{\theta^c}(\kappa_{\mathbf{i},n}) + Y_{\mathbf{i},n}, \quad \mathbf{i}\in\{\underline{\mathbf{1}},\ldots,\mathbf{N}\},$$

where $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}}$ is a square integrable centered process, $a = a_n, b = b_n \in \mathbb{R}^p$ are unknown with $a \neq b$. The unknown set θ is the focus of the following estimation methods. We divide $[0,1]^d$ into subrectangles

$$C_{\mathbf{i},n} := \left(\mathbf{N}^{-1}(\mathbf{i}-\underline{\mathbf{1}}), \mathbf{N}^{-1}\mathbf{i}\right]$$

such that each grid-point $\kappa_{\mathbf{i},n}$ is included in exactly one set $C_{\mathbf{i},n}$ and the volume of each set is $\lambda(C_{\mathbf{i},n}) = \prod_{j=1}^d N_j^{-1} = \operatorname{card}(\mathcal{I}_n)^{-1}$ and write the projection of a set $T\subset[0,1]^d$ on the $C_{\mathbf{i},n}$ as $T_{\mathcal{I}} = \bigcup_{\kappa_{\mathbf{i},n}\in T} C_{\mathbf{i},n}$. Writing

$$|T| = \operatorname{card}\{T\cap\mathcal{I}\},$$

this ensures that $|T| = |\mathcal{I}|\lambda(T_{\mathcal{I}})$. We identify θ with its projection on the grid $\theta = \theta_{\mathcal{I}}$.

Remark 1.1.

- *In the following, we will consider the asymptotics for $|\mathcal{I}_n|\to\infty$ as $n\to\infty$, i.e. a sequence of grids with $\max_{i=1,\ldots,d} N_i \to\infty$ and corresponding sequences of random fields $\{X_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$ and $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$. The assumption $|\mathcal{I}_n|\to\infty$ is sufficient for the general results presented here. However, it is sometimes necessary to make the stronger assumption $\min_{i=1,\ldots,d} N_i\to\infty$. For instance, this is the case when we want to switch from our general viewpoint of essentially equating subsets of $[0,1]^d$ with the grid nodes they contain to a continuous framework. In this setting, we view the projection of a set onto the grid as an approximation of that set which should become asymptotically finer (cf. Example 3.1).*

- *Since we consider observations on a sequence of grids, it is natural to model the stochastic part of the process as a sequence of random fields $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$ on the grids. However, an important special case which we keep in mind throughout this chapter is the following: Assume there is a centered random field $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ such* *that* $Y_{\mathbf{i},n} = Y_{\mathbf{i}}$ *for all* $n \in \mathbb{N}$, $\mathbf{i} \in \{\underline{\mathbf{1}}, \dots, \mathbf{N}\}$. *This model makes the uniform (in n) behavior of the random field — which is needed e.g. in Assumptions* (Y) *and* (Y(r)) *— more explicit and is the setting in which a lot of weak dependence concepts which allow the inference of our assumptions are defined (cf. Remark 1.7).*

In order to estimate θ, we choose a class $\mathcal{T} = \mathcal{T}_n \subseteq \mathcal{P}([0,1]^d)$ of candidate sets for θ such that each set $T \in \mathcal{T}$ is anchored on the grid, i.e. $T_{\mathcal{I}} = T$. Setting $\bar{X}_n = |\mathcal{I}_n|^{-1} \sum_{\kappa_{\mathbf{i},n} \in \mathcal{I}_n} X_{\mathbf{i},n}$, we use the statistic

$$\begin{aligned}
D_n(T) &= \frac{1}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T} (X_{\mathbf{i},n} - \bar{X}_n) \\
&= \frac{1}{|\mathcal{I}_n|} \left\{ \sum_{\kappa_{\mathbf{i},n} \in T} X_{\mathbf{i},n} - \frac{|T|}{|\mathcal{I}_n|} \left(\sum_{\kappa_{\mathbf{i},n} \in T} X_{\mathbf{i},n} + \sum_{\kappa_{\mathbf{i},n} \in T^c} X_{\mathbf{i},n} \right) \right\} \\
&= \frac{1}{|\mathcal{I}_n|} \left\{ \left(1 - \frac{|T|}{|\mathcal{I}_n|}\right) \sum_{\kappa_{\mathbf{i},n} \in T} X_{\mathbf{i},n} - \frac{|T|}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T^c} X_{\mathbf{i},n} \right\} \\
&= \frac{1}{|\mathcal{I}_n|} \left\{ \frac{|T^c|}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T} X_{\mathbf{i},n} - \frac{|T|}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T^c} X_{\mathbf{i},n} \right\} \\
&= \frac{|T|}{|\mathcal{I}_n|} \frac{|T^c|}{|\mathcal{I}_n|} \left\{ \frac{1}{|T|} \sum_{\kappa_{\mathbf{i},n} \in T} X_{\mathbf{i},n} - \frac{1}{|T^c|} \sum_{\kappa_{\mathbf{i},n} \in T^c} X_{\mathbf{i},n} \right\},
\end{aligned}$$

for $T \in \mathcal{T}$. (Here and in the rest of this chapter, $\sum_{\kappa_{\mathbf{i},n} \in T}$ is the sum over all $\kappa_{\mathbf{i},n} \in \mathcal{I}_n$ with $\kappa_{\mathbf{i},n} \in T$. If there are no such grid points, the sum is assumed to be zero.) Then

$$\hat{\theta}_n = \arg\max_{T \in \mathcal{T}_{\mathcal{I}}} \|D_n(T)\|_2$$

is our estimator for θ. Observing that

$$\begin{aligned}
\Delta_n(T) = ED_n(T) &= \frac{1}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T} \left(a I_\theta(\kappa_{\mathbf{i},n}) + b(1 - I_\theta(\kappa_{\mathbf{i},n})) - \frac{|\theta|}{|\mathcal{I}_n|} a - \left(1 - \frac{|\theta|}{|\mathcal{I}_n|}\right) b \right) \\
&= \frac{1}{|\mathcal{I}_n|} \sum_{\kappa_{\mathbf{i},n} \in T} \left(I_\theta(\kappa_{\mathbf{i},n}) - \frac{|\theta|}{|\mathcal{I}_n|} \right) (a - b) \\
&= \delta_n(T)(a - b),
\end{aligned}$$

with

$$\delta_n(T) = |\mathcal{I}_n|^{-2} \left\{ |T^c||T \cap \theta| - |T||T^c \cap \theta| \right\},$$

we further define $\rho_n(T) = |\delta_n(T)|$ and $B_n(T) = D_n(T) - \Delta_n(T)$. We assume for the rest of this chapter that $\sigma = \sigma_{\mathcal{I}} = |\mathcal{I}_n|^{-1} \min\{|\theta|, |\theta^c|\} > 0$, that is, that θ does not correspond to a trivial partition. As a distance measure on the grid, we define

$$d_n(T_1, T_2) = |\mathcal{I}_n|^{-1} |T_1 \triangle T_2|$$

where $\triangle$ denotes the symmetric difference of two sets. Using d_n as a measure for the distance between our estimator and the true set θ corresponds to a global approach where we count the total number of misclassified grid points. Since there is a-priori nothing in the model that distinguishes θ over θ^c, we also consider the question of estimating their common boundary $\partial\theta$. For this,

$$\partial_n(T,\theta) = \min\{d_n(T,\theta), d_n(T^c,\theta)\}$$

is the relevant grid-based distance measure.

With this notation, we obtain the following lemma, which gives a relation between δ_n and ρ_n and the distance between sets:

Lemma 1.1. *For any $T \subseteq [0,1]^d$ it holds that*

$$\sigma_{\mathcal{I}} d_n(T,\theta) \le \delta_n(\theta) - \delta_n(T) < d_n(\theta,T)$$

and

$$\sigma_{\mathcal{I}} \partial_n(T,\theta) \le \rho_n(\theta) - \rho_n(T) < \partial_n(\theta,T).$$

Remark 1.2. *In particular, if $d_n(T,\theta) \le 1/2$ for some $T \subseteq [0,1]^d$, we obtain*

$$\sigma_{\mathcal{I}} d_n(T,\theta) \le \rho_n(\theta) - \rho_n(T) < d_n(\theta,T),$$

since $d_n(T,\theta) \le 1 - d_n(T,\theta) = d_n(T^c,\theta)$ in this case.

Proof. The proof of the first assertion can be found in Ferger (2004) (Lemma A.1) (or also in Carlstein and Krishnamoorthy (1992)). The second assertion follows trivially from the first, since

$$\rho_n(\theta) - \rho_n(T) = \min\{\delta_n(\theta) - \delta_n(T), \delta_n(\theta) - \delta_n(T^c)\}$$

for any $T \subseteq [0,1]^d$. □

Remark 1.3. *As mentioned in the introduction, the results of this chapter were at least in part motivated by the need to have convergence rates of change-set estimators that could be used for the long-run variance estimation described in Bucchia and Heuser (2015). To do that, note that the general framework in Bucchia and Heuser (2015) can be viewed as a special case of the uniform grid model (where $N_1 = \cdots = N_d = n$), with a single process $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ such that $Y_{\mathbf{i}} = Y_{\mathbf{i},n}$ for $\mathbf{i} \in \{1,\ldots,n\}^d$. Since the long-run variance estimation was intended as part of a test for epidemic changes, Bucchia and Heuser (2015) assumed change-sets that are finite unions of m ($m \in \mathbb{N}$) rectangles with integer-valued edges. Because, for such sets, $\lambda(C_n) = |n^{-1}C_n|$, the (continuous) distance measure $\lambda(\hat{C}_n \triangle C_n)$ used in Bucchia and Heuser (2015) is equivalent to the (discrete) distance measure considered here, and all the results can be rewritten with $|\cdot|$ instead of $\lambda(\cdot)$. For the sake of simplicity, and since this assumption is needed to construct asymptotic tests, a functional central limit theorem was assumed in Bucchia and Heuser (2015). However, a careful reading of the proofs shows that what is actually used are the implied maximal inequalities (cf. Remark 2.2 in Bucchia and Heuser (2015))*

$$\max_{A\in\mathcal{A}_n} \left\| \sum_{\mathbf{i}\in A} Y_{\mathbf{i}} \right\| = \mathcal{O}_P(n^{d/2}m)$$

and

$$\max_{A\in\mathcal{A}_n}\left\|\sum_{\mathbf{i}\in A\cap N_{\mathbf{j}}} Y_{\mathbf{i}}\right\| = \mathcal{O}_P(n^{d/2}m),$$

where the latter follows from the fact that the sets $N_{\mathbf{j}} = \{\mathbf{k}\in\mathbb{N}^d : \underline{\mathbf{1}} \le \mathbf{k}, \mathbf{k}+\mathbf{j} \le \underline{\mathbf{n}}\}$ *are rectangles, and the same inequalities hold if we consider* A^c *instead of* A*. Since the specific form of the change-set was not needed otherwise, the results could be extended to any class* $\mathcal{A}_n$ *of sets such that maximal inequalities like the above are fulfilled. In particular, such maximal inequalities are derived in Section 3 (for the considered examples, maximal inequalities of type* (11) *can be derived analogously for sets of the form* $A\cap N_{\mathbf{j}}$ *or* $A^c\cap N_{\mathbf{j}}$*, since* $N_{\mathbf{j}}$ *is a rectangle).*

As mentioned above, the constants a, b as well as the number of grid points in the change-set depend on the underlying grid $\mathcal{I}$. As we will see in the following results, the change is easier to detect and estimate for large values of $\|a-b\|_2 = \|a_n - b_n\|_2$ and $\sigma_{\mathcal{I}}$, whereas change sizes that vanish asymptotically make stronger assumptions on the stochastic process necessary (cf. assumptions (7), (14) and (15) in Theorem 2.1 and Theorems 2.2 and 2.3). An important special case of size restrictions is the following:

Assumption (C)**.** *For all* $n\in\mathbb{N}$*, it holds that* $\|a_n - b_n\|_2 \ge \Delta > 0$ *for some* $\Delta > 0$*, which is independent of* n*. Analogously, it holds for the change-set that*

$$\sigma_{\mathcal{I}} = |\mathcal{I}_n|^{-1}\min\{|\theta|, |\theta^c|\} \ge \tilde{\sigma} > 0$$

for all $n\in\mathbb{N}$ *and some constant* $\tilde{\sigma} > 0$*, which is independent of* n*.*

We need some assumption on the best possible approximation of θ using candidate sets:

Assumption (T^*1)**.** *For any* $n\in\mathbb{N}$*, there is an element* $T^*\in\mathcal{T}$ *such that* $T^*\in\arg\min_{T\in\mathcal{T}}\partial_n(T,\theta)$ *and*

$$\partial_n(T^*,\theta) < \frac{1}{6}\sigma_{\mathcal{I}}|\mathcal{I}_n|^{-1}\alpha_n,$$

for some rate $\alpha_n = o(|\mathcal{I}_n|)$*.*

And analogously with d_n:

Assumption (T^*2)**.** *For any* $n\in\mathbb{N}$*, there is an element* $T^*\in\mathcal{T}$ *such that* $T^*\in\arg\min_{T\in\mathcal{T}} d_n(T,\theta)$ *and*

$$d_n(T^*,\theta) < \frac{1}{6}\sigma_{\mathcal{I}}|\mathcal{I}_n|^{-1}\alpha_n,$$

for some rate $\alpha_n = o(|\mathcal{I}_n|)$*.*

Finally, we introduce some assumptions on the underlying stochastic process:

Assumption (Y)**.** *For some* $K > 0$*, it holds for any* $M\subseteq[0,1]^d$ *and* $n\in\mathbb{N}$ *that*

$$E\left\|\sum_{\kappa_{\mathbf{i},n}\in M} Y_{\mathbf{i},n}\right\|_2^2 \le K|M|. \tag{1}$$

Remark 1.4. *Assumption* (Y) *is fulfilled e.g. if the autocovariances are absolutely summable, in the sense that*

$$\limsup_{n\to\infty} \max_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} \sum_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n} |\mathrm{Cov}(Y_{\mathbf{i},n}^{(k)}, Y_{\mathbf{j},n}^{(l)})| < \infty, \quad k,l \in \{1,\dots,p\}. \tag{2}$$

(Here, $k = l \in \{1,\dots,p\}$ would suffice to imply Assumption (Y)*.)*

Remark 1.5. *In the special case described in Remark 1.1, if $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ is weakly stationary, condition* (2) *is equivalent to*

$$\sum_{\mathbf{k}\in\mathbb{Z}^d} |\mathrm{Cov}(Y_{\underline{\mathbf{0}}}^{(i)}, Y_{\mathbf{k}}^{(j)})| < \infty, \quad i,j \in \{1,\dots,p\}. \tag{3}$$

Remark 1.6. *The properties* (2) *or* (3) *are often of use as part of central limit theorems and have therefore been proven under various weak dependence assumptions. Examples for this include (BL,θ)-dependence (cf. Bulinski and Shashkin (2007), Lemma 3.1.9 and Remark 3.1.10, in conjunction with Lemma 8 in Newman (1984)), mixing (cf. Guyon (1995), p. 110; Remark 2.3 and the proof of Lemma 2.2) and physical dependence (cf. El Machkouri et al. (2013), Proposition 2).*

Although Assumption (Y) is sufficient for the general results below, in applications we will often need the following stronger assumption, in which, for each $n \in \mathbb{N}$, $\mathcal{S} = \mathcal{S}_n \subseteq \mathcal{P}([0,1]^d)$ is a set of subsets of $[0,1]^d$ which are anchored on the pixels. (We will specify $\mathcal{S}$ as needed separately in each application.)

Assumption (Y(r))**.** *There are $r \geq 2$ and $K_r > 0$, so that for all $n \in \mathbb{N}$, it holds that $E\|Y_{\mathbf{i},n}\|_2^r < \infty$ for all $\kappa_{\mathbf{i},n} \in \mathcal{I}_n$ and*

$$E\left\|\sum_{\kappa_{\mathbf{i},n}\in S} Y_{\mathbf{i},n}\right\|_2^r \leq K_r|S|^{r/2}, \tag{4}$$

for all $S \in \mathcal{S}_n$.

Remark 1.7.

- *For a set $T \subset [0,1]^d$ which is anchored on the pixels, i.e. $T = T_{\mathcal{I}}$, it holds that*

  $$\mathbf{N}T = \bigcup_{\kappa_{\mathbf{i},n}\in T} \mathbf{N}C_{\mathbf{i},n} = \bigcup_{\kappa_{\mathbf{i},n}\in T} (\mathbf{i}-\underline{\mathbf{1}}, \mathbf{i}] \in \mathcal{B}(\mathbb{R}^d)$$

  and

  $$Z(\mathbf{N}T) = \sum_{\mathbf{i}\in\mathbb{Z}^d} \lambda(\mathbf{N}T \cap (\mathbf{i}-\underline{\mathbf{1}}, \mathbf{i}])Y_{\mathbf{i},n} = \sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n}$$

  as well as $\lambda(\mathbf{N}T) = |\mathcal{I}_n|\lambda(T) = |T|$. Therefore, the smoothed partial sum process is in this case identical to the unsmoothed process. Assumption (Y) *is then equivalent to*

  $$E\left\|\sum_{\mathbf{i}\in\mathbb{Z}^d} \lambda(\mathbf{N}T \cap (\mathbf{i}-\underline{\mathbf{1}}, \mathbf{i}])Y_{\mathbf{i},n}\right\|_2^2 \leq K\lambda(\mathbf{N}T) \tag{5}$$

  and analogously for Assumption (Y(r))*.*

- *For $r = 2$, Assumption* (Y(r)) *is implied by Assumption* (Y). *For $r > 2$, it is commonly used to prove the tightness in proofs of functional central limit theorems and can, for instance, be inferred from corresponding Rosenthal-inequalities. Let $Y_{\mathbf{i},n} = Y_{\mathbf{i}}$ for a centered real-valued random field $\{Y_{\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^d}$ (cf. Remark 1.1). Then Assumption* (Y(r)) *is fulfilled e.g. for mixing random fields (cf. Lin and Lu (1996), Lemma 6.2.3 (for $2 < r \leq 3$) or Zhang (1998), Theorem 2 and Bucchia and Wendler (2015), Lemma 4.1 for ρ-mixing; Lin and Lu (1996), Lemma 6.3.1 (for $2 < r \leq 3$) for nonuniform φ-mixing; Fazekas et al. (2000), Theorem 1, for (nonuniform) α-mixing), under physical dependence assumptions (cf. El Machkouri et al. (2013), Proposition 1) or positively or negatively associated (cf. Bulinski and Shashkin (2007), Theorems 2.3.1 and 2.3.3 in conjunction with Vronski (1998) and Christofides and Vaggelatou (2004)) random fields. In the special case of rectangular sets $\mathcal{S}$, Assumption* (Y(r)) *is also fulfilled under (BL,θ)-dependence (cf. Bulinski and Shashkin (2006), Theorem 1.1).*

2 Main results

To make this section easier to read, all proofs are relegated to Subsection 4.2 below.

2.1 Consistency

Theorem 2.1. *Let Assumption (T^*1) be fulfilled and let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence such that*

$$\max_{T\in\mathcal{T}} \|B_n(T)\|_2 = o_P(\xi_n). \tag{6}$$

Assume further that for all $\epsilon > 0$ there is some constant $\alpha > 0$ and a null sequence $(\beta_n)_{n\in\mathbb{N}}$ such that $\{T \in \mathcal{T} : \partial_n(T,\theta) < \beta_n\} \neq \emptyset$ for all $n \in \mathbb{N}$ and

$$\liminf_{n\to\infty} \xi_n^{-1} \|a - b\|_2 (\sigma_{\mathcal{I}}\epsilon - \beta_n) \geq \alpha. \tag{7}$$

Then $\hat{\theta}_n$ is a consistent estimator for the change-boundary, i.e.

$$\partial_n(\hat{\theta}_n, \theta) = o_P(1).$$

If Assumption (T^*2) and

$$\liminf_{n\to\infty} \xi_n^{-1} \|a - b\|_2 \left\{ \max_{T\in\mathcal{T}, d_n(T,\theta)<\beta_n} \rho_n(T) - \max_{T\in\mathcal{T}, d_n(T,\theta)\geq\epsilon} \rho_n(T) \right\} \geq \alpha, \tag{8}$$

hold instead of Assumption (T^*1) and (7), then $\hat{\theta}_n$ is a consistent estimator for the change-set, i.e.

$$d_n(\hat{\theta}_n, \theta) = o_P(1).$$

Remark 2.1.

1. *As shown in the proof of Theorem 2.1,*

$$\liminf_{n\to\infty} \xi_n^{-1} \left\{ \max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \right\} \geq \alpha$$

is implied by Lemma 1.1. An analogous argument for d_n needs an additional identifiability assumption on the candidate sets, which essentially guarantees that the estimated set is θ and not θ^c. (Since $\delta_n(T) = -\delta_n(T^c)$, the function $\rho_n(\cdot) = |\delta_n(\cdot)|$ is invariant with respect to taking the complement of a set.) The following assumption, combined with (7)*, yields* (8)*:*

Assumption (I)**.**

$$\liminf_{n\to\infty} \min_{T\in\mathcal{T}} d_n(T^c, \theta) > 0$$

Under Assumption (I)*, Lemma 1.1 implies* (8) *for any sequence $\beta_n = o(1)$ such that $\{T \in \mathcal{T} : d_n(T, \theta) < \beta_n\} \neq \emptyset$ and* (7) *is fulfilled. (cf. 4.2 for a proof)*

2. *Since $d_n(T, \theta) = 1 - d_n(T^c, \theta)$, Assumption* (I) *is equivalent to*

$$\limsup_{n\to\infty} \max_{T\in\mathcal{T}} d_n(T, \theta) < 1.$$

3. *Assume $\mathcal{T}$ is the projection of a class of measurable sets $\mathcal{A} \subseteq \mathcal{P}([0,1]^d)$ onto the grid:*

$$\mathcal{T} = \mathcal{A}_{\mathcal{I}} = \{T \subseteq [0,1]^d |\ \exists A \in \mathcal{A} : T = A_{\mathcal{I}}\},$$

where $\mathcal{A}$ fulfills the following conditions:

$$\begin{aligned} &\sup_{A\in\mathcal{A}} |d_n(A, \theta) - \lambda(A \triangle \theta)| \overset{n\to\infty}{\longrightarrow} 0 \\ \textit{and}& \\ &\sup_{A\in\mathcal{A}} \lambda(A \triangle \theta) < 1 \ \textit{(or equivalently } \inf_{A\in\mathcal{A}} \lambda(A^c \triangle \theta) > 0\textit{)} \end{aligned} \tag{9}$$

Then Assumption (I) *is fulfilled. (cf. 4.2 for a proof)*

4. *The argumentation described in 3. is based on the idea of approximating fixed sets $\mathcal{A}$ using the grid. Naturally, for* (9) *to be fulfilled, one needs the grid to become finer asymptotically. This often translates to requiring $\min_{i=1,\dots,d} N_i \to \infty$ as $n \to \infty$. For example, the first part of* (9) *is fulfilled if*

$$\sup_{A\in\mathcal{A}} \lambda(A(\delta)) \to 0 \qquad \textit{as } \delta \to 0, \tag{10}$$

$\min_{i=1,\dots,d} N_i(n) \to \infty$ *as $n \to \infty$ and $\theta \in \mathcal{A}$. Then, $\partial(A \triangle \theta) \subseteq \partial A \cup \partial\theta$ and thus*

$$(A \triangle \theta)(\delta) \subseteq A(\delta) \cup \theta(\delta), \quad \textit{for any } A \in \mathcal{A} \textit{ and } \delta > 0,$$

implies

$$\begin{aligned} \sup_{A\in\mathcal{A}} |d_n(A, \theta) - \lambda(A \triangle \theta)| &= \sup_{A\in\mathcal{A}} |\lambda((A \triangle \theta)_{\mathcal{I}}) - \lambda(A \triangle \theta)| \\ &\leq \sup_{A\in\mathcal{A}} \lambda\left((A \triangle \theta)\left((\min_{i=1,\dots,d} N_i)^{-1}\right)\right) \\ &\leq 2 \sup_{A\in\mathcal{A}} \lambda\left(A\left((\min_{i=1,\dots,d} N_i)^{-1}\right)\right) \overset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

As pointed out in Mallik (2013), one instance in which assumption (10) *is fulfilled concerns the case of nonempty, closed convex sets* $A \subseteq [0,1]^d$*, since for these sets there is a constant* $c > 0$ *(cf. Dudley (1984), pp. 62–63) such that*

$$\lambda(A(\delta)) \leq \lambda(A^\delta \setminus A_\delta) \leq c\delta, \quad \text{for any } 0 < \delta \leq 1,$$

where $A^\delta = \{\mathbf{x} : \rho(\mathbf{x}, A) < \delta\}$ *and* $A_\delta = \{\mathbf{x} : \rho(\mathbf{x}, A^c) \geq \delta\}$.

5. *Alternatively, since* $\limsup_{n\to\infty} d_n(T^\star, \theta) = 0$ *and* $d_n(T,\theta) \leq d_n(T^\star,\theta) + d_n(T^\star, T)$ *for any* $T \in \mathcal{T}$*, Assumption* (I) *is also fulfilled if*

$$\limsup_{n\to\infty} \max_{T_1,T_2\in\mathcal{T}} d_n(T_1,T_2) < 1$$

(i.e. the sets in $\mathcal{T}$ *do not "span" the whole set* $[0,1]^d$*). If* $\mathcal{T} = \mathcal{T}_\mathcal{I}$ *is the projection of a compact family* $\mathcal{A}$ *of sets onto the grid* $\mathcal{I} = \mathcal{I}_n$*, this can be achieved if the boundaries of sets in* $\mathcal{A}$ *are sufficiently smooth (such that the first part of* (9) *holds) and* $\lambda(A_1 \triangle A_2) < 1$ *for any* $A_1, A_2 \in \mathcal{A}$.

Concerning the assumption (6) on the stochastic process, note that

$$\begin{aligned}\max_{T\in\mathcal{T}} \|B_n(T)\|_2 &= \max_{T\in\mathcal{T}} \frac{1}{|\mathcal{I}_n|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T}(Y_{\mathbf{i},n} - \bar{Y}_n)\right\|_2 \\ &\leq \max_{T\in\mathcal{T}} \frac{1}{|\mathcal{I}_n|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n}\right\|_2 + \left\|\bar{Y}_n\right\|_2,\end{aligned}$$

where $\bar{Y}_n = |\mathcal{I}_n|^{-1} \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}$. Therefore proving (6) reduces to showing $\left\|\bar{Y}_n\right\|_2 = o_P(\xi_n)$, which is fulfilled e.g. for $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2})$ under Assumption (Y), and proving

$$\max_{T\in\mathcal{T}} \frac{1}{|\mathcal{I}_n|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n}\right\|_2 = o_P(\xi_n) \tag{11}$$

or, equivalently

$$\max_{T\in\mathcal{T}} \frac{1}{|\mathcal{I}_n|}\left|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i}}^{(l)}\right| = o_P(\xi_n), \quad \text{for all } l = 1,\ldots,p. \tag{12}$$

The equation (6) can therefore be seen as a type of uniform law of large numbers. For $\xi_n \equiv 1$, it is fulfilled e.g. in the following cases:

Lemma 2.1. *Assume that* $\mathcal{T}$ *has a sufficiently smooth boundary, i.e. it holds for* $r_n(\delta) = \max_{T\in\mathcal{T}} |\mathcal{I}_n|^{-1}|T(\delta)| = \max_{T\in\mathcal{T}} \lambda(T(\delta)_\mathcal{I})$ *that*

$$\lim_{\delta\to 0} \limsup_{n\to\infty} r_n(\delta) = 0. \tag{13}$$

Then the sequence of real-valued, centered random fields $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$ *fulfills* (6) *with* $\xi_n \equiv 1$ *if Assumption* (Y) *holds for* $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$ *and* $\{|Y_{\mathbf{i},n}| - \nu_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$*, where* $\nu_{\mathbf{i},n} = E|Y_{\mathbf{i},n}|$ *with* $\nu = \limsup_{n\to\infty} \max_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} \nu_{\mathbf{i},n} \in (0,\infty)$.

Remark 2.2. • *For instance, the assumption on the smoothness of the boundary is fulfilled if $\mathcal{T}$ is the projection of a class of sets $\mathcal{A}$ on the grid (i.e.*

$$\mathcal{T} = \{T \subseteq [0,1]^d | \, \exists A \in \mathcal{A}: \, A_{\mathcal{I}} = T\})$$

for a sequence of grids $\mathcal{I}_n$ with $\min_{l=1,\dots,d} N_l \to \infty$. Then, if (10) *holds, the assumption* (13) *is satisfied, because*

$$\begin{aligned} T(\delta) &= \{\mathbf{x} : \rho(\mathbf{x}, \partial T) < \delta\} \\ &\subseteq \{\mathbf{x} : \rho(\mathbf{x}, \partial A) < \delta + \rho(\partial A, \partial T)\} \\ &\subseteq \{\mathbf{x} : \rho(\mathbf{x}, \partial A) < \delta + (\min_{l=1,\dots,d} N_l)^{-1}\} \end{aligned}$$

for $A \in \mathcal{A}$ with $T = A_{\mathcal{I}}$, and thus

$$T(\delta)_{\mathcal{I}} = \bigcup_{\kappa_{\mathbf{i},n} \in T(\delta)} C_{\mathbf{i},n} \subseteq \bigcup_{\kappa_{\mathbf{i},n} \in A(\delta + (\min_{l=1,\dots,d} N_l)^{-1})} C_{\mathbf{i},n} \subseteq A\left(\delta + 2(\min_{l=1,\dots,d} N_l)^{-1}\right).$$

Therefore,

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \max_{T \in \mathcal{T}} \lambda(T(\delta)_{\mathcal{I}}) \leq \lim_{\delta \to 0} \limsup_{n \to \infty} \sup_{A \in \mathcal{A}} \lambda\left(A\left(\delta + \frac{2}{\min_{l=1,\dots,d} N_l}\right)\right) = 0.$$

• *Under suitable integrability assumptions, the assumptions of Lemma 2.1 on the stochastic processes are fulfilled for any weak dependence concept that implies* (3) *(or* (2)*) and is such that the dependence assumptions on $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n} \in \mathcal{I}}$ are inherited by the absolute value process. Remark 1.6 gives examples of situations in which (3) (or (2)) is fulfilled, namely under some (BL, θ)-dependence, mixing or physical dependence assumptions, and each of these dependence notions is inherited by the absolute value process.*

Finally, in order to facilitate the application of the results of this section (cf. Section 3), the following corollary gives sufficient conditions for the consistency:

Corollary 2.1. *Let Assumption* (T^*1) *be fulfilled and let $(\xi_n)_{n \in \mathbb{N}}$ be a sequence which fulfills $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2})$, $\liminf_{n \to \infty} \xi_n^{-1} \|a - b\|_2 \sigma_{\mathcal{I}} > 0$ and*

$$\max_{T \in \mathcal{T}} \frac{1}{|\mathcal{I}_n|} \left\| \sum_{\kappa_{\mathbf{i},n} \in T} Y_{\mathbf{i},n} \right\|_2 = o_P(\xi_n).$$

Then $\hat{\theta}_n$ is consistent with respect to ∂_n. If, additionally, Assumption (I) *is fulfilled and Assumption* (T^*2) *holds instead of Assumption* (T^*1)*, then $\hat{\theta}_n$ is consistent with respect to d_n.*

2.2 Rate of convergence

Theorem 2.2. *Let Assumptions* (Y) *and* (T^*2) *be fulfilled and let*

$$|\mathcal{I}_n|^{1/2} \sigma_{\mathcal{I}}^2 \|a - b\|_2 \overset{n \to \infty}{\longrightarrow} \infty. \tag{14}$$

Assume further that $\mathcal{T}$ and α_n are such that

$$\liminf_{n\to\infty} \sigma_{\mathcal{I}} \|a-b\|_2^2 \alpha_n > 0 \tag{15}$$

and

$$\forall \eta > 0 \forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \exists \alpha > 0 \forall n \geq n_0 :$$
$$P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta|} \left\| \sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} \right\|_2 > \varepsilon \sigma_{\mathcal{I}} \|a-b\|_2 \right) \leq \eta \tag{16}$$

Then if $\hat{\theta}_n$ is consistent with respect to d_n (i.e. $d_n(\hat{\theta}_n, \theta) = o_P(1)$) it converges with the following rate:

$$d_n(\hat{\theta}_n, \theta) = \mathcal{O}_P(|\mathcal{I}_n|^{-1}\alpha_n).$$

Here, the consistency is needed to have the identifiability. If we do not estimate θ, but $\partial\theta$, we do not need this restriction:

Theorem 2.3. *Let Assumptions* (Y) *and* (T^*1) *as well as* (14) *be fulfilled. For any $T \in \mathcal{T}$, choose $\bar{T} \in \{T, T^c\}$ such that $d_n(\bar{T}, \theta) \leq d_n(\bar{T}^c, \theta)$. Assume further that $\mathcal{T}$ and α_n are such that* (15) *is satisfied and*

$$\forall \eta > 0 \forall \varepsilon > 0 \exists n_0 \in \mathbb{N} \exists \alpha > 0 \forall n \geq n_0 :$$
$$P\left(\max_{T\in\mathcal{T}:|\bar{T}\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|\bar{T}\triangle\theta|} \left\| \sum_{\kappa_{\mathbf{i},n}\in \bar{T}} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} \right\|_2 > \varepsilon \sigma_{\mathcal{I}} \|a-b\|_2 \right) \leq \eta \tag{17}$$

Then $\hat{\theta}_n$ converges with respect to ∂_n at the rate

$$\partial_n(\hat{\theta}_n, \theta) = \mathcal{O}_P(|\mathcal{I}_n|^{-1}\alpha_n).$$

Theorems 2.2 and 2.3 reduce the question of finding a rate of convergence to proving a maximal inequality. In the next section, we will consider some examples where such maximal inequalities hold for special classes of $\mathcal{T}$. However, in order to give some idea of how to prove such inequalities for general classes of sets, we finish this section with an example of (16) under mixing conditions on Y. A straightforward method to prove such a maximal inequality (which is employed e.g. in Carlstein and Krishnamoorthy (1992)) is to first prove an exponential inequality for the set-indexed partial sums and to then ensure that the number of sets over which the maximum is taken is not "too large". In the special case when the grid is uniform (cf. Subsection 4.1) and the observations fulfill some mixing conditions, the following lemma gives an example for (16) using an exponential inequality. (Trivially, the same arguments could be used to obtain (17).) For simplicity, we restrict ourselves to non-vanishing changes as specified by Assumption (C).

Lemma 2.2. *Consider a uniform grid $\mathcal{I}_n$ with $N_1 = \cdots = N_d = n$. Let Assumptions* (C) *and* (T^*2) *be fulfilled for $\alpha_n = n^\eta$ ($\eta > 0$) and let $card(\mathcal{T}) = o(\exp(\mu n^\xi))$ for some $\mu > 0$ and $0 < \xi < \frac{1}{2}\eta$. Assume further that $\{\|Y_{\mathbf{i},n}\|_2^r\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n, n\in\mathbb{N}}$ is uniformly integrable for*

$$r > \frac{d}{\eta} \quad \textit{with} \quad r \geq \max\left\{2, \frac{1}{1-\frac{\eta-d}{\xi-d\delta}}\right\}$$

for $\delta \in (0,1)$ such that $\xi - d\delta < \eta - d$. If $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}$ is either

- *α-mixing with* $\alpha(x) = \mathcal{O}\left(e^{-2\log\left((2x)^{\frac{d\delta}{1-\delta}}\right)(2x)^{\frac{d\delta}{1-\delta}}}\right)$, *or*
- *nonuniform φ-mixing with* $\varphi(x) = \mathcal{O}(x^{-\gamma})$ *with* $\gamma > \max\{\frac{d}{1-\delta}, \frac{r}{r-1}(d-1)\}$,

then (16) *holds.*

Remark 2.3. *As seen in the proof of the Lemma, the specific size of the mixing coefficients is irrelevant as long as they fulfill*

$$p_n^d \theta_{q_n}\left(\frac{n}{2p_n}\right) = \begin{cases} p_n^d \alpha^{1/(1+p_n^d)}\left(\frac{n}{2p_n}\right), & \text{under } \alpha\text{-mixing} \\ 2^{-d} n^d \varphi\left(\frac{n}{2p_n}\right), & \text{under } \varphi\text{-mixing} \end{cases} = \mathcal{O}(1),$$

and are such that (2) *is satisfied. For instance, the latter is the case if*

$$\sum_{h=0}^{\infty} h^{d-1}\alpha(h)^{\frac{r-2}{r}} < \infty,$$

respectively $\sum_{h=0}^{\infty} h^{d-1}\varphi(h)^{\frac{r-1}{r}} < \infty$.

3 Examples

In order to illustrate the applicability of the main results, the following section contains examples of classes of sets for which the assumptions of the theorems are explicitly verified.

3.1 Example 1: Rectangles

Assume the candidate sets $\mathcal{T}$ as well as θ are rectangles, i.e. $\theta = (\boldsymbol{\theta}_1^0, \boldsymbol{\theta}_2^0]$ for $\underline{\mathbf{0}} < \boldsymbol{\theta}_1^0 < \boldsymbol{\theta}_2^0 < \underline{\mathbf{1}}$ and $\theta_{\mathcal{I}} = \mathbf{N}^{-1}(\lfloor \mathbf{N}\boldsymbol{\theta}_1^0 \rfloor, \lfloor \mathbf{N}\boldsymbol{\theta}_2^0 \rfloor] = \mathbf{N}^{-1}(\mathbf{k}_1^0, \mathbf{k}_2^0]$, and

$$\mathcal{T} = \mathcal{A}_{\mathcal{I}} = \{\mathbf{N}^{-1}(\mathbf{k}_1, \mathbf{k}_2] : \underline{\mathbf{0}} \le \mathbf{k}_1 < \mathbf{k}_2 \le \mathbf{N}\}$$

is the projection of $\mathcal{A} = \{(\mathbf{s}, \mathbf{t}] : \underline{\mathbf{0}} \le \mathbf{s} \le \mathbf{t} \le \underline{\mathbf{1}}\}$ onto the grid. Assume additionally that the $\mathcal{I}_n$ are such that $n \to \infty$ implies $\min_{i=1,\dots,d} N_i \to \infty$.

Note that in this model, Assumptions (T^*1) and (T^*2) are satisfied for any α_n, since $\theta_{\mathcal{I}} \in \mathcal{T}$.

Consistency: Let Assumption (Y(r)) be fulfilled for $\mathcal{S} = \mathcal{T}$. Then (6) is satisfied for

$$\xi_n^{-1} = \begin{cases} o\left(|\mathcal{I}_n|^{1/2}\left(\prod_{i=1}^d \log(N_i)\right)^{-1}\right), & \text{for } r = 2 \\ o(|\mathcal{I}_n|^{1/2}), & \text{for } r > 2. \end{cases}$$

If $|\mathcal{I}_n|^{1/2}\left(\prod_{i=1}^d \log(N_i)\right)^{-1}\|a-b\|_2\sigma_{\mathcal{I}} \to \infty$ (or, correspondingly, $|\mathcal{I}_n|^{1/2}\|a-b\|_2\sigma_{\mathcal{I}} \to \infty$ for $r > 2$), Corollary 2.1 implies consistency under both ∂_n and d_n.

Proof. For $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2})$, we know that (6) is fulfilled if (11) holds. The maximal inequality (11) follows from Assumption (Y(r)) and Corollaries 1 and 3 in Móricz (1983) (although Móricz (1983) considers real-valued random variables, his proofs can easily be extended to $\mathbb{R}^p$-valued observations, and indeed to observations in any normed space, since they rely on the triangle inequality rather than a specific property of $\mathbb{R}$):

$$\max_{\underline{\mathbf{0}}\le\mathbf{k}<\mathbf{m}\le\mathbf{N}} \left\| \sum_{\kappa_{\mathbf{i},n}\in\mathbf{N}^{-1}(\mathbf{k},\mathbf{m}]} Y_{\mathbf{i},n} \right\|_2 = \begin{cases} \mathcal{O}_P\left(|\mathcal{I}_n|^{1/2}\prod_{i=1}^d \log(N_i)\right), & r=2 \\ \mathcal{O}_P(|\mathcal{I}_n|^{1/2}), & r>2 \end{cases}$$

To apply Corollary 2.1, note that Assumptions (T^*1) and (T^*2) are satisfied for any α_n, and that the assumptions on $\|a-b\|_2\sigma_{\mathcal{I}}$ make it possible to choose $\xi_n = \|a-b\|_2\sigma_{\mathcal{I}}$. Assumption (I) is fulfilled as a special case of the general result in Remark 2.1: Since on the compact set $\{(\mathbf{s},\mathbf{t}) : \underline{\mathbf{0}} \le \mathbf{s} \le \mathbf{t} \le \underline{\mathbf{1}}\}$, the function $(\mathbf{s},\mathbf{t}) \mapsto \lambda((\mathbf{s},\mathbf{t}]\triangle(\boldsymbol{\theta}_1^0,\boldsymbol{\theta}_2^0])$ is continuous,

$$\begin{aligned} &\sup_{A\in\mathcal{A}} |d_n(A,\theta) - \lambda(A\triangle\theta)| \\ =& \sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} \left|\lambda\left(\mathbf{N}^{-1}\left((\lfloor\mathbf{N}\mathbf{s}\rfloor,\lfloor\mathbf{N}\mathbf{t}\rfloor]\triangle\left(\lfloor\mathbf{N}\boldsymbol{\theta}_1^0\rfloor,\lfloor\mathbf{N}\boldsymbol{\theta}_2^0\rfloor\right]\right)\right) - \lambda(A\triangle\theta)\right| \longrightarrow 0 \end{aligned}$$

for $\min_{i=1,\dots,d} N_i \to \infty$, and

$$\sup_{A\in\mathcal{A}} \lambda(A\triangle\theta) = \sup_{\underline{\mathbf{0}}\le\mathbf{s}\le\mathbf{t}\le\underline{\mathbf{1}}} \lambda((\mathbf{s},\mathbf{t}]\triangle(\boldsymbol{\theta}_1^0,\boldsymbol{\theta}_2^0]) < 1,$$

because $\lambda((\mathbf{s},\mathbf{t}]\triangle(\boldsymbol{\theta}_1^0,\boldsymbol{\theta}_2^0]) < 1$ for any $\underline{\mathbf{0}} \le \mathbf{s} \le \mathbf{t} \le \underline{\mathbf{1}}$. Finally, $\theta \in \mathcal{A}$ per assumption. □

Rate of convergence: Let Assumptions (C) and (Y(r)) with $r > 2$ and $\mathcal{S} = \mathcal{T}$ be fulfilled. Then we obtain (16) with a rate $\alpha_n = \alpha$ independent of n by using a result analogous to Bucchia and Heuser (2015). Theorem 2.2 therefore yields $d_n(\hat{\theta}_n,\theta) = \mathcal{O}_P(|\mathcal{I}_n|^{-1})$. Due to $\partial_n(\hat{\theta}_n,\theta) \le d_n(\hat{\theta}_n,\theta)$, this implies the same rate for ∂_n. (Since in this case the consistency with respect to d_n requires no additional assumptions and (T^*1) and (T^*2) are fulfilled by assumption, using Theorem 2.2 instead of Theorem 2.3 is no restriction.)

Remark 3.1. *Note that in contrast to Bucchia and Heuser (2015), we do not measure the symmetric difference of the sets directly, but rather the number of misclassified grid nodes. This allows us to derive an improved convergence rate, since in this setting, the perfect estimation of the change-rectangle (i.e. its anchoring on the pixels) is possible. Because the mean-function and long-run variance estimators in Bucchia and Heuser (2015) do not depend on the actual change-set estimator but only on the grid points contained within it, the convergence rate derived here can be used for the results in Bucchia and Heuser (2015).*

Proof. For $T \in \mathcal{T}$, write $T = \mathbf{N}^{-1}(\mathbf{k}_1,\mathbf{k}_2] = \mathbf{N}^{-1}R_{\mathbf{k}}$ and $\theta_{\mathcal{I}} = \mathbf{N}^{-1}(\mathbf{k}_1^0,\mathbf{k}_2^0] = \mathbf{N}^{-1}R_{\mathbf{k}^0}$. Then we observe the following:

1. $|T\triangle\theta| = \lambda(R_{\mathbf{k}} \setminus R_{\mathbf{k}^0}) + \lambda(R_{\mathbf{k}^0} \setminus R_{\mathbf{k}})$
2. $|T\triangle\theta| > 0 \Rightarrow \|\mathbf{k}-\mathbf{k}^0\|_\infty > 0$ and therefore also $\|\mathbf{k}-\mathbf{k}^0\|_\infty \ge 1$

3. Analogously to Bucchia and Heuser (2015), Lemma A.1, it can be proven by induction that
$$|T\triangle\theta| \geq C\prod_{l=1}^{d} N_l \left\|\frac{\mathbf{k}-\mathbf{k}^0}{\mathbf{N}}\right\|_\infty$$
for some $C > 0$ that may depend on d but is independent of n and T.

Now, the inequality follows analogously to the proof of Lemma A.2 in Bucchia and Heuser (2015) on the pages 125/126, where we replace $T_{\mathbf{k}_1,\mathbf{k}_2}$ by $|T\triangle\theta|$:

$$\begin{aligned}
&\max_{T\in\mathcal{T},|T\triangle\theta|\geq\alpha} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|} \\
&= \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}} \frac{\left\|\sum_{\mathbf{i}\in R_{\mathbf{k}}} Y_{\mathbf{i},n} - \sum_{\mathbf{i}\in R_{\mathbf{k}^0}} Y_{\mathbf{i},n}\right\|_2}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})} \\
&= \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}} \frac{\left\|\sum_{\boldsymbol{\epsilon}\in\{0,1\}^d}(-1)^{d-\sum_{l=1}^d \epsilon_l}\left(\sum_{\mathbf{i}\leq\mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)} Y_{\mathbf{i},n} - \sum_{\mathbf{i}\leq\mathbf{k}_1^0+\boldsymbol{\epsilon}(\mathbf{k}_2^0-\mathbf{k}_1^0)} Y_{\mathbf{i},n}\right)\right\|_2}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})} \\
&\leq \sum_{\boldsymbol{\epsilon}\in\{0,1\}^d} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}} \left\|\frac{1}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})}\sum_{\mathbf{i}\in(\underline{\mathbf{0}},\mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)]\setminus(\underline{\mathbf{0}},\mathbf{k}_1^0+\boldsymbol{\epsilon}(\mathbf{k}_2^0-\mathbf{k}_1^0)]} Y_{\mathbf{i},n}\right\|_2 \\
&+ \sum_{\boldsymbol{\epsilon}\in\{0,1\}^d} \max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}} \left\|\frac{1}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})}\sum_{\mathbf{i}\in(\underline{\mathbf{0}},\mathbf{k}_1^0+\boldsymbol{\epsilon}(\mathbf{k}_2^0-\mathbf{k}_1^0)]\setminus(\underline{\mathbf{0}},\mathbf{k}_1+\boldsymbol{\epsilon}(\mathbf{k}_2-\mathbf{k}_1)]} Y_{\mathbf{i},n}\right\|_2 \\
&= T_1 + T_2
\end{aligned}$$

We show the convergence for the term T_1, the proof for T_2 is analogous and is therefore omitted. We adopt the notation $M_{\mathbf{a}} = (\underline{\mathbf{0}},\mathbf{a}] \setminus (\underline{\mathbf{0}},\mathbf{a}^0]$ for

$$\mathbf{a} \in I_{\mathbf{k}_1,\mathbf{k}_2} = \{\mathbf{a}\in\mathbb{Z}^d:\ a^{(l)} \in \{k_1^{(l)}, k_2^{(l)}\},\ l = 1,\dots,d\},$$

(cf. Bucchia and Heuser (2015)) and note that

$$\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0}) \geq C\prod_{l=1}^{d} N_l \left\|\frac{\mathbf{k}-\mathbf{k}^0}{\mathbf{N}}\right\|_\infty \geq C\prod_{l=1}^{d} N_l \left\|\frac{\mathbf{a}-\mathbf{a}^0}{\mathbf{N}}\right\|_\infty.$$

Then an analogous proof to Bucchia and Heuser (2015), p. 123, yields

$$\lambda(M_{\mathbf{a}}) \leq c\prod_{l=1}^{d} N_l \left\|\frac{\mathbf{a}-\mathbf{a}^0}{\mathbf{N}}\right\|_\infty$$

for some $c > 0$. For $d = 1$, we obtain

$$\mathbb{P}\left(\max_{\substack{k_1<k_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}}\left\|\frac{1}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})}\sum_{i\in(0,a]\setminus(0,a^0]}Y_{i,n}\right\|_2\geq\varepsilon\right)$$

$$\leq\left(\frac{\varepsilon}{2}\right)^{-r}\mathbb{E}\left[\max_{1\leq a-a^0\leq\lfloor\alpha\rfloor}\left|\frac{1}{\alpha}\sum_{i=a^0+1}^{a}Y_{i,n}^{(l)}\right|^r\right]$$

$$+\left(\frac{\varepsilon}{2}\right)^{-r}\mathbb{E}\left[\max_{a-a^0\geq\alpha}\left|\frac{1}{a-a^0}\sum_{i=a^0+1}^{a}Y_{i,n}^{(l)}\right|^r\right]$$

$$\leq K_r\left(\left(\frac{\varepsilon}{2}\right)^{-r}\sum_{l=1}^{\lfloor\alpha\rfloor}\frac{1}{\alpha^{r/2}}+\left(\frac{\varepsilon}{2}\right)^{-r}\sum_{i=\lfloor\alpha\rfloor+1}^{\infty}\frac{1}{i^{r/2}}\right)$$

$$\leq K_r\left(\frac{\varepsilon}{2}\right)^{-r}\left(\alpha^{1-r/2}+\sum_{i=\lfloor\alpha\rfloor+1}^{\infty}\frac{1}{i^{r/2}}\right)\overset{\alpha\to\infty}{\longrightarrow}0,$$

where $\varepsilon > 0$ is some constant and we have used (4) and the fact that $|T\triangle\theta| \geq |a - a^0|$. For $d \geq 2$, we first observe that for $l \in \{1, \ldots, d\}$ and $h \in \{1, \ldots, \max_{j=1\ldots,d} N_j\}$

$$\#\{\mathbf{a} : \underline{\mathbf{0}} \leq \mathbf{a} \leq \mathbf{N},\ \|\mathbf{N}^{-1}(\mathbf{a}-\mathbf{a}^0)\|_\infty = N_l^{-1}\left|a^{(l)}-a^{0(l)}\right| = N_l^{-1}h\} \leq \tilde{C}\prod_{j\neq l}N_j,$$

for some constant $\tilde{C} > 0$. Now, Markov's inequality and (4) yield

$$\mathbb{P}\left(\max_{\substack{\mathbf{k}_1<\mathbf{k}_2\\ \lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})\geq\alpha}}\left\|\frac{1}{\lambda(R_{\mathbf{k}}\triangle R_{\mathbf{k}^0})}\sum_{\mathbf{i}\in M_{\mathbf{a}}}Y_{\mathbf{i},n}\right\|_2\geq\varepsilon\right)$$

$$\leq\mathbb{P}\left(\max_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\mathbf{N}\\ \|\mathbf{a}-\mathbf{a}^0\|_\infty\geq1}}\left\|\frac{1}{C\prod_{l=1}^d N_l\left\|\frac{\mathbf{a}-\mathbf{a}^0}{\mathbf{N}}\right\|_\infty}\sum_{\mathbf{i}\in M_{\mathbf{a}}}Y_{\mathbf{i},n}\right\|_2\geq\varepsilon\right)$$

$$\leq\varepsilon^{-r}\sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\mathbf{N}\\ \|\mathbf{a}-\mathbf{a}^0\|_\infty\geq1}}\tilde{c}\frac{1}{\left(\prod_{l=1}^d N_l\left\|\frac{\mathbf{a}-\mathbf{a}^0}{\mathbf{N}}\right\|_\infty\right)^{r/2}}$$

$$\leq\tilde{c}\varepsilon^{-r}\sum_{l=1}^{d}\sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\mathbf{N}\\ \|\mathbf{N}^{-1}(\mathbf{a}-\mathbf{a}^0)\|_\infty=N_l^{-1}|a^{(l)}-a^{0(l)}|>0}}\frac{1}{\left(\prod_{j\neq l}N_j\left|a^{(l)}-a^{0(l)}\right|\right)^{r/2}}$$

$$=\tilde{c}\varepsilon^{-r}\sum_{l=1}^{d}\sum_{h=1}^{N_l}\sum_{\substack{\underline{\mathbf{0}}\leq\mathbf{a}\leq\mathbf{N}\\ \|\mathbf{N}^{-1}(\mathbf{a}-\mathbf{a}^0)\|_\infty=N_l^{-1}|a^{(l)}-a^{0(l)}|=hN_l^{-1}}}\left(\prod_{j\neq l}N_j\right)^{-r/2}h^{-r/2}$$

$$\leq\varepsilon^{-r}\tilde{c}\tilde{C}\sum_{l=1}^{d}\left(\prod_{j\neq l}N_j\right)^{1-r/2}\sum_{h=1}^{N_l}\frac{1}{h^{r/2}}\to0,$$

for $\min_{l=1,\ldots,d} N_l \to \infty$, $\varepsilon > 0$ and some constant $\tilde{c} > 0$. Therefore, (16) is satisfied.

For the rate of convergence, note that Assumption (C) implies (14) and (15), and that the previous paragraph shows the consistency under d_n. □

3.2 Example 2: Unions of aggregated pixels

In order to consider change-sets with less form constraints, we introduce sets that result from arbitrary unions of "aggregated pixels", which serve as a coarser partition of $(0,1]^d$ than the $C_{\mathbf{i},n}$. This corresponds to the approach by Müller and Song (1994), who derived rates of convergence for change-boundaries of this form for i.i.d. real-valued observations. Such a choice of model has the added advantage that it lends itself to iterative algorithms for easier computation of the estimators (cf. Müller and Song (1994)).

Choose $\mathbf{M} = \mathbf{M}(n) \in \mathbb{N}^d$, $\mathbf{M} \leq \mathbf{N}$, and divide $[0,1]^d$ into sets $C_{\mathbf{j},n}^{(m)} = \mathbf{M}^{-1}(\mathbf{j} - \underline{\mathbf{1}}, \mathbf{j}]$, where $m = m_n = \prod_{i=1}^d M_i$ and $\mathbf{M}$ is chosen in such a way that the $C_{\mathbf{j},n}^{(m)}$ are anchored on the pixels (i.e. $C_{\mathbf{j},n,\mathcal{I}}^{(m)} = C_{\mathbf{j},n}^{(m)}$). Then $(C_{\mathbf{j},n}^{(m)})_{\underline{\mathbf{1}} \leq \mathbf{j} \leq \mathbf{M}}$ forms a partition of $(0,1]^d$ into subrectangles of equal size with

$$\lambda(C_{\mathbf{j},n}^{(m)}) = |\mathcal{I}_n|^{-1}|C_{\mathbf{j},n}^{(m)}| = \frac{1}{m}, \quad \mathbf{j} \leq \mathbf{M}.$$

Define a similar "anchoring mapping" on rectangles of aggregated pixels by setting

$$\phi_n^{(m)}(F) = \bigcup_{F_{\mathcal{I}} \cap C_{\mathbf{j},n}^{(m)} \neq \emptyset} C_{\mathbf{j},n}^{(m)}$$

for $F \in \mathcal{B}([0,1]^d)$. Denote the collection of aggregated pixels by

$$\mathcal{C}_n = \{C_{\mathbf{j},n}^{(m)} : \underline{\mathbf{1}} \leq \mathbf{j} \leq \mathbf{M}\}$$

and the collection of (nonempty) sets anchored on the aggregated pixels by

$$\mathcal{A}_n = \{A \subseteq [0,1]^d : \phi_n^{(m)}(A) = A \neq \emptyset\}.$$

Let $\mathcal{T} \subseteq \mathcal{A}_n$.

Consistency: Let Assumption (Y(r)) be fulfilled for $\mathcal{S} = \mathcal{C}_n$. Then (6) is satisfied for $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2} m_n^{-\left(\frac{1}{r}+\frac{1}{2}\right)})$. Under Assumption (T^*1), Corollary 2.1 implies consistency under ∂_n if $|\mathcal{I}_n|^{1/2} m_n^{-\left(\frac{1}{r}+\frac{1}{2}\right)} \|a-b\|_2 \sigma_{\mathcal{I}} \to \infty$, i.e. $m_n = o\left(\left(|\mathcal{I}_n|\sigma_{\mathcal{I}}^2\|a-b\|_2^2\right)^{\frac{r}{r+2}}\right)$. If Assumption (T^*2) instead of (T^*1) is satisfied and, additionally, Assumption (I) holds, we also have consistency under d_n in this setting. Assumption (I) is, for instance, satisfied if $\mathcal{T}$ is the projection of a set $\mathcal{A} \subseteq \mathcal{P}([0,1]^d)$ onto the aggregated pixels

$$\mathcal{T} = \phi_n^{(m)}(\mathcal{A}) = \{T \subseteq [0,1]^d | \; \exists A \in \mathcal{A} : T = \phi_n^{(m)}(A)\},$$

where $\mathcal{A}$ fulfills the following conditions:

(i) $\mathcal{A}$ has a sufficiently smooth boundary, i.e.

$$\sup_{A \in \mathcal{A}} |d_n(A,\theta) - \lambda(A \triangle \theta)| \overset{n\to\infty}{\longrightarrow} 0$$

and $\sup_{A\in\mathcal{A}} |\mathcal{I}_n|^{-1}|\mathcal{P}_{\mathcal{I}^{(m)}}(A)| \overset{n\to\infty}{\longrightarrow} 0$, where

$$\mathcal{P}_{\mathcal{I}^{(m)}}(A) = \bigcup_{\substack{\mathbf{j}: C_{\mathbf{j},n}^{(m)} \cap A_{\mathcal{I}} \neq \emptyset \\ C_{\mathbf{j},n}^{(m)} \cap A_{\mathcal{I}}^c \neq \emptyset}} C_{\mathbf{j},n}^{(m)}$$

denotes the perimeter of a set $A \in \mathcal{A}$.

(ii) $\theta \in \mathcal{A}$

(iii) $\sup_{A\in\mathcal{A}} \lambda(A\triangle\theta) < 1$ (or equivalently $\inf_{A\in\mathcal{A}} \lambda(A^c\triangle\theta) > 0$)

Remark 3.2.

- *The notion of the perimeter of a set was introduced by Carlstein and Krishnamoorthy (1992). Unlike our current setting, Carlstein and Krishnamoorthy (1992) use the Lebesgue measure of the difference of sets as a measure for the distance of their estimator to the true change-boundary. Since their estimator is by necessity anchored on the pixels, they need assumptions on the smoothness of the perimeter to ensure that the estimator could in theory be asymptotically close to the true change-set at a sufficient rate. They note that this kind of assumption is fulfilled, for instance, for two-dimensional boundaries that can be expressed as rectifiable curves (cf. Carlstein and Krishnamoorthy (1992), Theorem 3).*

- *Noting that*

$$\mathcal{P}_{\mathcal{I}^{(m)}}(A) \subseteq A\left(\frac{1}{\min_{i=1,\ldots,d} N_i} + \frac{1}{\min_{i=1,\ldots,d} M_i}\right),$$

we observe that the assumption (i) is fulfilled e.g. if $\theta \in \mathcal{A}$ and (10) hold, and

$$\min_{i=1,\ldots,d} N_i(n) \to \infty \quad \textit{as well as} \quad \min_{i=1,\ldots,d} M_i(n) \to \infty$$

as $n \to \infty$ (cf. also Remark 2.1).

Proof. For (6), it is sufficient to show (11). The latter is a simple consequence of Lemma 4.3:

$$\begin{aligned}
\max_{A\in\mathcal{A}_n} \frac{1}{|\mathcal{I}_n|}\left\|\sum_{\kappa_{\mathbf{i},n}\in A} Y_{\mathbf{i},n}\right\|_2 &\leq \frac{m_n}{|\mathcal{I}_n|}\max_{C\in\mathcal{C}_n}\left\|\sum_{\kappa_{\mathbf{i},n}\in C} Y_{\mathbf{i},n}\right\|_2 \\
&\leq \frac{m_n}{|\mathcal{I}_n|}\mathcal{O}_P\left(|\mathcal{I}_n|^{1/2} m_n^{\frac{1}{r}-\frac{1}{2}}\right) \\
&= \mathcal{O}_P\left(|\mathcal{I}_n|^{-1/2} m_n^{\frac{1}{r}+\frac{1}{2}}\right)
\end{aligned}$$

To apply Corollary 2.1, we again use the fact that the assumptions on $\|a-b\|_2\sigma_{\mathcal{I}}$ make it possible to choose $\xi_n = \|a-b\|_2\sigma_{\mathcal{I}}$. For the proof of Assumption (I) under the assumptions above, note that for any set $A \in \mathcal{A}$,

$$d_n(\phi_n^{(m)}(A),\theta) \leq d_n(\phi_n^{(m)}(A),A) + d_n(A,\theta) = |\mathcal{I}_n|^{-1}|\mathcal{P}_{\mathcal{I}^{(m)}}(A)| + d_n(A,\theta),$$

and therefore (i) implies

$$\sup_{A\in\mathcal{A}} |d_n(\phi_n^{(m)}(A),\theta) - \lambda(A\triangle\theta)| \overset{n\to\infty}{\longrightarrow} 0.$$

For $\mathcal{T} = \phi_n^{(m)}(\mathcal{A})$, we use $\max_{T\in\mathcal{T}} d_n(T,\theta) = \sup_{A\in\mathcal{A}} d_n(\phi_n^{(m)}(A),\theta)$ to obtain

$$\begin{aligned}
&\limsup_{n\to\infty} \max_{T\in\mathcal{T}} d_n(T,\theta)\\
=&\limsup_{n\to\infty} \sup_{A\in\mathcal{A}} (\lambda(A\triangle\theta) + (d_n(\phi_n^{(m)}(A),\theta) - \lambda(A\triangle\theta)))\\
\leq& \sup_{A\in\mathcal{A}} \lambda(A\triangle\theta) + \limsup_{n\to\infty} \sup_{A\in\mathcal{A}} |d_n(\phi_n^{(m)}(A),\theta) - \lambda(A\triangle\theta)|\\
=& \sup_{A\in\mathcal{A}} \lambda(A\triangle\theta) < 1.
\end{aligned}$$

□

Rate of convergence: Let Assumptions (T^*2) and (Y(r)) be fulfilled and suppose (15) holds. Then (16) is fulfilled for $m_n = o\left(\left(|\mathcal{I}_n|\sigma_{\mathcal{I}}^2\|a-b\|_2^2\right)^{\frac{r}{r+2}}\right)$. Replacing Assumption $((T^*2))$ by $((T^*1))$, we obtain (17) analogously. For this choice of m_n, if (14), (15) and Assumption (T^*1) hold, Theorem 2.3 yields the convergence rate $\alpha_n|\mathcal{I}_n|^{-1}$ for ∂_n. If Assumption (I) is also satisfied and we replace (T^*1) by (T^*2), Theorem 2.2 yields the same rate of convergence for d_n.

Proof. We show (16), noting that (17) can be proven the same way, since $T\in\mathcal{A}_n$ implies $\bar{T}\in\mathcal{A}_n$. We differentiate between two cases for the change-set and use the special structure of the sets in $\mathcal{A}_n$:
Case 1: $\theta_{\mathcal{I}} = \phi_n^{(m)}(\theta)$
Since $T\setminus\theta_{\mathcal{I}}, \theta_{\mathcal{I}}\setminus T, T\triangle\theta_{\mathcal{I}} \in \mathcal{A}_n$ for any $T\in\mathcal{T}$,

$$\begin{aligned}
&\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2\\
\leq& 2\max_{A\in\mathcal{A}_n} \frac{1}{|A|}\left\|\sum_{\kappa_{\mathbf{i},n}\in A} Y_{\mathbf{i},n}\right\|_2 = \mathcal{O}_P\left(|\mathcal{I}_n|^{-1/2} m_n^{\frac{1}{r}+\frac{1}{2}}\right),
\end{aligned}$$

by Lemma 4.3. Therefore, (16) is fulfilled for $m_n = o\left(\left(|\mathcal{I}_n|\sigma_{\mathcal{I}}^2\|a-b\|_2^2\right)^{\frac{r}{r+2}}\right)$.

Case 2: $\theta_{\mathcal{I}} \neq \phi_n^{(m)}(\theta)$, but $T^\star$ fulfills Assumption (T^*2)
Since $T^\star \in \mathcal{A}_n$, it holds that $T \setminus T^\star, T^\star \setminus T, T\triangle T^\star \in \mathcal{A}_n$ for any $T \in \mathcal{T}$, and therefore,

$$\begin{aligned}
&\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 \\
\leq &\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{|T\triangle T^\star|}{|T\triangle\theta|}\frac{1}{|T\triangle T^\star|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in T^\star} Y_{\mathbf{i},n}\right\|_2 \\
&+ \max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T^\star} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 \\
\leq &2\max_{T\in\mathcal{T}:|T\triangle T^\star|\geq(\alpha-1/6\sigma_{\mathcal{I}})\alpha_n} \frac{1}{|T\triangle T^\star|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in T^\star} Y_{\mathbf{i},n}\right\|_2 \\
&+ \frac{1}{\alpha\alpha_n}\left\{\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T^\star} Y_{\mathbf{i},n}\right\|_2 + \left\|\sum_{\kappa_{\mathbf{i},n}\in T^\star\setminus\theta} Y_{\mathbf{i},n}\right\|_2\right\},
\end{aligned}$$

where we have used $|T\triangle T^\star| \leq |T\triangle\theta| + |\theta\triangle T^\star| \leq 2|T\triangle\theta|$ and, by Assumption (T^*2), $|T\triangle T^\star| \geq |T\triangle\theta| - |\theta\triangle T^\star| \geq \alpha_n(\alpha - 1/6\sigma_{\mathcal{I}})$, for $T \in \mathcal{T}$ with $|T\triangle\theta| \geq \alpha\alpha_n$. By Assumptions (Y) and (T^*2), the second summand vanishes for $\alpha \to \infty$:

$$\begin{aligned}
&P\left(\frac{1}{\alpha\alpha_n}\left\{\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T^\star} Y_{\mathbf{i},n}\right\|_2 + \left\|\sum_{\kappa_{\mathbf{i},n}\in T^\star\setminus\theta} Y_{\mathbf{i},n}\right\|_2\right\} > \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right) \\
\leq &\frac{1}{\varepsilon^2\alpha^2\alpha_n^2}\sigma_{\mathcal{I}}^{-2}\|a-b\|_2^{-2}|\mathcal{I}_n|d_n(T^\star,\theta) \\
\leq &\frac{1}{\varepsilon^2}\frac{1}{6}\left(\sigma_{\mathcal{I}}\|a-b\|_2^2\alpha_n\right)^{-1}\frac{1}{\alpha^2} \overset{\alpha\to\infty}{\longrightarrow} 0,
\end{aligned}$$

for any $\varepsilon > 0$ and n large enough, since $\liminf_{n\to\infty}\sigma_{\mathcal{I}}\|a-b\|_2^2\alpha_n > 0$. The first summand can be treated analogously to case 1.

For the rate of convergence, note that Assumption (I) and the choice of m_n yield the consistency for d_n. □

3.3 Example 3: Nested sets

We now consider the special case when $\mathcal{T}$ is the disjoint union of finitely many classes of nested sets. Nested sets are of interest for instance if one considers parametric classes of sets that are defined by a location parameter and a scaling parameter such that for a fixed location, sets with different scaling are nested (e.g. circles, ellipses,...). To be precise, we consider the model in Corollary 2.4 of Ferger (2004), i.e.:

$$\exists M > 0, (v_n)_{n\in\mathbb{N}} \subseteq \mathbb{N}\ \forall n\in\mathbb{N} : v_n \leq M \quad \text{and} \quad \mathcal{T} = \sum_{j=1}^{v_n}\mathcal{T}_n^{(j)},$$

where each of the disjoint sub-classes $\mathcal{T}_n^{(j)}$, $1 \leq j \leq v_n$, is ordered.

Consistency: Let Assumption (Y(r)) with $\mathcal{S} = \mathcal{T}$ be fulfilled. Then (6) is satisfied with $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2} \log^{-1}(|\mathcal{I}_n|))$ for $r = 2$ and $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2})$ for $r > 2$. Under Assumption (T^*1), Corollary 2.1 implies consistency under ∂_n if

$$|\mathcal{I}_n|^{1/2} \log^{-1}(|\mathcal{I}_n|)\|a-b\|_2\sigma_{\mathcal{I}} \to \infty, \qquad \text{for } r=2,$$
$$|\mathcal{I}_n|^{1/2}\|a-b\|_2\sigma_{\mathcal{I}} \to \infty, \qquad \text{for } r>2.$$

If Assumptions (T^*2) and (I) are fulfilled, too, the consistency also holds under d_n. For Assumption (I), this is the case e.g. if $\liminf_{n\to\infty} \sigma_{\mathcal{I}} > \tilde{\sigma}$ for some $\tilde{\sigma} > 0$ and $\theta_{\mathcal{I}} \subseteq T$ or $T \subseteq \theta_{\mathcal{I}}$ holds for all $T \in \mathcal{T}$ and n.

Proof. Since $(v_n)_{n\in\mathbb{N}}$ is bounded, we can assume w.l.o.g. that $v_n \equiv v = 1$ for the proof of (6). Then we can assume w.l.o.g., that $\mathcal{T} = \mathcal{T}_n^{(1)} = \{T_{1,n}, \ldots, T_{m,n}\}$ for some $m \in \mathbb{N}$, with

$$\emptyset = T_{0,n} \subseteq T_{1,n} \subsetneq T_{2,n} \subsetneq \cdots \subsetneq T_{m,n} \subseteq T_{m+1,n} = [0,1]^d,$$

and $0 < |T_{1,n}| < |T_{2,n}| < \cdots < |T_{m,n}| \leq |\mathcal{I}_n|$. For $\xi_n^{-1} = o(|\mathcal{I}_n|^{1/2})$, we know that (6) is fulfilled iff (11) holds. An application of Lemma 4.4 yields

$$\max_{T\in\mathcal{T}} \frac{1}{|\mathcal{I}_n|} \left\| \sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} \right\|_2 = \begin{cases} \mathcal{O}_P(|\mathcal{I}_n|^{-1/2}\log(|\mathcal{I}_n|)), & r=2 \\ \mathcal{O}_P(|\mathcal{I}_n|^{-1/2}), & r>2 \end{cases}.$$

The applicability of Corollary 2.1 is arrived at the same way as in the previous subsections.

Assumption (I) follows from the following observation: For any set $T \in \mathcal{T}$, it holds that

$$\begin{aligned} d_n(T^c, \theta) &= |\mathcal{I}_n|^{-1}\left(|T^c \cap \theta^c| + |T \cap \theta|\right) \\ &= \begin{cases} |\mathcal{I}_n|^{-1}\left(|T^c| + |\theta|\right), & \text{if } \theta_{\mathcal{I}} \subseteq T \\ |\mathcal{I}_n|^{-1}\left(|\theta^c| + |T|\right), & \text{if } T \subseteq \theta_{\mathcal{I}} \end{cases} \\ &\geq |\mathcal{I}_n|^{-1}\min\{|\theta|, |\theta^c|\} = \sigma_{\mathcal{I}} \geq \tilde{\sigma}. \end{aligned}$$

□

Rate of convergence: Suppose Assumption (Y(r)) is fulfilled with $r > 2$ and $\mathcal{S} = \{T \setminus \theta : T \in \mathcal{T}\} \cup \{\theta \setminus T : T \in \mathcal{T}\}$. If $|\mathcal{I}_n|^{1/2}\sigma_{\mathcal{I}}^2\|a-b\|_\infty \to \infty$, (16) and (17) hold for any α_n such that $\liminf_{n\to\infty} \sigma_{\mathcal{I}}^2\|a-b\|_2^2\alpha_n > 0$. Under Assumption ($T^*1$), Theorem 2.3 then yields the rate of convergence $\alpha_n|\mathcal{I}_n|^{-1}$ for such α_n and ∂_n. If Assumptions (T^*2) and (I) are fulfilled, the same rates hold for d_n by Theorem 2.2.

Proof. It suffices to show (16) (the proof of (17) works analogously). We can again assume w.l.o.g. that $v_n \equiv v = 1$ and $\mathcal{T} = \mathcal{T}_n^{(1)} = \{T_{1,n}, \ldots, T_{m,n}\}$ as above. It holds that

$$\begin{aligned}
&\max_{T\in\mathcal{T},|T\triangle\theta|\geq\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|} \\
\leq &\max_{T\in\mathcal{T},|T\triangle\theta|\geq\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in T\setminus\theta} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|} + \max_{T\in\mathcal{T},|T\triangle\theta|\geq\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|} \\
\leq &\max_{T\in\mathcal{T},|T\setminus\theta|\geq\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in T\setminus\theta} Y_{\mathbf{i},n}\right\|_2}{|T\setminus\theta|} + \max_{T\in\mathcal{T},|T\setminus\theta|<\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in T\setminus\theta} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|} \\
&+ \max_{T\in\mathcal{T},|\theta\setminus T|\geq\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T} Y_{\mathbf{i},n}\right\|_2}{|\theta\setminus T|} + \max_{T\in\mathcal{T},|\theta\setminus T|<\alpha\alpha_n} \frac{\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T} Y_{\mathbf{i},n}\right\|_2}{|T\triangle\theta|}.
\end{aligned}$$

Due to the ordering of the $T_{j,n}$, we obtain

$$\theta \setminus T_{j+1,n} \subseteq \theta \setminus T_{j,n} \quad \text{and} \quad T_{j,n} \setminus \theta \subseteq T_{j+1,n} \setminus \theta$$

for each $j = 1, \ldots, m-1$. Let $J_k = \theta \setminus T_{m-k,n}$. We can assume w.l.o.g. that $|J_k| < |J_{k+1}|$ and therefore Assumption (Y(r)) yields

$$\begin{aligned}
&P\left(\max_{\substack{T\in\mathcal{T}\\|\theta\setminus T|\geq\alpha\alpha_n}} \frac{1}{|\theta\setminus T|}\left\|\sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T} Y_{\mathbf{i},n}\right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right) \\
=&P\left(\max_{\substack{0\leq k<m\\|J_k|\geq\alpha\alpha_n}} \frac{1}{|J_k|}\left\|\sum_{\kappa_{\mathbf{i},n}\in J_k} Y_{\mathbf{i},n}\right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right) \\
\leq&\frac{1}{\varepsilon^r\sigma_{\mathcal{I}}^r\|a-b\|_2^r} \sum_{\substack{0\leq k<m\\|J_k|\geq\alpha\alpha_n}} \frac{E\left\|\sum_{\kappa_{\mathbf{i},n}\in J_k} Y_{\mathbf{i},n}\right\|_2^r}{|J_k|^r} \\
\leq&\frac{K_r}{\varepsilon^r\sigma_{\mathcal{I}}^r\|a-b\|_2^r} \sum_{\substack{0\leq k<m\\|J_k|\geq\alpha\alpha_n}} |J_k|^{-r/2} \\
=&K_r\varepsilon^{-r}\sigma_{\mathcal{I}}^{-r}\|a-b\|_2^{-r}\alpha^{-r/2}\alpha_n^{-r/2} \sum_{\substack{0\leq k<m\\|J_k|\geq\alpha\alpha_n}} \left(\frac{|J_k|}{\alpha\alpha_n}\right)^{-r/2} \\
\leq&K_r\varepsilon^{-r}\sigma_{\mathcal{I}}^{-r}\|a-b\|_2^{-r}\alpha^{-r/2}\alpha_n^{-r/2} \sum_{\substack{0\leq k<m\\|J_k|\geq\alpha\alpha_n}} \left\lfloor\frac{|J_k|}{\alpha\alpha_n}\right\rfloor^{-r/2} \\
\leq&K_r\varepsilon^{-r}\left(\sigma_{\mathcal{I}}^2\|a-b\|_2^2\alpha_n\right)^{-r/2}\alpha^{-r/2}\sum_{k=1}^{\infty} k^{-r/2},
\end{aligned}$$

which becomes arbitrarily small for large α and n. Furthermore, since $|J_k| < |J_{k+1}|$, there are at most $\lfloor \alpha\alpha_n \rfloor + 1$ sets with $|\theta \setminus T_{k,n}| < \alpha\alpha_n$ and therefore Lemma 4.4 yields

$$\begin{aligned}
&P\left(\max_{\substack{T\in\mathcal{T}\\ |\theta\setminus T|<\alpha\alpha_n}} \frac{1}{\alpha\alpha_n} \left\| \sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T} Y_{\mathbf{i},n} \right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2 \right)\\
=&P\left(\max_{\substack{0\leq k<m\\ |J_k|<\alpha\alpha_n}} \frac{1}{\alpha\alpha_n} \left\| \sum_{\kappa_{\mathbf{i},n}\in J_k} Y_{\mathbf{i},n} \right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2 \right)\\
\leq&\frac{\alpha^{-r}\alpha_n^{-r}}{\varepsilon^r\sigma_{\mathcal{I}}^r\|a-b\|_2^r} E\left(\max_{\substack{0\leq k<m\\ |J_k|<\alpha\alpha_n}} \left\| \sum_{\kappa_{\mathbf{i},n}\in J_k} Y_{\mathbf{i},n} \right\|_2^r \right)\\
\leq&K_r^*\varepsilon^{-r}\sigma_{\mathcal{I}}^{-r}\|a-b\|_2^{-r}\alpha^{-r/2}\alpha_n^{-r/2}\\
=&K_r^*\varepsilon^{-r}\alpha^{-r/2}\left(\alpha_n\sigma_{\mathcal{I}}^2\|a-b\|_2^2\right)^{-r/2},
\end{aligned}$$

which also becomes arbitrarily small for large α and n. The same arguments with $\tilde{J}_k = T_{k,n}\setminus\theta$ and w.l.o.g. $|\tilde{J}_k| < |\tilde{J}_{k+1}|$, yield

$$P\left(\max_{\substack{T\in\mathcal{T}\\ |T\setminus\theta|\geq\alpha\alpha_n}} \frac{1}{|T\setminus\theta|} \left\| \sum_{\kappa_{\mathbf{i},n}\in T\setminus\theta} Y_{\mathbf{i},n} \right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2 \right) = \mathcal{O}\left(\left(\sigma_{\mathcal{I}}^2\|a-b\|_2^2\alpha_n\right)^{-r/2}\alpha^{-r/2} \right)$$

and

$$P\left(\max_{\substack{T\in\mathcal{T}\\ |T\setminus\theta|<\alpha\alpha_n}} \frac{1}{\alpha\alpha_n} \left\| \sum_{\kappa_{\mathbf{i},n}\in T\setminus\theta} Y_{\mathbf{i},n} \right\|_2 \geq \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2 \right) = \mathcal{O}\left(\left(\sigma_{\mathcal{I}}^2\|a-b\|_2^2\alpha_n\right)^{-r/2}\alpha^{-r/2} \right),$$

and therefore (16). Due to $|T^c\triangle\theta| = |T\triangle\theta^c|$,

$$\begin{aligned}
&\max_{T\in\mathcal{T}:|\bar{T}\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|\bar{T}\triangle\theta|} \left\| \sum_{\kappa_{\mathbf{i},n}\in\bar{T}} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} \right\|_2\\
\leq&\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta|} \left\| \sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} \right\|_2\\
&+\max_{T\in\mathcal{T}:|T\triangle\theta^c|\geq\alpha\alpha_n} \frac{1}{|T\triangle\theta^c|} \left\| \sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta^c} Y_{\mathbf{i},n} \right\|_2,
\end{aligned}$$

and therefore the same argumentation can be used to obtain (17).

In order to apply Theorems 2.2 and 2.3, note that $|\mathcal{I}_n|^{1/2}\sigma_{\mathcal{I}}^2\|a-b\|_\infty \to \infty$ and the assumption on α_n yield (14) and (15), and, in combination with Assumptions (T^*2) and (I), the consistency under d_n. □

4 Proofs

4.1 Preliminary results - some maximal inequalities

An exponential inequality under mixing assumptions

In this subsection, we derive a truncation result and an exponential inequality for a set-indexed partial sum process under mixing conditions. For simplicity, we consider the special case of uniform grids

$$\mathcal{I} = \mathcal{I}_n = \left\{ \kappa_{\mathbf{i},n} = \left(\frac{i_1}{n}, \ldots, \frac{i_d}{n} \right) : 1 \le i_j \le n \right\}$$

for $n \in \mathbb{N}$.

We first give a truncation result which will be used in the proof of Lemma 2.2 to obtain a bounded random field. For bounded random variables under mixing conditions, we then present a Bernstein type inequality by adapting the methods of proof employed in Lin and Lu (1996) (cf. proof of Theorem 6.2.3, p. 165 f.) and Valenzuela-Domínguez and Franke (2005) to our setting.

As a way to show the tightness of a smoothed version of the partial sum process employed here, Lin and Lu (1996) prove a truncation result and an exponential inequality for stationary real-valued random fields under nonuniform φ-mixing. As it is part of the proof of a functional central limit theorem, they only consider weights $b_n = n^{d/2}$. Valenzuela-Domínguez and Franke (2005) prove an exponential inequality for bounded, stationary, real-valued random fields under α-mixing conditions and the associated partial sums over rectangles. In both cases, the proof of the exponential inequality hinges on the fact that measurable transforms of the random variables inherit the mixing properties, which allows us to apply a covariance inequality to the exponential of the (truncated) process. Here, we aim to combine the two approaches by Lin and Lu (1996) and Valenzuela-Domínguez and Franke (2005) in order to give an extension of both of these results to multivariate random fields indexed on the grid with general weights b_n and the corresponding partial sums over general subsets of $[0,1]^d$. Since the truncation result employed by Lin and Lu (1996) requires strict stationarity, we adapt the proof by Goldie and Greenwood (1986) to our setting.

Given a sequence of $\mathbb{R}^p$-valued random fields $\{Y_{\mathbf{i},n}\}_{\kappa_{\mathbf{i},n} \in \mathcal{I}_n}$ and a family of sets $\mathcal{A} \subset \mathcal{B}([0,1]^d)$ (whose properties will be specified later), we consider set-indexed processes

$$Z_n(A) = \frac{1}{b_n} \sum_{\kappa_{\mathbf{i},n} \in A} (Y_{\mathbf{i},n} - EY_{\mathbf{i},n}),$$

$$Z_n(A, u, v) = \sum_{\kappa_{\mathbf{i},n} \in A} (\eta_{\mathbf{i},n}(u,v) - E\eta_{\mathbf{i},n}(u,v))$$

and

$$U_n(A, u, v) = \sum_{\kappa_{\mathbf{i},n} \in A} \|\eta_{\mathbf{i},n}(u,v)\|_2,$$

where $0 \le u < v \le \infty$, $(b_n)_{n \in \mathbb{N}}$ is a sequence of positive numbers and

$$\eta_{\mathbf{i},n}(u,v) = b_n^{-1} Y_{\mathbf{i},n} \, I(u \le f_n^{-1} \|Y_{\mathbf{i},n}\|_2 < v)$$

for some sequence $(f_n)_{n\in\mathbb{N}}$ with $f_n \to \infty$. Then it holds that

$$Z_n(A) = Z_n(A,0,a) + Z_n(A,a,\infty)$$

and

$$\|Z_n(A,a,\infty)\|_2 \leq U_n([0,1]^d,a,\infty) + EU_n([0,1]^d,a,\infty)$$

for any $a \in (0,\infty)$ and $A \in \mathcal{B}([0,1]^d)$.

Lemma 4.1. *For some $r \geq 2$, let $\{\|Y_{\mathbf{i},n}\|_2^r\}_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n, n\in\mathbb{N}}$ be uniformly integrable, i.e. let*

$$h(x) = \sup_{n\in\mathbb{N}} \max_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} E\left[\|Y_{\mathbf{i},n}\|_2^r I_{\{\|Y_{\mathbf{i},n}\|_2 \geq x\}}\right] \overset{x\to\infty}{\longrightarrow} 0.$$

Assume further that $n^d b_n^{-1} f_n^{1-r} = \mathcal{O}(1)$. Then $EU_n([0,1]^d,a,\infty) \to 0$ for any fixed $a > 0$ and $n \to \infty$ and therefore $Z_n(A) = Z_n(A,0,a) + o_P(1)$ for any set $A \in \mathcal{B}([0,1]^d)$ and even

$$\max_{A\in\mathcal{A}} \|Z_n(A)\|_2 = \max_{A\in\mathcal{A}} \|Z_n(A,0,a)\|_2 + o_P(1).$$

Proof. The following is essentially the proof of Theorem 1.1 in Goldie and Greenwood (1986), extended to general weights and adapted to the current setting. We give it here for ease of reading. For any fixed $a > 0$ and large n,

$$\begin{aligned}
EU_n([0,1]^d,a,\infty) &= \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} E\|\eta_{\mathbf{i},n}(a,\infty)\|_2 \\
&= \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} E\|b_n^{-1} Y_{\mathbf{i},n} I_{\{a\leq f_n^{-1}\|Y_{\mathbf{i},n}\|_2<\infty\}}\|_2 \\
&= \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} b_n^{-1} E\left[\|Y_{\mathbf{i},n}\|_2 I_{\{af_n\leq\|Y_{\mathbf{i},n}\|_2<\infty\}}\right] \\
&= \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} b_n^{-1} E\left[\|Y_{\mathbf{i},n}\|_2^r \|Y_{\mathbf{i},n}\|_2^{1-r} I_{\{af_n\leq\|Y_{\mathbf{i},n}\|_2<\infty\}}\right] \\
&\leq \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} b_n^{-1} f_n^{1-r} a^{1-r} \underbrace{E\left[\|Y_{\mathbf{i},n}\|_2^r I_{\{af_n\leq\|Y_{\mathbf{i},n}\|_2<\infty\}}\right]}_{\leq h(af_n)} \\
&\leq n^d b_n^{-1} f_n^{1-r} a^{1-r} h(af_n) = \mathcal{O}(1) a^{1-r} h(af_n) \overset{n\to\infty}{\longrightarrow} 0
\end{aligned}$$

By the Markov inequality and the nonnegativity of $U_n([0,1]^d,a,\infty)$, this also yields $U_n([0,1]^d,a,\infty) \overset{P}{\longrightarrow} 0$. □

Remark 4.1. *The assumptions on the sequences $(b_n)_{n\in\mathbb{N}}$ and $(f_n)_{n\in\mathbb{N}}$ are satisfied e.g. for $b_n = n^\eta$ $(0 < \eta < d)$ and $f_n = n^{\frac{d-\eta}{r-1}}$. Then $b_n^{-1} f_n^{1-r} = n^{-d}$, $f_n \to \infty$ and $f_n = b_n \cdot g_n$ with $g_n = n^{-\frac{r\eta-d}{r-1}} \overset{n\to\infty}{\longrightarrow} 0$, if $r > d/\eta$.*

Definition 4.1. *For two sets $I, J \subseteq [0,1]^d$, define*

$$dist_n(I,J) = \inf\{n\|\kappa_{\mathbf{i},n} - \kappa_{\mathbf{j},n}\|_\infty : \kappa_{\mathbf{i},n} \in I, \kappa_{\mathbf{j},n} \in J\}.$$

A sequence of random fields $\{\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$ $(n \in \mathbb{N})$ *is said to be* α*-mixing if there is a function* $\alpha(\cdot)$*, which is independent of* n *with* $\alpha(x) \to 0$ *as* $x \to \infty$*, and*

$$\sup_{\substack{A\in\sigma(\eta_{\mathbf{j},n}:\,\kappa_{\mathbf{j},n}\in I)\\ B\in\sigma(\eta_{\mathbf{j},n}:\,\kappa_{\mathbf{j},n}\in J)}} |P(A\cap B) - P(A)P(B)| \le \alpha(dist_n(I,J)),$$

for all $n \in \mathbb{N}$ *and* $I, J \subseteq [0,1]^d$.
$\{\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$ *is said to be nonuniform* φ*-mixing, if there exists a nonnegative function* $\varphi(\cdot)$*, which is independent of* n *with* $\varphi(x) \to 0$ *as* $x \to \infty$*, and*

$$\sup_{\substack{A\in\sigma(\eta_{\mathbf{j},n}:\,\kappa_{\mathbf{j},n}\in I)\\ B\in\sigma(\eta_{\mathbf{j},n}:\,\kappa_{\mathbf{j},n}\in J), P(B)>0}} |P(A|B) - P(A)| \le |I|\cdot\varphi(dist_n(I,J))$$

for all $n \in \mathbb{N}$ *and* $I, J \subseteq [0,1]^d$.

Now, we derive a Bernstein type inequality under either α- or nonuniform φ-mixing. Since an exponential inequality for the euclidean norm of a multivariate process can be inferred from corresponding inequalities for the coordinate processes (cf. the proof of Lemma 2.2), we only consider real-valued processes here.

Lemma 4.2. *Let* $\{\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$*,* $n \in \mathbb{N}$*, be a real-valued, centered random field with* $|\eta_{\mathbf{j},n}| \le C_{1,n}$ *and*

$$E\left[S_n(A)^2\right] \le C_{2,n}|A|,$$

for any $A \subseteq [0,1]^d$*, where* $S_n(A) = \sum_{\kappa_{\mathbf{j},n}\in A}\eta_{\mathbf{j},n}$*. Assume that* $\{\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$ *is either* α*-mixing or nonuniform* φ*-mixing. Let* $p_n \in \mathbb{N}$ *with* $p_n < n$*. Then it holds for any* $A \subseteq [0,1]^d$ *and* $\varepsilon > 0$ *that*

$$P\left(|S_n(A)| \ge \varepsilon\right) \le 2\exp(-\beta_n\varepsilon)\exp\left(2^{2d}\beta_n^2 C_{2,n} e|A| + (p_n^d - 1)\theta_{q_n}(m_n)\sqrt{e}\right),$$

where $m_n = \frac{n}{2p_n}$*,* $0 < \beta_n < \frac{1}{2^{d+1}C_{1,n}m_n^d e}$*,* $q_n = 1 + 1/p_n^d$ *and*

$$\theta_q(x) = \begin{cases} 10\alpha^{1-1/q}(x), & \text{under } \alpha\text{-mixing} \\ 2x^d\varphi(x), & \text{under } \varphi\text{-mixing} \end{cases}$$

for $q > 1$*,* $x > 0$.

Proof. The following proof is a mixture of the proofs of Theorem 6.2.3 in Lin and Lu (1996) and Theorem 3.1 in Valenzuela-Domínguez and Franke (2005), adapted to the current setting. The idea of subdividing the index set in the manner considered here stems from Lin and Lu (1996) (who adapted it from Goldie and Greenwood (1986)) and the covariance inequality as well as the general idea for iteration under α-mixing was taken from Valenzuela-Domínguez and Franke (2005). The present proof combines both approaches and extends them to not necessarily stationary sequences of random fields and — in the case of the result by Valenzuela-Domínguez and Franke (2005) — general classes of sets.

Divide $I^d = [0,1]^d$ in the following two ways: First, we divide it into p_n^d subrectangles $C_{\mathbf{l},p_n} = (p_n^{-1}(\mathbf{l} - \underline{\mathbf{1}}), p_n^{-1}\mathbf{l}]$, $\mathbf{l} \in \mathcal{J}_{p_n} := \{1,\ldots,p_n\}^d$, and secondly into the subrectangles

$C_{\mathbf{j},2p_n}$, $\mathbf{j} \in \mathcal{J}_{2p_n}$. Then each of the subrectangles $C_{\mathbf{l},p_n}$ contains 2^d subrectangles $C_{\mathbf{j},2p_n}$. Denoting the ith subrectangle $C_{\mathbf{j},2p_n}$ in $C_{\mathbf{l},p_n}$ by $I_{n,\mathbf{l},i}$, $i = 1, \ldots, 2^d$, we obtain

$$I_{n,i} = \bigcup_{\mathbf{l} \in \mathcal{J}_{p_n}} I_{n,\mathbf{l},i}, \quad i = 1, \ldots, 2^d.$$

Note that we do not need to assume that these subrectangles are anchored on the pixels. We observe that this partition yields $|I_{n,\mathbf{l},i}| \leq \left(\frac{n}{2p_n}\right)^d = m_n^d$ and

$$dist_n(I_{n,\mathbf{l},i}, I_{n,\mathbf{l}',i}) = n \inf\{\|\kappa_{\mathbf{j},n} - \kappa_{\mathbf{j}',n}\|_\infty : \kappa_{\mathbf{j},n} \in I_{n,\mathbf{l},i}, \kappa_{\mathbf{j}',n} \in I_{n,\mathbf{l}',i}\} \geq m_n$$

for $i \in \{1, \ldots, 2^d\}$ and $\mathbf{l}, \mathbf{l}' \in \mathcal{J}_{p_n}$. For each $i \in \{1, \ldots, 2^d\}$, we order the p_n^d subrectangles $I_{n,\mathbf{l}_u,i}$ for $\mathbf{l}_u \in \mathcal{J}_{p_n}$, $u = 1, \ldots, p_n^d$. It holds that

$$S_n(A) = \sum_{i=1}^{2^d} \sum_{u=1}^{p_n^d} S_n(A \cap I_{n,\mathbf{l}_u,i}) = \sum_{i=1}^{2^d} T(i, p_n^d).$$

In order to prove an exponential inequality for $S_n(A)$, we need only prove a corresponding inequality for each of the $T(i, p_n^d)$, since

$$\begin{aligned} E\left(e^{\beta S_n(A)}\right) &= E\left(e^{\beta \sum_{i=1}^{2^d} T(i,p_n^d)}\right) \\ &= E\left(\prod_{i=1}^{2^d} e^{\beta T(i,p_n^d)}\right) \\ &\leq 2^{-d} \sum_{i=1}^{2^d} E\left(e^{\delta T(i,p_n^d)}\right) \end{aligned}$$

by the inequality of arithmetic and geometric means, for $\beta > 0$ and $\delta = 2^d \beta$. For $r \leq p_n^d$ and $S(i, u) := S_n(A \cap I_{n,\mathbf{l}_u,i})$ $(u = 1, \ldots, p_n^d)$, it holds that

$$T(i, r) = \sum_{u=1}^{r} S(i, u) = T(i, r-1) + S(i, r)$$

and therefore

$$\begin{aligned} &E\left[e^{\delta T(i,r)}\right] \\ \leq &\left|E\left[e^{\delta T(i,r-1)} e^{\delta S(i,r)}\right] - E\left[e^{\delta T(i,r-1)}\right] E\left[e^{\delta S(i,r)}\right]\right| \\ &+ E\left[e^{\delta T(i,r-1)}\right] E\left[e^{\delta S(i,r)}\right]. \end{aligned}$$

Define $J(i, r-1) = I_{n,\mathbf{l}_1,i} \cup \cdots \cup I_{n,\mathbf{l}_{r-1},i}$. Since $T(i, r-1)$ is $\sigma(\eta_{\mathbf{j},n} : \kappa_{\mathbf{j},n} \in J(i, r-1))$-measurable, $S(i, r)$ is $\sigma(\eta_{\mathbf{j},n} : \kappa_{\mathbf{j},n} \in I_{n,\mathbf{l}_r,i})$-measurable and $dist_n(J(i, r-1), I_{n,\mathbf{l}_r,i}) \geq m_n$, the rest of the proof hinges on the following covariance inequality for either α- or φ-mixing random variables (cf. Doukhan (1994), Theorem 3, p. 9f.):

$$\begin{aligned} &\left|E\left[e^{\delta T(i,r-1)} e^{\delta S(i,r)}\right] - E\left[e^{\delta T(i,r-1)}\right] E\left[e^{\delta S(i,r)}\right]\right| \\ \leq &\theta_q(m_n) \|e^{\delta S(i,r)}\|_\infty \|e^{\delta T(i,r-1)}\|_q, \end{aligned}$$

with $q = 1 + 1/r$, where, in the φ-mixing case, we have used that $|A \cap I_{n,\mathbf{l}_r,i}| \leq m_n^d$. (For φ-mixing observations, $q = 1$ could be chosen in the inequality. However, in order to unify the proof for the two mixing assumptions, $q > 1$ is discussed here. The only difference this makes for the result is that $q^j = 1$, $j = 0, \dots, r$, for $q = 1$ and therefore e could be replaced by 1 in the following.)

Note that for any $1 \leq u \leq p_n^d$,

$$|S(i,u)| \leq \sum_{\kappa_{\mathbf{j},n} \in I_{n,\mathbf{l}_u,i}} |\eta_{\mathbf{j},n}| \leq C_{1,n} |I_{n,\mathbf{l}_u,i}| \leq C_{1,n} m_n^d$$

and therefore choosing $0 < \beta = \beta_n \leq \frac{1}{2^{d+1} C_{1,n} m_n^d e}$ yields $|\delta S(i,u)| \leq \frac{1}{2e}$ for $\delta = \delta_n = 2^d \beta_n$, and thus

$$\|e^{\delta D S(i,u)}\|_\infty \leq \sqrt{e} \tag{18}$$

for any $0 \leq D \leq e$. Therefore,

$$\begin{aligned} E\left[e^{\delta D S(i,r)}\right] &\leq 1 + \delta D E[S(i,r)] + \delta^2 D^2 E[S(i,r)^2] \\ &= 1 + \delta^2 D^2 E[S(i,r)^2] \\ &\leq e^{\delta^2 D^2 E[S(i,r)^2]} \leq e^{\delta^2 D^2 C_{2,n} |A \cap I_{n,\mathbf{l}_r,i}|} \end{aligned}$$

and, due to $q \geq 1$,

$$\begin{aligned} &E\left[e^{\delta T(i,r)}\right] \\ \leq &\theta_q(m_n) \|e^{\delta S(i,r)}\|_\infty \|e^{\delta T(i,r-1)}\|_q + E\left[e^{\delta T(i,r-1)}\right] e^{\delta^2 C_{2,n} |A \cap I_{n,\mathbf{l}_r,i}|} \\ \leq &\left(\theta_q(m_n) \|e^{\delta S(i,r)}\|_\infty + e^{\delta^2 C_{2,n} |A \cap I_{n,\mathbf{l}_r,i}|}\right) \|e^{\delta T(i,r-1)}\|_q. \end{aligned}$$

Using the fact that $1 \leq q^j \leq \left(1 + \frac{1}{r}\right)^r \leq e$ $(j = 0, \dots, r)$ for $q = 1 + 1/r$, we analogously obtain

$$\begin{aligned} &\|e^{\delta T(i,r-1)}\|_q \\ = &\left(E\left[e^{\delta q T(i,r-1)}\right]\right)^{1/q} \\ \leq &\left(\theta_q(m_n) \|e^{\delta q S(i,r-1)}\|_\infty + e^{\delta^2 q^2 C_{2,n} |A \cap I_{n,\mathbf{l}_{r-1},i}|}\right)^{1/q} \left(\|e^{\delta q T(i,r-2)}\|_q\right)^{1/q} \\ = &\left(\theta_q(m_n) \|e^{\delta q S(i,r-1)}\|_\infty + e^{\delta^2 q^2 C_{2,n} |A \cap I_{n,\mathbf{l}_{r-1},i}|}\right)^{1/q} \|e^{\delta T(i,r-2)}\|_{q^2} \end{aligned}$$

and iterating yields

$$\|e^{\delta T(i,r-j)}\|_{q^j} \leq \left(\theta_q(m_n) \|e^{\delta q^j S(i,r-j)}\|_\infty + e^{\delta^2 q^{2j} C_{2,n} |A \cap I_{n,\mathbf{l}_{r-j},i}|}\right)^{1/q^j} \|e^{\delta T(i,r-j-1)}\|_{q^{j+1}}$$

for $j = 0, \ldots, r-2$. Using this and (18), we obtain

$$\begin{aligned}
&E\left[e^{\delta T(i,r)}\right] \\
&\leq \prod_{j=0}^{r-2} \left(\theta_q(m_n)\sqrt{e} + e^{\delta^2 q^{2j} C_{2,n}|A\cap I_{n,\mathbf{l}_{r-j},i}|}\right)^{1/q^j} \|e^{\delta T(i,1)}\|_{q^{r-1}} \\
&\leq \left\{\prod_{j=0}^{r-2} e^{\delta^2 q^j C_{2,n}|A\cap I_{n,\mathbf{l}_{r-j},i}|} \left(1+\theta_q(m_n)\sqrt{e}\right)^{1/q^j}\right\} \|e^{\delta T(i,1)}\|_{q^{r-1}} \\
&= \prod_{j=0}^{r-2} \exp\left(\delta^2 q^j C_{2,n}|A\cap I_{n,\mathbf{l}_{r-j},i}| + \frac{1}{q^j}\ln\left(1+\theta_q(m_n)\sqrt{e}\right)\right) \|e^{\delta T(i,1)}\|_{q^{r-1}} \\
&\leq \prod_{j=0}^{r-2} \exp\left(\delta^2 q^j C_{2,n}|A\cap I_{n,\mathbf{l}_{r-j},i}| + \theta_q(m_n)\sqrt{e}\right) \|e^{\delta T(i,1)}\|_{q^{r-1}} \\
&\leq \left\{\prod_{j=0}^{r-2} \exp\left(\delta^2 e C_{2,n}|A\cap I_{n,\mathbf{l}_{r-j},i}|\right)\right\} e^{(r-1)\theta_q(m_n)\sqrt{e}} \|e^{\delta T(i,1)}\|_{q^{r-1}}.
\end{aligned}$$

Finally, by the same arguments as above,

$$\begin{aligned}
\|e^{\delta T(i,1)}\|_{q^{r-1}} &\leq \|e^{\delta T(i,1)}\|_e \\
&= \left(E\left[e^{\delta e S(i,1)}\right]\right)^{1/e} \\
&\leq \left(e^{\delta^2 e^2 E[S(i,1)^2]}\right)^{1/e} \\
&\leq \left(e^{\delta^2 e^2 C_{2,n}|A\cap I_{n,\mathbf{l}_1,i}|}\right)^{1/e} = e^{\delta^2 e C_{2,n}|A\cap I_{n,\mathbf{l}_1,i}|}
\end{aligned}$$

For $r = p_n^d$ and $q = q_n = 1 + \frac{1}{p_n^d}$, this yields

$$\begin{aligned}
&E\left[e^{\delta T(i,p_n^d)}\right] \\
&\leq \exp\left(\delta^2 e C_{2,n} \sum_{j=1}^{p_n^d} |A\cap I_{n,\mathbf{l}_j,i}|\right) e^{(p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}} \\
&\leq \exp\left(\delta^2 e C_{2,n}|A\cap I_{n,i}|\right) e^{(p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}}.
\end{aligned}$$

Since $\{-\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$ has the same mixing coefficients as $\{\eta_{\mathbf{j},n}\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$, combining the results and using the Markov inequality yields the statement of the lemma:

$$\begin{aligned}
P(|S_n(A)| \geq \varepsilon) &\leq P(S_n(A) \geq \varepsilon) + P(-S_n(A) \geq \varepsilon) \\
&\leq P\left(e^{\beta_n S_n(A)} \geq e^{\beta_n\varepsilon}\right) + P\left(e^{\beta_n(-S_n(A))} \geq e^{\beta_n\varepsilon}\right) \\
&\leq \exp(-\beta_n\varepsilon) 2^{-d} \sum_{i=1}^{2^d} \left(E\left[e^{\delta T(i,p_n^d)}\right] + E\left[e^{\delta(-T(i,p_n^d))}\right]\right) \\
&\leq 2\exp(-\beta_n\varepsilon)\exp\left(2^{2d}\beta_n^2 e C_{2,n}|A| + (p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}\right)
\end{aligned}$$

□

A maximal inequality for aggregated pixels

Consider the model of Subsection 3.2.

Lemma 4.3. *It holds that*

$$\max_{A\in\mathcal{A}_n}\frac{1}{|A|}\left\|\sum_{\kappa_{\mathbf{i},n}\in A}Y_{\mathbf{i},n}\right\|_2\leq\max_{C\in\mathcal{C}_n}\frac{1}{|C|}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2=\frac{m_n}{|\mathcal{I}_n|}\max_{C\in\mathcal{C}_n}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2.$$

If additionally Assumption (Y(r)) *is satisfied for* $\mathcal{S}=\mathcal{C}_n$*, it holds that*

$$\max_{C\in\mathcal{C}_n}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2=\mathcal{O}_P\left(|\mathcal{I}_n|^{1/2}m_n^{\frac{1}{r}-\frac{1}{2}}\right).$$

Proof. For any $A\in\mathcal{A}_n$, $C\in\mathcal{C}_n$, it holds that

$$\operatorname{card}(\{C\in\mathcal{C}_n : C\subseteq A\})\leq\frac{|A|}{|C|}=\frac{m_n|A|}{|\mathcal{I}_n|}$$

and therefore

$$\begin{aligned}&\max_{A\in\mathcal{A}_n}\frac{1}{|A|}\left\|\sum_{\kappa_{\mathbf{i},n}\in A}Y_{\mathbf{i},n}\right\|_2\\&\leq\max_{A\in\mathcal{A}_n}\frac{\operatorname{card}(\{C\in\mathcal{C}_n : C\subseteq A\})}{|A|}\max_{C\in\mathcal{C}_n,C\subset A}\frac{|\mathcal{I}_n|}{m_n|C|}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2\\&\leq\max_{C\in\mathcal{C}_n}\frac{1}{|C|}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2\end{aligned}$$

Under Assumption (Y(r)), the Markov inequality yields

$$\begin{aligned}P\left(\max_{C\in\mathcal{C}_n}\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2\geq\varepsilon\right)&\leq m_n\max_{C\in\mathcal{C}_n}P\left(\left\|\sum_{\kappa_{\mathbf{i},n}\in C}Y_{\mathbf{i},n}\right\|_2\geq\varepsilon\right)\\&\leq K_r m_n\max_{C\in\mathcal{C}_n}|C|^{r/2}\varepsilon^{-r}\\&=K_r\left(m_n^{\frac{1}{r}-\frac{1}{2}}|\mathcal{I}_n|^{1/2}\right)^r\varepsilon^{-r}.\end{aligned}$$

□

A maximal inequality for nested sets

Consider nested sets $T_1\subseteq\cdots\subseteq T_m$ that are anchored on the grid $\mathcal{I}_n$ with $0<b_1=|T_1|<\cdots<|T_m|=b_m$. Then we obtain the following maximal inequalities:

Lemma 4.4. *Suppose that Assumption* (Y(r)) *is satisfied for* $\mathcal{S} = \mathcal{T}$. *Then*

$$E\left(\max_{1\leq j\leq m}\left\|\sum_{\kappa_{\mathbf{i},n}\in T_j} Y_{\mathbf{i},n}\right\|_2^r\right) \leq \begin{cases} K_r^* b_m \log(b_m)^2, & r=2,\\ K_r^* b_m^{r/2}, & r>2,\end{cases}$$

where $K_r^* = 3K_r$ *for* $r=2$ *and* $K_r^* = \frac{5}{2}(1-2^{(1-r/2)/r})^{-r}K_r$ *for* $r>2$.

Proof. Choose a numbering of the b_m grid points in T_m such that the first 1 to b_1 points lie in T_1, the points b_1+1 to b_2 lie in $T_2 \setminus T_1$, and so on. Define $\{\psi_{j,n}\}_{j=1,\ldots,b_m}$ by

$$\psi_{j,n} = \begin{cases} 0, & j\notin\{b_1,\ldots,b_m\},\\ \sum_{\kappa_{\mathbf{i},n}\in T_k\setminus T_{k-1}} Y_{\mathbf{i},n}, & j=b_k\ (\exists k\in\{1,\ldots,m\})\end{cases},$$

where $T_0 = \emptyset$. Then $\{\psi_{j,n}\}_{j=1,\ldots,b_m}$ satisfies

$$\sum_{j=1}^{l}\psi_{j,n} = \begin{cases} 0, & l<b_1\\ \sum_{\kappa_{\mathbf{i},n}\in T_k} Y_{\mathbf{i},n}, & b_k\leq l< b_{k+1}\ (\exists k\in\{1,\ldots,m-1\})\\ \sum_{\kappa_{\mathbf{i},n}\in T_m} Y_{\mathbf{i},n}, & l=b_m\end{cases}$$

and therefore

$$E\left\|\sum_{j=1}^{l}\psi_{j,n}\right\|_2^r \leq K_r l^{r/2}, \quad \text{for all } l=1,\ldots,b_m.$$

Theorem and Corollary 1 in Móricz (1982) (which are applicable to $\mathbb{R}^p$-valued random variables as can be easily seen from their proof) imply

$$E\left(\max_{1\leq j\leq m}\left\|\sum_{\kappa_{\mathbf{i},n}\in T_j} Y_{\mathbf{i},n}\right\|_2^r\right) = E\left(\max_{1\leq l\leq b_m}\left\|\sum_{j=1}^{l}\psi_{j,n}\right\|_2^r\right) \leq \begin{cases} K_r^\star b_m\log(b_m)^2, r=2\\ K_r^\star b_m^{r/2}, r>2\end{cases}$$

□

4.2 Proofs of the main results

Proof of Theorem 2.1. The following proof follows the classical idea (cf. e.g. van der Vaart and Wellner (1996) or Kosorok (2010)) of dividing the process into a stochastic and a purely deterministic part and then showing that the stochastic part becomes asymptotically negligible and the deterministic part has a (well separated) maximum at the true change-set. However, unlike in most classical proofs, the latter is proven without using a fixed (pseudo-)metric on $\mathcal{B}([0,1]^d)$ (which would correspond to a limit for the grid dependent metric employed here) or explicitly deriving the limit function of the deterministic part and using continuity assumptions on the limit. Let $\epsilon>0$ be arbitrary. It holds that

$$\begin{aligned} &P\left(\partial_n(\hat\theta_n,\theta)\geq\epsilon\right)\\ \leq &P\left(\partial_n(\hat\theta_n,\theta)\geq\epsilon,\ \max_{T\in\mathcal{T}}\|B_n(T)\|_2<\xi\xi_n\right) + P\left(\max_{T\in\mathcal{T}}\|B_n(T)\|_2\geq\xi\xi_n\right)\\ = &P\left(\partial_n(\hat\theta_n,\theta)\geq\epsilon,\ \max_{T\in\mathcal{T}}\|B_n(T)\|_2<\xi\xi_n\right)+o(1)\end{aligned}$$

for any $\xi > 0$, and if we consider $\xi < \alpha/4$ and n large enough that $\beta_n < \epsilon$

$$
\begin{aligned}
&P\left(\partial_n(\hat{\theta}_n, \theta) \geq \epsilon,\ \max_{T\in\mathcal{T}} \|B_n(T)\|_2 < \xi\xi_n\right) \\
\leq &P\left(\overbrace{\max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|D_n(T)\|_2}^{\leq \max_{T\in\mathcal{T}}\|B_n(T)\|_2 + \max_{T\in\mathcal{T},\partial_n(T,\theta)\geq\epsilon}\|\Delta_n(T)\|_2} \geq \overbrace{\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\epsilon} \|D_n(T)\|_2}^{\geq \max_{T\in\mathcal{T},\partial_n(T,\theta)<\beta_n}\|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}}\|B_n(T)\|_2},\right. \\
&\qquad \left.\max_{T\in\mathcal{T}} \|B_n(T)\|_2 < \xi\xi_n\right) \\
\leq &P\left(\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \leq 2\max_{T\in\mathcal{T}} \|B_n(T)\|_2,\right. \\
&\left.\max_{T\in\mathcal{T}} \|B_n(T)\|_2 < \xi\xi_n\right) \\
\leq &P\left(\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \leq 2\xi\xi_n\right) \\
\leq &P\left(\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \leq \frac{\alpha}{2}\xi_n\right).
\end{aligned}
$$

Since

$$
\begin{aligned}
&\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \rho_n(T) - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \rho_n(T) \\
= &\min_{\substack{T\in\mathcal{T},\\ \partial_n(T,\theta)\geq\epsilon}} \max_{\substack{\tilde{T}\in\mathcal{T},\\ \partial_n(\tilde{T},\theta)<\beta_n}} \left(\rho_n(\tilde{T}) - \rho_n(T)\right),
\end{aligned}
$$

Lemma 1.1 yields

$$
\begin{aligned}
\rho_n(\tilde{T}) - \rho_n(T) &= \overbrace{\rho_n(\theta) - \rho_n(T)}^{\geq \sigma_{\mathcal{I}}\partial_n(T,\theta)} - \overbrace{(\rho_n(\theta) - \rho_n(\tilde{T}))}^{<\partial_n(\tilde{T},\theta)} \\
&\geq \sigma_{\mathcal{I}}\partial_n(T,\theta) - \partial_n(\tilde{T},\theta) \\
&\geq \sigma_{\mathcal{I}}\epsilon - \beta_n.
\end{aligned}
$$

Therefore, (7) implies

$$
\begin{aligned}
&\liminf_{n\to\infty} \xi_n^{-1} \left\{\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2\right\} \\
\geq &\liminf_{n\to\infty} \xi_n^{-1} \|a - b\|_2 \left(\sigma_{\mathcal{I}}\epsilon - \beta_n\right) \\
> &\alpha
\end{aligned}
\tag{19}
$$

and thus there is some $n_0 \in \mathbb{N}$ so that for any $n \geq n_0$,

$$
\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \geq \xi_n\alpha
$$

and finally

$$
P\left(\max_{T\in\mathcal{T}, \partial_n(T,\theta)<\beta_n} \|\Delta_n(T)\|_2 - \max_{T\in\mathcal{T}, \partial_n(T,\theta)\geq\epsilon} \|\Delta_n(T)\|_2 \leq \xi_n\alpha/2\right) \overset{n\to\infty}{\longrightarrow} 0.
$$

The convergence of $d_n(\hat{\theta}_n, \theta)$ can be proven analogously by replacing ∂_n by d_n in the proof and directly using (8) instead of (19). □

Proof of Remark 2.1. **Proof of 1:** By Assumption (I) and (7), we can choose $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$\min_{T \in \mathcal{T}} d_n(T^c, \theta) > \epsilon \quad \text{and} \quad \xi_n^{-1} \|a - b\|_2 \left(\sigma_{\mathcal{I}} \epsilon - \beta_n\right) > \alpha$$

for some $\epsilon > 0$ and $\alpha > 0$. As seen in the proof of Theorem 2.1,

$$\begin{aligned}
&\max_{T \in \mathcal{T}, d_n(T,\theta) < \beta_n} \rho_n(T) - \max_{T \in \mathcal{T}, d_n(T,\theta) \geq \epsilon} \rho_n(T) \\
\geq &\min_{\substack{T \in \mathcal{T}, \\ d_n(T,\theta) \geq \epsilon}} \max_{\substack{\tilde{T} \in \mathcal{T}, \\ d_n(\tilde{T},\theta) < \beta_n}} \left(\sigma_{\mathcal{I}} \partial_n(T, \theta) - \partial_n(\tilde{T}, \theta)\right) \\
\geq &\sigma_{\mathcal{I}} \min_{T \in \mathcal{T}, d_n(T,\theta) \geq \epsilon} \min\{d_n(T, \theta), d_n(T^c, \theta)\} - \beta_n \\
\geq &\sigma_{\mathcal{I}} \min\left\{\min_{T \in \mathcal{T}} d_n(T^c, \theta), \epsilon\right\} - \beta_n \\
= &\sigma_{\mathcal{I}} \epsilon - \beta_n.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\xi_n^{-1} \|a - b\|_2 \left\{\max_{T \in \mathcal{T}, d_n(T,\theta) < \beta_n} \rho_n(T) - \max_{T \in \mathcal{T}, d_n(T,\theta) \geq \epsilon} \rho_n(T)\right\} \\
\geq &\xi_n^{-1} \|a - b\|_2 \left(\sigma_{\mathcal{I}} \epsilon - \beta_n\right) \\
> &\alpha
\end{aligned}$$

for any $n \geq n_0$.
Proof of 3: Using $d_n(A, \theta) = d_n(A_{\mathcal{I}}, \theta_{\mathcal{I}})$, it holds that

$$\begin{aligned}
&\limsup_{n \to \infty} \max_{T \in \mathcal{T}} d_n(T, \theta) \\
= &\limsup_{n \to \infty} \sup_{A \in \mathcal{A}} (\lambda(A \triangle \theta) + (d_n(A, \theta) - \lambda(A \triangle \theta))) \\
\leq &\sup_{A \in \mathcal{A}} \lambda(A \triangle \theta) + \limsup_{n \to \infty} \sup_{A \in \mathcal{A}} |d_n(A, \theta) - \lambda(A \triangle \theta)| \\
= &\sup_{A \in \mathcal{A}} \lambda(A \triangle \theta) < 1
\end{aligned}$$

□

Proof of Lemma 2.1. The proof is analogous to the proof of Theorem 1.1 in Bass and Pyke (1984), with convergence in probability instead of almost surely. We include it here for ease of reading. For any $A \subseteq [0, 1]^d$, we write

$$S_n(A) = \sum_{\kappa_{\mathbf{i},n} \in A} Y_{\mathbf{i},n} \text{ and } \tilde{S}_n(A) = \sum_{\kappa_{\mathbf{i},n} \in A} (|Y_{\mathbf{i},n}| - \nu_{\mathbf{i},n}).$$

First, note that due to Assumption (Y),

$$|\mathcal{I}_n|^{-1} S_n((\underline{\mathbf{0}}, \mathbf{x}]) = \frac{|(\underline{\mathbf{0}}, \mathbf{x}]|}{|\mathcal{I}_n|} \frac{1}{|(\underline{\mathbf{0}}, \mathbf{x}]|} S_n((\underline{\mathbf{0}}, \mathbf{x}]) \xrightarrow{P} 0 \quad (n \to \infty)$$

for any $\mathbf{x} \in (0,1]^d$, since $|(\underline{\mathbf{0}}, \mathbf{x}]| = \prod_{l=1}^d \lfloor N_l x_l \rfloor \to \infty$ for $\max\limits_{l=1,\dots,d} N_l \to \infty$ and $\mathbf{x} > 0$. If a set T can be obtained by a finite number of unions and differences of rectangles $(\underline{\mathbf{0}}, \mathbf{x}]$, linearity implies

$$|\mathcal{I}_n|^{-1} S_n(T) \xrightarrow{P} 0 \text{ and analogously } |\mathcal{I}_n|^{-1} \tilde{S}_n(T) \xrightarrow{P} 0. \tag{20}$$

Fix an integer m and set $D_{\mathbf{j},m} = m^{-1}(\mathbf{j} - \underline{\mathbf{1}}, \mathbf{j}]$, $\mathbf{j} \in \{1, \dots, m\}^d$, and for any $T \subseteq [0,1]^d$, let

$$R_m^-(T) = \bigcup_{D_{\mathbf{j},m} \subseteq T} D_{\mathbf{j},m} \quad \text{and} \quad R_m^+(T) = \bigcup_{D_{\mathbf{j},m} \cap T \neq \emptyset} D_{\mathbf{j},m}.$$

The furthest any point of $R_m^+(T) \setminus R_m^-(T)$ can be from the boundary of T is the diameter of a cube with sides with a length of $1/m$. Hence,

$$|\mathcal{I}_n|^{-1} \max_{T \in \mathcal{T}} |T \setminus R_m^-(T)| \leq |\mathcal{I}_n|^{-1} \max_{T \in \mathcal{T}} |R_m^+(T) \setminus R_m^-(T)| \leq r_n(1/m)$$

Set $\mathcal{R}_m^- = \{R_m^-(T) : T \subseteq [0,1]^d\}$ and $\mathcal{R}_m^\triangle = \{R_m^+(T) \setminus R_m^-(T) : T \subseteq [0,1]^d\}$. For fixed m, these sets are finite and each set in $\mathcal{R}_m^\triangle$ or $\mathcal{R}_m^-$ can be obtained by a finite number of unions and differences of rectangles. Therefore,

$$\max_{B \in \mathcal{R}_m^\triangle} ||\mathcal{I}_n|^{-1} \tilde{S}_n(B)| = o_P(1) \quad \text{and} \quad \max_{B \in \mathcal{R}_m^-} ||\mathcal{I}_n|^{-1} S_n(B)| = o_P(1)$$

for $n \to \infty$ and any $m \in \mathbb{N}$, by (20). Note that for fixed m,

$$\begin{aligned}
&\max_{T \in \mathcal{T}} ||\mathcal{I}_n|^{-1} S_n(T) - |\mathcal{I}_n|^{-1} S_n(R_m^-(T))| \\
\leq &\max_{T \in \mathcal{T}} \left(|\mathcal{I}_n|^{-1} \tilde{S}_n(R_m^+(T) \setminus R_m^-(T)) + |\mathcal{I}_n|^{-1} \sum_{\kappa_{\mathbf{i},n} \in R_m^+(T) \setminus R_m^-(T)} \nu_{\mathbf{i},n} \right) \\
\leq &\max_{B \in \mathcal{R}_m^\triangle} ||\mathcal{I}_n|^{-1} \tilde{S}_n(B)| + \max_{\kappa_{\mathbf{i},n} \in \mathcal{I}_n} \nu_{\mathbf{i},n} |\mathcal{I}_n|^{-1} \max_{T \in \mathcal{T}} |R_m^+(T) \setminus R_m^-(T)| \\
\leq &\max_{B \in \mathcal{R}_m^\triangle} ||\mathcal{I}_n|^{-1} \tilde{S}_n(B)| + \nu r_n(1/m).
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
&\max_{T \in \mathcal{T}} ||\mathcal{I}_n|^{-1} S_n(T)| \\
\leq &\max_{T \in \mathcal{T}} ||\mathcal{I}_n|^{-1} S_n(T) - |\mathcal{I}_n|^{-1} S_n(R_m^-(T))| + \max_{T \in \mathcal{T}} ||\mathcal{I}_n|^{-1} S_n(R_m^-(T))| \\
\leq &\max_{B \in \mathcal{R}_m^\triangle} ||\mathcal{I}_n|^{-1} \tilde{S}_n(B)| + \max_{B \in \mathcal{R}_m^-} ||\mathcal{I}_n|^{-1} S_n(B)| + \nu r_n(1/m).
\end{aligned}$$

Finally, it follows for any $\varepsilon > 0$ and $m \in \mathbb{N}$ that

$$\begin{aligned}
&\limsup_{n \to \infty} P\left(\max_{T \in \mathcal{T}} ||\mathcal{I}_n|^{-1} S_n(T)| \geq \varepsilon \right) \\
\leq &\limsup_{n \to \infty} P\left(\max_{B \in \mathcal{R}_m^\triangle} ||\mathcal{I}_n|^{-1} \tilde{S}_n(B)| \geq \frac{1}{3}\varepsilon \right) + \limsup_{n \to \infty} P\left(\max_{B \in \mathcal{R}_m^-} ||\mathcal{I}_n|^{-1} S_n(B)| \geq \frac{1}{3}\varepsilon \right) \\
+ &\limsup_{n \to \infty} P\left(\nu r_n(1/m) \geq \frac{1}{3}\varepsilon \right) \\
= &\limsup_{n \to \infty} I_{\{r_n(1/m) \geq 1/3 \varepsilon \nu^{-1}\}}.
\end{aligned}$$

Since the left-hand side is independent of m and the right-hand side goes to zero for $m \to \infty$ by (13), this proves the Lemma. □

Proof of Corollary 2.1. Under Assumption (T^*1) (or, correspondingly, (T^*2) for d_n), $\beta_n = \frac{1}{6}\sigma_{\mathcal{I}}\alpha_n|\mathcal{I}_n|^{-1}$ can be chosen. With this choice of β_n, $\liminf_{n\to\infty} \xi_n^{-1}\|a-b\|_2\sigma_{\mathcal{I}} > 0$ implies (7), since the fact that $\alpha_n = o(|\mathcal{I}_n|)$ yields

$$\liminf_{n\to\infty} \xi_n^{-1}\|a-b\|_2(\sigma_{\mathcal{I}}\epsilon - \beta_n) = \liminf_{n\to\infty} \xi_n^{-1}\|a-b\|_2\sigma_{\mathcal{I}}\left(\epsilon - \frac{1}{6}\alpha_n|\mathcal{I}_n|^{-1}\right) > 0$$

for any $\epsilon > 0$. Therefore, the result follows from Theorem 2.1 and the additional comments in Subsection 2.1. □

Proof of Theorem 2.2. The following proof is inspired by the proof of Theorem 2.1 in Ferger (2004), modified to fit the current setting, where, in particular, no stochastic independence is assumed and the change-set itself is estimated, instead of its boundary. Focusing on the more general problem of a change in distribution, Ferger (2004) was able to employ an exponential inequality by Dümbgen (1991) to obtain bounds for error probabilities of the form $P(\partial(\hat{\theta}_n, \theta) > \varepsilon)$, $\varepsilon > 0$. While the current proof follows the general idea of Ferger (2004), it differs heavily in its execution, since no such inequality is available under our assumptions and our statistic is based on the partial sums instead of empirical measures.

Per definition of the estimator and Assumption (T^*2), it holds for any $\alpha > 1/6\sigma_{\mathcal{I}}$ that

$$\begin{aligned}\{|\mathcal{I}_n| d_n(\hat{\theta}_n, \theta) \geq \alpha\alpha_n\} &\subseteq \{\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \|D_n(T)\|_2 \geq \max_{T\in\mathcal{T}:|T\triangle\theta|<\alpha\alpha_n} \|D_n(T)\|_2\} \\ &\subseteq \{\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} (\|D_n(T)\|_2 - \|D_n(T^*)\|_2) \geq 0\}.\end{aligned}$$

Observe that for any $T \in \mathcal{T}$

$$\|D_n(T)\|_2 - \|D_n(T^*)\|_2 = \|D_n(T)\|_2 - \|D_n(\theta)\|_2 + \|D_n(\theta)\|_2 - \|D_n(T^*)\|_2$$

and

$$\begin{aligned}&D_n(T) \\ =&B_n(T) + \delta_n(T)(a-b) \\ =&B_n(T) - B_n(\theta) + \frac{\delta_n(\theta)}{\delta_n(\theta)}B_n(\theta) - \frac{\delta_n(T)}{\delta_n(\theta)}B_n(\theta) + \frac{\delta_n(T)}{\delta_n(\theta)}D_n(\theta) - \frac{\delta_n(T)}{\delta_n(\theta)}\delta_n(\theta)(a-b) \\ &+ \delta_n(T)(a-b) \\ =&B_n(T) - B_n(\theta) + \frac{\delta_n(\theta) - \delta_n(T)}{\delta_n(\theta)}B_n(\theta) + \frac{\delta_n(T)}{\delta_n(\theta)}D_n(\theta).\end{aligned}$$

Therefore,

$$\|D_n(T)\|_2 \leq \|B_n(T) - B_n(\theta)\|_2 + \frac{|\delta_n(\theta) - \delta_n(T)|}{\delta_n(\theta)}\|B_n(\theta)\|_2 + \frac{\rho_n(T)}{\delta_n(\theta)}\|D_n(\theta)\|_2.$$

If $d_n(T,\theta) \le 1/2$, Lemma 1.1 (cf. also Remark 1.2) and $\delta_n(\theta) = \rho_n(\theta)$ imply

$$\begin{aligned}
&\|D_n(T)\|_2 - \|D_n(\theta)\|_2 \\
\le&\|B_n(T) - B_n(\theta)\|_2 + \frac{|\delta_n(\theta) - \delta_n(T)|}{\delta_n(\theta)}\|B_n(\theta)\|_2 - \frac{\delta_n(\theta) - \rho_n(T)}{\delta_n(\theta)}\|D_n(\theta)\|_2 \\
\le&\|B_n(T) - B_n(\theta)\|_2 + \frac{d_n(T,\theta)}{\delta_n(\theta)}\|B_n(\theta)\|_2 - \frac{\sigma_{\mathcal{I}} d_n(T,\theta)}{\delta_n(\theta)}\|D_n(\theta)\|_2 \\
=&\|B_n(T) - B_n(\theta)\|_2 - \left\{\frac{\sigma_{\mathcal{I}}\|D_n(\theta)\|_2 - \|B_n(\theta)\|_2}{\delta_n(\theta)}\right\} d_n(T,\theta) \\
\le&\|B_n(T) - B_n(\theta)\|_2 - L d_n(T,\theta)
\end{aligned}$$

on the complement of the set

$$E = \left\{\frac{\sigma_{\mathcal{I}}\|D_n(\theta)\|_2 - \|B_n(\theta)\|_2}{\delta_n(\theta)} \le L\right\}$$

with $L = \sigma_{\mathcal{I}}/2\|a-b\|_2$. Furthermore,

$$\begin{aligned}
\|D_n(\theta)\|_2 - \|D_n(T^*)\|_2 &\le \|D_n(\theta) - D_n(T^*)\|_2 \\
&= \|B_n(\theta) - B_n(T^*) + (\delta_n(\theta) - \delta_n(T^*))(a-b)\|_2 \\
&\le \|B_n(\theta) - B_n(T^*)\|_2 + |\delta_n(\theta) - \delta_n(T^*)|\|a-b\|_2 \\
&\le \|B_n(\theta) - B_n(T^*)\|_2 + d_n(\theta, T^*)\|a-b\|_2.
\end{aligned}$$

Finally, we obtain

$$\begin{aligned}
&P(|\mathcal{I}_n| d_n(\hat{\theta}_n,\theta) \ge \alpha\alpha_n) \\
\le& P\left(|\mathcal{I}_n| d_n(\hat{\theta}_n,\theta) \ge \alpha\alpha_n, d_n(\hat{\theta}_n,\theta) \le 1/2, E^c\right) + P(E) + P(d_n(\hat{\theta}_n,\theta) > 1/2).
\end{aligned}$$

Consistency of the estimator implies

$$P(d_n(\hat{\theta}_n,\theta) > 1/2) \overset{n\to\infty}{\longrightarrow} 0. \tag{21}$$

It holds that

$$\begin{aligned}
E &= \left\{\frac{\sigma_{\mathcal{I}}\|D_n(\theta)\|_2 - \|B_n(\theta)\|_2}{\delta_n(\theta)} \le L\right\} \\
&\subseteq \left\{ \|B_n(\theta)\|_2 \ge \sigma_{\mathcal{I}} \underbrace{\|D_n(\theta)\|_2}_{\ge \delta_n(\theta)\|a-b\|_2 - \|B_n(\theta)\|_2} - \delta_n(\theta) L \right\} \\
&\subseteq \left\{ \|B_n(\theta)\|_2 \ge \frac{\delta_n(\theta)}{1+\sigma_{\mathcal{I}}} (\sigma_{\mathcal{I}}\|a-b\|_2 - L) \right\} \\
&\subseteq \left\{ \|B_n(\theta)\|_2 \ge \frac{\sigma_{\mathcal{I}}/2}{1+\sigma_{\mathcal{I}}} \sigma_{\mathcal{I}}/2\|a-b\|_2 = \frac{1}{4}\frac{\sigma_{\mathcal{I}}^2}{1+\sigma_{\mathcal{I}}}\|a-b\|_2 \right\},
\end{aligned}$$

since $L = \sigma_{\mathcal{I}}/2\|a-b\|_2$ and $\delta_n(\theta) \ge \sigma_{\mathcal{I}}/2$. Moreover,

$$\begin{aligned}
B_n(\theta) &= \frac{1}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\theta} (Y_{\mathbf{i},n} - \bar{Y}_n) \\
&= \frac{1}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} - \frac{|\theta|}{|\mathcal{I}_n|}\frac{1}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n},
\end{aligned}$$

which is $\mathcal{O}_P(|\mathcal{I}_n|^{-1/2})$ under Assumption (Y). Therefore, since

$$\frac{|\mathcal{I}_n|^{1/2}}{4}\frac{\sigma_{\mathcal{I}}^2}{1+\sigma_{\mathcal{I}}}\|a-b\|_2 \overset{n\to\infty}{\longrightarrow} \infty \quad \text{by (14)},$$

we obtain $P(E) \overset{n\to\infty}{\longrightarrow} 0$. Due to the arguments above,

$$\begin{aligned}
&P\left(|\mathcal{I}_n| d_n(\hat{\theta}_n,\theta) \geq \alpha\alpha_n, d_n(\hat{\theta}_n,\theta) \leq 1/2, E^c\right)\\
\leq& P\Big(\exists T\in\mathcal{T},\ |T\triangle\theta| \geq \alpha\alpha_n :\\
&0 \leq \|B_n(T)-B_n(\theta)\|_2 - L d_n(T,\theta) + \|B_n(\theta)-B_n(T^*)\|_2 + d_n(\theta,T^*)\|a-b\|_2\Big)\\
\leq& P\Bigg(\exists T\in\mathcal{T},\ |T\triangle\theta| \geq \alpha\alpha_n :\ L \leq \frac{|\mathcal{I}_n|}{|T\triangle\theta|}\|B_n(T)-B_n(\theta)\|_2\\
&\qquad + \frac{|\mathcal{I}_n|}{|T\triangle\theta|}\|B_n(\theta)-B_n(T^*)\|_2 + |\mathcal{I}_n|\alpha^{-1}\alpha_n^{-1} d_n(\theta,T^*)\|a-b\|_2\Bigg)\\
\leq& P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{|\mathcal{I}_n|\|B_n(T)-B_n(\theta)\|_2}{|T\triangle\theta|} \geq 1/3L\right)\\
&+ P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{|\mathcal{I}_n|\|B_n(\theta)-B_n(T^*)\|_2}{|T\triangle\theta|} \geq 1/3L\right)\\
&+ P(|\mathcal{I}_n|\alpha^{-1}\alpha_n^{-1} d_n(\theta,T^*)\|a-b\|_2 \geq 1/3L)\\
=:& P(A_1)+P(A_2)+P(A_3).
\end{aligned}$$

First, note that it follows from $\|a-b\|_2 \neq 0$ and Assumption (T^*2) on the approximation quality of the candidate sets $\mathcal{T}$ that

$$\begin{aligned}
A_3 &= \left\{|\mathcal{I}_n|\alpha^{-1}\alpha_n^{-1} d_n(\theta,T^*)\|a-b\|_2 \geq 1/6\sigma_{\mathcal{I}}\|a-b\|_2\right\}\\
&\subseteq \left\{d_n(\theta,T^*) \geq 1/6|\mathcal{I}_n|^{-1}\sigma_{\mathcal{I}}\alpha\alpha_n\right\} = \emptyset
\end{aligned}$$

for any $\alpha > 1$. For $T\in\mathcal{T}$ with $|T\triangle\theta| \geq \alpha\alpha_n$, it holds that

$$\begin{aligned}
&B_n(T)-B_n(\theta)\\
=&\frac{1}{|\mathcal{I}_n|}\left(\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \frac{|T|}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}\right) - \frac{1}{|\mathcal{I}_n|}\left(\sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n} - \frac{|\theta|}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}\right)\\
=&\frac{1}{|\mathcal{I}_n|}\left(\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right) + \frac{1}{|\mathcal{I}_n|}\left(\frac{|\theta|}{|\mathcal{I}_n|} - \frac{|T|}{|\mathcal{I}_n|}\right)\sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}
\end{aligned}$$

And therefore,

$$\begin{aligned}
&\frac{|\mathcal{I}_n|}{|T\triangle\theta|}\|B_n(T)-B_n(\theta)\|_2\\
\leq&\left\|\frac{1}{|T\triangle\theta|}\left(\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right)\right\|_2 + \frac{||\theta|-|T||}{|T\triangle\theta|}\left\|\frac{1}{|\mathcal{I}_n|}\sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}\right\|_2\\
\leq&\frac{1}{|T\triangle\theta|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 + \mathcal{O}_P\left(|\mathcal{I}_n|^{-1/2}\right)
\end{aligned}$$

by Assumption (Y). Thus, it can easily be seen that under Assumption (14), $P(A_1) \to 0$ for $\alpha \to \infty$ and $n \to \infty$ if (16) holds. Since $|T\triangle\theta| = |\mathcal{I}_n| d_n(T,\theta) \geq |\mathcal{I}_n| d_n(T^*,\theta) = |T^*\triangle\theta|$ per definition of T^*, analogous calculations imply

$$\begin{aligned}
&\frac{|\mathcal{I}_n|}{|T\triangle\theta|}\|B_n(T^*) - B_n(\theta)\|_2 \\
\leq &\frac{1}{|T\triangle\theta|}\left\| \sum_{\kappa_{\mathbf{i},n}\in T^*} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 + \mathcal{O}_P\left(|\mathcal{I}_n|^{-1/2}\right)
\end{aligned}$$

and therefore

$$\begin{aligned}
P(A_2) = &P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\geq\alpha\alpha_n} \frac{|\mathcal{I}_n|\|B_n(\theta) - B_n(T^*)\|_2}{|T\triangle\theta|} \geq 1/3L\right) \\
\leq &P\left(\frac{1}{\alpha\alpha_n}\left\| \sum_{\kappa_{\mathbf{i},n}\in T^*} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 \geq 1/6L\right) \\
&+ P\left(\frac{1}{|\mathcal{I}_n|}\left\| \sum_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n} Y_{\mathbf{i},n}\right\|_2 \geq 1/6L\right).
\end{aligned}$$

Now, Assumptions (Y) and $|\theta| \geq |\mathcal{I}_n|\sigma_{\mathcal{I}}$ yield

$$P\left(\frac{1}{|\theta|}\left\| \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 \geq 1/36L\right) \leq cK|\theta|^{-1}\sigma_{\mathcal{I}}^{-2}\|a-b\|_2^{-2} \leq cK\sigma_{\mathcal{I}}\left(|\mathcal{I}_n|^{1/2}\sigma_{\mathcal{I}}^2\|a-b\|_2\right)^{-2}$$

for some $c > 0$. Furthermore, it follows from Assumption (T^*2) that

$$\begin{aligned}
&P\left(\frac{1}{\alpha\alpha_n}\left\| \sum_{\kappa_{\mathbf{i},n}\in T^*} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 \geq 1/12L\right) \\
\leq &\frac{24^2}{L^2}\frac{1}{\alpha^2\alpha_n^2}\left(E\left\| \sum_{\kappa_{\mathbf{i},n}\in T^*\setminus\theta} Y_{\mathbf{i},n}\right\|_2^2 + E\left\| \sum_{\kappa_{\mathbf{i},n}\in\theta\setminus T^*} Y_{\mathbf{i},n}\right\|_2^2\right) \\
\leq &576\frac{K}{L^2}\frac{1}{\alpha^2\alpha_n^2}\left(\underbrace{|T^*\setminus\theta| + |\theta\setminus T^*|}_{=|T^*\triangle\theta|\leq 1/6\sigma_{\mathcal{I}}\alpha_n}\right) \\
\leq &\frac{384K}{\sigma_{\mathcal{I}}\|a-b\|_2^2\alpha_n}\frac{1}{\alpha^2}.
\end{aligned}$$

Therefore, under Assumptions (14) and (15), $P(A_2) \to 0$, for $\alpha \to \infty$ and $n \to \infty$. Moreover, using an analogous argumentation, it can easily be seen that under these assumptions, $P(A_1) \to 0$ for $\alpha \to \infty$ and $n \to \infty$ if (16) holds. □

Proof of Theorem 2.3. Note that

$$\|D_n(T)\|_2 = \|D_n(T^c)\|_2 = \|D_n(\bar{T})\|_2$$

and therefore

$$\begin{aligned}\{|\mathcal{I}_n|\partial_n(\hat{\theta}_n,\theta)\geq\alpha\alpha_n\}&\subseteq\{\max_{T\in\mathcal{T}:|\bar{T}\triangle\theta|\geq\alpha\alpha_n}\|D_n(\bar{T})\|_2\geq\max_{T\in\mathcal{T}:|\bar{T}\triangle\theta|<\alpha\alpha_n}\|D_n(\bar{T})\|_2\}\\&\subseteq\{\max_{T\in\mathcal{T}:|\bar{T}\triangle\theta|\geq\alpha\alpha_n}\left(\|D_n(\bar{T})\|_2-\|D_n(\bar{T}^*)\|_2\right)\geq 0\}.\end{aligned}$$

The remainder of the proof is identical to the proof of Theorem 2.2, where we simply replace T and T^* by $\bar{T}$ and $\bar{T}^*$ and d_n by ∂_n and note that since $d_n(\bar{T},\theta)=\partial_n(\bar{T},\theta)\leq 1/2$ for all $T\in\mathcal{T}$, the additional assumption of consistency is unnecessary. □

Proof of Lemma 2.2. We want to use the truncation result and exponential inequality of Subsection 4.1. First, note that the integrability assumption and either of the mixing assumptions imply (2), since there are constants $c_1,c_2>0$ (cf. Doukhan (1994)) such that

$$\begin{aligned}&\left|\mathrm{Cov}(Y_{\mathbf{i},n}^{(k)},Y_{\mathbf{j},n}^{(l)})\right|\\&\leq\begin{cases}c_1\alpha\left(dist_n(\{\kappa_{\mathbf{i},n}\},\{\kappa_{\mathbf{j},n}\})\right)^{(r-2)/r}\left(E|Y_{\mathbf{i},n}^{(k)}|^rE|Y_{\mathbf{j},n}^{(l)}|^r\right)^{1/r}, & \text{for }\alpha\text{-mixing}\\ c_2\varphi\left(dist_n(\{\kappa_{\mathbf{i},n}\},\{\kappa_{\mathbf{j},n}\})\right)^{(r-1)/r}\left(E|Y_{\mathbf{i},n}^{(k)}|^{r/(r-1)}\right)^{(r-1)/r}\left(E|Y_{\mathbf{j},n}^{(l)}|^r\right)^{1/r}, & \text{for }\varphi\text{-mixing}\end{cases}\\&\leq\begin{cases}c_1\alpha(\|\mathbf{i}-\mathbf{j}\|_\infty)^{(r-2)/r}D_1, & \text{for }\alpha\text{-mixing}\\ c_2\varphi(\|\mathbf{i}-\mathbf{j}\|_\infty)^{(r-1)/r}D_2, & \text{for }\varphi\text{-mixing}\end{cases}\end{aligned}$$

for any $k,l\in\{1,\dots,p\}$, $n\in\mathbb{N}$ and $\kappa_{\mathbf{i},n},\kappa_{\mathbf{j},n}\in\mathcal{I}_n$, where

$$D_1=\max_{k,l\in\{1,\dots,p\}}\sup_{n\in\mathbb{N}}\max_{\kappa_{\mathbf{i},n},\kappa_{\mathbf{j},n}\in\mathcal{I}_n}\left(E|Y_{\mathbf{i},n}^{(k)}|^rE|Y_{\mathbf{j},n}^{(l)}|^r\right)^{1/r}<\infty$$

and

$$D_2=\max_{k,l\in\{1,\dots,p\}}\sup_{n\in\mathbb{N}}\max_{\kappa_{\mathbf{i},n},\kappa_{\mathbf{j},n}\in\mathcal{I}_n}\left(E|Y_{\mathbf{i},n}^{(k)}|^{r/(r-1)}\right)^{(r-1)/r}\left(E|Y_{\mathbf{j},n}^{(l)}|^r\right)^{1/r}<\infty.$$

Therefore,

$$\begin{aligned}&\max_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}\sum_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}\left|\mathrm{Cov}(Y_{\mathbf{i},n}^{(k)},Y_{\mathbf{j},n}^{(l)})\right|\\&\leq c_1D_1\max_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}\sum_{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}}}\alpha(\|\mathbf{i}-\mathbf{j}\|_\infty)^{(r-2)/r}\\&=c_1D_1\max_{\underline{\mathbf{1}}\leq\mathbf{i}\leq\underline{\mathbf{n}}}\sum_{h=0}^{n}\sum_{\substack{\underline{\mathbf{1}}\leq\mathbf{j}\leq\underline{\mathbf{n}},\\ \|\mathbf{i}-\mathbf{j}\|_\infty=h}}\alpha(h)^{(r-2)/r}\\&\leq 2c_1D_1\sum_{h=0}^{\infty}h^{d-1}\alpha(h)^{(r-2)/r}<\infty\end{aligned}$$

under α-mixing with the assumed rate and analogously

$$\max_{\kappa_{\mathbf{i},n}\in\mathcal{I}_n}\sum_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}\left|\mathrm{Cov}(Y_{\mathbf{i},n}^{(k)},Y_{\mathbf{j},n}^{(l)})\right|\leq 2c_2D_2\sum_{h=0}^{\infty}h^{d-1}\varphi(h)^{(r-1)/r}<\infty$$

under φ-mixing (since $\gamma > \frac{r}{r-1}(d-1)$). For $b_n = \alpha\alpha_n = \alpha n^\eta$, $f(n) = \alpha n^{\frac{d-\eta}{r-1}} = b_n g_n$ with $g_n = n^{-\frac{r\eta-d}{r-1}}$, we consider the processes $Z_n(\cdot)$, $Z_n(\cdot,u,v)$ and $U_n(\cdot,u,v)$ $(0 \le u < v \le \infty)$ as in Lemma 4.1. For any $a > 0$, we obtain

$$\begin{aligned}
&P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{1}{|T\triangle\theta|}\left\|\sum_{\kappa_{\mathbf{i},n}\in T} Y_{\mathbf{i},n} - \sum_{\kappa_{\mathbf{i},n}\in\theta} Y_{\mathbf{i},n}\right\|_2 > \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
=&P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{b_n}{|T\triangle\theta|}\|Z_n(T)-Z_n(\theta)\|_2 > \varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
\le&P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{b_n}{|T\triangle\theta|}\|Z_n(T,0,a)-Z_n(\theta,0,a)\|_2 > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
+&P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{b_n}{|T\triangle\theta|}(U_n([0,1]^d,a,\infty)+EU_n([0,1]^d,a,\infty)) > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)
\end{aligned}$$

By Assumption (C) and Lemma 4.1,

$$P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{b_n}{|T\triangle\theta|}(U_n([0,1]^d,a,\infty)+EU_n([0,1]^d,a,\infty)) > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\overset{n\to\infty}{\longrightarrow} 0.$$

Since

$$\begin{aligned}
&P\left(\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n}\frac{b_n}{|T\triangle\theta|}\|Z_n(T,0,a)-Z_n(\theta,0,a)\|_2 > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
\le&\#\mathcal{T}\max_{T\in\mathcal{T}:|T\triangle\theta|\ge b_n} P\left(\frac{b_n}{|T\triangle\theta|}\|Z_n(T,0,a)-Z_n(\theta,0,a)\|_2 > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right),
\end{aligned}$$

we now consider a fixed $T \in \mathcal{T}$ with $|T\triangle\theta| \ge b_n$. Because of Assumption (C) and the equivalence of the norms $\|\cdot\|_2$ and $\|\cdot\|_\infty$, there is a constant $c > 0$ (which is independent of T and n) such that

$$\begin{aligned}
&P\left(\frac{b_n}{|T\triangle\theta|}\|Z_n(T,0,a)-Z_n(\theta,0,a)\|_2 > \frac{1}{2}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
\le&P\left(\frac{b_n}{|T\triangle\theta|}\|Z_n(T\setminus\theta,0,a)\|_2 > \frac{1}{4}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
&+P\left(\frac{b_n}{|T\triangle\theta|}\|Z_n(\theta\setminus T,0,a)\|_2 > \frac{1}{4}\varepsilon\sigma_{\mathcal{I}}\|a-b\|_2\right)\\
\le&p\left\{\max_{l=1,\dots,p} P\left(\tfrac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(T\setminus\theta,0,a)\right| > c\right) + \max_{l=1,\dots,p} P\left(\tfrac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(\theta\setminus T,0,a)\right| > c\right)\right\}.
\end{aligned}$$

It therefore suffices to consider the coordinate processes. Let $l \in \{1,\dots,p\}$ and set $S_n(A) = \frac{b_n}{|T\triangle\theta|}Z_n^{(l)}(A,0,a) = \sum_{\kappa_{\mathbf{j},n}\in A}\tilde{\eta}_{\mathbf{j},n}$ for $A \subseteq [0,1]^d$, with

$$\tilde{\eta}_{\mathbf{j},n} = \frac{b_n}{|T\triangle\theta|}(\eta_{\mathbf{j},n}^{(l)}(0,a) - E\eta_{\mathbf{j},n}^{(l)}(0,a)).$$

Note that the process $\{b_n\eta_{\mathbf{j},n}(0,a)\}_{\kappa_{\mathbf{j},n}\in\mathcal{I}_n}$ inherits the mixing and integrability properties of the original random field and therefore satisfies (2), which implies Assumption (Y).

Hence, it holds that

$$|\tilde{\eta}_{\mathbf{j},n}| \le \frac{2ab_n}{|T\triangle\theta|} g_n =: C_{1,n} \quad \text{and} \quad E\left[S_n(A)^2\right] \le \overbrace{K\frac{1}{|T\triangle\theta|^2}}^{=:C_{2,n}} |A|,$$

for some $K > 0$ which is independent of A or n. By Lemma 4.2,

$$\begin{aligned}&P\left(\frac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(T\setminus\theta,0,a)\right| > c\right)\\ \le& 2\exp(-\beta_n c)\exp\left(2^{2d}\beta_n^2 C_{2,n} e|T\setminus\theta| + (p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}\right),\end{aligned}$$

and analogously

$$\begin{aligned}&P\left(\frac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(\theta\setminus T,0,a)\right| > c\right)\\ \le& 2\exp(-\beta_n c)\exp\left(2^{2d}\beta_n^2 C_{2,n} e|\theta\setminus T| + (p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}\right),\end{aligned}$$

with $p_n = n^\delta$, $m_n = \frac{n}{2p_n}$, $q_n = 1 + 1/p_n^d$, $0 < \beta_n < \frac{1}{2^{d+1}C_{1,n}m_n^d e}$. It holds that

$$C_{1,n}m_n^d = \frac{2a}{|T\triangle\theta|} b_n g_n \frac{n^d}{2^d p_n^d} \le 2^{1-d} a n^{-\frac{r\eta-d}{r-1}+d-d\delta}$$

and therefore a small enough $a > 0$ can be chosen so that

$$\frac{1}{2^{d+1}C_{1,n}m_n^d e} \ge \frac{1}{4ea} n^{\frac{r\eta-d}{r-1}-d+d\delta} \ge \frac{\mu}{c} n^\xi,$$

since

$$\xi \le \frac{r\eta-d}{r-1} - d + d\delta \Leftrightarrow r \ge \frac{\xi-d\delta}{\xi-d\delta-\eta+d} = \frac{1}{1-\frac{\eta-d}{\xi-d\delta}}$$

for $\xi - d\delta < \eta - d$. Choose $\beta_n = \frac{\mu}{c}n^\xi$. It then holds that

$$\beta_n^2 C_{2,n}\max\{|T\setminus\theta|, |\theta\setminus T|\} \le K\left(\frac{\mu}{c}\right)^2 \frac{1}{|T\triangle\theta|} n^{2\xi} \le K\left(\frac{\mu}{c}\right)^2 \frac{1}{\alpha} n^{2\xi-\eta} \overset{n\to\infty}{\longrightarrow} 0$$

Finally, note that either of the two mixing conditions imply $p_n^d\theta_{q_n}(m_n) = \mathcal{O}(1)$, since

$$p_n^d \alpha^{1/(1+p_n^d)}(m_n) = \mathcal{O}\left(n^{\delta d} e^{-2\log\left(n^{\delta d}\right)\frac{n^{d\delta}}{n^{d\delta}+1}}\right) = \mathcal{O}(1),$$

and

$$p_n^d m_n^d \varphi(m_n) = \mathcal{O}(n^{d-\gamma(1-\delta)}) = \mathcal{O}(1).$$

Therefore, there is a $C > 0$ such that

$$\begin{aligned}&P\left(\frac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(\theta\setminus T,0,a)\right| > c\right)\\ \le& 2\exp(-\mu n^\xi)\exp\left(2^{2d} eK\left(\frac{\mu}{c}\right)^2 \frac{1}{\alpha} n^{2\xi-\eta} + (p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}\right)\\ \le& C\exp(-\mu n^\xi)\end{aligned}$$

and

$$
\begin{aligned}
&P\left(\frac{b_n}{|T\triangle\theta|}\left|Z_n^{(l)}(T\setminus\theta,0,a)\right|>c\right)\\
\leq&2\exp(-\mu n^{\xi})\exp\left(2^{2d}eK\left(\frac{\mu}{c}\right)^2\frac{1}{\alpha}n^{2\xi-\eta}+(p_n^d-1)\theta_{q_n}(m_n)\sqrt{e}\right)\\
\leq&C\exp(-\mu n^{\xi}).
\end{aligned}
$$

□

Chapter 6

Discussion

In this final chapter, the results of the thesis are discussed. In particular, we mention how they extend the already existing theory and discuss some possible combinations and extensions.

The main aim of this thesis was to extend results from change-point analysis to spatial data on a grid. Unlike articles such as Puri and Ruymgaart (1994), the results are not restricted to the case of two-parameter processes but give a unified treatment of random fields over d-dimensional space ($d \in \mathbb{N}$). While the theory of spatial change-point problems is well developed for independent observations, relatively few of the results have been extended to the dependent case. Additionally, a lot of results place restrictions on the distribution or assume a parametric form for the process (cf. e.g. Sharpnack and Arias-Castro (2014)). The main focus of the present work was therefore to present results in a nonparametric framework that would be applicable to a broad range of random fields under simple assumptions that are fulfilled for various dependence concepts.

For the problem of detecting epidemic changes in the mean of real-valued random fields, Bucchia (2014) presents an asymptotic change-point test that is consistent under the alternative, with no restriction on the sign of the change. For this, the approach described by Jarušková and Piterbarg (2011) is extended to weakly dependent data and general d-parameter processes. To make the results applicable to general kinds of weak dependence — including, as a special case, independence —, the inference in Bucchia (2014) is based on a weak invariance principle. As shown by the examples in the paper, one can then draw on a large number of such limit results from the literature. This approach has the further advantage that, aside from square integrability, no restrictions on the distribution of the random variables are necessary. In addition to showing how the continuous mapping theorem can be used to prove the weak convergence of change-point statistics for random fields, the critical values for the test for general dimension d are obtained.

The statistic used in Bucchia (2014) is a kind of scan statistic, where the size of the change-set is assumed to be unknown. Therefore, the results can also be viewed as a generalization of results by Sharpnack and Arias-Castro (2014), who tested for epidemic changes in a signal plus Gaussian white noise model and derived extreme value results for scan statistics under the null hypothesis of constant mean zero.

The procedure for the derivation of the weak limit of the statistic could in principle be used for any test statistic that allows an approximation by a continuous functional of the partial sum process, and thus, in particular, for some statistics with different weight functions for which trimming can be avoided. However, one might then need to resort to a different tail approximation since the theorem by Piterbarg (1996) that was used in Bucchia (2014) only covers fields with constant variance. In Bucchia (2014), it was applicable due to the fact that the specific weighting function used asymptotically corresponds to the variance of the Gaussian field. It would be interesting to look into how the tail approximations for nonhomogeneous Gaussian fields that were obtained e.g. by Piterbarg (1996) (section 8) might be applicable to this setting.

While Jarušková and Piterbarg (2011) focus on the asymptotics under the null hypothesis and assume a known variance, the consistency of the presented test in Bucchia (2014) was proven not only for constant change heights, but also for changes that might vanish asymptotically. Instead of assuming that the long-run variance is given, the results were derived under the assumption that a long-run variance estimator is available which is consistent under the null hypothesis and stochastically bounded under the alternative hypothesis. For this, a kernel-type estimator was discussed, which is consistent under the null (as was proven by Lavancier (2008)). It was noted that this estimator remains stochastically bounded under the alternative for bandwidth q_n and change heights b_n such that $b_n^2 q_n^d = \mathcal{O}(1)$ (cf. Chapter 2, Lemma 3.1). Since one needs $q_n \to \infty$ for the consistency under the null, this means that only changes that vanish asymptotically are allowed. As seen in the proof of consistency (cf. Chapter 2, Theorem 3.3), however, for the test to be consistent, it suffices that $n^{d/2}|b_n|\hat{\sigma}_n^{-1} \xrightarrow{P} \infty$, where $\hat{\sigma}_n$ is the long-run variance estimator. Therefore, the assumption of stochastic boundedness for the long-run variance estimator $\hat{\sigma}_n$ under the alternative could be relaxed while still retaining consistency.

However, too large values of the estimator might impact the power of the test negatively for finite sample sizes. To address this problem, Bucchia and Heuser (2015) extend a method previously employed in the change-point analysis of time series to the random field case. In passing, the consistency (under the null) of the classical long-run variance estimator, which had previously been investigated for univariate random fields, is extended to multivariate random fields. The paper presents error bounds for the kernel-type estimation with general weight function under the assumption that the mean of the random field takes on two values, one inside a change-set and one on the complement of the change-set. Although a data based heuristic for the choice of the bandwidth is presented, the optimal bandwidth choice remains an open problem.

Aside from its application as a scaling factor in test statistics for epidemic changes, the long-run variance estimator is also important for the bootstrap method developed in Bucchia and Wendler (2015). There, the residuals of the random field are weighted by a random field whose covariance structure can be described by a kernel function. As a result, the conditional covariances of the bootstrapped partial sums take the form of long-run variance estimators (cf. the proof of Theorem 2.2 in Chapter 4). Therefore, the joint weak convergence of the finite dimensional distributions of the original and the bootstrapped partial sum processes hinges on the consistency of the long-run variance estimation.

While the results presented in Bucchia (2014) treat real-valued random fields, an analogous argumentation might be used to generalize the results to multivariate observations. For the long-run variance estimation and the estimation of the change-points, this was explicitly done in Bucchia and Heuser (2015). Under the assumption of a multivariate functional central limit theorem, the same argumentation could in principle be employed to derive the limit distribution of a change-point statistic as presented in Jarušková and

Piterbarg (2011). After elimination of the long-run variance matrix in the limit, the limit distribution would be identical to the i.i.d. case and therefore the results by Jarušková and Piterbarg (2011) would yield critical values.

In Bucchia and Wendler (2015), the treatment of the epidemic change testing problem is extended to Hilbert space valued observations. In particular, the special case of multivariate data is included. For the latter, not only changes in the mean but also more generally testing procedures for changes in the marginal distribution are included in the analysis.

To avoid the estimation of the long-run variance operator and make the test applicable in practice, a sequential bootstrap method is introduced which mimics the asymptotic behavior of the partial sum process. To our knowledge, no comparable results — either for the epidemic change problem for Hilbert space valued random fields or the sequential bootstrap in this setting — exist. The bootstrap method we introduced is a variant of the dependent wild bootstrap method by Shao (2010). Aside from its easy computability, this method has the added advantage that, unlike the block bootstrap methods described e.g. in Lahiri (2003), no partition of the data into blocks is necessary. Such a partition requires the treatment of incomplete blocks and therefore leads to edge effects which increase in importance for growing dimension d.

In contrast to works such as Aston and Kirch (2012a), the analysis is not based on projections onto finite dimensional subspaces. This makes it possible to stay in the fully functional framework, avoiding the problems related to choosing suitable subspaces onto which to project the data (cf. e.g. Aston and Kirch (2012b) or Torgovitski (2015)). In particular, a functional central limit theorem for Hilbert space valued random fields is proven and the validity of the sequential bootstrap method is shown by proving the joint weak convergence of the original and the bootstrapped partial sum processes.

As a byproduct which is of independent interest, both a functional central limit theorem under mixing assumptions for multivariate random fields and a general characterization of such limit behavior was obtained. The latter is an extension to multivariate random fields of the functional central limit theorem derived by Deo (1975) (cf. Lemma 3 in Deo (1975)), which is based on a characterization of Brownian sheets and gives general conditions for the convergence which do not presuppose specific dependence assumptions.

In contrast to the general approach of this thesis, the results in Bucchia and Wendler (2015) are derived under specific mixing assumptions. This is due in part to the nature of the results but also to the scarcity of functional central limit theorems for Hilbert space valued or even multivariate random fields in the literature. It might therefore also be of independent interest to research possible generalizations to other types of dependence of the functional central limit theorems presented. From a technical viewpoint, the main ingredients necessary for the application to other types of dependence are the fact that the dependence is inherited by finite dimensional projections, the absolute summability of the (projected) autocovariance functions and the Rosenthal-type inequality for both the partial sums and the bootstrapped partial sums. For multivariate random fields, the general multivariate functional central limit theorem (Chapter 4, Lemma 4.2) might be used to obtain further results under different weak dependence conditions.

Although the theoretical analysis in Bucchia and Wendler (2015) focuses on the behavior of the statistics under the null hypothesis, both the sample mean and the mean estimator suggested by Bucchia and Heuser (2015) were considered for the bootstrap, since the critical values supplied by the empirical quantiles of the bootstrapped statistic are meant to be robust with respect to epidemic changes in the mean. As in the real-valued case, the simulations show that the changed mean estimation leads to tests with higher power under the alternative but also a higher false rejection rate under the null hypothesis. However, tests with either mean estimator do not always adhere to the nominal level and display good power against the alternatives considered. As it then turns out that the test has an over-rejection problem which is compounded by the use of a less than optimal mean estimator under the null, the test with the sample mean as an estimator for the mean function might therefore be preferable in this setting.

A large part of this thesis is concerned with rectangular change-sets. Rectangular sets or their unions are in a sense a natural fit for points on a grid with rectangular mesh. From a technical point of view, rectangular change-sets have several advantages. First, partial sums over any rectangle whose sides are parallel to the coordinate axes can be written as sums and differences of partial sums over rectangles whose lower edge is zero. This is not only useful for practical applications, where the form of the partial sums can be exploited for more efficient computation of the statistics, but also for the derivation of limit theorems. For the latter, sums over rectangles can be viewed as the increments of vector-indexed processes for which a well developed theory of weak convergence (cf. e.g. Neuhaus (1969) and Bickel and Wichura (1971)) and functional central limit theorems for various dependence concepts are available. In this context, note that the maximal inequalities by Móricz (1983) — which have been heavily applied in this thesis — are not only powerful tools to derive the tightness of such processes based on relatively weak assumptions on the moments of the partial sums, but are also more generally of use to bound the stochastic part of CUSUM statistics. Finally, the derivation of critical values for the test in Bucchia (2014) takes advantage of the fact that rectangular change-sets are uniquely defined by $\mathbb{R}^{2d}$-valued parameters in such a way that their volume and the volume of the symmetric difference of two such sets are continuously differentiable functions of the parameters. This makes it possible to view the limit of the test statistic as the maximum of a multiparameter Gaussian process and to give approximations of the local behavior of the covariance function of the limit process using a Taylor expansion.

The testing procedures described in this thesis could in principle be extended to more general classes of sets. For instance, Xie (1996) derives the weak convergence of his statistic for convex subsets of the unit cube and Brodsky and Darkhovsky (1993) present a functional central limit result for their change-set estimator that could be used for a corresponding testing procedure (cf. Theorem 6.1.2 in Brodsky and Darkhovsky (1993)). Hahubia and Mnatsakanov (1996) use a very general framework where, in particular, weak convergence in a generalization of the Skorohod space to set-indexed functions (cf. also Bass and Pyke (1985)) is used to obtain the limit of statistics for the change-set problem.

However, this more general approach creates both theoretical and computational difficulties. From a theoretical viewpoint, fewer functional central limit theorems for set-indexed partial sums are available, most of which concern smoothed partial sums. Part of the reason for this is the lack of a handy maximal inequality like in the rectangle case, which makes stronger dependence assumptions on the random field necessary to show the tightness of the process. For instance, Lin and Lu (1996) use an exponential inequality for truncated partial sums together with Bass's technique to obtain the tightness under mixing assumptions. It is unclear for which (if any) type of weak dependence the results by Hahubia and Mnatsakanov (1996) hold true.

The functional central limit theorems in Bucchia and Wendler (2015) were based on an extension of the weak convergence theory for multiparameter processes to Hilbert space valued processes. For more general classes of sets, both the lack of maximal inequalities and the general lack of theoretical background on which to build would need to be compensated for. For instance, the proof of Theorem 2.1 in Chapter 4 uses an inequality for multiparameter martingales. A thorough review of martingale inequalities of this type in the literature would be required to determine whether they have been — or could realistically be — applied to different classes of sets (for the concept of set-indexed martingales, cf. the monograph by Ivanoff and Merzbach (2000) and further works by these authors).

A further difficulty when extending the results to general classes of sets is the greatly increased computational complexity, which, in particular, complicates the identification of appropriate critical values. For instance, the examples cited above use a functional central limit theorem for set-indexed processes to obtain a functional of the set-indexed Brownian motion as a limit variable. However, since the quantiles of the limit are not tabulated, one would then be faced with the problem of finding appropriate critical values. None of the examples above give results on the computation of critical values for change-sets other than rectangles. For large classes of candidate sets, a Monte-Carlo approach to estimating the limit distribution is computationally intensive.

While there are results such as the approximation of tail probabilities in Piterbarg (1996) or Chan and Lai (2006) for random fields, much less is known about general set-indexed processes. A possible step towards extending the method by Jarušková and Piterbarg (2011) might be to consider other classes of sets that can be uniquely characterized by vector-valued parameters, such as circles or ellipses. As mentioned above, one would then have to deal with the problem of characterizing the local behavior of the resulting Gaussian process, leading to an analysis that is considerably more involved.

Since the derivation of critical values from the bootstrap in Bucchia and Wendler (2015) is essentially based on a Monte-Carlo simulation, the computational complexity would greatly increase for larger classes of change-sets.

In Bucchia and Heuser (2015), the assumption of rectangular change-sets was slightly relaxed to include finite unions of rectangles. As mentioned in Chapter 5, the proofs in Bucchia and Heuser (2015) can be generalized to other classes of change-sets, provided that suitable maximal inequalities are fulfilled (see Chapter 5, Remark 1.3). Additionally, this requires suitable change-set estimators for which rates of convergence are known. As a measure for the accuracy of the change-set estimation in this general context, it is not the Lebesgue measure of the symmetric difference (which was considered in Bucchia and Heuser (2015)) but rather the number of misclassified grid points that is of interest.

In order to obtain such estimators for more general change-sets, Chapter 5 contains additional material which discusses change-set estimation for weakly dependent multivariate observations with a change in the mean.

In Chapter 5, the observations are given on a grid with rectangular mesh where the scaling can be varied separately for each dimension. The aim was to stay in the discrete setting, which corresponds to the knowledge one actually has from the observations. This has the additional benefit that better rates can be derived since one is not trying to estimate a theoretically fixed change-set (i.e. one which is independent of the grid), which, naturally, can only be done up to a certain accuracy due to the coarseness of the grid.

In passing, Chapter 5 also extends the result obtained by Bucchia and Heuser (2015) for rectangular change-sets. For the special case of change-sets of this form, the change-set estimation can be reduced to the estimation of the edge points of the rectangle. This was done in Bucchia (2014) and Bucchia and Heuser (2015), where first the consistency for real-valued random fields and then a rate of convergence for the change-point estimation for multivariate random fields was derived. However, as mentioned above, for the application to long-run variance estimation, the quantity of interest is not the distance between the estimated and the true edge points or the Lebesgue measure of the symmetric difference of the sets, but the number of misclassified grid points. In Chapter 5, rectangular change-sets are directly estimated without a special focus on the estimation of the edge points. With respect to the discrete metric used, a higher rate of convergence than the one proven in Bucchia and Heuser (2015) is then derived.

The change-set and the related change-boundary estimation problem are considered in parallel. Additional identifiability assumptions required for the change-set estimation are discussed, including examples on conditions that can be used to verify these assumptions. Using a CUSUM-type statistic that is a set-indexed analogue of the change-point estimator employed before, general results are derived, which reduce the proofs of consistency and the rate of convergence to the availability of suitable maximal inequalities under relatively weak assumptions. Mainly, it is assumed that the class of candidate sets is rich enough and that the underlying stochastic process fulfills some moment inequality for partial sums. The analysis is made more generally applicable by explicitly considering the case of change heights and change-set sizes that depend on the number of grid points and allowing these to vanish asymptotically.

The applicability of the results is demonstrated by considering the case of rectangular sets, sets that can be approximated by unions of specific rectangular subsets and nested sets as examples. For each of these examples, the assumptions of the theorems are verified and resulting rates are given.

As a byproduct, maximal inequalities for these classes of sets are derived. Additionally, an exponential inequality is proven that combines and extends corresponding inequalities by Lin and Lu (1996) and Valenzuela-Domínguez and Franke (2005).

Such estimation problems have received a lot of attention in the literature — not least because aside from “classical” change-set problems, many questions of image analysis such as edge detection or the reconstruction of a multidimensional regression function from a noisy image can be framed in this context. For instance, for independent observations, Korostelev and Tsybakov (1993) obtained optimal rates of convergence for estimators of change-sets whose boundary is defined by a polynomial. Puri and Ruymgaart (1994) consider estimators for change curves that can be viewed as the graph of a sufficiently smooth monotonically nonincreasing function on $[0, 1]$. More recently, Khmaladze et al. (2006b) considered change-set estimators for changes in the conditional distribution of marks given locations in space. Alternative approaches to the estimation problem are, for instance, described by Wang (1998), who constructed estimators based on wavelets for change curves in an image. However, most of the results contain strong restrictions on the dependence structure of the observations (namely independence) or on the shape of the change-set. Neither of these restrictions are needed for the general discussion of the consistency or rate of convergence in Chapter 5 of this thesis. For the discrete setting considered, related results for the more general framework of a change in the marginal distribution were derived by Ferger (2004) for independent observations, whose proof of bounds for the error probability was the starting point of the analysis in Chapter 5.

The global viewpoint chosen here, where we do not try to classify single grid points but rather search for specific types of candidate sets, is based on the assumption that a priori information about the type of change-set is known, which allows the statistician to choose a suitable class of candidate sets. As noted in Carlstein and Krishnamoorthy (1994) (and seen in the proofs of Chapter 5), this involves the challenge of choosing a class of sets that is rich enough to contain a candidate that is close to the true set while also being small enough to make the derivation of maximal inequalities possible and keep the computation of the estimator feasible.

The examples discussed in Chapter 5 are just a few among the many types of change-set that could be considered. For instance, in Lemma 2.2 of Chapter 5, an exponential inequality was used to derive a maximal inequality for classes of sets whose cardinality fulfills certain restrictions. Using the model by Khmaladze et al. (2006b) for the candidate sets, where it is assumed that the change-set is part of a class of sets for which a δ-net ($\delta = \delta_n$) is available, one could prove corresponding rates of convergence. In contrast to the approach presented in Chapter 5, where the maximal inequality is obtained by taking the number of candidate sets times an exponential bound for each set, Khmaladze et al. (2006b) use the concept of local covering numbers to obtain tighter bounds. It would be interesting to see how their approach could be applied to the setting of Chapter 5.

In closing, we will now mention some further avenues for research. As mentioned above, the computational difficulties involved in both the testing and the estimation problem increase in parallel to the richness of the class of candidate sets and the size of the grid. This makes the derivation of algorithms that not only allow the efficient computation of the statistics, but also remain theoretically tractable, of high interest. To illustrate possible strategies, we now present some existing works in this direction. For the detection of multiple change-points in the mean of a series of independent random variables, Antoch and Jarušková (2013) developed a computation approach for the test statistic based on dynamic programming. Mallik (2013) modified an algorithm by Hartigan (1987) to efficiently deal with the computation of minima over closed convex subsets of $[0,1]^2$ (cf. Mallik (2013), Section 5.1.1). These algorithms have in common that they exploit the specific form of the test statistic. For instance, the algorithm described by Mallik (2013) relies on the fact that the statistic to be minimized is both nonnegative and additive with respect to the sets. For algorithms that are meant to be applicable to more general classes of sets, it might therefore be of interest to deliberately develop test statistics which recursive algorithms can be applied to. For epidemic changes, Sharpnack and Arias-Castro (2014) use a test whose critical values are adapted to the dimensions of the considered rectangle. Since this is computationally intensive, they suggest restricting the scan to an ϵ-covering of the class of rectangles, thereby reducing the number of sets that are considered. Similarly, for the estimation of change-sets that belong to more general classes of sets, Khmaladze et al. (2006b) consider estimators that involve maximization over a covering of the class. It might more generally be of interest to study the behavior of statistics that use not the original class of sets but an approximation thereof. Finally, another approach consists of modeling the change-set in a way that lends itself to easier (recursive) computation. This idea was used e.g. by Müller and Song (1994), who considered change-sets that could be built from unions of rectangular sets and used this property to develop a recursive algorithm for the computation of the resulting change-set estimators.

Bibliography

K.S. Alexander and R. Pyke. A uniform central limit theorem for set-indexed partial-sum processes with finite variance. *Ann. Prob.*, 14(2):582–597, 1986.

D.W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

J. Antoch and M. Hušková. Tests and estimators for epidemic alternatives. *Tatra Mt. Math. Publ.*, 7:311–329, 1996.

J. Antoch and D. Jarušková. Testing for multiple change points. *Comput. Stat.*, 28(5): 2161–2183, 2013.

J. Antoch, M. Hušková, and Z. Prášková. Effect of dependence on statistics for determination of change. *J. Stat. Plan. Infer.*, 60(2):291–310, 1997.

J.A.D. Aston and C. Kirch. Detecting and estimating changes in dependent functional data. *J. Multivar. Anal.*, 109:204–220, 2012a.

J.A.D. Aston and C. Kirch. Evaluating stationarity via change-point alternatives with applications to fMRI data. *Ann. Appl. Stat.*, 6(4):1906–1948, 2012b.

A. Aue and L. Horváth. Structural breaks in time series. *J. Time Ser. Anal.*, 34(1):1–16, 2013.

R.F. Bass and R. Pyke. A strong law of large numbers for partial-sum processes indexed by sets. *Ann. Prob.*, 12(1):268–271, 1984.

R.F. Bass and R. Pyke. The space D(A) and weak convergence for set-indexed processes. *Ann. Prob.*, 13(3):860–884, 1985.

I. Berkes and G.J. Morrow. Strong invariance principles for mixing random fields. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 57(1):15–37, 1981.

I. Berkes, L. Horváth, P. Kokoszka, and Q.-M. Shao. On discriminating between long-range dependence and changes in mean. *Ann. Statist.*, 34(3):1140–1165, 2006.

P.J. Bickel and D.A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.

P.J. Bickel and M.J. Wichura. Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.*, 42(5):1656–1670, 1971.

P. Billingsley. *Convergence of Probability Measures.* Wiley Ser. Probab. Math. Statist. Wiley, New York, 1968.

P. Billingsley. *Convergence of Probability Measures.* Wiley Ser. Probab. Statist. Probab. Statist. Wiley, New York, 2nd edition, 1999.

R.C. Bradley. *Introduction to Strong Mixing Conditions.* Kendrick Press, Heber City, Utah, 2007.

R.C. Bradley and W. Bryc. Multilinear forms and measures of dependence between random variables. *J. Multivar. Anal.*, 16(3):335–367, 1985.

B.E. Brodsky and B.S. Darkhovsky. *Nonparametric Methods in Change-Point Problems*, volume 243 of *Math. Appl.* Kluwer, Dordrecht, 1993.

B. Bucchia. Statistische Analyse epidemischer Strukturbrüche in Zufallsfeldern auf der Basis eines schwachen Invarianzprinzips. Master's thesis, University of Cologne, 2012. `http://www.mi.uni-koeln.de/~bbucchia/mthesis.pdf`[Online; accessed 23.03.2016].

B. Bucchia. Testing for epidemic changes in the mean of a multiparameter stochastic process. *J. Stat. Plan. Infer.*, 150:124–141, 2014.

B. Bucchia and C. Heuser. Long-run variance estimation for spatial data under change-point alternatives. *J. Stat. Plann. Infer.*, 165:104–126, 2015.

B. Bucchia and M. Wendler. Change-point detection and bootstrap for Hilbert space valued random fields. *arXiv preprint arXiv:1511.02609*, 2015.

A. Bücher and I. Kojadinovic. Supplement to "A dependent multiplier bootstrap for the sequential empirical copula process under strong mixing.". 2014. doi: 10.3150/14-BEJ682SUPP.

A. Bulinski. Statistical version of the central limit theorem for vector-valued random fields. *Math. Notes*, 76(4):455–464, 2004.

A. Bulinski and A. Shashkin. Strong invariance principle for dependent random fields. *Lect. Notes - Monogr. Ser.*, 48:128–143, 2006.

A. Bulinski and A. Shashkin. *Limit Theorems for Associated Random Fields and Related Systems*, volume 10 of *Adv. Ser. Stat. Sci. Appl. Probab.* World Scientific, New Jersey, 2007.

E. Carlstein and C. Krishnamoorthy. Boundary estimation. *J. Am. Stat. Assoc.*, 87(418): 430–438, 1992.

E. Carlstein and C. Krishnamoorthy. Practical considerations in boundary estimation: Model-robustness, efficient computation, and bootstrapping. In E. Carlstein, H.G. Müller, and D. Siegmund, editors, *Change-Point Problems*, volume 23 of *Lect. Notes - Monogr. Ser.*, pages 276–283. IMS, Hayward, Calif., 1994.

E. Carlstein, H.G. Müller, and D. Siegmund, editors. *Change-Point Problems*, volume 23 of *Lect. Notes - Monogr. Ser.* IMS, Hayward, Calif., 1994.

H.P. Chan and T.L. Lai. Maxima of asymptotically Gaussian random fields and moderate deviation approximations to boundary crossing probabilities of sums of random variables with multidimensional indices. *Ann. Prob.*, 34(1):80–121, 2006.

X. Chen and H. White. Central limit and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. *Econometric Theory*, 14(2): 260–284, 1998.

T.C. Christofides and E. Vaggelatou. A connection between supermodular ordering and positive/negative association. *J. Multivar. Anal.*, 88(1):138–151, 2004.

C.M. Crainiceanu and T.J. Vogelsang. Spectral density bandwidth choice: source of nonmonotonic power for tests of a mean shift in a time series. *Unpublished manuscript, Department of Economics, Cornell University*, 2001.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis.* Wiley, Chichester, 1997.

J. Davidson. *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford University Press, Oxford, 2002.

J. Dedecker, P. Doukhan, G. Lang, L.R.J. Rafael, S. Louhichi, and C. Prieur. *Weak Dependence: With Examples and Applications*, volume 190 of *Lect. Notes Stat.* Springer, New York, 2007.

H. Dehling, O.S. Sharipov, and M. Wendler. Bootstrap for dependent Hilbert space-valued random variables with application to von Mises statistics. *J. Multivar. Anal.*, 133:200–215, 2015.

C.M. Deo. A functional central limit theorem for stationary random fields. *Ann. Probab.*, 3(4):708–715, 1975.

C.M. Deo. A note on ϕ-mixing random fields. *Theory Prob. Appl.*, 21(4):867–870, 1976.

P. Doukhan. *Mixing: Properties and Examples.* Springer, New York, 1994.

P. Doukhan, G. Lang, A. Leucht, and M.H. Neumann. Dependent wild bootstrap for the empirical process. *J. Time Ser. Anal.*, 36(3):290–314, 2015.

R.M. Dudley. A course on empirical processes. In *École d'été de Probabilités de Saint-Flour XII–1982*, volume 1097 of *Lect. Notes Math.*, pages 1–142. Springer, Berlin, 1984.

L. Dümbgen. The asymptotic behavior of some nonparametric change-point estimators. *Ann. Statist.*, 19(3):1471–1495, 1991.

V.R. Eastwood. Some nonparametric methods for changepoint problems. *Can. J. Stat.*, 21(2):209–222, 1993.

M. El Machkouri. Nonparametric regression estimation for random fields in a fixed-design. *Stat. Inference Stoch. Process.*, 10(1):29–47, 2007.

M. El Machkouri, D. Volný, and W.B. Wu. A central limit theorem for stationary random fields. *Stochastic Process. Appl.*, 123(1):1–14, 2013.

I. Fazekas, A.G. Kukush, and T. Tómács. On the Rosenthal inequality for mixing fields. *Ukrainian Mathematical Journal*, 52(2):305–318, 2000.

D. Ferger. Boundary estimation based on set-indexed empirical processes. *J. Nonparametr. Statist.*, 16(1-2):245–260, 2004.

T. Garlipp and C.H. Müller. Robust jump detection in regression surface. *Sankhyā: The Indian Journal of Statistics*, 69(1):55–86, 2007.

L. Giraitis, P. Kokoszka, R. Leipus, and G. Teyssière. Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics*, 112(2):265–294, 2003.

C.M. Goldie and P.E. Greenwood. Variance of set-indexed sums of mixing random variables and weak convergence of set-indexed processes. *Ann. Prob.*, 14(3):817–839, 1986.

E. Gombay and L. Horváth. Change-points and bootstrap. *Environmetrics*, 10(6):725–736, 1999.

D.A. Griffith and L.J. Layne. *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets.* Oxford University Press, New York, 1999.

O. Gromenko and P. Kokoszka. Testing the equality of mean functions of ionospheric critical frequency curves. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 61(5):715–731, 2012.

O. Gromenko, P. Kokoszka, and M. Reimherr. Detection of change in the spatiotemporal mean function. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 2016.

X. Guyon. *Random Fields on a Network: Modeling, Statistics, and Applications.* Probability and its Applications. Springer Science & Business Media, New York, 1995.

T.S. Hahubia and R. Mnatsakanov. On the mode-change problem for random measures. *Georgian Mathematical Journal*, 3(4):343–362, 1996.

P. Hall. Resampling a coverage pattern. *Stochastic Process. Appl.*, 20(2):231–246, 1985.

J.A. Hartigan. Estimation of a convex density contour in two dimensions. *J. Am. Stat. Assoc.*, 82(397):267–270, 1987.

M. Holmes, I. Kojadinovic, and J.-F. Quessy. Nonparametric tests for change-point detection à la Gombay and Horváth. *J. Multivar. Anal.*, 115:16–32, 2013.

H. Huang, M. Gu, and H. Chao. An efficient method of license plate location in natural-scene image. In *Fifth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD '08)*, volume 4, pages 15–19, Washington, DC, USA, 2008.

M. Hušková. Estimators for epidemic alternatives. *Comment. Math. Univ. Carolinae*, 36 (2):279–292, 1995.

M. Hušková and C. Kirch. A note on studentized confidence intervals for the change-point. *Computation. Stat.*, 25(2):269–289, 2010.

I.A. Ibragimov. A note on the central limit theorems for dependent random variables. *Theory Probab. Appl.*, 20(1):135–141, 1975.

A. Inoue. Testing for distributional change in time series. *Econometric Theory*, 17(1): 156–187, 2001.

B.G. Ivanoff and E. Merzbach. *Set-Indexed Martingales*, volume 85 of *Monogr. Statist. Appl. Probab.* Chapman & Hall / CRC, Boca Raton, 2000.

B.G. Ivanoff and E. Merzbach. Optimal detection of a change-set in a spatial Poisson process. *Ann. Appl. Probab.*, 20(2):640–659, 2010.

D. Jarušková. Detecting non-simultaneous changes in means of vectors. *TEST*, 24(4): 681–700, 2015.

D. Jarušková. Detection of transient change in mean — a linear behavior inside epidemic interval. *Kybernetika*, 47(6):866–879, 2011.

D. Jarušková and V.I. Piterbarg. Log-likelihood ratio test for detecting transient change. *Statist. Probab. Lett.*, 81(5):552–559, 2011.

T. Juhl and Z. Xiao. Tests for changing mean with monotonic power. *J. Econometrics*, 148(1):14–24, 2009.

Z. Kabluchko and E. Spodarev. Scan statistics of Lévy noises and marked empirical processes. *Adv. Appl. Probab.*, 41(1):13–37, 2009.

M. Kejriwal. Tests for a mean shift with good size and monotonic power. *Econ. Lett.*, 102(2):78–82, 2009.

E. Khmaladze, R. Mnatsakanov, and N. Toronjadze. The change-set problem for Vapnik-Červonenkis classes. *Math. Methods Statist.*, 15(2):224–231, 2006a.

E. Khmaladze, R. Mnatsakanov, and N. Toronjadze. The change set problem and local covering numbers. *Math. Methods Statist.*, 15(3):289–309, 2006b.

D. Khoshnevisan. *Multiparameter Processes: An Introduction to Random Fields.* Springer Science & Business Media, New York, 2002.

K.I. Kim, K. Jung, and J.H. Kim. Color texture-based object detection: an application to license plate localization. In S.-W. Lee and A. Verri, editors, *Pattern Recognition with Support Vector Machines*, volume 2388 of *Lect. Notes Comput. Sci.*, pages 293–309. Springer, 2002.

T.S. Kim and E.Y. Seok. The invariance principle for ρ-mixing random fields. *J. Korean Math. Soc.*, 32(2):321–328, 1995.

M. Ko, H.C. Kim, and T. Kim. On functional central limit theorems for linear random fields with dependent innovations. *ANZIAM J.*, 49(4):533–541, 2008.

A.P. Korostelev and A.B. Tsybakov. *Minimax Theory of Image Reconstruction*, volume 82 of *Lect. Notes Stat.* Springer, New York, 1993.

M.R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference.* Springer Ser. Statist. Springer, New York, 2010.

S.N. Lahiri. *Resampling Methods for Dependent Data.* Springer Science & Business Media, New York, 2003.

F. Lavancier. The V/S test of long-range dependence in random fields. *Electron. J. Statist.*, 2:1373–1390, 2008.

B. Levin and J. Kline. The cusum test of homogeneity with an application to spontaneous abortion epidemiology. *Stat. Med.*, 4(4):469–488, 1985.

Z. Lin and C. Lu. *Limit Theory for Mixing Dependent Random Variables*, volume 378 of *Math. Appl.* Kluwer, Dordrecht, 1996.

C.R. Loader. Large-deviation approximations to the distribution of scan statistics. *Adv. in Appl. Probab.*, 23(4):751–771, 1991.

J.B. MacNeill and V.K. Jandhyala. Change-point methods for spatial data. In G.P. Patil and C.R. Rao, editors, *Multivariate Environmental Statistics*, volume 6 of *North-Holland Series in Statistics and Probability*, chapter 14, pages 289–306. Elsevier Science Publishers, Amsterdam, North-Holland, 1993.

A. Majumdar, A.E. Gelfand, and S. Banerjee. Spatio-temporal change-point modeling. *J. Stat. Plan. Infer.*, 130(1):149–166, 2005.

A. Mallik. *Topics on threshold estimation, multistage methods and random fields.* PhD thesis, University of Michigan, 2013.

D. Marinucci and S. Poghosyan. Asymptotics for linear random fields. *Statist. Probab. Lett.*, 51(2):131–141, 2001.

H. Moon, R. Chellappa, and A. Rosenfeld. Performance analysis of a simple vehicle detection algorithm. *Image Vision Comput.*, 20(1):1–13, 2002.

F. Móricz. A general moment inequality for the maximum of partial sums of single series. *Acta Sci. Math.*, 44:67–75, 1982.

F. Móricz. A general moment inequality for the maximum of the rectangular partial sums of multiple series. *Acta Math. Hungar.*, 41(3–4):337–346, 1983.

H.G. Müller and K.-S. Song. Cube splitting in multidimensional edge estimation. In E. Carlstein, H.G. Müller, and D. Siegmund, editors, *Change-Point Problems*, volume 23 of *Lect. Notes - Monogr. Ser.*, pages 210–223. IMS, Hayward, Calif., 1994.

H.G. Müller and K.-S. Song. A set-indexed process in a two-region image. *Stochastic Process. Appl.*, 62(1):87–101, 1996.

U.V. Naik-Nimbalkar and M.B. Rajarshi. Validity of blockwise bootstrap for empirical processes with stationary observations. *Ann. Statist.*, 22(2):980–994, 1994.

G. Neuhaus. *Zur Theorie der Konvergenz stochastischer Prozesse mit mehrdimensionalem Zeitparameter.* PhD thesis, Westfälische Wilhelms-Universität zu Münster, 1969.

Neuroonkologische Arbeitsgemeinschaft (NOA). Onko-Internetportal - Diagnose von Hirntumoren, 2012. http://www.krebsgesellschaft.de/onko-internetportal/basis-informationen-krebs/krebsarten/hirntumor/diagnose-von-hirntumoren.html [online, accessed 23.03.2016].

W.K. Newey and K.D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix. *Econometrica*, 55:703–708, 1987.

C.M. Newman. Asymptotic independence and limit theorems for positively and negatively dependent random variables. *Lect. Notes - Monogr. Ser.*, 5:127–140, 1984.

M. Peligrad. On the blockwise bootstrap for empirical processes for stationary sequences. *Ann. Probab.*, 26(2):877–901, 1998.

V.I. Piterbarg. *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, volume 148 of *Transl. Math. Monogr.* Amer. Math. Soc., Providence, 1996.

S. Poghosyan and S. Rœlly. Invariance principle for martingale-difference random fields. *Statist. Probab. Lett.*, 38(3):235–245, 1998.

D.N. Politis and J.P. Romano. Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivar. Anal.*, 47(2):301–328, 1993.

D.N. Politis and J.P. Romano. Limit theorems for weakly dependent Hilbert space valued random variables with applications to the stationary bootstrap. *Statist. Sinica*, 4(2): 461–476, 1994.

D.N. Politis and J.P. Romano. On flat-top kernel spectral density estimators for homogeneous random fields. *J. Stat. Plann. Infer.*, 51(1):41–53, 1996.

M.L. Puri and F.H. Ruymgaart. Change curves in the presence of dependent noise. In E. Carlstein, H.G. Müller, and D. Siegmund, editors, *Change-Point Problems*, volume 23 of *Lect. Notes - Monogr. Ser.*, pages 242–254. IMS, Hayward, Calif., 1994.

P. Qiu. *Image Processing and Jump Regression Analysis.* Wiley Ser. Probab. Statist. John Wiley & Sons, Hoboken, N.J., 2005.

D. Radulović. Another look at the disjoint blocks bootstrap. *TEST*, 18(1):195–212, 2009.

A. Rajaraman and J.D. Ullman. *Mining of Massive Datasets.* Camb. Univ. Press, Cambridge, 2012.

A. Račkauskas and C. Suquet. Hölder norm test statistics for epidemic change. *J. Stat. Plann. Infer.*, 126(2):495–520, 2004.

P.M. Robinson. Nonparametric spectrum estimation for spatial data. *J. Stat. Plan. Infer.*, 137(3):1024–1034, 2007.

X. Shao. The dependent wild bootstrap. *J. Amer. Statist. Assoc.*, 105(489):218–235, 2010.

O.S. Sharipov, J. Tewes, and M. Wendler. Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canad. J. Statist.*, 44(3):300–322, 2016.

J. Sharpnack and E. Arias-Castro. Exact asymptotics for the scan statistic and fast alternatives. *arXiv preprint arXiv:1409.7127*, 2014.

D. Siegmund and B. Yakir. Tail probabilities for the null distribution of scanning statistics. *Bernoulli*, 6(2):191–213, 2000.

D. Tjøstheim. Statistical spatial series modelling. *Adv. Appl. Probab.*, 10(1):130–154, 1978.

C. Tone. A central limit theorem for multivariate strongly mixing random fields. *Probab. Math. Statist.*, 30(2):215–222, 2010.

C. Tone. Central limit theorems for Hilbert-space valued random fields satisfying a strong mixing condition. *ALEA Lat. Am. J. Probab. Math. Stat.*, 8:77–94, 2011.

L. Torgovitski. Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components. *arXiv preprint arXiv:1509.07409*, 2015.

L. Truquet. *Propriétés théoriques et applications en statistique et en simulation de processus et de champs aléatoires stationnaires.* PhD thesis, Université Panthéon-Sorbonne-Paris I, 2008.

E. Valenzuela-Domínguez and J. Franke. A Bernstein inequality for strongly mixing spatial random processes. Technical report, DFG-Schwerpunktprogramm 1114, Mathematical methods for time series analysis and digital image processing ; Preprint 76, ISBN 3-88722-650-X, 2005.

A.W. van der Vaart and J.A. Wellner. *Weak Convergence and Empirical Processes.* Springer Ser. Statist. Springer, New York, 1996.

S. Vinson and L.D. Cohen. Multiple rectangle model for buildings segmentation and 3D scene reconstruction. In *Proc. of ICPR Int. Conf. on Pattern Recognition*, volume 3, pages 623–626, Québec, Canada, 2002.

S. Vinson, L.D. Cohen, and F. Perlant. Extraction of rectangular buildings in aerial images. In *SCIA '01*, pages 431–438, Bergen, Norway, 2001.

T.J. Vogelsang. Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series. *J. Econometrics*, 88(2):283–299, 1999.

M.A. Vronski. *Some limit theorems for associated random fields.* PhD thesis, Moscow State University, 1998.

Y. Wang. Change curve estimation via wavelets. *J. Am. Stat. Assoc.*, 93(441):163–172, 1998.

M.J. Wichura. Inequalities with applications to the weak convergence of random processes with multi-dimensional time parameters. *Ann. Statist.*, 40(2):681–687, 1969.

C.F.J. Wu. Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, 14(4):1261–1295, 1986.

L. Xie. *Detection and Estimation of Boundaries in Spatial Data.* PhD thesis, University of Western Ontario, 1996.

Q. Yao. Tests for change-points with epidemic alternatives. *Biometrika*, 80(1):171–191, 1993.

V. Zemlys. *Invariance principle for multiparameter summation processes and applications.* PhD thesis, Vilnius University, 2008.

L. Zhang. Rosenthal type inequalities for B-valued strong mixing random fields and their applications. *Sci. China Ser. A*, 41(7):736–745, 1998.

Y.F. Zhang and R.R. Bresee. Fabric defect detection and classification using image analysis. *Text. Res. J.*, 65(1):1–9, 1995.

J. Zhu and S.N. Lahiri. Bootstrapping the empirical distribution function of a spatial process. *Stat. Inference Stoch. Process.*, 10(2):107–145, 2007.

Y. Zhu, B. Carragher, D.J. Kriegman, R.A. Milligan, and C.S. Potter. Automated identification of filaments in cryoelectron microscopy images. *J. Struct. Biol.*, 135(3): 302–312, 2001.

Erklärung zum Eigenanteil

Der Artikel „*Testing for epdimic changes in the mean of a multiparameter process*“ ist in enger Zusammenarbeit mit Christoph Heuser entstanden. Dabei wurden sowohl die Entwicklung der mathematischen Theorie als auch die Durchführung von Simulationsstudien gemeinsam gemacht. Die so veröffentlichte Endfassung habe ich verfasst. Der Eigenanteil liegt daher bei ca. 55%.

Zur Zusammenarbeit mit Prof. Dr. Wendler am Artikel „*Change-point detection and bootstrap for Hilbert space valued random fields*“ ist anzumerken, dass Prof. Dr. Wendler für die Beschreibung und Herleitung der Resultate zum Bootstrap-Verfahren zuständig war und daher den entsprechenden Teil der Einleitung sowie den Beweis von Theorem 2.2 (sowie die dazu benötigten Hilfsresultate Lemma 4.5 und Lemma 4.6) beigetragen hat. Alle übrigen Resultate, Erläuterungen etc. stammen von mir und wurden von Prof. Dr. Wendler lediglich noch einmal gegengelesen. Die mathematischen Resultate, die nicht von mir hergeleitet wurden, wurden von mir überarbeitet und um weitere Anmerkungen ergänzt. Die Simulationsstudie wurde vollständig von mir, lediglich unter Rücksprache mit Prof. Dr. Wendler, durchgeführt. Der Eigenanteil bei der Herleitung und Formulierung der Resultate als Artikel liegt daher bei ca. 75%.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Professor i. R. Dr. Steinebach betreut worden.

Köln, den

(Béatrice Bucchia)

Teilpublikationen

Testing for epidemic changes in the mean of a multiparameter stochastic process, *Journal of Statistical Planning and Inference 150 (2014), 124-141.*
doi: 10.1016/j.jspi.2014.03.001
Long-run variance estimation for spatial data under change-point alternatives, *Journal of Statistical Planning and Inference 165 (2015), 104-126* (mit C. Heuser).
doi: 10.1016/j.jspi.2015.04.005
Change-point detection and bootstrap for Hilbert space valued random fields, *Journal of Multivariate Analysis.* eingereicht (unter Revision). (mit M. Wendler)