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 - Yota, Mesut, Marian (CERN, Yale)
- Run2 Performance web pages - Pritam, Dibakar, Tulika, Marian (Kolkata)
- PID calibration and dEdx optimization - 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)
- fastMCsalman - Marian, Federico (Oxford)
- Run3 TPCQA/ QC - Marian, Berkin
- data skimming - Mi, Mesut, Berkin
- TPC data volume studies - Marian, Marian Jr, Mesut

https://github.com/miranov25/RootInteractive/
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 $\rho_{\text{SC}}$ → Large space-charge distortions (dr, d$\phi$, dz) of measured space points O(5 cm)
- → Space-charge density and distortion fluctuations O(5 %) ~ 0.2 cm
- To be calibrated/corrected to $\sigma$ ~100 $\mu$m with granularity O(10^6) in space O(1-5 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*
Outlook

ALICE goals and challenges

NDimensional interactive analysis - **Seeing is believing**
- Motivation and Interactive n dimensional parameter optimization example

- 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)
- Triggers and data skimming (representative data selection) to enable interactive analysis

ND+RootInteractive functionality shown in real use cases - Tutorial (https://indico.cern.ch/event/1135398/)
- 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), trackCombinator(Kink/V0/Cascade/Cosmic/Looper)
- Highly customizable event display (Run3, ALICE3 reconstruction, high dEdx, spallation, (Magnetic monopole) tracking)
- Reconstruction Performance web page Data/MC for Run1,2
- MC/data remapping for the Machine learning
- fastMCKalman
- Data volume investigation
- QA/QC validation for the Run3
- ....

Interactive differential studies, physics analysis using skimmed/samled data
- Particle production as a function of combined multiplicity estimators (and event shape) in pp, pPb and A-A collisions with ALICE (UK Bratislava)
"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."

"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 ND 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
Content

ND - pipeline description and motivation on real live examples

- Multidimensional parameter optimization
- Non parametereric and parameteric/analytical models
- Functional composition examples
- Invariances example for the calibration and automatic QA
- RootInteractive/MultiInteractive session

RootInteractive functionality explained (data volume studies examples)

- 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 parameterization
- ONNX interface on client (starting)

Representative Sampling/skimming
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 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 ND local 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
Optimization of the digital signal processing (13 parameters in example) needed for particle identification 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 expert group, quickly converging to “expert” decision, generating new ideas
- standalone dashboards as a support material for internal/public notes

TPC baseline bias

TPC baseline fluctuation - rms

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
Physical model:

- dEdx resolution depends on 3 main variables: dEdx, track length ($\tan(\theta)$) and number of measurement ($N_{CR}$)
- 3 measurements in regions (i,j,k)

\[
RM_S^{Q_i} = \sqrt{RM_S^{Q_i/Q_j} + RM_S^{Q_i/Q_k} - RM_S^{Q_j/Q_k}/2}
\]

\[
RM_S^{ROC} \times \sqrt{N_{CR}} \approx p_0 \left( dEdx^{p_1} \times \sqrt{(1 + \tan(\theta))^{p_2}} \right)
\]

Input data pipeline:

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

At low IR agreement between dEdx intrinsic resolution and power low model as expected
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 $\sigma_{A_{\text{ref}}} > \sigma_{A_{\text{obj}}}$ is very often smaller by orders of magnitude

https://www.youtube.com/watch?v=a7LCTT7HKzc
Example data normalization - Hunting for SC distortion origin (2015)

\[ \sigma \vec{A} - \vec{A}_{ref} < \sigma \vec{A} (\pm) \sigma \vec{A}_{ref} \]

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**

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)

\[ \sigma \hat{A} \otimes \hat{A}_{ref} \leq \sigma \hat{A} (+) \sigma \hat{A}_{ref} \]

Analytical model - derivative of E field due line charge:

\[
\frac{N_{CI}(IR)}{N_{CI}(IR=0)} = \frac{(w + (\Delta r_\phi (r_\phi + w/2) - \Delta r_\phi (r_\phi - w/2)))}{w} \\
R = \left( \frac{\text{Occ}}{\text{Occ}_\text{ROC}} \right)_{IR} / \left( \frac{\text{Occ}}{\text{Occ}_\text{ROC}} \right)_{IR=0}
\]

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
Data should be compared with reference model/data in multidimension

- 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), B field symmetry
- data - non parameteric/parameteric physical model
- smoothness resp. local smoothness
- Outlier tagging e.g. (data-model) > N σ

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

**Aggregation/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-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

\[
\sigma \vec{A} \otimes \vec{A}_{ref} \leq \sigma \vec{A} (+) \sigma \vec{A}_{ref}
\]
Expert data preparation

- Agreement on data to collect
- Data sources
- Variables to import
- Pre-aggregation
- Data sampling

Data presentation

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

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

- Visualization and on client aggregation Python/TrueScript (RootInteractive)
- Machine learning wrappers Python
- PyRoot used to be able to use Root and RootInteractive together

- Explained on the TPC data volume calibration/QA Example - building application in your browser - calibration/QA ND viewer and track ND viewer

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 example - Calib/QA/QC RI+ML expert viewer

Figure tabs:
- 2D histo
- 2D projections (median|mean|rms)
- 3D Global XY view
- 3D LocalXY view

Selection tabs:
- selection
- histogram variable binning
- graphic properties

Functional composition, Machine learning regression, Interactive histogramming and interactive data aggregation in your browser (javascript).

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
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)

Parameters imported from the O2 CCDB and QCDB
Current implementation  TTree  query based

https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/ae4136c6f587e55482373252e2f1c4597fe4f606/IRRA/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)**

**Defining RDataFrame**

```c
ROOT::RDataFrame df(nTracks);
auto rdf = df.Define("qVector", "getQVector(i16)"
 .Define("loggVector", "ROOT::VecOps::log(qVector)"
 .Define("qStd", "StdDev(qVector)"
 .Define("qMean", "Mean(qVector)"
 .Define("qIStd", "StdDev(lgqVector)"
 .Define("qIMean", "Mean(lgqVector)"
 .Define("qMedian", "TMath::Median(qVector.size(), qVector.data())"
 .Define("qlMedian", "TMath::Median(qVector.size(), lgqVector.data())"
 .Define("qltrunc", "truncate(qVector)"
 .Define("loglTrunc", "ROOT::VecOps::log(lgTrunc);"
```

**Loading awkward array**

```python
In [7]:
array = ak.from_rdataframe(rdf,
columns=
    "loggTrunc",
    "loggVector",
    "qMean",
    "qMedian",
    "qStd",
    "qltrunc",
    "qVector",
    "qMedian",
    "qlMedian",
    "qlStd",
)
```

**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
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

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

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

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

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
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/ae4136c6f587e55482373252e2f1c4597e4f606/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
Customizable 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 weights

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)

Representative Sampling/skimming
Run1/2 skimming triggers

- **Data down-sampling** to prepare representative sample flat in variable of interest
  - Global T MLS fits used to estimate particle production
  - https://alice.its.cern.ch/jira/browse/ATO-465

Run1/2 topology horizontal down-sampling:
- Charged (AliESE,track) tracks down-sampling triggers
  - flat pt trigger, flat q/pt trigger, MB
- V0 trigger (Gamma, K, λ):
  - flat pt trigger, flat q/pt trigger, MB
- Nuclei (A>1)
  - primaries
  - down-sampled secondaries
- Cosmic track pairs:
  - "random cosmic" 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}

Run3 data to be sampled/skimmed in the similar way as Run1,2 (data down-sampled by factors 10^3)
- small server instead of farm to analyze the data
- Public node: https://alice-notes.web.cern.ch/node/1208

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://aliperfo.web.cern.ch/aliperfo/alice/2022
  - PassX/PassY, MC/data, PeriodX/PeriodY
  - MC/data tuning/remapping
  - Track matching/Efficiency/Inv.Mass/Material budget/Cross sections
- 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
RootInteractive Sampling - Run3 prototype studies

- 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
Example: Sampled/skimmed efficiency and MC trees

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

- $K^+, K^-, \pi^+, \pi^-$
- kinks
- $K^0, \gamma, \Lambda, \mu, \pi^0$
- $\Lambda, \Sigma, \Xi, \Omega$
- $\Lambda_c$ ...

**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 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
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 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
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 $\sigma_{A - A_{ref}}$ is very often smaller by orders of magnitude
Example data normalization - Hunting for SC distortion origin (2015)

\[ \sigma \bar{A} - \bar{A}_{ref} < \sigma \bar{A}(+)\sigma \bar{A}_{ref} \]

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

Laser calibration - Ion deposited on CE decrease work function → Increased emission of electrons during laser shots
Example data normalization - Hunting for distortion origin (2015)

\[ \sigma_{\Delta \Theta \Delta r_{ref}} \leq \sigma \Delta (+) \sigma_{\Delta r_{ref}} \]

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

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

Analytical model - derivative of E field due line charge:

\[ \frac{N_{CI}(IR)}{N_{CI}(IR=0)} = \frac{w + (\Delta r_{\phi} (r_{\phi} + w/2) - \Delta r_{\phi} (r_{\phi} - w/2))}{w} \]

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

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

Conclusion:

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

Analytical model - derivative of E field due line charge:

\[ \frac{N_{CI}(IR)}{N_{CI}(IR=0)} = \frac{w + (\Delta r_{\phi} (r_{\phi} + w/2) - \Delta r_{\phi} (r_{\phi} - w/2))}{w} \]

\[ R = \left( \frac{Occ}{<Occ_{ROC}>} \right)_{IR} \left/ \left( \frac{Occ}{<Occ_{ROC}>} \right)_{IR=0} \right. \]
QA consideration: ND analysis vs Shadow projections - alarms and invariants

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

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

Aggregation/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-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 & RootInteractive update and plans

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

ATO-563: Machine learning wrappers


ATO-575: uniform data sampling/skimming for Run3

PWGPP-613, fastMCKalman for the reconstruction 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

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.
Special wrappers for the ROOT input trees (aliTreePlayer), based on the old ROOT tree interface, 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

**Run3 data to be sampled/skimmed in the similar way as Run1,2 (data down-sampled by factors $10^3$)**
- small server instead of farm to analyze the data
- Public node: https://alice-notes.web.cern.ch/node/1208
Run3 prototyping studies (MI, Mesut, Berkin, Jens, Ruben)

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

- interactive histogram, scatters, sliders, summary aggregated information
Customizable event display (in Jupyter notebook or plain python):

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

[Links]
https://gitlab.cern.ch/alice-tpc-offline/alice-tpc-notes/-/blob/264a6fb497b0c1a601b7aaf65a45d25546441f/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": 59),
    ("name": "his_pion_CorrEtaPrime", "variables": ["pion_CorrEtaPrime"], "nbins": 59),
    ("name": "his_pion_CorrNoNorm", "variables": ["pion_CorrNo/pion_CorrNoRMS"], "nbins": 59),
    ("name": "his_pion_CorrEtaPrimeNorm", "variables": ["pion_CorrEtaPrime/pion_CorrEtaPrimeRMS"], "nbins": 59),
    ("name": "his_pion_CorrNoRMS", "variables": ["pion_CorrNoRMS"], "nbins": 59),
    ("name": "his_pion_CorrEtaPrimeRMS", "variables": ["pion_CorrEtaPrimeRMS"], "nbins": 59)
]
```

PID QA dashboard histogram part snapshot

PID QA dashboard summary for selected Mean, RMS, Sum

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

```json
histoArray = {
    "name": "histoPt", "variables": ["pt"], "nbins": 100, "histograms": {
        "entries": None,
        "weights": ["isPion"],
        "entriesPion": ["weights": ["isPion"],
                        "entriesProton": ["weights": ["isProton"],
                                          "entriesLambda": ["weights": ["isLambda"],
                                                             "entriesKoon": ["weights": ["isKoon"],
                                                                              "entriesKoon1": ["weights": ["isKoon"],
                                                                                 "entriesPion": ["weights": ["isPion"],
                                                                                    "entriesProton": ["weights": ["isProton"],
                                                                                          "entriesLambda": ["weights": ["isLambda"],
                                                                                               "]}
                           },
        "name": "histoPtWeighted", "variables": ["pt"], "nbins": 100, "histograms": {
            "entries": ["weights": "weight1"],
            "entriesPion": ["weights": ["isPion1"],
                            "entriesProton": ["weights": ["isProton1"],
                                            "entriesLambda": ["weights": ["isLambda1"],
                                                             "entriesKoon": ["weights": ["isKoon"],
                                                                              "entriesKoon1": ["weights": ["isKoon"],
                                                                                 "entriesPion": ["weights": ["isPion"],
                                                                                    "entriesProton": ["weights": ["isProton"],
                                                                                          "entriesLambda": ["weights": ["isLambda"],
                                                                                               "]}
                           }
```}

Function/alias declaration

```json
aliasArray = {
    "name": ["ratioPionAll", "variables": ["entries", "entriesPion"], "func": "return entriesPion / entries", "context": "histoPt"],
    "name": ["ratioLambdaAll", "variables": ["entries", "entriesLambda"], "func": "return entriesLambda / entries", "context": "histoPt"],
    "name": ["ratioProtonAll", "variables": ["entries", "entriesProton"], "func": "return entriesProton / entries", "context": "histoPt"],
    "name": ["ratioKoonAll", "variables": ["entries", "entriesKoon"], "func": "return entriesKoon / entries", "context": "histoPt"],
    "name": ["ratioPionErr", "variables": ["entries", "entriesPion"], "func": "return Math.sqrt(entriesPion) / entries", "context": "histoPt"],
    "name": ["ratioLambdaErr", "variables": ["entries", "entriesLambda"], "func": "return Math.sqrt(entriesLambda) / entries", "context": "histoPt"],
    "name": ["ratioProtonErr", "variables": ["entries", "entriesProton"], "func": "return Math.sqrt(entriesProton) / entries", "context": "histoPt"],
    "name": ["ratioKoonErr", "variables": ["entries", "entriesKoon"], "func": "return Math.sqrt(entriesKoon) / entries", "context": "histoPt"],
    "name": "downsamplingTrigger.C#L27"
}
```

- Array of histograms for different particle species
- Reweighting of spectra
- Spectra ratios/efficiency
- parametric cut efficiency
Performance diff - ALICE performance: DCA resolution/bias

http://aliperf0.web.cern.ch/aliperf0/alice/data/2018/LHC18c/kink_3sigma_CENT_pass2/dashboard/LHC16f_lownmult_pass2/fig0/compDefaultV0DCARLHC18c_kink_3sigma_CENT_pass2LHC16f_lownmult_pass2HistComp.html

<table>
<thead>
<tr>
<th>LHCh18c / LHC16f</th>
<th>Δ norm DCA</th>
<th>Δ DCA pull</th>
<th>σ norm DCA</th>
<th>σ DCA pull</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Histo: Data-Ref

Histo: (Data-Ref)/σ

Summary table: Data-Ref

Widgets for interactive ND selection

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

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)

<table>
<thead>
<tr>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ver0</td>
<td>Offline like</td>
</tr>
<tr>
<td>ver1</td>
<td>prim → V0</td>
</tr>
<tr>
<td>ver3</td>
<td>TPC → V0</td>
</tr>
<tr>
<td>ver4</td>
<td>OnTheFly if exist Offline comp</td>
</tr>
<tr>
<td>ver5</td>
<td>OnTheFly hybrid optimal</td>
</tr>
</tbody>
</table>

- Fast feedback →
- Very constructive discussion

https://indico.cern.ch/event/1088044/contributions/4574146/attachments/2334975/3979806/hpMass_ver5_kf1D_Dashboard.html
https://indico.cern.ch/event/1088044/contributions/4574146/attachments/2334975/3979807/hscH2_ver5_kf1D_Dashboard.html
https://indico.cern.ch/event/1135398/#sc-1-3-v0-and-cascade-finder-f
Run3 - **Space charge distortion** current granularity optimization (see Matthias, Ernst)

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/1135398/#sc-1-1-space-charge-distortion
Reducible, irreducible error and Probability density function
RootInteractive ML wrappers
Why Should we Care About Uncertainty?

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?

Machine learning wrappers

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

Bootstrapping CPU consuming

To speed up - to use the internal dispersion of the prediction in ensemble 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 ($\chi^2, N_{\text{Clusters}}, \sigma_{\text{DCA}}$), 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 limited

**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.
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.

An Adaptive kernel method

\[
q_i(x_0) = \frac{1}{B} \sum_{b=1}^{B} I(x_i \in L_b(x_0)) \left[ \frac{1}{|L_b(x_0)|} \right]
\]

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.

\[
\hat{\mu}(x_0) = \arg\min_{\mu, \theta} \left\{ \sum_{i \in T} a_i(x_0)(Y_i - \mu(x_i) - (x_i - x_0)\theta(x_0))^2 + \lambda||\theta(x_0)||^2 \right\}
\]

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 → for the moment postponed
\( f(A,B,C,D) = \text{norm} \times A \times \sin(n \times 2 \times \pi \times C) + B \times \text{noise} \)

4D Uniform input

\[
\text{df} = \text{pd.DataFrame(np.random.random_sample(size=(nPoints, 4)), columns=list('ABCD'))}
\]

\[
\text{df}['B'] = \text{df}['B'] + 0.5
\]

\[
\text{df}['\text{noise}'] = \text{np.random.normal(0, stdIn, nPoints)}
\]

\[
\text{df}['\text{noise}'] += (\text{np.random.random(nPoints)<outFraction}) \times \text{np.random.normal(0, 2, nPoints)}
\]

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/MIxgboostErrPDF_n2_stdIn0.2_nPoints200000.html

https://indico.cern.ch/event/1147231/contributions/4815612/attachments/2424564/4150688/MIxgboostErrPDF_back11042022.ipynb
f(A,B,C,D) = \text{norm} \cdot A \cdot \sin(n \cdot 2\pi C) + B \cdot \sigma_{\text{noise}}

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

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

**Parameter scan** to emulate statistics requirement

- number of points, function normalization, noise ($\sigma_{ired}$), $n_{\text{sin}}$

Making function variation small in respect to intrinsic noise ($\sigma_{ired}$), effective number of points increase $\rightarrow$ 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 ($\Delta I$)
  - ion current as white noise $\rightarrow$ individual FFT coefficient independent ($\mu=0$, $\sigma_i=\sigma$)
- Output: $\Delta$ distortion
- Convolution theorem $\rightarrow$ approximation response for individual FFT current harmonics
  - convolution in 3D space $\rightarrow$ multiplication in FFT space
  - Linear fit to approximate convolution kernel
  - 1 FFT as a LinearBase, 20 most important FFT

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 $\sim$ 80 microns $\sim$ 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**
Supporting references.

* Only part accessible to not ALICE members
References - articles

**Tracking articles:**  [https://twiki.cern.ch/twiki/bin/view/ALICE/TrackingReference](https://twiki.cern.ch/twiki/bin/view/ALICE/TrackingReference)

- [A0] CHEP2003: TPC tracking  [http://inspirehep.net/record/621229](http://inspirehep.net/record/621229)
- [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](https://indico.cern.ch/event/408139/contributions/979783/attachments/815694/1117684/MarianIvanovchep06.pdf)

**ALICE: Physics Performance Report, Volume II**


**TPC TDRs:**


**TRD TDR:**

[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/](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/](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](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](https://indico.cern.ch/event/1088044/?note=177737)
    • [https://indico.cern.ch/event/757761/contributions/3183222/attachments/1738216/2812589/TRDInTracking_PhysWeek2210.pdf](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](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](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/?note=162249))
    • [https://indico.cern.ch/event/991463/contributions/4343481/attachments/2235673/3790851/stat_photon_210429_TrackingMeeting.pdf](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](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/](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/2338476/3986854/2021-11-03_IDCs.pdf)
RootInteractive and ND pipeline:

  - https://indico.cern.ch/event/1091321/contributions/4612911/


- [P12] WP7 QA meeting (2020)

  - https://indico.cern.ch/event/806602/contributions/3379555/attachments/1824640/2995393/NDimensionalPipeline_OFFLWEENEEK05042019.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
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:

- 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/