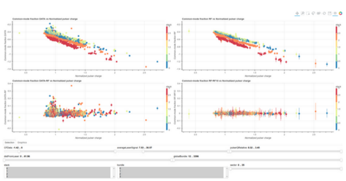
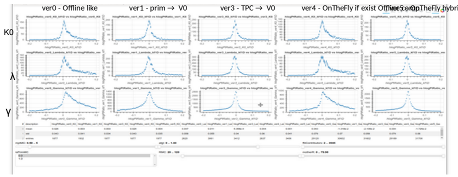
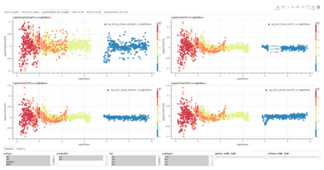
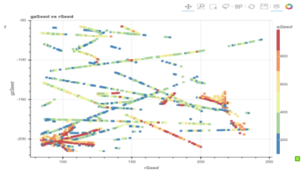








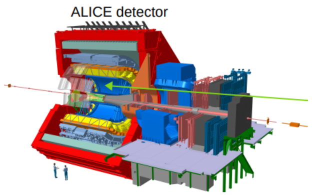
RootInteractive tool for multidimensional statistical analysis, machine learning and analytical model validation

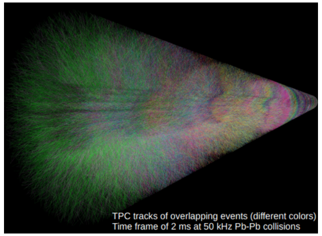
Seeing is believing

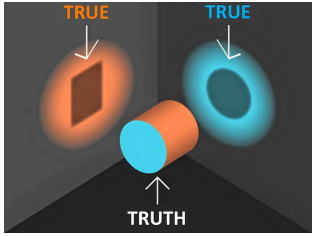
Marian I Ivanov (GSI), Marian Ivanov jr (UK Bratislava)

Alice Projects (Run1,2, Run3, ALICE 3)

- Run3 space charge calibration - Ernst, Mathias,Caitie, Marian (GSI,Frankurt)
- trackCombinator (V0Cascade/exotica) Marian (GSI+Heidelberg), Benedict (Frankfurt), Marian (UK Bratislava)
- CRU -Run3 digital signal processing - Yiota, Mesut,Marian (CERN, Yale)
- Run2 Performance web pages - Pritam, Dibakar, Tulika, Marian (Kolkota)
- PID calibration and dEdx optimzation - Tuba (Frankfurt),Mathias
- **High dEdx,spallation/Magnetic monopole - Marian, Timon (Wiena)**
- MC/Data remapping - Yale group, Marian, Marian
- **Particle production … combine estimator - Michal,Marian (UK Bratislava)**
- fastMCkalman - Marian, Federico (Oxford)
- Run3 TPCQA/ QC - Berkin
- data skimming - MI, Mesut, Berkin
- TPC data volume studies - Marian, Marian jr, Mesut







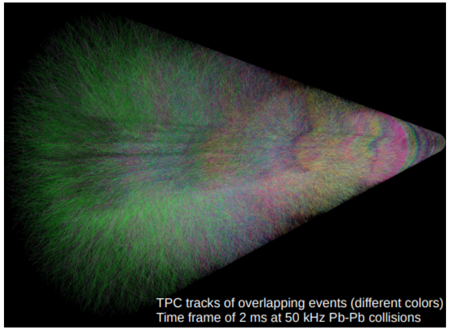
https://github.com/miranov25/RootInteractive/releases/tag/v0-01-09
https://indico.cern.ch/event/1135398/

Alice Run 3 - goals and challenges

Record large Pb-Pb minimum bias sample

Continuous readout at 50 kHz interaction rate in Pb-Pb collisions

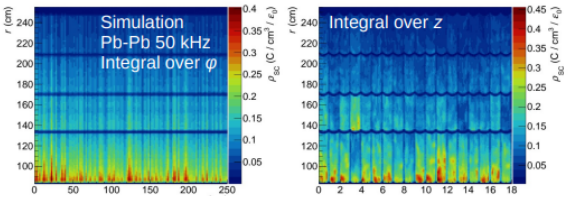
- No triggers or event rejection. Unknown time 0
- Reconstruction (in GPU) - Processing of time frames (TF, 10 - 20 ms) instead of events
- Events overlapping in TPC → substantial higher occupancy (~5 event)



New TPC GEM design → **space charge** in TPC

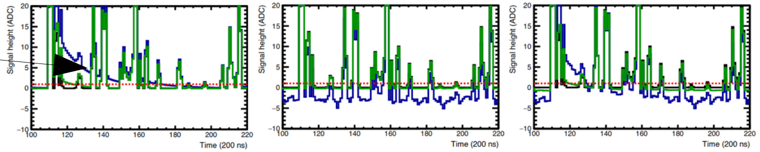
inside the drift volume **distorting** trajectories

- Non-uniform space-charge density ρ_{SC} ➔ Large space-charge distortions (dr, drφ, dz) of measured space points O(5 cm)
- ➔ Space-charge density and distortion fluctuations O(5 %) ~ 0.2 cm
- **To be calibrated/corrected to σ ~100 μm with granularity O(10^6) in space O(1-5 ms) in time**



Significant **baseline bias and baseline bias fluctuation** comparable with signal amplitude

- Online digital signal processing to recover baseline (in FPGA)
- To be corrected below internal noise level



A high interaction rate environment, pile-up, distortions fluctuation, etc. ... necessitates the use of advanced methods of data analysis. Experts and highly customisable tools are needed

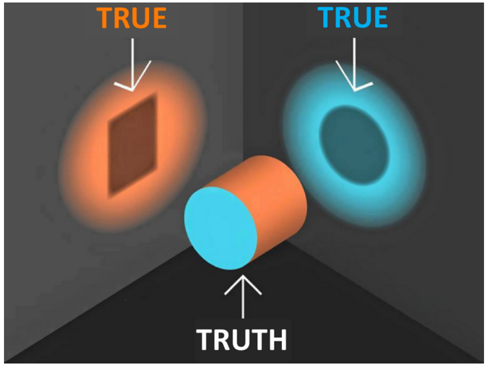
RootInteractive general purpose tool for ND statistical analysis

https://en.wikipedia.org/wiki/Occam%27s_razor

“Occam's razor is the problem-solving principle that "entities should not be multiplied without necessity",[1][2] or more simply, ***the simplest explanation is usually the right one****.”*

https://en.wikiquote.org/wiki/Albert_Einstein

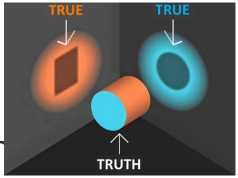
"Everything should be made ***as simple as possible, but no simpler****,"*



By oversimplifying in analysis level, the explanations tends to be more complex resp. wrong

Our goal to provide a tool to deal with multidimensional problems

- simplify data analysis in many (optimally all relevant) dimensions
- fit (ML regression) and visualise N-dimensional functions **including their uncertainties and biases**
- validate assumptions, approximations
- enable simple **functional composition for (non-parametric, parametric) functions and error propagation**
- aimed for **standard users** (Masters, PhD), not just computer experts for educational purposes
- **very fast feedback** from day one - **seconds instead of weeks,** to allow **interactive expert communication**
- for **multidimensional parameter optimisation** with fast convergence
- answering question “What happen if? (changing a paremeter, normalization)” within seconds → “Expert making”
- **Tool for Open data**



NDimensional interactive analysis - **Seeing is believing**

- general-purpose tool for multidimensional statistical analysis. It uses a declarative programming paradigm where it build the structure and elements of computer programs and express the logic of a computation without describing its control flow

NDimensional analysis pipeline (2015) & RootInteractive (2019)

- **Expert highly customizable** tool for multi-dimensional analysis and machine learning
- **Functional composition - non-analytical and analytical (physical model) functions**
- Software description - ALICE independent package
- **Interactive analysis** (ML, fits, histograming, data aggregation O(10^6-10^7)) **on server (Jupyter notebook) and on clients (browser)**
- Possible ideal tool for open data
- **Triggers and data skimming (representative data selection) to enable interactive analysis**

ND+RootInteractive functionality shown in real use cases - see subset in RootInteractive tutorials

- March 2022 -https://indico.cern.ch/event/1135398/
- December 2022 https://indico.cern.ch/event/1135398/

Interactive differential studies, physics analysis using skimmed/sampled data under preparation

- Particle production as a function of combined multiplicity estimators (and event shape) in pp, pPb and A-A collisions with ALICE
- Starting with Event generators

Content

ND - pipeline description and motivation example use cases

- Multidimensional parameter optimisation
- Non-parametric and parametric/analytical models
- Examples of functional composition
- Example of **invariances/symmetries** for calibration and automatic QA
- RootInteractive/MultiInteractive Session

RootInteractive/MultiInteractive functionality explained

- **Interactive ND histogramming**
- **TTree data loading**
    - **current and new interface RDataFrame->awkward**
- **Machine learning wrappers for local robust value estimators and local error estimators**
- **aggregated and derived information - functional composition**
- **parameterized functions on clients**
- **user defined figure transformation on client**
- **RootInteractive widget for selection, weights and parametererization**
- **ONNX interface on client (starting)**

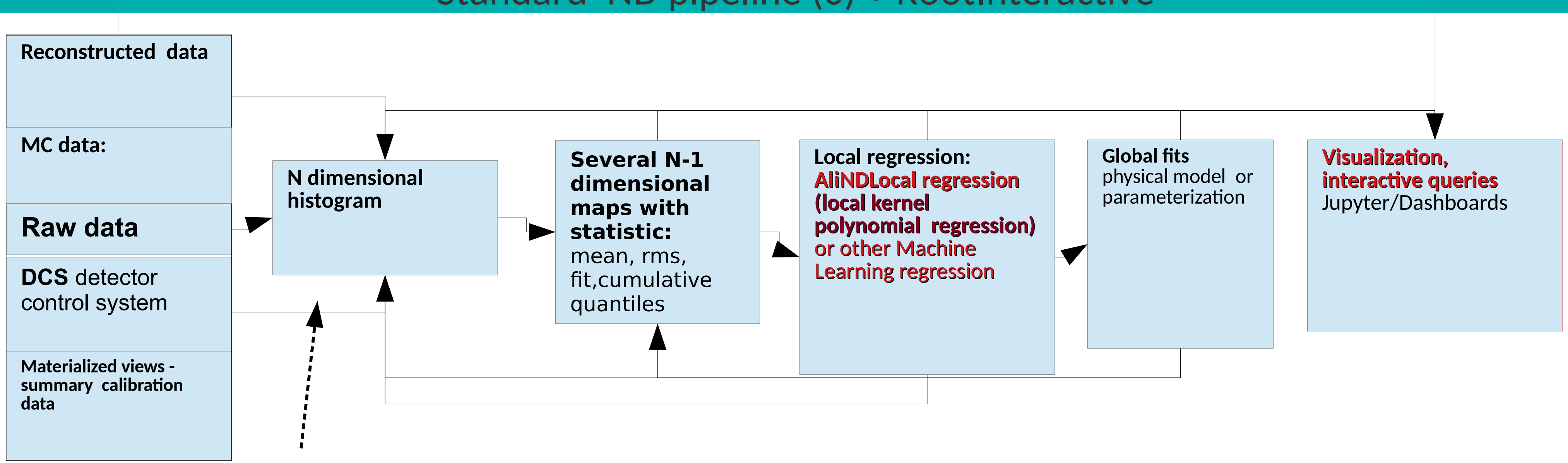
Representative Sampling/skimming

Work in progress

Multidimensional analysis pipeline (2015-…) & RootInteractive (2019 - ..)

- ND pipeline - libStat library written in C++98 (AliRoot+ROOT5)
- RootInteractive (Python+TypeScript+C++)
- NDpipeline - usage examples
  - Run2 distortion calibration, investigation of the space space charge origin and distortion physical model fits
  - TPC calibration, Tracking performance parameterization, MC/data parameters tuning
  - Detector and reconstruction QA
  - Toy MC, Digital signal processing optimization …

Standard ND pipeline (0) + RootInteractive



$$f(p_0, p_1, p_2, \ldots) \neq f_0(p_0) \oplus f_1(p_1) \oplus f_2(p_2) \oplus \ldots .$$

Standard calibration/performance maps and QA done and interpreted in multidimensional space

- dimensionality depends on the problem to study (and on available resources)
- **skimmed version of input data usually used in interactive or semi-interactive analysis**
- Data →Histogram → set of ND maps → set of NDlocal regression/TMVA → Global fits (physical model)
- Histogramming in case of non sparse data
- **ML for sparse (going to higher dimensions)**

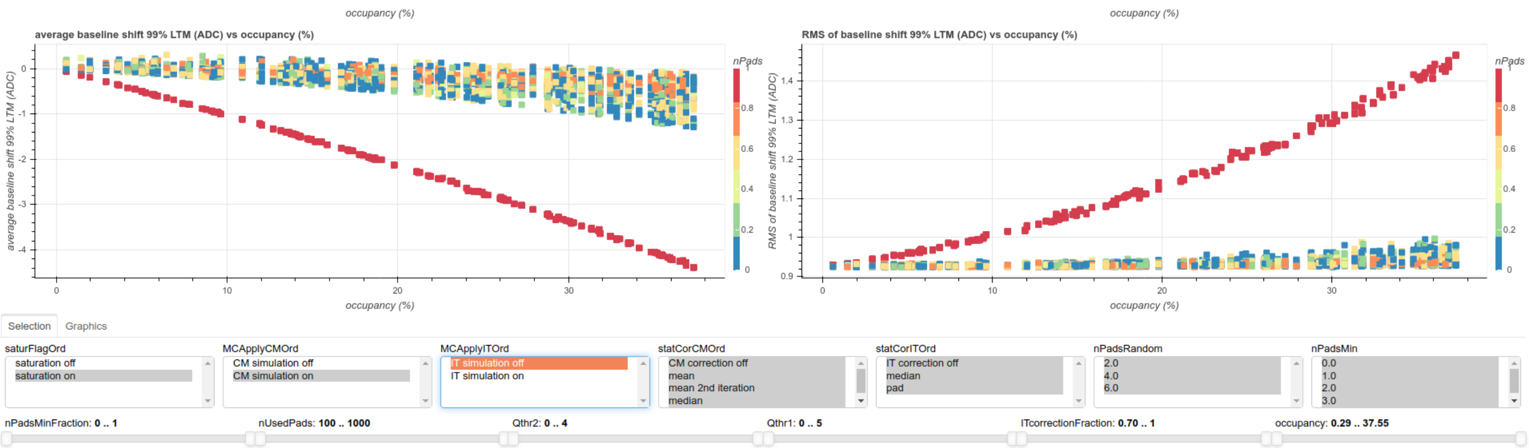
- **Generic "interactive" code. Minimizing amount of custom macros.**
- **"Declarative" programming - simple queries**
- **Non parametrical and parametrical functions physics models**

Mutidimensional parameter optimization example - digital signal processing (2021)

Digital signal processing (13 parameters in example) needed for particle identification and data volume optimization. **O(200000) parameter settings** simulated/generated on server

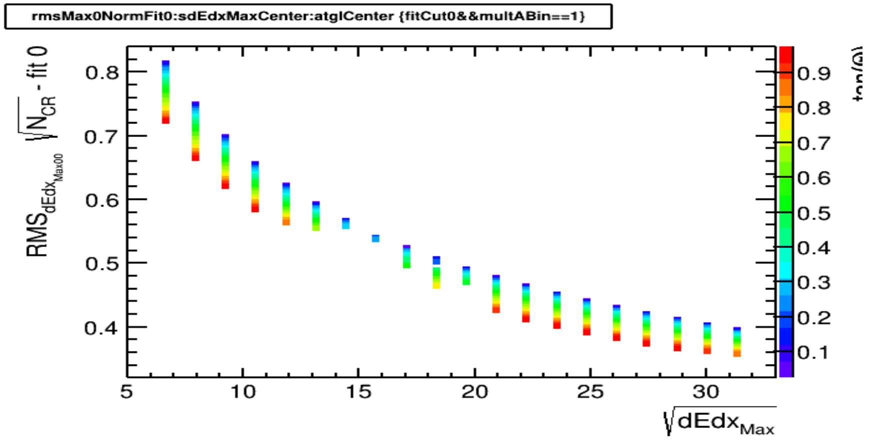
- parameters: effects (On/Off), algorithm (different version), parameters of individual algorithms
- simulation and visualization/aggregation (NDPipeline+RootInteractive) **done by master student**, very effective for education
- enabling very constructive interactive discussion within expert group, quickly converging to “expert” decision, generating new ideas,
- **FEEDBACK time for follow up questions O(seconds)**
- **standalone dashboards as a support material for internal/public notes**

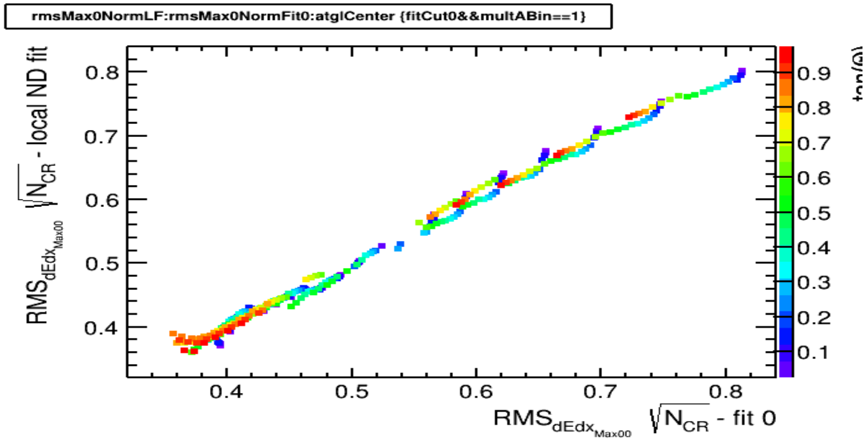
TPC baseline bias **TPC baseline fluctuation - rms**



Presentation, notebook, interactive dashboard and movie in RootInteractive tutorial:
https://indico.cern.ch/event/1135398/contributions/4764024/subcontributions/370740/attachments/2402507/4114272/CMITSimulGEMTPC_RootInteractiveTutorial10032022.pdf
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/master/JIRA/ATO-559/parameterScan.ipynb
https://indico.cern.ch/event/1073883/contributions/4588170/attachments/2334149/3986420/simulScan_02112021.html
https://indico.cern.ch/event/1135398/contributions/4764024/subcontributions/370740/attachments/2402507/4109039/CMITSimulationsGEMTPC.mp4

Non-parametric and parametric/analytical models example: TPC dEdx resolution





Physical model:

dEdx resolution depends on 3 main variable
dEdx, track length (tan(θ)) and number of measurement (N_{CR})
3 measurement in regions (i,j,k)

$$RMS_{Qi} = \sqrt{RMS^2_{Qi/Qj} + RMS^2_{Qi/Qk} - RMS^2_{Qj/Qk}/2}$$

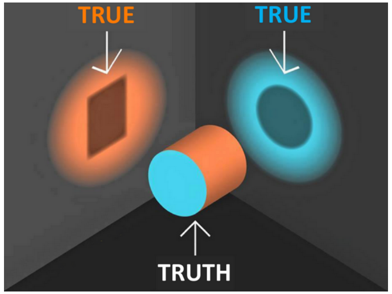
$$RMS_{ROC} \times \sqrt{N_{CR}} \approx p_0 \left(dEdx^{p_1} \times \sqrt{(1+\tan(\theta))^2}^{p_2} \right)$$

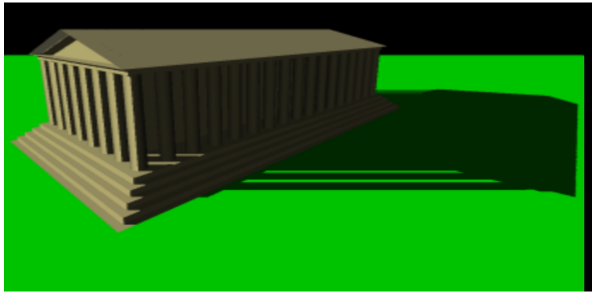
Input data pipeline:

skimmed data → 6x4D histograms of dEdx ratios in regions → 6x3D PDF maps (non parameteric) → local fits → global fit of physical model

Example conclusion: At low IR agreement between dEdx intrinsic resolution and power low model as expected

dEdx related studies in tutorials:
Run1: https://indico.cern.ch/event/1135398/#sc-3-2-dedx-calibration-nd-cor
Run3: https://indico.cern.ch/event/1221198/#7-optimization-of-the-tpc-and





https://www.youtube.com/watch?v=a7LCTT7HKzc

$$\sigma_{\vec{A}\ominus\vec{A}_{ref}} \leq \sigma_{\vec{A}}(+)\sigma_{\vec{A}_{ref}}$$

Object and reference objects should be compared optimally in the relevant ND space.

Shadow projection → Assumptions, imagination and rhetorical art in describing data needed

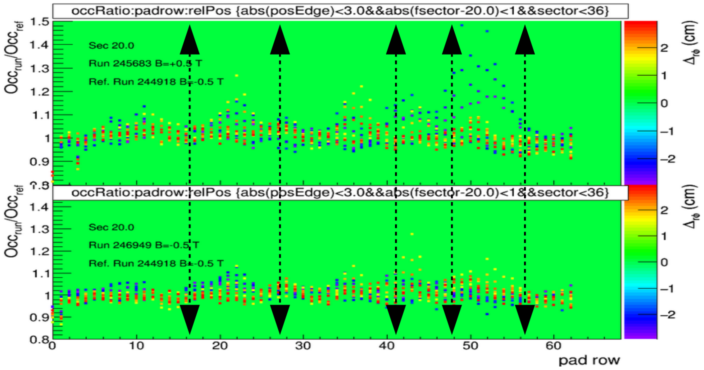
QA alarms, calibration validation, statements to be based on invariances or on normalized data - e.g. the difference between the object and the reference object

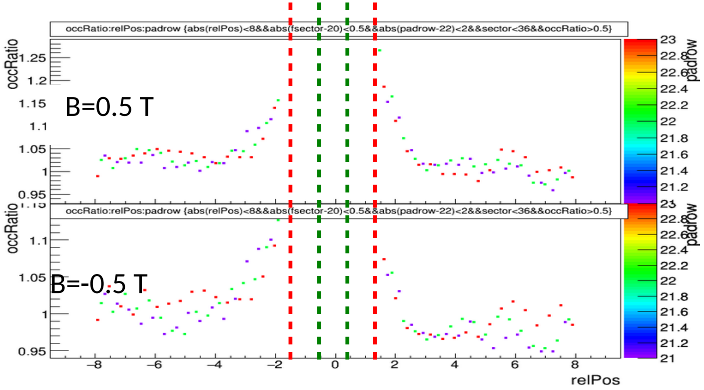
- After projection impossible
- In many typical cases variance $\sigma_{A\text{-}Aref}$ is very often smaller by orders of magnitude

Example functional composition - Hunting for distortion origin (2015)

$$\sigma_{\vec{A}\ominus\vec{A}_{ref}} \leq \sigma_{\vec{A}}(+)\sigma_{\vec{A}_{ref}}$$

Increase in normalized occupancy near the hot spot region due to space charge distortion





Analytical model - derivative of E field due line charge (analytical model) workin with "non-analytical" maps:

$$\frac{N_{Cl}(IR)}{N_{Cl}(IR=0)} = \frac{(w+(\Delta_{r\phi}(r_\phi+w/2)-\Delta_{r\phi}(r_\phi-w/2)))}{w}$$

$$R = \left(\frac{Occ}{< Occ_{ROC} >}\right)_{IR} / \left(\frac{Occ}{< Occ_{ROC} >}\right)_{IR=0}$$

Conclusion: **Distortion origin in the gap between sectors σ(mm) → No doubts → Hardware intervention approved**

Very precise measurement of the origin of the distortion - **measurement of the derivative of the distortion** with **sub-pad granularity**.
Without adequate normalisation to the reference (double ratio), the effect was invisible hided by other detector effects at sector boundaries → False conclusion by students in the first analysis

$$\sigma_{\vec{A}\ominus\vec{A}_{ref}} \leq \sigma_{\vec{A}}(+)\sigma_{\vec{A}_{ref}}$$

Data should be compared with reference model/data in multi-dimension

- RMS spread is much smaller (see ALICE performance example in next slide)

Invariance/symmetries

- in-variance in time (using e.g. reference/average run), in-variance in space (e.g. rotation, mirror symmetry) , B field symmetry
- data - non parameteric/parameteric physical model
- smoothness resp. local smoothness
- Outlier tagging with statistical significance - e.g. (data-model) > N σ

MC-Data comparison - should be done in N dimension not on projections

Aggregation/projections of normalized data in NDimension

Projections problems (hidden variables):

- **Information loss. Intrinsic spread of variable vectors A and A ref is usually significantly bigger than spread of $A-A_{ref}$**
  - noise map, DCA bias, resolution maps, occupancy maps, sigma invariant mass maps …. as function of 1/pt, θ, occupancy, dEdx
- **Projected vector A depends on the actual distribution of hidden variable**
  - Sometimes misleading results, non trivial interpretation of projected observation

Expert data preparation

- Agreement on data to collect and aggregate
- Data sources
- Variables to import - asking questions
    - Symmetries, invariances and possible alarms
- Pre-aggregation
- Data sampling
- Machine learning models
- Analytical models
- Re-iteation

Data presentation

- **Agenda: presentation, notebook, dashboard+ (optional)movie)**
- Goal
- Data preparation explained
- Variables description
- Observation highlights with snapshot from dashboard

Domain experts, participants in the meeting should be able to participate in decissions, resp. be able to interact with dashboard data based on desciption in presentation and

The data is presented in a multidimensional way. The aim is to answer all questions within one session If the information is not sufficient, new data sources to be agreed on.

RootInteractive General-purpose tool for multidimensional statistical analysis.

Declarative programming paradigm, where it build the structure and elements of computer programs and express the logic of a computation without describing its control flow

- Visualization and on client aggregation Python/TrueScript (RootInteractive)
- Machine learning wrappers in Python
- PyRoot used to be able to use Root and O2 and RootInteractive together
- **Explained on the TPC calibration/QA Example - building application in your browser - calibration/QA ND viewer and track ND viewer**

https://docs.google.com/presentation/d/10YPAwv8tdTZSPtX5IQ0T_0wEklOH6Hfl2aPy6l-YLJk/edit#slide=id.g1a652381433_0_0

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html
https://indico.cern.ch/event/1215913/contributions/5114796/attachments/2537879/4368596/tpcTracks.html

RootInteractive dashboard declarations

- User defined RootInteractive properties are required to get the html output:
  - **Alias array for derived variable definition - e.g defining status bitmask**
    - aliasArray=[("IDC0_OK","(0x2*(abs(IDC0_MeanRF0_LRatio)<sigmaRFCut0))|(0x4*(abs(IDC0_MeanRFL_LRatio)<sigmaRFCutL))"),..]
  - **Variable array**
    - variables=[..., ...]
  - **Parameter array - to control parameterized functions, selection and variable selection for ND histograms**
  - **Widget parameters**
    - [["select", ["varX"], {"name": "varX"}], ...]
  - **Widget layout dictionary**
    - {"Histograms": [["varX", ...], {'sizing_mode': 'scale_width'}], ...}
  - **Histogram array**
    - [{"name": "histoXYData", "variables": ["varX","varY"], "nbins":["nbinsX","nbinsY"]}, ...]
  - **Figure array**
    - [["bin_center_1"], ["bin_count"], {"source":"histoXYData", "colorZvar": "bin_center_0"}], ...]
  - **Figure layout dictionary**
    - {"histoXY":[[0,1], [2,3], {"plot_height":220}], ...}

- Links to the dashboard and the jupyter notebook:
  - https://indico.cern.ch/event/1221198/#8-ato-611-tpc-global-calibrati
  - https://indico.cern.ch/event/1221198/#7-optimization-of-the-tpc-and

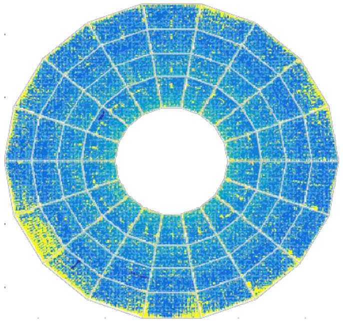
Machine learning - derived variables - RF regression - pad map calibration

```
statDictionary={"mean":None,"median":None, "std":None}

varListG=["lx","ly","GainMap","A_Side"]
varListLocal=["lx","ly","GainMap","roc"]
vars=[
   "NClusters_Clusters_Mean",'NClusters_Digits_Mean',
   'QMax_Clusters_Mean', 'QMax_Digits_Mean',
   'IDC0_Mean','SAC0_Mean'
]
```

Example derived variable for NClusters_Clusters:

- **NClusters_Clusters_Mean**
- NClusters_Clusters_MeanRF0,
- NClusters_Clusters_MeanRF0,
- NClusters_Clusters_MeanRFL,
- NClusters_Clusters_MeanRFL_Med
- NClusters_Clusters_MeanRFL_Std



ML models:

- varying parameter of models, input variables and local statistics

Global (varListG) and local regression (varListLocal) extracting for basic calibration and QA properties

- globa **φ symmetric** model, local model **without φ symmetry**

Robust local statistics - median and local std estimator for the outlier tagging

Functions on client, derived variables and functional composition

Predefined parametric javascipt function -

```
# here we can define derived variables -  to define some invariances eg abs(XX_Mean/XXXMedain)<
aliasArray=[
#    ("","dNprimdx*padLength"),      # ionization over pad
    ("Unit","1+roc*0"),
    ("phi","arctan2(gy,gx)"),
    ("QMax_Clusters_OK","(0x1*(NClusters_Clusters_Mean>minEntries))|(0x2*(abs(QMax_Clusters_MeanRF0_LRatio)<sigmaRFCut0))|(0x4*(abs(QMax_Clusters_MeanRFL_LRatio)<sigmaRFCutL))"),
    ("QMax_Digits_OK","(0x1*(NClusters_Digits_Mean>minEntries))|(0x2*(abs(QMax_Digits_MeanRF0_LRatio)<sigmaRFCut0))|(0x4*(abs(QMax_Digits_MeanRFL_LRatio)<sigmaRFCutL))"),
    ("SAC0_OK","(0x2*(abs(SAC0_MeanRF0_LRatio)<sigmaRFCut0))|(0x4*(abs(SAC0_MeanRFL_LRatio)<sigmaRFCutL))"),
    ("IDC0_OK","(0x2*(abs(IDC0_MeanRF0_LRatio)<sigmaRFCut0))|(0x4*(abs(IDC0_MeanRFL_LRatio)<sigmaRFCutL))"),
    #("IDC0_OK","1+(abs(IDC0_RMS/IDC0_Mean)<0.5)"),
    ("IDC0_MeanOK","0x1*(IDC0_RMS<5) |0x2*(IDC0_MeanLxCut)")
]
```

Anonymous function (used for example in ND histograms as weights or variable)

varX	varY	varYNorm	varZ	varZNorm
gx	gy	Unit	QMax_Digits_Mean	QMax_Clusters_Mean

```
{
    "name": "histoXYNormZData",
    "variables": ["varX","varY/varYNorm","varZ"],
    "nbins":["nbinsX","nbinsY","nbinsZ"], "axis":[1,2],"quantiles": [0.35,0.5],"unbinned_projections":True,
},
{
    "name": "histoXYZNormData",
    "variables": ["varX","varY","varZ/varZNorm"],
    "nbins":["nbinsX","nbinsY","nbinsZ"], "axis":[1,2],"quantiles": [0.35,0.5],"unbinned_projections":True,
},
```

Figure axis transformation

xAxisTransform	yAxisTransform
lambda x: log(1+x)	lambda x,y: y/x

Custom javascipt function (javascript function as a text)

```
# defining custom java script function to query  (used later in varaible list)
aliasArray+=[{
        "name": "funCustom0",
        "variables": [i for i in variables if  "ustom" not in i ],
        "func":"funCustomForm0",
    },
    {
        "name": "funCustom1",
        "variables": [i for i in variables if   "ustom" not in i],
        "func":"funCustomForm1",
    },
    {
        "name": "funCustom2",
        "variables": [i for i in variables if   "ustom" not in i],
        "func":"funcCustomForm2",
    },
```

padrow

Select | Custom | Histograms | Transform | Legend | Markers

return IDC0_RMS<2

funCustomForm0

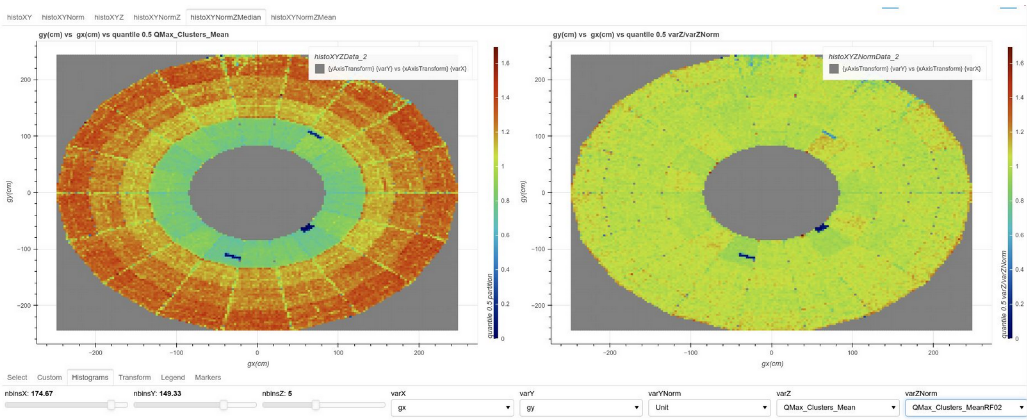
return IDC0_RMS/IDC0_Mean

Many different ways to define derived variables and functional composition. Dependency trees to resolve functional and data source dependencies.

Histogram declaration - calibration QA browser

Set of the 2D, 3D (ND) histograms declared

```
histoArray=[
    {
        "name": "histoXYData",
        "variables": ["varX","varY"],
        "nbins":["nbinsX","nbinsY"], "axis":[1],"quantiles": [0.35,0.5],"unbinned_projections":True,
    },
    {
        "name": "histoXYNormData",
        "variables": ["varX","varY/varYNorm"],
        "nbins":["nbinsX","nbinsY"], "axis":[1],"quantiles": [0.35,0.5],"unbinned_projections":True,
    },
    {
        "name": "histoXYZData",
        "variables": ["varX","varY","varZ"],
        "nbins":["nbinsX","nbinsY","nbinsZ"], "axis":[1,2],"quantiles": [0.35,0.5],"unbinned_projections":True,
    },
    {
        "name": "histoXYNormZData",
        "variables": ["varX","varY/varYNorm","varZ"],
        "nbins":["nbinsX","nbinsY","nbinsZ"], "axis":[1,2],"quantiles": [0.35,0.5],"unbinned_projections":True,
    },
    {
        "name": "histoXYZNormData",
        "variables": ["varX","varY","varZ/varZNorm"],
        "nbins":["nbinsX","nbinsY","nbinsZ"], "axis":[1,2],"quantiles": [0.35,0.5],"unbinned_projections":True,
    },
]
```



Anonymous function (used for example in ND histograms as weights or variables)

Parameterized histograms:

- Variables and wights could be any variable from data source (column, derived functions, anonymous function)
  - In the QA/calibration browser variables defined by user selecting (varX,varY, varZ)
  - Binning controlled by parameters (nbinsX, ...)
- Derived aggregated data exported as new data source
  - Declaring quantiles and projections
  - Projection could be binned (fast) and unbinned

Customizable Ndimensional histograms and projection. Example:

- X,y median profile of cluster charge map (left) and normalized to phi symmetric RF prediction

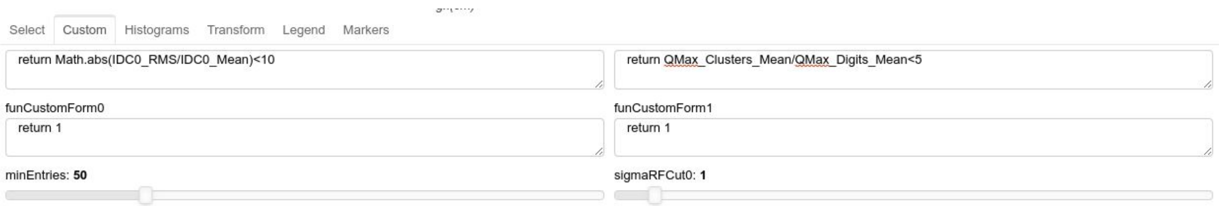
Parameter array - to control parameterized functions, selection, variable selection, graphics

Parameters:

- Histogram variable selection
- Histogram binning
- Custom axis transformation and normalization
- Custom function
- Variables - cut values for selection
- Predefined template parameter for graphic control

```
parameterArray = [
    {"name": "varX", "value":"padrow", "options":variables},
    {"name": "varY", "value":"padArea", "options":variables},
    {"name": "varYNorm", "value":"Unit", "options":variables},
    {"name": "varZ", "value":"partition", "options":variables},
    {"name": "varZNorm", "value":"Unit", "options":variables},
    {"name": "nbinsX", "value":30, "range":[10, 200]},
    {"name": "nbinsY", "value":30, "range":[10, 200]},
    {"name": "nbinsZ", "value":5, "range":[1,10]},
    #{"name": "sigmaNRel", "value":3.35, "range":[1,5]},
    #
    {"name": "exponentX", "value":1, "range":[-5, 5]},
    {'name': "xAxisTransform", "value":None, "options":[None, "sqrt", "lambda x: log(1+x)","lambda x: 1/sqrt(x)", "lambda x: x**exponentX","lambda x,y: x/y" ]},
    {'name': "yAxisTransform", "value":None, "options":[None, "sqrt", "lambda x: log(1+x)","lambda x: 1/sqrt(x)", "lambda x: x**exponentX","lambda x,y: y/x" ]},
    {'name': "zAxisTransform", "value":None, "options":[None, "sqrt", "lambda x: log(1+x)","lambda x: 1/sqrt(x)", "lambda x: x**exponentX" ]},
    # custom selection
    {'name': 'funCustomForm0', "value":"return 1"},
    {'name': 'funCustomForm1', "value":"return 1"},
    {'name': 'funCustomForm2', "value":"return 1"},
    # cut variables
    {"name": "minEntries", "value":50, "range":[5, 200]},
    {"name": "sigmaRFCut0", "value":1, "range":[0, 20]},
    {"name": "sigmaRFCutL", "value":1, "range":[0, 20]},
]

parameterArray.extend(figureParameters["legend"]['parameterArray'])
parameterArray.extend(figureParameters["markers"]['parameterArray'])
```



Functional composition, Histogramming, transformations, selection, graphics highly customizable, parameterizable by parameter array

Widgets declaration for data selection, parameters modification …

```
widgetParams=[
    ['multiSelect',["sector"],{"name":"sector"}],
    ['multiSelect',["partition"],{"name":"partition"}],
    ['multiSelect',["isEdgePad"],{"name":"isEdgePad"}],
    ['multiSelect',["A_Side"],{"name":"A_Side"}],
    #
    ['multiSelectBitmask',["QMax_Clusters_OK"],{"name":"QMax_Clusters_OK","mapping": {"Entries": 1, "RFLRatio0": 2, "RFLRatioL": 4}, "how":"all", "title": "QMaxCluster OK"}],
    ['multiSelectBitmask',["QMax_Digits_OK"],{"name":"QMax_Digits_OK","mapping": {"Entries": 1, "RFLRatio0": 2, "RFLRatioL": 4}, "how":"all", "title": "QMaxDigit OK"}],
    ['multiSelectBitmask',["SAC0_OK"],{"name":"SAC0_OK","mapping": {"Entries": 1, "RFLRatio0": 2, "RFLRatioL": 4}, "how":"all", "title": "SAC0 OK"}],
    ['multiSelectBitmask',["IDC0_OK"],{"name":"IDC0_OK","mapping": {"Entries": 1, "RFLRatio0": 2, "RFLRatioL": 4}, "how":"all", "title": "IDC0 OK"}],
    ['multiSelectBitmask',["IDC0_MeanOK"],{"name":"IDC0_MeanOK","mapping": {"RMS OK": 1, "lx OK": 2}, "how":"all", "title": "IDC0 Mean OK"}],
    #
    ['range',["sector"],{"name":"sector"}],
    ['range',["padrow"],{"name":"padrow"}],
    ['range',["lx"],{"name":"lx"}],
    ['range',["gx"],{"name":"gx"}],
    ['range',["gy"],{"name":"gy"}],
    ['textQuery', {"name": "customSelect0","value":"return Math.abs(IDC0_RMS/IDC0_Mean)<10"}],
    ['textQuery', {"name": "customSelect1","value":"return 1"}],
    ['textQuery', {"name": "customSelect2","value":"return 1"}],
    #
    ['text', ['funCustomForm0'], {"name": "funCustomForm0"}],
    ['text', ['funCustomForm1'], {"name": "funCustomForm1"}],
    ['text', ['funCustomForm2'], {"name": "funCustomForm2"}],
    #
    ['slider', ["minEntries"],{"name": "minEntries"}],
    ['slider', ["sigmaRFCut0"],{"name": "sigmaRFCut0"}],
    ['slider', ["sigmaRFCutL"],{"name": "sigmaRFCutL"}],
```

Widgets for data selection:

- **Select-** Sector, Size,partition
- **Bitmask -** calibration status
- **Range** -Position -lx,ly, gx,gy
- **texQuery -** custom selection as free text -javascipt
- **Text -** custom functions as free text - javascript
- **Slider (or spinner)** - cut values for parameterized selection

Widgets for parameter controls

```
    #
    ['select', ['varX'], {"name": "varX"}],
    ['select', ['varY'], {"name": "varY"}],
    ['select', ['varYNorm'], {"name": "varYNorm"}],
    ['select', ['varZ'], {"name": "varZ"}],
    ['select', ['varZNorm'], {"name": "varZNorm"}],
    ['slider', ['nbinsY'], {"name": "nbinsY"}],
    ['slider', ['nbinsX'], {"name": "nbinsX"}],
    ['slider', ['nbinsZ'], {"name": "nbinsZ"}],
    #
    ['spinner', ['exponentX'],{"name": "exponentX"}],
    #['spinner', ['sigmaNRel'],{"name": "sigmaNRel"}],
    ['select', ['yAxisTransform'], {"name": "yAxisTransform"}],
    ['select', ['xAxisTransform'], {"name": "xAxisTransform"}],
    ['select', ['zAxisTransform'], {"name": "zAxisTransform"}],
]

widgetParams.extend(figureParameters["legend"]["widgets"])
widgetParams.extend(figureParameters["markers"]["widgets"])
```

widgetsLayoutDesription declaration for widgets grouping

```
widgetLayoutDesc={
    "Select":[["sector","A_Side","partition","isEdgePad"],["QMax_Clusters_OK","QMax_Digits_OK","SAC0_OK","IDC0_OK","IDC0_MeanOK"],["padrow","lx","gx","gy"],],
    "Custom":[["customSelect0","customSelect1","customSelect2"],["funCustomForm0","funCustomForm1","funCustomForm2"],["minEntries","sigmaRFCut0","sigmaRFCutL"]],
    "Histograms":[["nbinsX","nbinsY", "nbinsZ", "varX","varY","varYNorm","varZ","varZNorm"], {'sizing_mode': 'scale_width'}],
    "Transform":[["exponentX","xAxisTransform", "yAxisTransform","zAxisTransform"],{'sizing_mode': 'scale_width'}],
    "Legend": figureParameters['legend']['widgetLayout'],
    "Markers":["markerSize"]
}
```

Widgets to control selection, custom selection, histogram parameters, transformation and graphics

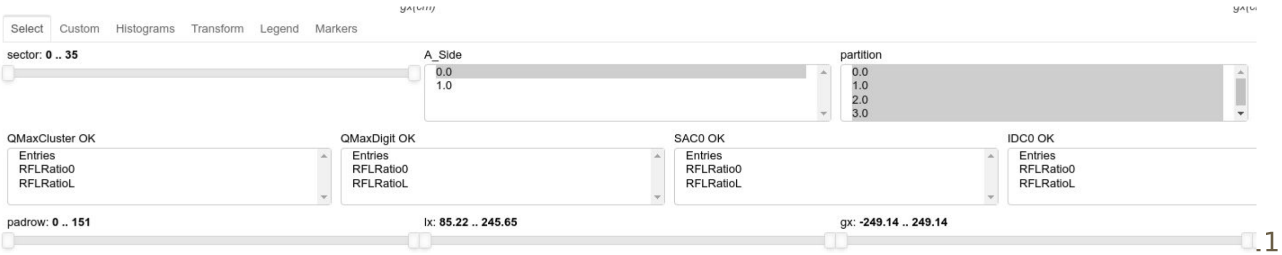


Figure and figure layout declaration

Figure array declaration example

```
figureArray = [
    #
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "bin_count", "source":"histoXYData"}],
    [["bin_center_1"], ["bin_count"], { "source":"histoXYData", "colorZvar": "bin_center_0"}],
    [["bin_center_0"], ["mean","quantile_1",], { "source":"histoXYData_1","errY":"std/sqrt(entries)"}],
    [["bin_center_0"], ["std"], { "source":"histoXYData_1","errY":"std/sqrt(entries)"}],
    #
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "bin_count", "source":"histoXYNormData"}],
    [["bin_center_1"], ["bin_count"], { "source":"histoXYNormData", "colorZvar": "bin_center_0"}],
    [["bin_center_0"], ["mean","quantile_1",], { "source":"histoXYNormData_1","errY":"std/sqrt(entries)"}],
    [["bin_center_0"], ["std"], { "source":"histoXYNormData_1","errY":"std/sqrt(entries)"}],
    #
    [["bin_center_0"], ["mean"], { "source":"histoXYZData_1","colorZvar":"bin_center_2","errY":"std/sqrt(entries)"}],
    [["bin_center_0"], ["quantile_0"], { "source":"histoXYZData_1","colorZvar":"bin_center_2","errY":"2*std/sqrt(entries)"}],
    [["bin_center_0"], ["quantile_1"], { "source":"histoXYZData_1","colorZvar":"bin_center_2","errY":"3*std/sqrt(entries)"}],
    [["bin_center_0"], ["std"], { "source":"histoXYZData_1","colorZvar":"bin_center_2","errY":"std/sqrt(entries)"}],
    #
    [["bin_center_0"], ["mean"], { "source":"histoXYNormZData_1","colorZvar":"bin_center_2","errY":"std/sqrt(entries)","yAxisTitle":"{varY}/{varYNorm}"}],
    [["bin_center_0"], ["quantile_0"], { "source":"histoXYNormZData_1","colorZvar":"bin_center_2","errY":"2*std/sqrt(entries)","yAxisTitle":"{varY}/{varYNorm}"}],
    [["bin_center_0"], ["quantile_1"], { "source":"histoXYNormZData_1","colorZvar":"bin_center_2","errY":"3*std/sqrt(entries)","yAxisTitle":"{varY}/{varYNorm}"}],
    [["bin_center_0"], ["std"], { "source":"histoXYNormZData_1","colorZvar":"bin_center_2","errY":"std/sqrt(entries)","yAxisTitle":"{varY}/{varYNorm}"}],
    # global XYZ profile - median
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "quantile_1", "source":"histoXYZData_2"}],
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "quantile_1", "source":"histoXYZNormData_2"}],
    # global XYZ profile - mean
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "mean", "source":"histoXYZData_2"}],
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "mean", "source":"histoXYZNormData_2"}],
    #
    figureGlobalOption
]
```

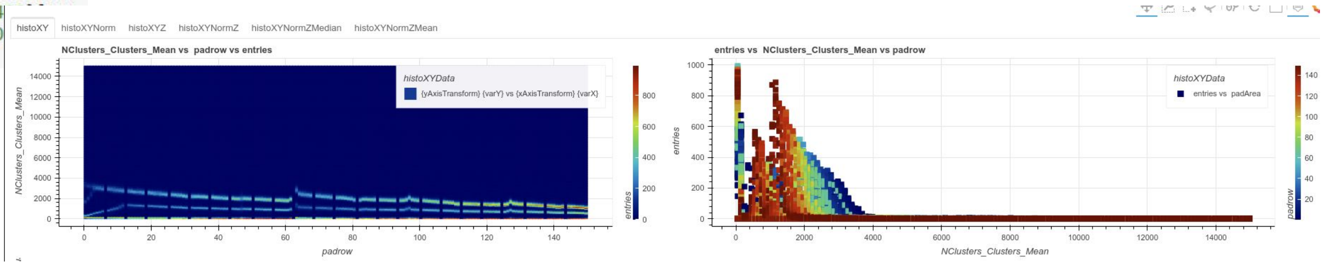
Figure array declaration example (figure id as number or using the names)

```
figureLayoutDesc={
    "histoXY":[[0,1],[2,3],{"plot_height":200}],
    "histoXYNorm":[[4,5],[6,7],{"plot_height":200}],
    "histoXYZ":[[8,9],[10,11],{"plot_height":200}],
    "histoXYNormZ":[[12,13],[14,15],{"plot_height":200}],
    "histoXYNormZMedian":[[16,17],{"plot_height":4
    "histoXYNormZMean":[[18,19],{"plot_height":400
}
```

Figure array declaration:

- [[<X array>],[<Yarray>],{options}]
  - E.g. Z color,source, titles
- Example:
  - 2D, 2D normalized,3D, 3D normalized, 3D heamaps
  - Heatmaps
- Data sources:
  - Unbinned data - original data
  - Histograms
  - Histogram projections
  - Any derived (anonymous) functions
  - Bin_count, bin_center,<stat>,<quantiles>, any function of columns in data souce

Different visulaization of the ND histogram content

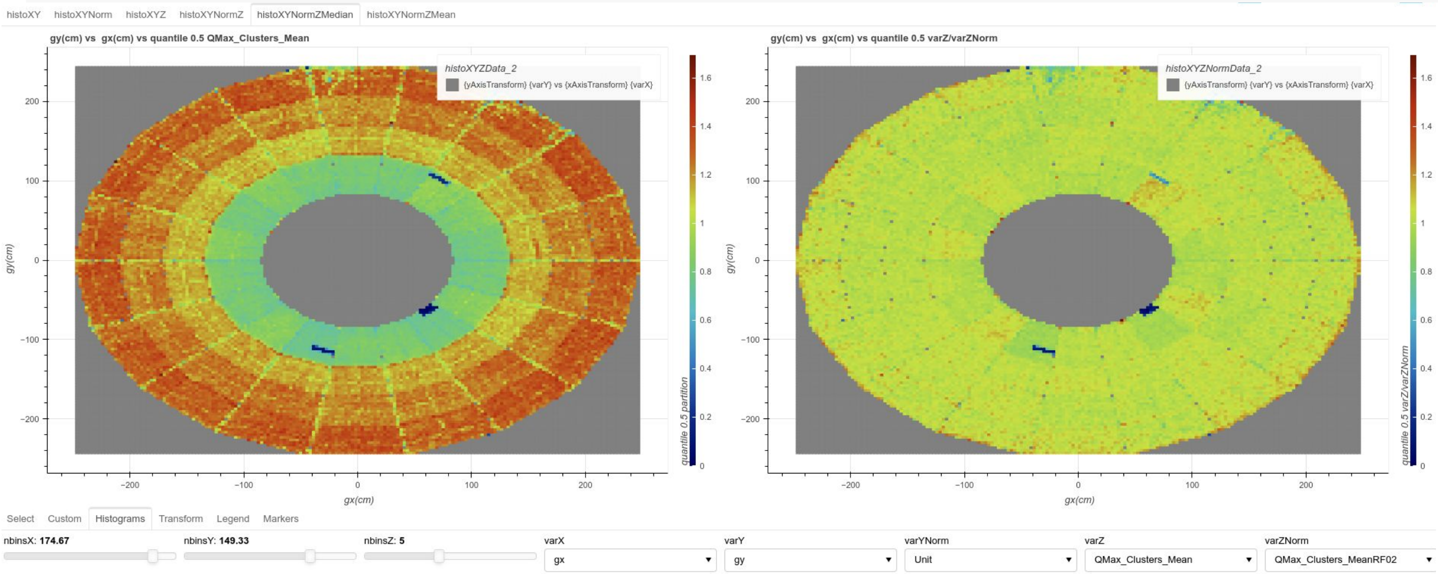


Calibration/QA/QC browser application

Dashboard declaration - as standalone application

```
output_file("QAQCcalPadSummary.html")
arrayCompression=arrayCompressionRelative10
dfSample=df.sample(frac=0.5).sort_index()
from RootInteractive.InteractiveDrawing.bokeh.palette import kBird256,kRainbow256

fig=bokehDrawSA.fromArray(dfSample, None, figureArray, widgetParams, layout=figureLayoutDesc, sizing_mode='scale_width', nPointRender=50000, widgetLayout=widgetLayoutDesc,
                          parameterArray=parameterArray, histogramArray=histoArray, rescaleColorMapper=True, arrayCompression=arrayCompression,aliasArray=aliasArray,palette=kRainbow256)
```

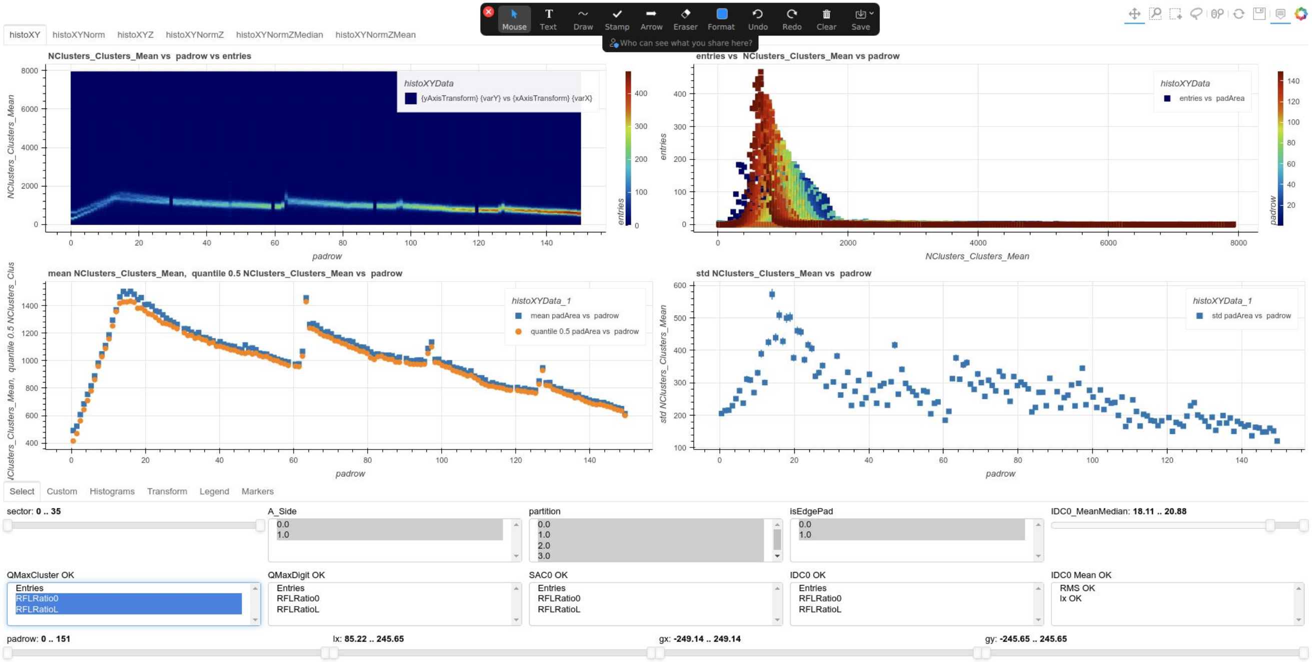


Creating application: using figure, widgets, histograms, alias, compression declaration. Dashboard either inside of **notebook** or as **standalone application**

Calibration/QA/QC browser application - layout

Figure tabs

- XY summary
- XY norm summary
- XYZ summary
- XYZ norm summary
- 3D projections



Widget tabs

- Default selection
- Custom selection and unction
- Histogram parameterization
- Transformation parameterization
- 3D projections

Highly customizable dashboard saved as standalone html application

RootInteractive - load input data - calibration/QA viewer example use case

Per pad calibration/properties:

- position/ids
- traceLength,padArea
- Pedestal/Noise
- Pulser
- Krypton gain
- … more aggregated info to be added …
  - space charge, IDC, time dependent gain correction

Per pad QC/QA:

- **NDigits occupancy**
- Qmax- digits (raw)
- **NClusters occupancy**
- Qmax, Qtot cluster (gain corrected)

Import variables and aliases from the tree to panda

```
%%time
varList=[
    "roc","ly","lx","gy","gx","row","pad","padArea", ## position
    "isEdgePad","partition",              ##
    "traceLength",                        ## trace length
    "GainMap",                            ## krypton gain map
    "Pedestals","Noise",                  ##
    "T0","Qtot",                          ## pulser properties
    "N_Digits","Q_Max_Digits",            ## digits occumancy and Qmax
    "N_Clusters","Q_Max","Q_Tot",         ## cluster Q-Max,Q_tot
    "fraction","expLambda",               ## ion tail prameters
    "Sigma_Time","Sigma_Pad",             ## should be mean sigma of cluster finder in pad and time direction
    "A_Side","C_Side"
]

dfScan=tree2Panda(tree,varList,"roc>=0",columnMask=[["_fElements",""]])
```

Tree Draw queries internally used:
Possibility to use functions, object (uproot not suitable)
variable lists using regular expression from branches, aliases, friends
column mask

Parameters imported from the ALICE O2 CCDB and QCDB

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html

RootInteractive - loading primary and derived data using - RDataFrame ↔ awkward (NEW)

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ab61ebe1148ebc7bd88c03f667aff0caa3b2b03e/JIRA/ATO-614/code/toydEdxSimul.C
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/611d5048fced221189c232cdc99eb953ab6bf470/JIRA/ATO-614/RDFtoAwkward.ipynb

Defining RDataFrame

```
ROOT::RDataFrame df(nTracks);
auto rdf = df.Define("qVector", "getQVector(160)")
                .Define("logqVector", "ROOT::VecOps::log(qVector)")
                .Define("qStd", "StdDev(qVector)")
                .Define("qMean", "Mean(qVector)")
                .Define("qlStd", "StdDev(logqVector)")
                .Define("qlMean", "Mean(logqVector)")
                .Define("qMedian", "TMath::Median(qVector.size(), qVector.data())")
                .Define("qlMedian", "TMath::Median(qVector.size(), logqVector.data())")
                .Define("qTrunc", "truncate(qVector);")
                .Define("logqTrunc", "ROOT::VecOps::log(qTrunc);");
```

Loading awkward array

In [7]:

```
%%time
array = ak.from_rdataframe(
        rdf,
        columns=(
            "logqTrunc",
            "logqVector",
            "qMean",
            "qMedian",
            "qStd",
            "qTrunc",
#            "qVector",
            "qlMean",
            "qlMedian",
            "qlStd",
        ),
    )
```

```
CPU times: user 1min 44s, sys: 884 ms, total: 1min 45s
Wall time: 10.2 s
```

dEdx optimization example

- Defining the data and derived function (C++) with native data representation
- loading the data → awkward array
- Execution scaling with number of cores (32 used in example)
- ML training/prediction → RDataFrame ()

Significant performance increase with parallel "RDataFrame ↔ awkward"
To be used extensively, e.g. in fastMCKalman (distortion simulation/correction) and in trackCombinator (V0,cascade,cosmic,loop finder, see tomorrow's presentation) prototyping use case studies

RootInteracitve - Derived variables: Regression/Normalization using RandomForest

Global regression Random Forest:

- localX, padArea, traceLength
- deepnes (14 bits)
- To define trivial properties

RF - Local regression:

- Sensitive to the properties in local neighborhood
- Robust local estimators exported
- RootInteractive local filters:
  - mean, median,std
- Many properties locally smooth (pad-gem distance, hit density) lead to smooth variation of derived variables (Gain, ion tail, occupancy)
- **Local & global based outlier tagging**

Random forest parameters

```
%%time
n_estimators=200
n_jobs=100
npoints=1000000
max_depthBase=14
max_samples=0.1
regressorBase = RandomForestRegressor(n_estimators =n_estimators,n_jobs=n_jobs,max_depth=max_depthBase,max_samples=max_samples)
regressorLocal = RandomForestRegressor(n_estimators =n_estimators,n_jobs=n_jobs,max_samples=max_samples)
```

Fit base and local propeties

- regressor with local X
- local filter regressor - mean,median,std filter

```
%%time
statDictionary={"mean":None,"median":None, "std":None}

varList=["lx","traceLength","padArea"]
varListLocal=["lx","ly","roc"]
vars=[
    "Noise","N_Digits","N_Clusters",
    "Q_Max", "Q_Tot", "GainMap","Q_Max_Digits",
    "fraction","expLambda"
]
for var in vars:
    # base regression limitted deep
    regressorBase.fit(dfScan[varList],dfScan[var])
    dfScan[f"{var}RF0"]= regressorBase.predict(dfScan[varList])
    dfScan[f"{var}RF0_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RF0"]
    # local regression
    regressorLocal.fit(dfScan[varListLocal],dfScan[var])
    statDictionaryOut=predictRFStatNew(regressorLocal,dfScan[varListLocal].astype('float32').to_numpy(),statDictionary,n_jobs)
    dfScan[f"{var}RFL"]= regressorLocal.predict(dfScan[varListLocal])
    dfScan[f"{var}RFL_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RFL"]
    dfScan[f"{var}RFL_Med"]=statDictionaryOut["median"]
    dfScan[f"{var}RFL_Std"]=statDictionaryOut["std"]
    dfScan[f"{var}RFLMed_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RFL_Med"]
    print(f"Fit {var}")
```

In Run1,Run2 QA based on local/global robust philter in fixed neighbourhood. Using RootInteractive ML wrappers, dynamically defined local neighborhood

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html

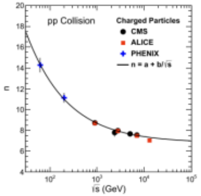
Representative Sampling/skimming

RootInteractive - Run3 - Data sampling/skimming

Run1/2 skimming triggers

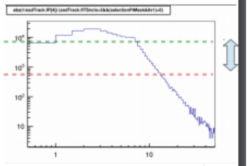
Data down-sampling to prepare representative sample flat in variable of interest

- Global Tsalis fits used to estimate particle production https://arxiv.org/pdf/1210.7464.pdf
- https://alice.its.cern.ch/jira/browse/ATO-465



Run1/2 topology horizontal down-sampling:

- **Charged (AliESDtrack) tracks down-sampling triggers**
  - flat pt trigger, flat q/pt trigger, MB
- **V0 trigger (Gamma, K_0, λ):**
  - flat pt trigger, flat q/pt trigger, MB
- **Nuclei (A>1)**
  - primaries
  - down-sampled secondaries
- **Cosmic track pairs:**
  - "random cosmics" for PID calibration
  - In Run3 → distortion characterization in regions not covered by ITS,TRD,TOF
- Others - under consideration (cascades, phi, D)
- **Event information - in Run2 not down-sampled -**
  - **small data volume (to be done for Run3)**



Data volume reduction determined by adjustable down-sampling factor

- Typically down-sampling for tracks O(10^-3) + additional derived information → data volume ~ O(10^-2)
- down-sampling factor adjusted base on statistics - e.g. in test production higher leveling
- In Run 3 ~ similar statistics to be stored - skimmed data volume can be reduced < 10 ^-3

17th March 2021 4

Run 1 and 2 PWGPP data skimming - example usage

RAA analysis and expert QA (in Run1)

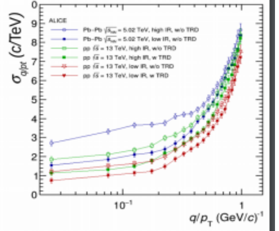
Almost all (my) reconstruction/PID debugging

- in case suitable information available

Tracking performance production parameterization

- see performance comparison web page http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/
- PassX/PassY , MC/data, PeriodX/PeriodY
- MC/data tuning/remapping
- Track matching/Efficiency/Inv.Mass/Material budget/Cross sections

$$\sigma_{p_T}/p_T = \sigma_{q/p_T} \times p_T$$



Reconstruction (TRD and pass2) commissioning/tuning

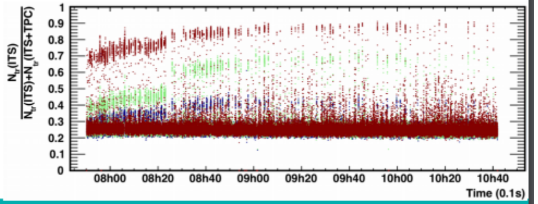
PID calibration and performance studies

- Pile-up correction, dEdx chi2

Event characteristic

- outliers and pile-up tagging

Time series for QA - outlier time interval tagging e.g. due space charge distortion fluctuation



17th March 2021 5

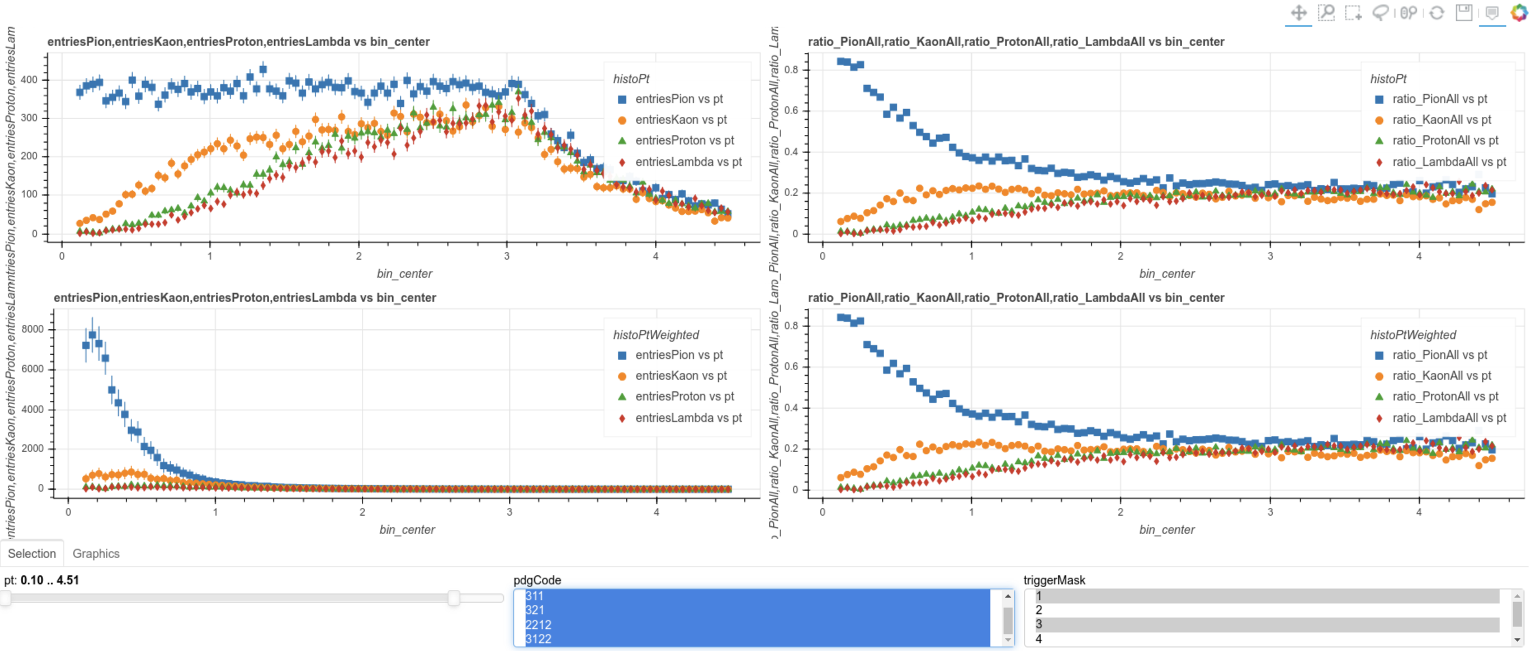
Run3 data to be sampled/skimmed in the similar way as Run1,2 (data down-sampled by factors 10^3)

- https://indico.cern.ch/event/1014566/contributions/4272119/attachments/2209987/3743263/ATO-465-DataSkimmingPerfCalPhysicsRun2Run3.pdf
- **small server instead of farm to analyze the data**

- Public node: https://alice-notes.web.cern.ch/node/1208

sampled spectra

re-weighted spectra



- Interactive ND histogramming and aggregation up to 10^7 points
- Using appropriate sampling interactive queries 10^9-10^10 representative points could be used
- Trigger optimization ongoing in dedicated fast MC studies
- Central sampling/skimming production to be run for the “ESD” and A02D data

https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164708/downsamplingTrigger.ipynb
https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164702/downsamplingTrigger.html
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/94435e925dd7b51f8753601f8ab2102587cf1702/JIRA/ATO-575/downsamplingTrigger.C#L27

Example: Sampled/skimmed efficiency and MC trees

MC true Efficiency and resolution trees for all particle species - fiducial volume determination - radial cut

- K+,K-, π+,π-
- kinks
- K0, γ,Λ,μ,π0
- Λ, Σ, Ξ, Ω
- Λc …

MC information down-sampled O(10^-3)

- **Run2 and Run3 version exist**
- Particle properties, MC track references, reconstructed information (track, V0, cascade)

Particle sampled/skimmed rough uniform distribution:

- rough **flat mass** down-sampling
- rough **flat pt** down-sampling
- exponential mass distribution assumed

Recursive down-sampling trigger:

- Particle sampled if the any of the mother particles in hierarchy sampled

Eff tree, track Tree, V0 tree, Cascade tree

Sampled data used to test different fit algorithm (resolution) and to test different selections (efficiency) e.g.

- Kalman Vertexer
- DCA selection
- Pointing selection

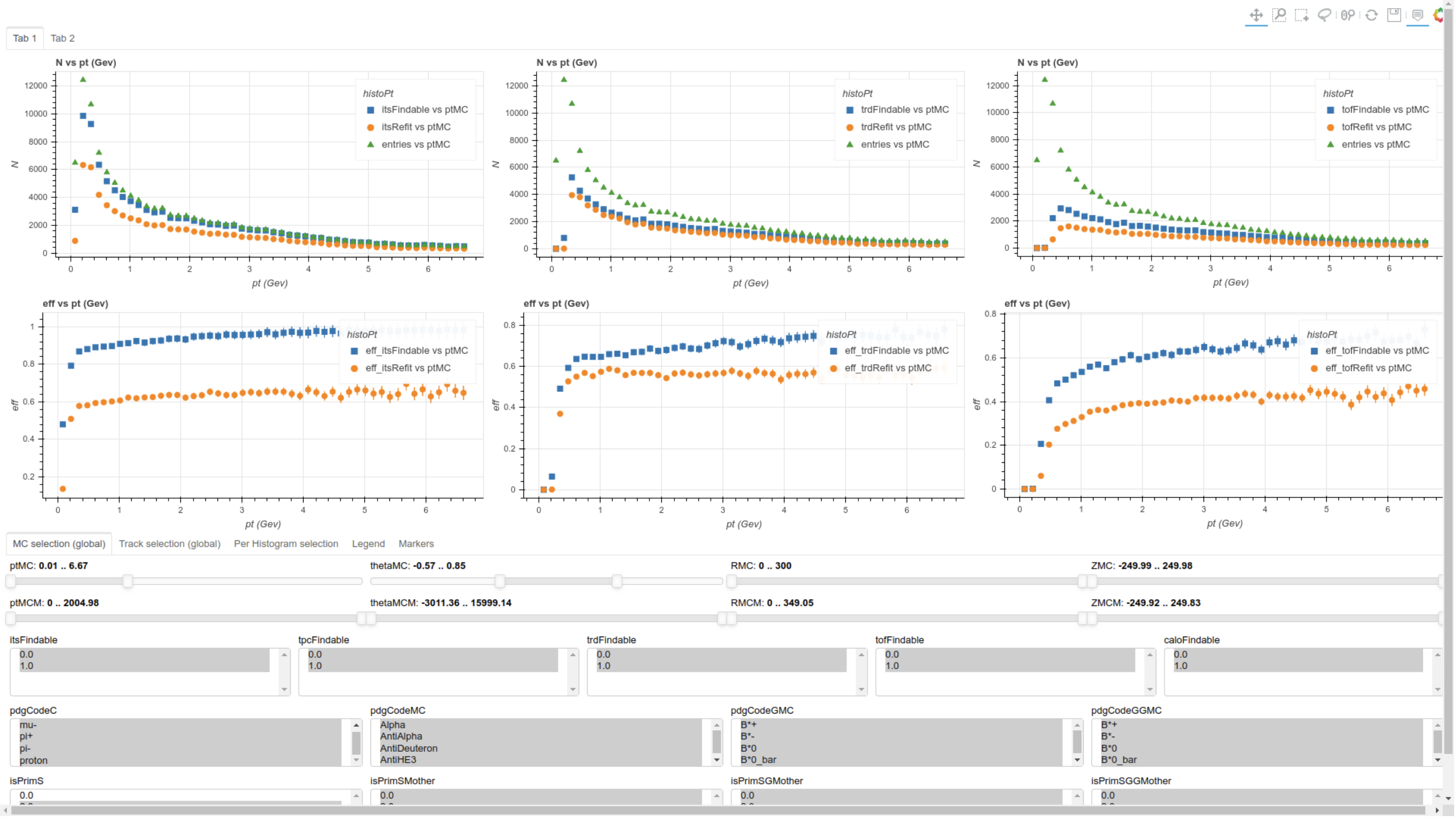
Saving computing resources. Enabling many iteration. Possibilities of interactive cases studies

Figure tabs ---------------►

- tab1 MC findable and reconstruction efficiency
- tab2 custom selection efficiency
- tab Findable
  - under preparation
- tab User custom
  - under preparation

Selection/widget and parameter tabs: --------►

- MC selection
- Global track selection
- Track selection for histogram
- Legend options
- Marker options



See configuration in Jupyter https://indico.cern.ch/event/1135398/#preview:4265612
https://indico.cern.ch/event/1135398/contributions/4950038/attachments/2474468/4282627/test_EffTrack_LHC16h8a.html

New tutorials to be prepared using simplified interface

- Expert version on 02.12.2022 - https://indico.cern.ch/event/1221198/
- Gallery, more example use cases will be added before public version om 16.12.2022

Installation recipe support for installation with Root, AliRoot, O2

- Recently problems with Python transitions.
- Installation recipes to be standartized
  https://github.com/miranov25/RootInteractive/blob/master/tutorial/README_WithROOTInstall.md#installing-only-as-virtual-environment-at-lxplus8

Analytical linear fits and convolution/deconvolution on client

Further development of the Machine learning part

- Parametric autoencoder, Linear regression forest …
- Aggregated statistics o client using autoencoders

Better support for the machine Learning on client → support for the ONNX functions on clients

RootInteractive conclusion

RootInteractive extensively used in many ALICE use cases for the multidimensional analysis

We plan to follow on simple example use cases, now mostly related the the TPC (calibration, simulation, QA) and global reconstruction/calibration (Run3,Run2 as a reference, Alice 3)

Pilot N dimensional physics analysis using sampled/skimmed data is in the queue

Backup

RootInteracitve - Derived variables: Regression/Normalization using RandomForest

Global regression Random Forest:

- localX, padArea, traceLength
- deepnes (14 bits)
- To define trivial properties

RF - Local regression:

- Sensitive to the properties in local neighborhood
- Robust local estimators exported
- RootInteractive local filters:
    - mean, median,std
- Many properties locally smooth (pad-gem distance, hit density) lead to smooth variation of derived variables (Gain, ion tail, occupancy)
- **Local & global based outlier tagging**

Random forest parameters

```
%%time
n_estimators=200
n_jobs=100
npoints=1000000
max_depthBase=14
max_samples=0.1
regressorBase = RandomForestRegressor(n_estimators =n_estimators,n_jobs=n_jobs,max_depth=max_depthBase,max_samples=max_samples)
regressorLocal = RandomForestRegressor(n_estimators =n_estimators,n_jobs=n_jobs,max_samples=max_samples)
```

Fit base and local propeties

- regressor with local X
- local filter regressor - mean,median,std filter

```
%%time
statDictionary={"mean":None,"median":None, "std":None}

varList=["lx","traceLength","padArea"]
varListLocal=["lx","ly","roc"]
vars=[
    "Noise","N_Digits","N_Clusters",
    "Q_Max", "Q_Tot", "GainMap","Q_Max_Digits",
    "fraction","expLambda"
]
for var in vars:
    # base regression limitted deep
    regressorBase.fit(dfScan[varList],dfScan[var])
    dfScan[f"{var}RF0"]= regressorBase.predict(dfScan[varList])
    dfScan[f"{var}RF0_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RF0"]
    # local regression
    regressorLocal.fit(dfScan[varListLocal],dfScan[var])
    statDictionaryOut=predictRFStatNew(regressorLocal,dfScan[varListLocal].astype('float32').to_numpy(),statDictionary,n_jobs)
    dfScan[f"{var}RFL"]= regressorLocal.predict(dfScan[varListLocal])
    dfScan[f"{var}RFL_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RFL"]
    dfScan[f"{var}RFL_Med"]=statDictionaryOut["median"]
    dfScan[f"{var}RFL_Std"]=statDictionaryOut["std"]
    dfScan[f"{var}RFLMed_Ratio"]=dfScan[f"{var}"]/dfScan[f"{var}RFL_Med"]
    print(f"Fit {var}")
```

In Run1,Run2 QA based on local/global robust philter in fixed neighbourhood. Using RootInteractive ML wrappers, dynamically defined local neighborhood

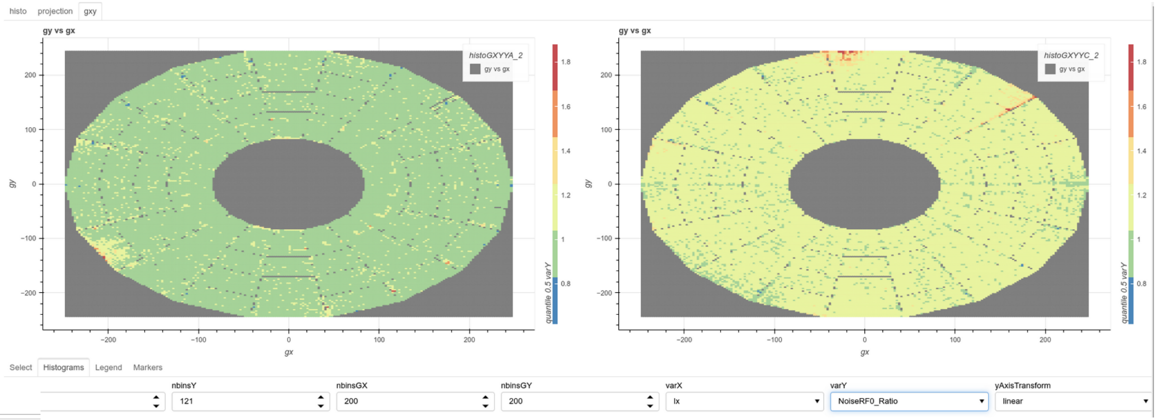
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html

RootInteractive ND histogramming and unbinned aggregation

Histogram array

- histogram user defined - X,Y
- number of bins user defined

```
histoArray=[
    {
        "name": "histoXY",
        "variables": ["varX", "varY"],
        "nbins":["nbinsX", "nbinsY"], "axis":[0,1],"quantiles": [.1, .5, .9],"unbinned_projections":True
    },
    #
    {"name": "histoX", "variables": ["varX"], "nbins":"nbinsX", "range": None},
    #
    {"name": "histoY", "variables": ["varY"], "nbins":"nbinsY", "range": None, },
    {
        "name": "histoGXYYA",
        "variables": ["gx", "gy","varY"],
        "nbins":["nbinsGX", "nbinsGY","nbinsY"], "axis":[2],"quantiles": [.5],"unbinned_projections":True, "weights":"A_Side"
    },
    {
        "name": "histoGXYYC",
        "variables": ["gx", "gy","varY"],
        "nbins":["nbinsGX", "nbinsGY","nbinsY"], "axis":[2],"quantiles": [.5],"unbinned_projections":True, "weights":"C_Side"
    }
]
```



Make parameters and widgets

```
variables=["lx","dPedestals","Noise","NoiseRF0","NoiseRF0_Ratio","N_Digits","N_DigitsRF0","N_DigitsRF0_Ratio","N_Clusters","N_ClustersRF0","N_ClustersRF0_Ratio",
           "Q_Max","Q_MaxRF0_Ratio","Q_Tot","Q_TotRF0_Ratio",
           "GainMap","GainMapRFL_Ratio","GainMapRF0_Ratio","GainMapRFL","Q_Max_Digits",
          "Sigma_Time","Sigma_Pad"
          ]


parameterArray = [
    {"name": "varX", "value":"Noise", "options": variables},
    {"name": "varY", "value":"N_Digits", "options": variables },
    {"name": "nbinsX", "value":30, "range":[5, 100]},
    {"name": "nbinsY", "value":30, "range":[5, 100]},
    #
    {"name": "nbinsGX", "value":50, "range":[30, 250]},
    {"name": "nbinsGY", "value":50, "range":[30, 250]},
    {"name": "yAxisTransform", "value":"linear", "options":["linear","sqrt","log"]},
]
```

Example:Median filter   Noise to expected noise(area/traceLength)

Create figure/application layout

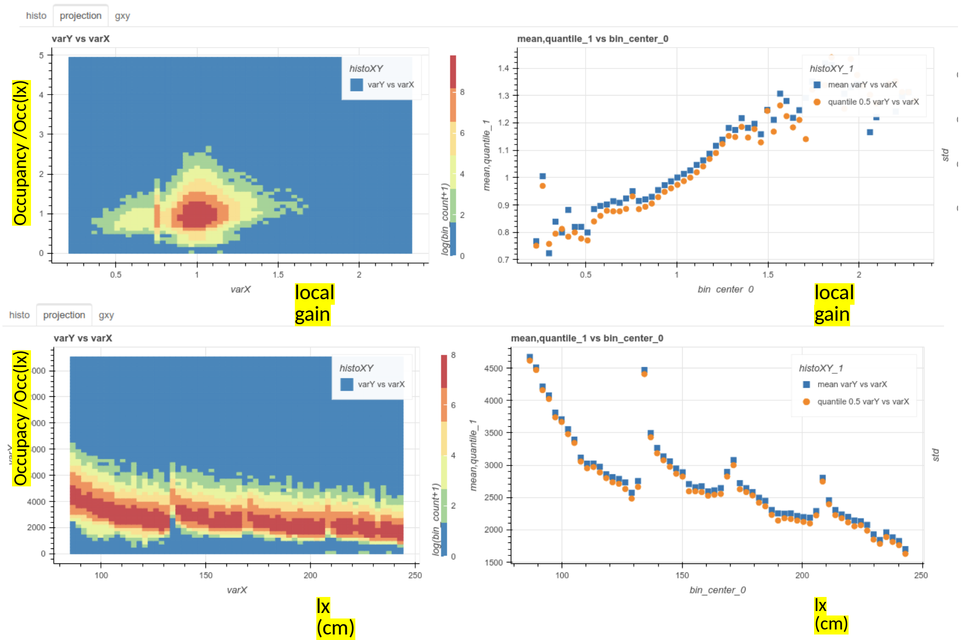
```
figureArray=[
    [["bin_center"],["entries"],{"source":"histoX","yAxisTitle":"N", "xAxisTitle":"varX", "errY": ["sqrt(entries)"]}],
    [["bin_center"],["entries"],{"source":"histoY","yAxisTitle":"N", "xAxisTitle":"varY", "errY": ["sqrt(entries)"]}],
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "log(bin_count+1)", "source":"histoXY"}],
    #
    [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "log(bin_count+1)", "source":"histoXY"}],
    [["bin_center_0"], ["mean","quantile_1"], { "source":"histoXY_1"}],
    [["bin_center_0"], ["std"], { "source":"histoXY_1"}],
    # global XY profile
     [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "quantile_0", "source":"histoGXYYA_2"}],
     [[("bin_bottom_0", "bin_top_0")], [("bin_bottom_1", "bin_top_1")], {"colorZvar": "quantile_0", "source":"histoGXYYC_2"}],

    figureGlobalOption
]
figureLayoutDesc={
    "histo":[[0,1,2],{"plot_height":350}],
    "projection":[[3,4,5],{"plot_height":350}],
    "gxy":[[6,7],{"plot_height":350}],
}
```

- **Creating application in your browser**
- User defined variables to histogram
- User defined binning/ranges

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html

RootInteractive example data interpretation - digit occupancy vs gain interpretation



Higher gain - bigger occupancy

Occupancy steps in segment boundaries (different pad areas)

Q/Gain in regions different

Data volume - occupancy observation:

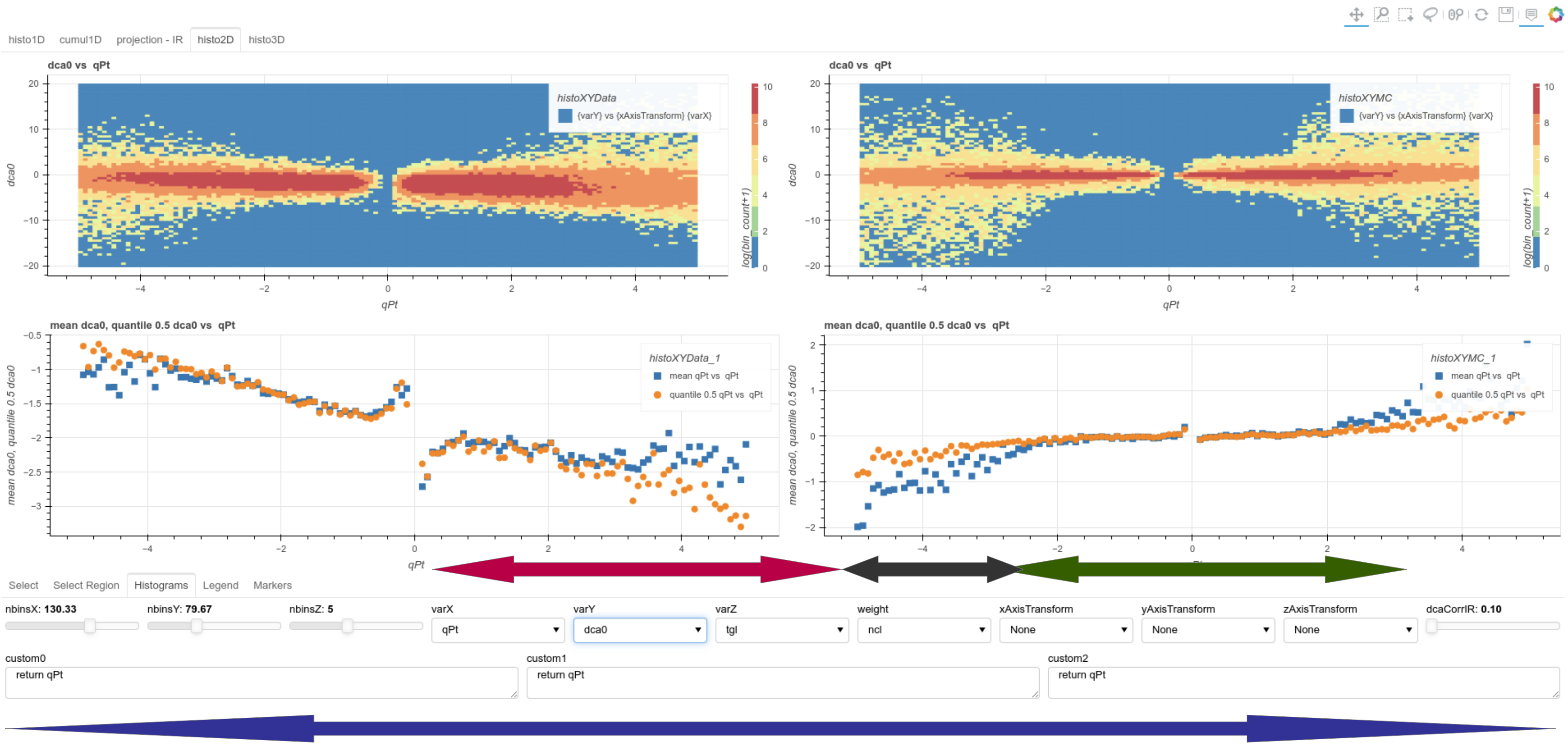
Digits occupancy linearly depends on the Gas gain (as it should in small local neighborhood) - **slope ~ 0.5**
Cluster occupancy less sensitive
Occupancy steps - **gain at higher ROCS (not in IROC) could be reduced**
Gaussian noise threshold properly adjusted
Noise problems only in well localized region with non gaussian induced error **at some sector boundaries**

Calibration observation:

Gain correction over-correcting
Preferable to make **gain correction only later and keep raw Q in clusters**

Optimally, complete information is provided to support the statements.
Dashboards, presentation with snapshots of dashboards, +link to films with statements.
Experts, collaborators and students can replay with customised selection

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/JIRA/ATO-611/tpcCalPadQA.ipynb
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319503/ATO-611-CalibPadViewer.pdf
https://indico.cern.ch/event/1126855/contributions/5057855/attachments/2511871/4319478/calibPad.html
https://indico.cern.ch/event/1135398/contributions/4764024/subcontributions/370740/attachments/2402507/4109039/CMITSimulationsGEMTPC.mp4

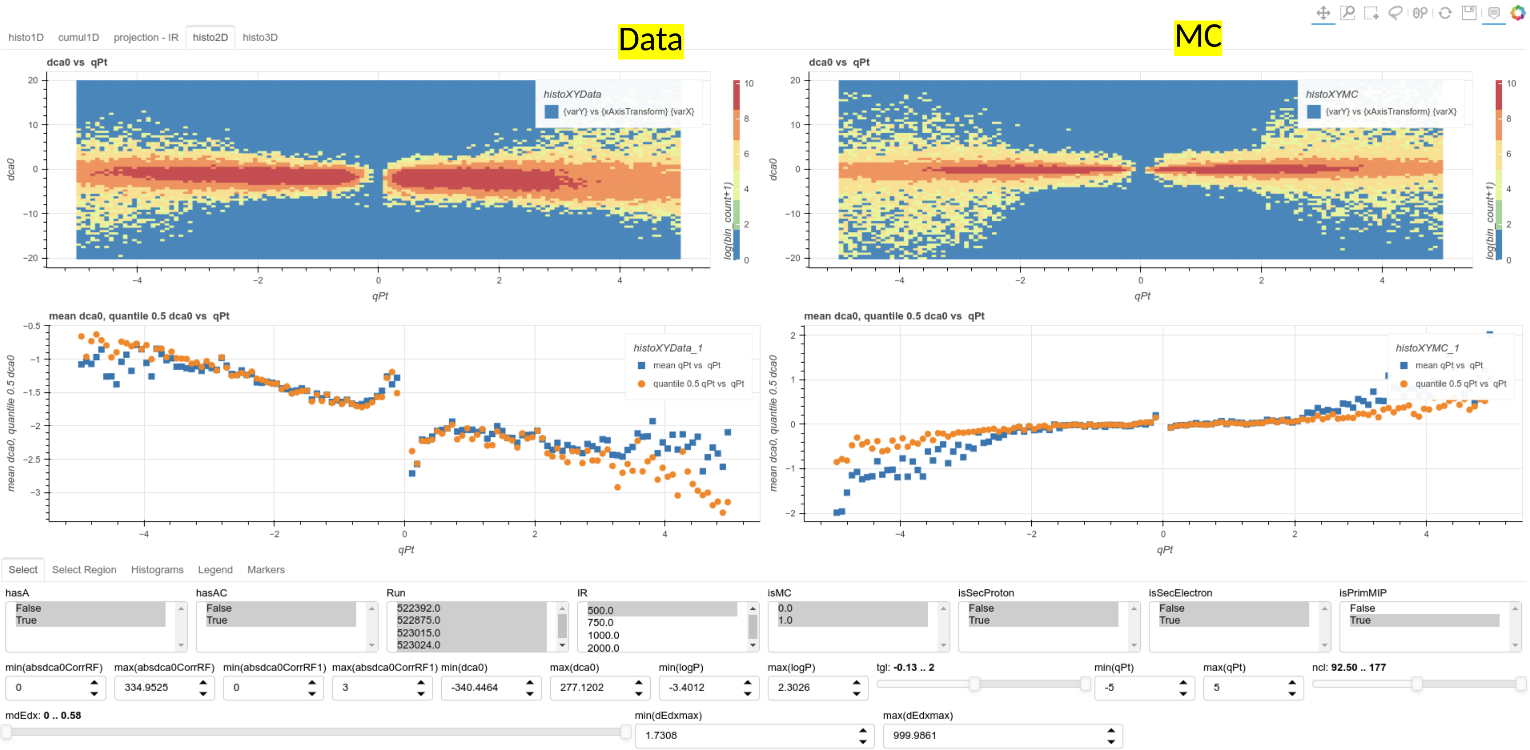


Customazible variables (multiselect) for the variables and histogram weights (multiselect)

Custom user defined functions as a text (all variables could be used), the same to be done for **weigthts**

User defined **axis transformation (none, sqrt, log ….) (points, errors and intervals)**

Widgets for custom selection/function definition/weight and histogram parameterization



Example dashboard dca bias due radial and rφ distortion
Low IR data

Select and multiselect, bitmask (&|) (e.g. for track selection, event selection,cluster selection),
sliders, range, custom selection, function
selection are toggle-able (could be enabled/disabled)
spinner, spinner range, axis range (min,max,nbins under preparation)

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/master/JIRA/ATO-609/trackClusterDumpDraw.ipynb

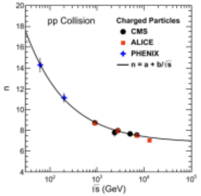
Representative Sampling/skimming

RootInteractive - Run3 - Data sampling/skimming

Run1/2 skimming triggers

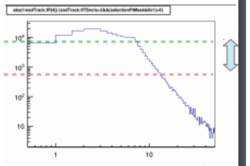
Data down-sampling to prepare representative sample flat in variable of interest

- Global Tsalis fits used to estimate particle production https://arxiv.org/pdf/1210.7464.pdf
- https://alice.its.cern.ch/jira/browse/ATO-465



Run1/2 topology horizontal down-sampling:

- **Charged (AliESDtrack) tracks down-sampling triggers**
  - flat pt trigger, flat q/pt trigger, MB
- **V0 trigger (Gamma, K_0, λ):**
  - flat pt trigger, flat q/pt trigger, MB
- **Nuclei (A>1)**
  - primaries
  - down-sampled secondaries
- **Cosmic track pairs:**
  - "random cosmics" for PID calibration
  - In Run3 → distortion characterization in regions not covered by ITS,TRD,TOF
- Others - under consideration (cascades, phi, D)
- **Event information - in Run2 not down-sampled -**
  - **small data volume (to be done for Run3)**



Data volume reduction determined by adjustable down-sampling factor

- Typically down-sampling for tracks O(10^-3) + additional derived information → data volume ~ O(10^-2)
- down-sampling factor adjusted base on statistics - e.g. in test production higher leveling
- In Run 3 ~ similar statistics to be stored - skimmed data volume can be reduced < 10 ^-3

17th March 2021 4

Run 1 and 2 PWGPP data skimming - example usage

RAA analysis and expert QA (in Run1)

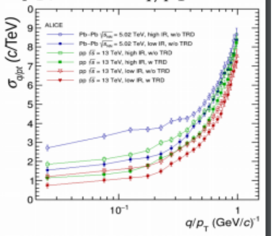
Almost all (my) reconstruction/PID debugging

- in case suitable information available

Tracking performance production parameterization

- see performance comparison web page http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/
- PassX/PassY , MC/data, PeriodX/PeriodY
- MC/data tuning/remapping
- Track matching/Efficiency/Inv.Mass/Material budget/Cross sections

$$\sigma_{p_T}/p_T = \sigma_{q/p_T} \times p_T$$



Reconstruction (TRD and pass2) commissioning/tuning

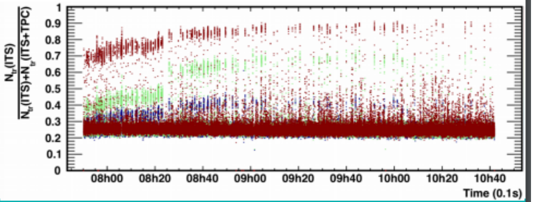
PID calibration and performance studies

- Pile-up correction, dEdx chi2

Event characteristic

- outliers and pile-up tagging

Time series for QA - outlier time interval tagging e.g. due space charge distortion fluctuation



17th March 2021 5

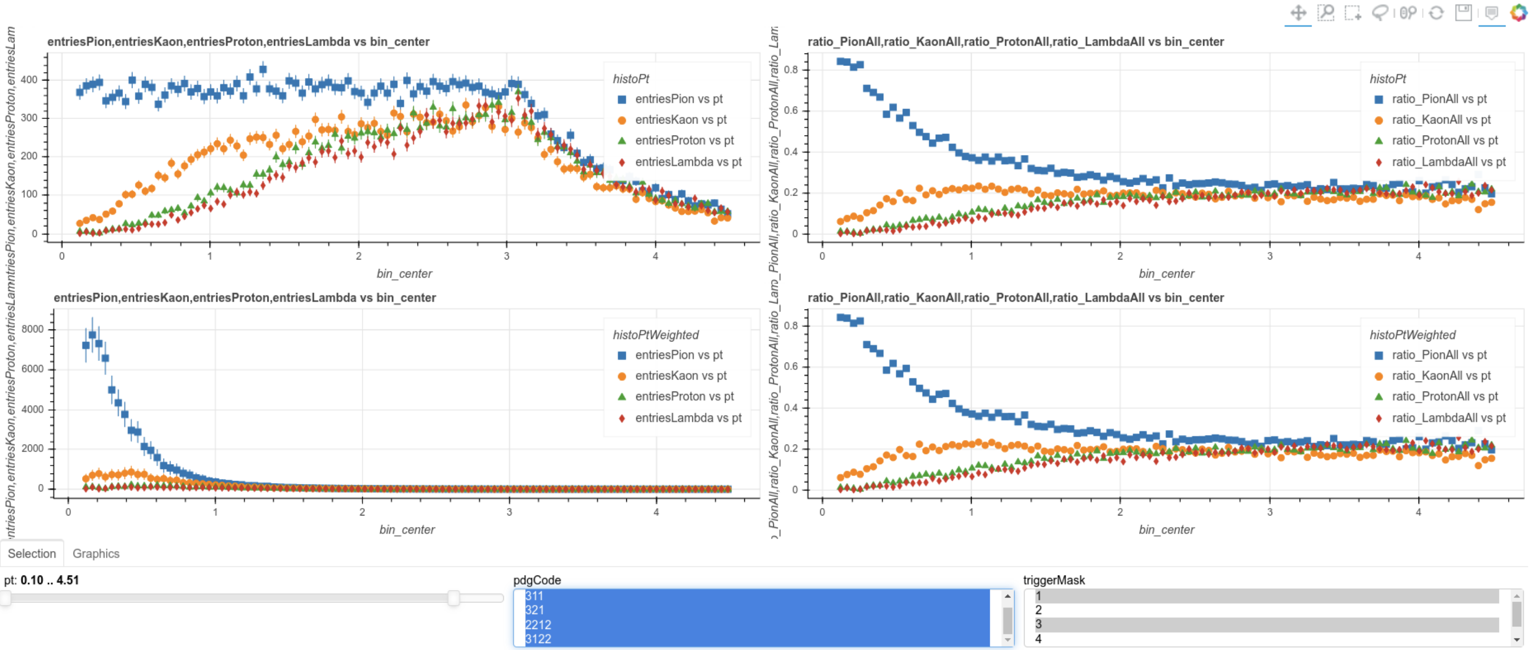
Run3 data to be sampled/skimmed in the similar way as Run1,2 (data down-sampled by factors 10^3)

- https://indico.cern.ch/event/1014566/contributions/4272119/attachments/2209987/3743263/ATO-465-DataSkimmingPerfCalPhysicsRun2Run3.pdf
- **small server instead of farm to analyze the data**

- Public node: https://alice-notes.web.cern.ch/node/1208

sampled spectra

re-weighted spectra



- Interactive ND histogramming and aggregation up to 10^7 points
- Using appropriate sampling interactive queries 10^9-10^10 representative points could be used
- Trigger optimization ongoing in dedicated fast MC studies
- Central sampling/skimming production to be run for the “ESD” and A02D data

https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164708/downsamplingTrigger.ipynb
https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164702/downsamplingTrigger.html
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/94435e925dd7b51f8753601f8ab2102587cf1702/JIRA/ATO-575/downsamplingTrigger.C#L27

Example: Sampled/skimmed efficiency and MC trees

MC true Efficiency and resolution trees for all particle species - fiducial volume determination - radial cut

- K+,K-, π+,π-
- kinks
- K0, γ,Λ,μ,π0
- Λ, Σ, Ξ, Ω
- Λc …

MC information down-sampled O(10^-3)

- **Run2 and Run3 version exist**
- Particle properties, MC track references, reconstructed information (track, V0, cascade)

Particle sampled/skimmed rough uniform distribution:

- rough **flat mass** down-sampling
- rough **flat pt** down-sampling
- exponential mass distribution assumed

Recursive down-sampling trigger:

- Particle sampled if the any of the mother particles in hierarchy sampled

Eff tree, track Tree, V0 tree, Cascade tree

Sampled data used to test different fit algorithm (resolution) and to test different selections (efficiency) e.g.

- Kalman Vertexer
- DCA selection
- Pointing selection

Saving computing resources. Enabling many iteration. Possibilities of interactive cases studies

RootInteractive development

Functions and functional composition on client. Simplified version of user interface developed recently.

- Standard javascript functions

New tutorials to be prepared using simplified interface

Code working “well” with AliRoot, O2, used for some ALICE3 prototypes

- Recently problems with Python transitions. Installation recipes to be standartized

Further development of the Machine learning part

- Parametric autoencoder, Linear regression forest …

Better support for the machine Learning on client → support for the ONNX functions on clients

RootInteractive conclusion

RootInteractive extensively used in many ALICE use cases for the multidimensional analysis

We plan to follow on simple example use cases, now mostly related the the TPC (calibration, simulation, QA) and global reconstruction/calibration (Run3,Run2 as a reference, Alice 3)

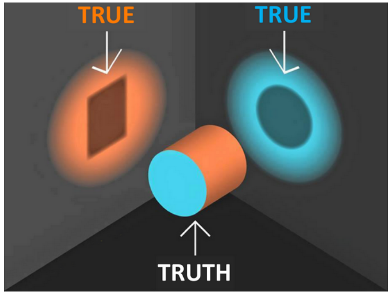
Pilot N dimensional physics analysis using sampled/skimmed data is in the queue

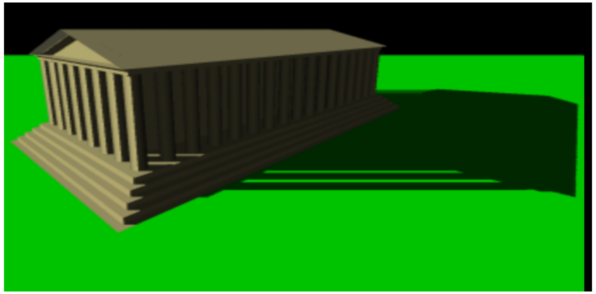
Backup

NDimensional pipeline and functional composition

- NDimensional pipeline code originally in C++ (Root/AliRoot)
- Visualization and on client aggregation Python/TrueScript (RootInteractive)
- Machine learning wrappers Python
- PyRoot used to be able to use Root and RootInteractive together

ND analysis vs shadow projections





https://www.youtube.com/watch?v=a7LCTT7HKzc

$$\sigma_{\vec{A}\ominus\vec{A}_{ref}} \leq \sigma_{\vec{A}}(+)\sigma_{\vec{A}_{ref}}$$

Object and reference objects should be compared optimally in the relevant ND space.

Shadow projection → Assumptions, imagination and rhetorical art in describing data needed

Example - QA alarms/statements to be based on invariance or on normalized data - e.g. the difference between the object and the reference object

- After projection impossible
- In many typical cases variance $\sigma_{A\text{-}Aref}$ is very often smaller by orders of magnitude

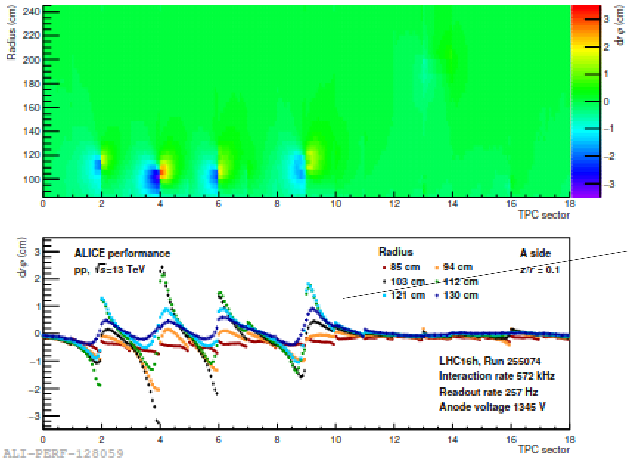
$$\sigma_{\vec{A}-\vec{A}_{ref}} < \sigma_{\vec{A}} (+) \sigma_{\vec{A}_{ref}}$$

2015 data crisis - Distortion in the TPC O (1-4 cm - Rate dependent)

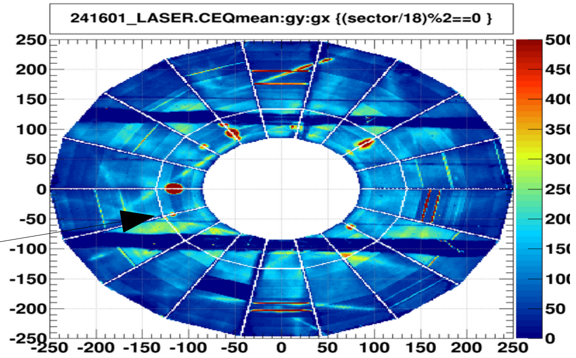
Center of gravity closer to sector gap **(inside)** than inner edge of affected chamber

Data normalized to reference data set (high rate/low IR rate data)

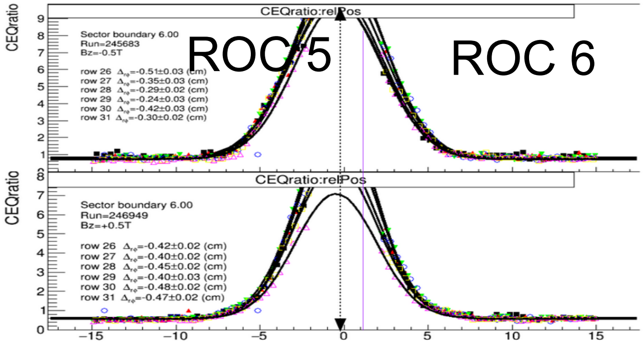
- fit indicates position of the space charge → distortion origin in gap inside
- for MB and TB - **result not yet convincing** for hardware intervention - **higher precision needed**



TPC distortion 2015



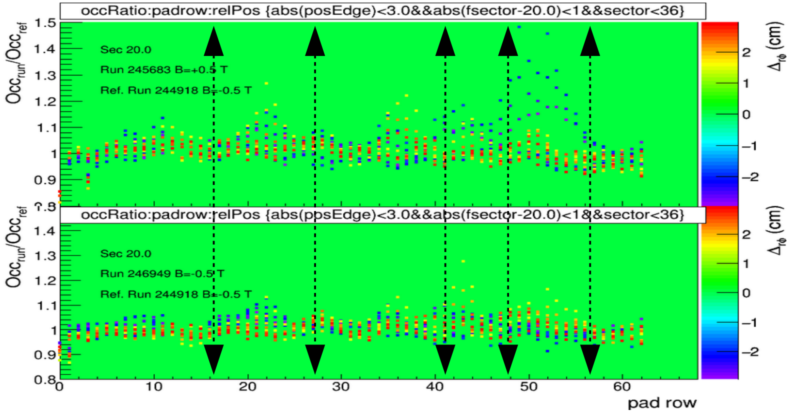
Laser calibration -Ion deposited on CE decrease work function → Increased emission of electrons during laser shots

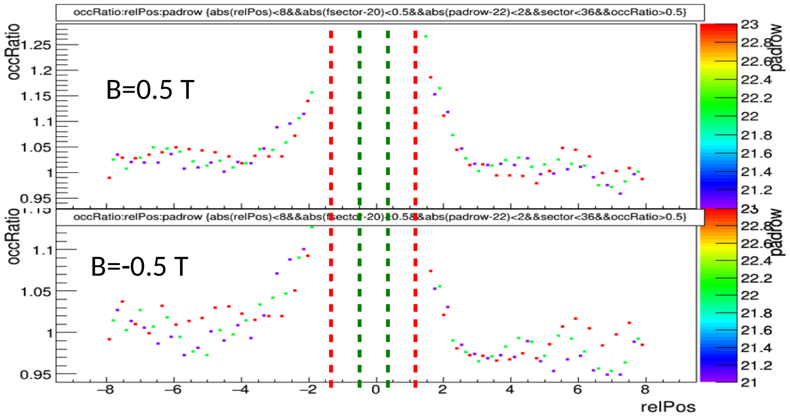


Distance to sector boundary (cm)

Example data normalization - Hunting for distortion origin (2015)

$$\sigma_{\vec{A} \ominus \vec{A}_{ref}} \leq \sigma_{\vec{A}} (+) \sigma_{\vec{A}_{ref}}$$





Analytical model - derivative of E field due line charge:

$$\frac{N_{Cl}(IR)}{N_{Cl}(IR=0)} = \frac{(w+(\Delta_{r\phi}(r_\phi+w/2)-\Delta_{r\phi}(r_\phi-w/2)))}{w}$$

$$R = \left(\frac{Occ}{< Occ_{ROC} >}\right)_{IR} / \left(\frac{Occ}{< Occ_{ROC} >}\right)_{IR=0}$$

$\overline{Z} \approx 125cm$ $\Delta r\phi$ (cm)

Conclusion: **Distortion origin in the gap between sectors - No doubts → Hardware intervention**

Increase in occupancy near the hot spot region due to space charge distortion
Very precise measurement of the origin of the distortion - **measurement of the derivative of the distortion** with **sub-pad granularity**.

Without proper normalization to reference (double ratio) effect was invisible →

Wrong concussion done by students in first analysis

$$\sigma_{\vec{A} \ominus \vec{A}_{ref}} \leq \sigma_{\vec{A}} (+) \sigma_{\vec{A}_{ref}}$$

Data should be compared with reference model/data

- RMS spread is much smaller (see ALICE performance example in next slide)

Invariance/symmetries in N dimensions (A ref model vector):

- in-variance in time (using e.g. reference/average run)
- in-variance in space (e.g. rotation, mirror symmetry)
- data - physical model
- TPC: A side/C side, B field symmetry
- smoothness resp. local smoothness

MC-Data comparison - should be done in N dimension not on projections

Aggreagation/projections of normalized data in NDimensions

Projections problems (hidden variables):

- **Information loss. Intrinsic spread of variable vectors A and A ref is usually significantly bigger than spread of $A\text{-}A_{ref}$**
  - noise map, DCA bias, resolution maps, occupancy maps, sigma invariant mass maps …. as function of 1/pt, θ, occupancy, dEdx
- **Projected vector A depends on the actual distribution of hidden variable**
  - Sometimes misleading results
  - Non trivial interpretation of projected observation

N-Dimensional pipeline & RootInteracive update and plans

- RootInteractive tutorial: https://indico.cern.ch/event/1135398/

ATO-563: Machine learning wrappers

- Learning to Discover workshop: https://indico.ijclab.in2p3.fr/event/5999/timetable/#21-alice-non-parametric-and-pa

ATO-575: uniform data sampling/skimming for Run3

PWGPP-613, fastMCKalman for the reconstuction performance/derivative parameterization

- PWGPP-722 fastMCKalman for the calculation of the track/vertex parameter numerical derivative in respect to distortion
- https://indico.cern.ch/event/1124104/contributions/4718877/attachments/2383900/4073706/PWGPP-613fastMCKalman_03022022.pdf

PWGPP-643 event shape estimators and parameteric autoencoder

- https://indico.cern.ch/event/1124104/contributions/4718877/attachments/2383900/4083301/PWGPP-643-eventProperties.pdf

Software description

NDimensional pipeline code originally in C++ (Root/AliRoot)

- libStat in AliRoot
- as a standalone Root library currently in the **fastMCKalman library**

RootInteractive - visualization and on data aggregation **Python/TrueScript/Bokeh**

- PyRoot used to be able to use Root libraries and RootInteractive together
- Fully independent of other ALICE software → used for Run2 and new Run3 studies
- Standalone client application -

Machine learning wrappers Python

- some wrapper for sklearn, tensorFlow (reducible, irreducible error)
- Work in progress:
  - generalization of the reducible and irreducible errors (PDF , Wrapper for auto-encoders and parametric auto-encoders)
  - see e.g. Distortion calibration presentation

Root/TTree interface wrappers:

- aliTreePlayer using old ROOT functionality (possibility to use C++ interface)
- RDataFrame ↔ uproot awkward - Work in progress https://github.com/scikit-hep/awkward-1.0/pull/1295

http://docs.bokeh.org/en/latest/ *Bokeh is a Python library for creating interactive visualizations for modern web browsers. It helps you build beautiful graphics, ranging from simple plots to complex dashboards with streaming datasets. With Bokeh, you can create JavaScript-powered visualizations without writing any JavaScript yourself.*

https://github.com/scikit-hep/awkward-1.0/pull/1295

https://github.com/scikit-hep/awkward-1.0/issues/588

jpivarski commented on Dec 10, 2020 Member

This issue is to collect my thoughts about how RDataFrame integration could be done. Such a thing would be useful because physicists could then mix analyses using Awkward Array, Numba, *and* ROOT C++ without leaving their environment. The benefits compound:

1. Data that are too complex to read from Uproot (efficiently or at all) can be loaded using `MakeRootDataFrame` and dumped into an Awkward Array.
2. Arbitrarily complex Awkward Arrays can be written to ROOT files by dumping the Awkward Arrays into an RDataFrame and taking a `Snapshot`.
3. Users can use ROOT C++ functions in an otherwise Awkward analysis at full speed. ("Full" in quotes; there is a conversion penalty, but it's compiled code, not so bad.)

Special wrappers for the ROOT input trees (aliTreePlayer), based on the old ROOT tree iterface, to allow the use of data and C++ functions without leaving the environment (many use cases in the agenda) - **current uproot not sufficient.**

In contact with scikit-hep in anticipation of RDataFrame ↔ awkward interface.

- https://github.com/scikit-hep/awkward-1.0/pull/1295
- awkward → RDataFrame implemented (26.04.2022)

Data skimming/representative samples and triggers

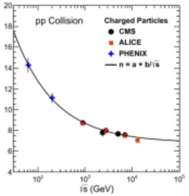
- To enable rapid development/feedback loop/interactivity, special representative data samples are usually used.

Run3 - Data sampling/skimming

Run1/2 skimming triggers

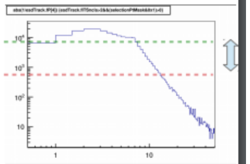
Data down-sampling to prepare representative sample flat in variable of interest

- Global Tsalis fits used to estimate particle production https://arxiv.org/pdf/1210.7464.pdf
- https://alice.its.cern.ch/jira/browse/ATO-465



Run1/2 topology horizontal down-sampling:

- **Charged (AliESDtrack) tracks down-sampling triggers**
  - flat pt trigger, flat q/pt trigger, MB
- **V0 trigger (Gamma, K_0, λ):**
  - flat pt trigger, flat q/pt trigger, MB
- **Nuclei (A>1)**
  - primaries
  - down-sampled secondaries
- **Cosmic track pairs:**
  - "random cosmics" for PID calibration
  - In Run3 → distortion characterization in regions not covered by ITS,TRD,TOF
- Others - under consideration (cascades, phi, D)
- **Event information - in Run2 not down-sampled -**
  - **small data volume (to be done for Run3)**



Data volume reduction determined by adjustable down-sampling factor

- Typically down-sampling for tracks O(10^-3) + additional derived information → data volume ~ O(10^-2)
- down-sampling factor adjusted base on statistics - e.g. in test production higher leveling
- In Run 3 ~ similar statistics to be stored - skimmed data volume can be reduced < 10 ^-3

17th March 2021 4

Run 1 and 2 PWGPP data skimming - example usage

RAA analysis and expert QA (in Run1)

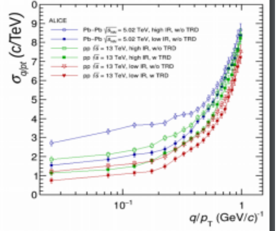
Almost all (my) reconstruction/PID debugging

- in case suitable information available

Tracking performance production parameterization

- see performance comparison web page http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/
- PassX/PassY , MC/data, PeriodX/PeriodY
- MC/data tuning/remapping
- Track matching/Efficiency/Inv.Mass/Material budget/Cross sections

$\sigma_{p_T}/p_T = \sigma_{q/p_T} \times p_T$



Reconstruction (TRD and pass2) commissioning/tuning

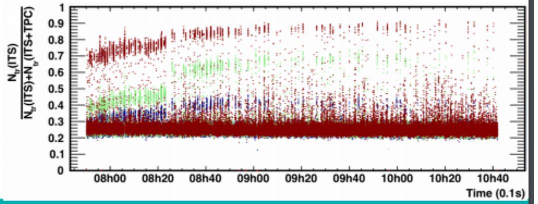
PID calibration and performance studies

- Pile-up correction, dEdx chi2

Event characteristic

- outliers and pile-up tagging

Time series for QA - outlier time interval tagging e.g. due space charge distortion fluctuation



17th March 2021 5

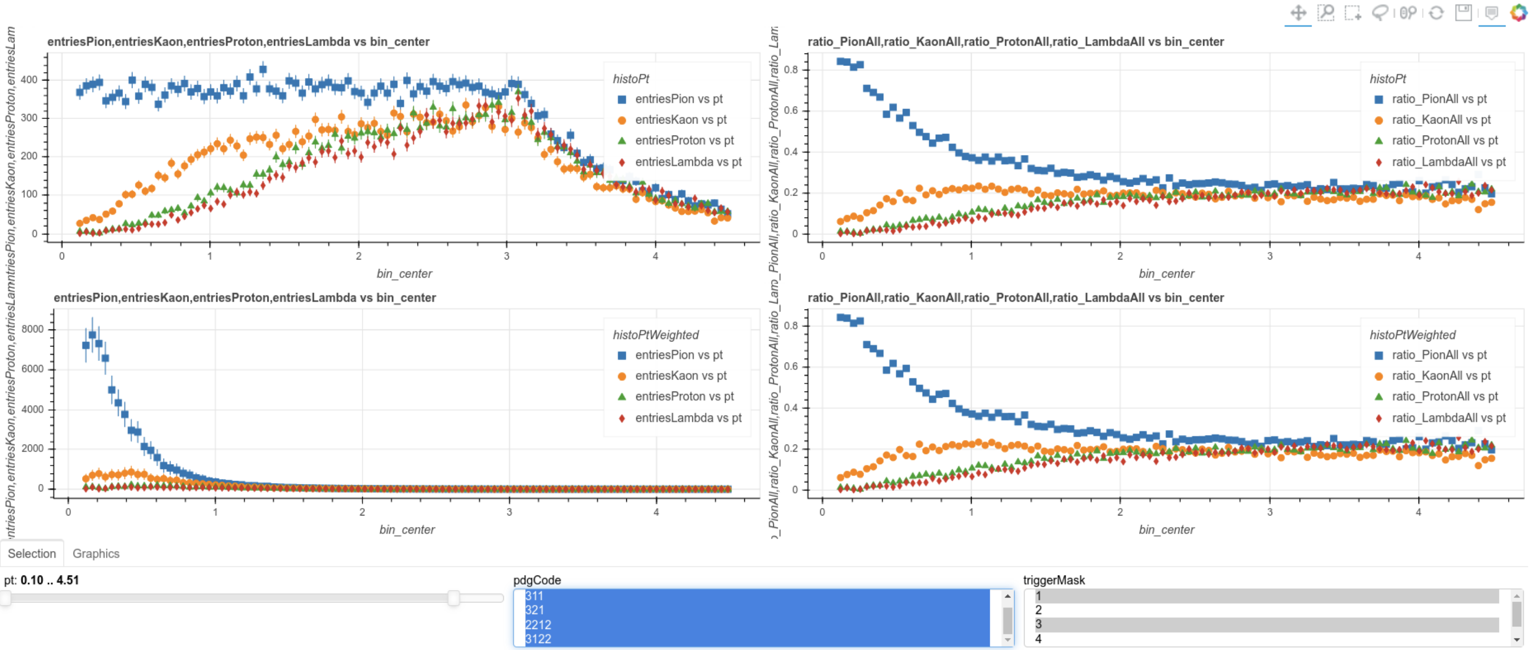
Run3 data to be sampled/skimmed in the similar way as Run1,2 (data down-sampled by factors 10^3)

- https://indico.cern.ch/event/1014566/contributions/4272119/attachments/2209987/3743263/ATO-465-DataSkimmingPerfCalPhysicsRun2Run3.pdf
- **small server instead of farm to analyze the data**

- Public node: https://alice-notes.web.cern.ch/node/1208

sampled spectra

re-weighted spectra

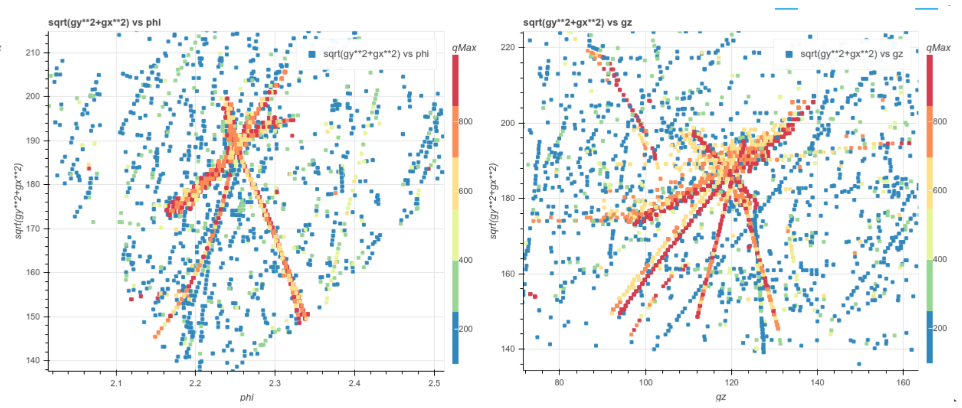


- Run1,Run2 sampling/skimming code extracted to standalone macro
  - To be integrated to the O2 template
- Trigger optimization ongoing in dedicated fast MC studies
  - Including event multiplicity sampling

https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164708/downsamplingTrigger.ipynb
https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164702/downsamplingTrigger.html
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/94435e925dd7b51f8753601f8ab2102587cf1702/JIRA/ATO-575/downsamplingTrigger.C#L27

ND+RootInteractive usage explained on real use cases

Spallation reconstruction event display



Customizable event display for magnetic spallation/monopole/high dEdx search reconstruction - triggered by tracks with saturated signal:

- interactive histogram, scatters, sliders, summary aggregated information

file:///lustre/alice/users/miranov/NOTESData/alice-tpc-notes/JIRA/ATO-432/AliRieman/production_22072021/dashboards/seedDisplay_dirClusters000.html

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/264a6fb497b05c1a601b7aaf6564a5d25546441f/JIRA/ATO-432/eventDisplay.ipynb

https://indico.cern.ch/event/989506/contributions/4225362/attachments/2186580/3694630/seed1DisplayRZPhi.html

Customizable display - Magnetic monopole search (1)

[9]:

```python
defaultCutTrack="entries>0"
output_file("seed1DisplayRZPhi.html")
histoArray = [
    {"name": "his_chi2N", "variables": ["chi2N"],"nbins":50},
    {"name": "his_fQMeanSeed1", "variables": ["fQMeanSeed1"],"nbins":50},
    {"name": "his_fQMedianSeed1", "variables": ["fQMedianSeed1"],"nbins":50},
    {"name": "his_fQSeed1Ratio", "variables": ["fQSeed1Ratio"],"nbins":50},
]

#dfQA=dfQA.sample(100000)
figureArray = [
    [['rSeed'], ["phiSeed"],  {"colorZvar":"qSeed"}],
    [['rSeed'], ["gzSeed"],  {"colorZvar": "qSeed"}],
    [['chi2N'],["his_chi2N"]],
    [['fQSeed1Ratio'],["his_fQSeed1Ratio"]],
    [['fQMeanSeed1'],["his_fQMeanSeed1"]],
    [['fQMedianSeed1'],["his_fQMedianSeed1"]],
    ["tableHisto", {"rowwise": True}],
    {"size": 5}
]
widgetParams=[
    ['range', ["sector"]],
    ['range', ['rSeed']],
    ['range', ['phiSeed']],
    ['range', ['gzSeed']],
    #
    ['range', ['chi2N']],
    ['range', ['fQSeed1Ratio']],
    ['range', ['fQRatio']],
    ['range', ['fQMeanSeed1']],
    ['range', ['fQMedianSeed1']],
    #
    ['range', ['qSeed']],
    ['range', ['seed1Tot']],
    ['range', ['eventID']],
]
tooltips = [("qSeed", "@qSeed"), ("fQMeanSeed1", "@fQMeanSeed1"), ("fQMedianSeed1", "@fQMedianSeed1"), ("eventID","@eventID"), ("sector","@sector"), ("rSeed","@rSeed")]
widgetLayoutDesc=[
    [0,1, 2,3],
    [4, 5,6,7,8],
    [9,10,11],
    {'sizing_mode':'scale_width'}
]
figureLayoutDesc=[
    [0,1,{'plot_height':450}],
    [2,3,4,5,{'plot_height':200}],
    [6,{'plot_height':25}],
    {'plot_height':240,'sizing_mode':'scale_width',"legend_visible":False}
]
fig=bokehDrawSA.fromArray(dfTrack.query("eventID>=0"), "chi2N>0&rSeed>0", figureArray, widgetParamsD,layout=figureLayoutDesc,tooltips=tooltips,sizing_mode='scale_width',
                          widgetLayout=widgetLayoutDescD,nPointRender=3000,rescaleColorMapper=True,arrayCompression=arrayCompressionRelative16,histogramArray=histoArray)
```

histogram array declaration

figure array declaration

widget array declaration

widget layout declaration

figure layout declaration

Customizable event display (in Jupyter notebook or plain python):

- interactive histogram, scatters, sliders, summary aggregated information
- input : TTree (or df) with C++ objects + functions

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/264a6fb497b05c1a601b7aaf6564a5d25546441f/JIRA/ATO-432/eventDisplay.ipynb

https://indico.cern.ch/event/989506/contributions/4225362/attachments/2186580/3694630/seed1DisplayRZPhi.html

file:///lustre/alice/users/miranov/NOTESData/alice-tpc-notes/JIRA/ATO-432/AliRieman/production_22072021/dashboards/seedDisplay_dirClusters000.html

Client side histogramming in bokeh interface - un-binned and binned data

- https://github.com/miranov25/RootInteractive/issues/90

Histogram derived information - efficiency/integral/mean/rms in user derived ranges resp. quantiles

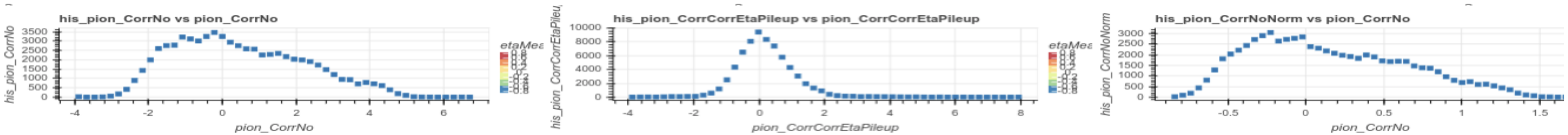
- https://github.com/miranov25/RootInteractive/issues/123

Example
PID QA interactive
histogram booking

```
histoArray = [
    {"name": "his_pion_CorrNo", "variables": ["pion_CorrNo"],"nbins":50},
    {"name": "his_pion_CorrCorrEtaPileup", "variables": ["pion_CorrCorrEtaPileup"],"nbins":50},
    #
    {"name": "his_pion_CorrNoNorm", "variables": ["pion_CorrNo/pion_CorrNoRMS"],"nbins":50},
    {"name": "his_pion_CorrCorrEtaPileupNorm", "variables": ["pion_CorrCorrEtaPileup/pion_CorrCorrEtaPileupRMS"],"nbins":50},
    #
    {"name": "his_pion_CorrNoRMS", "variables": ["pion_CorrNoRMS"],"nbins":50},
    {"name": "his_pion_CorrCorrEtaPileupRMS", "variables": ["pion_CorrCorrEtaPileupRMS"],"nbins":50}

]
```

PID QA dashboard
histogram part snapshot



PID QA dashboard
summary for selected
Mean, RMS, Sum

#	description	his_pion_CorrNo	his_pion_CorrCorrEtaPileup	his_pion_CorrNoNorm	his_pion_CorrCorrEtaPileupNorm
0	mean	0.456	0.194	0.144	0.064
1	std	1.733	0.876	0.492	0.235
2	entries	68072	68072	68072	68072
3	Σ(-3,3)	61780.555	67398.202	68072	68072
4	Σ_normed(-3,3)	0.908	0.99	1	1

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/master/JIRA/ATO-520/pidQAInteractiveRef.ipynb
https://indico.cern.ch/event/991451/contributions/4220782/attachments/2184007/3689893/qaPlotPion_Delta.html

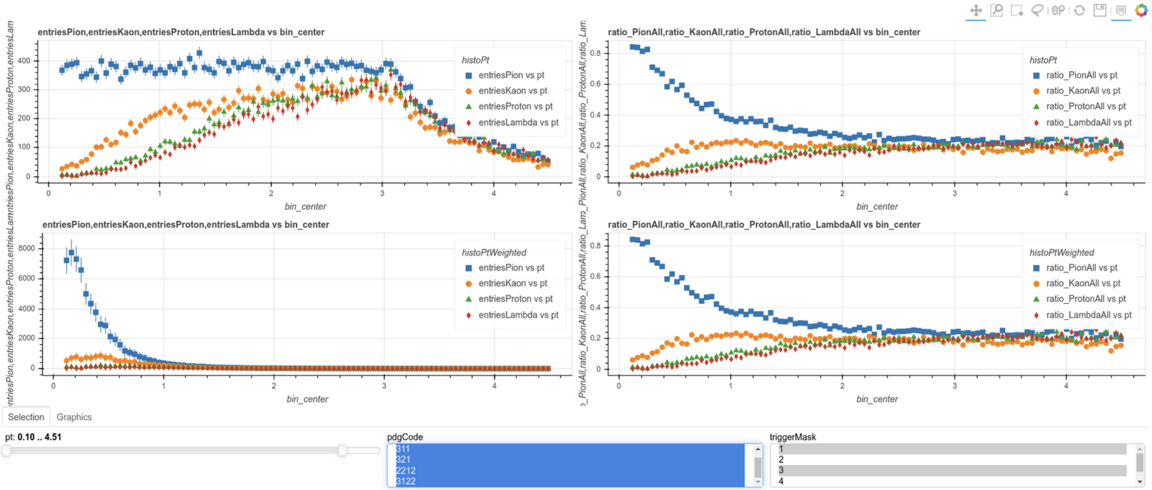
Histogram declaration

```
histoArray = [
    { "name": "histoPt", "variables": ["pt"], "nbins": 100, "histograms": {
            "entries": None,
            "entriesPion": {"weights": "isPion"},
            "entriesKaon": {"weights": "isKaon"},
            "entriesProton": {"weights": "isProton"},
            "entriesLambda": {"weights": "isLambda"},
        }
    },
    { "name": "histoPtWeighted", "variables": ["pt"], "nbins": 100, "histograms": {
            "entries": {"weights": "weight1"},
            "entriesPion": {"weights": "isPionW1"},
            "entriesKaon": {"weights": "isKaonW1"},
            "entriesProton": {"weights": "isProtonW1"},
            "entriesLambda": {"weights": "isLambdaW1"},
        }
    }
]
```

Function/alias declaration

```
aliasArray = [
    {"name": "ratio_PionAll", "variables": ["entries", "entriesPion"], "func": "return entriesPion / entries","context": "histoPt"},
    {"name": "ratio_KaonAll", "variables": ["entries", "entriesKaon"], "func": "return entriesKaon / entries","context": "histoPt"},
    {"name": "ratio_ProtonAll", "variables": ["entries", "entriesProton"], "func": "return entriesProton / entries","context": "histoPt"},
    {"name": "ratio_LambdaAll", "variables": ["entries", "entriesLambda"], "func": "return entriesLambda / entries","context": "histoPt"},
    #
    {"name": "entriesPionErr", "variables": ["entriesPion"], "func": "return Math.sqrt(entriesPion)","context": "histoPt"},
    {"name": "entriesKaonErr", "variables": ["entriesKaon"], "func": "return Math.sqrt(entriesKaon)","context": "histoPt"},
    {"name": "entriesProtonErr", "variables": ["entriesProton"], "func": "return Math.sqrt(entriesProton)","context": "histoPt"},
    {"name": "entriesLambdaErr", "variables": ["entriesLambda"], "func": "return Math.sqrt(entriesLambda)","context": "histoPt"},
    #
    {"name": "ratio_PionAll", "variables": ["entries", "entriesPion"], "func": "return entriesPion / entries","context": "histoPtWeighted"},
    {"name": "ratio_KaonAll", "variables": ["entries", "entriesKaon"], "func": "return entriesKaon / entries","context": "histoPtWeighted"},
    {"name": "ratio_ProtonAll", "variables": ["entries", "entriesProton"], "func": "return entriesProton / entries","context": "histoPtWeighted"},
    {"name": "ratio_LambdaAll", "variables": ["entries", "entriesLambda"], "func": "return entriesLambda / entries","context": "histoPtWeighted"},
    #
    {"name": "entriesPionErr", "variables": ["entriesPion"], "func": "return Math.sqrt(entriesPion*100)","context": "histoPtWeighted"},
    {"name": "entriesKaonErr", "variables": ["entriesKaon"], "func": "return Math.sqrt(entriesKaon*100)","context": "histoPtWeighted"},
    {"name": "entriesProtonErr", "variables": ["entriesProton"], "func": "return Math.sqrt(entriesProton*100)","context": "histoPtWeighted"},
    {"name": "entriesLambdaErr", "variables": ["entriesLambda"], "func": "return Math.sqrt(entriesLambda*100)","context": "histoPtWeighted"},
]
```

- Array of histograms for different particle species
- Reweighting of spectra
- Spectra ratios/efficiency
- parametric cut efifciency



https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164708/downsamplingTrigger.ipynb
https://indico.cern.ch/event/1146945/contributions/4843738/attachments/2431934/4164702/downsamplingTrigger.html
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/94435e925dd7b51f8753601f8ab2102587cf1702/JIRA/ATO-575/downsamplingTrigger.C#L27

Performance diff - ALICE performance: DCA resolution/bias

http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/LHC18c/kink_3sigma_CENT_pass2/dashboard/LHC16f_lowmult_pass2/fig0/compDefaultV0DCARLHC18c_kink_3sigma_CENT_pass2LHC16f_lowmult_pass2HistComp.html

LHCh18c / LHC16f — Δ norm DCA — Δ DCA pull — σ norm DCA — σ DCA pull

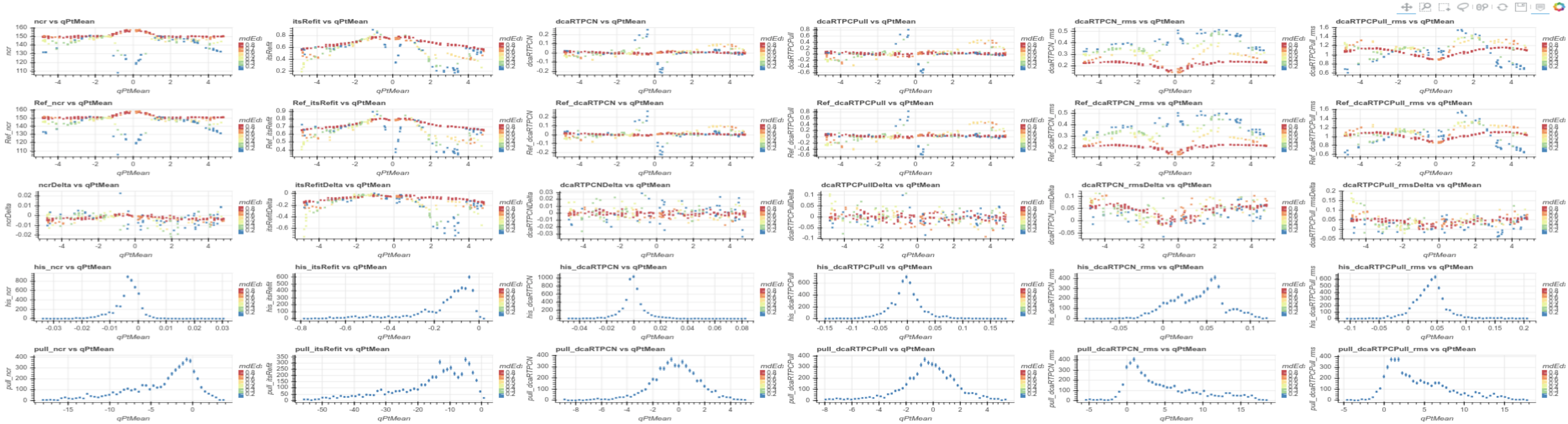
Data

Reference data

Data-Ref.

Histo:
Data-Ref

Histo:
(Data-Ref)/σ



Summary table:
Data-Ref

#	description	his_ncr	his_itsRefit	his_dcaRTPCN	his_dcaRTPCPull	his_dcaRTPCN_rms	his_dcaRTPCPull_rms	pull_ncr	pull_itsRefit	pull_dcaRTPCN	pull_dcaRTPCPull	pull_dcaRTPCN_rms	pull_dcaRTPCPull_rms
0	mean	-3.677e-3	-1.498e-1	-5.352e-4	-2.629e-3	0.04	0.038	-3.579e+0	-1.570e+1	-3.248e-1	-3.412e-1	4.618	4.365
1	std	0.005	0.141	0.009	0.031	0.03	0.031	4.085	11.926	1.876	2.035	4.831	4.048
2	entries	4990	4990	4990	4990	4990	4990	4990	4990	4990	4990	4990	4990

Widgets for
interactive
ND selection

qPtMean: **-4.82 .. 4.82** — tglMean: **-0.82 .. 0.82** — mdEdxMean: **0.10 .. 0.88** — contTPCMean: **249 .. 251**

ncr_entries: **404 .. 343651** — Ref_ncr_entries: **144 .. 106030** — dcaRTPCN_entries: **392 .. 343651** — Ref_dcaRTPCN_entries: **139 .. 106030**

ncrDelta: **-0.03 .. 0.03** — itsRefitDelta: **-0.80 .. 0.03** — dcaRTPCNDelta: **-0.05 .. 0.08** — dcaRTPCPullDelta: **-0.15 .. 0.18**

dcaRTPCN_rmsDelta: **-0.09 .. 0.12** — dcaRTPCPull_rmsDelta: **-0.11 .. 0.21**

Test data/Data, production/reference production, Period/Period, Data/MC
Production comparison in many dimensions (q/Pt,pz/pt,MIP/dEdx, mult)
Interactive browsing/histograms/aggregation in ND
Example above used for the B=0.2T (LHC18c) production preparation

Performance diff - ALICE performance MC/data: TPC+ITS QA

http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/LHC18c/pass2_CENT_syst_err/dashboard/LHC21a6_cent_kink5sigma/fig2/compDefaultV2LHC18c_pass2_CENT_syst_errLHC21a6_cent_kink5sigmaHistComp.html

LHC18c / MC LHC21a6

TPC ncls | TPC Cross. | TPC cl frac.. | TPC-ITScl refit | TPC chi2 | ITS chi2

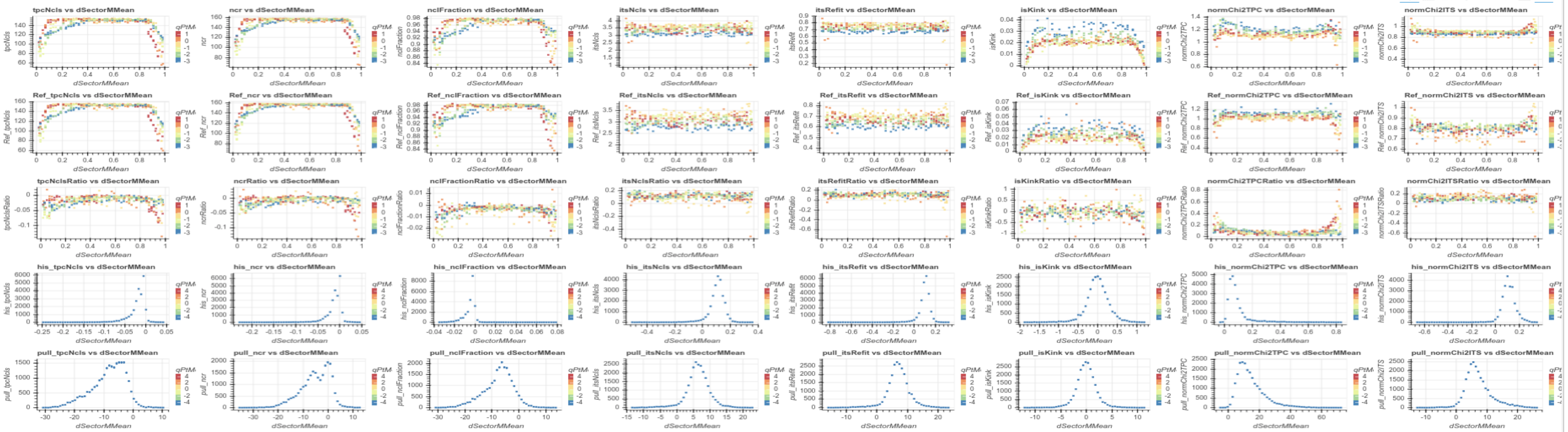
Data

Reference data

Data-Ref.

Histo: Data-Ref

Histo: (Data-Ref)/σ



Summary table: Data-Ref

#	description	his_tpcNcls	his_ncr	his_nclFraction	his_itsNcls	his_itsRefit	his_isKink	his_normChi2TPC	his_normChi2ITS	pull_tpcNcls	pull_ncr	pull_nclFraction	pull_itsNcls	pull_itsRefit	pull_isKink	pull_normChi2TPC	pull_normChi2ITS
0	mean	-1.978e-2	-1.146e-2	-4.836e-3	0.104	0.098	-1.021e-2	0.092	0.111	-8.388e+0	-5.066e+0	-7.592e+0	6.365	6.969	-5.572e-2	14.496	7.123
1	std	0.02	0.018	0.006	0.057	0.057	0.291	0.09	0.061	5.795	5.866	5.516	3.588	3.859	1.978	8.925	4.37
2	entries	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880	23880

Widgets for interactive ND selection

qPtMean: **-3.20 .. 1.60** tglMean: **-0.80 .. 0.80** dSectorMMean: **0.01 .. 0.99** contTPCMean: **249 .. 251**

tpcNcls_entries: **284 .. 23416** Ref_tpcNcls_entries: **193 .. 7783** itsNcls_entries: **249 .. 20298** Ref_itsNcls_entries: **170 .. 6656**

tpcNclsRatio: **-0.26 .. 0.04** ncrRatio: **-0.24 .. 0.05** nclFractionRatio: **-0.04 .. 0.08** itsNclsRatio: **-0.53 .. 0.38**

itsRefitRatio: **-0.84 .. 0.64** isKinkRatio: **-1.95 .. 1.25** normChi2TPCRatio: **-0.04 .. 0.86** normChi2ITSRatio: **-0.68 .. 0.55**

Test data/Data, production/reference production, Period/Period, Data/MC
Production comparison in many dimensions (q/Pt, pz/pt, sector distance,mult)
Interactive browsing/histograms/aggregation in ND
Tool to be used in ongoing service work with Yale group for Data ↔ MC remapping

Example comparison of the invariant mass performance for 5 different V0 finder scenario

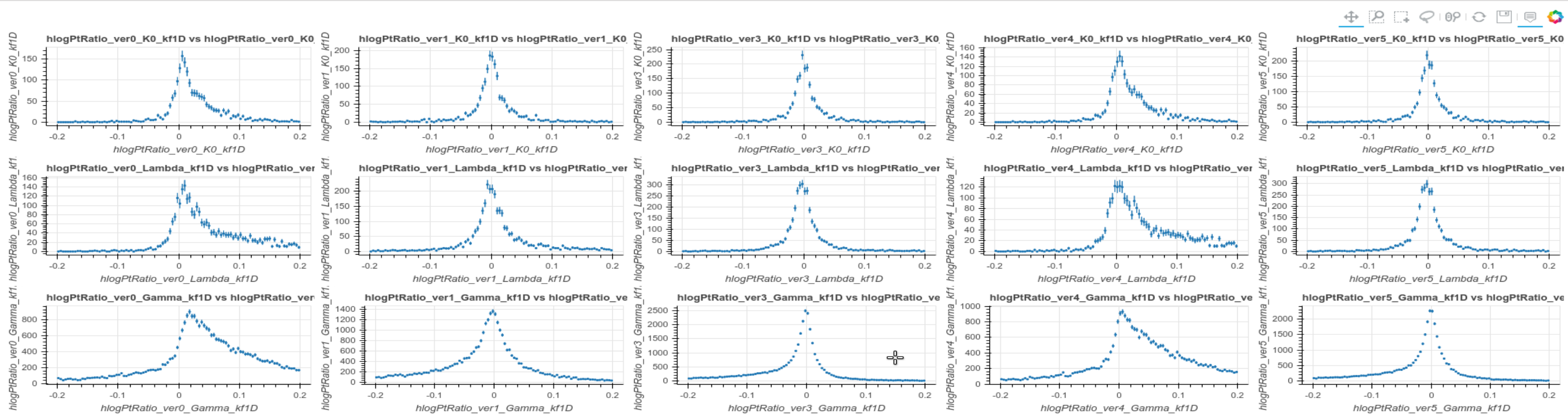
- providing summary dashboards as support material in agenda, expert hands-on session save several weeks iterations
- 6+1D (algorithm, is primary flag, 1/pt, multiplicity,pz/pt, decay radius, mother radius)
- **Optimal Hybrid V0/cascade finder** (proper material budget correction, optimal co-variance, causality information)

ver0 - Offline like | ver1 - prim → V0 | ver3 - TPC → V0 | ver4 - OnTheFly if exist Offline comp | ver5 - OnTheFly hybrid optimal

K0

λ

γ



#	description	hlogPtRatio_ver0_K0_	hlogPtRatio_ver1_K0_	hlogPtRatio_ver3_K0_	hlogPtRatio_ver4_K0_	hlogPtRatio_ver5_K0_	hlogPtRatio_ver0_Lar	hlogPtRatio_ver1_Lar
0	mean	0.026	0.003	0.003	0.025	0.004	0.047	0.011
1	std	0.043	0.041	0.034	0.043	0.035	0.058	0.055
2	entries	1877	1932	1977	1877	1977	2620	3061

#	description	hlogPtRatio_ver3_Lar	hlogPtRatio_ver4_Lar	hlogPtRatio_ver5_Lar	hlogPtRatio_ver0_Gai	hlogPtRatio_ver1_Gai	hlogPtRatio_ver3_Gai	hlogPtRatio_ver4_Gai	hlogPtRatio_ver5_Gai
0	mean	5.358e-4	0.044	0.001	0.043	-1.318e-2	-2.139e-2	0.034	-1.725e-2
1	std	0.04	0.06	0.041	0.079	0.07	0.058	0.079	0.06
2	entries	3412	2637	3406	29120	30602	31602	29189	31708

mptMC: **0.50 .. 5**　atgl: **0 .. 1.40**　fNContributors: **2 .. 3945**

isPrimMC: 0.0, 1.0　RMC: **20 .. 120**　motherR: **0 .. 79.98**

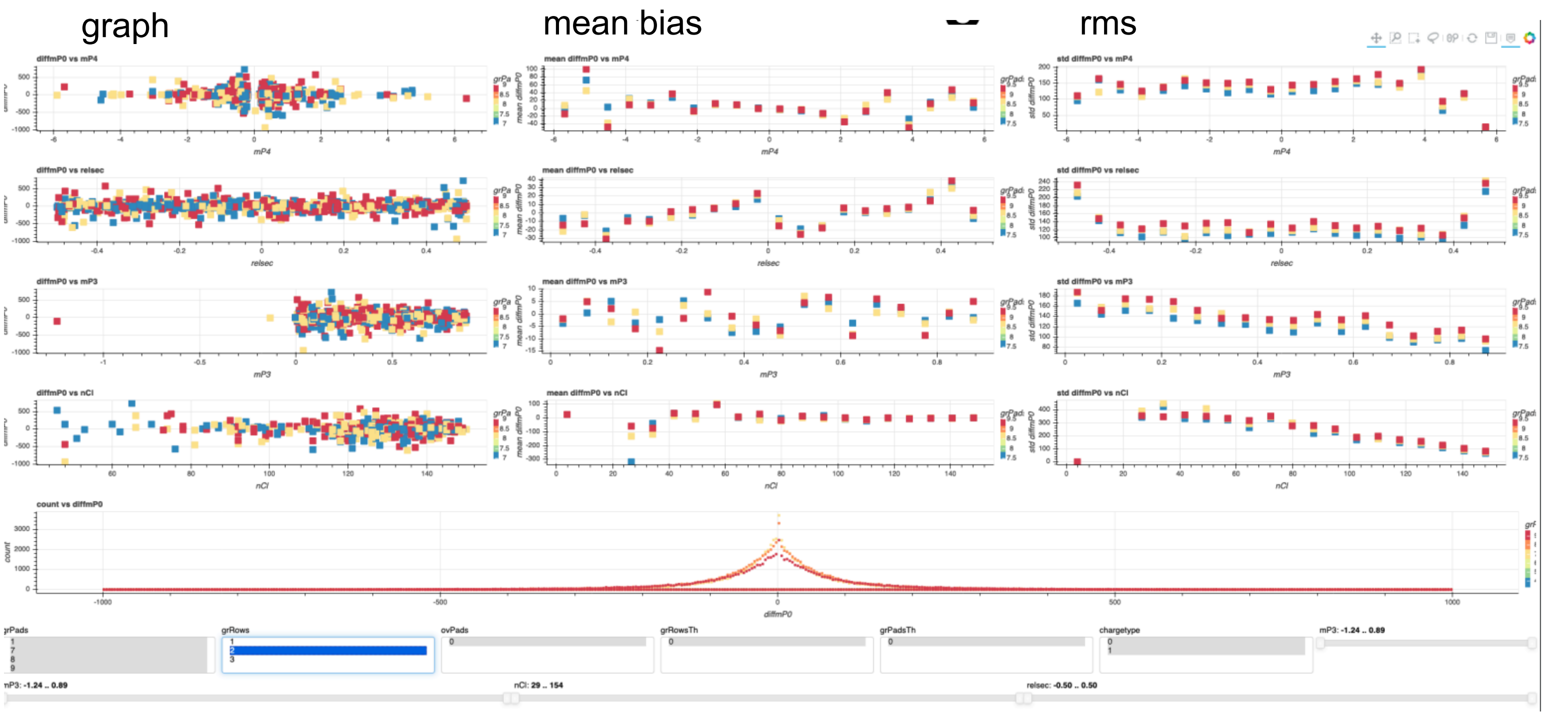
https://indico.cern.ch/event/1088044/contributions/4574146/attachments/2335010/3979885/ATO-544-HybridV0_27102021.pdf
https://indico.cern.ch/event/1088044/contributions/4574146/subcontributions/354933/attachments/2334975/3979831/hdMass_ver5_kf1D_Dashboard.html
https://indico.cern.ch/event/1088044/contributions/4574146/subcontributions/354933/attachments/2334975/3979832/hlogPtRatio_ver5_kf1D_Dashboard.html
https://indico.cern.ch/event/1088044/contributions/4574146/subcontributions/354933/attachments/2334975/3979806/hpMass_ver5_kf1D_Dashboard.html
https://indico.cern.ch/event/1088044/contributions/4574146/subcontributions/354933/attachments/2334975/3979807/hschi2_ver5_kf1D_Dashboard.html

https://indico.cern.ch/event/1135398/#sc-1-3-v0-and-cascade-finder-f

Fast feedback →
Very constructive discussion

Performance for different granularity of 3D ion currents - example of interactive 4D histograms and derived mean (bias) and rms (residual)

Using (12 D) dashboard - very constructive and effective discussion during meeting

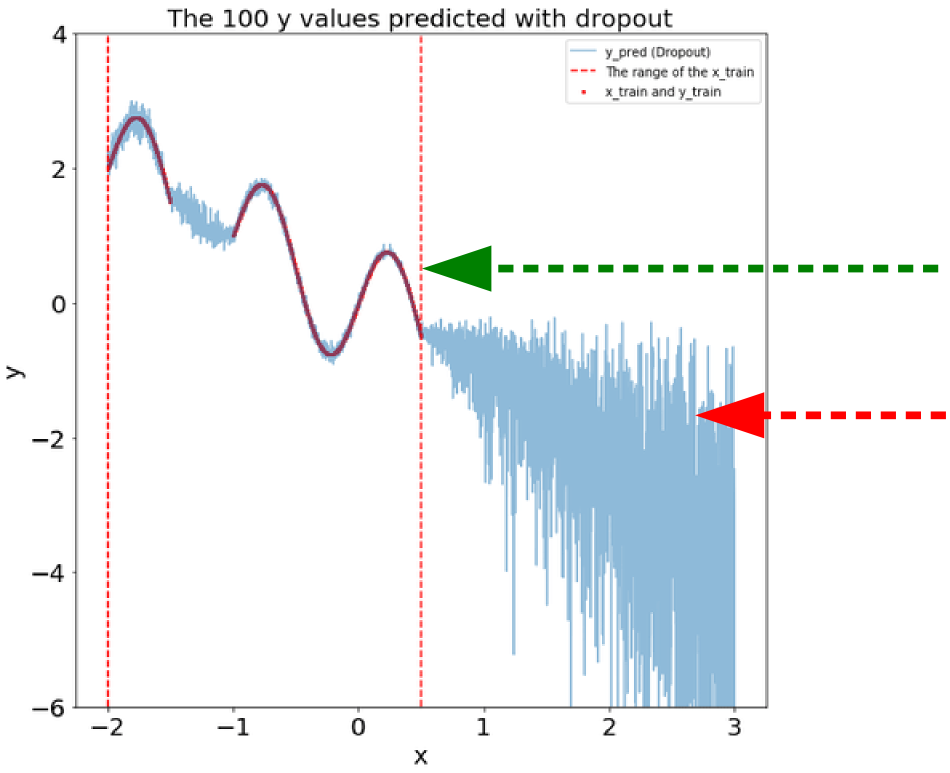


https://indico.cern.ch/event/1091510/
https://indico.cern.ch/event/1091510/contributions/4599999/attachments/2338476/3986580/residualTrackParam.html

https://indico.cern.ch/event/1135398/#sc-1-1-space-charge-distrotion

Reducible, irreducible error and **P**robability **d**ensity **f**unction
RootInteractive ML wrappers

https://fairyonice.github.io/Measure-the-uncertainty-in-deep-learning-models-using-dropout.html



Knowledge of errors and PDF crucial for data interpretation

- irreducible error intrinsic data fluctuation
- reducible error
- model error

ML non-parametric (non-constrained) models good for interpolation bad for extrapolation
Errors and PDF to be extracted locally

Combination of physical model and ML non parametric models preferable

What is the prediction error for non seen data ?

For data taken from a completely unknown distribution, a CI and errors can be calculated using a bootstrapping method (Efron, 1992; Johnson, 2001).

Bootstrappig CPU consuming

To speed up - to use the internal dispersion of the prediction in ensamble learning methods (random forest, xgboost)

Machine learning based regression algorithm for the non-parametric description of an unknown function:

- N-dimensional calibration, tracking performance parameterization ($\chi 2$, $N_{Clusters}$, σ_{DCA}), conditional PDF distribution

Provides wrappers for the standard ensemble learning method (Random forest, xgboost)

- Local error (reducible, irreducible) parameterization
- Automatic parameters adjustment to minimize reducible error
- Robust local estimator
- Conditional probability density function   and quantiles
- Linear Regression Forest - to reduce model error (Work in progress)

For the Neural net, error estimated using dropout prediction

- only prototype, not used yet in real use cases, model dependent

For the RandomForest - error estimated using decision trees RMS, for trees with and without max_depth

- ~irreducible error estimated using RMS of unbound trees
- ~reducible error estimated using RMS of prediction for trees with max_depth limitted

Irreducible local error could be strongly parameter dependent e.g.:

- e.g. bigger relative error of the ion tail for more noisy pads with smaller signal (signal/noise), multiplicity error proportional to sqrt(multiplicity), tracking relative pt resolution ~ (dEdx,L_{arm},)

Reducible error strongly depends on the granularity and on the function derivative and local density of points. Error of the extrapolation explodes.

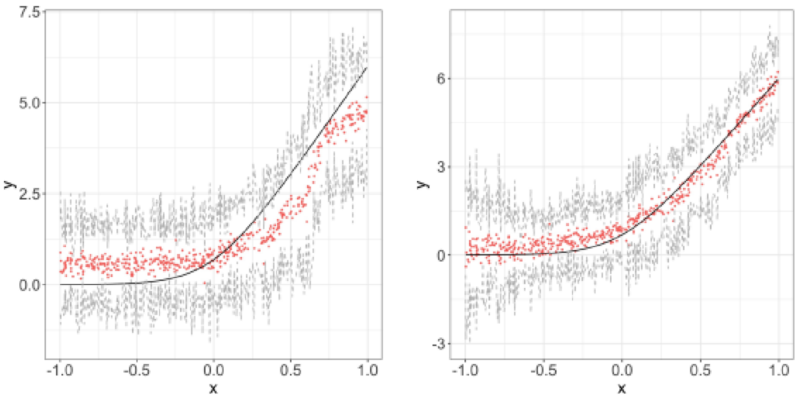
Local linear forest & adaptive kernel method

https://arxiv.org/pdf/1807.11408.pdf

*Random forests are a powerful method for non-parametric regression, but are limited in their ability to fit smooth signals. Taking the perspective of random forests as **an adaptive kernel method**, we pair the forest kernel with a local linear regression adjustment **to better capture smoothness**. The resulting procedure, local linear forests, enables us to **improve on asymptotic rates of convergence for random forests with smooth signals**, and provides substantial gains in accuracy on both real and simulated data.*

https://grf-labs.github.io/grf/articles/llf.html

An Adaptive kernel method



Random forest   Local linear forest

where the forest weight $\alpha_i(x_0)$ is the fraction of trees in which an observation appears in the same leaf as the target value of the covariate vector.

$$\alpha_i(x_0) = \frac{1}{B}\sum_{b=1}^{B}\frac{1\{x_i \in L_b(x_0)\}}{|L_b(x_0)|}$$

Local linear forests take this one step further: now, instead of using the weights to fit a local average at x_0, we use them to fit a local linear regression, with a ridge penalty for regularization. This amounts to solving the minimization problem below, with parameters: $\mu(x)$ for the local average, and $\theta(x)$ for the slope of the local line.

$$\begin{pmatrix}\hat{\mu}(x_0)\\ \hat{\theta}(x_0)\end{pmatrix} = \text{argmin}_{\mu,\theta}\left\{\sum_{i=1}^{n}\alpha_i(x_0)(Y_i-\mu(x_0)-(x_i-x_0)\theta(x_0))^2+\lambda||\theta(x_0)||_2^2\right\}$$

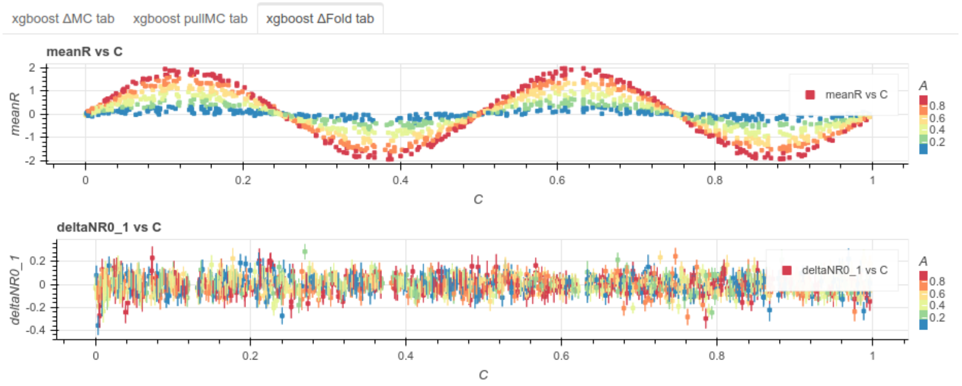
R package integrated within GeneralizedRandomForest (https://grf-labs.github.io/grf/)

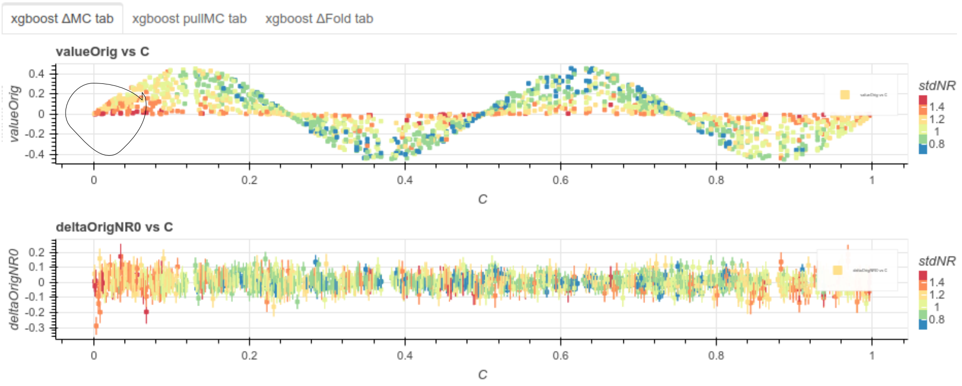
RootInteractive python implementation planned to be ready for workshop

- too slow (similar as in R package)

Cached version with approximation as used in our previous C++ implementation → fir the moment postponed

$$f(A,B,C,D) = norm*A*sin(n*2*pi*C) + B*noise$$





4D Uniform input

```
df = pd.DataFrame(np.random.random_sample(size=(nPoints, 4)), columns=list('ABCD'))
df["B"]=df["B"]+0.5
df["noise"] = np.random.normal(0, stdIn, nPoints)
df["noise"] += (np.random.random(nPoints)<outFraction)*np.random.normal(0, 2, nPoints)
```

Local reducible (color code) error increased at the boundaries

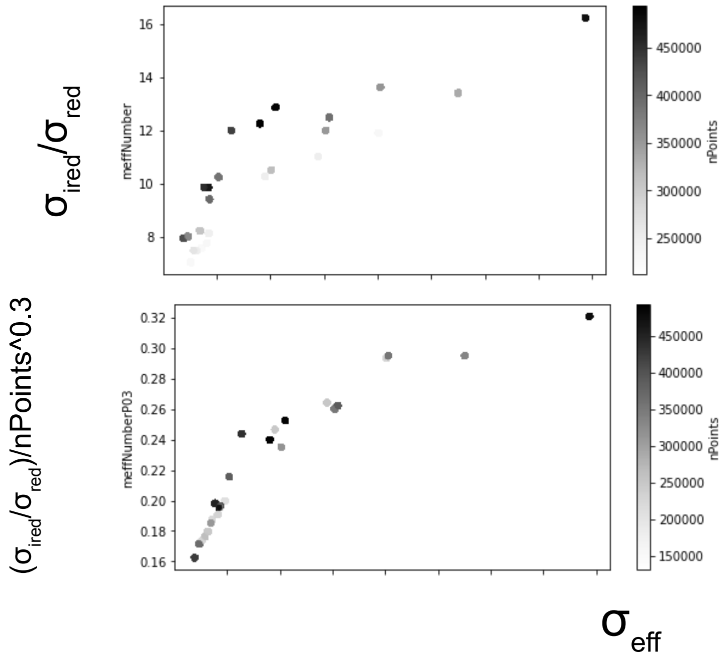
Reducible error estimated using spread of the xgboost in iterations after “early_stop”.
Keeping all parameters - reducing subsample and learning_rate

https://indico.cern.ch/event/1147231/contributions/4815612/attachments/2424564/4150687/MlxgboostErrPDF_n2_stdIn0.2_nPoints200000.html
https://indico.cern.ch/event/1147231/contributions/4815612/attachments/2424564/4150688/MlxgboostErrPDF_back11042022.ipynb

$f(A,B,C,D) = \mathbf{norm}*A*\sin(\mathbf{n}*2*pi*C) + B*\boldsymbol{\sigma}_{noise}$

$$\sigma_{eff} = \frac{\sigma_{ired}}{norm}$$

$$\sqrt{N_{eff}} = \frac{\sigma_{ired}}{\sigma_{red}}$$



Parameter scan to emulate statistics requirement

- number of points, function normalization, noise (σ_{ired}), n_{sin}

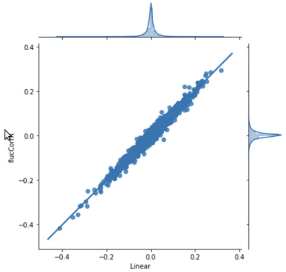
Making function variation small in respect to intrinsic noise (σ_{ired}), effective number of points increase → reducible error decrease

Making regression for delta model (observation - analytical approximation) is preferable

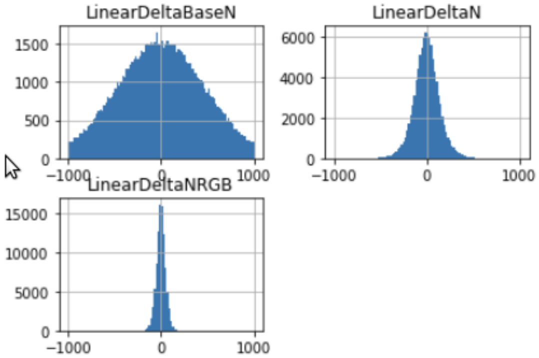
- **Used in many Alice use cases**

Global Linear fit - approximation of the physical model

- Input parameters:
  - local derivative of distortion , current in the TPC (ΔI)
  - ion current as white noise → individual FFT coefficient independent (μ=0, σ_i=σ)
- Output: Δ distortion
- Convolution theorem → approximation response for individual FFT current harmonics
  - convolution in 3D space → multiplication in FFT space
  - Linear fit to approximate convolution kernel
  - 1 FFT as a LinearBase , 20 most important FFT



ΔR at R<95, drift>0.5




```
flucCorrRN          1264.6
LinearDeltaBaseN     509.1
LinearDeltaN         153.4
```

Random forest and xgboost used with/without physical model as a prefilter

- Using physics models as prefilter significantly better residual resolution
  - for 10^6 training points ~ 80 microns ~ 40 microns
- Residual distortion after the LinearFit+XGB due 3D current fluctuation not used in the model

RootInteractive plans/wish list

Improve graphics customization:

- currently done by user parametization
    - e.g. marker size, axis variables content, variable transformation (e.g. log,linear,sqrt)
- template/default parameterization to be prototyped/provided

Data aggregation on client for interactive physics analysis

- interactive histogramming/efficiency/ratios exist
- interactive unfolding

Functions on client and ONNX:

- currently only standard java script function could be used on client
- ONNX to enable usage of ML on client

Parameteric autonecoders

Linear regression forest + adaptive kernel extraction

- test Mapie interface https://mapie.readthedocs.io/en/latest/index.html

Supporting references.

* Only part accessible to not ALICE members

- **Tracking articles:** https://twiki.cern.ch/twiki/bin/view/ALICE/TrackingReference
  - [A0] CHEP2003: TPC tracking http://inspirehep.net/record/621229
  - [A1] Time05 workshop: ALICE combined tracking and V0 finder http://www.sciencedirect.com/science/article/pii/S0168900206008126
  - [A2] CHEP2004 - ITS tracking integrated with V0 finder https://cds.cern.ch/record/688747/files/CERN-2005-002-V1.pdf
  - [A3] CHEP2004- BAYESIAN APPROACH FOR COMBINED PARTICLE IDENTIFICATION
  - [A4] CHEP2006 - TRD tracking https://indico.cern.ch/event/408139/contributions/979783/attachments/815694/1117684/MarianIvanovchep06.pdf

ALICE: Physics Performance Report, Volume II

- [A5] http://iopscience.iop.org/0954-3899/32/10/001/

TPC TDRs:

- [A6] TPC TDR 2000 chapter 7 - https://cds.cern.ch/record/451098/files/open-2000-183.pdf
- [A7] TPC TDR 2013 - chapter 7 (performance and space charge distortion/correction) https://cds.cern.ch/record/1622286/files/ALICE-TDR-016.pdf

TRD TDR:

- [A8] Chapter 6, local tracking performance and Digital cancellation of the tail in PASA signal https://cds.cern.ch/record/519145/files/cer-2275567.pdf

-
- [N1] Pass2 reconstruction modification - with big emphasis (but not only) on the dEdx and pileup correction
  - https://www.overleaf.com/project/61800f2b4ae921cb616ed79b
  - https://cernbox.cern.ch/index.php/s/R5beD9pcLOnTBqZ
- [N2] TPC digital signal processing
  - https://alice-notes.web.cern.ch/node/1207
  - https://www.overleaf.com/project/617b06fa5f8e42a110c21405
  - for non ALICE member - copy in the cernbox (Friday version): https://cernbox.cern.ch/index.php/s/R5beD9pcLOnTBqZ
-

- RD51 workshop (2020) - TPC:
  - [P1] TPC track reconstruction and PID https://indico.cern.ch/event/889369/contributions/4011353/(proceeding in preparation -[N1])
  - [P2] Common mode and ion tail analysis of the GEM upgrade of the ALICE TPC https://indico.cern.ch/event/889369/contributions/4044542/
- Reconstruction:
  - [P3] Performance of the hybrid V0 finder:
    - Presentation: https://indico.cern.ch/event/1088044/contributions/4574146/attachments/2335010/3979885/ATO-544-HybridV0_27102021.pdf
    - Minutes: https://indico.cern.ch/event/1088044/?note=177737
  - [P4] Physics week (October, 2018)- DPG/tracking: Combined TRD tracking in Run2.
    - https://indico.cern.ch/event/757761/contributions/3183222/attachments/1738216/2812589/TRDInTracking_PhysWeek2210.pdf
  - [P5] ALICE week (March 2020)- DPG/tracking: Reconstruction modification for the pass2/pass3 …
    - https://indico.cern.ch/event/876093/contributions/3784236/attachments/2002467/3343178/PWGPP-571-ReconstructionModification2018_1203.pdf
  - [P6] (DPG and AIM Meetings) Extended acceptance for tracking and TPC+PIXEL tracking
    - https://indico.cern.ch/event/876132/#1-extended-acceptnace-for-trac
  - [P7] DPG meeting - Invariant mass bias and pt bias calibration (https://indico.cern.ch/event/991463/?note=162249)
    - https://indico.cern.ch/event/991463/contributions/4343481/attachments/2235673/3790851/stat_photon_210429_TrackingMeeting.pdf

Distortion calibration:

  - [P8] Technical board (2017) - Distortion theoretical models, origin of space charge in Run2 and distortion mitigation
    - https://indico.cern.ch/event/605126/contributions/2538484/attachments/1441002/2218550/DistortionAnalitycalModelsForTB_06042017_v2.pdf
  - [P9] OFFLINE week (2020) TPC calibration: theoretical considerations and data driven approach
    - https://indico.cern.ch/event/888263/contributions/3784229/
  - [P10] SC meeting: Space charge IDC factorization and IDC grouping optimization:
    - https://indico.cern.ch/event/1091510/contributions/4599999/attachments/2338476/3986854/2021-11-03_IDCs.pdf
    - https://indico.cern.ch/event/1091510/contributions/4599999/attachments/2338476/3986449/ATO-494-Grouping_of_Pads_IDC_Workflow_SC_Meeting.pdf

Reference presentations

- **RootInteractive and ND pipeline:**
  - [P10] Offline weeek (2021) RootInteractive news
    - https://indico.cern.ch/event/1091321/contributions/4612911/
  - [P11] Offline week (2020)
    - https://indico.cern.ch/event/888263/contributions/3788628/attachments/2006705/3351619/PWGPP-485NDPipelineRootInteractive2003.pdf
  - [P12] WP7 QA meeting (2020)
    - https://indico.cern.ch/event/976023/contributions/4110642/attachments/2145661/3616562/PWGPP-485NDPipelineRootInteractive18112020.pdf
  - [P13] Offine week (2019) - Recent developments in ND-analysis pipeline (RootInteractive)
    - https://indico.cern.ch/event/806602/contributions/3379555/attachments/1824640/2995393/NDimensionalPipeline_OFFLINEWEEK05042019.pdf

References: interactive dashboards

- Support material for RCU note [N2] (Yiota, Marian, Mesut)
  - [D1] Visualization of the common-mode effect dependencies using ROOT interactive (11 Dimensions)
    - https://gitlab.cern.ch/aliceeb/TPC/-/blob/master/SignalProcessing/commonModeFractionML.html
  - [D2] Visualization of the ion-tail fit parameters and correction graphs using ROOT interactive (12 Dimensions)
    - https://gitlab.cern.ch/aliceeb/TPC/-/blob/master/SignalProcessing/ionTailFitParameters_sectorScan.html
  - [D3] Visualization of the toy MC results using ROOT interactive (13 Dimensions)
    - https://gitlab.cern.ch/aliceeb/TPC/-/blob/master/simulationScan/toyMCParameterScan.html

Support material for Hybrid V0 studies [P1] (Marian, Georgijs)

- [D4] Interactive invariant mass histogram dashboards (6+2 Dimensions)
  - https://indico.cern.ch/event/1088044/#sc-1-3-interactive-histograms
- [D5] Pt and invariant mass performance maps dashboards
  - https://indico.cern.ch/event/1088044/#sc-1-2-gamma-dashboards
  - https://indico.cern.ch/event/1088044/#sc-1-4-k0-dashboards

References: RootInteractive dashboards

QA and production preparation (service task students) :

- [D6] QA comparison of ongoing MC and raw data production (LHC18q,r, LHC18c,LHC16f,LHC17g..) See interactive dashboards in agenda of calibration/tracking meeting:
  - https://indico.cern.ch/event/991449/ , https://indico.cern.ch/event/991450/ , https://indico.cern.ch/event/991451/

- ## PID (Xiaozhi, Marian)
  - [D7] TPC PID calibration and QA
    - https://indico.cern.ch/event/983778
    - https://alice.its.cern.ch/jira/secure/attachment/53371/qaPlotPion_test1.html
    - https://indico.cern.ch/event/991451/contributions/4220782/attachments/2184007/3689893/qaPlotPion_Delta.html

- ## Fast MCkalman and event display (Timon, Marian)
  - [D8] Space charge distortion calibration (Run3) and performance optimization (Run2, Alice3) - [P9]
    - https://indico.cern.ch/event/1091510/contributions/4599999/attachments/2338476/3986580/residualTrackParam.html
    - https://indico.cern.ch/event/1087849/contributions/4577709/attachments/2331293/3973338/residual_track_parameter_Dist_GainIBF.html
  - [D9] High dEdx (spallation product) reconstruction and magnetic monopole tracking
    - https://indico.cern.ch/event/991452/contributions/4222204/attachments/2184856/3691411/seed1Display2.html

- ## Space charge distortion calibration (Matthias, Ernst, Marian)
  - [D10] digital current grouping and factorization studies
    - https://indico.cern.ch/event/1091510/
    - https://indico.cern.ch/event/1087849/