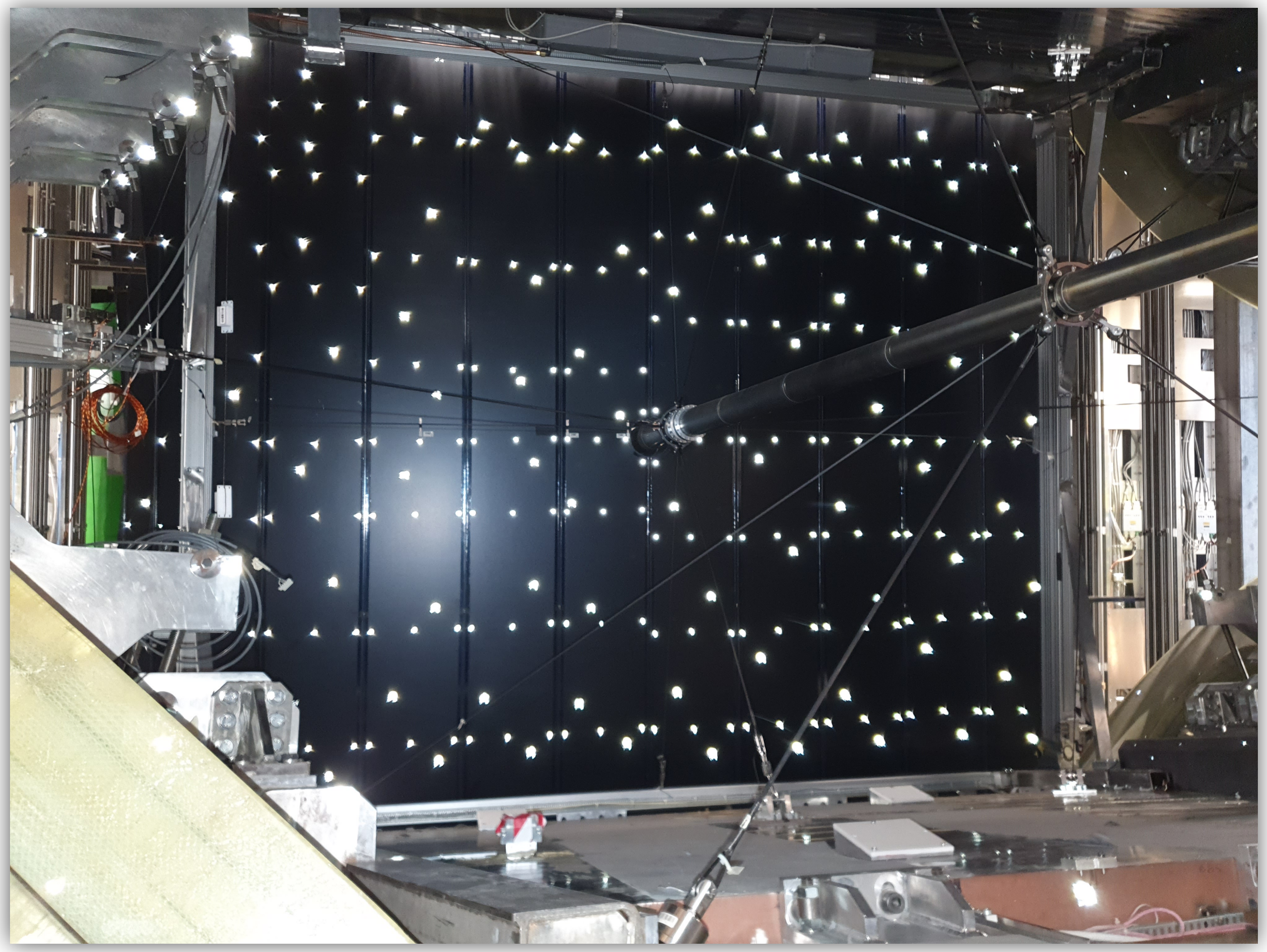
Data quality status

Elisabeth Niel

27/02/2023



Introduction

SciFi known challenges from last year relevant for efficiencies:

1. Fine and coarse time alignment shifts observed which degrades the efficiency → few percent change in efficiency from fine time align. and 100 % loss from BXID jumps (coarse time align.)
2. Tell40s and FE sending errors after some time taking data→100 % loss in a module
3. Calibration of the thresholds → reduce noisy channels
4. Other degrees of freedom :

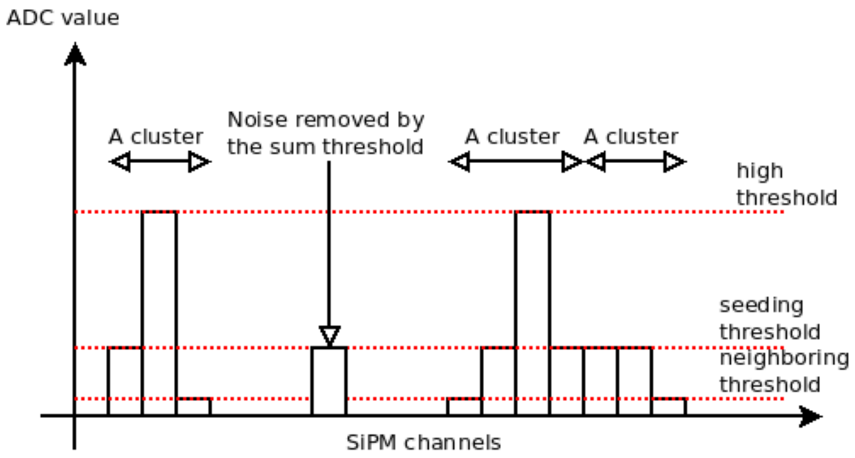
   + alignment
   + survey
   + change of operations conditions (warm)
   + …

These are the main points affecting efficiency and consequently tracking.

Data quality assessment will concern the first three points only.

To tag a run we look at **distribution of clusters.**

Clustering algorithm SciFi



Recipes

What can we change: thresholds and fine time alignment constants, which corresponds to single channel efficiency and whole front-end efficiency.

We have used **mainly two recipes in global data taking** which are:

1) **From 25/11/22** at around 03:48 new recipe with **latest time alignment** :PHYSICS_20221125

   → from beam timing scan by hand with steps here

2) Before: mainly PHYSICS_20221018 → from beam timing scan by hand with steps here

Thresholds:

pedestal are obtained from Cavern Threshold scan at – 40 C (2.5 pe, 3.5 pe and 4.5 pe), we have always used the same since October.

Different fine time alignment:

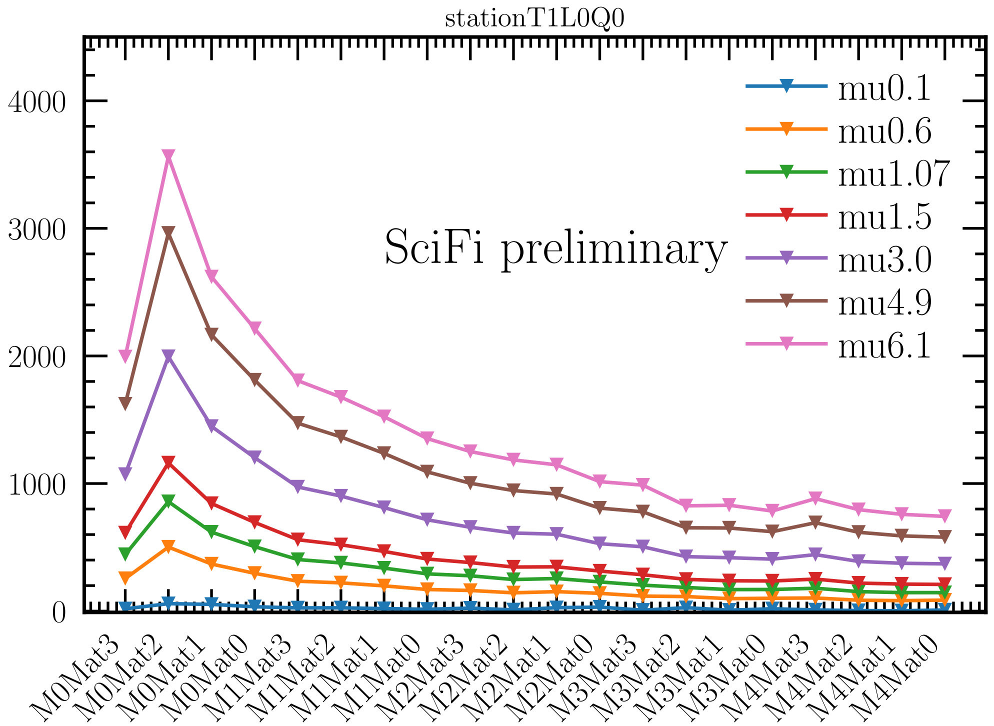
one could see differences in efficiencies at few percent level, we have 2 time alignments (one for each recipe above)

Our proxy: cluster profiles

- When you see holes in the distributions → we are not at maximum efficiency in that region.
- The holes can have different origins: tell40s in error, HalfROBs sending bad datas, bad time alignement.

→Here we discuss the consequences on the efficiencies.

For illustration purposes, a profile in *x*:



- This is a profile in the *x* direction obtained using SiPMs currents → no holes since there is no reconstruction involved

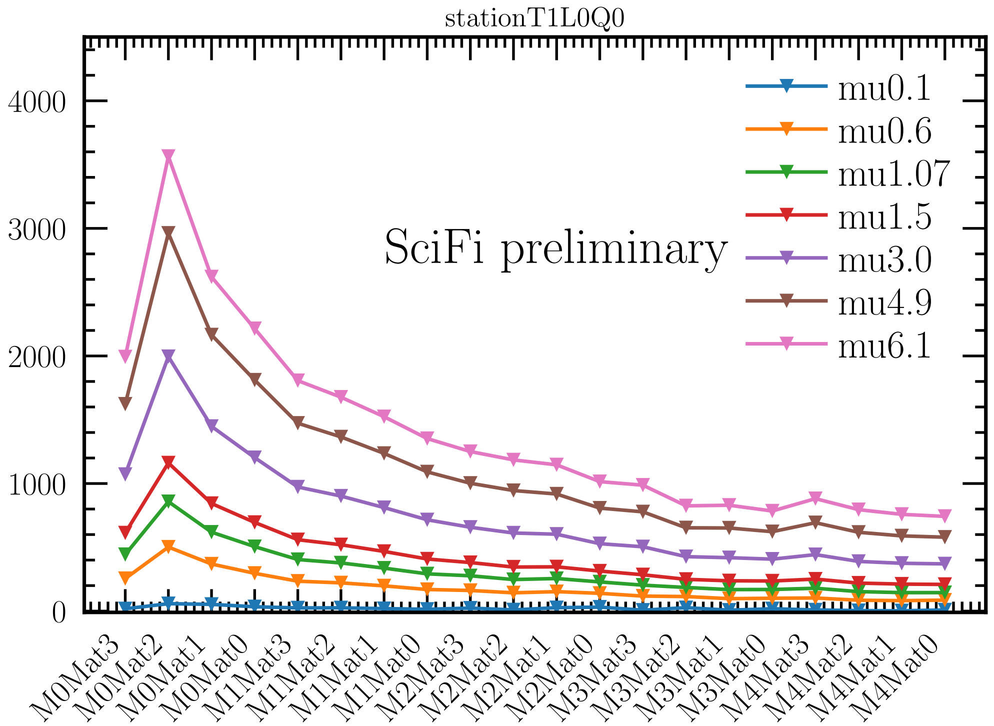
- The first mat of SciFi is cut (beam hole) → we expect lower number of clusters (current)

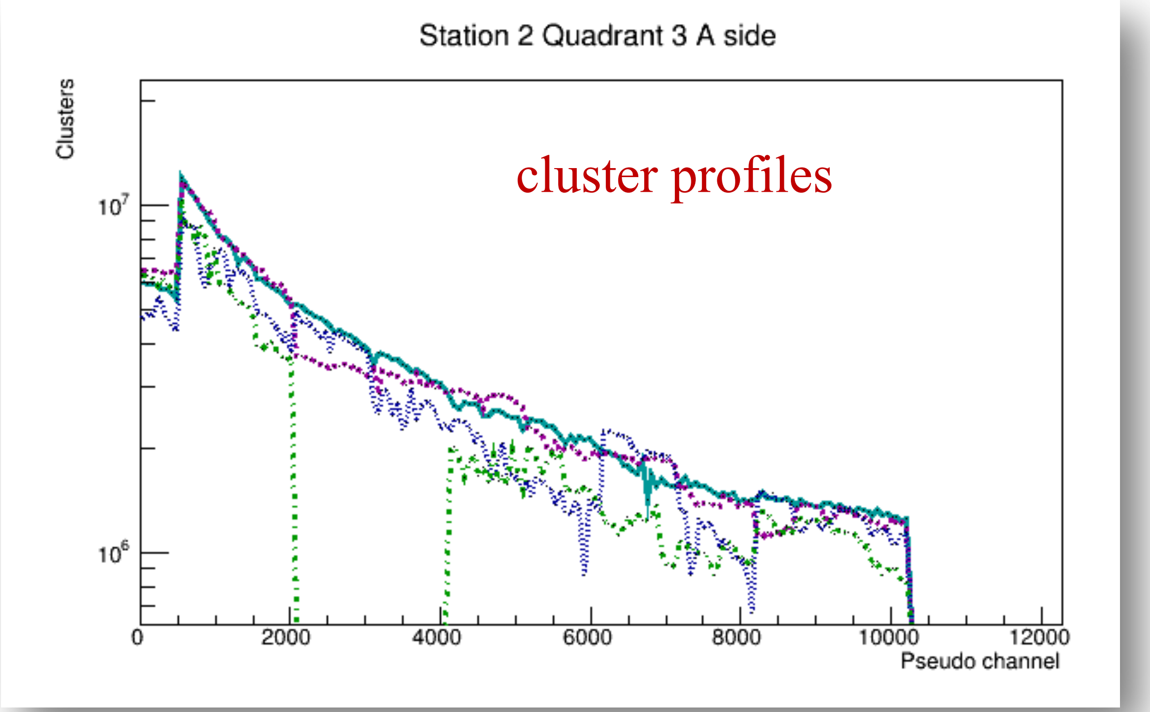
Our proxy: cluster profiles

- When you see holes in the distributions → we are not at maximum efficiency in that region.
- The holes can have different origins: tell40s in error, HalfROBs sending bad datas, bad time alignement.

→Here we discuss the consequences on the efficiencies.

For illustration purposes, a profile in *x*:

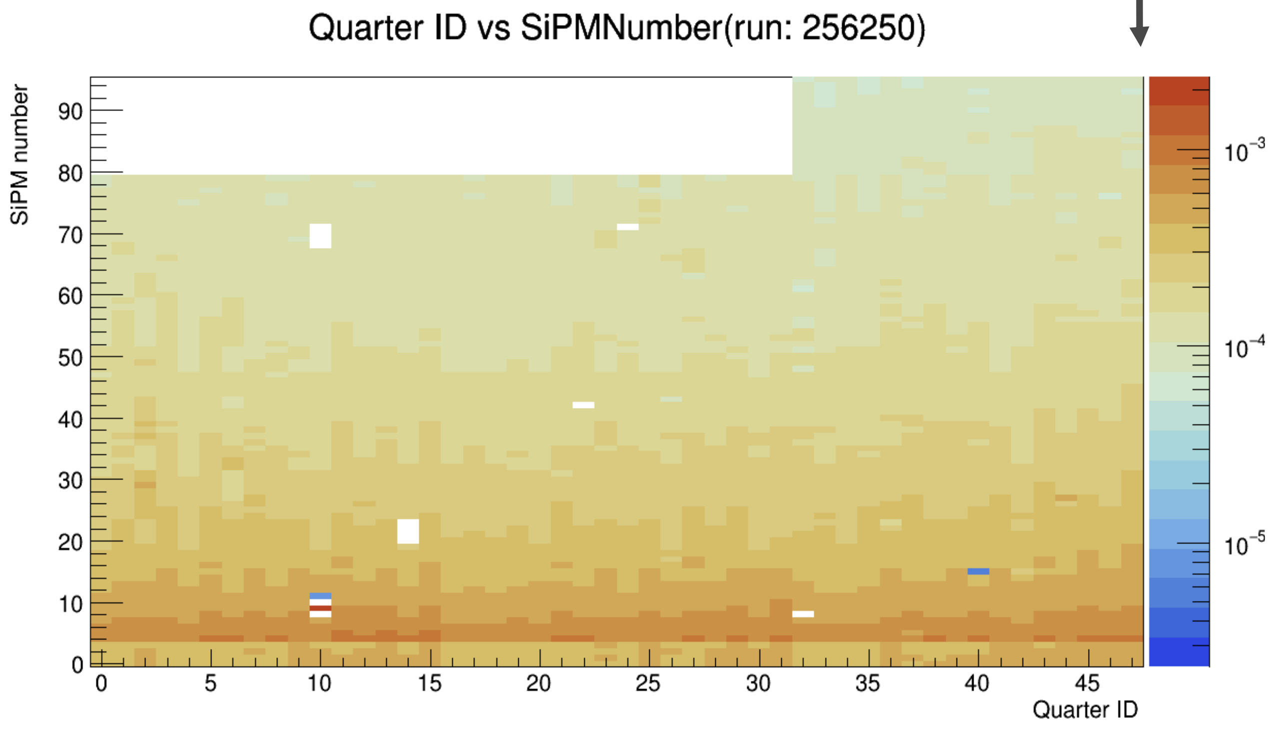




General Data Quality

Strategy to assess the data quality of a run B:

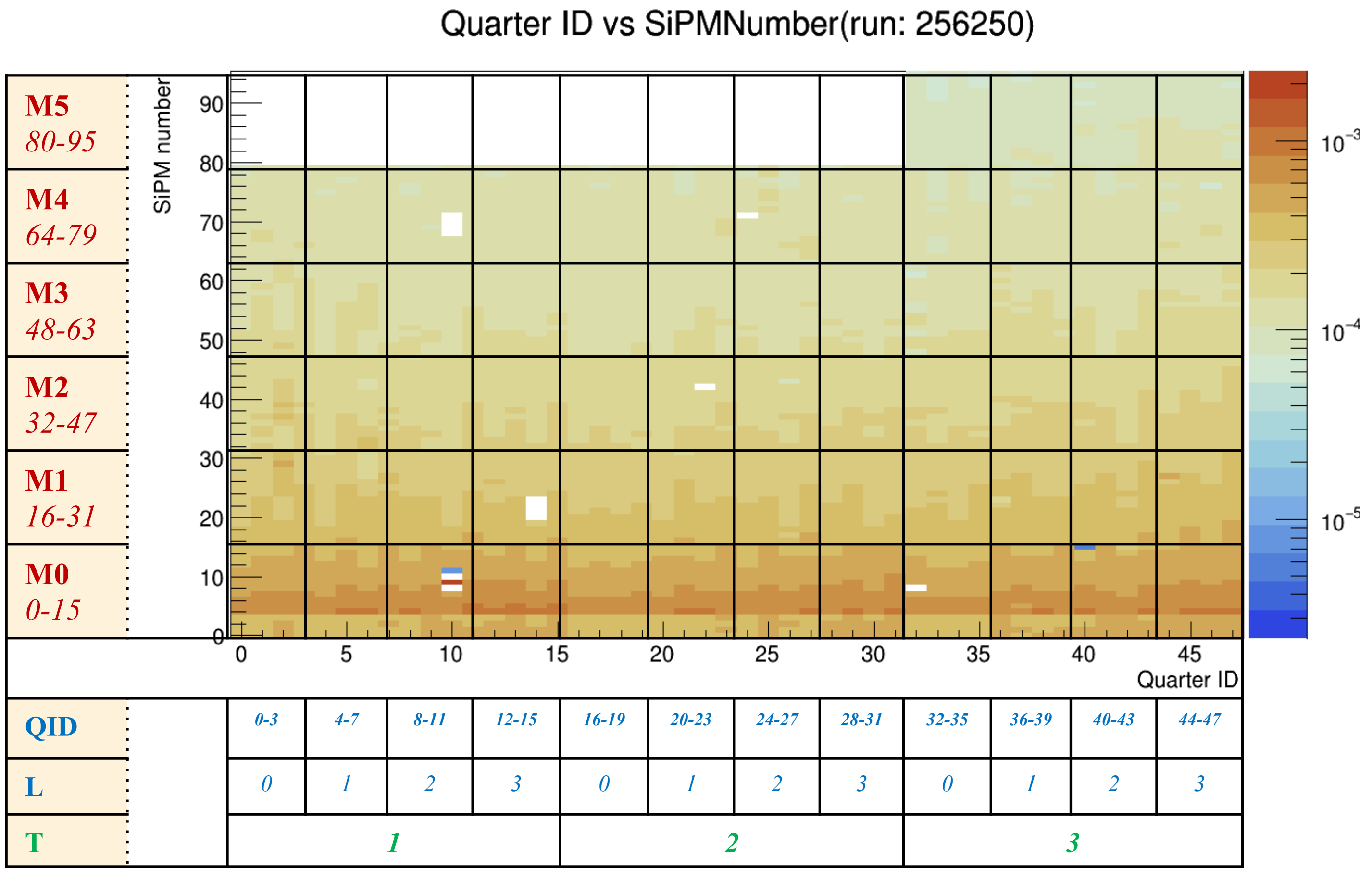
1. take a reference run A with reasonably uniform **2D clusters distribution**



General Data Quality

Strategy to assess the data quality of a run B:

1. take a reference run A with reasonably uniform **2D clusters distribution**



What is this plot?

Number of clusters seen by SciFi for SiPM Quarter ID.

- ❑ SiPM number:
  16 SiPM per module, max 6 modules 96 SiPMs
- ❑ Quarter ID:

  4 quarters per layer,
  4 layers per station per side
  → 16 quarter per station

General Data Quality

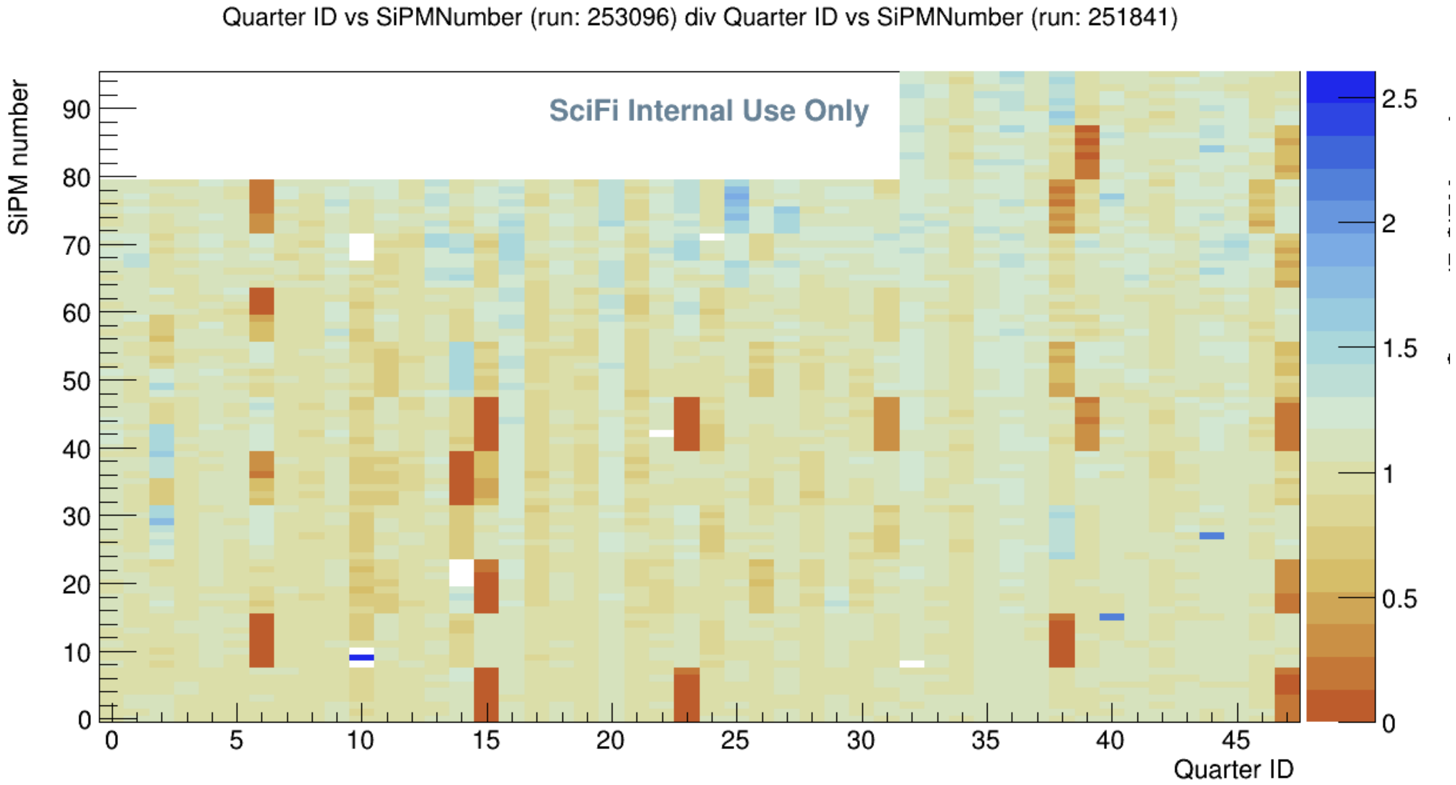
Strategy to assess the data quality of a run B:

1. take a reference run A with reasonably uniform **2D clusters distribution**
2. normalize 2D clusters distribution histogram for run A and run B

General Data Quality

Strategy to assess the data quality of a run B:

1. take a reference run A with reasonably uniform **2D clusters distribution**
2. normalize 2D clusters distribution histogram for run A and run B
3. take the ratio run B / run A



- Red region: groups of 8 SiPMs correlated to BXID jumps we experienced → loss of 100 % efficiency in this case

  See Blake's presentation at the SciFi general meeting for details

- White holes: broken fibers, these data links are always excluded

General Data Quality

Strategy to assess the data quality of a run B:

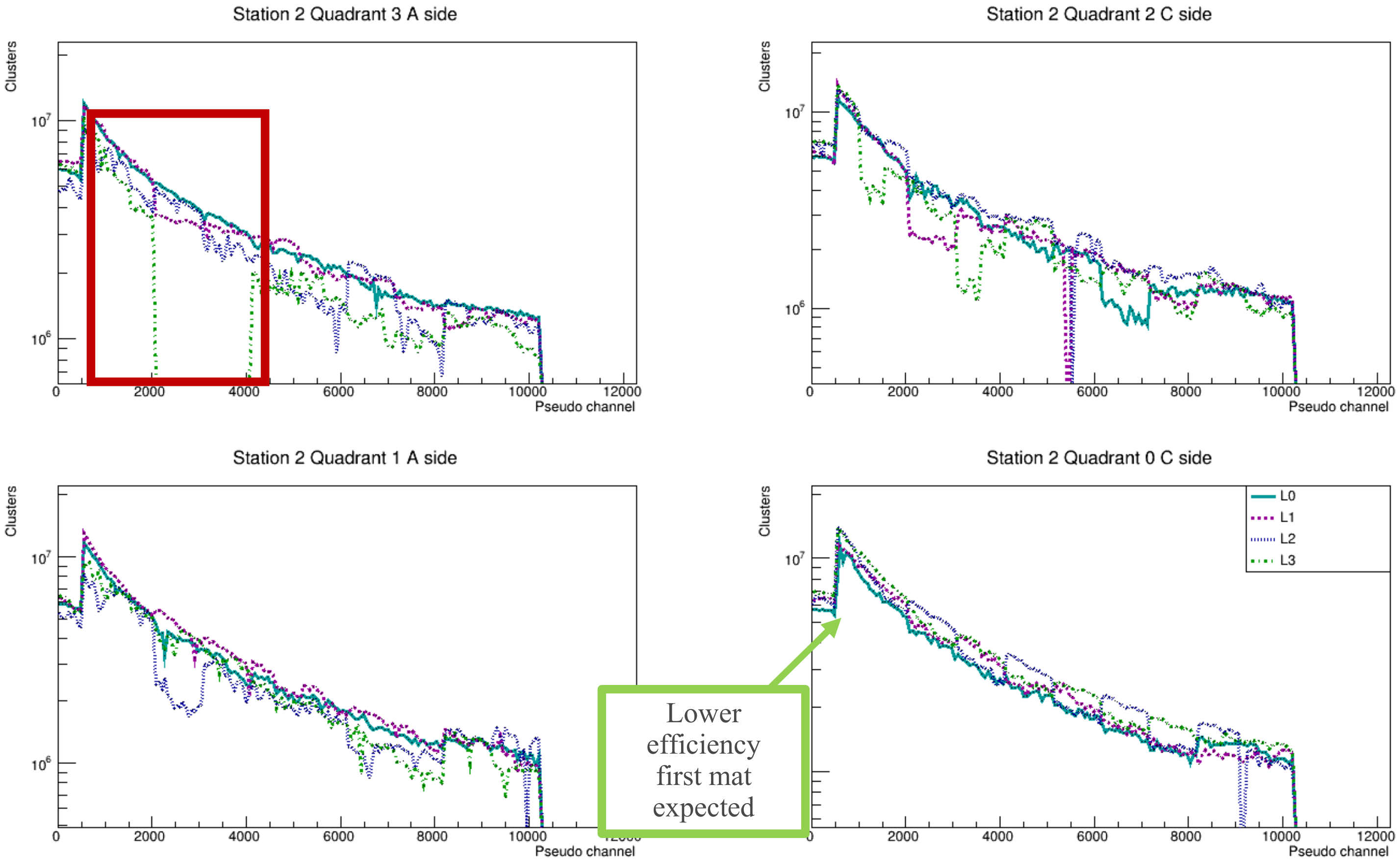
1. take a reference run A with reasonably uniform **2D clusters distribution**
2. normalize 2D clusters distribution histogram for run A and run B
3. take the ratio run B / run A
4. data quality tag:

Number of "clusters" in the ratio	**Global Data Quality Tag**
>= 90 %	HIGH QUALITY (GOOD)
>=80% <90 %	MEDIUM HIGH QUALITY
>=80% <70 %	MEDIUM LOW QUALITY
<70 %	LOW QUALITY

```
Data quality with respect to reference run 256250
Nb datalinks good (nb clusters >0.8 % wrt to reference) 86.74
Nb datalinks pass (nb clusters >0.65 % wrt to reference) =  0.54
Nb datalinks bad  (nb clusters <0.65 % wrt to reference) 1.28
Nb datalinks new zeros (missing links number) 2
Nb datalinks new zeros (missing links % over all SciFi) 0.04
Nb datalinks known_zeros (known zero clusters links) 11.39
Ratio with respect to reference run attached.
FINAL INFO:
Data quality tag:  MEDIUM HIGH QUALITY
The thresholds are [%]:  GOOD > 90, MEDIUM HIGH [90-80] , MEDIUM LOW [80-70] , BAD < 70 ).
```

Need one additional check if number of clusters with zero links is high. Message:
"NEED FURTHER CHECKING : number of links with zero clusters high, need human checking"

In addition: provide clusters profiles



Example for T2. Run 250798.

We provide a table for pseudochannel to geometrical position conversion.

- Hole in T2Q3L3 close to the region with high acceptance
- But, the other layers are reasonable
- Small clusters inefficiencies, mostly in outer regions far from the beam pipe

→Would expect decent tracking from this, but to be sure we need feedback from tracking !

This can be used **a posteriori** to help debugging tracking if SciFi related problems are found.

We will provide one plot per station for run of interest

Example of runs processed

Input: list of runs

```
255621
255949
255950
255956
255958
255964
255970
255971
255972
255974
255976
255977
255980
255982
255983
255998
255999
256000
256001
256003
```

1. (produce_root_files_from_list.sh)
2. process_dq_list_of_runs.sh
3. summarize_output_dq.py

```
Run , Tag
255621,MEDIUM LOW QUALITY
255983,BAD QUALITY
255998,NEED FURTHER CHECKING
255999,MEDIUM LOW QUALITY
256000,MEDIUM HIGH QUALITY
256001,MEDIUM HIGH QUALITY
256003,MEDIUM HIGH QUALITY
256004,MEDIUM HIGH QUALITY
256005,MEDIUM HIGH QUALITY
256007,MEDIUM LOW QUALITY
256008,MEDIUM HIGH QUALITY
256011,MEDIUM LOW QUALITY
256015,MEDIUM LOW QUALITY
256020,MEDIUM LOW QUALITY
256022,MEDIUM HIGH QUALITY
256026,MEDIUM HIGH QUALITY
256028,MEDIUM LOW QUALITY
256029,BAD QUALITY
256030,MEDIUM HIGH QUALITY
256032,MEDIUM HIGH QUALITY
255984,MEDIUM HIGH QUALITY
256105,MEDIUM HIGH QUALITY
256107,MEDIUM HIGH QUALITY
256110,MEDIUM HIGH QUALITY
256111,MEDIUM HIGH QUALITY
256112,MEDIUM HIGH QUALITY
256114,MEDIUM HIGH QUALITY
```

Conclusions

- ➢ Data quality validation procedure in place but still evolving
- ➢ Some extra general information can be found in the data quality txt summary (for instance: running warm, missing one quadrant…)
- ➢ We are digging into the type of inefficiency (tell40s, halRob etc…)
- ➢ The procedure is automated for the runs in **/hlt2/objects/LHCb/0000RunNumber**
  → for runs in **eos** need to compile the stack and produce root files using the grid
- ➢ We need feedback from RTA to see if our assessment is in line with the final tracking efficiency found.