N dimensional analysis pipeline Offline QC analysis in ALICE

Marian Ivanov

Outlook

Comparison of different methods:

- Thasseography and Shadow projections
- Multidimensional analysis

Multidimensional analysis using **ND pipeline** + MVA interface

Recent examples of pipeline

- Space charge distortion calibration and studies **(distortion mitigated)**
- Tracking performance characterization
- TRD tracking commissioning/optimization (**TRD in default tracking**)
- New event and track selection cuts (dEdx and chi2 QA analysis)

Visualization development

Work in progress and future plans

- MVA wrappers/algorithms + Visualization development (see next talk)
- Tuning of the simulation &reconstruction parameters
  - Example in Jupyter demo (Python invoking using C++ libStat)
- Tune on data and fast MC
- Optimization of the physics analysis (jets, J/psi, mass analysis using TOF)

Tasseography (1D)

Tasseography (also known as tasseomancy or tassology) is a divination or fortune-telling method that interprets patterns tea leaves, c or 1D histograms , or wine sediments.

Spring Pouchong tea (Chinese: 包種茶; pinyin: *Bāozhòngchá*) leaves that may be used for tasseography divination

An example of a tea leaf reading showing a dog and a bird on the side of the cup.

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

Tasseography (1D examples)

Example wrong statements:

"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 σ_{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 detector occupancy is outside of the range":

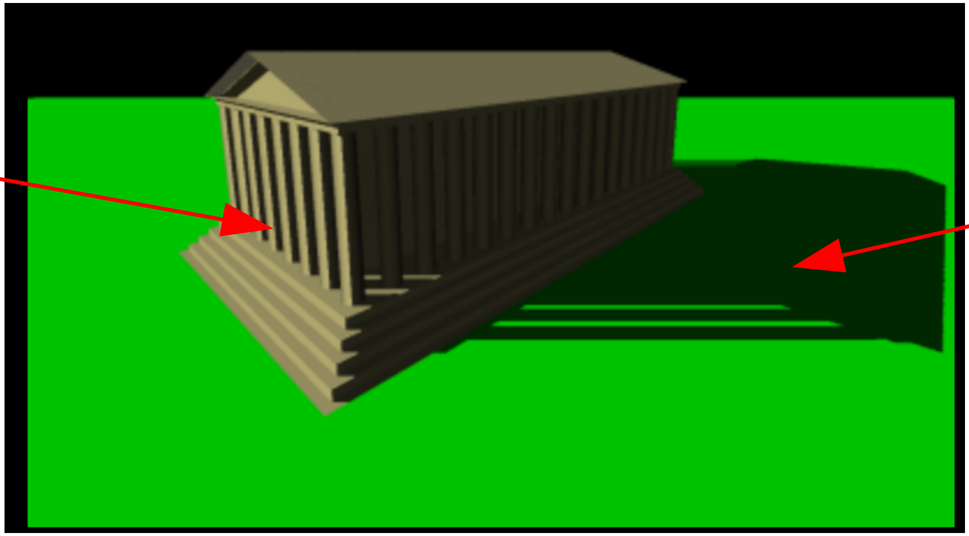
- **1D conclusion:** Mean TPC occupancy> limit
- **Reality could be:** occupancy increased because of increase of IR resp. beam background

"MC describes the data":

- **1D conclusion:** Mean mass bias in MC has the same sign as data
- **Reality could be:** The sign of bias is pure coincidence (50 % probability). Pt dependence not reproduced at all …

Shadow projections (2-Dimensional projections)

Our object
E.g
occupancy
map from
AMORE
(3D)



Our current
TPC DQM

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

Guessing from 2D projection more reliable than Tasseography

- some imagination to be involved (see next slides)

Alarms to be based on some invariance - e.g the difference between the object and referenced object

- after projection impossible
- in my typical cases variance $\sigma_{A\text{-}Aref}$ is very often smaller by orders of magnitudes

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

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

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

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

Usage of ND approach for distortion in finding of origin of the distortion

* distortion were later fully mitigated

Example: Distortion studies: Central Electrode analysis

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

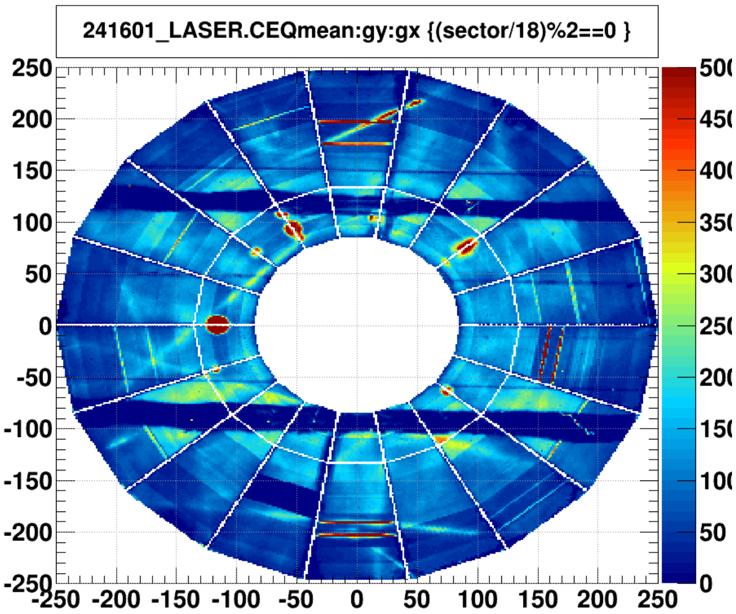
Ion deposited on CE decrease work function (online calibration/AMORE data as input)

- → Increased emission of electrons during laser shots

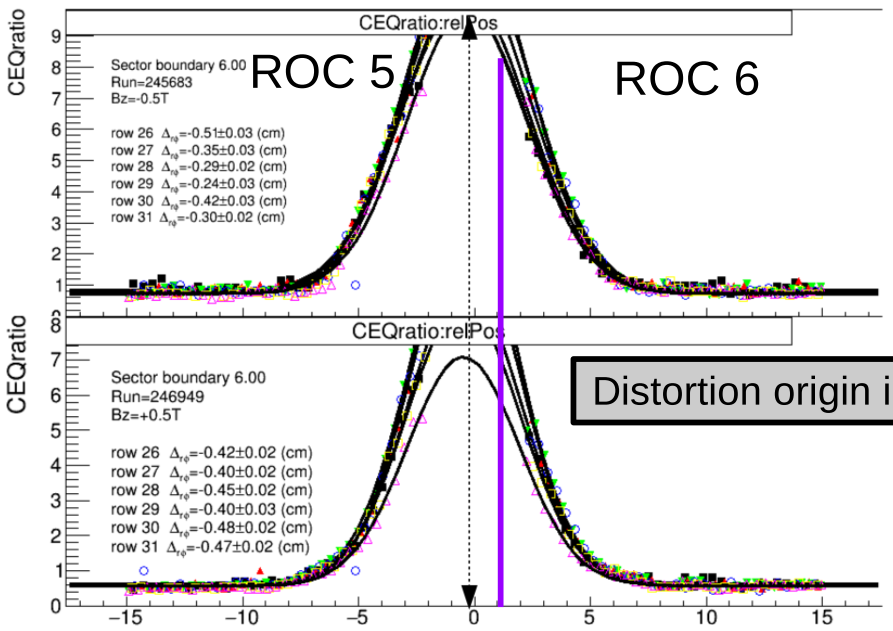
Center of gravity closer to sector gap **(inside)** than inner edge of affected chamber

Data had to be normalized to reference data set

- **after proper normalization to reference data set, fit indicates position of the space charge (inside the gap)**

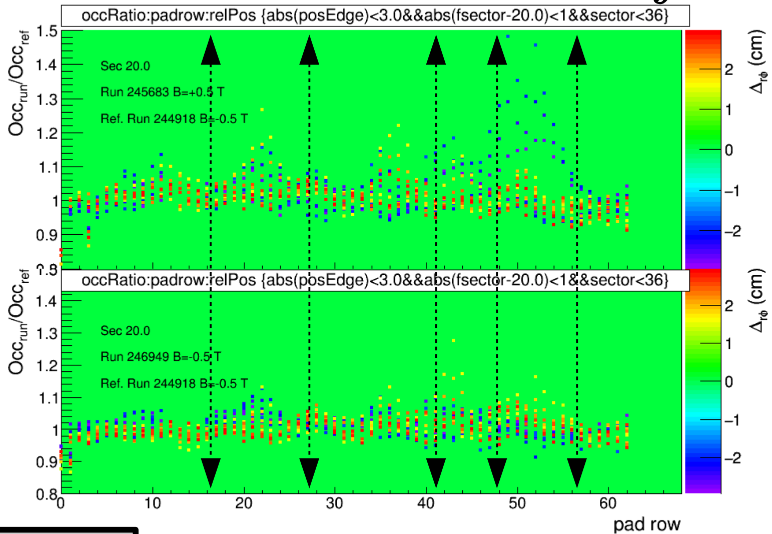


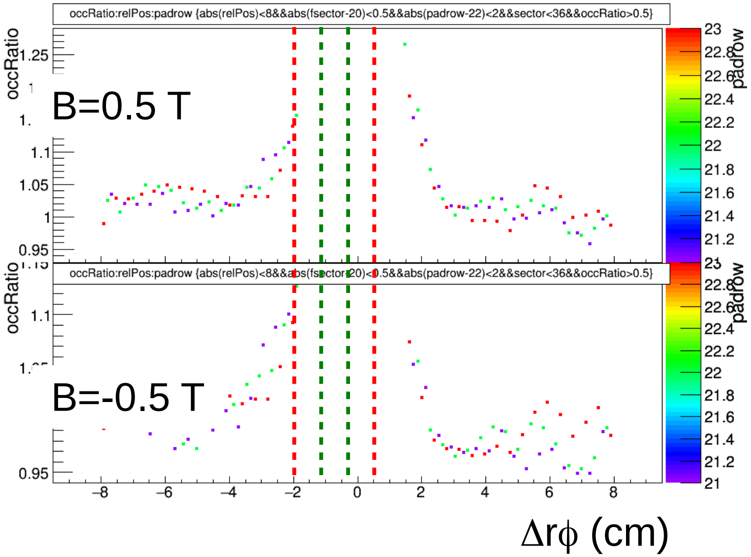
CE signal ratio - data/reference data



Example: Distortion studies. Occupancy analysis

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





Model:

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

$$\overline{Z} \approx 125cm$$

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

Conclusion: Distortion origin in the gap

Increase of occupancy close to the hot-spot region due to space-charge distortion

Very precise measurement of distortion origin - measuring **derivative of distortion** with

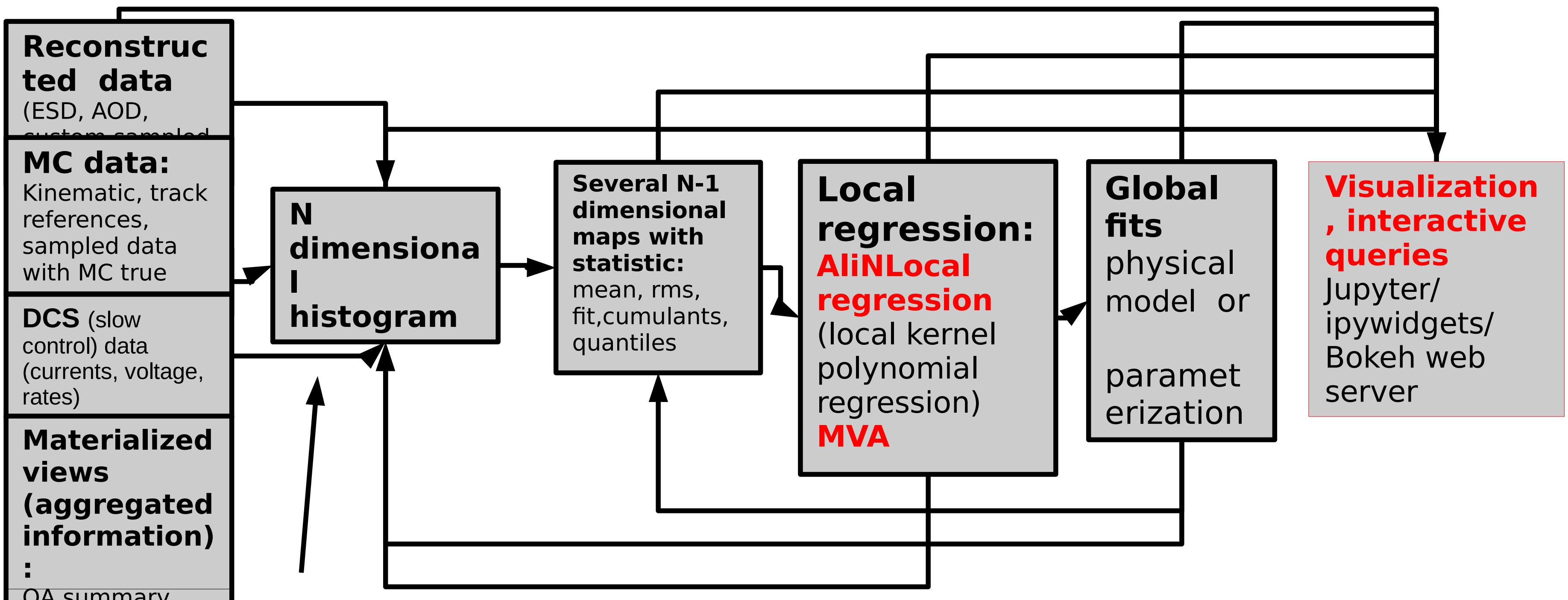
Without proper normalization to reference effect is invisible.

Wrong concussion was made in first analysis

Multidimensional analysis pipeline

- library (libStat) written in C++
- possible to use in Python

Standard ND pipeline (0)

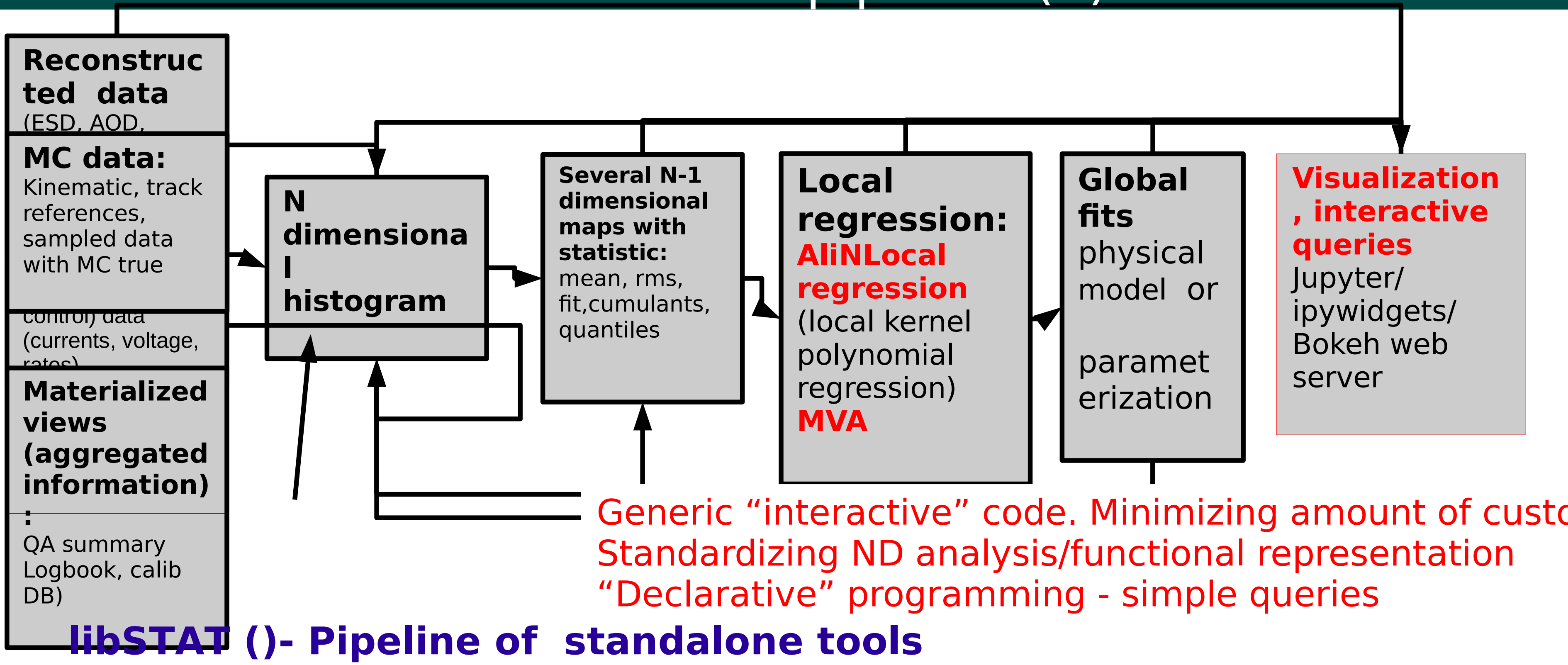


$$f(p_0, p_1, p_2, \ldots) \neq f_0(p_0) \oplus f_1(p_1) \oplus f_2(p_2) \oplus \ldots.$$

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

- dimensionality depends on the problem to study (and on available resources)
- Data →Histogram → set of ND maps → set of NDlocal regression/TMVA → Global fits
  - Some steps can be skipped, e.g local regression (MVA/AliNDLocal) can be done using unbinned input data
  - Histogramming in case of non sparse data
  - MVA for sparse (going to higher dimansions)

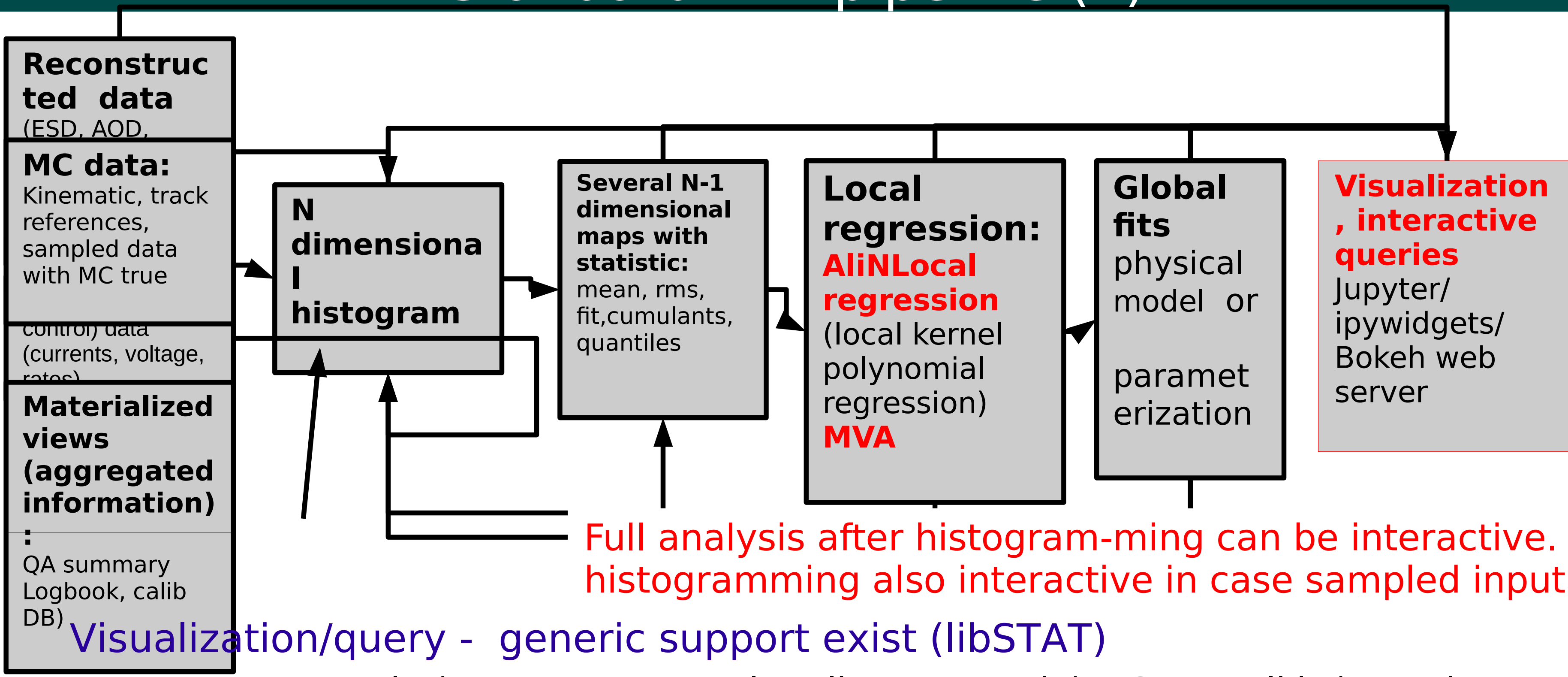
Standard ND pipeline (1)



libSTAT ()- Pipeline of standalone tools

- N dimensional histogramming
  - THn/THnSparse - part of root
  - AliTreePlayer::MakeHistogram ....
- THn → Map (tree)
  - part of the TStatToolkit
- Map(Tree) → Local regression
  - AliNDLocalRegression, TMVA+TMVA::Reader interface
- Map(Tree) → Global fits (physical models, parameterizations)
  - AliTMinuitToolkit

Standard ND pipeline (2)



Full analysis after histogram-ming can be interactive. N d
histogramming also interactive in case sampled input da

Visualization/query - generic support exist (libSTAT)

- TFormula (TF1, TTreeFormula, AliTreeFormula), TStatToolkit (Metadata describing input data)
- AliTreePlayer
- AliPainter and AliDrawStyle(CSS)

Declarative analysis
Integration with RDataFrame (JIT, paralelism) under development

Extensions:

- CSS library
- Multidimension drawing support (Histogram, TFormula, TTree)
  - interface inspired by some python libraries
- Web server (static and Jupyter/ipywidgets/Bokeh based)

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

When the dimensionality increases, the volume of the space increases so fast that **the available data become sparse**.

- The volume of a cube grows exponentially with increasing dimension
- The volume of a sphere grows exponentially with increasing dimension
- Most of the volume of a cube is very close to the (d − 1)-dimensional surface of the cube

Effect to be considered. Detector/reconstruction experts to be consulted

- Find relevant dimensions (2-6 dimensions)
- Proper selection of variables (smooth or linear behavior)
  - e.g q/pt instead of pt, occupancy/multiplicity instead of centrality
- Proper binning. In case proper selection of variables, few bins needed

In the following I'm considering properly designed dimensionality/binning of the space

MVA in case of the sparse data (too high dimensions, time series)

Curse of dimensionality (Example performance map)

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

When the dimensionality increases, the volume of the space increases so fast that the available data become sparse. This sparsity is problematic for any method that requires statistical significance.

Code fragment $AliPhysics_SRC/PWGPP/TPC/macros/performanceFiltered.C

```
//
hisString+=TString::Format("deltaP%d:qPt:tgl:logTracks5:#IsPrim4&&TPCOn&&ITSRefit&&ITSOn01&&nclCut>>hisDeltaP%d_Allv_qPt_tgl_logTracks5(400,%f,%f,200,-5,5,10,-1,1,10,0,10);",iPar,iPar,-rang
hisString+=TString::Format("deltaP%d:qPt:tgl:logTracks5:#IsPrim4&&TPCOn&&ITSRefit&&ITSOn01&&nclCut&&TRDOn>>hisDeltaP%d_TRDv_qPt_tgl_logTracks5(400,%f,%f,200,-5,5,10,-1,1,10,0,10);",iPar,iPa
hisString+=TString::Format("pullP%d:qPt:tgl:logTracks5:#IsPrim4&&TPCOn&&ITSRefit&&ITSOn01&&nclCut>>hisPullP%d_Allv_qPt_tgl_logTracks5(400,-8,8,200,-5,5,10,-1,1,10,0,10);",iPar,iPar);
hisString+=TString::Format("pullP%d:qPt:tgl:logTracks5:#IsPrim4&&TPCOn&&ITSRefit&&ITSOn01&&nclCut&&TRDOn>>hisPullP%d_TRDv_qPt_tgl_logTracks5(400,-8,8,200,-5,5,10,-1,1,10,0,10);",iPar,iPar);
 hisArray = AliTreePlayer::MakeHistograms(chain, hisString, defaultCut,0,maxEntries,200000,15);
```

For the tracking performance studies - histogramming is better option (at first stage)

Resolution/pulls as function of (q/pt,Q,mult) - O (20000) bins

Performance generator (jets flat in q/pt)

- 100 jobs x 50 events x 100 tracks (few hours at GSI farm)
    - Not sparse - O(25 tracks) per bin
    - more in case of bin grouping (parameter in map creation)

Interactive analysis using filtered trees (sampled input flat in pt)

Usage of n-dimensional pipeline

Pipeline with performance maps in N dimensions in form of generic function (TFormula).

- In many cases corresponding physical model or parameterization available

Usage:

- differential QA
- understand/remember detector behavior - physical models
- scan tuning of the reco. parameters (metric diff of perf. maps)
- scan tuning of the MC paraemeters (metric diff of perf. maps)
- compare differentialy data with MC
- provide recipes for optimal cut selections
- provide input/parameterizations for toy MC/fast MC
- feasibility studies
- enable tune on data in N-dimensions - remapping MC → Data
- **enable ML algorithms (tune on data)**

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

Recent usage of ND pipeline

- alarms example
- global/local fit example
- derived variable example
- consistency checks

Recent examples

- Space charge distortion calibration and studies
  - examples slides 7, 8
  - Distortion correlation studies
- Tracking performance characterization (TRD in combined tracking)
  - https://indico.cern.ch/event/710009/#53-trd-in-tracking-in-run2
- TPC/TRD tracking optimization
  - under development
- New event and track selection cuts (dEdx and chi2 QA analysis)
  - https://indico.cern.ch/event/710009/#62-proposal-for-new-event-and

TPC dEdx calibration macro.

Macro

```
void LoadChain(const char *filteredList="filtered.list",Int_t nChunks=1);
void MakeHistograms(Int_t nEvents=100000);
void MakeMaps();
void makeNDLocalFit(TString varName, TString customCut, TString errorVar);
```

Define variables and derived variable (e.g using local regression parameterization)

```
treeA->SetAlias("ratioTotMax0", "fTPCdEdxInfo.GetSignalTot(0)/fTPCdEdxInfo.GetSignalMax(0)");
treeA->SetAlias("ratioTotMax1", "fTPCdEdxInfo.GetSignalTot(1)/fTPCdEdxInfo.GetSignalMax(1)");
treeA->SetAlias("pullTotMax0","(ratioTotMax0-hisRatioTotMax0Dist.meanGFit)/hisRatioTotMax0Dist.rmsGFit");
treeA->SetAlias("pullTotMax1","(ratioTotMax1-hisRatioTotMax1Dist.meanGFit)/hisRatioTotMax1Dist.rmsGFit");
```

MakeHistograms

```
histoS+= "ratioTotMax0:atgl:multA:pileUpZ:sdEdxMax:#isSelected>>hisRatioTotMax0(100,0.5,2.0,20,0,1,20,0,20000,5,-
50,50,20,6,32);";
histS+= "ratioTotMax1:atgl:multA:pileUpZ:sdEdxMax:#isSelected>>hisRatioTotMax1(100,0.5,2.0,20,0,1,20,0,20000,5,-
50,50,20,6,32);";
AliTreePlayer::MakeHistograms(treeQ, histoS,
"nclCut&&chi2Cut&&downscale&&abs(qPt)<1.25&&nclCutGold&&pileUpCut", 0, nEvents, -1, 15);
```

makeNDLocalFit (local regression + register fit functor for later usage)

```
makeNDLocalFit("hisRatioTotMax1Dist.meanG","","sqrt(0.01+hisRatioTotMax1Dist.rmsG**2/hisRatioTotMax1Dist.entries)");
makeNDLocalFit("hisRatioTotMax2Dist.meanG","","sqrt(0.01+hisRatioTotMax2Dist.rmsG**2/hisRatioTotMax2Dist.entries)");
```

makeGlobalFit (dEdx resolution example)

```
AliTMinuitToolkit *fitterResolMaxTot = new AliTMinuitToolkit("AliTMinuitToolkitFitterResolMaxTot.root");
formulaResolMaxTot = new TFormula("formulaResolMaxTot", "sqrt( x[0]**2 + ([0])*(x[1]^([1]))*(x[2]^([2])) )");
fitterResolMaxTot->FillFitter(treeMap,"rmsMaxTot0" ,"rmsTot0NormFit0:(multAMean/10000.):(1/
((sdEdxMaxCenter**2)*sqrt(1+atglCenter**2)))" , "fitCut1&&outlierCut", 0, 10000000);
fitterResolMaxTot->Bootstrap(20, "report1");
treeMap->SetAlias("rmsMaxTot0.Fit", fitterResolMaxTot->GetFitFunctionAsAlias().Data());
```

Input - sampled data O(10^6-10^8) tracks (objects)- flat in q/pt resp ITS dEdx

22 - 5 dimensional histograms (TPC dEdx ratio region , tan(T), multiplicity, dEdx, pileup Z)

Pipeline example usage: dEdx resolution parametric fit.

```
makeNDLocalFit("hisRatioMax01Dist.rmsG","isOK","sqrt(0.01**2+hisRatioMax01Dist.rmsG**2/hisRatioMax01Dist.entries)");
makeNDLocalFit("hisRatioMax12Dist.rmsG","isOK","sqrt(0.01**2+hisRatioMax12Dist.rmsG**2/hisRatioMax12Dist.entries)");
makeNDLocalFit("hisRatioMax02Dist.rmsG","isOK","sqrt(0.01**2+hisRatioMax02Dist.rmsG**2/hisRatioMax02Dist.entries)");
treeMap->SetAlias("rmsMax0","sqrt((hisRatioMax01Dist.rmsGFit**2+hisRatioMax02Dist.rmsGFit**2-
hisRatioMax12Dist.rmsGFit**2)/2.)");
TFormula *formulaResol0 = new TFormula("formulaResol", "[0]*(x[0]^abs([1]))*(x[1]^abs([2]))");
fitterResol0->SetFitFunction(formulaResol0, kTRUE;
fitterResol0->FillFitter(treeMap, "rmsMax0", "(1./sdEdxMaxMean**2):(1./sqrt(1+atglMean**2))",
                                "fitCut0&&outlierCut", 0, 10000000);
fitterResol0->Bootstrap(20, "report0");
treeMap->SetAlias("rmsMax0Fit0", fitterResol0->GetFitFunctionAsAlias().Data());
```

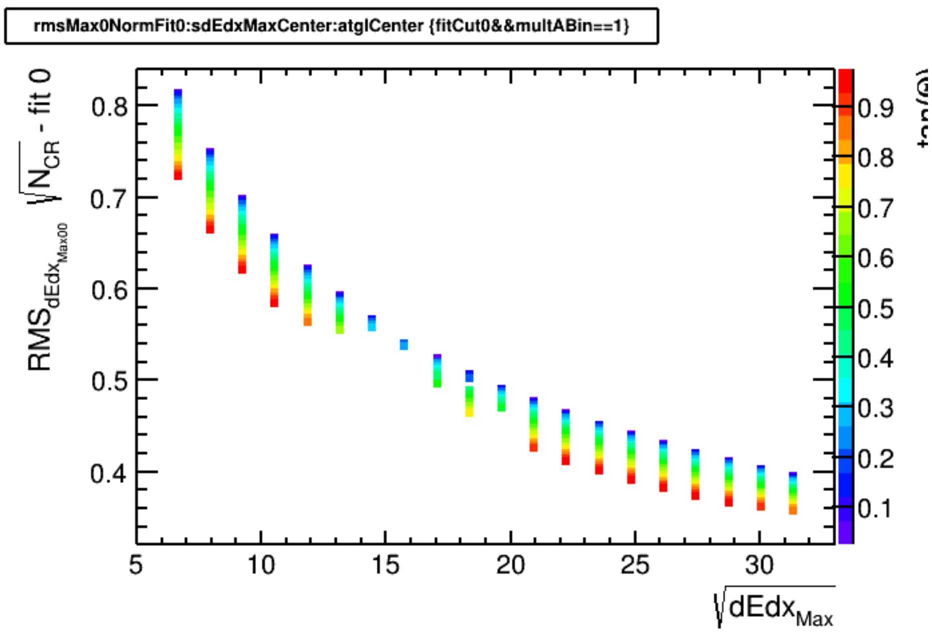
$$RMS_{Qi} = \sqrt{RMS^2_{Qi/Qj} + RMS^2_{Qi/Qk} - RMS^2_{Qj/Qk}/2}$$
$$RMS_{ROC} \times \sqrt{N_{CR}} \approx p_0 \left(dEdx^{p_1} \times \sqrt{(1+\tan(\theta))^2}^{p_2} \right) \quad (1)$$

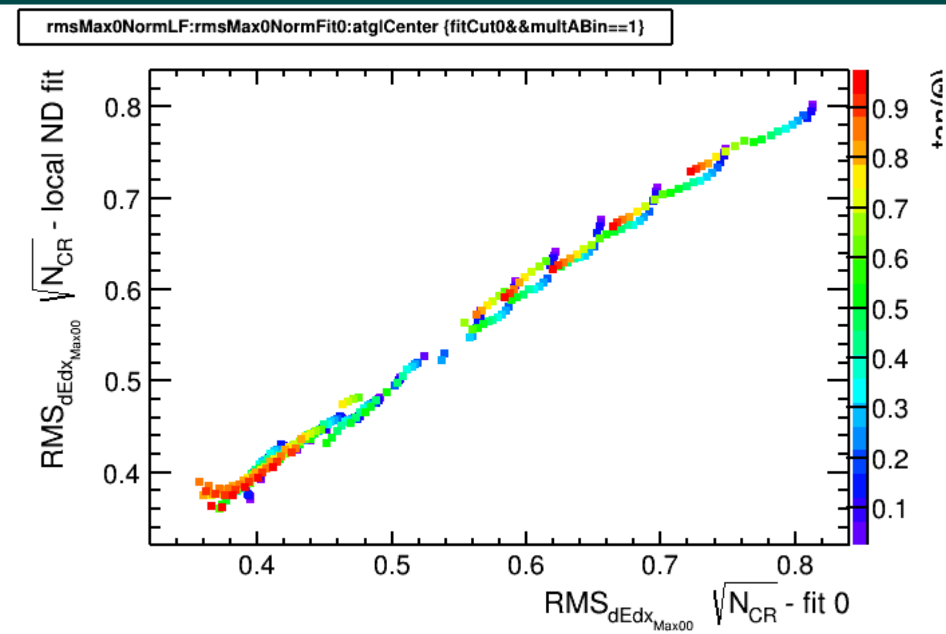
Derived information: Relative intrinsic resolution estimated from the RMS of the dEdx ratios in 3 regions of ALICE TPC detector

Model: At low multiplicity relative dEdx resolution - power law scaling with the dEdx and tracklet length

- Parameters p_0, p_1 and p_2 fitted in global fits
  - fit using AliTMinuitToolkit + bootstrap method
  - Huber loss used minimization (square, pseudo-huber in default, **user defined loss function possible**)
  - error estimator as obtained in bootstrap **(ransac and standard fit)**

Pipeline example usage: dEdx Q_{max} IROC0 at Low





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

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

Relative Intrinsic resolution for IROC as parameterized by global resp. local fit

At low IR agreement between dEdx resolution and power low model

- Parameters close to analytical expectation:
  - p_0 =1.96 +-0.01, p_1=0.233+-0.001, p_2=0.40+-0.02
- * axis automatically annotated using data source annotations (UserInfo of TTree)
  - ```
    TStatToolkit::AddMetadata(treeMap, "rmsMax0TotFit0.AxisTitle",
    "RMS_{dEdx_{Tot0}}  #sqrt{N_{CR}} local ND fit");
    ```
    ```

Consistency checks example:

```
void CheckInvariant(){

  treeMap->Draw("(exp(hisRatioMax01Dist.meanGFit)/exp(hisRatioTot01Dist.meanGFit))/(hisRatioTotMax1Dist.meanGFit/
hisRatioTotMax0Dist.meanGFit)","isOK&&entries>200");
  treeMap->Draw("(exp(hisRatioMax02Dist.meanGFit)/exp(hisRatioTot02Dist.meanGFit))/(hisRatioTotMax2Dist.meanGFit/
hisRatioTotMax0Dist.meanGFit)","isOK");
  treeMap->Draw("(exp(hisRatioMax12Dist.meanGFit)/exp(hisRatioTot12Dist.meanGFit))/(hisRatioTotMax2Dist.meanGFit/
hisRatioTotMax1Dist.meanGFit)","isOK");
}
```

$$Q_{\mathrm{det_0,type_0}}/Q_{\mathrm{det_1,type_1}} \neq const$$

Non invariant

$$Q_{\mathrm{det_0,type_0}}/Q_{\mathrm{det_1,type_1}} \approx \frac{dEdx \times T_{\mathrm{det_0,type_0}}}{dEdx \times T_{\mathrm{det_1,type_1}}} \quad (1)$$

$$T_{\mathrm{det_i,type_j}} \approx f(dEdx, \tan(\theta), \tan(), Gain, occu, thr.)$$

$$\frac{\frac{<Q_{max0/max1}>}{<Q_{to0/tot1}>}}{\frac{<Q_{tot0/max0}>}{<Q_{tot1/max1}>}} = const$$

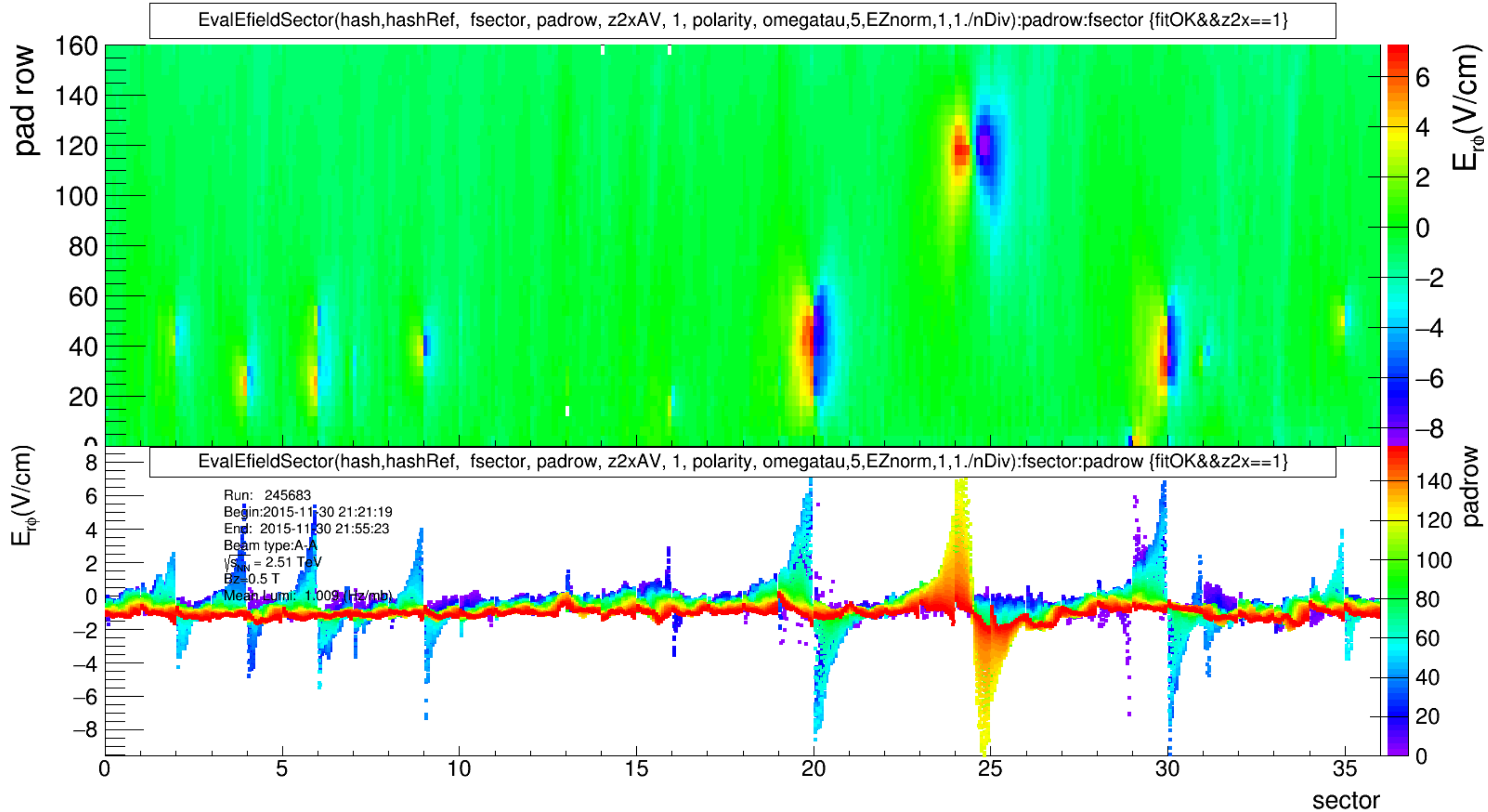
Performance map invariant

Set of 3-4D transfer function/maps created

Function to be used in user analysis (track and event cuts, tune on data)

Internal consistency of maps can be checked using invariants

E field distortion static 3D view → 1D



Run2:
Static view/projections without nominal map subtraction insensitive to the distortions
Run 3:
Width of the 1D projection distribution 20 cm. 1000 wider than requested resolution
Automatic alarms?

Alarm example: E field/distortion/current maps

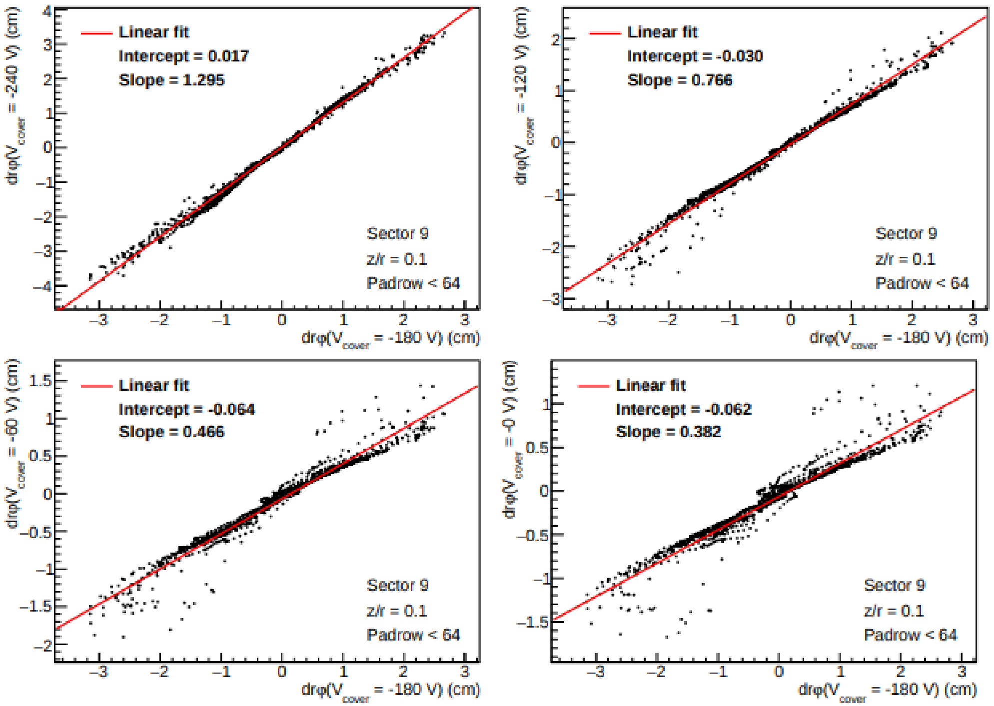


Figure 42: The four plots show the correlation between the distortions $dr\varphi$ with the nominal cover electrode voltage of $V_{\mathrm{cover}} = -180\,\mathrm{V}$ and the distortions with cover voltages of $-240\,\mathrm{V}$ (top left), $-120\,\mathrm{V}$ (top right), $-60\,\mathrm{V}$ (bottom left) and $0\,\mathrm{V}$ (bottom right) at sector 9. Each graph is fitted with a polynomial of first order. The fit parameters are quoted in the legend. The slope directly indicates the size of the distortions with the given cover electrode voltage relative to the nominal settings.

Alarms to be based on invariants - e.g. difference between maps and scaled reference maps

- Example: distortion maps correlation for different voltage setting
- Alarm criteria adjusted using models (sigmas) resp. history of past data

New use cases

Run3 - Distortion/current maps and alarms

Set of 3D distortion maps to be taken with O(s) granularity

Set of 3D digital (O(500000) pixels and analog currents maps to be taken with O(ms) time granularity

To be QA-ed with appropriate time granularity

- in case of problems later will be not recoverable
- decision/correction to be made online
- outliers will be always present
  - long tails - asymmetric distribution of number of pile-up tracks, ion cloud

Huge spread of distortion O(20 cm) - position and time (gain, epsilon, local density) dependent

Alarms to be based on the set of “invariants”

- Local smoothness in space and/or time
  - Smoothness assumption in time
    - steps in the IR are normal/ some double ratios to be used
    - left/right median filter
- Variable/local median
  - alarms on the mean/median/rms/cumulant of distributions

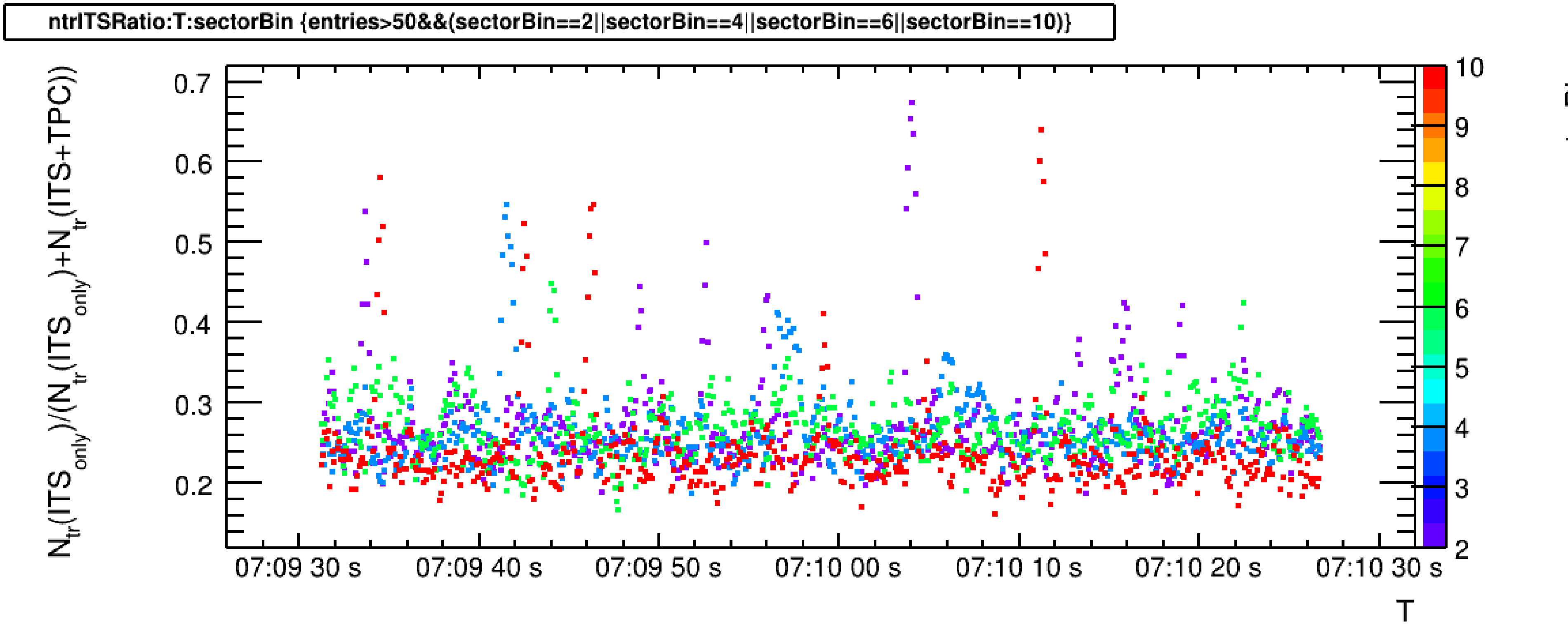
RUN3:

more complex than Run2 - good diagnostic needed
many decisions to be made on-line. In case lossy compression us
not afford mistakes

PbPb 2015 - time bin QA time series

Run2 example of ongoing investigation

Correlated TPC/ITS efficiency loss: **Normal run 246148**



Time series QA O(0.1s)

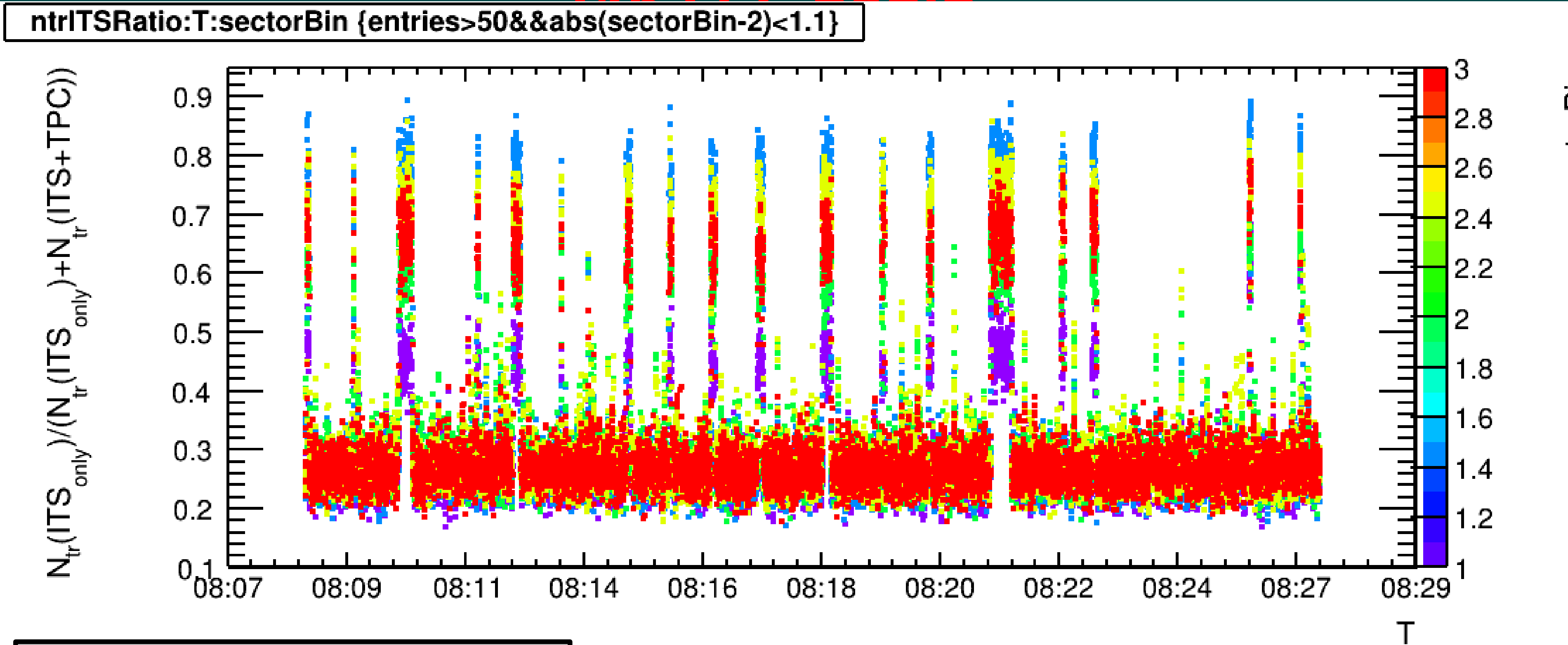
In particular time intervals **O(0.5) seconds** distortion increased

- locally worse resolution and matching efficiency

Distortion independent- see time position of spikes

- **sector bins 2, 4, 6, 10**

Correlated TPC/ITS efficiency loss. Problematic run 246272 begin



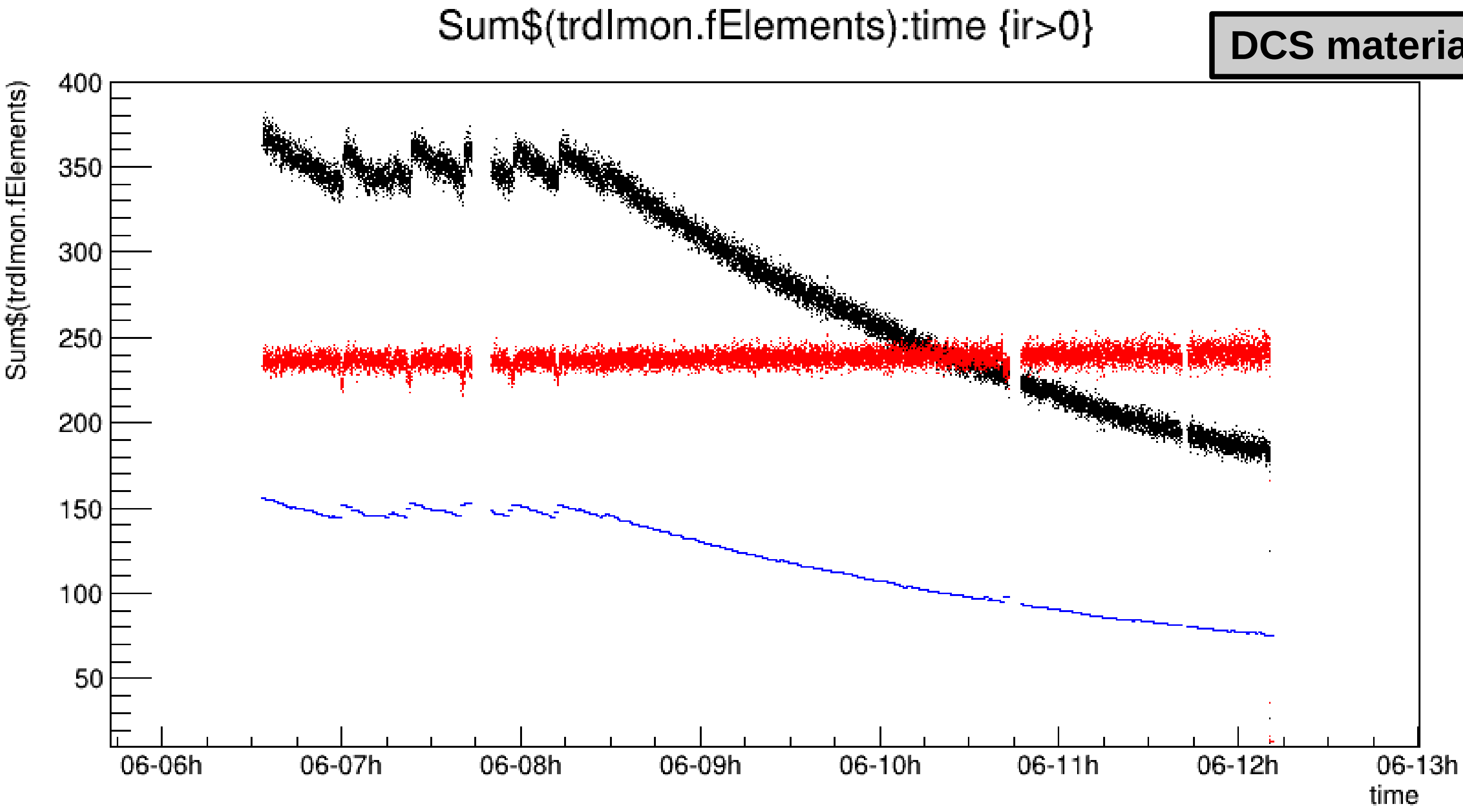
Time series QA O(0.1s)

Regular structure observed at sector boundary 2 in the run (246272 and also some others)

Outliers in matching efficiency related to time intervals

- Looks like regular position O(min) spacing
- Irregular amplitude, probability and duration
  - begin of run (higher IR) - bigger probability longer duration
- Possible origin - “onset of distortion”
  - 2 line charges holes in sector boundary 2 - one big one small

Rate and current monitoring



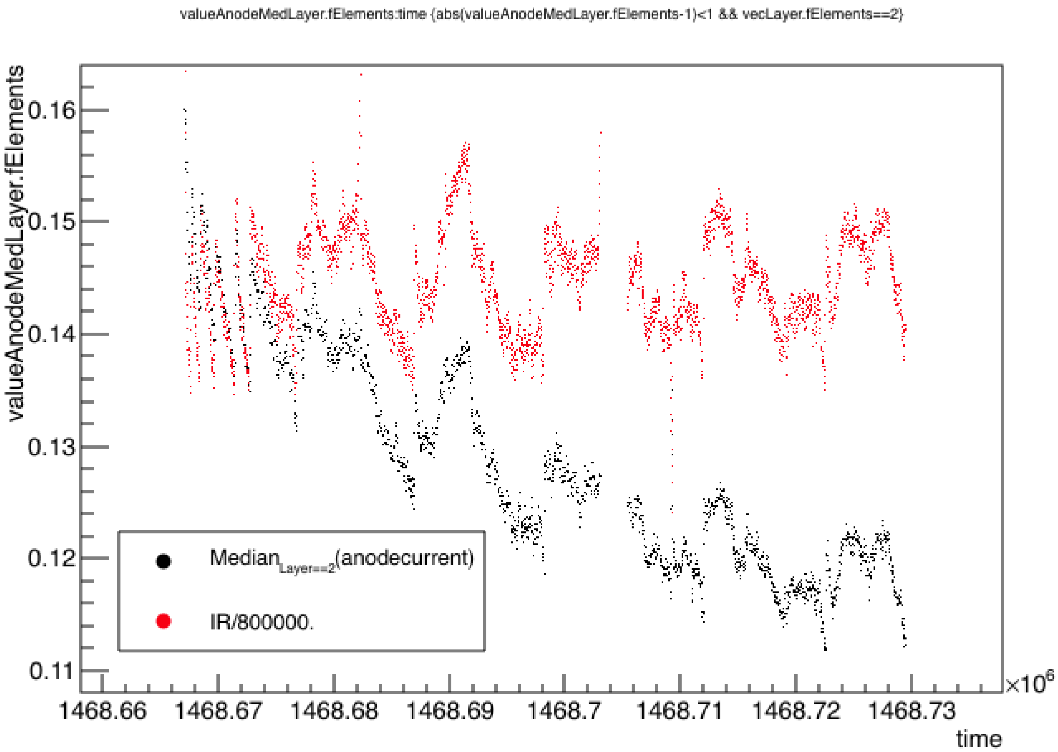
DCS materialized view

No indication of flux modulation with the granularity O(min)

Flux estimators (a.u) in fill with run 246272 - limited time granularity >> 1 s

- **Sum of TRD currents, interaction rate, ratio Current/IR**
- IR levelling in fill of interest ~ 15 minutes
- TRD currents and IR estimator from trigger detector agree (no indication of background)

Run2. TRD currents. What can go wrong? (pp -2016)



spikes in hardware currents and IR
not all time intervals usable
Model validity +-10 %

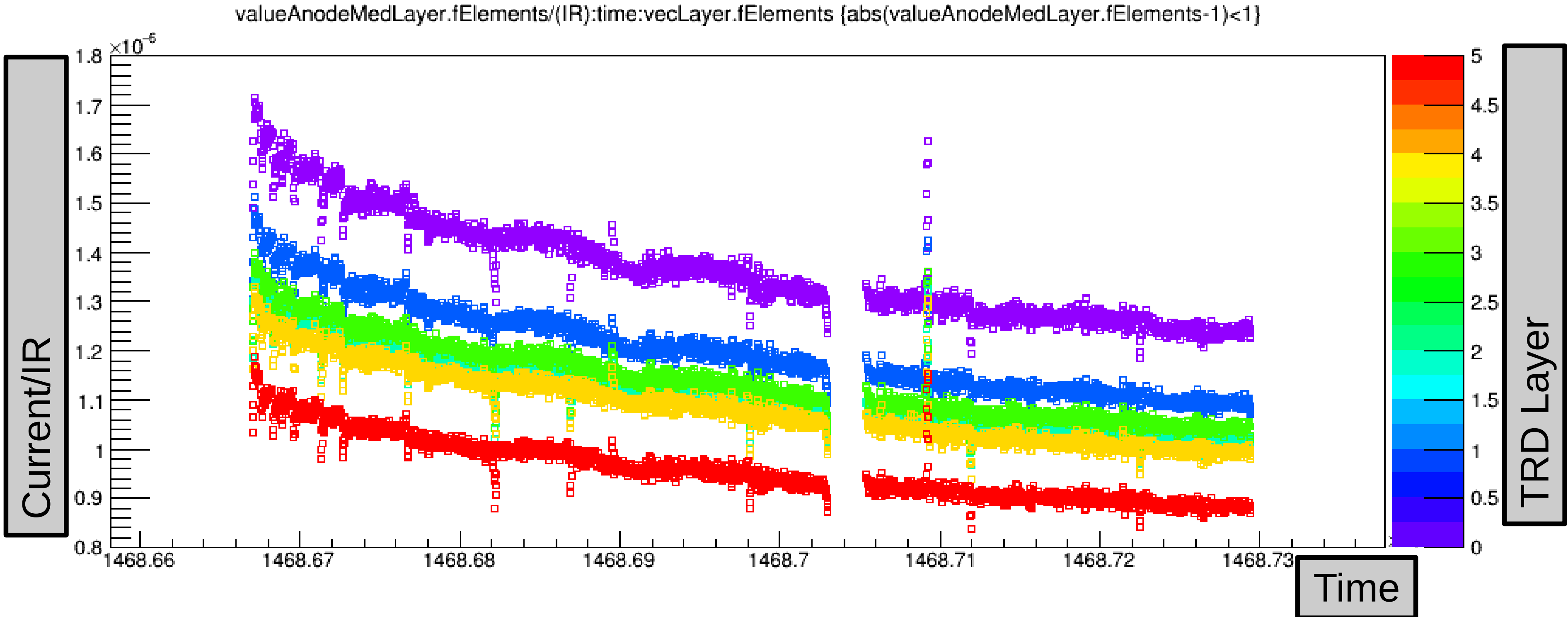
TPC distortion corrected using the IR resp. current estimator

- applicability of distortion scaling models limited to narrow ranges of possible values
- High time gradient → biased estimators

Better estimator to be obtained using TRD currents

- TRD currents ~ signal + beam background
- Weighting of currents → distortion scaling. Appropriate extrapolation function to TPC to be selected and controlled

TRD current to monitor TPC space charge ((pp - 2016))

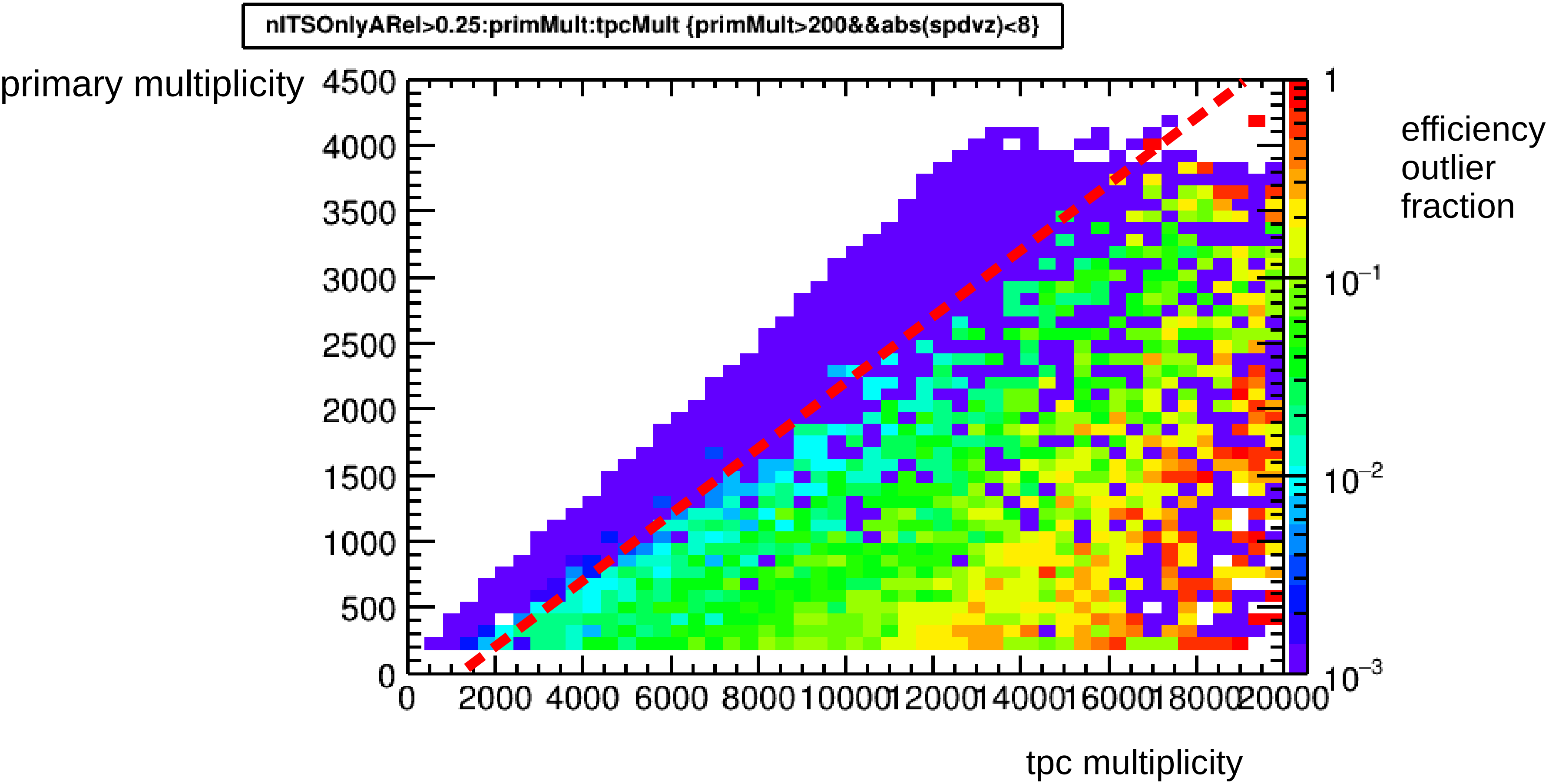


TPC distortion corrected using the IR estimator

Better estimator to be obtained using TRD currents

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

Analysis consideration: TPC → ITS matching outliers



- In current analysis of the PbPb LHC15o strong pile-up rejection cut used
- Suggestion - only very small part of pile-up events are outliers in matching efficiency
- Better tagging of outliers using mean event properties

ML for the time bin based QA

Classification problem:

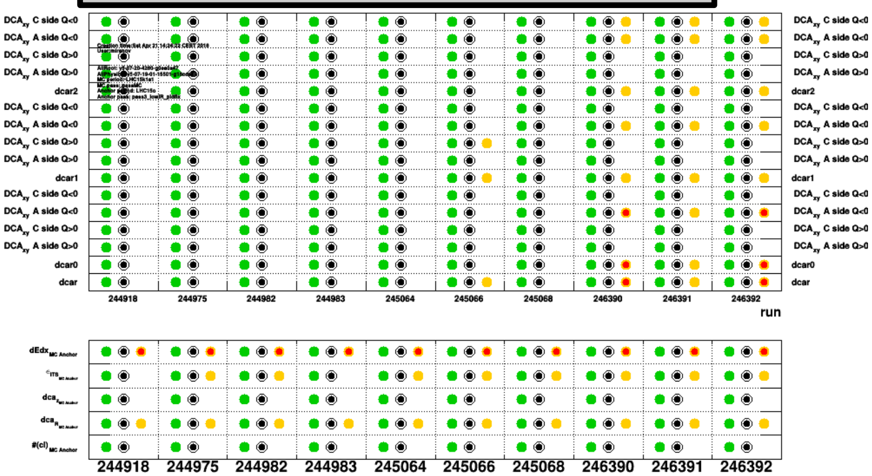
- Anomalies/Outliers in the performance QA
- Find most relevant features in the other observables
  - Maps: currents, distortion, local multiplicity, matching efficiency, chi2, Ncl, resolution ….
  - Derived “invariant” variables:
    - e.g RMS of current map/phi averaged map/scaled maps
    - distortion map/phi averaged map
    - local discontinuities in time and space
    - “physics” acceptable performance

Explain/find hardware origin of anomalies (currently hierarchy of alarms)

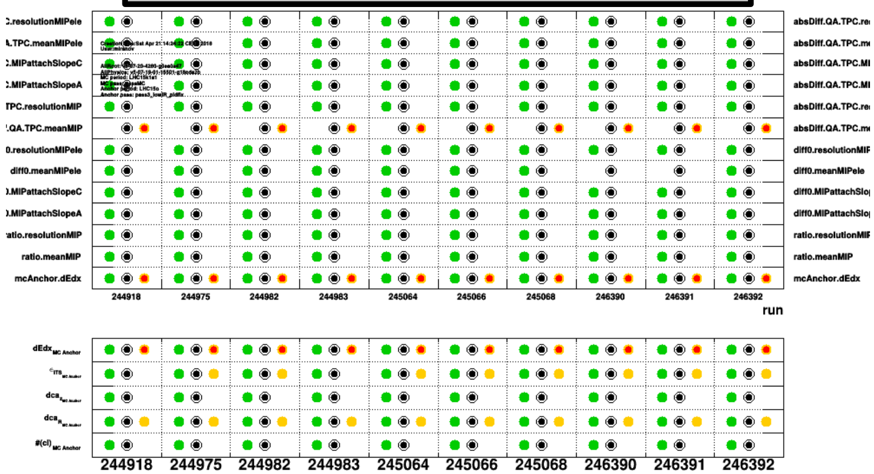
Training data:

- current maps, and distortion maps (at GSI) can be exported to alien
- time bin based QA currently available only for few run

DCAr alarms decomposition



dEdx alarms decomposition



Example: Jet analysis

Tracking performance not homogeneous in space

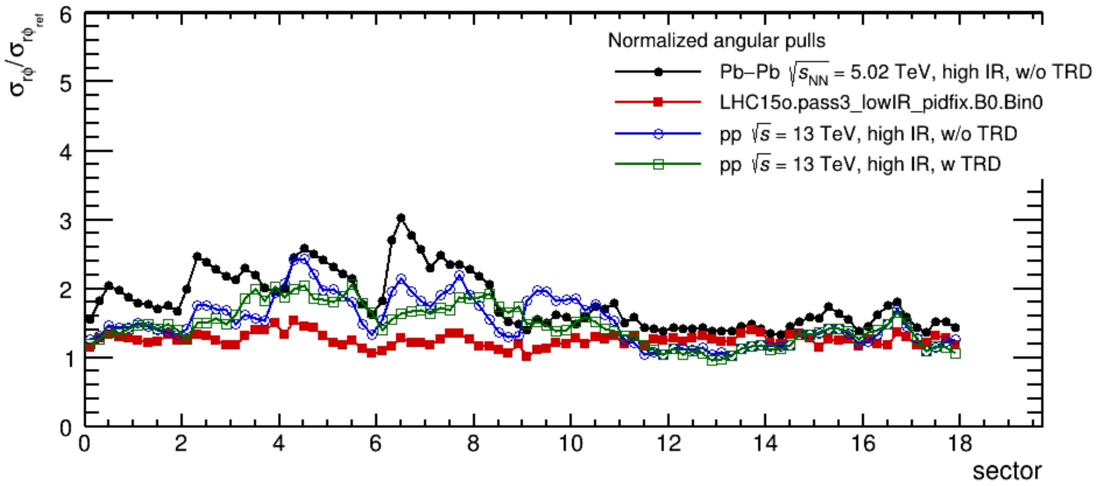
- discussed with Jet group (Leticia and Markus)

MC - using add hoc parameterization of the distortion fluctuation

- only partial description of performance deterioration
- physical model of fluctuation - no time to implement

Proposed solution:

- unfolding spectra using the 4-5D (phi eta performance maps)
- folding MC to e



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

Example: Fast MC

Fast MC developed for RUN3 (Ruben, Johannes ...) used for ...

- fast particle transport - trajectories, interaction, energy loss MS
- fast reconstruction of the data

Detector response **parameterized using N dimensional maps/parametrization** (Johannes, Marian):

- Distortion maps
- Dead zone
- Cluster smearing
  - TPC-ITS matching efficiency (https://alice.its.cern.ch/jira/browse/AOC-13)

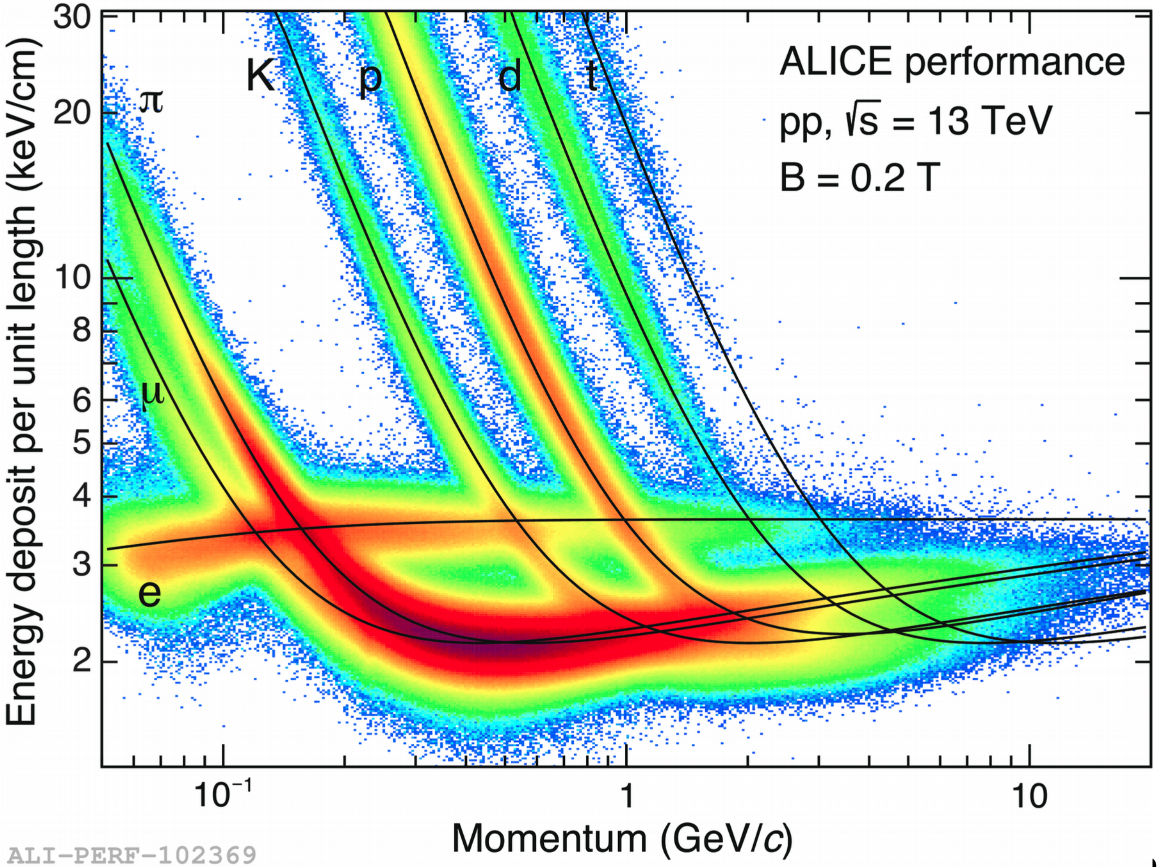
Using proper detector response parameterization more realistic performance than using current full MC expected

- Activities interrupted after Johannes left collaboration
- Reactivation (?) as one of the transport model

Several **other options** investigated in context of RUN3 activities. I consider N-Dimensional parameterization option as **more promising (either NDlocal regression or ML regression)**

- more exotic options (GANs) to be benchmarked in respect to “more” classical

Example outlier removal: Alice TPC dEdx performance (PDG)



Non standard tracking and

Tracking down to 30-40 MeV
TPC standalone without outer d
Pile-up rejection using dEdx inv
First usage of transfer function
Additional correction for th
Additional dEdx correction for t
and electron attachment
in particular run gain was 2 tim
smaller than nominal

https://aliceinfo.cern.ch/Figure/node/8731

C. Patrignani et al. (Particle Data Group), Chin. Phys. C, 40, 100001 (2016) and 2017 update.

- TPC standalone tracking and dEdx performance down to 30-50 MeV (ATO-245)
- Using **TPC dEdx invariant cut ("PID likelihood")** and 3D transfer function correction, TPC standalone analysis enabled

Usage/TO DO/Conclusion

Usage of n-dimensional pipeline

Pipeline with performance maps in N dimensions in form of generic function (TFormula).

- In many cases corresponding physical model or parameterization available

Usage:

- differential QA
- understand/remember detector behavior - physical models
- scan tuning of the reco. parameters (metric diff of perf. maps)
- scan tuning of the MC paraemeters (metric diff of perf. maps)
- compare differentialy data with MC
- provide recipes for optimal cut selections
- provide input/parameterizations for toy MC/fast MC
- feasibility studies
- enable tune on data in N-dimensions - remapping MC → Data
- **enable ML algorithms (tune on data)**

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

Conclussion

N-dimensional analysis pipeline developed within last years

Used in many recent successful projects (distortion mitigation, TRD commisioning)

- without ND approach project witll not converge

Active development:

- MVA interface wrappers - working prototype exist and is used (see next presentation)
    - errors
- Interactive N-Dimensional visualization of data (python)
    - Jupyter/ipywidget/Bokeh - see demo
    - to be merged with previous styling development development in (C++)

Demo

Browsing of extracted N-Dimensional performance maps

-

Distortion time series analysis studies

- MVA to describe distortion (+errors and quantiles) as function of …
- interactive widgets to browse