A Bayesian binary algorithm for RMS-based acoustic signal segmentation

Paulo Hubert,[1] Rebecca Killick,[2] Alexandra Chung,[1] and Linilson R. Padovese[1]
[1] *Department of Mechanical Engineering, Escola Politecnica - University of São Paulo, São Paulo, SP - Brazil*
[2] *Mathematics and Statistics Department, Lancaster University, Lancaster, UK*

(Dated: 25 June 2019)

Changepoint analysis (also known as segmentation analysis) aims to analyze an ordered, one-dimensional vector, in order to find locations where some characteristic of the data changes. Many models and algorithms have been studied under this theme, including models for changes in mean and / or variance, changes in linear regression parameters, etc. In this work, we are interested in an algorithm for the segmentation of long duration acoustic signals; the segmentation is based on the change of the RMS power of the signal. We investigate a Bayesian model with two possible parameterizations, and propose a binary algorithm in two versions, using non-informative or informative priors. We apply our algorithm to the segmentation of annotated acoustic signals from the Alcatrazes marine preservation park in Brazil.

 [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–10

I. INTRODUCTION

The problem of signal segmentation arises in different contexts[1–5]. The problem is broadly defined as follows: given a discretely sampled signal $y \in \Re^N$, divide it in contiguous sections that are internally homogeneous with respect to some characteristic. The segmentation is based on the premise that the signal structure changes one or many times during the sampled period, and one is looking for the times where the changes occur, i.e., the *changepoints*.

In this work we are interested in segmenting acoustic signals, more specifically underwater acoustic signals acquired off the Brazilian coast. Since 2010, the Acoustics and Environment Laboratory (LACMAM) at University of São Paulo has been designing equipment for underwater acoustic monitoring[6]; and over the past few years, we have acquired and stored over 2 years of acoustic recordings taken from different locations, amounting to more than 35 Tb of data.

The main challenge in exploring these data lies on the abundance of interesting events, and at the same time on the sparsity of such events. The sparsity of events makes the direct inspection of long duration signals a very demanding task, while the variety of potentially interesting events discourages the design and application of detection algorithms aimed at specific events, for they would potentially miss many unexpected (and for this exact reason, interesting) events.

Our approach is based on the hypothesis that the occurrence of an event induces an immediate change on the total sound pressure level, and that this change can be detected on the variance of the signal's amplitude. What we seek then is a **variance changepoint detection algorithm**.

A few algorithms to detect changes in signal's variance are available; in the next section we give a quick review on the signal segmentation and changepoint analysis literature. After that, section II defines the algorithm to be used for the segmentation; section III presents our results in the segmentation of both simulated and real acoustic signals, and section IV concludes the paper.

A. Changepoint analysis and signal segmentation

An interesting review on the signal segmentation analysis can be found in Theodorou *et al*[5]. The algorithms described by him have a few features in common:

1. The use of a more or less detailed parametric model to describe the signal;

2. The definition of frames, or windows, to characterize local behavior;

3. A peak detection or thresholding procedure applied to the collection of frames to obtain segments' boundaries.

These methods are well suited for the analysis of short to medium term signals (up to a few thousand data points), because the estimation step for the parametric models, be it a discrete Fourier or wavelet transform, and / or a filtering procedure, is usually computationally intensive. Also, the use of a detailed parametric model is adequate only when the additional structure imposed by the model over the original signal is well justified, i.e., when the phenomena causing the change in the signal's characteristics is reasonably well known.

Recent literature, however, proposes solutions for the problem that do not rely on detailed parametric models for the signal. Jackson[10], for instance, provides a general method based on dynamic programming that is able to find the global optimum of a fitness function, $V(P) = \sum g(B_m)$, where the sum is taken over m blocks, and g is the fitness function of a single block (usually a likelihood based on a probabilistic model), in $O(N^2)$ time.

In the same spirit, Killick *et al*[11] improve the work of Jackson by proposing a *Pruned Exact Linear Time* (PELT) algorithm that is able to optimize the global fitness function with complexity $\mathcal{O}(n)$ under reasonable conditions. Killick's method is general, and can be applied to any fitness function as long as it fulfills a mild condition on the relation between the fitness of an entire segment and the fitness of the same segment divided by one changepoint (for details, see the original paper[11]).

To the best of our knowledge, PELT is currently the state-of-the-art algorithm for signal segmentation. It suffers, however, from overfitting problems when applied to the analysis of long term signals with few changepoints. This overfitting also increases the computing time and memory requirements, since overfitting implies more changepoints to be tested and stores.

In this paper we propose a new, Bayesian binary algorithm, that is competitive when compared to PELT in the segmentation of short / medium size signals, but works better in long signals. Our algorithm approaches the problem of segmentation as one of sequential hypothesis testing. We adopt a binary strategy, first finding the best changepoint for the entire signal, and, if this changepoint is accepted, applying the procedure recursively to each segment obtained. In the next section, we define our model and the Bayesian binary algorithm.

II. A BAYESIAN ALGORITHM FOR VARIANCE CHANGEPOINT DETECTION

We start by assuming that the (discretely sampled) signal at time t, $y_t \in \Re$, has 0 mean amplitude for all t, and finite power σ_t^2. We adopt a Gaussian probabilistic model for the signal, $y_t \sim \mathcal{N}(0, \sigma_t^2)$.

We will assume that σ_t^2 is a piecewise constant function on t, and we are interested in estimating the localization of discontinuities or jumps in this function.

This is a very general signal model, which fits the main goal of our algorithm: to allow efficient analysis of long term signals, searching for sections that are likely to contain an event, regardless of the specific characteristics of the event.

If we do not consider the specific nature of the acoustical event, the best we can say about the signal after the event starts is that the total RMS power must increase (except if signal and noise are correlated, which we assume is not the case). Looking for changes in the signal's power is thus the most general segmentation model we can assume.

A. Binary algorithms

One of the simplest ways to tackle the changepoint location task is by using a binary algorithm. Given the entire signal, the first part of the algorithm looks for the single changepoint that is most likely or best in some sense. After obtaining this changepoint, the traditional binary approach will apply the same procedure recursively to the newly obtained segments. The stopping condition is usually based on a model selection criteria[12]. Figure 1 illustrates the binary segmentation process; in the figure, dashed vertical lines indicate the candidate changepoint at each step.

Our algorithm differs from the traditional binary strategy in the choice of the statistical hypothesis testing procedure to be applied at each step to decide if a given changepoint is valid (i.e., if there is enough evidence in the data that there is indeed a change at this point). After applying the procedure, and if the changepoint is considered valid, the algorithm continues to estimate new changepoints in the two segments obtained from the last iteration. If not, the execution is halted.

The binary segmentation algorithm is then based on a single changepoint model defined as follows:

$$y_t \sim \begin{cases} \mathcal{N}(0, \sigma_0^2) & \text{if } t \leq \bar{t} \\ \mathcal{N}(0, \sigma_1^2) & \text{if } t > \bar{t} \end{cases} \tag{1}$$

In this model, we use a Gaussian distribution for the signal, and we assume that the signal's variance (associated to the RMS power) changes abruptly when $t = \bar{t}$.

The first step of the algorithm involves estimating $\bar{t}$; this is done using Bayesian methods. We start by defining the log-likelihood function associated with the model

$$l(\bar{t}, \sigma_0^2, \sigma_1^2 | y) = -\frac{\bar{t}}{2} log\left(2\pi\sigma_0^2\right) - \frac{N-\bar{t}}{2} log\left(2\pi\sigma_1^2\right) - \frac{\sum_{t=1}^{\bar{t}} y_t^2}{2\sigma_0^2} - \frac{\sum_{t=\bar{t}+1}^{N} y_t^2}{2\sigma_1^2} \tag{2}$$

The likelihood function connects the data (signal) with our model; multiplying the likelihood by the prior, we arrive at the (unnormalized) **posterior** distribution for $\bar{t}$, i.e., the probability distribution for the parameter after seeing the data. In our model, this posterior will depend on $\bar{t}$, but also on σ_0 and σ_1 (the signal's variances before and after the changepoint). However, in estimating $\bar{t}$, these values are not important, i.e., they are *nuisance* parameters (we do not care what are the values of the variances, since we are just trying to estimate the moment at which they change). We thus eliminate this parameters from the posterior distribution, obtaining the marginal posterior of $\bar{t}$. To do that, we adopt Jeffreys' priors[13,14] for both σ_0 and σ_1, and integrate them out. This can be done analitically, yielding the marginal pos-

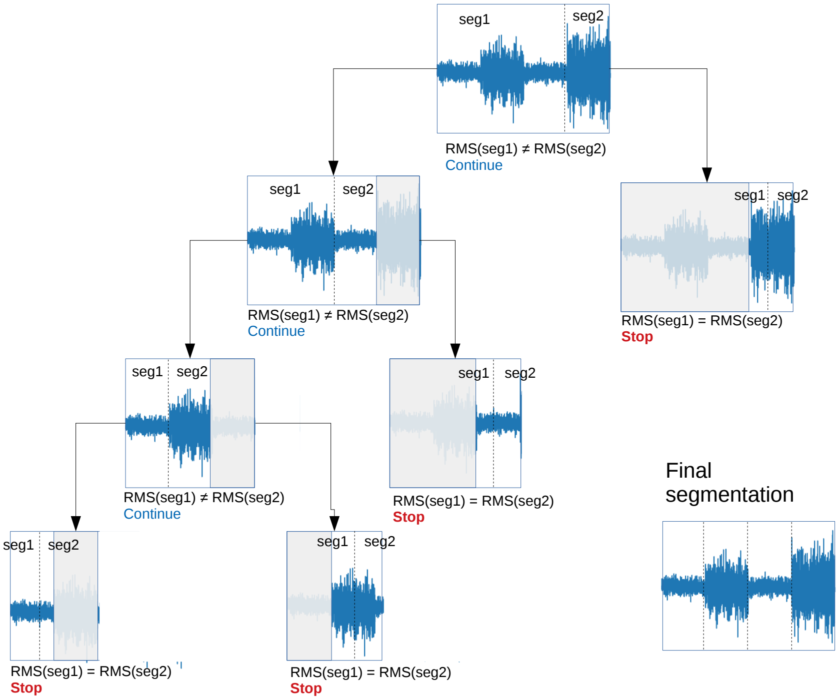

FIG. 1. Illustration of a binary segmentation approach

terior

$$P(\bar{t}|y) \propto \pi(\bar{t}) \cdot \left(\sum_{t=1}^{\bar{t}} y_t^2\right)^{-\frac{\bar{t}}{2}} \left(\sum_{t=\bar{t}+1}^{N} y_t^2\right)^{-\frac{(N-\bar{t})}{2}} \times \Gamma\left(\frac{\bar{t}}{2}\right)\Gamma\left(\frac{N-\bar{t}}{2}\right) \quad (3)$$

With this marginal posterior, the algorithm now must estimate the best unique changepoint for the current segment. We use the Maximum Posterior (MAP) estimate for the changepoint, i.e., we choose the candidate changepoint as the value of $\bar{t}$ that maximizes 3.

After estimating the current changepoint candidate, the algorithm must test its validity of the current changepoint. This will be done by using a full Bayesian test designed for precise (sharp) hypothesis. In the first author's PhD thesis[15] (in portuguese), different testing procedures are analyzed and compared, including Snedecor F's test, and a generalized likelihood procedure. The best results were obtained by using the Bayesian procedure we describe next.

B. Full Bayesian evidence measure

To be a valid changepoint, in the present context, means that the signal variances of the two segments are different. So this step requires an equality of variances test.

From the full model's likelihood 2, conditioning on $\bar{t}$ and multiplying by the joint prior on (σ_0, σ_1) yields the posterior

$$P(\sigma_0, \sigma_1|y, \bar{t}) \propto \pi(\sigma_0, \sigma_1) \cdot \mathcal{L}(\bar{t}, \sigma_0^2, \sigma_1^2|y) \quad (4)$$

Now, given the changepoint's location at $\bar{t}$, the goal is to test the equality of variances $H_0 : \sigma_0 = \sigma_1$.

It is important to note that the full model 4 is defined over a 2-dimensional parametric space, and that H_0 describes a lower (1-)dimensional manifold on this original space. Hypothesis that define lower dimensional manifolds on the parametric space are called *sharp* or *precise* hypothesis in the Bayesian literature[16].

These hypothesis are challenging to test in the usual Bayesian hypothesis testing frameworks, because the posterior measure over H_0 is by definition 0. Pereira and Stern[17], however, present a Bayesian evidence measure designed specifically for the test of sharp hypothesis; their measure is shown to be fully Bayesian (in the sense that it arrives directly from a particular cost function[18]), and to possess many desirable properties. This test has been succesfully applied to many problems involving sharp hypothesis testing[19–22].

Following the original authors, we call this measure the *e-value*, $ev(H_0)$ being the evidence value in favor of H_0. The full definition and analysis of the e-value is beyond the scope of this paper; however, to keep this work reasonably self-contained, we summarize the e-value in broad terms and refer the interested reader to the original paper by Pereira and Stern[17].

Given a full posterior model $P(\theta|x)$ with $\theta \in \Theta$, and a sharp hypothesis $H_0 : \theta \in \Theta_0$ with $dim(\Theta_0) < dim(\Theta)$, we obtain the maximum value of the full-posterior restricted to Θ_0

$$\theta^* = argmax_{\theta \in \Theta_0} P(\theta|x)$$
$$p^* = P(\theta^*|x)$$

Now define the tangent space or surprise set as

$$T(p^*) = \{\theta \in \Theta : P(\theta|x) > p^*\} \tag{5}$$

The tangent space is the set of all parameter values with higher posterior density than the maximum posterior under H_0. If this set has high posterior measure, it means that H_0 does not traverse regions of high posterior density, and the evidence in favor of H_0 must be low. In fact, define

$$ev(H_0) = 1 - \int_{T(p^*)} P(\theta|x) d\theta \tag{6}$$

to be the evidence in favor of H_0. The evidence will take the value 0 if the measure of the surprise set is 1 (i.e., if the maximum posterior value under H_0 is almost surely the minimum unrestricted posterior value), and conversly the evidence in favor of H_0 will be 1 if the measure of the surprise set is 0 (i.e., the maximum posterior under H_0 is almost surely the unrestricted maximum).

The definition of the test procedure finishes the construction of the binary algorithm. One full step of the algorithm will consist of two substeps: first, to estimate the segmentation point $\bar{t}$; second, to compare the variance of the segments, calculating a measure of evidence for the hypothesis $H_0 : \sigma_0 = \sigma_1$. A diagram illustrating the algorithm's flow can be seen in Figure 2.

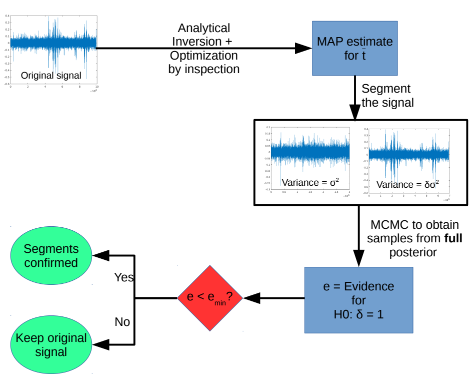

FIG. 2. One step of the sequential segmentation algorithm.

C. Informative priors and the power of the e-value

To calculate the e-value, from the segmentation model 4, all that is left to do is to pick a joint prior $\pi(\sigma_0, \sigma_1)$, and from then on follow the procedure delineated above.

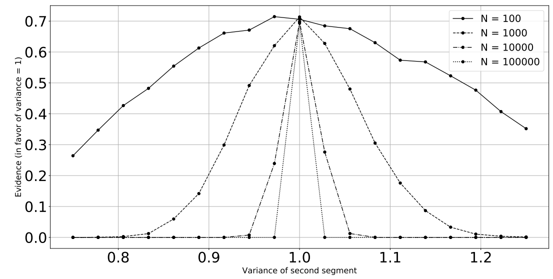

FIG. 3. Evidence value for H_0

One obvious choice for the priors is to adopt the product of Jeffreys' priors $(s_1 s_2)^{-1}$; by doing so, the model is treating both these parameters as completely unknown in advance, i.e., the algorithm will act as if it knows nothing about the segments' variances and the relation between them.

This choice gives the optimal value

$$\sigma_* = \frac{\sum_{t=1}^{N} y_t^2}{N+2} \tag{7}$$

for the signal's variance under H_0 (no changepoint). To calculate the evidence in favor of H_0, we estimate the integral of the posterior over the surprise set by the adaptive MCMC method of Haario[23].

To test the behavior of the e-value with this choice of priors, we simulate Gaussian signals with various sample sizes, divided into two segments, with the variance of the first segment set to 1, and that of the second segment varying in $[0.7, 1.3]$. Figure 3 shows the evidence in favor of H_0 (i.e., the evidence that variances are equal between segments) for several values of σ_1 and several sample sizes, where we have repeated each simulation 500 times. It is very important to take notice that the e-value is *not* a significance measure, i.e., it does not result from a *control type-I error* procedure. This implies that the sampling distribution of the e-value is not uniform; however, a transformation exists that changes the e-value into a significance measure[24]. Using this transformation, it is possible to fix the type-I error at 0.05 (for a single application of the test) and evaluate the resulting power. The result for different sample sizes and values of σ_1 is on figure 4, where the horizontal dashed line marks the 0.05 significance level. We ran 500 simulations for each combination of N and σ_1. The test based on the (transformed) e-value is quite powerful, as the simulations indicate. As expected, for a fixed type-I error, we can detect smaller changes as the sample size increases.

This is an important issue, especially in the segmentation algorithm where the test will be sequentially applied to the comparison of segments with different sample sizes. If we choose to keep α (probability of type-I error) fixed, the power of the test will change as the sample size changes. However, in a signal detection setup, usually one desires to balance both type-I and type-II error probabilities regardless of the size of the incoming signal.

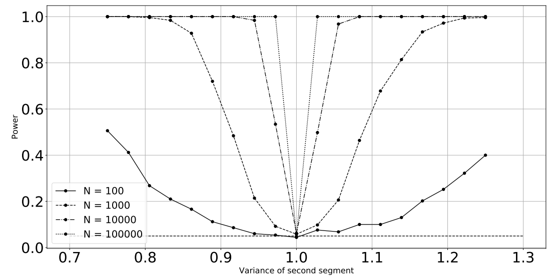

FIG. 4. Power of the test based on the e-value

D. Using informative priors

The relation between significance levels, test power and sample size is a deep and often discussed question in hypothesis testing[25,26]. Recent literature proposes to change the significance level as the sample size changes, to keep some relation $u(\alpha, \beta)$ between the probabilities of both error types at a constant value. This can be done by using adaptive significance levels (given by a function of the sample size n, see[25]) or by imposing an ordering on the parameter space based on Bayes factors[26].

For the segmentation task, however, and in our particular application (segmentation of large samples), the algorithm will have to work with segments of very different sizes (from 10000 to more than 9 million), and the adaptive significance level would also vary wildly. The consequence is that, for the larger segments, the algorithm would require very small significance values; and in a MCMC setting, higher precision for the probability estimates means longer chains, and longer chains mean higher execution times.

So instead of using an adaptive significance value, we propose instead to use a strongly informative prior, and use the hyperparameters to calibrate the power of the procedure.

This idea was first introduced in a previous paper[27]. The paper analyzes the binary algorithm for signal segmentation, but uses a different parameterization $\theta = (\sigma_0, \delta)$ where $\delta = \sigma_1/\sigma_0$. Independent priors for these two parameters are proposed, one that is uninformative on the value of σ_0, and strongly informative over δ. The advantage of working with (σ_0, δ) instead of (σ_0, σ_1) is that δ is a pure number, i.e., it does not depend on scale. It can be interpreted as the quotient between the power of any two contiguous segments.

There are however some difficulties in working with $\delta = \sigma_1/\sigma_0$, one of them being that, as σ_0 and σ_1 are nonnegative, δ must also be nonnegative. This limits the choice of priors for δ, and for this new, current version of the algorithm, we parameterize the problem using $\delta = log(\sigma_1/\sigma_0)$, and propose a Laplace prior with the form

$$p(\delta) = \frac{1}{2\beta} e^{-\frac{|\delta|}{\beta}} \tag{8}$$

The above Laplace distribution has a peak on $x = 0$, and the peak is sharper as the value of $\beta > 0$ decreases.

The segmentation algorithm works as above, except that now the e-value calculation uses the Laplace prior for δ. This prior, when β is close enough to 0, changes significantly the power of the test, and thus allows tuning of the algorithm's behavior.

Figure 5 shows the same estimation of power as in figure 4, but this time using the Laplace prior. The values of β were taken as $0.005, 0.0005, 0.00005$ for $N = 1000, 10000, 10000$ respectively (i.e., for $N = 1000$, $\beta = 0.005$, for $N = 10000$, $\beta = 0.0005$, and so on). Again, for each value of N and σ_1 we ran 500 simulations, and the horizontal dashed line marks the 0.05 significance level.

The effect of the highly informative prior is to lower the power of the test for all sample sizes. This is the case even when the hypothesis is true, i.e., when $\sigma_1 = 1$; in this case, it would be expected that the power be equal to the significance value (0.05 in the simulations). What happens, however, is that the prior evidence on the manifold $\sigma_1 = \sigma_0$ is so strong, that the evidence in the data is incapable of raising the evidence value above 0.

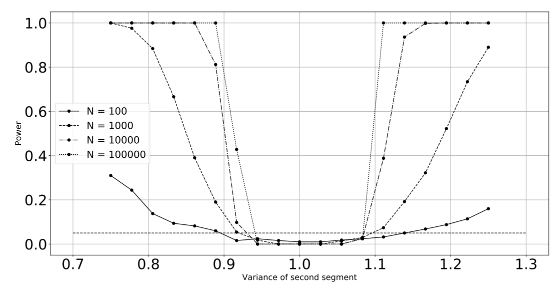

FIG. 5. Power of the test based on the e-value with strongly informative priors

Being able to control the power of the test will prove useful when segmenting underwater acoustic signals; in this setting, long segments with stationary power are not to be expected, even when the segment is capturing a single event. That is the case because both the background noise and the event's physical cause might be changing, due to many factors (including the weather, the movement of event's cause relative to the hydrophone, among others). With a high sampling rate (the data we use in this paper was sampled at $24KHz$) the e-value would give strong evidence against $H_0 : \sigma_0 = \sigma_1$ even inside a segment containing a uniform event, and this would lead to oversegmentation (overfitting). To control the power of the test using an informative prior will allow the algorithm's sensibility to be tuned to the goals of the analysis: if one is interested in capturing larger sections, that might suffer an internal power change that is small compared to the difference between the segment overall power and the background noise power, one only needs to adjust the hyperparameter accordingly.

E. The resolution parameter

The most demanding step in our binary algorithm is the optimization procedure that looks for the most likely changepoint at each step. This is done by a brute force procedure, that can be parallelized but nevertheless is costly, especially with long signals.

One way to increase the speed of our algorithm is to limit the search for the optimal changepoint: instead of calculating the objective function for all $i \in \{1, ..., N\}$, we can instead calculate the objective only for $i = lj, j \in \{1, ..., N/l\}$.

If the (discrete) posterior for $\bar{t}$, the changepoint parameter, is not very sharp around its maximum, and if the minimum expected segment length is also not too small, l above can be set to a high value, increasing the speed of the algorithm while still being able to identify the most probable changepoints at each step. This strikes a balance between the computational cost of achieving a segmentation and the accuracy of that segmentation.

However, and since the optimization step will be applied many times, to segments of different lengths, it is not advisable to pick a fixed integer value for l; imagine, for instance, that we fix $l = 1000$. In a signal of size $N = 1,000,000$, this value won't stop the algorithm from identifying a good approximation for the changepoint locations; however, for a signal of size $N = 10,000$, it is quite possible that using $l = 1000$ will cause the algorithm to miss the optimal point. For this reason, we adopt an adaptive resolution strategy: we pick a starting value for the resolution (say $l = 1000$), but as the algorithm starts obtaining new segments, it will keep the ratio l/N fixed at each step.

In the analysis of discretely sampled acoustic signals, the value of l can be converted to a time resolution: for instance, if the sampling rate is 1 kHz, $l = 1000$ means that the algorithm looks for candidate changepoints that are 1 second apart. When sampling rates are higher, as is usually the case, a value of $l = 1000$ will keep the time resolution sufficiently high when looking for candidate changepoints.

There is also one more important point about the time resolution parameter. It is only applied at the optimization step, i.e., in the search for the candidate changepoint. The equality of variances test is executed over the whole signal.

F. The PELT algorithm

As a basis of comparison to the Bayesian binary algorithm results, we use the PELT algorithm of Killick[11]; the PELT (*Pruned Exact Linear Time*) algorithm solves the dynamical optimization problem exactly, yielding the global optimum of the model. It does that with $\mathcal{O}(n^2)$ complexity in the worst case, but it can be shown to have $\mathcal{O}(n)$ complexity under mild conditions, which includes observing changepoints regularly throughout the data.

The algorithm is defined in terms of an additive cost function

$$C(\{t_i\}) = \sum_{i=1}^{m+1} \left[\mathcal{C}(y_{t_{i-1}+1:t_i}\right] + \beta f(m) \tag{9}$$

In the case of detection of variance changepoints

$$\begin{aligned}\mathcal{C}(y_{t_{i-1}+1:t_i}) = -\frac{|t_i - t_{i-1}|}{2} log\left(\sum_{j=t_{i-1}}^{t_i} y_j^2\right) + \\ log\left[\Gamma\left(\frac{|t_i - t_{i-1}|}{2}\right)\right]\end{aligned} \tag{10}$$

and $f(m)$ is the penalty or regularization function for the number of segments.

The penalty function is essential, since the direct optimization of the cost function will lead to overfitting (which, in this case, will mean oversegmentation). In our tests below, we adopt the MBIC penalty function[28], which is the penalty function used by default by the R package *changepoint* that implements the PELT algorithm[29].

For further comparison of our algorithm with other alternatives, we also run the binary segmentation algorithm of Scott[30], which is also implemented by the R package *changepoint*.

III. RESULTS

A. Simulated data

To analyze the performance of the Bayesian binary algorithm, we start by simulating Gaussian signals with constant mean and variance. We then simulate the changepoint process by using a geometric distribution to model the times between changepoints, and multiply the signal between changepoints for a given factor in order to obtain different variances.

It is clear that the effectiveness of a changepoint detection algorithm depends directly on both the size of the segments, and the magnitude of the jump in the process parameters. To observe the behavior of all algorithms with varying segment sizes, we will keep the expected number of changepoints fixed at 50 changepoints regardless of the signal's size. When the signal's size n changes, the expected length of the segments will change accordingly (linearly with n).

To simulate the magnitude of change in power between segments, we force the segments to alternate variances between 1.0 and 2.0. The simulation of the changepoint process was repeated ten times for each value of N, and we report the average results for each of these values.

The results appear in table I. The table reports the true number of changepoints in the simulated signal, the estimated total number of changepoints for each algorithm, and the F1 score. The F1 score is calculated as

$$F1 = \frac{precision * recall}{precision + recall}$$

TABLE I. Simulation results; see text for details

N	Algorithm	Time (s)	True k	Estimated k	F1 score
10,000	binseg	0.4072	34.3	2.4	0.0857
10,000	pelt	0.0378	34.3	5.1	0.2180
10,000	jeffreys	0.2104	34.3	4.0	0.1720
10,000	**laplace**	**0.2450**	**34.3**	**5.9**	**0.2365**
50,000	binseg	2.1517	46.1	15.9	0.4890
50,000	**pelt**	**0.1775**	**46.1**	**30.7**	**0.7939**
50,000	jeffreys	1.6281	46.1	28.6	0.7017
50,000	laplace	1.5635	46.1	34.1	0.7613
100,000	binseg	4.2698	45.9	29.5	0.7725
100,000	**pelt**	**0.3332**	**45.9**	**38.2**	**0.9074**
100,000	jeffreys	2.6243	45.9	37.3	0.8409
100,000	laplace	2.3943	45.9	41.7	0.8728
500,000	binseg	20.9543	50.8	42.6	0.8708
500,000	**pelt**	**1.9974**	**50.8**	**49.1**	**0.9817**
500,000	jeffreys	4.5585	50.8	50.2	0.8886
500,000	laplace	4.0887	50.8	49.9	0.8285
1,000,000	binseg	20.6610	51.8	40.0	0.3720
1,000,000	**pelt**	**3.9094**	**51.8**	**50.0**	**0.9826**
1,000,000	jeffreys	6.2435	51.8	53.7	0.9246
1,000,000	laplace	5.9118	51.8	56.5	0.9215

where *precision* is the number of true positives divided by the total number of changepoints identified, and *recall* is the number of true positives divided by the total number of true changepoints. To accept an estimated changepoint as a true one, it must be between $N/100$ points of a true changepoint.

The value of α for the Jeffreys prior, and the values of both α and β for the Laplace prior were selected using the Bayesian Information Criterion (BIC); both the PELT and the BinSeg algorithms utilized the Modified BIC of Zhang[28]. The time resolution parameter for the Bayesian binary algorithm was kept fixed and with value $l = 1$.

The PELT algorithm was the quickest and also the most accurate algorithm on average for all signal sizes, except for $N = 10,000$ where the Bayesian binary algorithm with the Laplace prior showed a higher F1 score. The binary algorithm of Scott[30] was always the slowest and less precise; also, since it is implemented recursively, for longer signals there was an operational system error related to the stack size that stopped the algorithm from running in many simulations.

The Bayesian binary segmentation can be seen to be competitive with PELT in accuracy, even though PELT runs considerably faster in all cases. The use of an informative (Laplace) prior improved the accuracy in almost all scenarios.

In the next section, we apply the Bayesian binary algorithm and PELT to real underwater acoustic signals; the binary algorithm won't be tested because it is unpractical for signals of the size we will be using.

B. Underwater acoustic signals

Now we apply the three algorithms to the segmentation of real underwater acoustic signals. These signals were obtained by the LACMAM's team on 2017, in the region of Alcatrazes, an archipelago 35 km off the Brazilian coast, in the city of São Sebastião, SP. More information about the data and the experiment can be found in the work of Sanchez-Gendriz and Padovese[31].

One of the main goals in acquiring these samples is the study of acoustical signatures of boats. Alcatrazes is a marine ecological reserve, the second largest in Brazil, and as such fishing is prohibited in the archipelago's area. As passive acoustic monitoring is cheap, efficient algorithms for boat detection using hydrophone data are a valuable resource to the reserve's fiscalization authorities.

The laboratory has, by January, 2019, collected almost two years of acoustic signals from the reserve's region. In these signals, many events can be found: the passage of boats, but also fish and whales' vocalizations, and other events with both biological and anthropogenic sources. These events, however, are scarce, making the direct inspection and annotation of the signal a demanding task. The segmentation algorithm will be used to aid in this inspection, by first separating sections of the signal that are likely to contain any significant event.

To test the segmentation algorithms, we have chosen two 15 minutes long samples where visual inspection of the spectrogram shows many short duration events. After examination of the spectrograms, the samples were listened to and the start and finish times of all events were annotated by an expert. A total number of 32 changepoints were detected, all of them caused by the passage of boats. What we expect is that the segmentation algorithm will be able to correctly identify the boundaries of these events.

One disclaimer is due at this point. The inspection of the samples was aimed at the separation of samples of the acoustic signal generated by the passage of boats. The researcher responsible for the annotation, thus, was not looking to annotate changes in the signal power. For that reason, it is not expected that any algorithm will get high measures of precision or recall, due to other features in the data that will present themselves as changes in variance.

The sampling rate of these files is 24 kHz, resulting in signals with size $21,600,000$ for 15 minutes recordings. To reduce this signal size, it is possible to arbitrarily break the 15 minutes signal into smaller pieces, or to downsample the signal. The arbitrary separation of smaller pieces seem the least desirable approach, since it introduces the problem of deciding where to separate the pieces.

For the following tests, however, no downsampling was adopted, and the reported results refer to the segmentation of the full $21,600,000$ points signal.

For the Bayesian binary algorithm with the Laplace prior, the selection of the β value is done based on an elbow plot of the BIC criterion, i.e., we select the greater β for which the plot $BIC = f(\beta)$ shows a pronounced decrease when compared to the previous β value (i.e., the elbow method in scree plots). For the PELT algorithm, the MBIC criterion is applied, using the default penalty value. Methods such as the scree plot could be applied to the selection of PELT's penalty value, but this would be unpractical regarding total computation time.

In the results in table II, the execution time for the Bayesian binary algorithm with Laplace prior includes all the runs necessary to obtain the best β. In order to assess the effect of using strongly informative priors in our algorithm, we also included the results for the Bayesian binary algorithm using the Jeffreys' (non-informative) prior.

As seen in table II, the Bayesian binary algorithm showed superior results to PELT in the segmentation of real samples. The first thing to notice is that PELT resulted in an excessive number of changepoints; that is the case because PELT works with the exact optimization of a cost function that is based on a (Gaussian) likelihood, and even with the regularization induced with the MBIC criterion, a higher number of changepoints gives a better fit. The same happens with the Bayesian binary algorithm using non-informative priors, i.e., with uncontrolled power of the test based on the e-value.

With the Bayesian binary algorithm, on the other hand, the value of β helps to control the power of the test based on the e-value, avoiding oversegmentation.

In figures 6 and 7, the changepoints estimated by the Bayesian binary algorithm are plotted over the spectrogram of the samples. It is noticeable that the boundaries of the most prominent events are correctly captured by the algorithm, while at the same time sections with no important events (as can be seen by direct inspection of the spectrogram) are kept unsegmented.

IV. CONCLUSION

The segmentation of acoustic signals is an important task, especially in the retrospective analysis of long duration signals.

Among the many possible criteria for the segmentation, the RMS-based segmentation is particularly interesting when one is mainly interested in separating sections with background noise only, from sections composed of background noise plus some (possibly) interesting event.

In this paper, we present a Bayesian binary algorithm for RMS-based acoustic signal segmentation. We show that this algorithm is precise, and robust to violations on the basic assumptions: normality of background noise, and a stepfunction for the RMS in the different segments. We claim that this robustness is mainly due to two characteristics of our algorithm: first, the use of a marginal posterior for the selection of candidate changepoints; and second, the use of strongly informative priors.

By comparing our algorithm with other alternatives from the literature, we showed that it is competitive with the current state-of-the-art changepoint algorithm (PELT), and sensibly superior to previous binary algorithms in simulated data. When analyzing real data, we showed that our algorithm can have superior results even when compared to PELT, if we use the strongly informative (Laplace) prior on the log-ratio of variances between segments.

The hyperparameter of the Laplace prior can be efficiently selected using model selection criteria such as the Bayesian Information Criterion (BIC).

Further work will analyze other possibilities for the model selection problem in this setting. We are also working on a hybrid version of our algorithm and the PELT algorithm, by using a version of our marginal posterior as the cost function to be optimized with PELT.

Our algorithm is written in *cython*, is open sourced an can be downloaded at http://github.com/paulohubert/bayeseg, along with some sample acoustic data and some illustrative *IPython* notebooks. The signals used in this paper are available upon request.

[1] R. Makowsky and R. Hossa, "Automatic speech signal segmentation based on the innovation adaptive filter," Int. J. Appl. Math. Comput. Sci. **24**(2), 259–270 (2014).

[2] A. Ukil and R. Zivanovic, *Automatic Signal Segmentation based on Abrupt Change Detection for Power Systems Applications* (Power India Conference, 2006).

[3] A. Schwartzman, Y. Gavrilov, and R. J. Adler, "Multiple testing of local maxima for detection of peaks in 1d," The Annals of Statistics **39**(6), 3290–3319 (2011).

[4] S. Kuntamalla and L. R. G. Reddy, "An efficient and automatic systolic peak detection algorithm for photoplethysmographic signals," International Journal of Computer Applications **97**, 19 (2014).

[5] T. Thedorou, I. Mporas, and N. Fakotakis, "An overview of automatic audio segmentation," I.J. Information Technology and Computer Science **11**, 1–9 (2014).

[6] M. Caldas-Morgan, A. Alvarez-Rosario, and L. R. Padovese, "An autonomous underwater recorder based on a single board computer," PLos One **10** (2015).

[7] S. A. Shaban, "Change point problem and two-phase regression: An annotated bibliography," International Statistical Review **48**(1), 83–93 (1980) http://www.jstor.org/stable/1402408.

[8] S. Aminikhanghahi and D. Cook, "A survey of methods for time series change point detection," Knowledge and Information Systems **51**(2), 339–367 (2017) doi: 10.1007/s10115-016-0987-z.

[9] B. A. Tan, P. Gerstoft, C. Yardim, and W. S Hodgkiss, "Change-point detection for recursive bayesian geoacoustic inversions," The Journal of the Acoustical Society of America **137**, 1962 (2015) doi: 10.1121/1.4916887.

[10] B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulchareon, L. Tan, and T. T. Tsai, "An algorithm for optimal partioning of data on an interval," IEEE Signal Processing Letters **12**(2) (2005).

[11] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," Journal of the American Statistical Association **107**(500), 1590–1598 (2012).

[12] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application* (Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993).

[13] E. T. Jaynes, "Prior probabilities," IEEE Transactions On Systems Science and Cybernetics **4**(3), 227–241 (1968).

[14] H. Jeffreys, "An invariant form for the prior probability in estimation problems," Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences **186**(1007), 453–461 (1946).

[15] P. Hubert, "Um sistema para segmentao e caracterizao nosupervisionada de eventos em sinais acsticos," Ph.D. thesis, University of São Paulo, 2018.

[16] J. M. Dickey and B. P. Lientz, "The weighted likelihood ratio, sharp hypotheses about chances, the order of a markov chain,"

TABLE II. Results on real samples; see text for details

Sample	Method	Time (s)	Beta	Expert's k	Estimated k	Precision	Recall	F1
A	jeffreys	1239.59	-	12	42074	0.03%	100%	0.0003
B	jeffreys	1329.73	-	20	45277	0.04%	100%	0.0004
A	**laplace**	**27.41**	**3.3e-5**	**12**	**28**	**17.9%**	**41.7%**	**0.1250**
B	**laplace**	**30.89**	**1.6e-5**	**20**	**21**	**30.0%**	**30.0%**	**0.1500**
A	pelt	205.41	-	12	39170	0.03%	100%	0.0003
B	pelt	205.38	-	20	38274	0.05%	100%	0.0005

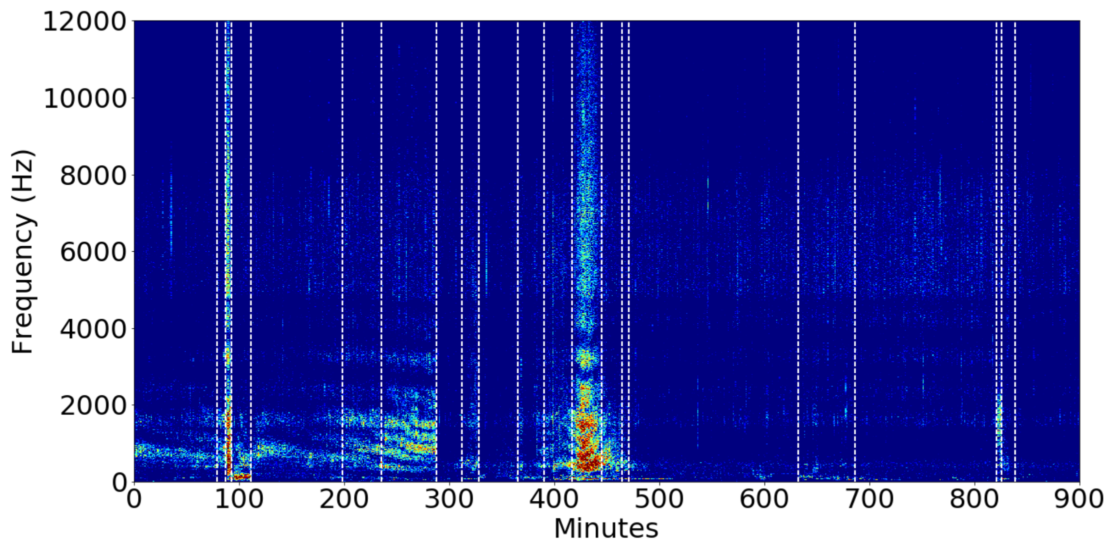

FIG. 6. Spectrogram of sample A with changepoints estimated by the Bayesian binary algorithm

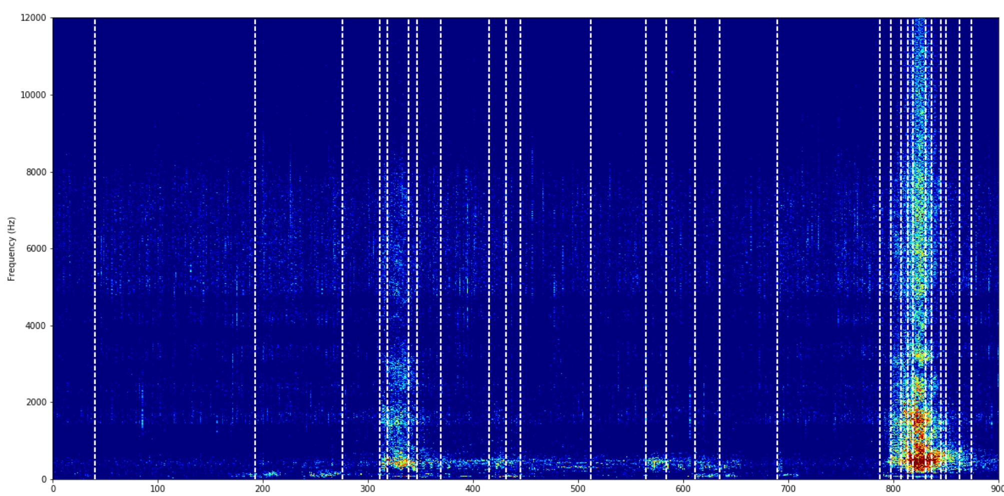

FIG. 7. Spectrogram of sample B with changepoints estimated by the Bayesian binary algorithm

The Annals of Mathematical Statistics **41**(1), 214–226 (1970) http://www.jstor.org/stable/2239734.

[17]C. A. B. Pereira and J. M. Stern, "Evidence and credibility: full Bayesian significance test for precise hypotheses," Entropy **1**, 99–110 (1999).

[18]M. R. Madruga and S. W. L. G. Esteves, "On the bayesianity of pereira-stern tests," Test **10**(2), 291–299 (2001).

[19]P. Hubert, M. Lauretto, and J. M. Stern, "Fbst for the generalized poisson distribution," AIP Conference Proceedings **1193**(210) (2009).

[20]M. Diniz, C. A. B. Pereira, and J. M. Stern, "Unit roots: Bayesian significance test," Communications in Statistics - Theory and Methods **40**(23), 4200–4213 (2012).

[21]D. Chakrabarty, "A new bayesian test to test for the intractability-countering hypothesis," Journal of the American Statistical Association **112**(518), 561–577 (2017) https://doi.org/10.1080/01621459.2016.1240684 doi: 10.1080/01621459.2016.1240684.

[22]P. Hubert, L. Padovese, and J. M. Stern, "Full bayesian approach for signal detection with an application to boat detection on underwater soundscape data," 37th Maximum Entropy Methods in Science and Engineering (2017).

[23]H. Haario, E. Saksman, and J. Tamminen, "An adaptive metropolis algorithm," Bernoulli (2001) doi: https://doi.org/10.2307/3318737.

[24]J. M. Stern, "Cognitive constructivism and the epistemic significance of sharp hypothesis," (2008).

[25]L. Pericchi and C. A. B. Pereira, "Adaptative significance levels using optimal decision rules: Balancing the error probabilities," Brazilian Journal of Probability and Statistics **30**(1), 70–90 (2016).

[26]C. A. B. Pereira, E. Y. Nakano, V. Fossaluza, L. G. Esteves, M. A. Gannon, and A. Polpo, "Hypothesis tests for binomial experiments: Ordering the sample space by Bayes factors and using adaptive significance levels for decisions," Entropy **19**(12) (2017).

[27]P. Hubert, L. Padovese, and J. M. Stern, "A sequential algorithm for signal segmentation," Entropy **20**(1), 44 (2018).

[28]N. R. Zhang and D. O. Siegmund, "A modified bayes information criterion with applications to the analysis of comparative genomic hybridization data," Biometrics **63**, 22–32 (2007).

[29]R. Killick and I. A. Eckley, "changepoint: an r package for changepoint analysis," Journal of Statistical Software **58**(3) (2014).

[30]A. J. Scott and M. Knott, "A cluster analysis method for grouping means in the analysis of variance," Biometrics **30**(3), 507–512 (1974).

[31]I. Snchez-Gendriz and L. Padovese, "A methodology for analyzing biological choruses from long-term passive acoustic monitoring in natural areas," Ecological Informatics **41**, 1–10 (2017) doi: 10.1016/j.ecoinf.2017.07.001.