Preprint, Institute of Statistics, RWTH Aachen University

DECOUPLING CHANGE-POINT DETECTION BASED ON CHARACTERISTIC FUNCTIONS: METHODOLOGY, ASYMPTOTICS, SUBSAMPLING AND APPLICATION

Ansgar Steland[1] and Ewaryst Rafajłowicz

Institute of Statistics
RWTH Aachen University
Wüllnerstr. 3, D-52056 Aachen, Germany
steland@stochastik.rwth-aachen.de
and
Department of Computer Engineering, Control and Robotics
University of Technology Wrocław
Wrocław, Poland
ewaryst.rafajlowicz@pwr.wroc.pl

We study the problem of detecting changes in a location scale model. Our novel detector is based on sequential estimates of two indicators derived from the characteristic function (ch.f.), which allow to decouple the location from the scale problem. The asymptotic theory treats general weighted integrals of nonlinear functions of the real and imaginary parts of (sequential) estimates of the characteristic function and covers (functional) central limit theorems as well as the corresponding subsampling versions, where the latter allow for resampling-based estimation of control limits. In this way, we provide a unifying approach and provide a base for practical implementations of the procedures. Our results also reveal that the estimated indicators have different convergence rates. This explains the decoupling and clustering effects observed in practice and is also in contrast to the case of the sample mean and sample variance, which share the same convergence rate. Monte Carlo simulations show that the effect is also present in finite samples and that the proposed monitoring procedures are powerful, especially for small shifts. Our simulations also show that subsampling with calibration leads to accurate estimation of control limits even in small samples. Lastly, we illustrate our procedure by applying it to the monitoring of intraday climate data.

Keywords: Change detection, climate data, control chart, functional data, location-scale model, subsampling, time series.

1. INTRODUCTION

Let $Y_j \sim F_j,\ j = 1, 2, \dots$ be a sequence of random variables. When the observations arrive sequentially, thus forming a data stream, one is often interested in detecting changes of the location and scale by applying monitoring procedures (control charts). The problem arises in diverse areas of applications such as econometrics, finance, environmetrics, physics and engineering, and we shall illustrate our findings by an application to monitoring of environmental data. The approach proposed in this paper has the advantage that it is based on characteristics of the underlying distribution and associated estimates, which decouple the location and scale problem. At this point it is worth recalling that that many test classic statistics and control charts are based on the statistic $T = |\overline{Y} - \mu_0|/S$, where μ_0 denotes the nominal

[1]corresponding author

(in-control or target) location parameter, while $\overline{Y}$ and S are the standard estimates of expectation and dispersion, respectively. The statistic T serves as an estimator of the normalized shift, $|\mu - \mu_0|/\sigma$, where μ is the true mean and σ the scale parameter (usually the dispersion). If we admit simultaneous changes in location and scale, it may happen that the indicator statistic T remains unchanged, since location changes are partly masked by changes of the scale. Further, the control limits of frequently used charts depend on nuisance parameters, whereas the control statistic of our proposal to monitor σ is a pivot.

The empirical ch.f., the related characteristic process and its application to tests date back to the seminal works of [23] and [12] and have been investigated for various statistical problems such as tests for symmetry, see [12] or [18] and the discussion given there, or tests for dependence, see e.g. [11]. Change-point detection and tests based on empirical ch.f.s has been studied recently by [19, 20, 21]. For a general and extensive monograph on ch.f.s we refer to [48]; for treatments of the general methodology of detection procedures, sequential (on-line) as well as non-sequential (off-line), and other recent results we refer to [6], [33], [8], [34], [32] and [45, ch.9]. Further mathematical background and results on sequential estimation can be found in [13] and [51], amongst others.

The method we shall introduce in the present paper uses a novel idea that relies on the fact that in a location scale family with symmetric generic distribution function (d.f.) the ratio of the real and imaginary part of the ch.f. depends solely on the location parameter, whereas its absolute value is a function of the scale parameter; details will be given in Section 2. In this way, we can nicely decouple the location and scale problems. The asymptotic theory developed in this paper shows that the proposed nonparametric detection statistics have different convergence rates, which helps explaining the decoupling and clustering effect observed in practice, see Section 2.2. This is in contrast to the case of the sample mean and variance. Here the underlying parameters are coupled, since the variance depends on the mean and they share the convergence rate $1/\sqrt{n}$. From an applied viewpoint, our method has the advantage that it is not necessary to collect rational subgroups at each time point or, alternatively, group observations in subgroups of size l and apply a chart to these subgroups leading to a substantially coarser time scale of the detector, as it is the case for classic methods such as the R (range) or S charts or CUSUM schemes recently discussed in [16].

Monitoring procedures should meet well defined statistical criteria such as a nominal significance level or a bounded average run length (ARL) when the underlying series (process) is *in-control*, i.e. if the null hypothesis of no change holds true. We propose to rely on the subsampling approach of [30], see also [31], that provides consistent resampling approximations under weak regularity assumptions. Here the procedure is recalculated from shorter subseries to approximate its distribution. In this way, one can easily design the proposed monitoring procedures.

To introduce the change-point model, let us tentatively assume that the observations are independent. We shall later relax that assumption substantially and allow for a large class of nonlinear time series of the form $Y_n = f(\epsilon_n, \epsilon_{n-1}, \epsilon_{n-2}, \dots)$ for an i.i.d. noise process $\{\epsilon_n\}$ and a measurable function f. Such nonlinear models appear naturally when nonlinear filters are applied to preprocess the data, which is a widely spread technique in imaging as well as time series analysis.

We are interested in the following basic (at most one change-point point) detection problem of sequential analysis. Suppose that random variables $Y_1, Y_2, \dots$ with distribution functions (d.f.s) $F_1, F_2, \ldots$ arrive sequentially and assume the change-point model

$$F_1 = \cdots = F_{q-1} \neq F_q = \cdots = F_N$$

where $N \in \mathbb{N}$ denotes the maximum sample size. F_1 denotes the d.f. of the measurements when the process is in control and F_q stands for the out-of-control distribution. If $F_1 \neq F_q$ and $1 < q$, there is a change in the distribution starting at observation q. The goal is to detect the occurrence of the change-point during the monitoring period $1, \ldots, N$ as early as possible by applying a control chart defined by a stopping time. If the chart provides a signal, one concludes that there was a change-point (structural break). An important special case of the general change-point problem on which we shall focus is the location-scale problem

$$F_1 = F \qquad \text{and} \qquad F_q = F\left(\frac{\cdot - \mu}{\sigma}\right) \tag{1.1}$$

for some fixed d.f. F with $F(x) = 1 - F(-x)$, $x \in \mathbb{R}$.

The classic problem formulation in quality control assumes that F and therefore the corresponding characteristic function (ch.f.) is known to us. However, our results also cover the case when the ch.f. is unknown and estimated from either a learning sample of in-control observations right before monitoring starts, or when it is estimated sequentially using all available observations except the most recent ones. The idea behind the latter approach is that when there was no signal yet, one may use almost all of them to estimate unknown quantities depending on the in-control distribution, since those measurements are likely distributed according to F_1.

We approach the problem as follows: When studying procedures based on (estimated) ch.f.s, it is natural to associate to each observation Y_j the random function

$$e^{itY_j} = \cos(tY_j) + i\sin(tY_j), \qquad t \in \mathbb{R}, \tag{1.2}$$

yielding a *functional data set* of continuous functions. Indeed, taking this standpoint and using methods which can cope with functional data will be the key for general results which reveal the analytic reasons why the proposed statistics for location and scale behave quite differently. Indeed, we found that they have different convergence rates and follow different types of asymptotic laws.

Behind our procedures are unbiased nonparametric estimates $\widehat{C}_l^k(t)$ of the function $C(t) = (E\cos(tY_1), E\sin(tY_1)))'$, $t \in \mathbb{T}$, defining the ch.f., which depend on the subsample $Y_l, \ldots, Y_k$, for integers $l \leq k$. Given those estimates, we consider the class of control statistics given by

$$\int_{\mathbb{T}} \psi(\widehat{C}_l^k(t))w(t)\,dt - \int_{\mathbb{T}} \psi(C(t))w(t)\,dt, \qquad l \leq k \leq N, N \in \mathbb{N}, \tag{1.3}$$

for a suitable smooth function ψ defined on $\cap_{l\leq k\leq N, N\geq 1}\Im(\widehat{C}_k^l) \cap \Im(C)$ and a nonnegative weighting function $w(t)$ defined on some appropriately chosen domain $\mathbb{T}$ such that the above integrals and $\int w(t)\,dt$ exist. For sake of brevity of notation, here and in what follows we omit the area of integration. Here $\Im(f)$ denotes the image of some function f defined on $\mathbb{T}$. For the specific ψ-functions we shall use to handle the location scale detection problem, further details will be discussed later.

The outline of the paper is as follows. In Section 2, we review some basic facts on characteristic functions and introduce the decoupling indicators, which are sensitive with respect to location and scale changes, respectively, and unique up to smooth transformations. We also introduce the proposed sequential estimates yielding our control statistics. Section 3 introduces the nonlinear time series model to which our results basically apply, discusses how the treatment of the proposed class of detectors can be traced back to well known characteristic process and presents our contributions to its study. In Section 4, we present and discuss our theoretical results on the monitoring procedures under the null hypothesis of no change. We show that and explain why these procedures have qualitatively different limits. Section 5 discusses the subsampling approach in order to estimate the distribution of the procedure, in particular of the stopping time defining the detector. Section 6 complements the theoretical results by providing the asymptotic theory under local alternatives and identifies the right order to obtain non-trivial limits. The performance of the proposed detectors is studied by simulations in Section 7. We study to which extent the theoretical decoupling effect carries over to statistical practice, compare our proposals with some competitors and investigate the accuracy of the subsampling approximations. Lastly, we illustrate in Section 8 our procedures by applying them to the monitoring of intraday SO_2 measurements, motivated by the role of SO_2 as an air pollutant that has an climate effect, see [25].

2. INDICATORS FOR LOCATION AND SCALE SHIFTS AND RELATED CONTROL CHART DETECTION PROCEDURES

This section provides more details on the basic change-point model of interest and discusses the proposed indicators to measure changes in the location and scale of symmetric distributions, respectively. To the best of our knowledge, the location indicator has not yet been studied in the literature. We show that both indicators are unique up to C^1 transformations, which provides a rigorous mathematical justification. We also demonstrate general their potential for data analysis when estimating them nonparametrically, particularly their ability to decouple changes in the mean and scale of measurements. Moving window (sequential) versions of those nonparametric estimators shall form the base for the control statistics of our detectors.

After those preliminaries, we describe in detail the construction of appropriate control chart detection procedures for monitoring. They will be based on stopping times based on nonparametric moving window estimates of the indicators.

2.1. *Preliminaries and Decoupling Indicators for Location and Scale*

Let us start by introducing further notation and reviewing some elementary facts, which will lead to our proposal for a control chart to monitor changes in the location or scale of a sequence of observations. Recall that the ch.f. of a random variable X is given by

$$\varphi_X(t) = E\exp(itX) = R_X(t) + iI_X(t), \qquad t \in \mathbb{R},$$

where $i^2 = -1$ and $R_X(t) = E\cos(tX),\ I_X(t) = E\sin(tX),\ t \in \mathbb{R}$, are the real and imaginary parts, respectively. The location-scale model above can be re-phrased as follows. Assume that X satisfies $X \stackrel{d}{=} -X$ and let $Y_1, Y_2, \dots$ be a sequence of independent random variables such that $Y_1, \dots, Y_{q-1}$ follow the distribution of X and $Y_q, Y_{q+1}, \dots$ follow a location-scale model, $Y = \mu + \sigma X$, for constants $\mu \in \mathbb{R}$ and $\sigma > 0$. If $\mu \neq 0$ or $\sigma \neq 1$, then q is a change-point (structural break point) where the location and/or scale of the measurements changes. We assume that the change occurs after a certain fraction of the maximum sample size N, i.e.

$$q = \lfloor N\tau \rfloor \qquad \text{for some } \tau \in (s_0, 1),$$

where s_0 determines the start of monitoring on the $[0,1]$ time scale, see below. As common in sequential detection and its applications, specifically in quality control, we tentatively assume that the distribution of X and therefore its ch.f., φ_X, is known to us. We discuss at the end of the present section how to handle the case when φ_X is unknown. The ch.f. of the observations Y_j, $j \geq q$, after the change is $\varphi_Y(t) = R_Y(t) + iI_Y(t)$ with

$$R_Y(t) = \cos(\mu t)\varphi_X(\sigma t) \qquad \text{and} \qquad I_Y(t) = \sin(\mu t)\varphi_X(\sigma t),$$

by symmetry of X which implies $I_X = 0$. Now observe that when taking the ratio, $\varphi_X(\sigma t)$ cancels and we are left with $\tan(\mu t)$. This motivates to introduce the indicators

$$M(Y)_t = \frac{I_Y(t)}{R_Y(t)} = \tan(\mu t) \qquad \text{and} \qquad S(Y)_t = R_Y^2(t) + I_Y^2(t) = |\varphi_X(\sigma t)|^2,$$

which are well defined on the real line under the no change hypothesis, whereas under a fixed alternative given by $\mu > 0$ one may select the domain

$$\mathbb{T} = \cup_k \{(-\pi/(2\mu) + \delta, \pi/(2\mu) - \delta) + (2/\mu)\pi k\},$$

for $t \in \mathbb{T}$, where $\delta > 0$ is fixed. Notice that under a sequence of alternatives $\mu = \mu_N \downarrow 0$, which we shall study in Chapter 6, $(-\pi/(2\mu_N) + \delta, \pi/(2\mu_N) - \delta) \to (-\infty, \infty)$, as $N \to \infty$, such that the choice of $\mathbb{T}$ poses no problem in that setting. From now on we assume that $\mathbb{T} = [T_0, T_1]$ for $0 < T_0 < T_1 < \infty$, and all integrals with respect to dt are taken over the domain of integration $[T_0, T_1]$.

For our change-point model, we have the explicit formulas

$$M(Y_i)_t = \tan(\mu t) 1(i \geq q), \tag{2.1}$$

$$S(Y_i)_t = |\varphi_X(t)|^2 1(i < q) + |\varphi_X(\sigma t)|^2 1(i \geq q), \tag{2.2}$$

where $1(A) = 1$ if A holds true and $= 0$ otherwise. Notice that then the associated integrals

$$M = \int \left(\frac{I_Y(t)}{R_Y(t)} \right)^2 w(t)\, dt, \qquad S = \int (R_Y^2(t) + I_Y^2(t)) w(t)\, dt$$

corresponding to the ψ-functions $\psi_L(z) = z_2/z_1$ and $\psi_S(z) = z_1^2 + z_2^2$, $z = (z_1, z_2) \in \mathbb{R}$, respectively, are well defined for a large class of weighting functions $w(t)$ including the identity, the indicator on the interval $[-a, a]$ for some $a > 0$ and the choice $w(t) = e^{-w_0 t}$ frequently used in the literature. Such weight functions are used to truncate or dampen the influence of the tails of the characteristic function which are hard to estimate. The latter choice of an exponential avoids sudden truncation. Further interesting candidates for such weighting functions could be functions that are commonly used in signal processing under the names: Bartlett's, Hamming's and Blackman's windows, see, e.g., [35].

Equations (2.1) and (2.2) show that the functional $M(Y_i)$ can be used to look for a change in the mean even in the presence of a change of the scale, whereas $S(Y_i)$ is sensible with respect to a change in scale but remains unaffected when the location changes.

The question arises whether or not the indicators $M(Y)_t$ and $S(Y)_t$ are unique up to transformations. The following theorem provides the answer: Any pair of smooth functions of the ch.f. turns out to be a pair of smooth transformations of the indicators $M(Y)_t$ and $S(Y)_t$. This result also justifies our choice of indicators among the class of functionals which depend on the underlying distribution only through the characteristic function.

THEOREM 2.1 *Assume that R_Y and I_Y are differentiable and $R_Y \neq 0$ on some domain $\mathbb{T}$. Let $W(\chi, \eta)$ and $U(\chi, \eta)$, $(\chi, \eta) \in \mathbb{R}^2$, be C^1 transformations such that*
(i) $W(I_Y(t), R_Y(t))$, $t \in \mathbb{T}$, depends only on σ but not on μ and
(ii) $U(I_Y(t), R_Y(t))$, $t \in \mathbb{T}$, depends only on μ but not on σ.
Then C^1 transformations W and U for which (i) and (ii) hold have the following form

$$W(I_Y(t), R_Y(t)) = \Phi(R_Y^2(t) + I_Y^2(t)) \quad \textit{and} \quad U(I_Y(t), R_Y(t)) = \Psi(I_Y(t)/R_Y(t))$$

for all $t \in \mathbb{T}$, where Φ and Ψ are non-constant C^1 functions.

Theorem 2.1 tells us that any transformations of $R_Y(t)$ and $I_Y(t)$, for which properties (i) and (ii) hold, must be univariate functions of our indicators. The proof is based on finding the general solutions of certain partial differential equations for W and V and is given in the appendix.

2.2. *Estimates of the indicators and the decoupling effect*

Given an i.i.d. sample $Y_1, \dots, Y_n$ define for $t \in \mathbb{T}$ the nonparametric estimators

$$\widehat{R}^n(t) = \frac{1}{n}\sum_{j=1}^{n} \cos(t\,Y_j) \qquad \text{and} \qquad \widehat{I}^n(t) = \frac{1}{n}\sum_{j=1}^{n} \sin(t\,Y_j) \tag{2.3}$$

for $R_Y(t)$ and $I_Y(t)$, respectively. For simulation purposes select tentatively $w(t) = 1$ for $|t| \leq a$, $a > 0$ and zero, otherwise. Divide $[-a,\, a]$ into $2\,\ell + 1$ intervals of length $\Delta > 0$. We shall estimate M and S as follows:

$$\widehat{M}^n = \Delta \sum_{l=-\ell}^{\ell} \left[\frac{\widehat{I}^n(l\,\Delta)}{\widehat{R}^n(l\,\Delta)} \right]^2, \quad \widehat{S}^n = \Delta \sum_{l=-\ell}^{\ell} \left[\widehat{R}^n(l\,\Delta)^2 + \widehat{I}^n(l\,\Delta)^2 \right].$$

The following observation is simple but important and worthwhile mentioning, also cf. [48, p. 233]. Recall that a statistic T_n is called a pivot w.r.t. a parameter ϑ, if its distribution does not depend on ϑ.

PROPOSITION 2.1 *If $Y_i = \mu + \sigma X_i$ for all $i = 1, \dots, n$, then*

$$\widehat{R}^n(t)^2 + \widehat{I}^n(t)^2 = n^{-1} + n^{-2} \sum_{i \neq j} \cos(t\,\sigma(X_i - X_j)). \tag{2.4}$$

Therefore, in our setting $\widehat{S}^n$ is a pivot w.r.t. μ.

For the sake of illustration, we have performed the following experiments. Choosing the Gaussian $N(0,\, 0.1)$ distribution as the nominal (in-control) one, for a sample of size $n = 100$ the pair $(\widehat{M}^n,\, \widehat{S}^n)$ was calculated and plotted as one point in Fig. 1. Then, these steps were repeated 1000 times, providing the cluster of points marked in Fig. 1 as "In Control". Similarly, the cluster "Var. Change" corresponds to a $N(0, 0.3)$ distribution and the cluster "Mean Shift" to $N(0.5, 0.1)$, respectively. The largest cluster ("Mean & Var. Change") was obtained for a $N(0.5, 0.3)$, i.e. when both parameters were changed. Fig. 1 reveals that all four clusters can easily be distinguished and separated.

It it also interesting to note that the dispersion for the location indicator when the observations have mean 0 is substantially smaller than the dispersion of the scale indicator, which substantially contributes to the clustering effect. Our asymptotic results provide an explanation: We find that the convergence rate of the location indicator is $1/n$ if the observations have mean $\mu = 0$, whereas it is $1/\sqrt{n}$ otherwise. Contrary, for the scale indicator the convergence rate is $1/\sqrt{n}$ for all $\mu \in \mathbb{R}$ and $\sigma > 0$.

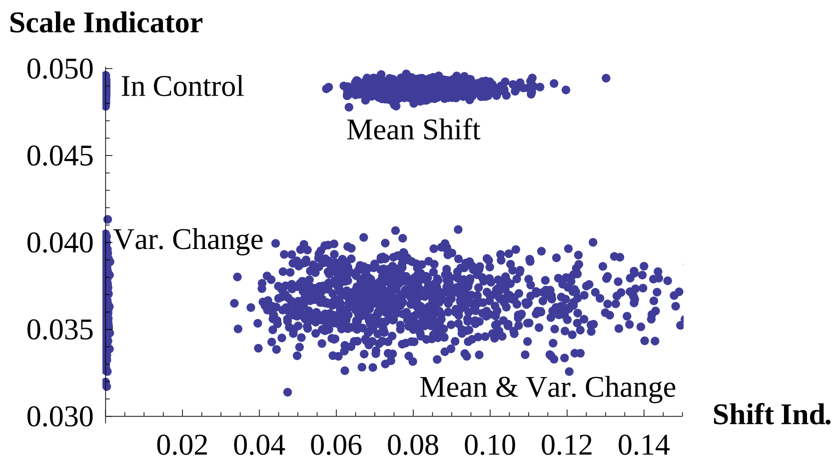

FIGURE 1.— Experimental data to illustrate, firstly, the separability of the sample indicators $\widehat{M}^n$ and $\widehat{S}^n$ to decouple changes in the mean and scale and, secondly, their different convergence rates.

2.3. *Control charts based on sequential window estimates*

The theoretical as well as practical insights of the previous subsections motivate to construct control charts for monitoring (surveillance) based on the above nonparametric estimators for the characteristic function. We shall combine them with a sequential moving window approach, such that only the most recent observations are used by the detectors. This approach generally improves the detection power for late changes. After introducing these sequential estimators, we describe the proposed detection procedures and related computational aspects.

The basic idea to obtain (sequential) estimates of M and S based on the recent data $Y_i, \dots, Y_n$, $i \leq n$, at time n is to plug in the canonical estimates and to combine this with a rolling-window approach where a certain fraction of the most recent observations is used. These sequential estimates are then used to construct the detectors. Thus, in the sequel of the paper we consider the estimators

$$\widehat{R}_i^n(t) = \frac{1}{n-i+1} \sum_{j=i}^{n} \cos(tY_j), \tag{2.5}$$

and

$$\widehat{I}_i^n(t) = \frac{1}{n-i+1}\sum_{j=i}^{n}\sin(tY_j). \tag{2.6}$$

Clearly, $E\widehat{R}_i^n(t) = R_Y(t)$ and $E\widehat{I}_i^n(t) = I_Y(t)$. In addition, put

$$\widehat{C}_i^n(t) = (\widehat{R}_i^n(t), \widehat{I}_i^n(t))', \qquad C_Y(t) = (R_Y(t), I_Y(t))'$$

and notice that $C_Y = C_X = (R_X, I_Y)'$ if there is no change. The plug-in principle now leads us to the estimators

$$M_i^n = \int \left(\frac{\widehat{I}_i^n(t)}{\widehat{R}_i^n(t)} \right)^2 w(t)\,dt, \qquad n \in \mathbb{N}, \tag{2.7}$$

$$S_i^n = \int [\widehat{R}_i^n(t)^2 + \widehat{I}_i^n(t)^2] w(t)\,dt, \qquad n \in \mathbb{N}. \tag{2.8}$$

For S_i^n an explicit formula avoiding integration is known, see [12] and [48, p.233] that eases computations provided the ch. f. of $W = \int g(u)\,du$ is known in closed form.

We are now in a position to define the monitoring procedures in detail. In order to detect a change, one should not use all but only the most recent data. Using the most recent k observations leads to the choice $i = n-k+1$. We assume that the effective sample size k is chosen as a fraction of the maximum sample size, i.e.

$$k = \lfloor N\theta \rfloor, \qquad \text{for some } \theta \in (0,1). \tag{2.9}$$

Now monitoring can be initiated at any time n_0 satisfying $k \le n_0$. We assume

$$n_0 = \lfloor N s_0 \rfloor, \qquad \text{for some } s_0 \in (0,1).$$

This yields the sequence of rolling window control statistics $L_{nk}, n = n_0, n_0+1, \ldots,$ where

$$L_{nk} = k \int \left(\frac{\widehat{I}_{n-k+1}^n(t)}{\widehat{R}_{n-k+1}^n(t)} \right)^2 w(t)\,dt. \tag{2.10}$$

Notice that $E(L_{nk}) \ge 0$. Nevertheless, we expect that the right asymptotic centering term for L_{nk} when the observations are identically distributed with ch.f. given by $C_Y = (R_Y, I_Y)$ is

$$\int \left(\frac{I_Y(t)}{R_Y(t)} \right)^2 w(t)\,dt = \int \tan^2(\mu t) w(t) dt, \tag{2.11}$$

which vanishes under the null hypothesis, since then $C_Y = C_X$ and $\mu = 0$. The factor k appearing in the definition of L_{nk} anticipates the correct convergence rate according to our asymptotic results, which also confirms that we do not have to center L_{nk} at its expectation. The corresponding control chart to detect a change in the location is given by the stopping time

$$T_L = T_{L,N} = \min\{n_0 \le n \le N : L_{nk} > c_L\} \tag{2.12}$$

for some control limit c_L and a time horizon N. Let us agree on the convention $\inf \emptyset = \min \emptyset = \infty$. The event $\{T_L \leq N\}$ is interpreted as a signal at time T_L indicating a change; monitoring stops latest at time instant N when the maximum number of observations (time horizon) is reached. The procedure to detect a change in the scale is based on

$$S_{nk} = \sqrt{k}\left\{\int \left(\widehat{I}^n_{n-k+1}(t)^2 + \widehat{R}^n_{n-k+1}(t)^2\right) w(t)\,dt - \int |\varphi_X(t)|^2 w(t)\,dt\right\}. \tag{2.13}$$

We stop and decide in favor of a change at time

$$T_S = T_{S,N} = \min\{n_0 \leq n \leq N : |S_{nk}| > c_S\} \tag{2.14}$$

for a control limit c_S. In case that the in-control distribution and thus the centering term appearing in (2.13) are unknown, it may be estimated from a learning sample. We discuss that issue below.

Notice that from the definitions it is obvious that up to the scaling factors of the form k^γ, $\gamma > 0$, L_{nk} and S_{nk} are of the general form (1.3).

The control limits of the above detectors should be selected to ensure well-defined (nominal) statistical properties such as a minimal ARL, a_0, e.g.

$$ARL_0(T) = E_0(T) \geq a_0, \qquad T \in \{T_L, T_S\},$$

or a controlled type I error rate,

$$P_0(T \leq N) \leq \alpha, \qquad T \in \{T_L, T_S\},$$

for some given significance level $\alpha \in (0,1)$, when there is no change. Here P_0 and E_0 denote probability and expectation, respectively, when there is no change. For independent data and a known in-control model, one may calculate the control limits by Monte Carlo simulation. In the general case, one may simulate trajectories from the limit processes derived in Section 3 with estimated covariance functions. However, we shall show that subsampling leads to consistent approximations, thus providing us with a feasible and powerful approach to estimate control limits in applications. Frequently, and we shall discuss this issue below in greater detail, the observations $1, \dots, n_0 - 1$ can be used as an in-control learning sample in order to estimate unknown quantities. In particular, we may subsample from the learning sample to estimate the control limit. One may also go beyond this and re-estimate the control limit when running the monitoring procedure by subsampling from all available observations. This sequential resampling approach originates in the work of [44], where it has been developed to design a fast sequential bootstrap scheme for the detection of a change in an AR(1) model.

2.4. *Case: φ_X unknown*

In order to calculate S_{nk}, we need to know the ch.f. $\varphi_X(t)$. If $\varphi_X(t)$ is unknown, we may use the learning sample of size L,

$$Y_1, \dots, Y_L, \qquad L = n_0 - d,$$

which is distributed as $X_1, \dots, X_L$. $d > k$ is used to ensure that dependencies between the learning sample and the sample $Y_{n_0}, Y_{n_0+1}, \dots$ used for monitoring are sufficiently small. For independent data, one may put $d = 1$; for a $MA(m)$ time series, $d = m$ is a reasonable choice. $\varphi_X(t)$ can now be estimated by

$$\widehat{\varphi}_X(t) = \frac{1}{L}\sum_{j=1}^{L} e^{itY_j} = \widehat{R}_1^L(t) + i\widehat{I}_1^L(t), \qquad t \in [T_0, T_1]. \tag{2.15}$$

Applying the control statistic at current time n implies that there was not yet a signal, i.e. the data $Y_1, \dots, Y_n$ are classified as being distributed as the X-process. Then one may also use the past data reduced by the most recent $100 \cdot \varepsilon$ percent, i.e. use the estimator (2.15) with L replaced by $\lfloor n - d \rfloor$, where now $d = N\varepsilon$, for some $\varepsilon > 0$.

Remark 2.1 *We parametrize the relevant quantities (time horizon, size of the learning sample, start of monitoring etc.) by N, the time horizon. The asymptotic results shall rescale the time from $[0, N]$ to the unit interval $[0, 1]$, and their extension to $[0, xN]$ corresponding to $[0, x]$, $0 < x < \infty$ arbitrary, is then straightforward and also provides the asymptotics for the non-negative real line. Equivalently, one can use the size of the learning sample $L = \lfloor N(s_0 - \varepsilon) \rfloor$ as the parameter, such that $N = \varrho_N L \sim \varrho L$ with $\varrho_N = N/L \to \varrho = (s_0 - \varepsilon)^{-1}$.*

3. TIME SERIES MODEL AND PRELIMINARIES ON THE SEQUENTIAL CHARACTERISTIC PROCESS

Our approach to establish asymptotic probabilistic properties for the proposed change-point procedures is to trace them back to a certain basic underlying sequential empirical processes, the sequential characteristic process. After explaining those reduction steps, we introduce and discuss our main assumption, a weak invariance principle for that process. Indeed, it already provides us with limiting distributions. Nevertheless, we shall specialize to a general nonlinear time series model that covers many stochastic models used in present day applications and for which it is known that the invariance principle holds true.

The first reduction step is to observe the representations

$$T_L = \inf\{n_0/N \le u \le 1 : L_N(u) > c_L\},$$
$$T_S = \inf\{n_0/N \le u \le 1 : |S_N(u)| > c_S\},$$

of our control charts, if we define the continuous time control processes

$$L_N(u) = L_N^{(k)}(u) = L_{\lfloor Nu \rfloor, k} \qquad \text{and} \qquad S_N(u) = S_N^{(k)}(u) = S_{\lfloor Nu \rfloor, k}, \qquad u \in [0, 1]. \tag{3.1}$$

By continuity of the inf-type stopping times when considered as functionals on the Skorohod space $D([s_0, 1]; \mathbb{R})$, weak convergence results for the processes L_N and S_N with a.s. continuous limits yield central limit theorems for T_L and T_S. Further, studying the weak limits of L_N and S_N is of interest in its own right as well.

The second reduction is obtained by observing that the probabilistic properties of L_N and S_N are driven by a process mathematically formalizing the rolling window estimators which in turn are governed by a multiparameter empirical process, the *sequential empirical characteristic process*

$$\mathcal{C}_N(s, t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} Z_j(t), \ \text{ with } Z_j(t) = \begin{pmatrix} \cos(tY_j) - R_X(t) \\ \sin(tY_j) - I_X(t) \end{pmatrix}, \qquad s \in [s_0, 1], t \in [T_0, T_1],$$

associated to the functional data set (1.2), which we shall study under the no-change hypothesis such that the expectation of $Z_j(t)$ is 0, since then $(R_Y, I_Y) = (R_X, I_Y)$. The reduction steps are carried out in the proofs provided in detail in the appendix. Indeed, the proofs only require the following weak invariance principle assumed throughout the article. For a discussion of this approach see, e.g., [41] and [45, ch.9].

Assumption (Weak Invariance Principle): The sequential characteristic process $\mathcal{C}_N$ converges weakly to some Gaussian process $\mathcal{C}$ with a.s. continuous paths.

REMARK 3.1 *$\mathcal{C}_N(s,t)$ is the sequential generalization of the classic characteristic process $\mathcal{C}'_N(t) = \mathcal{C}_N(1,t)$ that is particularly well studied for i.i.d. observations. The weak convergence of $\mathcal{C}'_N(t)$ to a sample continuous Gaussian process $\mathcal{C}'(t)$, in the space $C([0,1];\mathbb{R})$ has been studied by [12], [27] and [9]. Introduce $m(y) = \lambda(\{h \in (-1/2, 1/2) : \varphi(h) < y\})$, where $\varphi(h) = (1 - R_Y(h))^{1/2}$ and λ denotes Lebesgue measure, and define the nondecreasing rearrangement of $\varphi(h)$ by $\bar{\varphi}(h) = \sup\{y : m(y) < h\}$. Then a separable version of $\mathcal{C}'(t)$ is a.s. continuous, if and only if*

$$\int \bar{\varphi}(e^{-x^2})\, dx < \infty, \tag{3.2}$$

and, as shown by [27], $\mathcal{C}'_n(t) \Rightarrow \mathcal{C}'(t)$ in $C([-1/2,1/2];\mathbb{R})$, if and only if that condition is satisfied. For a symmetric d.f. $F(x)$ which is concave for $x > x_0 > 0$, a characterization in terms of the tails of F is as follows: A separable version of $\mathcal{C}'$ is a.s. continuous, if and only if

$$\int_{x_0}^{\infty} \frac{(1 - F(x))^{1/2}}{x(\log x)^{1/2}}\, dx < \infty. \tag{3.3}$$

A further sufficient condition is

$$x^\alpha F(-x) + x^\alpha (1 - F(x)) = O(1), \quad \text{as } x \to \infty \text{ for some } \alpha > 0. \tag{3.4}$$

The latter condition is known to be valid for stable laws, cf. the discussion in [49]. Hence, in the i.i.d. case the weak convergence particularly holds true for the large class of stable distributions.

As an important large class of time series to which we want to specialize, let us assume that Y_n is a weakly stationary time series

$$E(Y_n) = 0, \qquad \text{for all } n, \tag{3.5}$$

which is of the form

$$Y_n = f(\epsilon_n, \epsilon_{n-1}, \dots), \qquad n \in \mathbb{N}, \tag{3.6}$$

for some measurable function $f : \mathbb{R}^\infty \to \mathbb{R}$ and an i.i.d. noise process $\{\epsilon_n : n \in \mathbb{Z}\}$, such that $\{Y_n\}$ is also strictly stationary.

Time series models of this type play a prominent role in various scientific areas including financial econometrics, economics, signal processing and image processing. They arise naturally when the data under consideration is preprocessed by nonlinear filters. For example, the well known class of GARCH models appears as a special case. Having in mind signal processing and time series analysis in general, it is worth mentioning that in recent years several nonlinear filters, known under the names *bilateral filters* or *vertically weighted filters*, as studied by [42], [43], [14], [1], [7], and [37], amongst others, have been proposed that are generalizations of the well known sigma filter [24]. Their application is wide spread in image processing and also common in the analysis of univariate time series, although there is only limited rigorous knowledge on their asymptotic properties.

Nonlinear time series of the form (3.6) are known to be S-mixing and for such S-mixing series the validity of the weak invariance principle as formulated above has been shown in [40]. Here a series $\{Z_n\}$ is called S-mixing, if for any $k \in \mathbb{Z}$ and $m \in \mathbb{N}$ one can find a random variable Z_{km} such that

(i) there are sequences $\gamma_m, \delta_m = o(1)$ such that $P(|Z_k - Z_{km}| \geq \gamma_m) \leq \delta_m$ for $k \in \mathbb{Z}, m \in \mathbb{N}$, and

(ii) for all disjoint intervals $I_1, \dots, I_r \subset \mathbb{Z}$, $r \geq 2$, and $m_1, \dots, m_r \in \mathbb{N}$, the random vectors $(Z_{jm_1})_{j \in I_1}, \dots, (Z_{jm_r})_{j \in I_r}$ are independent, if the lag between I_k and I_l is greater than $m_k + m_l$.

By applying the results of [40], we obtain the following basic result for the sequential characteristic process.

THEOREM 3.1 *If $\{Y_t\}$ satisfies (3.6) and (3.5) with i.i.d. errors $\{\epsilon_n : n \in \mathbb{Z}\}$, then*

$$\mathcal{C}_N(s,t) \Rightarrow \mathcal{C}(s,t),$$

as $N \to \infty$, in the space $D([s_0,1] \times \mathbb{T}; \mathbb{R}^2)$, for some Gaussian process with mean zero and covariance structure given by

$$Cov(\mathcal{C}^{\cos}(s_1,t_1), \mathcal{C}^{\sin}(s_2,t_2)) = (s_1 \wedge s_2) \left\{ \gamma_{cs}(0;t_1,t_2) + 2 \sum_{l=1}^{\infty} \gamma_{cs}(l;t_1,t_2) \right\},$$

$$Cov(\mathcal{C}^{\cos}(s_1,t_1), \mathcal{C}^{\cos}(s_2,t_2)) = (s_1 \wedge s_2) \left\{ \gamma_{\cos}(0;t_1,t_2) + 2 \sum_{l=1}^{\infty} \gamma_{\cos}(l;t_1,t_2) \right\},$$

$$Cov(\mathcal{C}^{\sin}(s_1,t_1), \mathcal{C}^{\sin}(s_2,t_2)) = (s_1 \wedge s_2) \left\{ \gamma_{\sin}(0;t_1,t_2) + 2 \sum_{l=1}^{\infty} \gamma_{\sin}(l;t_1,t_2) \right\},$$

$s_1, s_2 \in [s_0,1]$, $t_1, t_2 \in \mathbb{T}$, where

$$\begin{aligned} \gamma_{\sin}(l;t_1,t_2) &= Cov(\sin(t_1 Y_{1+l}), \sin(t_2 Y_1)), \\ \gamma_{\cos}(l;t_1,t_2) &= Cov(\cos(t_1 Y_{1+l}), \cos(t_2 Y_1)), \\ \gamma_{cs}(l;t_1,t_2) &= Cov(\cos(t_1 Y_{1+l}), \sin(t_2 Y_1)), \end{aligned}$$

for $l = 0, 1, \dots$ and $t_1, t_2 \in \mathbb{T}$

REMARK 3.2 *Note that for fixed $t \in \mathbb{T}$ $\mathcal{C}(s,t)$ is a Brownian motion. It has the scaling property*

$$\{\mathcal{C}(\lambda s, t) : s \in \lambda[s_0,1], t \in \mathbb{T}\} \stackrel{d}{=} \{\lambda^{1/2} \mathcal{C}(s,t) : s \in [s_0,1], t \in \mathbb{T}\},$$

and for $s_1, \dots, s_k \in [s_0,1]$ with $s_1 < \dots < s_k$ the increments $\mathcal{C}(s_2,t) - \mathcal{C}(s_1,t), \dots, \mathcal{C}(s_k,t) - \mathcal{C}(s_{k-1},t)$ are independent.

We are now in a position to treat the process corresponding to the rolling window estimates (2.5) and (2.6). We rescale time using the mapping $u \mapsto Nu$, $u \in [0,1]$. Let us define the sequential rolling window process centered at expectations under the null hypothesis,

$$C_N^{(s,\theta)}(t) = \sqrt{\lfloor N\theta \rfloor}\left(\widehat{R}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - R_X(t), \widehat{I}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - I_X(t) \right)',$$

for $0 \le \theta \le s \le 1, t \in [T_0, T_1]$, and $C_N^{(s,\theta)}(t) = 0$ if $\theta > s$. Notice that $n = \lfloor Ns \rfloor$ and $k = \lfloor N\theta \rfloor$ yields $\sqrt{k}[\widehat{C}^n_{n-k+1}(t) - C(t)]$. The following result provides the asymptotics of the sequential rolling windows process $C_N^{(s,\theta)}(t)$ under the null hypothesis of no change.

Corollary 3.1 *We have*

$$C_N^{(s,\theta)}(t) \Rightarrow \mathcal{C}_\theta(s,t), \qquad N \to \infty,$$

in the Skorohod space $D([s_0,1]^3; \mathbb{R}^2)$*, where*

$$\mathcal{C}_\theta(s,t) = \left(\mathcal{C}_\theta^{(1)}(s,t), \mathcal{C}_\theta^{(2)}(s,t) \right)' = \theta^{-1/2}[\mathcal{C}(s,t) - \mathcal{C}(\theta,t)] 1_{\{s<\theta\}}, \qquad s_0 \le s,t \le 1. \tag{3.7}$$

Remark 3.3 *Notice that, for sake of generality, in Corollary 3.1 the process* $C_N^{(s,\theta)}(t)$ *is studied as a process defined for* (s,t,θ)*. However, according to Assumption (2.9)* θ *is fixed in our treatment. Thus, in what follows, that process is studied as a process attaining values in* $D([s_0,1]^2; \mathbb{R}^2)$*.*

4. LIMIT THEORY WHEN THERE IS NO CHANGE

The asymptotic properties of the detectors T_L and T_S when there is no change can now be investigated using the general results of the previous section. First we show that the proposed location detector exhibits a qualitatively different type of asymptotics than the scale detector and, in particular, attains a different convergence rate. The analysis is based on a functional Taylor expansion which shows that, in general, the asymptotics depends on the ψ-function and its analytic properties. The ψ-function used for the location detector has a vanishing first order derivative under the no-change hypothesis such that the linear term of the expansion vanishes, which is in contrast to the analytic properties of the ψ function used for the scale detector.

4.1. *Change-point asymptotics - Case I: φ_X known*

Notice that, as indicated in Section 2, the detectors T_L and T_S are both of the general form

$$T_{Nk}(\psi) = \min\{n_0 \le n \le N : |L_{nk}(\psi)| > c_\psi\}, \qquad c_\psi \text{ a constant},$$

where the control statistic is given by

$$L_{nk}(\psi) = k^\gamma \left\{ \int \psi(\widehat{C}^n_{n-k+1}(t)) w(t)\, dt - \int \psi(C(t)) w(t)\, dt \right\}, \qquad 1 \le k \le n, n \in \mathbb{N},$$

for some smooth function $\psi : \mathbb{R}^2 \to \mathbb{R}$ and a rate parameter $\gamma > 0$. Here $C = C_X$ is known to us under the no-change hypothesis. To $L_{Nk}(\psi)$, $1 \le k \le N$, we associate the continuous time process $\mathcal{L}_N(\psi)$ defined by

$$\mathcal{L}_N(\psi)(s) = L_{\lfloor Ns \rfloor, \lfloor N\theta \rfloor}(\psi), \qquad s \in [s_0,1]. \tag{4.1}$$

cf. (2.9), which links N and k via the parameter θ.

Let us denote the first derivative (or gradient) of a function f by $\dot{f}$ and the second derivative (or Hessian) by $\ddot{f}$. As we shall see below, the case

$$\dot{\psi}(C(\cdot))\big|_{[T_0,T_1]\cap W} = 0, \qquad wd\lambda - a.e.,$$

where W denotes the support of the weighting function, requires a special treatment and leads to another type of asymptotics. But let us first discuss ψ-functions where this is not the case. As we shall show now, the proposed procedure to detect a change in the scale belongs to these ψ-functions.

Recall the definition of S_{nk} in (2.13) and the definition of S_N in (3.1) which, when combined with (4.1), give rise to the representations

$$S_{nk}(\psi) = L_{nk}(\psi_S), \qquad \text{and} \qquad S_N = \mathcal{L}_N(\psi_S),$$

if we select the ψ-function

$$\psi_S(z) = z_1^2 + z_2^2, \tag{4.2}$$

for $z = (z_1, z_2)' \in \mathbb{R}^2$, and $\gamma = 1/2$. Clearly, $\dot{\psi}_S(z) = 2z$ and $\ddot{\psi}_S(z) = 2I$, I denoting the identity matrix of dimension 2, such that $\dot{\psi}_S(C)(t) = 2C(t) \neq 0$ for λ-almost all $t \in [T_0, T_1]$. The general result for such cases is as follows.

THEOREM 4.1 *Let ψ be continuously differentiable with bounded derivative such that $\dot{\psi}(C) \neq 0$ except on a $wd\lambda$-null set and put $\gamma = 1/2$. Under the null hypothesis we have in the Skorohod space $D([s_0, 1]; \mathbb{R})$*

$$\mathcal{L}_N(\psi)(s) \Rightarrow \mathcal{L}_1(\psi)(s), \qquad N \to \infty,$$

for the a.s. continuous process

$$\mathcal{L}_1(\psi)(s) = \int \dot{\psi}(C(t))'\mathcal{C}_\theta(s,t)w(t)\,dt, \qquad s \in [s_0, 1],$$

provided the random element on the right side of the above display is not concentrated in 0.

Due to the fact that $\mathcal{C}_\theta(s,t)$ and $\dot{\psi}(C(t))'\mathcal{C}_\theta(s,t)$ are Gaussian processes and since integrals over Gaussian processes have normal laws, we obtain the following sufficient condition for the non-degeneracy of $\mathcal{L}_1(\psi)(s)$. Let

$$K_s(t_1, t_2) = \operatorname{Cov}\left(\dot{\psi}(C(t_1))'\mathcal{C}_\theta(s, t_1), \dot{\psi}(C(t_2))'\mathcal{C}_\theta(s, t_2)\right), \qquad t_1, t_2 \in [s_0, s],$$

for fixed $s \in (s_0, 1]$. Suppose that

$$\sigma_s^2 = \int_{s_0}^{s} \int_{s_0}^{s} K_s(t_1, t_2) w(t_1) w(t_2)\, dt_1 dt_2 \in (0, \infty).$$

Then, cf. [39],

$$\mathcal{L}_1(\psi)(s) \sim N(0, \sigma_s^2)$$

such that $\mathcal{L}_1(\psi)(s)$ cannot be concentrated in 0.

The asymptotics of the scale-detector now appears as a collorary.

Corollary 4.1 (The Change-in-Scale Detector)
Under the above assumptions, we have

$$S_N(s) \Rightarrow \mathcal{L}_S^*(s) = 2\int [R_X(t)\mathcal{C}_\theta^{(1)}(s,t) + I_X(t)\mathcal{C}_\theta^{(2)}(s,t)]w(t)\,dt,$$

where $\mathcal{C}_\theta = (\mathcal{C}_\theta^{(1)}, \mathcal{C}_\theta^{(2)})'$ *has been defined in (3.7), and therefore*

$$T_S/N \xrightarrow{d} \inf\{s_0 \le s \le 1 : |\mathcal{L}_S^*(s)| > c_S\},$$

as $N \to \infty$.

Let us now consider the proposed method to detect a change in the location, where we have to put $\gamma = 1$ and obtain

$$L_N = \mathcal{L}_N(\psi_L), \quad \text{if} \quad \psi_L(z) = (z_2/z_1)^2,$$

for $z = (z_1, z_2)' \in \mathbb{R}^2$ with $z_1 \neq 0$. Since

$$\dot\psi_L(z) = 2\psi_L(z)(-z_2 z_1^{-2}, z_1^{-1}) \quad \text{and} \quad \psi_L(C_Y(t)) = (\tan(\mu t))^2, \quad \text{if } Y_1 \sim \mu + \sigma X, \tag{4.3}$$

we obtain $\dot\psi_L(C(t)) = 0$ under the null hypothesis $\mu = 0$. The general result for such cases is as follows.

Theorem 4.2 *Let* ψ *be three times continuously differentiable with bounded derivatives such that* $\dot\psi(C(\cdot)) = 0$ *except on a* $w d\lambda$*-null set. Put* $\gamma = 1$. *Under the null hypothesis of no-change*

$$\mathcal{L}_N(\psi)(s) \Rightarrow \mathcal{L}_2(\psi)(s) = \frac{1}{2}\int \mathcal{C}_\theta(s,t)'\ddot\psi(C(t))\mathcal{C}_\theta(s,t)w(t)\,dt,$$

as $N \to \infty$, *in* $D([s_0,1];\mathbb{R})$, *provided* $\mathcal{L}_2(\psi)$ *is not concentrated in* 0.

Again, we have the following sufficient criterion for non-degeneracy of $\mathcal{L}_2(\psi)$. For $s \in (s_0, 1]$ let

$$G_s(t_1,t_2) = \operatorname{Cov}\left(\mathcal{C}_\theta(s,t_1)'\ddot\psi(C(t_1))\mathcal{C}_\theta(s,t_1), \mathcal{C}_\theta(s,t_2)'\ddot\psi(C(t_2))\mathcal{C}_\theta(s,t_2)\right),$$

for $t_1, t_2 \in [s_0, 1]$. If

$$\int_{s_0}^{s}\int_{s_0}^{s} K_s(t_1,t_2)w(t_1)w(t_2)\,dt_1 dt_2 \in (0,\infty),$$

then $\mathcal{L}_2(\psi)(s)$ follows a normal law and is therefore not concentrated in 0.

Let us compute the limit process for the change-in-location detector $L_N = \mathcal{L}_N(\psi_L)$. Since

$$\ddot\psi_L(C(t)) = \begin{bmatrix} 0 & 0 \\ 0 & 2/R_X(t) \end{bmatrix},$$

we see that the limiting process of L_N depends on $\mathcal{C}_\theta(s,t)$ only through the second coordinate $\mathcal{C}_\theta^{(2)}(s,t)$. The main result for the change-in-location detector is now as follows.

COROLLARY 4.2 (THE CHANGE-IN-LOCATION DETECTOR)
Under the null hypothesis of no change we have

$$L_N(s) \Rightarrow \mathcal{L}_L^*(s) = \int \left(\frac{C_\theta^{(2)}(s,t)}{R_X(t)} \right)^2 w(t)\, dt,$$

in $D([s_0,1];\mathbb{R})$ *and therefore*

$$T_L/N \xrightarrow{d} \inf\{s_0 \le s \le 1 : \mathcal{L}_L^*(s) > c_L\},$$

as $N \to \infty$.

4.2. *Change-point asymptotics - Case II:* φ_X *unknown*

Whereas for the location detector the centering term vanishes, cf. (2.11), Theorem 4.2 and Corollary 4.2, we need to know the in-control ch.f. φ_X in order to calculate the scale detector, which can be restrictive for certain applications. Then it is natural to estimate the ch.f. from the learning sample, cf. Section 2. For asymptotics, we assume that the learning sample is

$$X_1, \dots, X_L, \qquad \text{with } L = L_\varepsilon = \lfloor N(s_0 - \varepsilon) \rfloor,$$

for some $0 < \varepsilon < s_0$, which serves to ensure a sufficient gap between the learning sample and the sample used for monitoring such that dependencies die out. The case that $L = L_s = \lfloor N(s-\varepsilon) \rfloor$ is treated analogously and therefore omitted.

We wish to replace the centering term $\int \psi(C(t))w(t)\,dt$ appearing in $\mathcal{L}_N(\psi)$ by its estimate $\int \psi(\widehat{C}_1^L(t))w(t)\,dt$ based on the learning sample. Let us describe the approach for the case $\dot{\varphi}(C) \neq 0$; a vanishing first order derivative can be handled analogously. Then the above ideas give rise to the definitions

$$\widetilde{\mathcal{L}}_N(\psi)(s) = \lfloor N\theta \rfloor^{1/2} \left\{ \int \psi(\widehat{C}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}^{\lfloor Ns \rfloor}(t))w(t)\,dt - \int \psi(\widehat{C}_1^L(t))w(t)\,dt \right\},$$

for $s \in [s_0, 1]$, cf. (2.10) and (4.1), and (for fixed θ)

$$\mathcal{C}_N^{(0)}(t) = \sqrt{\lfloor N\theta \rfloor}[\widehat{C}_1^L(t) - C_X(t)], \qquad t \in [T_0, T_1].$$

It is clear that

$$\mathcal{C}_N^{(0)}(t) \Rightarrow \mathcal{C}_\theta^{(0)}(t) = \theta^{-1/2}\mathcal{C}_\theta(1,t),$$

as $N \to \infty$, in $D([T_0,T_1];\mathbb{R}^2)$.

Our assumptions on the nonlinear time series $\{Y_n\}$ are focused on time series defined via innovation series $\{\epsilon_n\}$ and do not automatically imply a mixing property. However, if $\{Y_n\}$ is also α-mixing, which is often the case under certain regularity conditions, the process $\mathcal{C}_N^{(0)}(t)$, corresponding to the additional estimation step using the learning sample, converges jointly with the processes $\mathcal{C}_N^{(0)}(t)$ and $\mathcal{C}_N^{(s,\theta)}(t)$ treated in Theorem 3.1 and Corollary 3.1.

To proceed recall that a process $\{Z_t\}$ is called α-mixing, if

$$\sup_t \sup_{A \in \mathcal{F}^t, B \in \mathcal{F}_{t+k}} |P(A \cap B) - P(A)P(B)| \to 0,$$

as $k \to \infty$, where $\mathcal{F}^t = \sigma(Z_s : s \le t)$ and $\mathcal{F}_{t+k} = \sigma(Z_s : s \ge t+k)$.

LEMMA 4.1 *Assume that, in addition to the conditions of Theorem 3.1, the process $\{Y_n\}$ is α-mixing. Then we have for $N \to \infty$,*

(i) $(\mathcal{C}_N^{(0)}, \mathcal{C}_N) \Rightarrow (\mathcal{C}_\theta^{(0)}, \mathcal{C})$, *in* $D([T_0, T_1]; \mathbb{R}^2) \times D([s_0, 1] \times [T_0, T_1]; \mathbb{R}^2)$,

(ii) $(\mathcal{C}_N^{(0)}, \mathcal{C}_N^{(\cdot,\theta)}) \Rightarrow (\mathcal{C}_\theta^{(0)}, \mathcal{C}_\theta)$, *in* $D([T_0, T_1]; \mathbb{R}^2) \times D([s_0, 1] \times [T_0, T_1]; \mathbb{R}^2)$.

Given the previous results, Lemma 4.1 allows us to derive easily the asymptotics of $\widetilde{\mathcal{L}}_N$.

THEOREM 4.3 *Assume that, in addition to the conditions of Theorem 3.1, the process $\{Y_n\}$ is α-mixing. Then*

$$\widetilde{\mathcal{L}}_N(\psi)(s) \Rightarrow \widetilde{\mathcal{L}}(\psi)(s) = \int \dot{\psi}(C(t))'[\mathcal{C}_\theta(s,t) + \mathcal{C}_\theta^{(0)}(t)]w(t)\,dt,$$

in the Skorohod space $D([s_0, 1]; \mathbb{R})$, as $N \to \infty$. The corresponding control chart,

$$T_{\widetilde{L}} = \inf\{n_0 \le n \le N : \widetilde{\mathcal{L}}_N(\psi)(n/N) > c_{\widetilde{L}}\},$$

satisfies

$$T_{\widetilde{L}}/N \xrightarrow{d} \inf\{s_0 \le s \le 1 : \widetilde{\mathcal{L}}(\psi)(s) > c_{\widetilde{L}}\},$$

as $N \to \infty$.

5. SUBSAMPLING APPROXIMATIONS FOR DETECTORS

The statistical design of a monitoring procedure that gives a signal when a control statistic exceeds a control limit, e.g. in order to attain a given (nominal) significance level $\alpha \in (0,1)$, requires to select the control limit to meet such a criterion. In general, the control limit then depends on the (asymptotic) distribution of the associated (normed) stopping time r_N, in our case

$$r_N \in \{T_L/N, T_{\widetilde{L}}/N, T_S/N\}.$$

Since the limiting distributions are rather involved for the proposed detectors, we propose to use subsampling in order to obtain approximations of the required distributions. Subsampling is a resampling technique developed for time series. It is easy to apply, usually works under much weaker assumptions than other resampling techniques such as the block bootstrap, and does not require case-by-case analyses as block bootstrapping. The ideas behind resampling techniques such as subsampling and bootstrapping can be traced back to the works of [36] and [47]. For an extensive exposition see [30].

Actually, we may even subsample the whole trajectories of our control statistics, but we shall start our exposition with the real-valued stopping times.

In what follows, let us indicate the dependence of r_N on the random variables $Y_1, \dots, Y_N$ by writing $r_N(\mathcal{Y}_N)$ with $\mathcal{Y}_N = \{Y_1, \dots, Y_N\}$. The basic idea of subsampling is to construct appropriate replicates of the control statistic and stopping time, respectively, by calculating those quantities from subseries of length $b < L$ from the learning sample of size L. We describe the subsampling scheme for the more involved case $r_N = T_{\widetilde{L}}/N$, the required modifications for the other procedures are then straightforward. First, construct $L - b + 1$ subsamples

$$\mathcal{Y}_{\ell b} = (Y_\ell, \dots, Y_{\ell+b-1}), \qquad \ell = 1, \dots, L - b + 1,$$

of length $b = b_N \le L$ and calculate the corresponding subsampled replicates of $\widetilde{\mathcal{L}}_N(\psi)(s)$,

$$\widetilde{\mathcal{L}}_{b\ell}(\psi)(s) = \lfloor b\theta \rfloor^{\gamma} \left\{ \int \psi(\widehat{C}(\mathcal{Y}_{\ell b})_{\ell \vee (\ell + \lfloor bs \rfloor - \lfloor b\theta \rfloor + 1)}^{\ell + \lfloor bs \rfloor}(t)) w(t)\, dt - \int \psi(\widehat{C}(\mathcal{Y}_{\ell b})_{\ell}^{\ell + \lfloor bs_0 \rfloor - 1}(t)) w(t)\, dt \right\},$$

where $\widehat{C}(\mathcal{Y}_{\ell b})_u^v = (\widehat{R}(\mathcal{Y}_{\ell b})_u^v, \widehat{I}(\mathcal{Y}_{\ell b})_u^v)$ with

$$\widehat{R}(\mathcal{Y}_{\ell b})_u^v = \frac{1}{v-u+1} \sum_{j=u}^{v} \cos(tY_j) \qquad \text{and} \qquad \widehat{I}(\mathcal{Y}_{\ell b})_u^v = \frac{1}{v-u+1} \sum_{j=u}^{v} \sin(tY_j),$$

for $\ell = 1, \dots, L-b+1$. Next, one calculates the associated replicates of the stopping times,

$$T^*_{\widetilde{L}\ell}(\mathcal{Y}_{\ell b}) = \inf\{\lfloor bs_0 \rfloor \le n \le b : \widetilde{\mathcal{L}}^*_{b\ell}(\psi)(n/b) > c_{\widetilde{L}}\},$$

leading to the subsampled replicates

$$r_{b\ell} = T^*_{\widetilde{L}\ell}(\mathcal{Y}_{\ell b})/b, \qquad \ell = 1, \dots, L-b+1.$$

Given those replicates, the unknown d.f. F_{r_N} of r_N is then estimated by the e.d.f. of the replicates,

$$\widehat{F}_{r_N}(x) = \frac{1}{L-b+1} \sum_{\ell=1}^{L-b+1} 1(r_{b\ell} \le x), \qquad x \in \mathbb{R}.$$

More generally, the unknown law $P_{\widetilde{\mathcal{L}}_N(\psi)}$ of the trajectory $s \mapsto \widetilde{\mathcal{L}}_N(\psi)(s)$ is estimated by the empirical measure

$$\widehat{P}_{\widetilde{\mathcal{L}}_N(\psi)} = \frac{1}{L-b+1} \sum_{\ell=1}^{L-b+1} \delta_{\widetilde{\mathcal{L}}_{b\ell}(\psi)},$$

where δ_x denotes the Dirac measure in a point x.

An application of the general subsampling theorem of [30] now provides us with the following subsampling central limit theorems that show the weak consistency of the subsampling approximations, since our results ensure that

$$r_N \xrightarrow{d} r,$$

as $N \to \infty$, for some non-degenerated random variable r, as well as

$$\widetilde{\mathcal{L}}_N(\psi) \Rightarrow \widetilde{\mathcal{L}}(\psi),$$

as $N \to \infty$. To formulate the result, recall that the bounded Lipschitz metric $d_{BL}(P, Q)$ for probability laws P and Q defined on a metric space S endowed with a σ-field $\mathcal{A}$ is defined as $d_{BL}(P, Q) = \sup\{|\int f dP - \int f dQ|\}$, where the supremum is taken over all measurable $\mathcal{A}$-measurable functions f with $|f(x) - f(y)| \le d(x, y)$ and $\sup_{x \in S} |f(x)| \le 1$. For the case of subsampling the trajectories, we take the Skorohod space $D([s_0, 1]; \mathbb{R})$. Convergence of a sequence $\{P_n, P\}$ of probability measures in the bounded Lipschitz metric implies weak convergence, if the limit P is concentrated on a separable subset, i.e. on the set $C([s_0, 1]; \mathbb{R})$ of continuous functions in the case of $D([s_0, 1]; \mathbb{R})$.

THEOREM 5.1 *If the learning sample* $\{Y_1, \dots, Y_L\}$, $L = \lfloor N(s_0 - \varepsilon) \rfloor$, *satisfies* (3.6) *and* (3.5) *with i.i.d. errors* $\{\epsilon_n : n \in \mathbb{Z}\}$ *and is* α*-mixing, then*

$$d_{BL}(\widehat{F}_{r_N}, F_r) \to 0,$$

as $N \to \infty$, *in probability, and*

$$d_{BL}(\widehat{P}_{\widetilde{\mathcal{L}}_N(\psi)}, P_{\widetilde{\mathcal{L}}(\psi)}) \to 0,$$

as $N \to \infty$, *in probability, where* d_{BL} *denotes the bounded Lipschitz metric, provided*

$$\frac{b}{N} \to 0, \qquad b \to \infty,$$

as $N \to \infty$.

The above result can be used to design the proposed monitoring procedures. In particular, by calculating the $(1-\alpha)$-quantile of the subsampled maxima of the trajectories we obtain an estimate for the control limit corresponding to a monitoring procedure with a nominal type I error equal to α.

REMARK 5.1 *If* $L - b + 1$ *is very large, one may speed up computations by drawing (with replacement)* B *times from the set of all subsamples* $\{\mathcal{Y}_{\ell b} : \ell = 1, \dots, L - b + 1\}$ *and use the e.d.f. formed of the corresponding replicates of* r_N, *which is very close to block bootstrap approach, where, however, one usually chooses the block lengths relatively small and constructs a longer time series by putting the blocks side by side.*

6. ASYMPTOTICS UNDER ALTERNATIVES

This section is devoted to a detailed study of the asymptotic probabilistic behavior of the procedures under a sequence of local alternatives. Thus, let us modify the mathematical framework and consider a sequence of local alternative models where location (drift) and scale approach 0 and 1, respectively, at certain rates. This means, as the maximum sample size increases, it gets harder for the method to detect the change. When using the correct rate, one obtains non-trivial limits which provide insights into the performance properties in large samples. For brevity of presentation, we confine ourselves to ψ-functions such that $\int \dot{\psi}(C(t))' \mathcal{C}_\theta(s,t) w(t) \, dt$ is a non-degenerated random variable. As shown in the previous section, this holds true for the ψ-function corresponding to the location detector.

To proceed suppose we are given a time series $\{X_n\}$ and an array of random variables $\{Y_{Ni} : 1 \le i \le N, N \in \mathbb{N}\}$ defined on a common measurable space $(\Omega, \mathcal{F})$ equipped with a sequence of probability measures $\{P_N\}$, such that

(i) $\{X_n\}$ is a stationary series with $X_n \stackrel{d}{=} -X_n$ for all n under all P_N . Further, the results of Section 3 hold true when the series $\{Y_n\}$ is replaced by $\{X_n\}$, and,

(ii) for each N, under P_N,

$$\{Y_{Nt} : t = 1, \dots, q_N - 1\} \stackrel{d}{=} \{X_1, \dots, X_{q-1}\} \tag{6.1}$$

whereas

$$\{Y_{Nt} : t = q_N, q_N + 1, \dots\} \stackrel{d}{=} \{\mu_N + \sigma_N X_t : t = q_N, q_N + 1, \dots\}, \tag{6.2}$$

with

$$\mu_N = \Delta_\mu N^{-1/2}, \qquad \sigma_N = 1 + \Delta_\sigma N^{-1/2}, \tag{6.3}$$

for two constants Δ_μ and Δ_σ such that $(\Delta_\mu, \Delta_\sigma) \neq (0,1)$. The change-point q_N is given by

$$q = q_N = \lfloor N\tau \rfloor$$

for some $\tau \in (0,1)$.

THEOREM 6.1 *Suppose (6.1)-(6.3) and $E|X|^2 < \infty$, such that φ_X is of the class $C^2(A)$ with $\sup_{t\in A} |\varphi_X''(t)| < \infty$ for any compact set A. Additionally, assume that $\mathbb{T}$ is bounded. Then*

$$\mathcal{C}_N(s,t) \Rightarrow \mathcal{C}_\theta(s,t) + c(s,t), \qquad N \to \infty,$$

under the sequence $\{P_N\}$ of probability measures, where the process $\mathcal{C}_\theta(s,t)$ is as in Corollary 3.1 and

$$c(s,t) = 1_{\{s>\tau\}}(s-\tau)(\varphi_X'(t)t\Delta_\sigma - \Delta_\mu^2 t^2 \varphi_X(t)/2, \Delta_\mu \varphi_X(t))'.$$

Further, under $\{P_N\}$ we have

$$C_N^{(s,\theta)}(t) \Rightarrow \mathcal{C}_\theta^c(s,t) = \theta^{-1/2}[\mathcal{C}_\theta(s,t) - \mathcal{C}_\theta(\theta,t) + c(s,t) - c(\theta,t)],$$

as $N \to \infty$.

The above theorem shows that, asymptotically, the local alternative model affects the process $\mathcal{C}_\theta(s,t)$ by an additive drift which depends on s.

To discuss the resulting asymptotics of a detector based on $\mathcal{L}_N(\psi)$ for a general ψ-function such that $\dot\psi(C) \neq 0$ except on a $wd\lambda$-null set, recall that the control statistic process is then given by

$$\mathcal{L}_N(\psi)(s) = \lfloor N\theta \rfloor^{1/2} \left\{ \int \psi(\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t)) w(t)\,dt - \int w(C_X(t)) w(t)\,dt \right\},$$

where $C_X = (R_X, I_Y)'$ is assumed to be known and $\widehat{C}_i^n = (\widehat{R}_i^n, \widehat{I}_i^n)$ with

$$\widehat{R}_i^n(t) = \frac{1}{n-i+1} \sum_{j=i}^n \cos(tY_{Nj}), \quad \widehat{I}_i^n(t) = \frac{1}{n-i+1} \sum_{j=i}^n \sin(tY_{Nj}), \quad i \le n.$$

THEOREM 6.2 *Suppose the assumptions of Theorem 6.1 hold. We have under the sequence $\{P_N\}$*

$$\mathcal{L}_N'(\psi)(s) \Rightarrow \int \dot\psi(C(t))' \mathcal{C}_\theta^c(s,t) w(t)\,dt, \qquad N \to \infty,$$

in the Skorohod space $D([0,1];\mathbb{R})$, as $N \to \infty$.

7. SIMULATIONS

We investigated the control charts proposed in Section 2, in order to figure out their merits and, in particular, to clarify to which extent the decoupling property applies in practice. One can not expect that empirical means will behave exactly in the same way as the theoretical expectations that appear in the definitions of our indicators. Thus, in practice, a certain impact of changes in the mean on the variance and vice versa is unavoidable, because to some extent random fluctuations must spoil the theoretical decoupling property. For the simulations addressing the cases of known and unknown in-control distribution, the control limit was obtained by simulation and chosen to control the ARL, in order to compare the results with other studies. Our investigation of the accuracy of the subsampling approximation looks at the type I error rate.

7.1. *Simulations when the in-control distribution is known*

We were interested in analyzing the statistical properties of the proposed procedures under the Gaussian benchmark model using simulated control limits, in order to obtain valid results on their performance as well as comparisons with existing methods. We study an early change model where the shift in the mean (jump) is located at time zero. To ensure that the impact of N is negligible, we put $N = 10000$. 64 pre-run observations were used to fill the buffer of the detectors. All results reported below are obtained by means of 10000 simulation runs.

The thresholds c_L and c_S of our charts were selected by solving the nonlinear equation $ARL = 435$ numerically. The control limit for the scale detector, T_S, where ϕ_X is unknown was obtained in the same way.

The results of the simulation studies for detecting a shift in the mean can be summarized as follows (see Tab. I).

- The shift indicator behind our location chart T_L provides much shorter out-of-control ARLs than that of a CUSUM chart for small shifts in the mean, i.e. $0.1\,\sigma$, $0.25\,\sigma$. Also the dispersions of the run lengths (RLs) of the shift indicator of T_L chart are smaller than those of CUSUM chart [15].
- For a shift $0.5\,\sigma$ the ARLs of both charts are comparable although the dispersions of RLs are still smaller for the shift indicator of the T_L chart. For larger shifts, starting from $0.75\,\sigma$, the CUSUM chart outperforms the T_L chart, providing shorter ARLs. We have observed a similar behavior in an essentially different chart proposed in [38].
- The above statements hold as well when the shift indicator of the T_L chart is compared with other charts such as the optimal EWMA, Shewhart-EWMA, GEWMA and GLR charts, which were studied and compared by simulations in [15].

REMARK 7.1 *Notice that the T_L chart reacts faster than the CUSUM chart, which is a bit surprising in view of its optimality properties, cf. [26], [28]. A possible explanation is that our T_L chart uses observations following the in-control distribution right from the start. Thus, our chart is favoured by its construction, which is reflected in its behavior for small shifts.*

The aim of the second series of simulation studies was to verify to which extent the shift in the mean indicator used in the location chart, T_L, is insensitive to changes in the variance of the observations when a shift in the mean is present. Heteroskedasticity in the sense of a temporarily increasing and decreasing variance is present in many real data such as financial returns. To this end, the variance was changed according to the schedules shown in Fig. 2 and

T_L Chart			CUSUM		
Jump	ARL	RL Disp,	Jump	ARL	RL Disp,
0	432.02	431.21	0	434	436
0.1	244.08	244.39	0.1	326	323
0.25	79.16	62.14	0.25	132	123
0.5	35.6	16.23	0.5	37.2	30.4
0.75	24.07	10.71	0.75	16.7	10.8
1	18.79	8.31	1	10.3	5.45

TABLE I

COMPARISON OF ARLS OF THE TWO CONTROL CHARTS WITH ARL EQUAL TO 435 AND INDEPENDENT SIMULATION TRIALS WITH $N(0,1)$ ERRORS. LEFT PANEL: T_L CHART FOR A CHANGE IN THE MEAN USING A BUFFER LENGTH=64. RIGHT PANEL: CUSUM CHART (DATA FROM [15]).

T_L Chart			T_L chart plus SC1		T_L chart plus SC2	
Jump	ARL	RL Disp,	ARL	RL Disp,	ARL	RL Disp,
0	432.02	431.21	437.77	435.23	416.8	412
0.1	244.08	244.39	238.24	238.1	247.2	240.7
0.25	79.16	62.14	78.1	61.75	78.8	62.4
0.5	35.6	16.23	33.4	16.7	33.3	16.2
0.75	24.07	10.71	23.07	10.91	23.7	10.6
1	18.79	8.31	18.1	8.1	18.7	8.25

TABLE II

LEFT PANEL: ARLS OF THE T_L CHART REPRODUCED FROM TAB. I. MIDDLE AND RIGHT PANEL: ARLS OF THE T_L CHART FOR A VARIANCE VARYING ACCORDING TO SCENARIOS SC1 AND SC2 (FIG. 2), RESPECTIVELY.

all the rest of the conditions were unchanged. The results are provided in the middle and the right panel of Tab. II. The left panel of this table is repeated from Tab. I.

As one can notice, the location detector, T_L, does not change its ARLs when fluctuations of the variance are present, even when they are as large as here, i.e. ranging from 0.05 to 1.95 (cf. schedule scheme SC2 in Figure 2). However, when changes of the variance are large and permanent (i.e. not fluctuating), then the ARLs of the T_L chart are still relatively insensitive to them. However, permanent changes in the variance reduce ARLs, which is likely due to the fact that the chart is not based on a pivot statistic.

The aim of the simulations summarized in Table III was to determine ARLs to detect

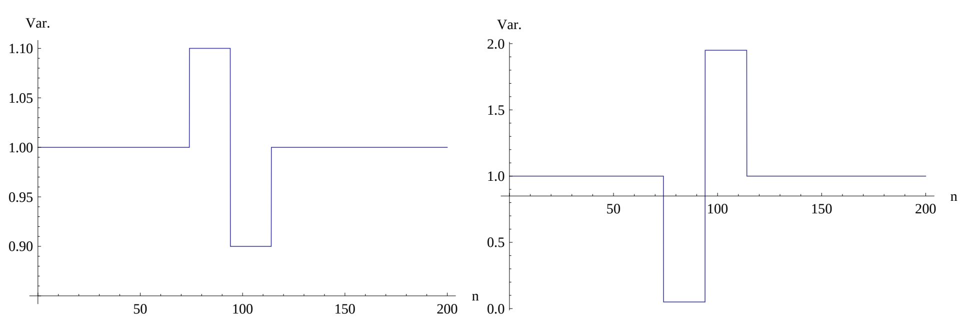

FIGURE 2.— The schedule of changing variance when shift in the mean is to be detected: small changes (SC1) – left panel, large changes (SC2) – right panel.

T_S chart – nominal			T_S chart plus shift		
Scale	ARL	RL Disp,	Scale	ARL	RL Disp,
1	436.4	438.1	1	411.5	415.4
1.25	54.8	32.1	1.25	55.25	28.1
1.5	49.5	18.7	1.5	47.5	18.5
1.75	42.9	14.5	1.75	40.8	1 3.7
2	39.5	13.6	2	38.4	13.1

TABLE III

LEFT PANEL: ARLS OF THE T_S CHART (WITH SCALE=1, I.E. IN-CONTROL STATE). RIGHT PANEL: ARLS OF THE T_S CHART FOR A PERMANENT SHIFT IN THE MEAN OF HEIGHT $0.8\,\sigma$ OCCURRING SIMULTANEOUSLY WITH SCALE CHANGES.

Chart	Prob.	Remarks
T_L	0.051	
T_L	0.054	scale SC2
T_S	0.053	
T_S	0.055	mean shift =1.6

TABLE IV

ESTIMATED PROBABILITIES OF FALSE ALARM IN EACH TIME INSTANT FOR T_L AND T_S CHARTS. THE SECOND ROW PROVIDES THE SAME PROBABILITIES WHEN ADDITIONAL SCALE CHANGES ACCORDING TO SC2 ARE PRESENT, THE FOURTH ROW THE PROBABILITIES OF A FALSE ALARM GIVEN A PERMANENT SHIFT IN THE MEAN.

changes in the scale (left table) when using the scale detector, T_S, assuming $\sigma = 1$ for the in-control process. Additionally, the r.h.s. subtable of Table III provides ARLs that were obtained when, simultaneously with the scale change, there was also a shift in the mean. Again, as expected from the theoretical insights, the scale indicator of the ch.f. chart, which is based on a pivot statistic, is to a large extent insensitive to shifts in the mean when changes in the scale are detected.

Similar conclusions can be drawn from estimated false alarm probabilities that are shown in Tables IV and V. Those probabilities of false alarms for both charts increase by about 0.02 when additional changes in the scale (in the mean, respectively) are present (see Table IV). In Table V the estimated probabilities of detecting changes in the location (in the scale, respectively) with a delay of at least 30 observations are given. One can notice that the presence of of additional changes in the scale (in the location, respectively) increases the corresponding probabilities only slightly, except when small changes are to be detected. But even then the increase of the ARL is small (around 0.04).

We do not provide comparisons of the scale indicator ARL's with the well known R or S charts, since they can not be run on individual observations.

7.2. *Simulations for unknown in-control distribution*

In the simulation experiments reported above we have assumed that the reference point $\int |\phi_X(t)|^2\, w(t)\, dt$ for the T_S chart is known. This is frequently the case when runs of a production process are well documented. However, it may happen that this reference point is unknown and has to be estimated from a relatively small learning sample, cf. Section 2.4. In our simulations, the reference point was estimated by calculating at time n the statistic S_{nk} from n, $(n-1)$, $\ldots$, $(n-63)$. The estimator $\widehat{\varphi}_X(t)$ yielding the estimated reference point $\int |\widehat{\varphi}_X(t)| w(t)\, dt$ was estimated from the samples $(n-64)$, $(n-65), \ldots$, $(n-128)$, but only

jump	Prob. n.d.	Prob. n.d. CS1
0.2	0.227	0.229
0.4	0.159	0.201
0.6	0.109	0.119
0.8	0.066	0.057
1.	0.029	0.026
1.2	0.009	0.009
1.4	0.004	0.002

scale	Prob. n.d.	Prob. n. d. shift=0.8
1.5	0.209	0.252
1.7	0.071	0.070
1.9	0.013	0.016
2.1	0.007	0.006
2.3	0.	0.001
2.5	0.	0.

TABLE V

Left panel: Estimated probabilities that the T_L chart detects a jump with a delay of 30 units of time or later. The third column provides the same probabilities given scale changes according to SC1.
Right panel: Estimated probabilities that the T_S chart detects scale with a delay of 30 units of time or later. The third column provides those probabilities given a permanent shift change of size 0.8.

T_S Chart, ref. estimated			T_S Chart, ref. estimated + shift		
Scale	ARL	RL Disp,	Scale	ARL	RL Disp,
1	433.3	418.0	1	440.9	498.0
1.25	71.0	59.0	1.25	72.3	54.0
1.5	61.8	32.5	1.5	55.8	34.5
1.75	40.7	18.8	1.75	38.6	17.8
2	36.8	17.5	2	36.6	16.7

TABLE VI

ARLs of T_S chart (scale=1, i.e. in control) when the reference level is estimated: shift=0 (left panel), permanent shift=0.8σ (right panel).

once for each simulation run, i.e., at time instant $n = 129$. This way of using previous in-control samples ensures the independence of S_{nk} and $\widehat{\varphi}_X(t)$. The price for this is the need of having 128 pre-run samples.

The results are summarized in Table VI. As one can notice, out of control ARLs essentially increased in comparison to the case when the reference point is know. The reason is that the estimation of the reference point adds additional variation to the ARLs. As it was expected, the dispersions of run lengths increased as well, but the increase was not as large as that of the ARLs. On the other hand, the r.h.s. panel of this table indicates that no essential changes of ARLs and RL dispersions occur when additionally a shift in the mean is present.

7.3. *Accuracy of the subsampling approach*

Lastly, we investigate by a small simulation experiment the accuracy of the subsampling procedure. Our experiment has the real data analysis of the next section in mind. Thus, we fixed the time horizon at $N = 480$ and the size of the learning sample at $L = 240$. Instead of selecting the control limit to achieve a certain in-control ARL, we now address the type I error rate, cf. our discussion in Section 2.

For the subsampling approach, one has to choose the length b of the subseries. We analyzed which choice of the block length is appropriate to achieve a type I error rate when running the detector starting at $n_0 = 241$ and estimating unkowns from the learning sample. To simulate the resulting accuracy, for a simulated time series of length N the subsampling procedure uses the first $L = \lfloor N s_0 \rfloor$ observations as a learning sample to generate subsamples. Our ansatz

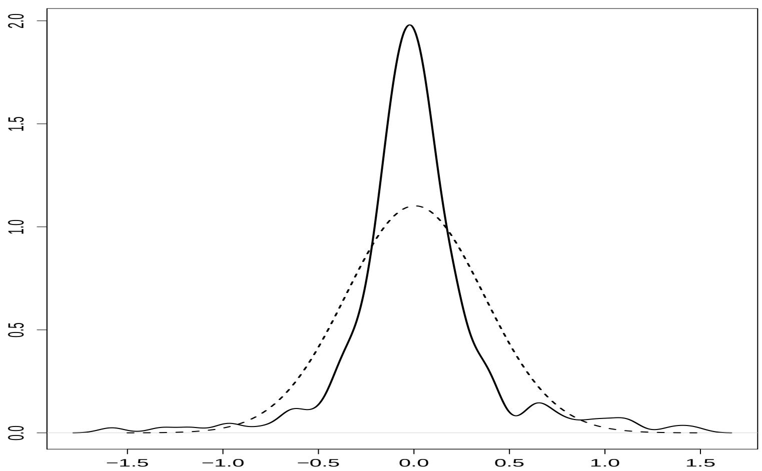

FIGURE 3.— Kernel density estimate for SO_2 data and estimated Gaussian density.

for the block length is $b = \lfloor L^\gamma \rfloor$. The detector is then applied to the subsamples in order to estimate the control limit that is chosen to control the relative frequency of a signal during the time span from $\lfloor s_0 b \rfloor$ to b within the subsamples. By consistency of subsampling, we control for the asymptotic significance level in this way. In practice, the assumption of centered in-control observations is sometimes not ensured and practitioners then subtract the mean of the observations from the learning sample. Thus, we took that into account. We simulated the procedure in this way for measurements following a standard normal distribution and for observations generated by bootstrapping from the errors of the learning sample of the SO_2 data studied in the next section, which also motivated the choice $s_0 = 1/2$. Figure 3 provides a kernel density estimate of the SO_2 data with cross-validated bandwidth choice. It is obvious that the errors are symmetric and non-normal, which is also confirmed by applying a Shapiro-Wilk test and the sign test for the median.

Table VII reports some simulations addressing the location chart that is based on our novel characteristic function approach not yet studied in the literature. To improve accuracy, we adopted the calibration procedure discussed in [31, p. 195]. We fixed $\gamma = 0.8$ and $N = 480$ and determined a calibration function for $N(0,1)$ errors by simulating the real rejection rates for various nominal type I error rates α and fitting a polynomial of degree 2. Table VII shows the results for $N(0,1)$ errors as well as for the SO_2 errors; each setting as well as each simulation step in the calibration procedure is based on $10,000$ repetitions. As one can notice, the calibrated subsampling procedure provides accurate results even when confronted with a highly nonnormal distribution. The table entries for $N = 240$ shed some additional light on the robustness of the calibrated subsampling approximation, when using a calibration function determined for a different maximal sample size.

8. APPLICATION TO CLIMATE DATA

For sake of illustration, we applied our procedures to a real data set of preprocessed intraday sulfur dioxide (SO_2) measurements taken every 30 minutes. Sulfur dioxide emission due to industrial production, ship engines and volcanoes is regarded a major source of acid rain. It is

N	Simulation from $N(0,1)$			Bootstrap from SO_2 errors		
	1%	5%	10%	1%	5%	10%
240	1.98%	5.05%	9.26%	2.95%	6.55%	9.58%
480	1.55%	5.10%	9.1%	1.62%	5.08%	9.23%

TABLE VII

Simulated level of the location detector with calibrated subsampling control limits.

also related to the creation of aerosols that have an effect on the climate, [25]. A major source of the aerosols in the tropical tropopause layer is the chemical reaction of SO_2 with OH, the chemical product condensing onto particles with water, thus creating new particles at a nanometer scale, see [22]. We set up appropriate monitoring schemes in order to detect changes in the location and scale of such SO_2 measurements. We analyzed a historic sample of $N = 480$ observations at the ground level corresponding to a time span of 10 days. The first $L = 240$ observations were regarded as in-control measurements and used as the learning sample, cf. Figure 4 that depicts the preprocessed measurements. Thus, a time span corresponding to 5 days is monitored. We aim at designing the procedures in such a way that the type I error rate is controlled at a fixed level α. Then a signal during the monitoring period can be regarded as being statistically significant.

For sake of simplicity, we use the weighting function $w(t) = t$ and $\mathbb{T} = [0, a]$ with $a = 1/2$. For the choice of the upper limit of integration a, which could be delicate in some cases, one could also rely on estimates of the first positive zero, see [48, chapter 3.3] and [17]. The integrals were approximated by Riemann sums with 240 grid points. The control limits were estimated from the learning sample by the subsampling approach with calibration using the block length $b = \lfloor L^f \rfloor$ with $f = 0.8$, in order to attain an asymptotic type I error rate equal to $\alpha = 5\%$. The monitoring procedures analyze the data located in a moving window defined by the parameter k, cf. the definitions of the detector statistics (2.10) and (2.13) as well as the definitions of the stopping times (2.12) and (2.14). We used $k = 48$ such that the procedures look back one day.

Figure 5 depicts the charts, i.e. the control statistics L_{nk} and S_{nk} as well as the associated control limits. The example confirms that, when using an effective sample size k for estimation, one should ignore the first k values of the control statistics, since with too few observations the estimates are unstable and lead to false signals. Ignoring this start up behavior, there are only a few signals for very short periods. The first part of the change in location starting around the $340th$ time instant and lasting for ca. 40 observations is clearly detected by the location chart, whereas the scale chart still does not react. The constant trend starting around the 400th observation is quickly detected by our location chart and results in a more or less diverging control statistic. The scale chart first reacts, since, in general, the first part of the trend can not be distinguished from an increase of variability, but the duration where the chart gives signals is in agreement with the memory parameter $k = 48$. However, the signals at the end of the monitoring period could be interpreted as a certain increase of the variability of SO_2 measurements.

ACKOWLEDGMENTS

Part of this work was undertaken during two research stays of the second author at Wrocław University of Technology. The authors thank anonymous referees for their helpful comments which improved the presentation of the results.

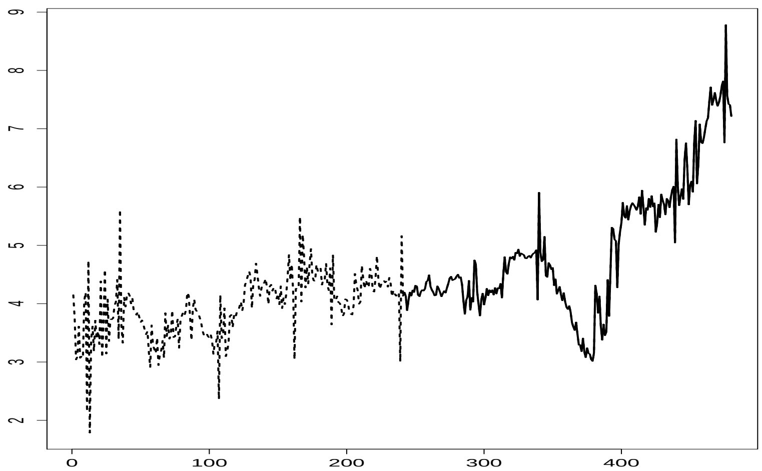

FIGURE 4.— SO_2 readings. The first 240 observations form the learning sample (dashed) and monitoring starts at observation 241. There is a visible and substantial increase of the SO_2 exposition which is followed by a certain increase of the variability.

APPENDIX A: PRELIMINARIES

Let us denote by $(\Omega, \mathcal{F}, P)$ the underlying probability space on which all random variables are defined. We are interested in the Skorohod space $D([s_0, 1] \times [T_0, T_1]; \mathbb{R}^l)$, but in principle one can assume that $s_0 = T_0 = 0$ and $T_1 = 1$. Thus, for the following review of some important facts on their definition and the concept of weak convergence (i.e. convergence in distribution) of stochastic processes attaining values in those spaces, let us consider the function spaces $D([a, b]^d, \mathbb{R}^l)$ for dimensions $d, l \in \mathbb{N}$ and real numbers a, b we assume to be $a = 0, b = 1$ in what follows. More information can be found in the monographs [2], [50] and the articles [29] and [46]. As a set of functions, the space $D([0, 1]^d; \mathbb{R}^l)$ is, basically, the uniform closure of simple functions that attain a constant value on each cell of a partition of $[0, 1]^d$, where one considers as partitions all partitions $\{A_k\}$ of $[0, 1]^d$ in cells given by the grid defined by the grid points $t_j = j/2^k$, $j = 1, \ldots, 2^k$, on each coordinate and all partitions $\{\lambda(A_k)\}$, $\lambda \in \Lambda^d$, where Λ is the class of continuous functions $[0, 1] \to [0, 1]$ that have a continuous inverse (i.e. homeomorphisms).

The space $D([0, 1]^d; \mathbb{R}^l)$ can be equipped with the following Skorohod metric, see [46]. For functions $\lambda \in \Lambda$ define the modified slope norm

$$\|\lambda\|^0 = \sup_{s,t \in [0,1], s \neq t} \left| \log \frac{|\lambda(s) - \lambda(t)|}{|s - t|} \right| + \|\lambda - \mathrm{id}\,\|_\infty.$$

Notice that Λ^d defines a group of homeomorphisms on $[0, 1]^d$ and

$$\|\lambda\|^0 = \max_{1 \leq j \leq d} \|\lambda_j\|^0, \qquad \lambda = (\lambda_1, \ldots, \lambda_d)' \in \Lambda^d,$$

defines a complete metric space (Λ^d, d_Λ) where the metric d_λ is induced by the norm $\| \bullet \|^0$. After these preliminaries, we may now equip the Skorohod space with the metric

$$d(f, g) = \inf\{\varepsilon > 0 : \exists \lambda \in \Lambda^d : \|\lambda\|^0 < \varepsilon, \|f - g \circ \lambda\|_\infty < \varepsilon\},$$

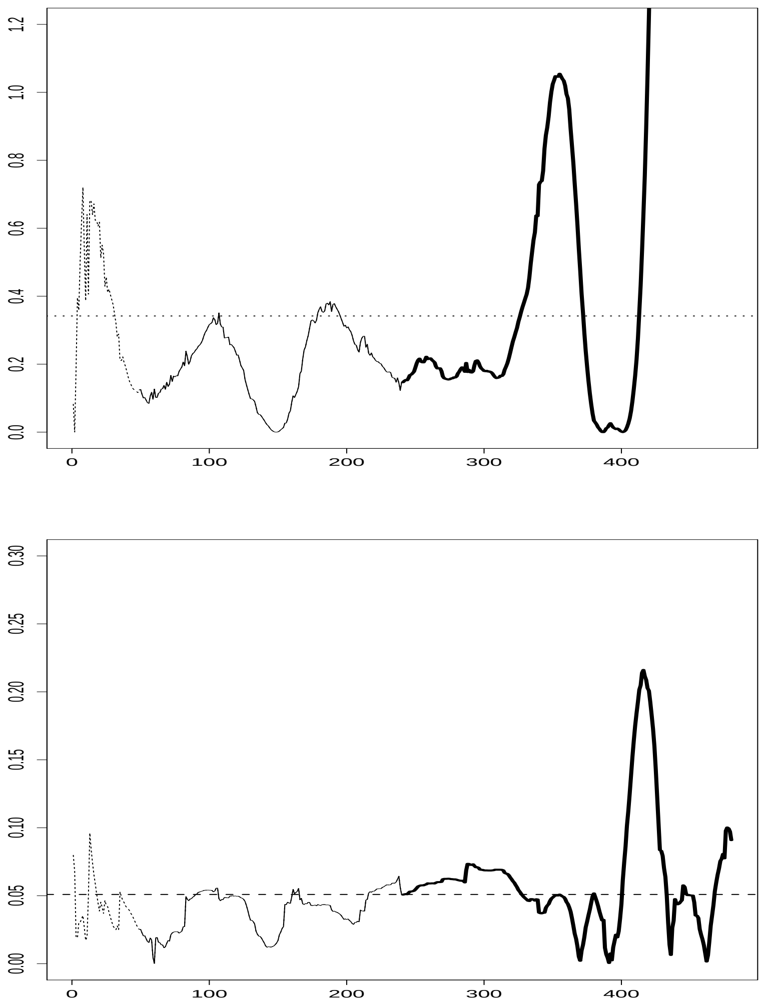

FIGURE 5.— Monitoring charts with control limits (dashed lines). The first 48 values are dashed, since they correspond to unstable estimation and should be ignored. The upper panel shows the location chart, the lower panel depicts the scale chart for the SO_2 intraday data.

for $f, g \in D([0,1]^d; \mathbb{R}^l)$ that makes it a complete metric space. Clearly, the supnorm of $f \in D([0,1]^d; \mathbb{R}^l)$ is defined here as $\|f\|_\infty = \sup_{t_1,\dots,t_d \in [0,1]} \|f(t_1,\dots,t_d)\|$, $\|a\|$, $a \in \mathbb{R}^l$, being an arbitrary vector norm. We have $d(f,g) \le \|f-g\|_\infty$ such that uniform convergence implies convergence in the Skorohod metric.

Let $\{X, X_n : n \in \mathbb{N}\}$ be a sequence of $D([0,1]^d, \mathbb{R}^l)$-valued processes, i.e. $X_n = (X_{n1}, \dots, X_{nl})'$, $n \ge 1$, and $X = (X_1, \dots, X_l)'$, where $X_{nj}, X_j : [0,1] \to \mathbb{R}$ are the corresponding coordinate mappings. Then the finite dimensional distributions (fidis) are given by the distributions of the matrix-valued random elements

$$(X_{n1}(t_r), \dots, X_{nl}(t_r))_{r=1}^k \, , n \ge 1, \qquad (X_1(t_r), \dots, X_l(t_r))_{r=1}^k \, ,$$

where $t_1, \dots, t_k \in [0,1]$, $k \in \mathbb{N}$, are fixed time points. Since a sequence of random matrices of fixed dimensions $k \times l$ converges in distribution if and only if the corresponding random vector obtained by stacking the columns converges in distribution in the Euclidean space $\mathbb{R}^{kl}$, the fidi convergence $X_n \stackrel{fidi}{\to} X$, as $n \to \infty$, holds, if one shows that $\sum_{r=1}^k \sum_{j=1}^l \lambda_{jr} X_{nj}(t_r) \stackrel{d}{\to} \sum_{r=1}^k \sum_{j=1}^l \lambda_{jr} X_j(t_r)$ for any choice of real numbers λ_{jr}, $1 \le r \le k$, $j = 1, \dots, l$ and for all $t_1, \dots, t_k \in [0,1]$, $k \in \mathbb{N}$. The weak convergence of a sequence $\{X, X_n\}$ of random functions $X, X_n \in D([0,1]^d; \mathbb{R}^l)$ is defined as the weak convergence of the measures P_{X_n} to P_X, as $n \to \infty$. This is equivalent to convergence of the fidis plus tightness. However, we shall make use of the Skorohod/Dudley/Wichura representation theorem (SDW theorem) in metric spaces that we have at our disposal, see [39, Th. 4, p. 47], applied to our basic assumption on the validity of a weak invariance principle for the sequential characteristic process. It asserts that for a weakly convergent sequence there exists equivalent (in distribution) versions on a new probability space that converge a.s. with respect to the Skorohod metric.

APPENDIX B: PROOFS

Proof of Theorem 2.1. The result follows immediately by observing that properties (i) and (ii) lead to the following two partial differential equations (PDE). If W satisfies (i), then the PDE

$$-\eta W_\chi(\chi, \eta) + \chi W_\eta(\chi, \eta) = 0$$

follows. Analogously, any C^1 transformation satisfying (ii) yields the PDE

$$\chi U_\chi(\chi, \eta) + \eta U_\eta(\chi, \eta) = 0.$$

The general solutions for these PDEs are well known and of the form

$$W(\chi, \eta) = \Phi(\chi^2 + \eta^2), \quad U(\chi, \eta) = \Psi(\eta/\chi),$$

for non-constant C^1 transformations Φ and Ψ, which completes the proof.

Proof of Theorem 3.1. For the i.i.d. case note that the CLT for random functions $X(t), X_i(t)$, $i = 1, \dots, n$, with continuous trajectories such that $|X(t) - X(s)| = O(|t-s|)$ satisfy the CLT. Precisely, a sufficient condition for the validity of the CLT for i.i.d. centered random cadlag functions $X(t), X_i(t)$, $i = 1, \dots, n$, with $EX_i(t)^2 < \infty$ with a limiting Gaussian process with continuous sample paths is the existence of non-negative functions f, g on $[0, \infty)$, which are nondecreasing near 0, and increasing continuous functions F, G defined on $[0,1]$, such that

$$E(X(t)-X(s))^2 \leq g(G(t)-G(s)), \tag{B.1}$$

$$E(|X(s)-X(t)| \wedge |X(t)-X(u)|)^2 \leq f(F(u)-F(s)), \tag{B.2}$$

for all $s \leq t \leq u$ with $u-s$ small, and $\int_0^1 f^{1/2}(u)u^{-3/2}\,du < \infty$ as well as $\int_0^1 g^{1/2}(u)u^{-5/4}\,du < \infty$, see [3, Th. A] and [10]. Furthermore, if the $X_i(t)$ are not centered, by virtue of [3, Prop. 1.2] it suffices to check those conditions for the $X_i(t)$ instead the centered versions. But in our case, $X_i(t) = \cos(tY_i)$ and $= \sin(tY_i)$, respectively, such that those conditions are easily verified, since, e.g.,

$$E(|\cos(tY_1)-\cos(tY_1)| \wedge |\cos(tY_1)-\cos(uY_1)|)^2 \leq E((|t-s||Y_1|) \wedge (|t-u||Y_1|))^2,$$

where the right-hand side can be bounded by $\max(|s-t|,|u-t|)E|Y_1|^2 \leq |u-s|E|Y_1|^2$. Now the FCLT follows from [4, Prop. 7] that provides the equivalence of the CLT and the FLCT. For further discussions on the CLT in general spaces, we refer to see [5].

The case of a general nonlinear time series can be handled as follows. By boundedness of sine and cosine and the fact that they have bounded variation over compact sets, the results in [40] apply and show that both coordinate processes of $\mathcal{C}_N(s,t)$ can be represented as linear functionals of a single underlying process that converges weakly for a nonlinear time series $\{Y_t\}$ satisfying (3.6) and (3.5) with i.i.d. errors $\{\epsilon_n : n \in \mathbb{Z}\}$. Thus the weak convergence of $\mathcal{C}_N(s,t)$ can be shown by an application of the SDW theorem. For sake of brevity, we omit the technical details. The covariance structure follows by a direct calculation. For example, for $s_1 \leq s_2$ and $t_1, t_2 \in \mathbb{T}$ we have

$$\begin{aligned}
\operatorname{Cov}(\mathcal{C}^{\cos}(s_1,t_1), \mathcal{C}^{\sin}(s_2,t_2)) &= \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{\lfloor Ns_1\rfloor} \sum_{k=1}^{\lfloor Ns_2\rfloor} \operatorname{Cov}(\cos(t_1Y_j), \sin(t_2Y_k)) \\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{\lfloor Ns_1\rfloor} \sum_{k=1}^{\lfloor Ns_2\rfloor} \operatorname{Cov}(\cos(t_1Y_{j-k}), \sin(t_2Y_0)) \\
&= s_1 \left\{ \gamma_{cs}(0;t_1,t_2) + 2\sum_{l=1}^{\infty} \gamma_{cs}(l;t_1,t_2) \right\}.
\end{aligned}$$

Proof of Corollary 3.1. By virtue of the SDW theorem in metric spaces, we may assume that $\mathcal{C}_N \to \mathcal{C}$, as $N \to \infty$, in the uniform topology on $D([s_0,1] \times [T_0,T_1]; \mathbb{R}^2)$. Note that for $0 < s_0 \leq \theta \leq s \leq 1$ we have $\sup_{s_0 \leq \theta \leq s \leq 1} |\lfloor Ns\rfloor/\lfloor N\theta\rfloor - s/\theta| \to 0$, as $N \to \infty$. By definition of $\widehat{R}^n_{n-i+1}$ and $\widehat{I}^n_{n-i+1}$, $1 \leq i \leq n$, $n \geq 1$, we obtain

$$\begin{aligned}
C_N^{(s,\theta)}(t) &= \sqrt{\lfloor N\theta\rfloor}\left(\widehat{R}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t) - R_X(t), \widehat{I}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t) - I_X(t)\right)' \\
&= \frac{1}{\sqrt{\lfloor N\theta\rfloor}} \sum_{j=\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}^{\lfloor Ns\rfloor} \left(\cos(tY_j) - E\cos(tY_j), \sin(tY_j) - E\sin(tY_j)\right)' \\
&= \sqrt{\frac{N}{\lfloor N\theta\rfloor}}\left[\mathcal{C}_N(s,t) - \mathcal{C}_N(\theta,t)\right] \\
&\to \theta^{-1/2}[\mathcal{C}(s,t) - \mathcal{C}(\theta,t)] = \mathcal{C}_\theta(s,t),
\end{aligned}$$

as $N \to \infty$, by the triangle inequality almost surely in the uniform topology. The second half of the SDW theorem now yields the weak convergence of the original versions.

PROOF OF THEOREM 4.1. Pick a function ψ satisfying our assumptions and consider

$$\begin{aligned}\mathcal{L}_N(\psi)(s) &= L_{\lfloor Ns\rfloor,\lfloor N\theta\rfloor}(\psi)\\ &= \sqrt{\lfloor N\theta\rfloor}\int\left[\psi\left(\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)\right)-\psi(C(t))\right]w(t)\,dt.\end{aligned}$$

Using the definition of the process $\mathcal{C}_N^{(s,\theta)}(t)$, a Taylor expansion leads us to

$$\begin{aligned}\mathcal{L}_N(\psi)(s) &= \int\dot\psi(\bar c_N(t))'\sqrt{\lfloor N\theta\rfloor}[\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)-C(t)]w(t)\,dt\\ &= \int\dot\psi(\bar c_N(t))'C_N^{(s,\theta)}(t)w(t)\,dt\end{aligned}$$

for some random point $\bar c_N(t)$ between $\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)$ and $C(t)$. By Corollary 3.1, we know that the moving average process converges weakly,

$$C_N^{(s,\theta)}(t) = \sqrt{\lfloor Ns\rfloor}\left(\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)-C(t)\right)\Rightarrow\mathcal{C}_\theta(s,t),$$

as $N \to \infty$, yielding

$$\sup_{s,t}|C_N^{(s,\theta)}(t)| = O_P(1).$$

But this implies $\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)\to C(t)$ uniformly in t, in probability. Noticing that $C(t)$ is a continuous and deterministic function, by virtue of [2, Th. 4.1], we may conclude that

$$\left(\sqrt{\lfloor Ns\rfloor}(\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}(t)-C(t)),\bar c_N(t)\right)\Rightarrow(\mathcal{C}_\theta(s,t),C(t)),$$

as $N \to \infty$, jointly, in the product space $D([s_0,1]\times[T_0,T_1];\mathbb{R}^2)\times D([T_0,T_1];\mathbb{R}^2)$. By virtue of SDW theorem, we may assume that for equivalent processes defined on a new probability space a.s. convergence in the Skohorod metric and, by continuity of the limit processes, also in the uniform topology holds true. For brevity of notation, in what follows we shall use the same symbols for the equivalent processes. This means,

$$\sup_{s,t}\left\|\left(\sqrt{\lfloor Ns\rfloor}(\widehat{C}^{\lfloor Ns\rfloor}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}-C(t)),\bar c_N(t)\right)-(\mathcal{C}_\theta(s,t),C(t))\right\|\overset{a.s.}{\to}0, \tag{B.3}$$

as $N \to \infty$.

Next consider the operator $\phi : D([s_0,1]\times[T_0,T_1];\mathbb{R}^2)\times D([s_0,1];\mathbb{R}^2)\to D([s_0,1];\mathbb{R}^2)$ given by

$$\phi(f,\xi)(s) = \int\dot\psi(\xi(t))'f(s,t)w(t)dt,\qquad s\in[0,1], \tag{B.4}$$

for $(f,\xi)\in D([s_0,1]\times[T_0,T_1];\mathbb{R}^2)\times D([T_0,T_1];\mathbb{R}^2)$. Let us agree to denote the above operator by $\phi(f,\xi)$. Take a sequence $\{(f_n,\xi_n),(f,\xi)\}\subset D([s_0,1]\times[T_0,T_1];\mathbb{R}^2)\times D([T_0,T_1];\mathbb{R}^2)$ with $f\in C([s_0,1]\times[T_0,T_1];\mathbb{R}^2)$ and $\xi\in C([s_0,1];\mathbb{R}^2)$, which converges in the Skorohod (product) topology, and therefore also in the uniform (product) topology, i.e. $\|f_n - f\|_\infty + \|\xi_n - \xi\|_\infty \to 0$, as $n \to \infty$. Consider the decomposition

(B.5)

$$\phi(f_n,\xi_n)(s) - \phi(f,\xi)(s) = \int [\dot\psi(\xi_n(t))' - \dot\psi(\xi(t))'] f_n(s,t) w(t)\,dt + \int \dot\psi(\xi(t))'(f_n(s,t) - f(s,t)) w(t)\,dt.$$

The first term can be bounded as follows.

$$\begin{aligned}\sup_s \left| \int [\dot\psi(\xi_n(t))' - \dot\psi(\xi(t))'] f_n(s,t) w(t)\,dt \right| &\le \int \|\dot\psi(\xi_n(t)) - \dot\psi(\xi(t))\| \sup_s \|f_n(s,t)\| w(t)\,dt \\ &= O\left(\int \|\dot\psi(\xi_n(t)) - \dot\psi(\xi(t))\| w(t)\,dt \right) = o(1),\end{aligned}$$

by the dominated convergence theorem, since $\int w(t)\,dt < \infty$. The supnorm of the second term in (B.5) is not larger than $\|\dot\psi\|_\infty \|f_n - f\|_\infty \int w(t)\,dt = o(1)$. This shows that the operator ϕ is continuous in the sense that

$$\phi(f_n,\xi_n) \to \phi(f,\xi), \qquad n \to \infty, \tag{B.6}$$

both in the Skorohod metric of the the space $D([0,1]\times[T_0,T_1];\mathbb{R}^2) \times D([T_0,T_1];\mathbb{R}^2)$ and the corresponding uniform topology. Now apply this fact with the almost surely convergent sequences

$$f_N(s,t) = \sqrt{\lfloor Ns \rfloor}(\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - C(t)) \quad \text{and} \quad \xi_N(t) = \bar{c}_N(t)$$

to conclude that

$$\mathcal{L}_N(\psi)(s) = \int \dot\psi(\bar{c}_N(t))' C_N^{(s,\theta)}(t) w(t)\,dt \to \int \dot\psi(C(t))' \mathcal{C}_\theta(s,t) w(t)\,dt \tag{B.7}$$

almost surely in the supnorm, as $N \to \infty$. By virtue of the second half of the SDW theorem, the almost sure convergence (B.7) implies the weak convergence of the orginal processes on the probability space $(\Omega, \mathcal{F}, P)$. Thus we have shown

$$\mathcal{L}_N(\psi)(s) \Rightarrow \int \dot\psi(C(t))' \mathcal{C}_\theta(s,t) w(t)\,dt,$$

as $N \to \infty$.

Proof of Theorem 4.2. Our starting point is the representation

$$\mathcal{L}_N(\psi)(s) = \lfloor N\theta \rfloor \int \left[\psi(\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t)) - \psi(C(t)) \right] w(t)\,dt.$$

Note that

$$\begin{aligned}&\lfloor N\theta \rfloor \left[\psi(\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t)) - \psi(C(t)) \right] \\ &\quad = \sqrt{\lfloor N\theta \rfloor} \dot\psi(C(t))' \sqrt{\lfloor N\theta \rfloor} [\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - C(t)] \\ &\quad + \frac{1}{2} \sqrt{\lfloor N\theta \rfloor} \left[\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - C(t) \right]' \ddot\psi(C(t)) \sqrt{\lfloor N\theta \rfloor} \left[\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t) - C(t) \right] \\ &\quad + O_P(1/\sqrt{N}),\end{aligned}$$

where the O_P term is uniform in t and s, since $\sqrt{\lfloor N\theta \rfloor}[\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t)) - C(t)]$ converges weakly to a continuous process under the imposed assumptions. Using $\dot{\psi}(C(\cdot)) = 0$ except on a $wd\lambda$-null set, we arrive at

$$\mathcal{L}_N(\psi)(s) = \frac{1}{2} \int [C_N^{(s,t)}]' \ddot{\psi}(C(t)) C_N^{(s,t)} w(t)\, dt + o_P(1).$$

Now one argues as in the proof of Theorem 4.1 to conclude that

$$\mathcal{L}_N(\psi) \Rightarrow \frac{1}{2} \int \mathcal{C}_\theta(s,t)' \ddot{\psi}(C(t)) \mathcal{C}_\theta(s,t) w(t)\, dt,$$

as $N \to \infty$.

Proof of Lemma 4.1. It suffices to show (i). Let $A \subset D([T_0, T_1]^2; \mathbb{R}^2)$ and $B \subset D([0,1]^2; \mathbb{R}^2)$ be measurable subsets such that A is $\mathcal{C}_N^{(0)}$-continuous and B is $\mathcal{C}_N$-continuous. Since $\{\mathcal{C}_N^{(0)} \in A\} \in \mathcal{F}_1^L = \sigma(X_1, \dots, X_L)$ and $\{\mathcal{C}_N \in B\} \in \mathcal{F}^\infty_{\lfloor N s_0 \rfloor} = \sigma(X_{\lfloor N s_0 \rfloor}, \dots)$, we obtain

$$\begin{aligned} P(\mathcal{C}_N^{(0)} \in A, \mathcal{C}_N \in B) &= P(\mathcal{C}_N^{(0)} \in A) P(\mathcal{C}_N \in B) + \alpha(\lfloor N s_0 \rfloor - \lfloor N(s_0 - \varepsilon) \rfloor) \\ &= P(\mathcal{C}_\theta^{(0)} \in A) P(\mathcal{C}_\theta \in B) + o(1), \end{aligned}$$

since $\mathcal{C}_N^{(0)} \Rightarrow \mathcal{C}_\theta^{(0)}$ and $\mathcal{C}_N \Rightarrow \mathcal{C}_\theta$, if $N \to \infty$, imply $P(\mathcal{C}_N^{(0)} \in A) = P(\mathcal{C}_\theta^{(0)} \in A) + o(1)$, and $P(\mathcal{C}_N \in B) = P(\mathcal{C}_\theta \in B) + o(1)$, as $N \to \infty$. This shows $P_{(\mathcal{C}_N^{(0)}, \mathcal{C}_N)} \Rightarrow P_{(\mathcal{C}_\theta^{(0)}, \mathcal{C}_\theta)}$, as $N \to \infty$, in the product space, cf. [2, p. 27].

Proof of Theorem 4.3. Notice the decomposition $\widetilde{\mathcal{L}}_N = \mathcal{L}_N + \mathcal{L}_N^{(0)}$. where

$$\mathcal{L}_N^{(0)}(s) = \lfloor N\theta \rfloor^{1/2} \left\{ \int \psi(\widehat{C}^{\lfloor Ns \rfloor}_{\lfloor Ns \rfloor - \lfloor N\theta \rfloor + 1}(t)) w(t)\, dt - \int \psi(C(t)) w(t)\, dt \right\}$$

Now the result follows easily, since by linearity for $N \to \infty$

$$\begin{aligned} \widetilde{\mathcal{L}}_N &= \int \dot{\psi}(C(t))' \mathcal{C}_N(s,t) w(t)\, dt + \int \dot{\psi}(C(t))' \mathcal{C}_N^{(0)}(t) w(t) dt + o_P(1) \\ &\Rightarrow \int \dot{\psi}(C(t))' [\mathcal{C}_\theta(s,t) + \mathcal{C}_\theta^{(0)}(t)] w(t)\, dt \end{aligned}$$

Proof of Theorem 5.1. Since by assumption $\{Y_n : 1 \leq n \leq L\}$ is α-mixing with mixing coefficients tending to 0, we may directly apply a general subsampling result, [30, Prop. 3.1], if we let $X_i := Y_i$ and define the roots as $R_b(X_\ell, \dots, X_{\ell - b + 1}; \theta(Q)) = r_{b\ell}$ and $R_b(X_\ell, \dots, X_{\ell - b + 1}; \theta(Q)) = \widetilde{\mathcal{L}}_{b\ell}(\psi)$, respectively, for $\ell = 1, \dots, L - b + 1$. In the latter case, one uses the separable Skorohod space $D([s_0, 1]; \mathbb{R})$.

Proof of Theorem 6.1. Put

$$R_{YN}^{(1)}(t) = E_N \cos(tY_{Nq}), \qquad I_{YN}^{(1)}(t) = E_N \sin(tY_{Nq})$$

and notice that $R_{YN}^{(1)}(t) + i I_{YN}^{(1)}(t)$ is the ch.f. of Y_j for $j \geq q = \lfloor N\tau \rfloor$, whereas the ch.f. of Y_j for $j < q$ equals $\varphi_X(t)) = R_X(t) + i I_X(t)$. Here and in the sequel, E_N denotes the expectation under P_N. Observe that in the characteristic process

$$\mathcal{C}_N(s,t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} [(\cos(tY_{Nj}), \sin(tY_{Nj}))' - (R_X(t), I_X(t))']$$

the terms are centered at their expectation under the no-change hypothesis. We have

$$\begin{aligned}
\mathcal{C}_N(s,t) &= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \begin{pmatrix} \cos(tY_{Nj}) - E_N \cos(tY_{Nj}) \\ \sin(tY_{Nj}) - E_N \sin(tY_{Nj}) \end{pmatrix} + \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \begin{pmatrix} E_N \cos(tY_{Nj}) - R_X(t) \\ E_N \sin(tY_{Nj}) - I_X(t) \end{pmatrix} \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} \begin{pmatrix} \cos(tY_{Nj}) - E_N \cos(tY_{Nj}) \\ \sin(tY_{Nj}) - E_N \sin(tY_{Nj}) \end{pmatrix} \\
&\quad + \frac{\lfloor Ns \rfloor - \lfloor N\tau \rfloor + 1}{\sqrt{N}} \begin{pmatrix} R_{YN}^{(1)}(t) - R_X(t) \\ I_{YN}^{(1)}(t) - I_X(t) \end{pmatrix} 1(\lfloor Ns \rfloor \geq \lfloor N\tau \rfloor) \\
&= \mathcal{C}_N^{(0)}(s,t) + c_N(s,t).
\end{aligned}$$

Note that

$$\begin{aligned}
\cos(\mu_N t) - 1 &= -\Delta_\mu^2 / 2N^{1/2} + O(N^{-1}), \\
\varphi_X(\sigma_N t) &= \varphi_X(t) + \varphi_X'(t) t \Delta_\sigma N^{-1/2} + O(N^{-1}).
\end{aligned}$$

Now one easily checks that

$$I_{YN}^{(1)}(t) - I_X(t) = \sin(\mu_N t) \varphi_X(\sigma_N t) = \frac{\Delta_\mu}{N^{1/2}} t \varphi_X(t) + O\left(\frac{1}{N} \right)$$

and

$$\begin{aligned}
R_{YN}^{(1)}(t) - R_X(t) &= [\cos(\mu_N t) - 1][\varphi_X(\sigma_N t) - \varphi_X(t)] + [\varphi_X(\sigma_N t) - \varphi_X(t)] \\
&\quad + [\cos(\mu_N t) - 1] \varphi_X(t) \\
&= [\varphi_X'(t) t \Delta_\sigma - \Delta_\mu^2 t^2 \varphi_X(t)/2] N^{-1/2} + O(N^{-1/2})
\end{aligned}$$

where the O terms are uniform in $t \in A$ for any compact set A. Thus,

$$\begin{aligned}
c_N(s,t) &= \frac{\lfloor Ns \rfloor - \lfloor N\tau \rfloor}{N} \sqrt{N} \begin{pmatrix} R_{YN}^{(1)}(t) - R_X(t) \\ I_{YN}^{(1)}(t) - I_X(t) \end{pmatrix} 1(\lfloor Ns \rfloor \geq \lfloor N\tau \rfloor) \\
&\to (s - \tau) \begin{pmatrix} \Delta_\mu t \varphi_X(t) \\ \varphi_X'(t) t \Delta_\sigma - \Delta_\mu^2 \varphi_X(t)/2 \end{pmatrix} 1(s > t) \\
&= c(s,t),
\end{aligned}$$

as $N \to \infty$, in the Skorohod metric. To verify $\mathcal{C}_N^{(0)} \Rightarrow \mathcal{C}$, as $N \to \infty$, under the sequence P_N of probability measures, we interpret $\mathcal{C}_N^{(0)}(s,t)$ as a random variable taking values in $\mathbb{C}$. Notice that by assumption $(Y_{N1}, \dots, Y_{NN})$ is equal in distribution to $(X_1, \dots, X_{\lfloor N\tau \rfloor - 1}, \mu_N + \sigma_N X_{\lfloor N\tau \rfloor}, \dots, \mu_N + \sigma_N X_N)$, under P_N. We have for $s > t$

$$\begin{aligned}
\mathcal{C}_N^{(0)}(s,t) &= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor Ns \rfloor} [e^{itY_{Nj}} - E_N(e^{itY_{Nj}})] \\
&= \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor N\tau \rfloor - 1} [e^{itY_{Nj}} - \varphi_X(t)] + \frac{1}{\sqrt{N}} \sum_{j=\lfloor N\tau \rfloor}^{\lfloor Ns \rfloor} [e^{itY_{Nj}} - e^{it\mu_N} \varphi_X(\sigma_N t)] \\
&\stackrel{d}{=} \frac{1}{\sqrt{N}} \sum_{j=1}^{\lfloor N\tau \rfloor - 1} [e^{itY_{Nj}} - \varphi_X(t)] + \frac{e^{it\mu_N}}{\sqrt{N}} \sum_{j=\lfloor N\tau \rfloor}^{\lfloor Ns \rfloor} [e^{itY_{Nj}} - \varphi_X(\sigma_N t)]
\end{aligned}$$

This leads to the representation

$$\mathcal{C}_N^{(0)}(s,t) \stackrel{d}{=} \widetilde{\mathcal{C}}_N(\tau-1/N,t) + e^{it\mu_N}[\widetilde{\mathcal{C}}_N(s,\varphi_N(t)) - \widetilde{\mathcal{C}}_N(\tau-1/N,\varphi_N(t))]1(\lfloor Ns\rfloor > \lfloor Nt\rfloor),$$

if we put $\widetilde{\mathcal{C}}_N(s,u) = N^{-1/2}\sum_{j=1}^{\lfloor Ns\rfloor}[e^{iuX_j} - \varphi_X(t)]$ for $s \in [s_0,1]$, $t \in \mathbb{T}$ and $\varphi_N(t) = \sigma_N t$ for $t \in \mathbb{T}$. The next step is to verify that the three processes on the right-side of the last display converge jointly. Observe that

$$\begin{pmatrix} \widetilde{\mathcal{C}}_N(\tau-1/N,t) \\ \widetilde{\mathcal{C}}_N(s,\varphi_N(t)) \\ \widetilde{\mathcal{C}}_N(\tau,\varphi_N(t)) \end{pmatrix} = \Psi_N \begin{pmatrix} T_N(\widetilde{\mathcal{C}}_N(s,t)) \\ \widetilde{\mathcal{C}}_N(s,\varphi_N(t)) \\ T_N(\widetilde{\mathcal{C}}_N(s,\varphi_N(t))) \end{pmatrix}$$

where for $N \in \mathbb{N}$ the mappings $\Psi_N : F \to F$, $F = D([T_0,T_1];\mathbb{R}^2) \times D([s_0,1]\times[T_0,T_1];\mathbb{R}^2) \times D([T_0,T_1];\mathbb{R}^2)$, are given by

$$\Psi_N(f_1,f_2,f_3)(t) = (f_1(s,t), f_2(s,\varphi_N(t)), f_3(s,\varphi_N(t)))$$

for $(f_1,f_2,f_3) \in F$, and $T_N : D([s_0,1]\times[T_0,T_1];\mathbb{R}^2) \to D([T_0,T_1];\mathbb{R}^2)$ by $T_N(f)(s,t) = f(\tau-1/N,t)$. Clearly, $\|\varphi_N - \mathrm{id}\|_\infty \to 0$, as $N \to \infty$, as well as $\|T_N - \mathrm{id}\|_\infty \to 0$, $N \to \infty$, holds true on the subspace of continuous functions. Since the weak limit $\mathcal{C}$ of $\mathcal{C}_N$ (under P_N) is a.s. continuous, the continuous mapping theorem entails the joint weak convergence. Therefore, combining this fact with $(\mu_N,\sigma_N) \to (0,1)$, we may conclude that

$$\begin{aligned}\mathcal{C}_N^{(0)}(s,t) &\stackrel{d}{=} \widetilde{\mathcal{C}}_N(\tau-1/N,t) + e^{it\mu_N}[\widetilde{\mathcal{C}}_N(s,\varphi_N(t)) - \widetilde{\mathcal{C}}_N(\tau-1/N,\varphi_N(t))] \\ &\Rightarrow \mathcal{C}(\tau,t) + e^0[\mathcal{C}(s,t) - \mathcal{C}(\tau,t)] = \mathcal{C}(s,t),\end{aligned}$$

as $N \to \infty$, which also implies $\mathcal{C}_N \Rightarrow \mathcal{C} + c$, as $N \to \infty$. The result for the moving-window process $\mathcal{C}_N^{(s,\theta)}(s,t)$ follows now easily from the algebraic identity

$$\mathcal{C}_N^{(s,\theta)}(s,t) = \sqrt{\frac{N}{\lfloor N\theta\rfloor}}[C_N(s,t) - C_N(\theta,t)].$$

Proof of Theorem 6.2. We may argue as in the proof of Theorem 4.1 and therefore give only a sketch. As shown in Theorem 6.1,

$$\mathcal{C}_N(s,t) \to \mathcal{C}(s,t) + c(s,t), \qquad \text{and} \qquad C_N^{(s,\theta)}(t) \to \mathcal{C}_\theta^c(s,t),$$

as $N \to \infty$, and this convergence can be understood as a.s. convergence in the uniform topology for equivalent versions on a new probability space. Thus, recalling that $C = C_X$, for some random point $\bar{c}_N(t)$ between $\widehat{C}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}^{\lfloor Ns\rfloor}(t)$ and $C(t)$, we obtain

$$\begin{aligned}\mathcal{L}_N(\psi)(s) &= \sqrt{\lfloor N\theta\rfloor}\left\{\int \psi(\widehat{C}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}^{\lfloor Ns\rfloor}(t))w(t)\,dt - \int \psi(C(t))w(t)\,dt\right\} \\ &= \sqrt{\lfloor N\theta\rfloor}\int[\psi(\widehat{C}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}^{\lfloor Ns\rfloor})(t) - \psi(C(t))]w(t)\,dt \\ &= \int \dot\psi(\bar{c}_N(t))'\sqrt{\lfloor N\theta\rfloor}[\widehat{C}_{\lfloor Ns\rfloor-\lfloor N\theta\rfloor+1}^{\lfloor Ns\rfloor}(t) - C(t)]w(t)\,dt \\ &\Rightarrow \int \dot\psi(C(t))'\mathcal{C}_\theta^c(s,t)w(t)\,dt,\end{aligned}$$

as $N \to \infty$, by virtue of the results of Theorem 6.1.

REFERENCES

[1] D. Barash. A fundamental relationship between bilateral filtering, adaptive smoothing and the nonlinear diffusion equation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(6):844–847, 2002.

[2] P. Billingsley. *Convergence of Probability Measures.* John Wiley & Sons Inc., New York, 1968.

[3] M. Bloznelis and V. Paulauskas. Central limit theorem in skorokhod spaces and asymptotic strength distribution of fiber bundles. pages 75–87, 1993.

[4] M. Bloznelis and V. Paulauskas. On the central limit theorem for multiparameter stochastic processes. In *Probability in Banach spaces, 9 (Sandjberg, 1993)*, volume 35 of *Progr. Probab.*, pages 155–172. Birkhäuser Boston, Boston, MA, 1994.

[5] D. Bosq. *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics.* Springer-Verlag, New York, 2000. Theory and applications.

[6] B.E. Brodsky and B.S. Darkhovsky. *Non-Parametric Statistical Diagnosis. Problems and Methods.* Mathematics and its Applications (Dordrecht). 509. Dordrecht: Kluwer Academic Publishers. xvi, 452 p. .

[7] Coll B. Buades, A. and J.M. Morel. Neighborhood filters and PDEs. *Numerical Mathematics*, 1:1–34, 2006.

[8] Miklós Csörgő and Lajos Horváth. *Limit Theorems in Change-Point Analysis.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1997. With a foreword by David Kendall.

[9] S. Csörgő. Limit behaviour of the empirical characteristic function. *Ann. Probab.*, 9(1):130–144, 1981.

[10] X. Fernique. Les fonctions aléatoires càdlàg, la compacité de leurs lois. *Liet. Mat. Rink.*, 34(3):288–306, 1994.

[11] A. Feuerverger. A consistent test for bivariate dependence. *International Statistical Review*, 61(3):419.

[12] A. Feuerverger and R. A. Mureika. The empirical characteristic function and its applications. *Ann. Statist.*, 5(1):88–97, 1977.

[13] Malay Ghosh, Nitis Mukhopadhyay, and Pranab K. Sen. *Sequential Estimation.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, 1997. A Wiley-Interscience Publication.

[14] Lambert A. Gijbels, I. and P. Qiu. Edge-preserving image denoising and estimation of discontinuous surfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(7):1075–1087, 2006.

[15] D. Han and F. Tsung. A generalized EWMA control chart and its comparison with the optimal EWMA, CUSUM and GLR schemes. *Ann. Statist.*, 32(1):316–339, 2004.

[16] D. M. Hawkins and K.D. Zamba. A change-point model for a shift in variance. *J. Qual. Tech.*, 37(1):21–31.

[17] C. R. Heathcote and J. Hüsler. The first zero of an empirical characteristic function. *Stochastic Process. Appl.*, 35(2):347–360, 1990.

[18] N. Henze, B. Klar, and S. G. Meintanis. Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. *J. Multivariate Anal.*, 87(2):275–297, 2003.

[19] M. Hušková and S. G. Meintanis. Change point analysis based on empirical characteristic functions. *Metrika*, 63(2):145–168, 2006.

[20] M. Hušková and S. G. Meintanis. Testing procedures based on empirical functions ii: k-sample problem, change point problem. *Tatra Mt. Math. Publ.*, 39:235–243, 2008.

[21] M. Hušková and S. G. Meintanis. Tests for the multivariate k-sample problem based on the empirical characteristic function. *J. Nonparametr. Stat.*, 20(3):263–277, 2008.

[22] S. Fueglistaler D. Weisenstein M. Rex M.G. Lawrence H. Bingemer I. Wohltmann T. Corti T. Warneke R. von Kuhlmann J. Notholt, B.P. Luo and T. Peter. Influence of tropospheric SO2 emissions on particle formation and the stratospheric humidity. *Geophysical Research Letters*, 32, 2005.

[23] John T. Kent. A weak convergence theorem for the empirical characteristic function. *J. Appl. Probability*, 12(3):515–523, 1975.

[24] J.S. Lee. Digital image smoothing and the sigma filter. *Computer Vision, Graphics, and Image Processing*, 24:255–269, 1983.

[25] J. Lelieveld and J. Heintzenberg. Sulfate cooling effect on climate through in-cloud oxidation of anthropogenic SO2. *Science*, 258(5079):117–120, 1992.

[26] G. Lorden. Procedures for reacting to a change in distribution. *Ann. Math. Statist.*, 42:1897–1908, 1971.

[27] M. B. Marcus. Weak convergence of the empirical characteristic function. *Ann. Probab.*, 9(2):194–201, 1981.

[28] G. V. Moustakides. Optimal stopping times for detecting changes in distributions. *Ann. Stat.*, 14:1379–1387, 1986.

[29] G. Neuhaus. On weak convergence of stochastic processes with multidimensional time parameter. *Ann. Math. Statist.*, 42:1285–1295, 1971.

[30] Dimitris Politis, Joseph P. Romano, and Michael Wolf. Weak convergence of dependent empirical measures with application to subsampling in function spaces. *J. Statist. Plann. Inference*, 79(2):179–190, 1999.
[31] Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer Series in Statistics. Springer-Verlag, New York, 1999.
[32] Aleksey S. Polunchenko and Alexander G. Tartakovsky. On optimality of the Shiryaev-Roberts procedure for detecting a change in distribution. *Ann. Statist.*, 38(6):3445–3457, 2010.
[33] H. Vincent Poor. *An Introduction to Signal Detection and Estimation*. Springer, New York, 1994.
[34] H. Vincent Poor and Olympia Hadjiliadis. *Quickest Detection*. Cambridge University Press, New York, 2009.
[35] W.K. Pratt. *Digital Image Processing*. John Wiley & Sons, Ltd, New York, 2001.
[36] M. H. Quenouille. Approximate tests of correlation in time-series. *J. Roy. Statist. Soc. Ser. B.*, 11:68–84, 1949.
[37] E. Rafajłowicz, M. Pawlak, and A. Steland. Nonlinear image processing and filtering: a unified approach based on vertically weighted regression. *Int. J. Appl. Math. Comput. Sci.*, 18(1):49–61, 2008.
[38] E. Rafajłowicz and A. Steland. A binary control chart to detect small jumps. *Statistics*, 43(3):295–311, 2009.
[39] G.R. Shorack and J.A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.
[40] A. Steland. Asymptotics for random functions moderated by dependent noise. *Preprint*.
[41] A. Steland. Discussion on 'Quickest Detection Problem - Fifty Years Later' by Professor A.N. Shiryaev, Steklov Institute, invited. *Sequential Analysis*, 29(3):403–407.
[42] A. Steland. Jump-preserving monitoring of dependent time series using pilot estimators. *Statist. Decisions*, 21(4):343–366, 2003.
[43] A. Steland. On the distribution of the clipping median under a mixture model. *Statist. Probab. Lett.*, 71(1):1–13, 2005.
[44] A. Steland. A bootstrap view on Dickey-Fuller control charts for AR(1) series. *Austrian Journal of Statistics*, 35:339–346, 2006.
[45] A. Steland. *Financial Statistics and Mathematical Finance: Methods, Models and Applications*. John Wiley & Sons, Ltd, 2012.
[46] M. L. Straf. Weak convergence of stochastic processes with several parameters. Proc. 6th Berkeley Sympos. math. Statist. Probab., Univ. Calif. 1970, 2, 187-221 (1972)., 1972.
[47] J.W. Tukey. Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics*, 29:614, 1958.
[48] N. G. Ushakov. *Selected Topics in Characteristic Functions*. Modern Probability and Statistics. VSP, Utrecht, 1999.
[49] M. T. Wells. The weak approximation of the empirical characteristic function process when parameters are estimated. *Stochastic Process. Appl.*, 40(1):83–102, 1992.
[50] W. Whitt. *Stochastic-Process Limits. An Introduction to Stochastic-Process limits and Their Application to Queues*. Springer Series in Operations Research. Springer-Verlag, New York, 2002.
[51] S. Zacks and N. Mukhopadhyay. Distributions of sequential and two-stage stopping times for fixed-width confidence intervals in Bernoulli trials: application in reliability. *Sequential Anal.*, 26(4):425–441, 2007.