

Micro-Macro Changepoint Inference for Periodic Data Sequences

Anastasia Ushakova

Department of Mathematics and Statistics, Lancaster University

Simon A. Taylor

School of Mathematics, University of Edinburgh

Rebecca Killick

Department of Mathematics and Statistics, Lancaster University

May 4, 2022

Abstract

Existing changepoint approaches consider changepoints to occur linearly in time; one changepoint happens after another and they are not linked. However, data processes may have regularly occurring changepoints, e.g. a yearly increase in sales of ice-cream on the first hot weekend. Using linear changepoint approaches here will miss more global features such as a decrease in sales of ice-cream due to other product availability. Being able to tease these global changepoint features from the more local (periodic) ones is beneficial for inference. We propose a periodic changepoint model to model this behavior using a mixture of a periodic and linear time perspective. Built around a Reversible Jump Markov Chain Monte Carlo sampler, the Bayesian framework is used to study the local (periodic) changepoint behavior. To identify the optimal global changepoint positions we integrate the local changepoint model into the pruned exact linear time (PELT) search algorithm. We demonstrate that the method detects both local and global changepoints with high accuracy on simulated and motivating applications that share periodic behavior. Due to the micro-macro nature of the analysis, visualization of the results can be challenging. We additionally provide a unique perspective for changepoint visualizations in these data sequences. Supplementary Materials for this article are available online.

Keywords: PELT, segmentation, periodicity, level shift, house price index, North Atlantic Oscillation

1 Introduction

In a world constantly challenged by change, identifying the points of change is of crucial importance for researchers in all fields including health, finance, and the environment. However, what if some processes exhibit a periodic (e.g. yearly) change and understanding how constant this periodicity is might be of interest? Such non-stationary data sequences experiencing sudden change in their structure can be studied using changepoint analysis, where often the changes can be described by changes in means, variance or other distributional characterizations. The roots of research in changepoint analysis date back to 1950s in the area of statistical process control (Page, 1955). Since then, changepoint based problems can be found far beyond statistical sciences and include engineering, environmental sciences, linguistics and many others.

There are a number of changepoint methods developed for both frequentist and Bayesian inference. Traditional multiple changepoint search algorithms consider a one scale perspective where a single data sequence is partitioned into a number of segments, that best describe the characteristics of interest, per segment. Among the most notable frequentist changepoint search algorithms are binary segmentation (BS) (Scott and Knott, 1974) and pruned exact linear time (PELT) (Killick et al., 2012), with Bayesian alternatives represented by filtering mechanism (Fearnhead, 2006) and hidden Markov models (Reeves et al., 2007). For more comprehensive overview of modeling techniques and assumptions please see Eckley et al. (2011); Reeves et al. (2007).

The majority of changepoint algorithms evaluate from a sequential time perspective; treating the process prior to the change as independent from the process after the change. This approach is statistically convenient but the assumption is often not satisfied in real-world applications. In this paper we want to challenge this idea by introducing a nested or dual-level structure. Specifically, we consider that the data generating process possesses a macro and micro changepoint structure. The structure of the macro-level represents the aforementioned overall pre- and post- changepoint independence. The micro-level, however, represents a periodic structure whereby there is a reoccurring pattern of abrupt changes over a given macro-level segment. For example, a 24/7 customer support service may ob-

serve a regular daily changepoint pattern in demand, but the introduction of a new product or a failure could impact the level of demand in one or more daily segments, and perhaps simultaneously change when the transitions occur. The introduction of changepoint periodicity at the micro-level introduces dependence as one would expect the same underlying process at a given point within the period every time that period occurs until the next macro-level changepoint event. This gives a parsimonious representation as each period has the same underlying distribution.

As a motivating application consider the UK house sales price index (HPI) dynamics (Figure 1). The data is expected to have changes in mean within the year (micro behavior) as well as macro changes due to, for example, recessions. There are a total of 216 time points which are split in 18 periods of length 12. A quick observation would note at least one potential changepoint position at around the year 2010.

Rather than the overall index, Fryzlewicz (2020) consider averages of “inner” and “outer” London boroughs. They argue that there are considerable numbers of changepoints in both series. Upon further inspection one can ascertain that these may be exhibiting periodic changepoint behavior characterizing the larger and smaller values on a yearly basis.

Considering again the overall sequence of HPI, there is arguably a periodic nature to the highs and lows each year in parts of the series. Using a linear time solution might pose an identification problem for the micro scale events as a high number of changepoints, with short segments, would need to be estimated to describe the observed data, as in Fryzlewicz (2020). Furthermore, even if all changepoints are identified, as these are estimated independently they are unlikely to align within each period as the periodic nature is not utilized. Practitioners also would like to separate out the year-to-year expected variation from the macro-level changes. This is not possible within the linear time framework.

Fitting two linear time changepoint algorithms, PELT (Killick et al., 2012) and WBS2 (Fryzlewicz, 2020) with changes in mean and variance and default penalties (Figure 1), they identify a few changepoints separating the series into distinct segments, with some, but not all, shorter segments as they cannot take into account the micro periodic structure.

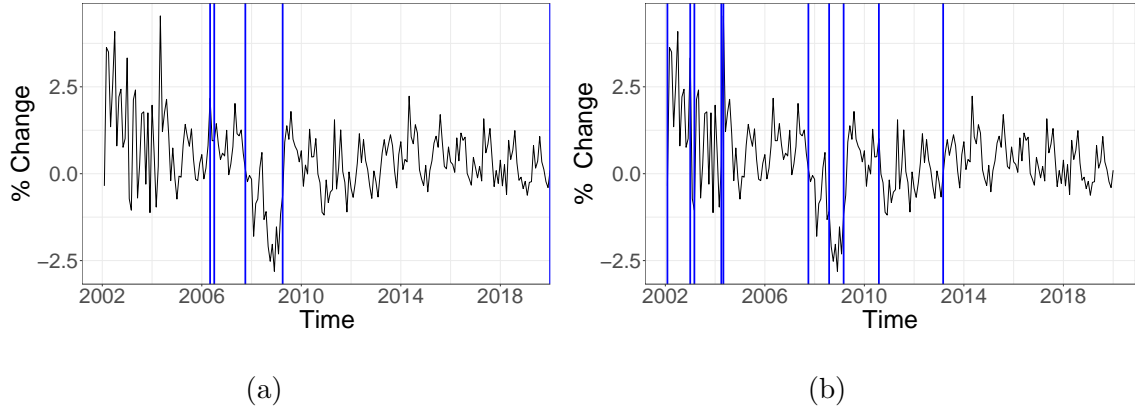


Figure 1: (a) Changepoints identified using Normal likelihood with PELT search algorithm and default penalty (b) Changepoints identified using mean change with WBS2 search algorithm and default parameters.

In this paper, we propose a novel method that permits the investigation of the micro-level periodic changepoints alongside the overall macro-level changes. We extend the previous work of Taylor et al. (2021) for Binary observations to continuous observations and combine it with the method of Killick et al. (2012) to create a single inference scheme. This novel approach simultaneously estimates both the micro and macro-level changes, directly allowing them to influence each other. As the micro-level changepoints are characterized as periodic, they require fewer parameters than traditional linear changepoint detection. Therefore we explicitly identify, in a parsimonious way, changepoints that occur regularly and those which are more structural. Alongside this novel micro-macro changepoint methodology, we simultaneously propose innovative visualizations to assess the results, helping practitioners make inferences and take decisions.

The structure of the paper is as follows. Section 2 describes our novel approach with 2.2 and 2.3 covering the micro and macro methodology respectively. This section pays particular attention to derivation of the posterior distribution and overview of the prior distribution for parameters that are used for the inference mechanism in the Bayesian framework. The method is evaluated in Section 3 using simulated scenarios and applications in Section 4. Section 5 concludes with a discussion of our approach.

2 Materials and Methods

2.1 Dual-level changepoint structure

Let y_1, \dots, y_n denote a real-valued sequence of observations that follows a micro-macro changepoint structure where the micro-level has periodicity of known length. Specifically, let t_1, \dots, t_q denote the times of the q macro-level changepoint events that partition the observation sequence into $q+1$ macro-segments, with convention that $t_0 = 0$ and $t_{q+1} = n$. We denote the length of the macro-segments by $n_i = t_i - t_{i-1}$, and the set of observations in the i th macro-segment as $y_{(t_{i-1}, t_i)} = \{y_{t_{i-1}+1}, \dots, y_{t_i}\}$.

The observation sequence within a given macro-level segment are further partitioned according to the micro-level changepoints. Due to the periodicity of the data, with known length N , we need only notate the changepoint events for a single period and the other events are subsequently derived. For the i th macro-level segment, let $0 = t_{i,1}, \dots, t_{i,m_i} < N$ denote the micro-level changepoints that partition the N -period into m_i segments such that $t_{i,m_i+1} \bmod N = t_{i,1}$. Note that if $m_i = 1$ then the period contains one single segment and so no micro-level changepoints exist. For $m_i > 1$, the micro-level changepoints are derived according to $t_{i,j,k} = t_{i,j} + kN$ for all integers k such that $t_{i,j,k} \in (t_{i,j}, t_{i,j+1}]$. See Figure 2 for a visual depiction of this setup.

Let $B_i = (t_{i-1}, t_i]$ denote the macro-level interval for the i th segment and $B_{i,j}$ denote the sequence of micro-level repeated intervals relating to the micro-level j th segment contained within B_i :

$$B_{i,j} = B_i \setminus \bigcup_{k \in \mathbb{Z}} (t_{i,j-1,k}, t_{i,j,k}] \quad (1)$$

The doubly indexed micro-level intervals are disjoint and their union covers the whole period: $\bigcup_{i=1}^{q+1} \bigcup_{j=1}^{m_i} B_{i,j} = \{1, \dots, n\}$. It is important to emphasize that there need not be alignment between the macro and micro-level changepoints, that is to say the end of the current macro-level segment need not align with the end of a micro-level segment and the beginning of the next macro-level segment does not dictate the beginning of a new micro-level changepoint event.

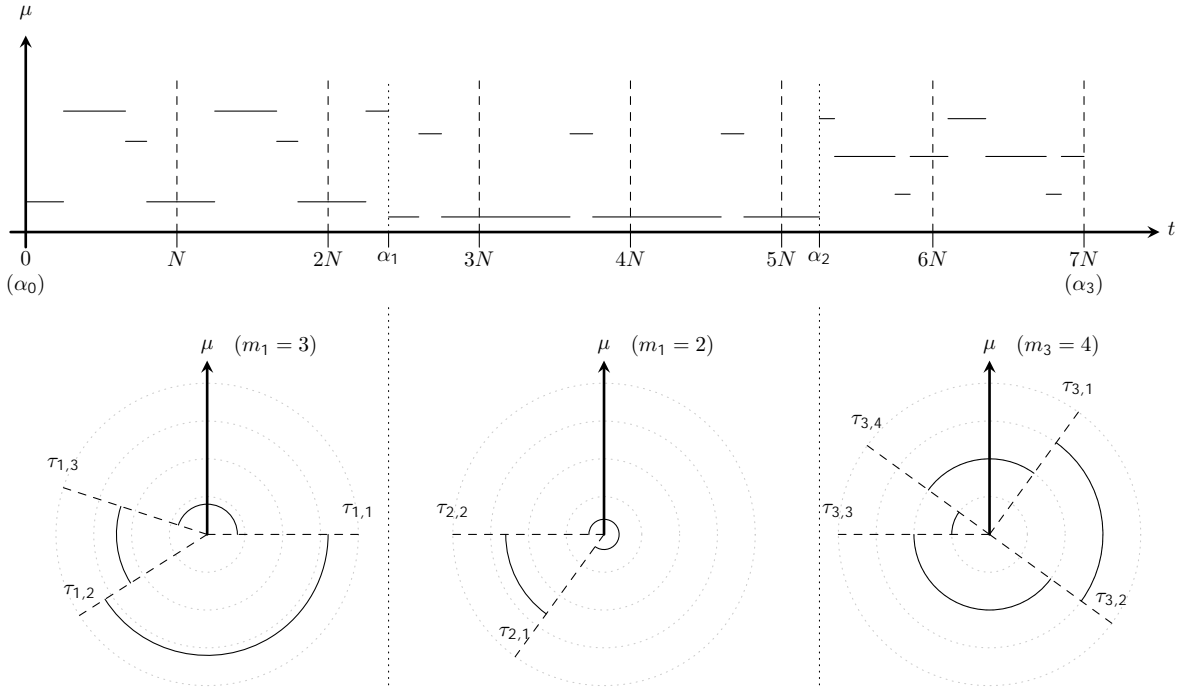


Figure 2: Visual depiction of the Micro-Macro problem setup. With the linear time perspective on the top graphic and circular time perspective within each macro-segment in the lower graphics. Dotted lines (α_i) in the top graphic are macro-changes and in the bottom are τ_{ij} micro-changes.

We assume that the data, y_1, \dots, y_n is well represented by the following Gaussian model:

$$y_t \stackrel{iid}{\sim} N(\mu_{ij}; \frac{\sigma^2}{i_{ij}}); \text{ for } t \in B_{ij}; \quad (2)$$

where μ_{ij} and $\frac{\sigma^2}{i_{ij}}$ respectively denote the unknown mean and variance on the corresponding set of micro-level intervals. See Figure 2 for a visual depiction of our micro-macro changepoints for the mean parameter. This assumption can be relaxed or changed to other modeling structures provided that appropriate micro-level changepoint methods are available.

The complete model is described by two key components: a) a micro-level periodic changepoint model; b) a macro-level linear changepoint model.

2.2 Periodic changepoint model with macro-level inference

Here we provide details about the Bayesian within period changepoint model and inference procedure of Taylor et al. (2021). Although developed to investigate changepoints within period binary observations, here we extend the method to other sampling distributions. We emphasize where these adaptations occur for our purposes. The discussion in this section focuses on the micro-level variation for a given macro-level segment. For brevity of notation, we define $\mathbf{y}_i = \mathcal{Y}_{(i-1+1):i}$ for the set of observations on the current macro segment, and temporarily suppress the index i on all other parameters and terms.

2.2.1 Within Period Changepoint Model

The observations on a given macro-segment \mathbf{y}_i is assumed to be periodic with known period length N . It is important to re-emphasize that the first observation in \mathbf{y}_i can occur at any point within the period cycle and need not dictate its starting position. The period is partitioned according to the changepoint events τ_1, \dots, τ_m . The duration between events, or the length of the segment, is assumed to be larger than a pre-specified minimum: $\tau_j - \tau_{j-1} \geq l_1$. In practice, the minimum segment length l_1 represents how long any particular regime structure must be sustained before transitioning to the next. In addition, this condition also ensures there is a sufficient number of observations for parameter estimation on short segments, especially when the duration of \mathbf{y}_i is also small.

The number of micro-level segments, m , is given a Poisson prior with rate 1 that is truncated to the set $\{1, \dots, \lfloor N/l_1 \rfloor\}$; ranging from the single segment where there are no micro-level changes, up to the case where as many changepoints as possible can occur given the minimum segment length constraint. As mentioned in Section 2.1, the case $m = 1$ describes a period containing only one segment with no abrupt changes and therefore the micro-level changepoint events are undefined. Otherwise, we denote the vector of micro-level changepoint events by $\boldsymbol{\tau} = (\tau_1, \dots, \tau_m)$.

Conditional on the number of segments, define $\tilde{\tau}_j = \tau_j - \tau_{j-1} - l_1$ to be the excess segment length of the j th segment after accounting for the minimum length constraint, whereby $\sum_{j=1}^m \tilde{\tau}_j = N - ml_1$. The prior for the spacing between micro-level changepoints

is given by the Dirichlet-Multinomial distribution:

$$j|m > 1 \sim \text{DirMult}(N \cdot m \mathbf{1}_1; \mathbf{1}_m); \quad (3)$$

with positive shape hyperparameter α and $\mathbf{1}_m$ denoting a ones vector of length m . The shape parameter controls the prior degree of clustering, $\alpha < 1$, or repulsion, $\alpha > 1$, between neighboring changepoint events. To determine the exact changepoint positions, the transformation is reversed conditional on the position of one changepoint that initializes the calculation. The prior for this initial changepoint is uniform over the set $\{0, \dots, N-1\}$. As a consequence, it is not possible *a priori* to be overly specific on the degree of clustering or repulsion of selective neighboring changepoint pairs, hence the same hyperparameter in (3) is applied to all components of \mathbf{c} . If necessary, we re-label the events to ensure that $c_1 = \min_j (c_j \bmod N)$.

Similarly to Taylor et al. (2021), we specify conditional conjugate priors for the sampling distribution given the changepoint events. For the Gaussian structure in (2), the following conditional conjugate normal-inverse-gamma prior is defined for the segment parameters:

$$j|j > m \sim \text{InvGamma}(a; b); \quad j|j > j > m \sim \text{Normal}(d; c \cdot j^2);$$

with the same positive shape (a), scale (b, c) and real valued location (d) hyper-parameters for each segment.

From Bayes' Theorem, the posterior distribution for the micro-level changepoints and segment parameters, up to a constant of proportionality, given the observations on the current macro-level segment is:

$$(m; \alpha; \beta; \gamma | \mathbf{y}_i) \propto \prod_{j=1}^m \prod_{t \in B_{i,j}} f(y_t | j; \mu_j, \sigma_j^2) (\mu_j; \alpha; \beta | m; \gamma) (\mu_j, \sigma_j^2 | m); \quad (4)$$

Here $f(\cdot)$ is the probability density function for the data (Normal distribution) and (\cdot) are the prior densities for the parameters. We can leverage the conjugate structure of the segment parameters to obtain a marginalized form from (4) that focuses only on the

change point events:

$$\begin{aligned}
p(m; \mathbf{y}_i) &= \int \int (m; \mathbf{y}_i; \mathbf{t}_i) d \mathbf{t}_i d \mathbf{y}_i \\
&\propto \int \int \prod_{j=1}^m \prod_{t \in B_{i,j}} f(y_{tj}; \mathbf{t}_j) (\mathbf{t}_j; \mathbf{y}_{i,j}; m_j) (\mathbf{y}_i; m) d \mathbf{t}_j d \mathbf{y}_j \\
&= \prod_{j=1}^m \int \int \prod_{t \in B_{i,j}} f(y_{tj}; \mathbf{t}_j) (\mathbf{t}_j; \mathbf{y}_{i,j}; m_j) d \mathbf{t}_j d \mathbf{y}_j (\mathbf{y}_i; m) \\
&= \prod_{j=1}^m f_j(\mathbf{y}_{i,j}; m_j) (\mathbf{y}_i; m); \tag{5}
\end{aligned}$$

where the marginalized sampling distribution for the j th segment, $f_j(\mathbf{y}_{i,j}; m_j)$, is analytically available in closed form. The principle advantage of doing this is to reduce the dimensionality of the posterior to a finite, albeit potentially large, support such that the MCMC sampling scheme focuses on the change point events (see Section 2.2.2). Posterior inference for the segment parameters, either marginalized or conditionally on a particular sampled change point event, can easily be evaluated in a post processing step.

2.2.2 MCMC Sampler

Methods for performing inference for the micro-level change points need to account for the variable dimension to allow more or fewer events to be specified to describe the underlying process. A common approach to perform inference in this setting is Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) where one can explore the variable dimension of the micro-level change point vector by considering a proposal that either: partitions a randomly selected segment into two by adding a new micro-level change point; combines a sampled segment with the next and remove a change point; or, update the position of a change point and maintain the current length of the vector. In addition to the birth-death-move proposal, in Section 4.3 of Green (1995) considered a fourth possible transition that provides an update to the parameters of sampled segment, but by focusing inference on the marginal posterior (5) makes this case redundant for our application.

The sample space for the proposal distribution in Taylor et al. (2021) considers a slight

variation on the birth and death transitions to account for restrictions in navigating the sample space imposed by the minimum segment length. For a death transition a sampled segment is combined with the following segment, but position of the changepoint for the new combined segment is allowed to vary up to l_1 time-steps earlier. A corresponding adaptation to the birth transition is made where the changepoint of the sampled segment can be moved up to l_1 later time-steps before determining the position of a new changepoint that partitions the potentially extended segment. This adaptation to the proposal distribution increases the number of possible scenarios that the chain can move to at each iteration and reduces the possibility for the chain getting stuck in a minor mode. For a more detailed explanation, please refer to Section 3 of Taylor et al. (2021).

The support for the proposal distribution under the described extension to the range of transitions contains fewer than $N(l_1 + 1)$ unique changepoint vectors. As this set is of modest size, we define the proposal distribution to be proportional to the targeted posterior over this set. Specifically, for the current sample $(m; \mathbf{y})$ we first uniformly sample a segment j^θ from the range $1:::m$ and subsequently define the support for the proposal distribution under the extended birth-death-move transitions by the set $P_j(m; \mathbf{y})$. The proposal $(m^\theta; \mathbf{y}^\theta)$ is thereby sampled from:

$$g(m^\theta; \mathbf{y}^\theta | m; \mathbf{y}; j^\theta) = \frac{(m^\theta; \mathbf{y}^\theta | j^\theta)}{K_{j^\theta}(m; \mathbf{y})}; \quad \text{for } (m^\theta; \mathbf{y}^\theta) \in P_j(m; \mathbf{y}); \quad (6)$$

with $K_{j^\theta}(m; \mathbf{y})$ denoting the normalizing constant. If the sampled proposal is the same as the current sample, then the chain is updated with the proposal with acceptance probability of 1 but the value of the chain remains unchanged. Otherwise the proposal is accepted with acceptance probability that is a function of the normalizing constants:

$$A(m^\theta; \mathbf{y}^\theta | m; \mathbf{y}; j^\theta) = \min \left\{ 1; \frac{m K_{j^\theta}(m; \mathbf{y})}{m^\theta K_{j^{\theta\theta}}(m^\theta; \mathbf{y}^\theta)} \right\}; \quad (7)$$

where $j^{\theta\theta}$ denotes the unique and only segment under the proposed sample that gives a valid returning path to the current sample such that $g(m; \mathbf{y} | m^\theta; \mathbf{y}^\theta; j^{\theta\theta})$ is non-zero.

An added complexity to the aforementioned MCMC scheme regards the transitioning

into and out of the single segment model, $m = 1$. This case describes a constant data sampling scheme throughout the period where no micro-level changepoints need defining, whereas the neighboring two segment model $m = 2$ requires the definition of two changepoints (the intuition here is analogous to cutting a slice of pizza/cake - you need two cuts, not one). This means that the birth transition from the single segment model could be to any one of the $N(N - 1) = 2$ two segment models. As a result, the size of the sample space of the proposal distribution for this unique case is significantly increased, but not overly too large for many practical period lengths to have a substantial negative impact on the performance of the MCMC scheme. One final difficulty related to the single segment case is that it does not matter whether the first or second segment is sampled when proposing a death transition from the two segment model. This can easily be rectified by applying a minor adaptation to (7) the defining the acceptance probability for transitioning into the single segment model by:

$$A(1; j|2; \cdot) = \min \left\{ 1; \frac{2K_1(2; \cdot)K_2(2; \cdot)}{K_1(1; \cdot)[K_1(2; \cdot) + K_2(2; \cdot)]} \right\}; \quad (8)$$

The corresponding probability for the inverse action of transitioning from the single segment case, $A(2; \cdot|1; \cdot)$, is given by inverting the ratio inside (8). This ensures that all the properties required for RJMCMC are satisfied.

From a practical level, in order to assess convergence of our RJMCMC sampler, we initiate two chains, one under the assumption of no within-period changepoints and the other from a saturated case based on the minimum segment length condition. By setting the chains from two opposite and extreme scenarios we expect the true posterior distribution lies between these two cases and so aim to converge from opposite directions. Convergence is deemed to be acceptable if samples pass the two-sample Kolmogorov-Smirnov test for both the number of changepoints and separately for their positions. Whilst we could use a more sophisticated test of discrete sample chains such as those described in Deonovic and Smith (2017), the additional computational time to run the test is prohibitive as we run these chains for each potential macro-level changepoint combination.

2.3 Macro-level changepoint model

The ability to identify macro-level changepoints crucially depends on the underlying model of micro-level changes. These are necessarily inter-twinned. The goal of the macro-level changepoint detection is to identify when, in linear time, the parameters of the underlying micro-level changepoint model have changed. This could be a change in the mean/variance in the micro-level segments, and/or a change in the location/number of micro-level changepoints. We do not prescribe the type of change we expect to see. In this way, the macro-level changepoint detection is identical to the aims of traditional linear time multiple changepoint approaches. The only difference is that the complexity of the model within each segment is more complicated (now being the micro-level changepoint model).

To identify multiple changepoints, a common approach is use a cost function $\mathcal{C}(\cdot)$ that provides a measure of fit for the observations on a given segment, where the optimal case lies at the minimum. The aim is to identify the changepoint events that minimize:

$$\sum_{i=1}^{q+1} [\mathcal{C}(y_{(i-1)+1:s_i})] + q \quad (9)$$

Here, the cost function is evaluated on each macro-level segment and added together with q representing a linear penalty term in respect to the number of macro-level changepoint events to guard against over-fitting. Assuming conditional independence of data across the segments, we can calculate the cost associated with the identified segment, using the micro-level changepoint model, and consequently, use the results as the measure of fit. The parameters associated with each segment observed in the micro-level of the algorithm are assumed to be independent. The data points within the segment, given set of parameters, $\hat{\mu}_{ij}, \hat{\sigma}_{ij}^2$ are also independent and identically distributed and represented by the density function $f(y_j^{\hat{\mu}}) = f(y_j | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2)$.

The cost function for the current segment, indexed by i , following from the last proposed macro event $i-1$ up to time S is given by the micro-level maximum a posteriori likelihood:

$$\mathcal{C}(y_{(i-1)+1:s}) = 2 \sum_{j=1}^{m_i} \sum_{t \in B_{ij}} \log f(y_t | \hat{\mu}_{ij}, \hat{\sigma}_{ij}^2)$$

where $B_{i,j}$ denotes the set of intervals given by (1) based on the segment according to the proposed macro changepoints $i-1$ and s , and the estimated micro-level changepoints \hat{i} . The micro-level estimates \hat{m}_i , \hat{i} and $\hat{f}_{i,j}^2 \hat{g}_{i,j}^{\hat{m}_i}$ denote the posterior mode estimates from the inference detailed in Section 2.2 based on the available data for the given macro-level segment. Specifically:

$$\hat{f}_{i,j}^2 \hat{g}_{i,j}^{\hat{m}_i} = \arg \max_{(m_i, j) \in (i-1+1):s} f(y_{(i-1+1):s}; m_i, j) \quad \text{and} \quad (10)$$

$$\hat{f}_{i,j}^2 \hat{g}_{i,j}^2 = \arg \max_{(i, j) \in (i-1+1):s} f(y_{(i-1+1):s}; \hat{m}_i, \hat{i}, j) \quad \text{for } j = 1, \dots, \hat{m}_i \quad (11)$$

The log $f(y_j^{\hat{i}})$ are observed as we pass through the data with minimum segment length being equivalent to the length of two periods or $2N$. We recommend using $2N$ as a default as it is vital to pass through enough information for the micro-level model to fit adequately (i.e. 2 or more periods) before suggesting a macro-level changepoint.

Now that we have framed the global optimization procedure for both macro- and micro-level changepoints any generic multiple changepoint search algorithm could be used to identify the macro-level changepoints with the micro-level optimization nested within. Examples of suitable macro-level approaches include, Binary Segmentation (Scott and Knott, 1974), PELT (Killick et al., 2012), SMUCE (Frick et al., 2014). Note that due to the complexity of the likelihood for each segment it is not clear how to apply Wild Binary Segmentation as the combination process relies on results that do not apply to this setting. In the remainder of this paper we embed the micro-level changepoint identification within the PELT algorithm. We choose PELT as it provides an exact optimization of (9) within a computationally efficient framework. However, we wish to stress that this is merely an example and any of the aforementioned algorithms could be applied. Our final algorithm for identifying both micro- and macro-level changepoints within the PELT algorithm is given in Algorithm 1.

Algorithm 1 Macro-Micro Circular Changepoint Algorithm

Input: A time series of the form, (y_1, y_2, \dots, y_n) where $y_i \geq \mathbb{R}$.
A period length N .
A measure of micro-level fit $\mathcal{C}(\cdot)$ dependent on the data.
A penalty β independent of the number and location of changepoints.

MICRO-LEVEL

Initialize: Set $(m^{(0)}, \cdot^{(0)})$ from the sample space, differently for 2 or more chains.

Iterate per chain For $r = 1, 2, \dots$ until chain convergence

1. Sample segment j' uniformly from $f_1, \dots, m^{(r-1)}g$.
2. Draw proposal (m', \cdot') from $g(m', \cdot' | m^{(r-1)}, \cdot^{(r-1)}, j')$.
3. With probability $A(m', \cdot' | m^{(r-1)}, \cdot^{(r-1)})$:
Accept proposal and set $(m^{(r)}, \cdot^{(r)}) = (m', \cdot')$, or,
Reject proposal and set $(m^{(r)}, \cdot^{(r)}) = (m^{(r-1)}, \cdot^{(r-1)})$.

Output: Maximum a posteriori estimates \hat{m} , $\hat{\cdot}$ and $\hat{f}\hat{\mu}_j, \hat{\sigma}_j^2 g_{j=1}^n$ from post burn-in samples.

MACRO-LEVEL

Initialize: Let $n =$ length of time series
Set $F(0) = \beta$, $cp_{micro}(0) = cp_{macro}(0) = 0$, $R_{2N} = f_0, Ng$.

Iterate For $\alpha^* = 2N, \dots, n$

1. Calculate $\mathcal{C}(y_{(\cdot+1): \cdot})$ and $\hat{\cdot}(\alpha)$ for each $\alpha \geq R^*$ using the MICRO-LEVEL algorithm.
2. Calculate $F(\alpha^*) = \min_{\alpha \in R_{\alpha^*}} [F(\alpha) + \mathcal{C}(y_{(\cdot+1): \cdot}) + \beta]$.
3. Let $\alpha^1 = \arg \{ \min_{\alpha \in R_{\alpha^*}} [F(\alpha) + \mathcal{C}(y_{(\cdot+1): \cdot}) + \beta] \}$.
4. Set $cp_{macro}(\alpha^*) = [cp_{macro}(\alpha^1), \alpha^1]$ and $cp_{micro}(\alpha^*) = [cp_{micro}(\alpha^1), \hat{\cdot}(\alpha^1)]$.
5. Set $R_{\cdot+1} = \alpha^* \setminus f\alpha \geq R^* : F(\alpha) + \mathcal{C}(y_{s+1:T}) = F(\alpha^*)g$.

Output: The change points recorded in $cp_{micro}(n)$ and $cp_{macro}(n)$.

3 Simulation Study

When considering designing simulations there are several possible factors to consider. The structure of the micro-level and the ability to detect these changes as well as the macro-level changes which could result from a) micro-level changepoints moving; b) the means and variances of the micro-level segments varying; c) additions or deletions of micro-level changes. We design scenarios to cover these possibilities systematically to allow differentiation between the effects of each.

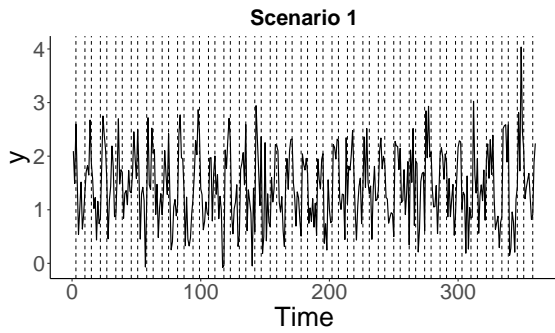
Twenty cases with varying positions of micro and macro changepoints as well as varying periodic means and variances were simulated according to the details in Appendix A. For

all scenarios we simulate 30 years of data, recorded monthly, giving us a total of $n = 360$ for each scenario and $N = 12$. For many changepoint data sets these may be considered as small data lengths but these were chosen to mimic the applications we have commonly seen in practice and consider in Section 4. In all the figures depicting micro and macro changes, dashed lines depict the periodic positions, with bold lines representing macro changes. As there are no existing micro-macro methods to compare to, we compare to the PELT search algorithm assuming a Normal distribution with varying mean and variance per segment, and the WBS2 method designed for identifying frequent changepoints in the mean structure only. The PELT approach with SIC penalty is fit using the `changepoint` R package (Killick et al., 2016) and the WBS2 with recommended Steepest Drop to Low Levels (SDLL) penalty using the `breakfast` R package (Anastasiou et al., 2021).

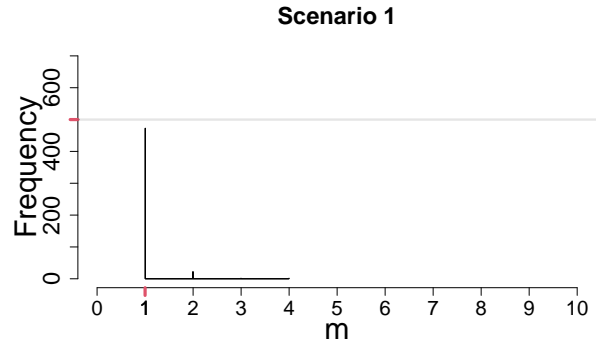
We present the summary of accuracy of changepoints identification given our set up for distinct scenarios using histograms that report the number of resulting segments and respective positions observed on 500 simulations. We focus below on four scenarios, namely Scenarios 1, 2, 9, and 20 as examples of the types of scenarios we encounter. Remaining scenarios can be found in Supplementary Materials. Note that on all graphics, the axis ticks and light gray horizontal lines indicate the true values and frequencies that were simulated.

From the simulation results, one can note that among the 500 cases, the true number of segments were picked up accurately, on average, with a slight tendency to suggest a greater number of segments (+1=2). With regards to the exact positions of macro- and micro-level changepoints, there are small uncertainties around values of the exact position and the modes are around the true values as expected. Similarly, we can see a satisfactory behavior of micro-level changepoint positions with a small uncertainty in cases where the shift is less drastic (i.e. micro-level positions shift from (2,10) to (3,11)).

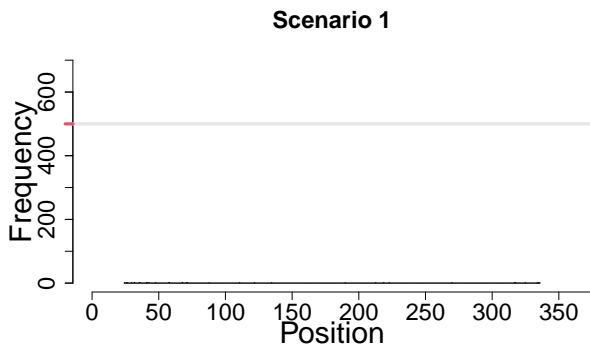
For the first scenario where no macro-level changepoints are simulated, it is evident that the micro-level model is capable of picking up the truth consistently (Figure 3). Looking at the scenarios with the micro-level mean changes, the behavior is also very consistent with the truth. Furthermore, the overall performance with respect to position of a macro-level changepoint is improved when both micro-level changepoints and the segment means have



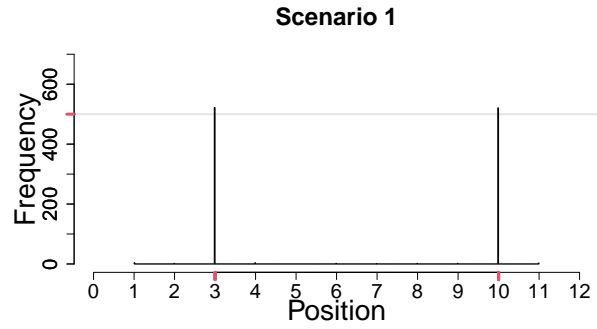
(a) Realization



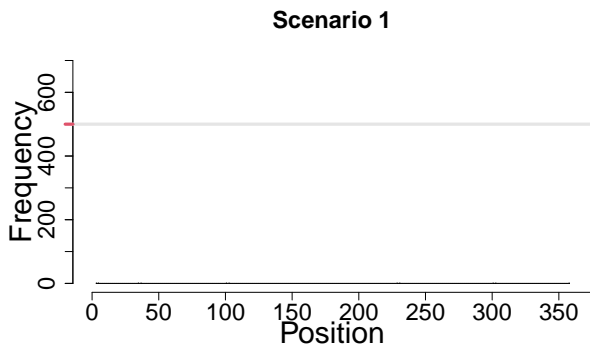
(b) Number of Segments



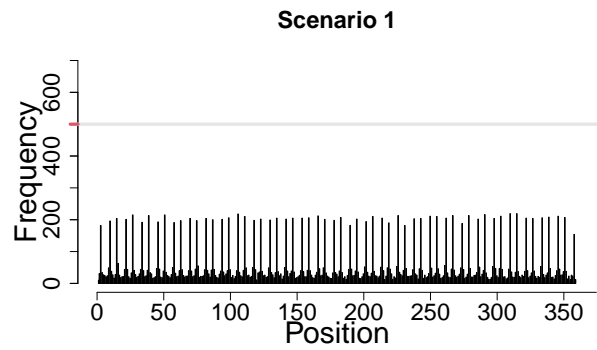
(c) Macro Positions



(d) Micro Positions

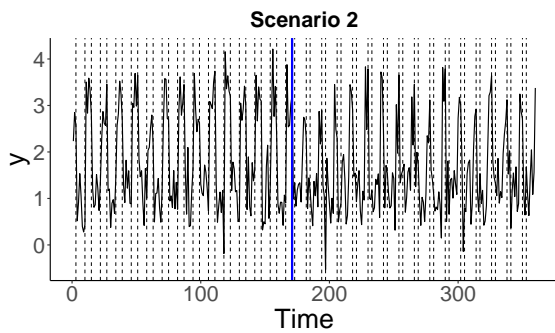


(e) Cpt positions (Normal Mean & Var + PELT)

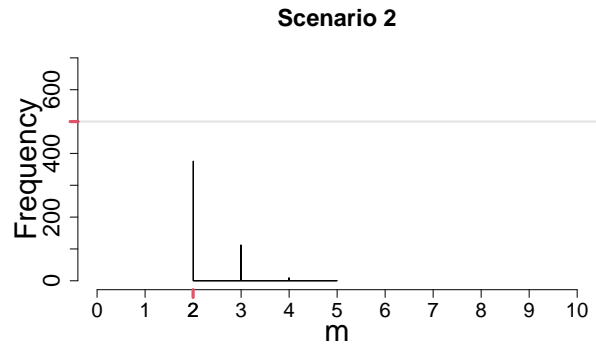


(f) Cpt positions (Mean + WBS2)

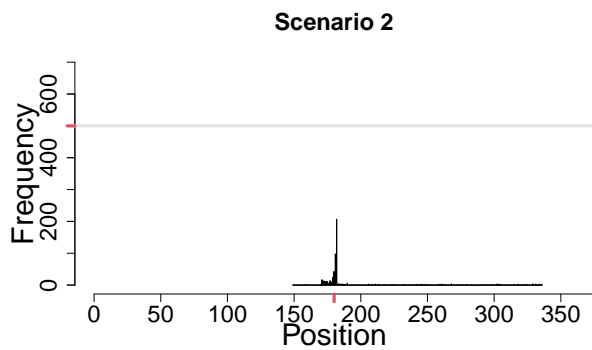
Figure 3: Simulation results for scenario 1: no macro-changes, micro-changes at months 3 and 10



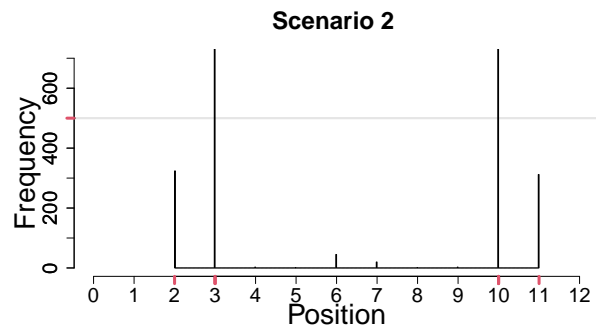
(a) Realization



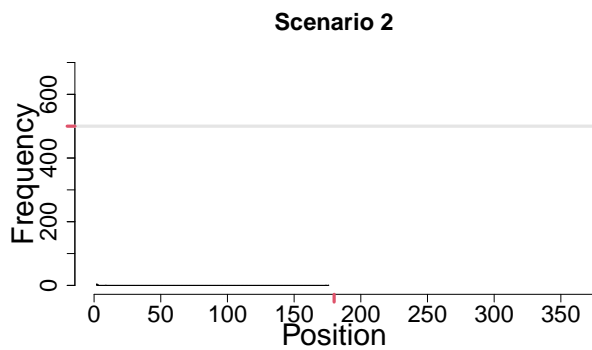
(b) Number of Segments



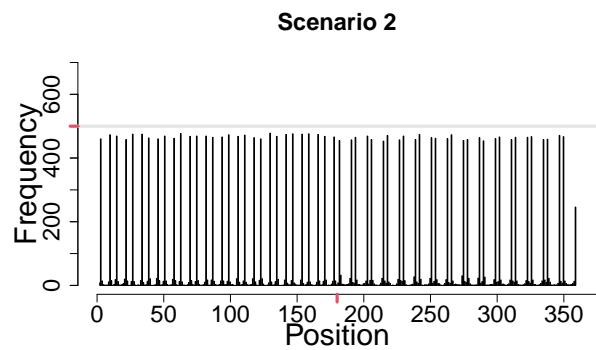
(c) Macro-change Positions



(d) Micro-change Positions

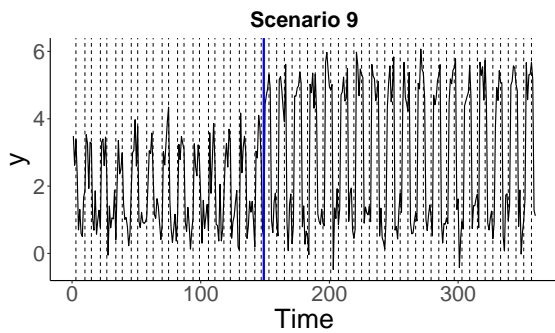


(e) Cpt positions (Normal Mean & Var + PELT)

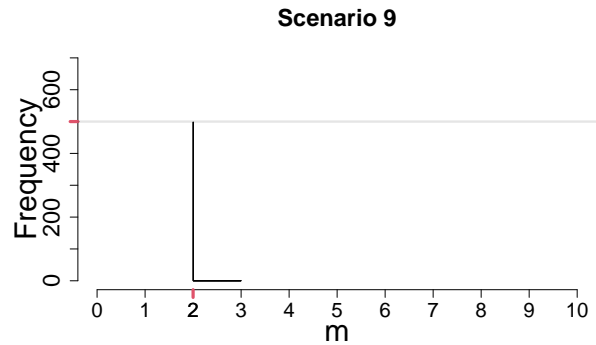


(f) Cpt positions (Mean + WBS2)

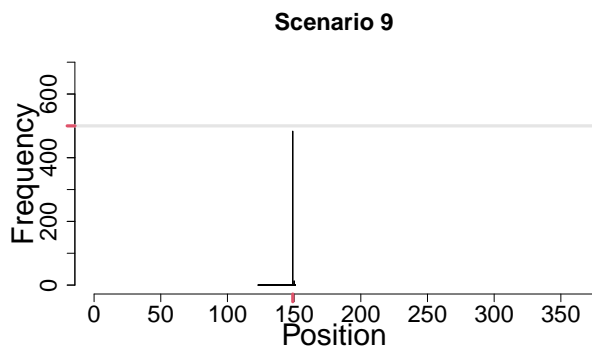
Figure 4: Simulation results for scenario 2: 1 macro-change at 15th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



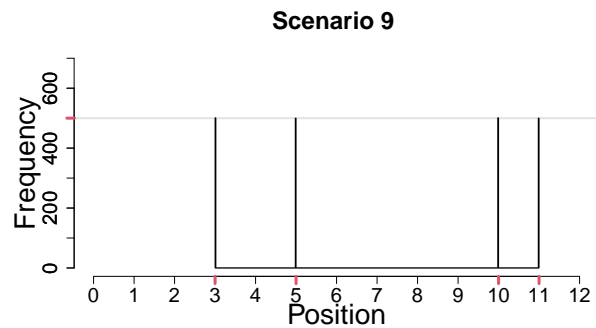
(a) Realization



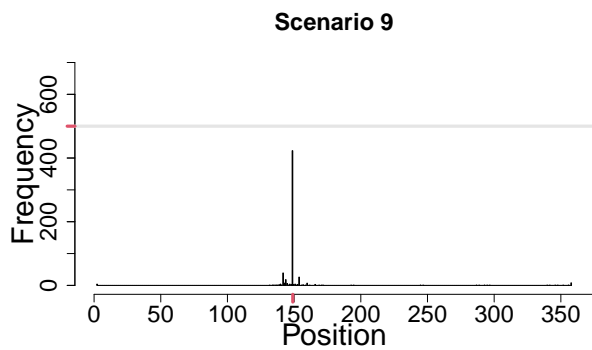
(b) Number of Segments



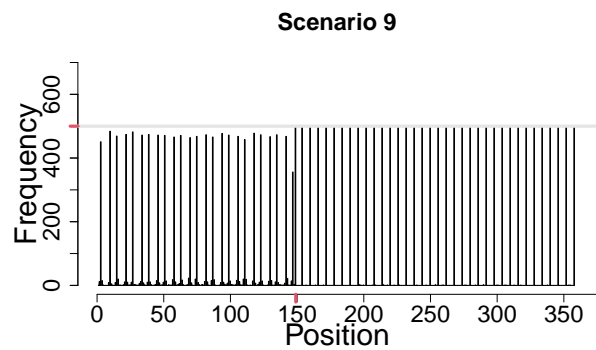
(c) Macro-change Positions



(d) Micro-change Positions



(e) Cpt positions (Normal Mean & Var + PELT)



(f) Cpt positions (Mean + WBS2)

Figure 5: Simulation results for scenario 9: 1 macro-change at 5th month of 13th period, micro-changes at months (3,10) and (5,11) for segments 1 and 2 respectively.

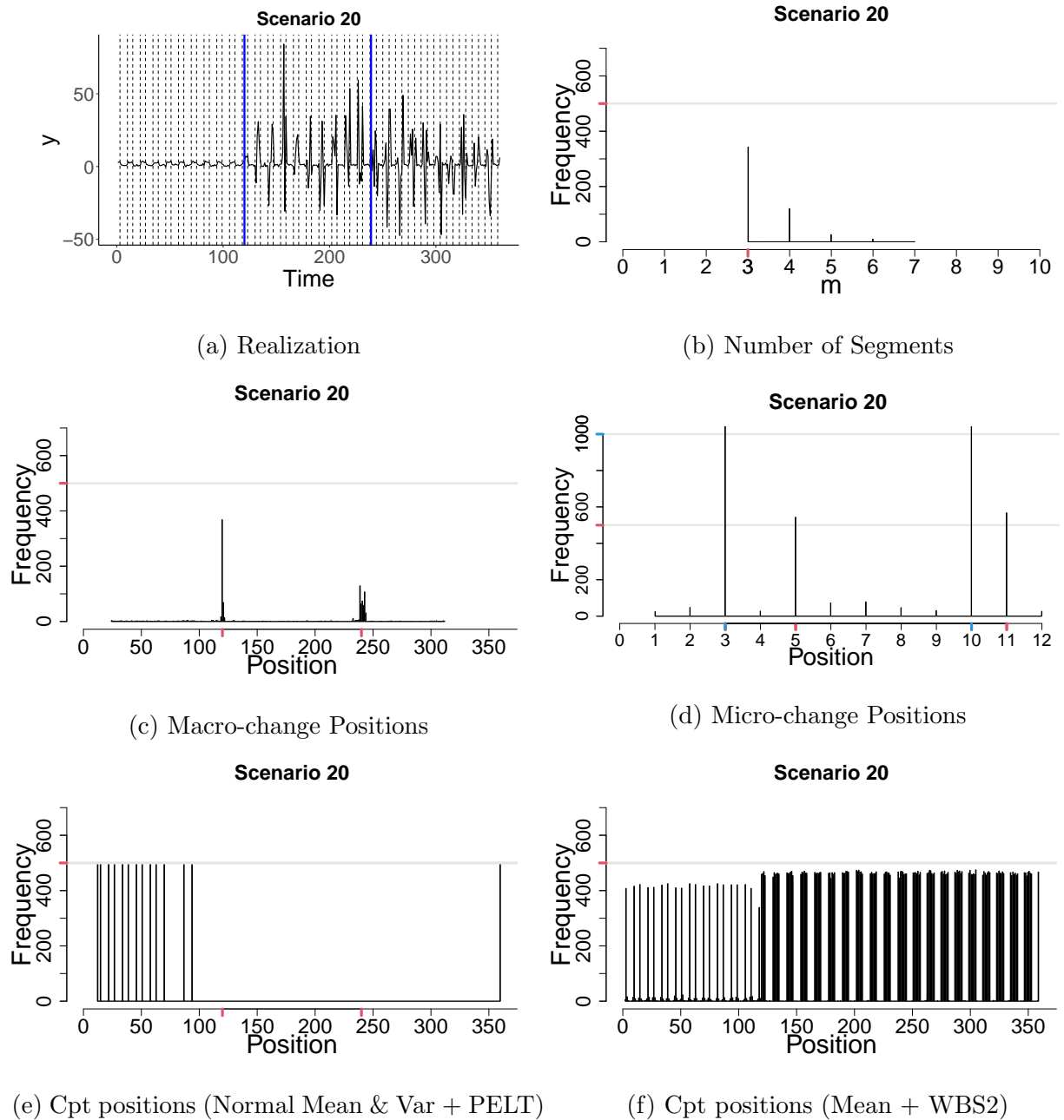


Figure 6: Simulation results for scenario 20: 2 macro-change points at 15th period and 20th period, micro-changes at months (3,10) for segments 1, 2, and (5,11) for the 3rd segment. The ticks on the axes indicate which true positions should occur with which frequency.

changed (Scenario 9 in Figure 5. This is to be expected as it signifies a larger change in the distribution than a change in either the positions or the means as with the cases where both mean and variance change (Scenario 20 in Figure 6).

Where one expects a macro-level changepoint due to a change in micro-level changepoint positions without changes in segment means, we find a slightly larger uncertainty around the exact positions of the macro-level changepoint (Scenarios 2-4). For instance, when we have a relatively small micro-level change, from (2,10) to (3,11), the model is likely to pick up (3,10) as a new micro-level changepoint. Where the distance between micro-level changepoints is much more visible or abrupt (e.g., (3,10) to (5,11), and is accompanied by changes in mean in Scenario 9), the method correctly identifies the changepoint positions.

In scenarios where we have multiple macro-level changepoints (e.g. Scenario 20), the effect of a micro-level mean and variance change compared to varying micro-level changepoint locations is less visible and performance in those cases is fairly similar with respect to both macro- and micro-level changepoint identification.

It is crucial to note, that in all of the scenarios considered, the smallest number of segments identified by the model is consistently at the true value. Such implies that even though we occasionally have cases of false positives, there are no occurrences of false negatives.

Each set of simulations were also studied using PELT with Normal mean & variance cost function (Normal MV+PELT) and WBS2 for mean change (Mean+WBS2) approaches. In contrast to our approach the Normal MV+PELT and Mean+WBS2 linear time methods highlight different changes. Normal MV+PELT tends to favor identifying the macro changes in most scenarios whilst Mean+WBS2 the micro-changes. Mean+WBS2 fails to identify the changes in scenario 1 and tends to falter around the macro changepoint locations, likely due to the disruption between micro and macro changes. Nevertheless, Mean+WBS2 can often identify the micro-changes fairly accurately, this is not taking into account the periodic nature of the data and so the structure is not estimated in a parsimonious way - estimating c.120 parameters where the micro-macro approach would estimate 8, for example.

4 Applications

4.1 North Atlantic Oscillation

One of the motivating applications in the development of our approach is the North Atlantic Oscillation dynamics and similar environmental data. The North Atlantic Oscillation (NAO) is a meteorological phenomenon and is measured by obtaining the differences in atmospheric pressures at sea level between the Icelandic Low and the Azores High. We obtained data on the NAO, aggregated to the monthly level, from <https://crudata.uea.ac.uk/cru/data/nao/> Climatic Research Unit (2019). The data is depicted in Figure 7a.

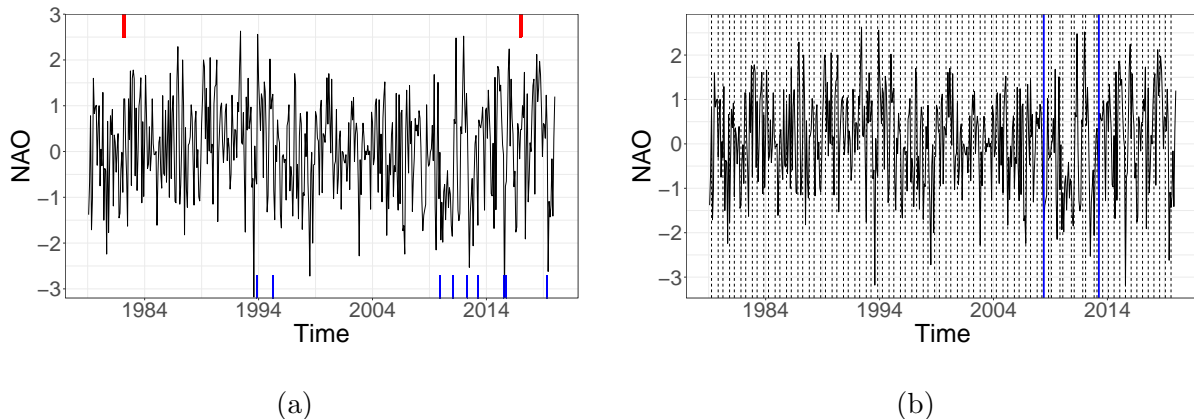


Figure 7: (a) NAO data with Normal MV+PELT (upper ticks) and Mean+WBS2 (bottom ticks) changepoints and (b) estimated micro-macro changepoint positions. The Normal MV+PELT and micro-macro approach use $pen = 4 \log(n)$ and Mean+WBS2 the data-adaptive SDLL penalty.

Recent research suggests that it might be appropriate to characterize the NAO by winter/summer differences (Sansom et al., 2019). There is a further evidence of decadal trends presented in NAO fluctuations studied in Stephenson et al. (2000). However, with respect to changes in periodic nature, there are mixed views on whether such change is random or may have a pattern.

We fit the yearly winter/summer pattern using a change in mean and variance using Algorithm 1. For real-world applications the assumptions made in simulations may not hold. Thus it is prudent to conduct a sensitivity analysis on the penalty value used. Figure

25a (Appendix) depicts this sensitivity analysis plot for the likelihood as a function of varying penalty. Considering both the sensitivity analyses, standard penalty values, and conversations with Environmental Scientists, we choose two macro-level changepoints for the NAO index. The resulting segmentation is given in Figure 7b.

Macro-level segment	Micro-level changepoints
1979-2008	April and October
2008-2013	May and August
2013-2019	April and October

Table 1: Estimated macro- and micro-level changepoint positions for the NAO index.

We note the full changepoint segmentation in Table 1. Here we can observe expected periodicity (switch between early spring/early autumn) for a period throughout 1979-2008. The change in periodicity to late spring and late summer is observed during the period 2008-2013, just before switching back to April and October for the remaining years up to 2019. The last change in the periodicity (2013-2019) has received attention in climate literature indicating that there is evidence for winter variability change in NAO in the recent decade (Hanna et al., 2015). With respect to decadal trend the last two macro-level changepoints are in line with evidence discussed in Stephenson et al. (2000).

The methodology we propose offers a novel angle to analyze the dynamics of such periodic data based on meteorological observations. This is an important contribution to address the evidence of changes in the reoccurring structure of the data (Stephenson et al., 2000).

4.2 House Price Index

For the second case study, we use data for monthly House Price Index (HPI) in the UK for almost 20 years (2002-2019) available freely from UK Land Registry Registry (2020) <https://landregistry.data.gov.uk/app/ukhpi>. The data is displayed in Figure 8a and is known to exhibit periodic behavior (summer/winter) as well as changes in the overall data sequence structure due to external influences, such as macro economic factors or changes in the housing market. The proposed methodology offers inference across multiple levels

in order to capture macro-level changepoints and micro-level changepoints to provide a comprehensive modeling framework. Further background and details on the challenges in modeling of house prices dynamics can be found in Yusupova (2016).

We fit the yearly winter/summer pattern using a change in mean and variance using Algorithm 1. The macro- and micro-level changepoints are depicted in Figure 8b with exact changepoint positions given in the Table 2.

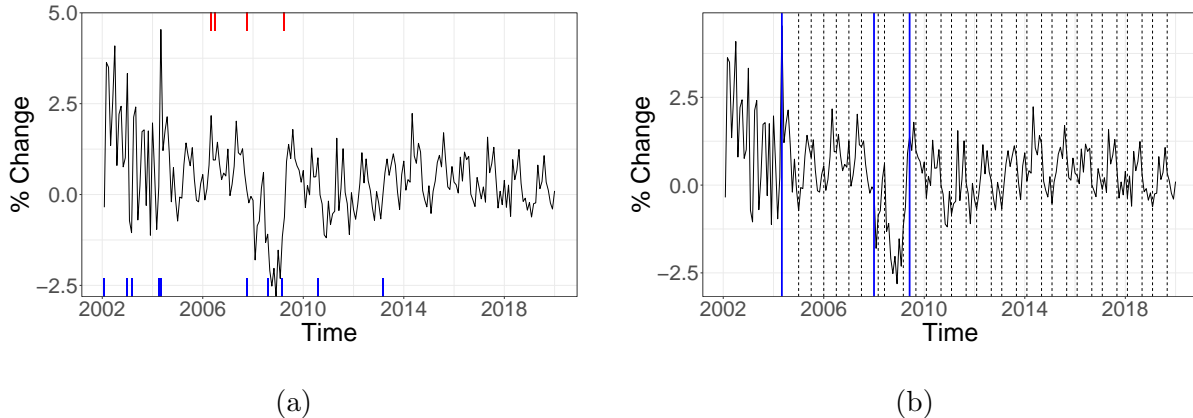


Figure 8: (a) HPI data with Normal MV+PELT (upper ticks) and Mean+WBS2 (bottom ticks) and (b) estimated micro-macro changepoints. The Normal MV+PELT and micro-macro approach use $pen = 3 \log(n)$ (SIC) and Mean+WBS2 the data-adaptive SDLL penalty.

The results suggest either no or two changes within the year alternating mostly between late winter/early spring and late summer. For example, in the years up to 2004, there is no evidence of micro-level periodicity. However, during 2004-2008 there is a proposal for two periods within the year (February and August). The sensitivity analyses for penalty values and likelihood are provided in Appendix (Figure 25b).

Macro-level segment	Micro-level changepoints
2002-2004	No micro changepoint
2004-2008	February and August
2008-mid 2009	February and May
mid 2009-2019	March and August

Table 2: Estimated macro- and micro-level changepoint positions for House Price Index Data.

The above period also precedes a significant break that is easy to observe in the index.

This break is likely to be attributed to 2008 recession with the house prices falling by around 16% in 2008. The proceeding macro-level segment is 2008 to mid-2009 with February and May changepoints. After 2009 the market recovers and the overall dynamic resumes a new regular pattern characterized by two micro-level changepoints in late winter and late summer (March and August).

The findings correspond with evidence on House Price Index dynamics found in the economics literature (Beghazi and Katsiampa, 2019). The results also align with the mean pattern identified in London boroughs in Fryzlewicz (2020) but with considerably fewer parameters and changepoints to estimate. In comparison to Normal MV+PELT and Mean+WBS2 methods, the 2008 and 2009 macro-level changes (driven by the recession and recovery) are complemented by the insights into periodic structures pre and post recession.

5 Conclusion

This paper has provided a novel approach for evaluation of the multi-scale changepoint structure, introducing micro- and macro-level changepoint identification. The central contribution of the method is an attempt to solve a common trade-off between having too many changepoints or too few, focusing on application that have micro- and macro-level structures to account for periodic changepoints. Such analysis is vital where one is interested in the stability and consistency of periodic changepoint behavior across time. Taking this micro-macro perspective can be crucial for understanding precise changes in the processes, and periodicity shifts which may be a sign of a much more complex long-term change whilst allowing parsimonious modeling. We envisage application areas to include sensor-based activities (i.e., smart meter data, movement data), environmental processes (e.g., moth counts, algae blooms) and economic applications (e.g., employment and production data).

Experimental results have illustrated that the method is accurate in identifying the true number of changepoints as well as their respective positions at micro and macro-levels.

Future research may consider theoretical dimension to the method presented taking

either fully frequentist or fully Bayesian setting. Another interesting dimension to explore would be uncertainty measures for macro-level changepoint estimation taking account of the uncertainty at the micro-level.

Supplemental Materials

R-code: The supplemental files for this article include R-code to obtain the micro-macro changepoints for given datasets along with functions to aid plotting. We include an R-script that generates all the plots for the simulated data examples and the applications. Please consult the file README contained in the zip file for more details.

Appendix: The Appendix contains full simulation details, additional simulation results and sensitivity analyses for our applications. (Appendix.pdf)

References

- Anastasiou, A., Chen, Y., Cho, H., and Fryzlewicz, P. (2021), *breakfast: Methods for Fast Multiple Change-Point Detection and Estimation*, R package version 2.2.
- Beghazi, K., and Katsiampa, P. (2019), “Modelling UK house prices with structural breaks and conditional variance analysis,” *The Journal of Real Estate Finance and Economics*, 58, 290–309.
- Climatic Research Unit, U. o. E. A. (2019), “NAO Index,” <https://crudata.uea.ac.uk/cru/data/nao/nao.dat>.
- Deonovic, B. E., and Smith, B. J. (2017), “Convergence diagnostics for MCMC draws of a categorical variable,” *arXiv*, 1706.04919.
- Eckley, I. A., Fearnhead, P., and Killick, R. (2011), in *Bayesian Time Series Models*, eds. Barber, D., Cemgil, T., and Chiappa, S., Cambridge University Press.

- Fearnhead, P. (2006), “Exact and efficient Bayesian inference for multiple changepoint problems,” *Statistics and Computing*, 16, 203–213.
- Frick, K., Munk, A., and Sieling, H. (2014), “Multiscale change point inference,” *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 76, 495–580.
- Fryzlewicz, P. (2020), “Detecting possibly frequent change-points: Wild Binary Segmentation 2 and steepest-drop model selection.” *J. Korean Stat. Soc.*, 49, 1027–1070.
- Green, P. J. (1995), “Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,” *Biometrika*, 82, 711–732.
- Hanna, E., Cropper, T. E., Jones, P. D., Scaife, A. A., and Allan, R. (2015), “Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index,” *International Journal of Climatology*, 35, 2540–2554.
- Killick, R., Fearnhead, P., and Eckley, I. A. (2012), “Optimal detection of changepoints with a linear computational cost,” *Journal of the American Statistical Association*, 107, 1590–1598.
- Killick, R., Haynes, K.— (2016), *changepoint: An R package for changepoint analysis*, R package version 2.2.2.
- Page, E. (1955), “A test for a change in a parameter occurring at an unknown point,” *Biometrika*, 42, 523–527.
- Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007), “A review and comparison of changepoint detection techniques for climate data,” *Journal of Applied Meteorology and Climatology*, 46, 900–915.
- Registry, U. L. (2020), “UK House Price Index,” <https://landregistry.data.gov.uk/app/ukhpi>, [Online; accessed 20-April-2020].

- Sansom, P. G., Williamson, D. B., and Stephenson, D. B. (2019), “State space models for non-stationary intermittently coupled systems: an application to the North Atlantic oscillation,” *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 68, 1259–1280.
- Scott, A. J., and Knott, M. (1974), “A Cluster Analysis Method for Grouping Means in the Analysis of Variance,” *Biometrics*, 30, 507–512.
- Stephenson, D. B., Pavan, V., and Bojariu, R. (2000), “Is the North Atlantic Oscillation a random walk?” *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 20, 1–18.
- Taylor, S. A. C., Killick, R., Burr, J., and Rogerson, L. (2021), “Assessing Daily Patterns using Home Activity Sensors and Within Period Changepoint Detection,” To Appear.
- Yusupova, A. Y. (2016), “An econometric analysis of UK regional real estate markets,” Ph.D. thesis, Lancaster University.

A Simulation Details

Scenario	$i(\text{Time} D)$	$i(\text{Period} D)$	i_1	i_2	i_3	i_1	i_2	i_3
1	0	0	3,10	3,10		2,1	2,1	
2	180	15	3,10	2,11		3,1	3,1	
3	149	12(5)	3,10	2,11		3,1	3,1	
4	180	15	3,10	3,10		3,1	5,1	
5	149	12(5)	3,10	3,10		3,1	5,1	
6	180	15	3,10	2,11		3,1	2,1	
7	149	12(5)	3,10	2,11		3,1	2,1	
8	180	15	3,10	5,11		3,1	5,1	
9	149	12(5)	3,10	5,11		3,1	5,1	
10	120,240	10,20	3,10	2,11	5,11	3,1	3,1	3,1
11	186,262	15(6), 21(10)	3,10	2,11	5,11	3,1	3,1	3,1
12	120,240	10,20	3,10	3,10	3,10	3,1	5,1	2,1
13	186,262	15(6), 21(10)	3,10	3,10	3,10	3,1	5,1	2,1
14	120,240	10,20	3,10	2,11	5,11	3,1	5,1	2,1
15	186,262	15(6), 21(10)	3,10	2,11	5,11	3,1	5,1	2,1
16	120,240	10,20	3,10	3,10	5,11	3,1	5,1	5,1
17	186,262	15(6), 21(10)	3,10	3,10	5,11	3,1	5,1	5,1

Table 3: Description of simulated yearly scenarios at monthly resolution for varying change-points positions and respective means. Each scenario was simulated 500 times. For the period index of the macro-level changepoints i $X(Y)$ indicates that it is the Y^{th} month within the X^{th} year.

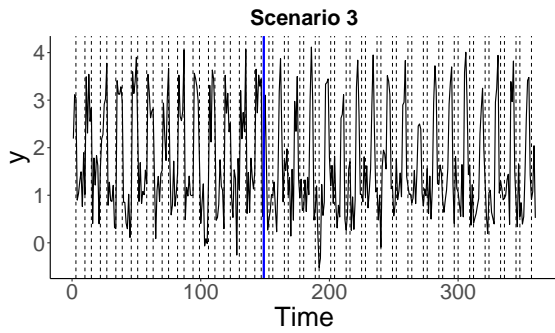
Scenario	Time	Period	i_1	i_2	i_3	i_1	i_2	i_3	$\sigma^2_{i_1}$	$\sigma^2_{i_2}$	$\sigma^2_{i_3}$
18	180	15	3,10	2,11		3,1	2,1		0.5,0.5	25,0.5	
19	149	12(5)	3,10	5,11		3,1	5,1		0.5,0.5	25,0.5	
20	120,240	10,20	3,10	3,10	5,11	3,1	5,1	5,1	0.5,0.5	25,0.5	25,0.5

Table 4: Description of simulated yearly scenarios at monthly resolution for varying change-points positions and respective means and variances. Each scenario was simulated 500 times.

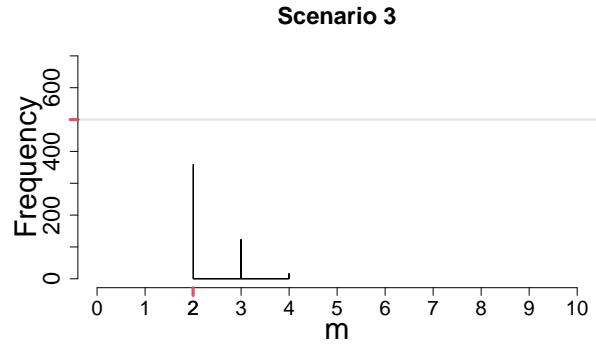
Each of the simulated scenarios shares constant characteristics with respect to the following parameters. The data are simulated from a Normal distribution with a changing mean as detailed in Table 3. The minimum segment length is fixed at $2N$ and the penalty value, λ , is fixed at $2 \log(n)$ for mean change only and at $3 \log(n)$ for mean and variance change, commonly used in changepoint identification. We simulate 30 years of data,

recorded monthly, giving us a total of $n = 360$ for each scenario and $N = 12$. The variance is held constant at $\sigma^2 = 0.5$ for scenarios where positions and means were varied (Table 3. The changes in both mean and variance were evaluated using extensions to scenarios 6, 9, 16. These are described in Table 4

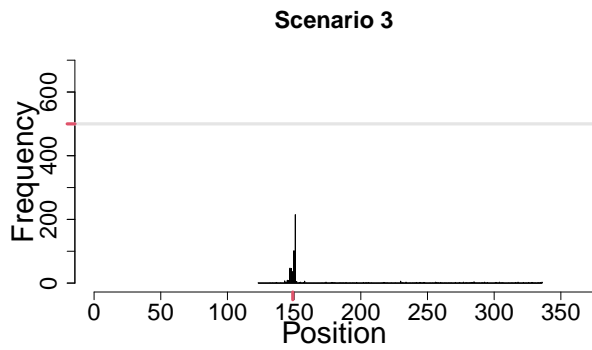
B Results from remaining simulation scenarios



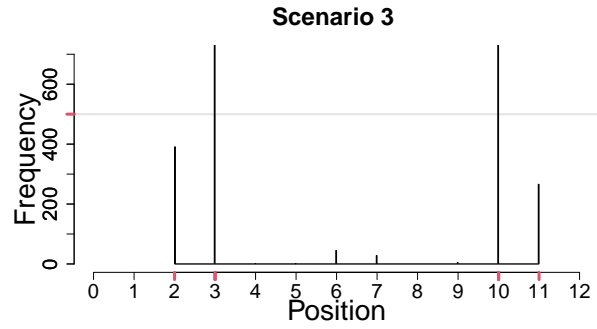
(a) Realization



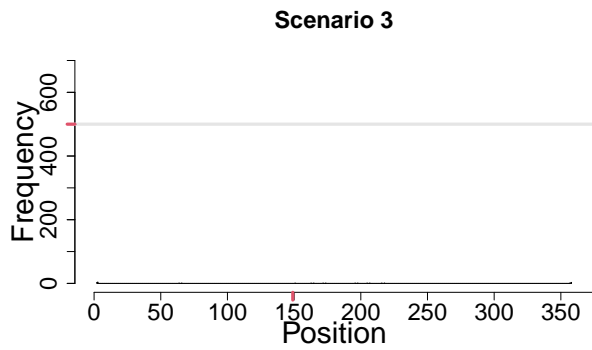
(b) Number of Segments



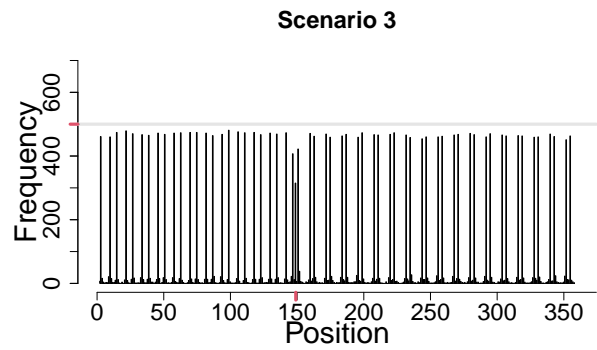
(c) Macro-change Positions



(d) Micro-change Positions

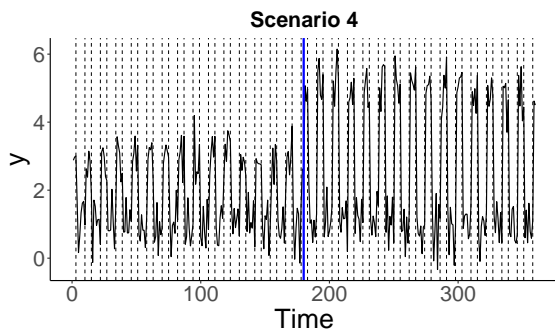


(e) Cpt positions (Normal Mean & Var + PELT)

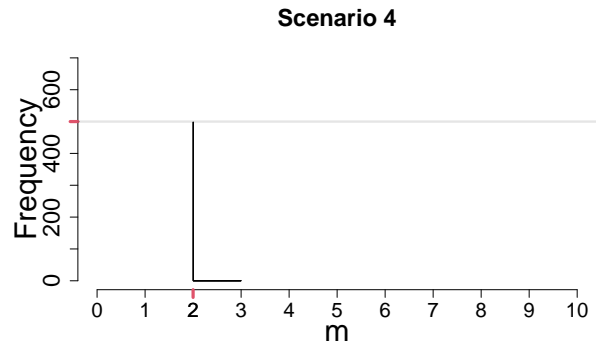


(f) Cpt positions (Mean + WBS2)

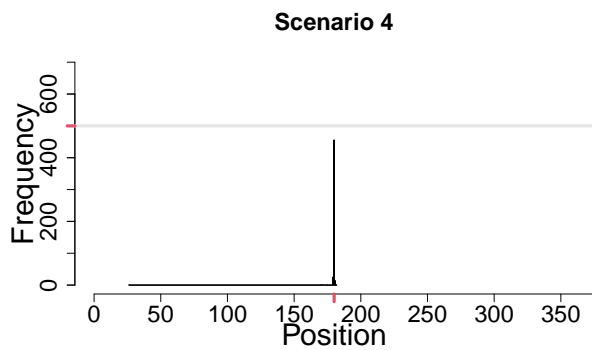
Figure 9: Simulation results for scenario 3: 1 macro-change at 15th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



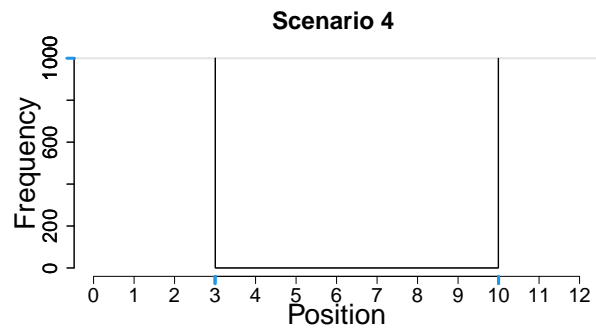
(a) Realization



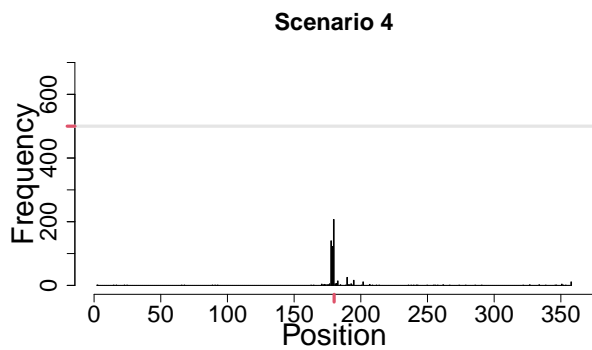
(b) Number of Segments



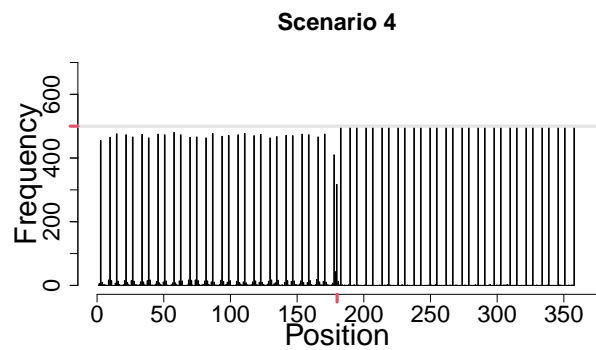
(c) Macro-change Positions



(d) Micro-change Positions

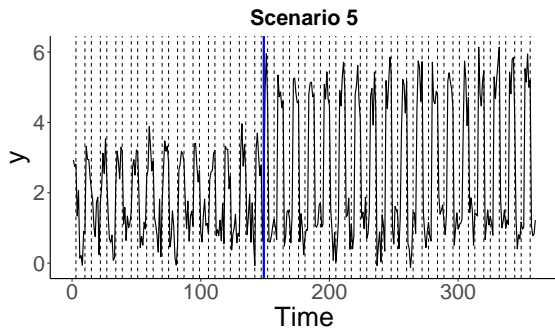


(e) Cpt positions (Normal Mean & Var + PELT)

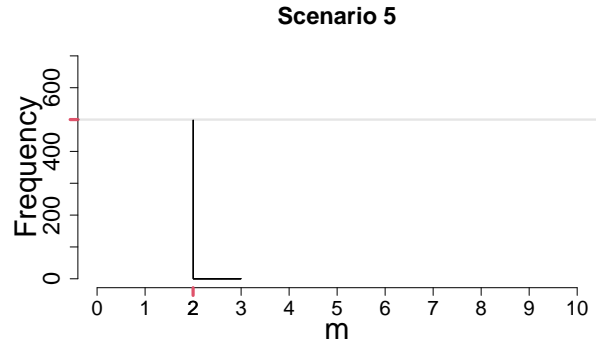


(f) Cpt positions (Mean + WBS2)

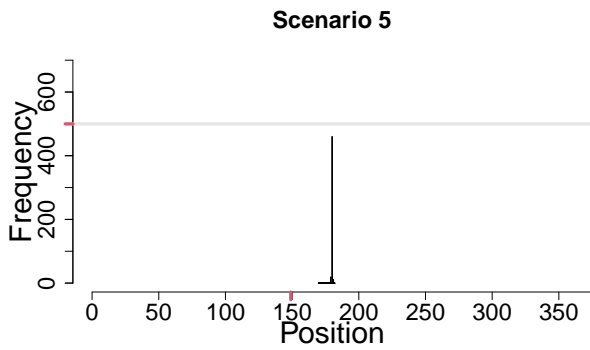
Figure 10: Simulation results for scenario 4:1 macro-change at 5th month of 13th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



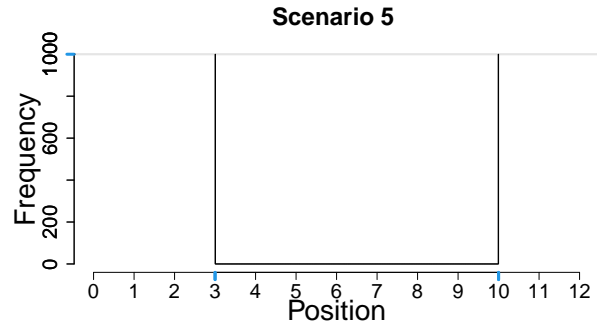
(a) Realization



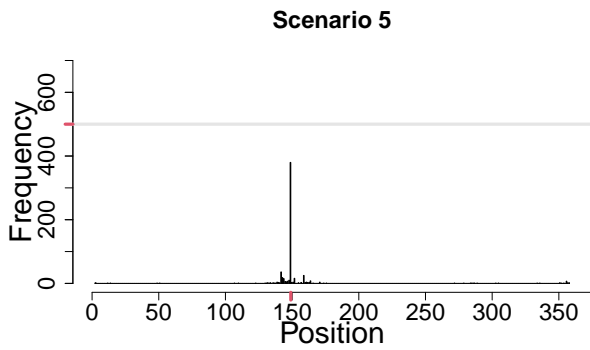
(b) Number of Segments



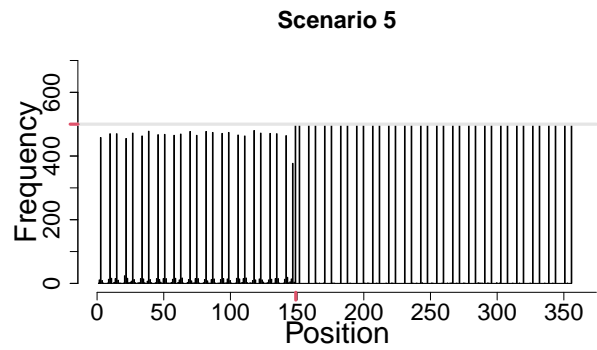
(c) Macro-change Positions



(d) Micro-change Positions

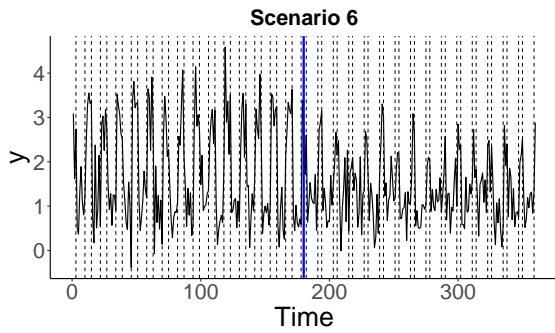


(e) Cpt positions (Normal Mean & Var + PELT)

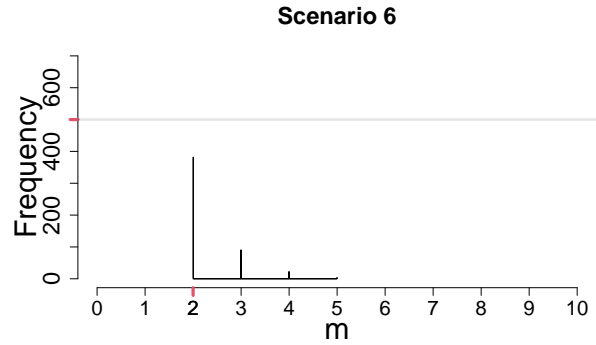


(f) Cpt positions (Mean + WBS2)

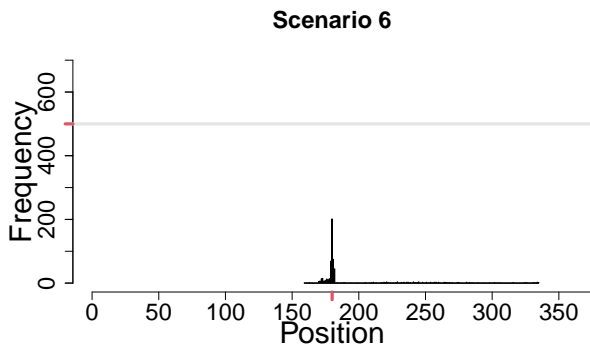
Figure 11: Simulation results for scenario 5: 1 macro-change at 5th month of 13th period, micro-changes at months 3 and 10 for both segments.



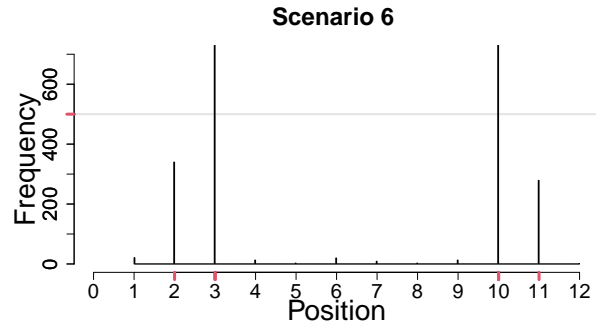
(a) Realization



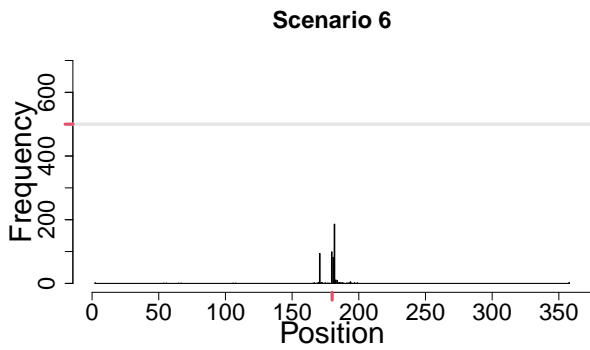
(b) Number of Segments



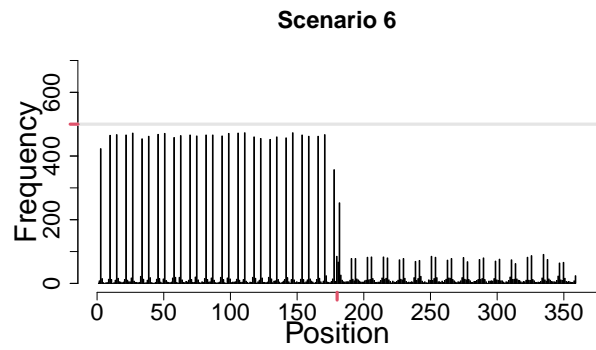
(c) Macro-change Positions



(d) Micro-change Positions

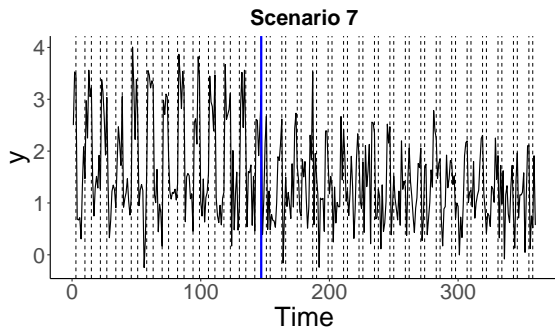


(e) Cpt positions (Normal Mean & Var + PELT)

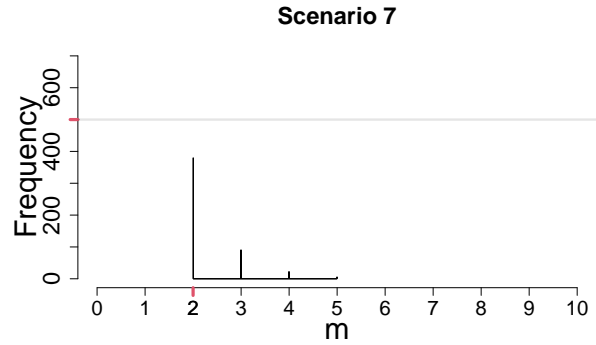


(f) Cpt positions (Mean + WBS2)

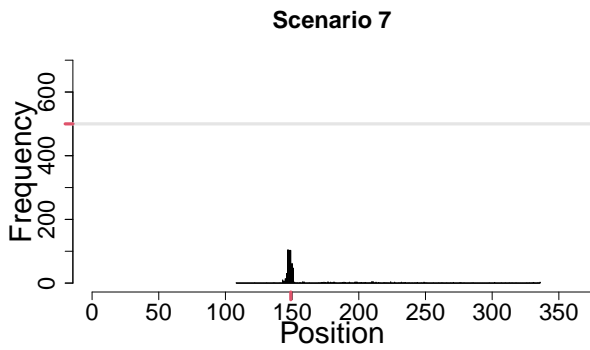
Figure 12: Simulation results for scenario 6: 1 macro-change at 15th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



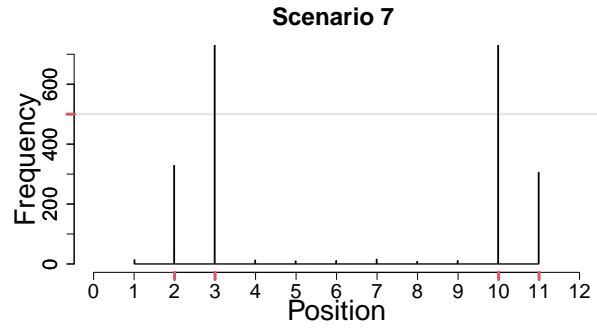
(a) Realization



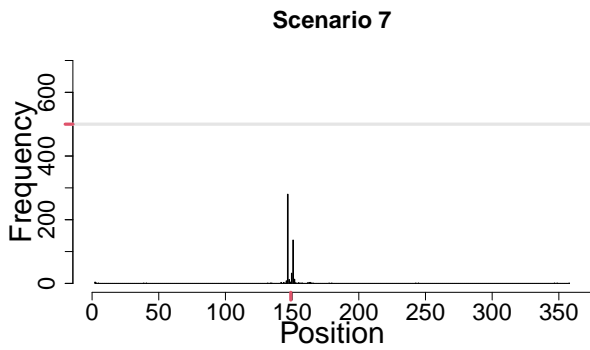
(b) Number of Segments



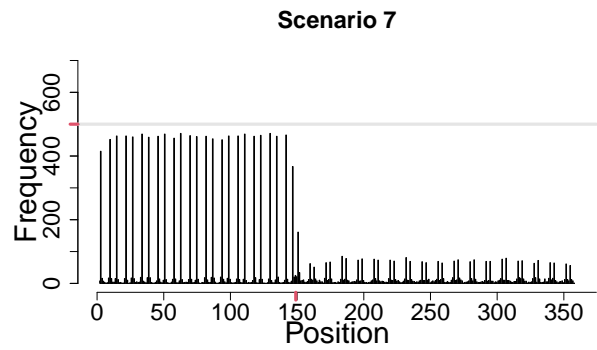
(c) Macro-change Positions



(d) Micro-change Positions

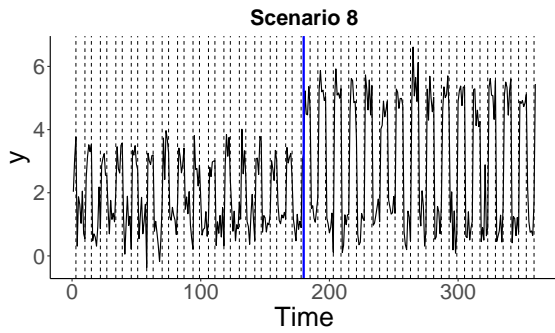


(e) Cpt positions (Normal Mean & Var + PELT)

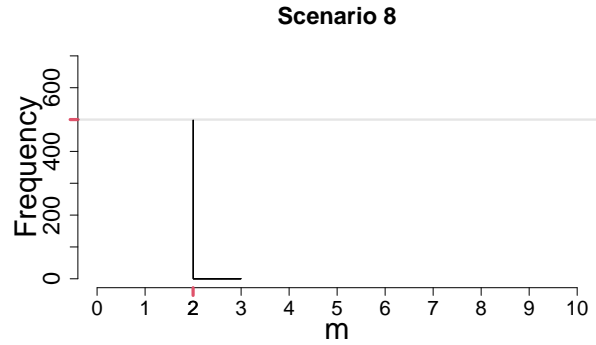


(f) Cpt positions (Mean + WBS2)

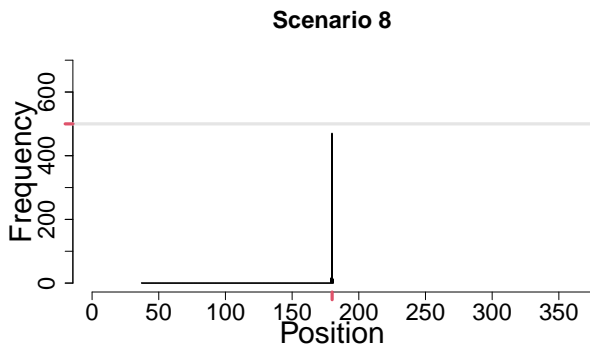
Figure 13: Simulation results for scenario 7: 1 macro-change at 5th month of 13th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



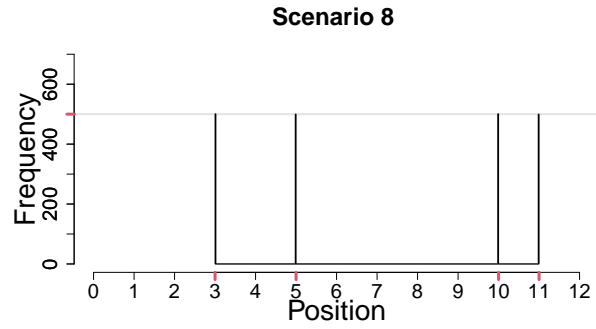
(a) Realization



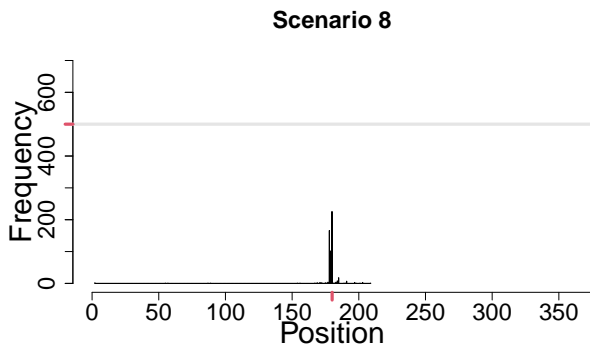
(b) Number of Segments



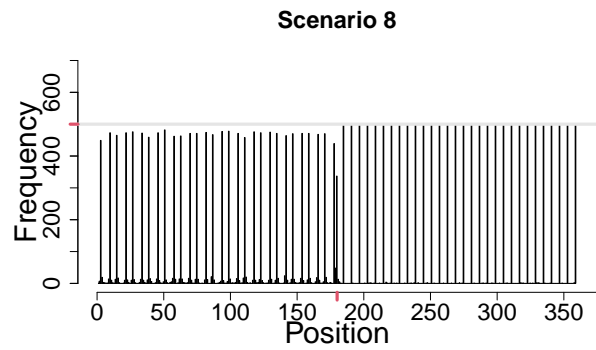
(c) Macro-change Positions



(d) Micro-change Positions

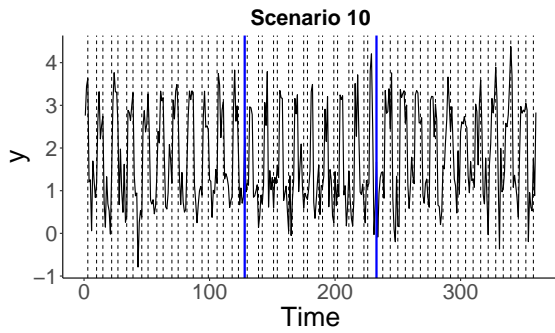


(e) Cpt positions (Normal Mean & Var + PELT)

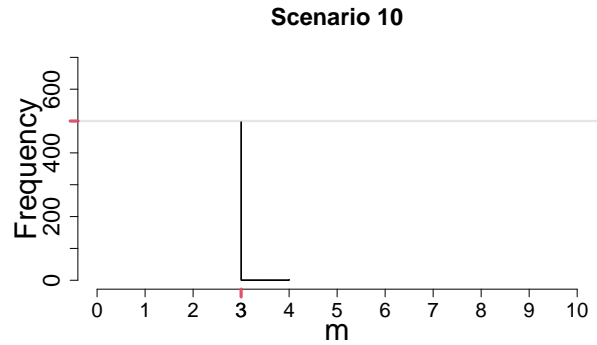


(f) Cpt positions (Mean + WBS2)

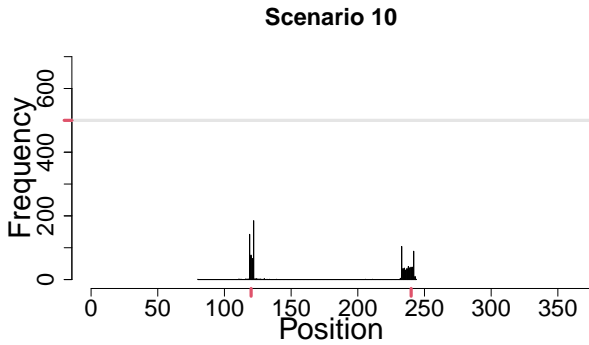
Figure 14: Simulation results for scenario 8: 1 macro-change at 15th period, micro-changes at months (3,10) and (5,11) for segments 1 and 2 respectively.



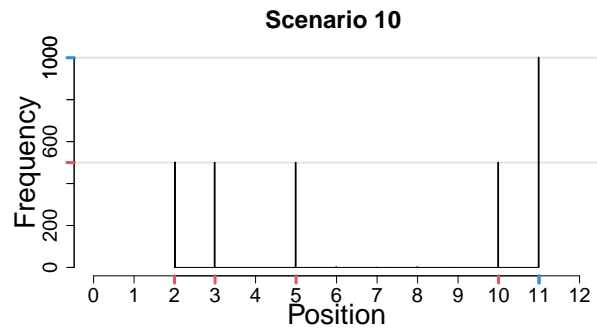
(a) Realization



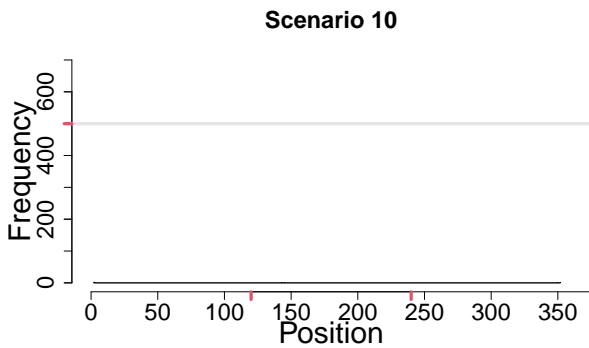
(b) Number of Segments



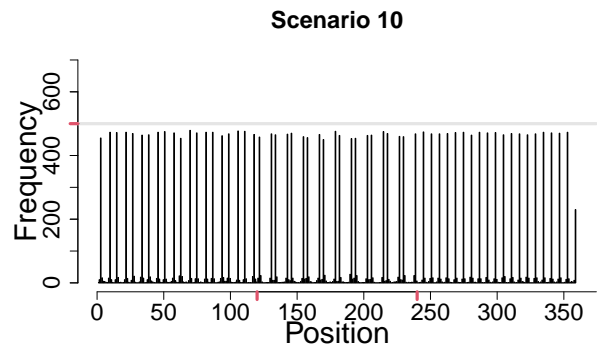
(c) Macro-change Positions



(d) Micro-change Positions



(e) Cpt positions (Normal Mean & Var + PELT)



(f) Cpt positions (Mean + WBS2)

Figure 15: Simulation results for scenario 10: 2 macro-change points at 10th period and 20th period, micro-changes at months (3,10) for segment 1 and (2,11), (5,11) for segments 2 and 3 respectively. The ticks on the axes indicate which true positions should occur with which frequency.

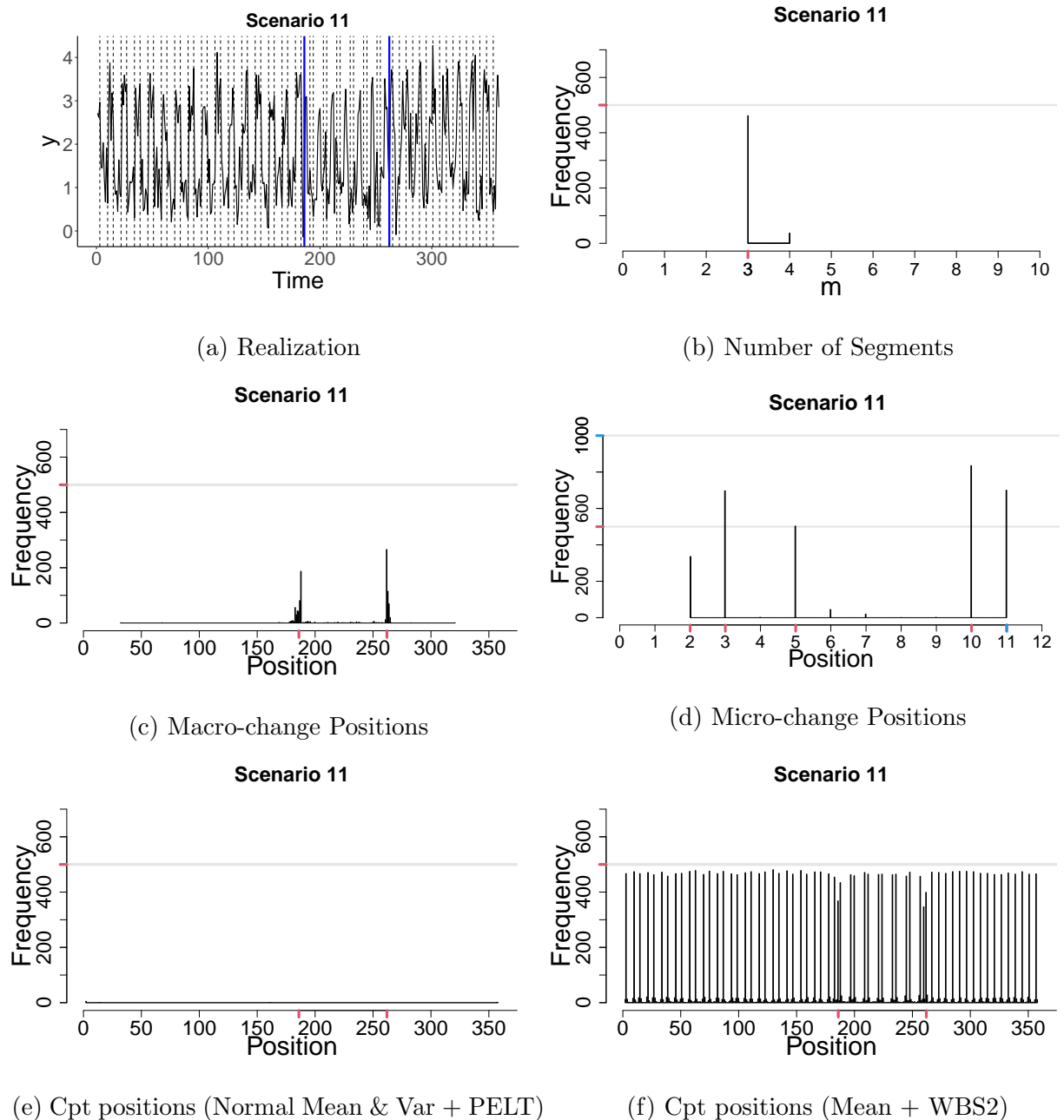
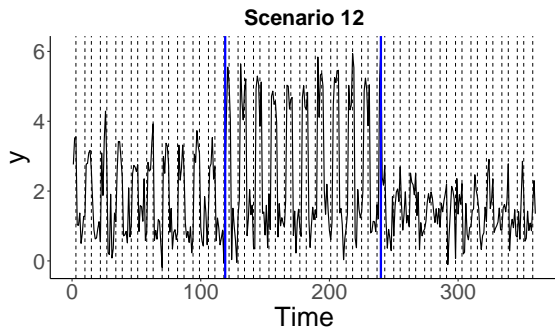
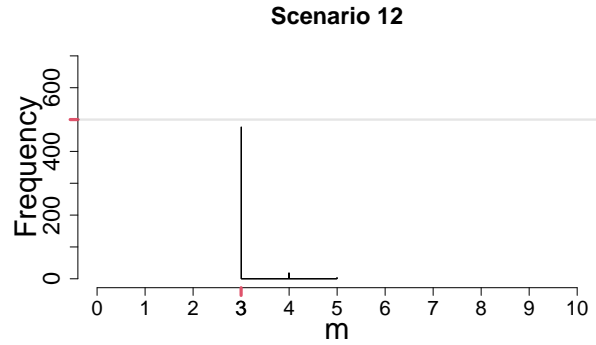


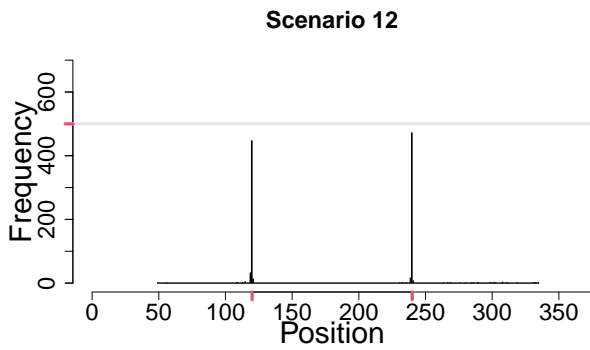
Figure 16: Simulation results for scenario 11: 2 macro-change points at 6th month of 16th period and 10th month of 22nd period, micro-changes at months (3,10) for segment 1, and (2,11), (5,11) for the 2nd and 3rd segments respectively. The ticks on the axes indicate which true positions should occur with which frequency.



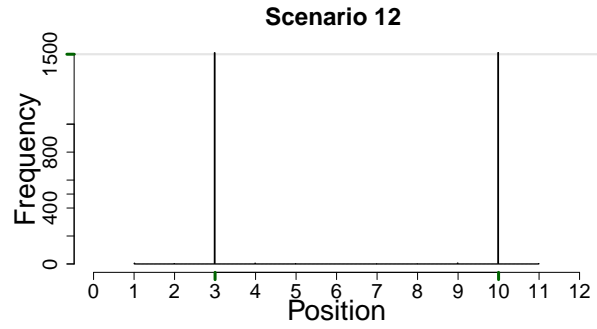
(a) Realization



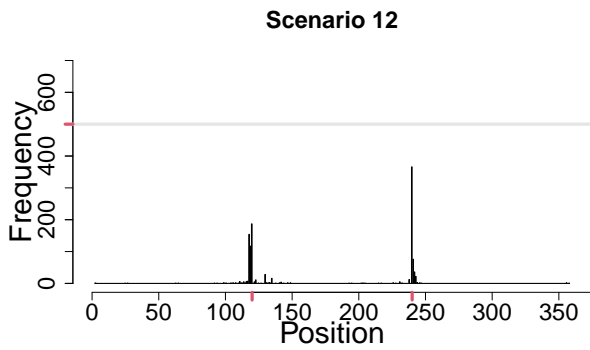
(b) Number of Segments



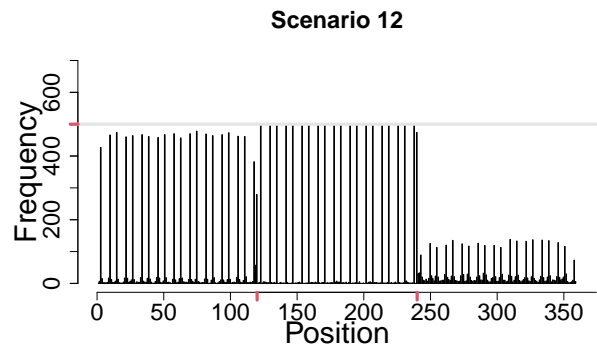
(c) Macro-change Positions



(d) Micro-change Positions



(e) Cpt positions (Normal Mean & Var + PELT)



(f) Cpt positions (Mean + WBS2)

Figure 17: Simulation results for scenario 12: 2 macro-change points at 10th period and 20th period, micro-changes at months (3,10) for segments 1, 2 and 3.

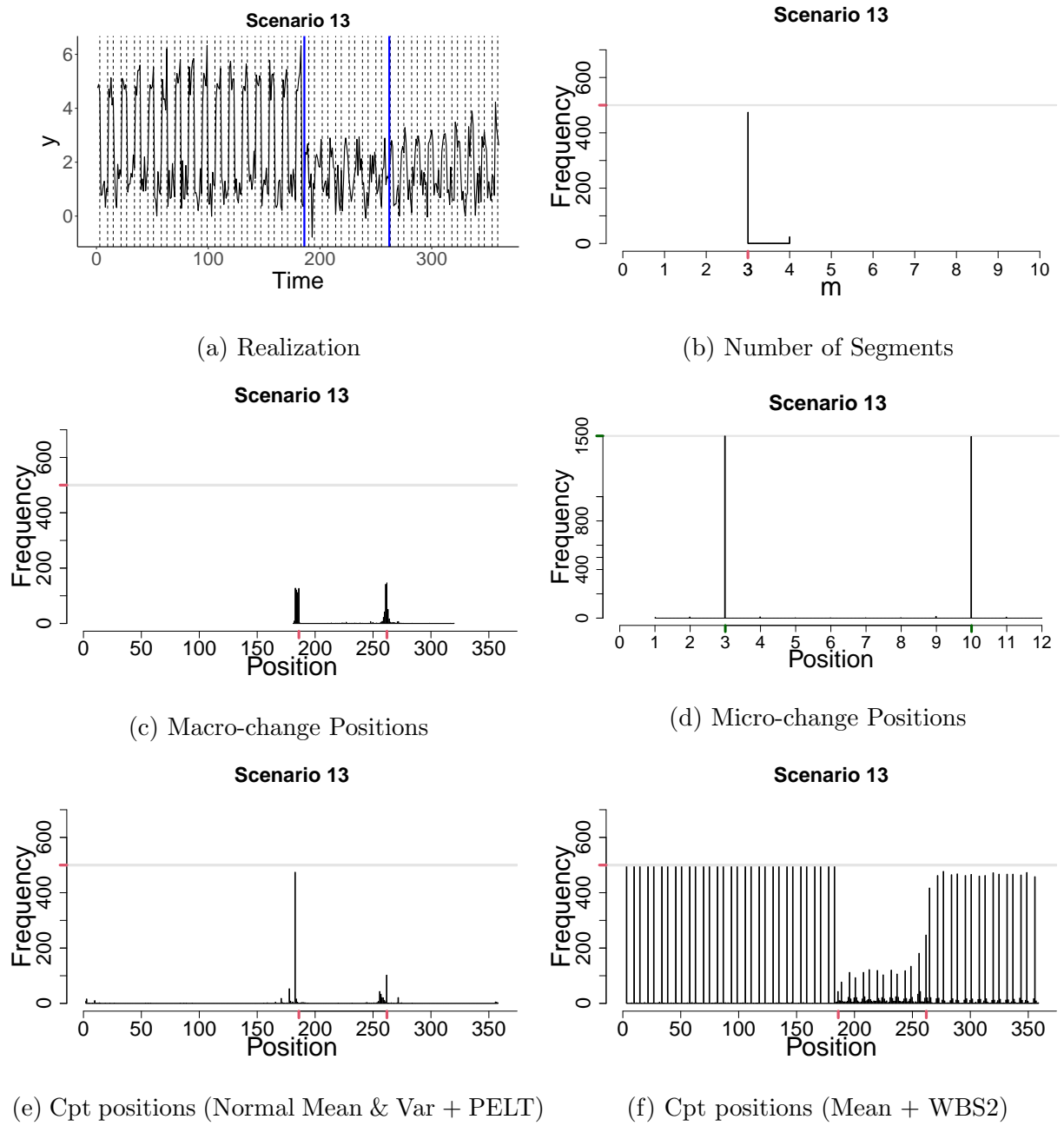


Figure 18: Simulation results for scenario 13: 2 macro-change points at 6th month of 16th period and 10th month of 22nd period, micro-changes at months (3,10) for segments 1, 2 and 3. The ticks on the axes indicate which true positions should occur with which frequency.

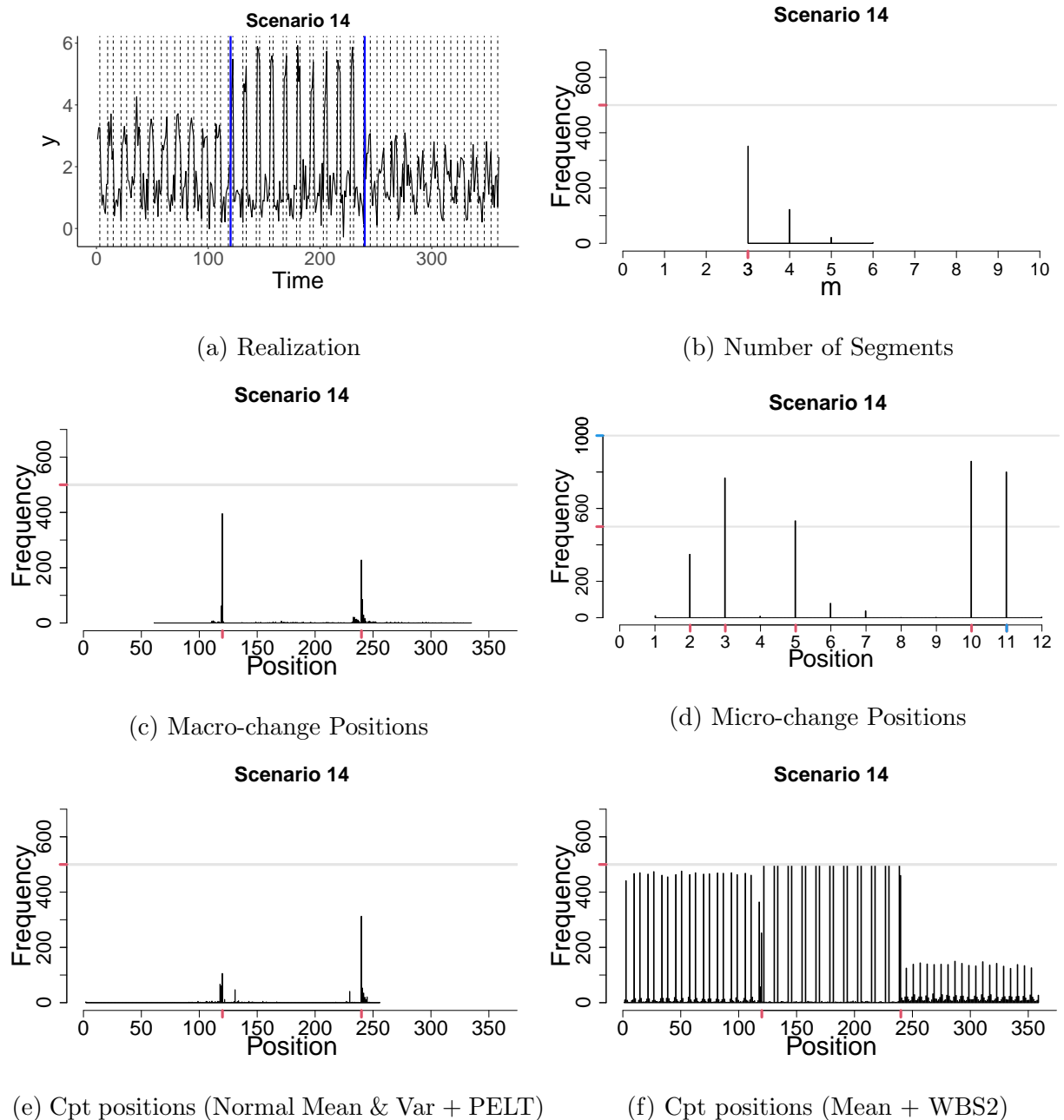


Figure 19: Simulation results for scenario 14: 2 macro-change points at 10th period and 20th period, micro-changes at months (3,10) for segment 1 and (2,11), (5,11) for the 2nd and 3rd segments respectively. The ticks on the axes indicate which true positions should occur with which frequency.

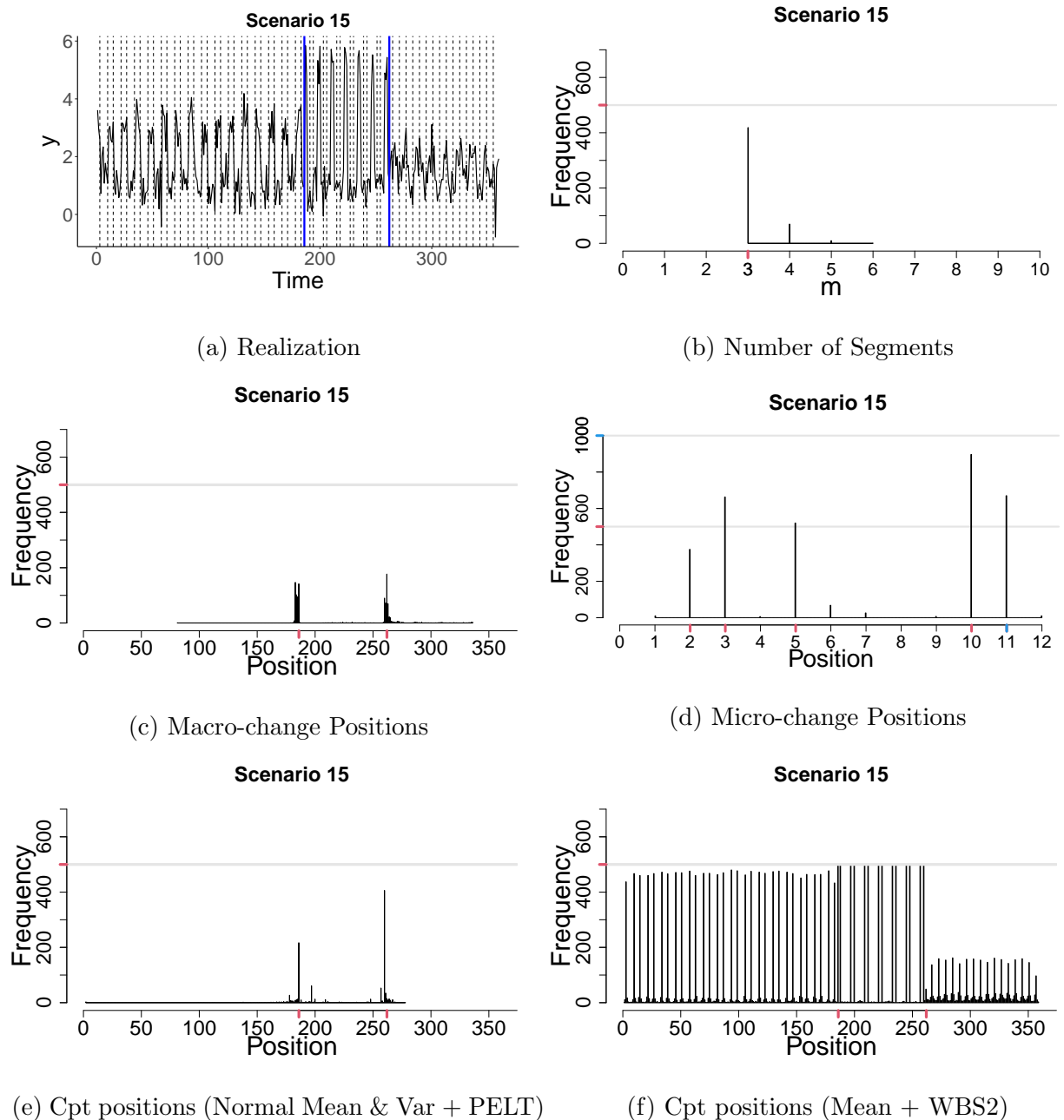


Figure 20: Simulation results for scenario 15: 2 macro-change points at 6th month of 16th period and 10th month of 22nd period, micro-changes at months (3,10) for segments 1, 2, and (5,11) for the 3rd segment. The ticks on the axes indicate which true positions should occur with which frequency.

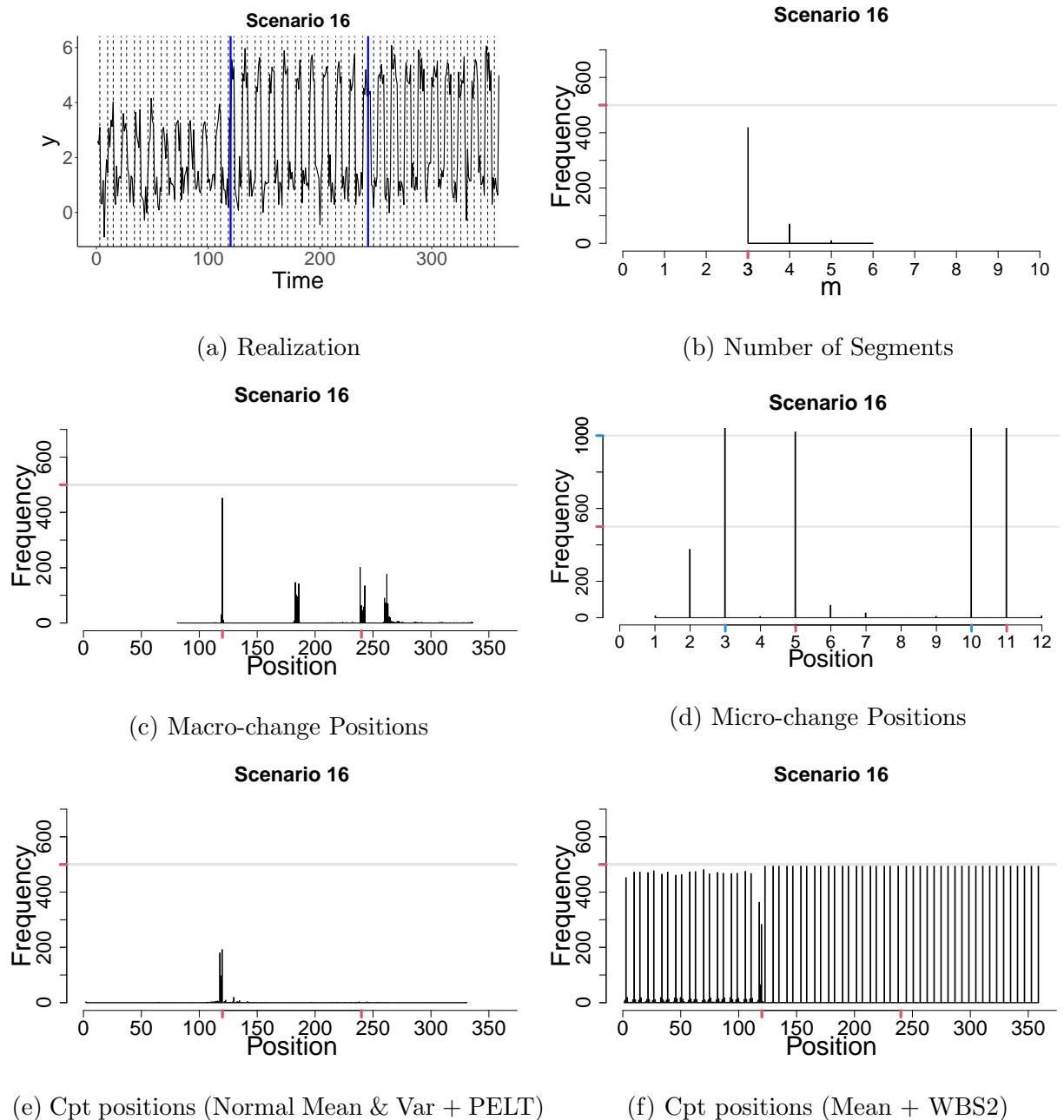


Figure 21: Simulation results for scenario 16: 2 macro-change points at 15th period and 20th period, micro-changes at months (3,10) for segments 1, 2, and (5,11) for the 3rd segment. The ticks on the axes indicate which true positions should occur with which frequency.

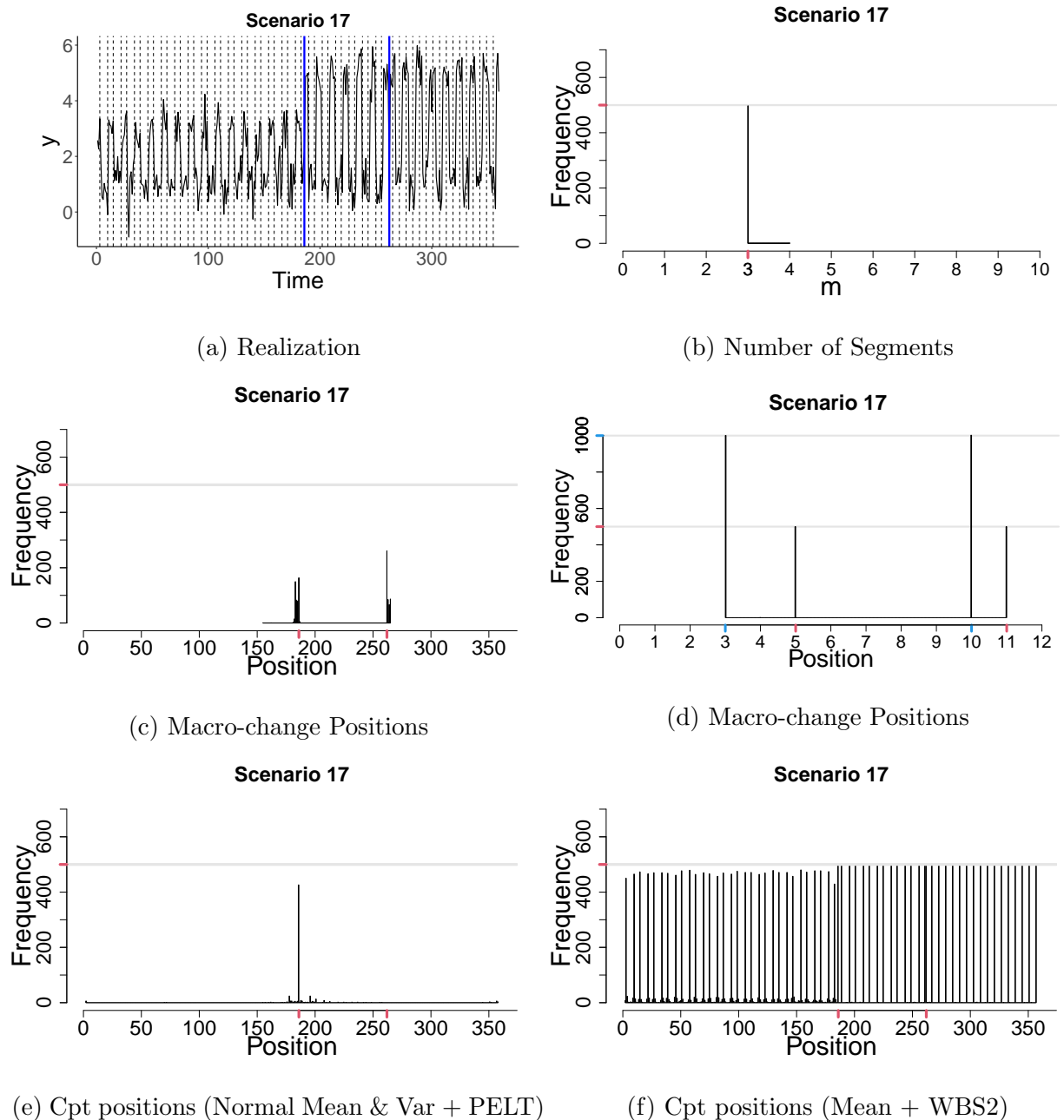
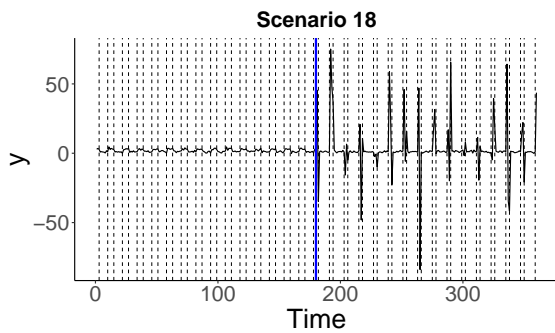
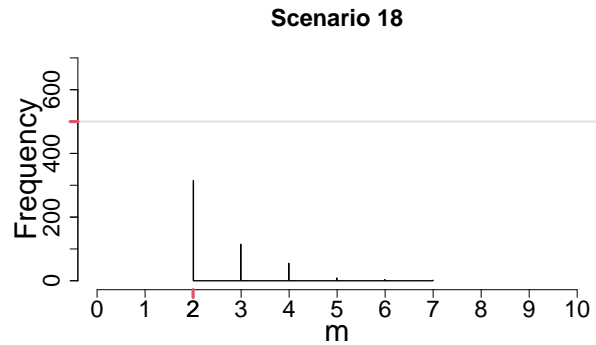


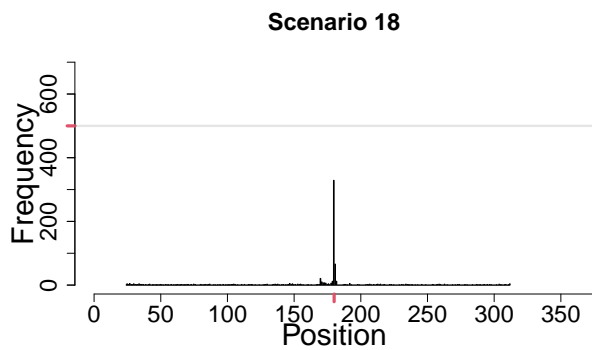
Figure 22: Simulation results for scenario 17: 2 macro-change points at 6th month of 16th period and 10th month of 22nd period, micro-changes at months (3,10) for segments 1, 2, and (5,11) for the 3rd segment. The ticks on the axes indicate which true positions should occur with which frequency.



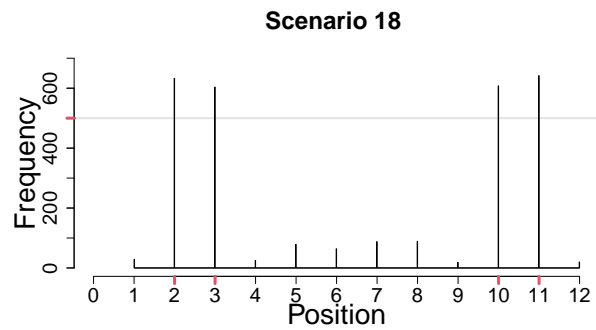
(a) Realization



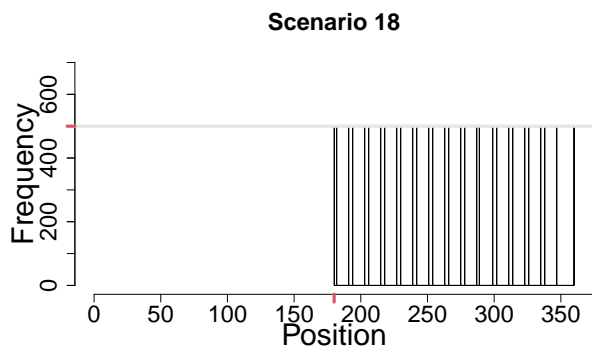
(b) Number of Segments



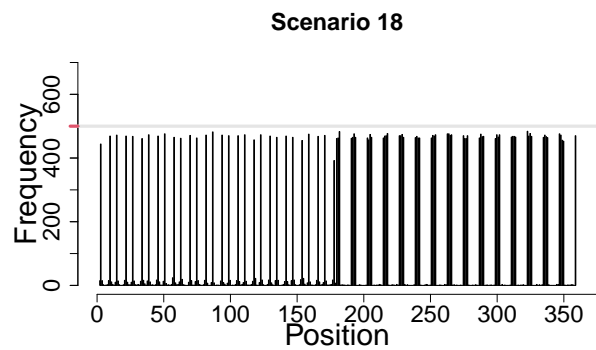
(c) Macro-change Positions



(d) Micro-change Positions

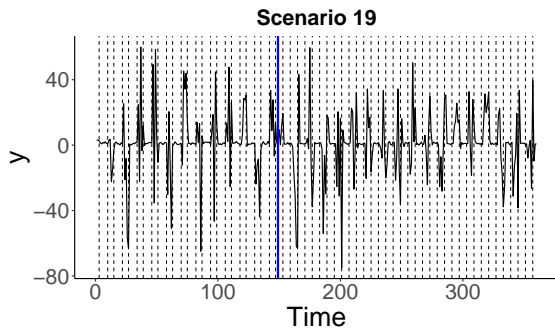


(e) Cpt positions (Normal Mean & Var + PELT)

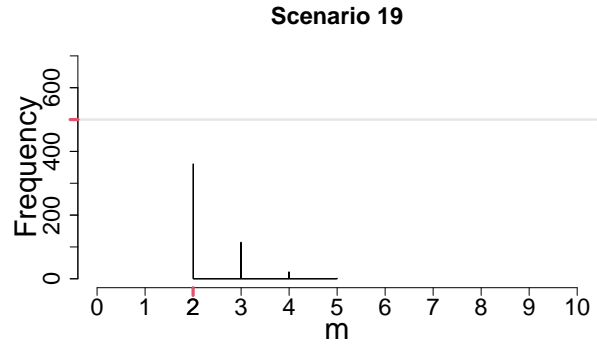


(f) Cpt positions (Mean + WBS2)

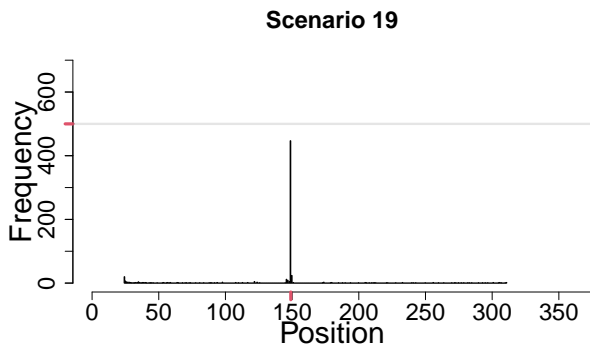
Figure 23: Simulation results for scenario 18: 1 macro-change at 15th period, micro-changes at months (3,10) and (2,11) for segments 1 and 2 respectively.



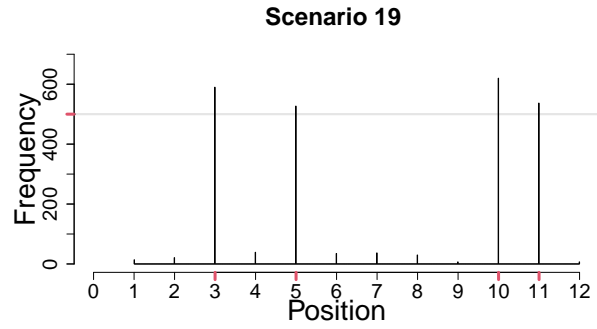
(a) Realization



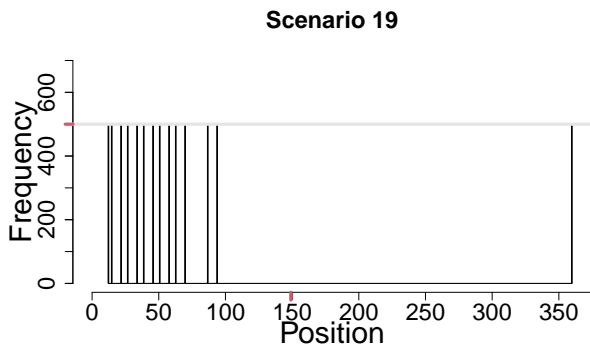
(b) Number of Segments



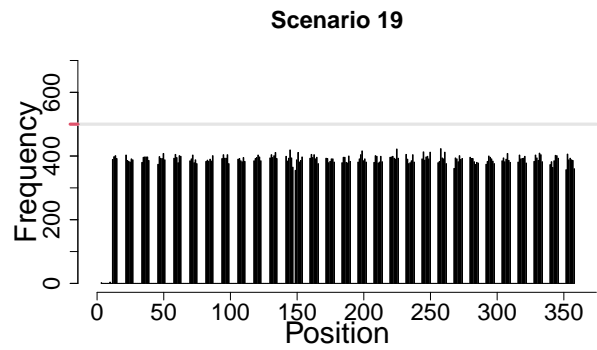
(c) Macro-change Positions



(d) Micro-change Positions



(e) Cpt positions (Normal Mean & Var + PELT)



(f) Cpt positions (Mean + WBS2)

Figure 24: Simulation results for scenario 19: 1 macro-change at 5th month of 13th period, micro-changes at months (3,10) and (5,11) for the segments 1 and 2 respectively.

C Applications

C.1 Sensitivity Analyses

These “scree” type plots (Figure 25) are commonly used by practitioners to decide on an appropriate penalty for a specific time series which may not satisfy all the theoretical assumptions. We expect that as we decrease the penalty value we are first adding true changepoints into our model, these will be the larger changes that make a big different to our model fit (likelihood). As we decrease the penalty further we will start to pick up smaller, less obvious changes and eventually, as we decrease to a penalty of zero we will just be picking up changes due to noise. These less obvious and noisy changes will not improve our fit as much as the obvious changes. Thus we are looking to identify the elbow in the curve where we move from larger changes in the likelihood values, to smaller changes due to noise. This is a judgment that is typically based on identifying a region for the elbow and then looking at segmentations within that region to identify a plausible solution.

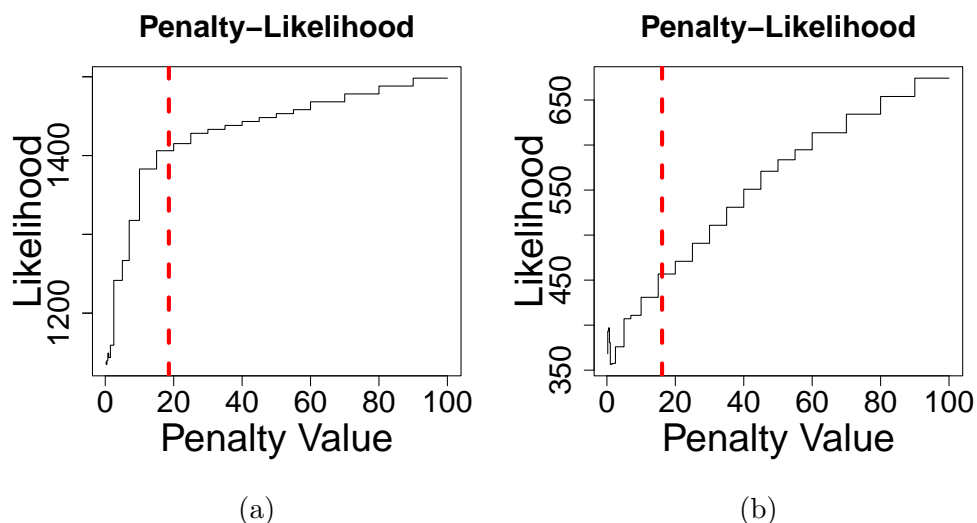


Figure 25: Likelihood versus penalty analyses for (a) NAO data and (b) HPI data. The dashed lines represent chosen penalty values: $4\log(n)$ and $3\log(n)$ respectively for (a) and (b).