Statistical analysis of operational data of the GTS-LHC ion source

D. Küchler, CERN, BE/ABP/HSL, 1211 Geneva 23, Switzerland
M. Mihailescu, Rheinische Friedrich-Wilhelms-Universität Bonn, D-53113 Bonn, Germany

Abstract

The GTS-LHC ECR ion source delivers ion species for the physics programme of the CERN accelerator complex. Stability and reproducibility are key parameters for a successful operation. Therefore the source requires regularly intervention by an expert.

The integrated functionality of data logging in the control system of the CERN accelerator complex allows to store all relevant control values and beam properties for the different machines. For the years 2015, 2016 and 2018, the data for the lead ion operation of the GTS-LHC ion source were extracted from the logging and treated with various statistical methods to identify recurring patterns in the operational settings of the source. If such patterns exist, they may be used in the future for the control of the source by semi-automatic feedback loops.

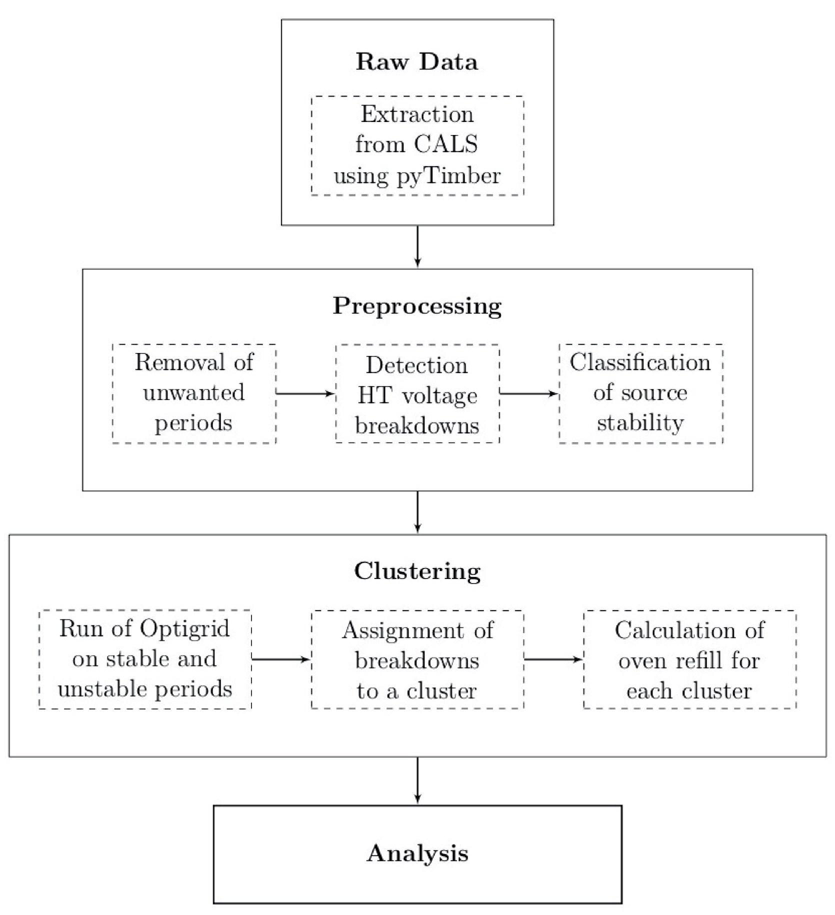
The main study performed was based on a cluster analysis using the Optigrid algorithm, which is well suited for problems with a high number of parameters.

Introduction

The setting up and tuning of the source depends presently on the experience of a team of source experts. As this team is very small it is a single point of failure for the operation. To allow a more person independent and more standardized operation a request to develop software tools to support the operation of the source was made (GTS-LHC operation support – GHOST). These tools should help to improve the source stability and reliability and to reduce the necessity of the intervention of an operator or source expert.

The settings and acquisitions used in this analysis are the most commonly adjusted settings (source solenoids, microwave power, gas injection, oven power, bias disk voltage). The acquisitions are values from the beam current transformers along the linac (one directly after the charge separation and one at the end of the linac) and the current of the high voltage power supply of the source.

The settings of the two ovens were not used in the clustering itself, but were included in the analysis of the resulting clusters. Since the need for ramping up the oven during one lead cycle is already well known, the oven power was excluded from the clustering input parameters to reduce the dimensionality and increase interpretability of the results.



Flowchart of the different steps of the data analysis

The clustering algorithm

Clustering is the process of grouping together data points that are *similar*. Clustering is an unsupervised learning method, i.e. it has to discover unknown classes in the data. The goal is to find similar settings of the source using clustering. So one will consider two points as similar if they represent the same setting of the source up to some variation that accounts for small changes or noise.

When it comes to the actual clustering, there is a wide range of algorithms to chose from, and while there exist several measures to grade clustering results it is generally not possible to say which algorithm will be the better one, as this depends on the actual use case and definition of similarity. In the presented case clustering for unsupervised learning is used, meaning that the output classes are not known a priori. Hence the results have to be evaluated to see if the determined clusters fit the requirements.

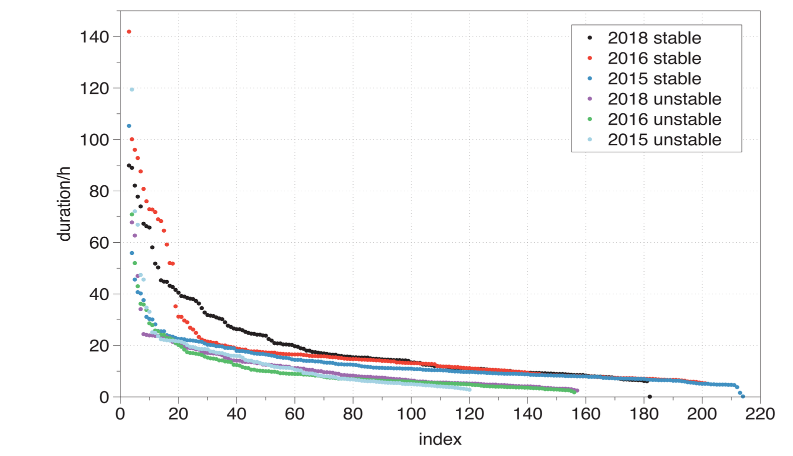
For this analysis the Optigrid algorithm developed by Hinndeburg an Keim was selected. This algorithm was developed to solve the problem of clustering in high dimensional data (known as the „Curse of Dimensionality").

The Optigrid algorithm partitions the input space into a grid, where each cell represents one cluster. Each cluster is described by its median and standard deviation in percent of mean, to capture the centre and dispersion of a cluster. In the present case, the two values can be interpreted as the setting that is represented by the cluster, and the variability of the setting, i.e. whether it was almost constantly on one value or if there were fluctuations.

After each point is assigned to one cluster, it is possible to calculate attributes for each cluster, e.g. its duration and the number of high voltage breakdowns while it was active.

General observations and cluster size

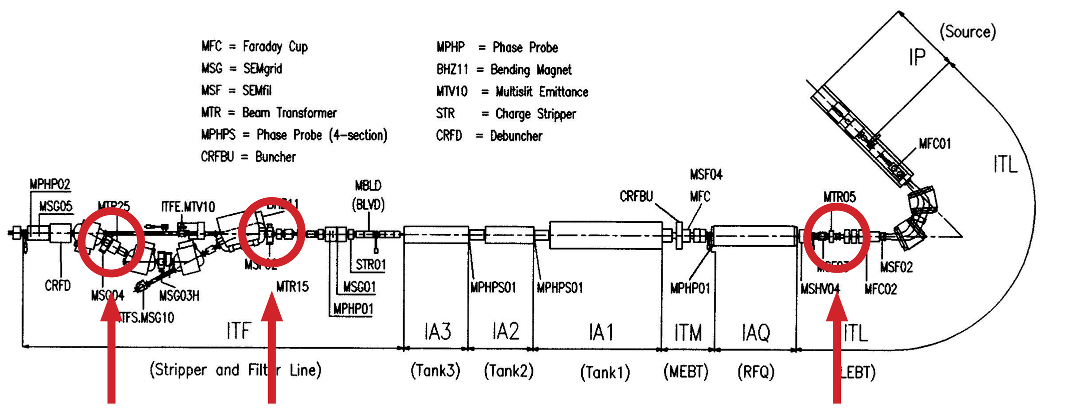
The calculation delivers a vast number of clusters which cannot easily combined to bigger clusters without losing information. But it can also be seen that if one orders the clusters by there duration, the size of the clusters drop quickly.



There are few large clusters and many small ones. There are only small differences between the years and the operation modes. The stable operation mode in 2016 seems to have the highest number of long clusters. The longest one has a duration of 142 hours. But this is still short compared to the yearly operational period of around 6000 hours or compared to the time between two oven refills (two weeks = 336 hours). The duration accumulated per cluster is for most of the cases less than 30 hours. Compared to the yearly operational period this shows that the settings are varied quite often. There is no dominant setting.

A detailed look at the internal distribution of the clusters in time also reveals that some clusters are continuous in time, while other are split up into many small fragments.

Layout of Linac3 and positions of the intensity measurement



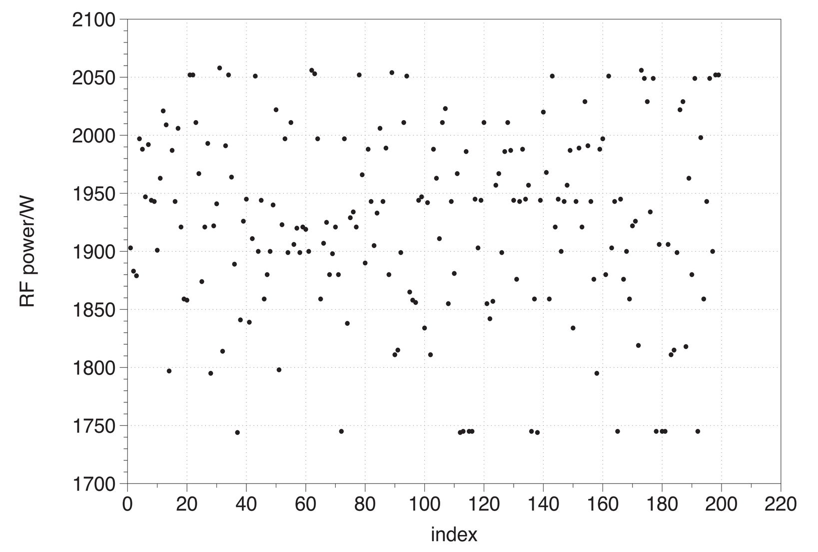
ITF.BCT25
after filter
only Pb^{54+} left

ITF.BCT15
after after stripper
several charge states
centrered around Pb^{54+}

ITL.BCT05
after source spectrometer
before RFQ
Pb^{29+}

Cluster and operational parameter

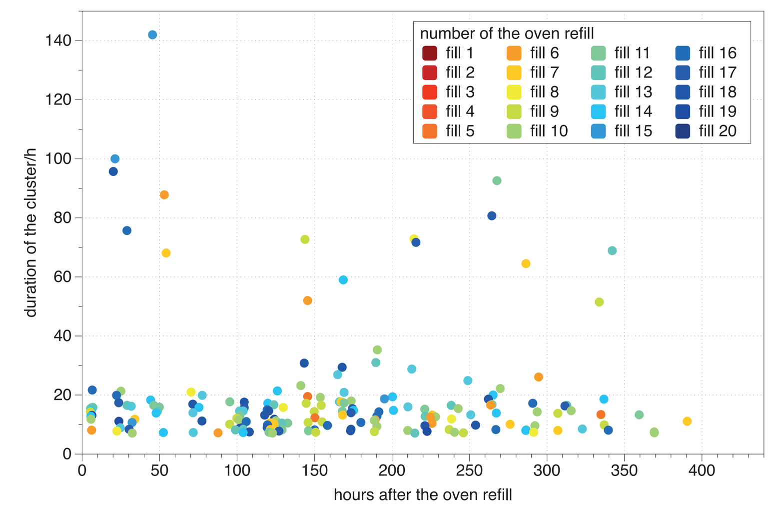
The RF power was chosen to give an impression of the distribution of the medians of the clusters. The Figure shows the medians of the clusters for RF power of the stable period in 2016.



The clusters are ordered by size. The lower index represents the larger cluster. One can see that the used settings do not accumulate around a certain value but only lie in a (relatively broad) range.

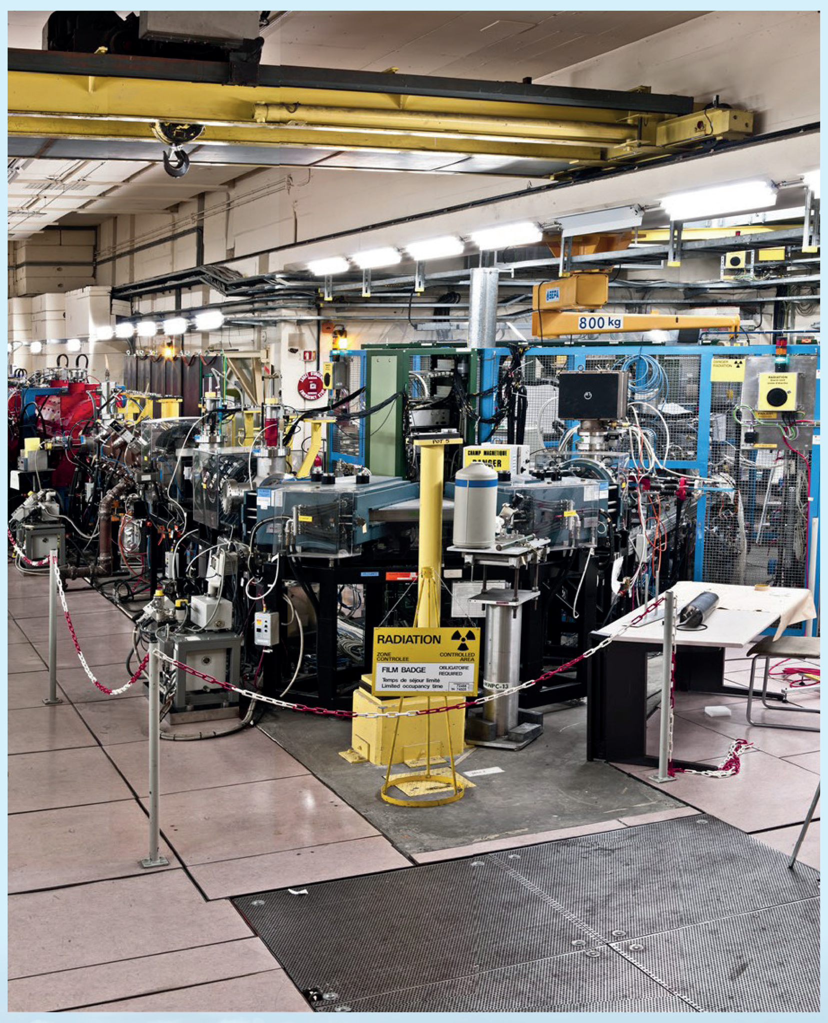
Cluster and oven operation

Between two oven refills there is normally a period of two weeks (14 days = 336 hours). In the Figure the duration of the clusters is plotted in relation to the time after the last oven refill for the stable operation periods in 2016. With the colour the number of the fill in the respective year is coded. There does not appear to be a relation between duration and time after oven refill.



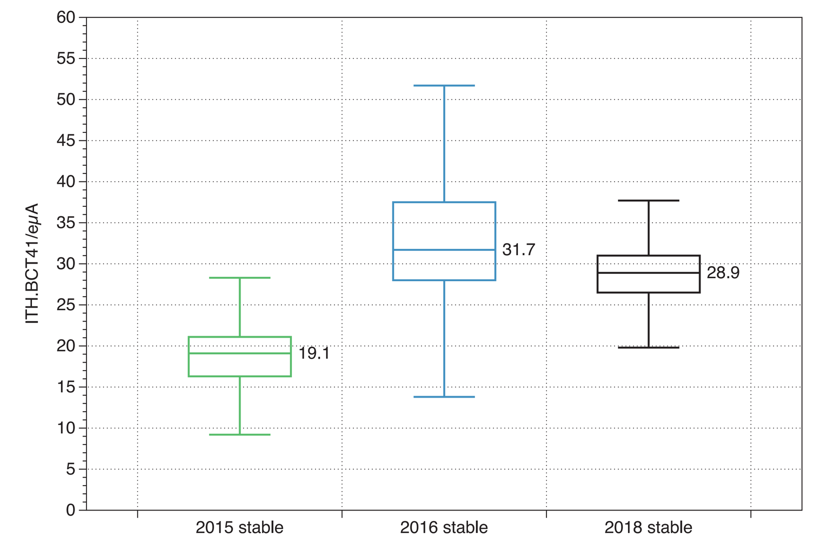
The first week ($t < 168$ hours) after an oven refill the source needs normally very little tuning. So one could expect to find longer clusters in this time periods. But large clusters can be found at any time of the oven running period. No moment is preferred for longer clusters. In the second week ($t > 168$ hours) after an oven refill the source needs to be tuned quite often to keep the intensity and the stability. So one would expect to see a more short clusters there. This could not be confirmed by looking only at the cluster duration.

After an oven refill there is normally only little tuning on the source. So one may have the expectation that the two time periods before and after an oven refill fall into the same cluster. But also this expectation could not be confirmed.



Source and linac performance

The Figure below shows the ion intensity of the Pb^{54+} beam at the end of the linac for the stable periods of the three different years. Stability was measured by defining threshold in beam intensity and variation in time windows of several hours.

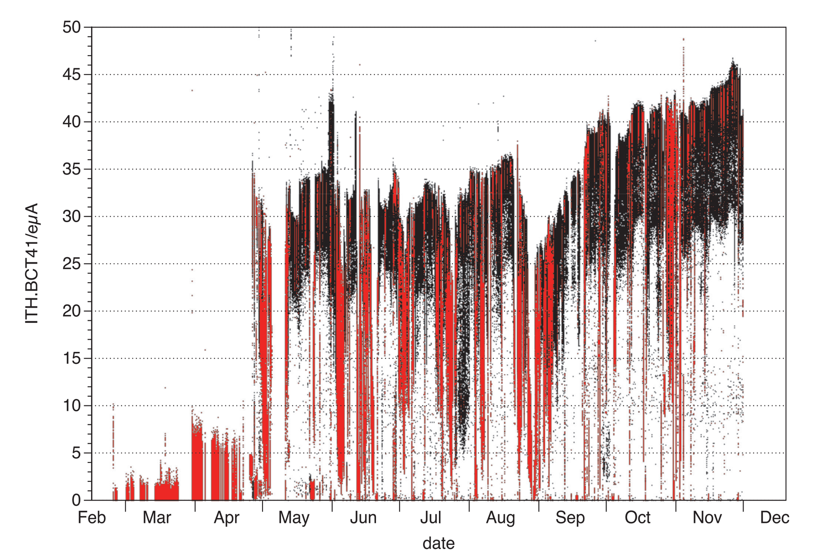


There are clear differences between the years visible. The increase from 2015 to 2016 was the result of the modification of the low energy beam transport.

During 2016 the best average intensity could be delivered. In 2018 the intensity was lower due to problems with the stripper foils (the beam at the end of the linac is stripped from Pb^{29+} to Pb^{54+} in a thin carbon stripper foil).

The very high values (currents > 40 eµA) for the year 2016 are an artefact, as for degrading stripper foils the charge states Pb^{53+} and Pb^{55+} can sneak through the charge state filtering after the stripper and so the integral current measured in the beam current transformer becomes higher.

In the following Figure the ion beam intensity at the exit of the linac is plotted for 2016. Stable periods are marked in black, unstable periods are marked in red.



If one compares the integral of stable with the unstable periods per year one get 62.8 % for 2015, 70 % for 2016 and 68% for 2018. One can see that in 2016 there were long periods of stable beam operations, in particular during the month of November, when the 2016 heavy ion run took place.

Conclusions and outlook

This present cluster analysis confirmed the qualitative experience of the operators. No dedicated recurring pattern in the source settings could be found. Neither between the years and nor even within the years.

All source settings are within certain boundaries, but there is no single set or a couple of sets of settings that can be used to get directly to a stable and reliable source operation. None of the found clusters can be used directly to create one or several GHOST modules which can run the source at an automatised or semi-automatised mode.

More sophisticated time series analysis tools, e.g. LSTM neural networks, could potentially discover more complex relations than investigated in the study, and could in the future be used for (mid- or long-term) beam quality prediction.

At the moment in all these studies the source is treated as a black box. Additional information about the interior of the box beyond the high voltage drain current and the ion current will help to improve the understanding of the source behaviour and the operation. A way to gain knowledge about the processes inside the source (inside the plasma) could be optical emission plasma spectroscopy. This can be performed parasitically. A corresponding study is in preparation.

Contact information

Dr. Detlef Küchler
CERN
BE Department
CH-1211 Geneva 23
Switzerland
detlef.kuchler@cern.ch