

Calibration of Two Novel Segmentation Approaches by Synthetic Road Vehicle Vibrations

László Róbert HÁRI*, Péter FÖLDESI

Abstract: Simulation of non-stationary random vibrations has motivated Packaging vibration testing for decades. Often, an event-detection algorithm decomposes Road vehicle vibrations when analyzing the recorded series. However, heuristics and subjective justifications are often in the papers, whereby the foremost concern is the validation of the non-stationarity of simulated signals. Furthermore, if a changepoint detection is inherent to the procedure, it is recommended to calibrate the detector. The current paper concerns the Receiver operating characteristics (ROC) of two novel algorithms and provides contextual support by Segment length distributions (SLD).

Keywords: packaging vibration testing; receiver operating characteristic; road vehicle vibration; segmentation; spectrogram

1 INTRODUCTION

It is widely acknowledged that Road vehicle vibrations (RVV) are non-stationary, hence non-Gaussian in nature. Thus, Packaging vibration testing (PVT) introduced various efforts to dissect RVV into homogenous sections, so the yielded segments could be simulated separately. Following the concatenation of synthesized parts, a total non-stationary vibration signal is obtained-overcoming the limitation of testing via stationary signals. Other techniques may modulate one series of stationary Gaussian signals on various periods to obtain non-stationarity. The following paragraph briefly introduces the main segmentation methods, commenting on the corresponding evaluation form. The bar " | " denotes borders of ideas. Moving statistics for detection tasks offer numerically cheap computations, though the window length and detection threshold should be justified. Possible predictors can be moving mean, RMS, crest factor, or kurtosis [1]. Ibid. was further developed to extract transients from RVV; after that, RMS variations were algorithmically looked for [2]. Ideally, the robustness of subjective parameters is confronted. Bayesian changepoint detection is a promising step toward objectivity under probabilistic environments, e.g., as presented on road profile data [3]. A similar approach might be used for the severely finer sampled vertical accelerations. Critiques on the stationarity of a single Power spectral density (PSD) may be addressed by providing a number of PSD, as presented by the Split spectra method [4]. However, cut-offs separating PSD into low- and high RMS groups are somewhat arbitrary. Speculatively, it may be worth asserting the thresholds according to an individual products'fatigue limits. As an extension, more thresholds are idealized by Probability split spectra (an unpublished work had been cited in the source, see Ref. 4 in [4]: Root, Dale "Transportation Severity Report, Leonardo Project", Lansmont Corporation, 2008. The reference has not been found here). The PSD level is accounted for each frequency to "represent the probability that an encountered PSD level will be at or below the profile based on all data events recorded". Supra-threshold procedures are also applied for shock and vibration environments of railcars [5]; in the same paper, alternatively, filtering is proposed for separating the rigid-body motion from structural high-frequency bursts. While a cut-frequency seems

helpful in the presentation, the applicability to road vehicle vibration-instead of rail-must undergo thorough testing. For similar purposes, Hilbert-Huang transform and Empirical mode decomposition are introduced to railcar vibrations [6], whereby the number of Intrinsic mode functions (IMF) can demarcate rigid-body- and structural vibrations, as exemplified in plots. Road roughness measurements with artificial data manipulation (for presentation purposes) are analyzed within the framework of wavelet theory [7]. The main contribution is the introduction of wavelets' detection capability to PVT-and in those terms, it is very successful. Wavelet-based Gaussian decomposition iteratively decomposes an RVV into Gaussian components by complex wavelets in the translation-scale (thus, in time-frequency-) domain [8]. Scuff damage experiments were presented as a unique validation of the simulation. Cumulative sum (CUSUM) schemes have gained particular attention, which a form of bootstrapping can enhance. Subsequent presentations follow the segmentation of the instantaneous magnitude of RVV [9]. Machine learning classifiers also appeared in PVT. The applications may use several predictors simultaneously, such as moving RMS, crest factor, kurtosis, DWT, and HHT [10]. Special considerations also by ROC curves and the purpose-built "Pseudo-energy ratio/fall-out curve" are presented. One can find validation examples not only by simulated but also by measured signals [11]. In summary, an ever-expanding and heterogeneous toolset is available for now. Still, only a few works introduced a form of the detection theory's verification. The following discourse presents the evaluation of a CUSUM-type changepoint detection (CpD) and the Hypothesis-based spectrogram segmentation (3S), the latter complemented by Bonferroni correction (3SB). If 3S is concerned with or without Bonferroni correction, the 3S (B) notion is used. Also, Bonferroni and post hoc correction are used interchangeably throughout the paper. Section 2 discusses the applied algorithms, and Section 3 introduces the common ground of evaluations-hence a two-fold derivation is unnecessary. Section 4 illustrates the results again per subsection as of application, and Section 5 discusses them similarly in a dual manner. The general conclusions of the last section conclude the paper. Albeit the Nomenclature consists of frequent abbreviations, the reader is advised to note the subtle difference between changepoints, Cp, and condition positives, CP.

2 COMPARED DETECTORS

2.1 CUSUM-type Changepoint Detection (CpD)

The CpD approach-realized after [12] is applied to the first four spectral moments of STFT, denoted by μ_{ij} for $i = [1; 2; 3; 4]$ and $j = 1, 2, 3, \dots, J$ seconds of the time-frequency domain. The four series of changepoints (Cp) are unified, dissecting a spectrogram at specified instances. The CpD uses the cumulative sum of differences between the sample element and the total mean, such as

$$S_{ij} = \sum_{j=1}^J (\mu_{ij} - \bar{\mu}_{ij}) \quad (1)$$

A sudden change in the direction of S_{ij} at $\hat{S}_{ij} = \max_j |S_{ij}|$ indicates a possible change in the mean tendency. At each candidate Cp, a type of bootstrapping can evaluate the significance of the test statistic. Here, a permutation loop is implemented for $r = 1, 2, 3, \dots, N$ rearrangements, such that $S_{ij} \rightarrow R_{ikr}$, j not necessarily equalling k per r . The (resampled) test-statistic \hat{R}_{ikr} is accounted at each CP. Simply put, $\tilde{p} = 0$ confidence is voted for \hat{S}_{ij} before resampling, each time counterevidence is found, \tilde{p} increases with $1/N$. Formally, the approximated p -value is:

$$\tilde{p} = \frac{\#\left[\hat{R} \mid \hat{R} > \hat{S}\right]}{N} \quad (2)$$

where $\#$ denotes the number of elements. If \tilde{p} stays smaller than a threshold, the Cp is considered significant, and the series is bisected. Then, each sub-section is introduced to the same procedure until no more Cp is found, a minimum segment length is reached, or exit conditions of an infinite while loop are encountered.

2.2 Hypothesis-based Spectrogram Segmentation (3S)

In previous works, consecutive DFT profiles of the STFT had been segmented by significant changes according to (one-sample) paired t -tests [13] or (unpaired) two-sample t -tests [14]. The STFT spectrogram was transformed here by \log_{10} operation to bring the normality-assumptions closer. The paired t -test among consecutive spectrums j and $j + 1$ is used here on reasons discussed in Sec.5.1. For x and y arbitrary paired vectors, the corresponding hypothesis can be formulated, such as:

H_0 : data in $x - y$ has a mean equal to 0,

H_A : data in $x - y$ does not have a mean equal to 0.

The frequency axis ensures the pairing between neighbour DFT vectors. A Cp is inserted if there is enough evidence against H_0 . However, inflating type I. error is accountable in multiple hypothesis testing; thus, post hoc correction might compensate for the inflation.

2.3 Hypothesis-based Spectrogram Segmentation with Bonferroni Correction (3SB)

Familywise error rate, α_f , increases in multiple hypothesis testing, i.e., an increasing probability of at least one type I. error is accountable. The Bonferroni correction expresses for m comparisons conducted at the same preliminary significance level, $\alpha_{(0)}$, that the probability of no α -error for the overall family of m tests is [15]: $\alpha_f = 1 - [1 - \alpha_{(0)}]^m$, where m equals the number of comparisons, i.e., the number of spectrums minus one. With post hoc correction, individual decisions are overruled by the new

$$\alpha_B = \alpha_{(0)} / m \quad (3)$$

Bonferroni significance limit.

3 METHODS OF COMPARISON

The investigations resemble the same logic, such as 3.1 the establishment of samples, 3.2 derivations of technological windows and the choice of Points of Interest (POI), 3.3 the Receiver operating characteristics (ROC), and 3.4 the Segment length distributions (SLD).

3.1 Establishment of Samples

Promising results are obtained by ongoing research on the synthesis of RVV, simulated directly in the time-frequency domain. The method simulates statistically well-aligned DFT profiles per second of segments, and in each segment, a different profile is chosen. The data-driven approach relies on real-world RVV; thus, it is understood as a mixture of artificial and natural RVV. Noteworthy, however, those changepoints (Cp) are known in advance; therefore, the condition positives (CP) are also known. Two ensembles are used, not being on view for reasons of space. The first ensemble of four series (A, B, C, D) with different prevalence is subjected to CpD firsthand. Sample A is depicted in Fig. 1. The same ensemble is used later in Section 4.3, confronting CpD and 3S methods. The second ensemble is from three series (E, F, G) with a high prevalence for 3S (B) procedures. Furthermore, each of the three samples is trisected to supply similar prevalence but on different lengths, J .

3.2 Operation Surfaces and Points of Interest (POI)

Eq. (2) expresses that a given number of permutations is necessary (at least as extreme as the un-resampled test-statistic) to consider a candidate Cp insignificant. This way, the concept of Significance reserve can be introduced.

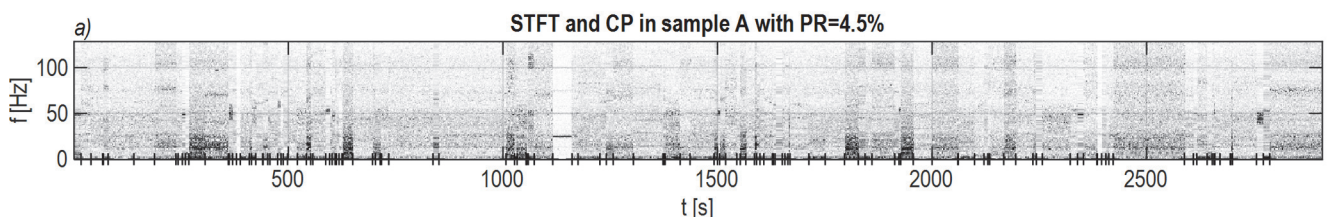


Figure 1 Short-time Fourier transform of the sample, A, changepoints, i.e., condition positives indicated by bars, on the horizontal axes

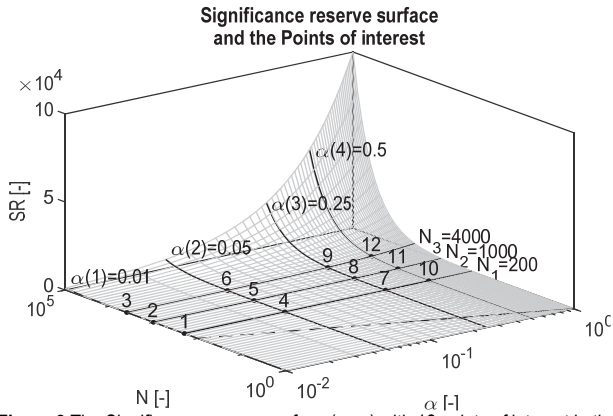


Figure 2 The Significance reserve surface (grey) with 12 points of interest in the cross-sections of the iso-permutation curves $N = \{200; 1000; 4000\}$ and iso-significance limits $\alpha = \{0,01; 0,05; 0,25; 0,50\}$. Each point is above the minimum number of permutations (dashed)

ROC snapshots of CPD at 12 POI

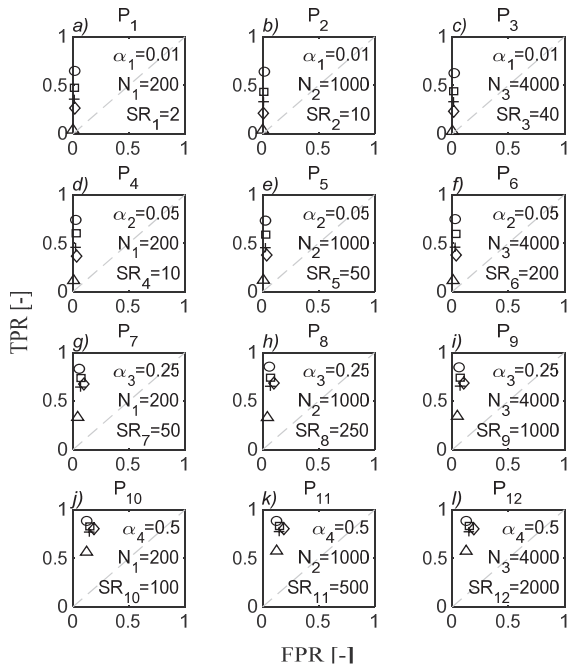
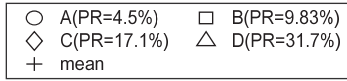


Figure 3 ROC snapshots of the CUSUM-type changepoints detection (CPD) on four different samples (denoted by symbols) with different prevalence at 12 POI. Rows of panes correspond to iso-significance limits, α ; columns of panes relate to iso-permutations, N . Note that the significance reserve, SR , varies per POI

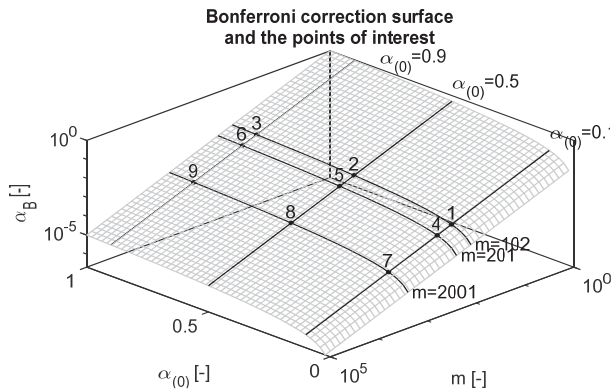


Figure 4 The Bonferroni correction surface (grey) along the nine points of interest in the cross-sections of the iso-curves corresponding to the number of hypotheses to test $m = \{102; 201; 2001\}$ and preliminary iso-significance limits $\alpha_{(0)} = \{0,10; 0,50; 0,90\}$. Samples E, F, G are tested at each POI

ROC snapshots of 3S(B) procedures at 9 POI

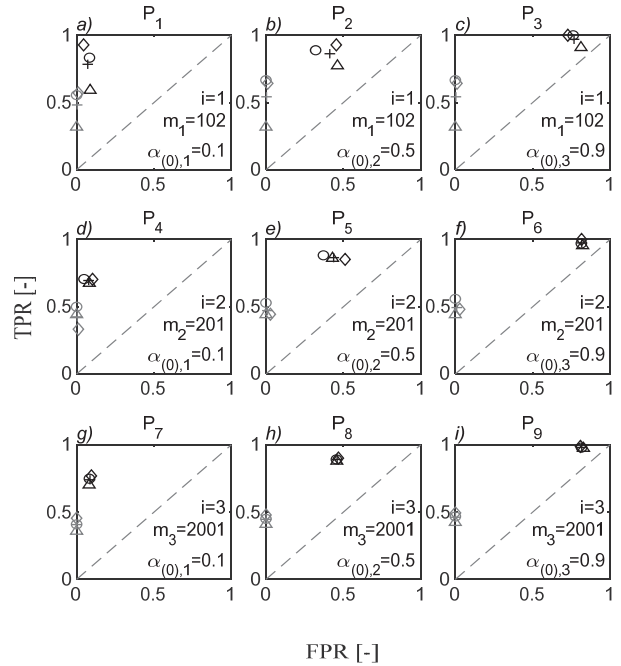
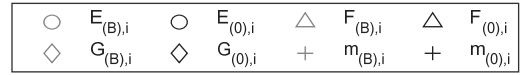


Figure 5 ROC snapshots of Hypothesis-based segmentation without Bonferroni correction (3S) denoted by black symbols and (0) subscripts; and 3S with Bonferroni correction (3SB) denoted by grey symbols and (p) subscripts on three different samples (E, F, G) at 9 POI. Rows of panes correspond to iso-significance limits, α ; columns of panes relate to number of spectrums to test

SLD of CpD by Samples of samples A, B, C, D

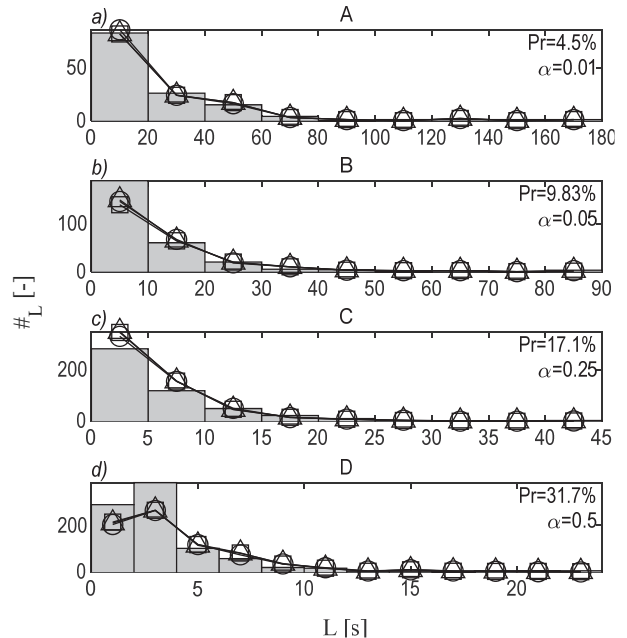
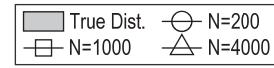


Figure 6 Segment length count-distributions from the CUSUM-type changepoints detection (CPD) on four samples. Bars denote the true distributions; lines correspond to the number of permutations. Panes correspond to different samples, each tested on iso-significance limits at different permutations. Note how different permutations yielded overlapping lines

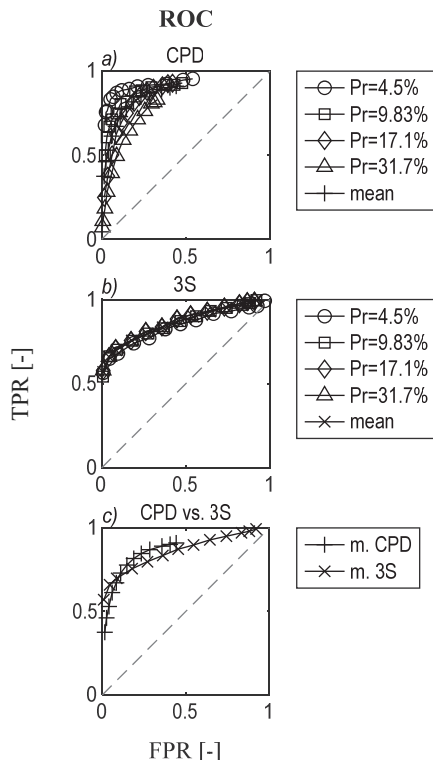


Figure 7 Comparison of CUSUM- and Hypothesis-based segmentation by receiver operating characteristics on four samples of (A, B, C, D) with different prevalence (Pr). Pane a) indicates ROC of CPD and the mean of statistics (+); pane b) shows ROC of 3D for the same samples, with mean statistics denoted by the symbol ×. Finally, pane c) repeats the mean ROC statistics overlaid.

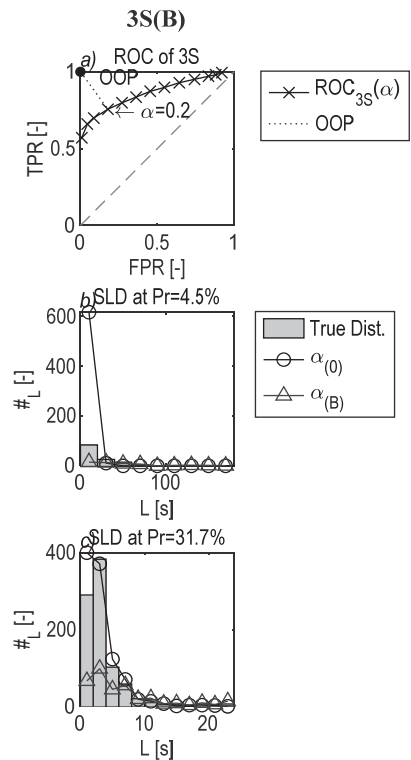


Figure 9 Comparison of segment length distributions by hypothesis-based segmentation without- and with Bonferroni correction, 3S (circles) and 3SB (triangles), respectively. Pane a) illustrates the derived optimal operating point (OOP) closest to [0; 1] coordinate from Fig. 7c) by crosses, yielding $\alpha = 0, 10$. Pane b) shows the true distribution with bars and corresponding SLD from 3S and 3SB at optimal α for the low- prevalence sample, A; and pane c) similarly for the high- prevalence sample, D. Note that OOP is derived from ROC of 3S, whereby an OOP of 3SB is not applicable according to Fig. 7.

SLD of 3S by samples E_i, F_i, G_i

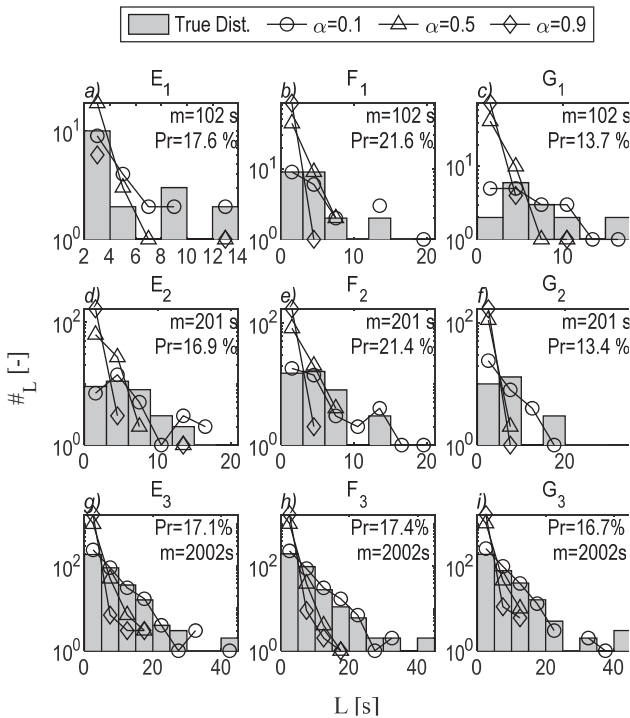


Figure 8 Segment length count-distributions from the Hypothesis-based segmentation (3S) on nine subsamples. Bars denote the true distributions; lines correspond to preliminary significance limits. Columns of panes correspond to primary samples (E, F, G), and rows of panes relate to the i -th fraction of main samples (E_i, F_i, G_i).

Definition. The significance reserve (SR) denotes the number of permutations necessary to consider a candidate changepoint non-significant in a resampling scheme when a rearrangement of $x \rightarrow r$ yields at least as extreme test-statistic, \hat{r} as the un-rearranged sample, \hat{x} , for $n = 1, 2, \dots, N$ resampling. The fraction of α/N significance portion is accumulated for $\tilde{p} = 0$, each time $\hat{r} > \hat{x}$. In other words, the significance reserve describes a reserve of staying significant for a candidate changepoint. While the rearrangement of elements is a matter of randomness, the closest parameters of influence are α and N , which can be expressed as depicting the Significance reserve surface (SRS), illustrated in Fig. 2. The POI avoid the lower limit of $SRS = 1$, also, $N = 4000$ permutations constitute an upper limit due to computational practicality.

$$SRS = \frac{\alpha}{1/N} \quad (4)$$

In the case of Bonferroni correction, Eq. (3) is illustrated in Fig. 4. Special considerations worth mentioning. The number of hypotheses to test is not always a matter of choice; here, the length of samples skeletonizes different scenarios. Also, few investigations are available on the number of Cp in RVV, while the number of methodologies arises. At last, $\alpha_{(0)} > 0,10$ significance limits are atypical, but this paper explores the operation from large perspectives due to the novelty of 3S(B).

3.3 Receiver Operating Characteristic

Consider a population with a priori prevalence, PR, of an arbitrary condition. The condition may be the real presence of disease (condition positives, CP) among healthy individuals (condition negatives, CN). In time-series analysis, e.g., shocks can be present in a random vibration signal, as in Fig. 10a). Let us suppose an RVV is recorded, and a particular statistic is obtained per second. Let us further assume that we could precisely discern shocks (CP) from stationary Gaussian random vibrations (CN) possessing sufficiently good values of the statistic. Additionally, we aim to investigate different thresholds for the given statistic, by which we could distinguish shocks (S) from shock-free vibrations (V) in future observations. A binary logistic model is fitted to 100 shock- and 100 vibration statistics data in the same pane.

$$PR = CP / (CP + CN) \tag{6}$$

Thus, if then 50% prevalence is present in the example. Let the rhombus symbol denote our first guess of a threshold in the same plot. Then, most shocks are correctly considered positive answers above the threshold, contributing to true positives (TP). On the other hand, shocks below the threshold are incorrectly classified as vibration; hence, false negatives (FN) are obtained.

Receiver operating characteristic from an exemplary Binary logistic model

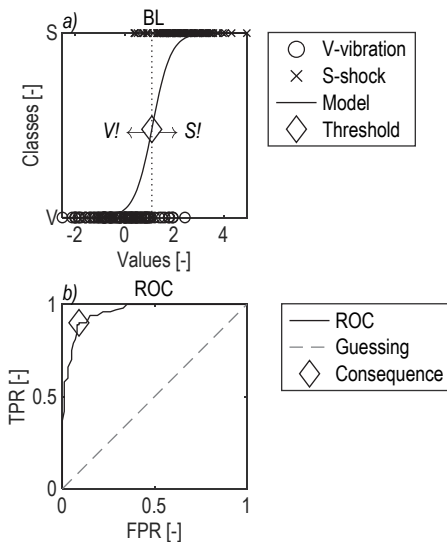


Figure 10 An exemplary classification task. Pane a) binary logistic model (BL) of vibration (V) and shock (S) events with corresponding values and an arbitrary threshold denoted by rhombus. Events with values lower than the threshold are categorized as vibration (V!); conversely, supra threshold values are considered shocks (S!). Pane b) receiver operating characteristic as of thresholds, also highlighting the consequence of the initial choice. The ROC curve suggests that the BL model is better than guessing here

Table 1 Confusion matrix for a binary response

		Prediction		Metrics
		DP	DN	
Reality	CP	TP	FN	TPR = TP / CP; FNR = FN / CP
	CN	FP	TN	

DP: positive by decision; DN: negative by decision;
FNR: false negative rate; TNR: true negative rate.

Table 2 Prevalence according to samples (A, B, C) per the $i = 1, 2, 3$ subsamples and corresponding points of interest per row for the hypothesis-based segmentation

i	E	F	G	POI
1	Pr(E_1) = 17,65%	Pr(F_1) = 21,57%	Pr(G_1) = 13,73%	1;2;3
2	Pr(E_2) = 16,92%	Pr(F_2) = 21,39%	Pr(G_2) = 13,43%	4;5;6
3	Pr(E_3) = 17,08%	Pr(F_3) = 17,43%	Pr(G_3) = 16,68%	7;8;9

Typically, the control group is investigated similarly. Vibration values below the threshold are correctly classified as vibration constituting true negatives (TN), and vibrations above the cut-off are falsely classified as a shock, yielding false positives (FP). It is in a common interest to evaluate the true positive rate, $TPR = TPR(t)$ being a function of threshold, t ; in other terms, the *sensitivity*, such that which is "the probability that a person will test positive for a disease, given the person actually has the disease" [16] p.484. It may be translated as "the probability that a classifier will classify a pattern as a target when it really is a target" [17] p. 378.

$$TPR = TP / CP \tag{7}$$

In the low base rate problem, with the rate corresponding to a condition or behaviour, tests designed to discern rare conditions can find TP beneath an unexpectedly large number of FP. Thus, it is wise to measure the true negative rate, $TNR = TNR(t)$, namely specificity, which is "the probability that a person will test negative for a disease, given the person actually does not have the disease" [16]; in other terms "the probability that a classifier will correctly classify the true nontarget cases" [17]. It is commonly expressed as $TNR = TN / CN$, and it occurs that the false positive rate $FPR = 1 - TNR$, or directly denoting the ratio of false positives among the (truly) condition negatives. Tab. 1 summarises the confusion matrix and frequent statistics necessary here. The ROC curve for the introductory examples can be seen in Fig. 10b.

$$FPR = FP / CN \tag{8}$$

3.4 Segment Length Distribution

The series of C_p indicate the borders of segments in the STFT at given instances denoted by " | " symbols at $f = 0$ Hz in Fig. 1. The C_p is considered simultaneously present in every frequency bin. The segment length is in the centre of interest of the horizontal distance between two C_p , which is investigated on count-based Segment length distributions (SLD).

4 RESULTS

4.1 Evaluation of Change-point Detection

The C_pD is run on the four samples of different prevalence classes, evaluated at the 12 POI on the SRS in Fig. 2. The ROC statistics are here snapshots of the ROC curve in Fig. 3. The General observations are:

1. the prevalence has much influence at low α ; with increasing PR, the effect decreases.
2. low PR tends to yield higher TPR; conversely, high PR can deliver low TPR.

3. the number of permutations produced minor differences at iso-significance settings; see rows of Fig. 3.
4. it is hard to approach FPR regions of 50%; see the last row of Fig. 3.

Segment length distributions in Fig. 6 directly show the number of segments, #, in the bins. It supports that

5. the number of resampling, N , produced few differences in the obtained distribution.

Thus, further investigations concentrate on PR and α . For reasons of space, only such PR & POI pairs are plotted in Fig. 6, which showed proper alignments. For the same reason, it is just deductively introduced that:

6. false positives contribute to over-segmentation, and
 7. false negatives contribute to under-segmentation.
- Furthermore, Fig. 6 suggests that.
8. there exists an α threshold for each PR, minimizing the erroneous statistical inferences.

4.2 Evaluation of Hypothesis-based Segmentation

The 3S(B) is applied to the three samples E, F, G aimed to possess a similar prevalence of $\sim 17\%$. Each sample is divided into subseries $i = [1, 2, 3]$ of length $\sim \{100, 200, 300\}$ sec corresponds to nine POI. As expected, the subdivision changed the sample-wise PR summarised in Tab. 2 with the corresponding POI in Fig. 4. Also, Fig. 5 shows the 3S method without correction (black) and with post hoc correction (grey). Similarly, the ROC snapshots of 3S(B) allow the following observations according to:

9. low $\alpha_{(0)}$ yielded low FPR and medium to high TPR,
 10. medium $\alpha_{(0)}$ yielded medium FPR and high TPR,
 11. high $\alpha_{(0)}$ yielded high FPR and high TPR,
- which is a good justification for the expected ROC.

Regarding sample sizes,

12. the longer the sample, the more rigorous the TPR & FPR estimations are.

With post hoc setup in Fig. 5 (grey symbols),

13. minimal FPR is registered, the mean TPR being dispersed around 0.5.

Fig. 8 compares the count-based SLD after 3S, each pane illustrating a given subsample with three different preliminary significance limits. It is concluded that:

14. high thresholds are likely to yield over-segmentation,
15. low α is the best guess in general, given the current ensemble,

which success must be accredited to the paired t -test. Results of 3SB in the probability domain are omitted here; further explanations are given in the Discussion. The following subsection directly compares CpD and 3S.

4.3 Competitive Comparison of Methods CpD and 3S

So far, the behaviour of CpD, 3S, and 3SB has been explored, but a direct comparison could not be conducted without the above observations. The following experiences are echoed here:

- the concerned CpD realization does not necessarily require immense iterations per candidate Cp,
- the sample length is indifferent to significance limits in CpD and 3S-not so in 3SB,

- Bonferroni correction may be overlooked, and directly 3S results can also be helpful,
- short recordings with high Pr were necessary for 3S for demonstration purposes. (Note that the long sample-low prevalence setting operation could not be shown.)

Thus, CpD and 3S are rerun on samples A, B, C, D with $N = 100$ replications in case of CpD; and without post hoc correction of 3S. In both cases, mutual significance levels are applied ($\alpha = \{0.01; 0.05; 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 0.99\}$). The ROC characteristics are displayed in Fig. 7 by the four different PR classes. Therein, pane a) confirms that an increase in PR can flatten the ROC curve of CpD. Remarkably, CpD can hardly pass the FPR = 0.5 limit with increasing α , which must be accredited to the exit conditions. Pane b), however, depicts 3S as quasi prevalence-independent, which could not reach below TPR = 0.5 in current samples. However, this satisfactory performance is not a general conclusion, as Fig. 5a) black rhombus instantly falsifies that.

5 DISCUSSIONS

In brief, it was aimed to find all real changepoints, TP, beneath not-being oversensitive, i.e., with consistency on TN. It is shown that FP often corrupts these aims leading to over-segmentation. When false alarms occur, the obtained distribution can overestimate short segments and underestimate long ones. In contrary terms, the modest sensitivity can increase FN, which overlooking of real changes can result in under-segmentation; thus, SLD might be under-represented in lower bins and over-estimated in larger bins.

5.1 Arguments for Detection Methods

The number of resampling applied to the given samples did not yield dramatically different results, suggesting that CpD can be robust enough also at a moderate number of permutations. However, the analyst is advised that the recursion can suffer a severe failure when an FN is incurred in one of the first p -value approximations. Then, simply due to randomness, a prominent changepoint is left out, the loop exits, and a dramatical under-segmentation is obtained. Let us suppose that the prevalence is precisely 50%, i.e., every 2-nd sec. is different from neighboring ones. Then, in the hypothetical spectrogram of [a; b; a; b; a; ...], with DFT vectors sufficiently different a and b, the CpD may encounter that the absolute of CUSUM of a spectral moment does not have a maximum. Above 50% prevalence, however, one must consider that (a) either transients are occasionally present in an underlying slowly varying process, or (b) mainly successive transients describe the process. While (a) is more straightforward to imagine under everyday circumstances, such as traveling on roads and occasionally meeting some potholes, setting (b) can arise from driving on off-road tracks, for example. It is expected that 3S can overcome such limitations incurred in RVV during the latter case (b). 3S (B) methods rely on the one-sample t -test here, which omits a data-driven resampling; thus, it can be performed under every prevalence scenario. It shall be appreciated that the one-sample t -test derives its

test-statistic still from two samples: precisely, the differences between two paired samples. Therefore, the differences between two DFT profiles along the whole bandwidth of the spectrum contribute to the estimation. The two-sample (or unpaired) t -test accounts for the two samples differently. Let the following minimum-example use DFT vectors of $a = [1; 1; 2]$ and $b = [2; 1; 1]$ ordered as of $[f_1; f_2; f_3]$ frequency components, depicting a frequency modulation from a to b . It is apparent that the vectors a & b are the "same" for the two-sample t -test, which does not take the sequence-hence such a frequency modulation-into account. Instead, the two-sample t -test is better to examine overall magnitude modulations by the scenario of, e.g., increasing background noise such as from $a = [1; 1; 2]$ to $b = [1; 2; 2]$. Despite the Bonferroni correction being a mature method, contributing to a whole branch of statistics, its application remained more or less chaotic, e.g., as pointed out in the ophthalmic discipline [18]. Some thoughts of *ibid.* will be borrowed here and put into the current context. Foremost to highlight here the subtle narrative, the use of Bonferroni correction sometimes depends on the intentions-though not the wishes of the analyst. For example, if it is aimed to assess a universal null hypothesis that all tests are non-significant, i.e., there is no change (at all) in the RVV, then Bonferroni correction can improve the reliance of the general H_0 . Similarly, if it is imperative to avoid a type I. error, that is, important costs can be attributed to inserting a changepoint, then applying the post hoc correction can also be imperative. But as shown in Fig. 5, CP can be left out, as well. Conversely, no correction had been suggested in exploratory studies concerning the search for *any* significant difference; or if the individual tests' results are important. The numerical investigations presented in Fig. 7. also clearly support "reducing the chance of a type I error but at the expense of a type II. error" (*ibid.*). At this point, it seems that the Bonferroni correction does mitigate type I. errors, but at consequences of increasing false negatives; hence, it cannot provide SLD reasonably well-aligned. Still, the post hoc test may be beneficially implemented in general null testing.

5.2 On the Receiver Operating Characteristics

ROC curves are mainly used on two occasions: comparing classifiers and deriving a classifiers' optimal operating point (OOP). Numerous proposals have been made so far on OOP [19]; also, an OOP shall be justified in the given context. The interested reader is referred to [20] for examples of justifications. The simple "closest to (0, 1) criteria" is presented here, referring to the point closest to the top-left corner on the ROC plot. Formally,

$d_E = \sqrt{\text{FPR}^2 + (1 - \text{TPR})^2}$ denotes the Euclidean distances between the ROC curve and the (0; 1) coordinate. It is found in Fig. 9a) that the point of (0.09; 0.72) relates to the minimum d_E corresponding to $\alpha = 0.10$ significance limit in 3S-according to the given rudimental curve. The corresponding SLD suggest, that a low prevalence sample on pane b) could not be adequately modelled. However, high prevalence sample on pane c) depicts the reasonable well-alignment of 3S; also, the over-conservatism of the Bonferroni correction is presented in either case, b-c).

6 CONCLUSIONS

3S(B) and CpD rely on established methods. 3S(B) are comparatively faster than CpD. 3S(B) are easier to implement. If CpD exits suspiciously early, it contributes to under-segmentation. 3S can be run on the contrary. The Bonferroni correction again proved to be too conservative. Current challenge points toward the estimation of prevalence. In a low base rate problem, CpD is suggested and can be suited for high PR classes. High prevalence samples can be beneficially analyzed also by 3S if resampling schemes are not favourable. It is generally advised to strive for robustness analysis of detectors in changepoint-indexing tasks.

7 REFERENCES

- [1] Bruscella, B., Rouillard, V., & Sek, M. (1999). Analysis of road surface profiles. *Journal of Transportation Engineering*, 125(1), 55-59. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1999\)125:1\(55\)](https://doi.org/10.1061/(ASCE)0733-947X(1999)125:1(55))
- [2] Bruscella, B. (1997). Analysis and simulation of the spectral and statistical properties of road roughness for package performance testing. *Master of Engineering in Mechanical Engineering*. Victoria University of Technology.
- [3] Thomas, F. (2005). Automated road segmentation using a Bayesian algorithm. *Journal of Transportation Engineering*, 131(8), 591-598. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2005\)131:8\(591\)](https://doi.org/10.1061/(ASCE)0733-947X(2005)131:8(591))
- [4] Kipp, W. I. (2008). Random vibration testing of packaged-products: Considerations for methodology improvement. Bangkok, Thailand, 1-12.
- [5] Rouillard V. & Richmond, R. (2007). A novel approach to analysing and simulating railcar shock and vibrations. *Packaging Technology and Science*, 20(1), 17-26. <https://doi.org/10.1002/pts.739>
- [6] Rouillard V. & Sek, M. A. (2005). The use of intrinsic mode functions to characterize shock and vibration in the distribution environment. *Packaging Technology and Science*, 18(1), 39-51. <https://doi.org/10.1002/pts.677>
- [7] Wei, L., Fwa, T. F., & Zhe, Z. (2005). Wavelet analysis and interpretation of road roughness. *Journal of Transportation Engineering*, 131(2), 120-130. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2005\)131:2\(120\)](https://doi.org/10.1061/(ASCE)0733-947X(2005)131:2(120))
- [8] Griffiths, K. R., Hicks, B. J., Keogh, P. S., & Shires, D. (2016). Wavelet analysis to decompose a vibration simulation signal to improve pre-distribution testing of packaging. *Mechanical Systems and Signal Processing*, 76-77, 780-795. <https://doi.org/10.1016/j.ymssp.2015.12.035>
- [9] Rouillard, V. (2007). On the statistical distribution of stationary segment lengths of road vehicles vibrations. *Proceedings of the World Congress on Engineering*, 549-550. https://doi.org/10.1007/978-1-4020-6239-1_272
- [10] Lepine, J. & Rouillard, V. (2018). Evaluation of shock detection algorithm for road vehicle vibration analysis. *Vibration*, 1(2), 220-238. <https://doi.org/10.3390/vibration1020016>
- [11] Lepine, J., Rouillard, V., & Sek, M. (2019). Evaluation of machine learning algorithms for detection of road induced shocks buried in vehicle vibration signals. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, 233(4), 935-947. <https://doi.org/10.1177/0954407018756201>
- [12] Taylor A. W. (2000). *Change-point analysis: A powerful new tool for detecting changes*. Taylor Enterprises, Inc.
- [13] Hári, L. R. (2021). Segmentation of a road vehicle vibration signal using multiple comparison procedures between paired

- samples. *Tehnicki vjesnik - Technical Gazette*, 28(5), 1597-1604. <https://doi.org/10.17559/TV-20200122112551>
- [14] Hári, L. R. (2021). Multiple hypothesis testing by unpaired samples for indexing changepoints in a road-induced vibration signal. *U.P.B. Sci. Bull.*, 83(3), 17-28.
- [15] Kim, H.-Y. (2015). Statistical notes for clinical researchers: post-hoc multiple comparisons. *Restorative Dentistry & Endodontics*, 40(2), 172-176. <https://doi.org/10.5395/rde.2015.40.2.172>
- [16] Salkind, N. J. (2010). *Encyclopedia of research design*. SAGE Publications, Thousand Oaks, Calif. <https://doi.org/10.4135/9781412961288>
- [17] Martinez W. L. & Martinez, A. R. (2015). *Computational Statistics Handbook with MATLAB, 3rd Edition*. New York: Chapman and Hall/CRC.
- [18] Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5), 502-508. <https://doi.org/10.1111/opo.12131>
- [19] Unal, I. (2017). Defining an optimal cut-point value in ROC Analysis: An alternative approach. *Computational and Mathematical Methods in Medicine*, 3762651, 1-14. <https://doi.org/10.1155/2017/3762651>
- [20] Aggarwal, R. & Ranganathan, P. (2018). Understanding diagnostic tests - Part 3: Receiver operating characteristic curves. *Perspectives in Clinical Research*, 9(3), 145-148. https://doi.org/10.4103/picr.PICR_87_18

FREQUENT ABBREVIATIONS

3S	Statistical spectrogram segmentation
3SB	3S with Bonferroni correction
3S(B)	3S with or without Bonferroni correction
Cp	Changepoint
CP	Condition positive
CpD	Changepoint detection
CN	Condition negatives
FN	False negative
FP	False positive
FPR	False positive rate
POI	Point of interest
PR	Prevalence
ROC	Receiver operating characteristics
SLD	Segment length distribution
SRS	Significance reserve surface
TN	True negative
TP	True positive
TPR	True positive rate

Contact information:

László Róbert HÁRI, MSc.
(Corresponding author)
Department of Mechanical Engineering and Energetics,
University of Dunaújváros,
Táncsics M. u. 1/A, H-2400Dunaújváros, Hungary
E-mail: harilaszl@uniduna.hu

Péter FÖLDESI, Prof. Dr.
Department of Logistics and Forwarding, University of Győr,
Széchenyi István University,
Egyetem tér 1., H-9026 Győr, Hungary
E-mail: foldesi@sze.hu