

Detection of anomalous sequences in crack data of a bridge monitoring

Sermad Abbas, Roland Fried, Jens Heinrich, Melanie Horn, Mirko Jakubzik, Johanna Kohlenbach, Reinhard Maurer, Anne Michels, and Christine H. Müller

Abstract For estimating the remaining lifetime of old prestressed concrete bridges, a monitoring of crack widths can be used. However, the time series of crack widths show a strong variation mainly caused by temperature and traffic. Additionally, sequences with extreme volatility appear where the cause is unknown. They are called anomalous sequences in the following. We present and compare four methods which aim to detect these anomalous sequences in the time series. Volatilities caused by traffic should not be detected.

Sermad Abbas (responsible for Method LCP)
Department of Statistics, TU University Dortmund, e-mail: sermad.abbas@uni-dortmund.de

Roland Fried (responsible for Method LCP)
Department of Statistics, TU University Dortmund, e-mail: biostat@statistik.tu-dortmund.de

Jens Heinrich (responsible for experimental setup and bridge monitoring)
Department of Architecture and Civil Engineering, TU University Dortmund, e-mail: jens.heinrich@tu-dortmund.de

Melanie Horn (responsible for data transformation)
Department of Statistics, TU University Dortmund, e-mail: mhorn@statistik.tu-dortmund.de

Mirko Jakubzik (responsible for Method CMAD)
Department of Statistics, TU University Dortmund, e-mail: jakubzik@statistik.tu-dortmund.de

Johanna Kohlenbach (responsible for Method CMAD)
Department of Statistics, TU University Dortmund, e-mail: johanna.kohlenbach@tu-dortmund.de

Reinhard Maurer (responsible for experimental setup and bridge monitoring)
Department of Architecture and Civil Engineering, TU University Dortmund, e-mail: reinhard.maurer@tu-dortmund.de

Anne Michels (responsible for Method VCP, MVCP)
Department of Statistics, TU University Dortmund, e-mail: anne2.michels@tu-dortmund.de

Christine H. Müller (responsible for Method VCP, MVCP and the general text)
Department of Statistics, TU University Dortmund, e-mail: cmueller@statistik.tu-dortmund.de

1 Introduction

The accident in Genoa on August 14th, 2018, showed that the safety of bridges is a severe problem. This is because many bridges were built in the fifties and sixties of the last century. They were designed and constructed for much lower than today's traffic loads in reality. Furthermore, the principles for designing and detailing had been advanced. Hence, an important task is to predict the remaining lifetime of old prestressed concrete bridges. This can be done by monitoring cracks in the bridges. [8], [9], [16] showed how this can be done in lab experiments with prestressed concrete beams. Crucial in these approaches is to detect the breaking of tension wires. This was mainly done by acoustic measurements. Thereby, it turned out that each break causes a distinct increase of the width of an initial crack.

These acoustic measurements are not possible in bridge monitoring. Only the crack widths can be monitored. However, they are influenced by the traffic and, more heavily, by temperature. The left-hand side of Fig. 1 shows crack width measurements together with two temperature measurements. The traffic load of the tram and other heavy vehicles are visible as isolated peaks. In particular, there is a large isolated peak at 6:30 a.m. which was caused by a heavy test load using a mobile crane of 48 tons. Moreover, the measurements of the crack widths are disturbed by strong anomalous sequences. Such an anomalous sequence is shown on the right-hand side of Fig. 1. It provides a zoom around 12:48 p.m. of the left-hand plot.

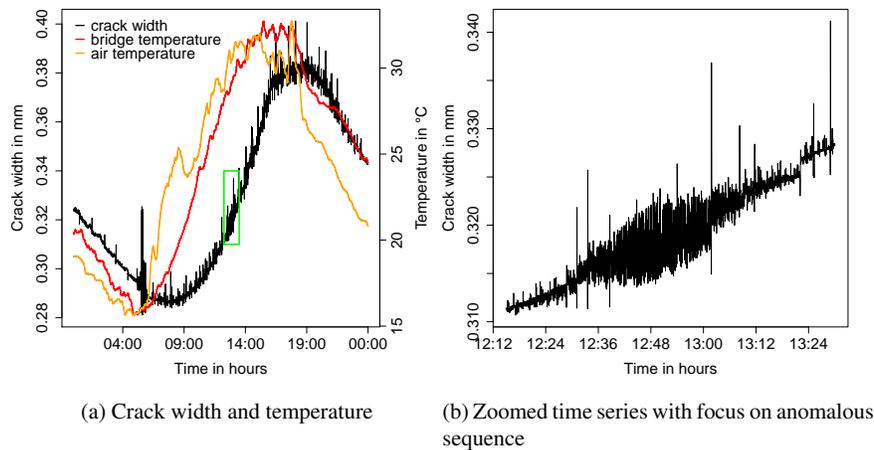


Fig. 1: Left: crack width and temperature measured above (air temperature) and below (bridge temperature) the bridge on June 5th, 2016, for sensor WWS4. Right: zoom of the crack width curve given by the green box on the left-hand side.

Anomalous sequences show an extreme volatility. It is unclear whether these phenomena are caused by beat effects or if there are other causes. In any case, they disturb the analysis severely so that a first step is to extract these anomalous sequences. Thereby, it is important that peaks caused by traffic are not extracted.

We present four methods for detecting the anomalous sequences. All of them are based on filtering the time series of crack width data. Three of them use additionally special change point detection methods. A location change point method is used by the LCP (Location Change Point Detection using Squared Residuals) method, in which a two-sample test based on the Hodges-Lehmann estimator for shift is applied on the logarithm of squared residuals in a moving time window, see [1] and [7]. The VCP (Variance Change Point Detection) and MVCP (Modified Variance Change Point Detection) methods use the PELT (Pruned Exact Linear Time) method of [13] for variance change point detection. The PELT method aims to find the global minimum of a cost function with a penalty for several change points where the number of evaluations is linear in the number of observations. The fourth method, called CMAD (Clustering of MAD filtered data), uses a MAD (Median Absolute Deviation from the median) filter and two times a median filter. It clusters the MAD filtered data by the k -means algorithm.

In Section 2, the experimental setup and the data are described. The statistical methods are explained in Section 3. Section 4 provides a comparison of the four methods. Finally, in Section 5 a discussion is given.



Fig. 2: Bridge view during the installation of the monitoring.

2 Experimental Setup and Data Description

2.1 Bridge Monitoring

2.1.1 Motivation

In June 2016, a bridge monitoring on an existing post-tensioned prestressed concrete bridge in Bochum built in 1961 (Fig. 2) has been installed and started. The reason was an assessment of this bridge by recalculation. It showed a significant deficit of fatigue strength according to the current standards in certain areas of the

superstructure. Furthermore, in those critical areas, several bending cracks showing widths of more than 0.5 mm have been detected during a routine bridge inspection. Therefore, it was decided to carry out a crack monitoring with sensors at 16 measurement points to observe a potential increase of crack widths which indicates a damage due to fatigue in the cracked areas.

2.1.2 State of the building

In the course of a design recalculation, deficits with regard to the limit of the state of decompression and fatigue strength were determined. The assessment was based on the *German recalculation guideline for existing buildings* [3]. An important aspect of this guideline is the determination of a calculated remaining service life of existing structures under traffic. It was possible to estimate a remaining service life due to fatigue until the year 2019.

Considering this low remaining service life and also the large crack widths, which could already point at a preexisting damage, it has been decided to build a new bridge. Furthermore, additional measures were taken to prevent the collapse of the structure. In addition to the restriction of a maximum weight of 24 tons per vehicle and the limitation to one lane in each direction, a crack monitoring was planned until the demolition of the existing bridge.

2.1.3 Objectives of the monitoring

The primary objective of the monitoring was the early and timely detection of crack width increases, which could indicate failure of prestressing steel due to fatigue. In case of a successive crack width increase as a result of prestressing steel failure, further emergency measures, such as bridge closure and temporary support, can be initiated.

2.1.4 Installation of the monitoring

During continuous monitoring of the structure, high-frequency crack measurements by means of inductive displacement transducers are used, which were applied at the 16 areas of cracking. These transducers were denoted by WON1 to WON4, WWN1 to WWN4, WOS1 to WOS4, WWS1 to WWS4, where "O" stands for "east", "W" for "west", "N" for "north", and "S" for "south". In addition to the installation of the transducers, three measuring points for temperature recording were installed.

2.2 Data Description and Data Transformation

Since June 1st, 2016, the bridge monitoring is put into practice. The monitoring is carried out at a total of 16 measuring points for the crack width and three measuring points for the temperature. The measuring frequency is 0.5 Hz, so that the crack width and the temperature is recorded every two seconds. After each day, the collected data are stored in a *Microsoft Access Add-in* data file (.mdt). This results in a total data volume of 43 200 observations per measuring point and day and adds up to a file size of about 1.6 Megabyte per day. As we will analyze the data with R, we have to make the data accessible for it. Converting the MDT files to CSV files for using the standard R-functions is not practical for long time intervals because the data of a single day in CSV format add up to a size of 11.8 Megabyte.

For this reason, we wrote a new R-function to read the data from the MDT file. An MDT file consists of three parts: an ASCII-coded header with information about the data, the measurements stored as signed two-byte integers, and an ASCII-coded sentence at the end of the file. The header provides useful information, for example the day and the starting time of the measurements, but it also includes a table with, among other things, the name of the point of measurement and a numerical *offset* and *factor*. The offsets and factors are necessary to transform the two-byte integers of the crack width and temperature measurements to their real values. Let x be a two-byte integer. Then, the real value is obtained by $(x - \text{offset}) \cdot \text{factor}$. Thus, in R we can read the table in the header with the help of the function `read.table()` and the measurements with `readBin()` and after transforming the integers to their real values and adding some aesthetics like names, we obtain a `data.frame` with 20 columns (16 crack widths, three temperatures and one time column) and about 43 200 rows. If we store the obtained data in a RData-file, the file has a size of 1.0 Megabyte.

3 Statistical Methods

In the following, four methods for finding anomalous sequences are described. Therefore, time series $(X_t)_{t \in T}$ of crack width measurements are examined where $T = \{1, \dots, N\}$ is the time index set. For simplification the notation $X = (X_t)_{t \in T}$ is used.

All methods use a filtration in the sense of Brockwell and Davis [2]. It transforms the time series through a time-invariant but not necessarily linear filter $f: \mathbb{R}^{2w+1} \rightarrow \mathbb{R}$, $w \in \mathbb{N}$, given by

$$X_t^{(f,w)} = (X_t^{(f,w)})_{t \in T} = f(X_{t-w}, \dots, X_{t+w}). \quad (1)$$

Then $X^{(f,w)}$ is the time series filtered through f with window width $2w + 1$. In this article, the filters $f = \text{med}$ (median) and $f = \text{MAD}$ (Median Absolute Deviation from the median) are applied.

In particular, all methods use the filter $f = med$ for trend adjustment of X_t . The window width w varies from method to method. The trend-adjusted crack width process is then given by

$$\tilde{X}_t = X_t - X_t^{(med,w)}, \quad t \in T. \quad (2)$$

The calculations for all methods were performed with the statistical software *R* [15].

3.1 Location Change Point Detection using Squared Residuals (LCP)

The Location Change Point Detection using Squared Residuals (LCP) approach consists of several steps which we will describe in the following. We use Fig. 3 to illustrate them on the time series presented in Fig 2.

First, we calculate a trend-adjusted time series as in (2). The estimated trend function is depicted in Fig. 3 b) for the parameter $w = 75$. This choice of w is rather intuitive and not the result of a thorough optimization. In general, w should be chosen small enough so that the level of the time series is nearly constant and large enough so that the filter withstands a few large observations and to ensure a satisfying efficiency. In our case, it is only important that no relevant anomalous sequences are smoothed by the filter. The resulting trend-adjusted time series of our example is represented by the black line in Fig. 3 c) and 3 d).

To detect volatility changes, we transform the trend-adjusted time series by computing the logarithm of \tilde{X}_t^2 . For our example, the result is shown in Fig. 3 c). The volatility changes are visible as location shifts. In the trend-adjusted time series, about 33% of the values equal zero. That is why we add random noise from a uniform distribution to the observations before applying the logarithm. The range of the random numbers is by the factor ten smaller than the scale of the data.

Sequential two-sample tests in a moving time window can now be applied to check for a location shift at each time point t . This approach is, for example, described in [1]. At each time point $t \geq k$, we consider a time window $X_t = (X_{t-k+1}, \dots, X_t, \dots, X_{t+k})'$ of width $n = 2k + 1$ centred at t . We test for a location shift between t and $t + 1$ by splitting the window into two subwindows

$$\tilde{X}_{t-} = (\tilde{X}_{t-k+1}, \dots, \tilde{X}_t)' \quad \text{and} \quad \tilde{X}_{t+} = (\tilde{X}_{t+1}, \dots, \tilde{X}_{t+k})',$$

each of length k . The window width k controls the robustness against outliers and the power of the test. The latter can be improved by using a large window. However, it is crucial that the assumption of a constant underlying signal within the window can be justified. Using this local approach avoids making, possibly unrealistic, global parametric assumptions. Moreover, the methods are able to adapt to a slowly varying signal and lessen the impact of a possible variation due to the natural behaviour

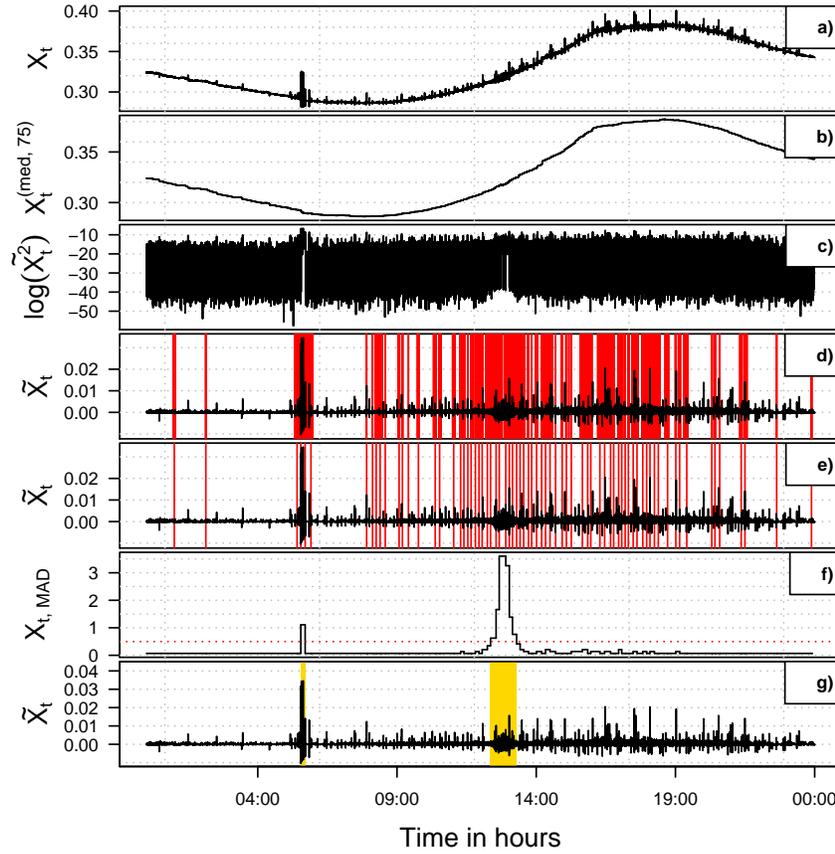


Fig. 3: Presentation of LCP method using the crack widths from June 5th, 2016, at measuring point WWS4. a) Original time series, b) estimated trend after median filtering, c) logarithm of squared residuals, d) trend-adjusted time series with change point candidates after applying asymptotical HL22-test (annotated in red), e) estimated change points after using method of [17], f) MAD within candidate intervals on standardized residual time series for anomalous sequences (red dotted line denotes the threshold $\gamma = 0.5$), g) parts that are detected as anomalous sequences (highlighted in yellow).

of the process [6] which may not have been completely eliminated by the trend adjustment.

Assuming that the underlying distributions in X_{t+} and X_{t-} differ at most in location allows for a comparison of both subwindows by a two-sample location test.

Generally, any two-sample location test could be used within this framework. In the following, we use an asymptotic test based on the robust and rather efficient

two-sample Hodges-Lehmann estimator for shift (HL2 estimator) which is given by

$$\hat{\Delta}_t = \text{med}\{\tilde{X}_{t+j} - \tilde{X}_{t+i} : j = 1, \dots, k, i = -k+1, \dots, 0\}$$

at time $t = 1, \dots, N$. The test statistic is

$$T_t^{(\text{HL2})} = \sqrt{3} \int g^2(x) dx \cdot \sqrt{n} \cdot \hat{\Delta}_t \quad (3)$$

with $n = 2 \cdot k$ and asymptotically follows the standard normal distribution. Here, g denotes the density of the distribution underlying the \tilde{X}_t under $H_0 : \Delta = 0$ [10]. Let h be the density of $\tilde{X}_{t+} - \tilde{X}_{t-}$ under H_0 . Then, it can be shown that $h(0) = \int f^2(x) dx$. Estimation of this value is possible by applying a kernel-density estimator applied to the set of pairwise differences between the elements of \tilde{X}_{t+} and \tilde{X}_{t-} [4].

We use an asymptotic test because some prior studies showed that large windows lead to better results on the herein analyzed data. In the simulation study in [4], the HL2-test turns out to be similarly powerful as the t -test, the standard method for two-sample location problems, while providing a strong protection against outliers.

We choose the rather small value $\alpha = 0.005$ for the significance level to ensure that the number of false alarms is not too large. We obtain about 15% of test decisions that lead to a rejection of H_0 . As Fig. 3 d) shows for $\alpha = 0.005$ by the vertical red lines, they are scattered all over the time range. This implies a large amount of false alarms.

The figure shows that many alarms occur subsequently. This is because of the moving-window nature of this approach. If an alarm is triggered at some time point t , it is likely that it will also be given at time points prior and after t since the samples are only different by a few observations. Hence, it is likely that the same cause is responsible for several subsequent alarms. To thin the number of alarms out, we use a method proposed by Wu and Chu [17] to estimate the true change point out of subsequent alarms. Each time point at which H_0 is rejected can be seen as a change point candidate. We determine the time index t^* which belongs to the smallest p -value among the set of candidate points. Within a neighborhood $\{t^* - k + 1, t^* - k + 2, \dots, t^* - 1, t^* + 1, \dots, t^* + k\}$ of t^* , we remove all candidate points. This process is repeated until no candidates are left. The remaining time points $t_1^* < t_2^* < \dots < t_m^*$ are estimators for the change points. The result for our example is depicted in Fig. 3 e), where we now have rejections at approximately 0.02% of the testing time points. However, they are still distributed over the complete time range.

Let now $\{t_1^*, t_1^* + 1, \dots, t_2^*\}$, $\{t_2^*, t_2^* + 1, \dots, t_3^*\}$, ..., $\{t_{m-1}^*, t_{m-1}^* + 1, \dots, t_m^*\}$ be the sets of time points between two estimated change points. To identify potential anomalous sequences, we estimate the variability for the observations in the trend-adjusted time series within each set. Except for t_1^* and t_m^* , each estimated change point can be the starting or the end point of a potential anomalous sequence under the assumption that we did not miss any relevant change point before t_1^* and after t_m^* . Hence, subsequent sets share their first and last element. We estimate the variability within each set by using the MAD on the globally standardized trend-adjusted observations. This means that we first subtract the mean and divide this difference by

the standard deviation, both calculated from all observations. We use a threshold approach with which we remove the sets for which the estimated variability is smaller than a value $\gamma > 0$. By standardizing, we want to make sure that a single threshold can be used for different time series. The MAD ensures that single large or small crack widths do not affect the estimation. Fig. 3 f) shows the results for the MAD and the threshold (horizontal line) of $\gamma = 0.5$.

In the last step, we combine connected sets for which the threshold is exceeded. By doing so, we assume that they belong to the same anomalous sequence. In Fig. 3 g), two sequences are identified, highlighted by a yellow background. The first is caused by a test load from a 48-ton mobile crane. The second is an anomalous sequence of interest.

3.2 Variance Change Point Detection (VCP)

Variance Change Point Detection (VCP) uses variance change point analysis to detect anomalous sequences. For detection of anomalous sequences, the trend adjusted time series \tilde{X}_t is used. Therefore, the difference of the original time series and the running median (with $w = 150$) of the time series is calculated (see Fig. 4 a)). On this transformed data set, changes in variance are detected using change point analysis with the method `cpt.var()` from the R package `changept` [14].

The R method `cpt.var()` uses the Pruned Exact Linear Time (PELT) method of Killick et al. (2012) [13] which is based on the Optimal Partitioning (OP) approach of Jackson et al. (2005) [12]. This approach is able to find several change points $1 \leq t_1^* < t_2^* < \dots < t_M^* < N$ in a time series $(Y_t)_{t \in \{1, \dots, N\}}$ where the number M of change points is estimated as well. It is based on a cost function \mathcal{C} with the property

$$\mathcal{C}(Y_{t_{m-1}:t_m}) + \mathcal{C}(Y_{(t_m+1):t_{m+1}}) = \mathcal{C}(Y_{t_{m-1}:t_{m+1}}) \quad (4)$$

if no change point exists in the interval $[t_{m-1}, t_{m+1}]$ and

$$\mathcal{C}(Y_{t_{m-1}:t_m}) + \mathcal{C}(Y_{(t_m+1):t_{m+1}}) < \mathcal{C}(Y_{t_{m-1}:t_{m+1}}) \quad (5)$$

if only one change point exists at $t_m \in [t_{m-1}, t_{m+1})$ so that the change point is lying in the interval $(t_m, t_m + 1]$. Thereby, $Y_{t_k:t_l}$ stands for $(Y_t)_{t \in \{t_k, t_k+1, \dots, t_l-1, t_l\}}$ for $t_k \leq t_l$, $t_k, t_l \in \{1, \dots, N\}$.

Here the adjusted time series \tilde{X}_t is used for Y_t so that we can assume that its mean is zero. Then an appropriate cost function for variance changes satisfying (4) and (5) is twice the negative log-likelihood function based on the maximum likelihood estimator for the variance with known mean equal to zero. To avoid too many change points, a term $\log(t_m - t_{m-1})$ is added, i.e. the cost function is given by

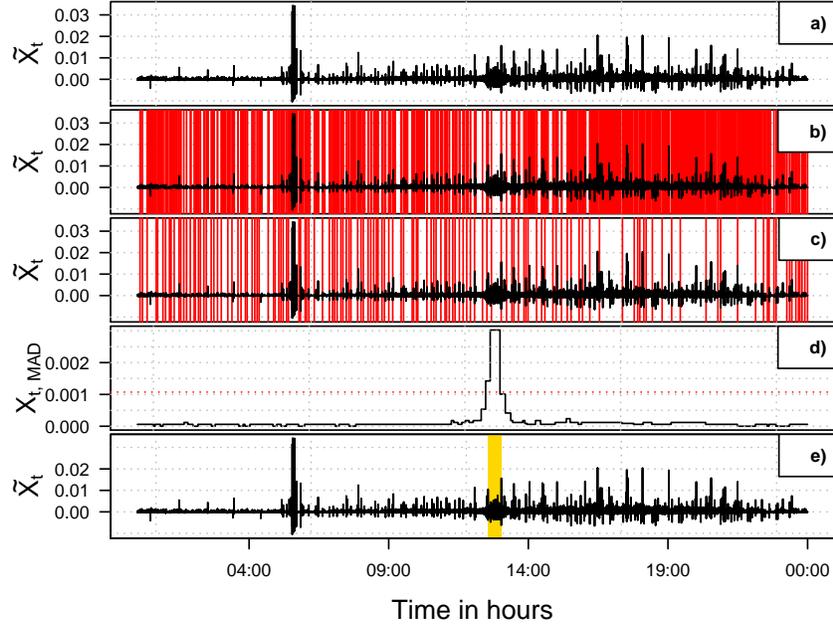


Fig. 4: Presentation of VCP method using the crack widths from June 5th, 2016, at measuring point WWS4. a) The trend adjusted time series \tilde{X}_t , b) \tilde{X}_t with change point detection after applying `cpt.var()` (annotated in red), c) \tilde{X}_t with change points that have a distance of at least 80 observations to the next change point, d) MAD between the change points from c) where the red dotted line marks the threshold between clusters, e) the part that is detected as anomalous sequence by VCP (highlighted in yellow).

$$\begin{aligned} & \mathcal{C}(Y_{(t_{m-1}+1):t_m}) \\ &= (t_m - t_{m-1}) \left(\log(2\pi) + \log \left(\frac{\sum_{j=t_{m-1}+1}^{t_m} (Y_j)^2}{t_m - t_{m-1}} \right) + 1 \right) + \log(t_m - t_{m-1}). \end{aligned} \quad (6)$$

Given a cost function \mathcal{C} , then the change points are determined as the vector $\tau_N^* := \tau_{N,M^*}^* := (t_0^*, t_1^*, t_2^*, \dots, t_{M^*}^*, t_{M^*+1}^*)$ which solve the following minimization problem

$$\min_{M=0,1,\dots,N-1} \min_{\tau_M \in \mathcal{T}_{N,M}} \left(M\beta + \sum_{m=1}^{M+1} \mathcal{C}(Y_{(t_{m-1}+1):t_m}) \right) \quad (7)$$

where β is a penalty and

$$\mathcal{T}_{N,M} := \{(t_0, t_1, t_2, \dots, t_M, t_{M+1}) \in \mathbb{N}^{M+2}; 0 = t_0 < t_1 < t_2 < \dots < t_M < t_{M+1} = N\}.$$

For calculating the solution of (7), the OP approach of Jackson et al. (2005) [12] uses the following recursion for $I = 1, \dots, N$

$$\begin{aligned} F(I) &:= \min_{M=0,1,\dots,I-1} \min_{\tau_{I,M} \in \mathcal{T}_{I,M}} \sum_{m=1}^{M+1} (\mathcal{C}(Y_{(t_{m-1}+1):t_m}) + \beta) \\ &= \min_{J=0,1,\dots,I-1} \min_{M=0,1,\dots,J-1} \min_{\tau_{J,M} \in \mathcal{T}_{J,M}} \left[\sum_{m=1}^{M+1} (\mathcal{C}(Y_{(t_{m-1}+1):t_m}) + \beta) + (\mathcal{C}(Y_{(J+1):I}) + \beta) \right] \\ &= \min_{J=0,1,\dots,I-1} [F(J) + (\mathcal{C}(Y_{(J+1):I}) + \beta)] \end{aligned}$$

with $F(0) := -\beta$. The idea for the OP algorithm is then to calculate recursively

$$F(I) = \min_{J=0,1,\dots,I-1} [F(J) + (\mathcal{C}(Y_{(J+1):I}) + \beta)] \quad (8)$$

and

$$J(I) := \arg \min_{J=0,1,\dots,I-1} [F(J) + (\mathcal{C}(Y_{(J+1):I}) + \beta)] \quad (9)$$

for $I = 1, \dots, N$ starting at $I = 1$ and to store all $F(1), \dots, F(N)$ and $J(1), \dots, J(N)$. Then backtracking is used to obtain

$$I_0 := N, \quad I_1 := J(N), \quad I_m := J(I_{m-1}) \quad \text{as long as } I_{m-1} > 0,$$

so that $\tau_N^* = (t_0^*, t_1^*, t_2^*, \dots, t_{M^*}^*, t_{M^*+1}^*) = (I_{M^*+1}, I_{M^*}, I_{M^*-1}, \dots, I_1, N)$ with $I_{M^*+1} = 0$ and $I_{M^*} > 0$ is a solution of (7). This requires $O(N^2)$ evaluations. The PELT algorithm reduces the number of evaluations to $O(N)$ so that it is linear. This is done by pruning the set $\{0, 1, \dots, I-1\}$ for calculating $F(I)$ and $J(I)$ in (8) and (9), respectively. The pruned set is given by

$$R_I := \{I-1\} \cup \{i \in R_{I-1} : F(i) + \mathcal{C}(Y_{(i+1):(I-1)}) < F(I-1)\}$$

with $R_1 = \{0\}$. Hence $J \in R_I$ is used instead of $J \in \{0, 1, \dots, I-1\}$ in (8) and (9). This pruning is possible since i with $F(i) + \mathcal{C}(Y_{(i+1):(I-1)}) \geq F(I-1)$ cannot be the last change point due to (4) and (5). That this pruning reduces the set $\{0, 1, \dots, I-1\}$ drastically is not so easy to see, but the proof is given in [13].

For the VCP method, `cpt.var()` is used with its default values. Thereby, the penalty β is the modified Bayes Information Criterion given by $3 \log(N)$. Moreover, a step length of 2 is used as default so that the results for a reverted time series differ from the non-reverted time series although the PELT algorithm should find the global minimum.

With this method, nearly every change in variance of the crack width is labeled as change point (see Fig. 4 b)). Thereby also changes of the variability within an anomalous sequence are labelled as change point. Moreover, changes in crack width triggered by traffic are labelled as change points as well.

To separate changes in crack width triggered by traffic from anomalous sequences and to reduce the change points within anomalous sequences, the MAD is calculated for observations between adjacent change points. All MADs which are based on at least 80 observations are clustered by k-means into two clusters, i.e. in a cluster with high MAD and in a cluster with small MAD. Fig. 4 c) shows only those changepoints that are used for clustering and in part d) the corresponding MADs are pictured. If connected sequences of the data are classified to the cluster with high MAD then they are joined to one sequence and this sequence is then a detected anomalous sequence. Then an anomalous sequence is followed by a nonanomalous sequence and this sequence is followed by anomalous sequence and so on. Sequences with less than 80 observations and high MAD are not considered as anomalous sequence here since the high MAD may be caused by the traffic. The anomalous sequence located by this procedure is shown in Fig. 4 e).

A modification of the above method is to apply the k-means clustering on all MADs, also on those with less than 80 observations. However, then the cluster with high MADs contains mainly very small sequences and these sequences are mainly triggered by traffic. If afterwards the sequences with less than 80 observations are removed then often no anomalous sequences are found. Hence, this modification is not a good alternative.

3.3 Modified Variance Change Point Detection (MVCP)

The method `cpt.var()` from the `changepoint`-package [14] detects changes in variance at different points in the time series if the reversed time series is used. This is due to the dependence on the past in the change point identification process which is caused by the fact that `cpt.var()` uses a step length of 2 by default so that shifts can appear in the detected anomalous sequences. Therefore, the Modified Variance Change Point Detection (MVCP) method uses the VCP method on the original time series and on the reversed time series. Only those observations which are identified as anomalous sequences in both runs get labelled as anomalous sequences by MVCP.

3.4 Clustering of MAD filtered data (CMAD)

At first a trend adjustment is conducted on X_t . Therefore the filter $f = med$ with window width $w = 15$ is applied on the time series X_t . The trend adjusted crack width process is shown in Fig. 5 a).

For the further process, it is assumed that anomalous sequences certainly exist in adequate large time intervals. This was confirmed in preliminary studies.

The next step is based on the assumption that anomalous sequences are marked by a locally increased variability. Application of the filter $f = MAD$ with window

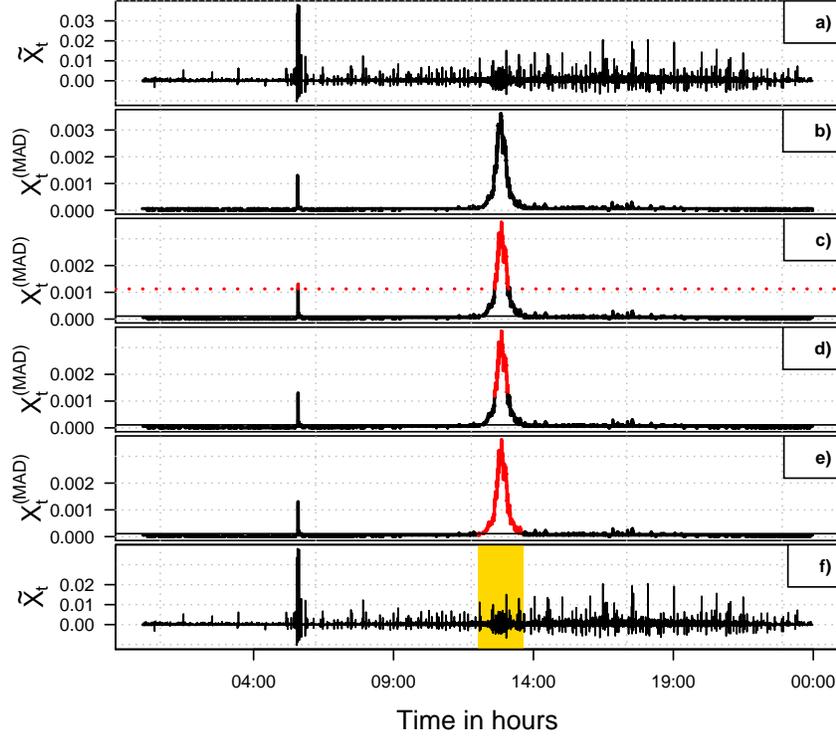


Fig. 5: Presentation of CMAD method using the crack widths from June 5th, 2016, at measuring point WWS4. a) The trend adjusted time series \tilde{X}_t , b) $X_t^{(MAD)}$, the MAD smoothed, trend adjusted time series, c) $X_t^{(MAD)}$ with red marked candidates where $P_t = 1$, d) $X_t^{(MAD)}$ with red marked candidates where $P_t^{(med)} = 1$, e) $X_t^{(MAD)}$ with extension of sequences where $P_t^{(med)} = 1$, f) \tilde{X}_t with anomalous sequence (highlighted in yellow).

width 91 on the trend adjusted time series \tilde{X}_t , i.e.

$$X_t^{(MAD)} = \tilde{X}_t^{(MAD,45)}, \quad t \in T, \quad (10)$$

allows the detection of anomalous sequences by finding clusters of higher variability. In Fig. 5 b) peaks are visible at 06:00 o'clock and 12:00 o'clock.

After that, all observations of the MAD smoothed time series $X_t^{(MAD)}$ are divided into two clusters by the k -means algorithm. Because of the target variable being one dimensional and k being 2, a threshold is the result so that every observation above this threshold is classified as a candidate for an anomalous sequence. A time series $P_t = (P_t)_{t \in T}$ is constructed, whereby $P_t = 1$ means that $X_t^{(MAD)}$ is above the thresh-

old and $P_t = 0$ stands for a value below this threshold. The candidates for anomalous sequences are shown in Fig. 5 c) where $X_t^{(MAD)}$ is shown with red marked candidates for anomalous sequences based on P_t .

To prevent special occurrences like the test load being classified as anomalous sequences, the filter $f = med$ with window width 301 is applied on P_t . The resulting time series is denoted $P_t^{(med)}$. In Fig. 5 d), $X_t^{(MAD)}$ is represented with red marked candidates for anomalous sequences where $P_t^{(med)} = 1$. The peak at 06:00 o'clock is no longer classified as an anomalous sequence.

Furthermore, it is apparent that start and end of an anomalous sequence are not included in the cluster. Hence, the boundary points are examined. These points form the transition between a normal and an anomalous sequence. So the boundary points mark the beginning and ending of an anomalous sequence, but are not classified as one yet. For this, the average $\overline{X_t^{(MAD)}}$ of $X_t^{(MAD)}$ is calculated. The values of the boundary points are compared with this value. If they are larger than $\overline{X_t^{(MAD)}}$, the points are assigned to the candidates for the anomalous sequence. Then the new boundary points are examined iteratively. This process is stopped as soon as the values of the boundary points are smaller or equal to $\overline{X_t^{(MAD)}}$ and hence the maximal extension of the anomalous sequence is reached. This is shown in Fig. 5 e).

In Fig. 5 f), \tilde{X}_t is pictured with anomalous sequences underlaid by yellow boxes.

4 Comparison of the Methods

In this section, we investigate the performance of the different approaches on selected time series. The available data sets do not contain information about which sequences fit our definition of anomalous. A categorization was performed manually for a limited number of time series. Thus, a thorough parameter tuning is not possible. Therefore, the following results should only be considered as a descriptive performance study. We use the parameter settings mentioned in the corresponding subsections in Section 3.

A major difference between the procedures is that, in contrast to VCP, MVCP, and CMAD, the LCP method is based on the idea of a sequential online monitoring. It does not use the complete time series for detecting the change points. Moreover, the threshold to identify intervals with a large variability is a fixed value for LCP, whereas the other methods choose it for each data set separately due to clustering on the complete time series.

As illustrated by Figs. 3, 4, and 5 in Section 3, the approaches do not necessarily identify the same sequences in a time series as being anomalous. For the time series used in Section 3, only LCP categorizes the short sequence of large variability at around 6:30 a.m. which was caused by a heavy test load using a mobile crane. Therefore, we do not consider this short peak caused by traffic as an anomalous sequence of interest. The methods VCP, MVCP, and CMAD are not identifying this sequence so that, in this case, they are able to distinguish between short variance

changes caused by traffic from large variance changes where the cause is unknown. The reason for this is that VCP, MVCP, and CMAD, unlike LCP, use an assumption on the minimal length of an anomalous sequence.

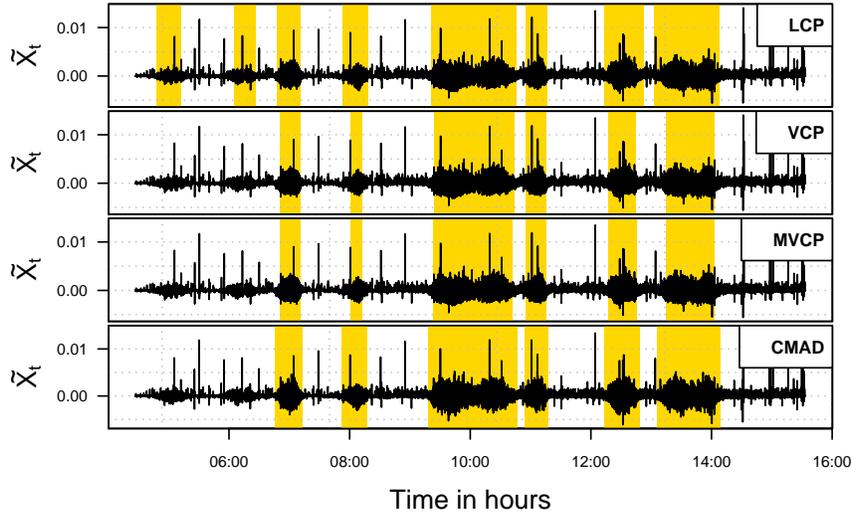


Fig. 6: Comparison of the anomalous sequences detected by the four methods (highlighted in yellow) on June 12th, 2016, of WOS4. The figure shows the trend-adjusted time series for each method.

To point out another difference, we look at the measurements from June 12th, 2016, at WOS4. Figure 6 shows the trend-adjusted time series for each procedure. VCP, MVCP, and CMAD return the same six sequences. The sequences returned by VCP and MVCP are similar and shorter than those returned by CMAD. LCP also returns these six sequences with widths comparable to those of CMAD. In addition, LCP categorizes two sequences, starting at around 4 o'clock and 6 o'clock, as anomalies. A closer inspection of these time intervals shows that the variability of the process is slightly increased compared to the time points before and after. This variability is smaller than for the other returned sequences. Thus, because of the clustering, VCP, MVCP, and CMAD are able to adapt to the specific variability of a time series so that they lead to more reasonable results.

The two aforementioned highlighted regions for LCP do not cover the complete sequences with slightly increased variability. Again, this is because the variability in the sequence next to the detected ones is not larger than the threshold. Thus, they are dropped before the combination step.

We now compare the procedures on three other time series with respect to their error rates. The measurements were taken on August 5th, 2016, at WOS2, January

1st, 2017, at WOS2, and February 22nd, 2017, at WWS4. The error rates are computed by comparing the returned sequences to a manually performed categorization. As it is sometimes difficult for the human eye to detect regions with a higher variability, two persons classified the time series into anomalous and non-anomalous sequences. Both decided for each time point whether the corresponding observation belongs to an anomalous sequence. The results for each time point are captured in binary vectors called Counting 1 and Counting 2. Similar vectors are obtained by the different methods. Each counting is compared with each result vector. The error rate is defined as the fraction of elements that are different. Table 1 shows the rates and Figure 7 illustrates them by a bar plot.

Table 1: Error rates of the four methods compared with Counting 1 and 2 and between Counting 1 and 2.

Day	Error rates between Methods and Countings								Error rates between Counting 1 and 2
	LCP		VCP		MVCP		CMAD		
	1	2	1	2	1	2	1	2	
5th Aug 2016	0.009	0.014	0.010	0.007	0.010	0.007	0.002	0.006	0.0063
1st Jan 2017	0.004	0.005	0.008	0.008	0.008	0.008	0.006	0.006	0.0003
22nd Feb 2017	0.099	0.155	0.070	0.079	0.095	0.048	0.077	0.046	0.1084

The last column in Table 1 quantifies differences between the two manual countings. The large value of about 10% for February 22nd, 2017, indicates that the classification in anomalous and non-anomalous sequences can be ambiguous. Here, the time series has in general a larger variability which makes it harder to distinguish between normal process behavior and anomalous sequences.

For January 1st, the differences between the procedures are negligible. Some larger differences can be noted for August 5th. Here, the smallest error rate of 0.2% belongs to CMAD on Counting 1 while LCP leads to the largest rate of 1.4%. For the time series from February 22nd, the rates are, compared to those for the other two time series, quite large for all procedures. Differences are mostly visible for Counting 2. LCP has the largest value with 15.5%, whereas MVCP and CMAD perform similarly with about 5%.

There are several reasons for these observations. The countings are, as they were performed by human beings, subjective to a certain degree. This can be seen from the differences between them. Thus, we could not compare our results to the true sequences but only to those that we assume to be plausible. The differences between the countings illustrate that it is not easy to identify anomalous sequences on some time series, for example due to a generally larger variability in the complete series.

Another reason is that the methods may detect the same sequences, but their lengths are different. This is the case for August 5th and January 1st. Only on February 22nd, LCP discovers more anomalous sequences than the other procedures.

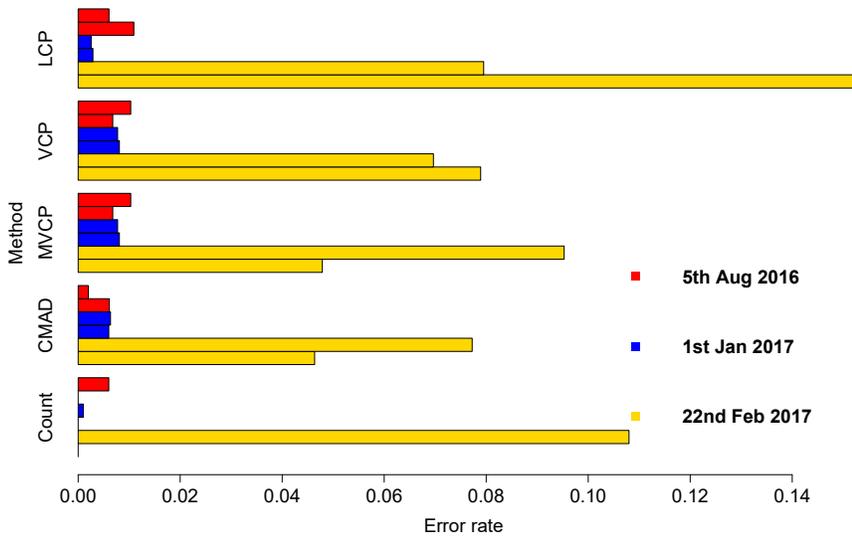


Fig. 7: Comparison of the error rates of the four methods for three different days with respect to both countings and the error rate between Counting 1 and 2 (denoted by "Count"). The upper bar of a day is the result of the comparison with Counting 1 and the lower one is the result of the comparison with Counting 2.

To sum up, in general, the procedures lead to quite similar results when the anomalous sequences express themselves as intervals in which the variability is largely different from the general process behavior. Differences seem to arise when the variability in the process is generally large so that it might be more difficult to identify the anomalous sequences.

However, the results should be treated with caution. In order to perform a thorough parameter tuning, more of the available data should be used. This was not done up to now because more manually analysed data would be necessary.

Nevertheless, the obtained results look promising and further refinement of the procedures, for example adding a threshold for the minimal sequence length in the LCP method, might lead to a severe improvement.

5 Discussion

A monitoring of cracks in a bridge provided many time series of crack width data. The final aim is to decide whether there is a significant increase of the crack widths.

This is not an easy task since the time series show a strongly varying behavior. Some variation is caused by temperature and traffic. In addition, there are so called anomalous sequences with much more extreme volatility for which the cause is unknown. Therefore, a further analysis of these time series is only possible if the anomalous sequences are removed. We present four methods to detect them automatically.

The LCP (Location Change Point Detection using Squared Residuals) method is based on applying a two-sample location test in a moving time window to a transformed time series. VCP (Variance Change Point Detection) and MVCP (Modified Variance Change Point Detection) are based on the PELT (Pruned Exact Linear Time) method. CMAD (Clustering of MAD filtered data) uses a classification approach.

We evaluate the procedures on a small set of test time series, all four procedures lead to similar results. Differences occur especially in the length of the sequences. Moreover, VCP, MVCP, and CMAD, using an assumption on the minimal length of an anomalous sequence, are able to distinguish between short sequences which may be caused by traffic and long sequences where the cause is unknown. Hence these methods, unlike LCP, do not identify short sequences as anomalous sequences.

When the variability in the time series is generally large, it turns out to be difficult to identify anomalous sequences. Again, LCP turns out to provide more sequences than the other procedures. However, this seems to be primary due to the chosen parameters. LCP does not use knowledge on the minimal length of a sequence. Moreover, categorizing sequences into classes, one with large and one with small variability, is done by a fixed threshold by LCP. The other approaches use a clustering algorithm on each time series and thus consider the variability of each time series individually. This may be advantageous since one does not have to specify a threshold value globally. However, it may be prone to errors if there are time series without anomalous sequences. On the other hand, a fixed value needs proper tuning over several time series with a similar variability structure.

The presented analysis concentrates on a few days only. This is mainly because only measurements for three days were visually inspected by two persons on one of the 16 measurement locations on the bridge. Thus, conclusions drawn from our analysis should be treated with caution. However, they show that it is worth to study the procedures further in a larger setting.

In further studies, the methods should be applied to a larger set of time series. This would help to make reliable statements on their advantages and disadvantages. An important aspect is a proper parameter tuning. Moreover, the influence of the method for change point detection should be investigated.

Although we apply the procedures retrospectively, that is under the assumption that all data are available, in practice it may be important to analyze the time series online and detect unusual sequences with only little time delay. Here, it would be interesting how the procedures can be adapted to this online scenario. For example, LCP is based on sequential two-sample testing that can easily be used in the online analysis.

Acknowledgments

The authors gratefully acknowledge support from the Collaborative Research Center "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823, B5, C3) of the German Research Foundation (DFG).

References

1. Abbas, S., Fried, R.: Control charts for the mean based on robust two-sample tests. *Journal of Statistical Computation and Simulation* 87(1), 138 - 155 (2017)
2. Brockwell, P.J., Davis, R.A.: *Time Series: Theory and Methods*. Second Edition, Springer, New York (2006)
3. Bundesministerium für Verkehr, Bau und Stadtentwicklung: *Richtlinie zur Nachrechnung von Straßenbrücken im Bestand (Nachrechnungsrichtlinie)*. Berlin (2011)
4. Fried, R., Dehling, H.: Robust nonparametric tests for the two-sample location problem. *Statistical Methods and Applications*, 20(4), 409 - 422 (2011)
5. Fried, R.: On the robust detection of edges in time series filtering. *Computational Statistics and Data Analysis*, 52(2), 1063 - 1074 (2007)
6. Fried, R., Gather, U.: On rank tests for shift detection in time series. *Computational Statistics and Data Analysis*, 52(1), 221 - 233 (2007)
7. Fried, R.: On the online estimation of local constant volatilities. *Computational Statistics and Data Analysis*, 56(11), 30803090 (2012)
8. Heeke, G., Heinrich, J., Maurer, R., Müller, C.H.: Neue Erkenntnisse zur Ermüdungsfestigkeit und Prognose der Lebensdauer von einbetonierten Spannstählen bei sehr hohen Lastwechselzahlen. In *Tagungshandbuch 2016, 2. Brückenkolloquium*, Technische Akademie Esslingen, Stuttgart, 529-539 (2016)
9. Heinrich, J., Maurer, R., Hermann, S., Ickstadt, K., Müller, C.: Resistance of Prestressed Concrete Structures to Fatigue in Domain of Endurance Limit. In: *Proceedings of fib Symposium Maastricht, High Tech Concrete: Where Technology and Engineering Meet, 12th - 14th June 2017*, (2017)
10. Hodges, J.L., Lehmann, E.L.: Estimates of location based on rank tests. *The Annals of Mathematical Statistics*, 34(2), 598 - 611 (1963)
11. Høyland, A.: Robustness of the Hodges-Lehmann estimates for shift. *The Annals of Mathematical Statistics*, 36(1), 174 - 197 (1965)
12. Jackson, B., Scargle, J. D., Barnes, J., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., Tsai, T. T.: An Algorithm for Optimal Partitioning of Data on an Interval. *IEEE Signal Processing Letters*, 12 (2), 105-108 (2005)
13. Killick, R., Fearnhead, P., Eckley, I.A.: Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:500, 1590-1598, (2012), DOI: 10.1080/01621459.2012.737745
14. Killick, R., Heynes, K. und Eckley, I. A. (2016): changepoint: An R package for changepoint analysis, R package version 2.2.2, <https://CRAN.R-project.org/package=changepoint>
15. R Development Core Team: *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, <http://cran.r-project.org/> (2017)
16. Szugat, S., Heinrich, J., Maurer, R., Müller, C.H.: Prediction intervals for the failure time of prestressed concrete beams. *Advances in Materials Science and Engineering*, (2016) DOI:10.1155/2016/9605450
17. Wu, J.S., Chu, C.K.: Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics*, 21(3), 1545 - 1566 (1993)