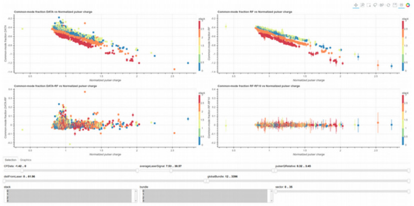
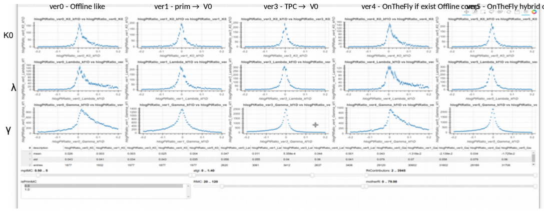
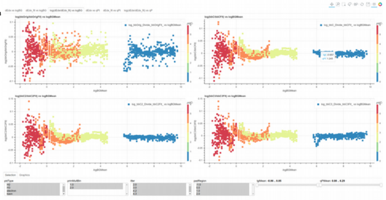




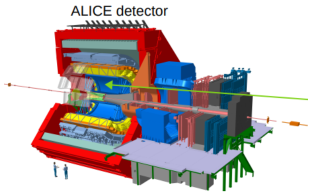


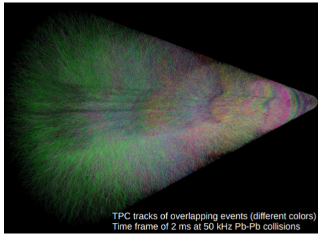
ALICE: RootInteractive tool for N dimensional interactive analysis: **Seeing is believing**

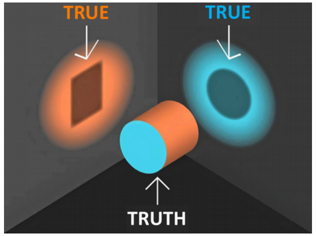
Developers: Marian Ivanov (CERN,GSI Darmstadt, Uni. Heidelberg), Marian Ivanov jr.

Alice Projects:

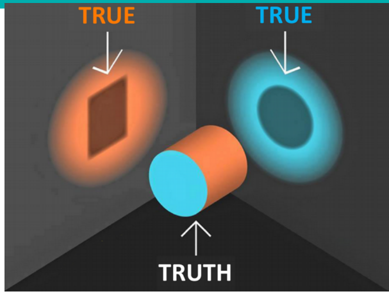
- Run3 space charge calibration - Ernst, Mathias,Caitie, Marian (GSI,Frankurt)
- CRU -Run3 digital signal processing - Yiota, Mesut,Marian (CERN, Yale)
- Performance web pages - Pritam, Dibakar, Tulika, Marian (Kolkota)
- PID calibration - Tuba (Frankfurt),Mathias
- **High dEdx,spallation/Magnetic monopole - Marian, Timon (Wiena)**
- V0, Cascade, exotica reconstruction - Marian (GSI+Heidelberg),Georgijs(in past) - Benedict(to join)
-
- MC/Data remapping - Yale group
- **Particle production … combine estimator - Michal,Marian (UK Bratislava)**
- fastMCkalman - Marian, Federico (Oxford)







https://github.com/miranov25/RootInteractive/



ALICE goals and challenges

NDimensional interactive analysis - **Seeing is believing**

- Motivation and Interactive n dimensional parameter optimization example

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) **on server (Jupyter notebook) and on clients (browser)**
- Triggers and data skimming (representative data selection) to enable interactive analysis

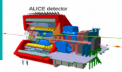
ND+RootInteractive functionality shown in real use cases:

- **Run3 Calibration and calibration QA** (CRU digital signal processing, ion tail, common mode, Space charge distortion calibration)
- **Run2,3 calibration and calibration QA (PID)**
- **Run3 space charge distortion ML studies**
- **Hybrid V0 and cascade finder (Run2->Run3)**
- **Highly customizable event display** (Run3, ALICE3 reconstruction, high dEdx,spallation, (Magnetic monopole) tracking)
- **Reconstruction Performance web page** Data/MC
- **MC/data remapping for the Machine learning**
- **fastMCKalman**

Interactive differential studies, physics analysis

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

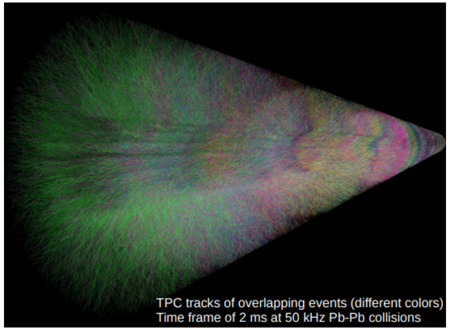
Alice Run 3 - goals and challenges



Record large Pb-Pb minimum bias sample

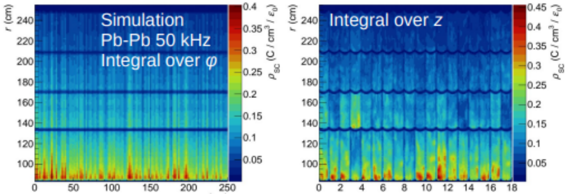
Continuous readout at 50 kHz interaction rate in Pb-Pb collisions

- 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)



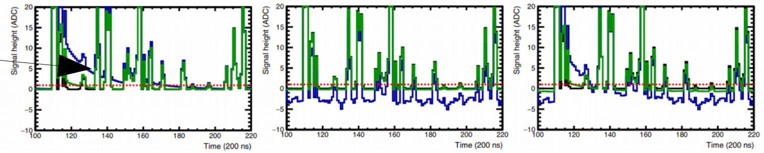
New TPC GEM design → **space charge** in TPC inside the drift volume **distorting** trajectories

- 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(5 ms) in time



Significant **baseline bias** 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, etc. ... necessitates the use of advanced methods of data analysis. Experts needed

N dimensional interactive analysis:

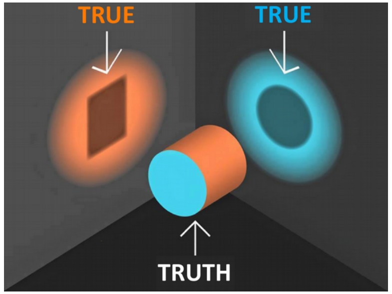
Seeing is believing

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 tool

- to simplify data analysis in many (optimally all relevant) dimensions
- to fit (ML regression) and visualise N-dimensional functions
- to validate assumptions
- to enable simple **functional composition for (non-parametric, parametric) functions**
- aimed **at standard users** (Masters, PhD), not just computer experts for educational purposes
- to give students **very fast feedback** from day one (seconds instead of weeks), to allow **interactive expert communication** in working groups
- for **multidimensional parameter optimisation** for convergence to the optimal solution
- to simplify management ↔ domain expert negotiation

https://en.wikipedia.org/wiki/Lie-to-children

"A lie-to-children (plural lies-to-children) is a simplified explanation of technical or complex subjects as a teaching method for children and laypeople.

Simple fairy tales have an important role in teaching messages to children, but Wikipedia is not a children's encyclopedia. Editors should not oversimplify material to try to make it easier to understand.

Real example cases of 1D over-simplification:

"Detector noise did not change":

- 1D conclusion: 1D mean and rms is "in range"
- Reality could be : relative increase of the noise in critical/noisy regions by factor 2-3 not spotted

"DCA resolution is fine":

- 1D conclusion: TPC s DCA is 1 cm as usual
- Reality could be:
  - DCA resolution at high pt 3-4 times worse (3-4 mm instead of the 1 mm)
  - DCA is biased as function of phi

"TPC-ITS matching is fine":

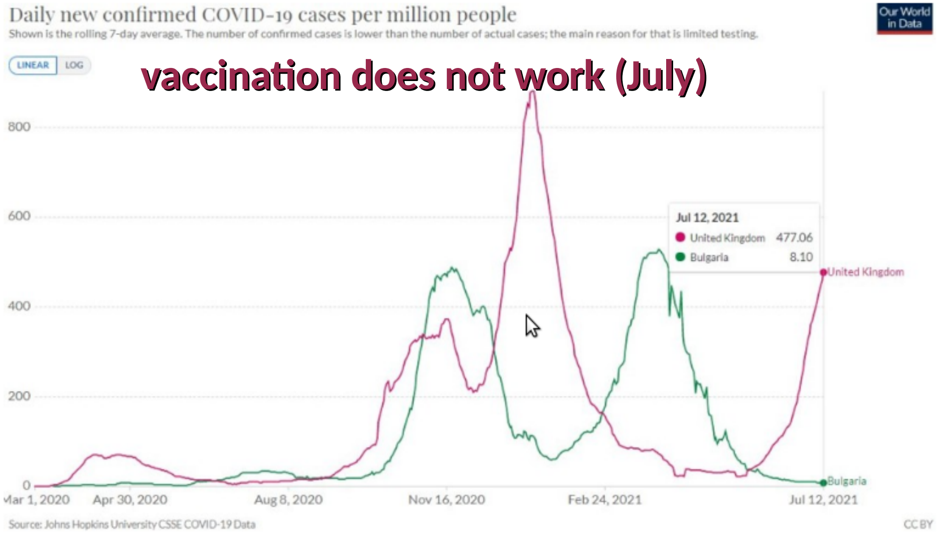
- 1D conclusion: mean TPC-ITS matching is stable
- Reality could be: local spikes in time and space. LHC15o sector 2 misbehaving (closeto zero TPC → ITS matching) in some time intervals - due distortion fluctuation

"dEdx bias for pile-up event is small":

- 1D conclusion: 0.2 sigma bias
- Reality (4D maps) could be: up to 2 sigma bias - data not suitable for some categories of analysis

vaccination works everywhere

vaccination does not work (July)



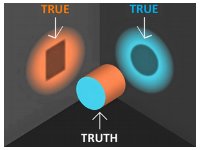
Mesíkovo porovnanie: Počet prípadov covid-19 – Veľká Británia vs. Bulharsko v júli 2021. Všimnite si, že Bulharsku práve skončila jedna vlna – delta tam ešte neprišla. Na druhej strane, vo Veľkej Británii už začínala zúriť delta. Ak by sme rovnaké porovnanie spravili pár mesiacov skôr, Bulharsko by bolo na vrchole a Británia by práve mala počty nízke.

Left side: In vaccinated UK more COVID cases than Bulgaria → vaccination does not work

Right side: differential comparison (countries, age, time,norm.) → vaccination works (everywhere)

Data are correct in both cases (+-error). Interpretation on left is wrong (as seen in November)

Properly normalized data in many dimensions → proper conclussion



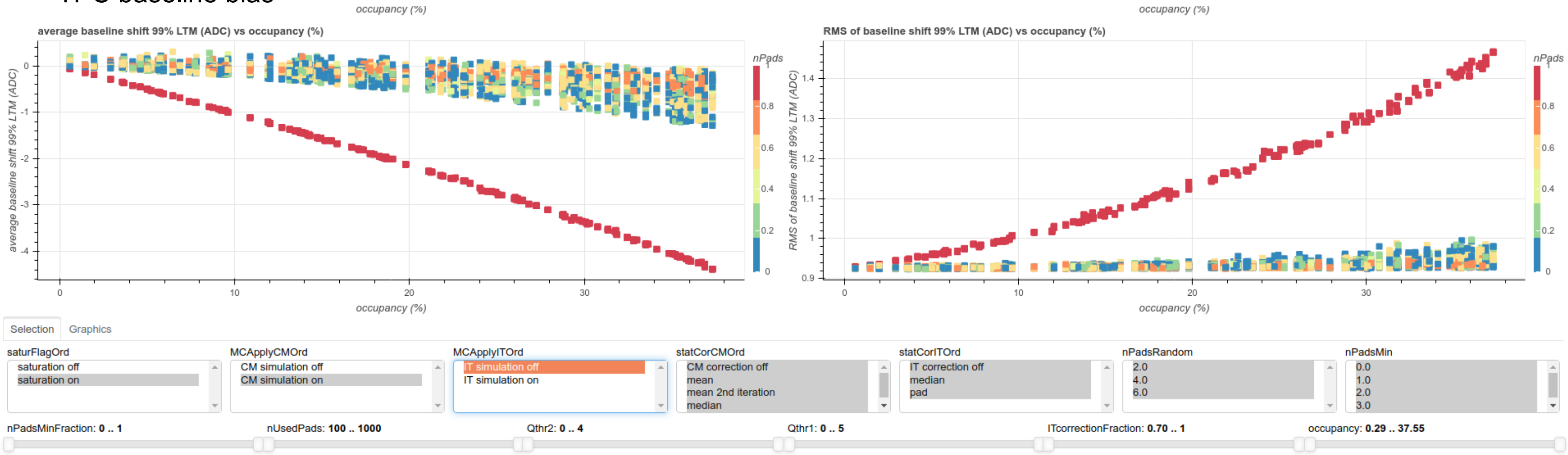
Source: https://dennikn.sk/blog/2636899/vakciny-funguju-vsade/

Optimization of the CRU digital signal processing (13 parameters in example) needed for PID and data volume optimization O(200000) settings simulated/generated

- parameters: effects (On/Off), algorithm (different version), parameters of individual algorithms
- simulation and visualization job done by master student, very effective for education
- enabling very constructive interactive discussion within CRU group, quickly converging to “expert” decision, generating new ideas
- editorial board - dashboards to be used as a support material for internal/public notes

TPC baseline bias

TPC baseline fluctuation - rms



parameters widgets

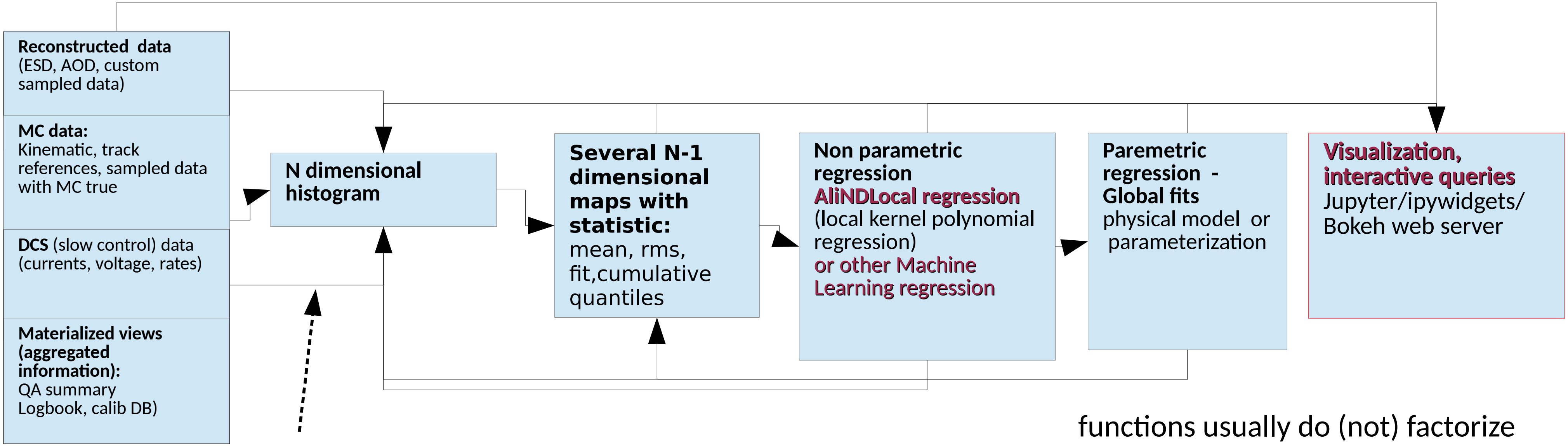
https://indico.c
https://gitlab.cern.ch/aliceeb/TPC/-/blob/master/simulationScan/toyMCParameterScan.html
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/master/JIRA/ATO-559/parameterScan.ipynb
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/master/JIRA/ATO-559/ionTailFitsAnalysis.ipynb
https://indico.cern.ch/event/1073883/contributions/4588170/attachments/2334149/3986420/simulScan_02112021.html

https://indico.cern.ch/event/1135398/#sc-3-4-digital-signal-processi

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

Standard ND pipeline (0) + RootInteractive



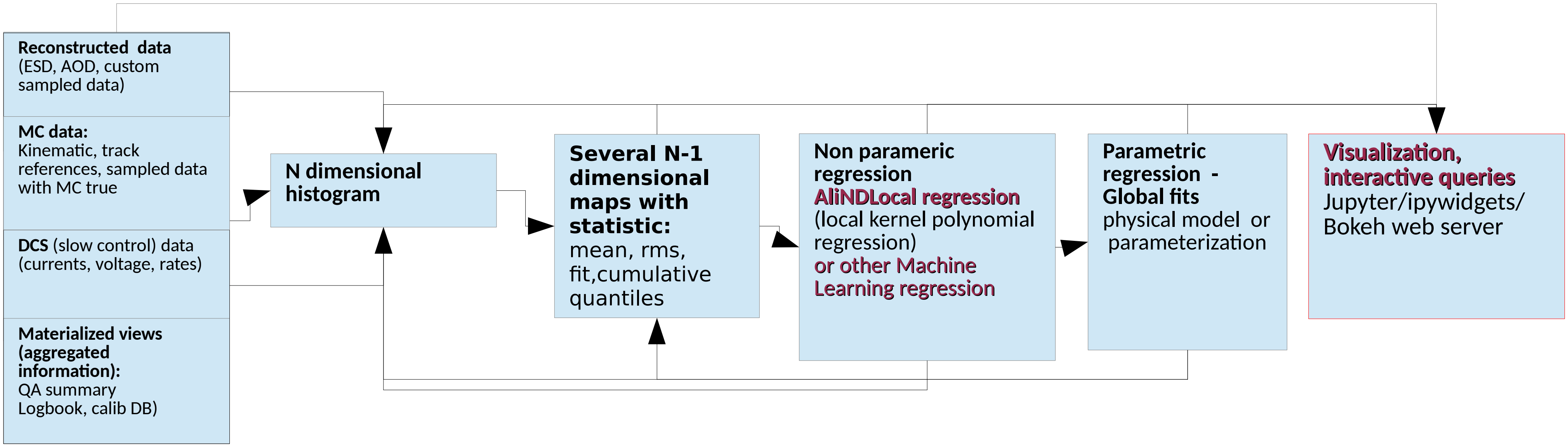
functions usually do (not) factorize

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

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
  - Some steps can be skipped, e.g local regression (Pytho ML/AliNDLocal) can be done using un-binned input data
  - Histogramming in case of non sparse data
- **ML for sparse (going to higher dimensions)**

Standard ND pipeline (0) + RootInteractive



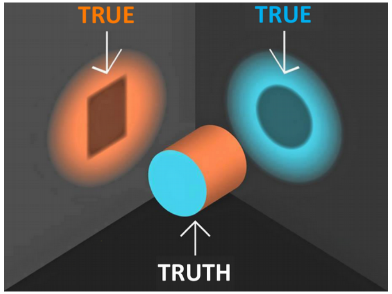
Pipeline of standalone tools

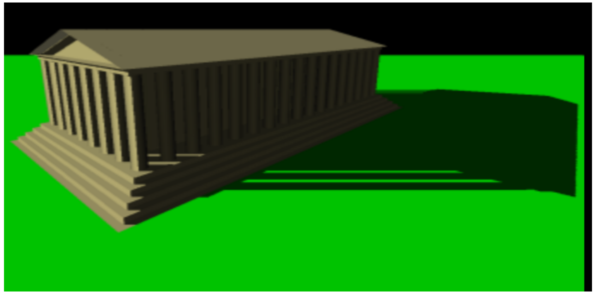
- N dimensional histogramming
  - (default)C++ and (experimental)RootInteractive
- Histogram → PDF Map (tree,panda)
  - (default) C++ and (experimental) RootInteractive
- Map(Tree) → Local regression
- Map(Tree) → Global fits (physical models, parameterizations)
  - AliTMinuitToolkit (working on TensorFlow, PyTorch)

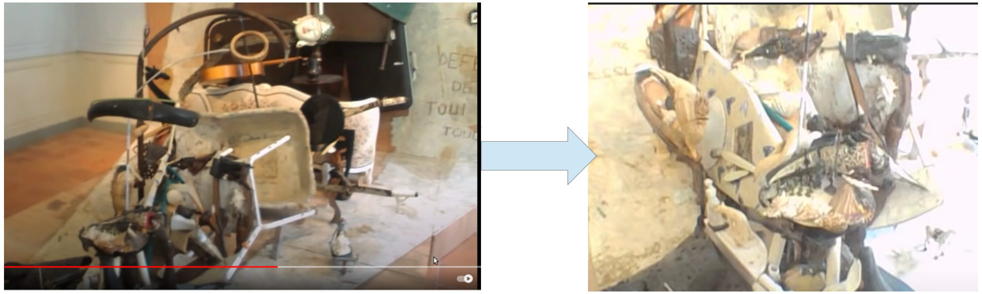
Generic “interactive” code. Minimizing amount of custom macros.
Standardizing ND analysis/functional representation
“Declarative” programming - simple queries

Working with function and functional composition
Interchanging between non parametric and parametric representation (physics models)

ND analysis vs shadow projections







Expo Bernard PRAS - Palais du Facteur Cheval
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

Let the data speak for itself

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

Example data normalization - Hunting for SC distortion origin (2015)

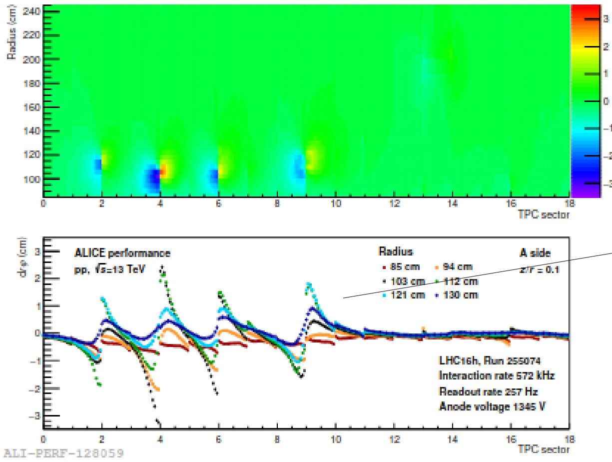
$$\sigma_{\vec{A}-\vec{A}_{ref}} < \sigma_{\vec{A}} (+) \sigma_{\vec{A}_{ref}}$$

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

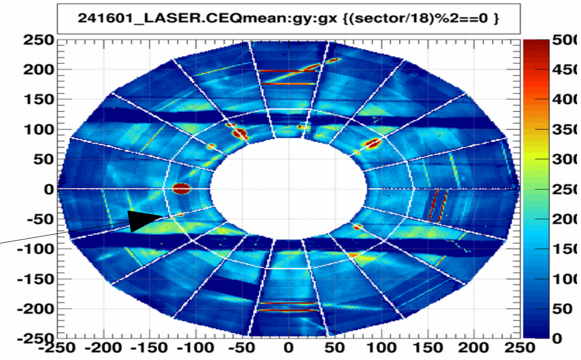
Center of gravity closer to sector gap **(inside)** than inner edge of affected chamber

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

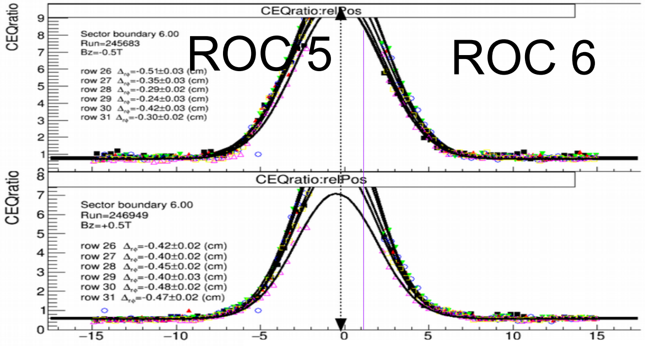
- fit indicates position of the space charge → distortion origin in gap inside
- for MB and TB - **result not yet convincing** for hardware intervention - **higher precision needed**



TPC distortion 2015



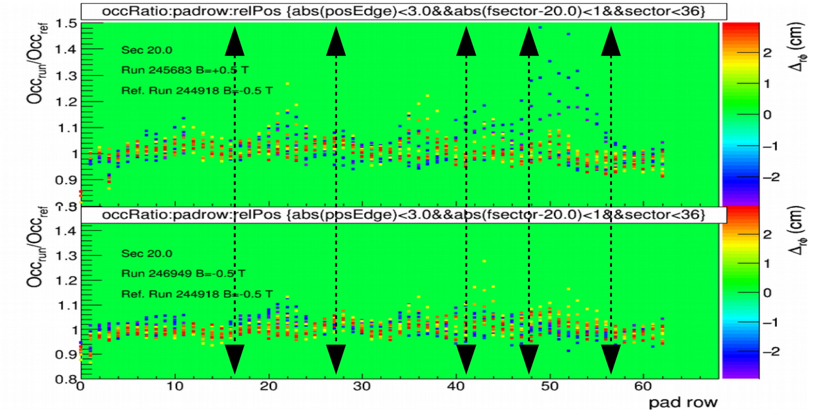
Laser calibration -Ion deposited on CE decrease work function → Increased emission of electrons during laser shots

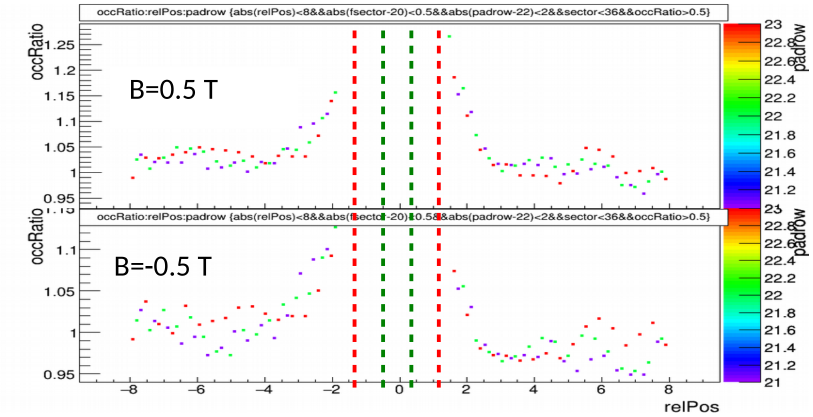


Distance to sector boundary (cm)

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:

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

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

$\overline{Z} \approx 125cm$ Δrφ (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

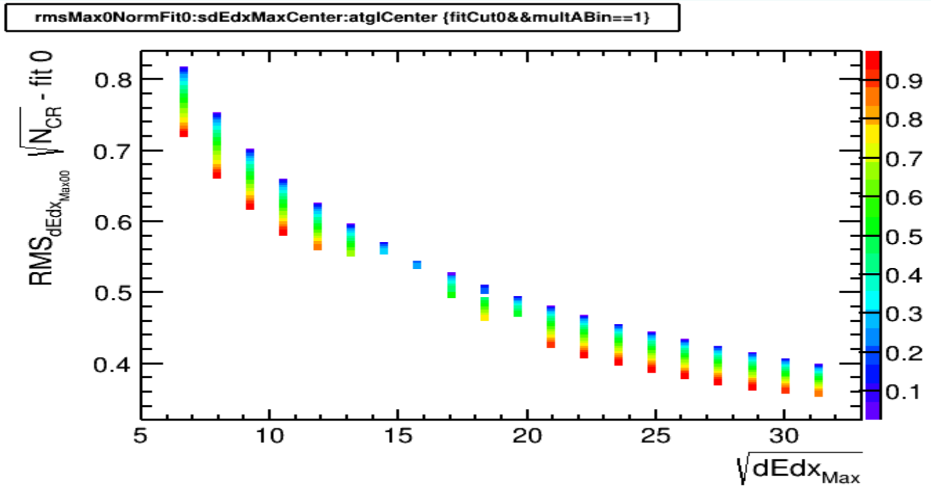
MC-Data comparison - should be done in N dimension not on projections

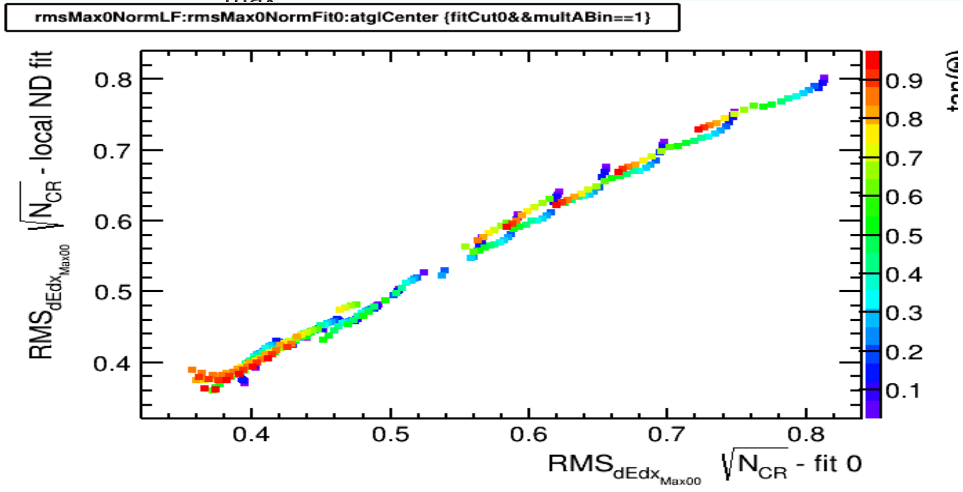
Aggreagation/projections of normalized data in NDimensions

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

PID Derived variables example: dEdx Q_{max} ROC0 at Low occu.





Physical model:

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

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

Input data pipeline:

skimmed data → 3x4D histograms of dEdx ratios in regions → 3x3D resolution maps (non parametric) → local fits → global fit of physical model

At low IR agreement between dEdx intrinsic resolution and power low model

- p0 =1.96 +-0.01, p1=0.233+-0.001, p2=0.40+-0.02

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

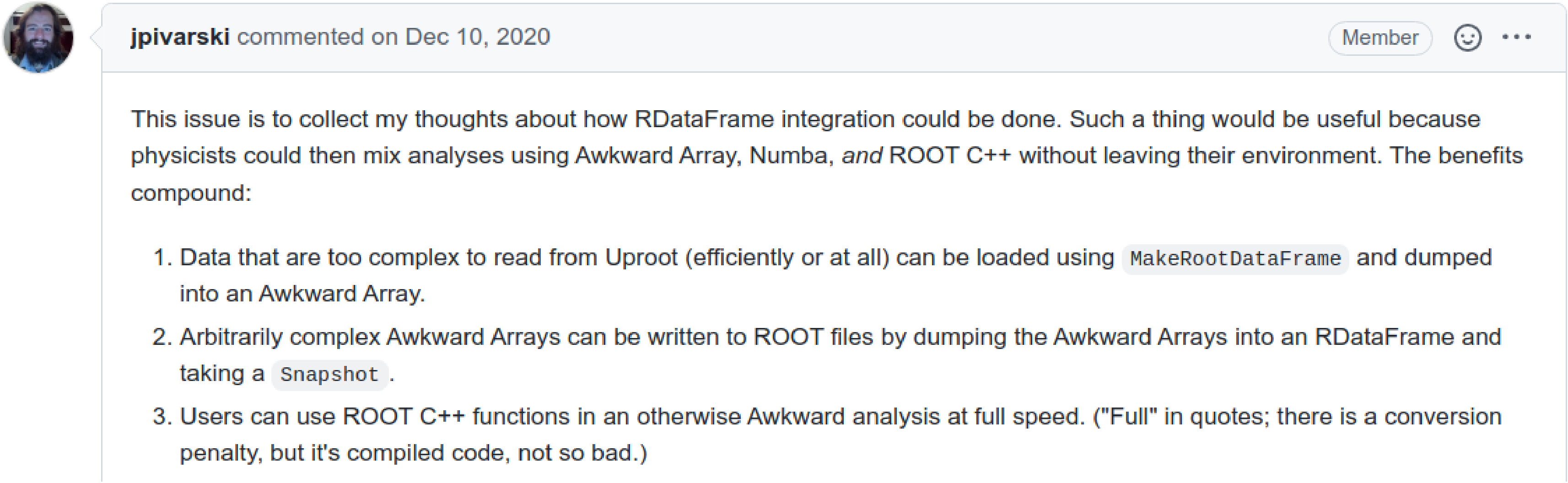
Root/TTree interface wrappers:

- 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 (beta tester) in anticipation of RDataFrame ↔ awkward interface.
  - https://github.com/scikit-hep/awkward-1.0/pull/1295

Data skimming/representative samples and triggers

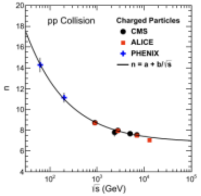
- To enable rapid development/feedback loop/interactivity, special representative data samples are usually used.
- Not part of RootInteractive, but often used as input in our typical use cases.

Run3 - Data skimming -

Run1/2 skimming triggers

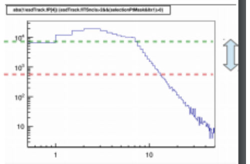
Data down-sampling to prepare representative sample flat in variable of interest

- 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)

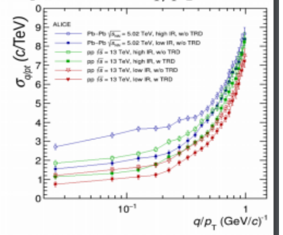
Almost all (my) reconstruction/PID debugging

- 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

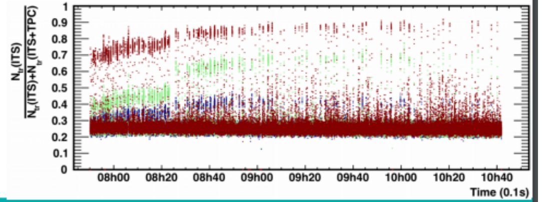
PID calibration and performance studies

- 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

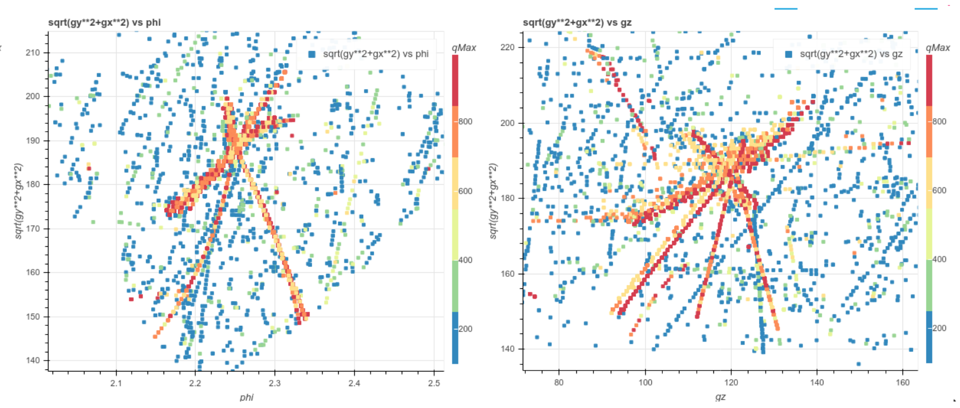
Run3 data to be skimmed in the similar way as Run1,2 (data down-sampled by factors 10^3)

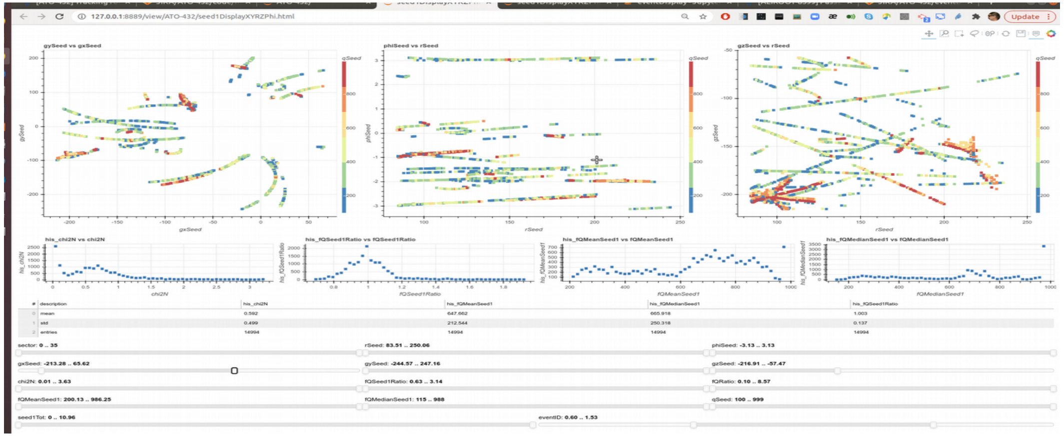
- https://indico.cern.ch/event/1014566/contributions/4272119/attachments/2209987/3743263/ATO-465-DataSkimmingPerfCalPhysicsRun2Run3.pdf
- **small server instead of farm to analyze the data**

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

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

Interactive histogramming - PID QA

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

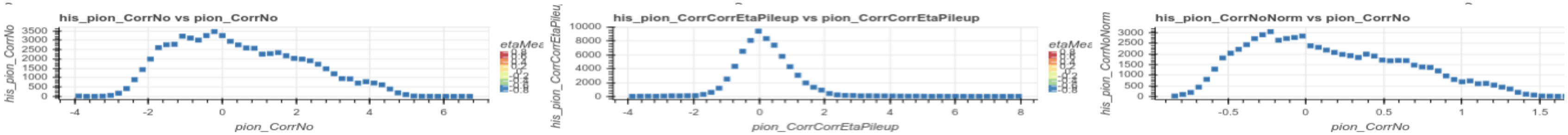
- https://github.com/miranov25/RootInteractive/issues/123

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

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

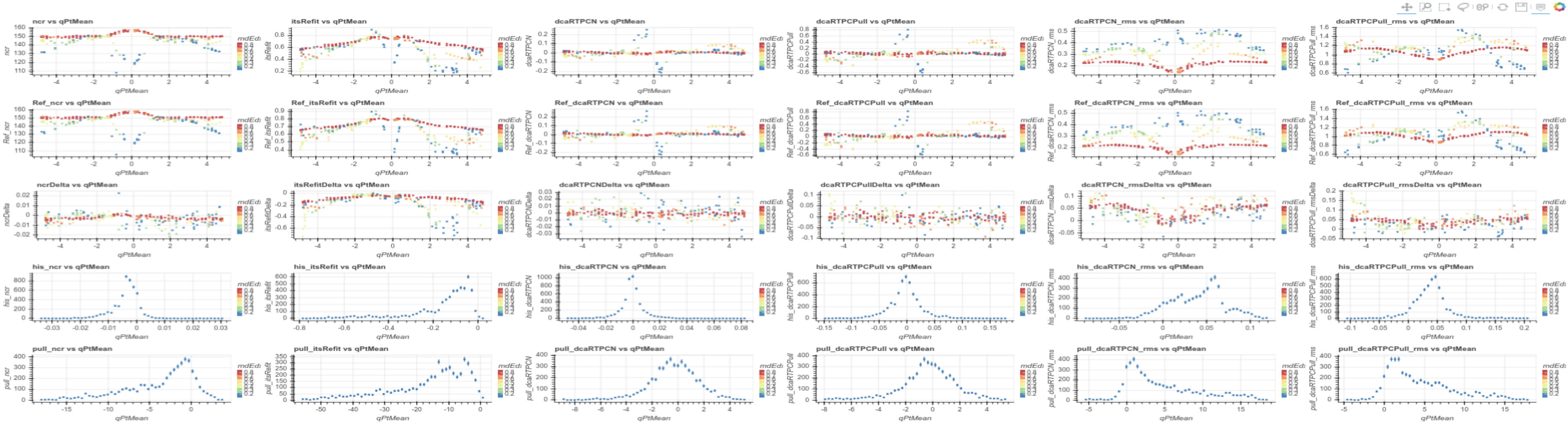
Data

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

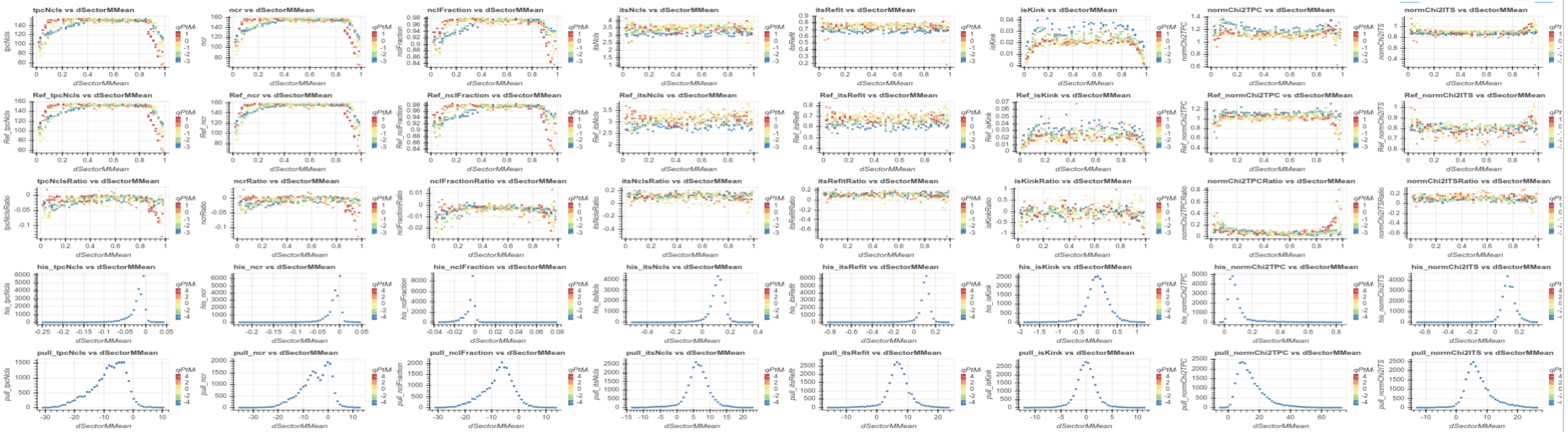
Data

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

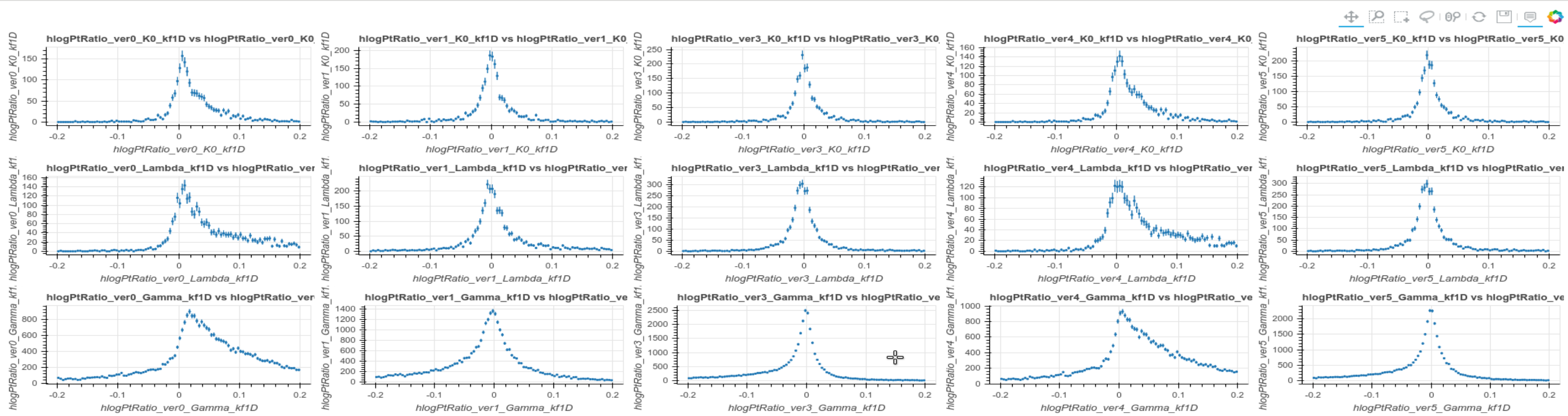
- providing summary dashboards as support material in agenda, expert hands-on session save several weeks iterations
- 6+1D (algorithm, is primary flag, 1/pt, multiplicity,pz/pt, decay radius, mother radius)
- **Optimal Hybrid V0/cascade finder** (proper material budget correction, optimal co-variance, 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	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	0.026	0.003	0.003	0.025	0.004	0.047	0.011	5.358e-4	0.044	0.001	0.043	-1.318e-2	-2.139e-2	0.034	-1.725e-2
1	std	0.043	0.041	0.034	0.043	0.035	0.058	0.055	0.04	0.06	0.041	0.079	0.07	0.058	0.079	0.06
2	entries	1877	1932	1977	1877	1977	2620	3061	3412	2637	3406	29120	30602	31602	29189	31708

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

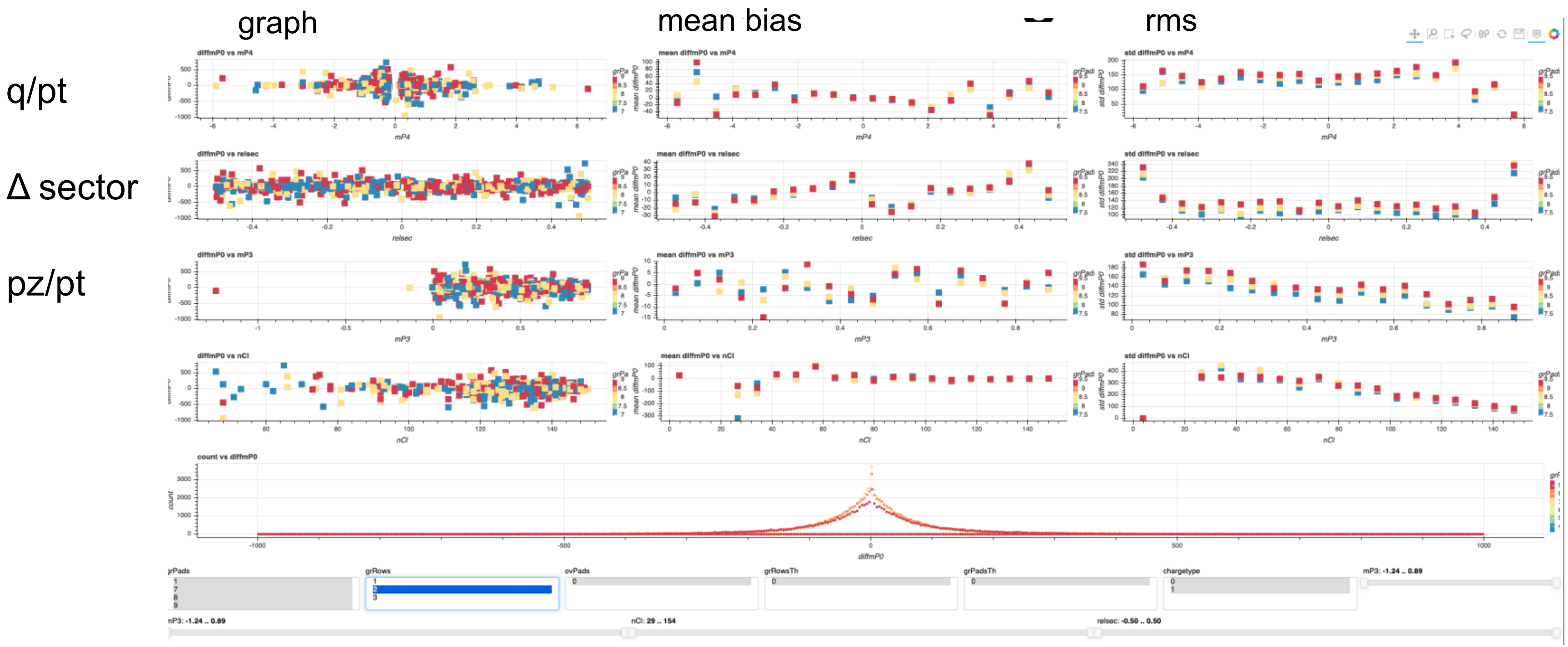
isPrimMC: 0.0, 1.0   RMC: **20 .. 120**   motherR: **0 .. 79.98**

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

Fast feedback →
Very constructive discussion

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

- 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



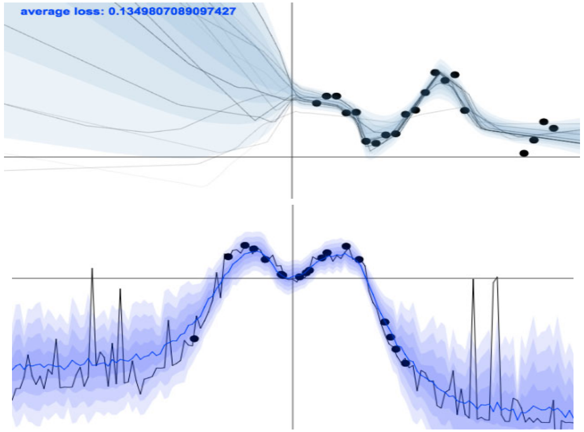
https://indico.cern.ch/event/1091510/
https://indico.cern.ch/event/1091510/contributions/4599999/attachments/2338476/3986580/residualTrackParam.html

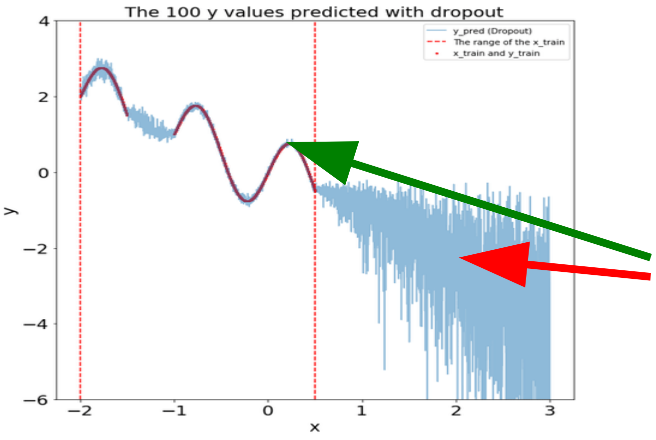
https://indico.cern.ch/event/1135398/#sc-1-1-space-charge-distrotion

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

http://www.cs.ox.ac.uk/people/yarin.gal/website/blog_images/reg_demo_small.jpg

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

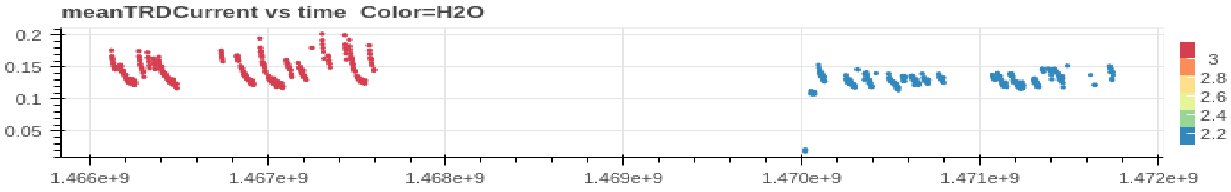


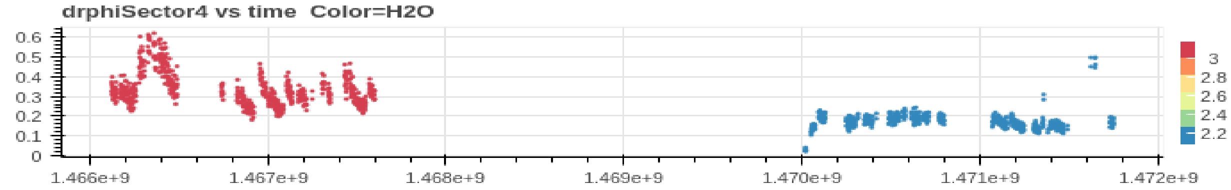


Knowledge of errors and PDF critical for data interpretation

- 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





Example:
Alice example time series flux, gas composition and distortion

What is the prediction error for non seen data ?

Current code: Uncertainties estimator using dropout resp, mean, median, rms, PDF from random trees

- Used in part of the analysis (see e.g distortion calibration contribution)

New developments for the **generalized regression forest**

- Extending standard interface for predict
- Using local linear forest to get smoother representation
  - needed e.g. for the current → space charge, or multiplicity estimation (V0, T0 → Pixel)
- generalized random forests R → python
  - https://grf-labs.github.io/grf/index.html
  - https://github.com/miranov25/RootInteractive/blob/940e483ae455e98e521b6c5597799f8115b307e3/RootInteractive/MLpipeline/local_linear_forest.py#L128

Applying for the ML workshop - Dealing with Uncertainties

- https://indico.ijclab.in2p3.fr/event/5999/

Ongoing analysis:

- Distortion

Error estimates - Work in progress

Error estimates and PDF extracted locally

- data std = std(data-prediction)
  - tests OK
- reducible error sqrt(f)/(1-f) std(tree permutation)
  - estimated using permutation of subset (fraction f) of trees trained on independent samples
  - statistical error of the prediction - test OK
- irreducible error = sqrt(std^2-sred^2)
  - estimator of intrinsic noise
  - test OK
- model error
  - to define

User interface:

- predicRFStat(X,Y,statDictionary,)
  - “mean”,”median,”rms”
  - reducible error
- https://github.com/miranov25/RootInteractive/blob/940e483ae455e98e521b6c5597799f8115b307e3/RootInteractive/MLpipeline/MIForestErrPDF.py#L142

Extract PDF for RF residuals (RootInteractive wrapper)

In [267]:
```
%%time
statDictionary={"mean":0, "std":0, "median":0, "quantile":[0.1,0.5,0.9]}
#dfSample=df.sample(200000).sort_index()
#xxx = predictRFStat(regressorRFF0,dfSample[variablesRF].to_numpy(dtype=np.float32),statDictionary,100)
predictRFF0=predictRFStat(regressorRFF0,df[variablesRF].to_numpy(dtype=np.float32),statDictionary,n_jobs*2)
df["flucCorrR_Mean"]=predictRFF0["mean"]
df["flucCorrR_Median"]=predictRFF0["median"]
df["flucCorrR_std"]=predictRFF0["std"]
df["flucCorrR_F10"]=predictRFF0["quantiles"][0.1]
df["flucCorrR_F90"]=predictRFF0["quantiles"][0.9]
predictRFF1=predictRFStat(regressorRFF1,df[variablesRF].to_numpy(dtype=np.float32),statDictionary,n_jobs*2)
df["LinearDeltaBase_Mean"]=predictRFF1["mean"]
df["LinearDeltaBase_Median"]=predictRFF1["median"]
df["LinearDeltaBase_std"]=predictRFF1["std"]
df["LinearDeltaBase_F10"]=predictRFF1["quantiles"][0.1]
df["LinearDeltaBase_F90"]=predictRFF1["quantiles"][0.9]
predictRFF2=predictRFStat(regressorRFF2,df[variablesRF].to_numpy(dtype=np.float32),statDictionary,n_jobs*2)
df["LinearDelta_Mean"]=predictRFF2["mean"]
df["LinearDelta_Median"]=predictRFF2["median"]
df["LinearDelta_std"]=predictRFF2["std"]
df["LinearDelta_F10"]=predictRFF2["quantiles"][0.1]
df["LinearDelta_F90"]=predictRFF2["quantiles"][0.9]
print(df["flucCorrR_std"].mean(), df["LinearDeltaBase_std"].mean(), df["LinearDelta_std"].mean())
```

```
CPU times: user 5min 1s, sys: 6.57 s, total: 5min 7s
Wall time: 1min 23s
```

Example usage:

https://indico.cern.ch/event/1135398/#sc-3-1-tpc-space-charge-1d-cur

Supporting references.

* Only part accessible to not ALICE members

References - articles

- ## 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

References - presentation

- 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/