Classifiers for centrality determination in proton-nucleus and nucleus-nucleus collisions

Igor Altsybeev, Vladimir Kovalenko
Saint-Petersburg State University

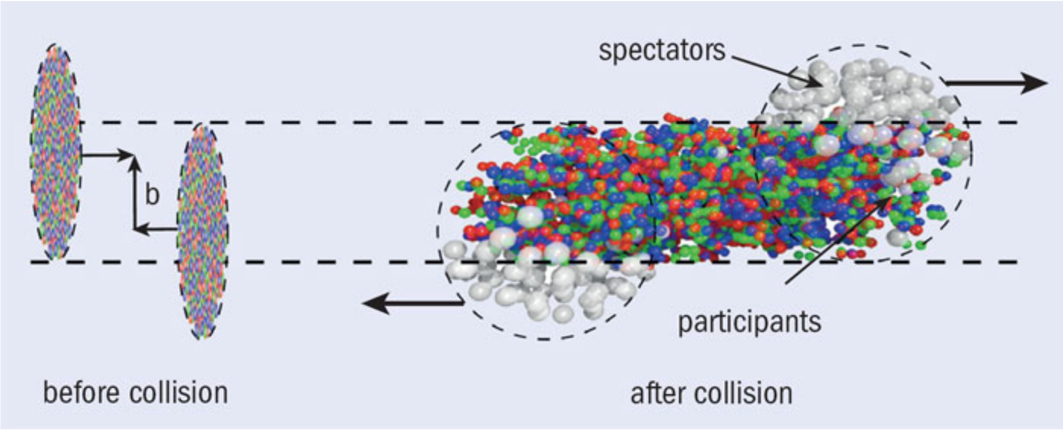
Heavy Flavour Data Mining workshop
Zurich, February 19, 2016

Prologue

- ML usage in HEP so far:
  search for rare decays, detector response optimization
- Interesting to find new applications
  - try to address some physics of heavy ion events

The collision centrality

The centrality is a key parameter in the study of the properties of QCD matter at extreme temperature and energy density, because it is directly related to the initial overlap region of the colliding nuclei.



The *impact parameter (b)* is the distance between the centers of the two colliding nuclei.

The centrality determination in experiment

Receipt:

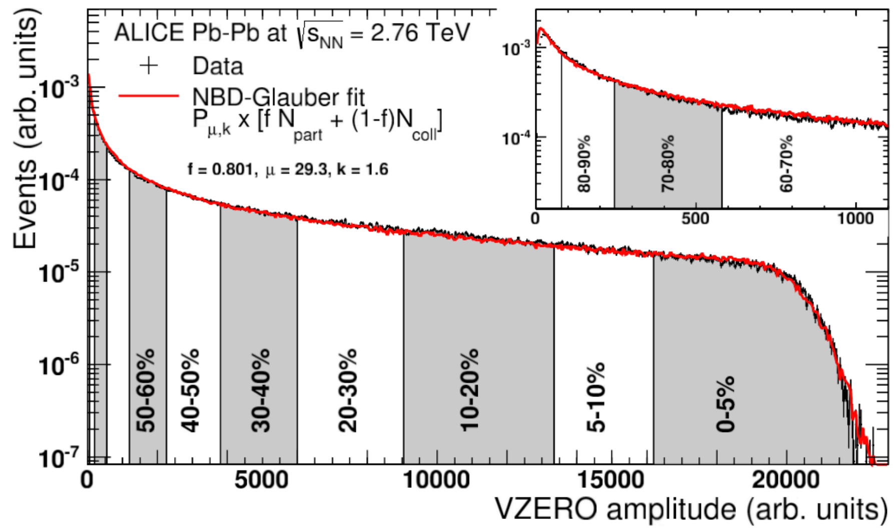
- Use signal distribution in some detector
- Fit with some geometry-based model
- Split into *centrality classes*.

In ALICE experiment:

main method uses *multiplicity distribution* in (semi-central) VZERO detector fitted with the Glauber model.

Close to 0% → most central events.

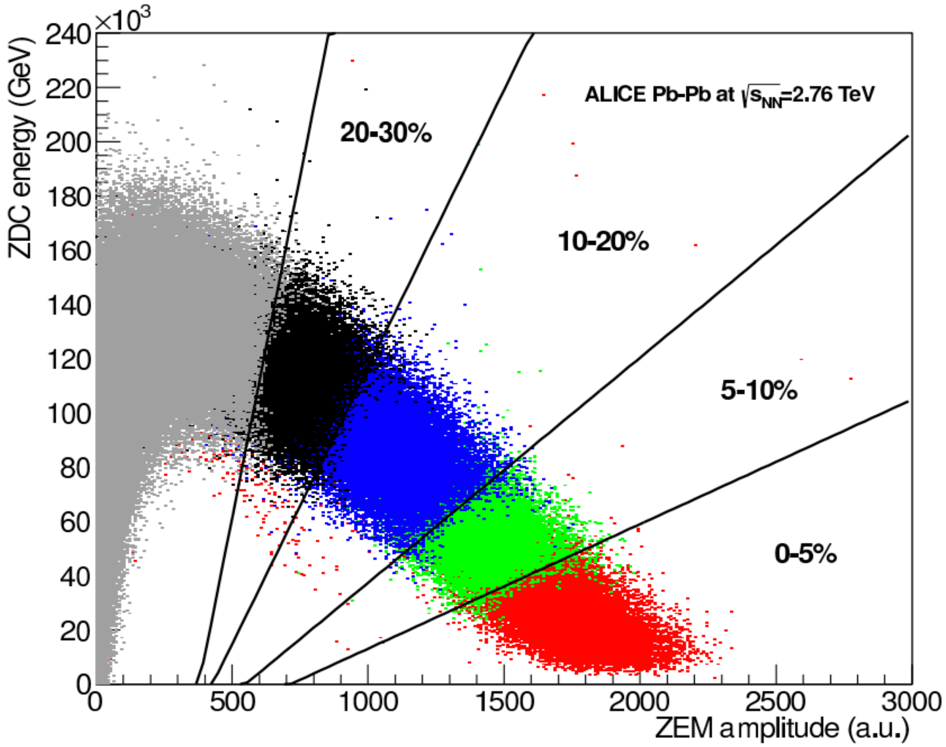
arXiv:1301.4361



→ b_{impact}, N_{part}, N_{coll}, N_{spec} are deduced from the Glauber model.

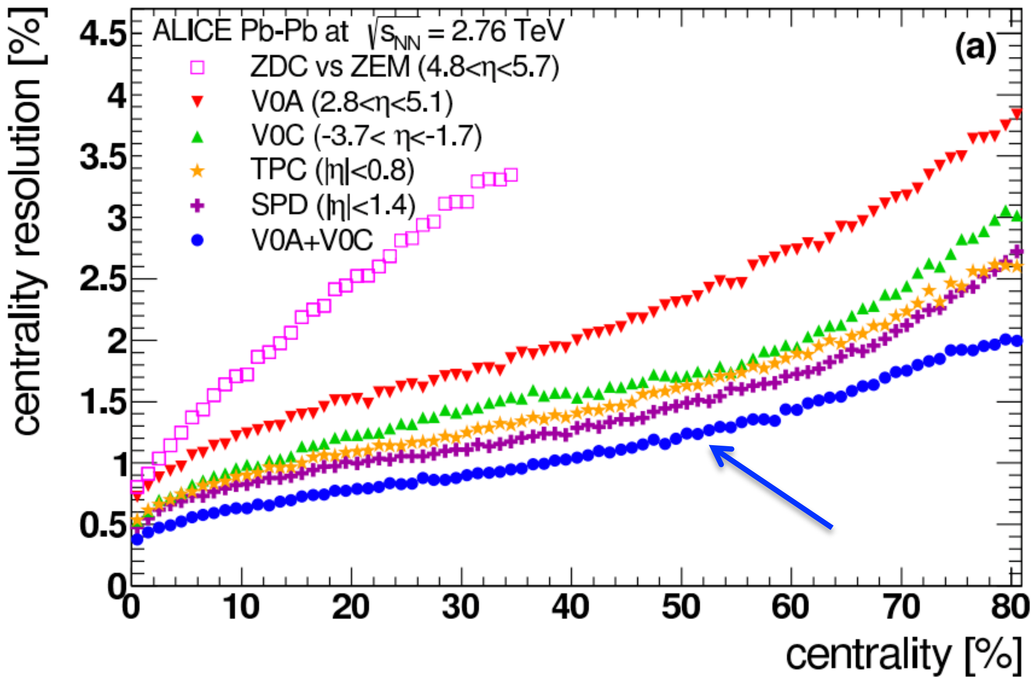
The centrality determination in experiment

Second method of ALICE uses energy from neutrons-spectators in Zero-degree calorimeters (ZDC) coupled with signals in electromagnetic calorimeters (ZEM).



The centrality determination in experiment

Best centrality resolution is achieved by usage of the VZERO estimator:



→Can we perform better using multiple detectors simultaneously?
Try machine-learning techniques for that.

Centrality determination ML task

- Use AMPT monte-carlo generator to simulate Pb-Pb events at 2.76 TeV (without detector response) (400k events)
- 5+3 features are selected in correspondence with the subsystems of the ALICE detector:

ZDC-C

N neutrons

VZERO-C

multiplicity

$-3.7 < \eta < -1.7$

TPC

multiplicity

$|\eta| < 0.8$

Additional features:

- average p_T
- kaon/pion ratio
- proton/pion ratio

VZERO-A

multiplicity

$2.8 < \eta < 5.1$

ZDC-A

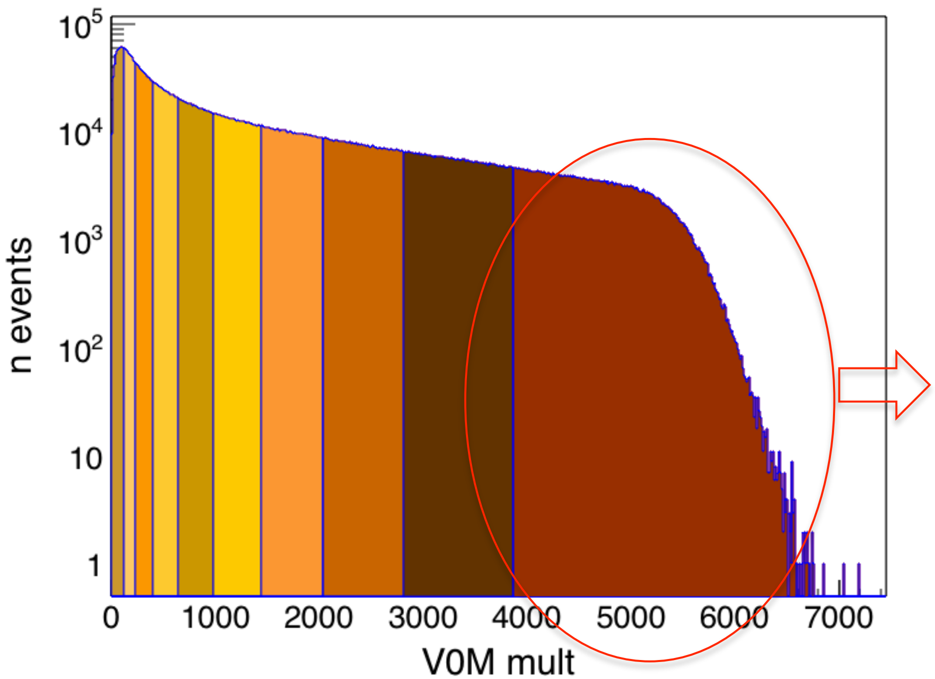
N neutrons

Using TMVA Version 4.2.0

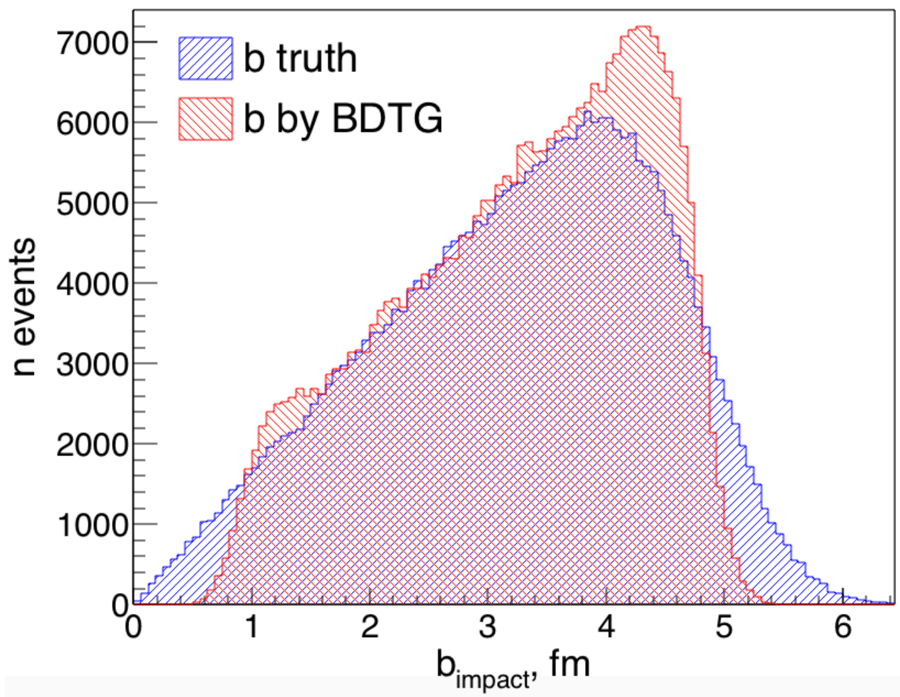
Target for regression: impact parameter *b*

Pre-selection:

events with 0-10% V0M centrality



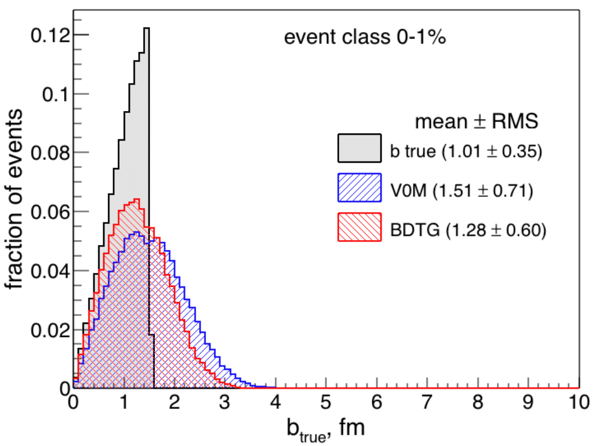
True *VS* estimated by BDTG:

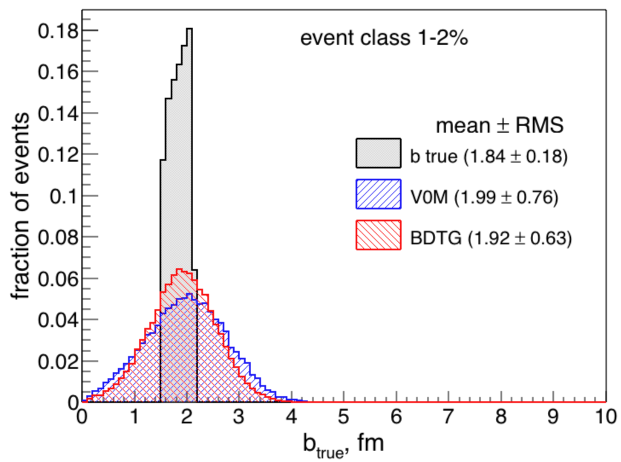


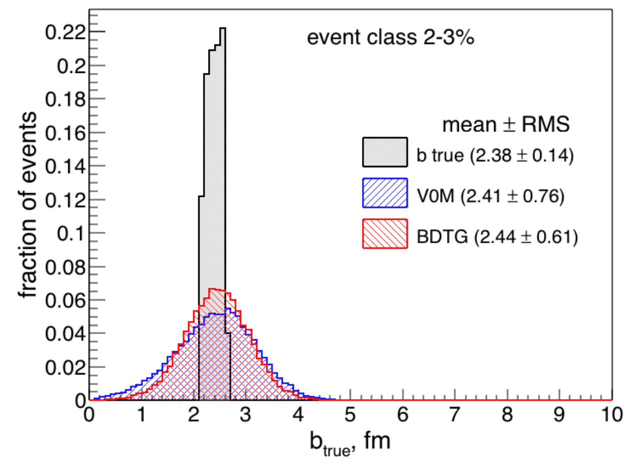
- Estimator: unavoidable "deformation" of the *b*-distribution at the edges.

Estimator performance *vs* V0M-based selection

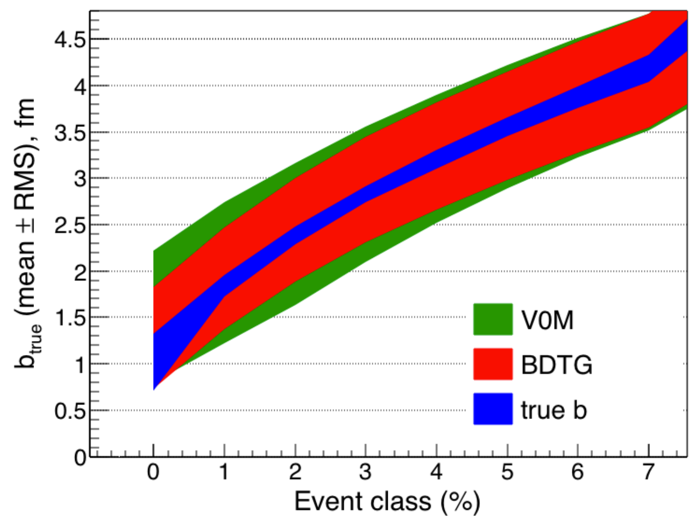
- Split the regression output of the estimator into centrality classes
- Look at *b impact* distributions:







Mean +/- RMS of b impact within centrality classes:

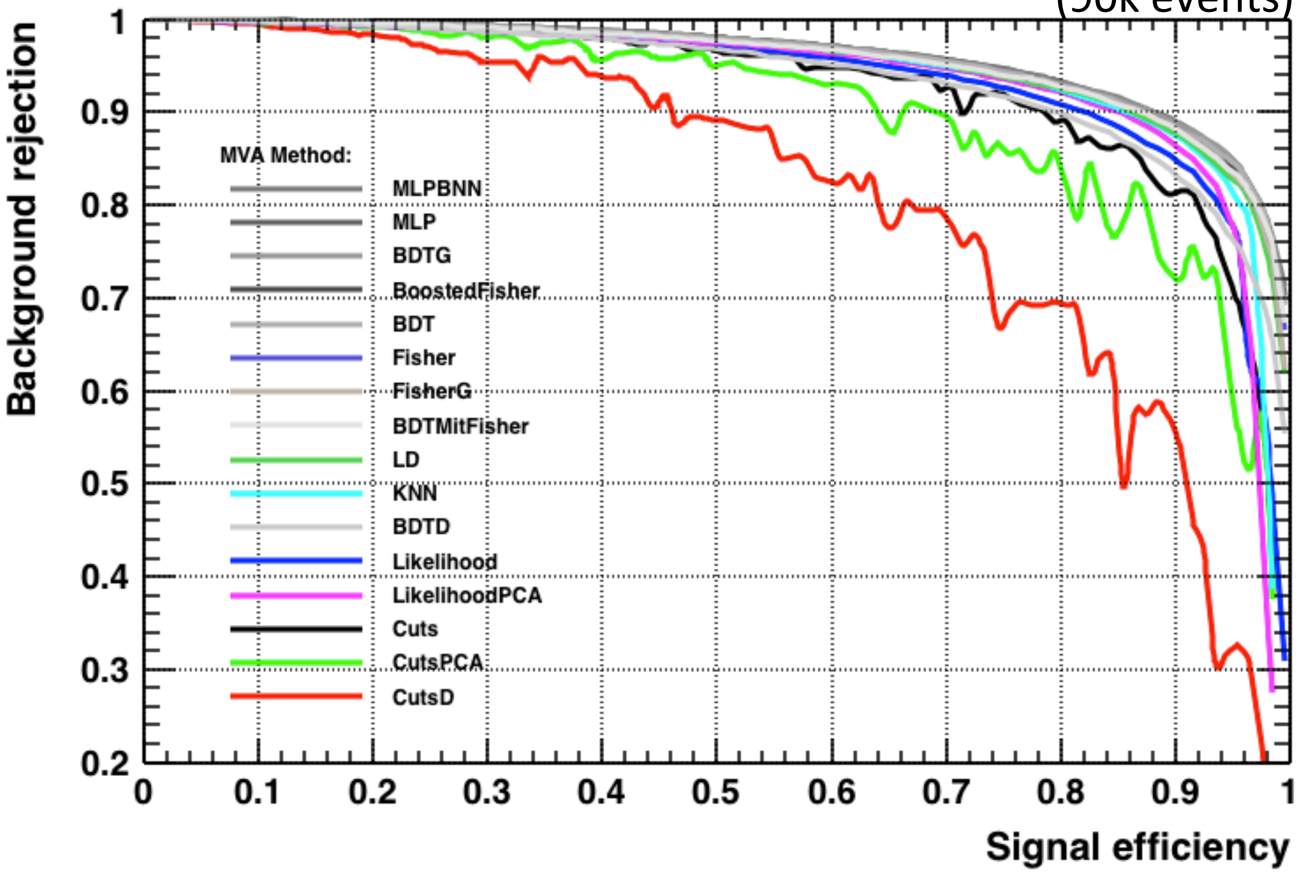


- smaller RMS by estimator than by V0M
- most central 0-1% events by estimator are closer to truth b

Classification task: find *most central* within 0-10% events

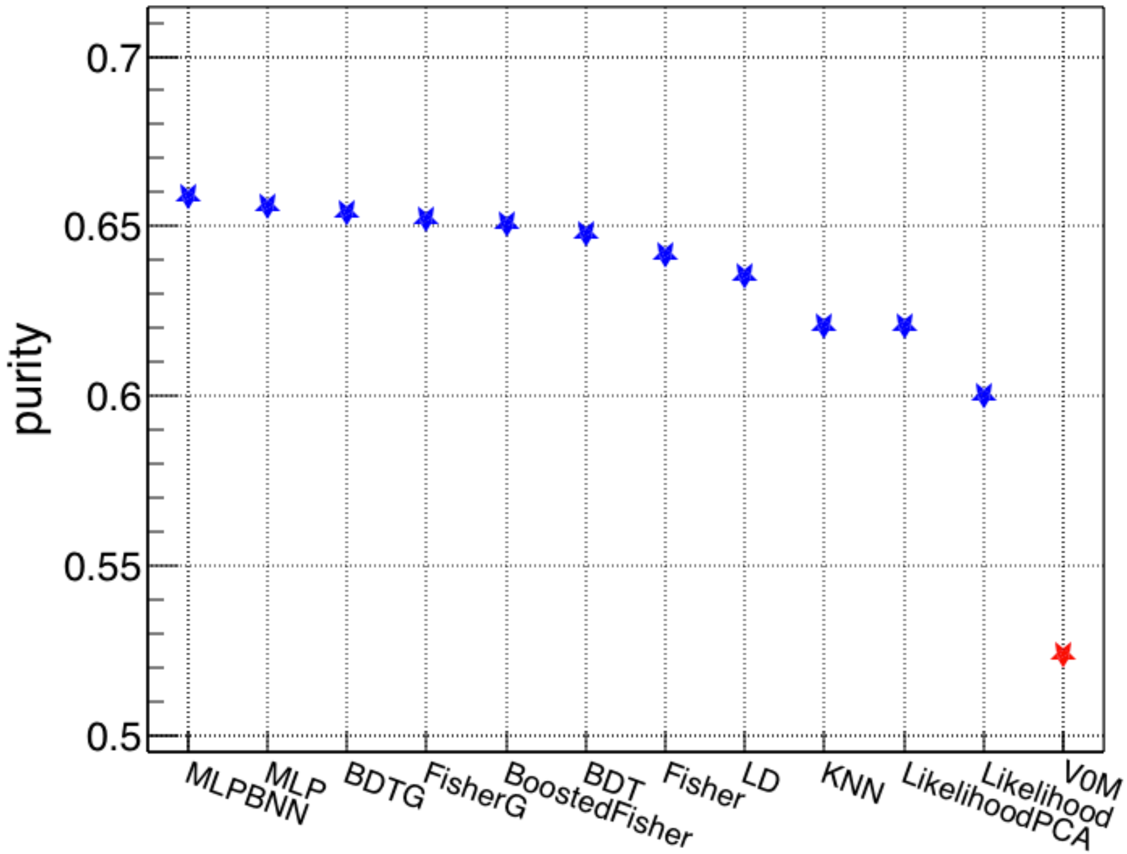
signal = 0-1% most central events (b_{impact} < 1.51 fm)
background = 1-10%

(90k events)



- Cut-based classifiers perform worse than others

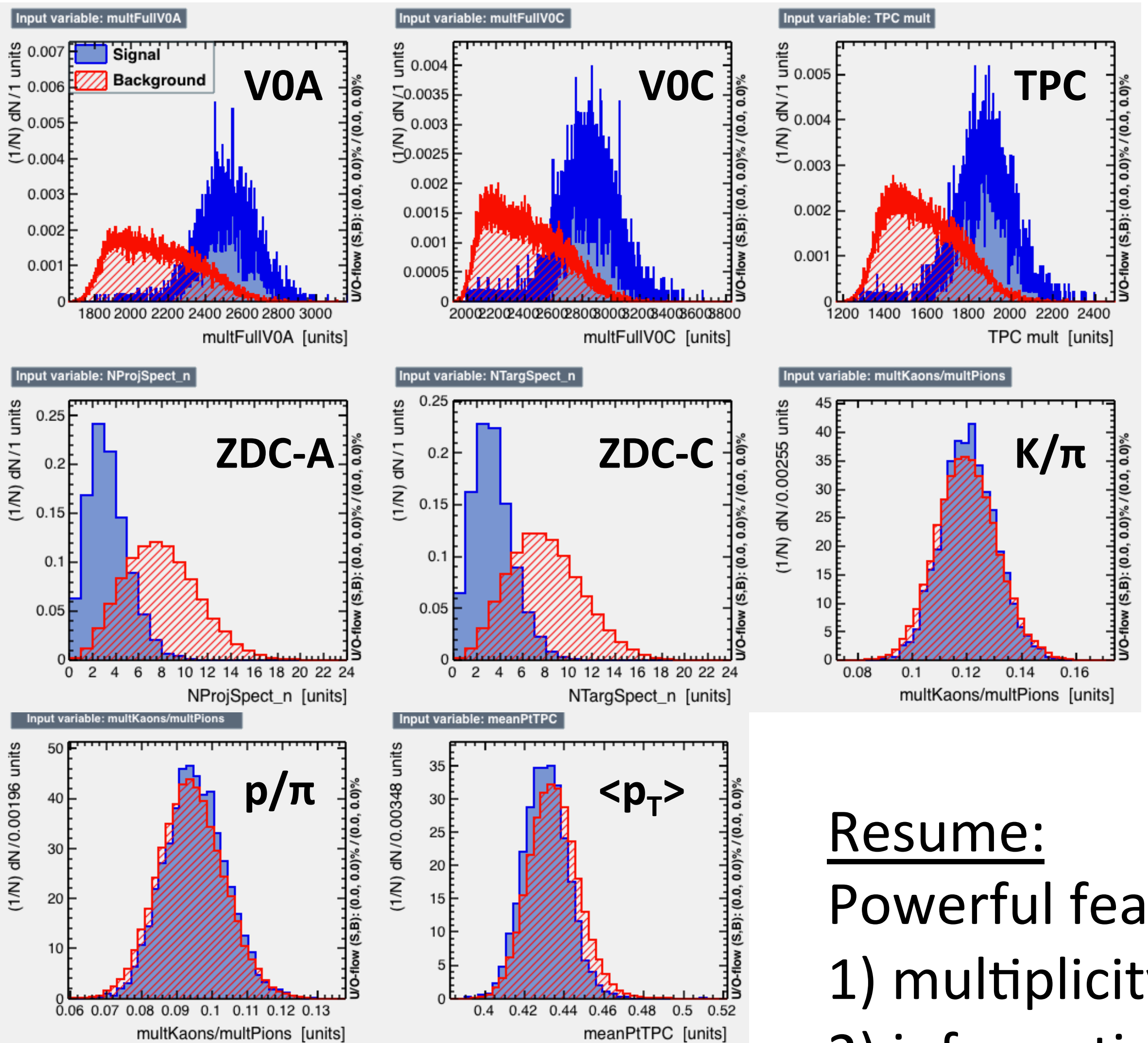
Fraction of signal events in 0-1% class selected by classifiers output

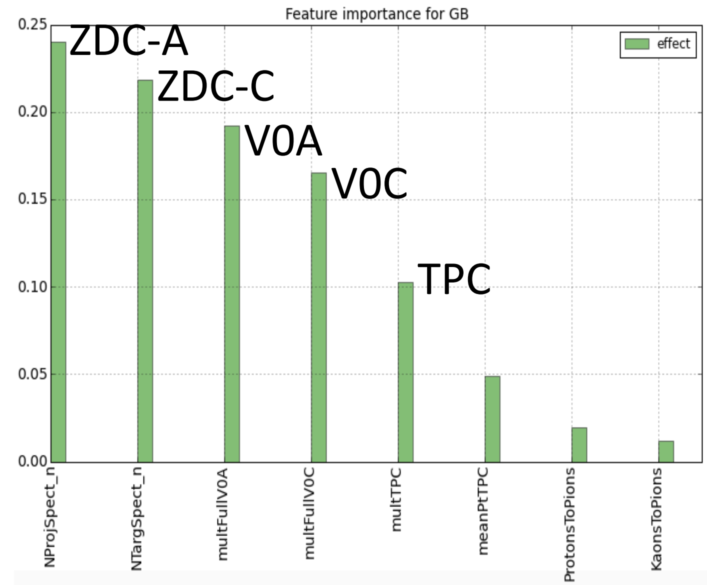


Purity increased by ~13%.

Significant gain of combined usage of information from several sub-detectors.

Which features are the most relevant?





Resume:

Powerful features are:

1) multiplicity in large rapidity ranges

2) information about spectators

Additional features didn't help.

Theses

- we want to go back to native definition of the notion of "centrality": through the *impact parameter.*
  (possible to perform similar study for N participants)
- we want to keep same statistics within centrality classes

  - Machine learning builds decision boundaries for us
  - MC vs Data – how to match?

- Accurate centrality is a baseline for many physics analysis:
  - crucial, for example, in fluctuations and correlations studies.