Bayesian Online Changepoint Detection

Max Mihailescu

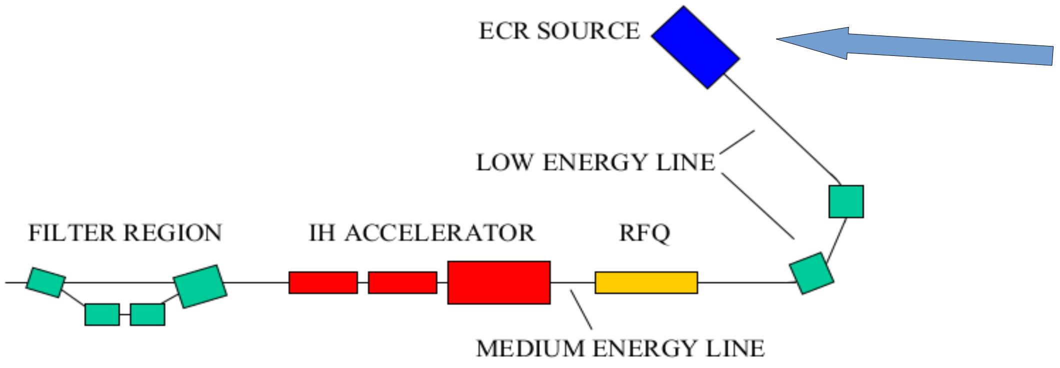
07.08.2020

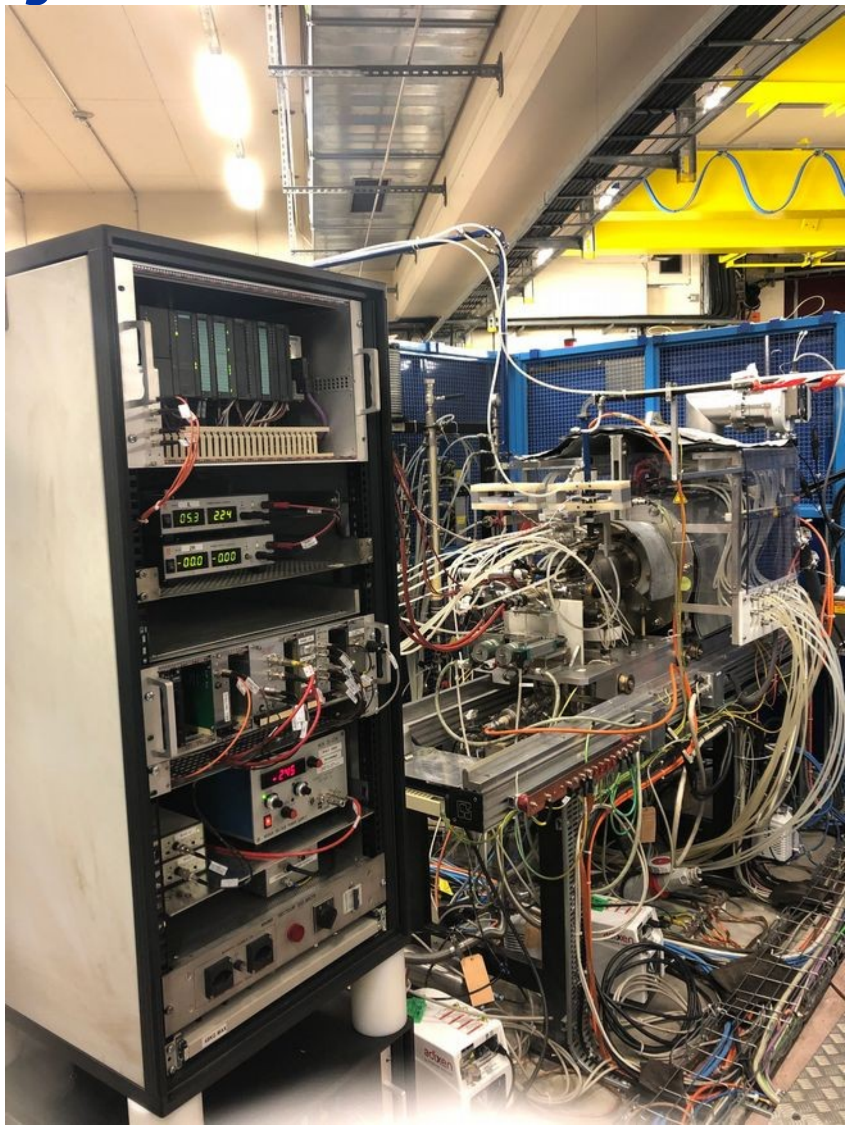
Outlook

- Motivation at Linac3
- Presentation of the Algorithm
- Code example

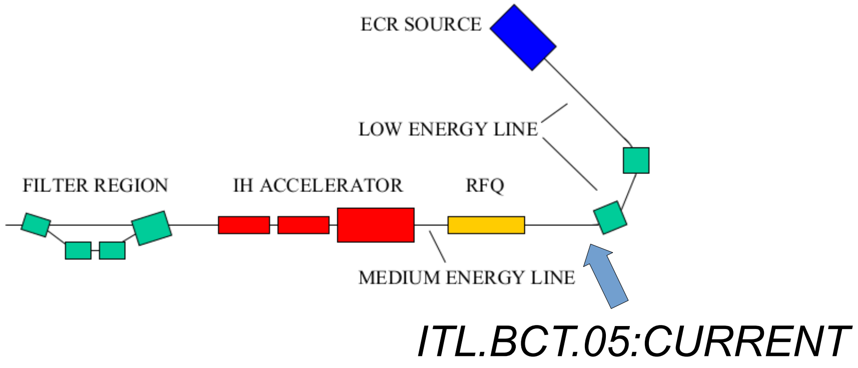
Layout of LINAC3 and the objective

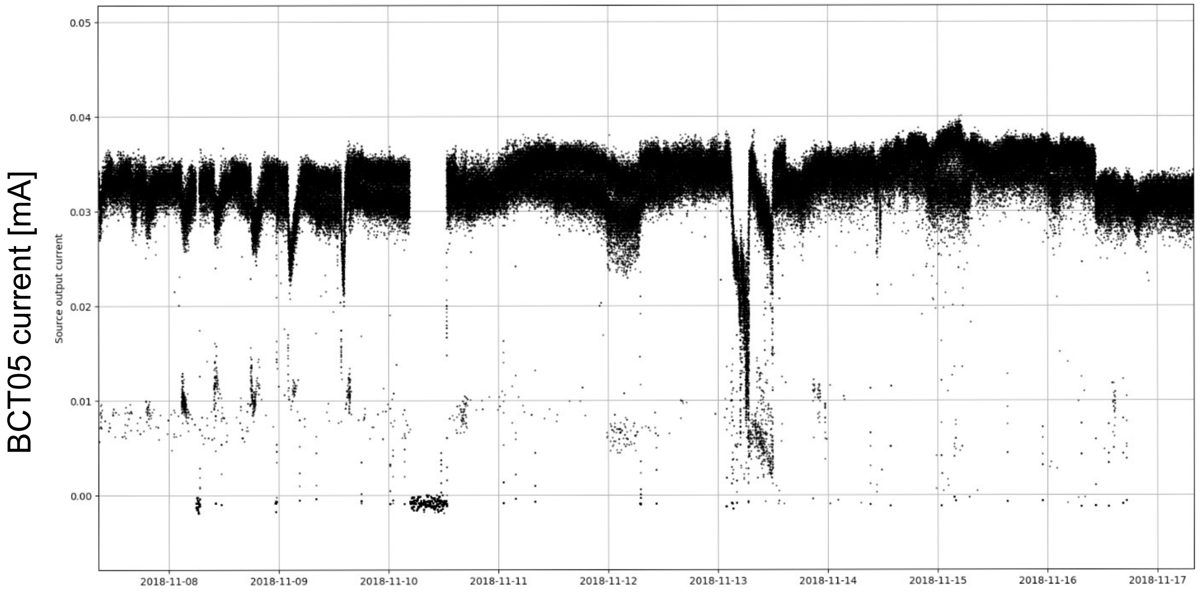
Linac3 produces heavy ions, mainly Pb.





Layout of LINAC3 and the objective





There is a lot to tweak at the source: Oven Power, RF, three solenoids, gas inflow, …

This leads to a very complex system which can produce results that are very difficult to predict.

➡ The goal is to quickly identify regime changes in the output beam current

Online Changepoint Detection

Given an ordered collection of data points (a time series)...

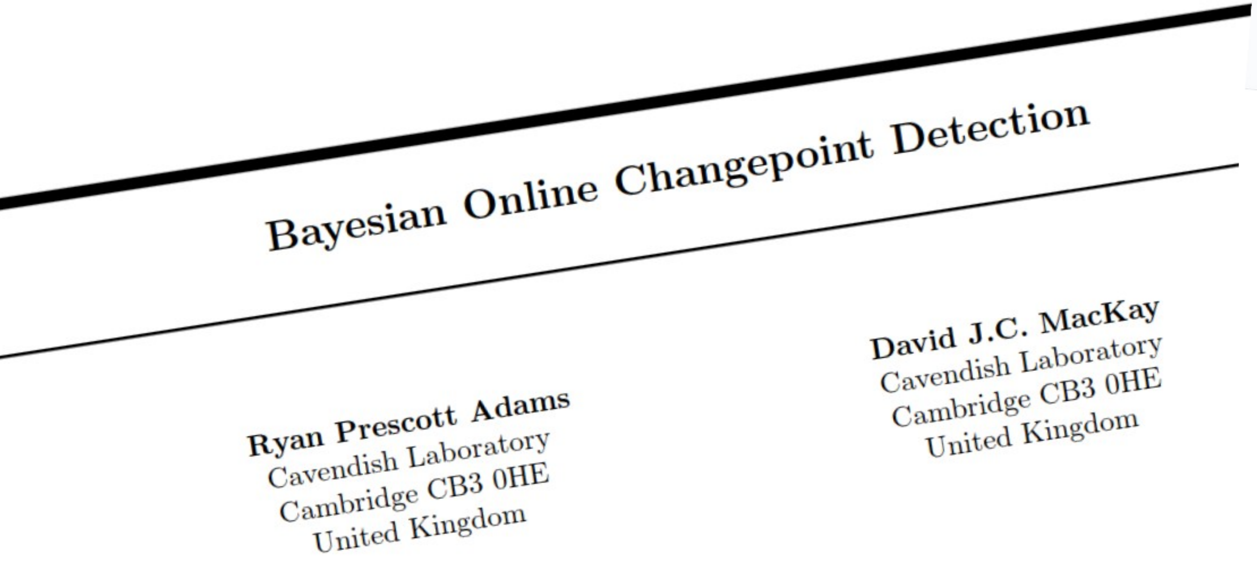
...changepoint detection is the task of detecting qualitative changes in the observed time series.

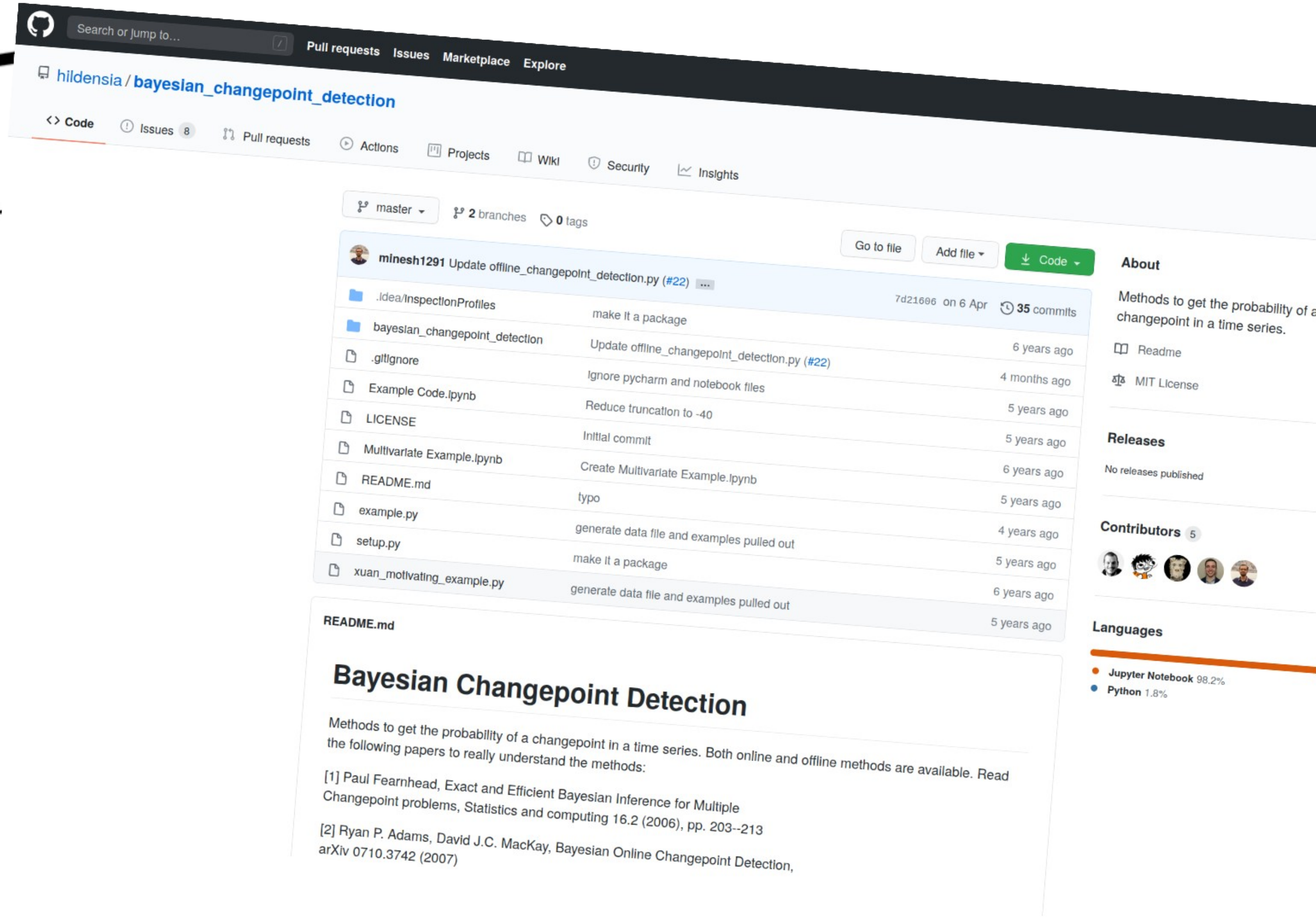
- Probabilistic models search for changes in distribution of data
- Other models minimize a cost function (e.g. piece wise linear regression)

...online changepoint detection performs analysis on streaming data.

- The full time series is typically not known
- Whenever new data is observed, decide if a changepoint occurred
- update model to include new knowledge (optional)

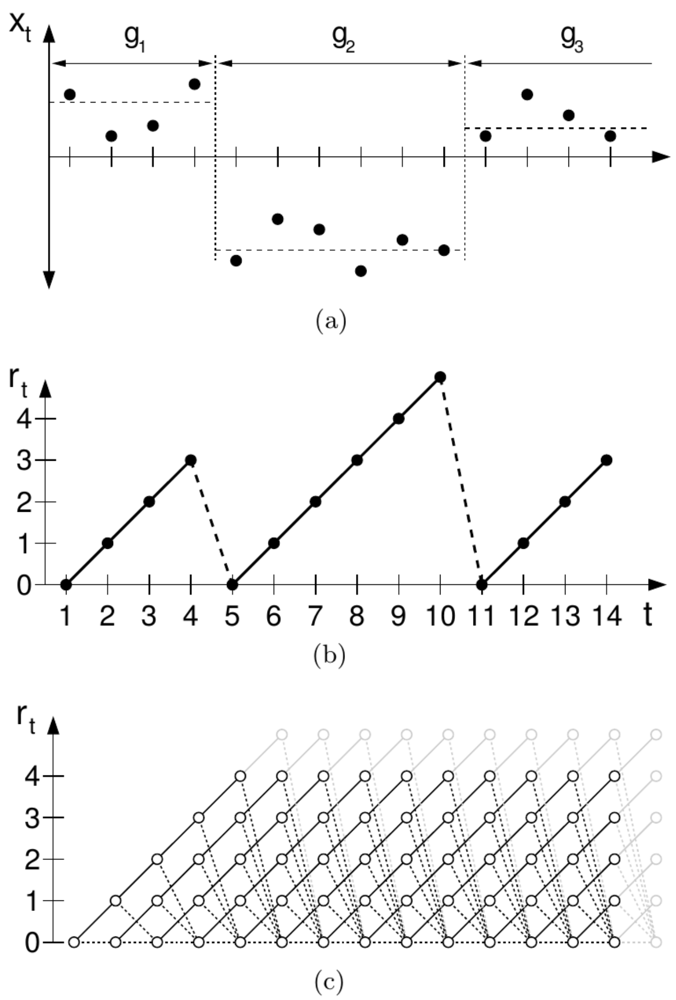
Paper by Adams and MacKay (2007)





Published on arXiv: https://arxiv.org/abs/0710.3742
Python implementation (MIT license):
https://github.com/hildensia/bayesian_changepoint_detection

The concept of run lengths



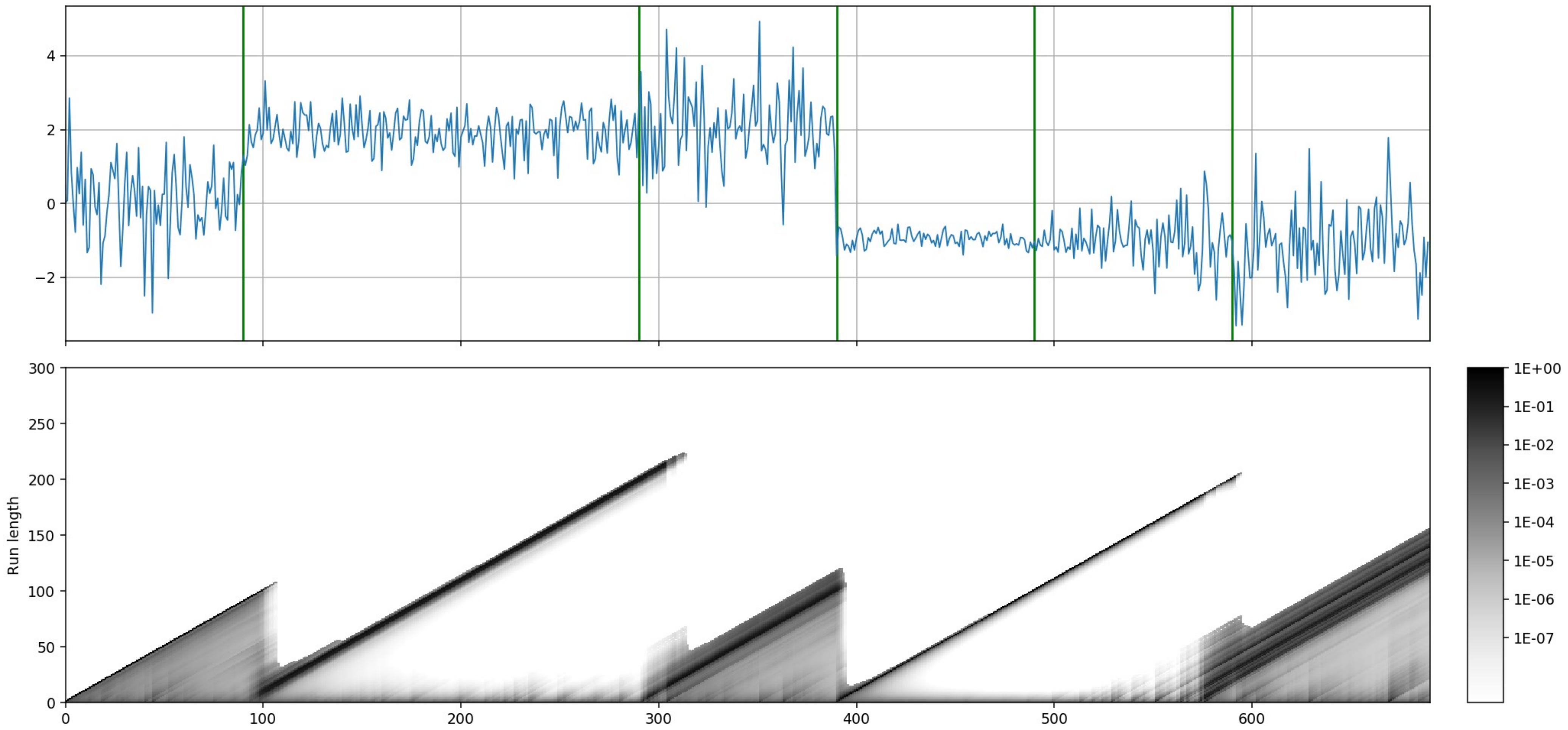
Consider an ordered series of points where the observations are drawn i.i.d. from different distributions.

The run length r_t indicates how many time steps have passed since the last changepoint occurred.

Build a probabilistic model with probabilities for all possible run lengths at every time step.

Graphic from Bayesian Online Changepoint Detection, Adams & MacKay (2007)

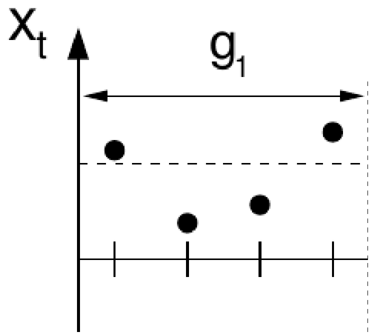
Example output on simulated data



Simulated time series with true changepoints in green (top),
Run length probabilities (bottom) – darker pixels mean represent higher probability

Mathematical formulation of the model

We are interested in every time a new observation arrives



$$P(r_t \mid x_{1,} \dots, x_t) = \frac{P(r_t, x_{1,} \dots, x_t)}{P(x_{1,} \dots, x_t)}$$

Expansion leads to recursive term

$$P(r_t, x_{1,} \dots, x_t) = \sum_{r_{t-1}} P(r_t \mid r_{t-1}) P(x_t \mid r_{t-1}, x_{1,} \dots, x_{t-1}) P(r_{t-1}, x_{1,} \dots, x_{t-1})$$

Changepoint prior

Predictive probability
of new observation

Previous run length
probabilities

Afterwards update model parameters to reflect information gained from x_t

Graphic from Bayesian Online Changepoint Detection, Adams & MacKay (2007)

The Changepoint Prior

$$\sum_{r_{t-1}} \boxed{P(r_t \mid r_{t-1})} P(x_t \mid r_{t-1}, x_{1,} \ldots, x_{t-1}) P(r_{t-1}, x_{1,} \ldots, x_{t-1})$$

Given r_{t-1} there are only two possible outcomes for r_t:

- Changepoint happened: $r_t = 0$
- No changepoint: $r_t = r_{t-1} + 1$

$$P(r_t \mid r_{t-1}) = \begin{cases} H(r_{t-1}+1) & \text{if} \quad r_t = 0 \\ 1 - H(r_{t-1}+1) & \text{if} \quad r_t = r_{t-1}+1 \\ 0 & \text{otherwise} \end{cases}$$

H is called hazard function and encodes how the changepoint probability relates to the time passed this the last change

e.g. a mechanic piece that is more likely to break as it gets older

If changepoints are equally likely at any time: $H = 1/\lambda$ with $\lambda \approx$ expected duration of segment

A word on the predictive probability

$$\sum_{r_{t-1}} P(r_t \mid r_{t-1}) \boxed{P(x_t \mid r_{t-1}, x_{1,} \ldots, x_{t-1})} P(r_{t-1}, x_{1,} \ldots, x_{t-1})$$

Lets assume our data is in a segment follows a distribution **D** with (unknown) parameters **θ**.
Then, one could say

$$x_t \mid r_{t-1}, x_{1,} \ldots, x_{t-1} \sim D(\bar{\theta})$$

With an estimate $\bar{\theta}$ of the parameter.
But, we are unsure about this estimate

$$\Rightarrow p(x_t \mid r_{t-1}, x_{1,} \ldots, x_{t-1}) = \int_{\theta} p(x_t \mid \theta) * p(\theta \mid r_{t-1}, x_{1,} \ldots, x_{t-1}) d\theta$$

Likelihood (how likely is the observation?)

Prior (how well does the parameter explain the previous observations)

➡ In general, this integral needs to be evaluated numerically (using MCMC methods)

A word on the predictive probability

In our case we can make our lives easier using conjugate priors!

$$\int_{\mu,\sigma^2} N\left(x_t \mid \mu,\sigma^2\right) * p\left(\mu,\sigma^2 \mid r_{t-1}, x_{1,}\ldots,x_{t-1}\right) d\left(\mu,\sigma^2\right)$$

This is a normal with unknown mean and variance

We can choose prior distribution to get an algebraic expression

The conjugate prior has four parameters that we update every time a new data point *x* is observed.

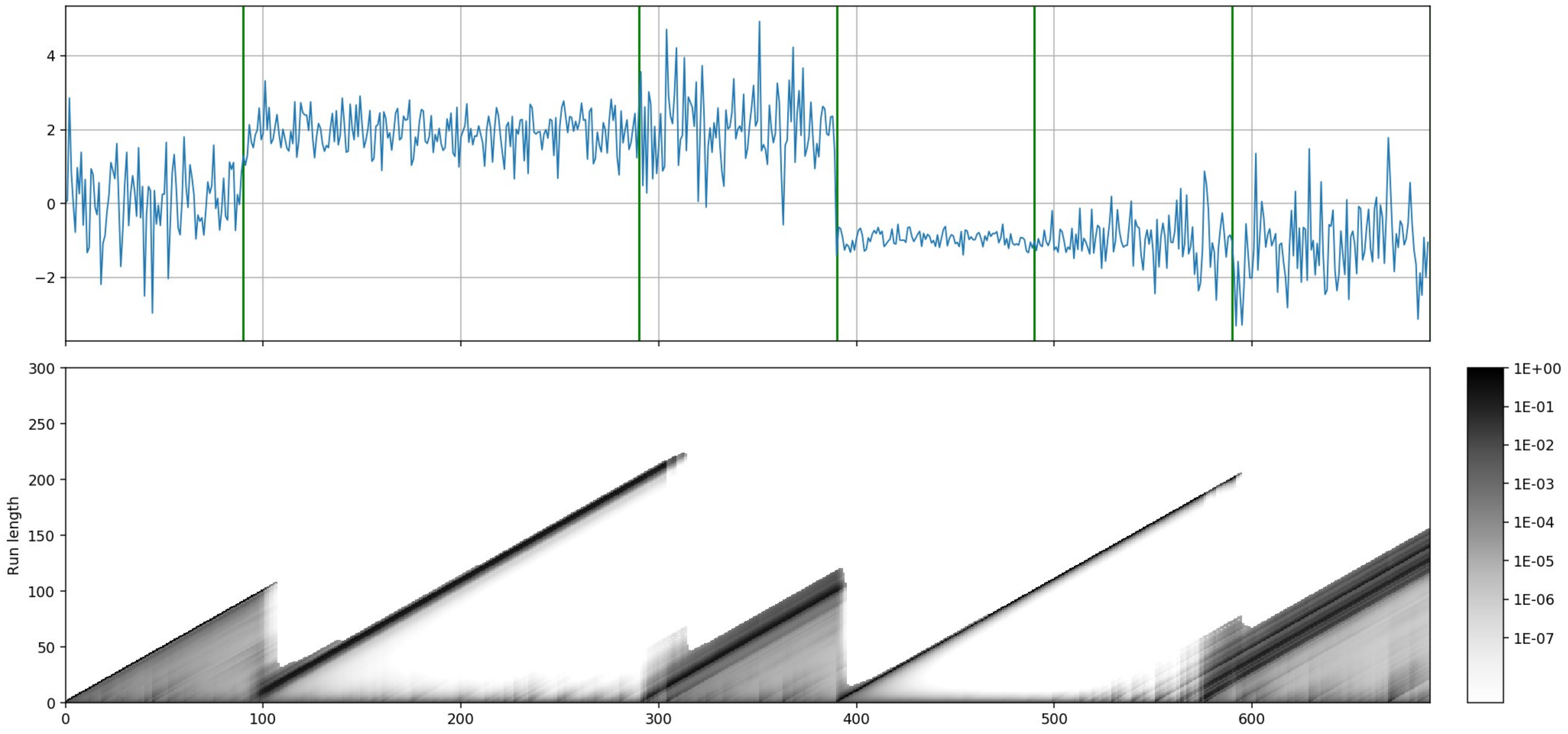
$$\kappa \rightarrow \kappa+1 \qquad \alpha \rightarrow \alpha+0.5$$

$$\hat{\mu} \rightarrow \frac{\kappa\mu+x}{\kappa+1} \qquad \beta \rightarrow \beta + \frac{\kappa}{\kappa+1}\frac{(x-\hat{\mu})^2}{2}$$

Without going into detail, the integral equals the probability density of a Student-T distribution

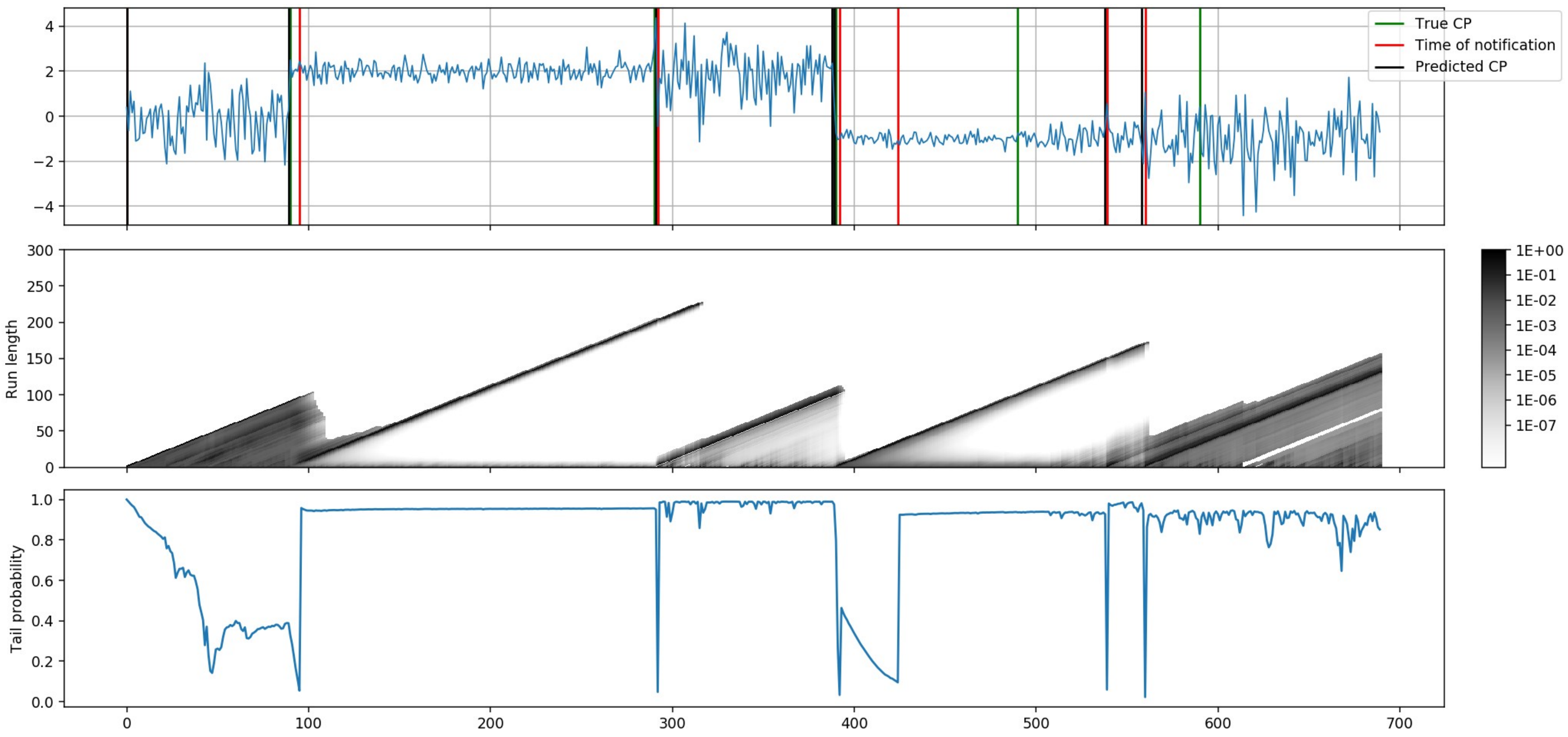
$$t_{2\alpha}\left(\hat{\mu}, \frac{\beta(\kappa+1)}{\alpha\kappa}\right)$$

And now?



Simulated time series with true changepoints in green (top),
Run length probabilities (bottom) – darker pixels mean represent higher probability

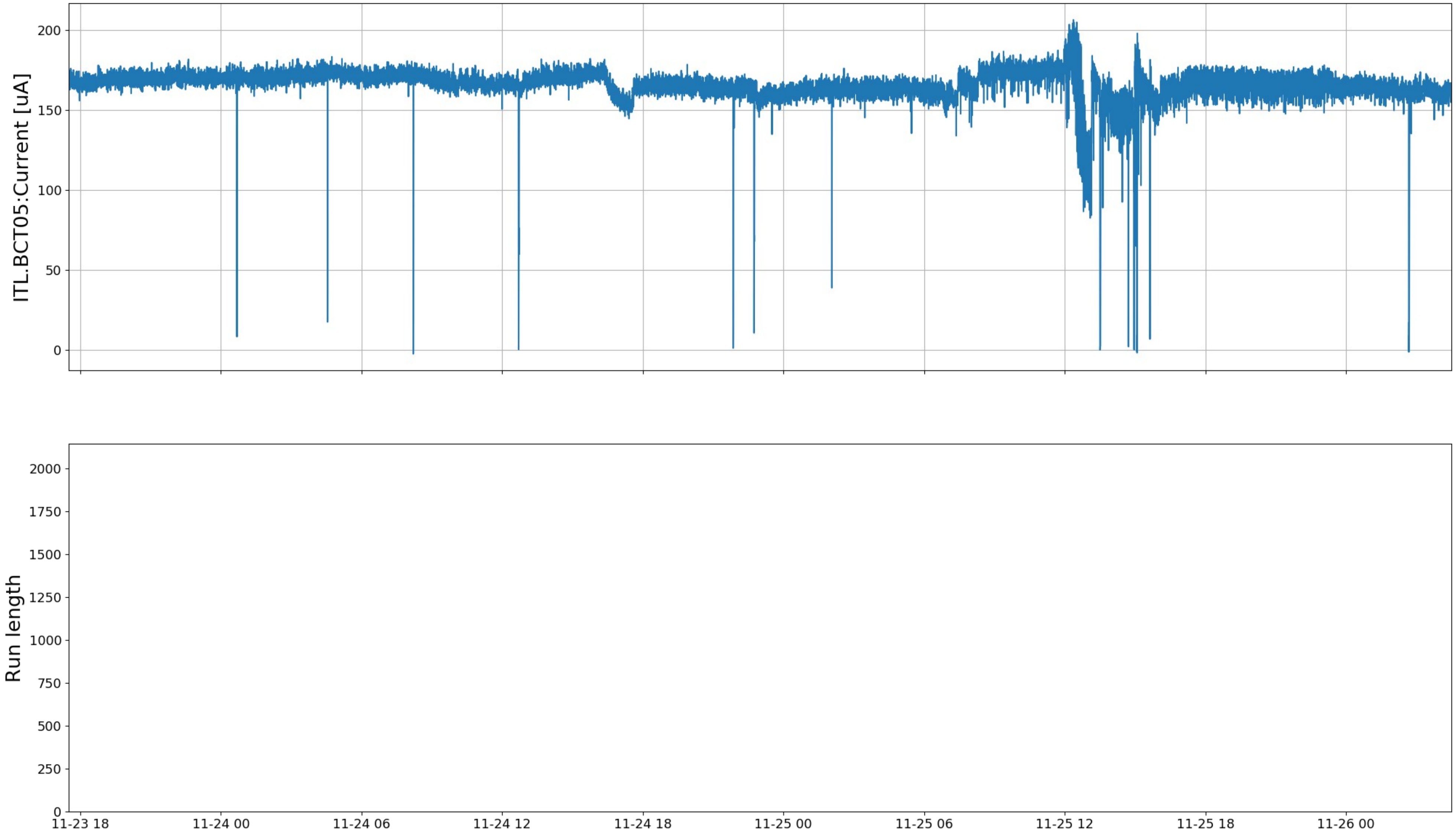
And now?



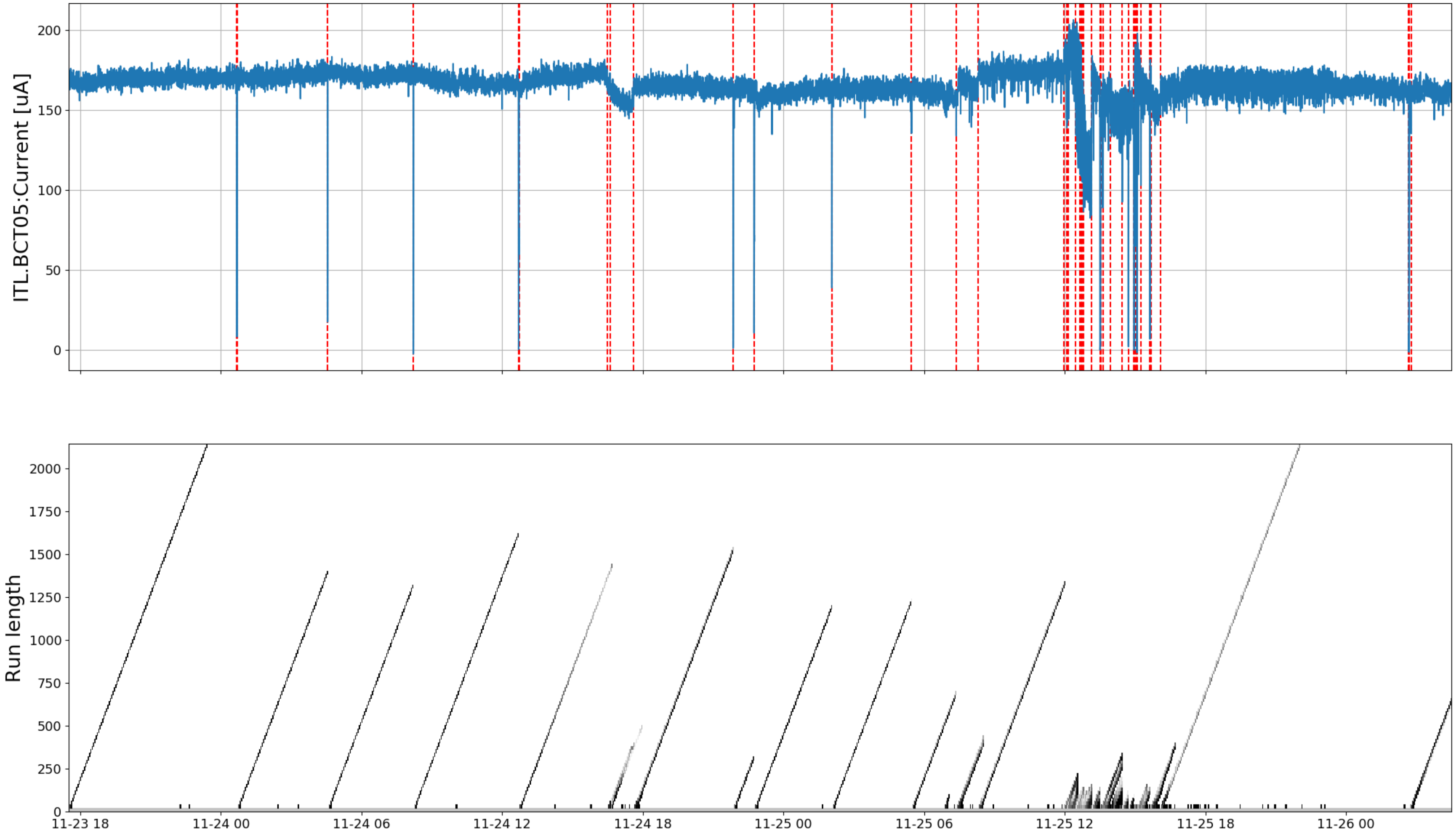
How to translate probabilities to actionable information?
Every time new data arrives, ask “Is it likely that the new data point is in the same segment?“

➡ Sum up the probability of the longest 20% run lengths and check if sum is smaller than 0.1

First results on linac3 data



First results on linac3 data



CERN