WWW.OEAW.AC.AT/SMI

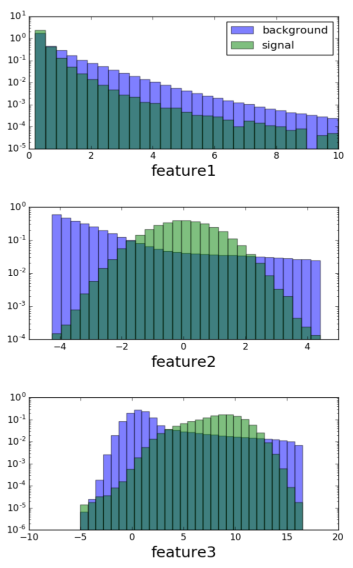
Electron Particle Identification with Boosted Decision Trees for a dielectron analysis in p-Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV with ALICE

Aaron Capon*
(on behalf of the ALICE collaboration)
*Stefan Meyer Institute for Subatomic Physics, Vienna, Austria

ÖAW AUSTRIAN ACADEMY OF SCIENCES

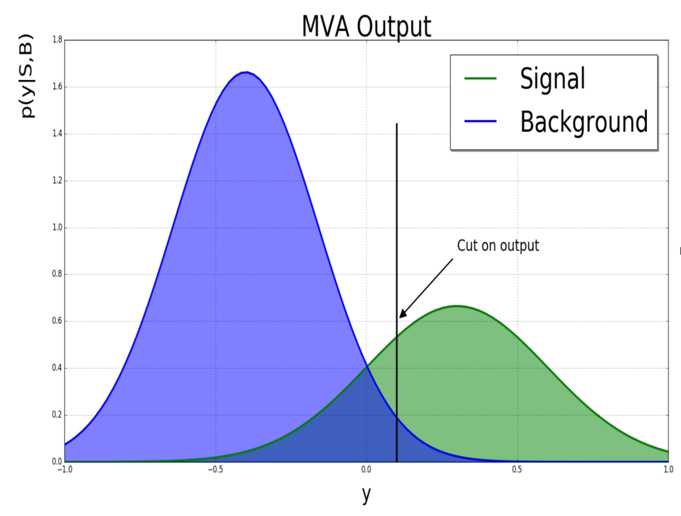
Particle identification within ALICE is typically performed using the "traditional" cut based method, where one or two variables are cut on at a time, in order to select out a desired particle species. This approach typically suffers from a high trade-off between efficiency and purity.
In past years an interest in a Bayesian approach has also been investigated, and shown to improve the resulting purity and efficiencies.
More recently, the ALICE collaboration has begun looking towards multivariate methods (MVA) as a way of improving a multitude of steps within the analysis chain. One such area that shows promising results, is the application of a MVA approach to particle identification. This work aims for an invariant low-mass dielectron spectrum, and therefore aimed at electron particle identification via Boosted Decision Trees (BDT). A comparison between this new method and a set of "traditional" cuts is presented using data from p-Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV.

Multivariate Classification



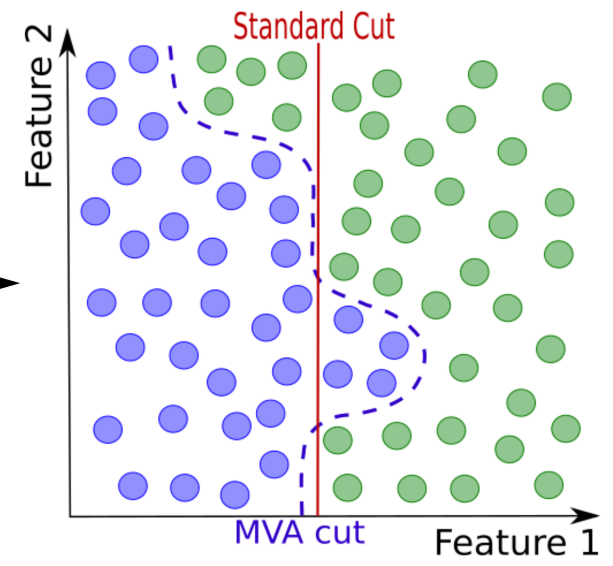
Aim: Find a mapping from a D-dim. input feature space to a 1-dim. output (y), such that signal and background have the least overlap.

The mapping is found in a training phase which utilises Monte Carlo (MC) simulations.



A single cut is then applied on y (MVA cut).

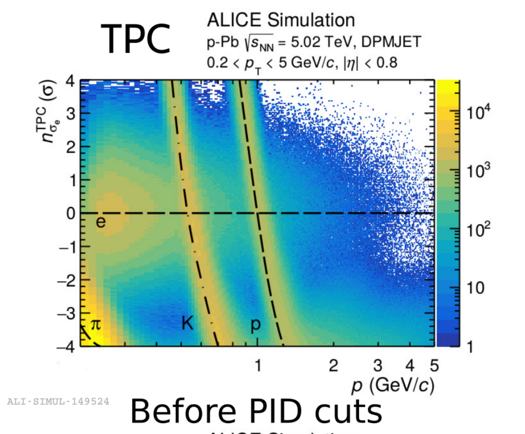
The cut, which corresponds to y = const, defines the decision boundary in the D-dim. feature space.



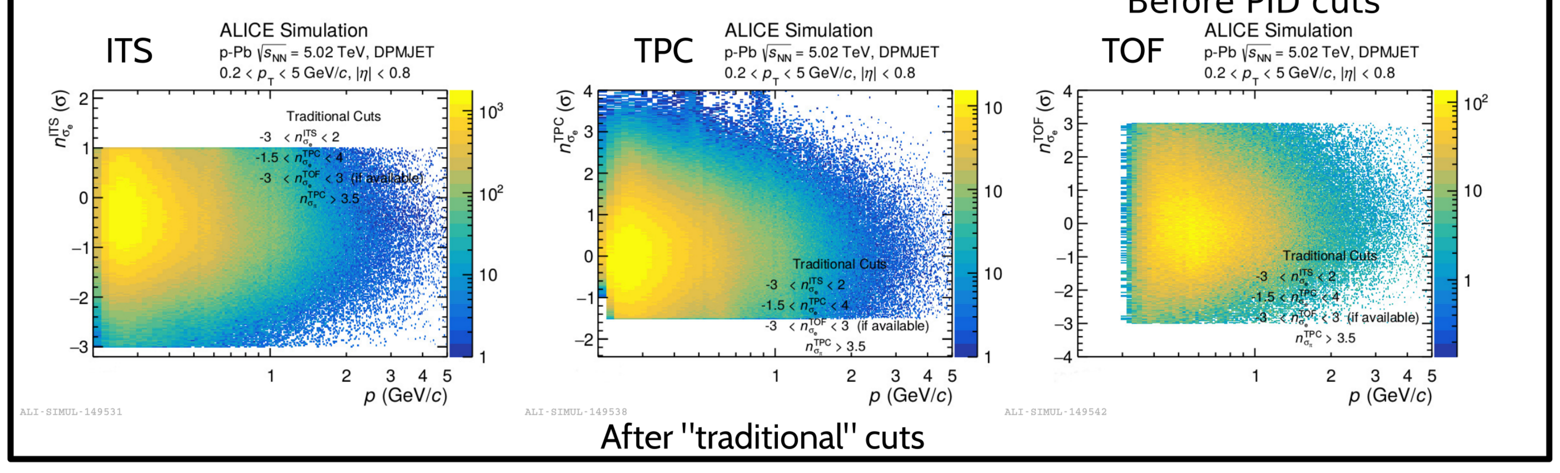
Traditional Particle Identification

Electrons are identified by using their specific energy loss within the inner tracking system (ITS) and time projection chamber (TPC), as well as their time-of-flight (TOF), which are represented as the number of standard deviations, nσ, away from the expected value for a given species and *p*.

The standard approach then sequentially cuts on these values in order to select a pure sample of electrons.



Before PID cuts



After "traditional" cuts

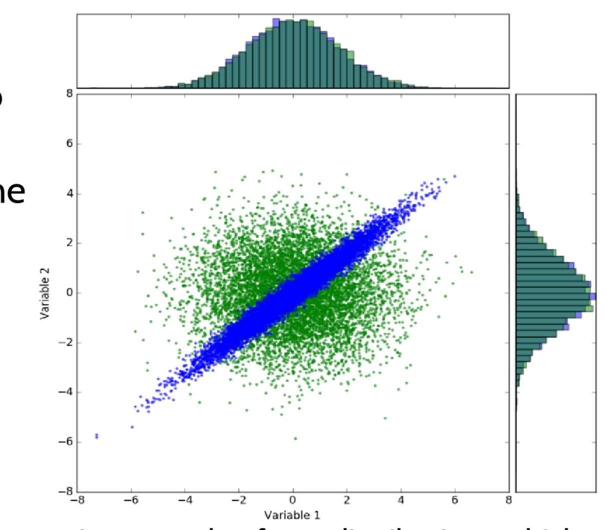
Monte Carlo Calibration

A multivariate approach requires a greater level of agreement between the data (RD) and the Monte Carlo (MC) (see example in plot to the right).
A technique that uses Boosted Decision Trees to find the appropriate weights, $\omega(x)$, to apply to the Monte Carlo was utilised [1].

$$p_{RD}(x) \neq p_{MC}(x) \rightarrow p_{RD}(x) \approx \omega(x) p_{MC}(x)$$

The weights are the ratios of RD and MC tracks in each hypercube,

$$\omega(x) = \frac{\omega_{RD}}{\omega_{MC}}$$



An example of two distributions which agree in their projections, but not in their full dimensional space.

The BDT chooses cuts based on the optimisation of the χ^2 function, and the procedure is repeated until the distributions reach the desired level of agreement.

$$\chi^2 = \sum \frac{(\omega_{MC} - \omega_{RD})^2}{(\omega_{MC} + \omega_{RD})}$$

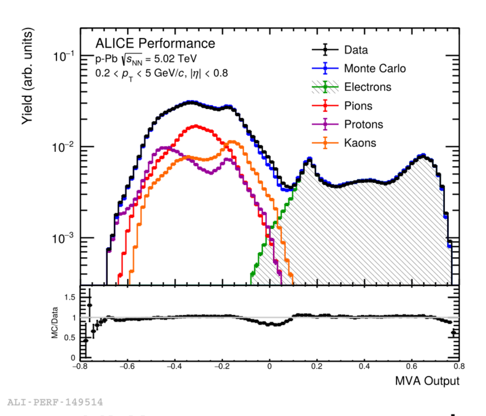
[1] Reweighting with Boosted Decision Trees -> http://arxiv.org/abs/1608.05806

MVA Particle Identification

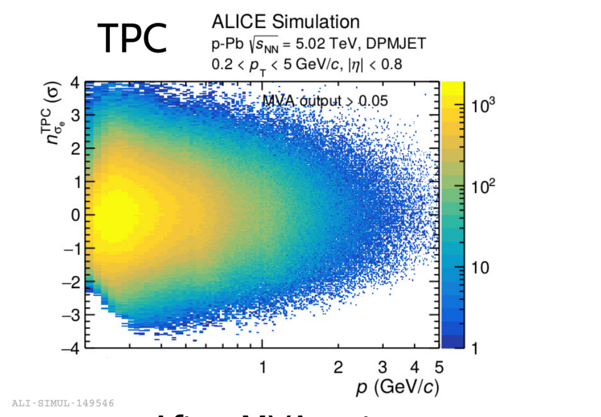
A boosted decision tree was trained with the same features used for the traditional approach (see box above for features).

Cutting on the MVA output now exploits the interplay of all detector features simultaneously, and therefore a much higher signal efficiency can be obtained.

The ratio between the MVA output from RD and MC is always compared in order to ensure that the overall MC distribution is representative of the RD.



MVA outputs compared



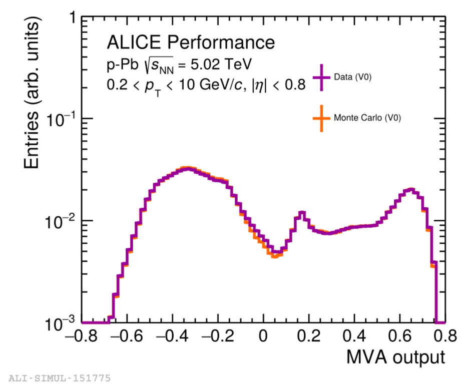
After MVA cut

Calibration Validation I

To check the effect of reweighting on the signal (electrons), a data driven test is being developed, which uses long lived neutral decay particles (V0 particles).

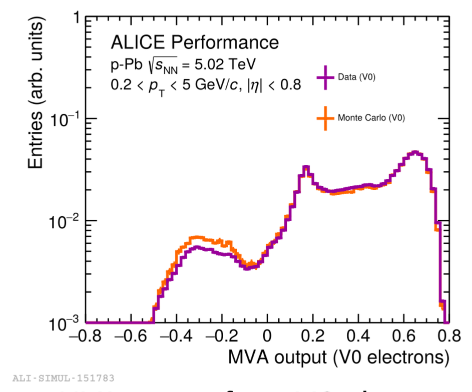
$$p^i_{RD}(x) \approx \omega(x) p^i_{MC}(x) \text{ for } i \in \{e, \pi, p, K\}$$

1) Select V0 decay particles (e's. π's, and p's) from the same set of events in RD and MC.
2) Perform the calibration (reweighting) step.
3) Classify the data sets with the analysis BDT.
4) Compare MVA outputs from V0 RD and MC, as done in analysis (see left plot).
5) Extract electrons, via decay kinematics, and compare MVA outputs (see right plot).



MVA output from V0 data

Extract electrons via their decay kinematics



MVA output from V0 electrons

Calibration Validation II

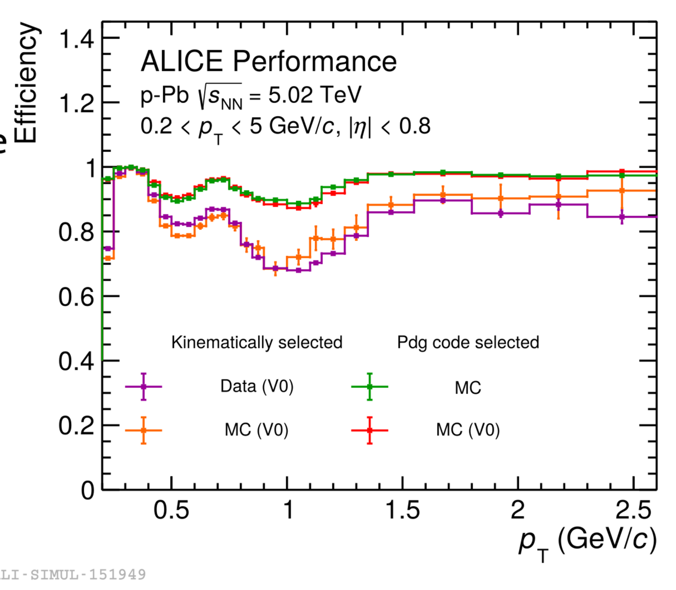
For the calibration tests to be applicable to the analysis data, two requirements must be met. Both of which can be shown via an efficiency comparison.

1) The electron selection criteria, via the decay kinematics, was correctly chosen such that the V0 MVA output comparisons between RD and MC are fair:
-> Compare purple (RD) and orange (MC)

2) The distribution of electrons between the analysis and the V0 data sets are similar after reweighting.

$$p^{e\pm}_{analysis}(x) \approx p^{e\pm}_{V0}(x)$$

-> Compare green (analysis) and red (V0)

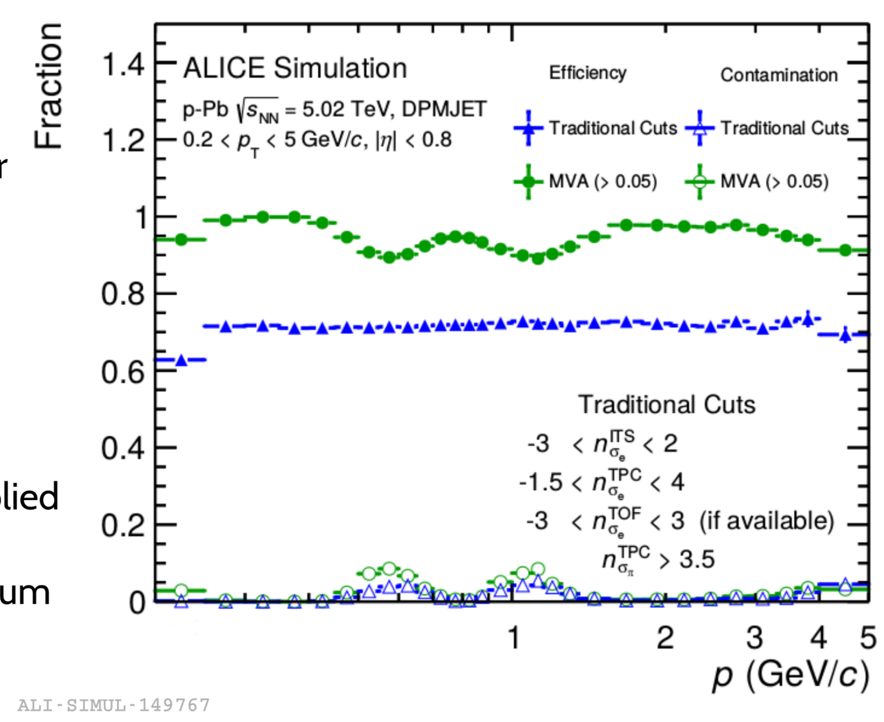


Note: the difference in efficiencies between the two criteria is due to some contamination that is introduced when selecting electrons via their kinematics.

Results

For the task of identifying electrons the Boosted Decision Tree can maintain the required high purity, while obtaining a significantly higher efficiency than that obtained with "traditional" cuts

The MVA technique will now be applied to a multiplicity differential study of the dielectron invariant mass spectrum in p-Pb collisions at $\sqrt{s_{NN}}$ = 5.02 TeV.



ALICE

FTE NATIONALSTIFTUNG FORSCHUNG | TECHNOLOGIE | ENTWICKLUNG

∫dkΠ Doktoratskolleg Particles and Interactions

SMI – STEFAN MEYER INSTITUTE FOR SUBATOMIC PHYSICS