Automatic DQ checks

Steffen Schaepe
for the TRT DQ team

Physikalisches Institut
Universität Bonn

PHYSIKALISCHES
INSTITUT

June 22, 2010

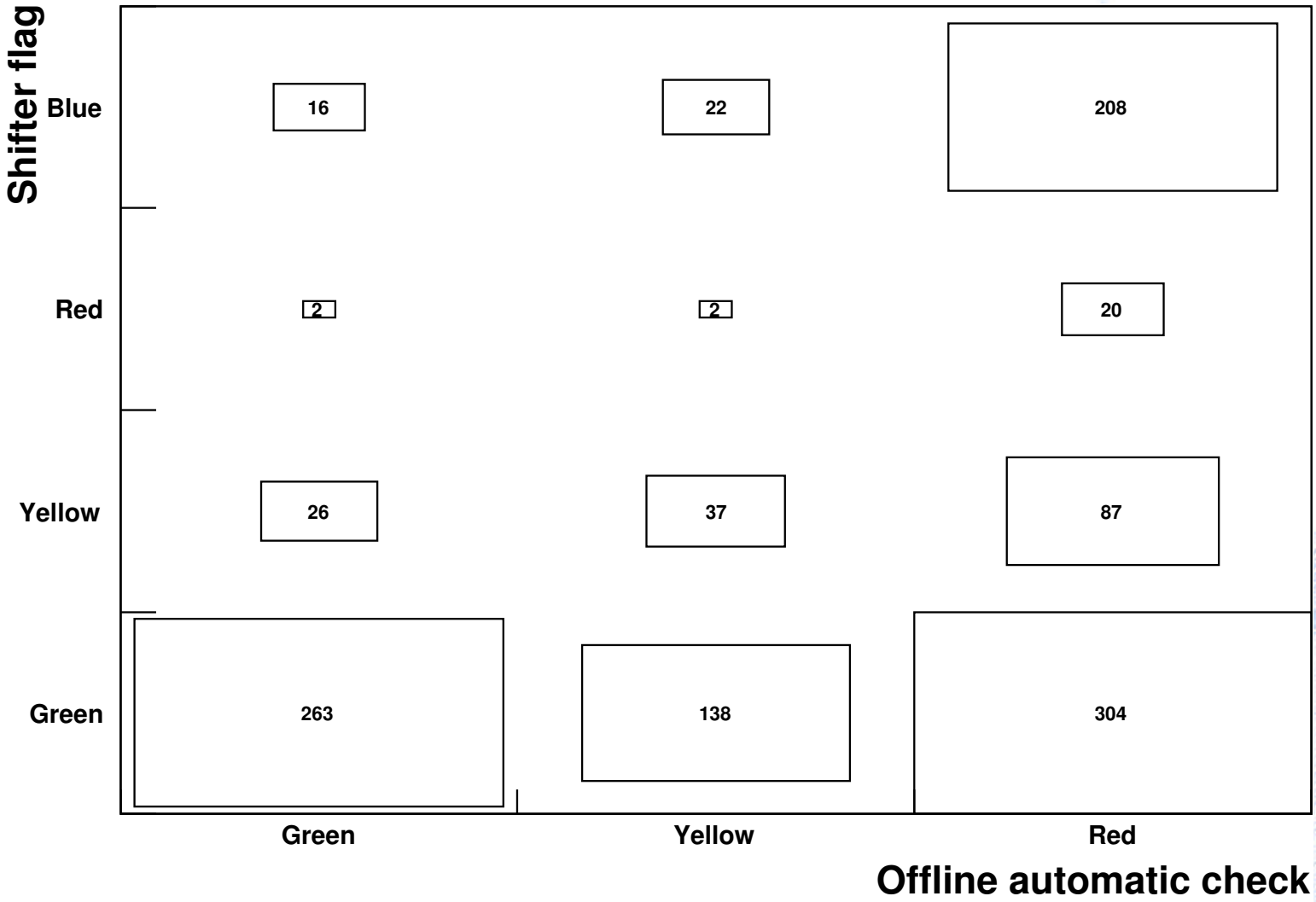
Introduction

Automatic DQ checks...

- ...are a set of algorithms defined in the ATLAS Data Quality framework
- ...are used to check control histograms generated for each run which (hopefully) reflect the quality of data
- ...report a set of parameters on which thresholds can be set
- ...serve as an input for a human shifter (man over machine)
- ...are run online during datataking on the data as it's taken and offline on all trigger streams

Some statistics

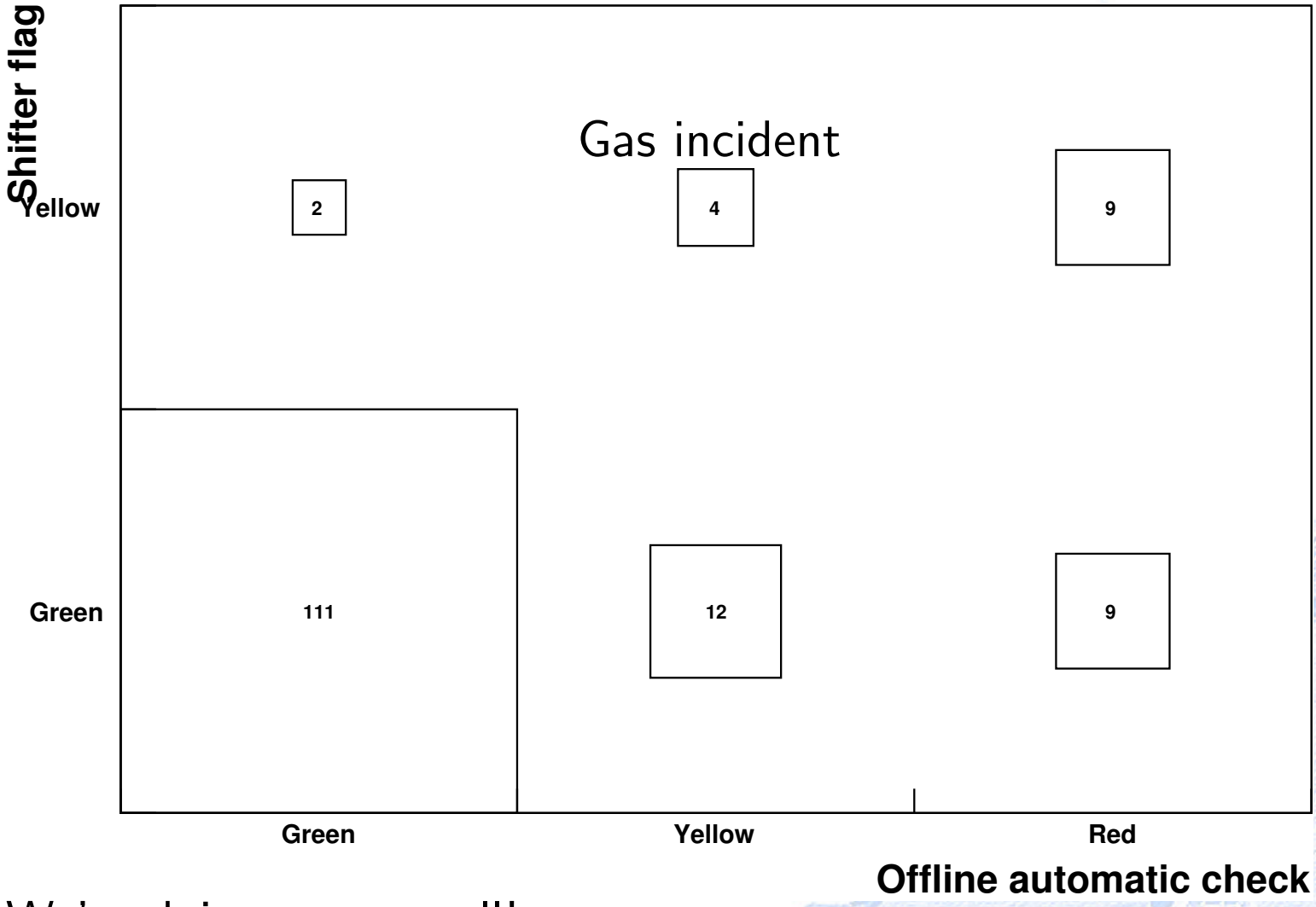
All runs in 2010



In general: stay on the safe side, rather be pessimistic.

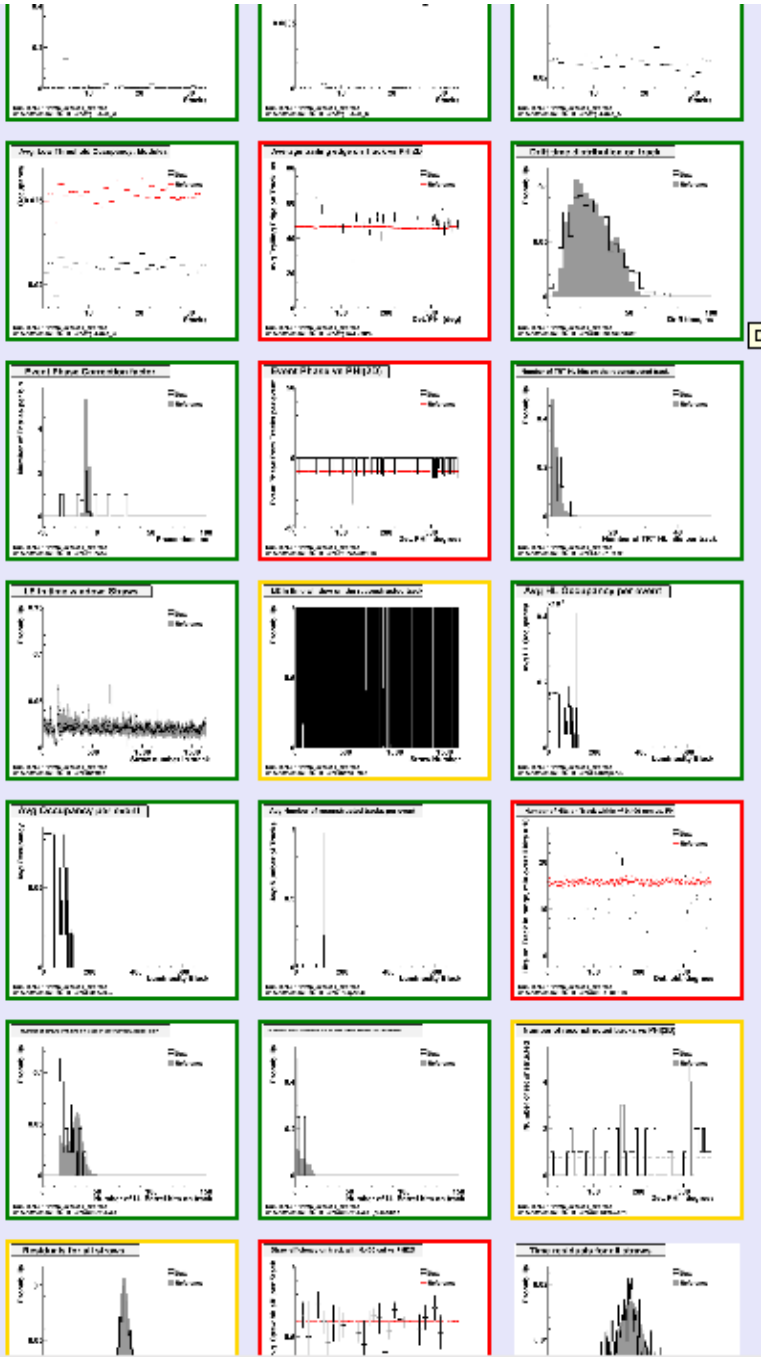
Some statistics

Runs with stable beams in 2010



We're doing pretty well!

Some statistics



For each run:

- 390 histograms checked online (70 shifter + 320 expert histograms)
- ...using 8 different algorithms
- 114 histograms checked offline
- ...using 9 different algorithms
- (generally should be the same on- and offline)

BinContentComp

Status: Green
Algorithm: BinContentComp
Num. of Entries: 14224.0

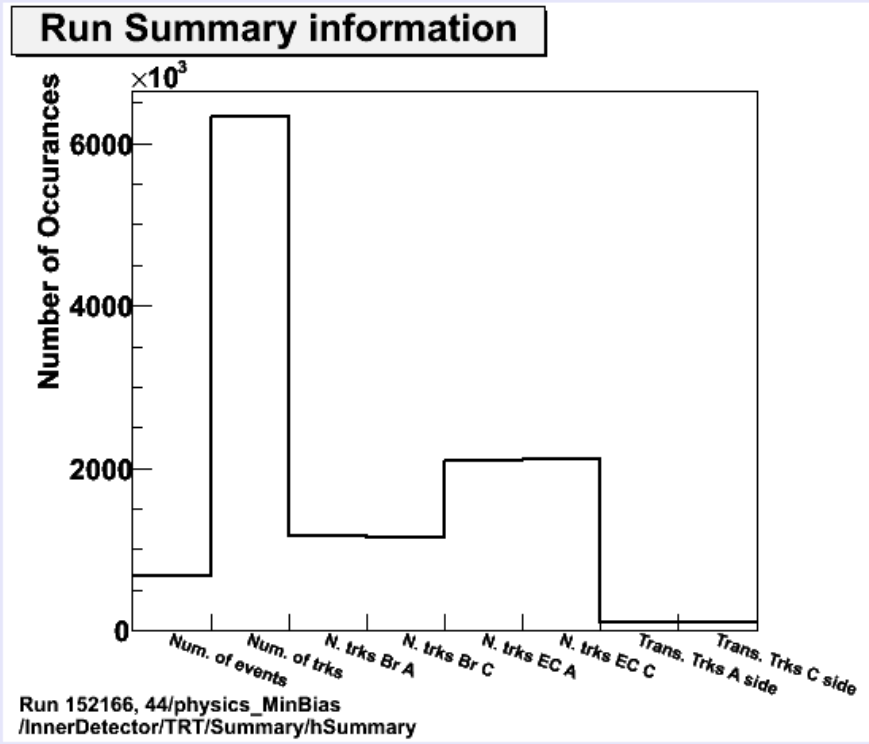
Configuration Parameters:

MinStat: 0.0
NSigma: 0.0
PublishBins: 1.0
Value: 0.0

NBins
XXXXXXXI XXXXXXXI XXXXXXX
8.0 8.0

Results:

N. trks Br A(2.500000): 1158000.0
N. trks Br C(3.500000): 1154000.0
N. trks EC A(4.500000): 2097000.0
N. trks EC C(5.500000): 2119000.0
NBins: 8.0
Num. of events(0.500000): 675300.0
Num. of trks(1.500000): 6326000.0
Trans. Trks A side(6.500000): 98920.0
Trans. Trks C side(7.500000): 102500.0



- Originally compares histogram with reference bin by bin
- In TRT algorithm is used to display the content of bins with non-zero entries

Bins_GreaterThan_Threshold

Assessment Details:

Name: hChipBSErrorsVsLB_EA
Status: Yellow
Algorithm: Bins_GreaterThan_Threshold
Num. of Entries: 479282.0

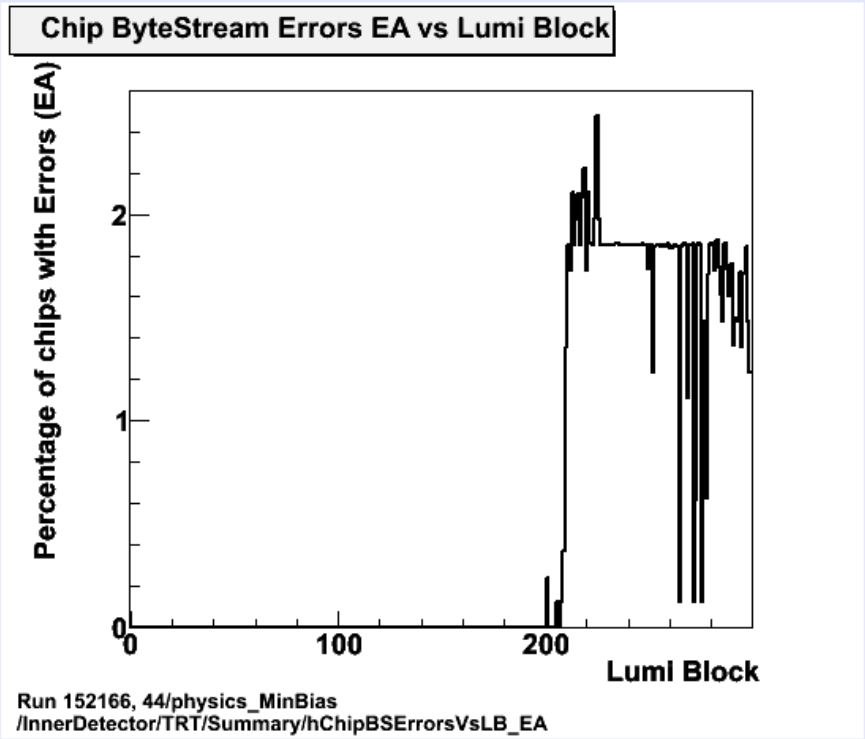
Configuration Parameters:

BinThreshold: 2.0
MaxPublish: 100.0
PublishBins: 1.0

NBins
XXXXXXX| XXXXXXX| XXXXXXX
2.0 20.0

Results:

(213.000000): 2.105
(216.000000): 2.096
(218.000000): 2.084
(219.000000): 2.219
(221.000000): 2.102
(225.000000): 2.474
NBins: 6.0



- Compares each bin of the histogram to be checked with a given threshold value

CheckPeakSimple

Assessment Details:

Name: hNumSwLLWoT
Status: Green
Algorithm: TRTCheckPeakSimple
Num. of Entries: 2312634.0

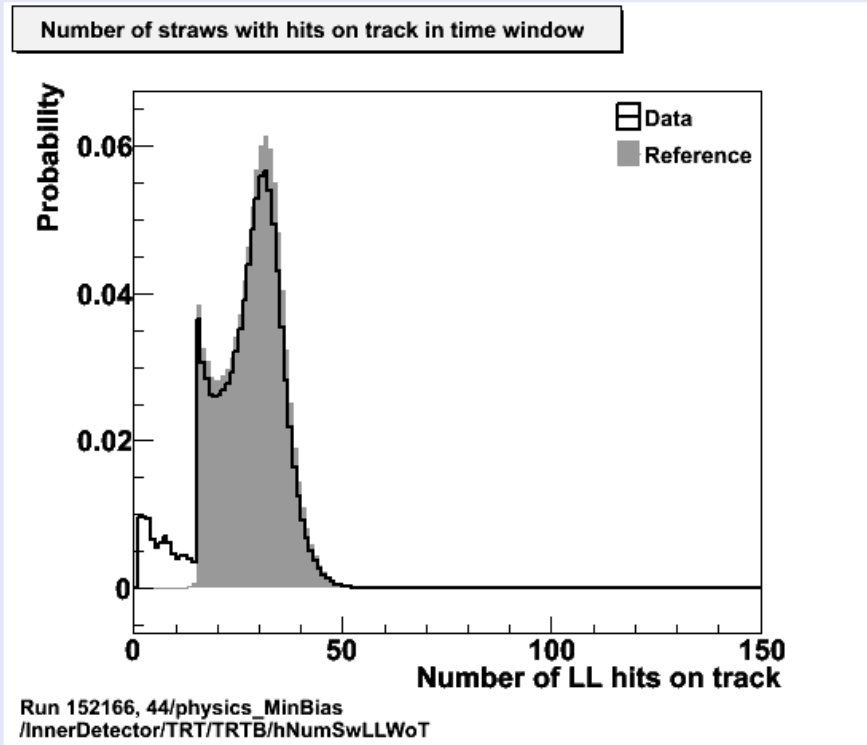
Configuration Parameters:

MinStat: 100.0

PeakPosition
XXXXXXXI XXXXXXXI XXXXXXX
5.0 10.0

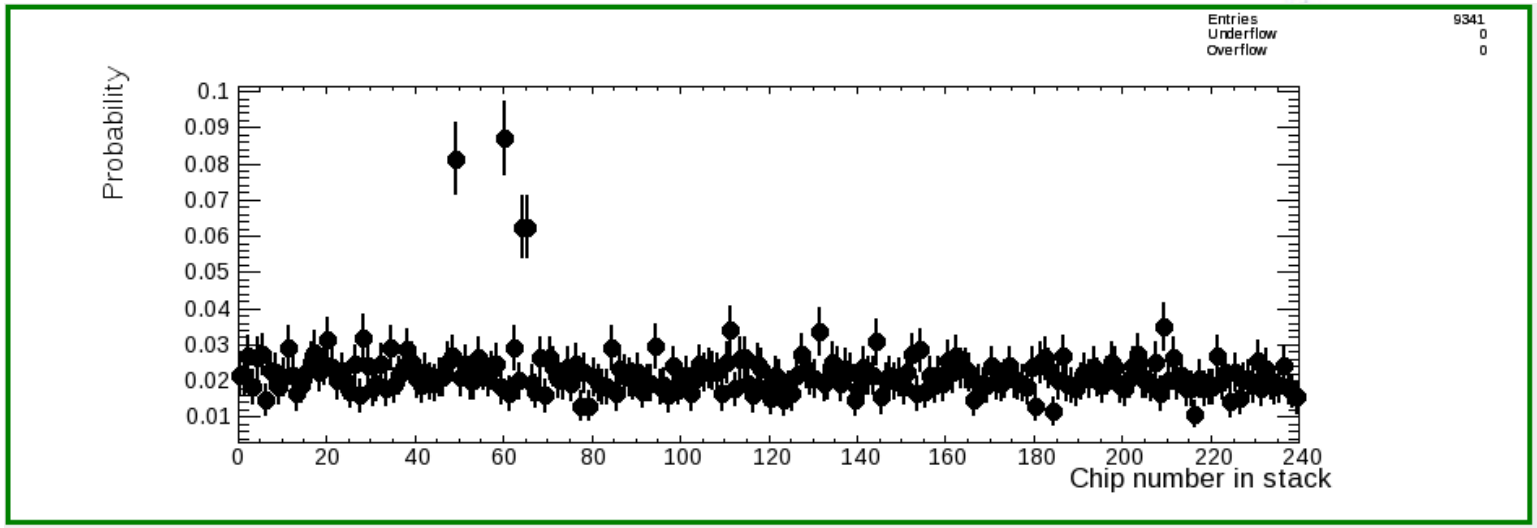
Results:

PeakBin: 31.99
PeakPosition: 31.49
Weighted_mean: 26.94



- Algorithm finds a peak and its position in a distribution

Check_Outlier_Bins



Description | Results | Troubleshooting | Configuration

Make e-log entry

Add to scratch pad

Arithmentic_mean	0.0221146
Number_of_bins_equal_zero	0
Number_of_outlier_bins	4
Standard_deviation	0.00787405

- Checks for outliers in an “to be flat“ distribution
- Outliers can be defined either in terms of numbers of standard deviations or by absolute deviations from the mean
- Counts numbers of bins fulfilling the given outlier definition
- Is used online for e.g. phi distributions

OutlierAndFlatnessTest

Status: Red
Algorithm: OutlierAndFlatnessTest
Num. of Entries: 56943.0

Configuration Parameters:

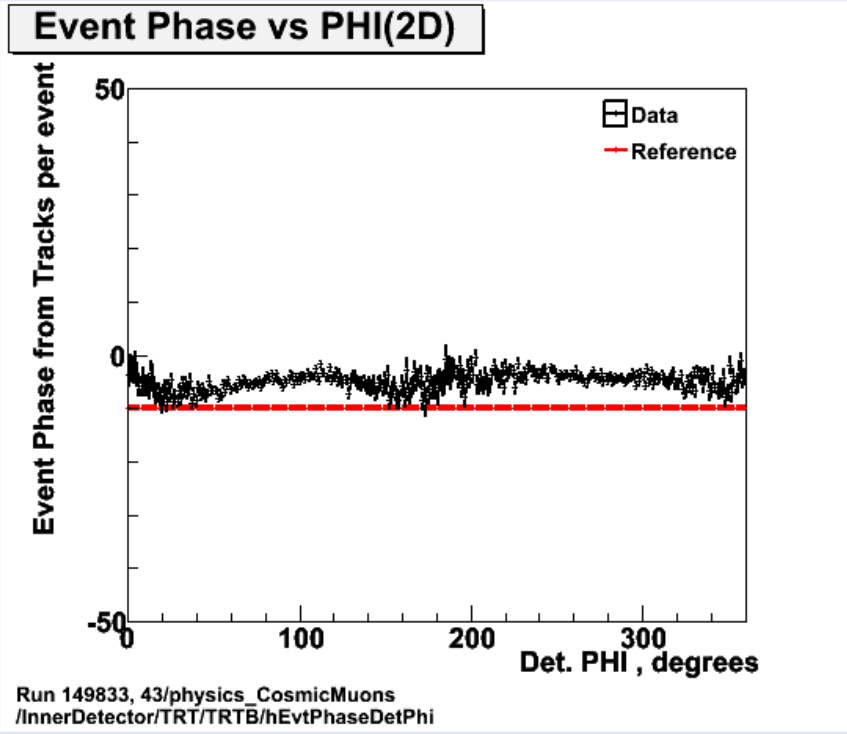
CheckSigmaDev: 1.0
FitCircular: 1.0
Ignore0: 1.0
MinStat: 1.0
SigmaDev: 4.0

Corrected_standard_deviation
XXXXXXXI XXXXXXXI XXXXXXX
1.0 1.5

Max_rel_asym_deviation
XXXXXXXI XXXXXXXI XXXXXXX
0.1 0.2

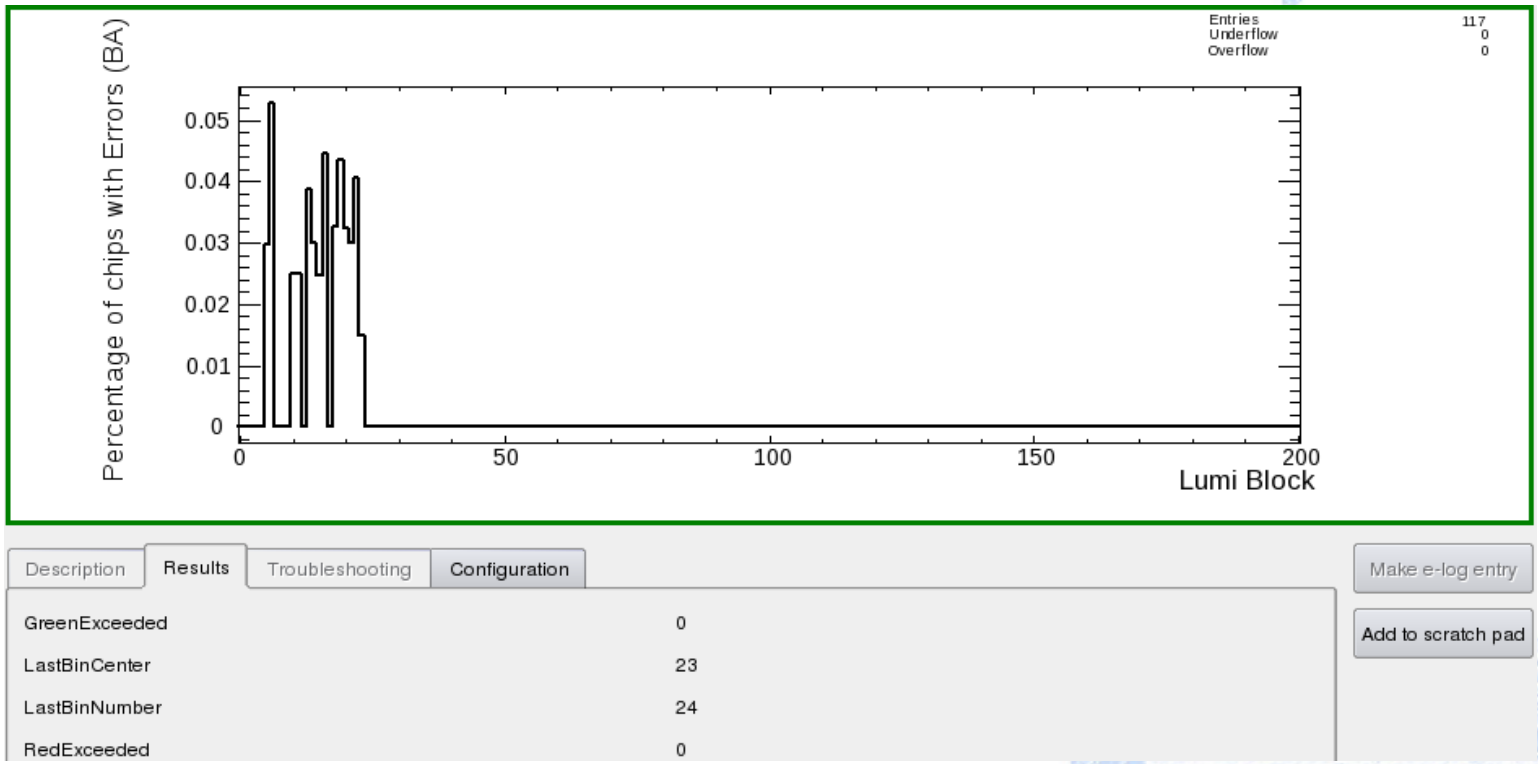
Max_rel_sym_deviation
XXXXXXXI XXXXXXXI XXXXXXX
0.1 0.2

Number_of_outlier_bins
XXXXXXXI XXXXXXXI XXXXXXX
5.0 10.0



- Finds, counts and removes iteratively all outliers
- The remaining distribution is tested for flatness
- Standard offline algorithm for all distributions expected to be flat

LastBinThreshold



- Algorithm compares the last non-zero bins if a histogram with a threshold value
- Is used for plots over time (LBs) continuously filled during run

SideBand_Relative

Assessment Details:

Name: hWireToTrkPosition_A
Status: Green
Algorithm: SideBand_Relative
Num. of Entries: 52318120.0

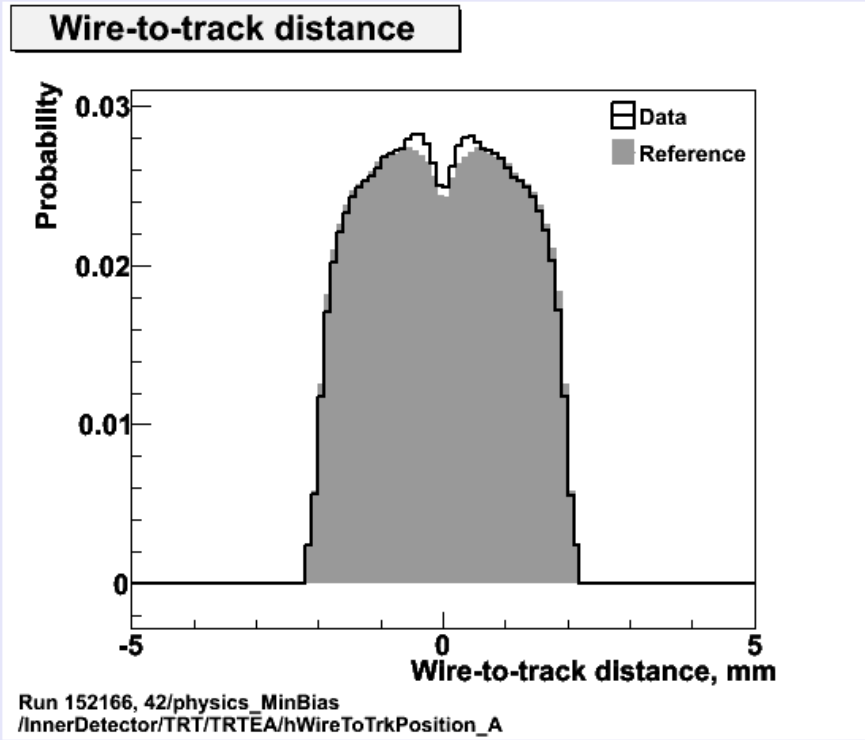
Configuration Parameters:

Max: 2.0
Min: -2.0
MinStat: 100.0
UseOverflow: 1.0
UseUnderFlow: 1.0

Threshold
XXXXXXXI XXXXXXXI XXXXXXX
0.1 0.2

Results:

SideBands: 0.01052
TotalIntegral: 1.0



- Check compares the integral of a distribution inside and outside a given range
- It's used to make sure there are not too many hits "outside" the physical tubes

Simple_gaus_Fit

Assessment Details:

Name: hResidual_C
Status: Yellow
Algorithm: Simple_gaus_Fit
Num. of Entries: 27821998.0

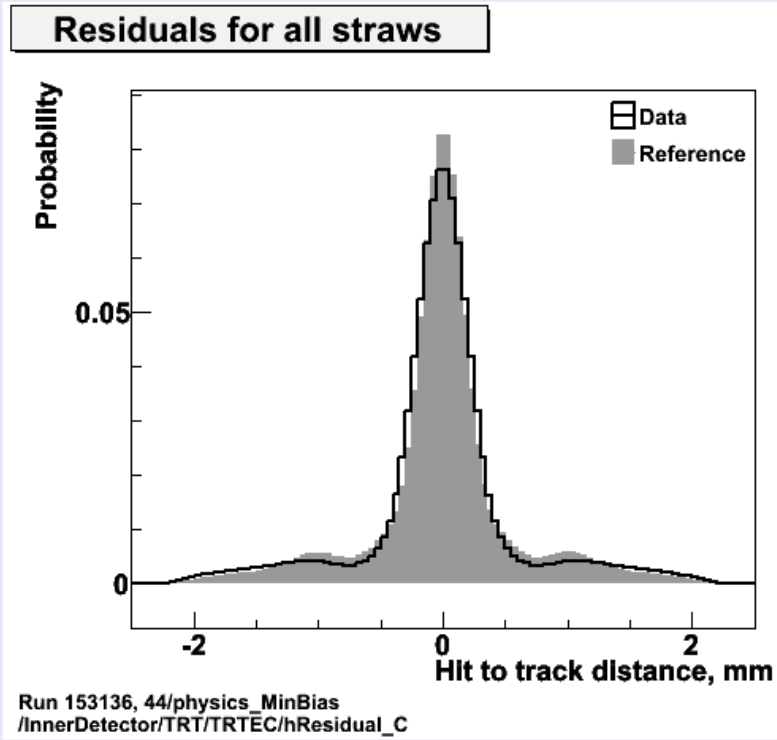
Configuration Parameters:

MinStat: 1000.0
xmax: 0.3
xmin: -0.3

Sigma
XXXXXXXI XXXXXXXI XXXXXXX
0.2 0.3

Results:

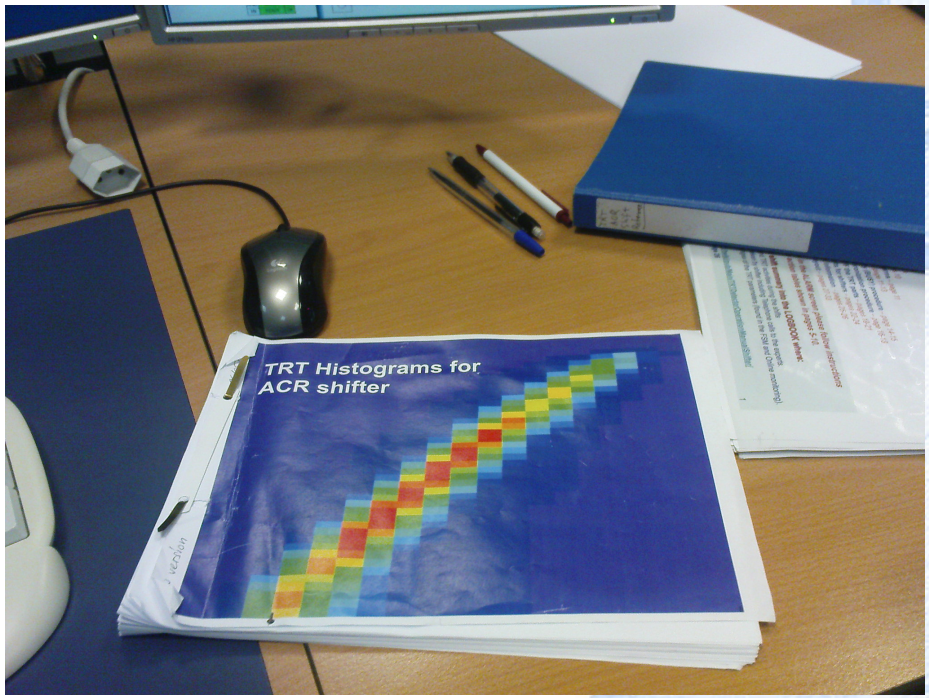
Constant: 0.07576
Constant Error: 2.44e-05
Mean: 0.0002085
Mean Error: 6.611e-05
Sigma: 0.207
Sigma Error: 8.832e-05



- Performs a fit of a Gaussian in a given range
- Reports fit parameters

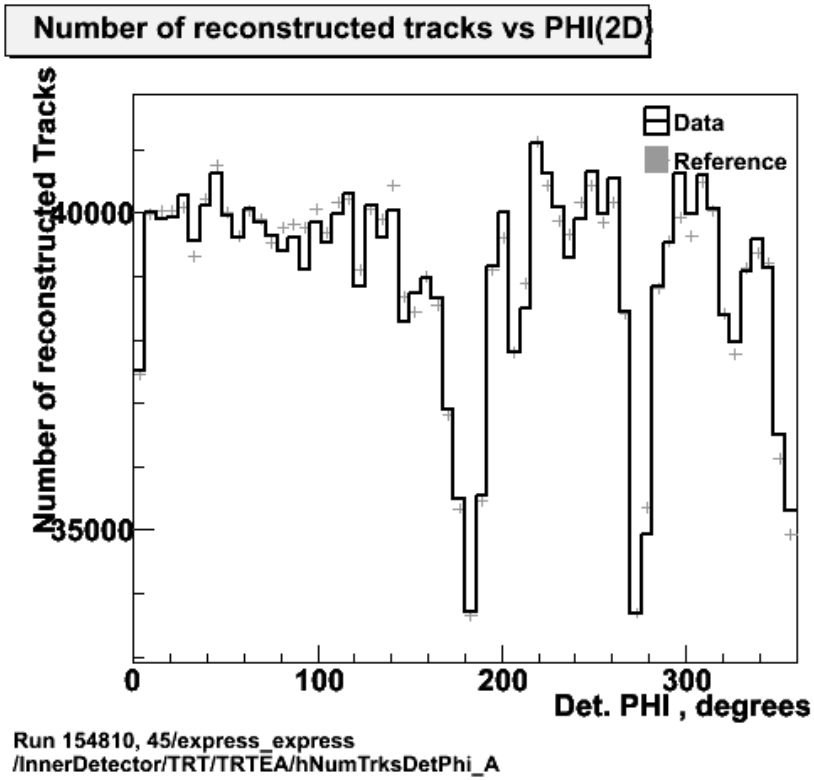
Documentation

- We put together a documentation about the histograms being checked and the Algorithm used for that
- It's available on the TRT desk in ACR and online `https://atlasop.cern.ch/atlas-point1/twiki/pub/Main/TRTDetectorOperationManualShifter/DQHistogramsOnline.ppt`



Outlook

Where we have problems...



- In principle this should be flat...
- Dips are due to known inefficiencies: Expected and OK

- How to treat such histograms?
  - loosen thresholds? $\leftrightarrow$ loss of sensitivity
  - mask bins? $\leftrightarrow$ blind to problems in that area
  - compare to reference? $\leftrightarrow$ "bad" reference looks like problem

Outlook

Where we need some changes...

Name: hDriftTimeonTrkDist_A
Status: Green
Algorithm: TRTCheckPeakSimple
Num. of Entries: 20964830.0

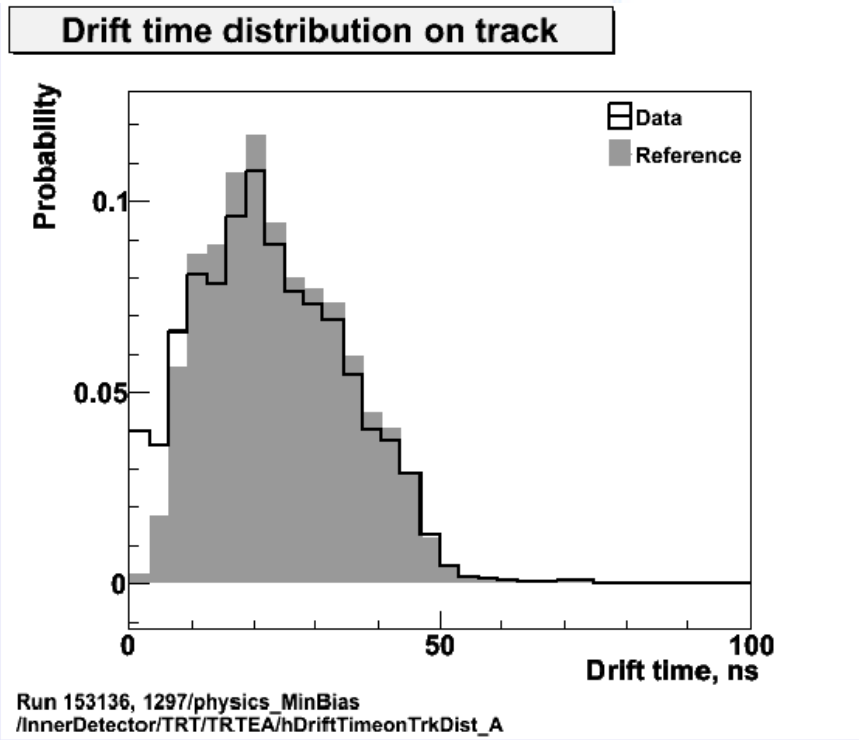
Configuration Parameters:

MinStat: 10.0
XMax: 100.0
XMin: 1.0

PeakPosition
XXXXXXXI XXXXXXXI XXXXXXX
3.125 15.62

Results:

PeakBin: 6.976
PeakPosition: 20.24



- Not the right check for this histogram
  $\Rightarrow$ In principle we don't care for the peak position
- Would rather make sure that distribution does not run "out of borders" (this example should not be flagged green!)

Outlook

What we are missing...

- Algorithm to set relative thresholds on the content of the "outermost" bins
  $\Rightarrow$ Modified sideband check usable?
- Tool to deal with "flat up to single known dips" distributions
- Algorithm to spot sudden changes in "value over time" histograms
  $\Rightarrow$ Derivative test
- A tool to determine the width of a gaussian shaped core of an otherwise non-gaussian distribution
  $\Rightarrow$ Gauss fit with iterative narrowing of the fit range
- We have no idea about heavy ion running yet

Generally: Need constantly evaluation and tuning of parameters and thresholds!

Thanks for your attention