Trabalho Séries Temporais

Bruno Grillo bruno.grillo@ufrgs.br
193827

1. Introduction

The purpose of this text is to compare distinct change-point detection methods on simulated and real time series. The simulated data helps to understand the methods properties of determining the amount of change-points in a given time series and their location under differente scenarios. The scenarios might include heavy tail residual, change in mean and/or variance, trend and seasonality. The methods employed in the comparison are: (i) Narrowest Over Threshold proposed by R. Baranowski et al (2019), (ii) Tail-greedy proposed by P. Fryzlewicz (2018) and (iii) Wild Binary Segmentation proposed by P. Fryzlewicz (2014). These methods are based on information criteria, which means they decide the location and amount of change-points by employing an information criteria. The methods are implemented in the following R packages: (i) *breakfast*, (ii) *not*.

After assessing the methods performance under different scenarios in synthetic data, they will be applied at real time series that represents the traded volume of goods belonging to four distinct NCM, an acronym for *Nomenclatura Comum do Mercosul*, an standardised reference for goods traded by countries belonging to Mercosur. The series are aggregated by month from January 1997 to April 2021 and consists of the total imports to and exports from Rio Grande do Sul measured in tons. The selected goods are (i) imports of combine harvesterm (ii) imports of wheat, (iii) exports of polyethylene and (iv) exports of tobacco leaves.

One useful application of change-point detection is to assess what part of the data brings relevant information to the model. If there are structural changes in the data, it makes sense to employ the most recent data that follows the same data generation process.

2. Simulated Data

There are 4 simulated time series to test the methods properties. To compare the methods, each time series is simulated 1000 times and the competing methods are employed. The purpose is to compare two metrics among the candidate methods: (i) sensitivity and (ii) precision. The sensitivity is calculated as

$$Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative}$$

The sensitivity provides the portion of the true change-points that were correctly classified. It answers the following question: of all change-points, how many were detected? The precision is calculated as

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

and it provides the portion of the data points classified as change-point that are actually a change-point. It answers the following question: of all points classified as change-point, how many were actually change-points?
The change-point is considered to be correct if it is at most two points away from the actual change-point, a choice that will produce greater values for the evaluated metrics than exact location of change-point.
The Narrowest Over Threshold has several available contrast functions, while the other methods do not allow such parametrization. On choosing the not candidate, all contrast function possibilities will be calculated and the one that minimizes the information criteria will be selected. It is important to notice that all methods employ an information criteria to determine the amount of points and their location, however, only not has such structure.

2.1 Piece Wise Constant Mean and Gaussian Noise

In this subsection we evaluate the methods performance for a time series vector that has piece-wise constant mean and Gaussian noise. It means that each data interval is generated by repeating a value and adding a Gaussian noise. The data has 80 observations and contains three change-points and, consequently, four subsamples. The change-points are located at locations 21, 41 and 61.
For each simulation, the subsamples are constructed by the following script:

```
x <- c(rep(0, 20), rep(1,20), rep(-1, 20), rep(0.5, 20)) + rnorm(80)

```

The following histogram displays the distribution of the amount of change-points detected by each method. The dotted vertical lines represent the actual amount of change-points.

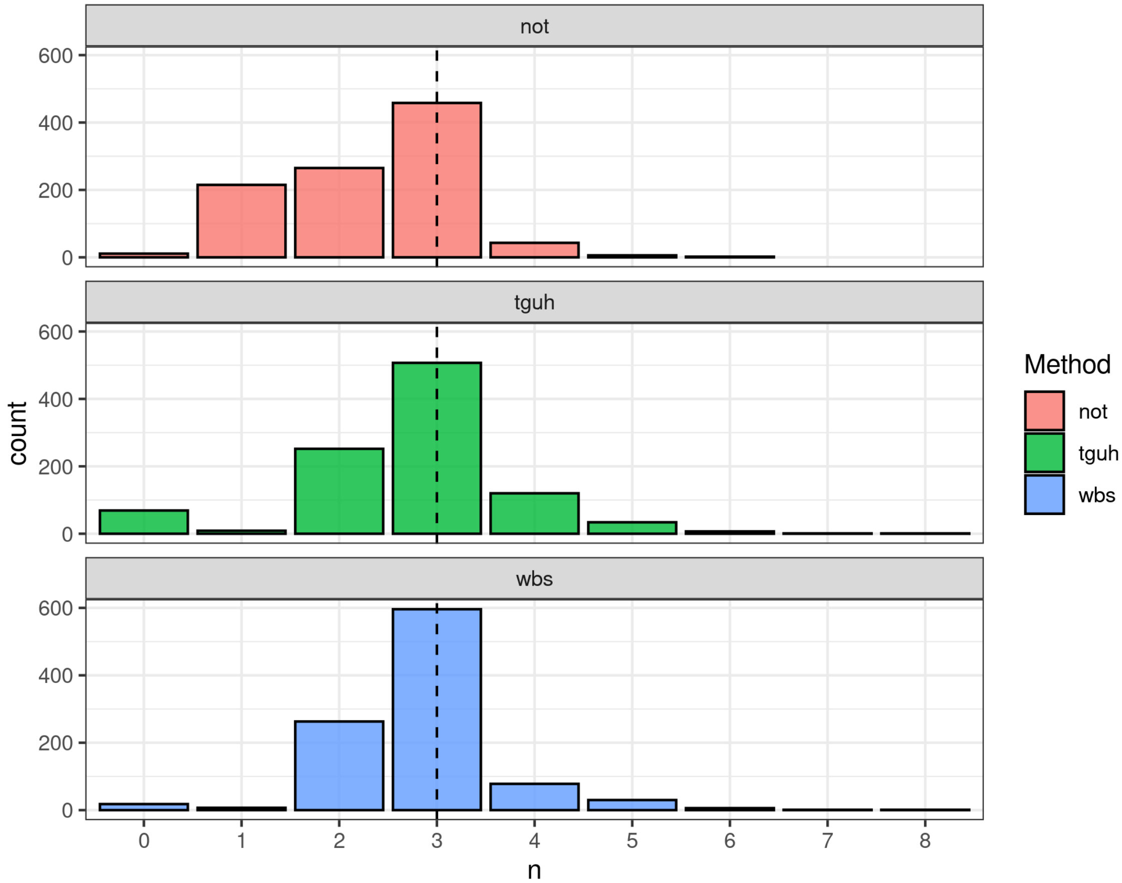

Figure 1: Histogram Amount of Detected Change-Points

To evaluate the location, consider the following histogram that shows the frequency there was a change-point detected in the specific time. The dotted vertical line represents the actual location of the change-points.

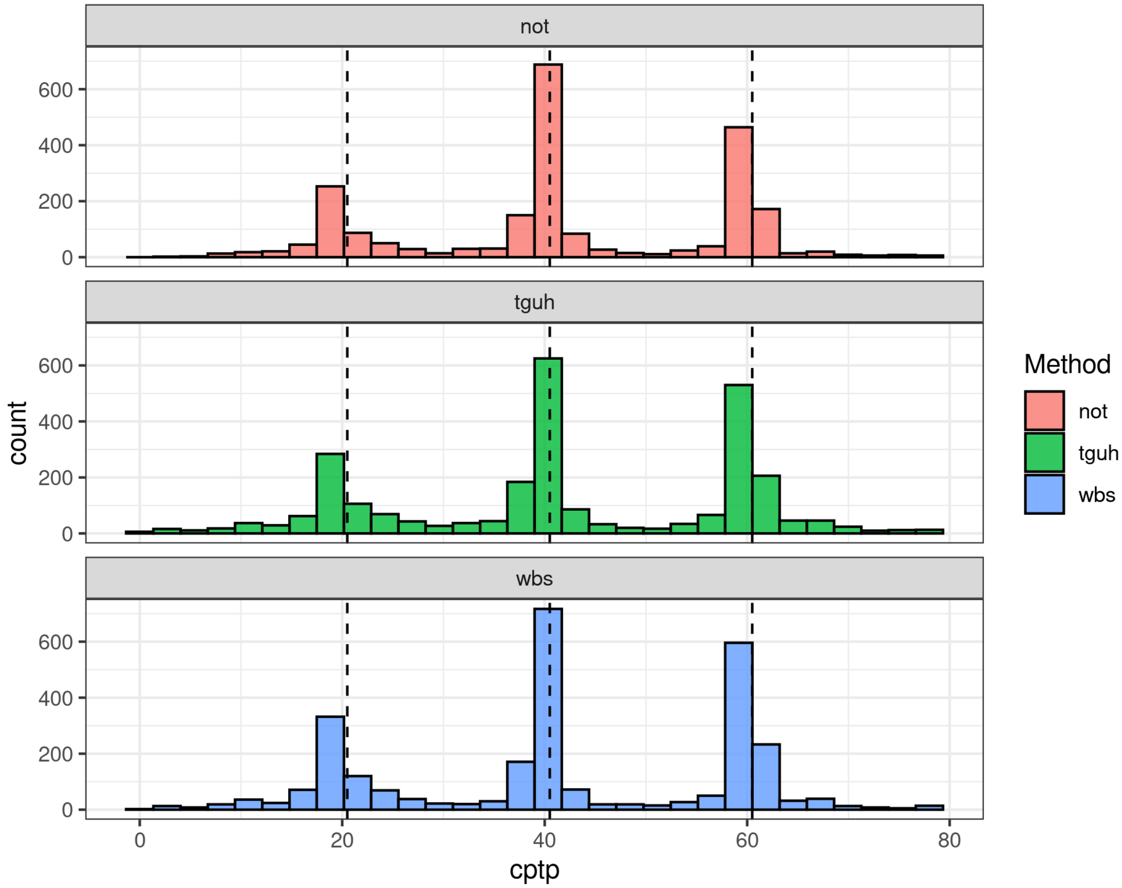

Figure 2: Change-Points Location

Considering the histogram of correct amount of change-points, wild binary segmentation produces better results. Considering the performance metrics in the following table, the wild binary segmentation is the clear winner because it has the greatest sensitivity among candidates and the precision is very close to the highest. Tail greedy detects more of the actual change-points (64,2%) than narrowest over threshold (60,5%), however, it does so by labelling more data points as change-points incorrectly (measured by 1-precision). Considering the histogram of change-point locations, all methods have lower detection rate for the first change-point (located at position 21), which has the smallest deviation in terms of mean from the neighbor subsample.

Table 1: Performance Evaluation

Metric	not	tg	wb
Sensitivity	0.605	0.642	0.720
Precision	0.774	0.686	0.758

2.2 Piece-wise Constand Mean and Heavy Tail

In this subsection we evaluate the methods performance for a time series vector that has piece-wise constant mean and noise from a T-Student distribution with 2 degrees of freedom. It means that each data interval is generated by repeating a value and adding a heavy tail noise. The data has 80 observations and contains three change-points and, consequently, four subsamples. The change-points are located at locations 21, 41 and 61.
For each simulation, the subsamples are constructed by the following script:

```
x <- c(rep(0, 20), rep(1,20), rep(-1, 20), rep(0.5, 20)) + rt(80, df = 2)

```

The following histogram displays the distribution of the amount of change-points detected by each method. The dotted vertical lines represent the actual amount of change-points.

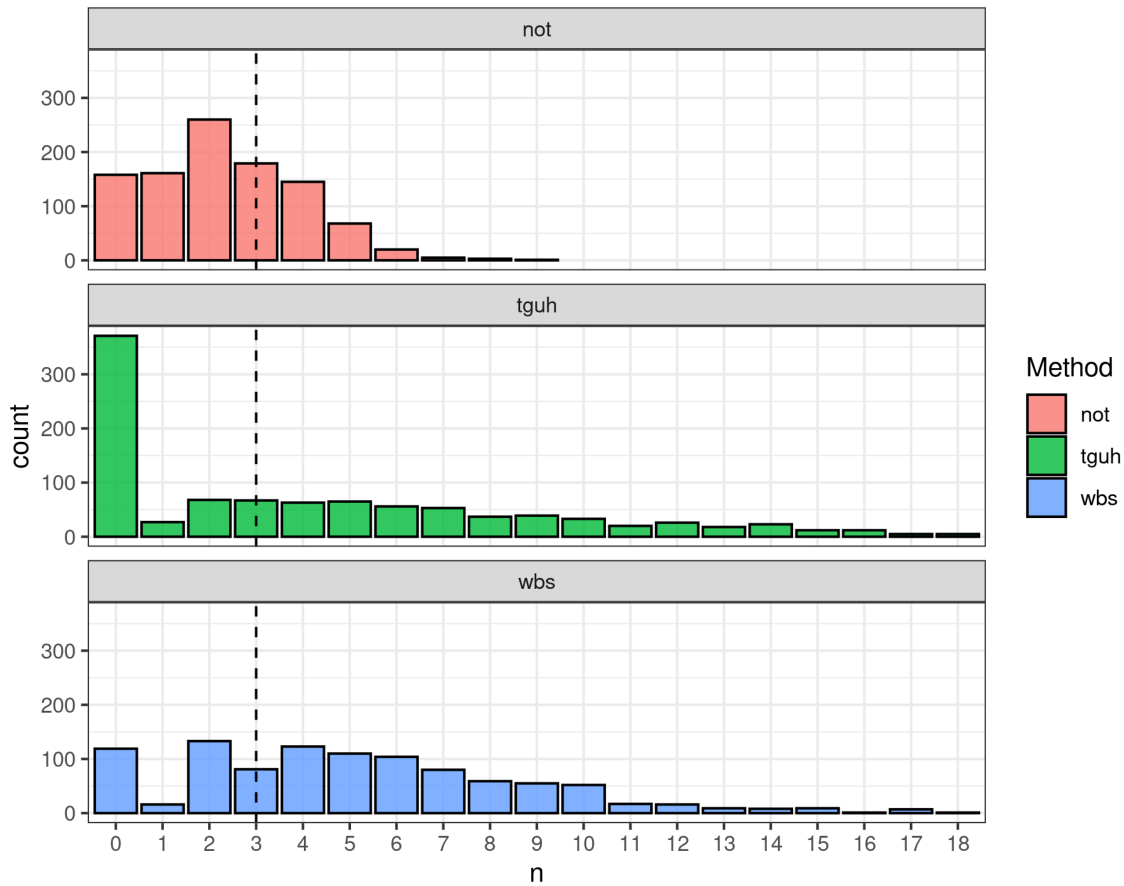

Figure 3: Histogram Amount of Detected Change-Points

To evaluate the location, consider the following histogram that shows the frequency there was a change-point detected in the specific time. The dotted vertical line represents the actual location of the change-points.

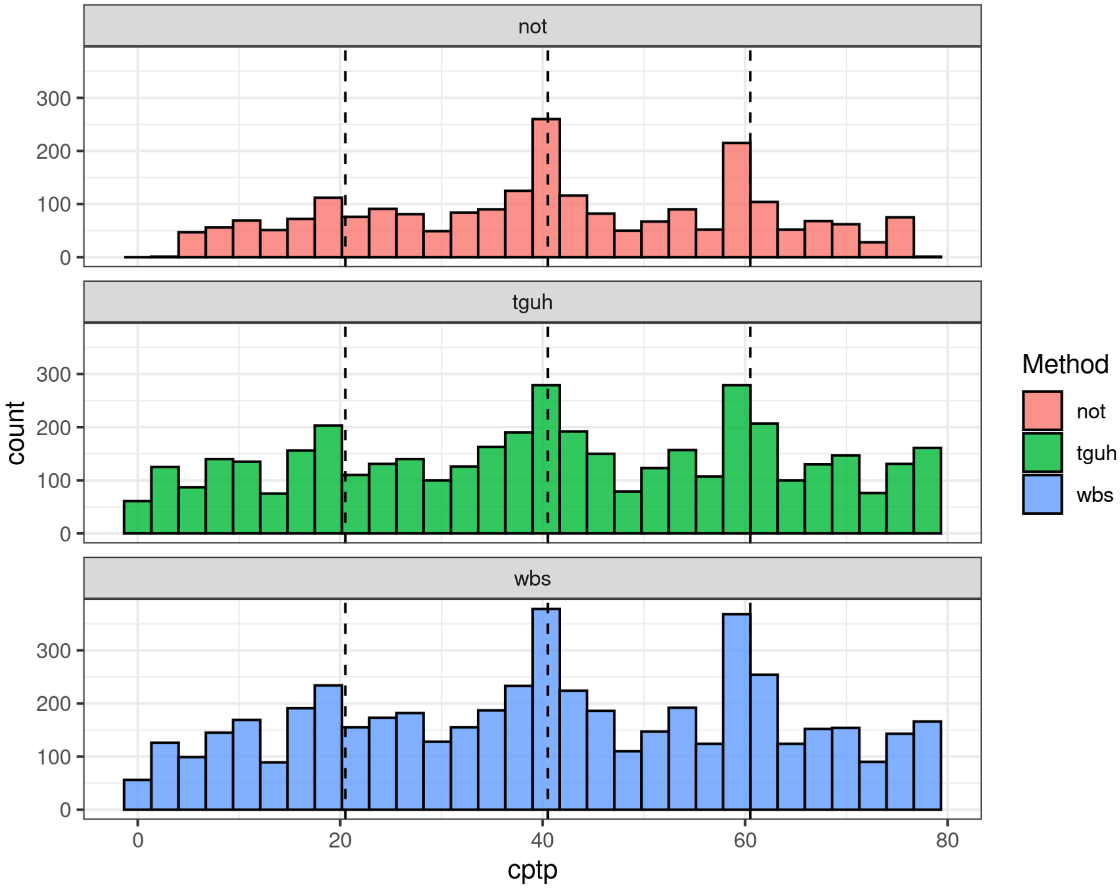

Figure 4: Change-Points Location

Considering the histogram of correct amount of change-points, narrowest over threshold produces better results for data under constant piece-wise mean and heavy tail. Considering the performance metrics in the following table, there is no clear winner because all metrics display low values. Wild binary detects more of the actual change-points (52,6%), however, it does so by labelling many data points as change-points. Only 30% of the points it labels as change-points are actually change-points. Depending on the application, narrowest over threshold might be a better choice because the points it labels as change-points are more likely to be actual change-points (35,3%). Considering the histogram of change-point locations, all methods have lower detection rate for the first change-point (located at position 21), which has the smallest deviation in terms of mean from the neighbor subsample.

Table 2: Performance Evaluation

Metric	not	tg	wb
Sensitivity	0.292	0.407	0.526
Precision	0.353	0.263	0.301

2.3 Piece-Wise Constant Mean and Piece-Wise Constant Variance

In this subsection we evaluate the methods performance for a time series vector that has piece-wise constant mean and piece-wise constant variance. The data has 80 observations and contains three change-points and, consequently, four subsamples. The change-points are located at locations 21, 41 and 61. The first change-point represents a change in the variance, the second change-point represents a change in the mean and the last change-point represents a change in the variance. For each simulation, the subsamples are constructed by the following script:

```
x <- c(rnorm(20, 5, 1), rnorm(20, 5, 2), rnorm(20, 4, 2), rnorm(20, 4, 1))

```

The following histogram displays the distribution of the amount of change-points detected by each method. The dotted vertical lines represent the actual amount of change-points.

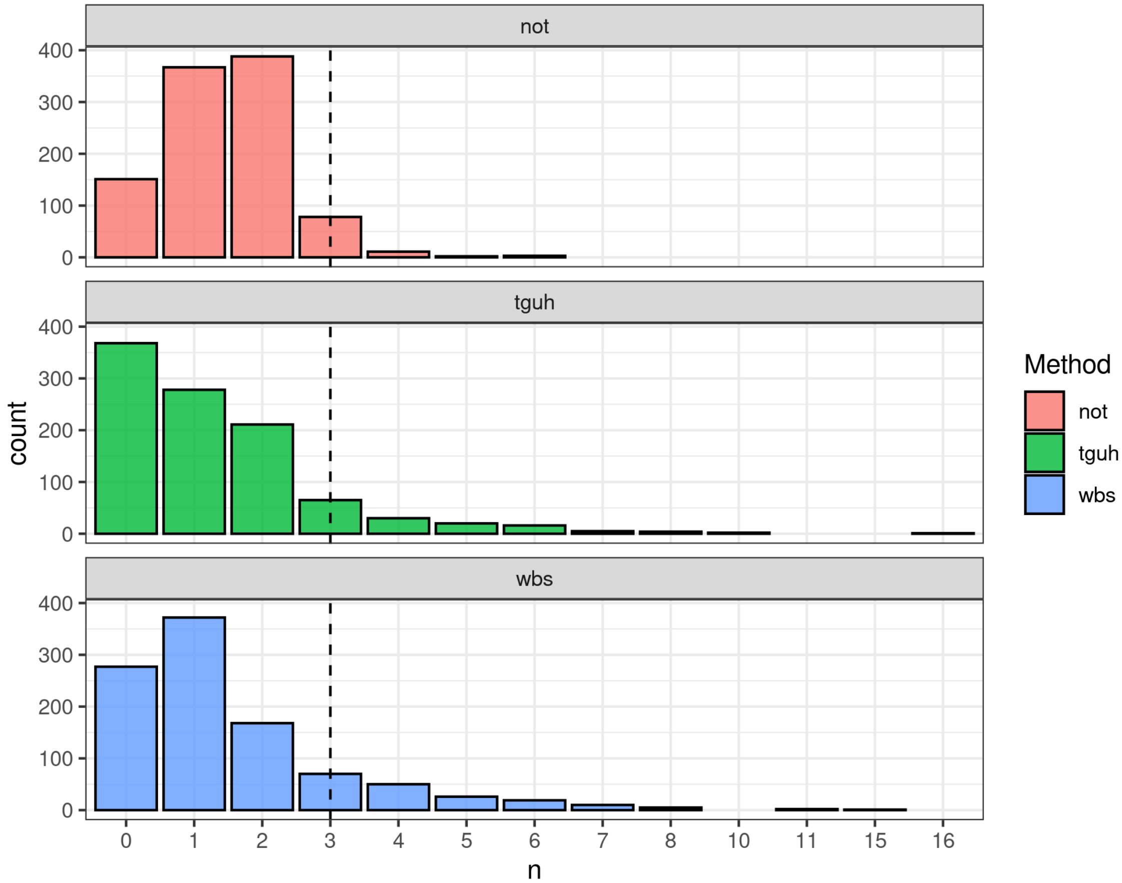

Figure 5: Histogram Amount of Detected Change-Points

To evaluate the location, consider the following histogram that shows the frequency there was a change-point detected in the specific time. The dotted vertical line represents the actual location of the change-points.

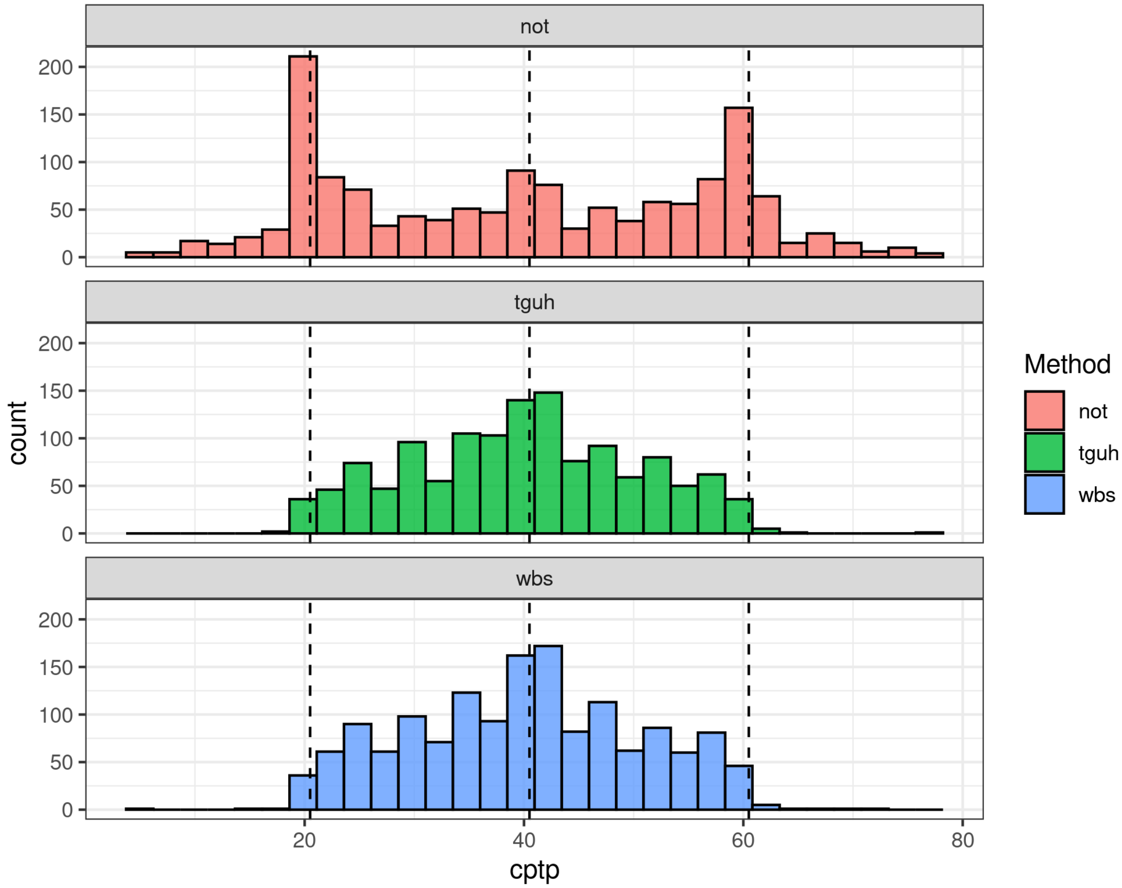

Figure 6: Change-Points Location

Considering the histogram of correct amount of change-points, narrowest over threshold produces better results for data with constant piece-wise mean and variance. Considering the performance metrics in the following table, narrowest over threshold is the clear winner since it produces the highest value for both metrics. However, all methods produce low values for the performance metrics. Narrowest over threshold correctly identifies only 22,4% of the change-points, and those that it labels as change-points, only 41,9% are true change-points. Depending on the application purpose, it might be helpful. By analysing the histogram of change-point location, it is clear that narrowest over threshold has an advantage in identifying the first and third change-points (variance changes), while the other methods fail poorly. For the second change-point (mean change), the other methods produce better results.

Table 3: Performance Evaluation

Metric	not	tg	wb
Sensitivity	0.224	0.140	0.157
Precision	0.419	0.250	0.264

2.4 ARIMA Process, Linear and Quadratic Trend, Heavy Tail

In this subsection we evaluate the methods performance for a time series vector that is generated by a complex construction. The data has 80 observations and contains three change-points and, consequently, four subsamples. The change-points are located at locations 21, 41 and 61. The first 20 observations come from an ARIMA(1,1,0) with $\phi_1 = -0.07$ and noise from a T-Student distribution with 7 degrees of freedom. The following 20 observations have constant mean and Gaussian noise, the mean of this subsample is the sum of the (i) the mean of the last four observations of the ARIMA process and (ii) 2 times the standard deviation of the last four observations of the ARIMA process. The third subsample of 20 observations is a linear trend with noise from a T-Student distribution with 5 degrees of freedom. The intercept is the mean of the last four observations of the second subsample and the angular coefficient is $0.15t$, where t is the index of the observations (in this case, it ranges from 41 to 60). The last subsample has a quadratic trend with Gaussian noise. The intercept is the mean of the 10 last observations from the third subsample and the angular coefficient is -0.0008t^2. For each simulation, the subsamples are constructed by the following script:

```
x <- arima.sim(n = 20,
               model = list(order = c(1,1,0), ar = -0.07),
               start.innov = 5, n.start = 1, innov = rt(20, df = 7))[-1]

x <- c(x, rep(mean(tail(x, 4)) + 2 * sd(tail(x, 4)), 20) + rnorm(20))

x <- c(x, (mean(tail(x, 4))) + 0.15 * seq(41, 60) + rt(20, df = 5))

x <- c(x, mean(tail(x, 10)) - 0.0008 * seq(61, 80) ** 2 + rnorm(20))
x # FINAL
```

The following histogram displays the distribution of the amount of change-points detected by each method. The dotted vertical lines represent the actual amount of change-points.

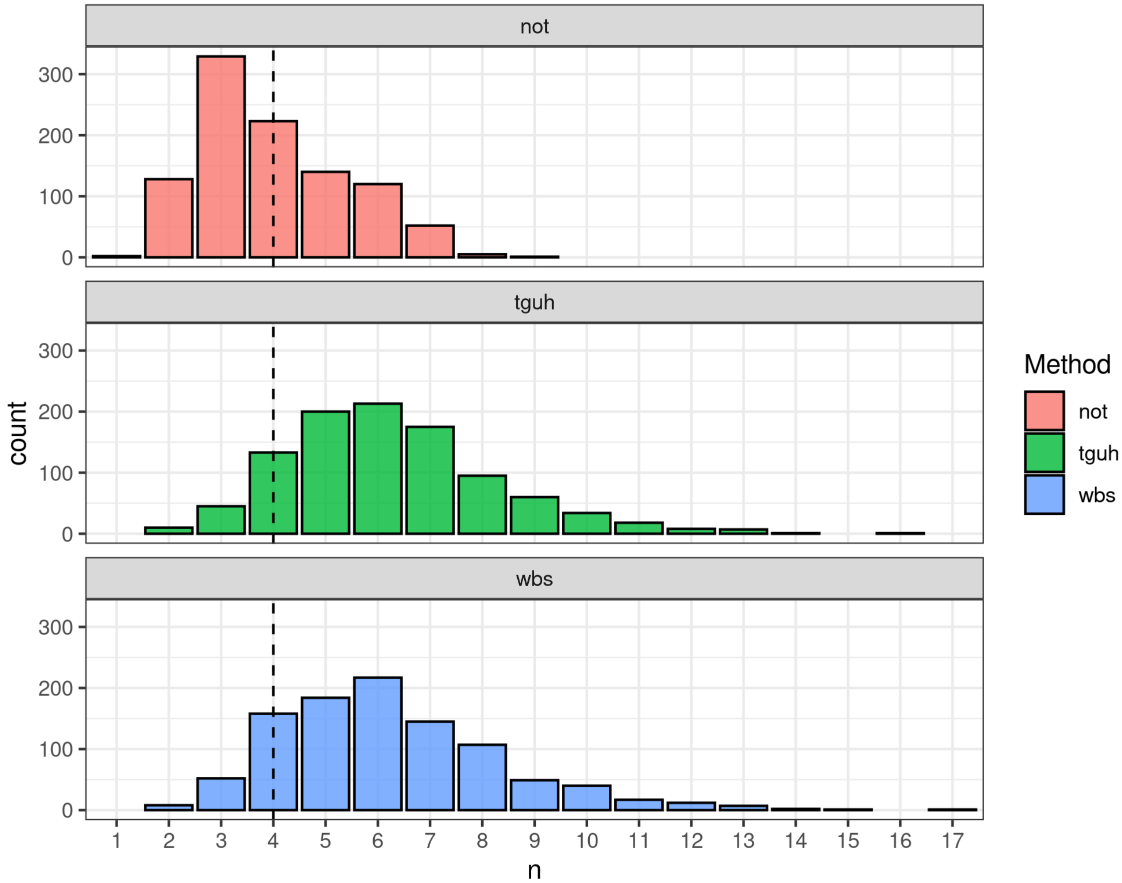

Figure 7: Histogram Amount of Detected Change-Points

To evaluate the location, consider the following histogram that shows the frequency there was a change-point detected in the specific time. The dotted vertical line represents the actual location of the change-points.

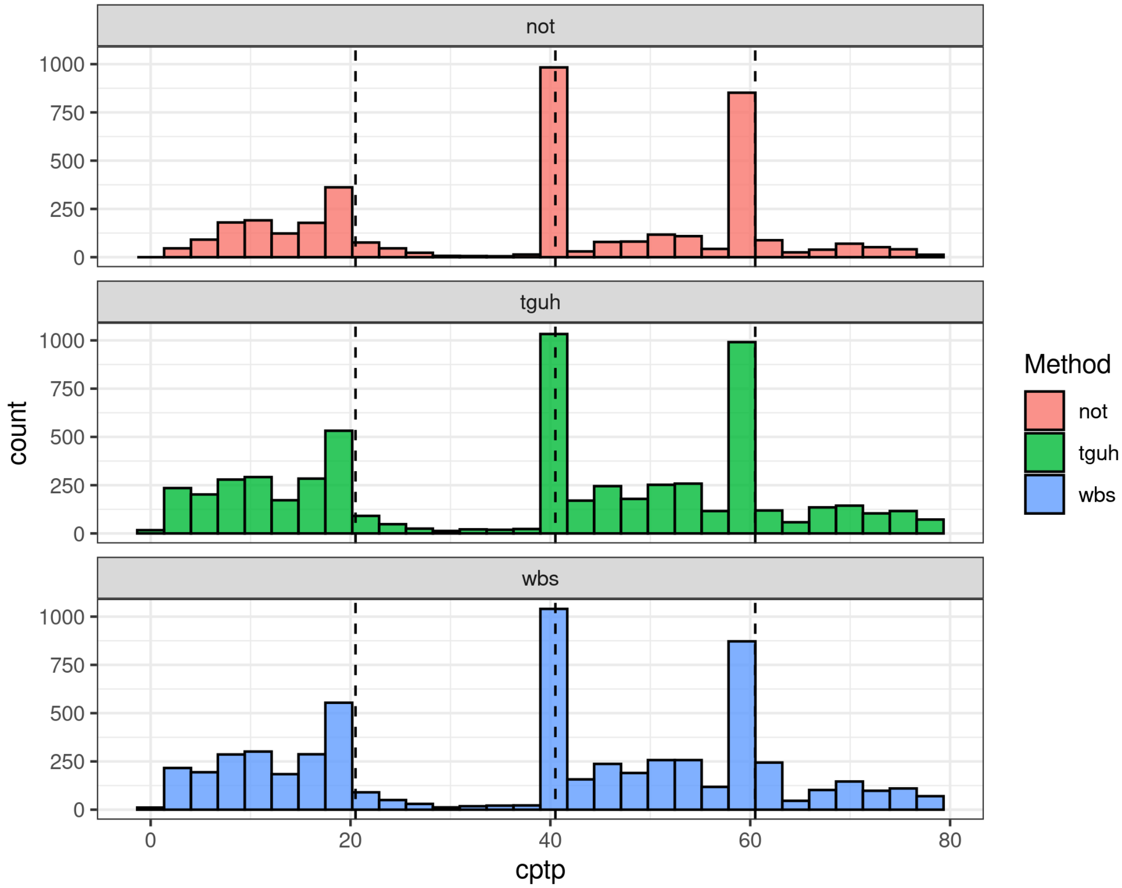

Figure 8: Change-Points Location

Considering the histogram of correct amount of change-points, narrowest over threshold produces better results (center of mass closer to the actual value and not suggesting a huge amount of change-points) for this series. Considering the performance metrics in the following table, there is no clear winner, however, the narrowest over threshold and wild binary (slightly better than tail greedy) are the best candidates. Narrowest over threshold correctly identifies 78,9% of the change-points, and of those that it labels as change-points, 59,6% are true change-points. The wild binary correctly identifies 94,3% of the change-points, and of those labeled as change-points, 45,5% are true change-points. By analysing the histogram of change-point location, narrowest over threshold has a lower detection rate for first change-point (from ARIMA process to constant mean with Gaussian noise). All methods have high detection rate for the second and third change-points.

Table 4: Performance Evaluation

Metric	not	tg	wb
Sensitivity	0.789	0.935	0.943
Precision	0.596	0.449	0.455

3. Real Data

In this section, the main goal is to see how these methods perform in real data described in the introduction. However, it is not possible to evaluate the performance as in the previous section, since there is no information on the true change-points. The only comparison is the presence or lack of convergence among the distinct methods.

3.1 Tobacco Leaves Exports

The following plot has the same time series plotted on each facet. The vertical lines are the change-points detected by each method. For this time series with strong seasonality, the methods converge in the amount of change-points (one) and in the location.

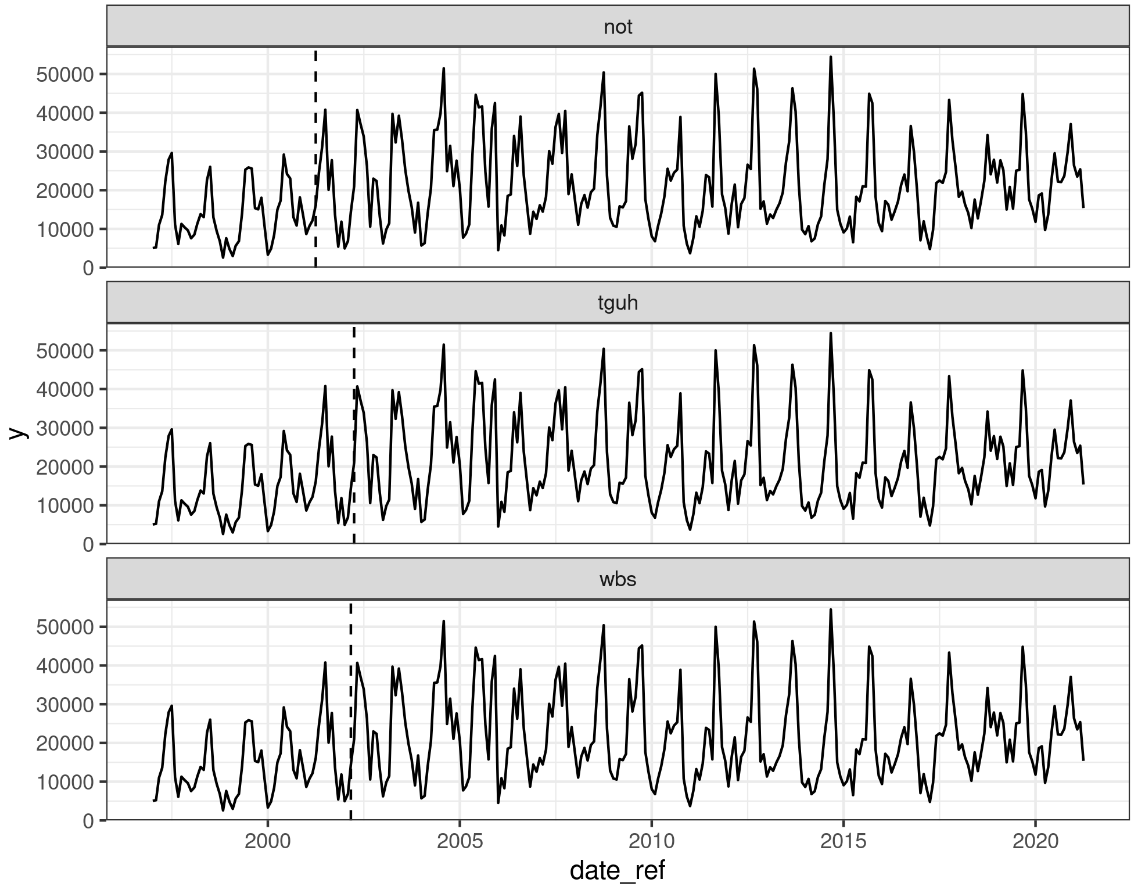

Figure 9: Tobacco Leaves Exports

3.2 Polyethylene Exports

The following plot has the same time series plotted on each facet. The vertical lines are the change-points detected by each method. For this time series, the methods diverge in the amount of change-points and their locations. NOT suggests seven change-points, WBS suggests 13 and TG suggests only one. Considering the position of the change-points detected by NOT and WBS, there is a considerable overlap.

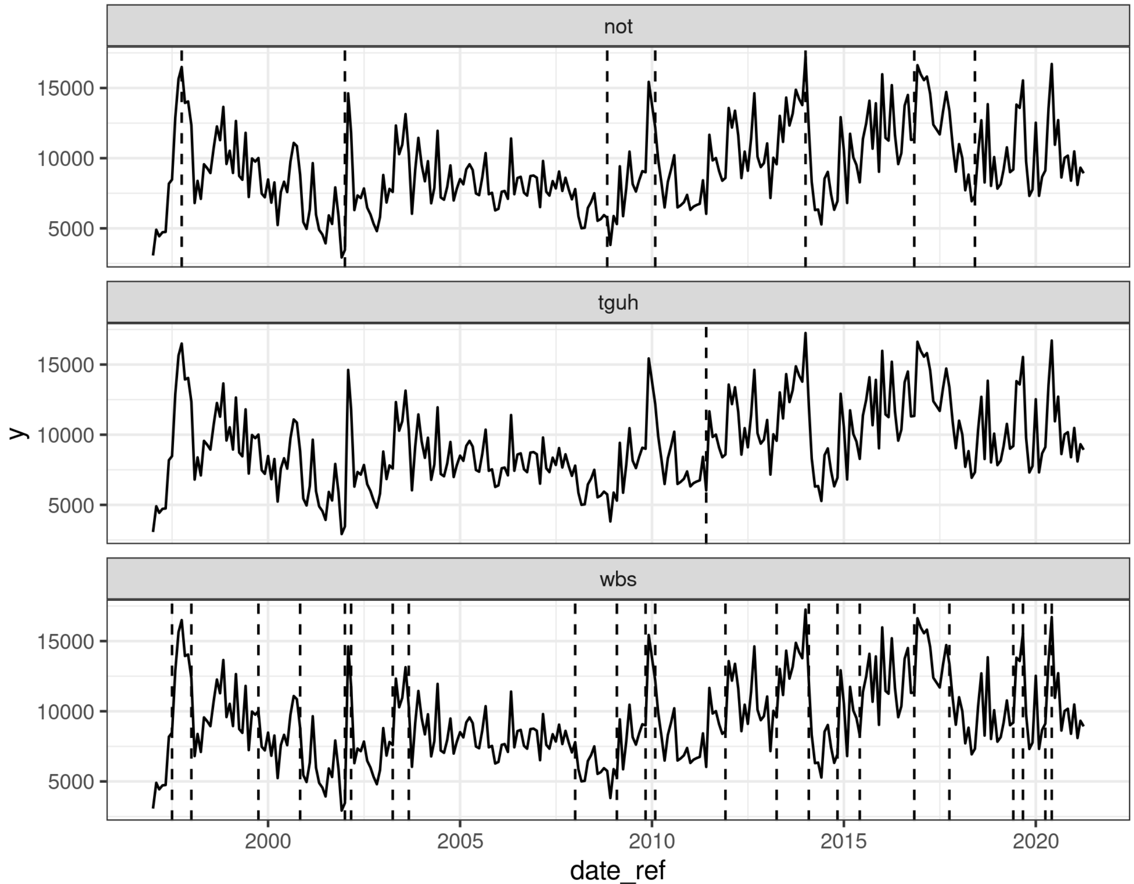

Figure 10: Polyethylene Exports

3.3 Wheat Import

The following plot has the same time series plotted on each facet. The vertical lines are the change-points detected by each method. All methods converge both in the amount of change-points (nine) and location. The points considered change-points by (i) NOT are $[7, 24, 58, 118, 129, 137, 176, 193, 226]$, (ii) TG are $[6, 24, 58, 118, 129, 137, 184, 193, 225]$ and (iii) WBS are $[7, 24, 44, 118, 129, 137, 176, 192, 225]$.

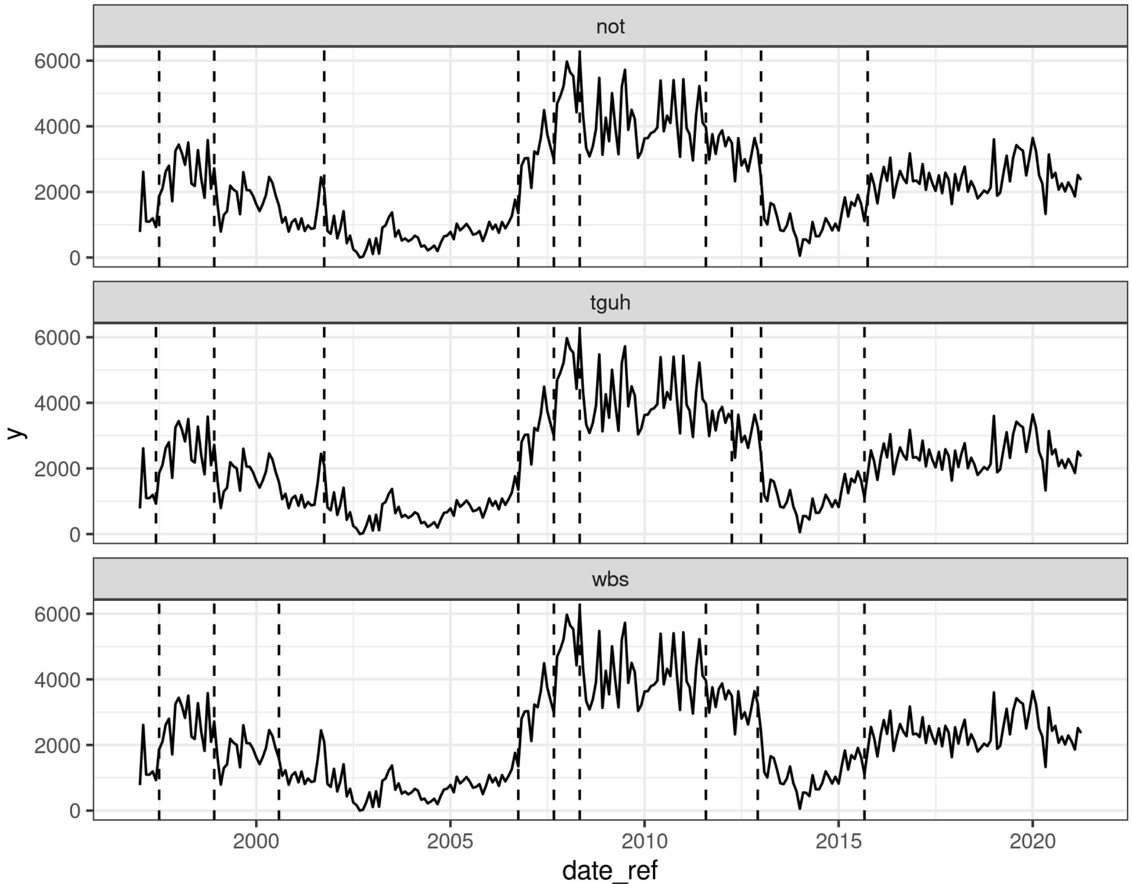

Figure 11: Wheat Imports

3.4 Combine Harvester Import

The following plot has the same time series plotted on each facet. The vertical lines are the change-points detected by each method. For this time series, the methods diverge in the amount of change-points and their locations. NOT suggests six change-points, WBS suggests 26 and TG suggests seven.

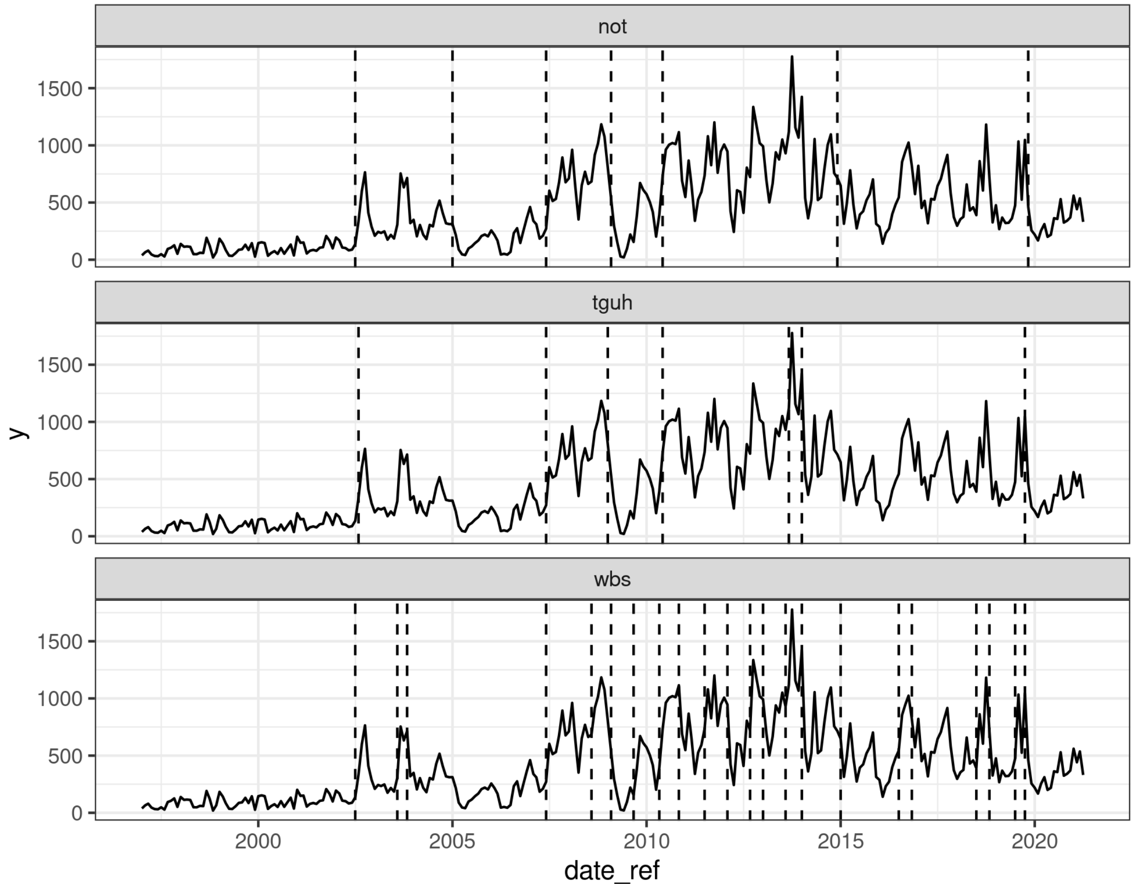

Figure 12: Combine Harvester Imports

4. Conclusion

For the simulated data considered in this work, the narrowest over threshold has produced better results for change-points with variance change. Furthermore, among the candidate methods, it finds less change-points than the competitor, however, the detected change-points are more likely to be true change-points. The wild binary segmentation method has systematically produced better results than tail greedy.

For change-point detection applications, the choice between narrowest over threshold and wild binary segmentation will depend on the cost of each error. If there is a high cost associated with falsely labeling a data as change-point, the narrowest over threshold seems to be preferable, but if there is a high cost associated without identifying a change-point, the wild binary seems to be preferable.

5. References

R. Baranowski, Y. Chen and P. Fryzlewicz (2019). Narrowest-Over-Threshold detection of multiple change-points and change-point-like features. Journal of the Royal Statistical Society Series B, 81, 649-672.
P. Fryzlewicz (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. Annals of Statistics, 46, 3390-3421.
Fryzlewicz, P. (2014). Wild Binary Segmentation for multiple change-point detection. Annals of Statistics 42 2243–2281.