STOR-i

excellence with impact

Statistics and Operational Research
Doctoral Training Centre
Lancaster University

Changepoints for a **R**ange **O**f **P**enaltie**S** using the `changepoint` R package

Kaylea Haynes

Supervised by: Idris Eckley, Paul Fearnhead and Ralph Mansson

useR! 2015 Aalborg

July 2015

[dstl]

EPSRC
Engineering and Physical Sciences Research Council

Overview

STOR-i
excellence with impact

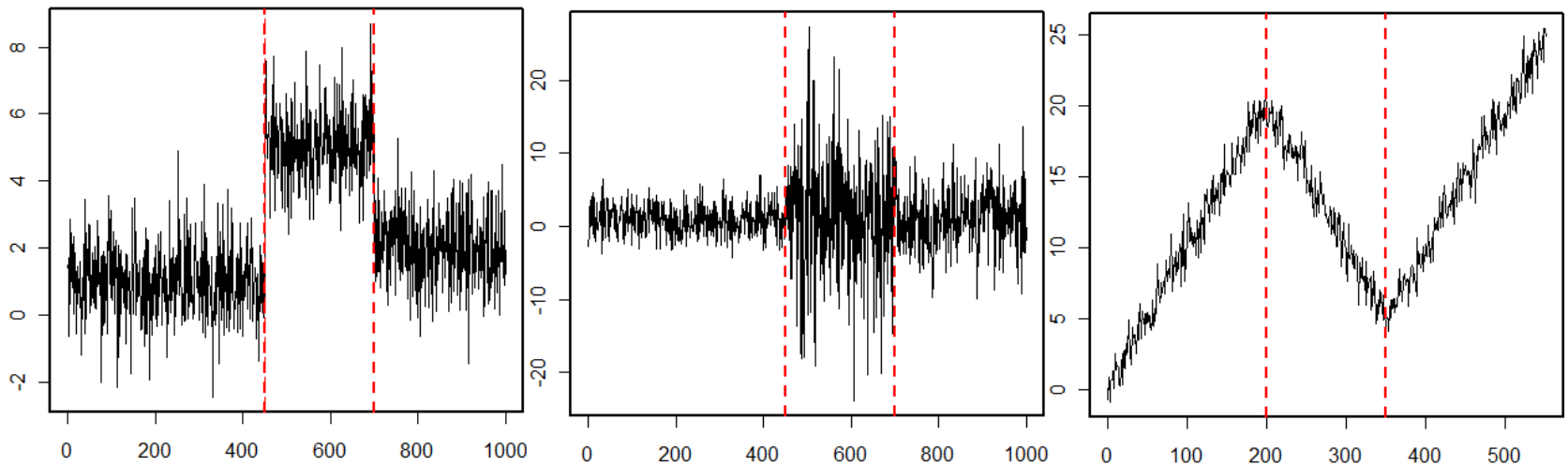
- `changepoint`: An R package for changepoint analysis
  - Version: 2.0
  - Author: Rebecca Killick [aut, cre], Kaylea Haynes [aut], Idris Eckley [aut] and Paul Fearnhead [ctb]
- Multiple changepoint search methods
  - Binary Segmentation
  - Pruned Exact Linear Time (PELT)
  - Changepoints for a Range of Penalties (CROPS)

(Killick et al. 2014, Killick and Eckley 2014)

Changepoint Detection

STOR-i
excellence with impact

For data $y_1, ..., y_n$, a changepoint is a location τ where the statistical properties of $y_1, ..., y_\tau$ differ from those of $y_{\tau+1}, ..., y_n$.



Segment Costs

STOR-i
excellence with impact

Common costs

$$\sum_{i=1}^{m+1} [\mathcal{C} y_{(\tau_{i-1}+1):\tau_i}] + \beta m$$

- `cpt.mean(data,...)`
- `cpt.var(data,...)`
- `cpt.meanvar(data,...)`

`test.stat: "Normal", "Gamma", "Exponential" and "Poisson".`

Optimal Partitioning

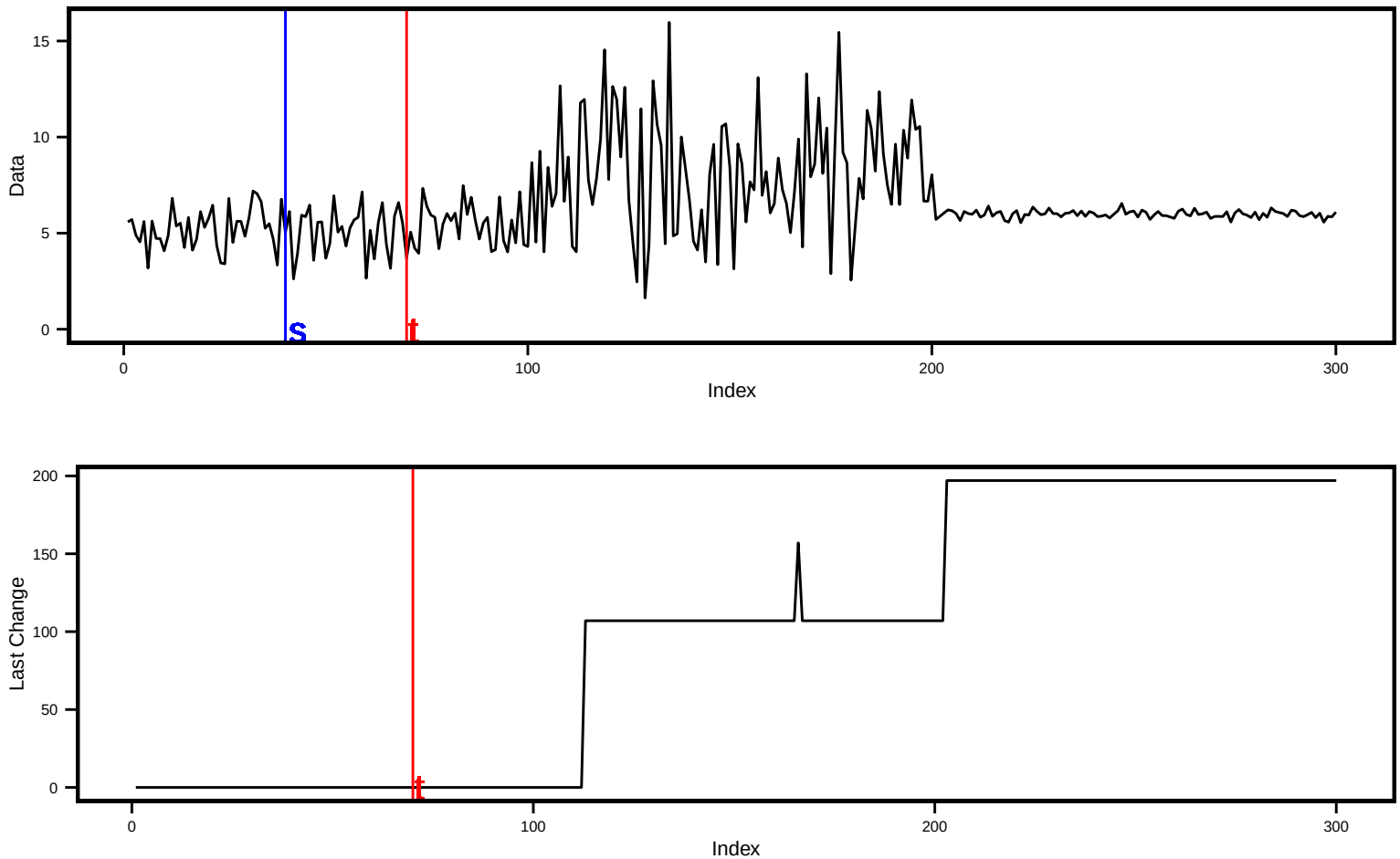
STOR-i
excellence with impact

Recursion

$$F(t) = \min_{\tau \in \boldsymbol{\tau_t}} \left\{ \sum_{i=1}^{m+1} [\mathcal{C}(y_{(\tau_{i-1}+1):\tau_i}) + \beta] \right\}$$
$$= \min_{s \in \{0,\ldots,t-1\}} \{F(s) + \mathcal{C}(y_{(s+1):n}) + \beta\}$$

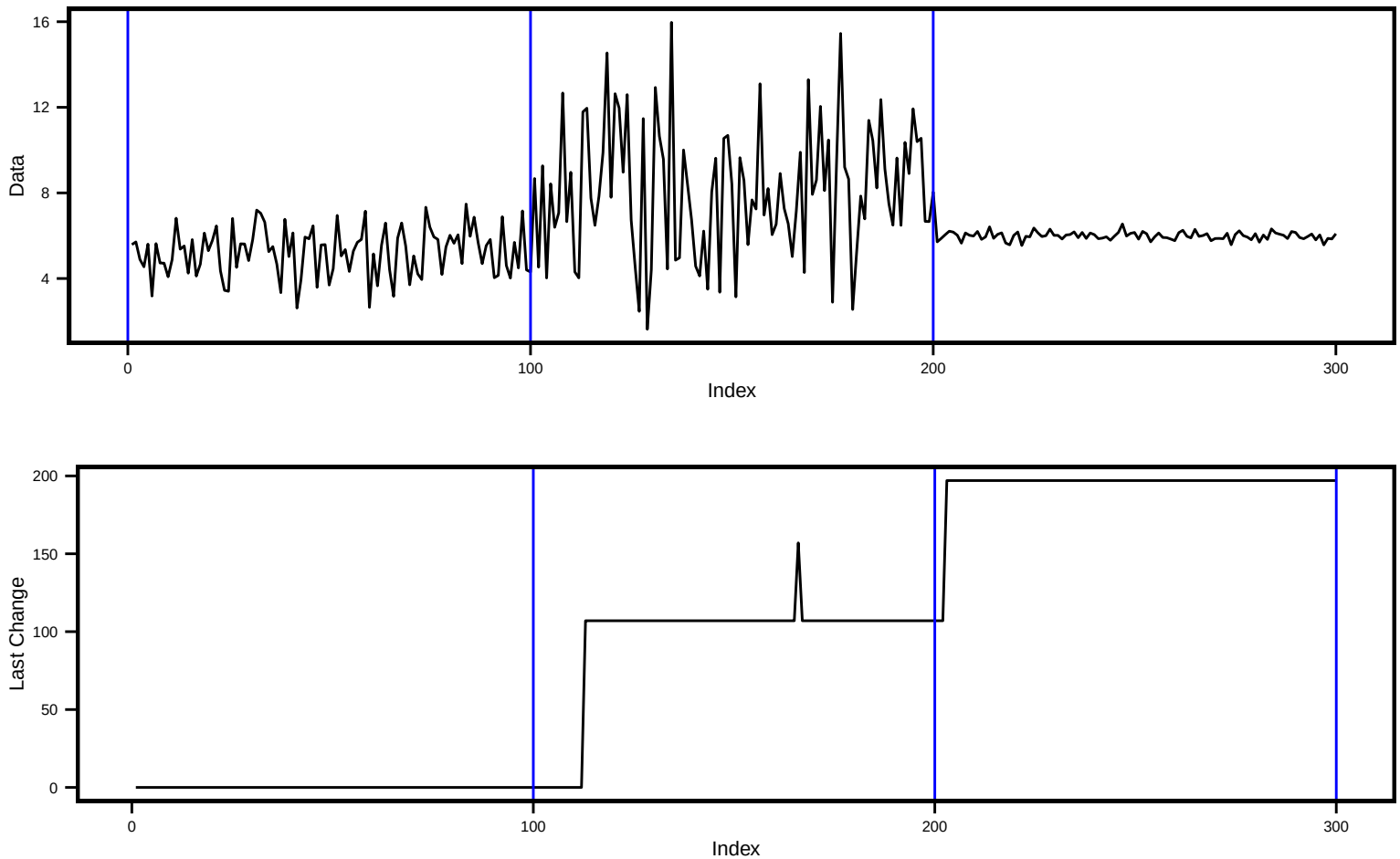
Optimal Partitioning

STOR-i
excellence with impact



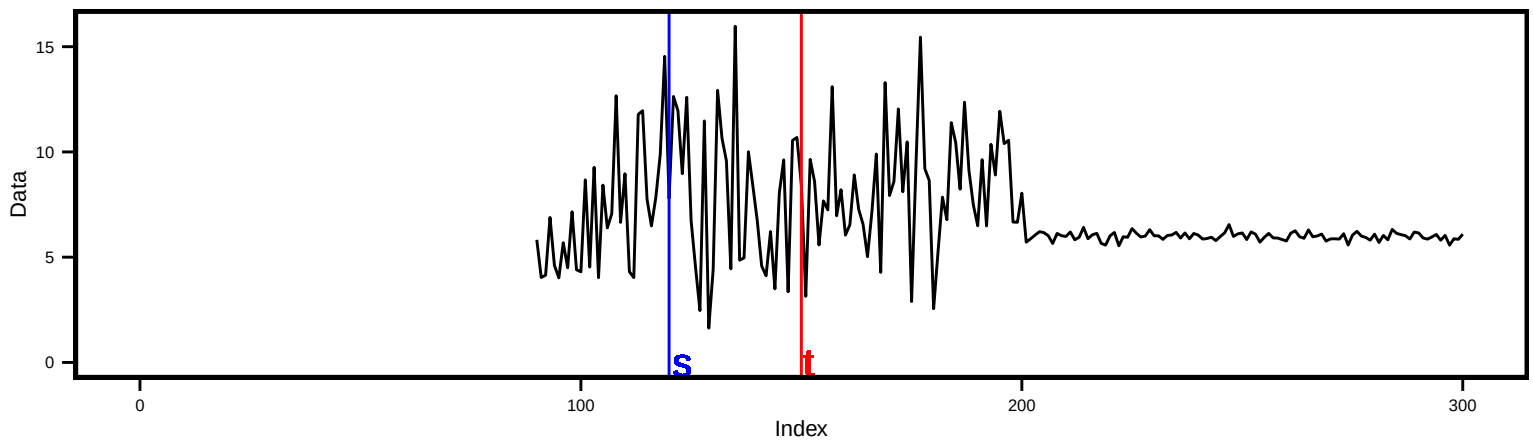
Optimal Partitioning

STOR-i
excellence with impact



PELT

STOR-i
excellence with impact



PELT

STOR-i
excellence with impact

Pruning

If there exists a constant K such that for all $s < t < T$

$$\mathcal{C}(y_{(s+1):t}) + \mathcal{C}(y_{(t+1):T}) + K \leq \mathcal{C}(y_{(s+1):T}),$$

and for $t > s$, if

$$F(s) + \mathcal{C}(y_{(s+1):t}) + K \geq F(t),$$

then at a future time $T > t$, s can never be the optimal last changepoint prior to T.

```
cpt.var(data, penalty, pen.value, test.stat, method =
"PELT", minseglen)
```

Penalty

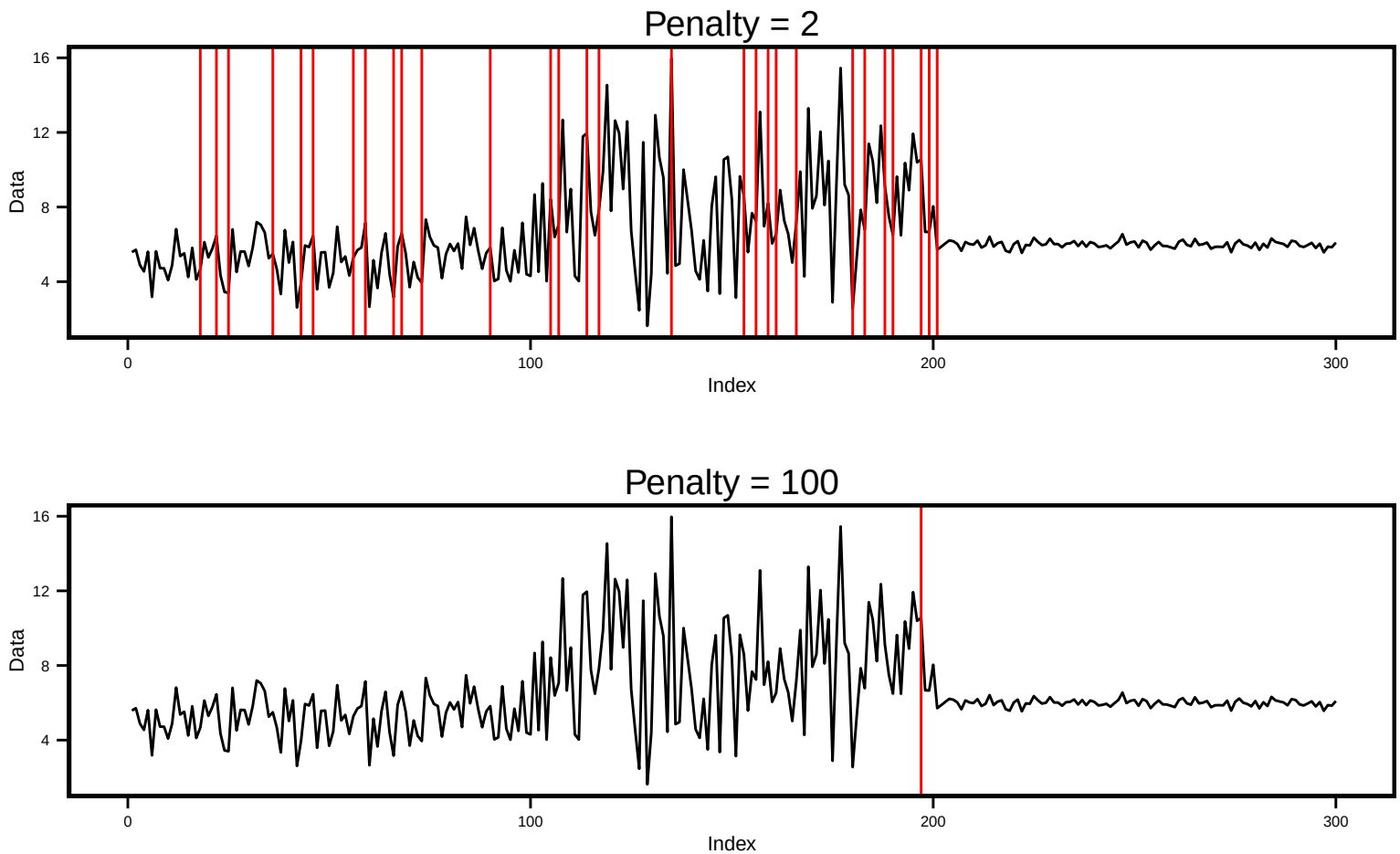
STOR-i
excellence with impact

`cpt.var(data, penalty, pen.value, test.stat, method, minseglen)`

- "None"
- "SIC" or "BIC"
- "MBIC"
- "AIC"
- "Hannan-Quinn"
- "Asymptotic"
- "Manual": `pen.value` $= ?$

Penalty

STOR-i
excellence with impact



Changepoints for a Range of PenaltieS (CROPS).

Step 1: Run algorithm* for β_{min} and β_{max}

STOR-i
excellence with impact

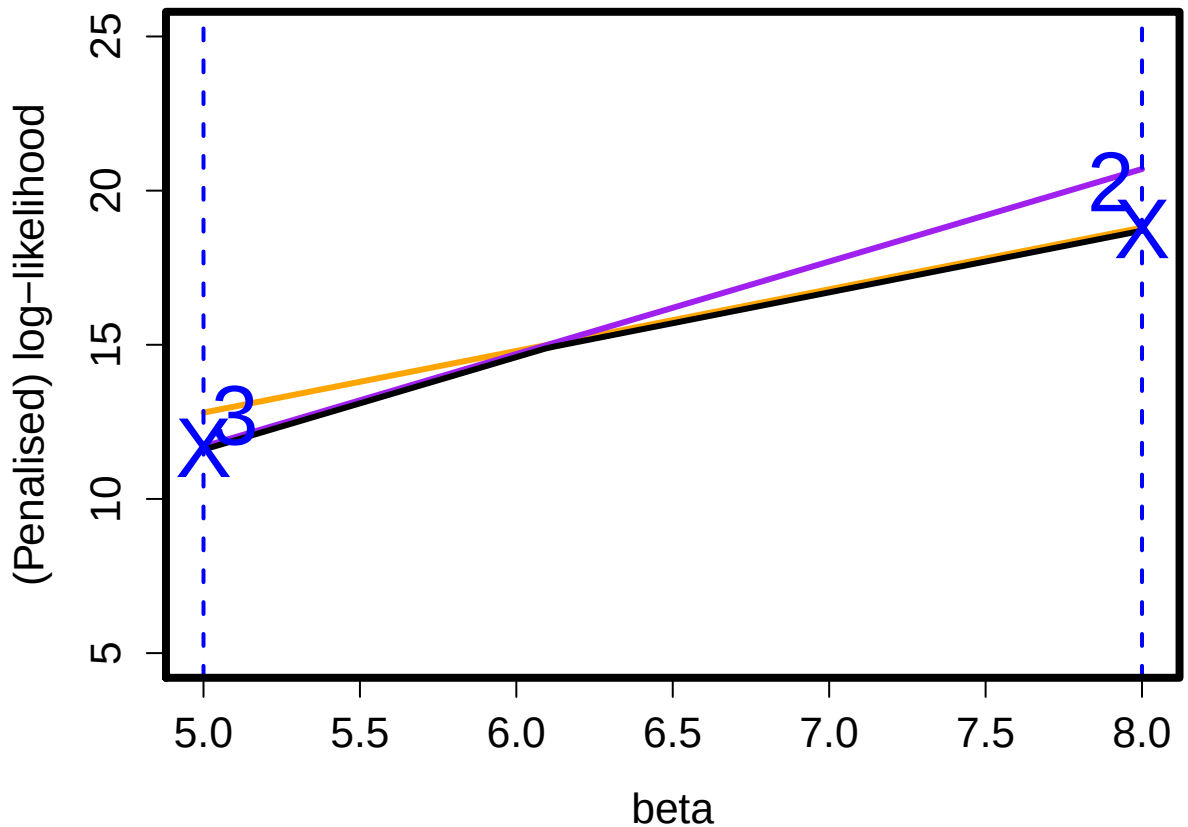
If both the segmentations have the same number of changepoints: **STOP**.
The optimal number of changepoints decreases as β increases. So all values of $\beta \in [\beta_{min}, \beta_{max}]$ will have the same optimal segmentation.

* Optimal Partitioning (Jackson et al 2005) or PELT (Killick et al (2012)).

Step 2: Compare the number of changepoints

STOR-i
excellence with impact

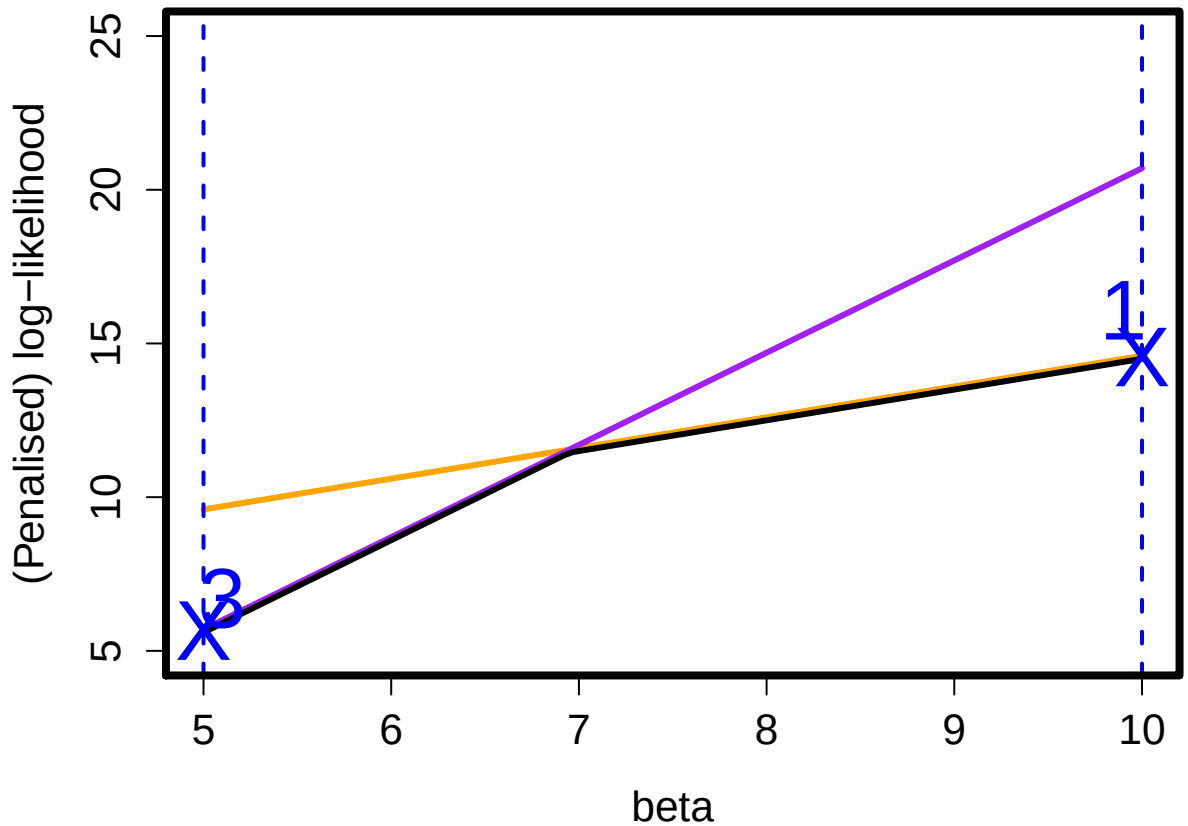
If the number of changepoints differs by 1, $m(\beta_{min}) = m(\beta_{max}) + 1$: **STOP**.



Step 2: Compare the number of changepoints

STOR-i
excellence with impact

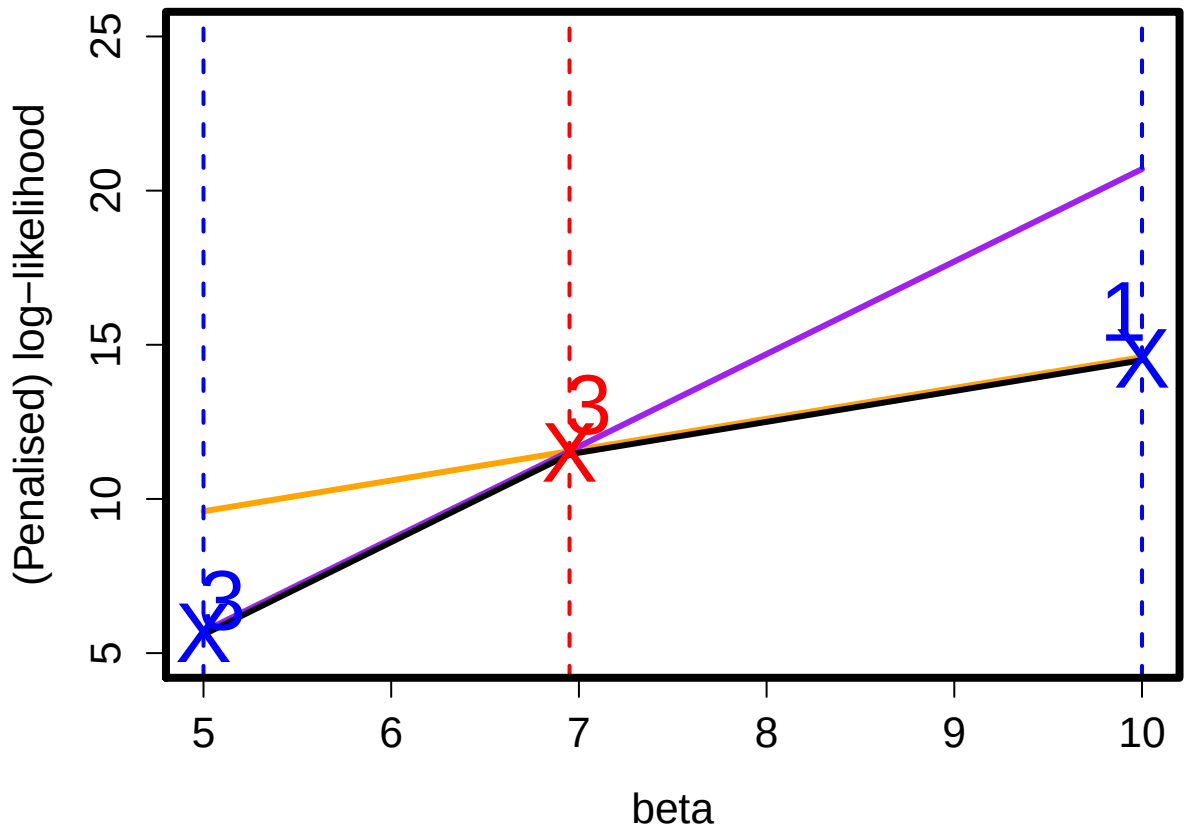
Otherwise calculate the point of intercept β^*, and run CPD algorithm with this value.



Step 3: Check if new Optimal Segmentation

STOR-i
excellence with impact

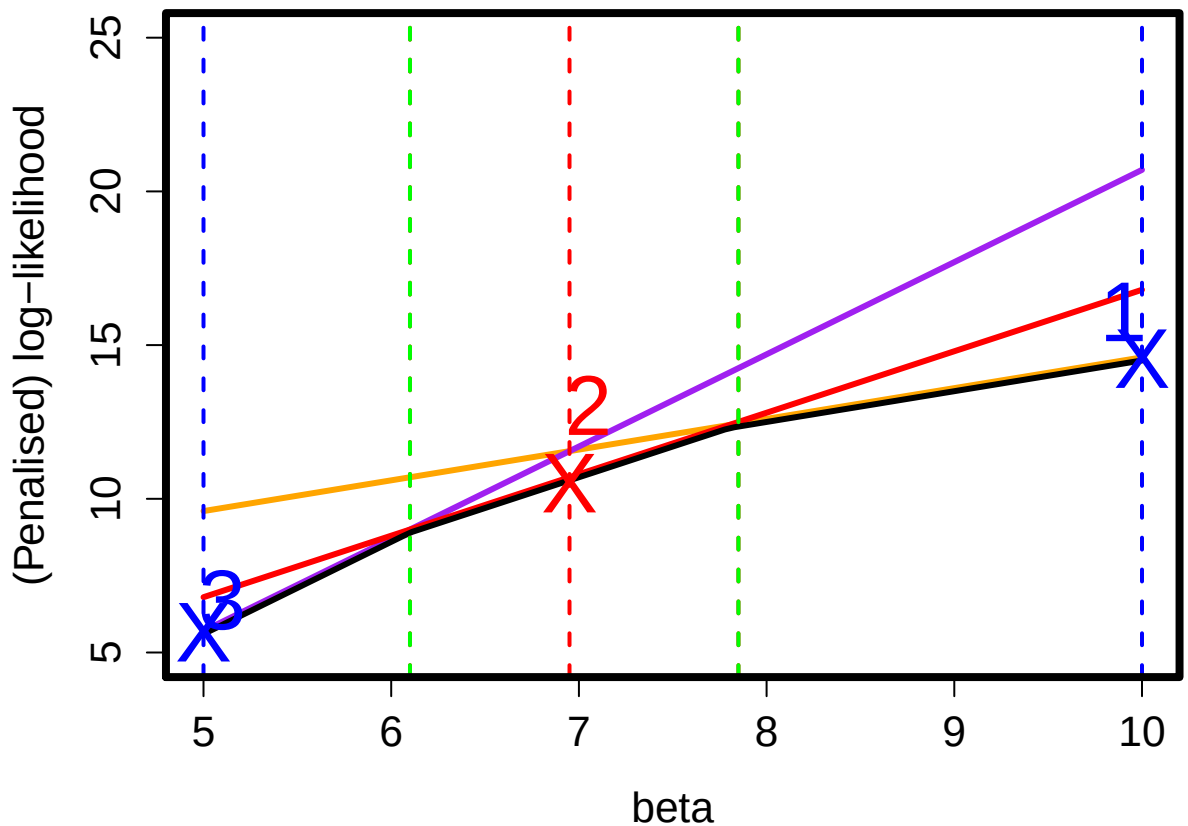
If the optimal segmentation for β^* is the same as for β_{min} or β_{max}: **STOP**.



Step 3: Check if new Optimal Segmentation

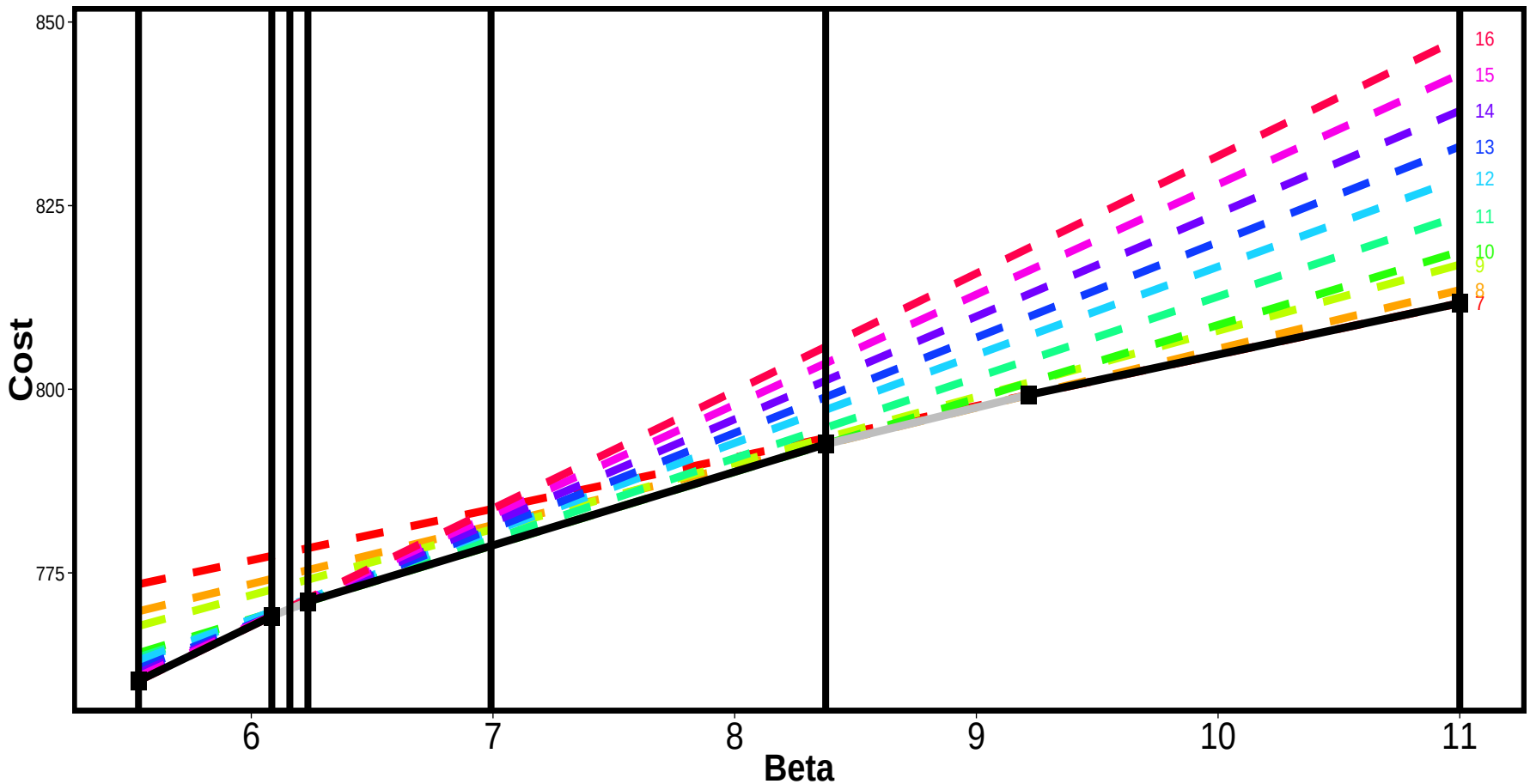
STOR-i
excellence with impact

Otherwise, treat the intervals $[\beta_{min}, \beta^*]$ and $[\beta^*, \beta_{max}]$ separately and process using Step 2.



Final Solution

STOR-i
excellence with impact



CROPS

STOR-i
excellence with impact

```
cpt.mean(data, penalty, pen.value = c(0,10), test.stat,
method = "PELT", minseglen)
```

if (length(pen.value == 2){
return(CROPS(data, pen.value, test.stat, minseglen, func="mean"))
}

(Haynes et al. (2014))

CROPS - Example output

STOR-i
excellence with impact

```
> cpt.meanvar(data, penalty = "CROPS", pen.value = c(11,200), class = TRUE)
[1] "Maximum number of runs of algorithm = 11"
[1] "Completed runs = 2"
[1] "Completed runs = 3"
[1] "Completed runs = 4"
[1] "Completed runs = 6"
[1] "Completed runs = 8"
[1] "Completed runs = 9"
Class 'cpt' : Changepoint Object
      ~~   : S4 class containing 13 slots with names
             cpts.full pen.value.full data.set cpttype method test.stat pen.type pen.value minseglen cpts ncpts.max param.est date

Created on  : Fri Jun 26 19:31:26 2015

summary(.)  :
----------
Changepoint type      : Change in mean and variance
Method of analysis    : PELT
Test Statistic  : Normal
Type of penalty       : CROPS with value, 11 200
Minimum Segment Length : 2
Maximum no. of cpts   : Inf
Changepoint Locations :
Number of segmentations recorded: 6  with between  2  and  11 changepoints.
 Penalty value ranges from: 11  to  154.7935
```

CROPS - Example output

STOR-i
excellence with impact

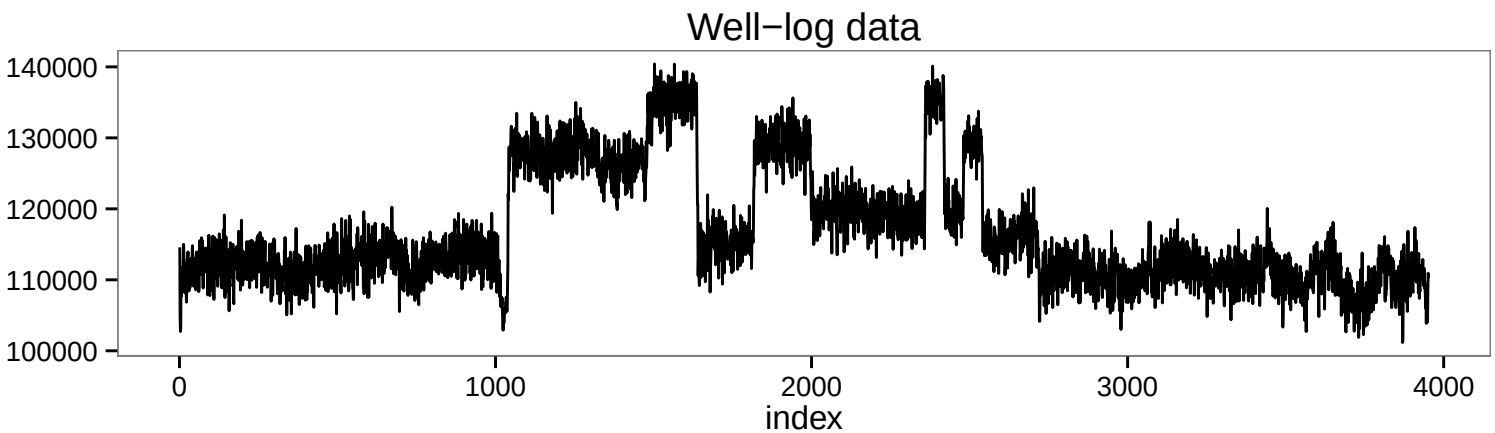
```
> cpt.meanvar(data, penalty = "CROPS", pen.value = c(11,200), class = FALSE)
[1] "Maximum number of runs of algorithm = 11"
[1] "Completed runs = 2"
[1] "Completed runs = 3"
[1] "Completed runs = 4"
[1] "Completed runs = 6"
[1] "Completed runs = 8"
[1] "Completed runs = 9"
$cpt.out
                          [,1]      [,2]      [,3]      [,4]      [,5]     [,6]
beta_interval          11.0000  11.67736  11.94004  12.14062  14.99827 154.7935
numberofchangepoints   12.0000  10.00000   9.00000   6.00000   4.00000   3.0000
penalised_cost        654.4173 677.77199 689.71203 726.13388 756.13041 910.9239

$changepoints
$changepoints[[1]]
 [1]   0  27  29  98 100 195 197 199 201 229 231 300

$changepoints[[2]]
 [1]   0  98 100 195 197 199 201 229 231 300

$changepoints[[3]]
[1]   0 100 195 197 199 201 229 231 300

$changepoints[[4]]
[1]   0 100 200 229 231 300

$changepoints[[5]]
[1]   0 100 200 300

$changepoints[[6]]
[1]   0 200 300
```

Example

Example

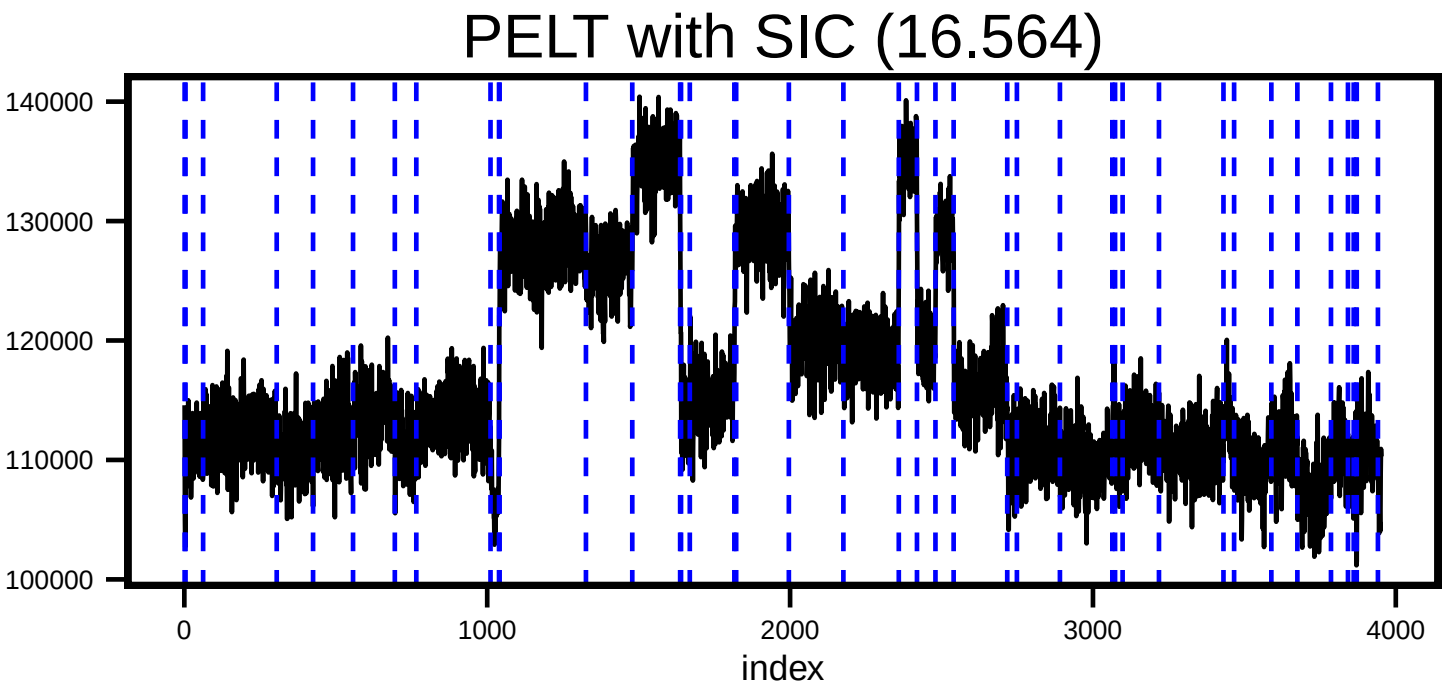
STOR-i
excellence with impact

- Changes in nuclear magnetic response as a probe is lowered down a bore-hole into the Earth's surface.
- Change in mean with $\hat{\sigma}^2 =$ median absolute distance



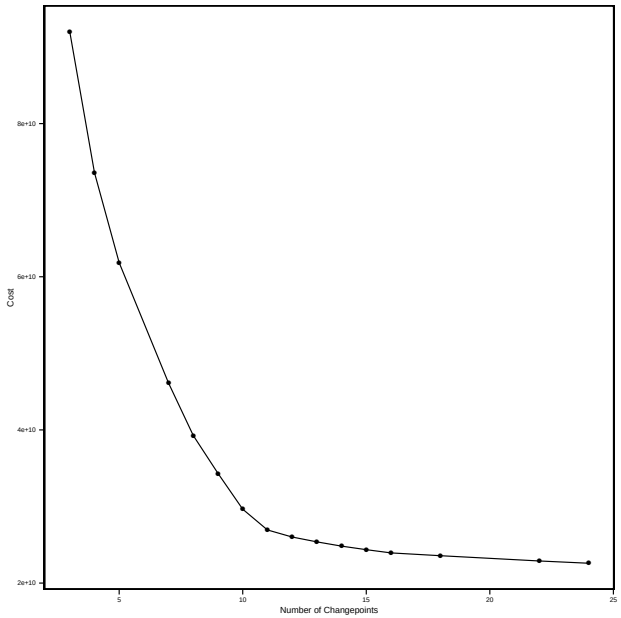
Example

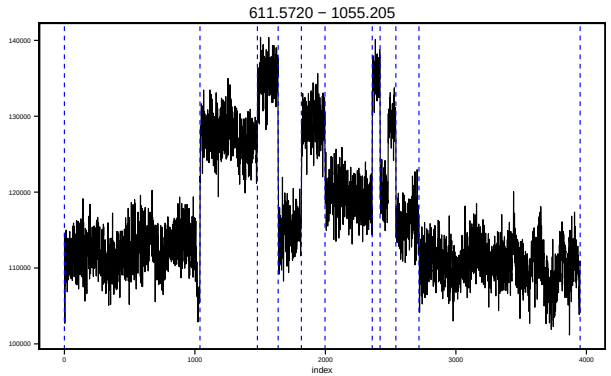
STOR-i
excellence with impact

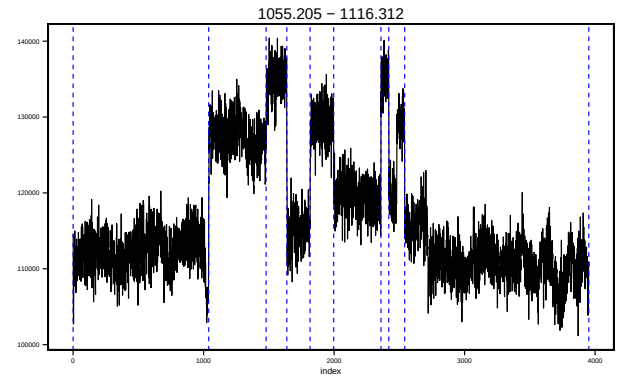


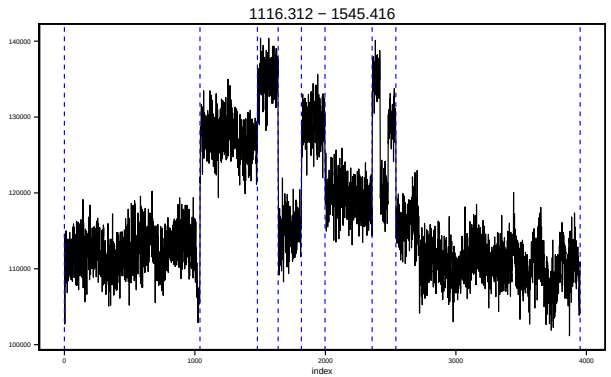
Example

STOR-i
excellence with impact









Summary

STOR-i
excellence with impact

- changepoint package has been updated to version 2.0.
- Penalty choice
  - Default changed to MBIC
  - Inclusion of CROPS - more appropriate to use a range of segmenations rather than just using a single choice
  - Popular choices such as the SIC lack robustness in the presence of model mis-specification

Thank you for listening.

References

STOR-i
excellence with impact


K. Haynes, I.A. Eckley and P. Fearnhead.
Efficient penalty search for multiple changepoint problems.
In Submission, arXiv:1412.3617, (2014)


B. Jackson and J.D. Sargle.
An Algorithm for optimal partitioning of data on an interval.
IEEE, Signal Processing Letters, 12(2):105–108, 2005.


R. Killick, P. Fearnhead and I.A. Eckley.
Optimal detection of changepoints with a linear computational cost.
JASA, 107, 1590-1598, 2012.


R. Killick, and I.A. Eckley.
changepoint: An R package for changepoint analysis.
Journal of Statistical Software, 58(3):1–19, 2014.


R. Killick, I.A. Eckley and K. Haynes.
changepoint: An R package for changepoint analysis.
R package version 2.0, 2015 (To appear very soon).