Detector requirements for QA tools - TPC case

Outlook

Current RUN1 QA

Requirement for RUN2 and Run3

QA/calibration tools and ongoing activities

Elastic stack consideration

Murphy's law is an adage or epigram that is typically stated as:

Anything that can go wrong, will go wrong.

https://en.wikipedia.org/wiki/Murphy%27s_law

Douglas Adams ***http://www.quotationspage.com/quote/27059.html***

The major difference between a ***thing that might go wrong*** *and a* ***thing that cannot possibly go wrong*** *is that when a thing that cannot possibly go wrong goes wrong it usually turns out to be impossible to get at or repair.*

Quality assurance

https://en.wikipedia.org/wiki/Quality_assurance

Definition:

Quality assurance (QA) *is a way of preventing mistakes or defects in manufactured products and avoiding problems when delivering solutions or services to customers; ...*

This defect prevention in quality assurance differs subtly from defect detection and rejection in ***quality control****, and has been referred to as a shift left as it focuses on quality earlier in the process.[2]*

Two principles included in quality assurance are:

"*Fit for purpose*" *(the product should be suitable for the intended purpose);*

"*right first time*" *(mistakes should be eliminated).*

QA and release validation.
Run1

ALICE QA histogram

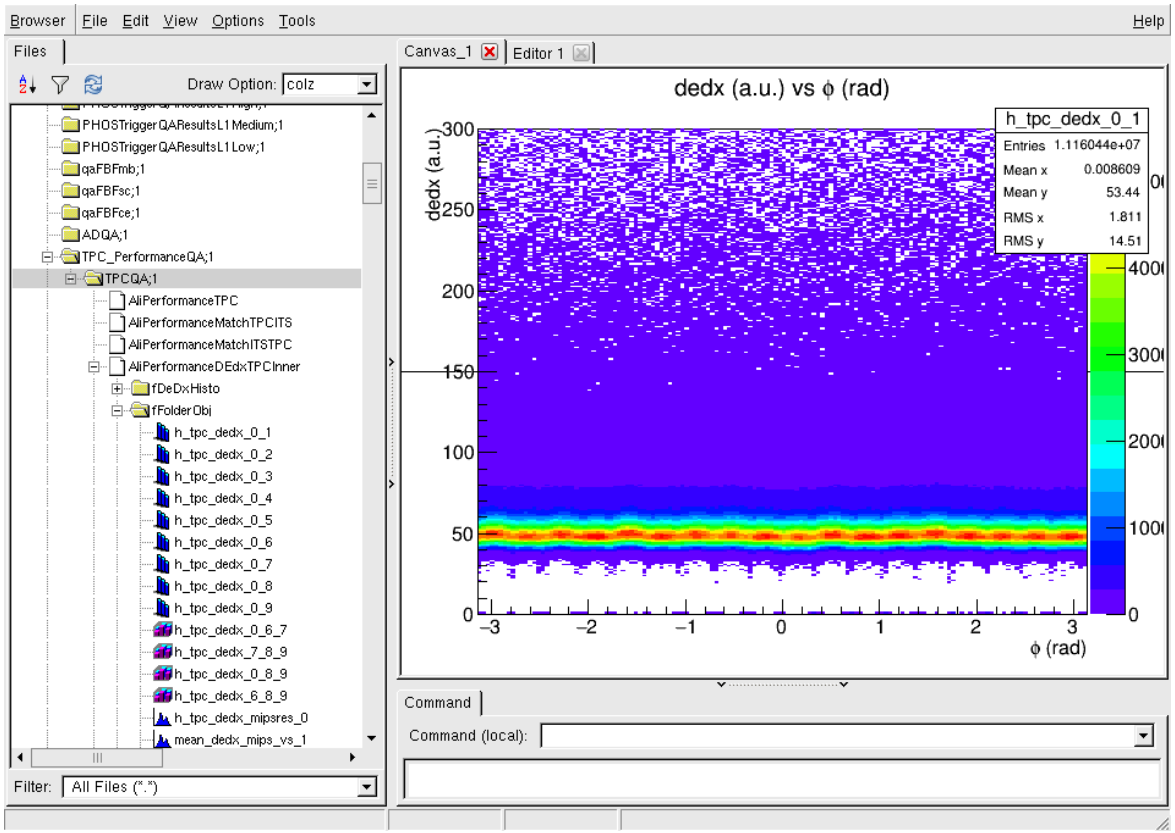
QA histograms and QA summary information stored in folder structure.

- Same naming convention and folder structure as **AliEn file catalog.**
- **No external DB.**

$PATH=$prefix/$datatype/$year/$period/$run/$recopass/$suffix

RAW QA histograms browsable from file catalog (e.g alien file catalog for central production)
O(150MBy) per run

```
TFile::Open("alien:///alice/data/2015/LHC15o/000244917/cpass1_pass2_lowIR/QAresults_barrel.root")
```

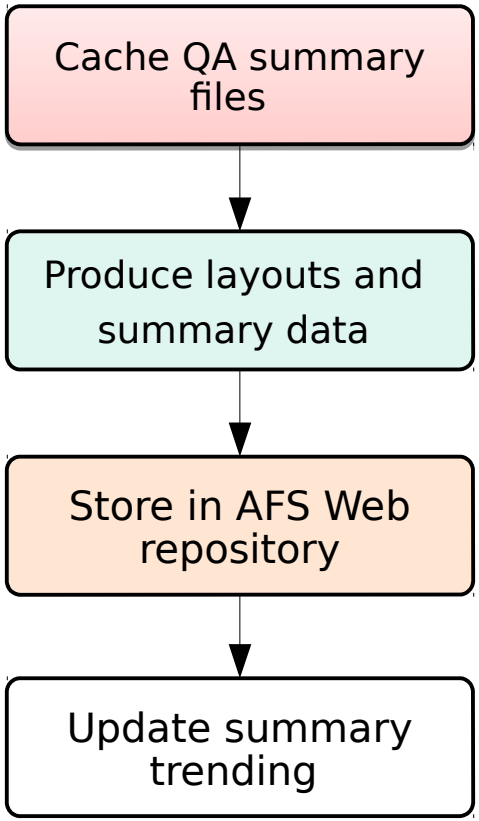


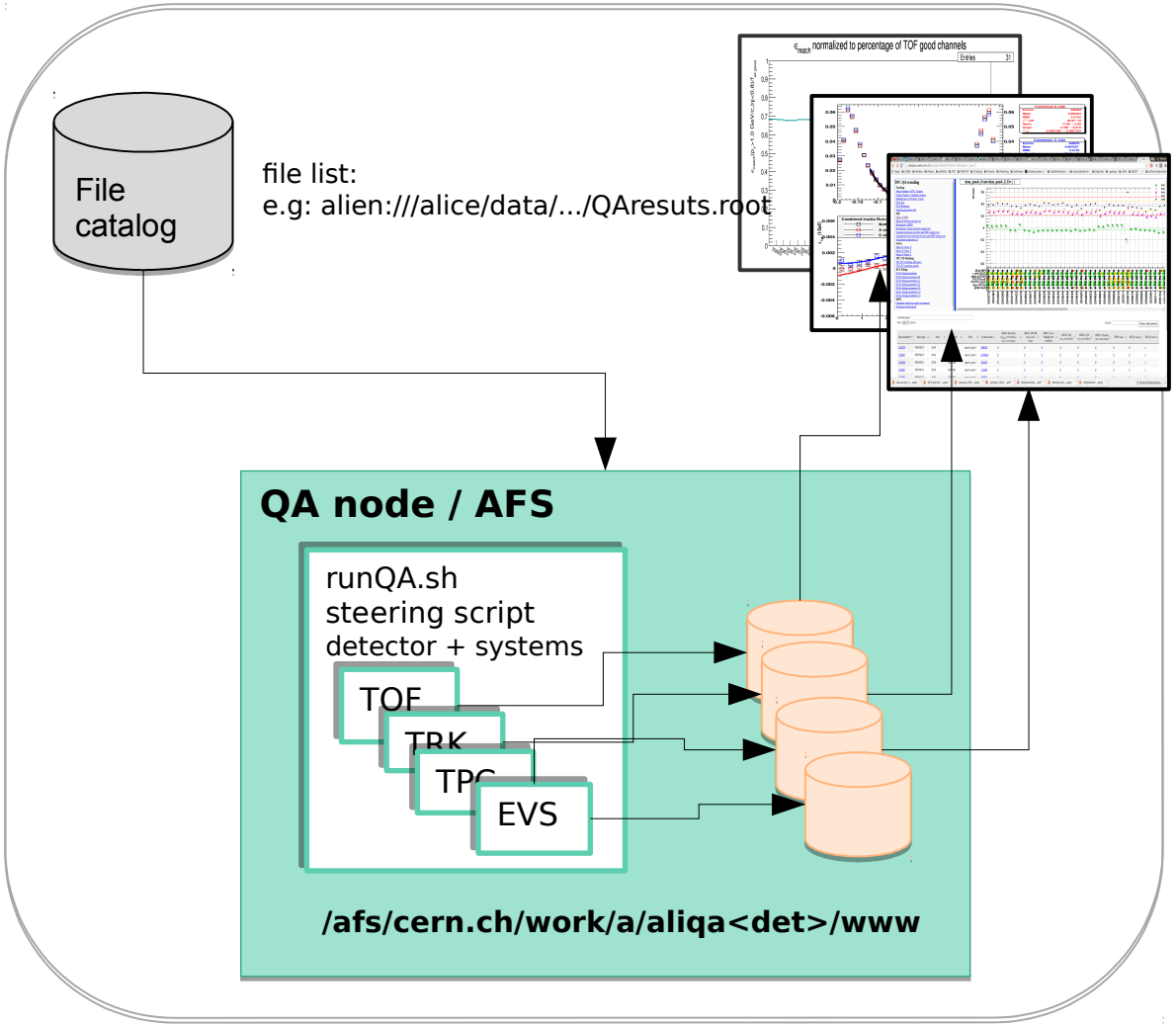
Example:
TPC dEdx vs ϕ(sector)

Automatic post processing QA data flow

Data layouts and high level summary data automatically processed - the same code used:

- **central production QA**
- **release** (CPass0/CPass1/QA) **validation**
- **user validation** (CPass0/CPass1/QA) - automatic QA checks and **residual calibration** check

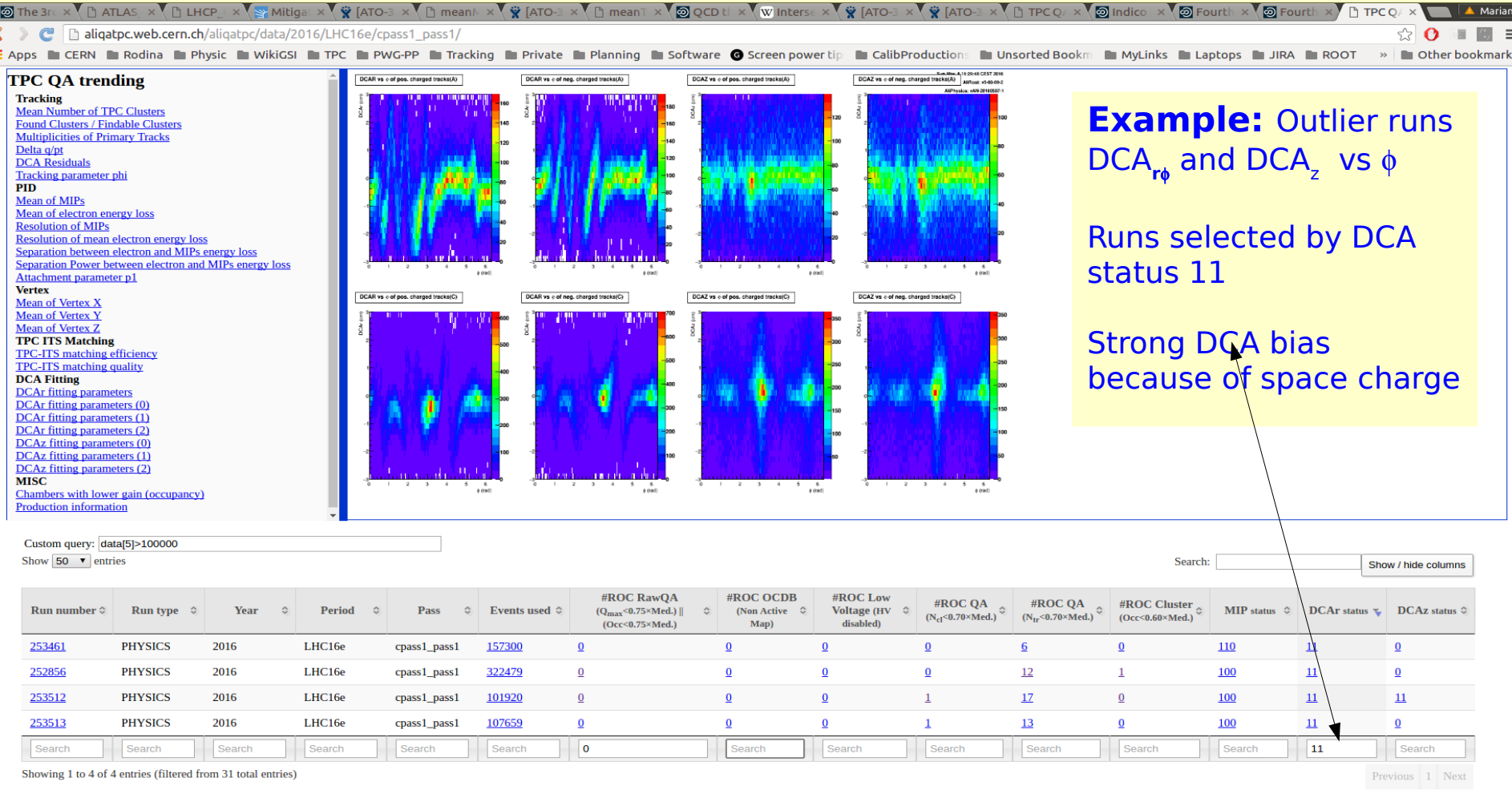




CPass1 - Run QA properties example

Interactive Outliers browsing example: http://aliqatpc.web.cern.ch/aliqatpc/data/2016/LHC16e/cpass1_pass1/

- Based on DataTables plug-in for the jQuery Javascript
  - Histograms organized into layouts to bundle related information
  - Status flag: defined by user defined logical expression (TTree query) using summary information (absolute bands, nσ bands, &&,||)



TPC run trending and alarms example

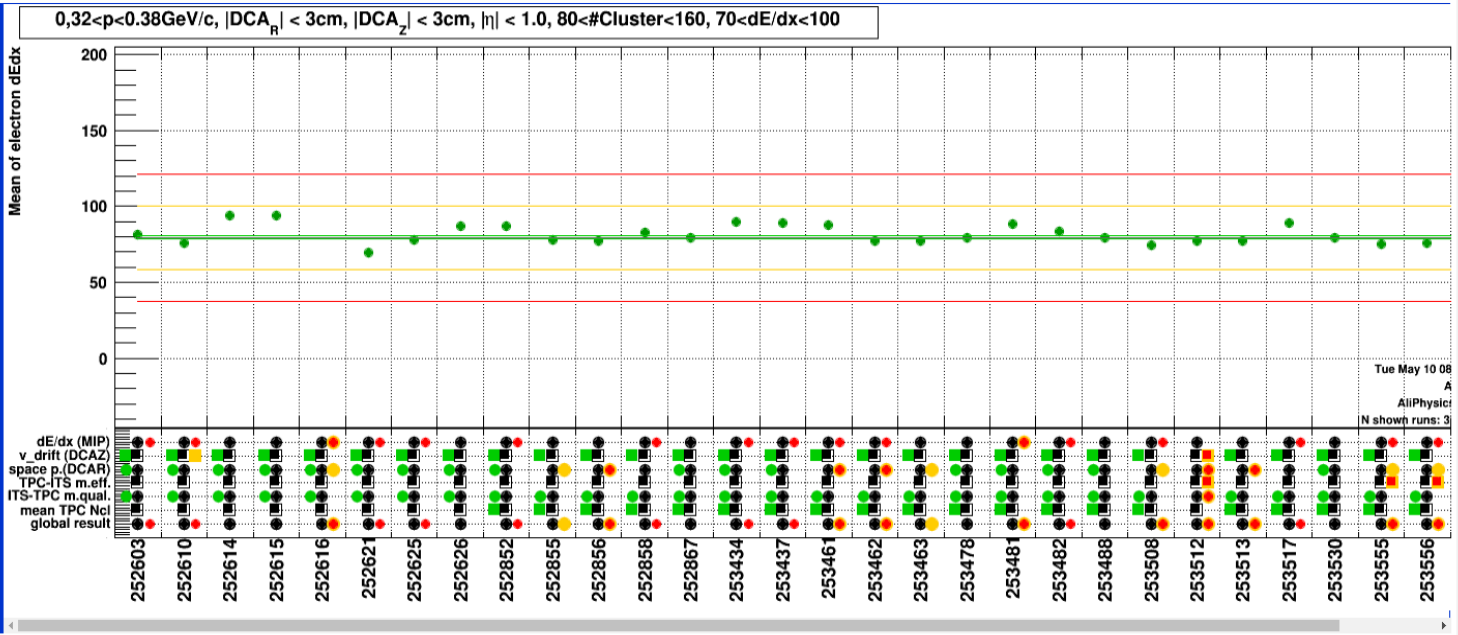
Run status browsing example:

- Based on DataTables plug-in for the jQuery Javascript
- Example: CPass1 calibration pass with failed <dEdx> calibration

aliqatpc.web.cern.ch/aliqatpc/data/2016/LHC16e/cpass1_pass1/

TPC QA trending

Tracking
Mean Number of TPC Clusters
Found Clusters / Findable Clusters
Multiplicities of Primary Tracks
Delta q/pt
DCA Residuals
Tracking parameter phi
PID
Mean of MIPs
Mean of electron energy loss
Resolution of MIPs
Resolution of mean electron energy loss
Separation between electron and MIPs energy loss
Separation Power between electron and MIPs energy loss
Attachment parameter p1
Vertex
Mean of Vertex X
Mean of Vertex Y
Mean of Vertex Z
TPC ITS Matching
TPC-ITS matching efficiency
TPC-ITS matching quality
DCA Fitting
DCAr fitting parameters
DCAr fitting parameters (0)
DCAr fitting parameters (1)
DCAr fitting parameters (2)
DCAz fitting parameters (0)
DCAz fitting parameters (1)
DCAz fitting parameters (2)
MISC
Chambers with lower gain (occupancy)
Production information



Custom query: data[5]>100000

Show 50 entries

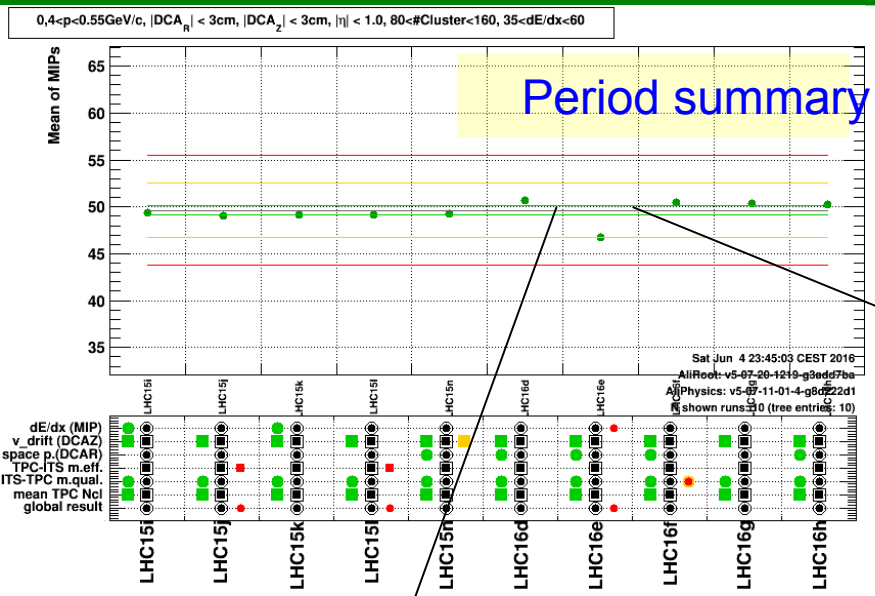
Search: Show / hide columns

Run number	Run type	Year	Period	Pass	Events used	#ROC RawQA (Q_{max}<0.75×Med.) || (Occ<0.75×Med.)	#ROC OCDB (Non Active Map)	#ROC Low Voltage (HV disabled)	#ROC QA (N_{cl}<0.70×Med.)	#ROC QA (N_{tr}<0.70×Med.)	#ROC Cluster (Occ<0.60×Med.)	MIP status	DCAr status	DCAz status
253461	PHYSICS	2016	LHC16e	cpass1_pass1	157300	0	0	0	0	6	0	110	11	0
252856	PHYSICS	2016	LHC16e	cpass1_pass1	322479	0	0	0	0	12	1	100	11	0
253512	PHYSICS	2016	LHC16e	cpass1_pass1	101920	0	0	0	1	17	0	100	11	11
253513	PHYSICS	2016	LHC16e	cpass1_pass1	107659	0	0	0	1	13	0	100	11	0
Search	Search	Search	Search	Search	Search	0	Search	Search	Search	Search	Search	Search	11	Search

Showing 1 to 4 of 4 entries (filtered from 31 total entries)

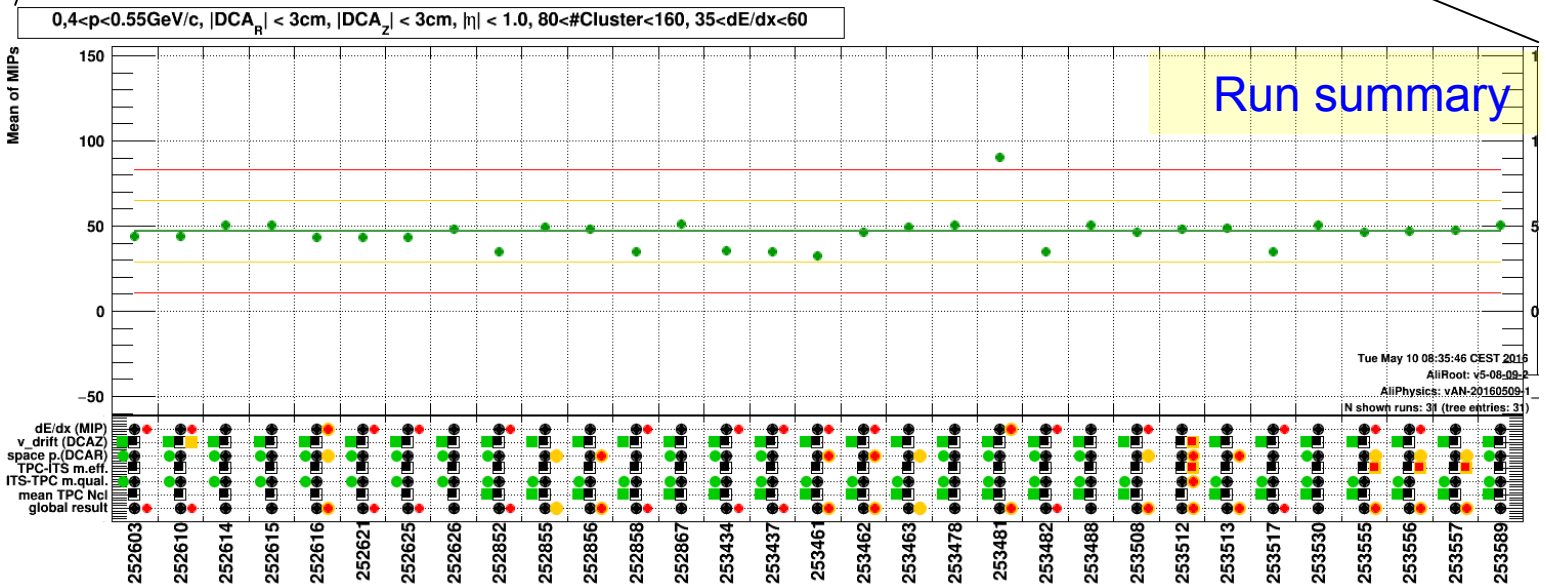
Previous 1 Next

Run and period trending of summary data



Automatic trending of high level summary data

- x axis regression - period, pass, AliRoot
- Example status bar (period with failed
- dEdx calibration):
  - dEdx
  - vdrift
  - DCA (distortion)
- matching efficiency



What can go wrong? TStatToolkit: Status and alarms

Supported alarms (defined as queries - TTreeFormulas)

- Warnig/Outlier/PhsAcc in respect to expectation
  - fixed range
  - absolute/relative
- Warnig/Outlier/PhysAcc in respect to reference
  - Period statistic (Mean, Median, RobustMean, LTS)
- Warnig/Outlier/PhysAcc in respect to model
  - e.g resolution as function of the IR
  - access to external information sources essential

Ongoing activities

- Warnig/Outlier/PhysAcc in respect to Anchor production
  - see ***Demonstration of the TPC QA tools***
- Time series Warnig/Outlier/PhysAcc in respect to moving robust estimators

What can go wrong? dEdx alarm example

Any calibration can fail. Logical or of outliers/warning/PhysAcc

- QA.TPC.trending.meanMIP_Outlier
- QA.TPC.trending.resolutionMIP_Outlier
- QA.TPC.trending.MIPattachSlopeA_Outlier
- QA.TPC.trending.MIPattachSlopeC_Outlier
- QA.TPC.trending.meanMIPele_Outlier
- QA.TPC.trending.resolutionMIPele_Outlier
- QA.TPC.trending.electroMIPSeparation_Outlier
- QA.TPC.trending.MIPattachSlope_comb2_Outlier
- QA.TPC.trending.MIPquality_Outlier
- QA.TPC.trending.PIDSepPow_comb2_Outlier
- **Try: AliTreePlayer::printSelectedTreeInfo(tree,"alias","QA.TPC","MIP.*Outlier$",1);**

Calibration parameters changing in time with O(min) granularity

“Fighting chambers” with lower HV O (s)

Calibration per sector

Simulation anchoring:

- OCDB parameters mismatch …
- BetheBloch
- Gain

RUN2 and Run3

Run2: Distortion monitoring. Data sources under preparation (0)

Time series DB based on root trees under preparation

- distortion fluctuating O(0.05 s)
- IR and background fluctuating
- Several detectors TPC + Reference (ITS,TRD,TOF)
- Different sources, time granularity
    - IMPORTANT aspect, Software should enable cross queries

Tracking QA parameters into DB:

- TPC DCA bias/resolution/pulls/outliers phi map (18-36 bins) A/C side 2-3 pt bins
    - Time granularity **O(min)**
- TPC DCA bias/resolution/pulls/outliers integral
    - Time granularity **O(min)**
- Constrain/Combined angular resolution
    - Time granularity **<1 min**

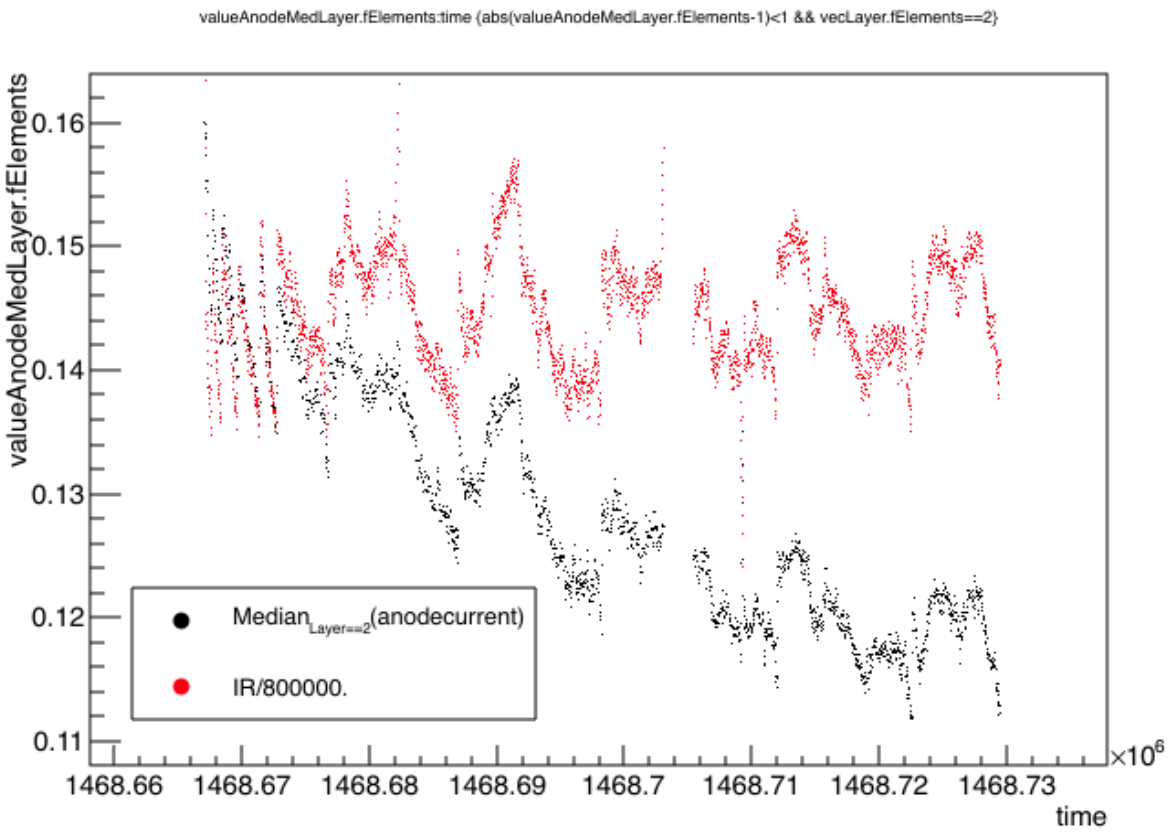
Distortion parameters into DB

- Time granularity per map **O(20-40 min)**
    - Mean distortion per hotspot
    - RMS fluctuation per hotspot

Space charge estimators into DB

- TRD currents **O(s)**
- IR trigger+background monitoring **O(s)**

TRD current to monitor TPC space charge

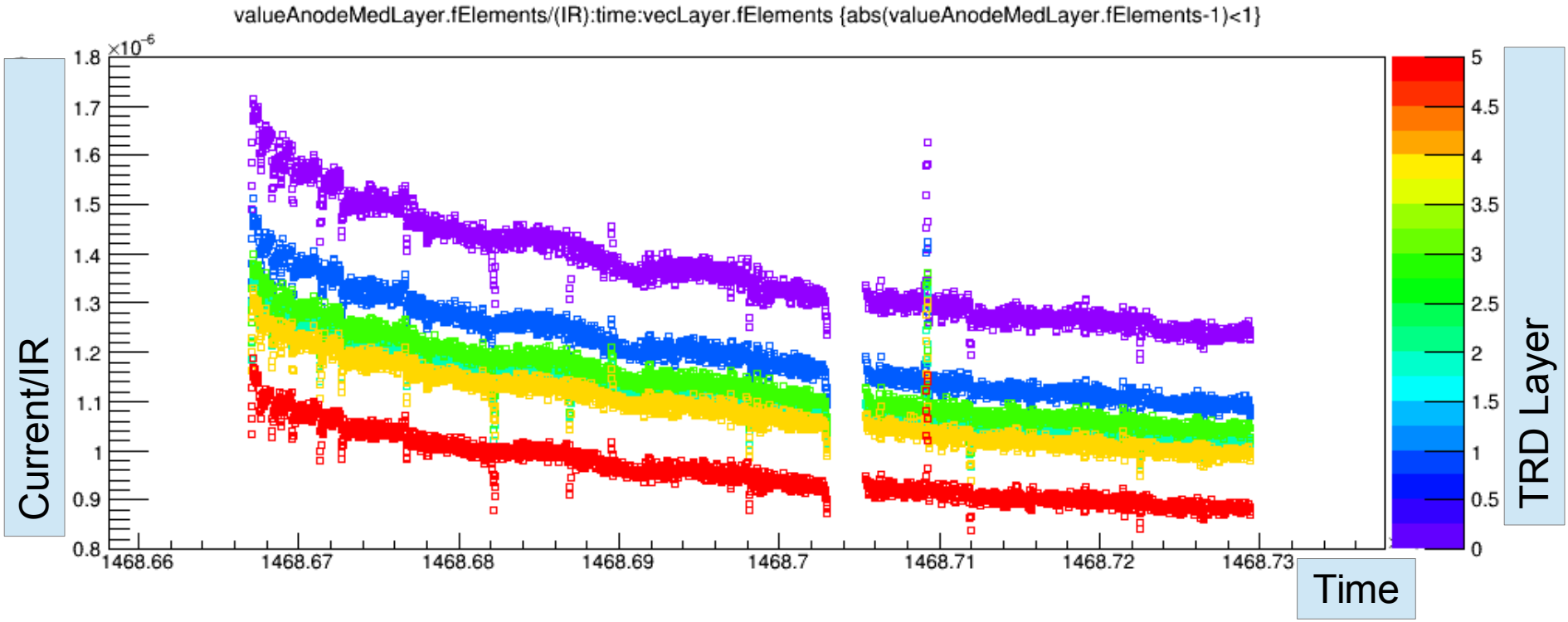


TPC distortion corrected using the IR estimator

Better estimator to be obtained using TRD currents

- TRD currents ~ signal + beam background
  - outer ALICE background monitors depends on fiducial volume
  - TPC currents not usable to estimates space charge (gating grid), current depends on triggering schema
- How to interpolate space charge estimator (TRD current) to the TPC volume ?

TRD current to monitor TPC space charge (Ernst, MI)

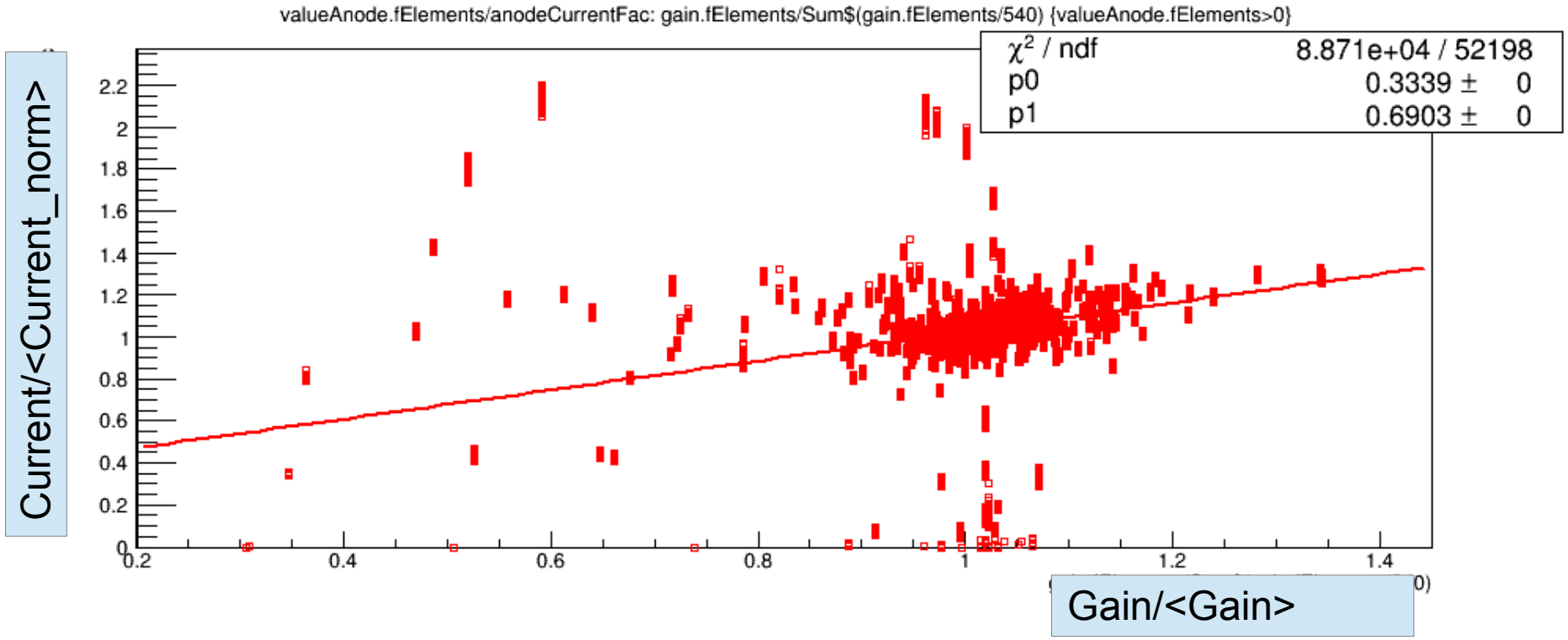


TPC distortion corrected using the IR estimator

Better estimator to be obtained using TRD currents

- Current normalized to IR from scalers in individual layers (color decreasing as function of R)
- decrease as expected - background is decreasing IR is leveled
- Steps at region with high gradient of IR 30 %
  - Time intervals to be removed from analysis? - time series analysis needed to decide

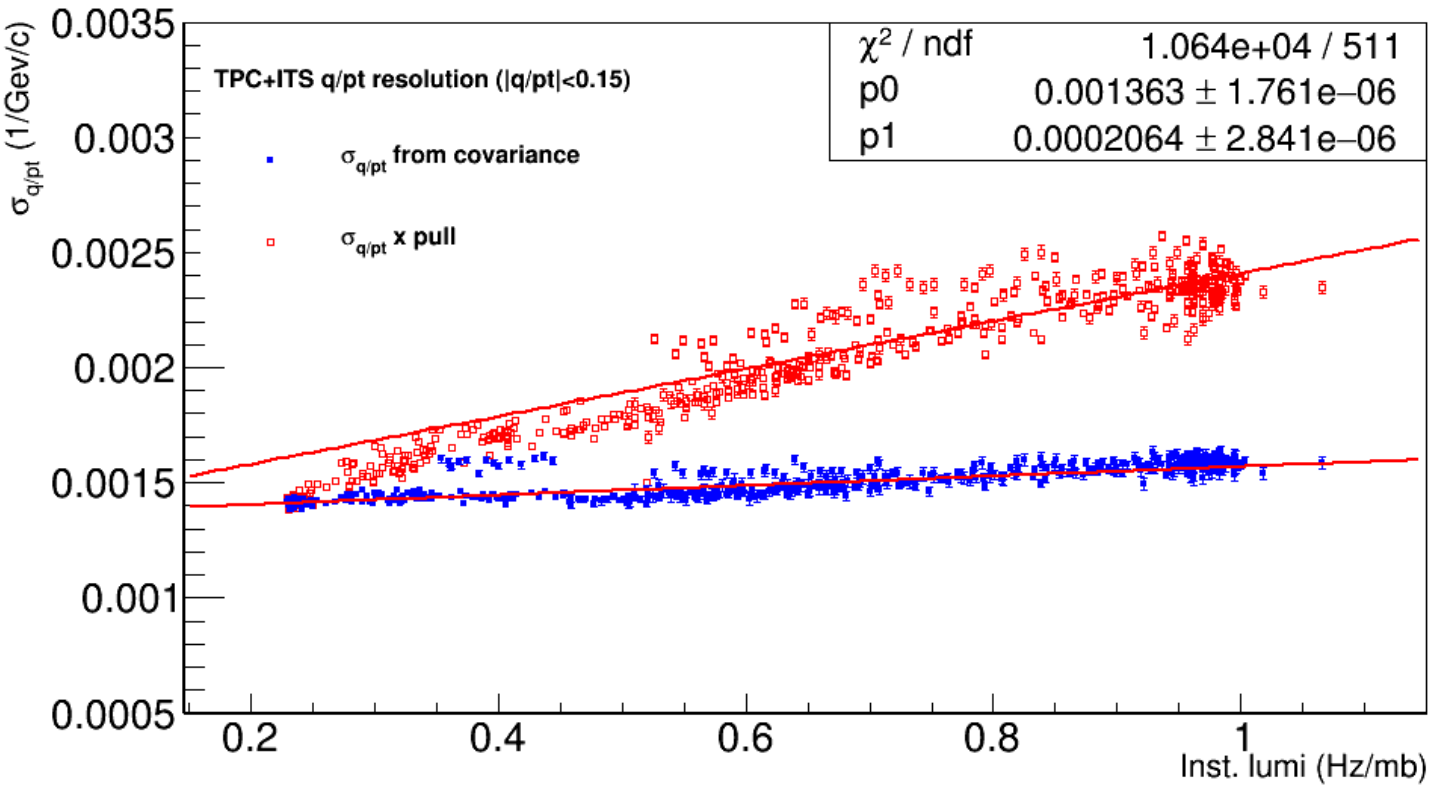
TRD current to monitor TPC space charge (Ernst, MI)



TRD current calibration and outlier removal:

- Current normalized to factorized as function of normalized gain
- Inorm(layer,stack,sector)=Ilayer(layer)*Istack(stack)*Isector(sector)

Run2 Monitoring: Angular pulls for constrained track



Is covariance matrix error estimator correct ? NO

angular pulls for TPC constrained tracks at high pt deviates from one (point every 5 minutes)

- Interpretation work in progress (analytical formula to confirm)
- To come - time series - pulls/resolution/bias at individual “hot regions”

Run3: Distortion monitoring. Data sources

Goal - fully automatic QA

- O(10^4 -10^5) data-points per 0.01 s to be monitored, correlated, calibrated

3D Distortion maps (~180x20x20)

- O (0.01 s) following fluctuations scenario
- O (minutes) mean correction scenario

3D digital currents (~180x20x20)

- O(0.01 s) for distortion fluctuation monitoring/correction
- analog currents also to be read but with worse space and time granularity

QA residual monitoring

- O(0.01 s) ?

Ongoing activities

QA generalization and time series

[JIRA] (ATO-46) Provide access to external info.

[JIRA] (ATO-382) Time series query support for tree using AliTreePlayer

[JIRA] (PWGPP-163) Acceptance cut based on the custom parameterization using AliNDLocalRegression

[JIRA] (ATO-373) TTree -> html table, csv, json, xml (SELECT FROM WHERE ORDER BY) + Exporting metadata

[JIRA] (PWGPP-163) Acceptance cut based on the custom parameterization using AliNDLocalRegression

[JIRA] (ATO-360) Generalization of the TPC detector QA trending (run, period,time) for other detectors and actions (tracking, distortion calibration)

[JIRA] (ATO-361) Generalization of the TPC QA web page for other detectors/calibration/monitoring (Code based on the Datatables)

[JIRA] (ATO-372) Elastic search and Kibana - investigation

Time series examples

[JIRA] (ATO-348) Hardware current studies and usage for space charge correction - TPC and TRD

[JIRA] (PWGPP-221) Performance plots to benchmark new Space point distortion calibration

[JIRA] (PWGPP-209) QA of LHC15o pass1 High IR runs

QA development (0)

[JIRA] (ATO-46) Provide access to external info. (Jens, MI, Carsten)

- https://alice.its.cern.ch/jira/browse/ATO-46
- Standartized cross queries (e.g QA.TPC,Lobook, QA.TRD, QA.EVS) used in user queries
- Inner joins implemented (https://en.wikipedia.org/wiki/Join_(SQL)
- Left and right join to be added ?
    - logical or in case of sparse data
    - E.g more runs in RCT than in the QA.TPC

[JIRA] (ATO-373) TTree -> html table, csv, json, xml (SELECT FROM WHERE ORDER BY) + Exporting metadata (MI + Hans)

- https://alice.its.cern.ch/jira/browse/ATO-373
    - Elastic json
    - JQuery, Datatables
- first version committed - standardization of tree → html support
- meta-data support
    - filtering of the columns (O(100s)) necessary
    - automatic annotations
- Support for time series

QA development (1)

[JIRA] (ATO-382) Time series query support for tree using AliTreePlayer (Hans Beck, MI)

- https://alice.its.cern.ch/jira/browse/ATO-382
- Tree as a time series DB with standard queries
- To simplify, generalize time queries in QA and calibration
  - e.g outliers and steps detection

[JIRA] (PWGPP-163) Acceptance cut based on the custom parameterization using AliNDLocalRegression (Philipp Luetig???, MI)

- https://alice.its.cern.ch/jira/browse/PWGPP-163
- Development staled - to be restarted

For some regions with worse performance in space and time to be removed in Analysis

Time series support (1)

[JIRA] (ATO-382) Time series query support for tree using AliTreePlayer

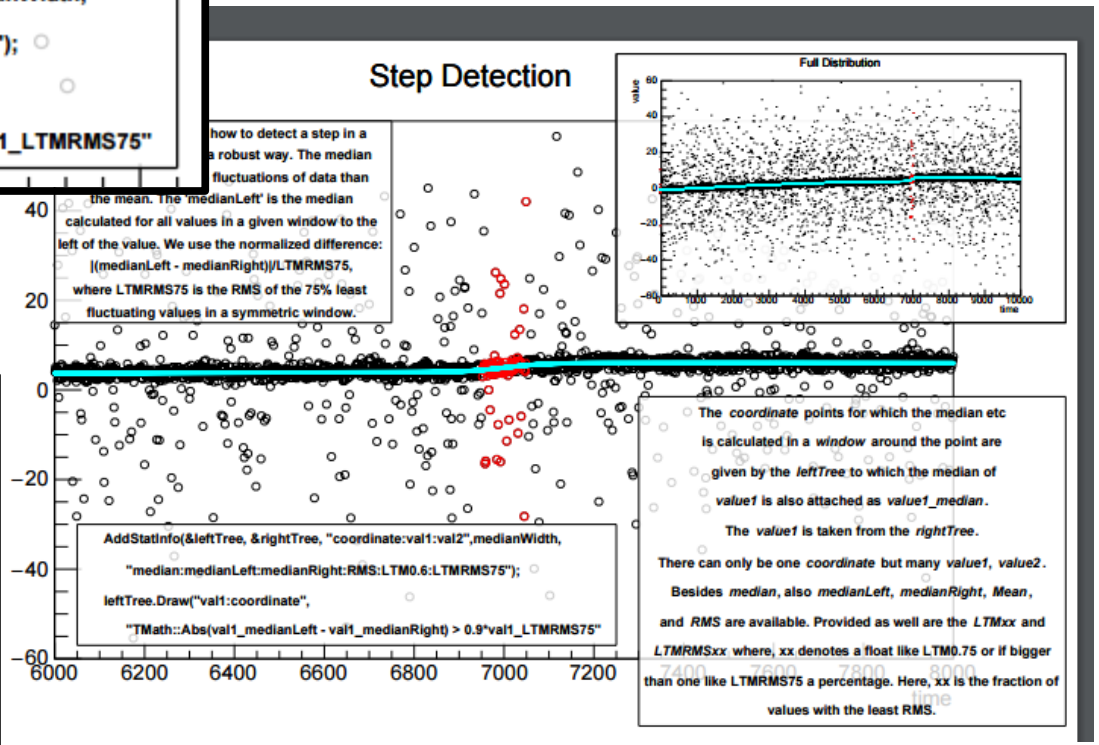
- Working with simple data structures primitives (TLeaf) and array of primitives

•Example local moving statistic from unit test (in TPC git)

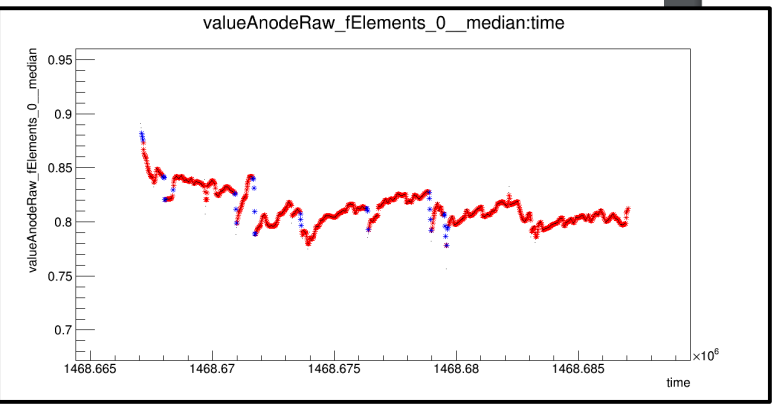
Step detection example using median left and right estimators:

```
AddStatInfo(&leftTree, &rightTree, "coordinate:val1:val2",medianWidth,
    "median:medianLeft:medianRight:RMS:LTM0.6:LTMRMS75");
leftTree.Draw("val1:coordinate",
    "TMath::Abs(val1_medianLeft - val1_medianRight) > 0.9*val1_LTMRMS75"
```

Unit test simulation



TRD current real live example



QA development (2)

[JIRA] (ATO-360) Generalization of the TPC detector QA trending (run, period,time) for other detectors and actions (tracking, distortion calibration)

- https://alice.its.cern.ch/jira/browse/ATO-360
- part of TPC Jupyter Tutorial today 14:20

[JIRA] (ATO-361) Generalization of the TPC QA web page for other detectors/calibration/monitoring (Code based on the Datatables)

- https://alice.its.cern.ch/jira/browse/ATO-361

[JIRA] (ATO-372) Elastic search and Kibana - investigation

- https://alice.its.cern.ch/jira/browse/ATO-372

QA development (3)

[JIRA] (ATO-367) OCDB toolkit make tables (html and JIRAs)

- https://alice.its.cern.ch/jira/browse/ATO-367

- [JIRA] (ATO-15) Adding production information and OCDB information snapshot to the QA script. OCDB.Print . OCDB.Dump
  - https://alice.its.cern.ch/jira/browse/ATO-15

Elastic stack consideration

Current version of the Elastic and Kibana not suitable for the QA and calibration monitoring

- 3rd party products currently needed
- More detailed discussion related to Elastic in following talks
  - 14:00 QA tools overview
  - 14:40 Elasticsearch+Kibana/Kibi as QA tools

In table bellow I show functionality which is currently missing in order to provde a qa in forms like we provide now in the TPC QA web page and soon in other QA web pages.
http://aliqatpc.web.cern.ch/aliqatpc/data/2015/LHC15o/pass1/

See also summary presentation from QA meeting - June 2016 QA tools : https://indico.cern.ch/event/545248/timetable/

Item		status	git request
versioning	+		
joins	-	***A single document should contain all of the information that is required to decide whether it matches a search request.***	see joins and performance discussion in source 3-party product - based on Elastic - crate see crate web page emulate SQL functionality, but require additional layer. Unfortuantalley Kibana does not support crate ...
tabular metadata	-	not supported	not found closest git request in https://github.com/elastic/kibana/issues/791
custom axis description	-	not supported	request exist (in git since 1.5 year) https://github.com/elastic/kibana/issues/2386
html preview	-	not supported	request exist but not accepted (link to add here)
query formula	-	too complex in comparison with SQL or Root TFormula - see discussion	not found request for simplification -> third product ***crate*** can be used to avoid syntax and also to iplement joins https://crate.io/a/how-is-crate-data-different-than-elasticsearch/
data volume	-		22 bytes per point representation (sensors reading benchmark example) vs 1-2 bytes in alternative Time series DB - see comment
data volume indexing	-	big decirepacny in data volume also for the fscraclawler file indexing example comparing to simple ncdu or du	see discussion in fscrawler JIRA ticket https://alice.its.cern.ch/jira/browse/ATO-375 . E.g example queries[https://alice.its.cern.ch/jira/browse/ATO-375?focusedCommentId=180291&page=com.atlassian.jira.plugin.system.issuetabpanels:comment-tabpanel#comment-180291]

Slide 15

Elastic Kibana/ Plans

We will prepare within TPC and DPG use cases investigated (metrics and queries) for current data

- QA/Calibration/Production/Performance (run properties) + new time calibration+QA series
- several independent sources of information with different time granularities (period,fill, run, time (O(1hour),O(min), O(s), O (~10 ms)?)

Expected use cases for RUN3

- Convert our data from root format to json format and populate cluster proposed by (Pablo)
- Exercise queries (within our ALICE DPG - Data preparation group)

Solve missing relation problem - which technique to use ?

- https://www.elastic.co/guide/en/elasticsearch/guide/current/relations.html
- Parent-child ? https://www.elastic.co/guide/en/elasticsearch/guide/current/parent-child.html
- Data denormalization ?
- 3rd party

Conclusion

TPC complex detector with many degrees of freedom

All aspects of operation to be automatically monitored, QAed

- **"*Fit for purpose*"** *(the product should be suitable for the intended purpose);*
- **"*right first time*"** *(mistakes should be eliminated).*

RUN2 and RUN3 enormous increase in number of parameters to monitor

- *Run 3 → fully automated calibration and QA essential (50 kHz Pb-Pb, no second chance, must be right the first time)*
  - *—> Run2 2 is testbed for all of that*
- *Use cases to be exercised*
- *Proper tools to be chosen*

QA tools tutorial

- *Demonstration of the TPC QA tools 14:20*