Gamma/hadron separation for water Cherenkov EAS detectors on the basis of multidimensional feature space using non parametric approach

V.Grabski[1], A.Vardanyan[2], A.Chilingarian[2]
1 Universidad Nacional Autónoma de Mexico, Mexico
2 Yerevan Physics Institute, Armenia

Introduction

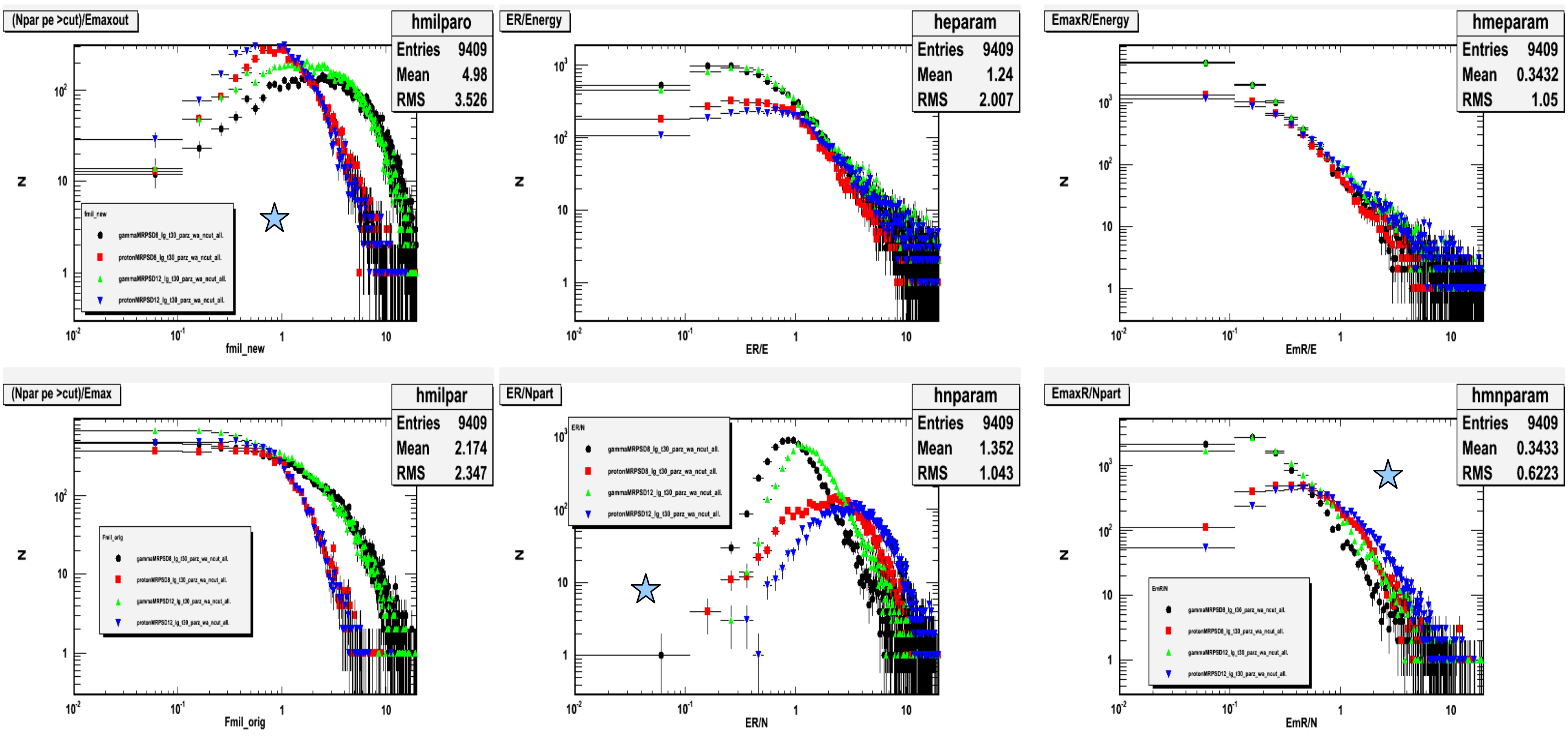
Two main methods for the gamma hadron separation could be mentioned:

- Based on the difference of gamma and hadronic showers topology(the size and particle density).
- Detection of muons inside the shower (this is effective above a few TeVs energy); below 1TeV 10-15% hadronic showers don't include muons (From CORSIKA).
- ✓ Tibet AS MD: both methods can be combined, as they will have clean muon detection.
- ✓ HAWC: there is no clean muon detection, but a large signal outside the shower core mostly comes from the muons.

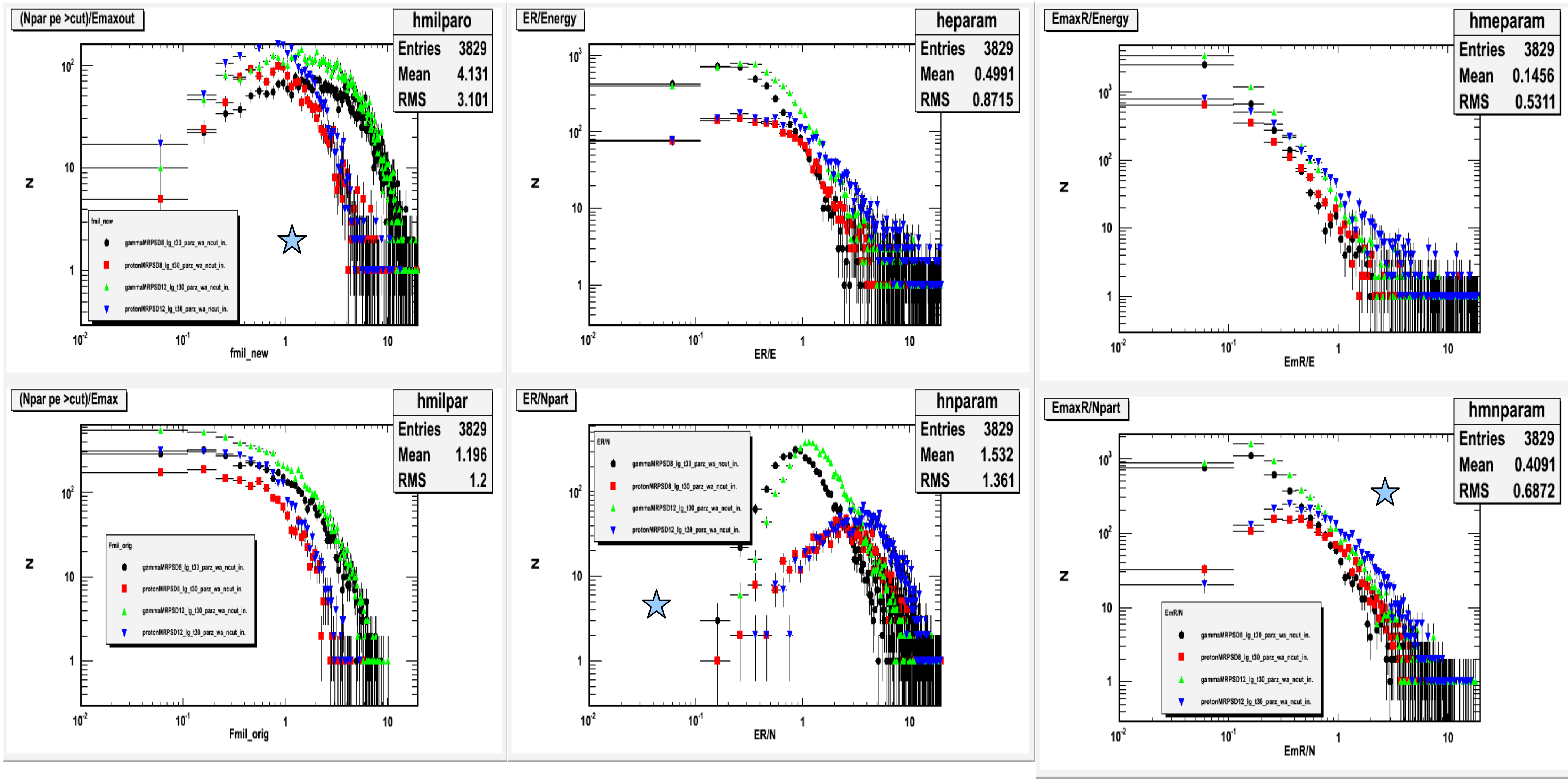
Variables for the gamma/proton separation

- Simulation and data processing was described in previous presentation. Here it should be mentioned that for the core position was used the Parzen algorithm (see previous presentation).
- Milagro – Ntank(PE>cut)/PEmax(R>40m)-large signals outside the core region; assuming that large signals come from muons. In parallel also Ntank(PE>cut)/PEmax(R>0m) has been considered.
- ERN – Σ(R*PE(R>40m))/Ntank(PE>cut): one can expect more large signals outside 40m radius circle for hadron showers due to muons and π^0. Division on Ntank for the normalization purpose.
- EmRN – Rmax*PEmax(R>40m)/Ntank(PE>cut) this is similar to milagro variable only it amplifies if large signal has large distances(again due to muons and π^0).

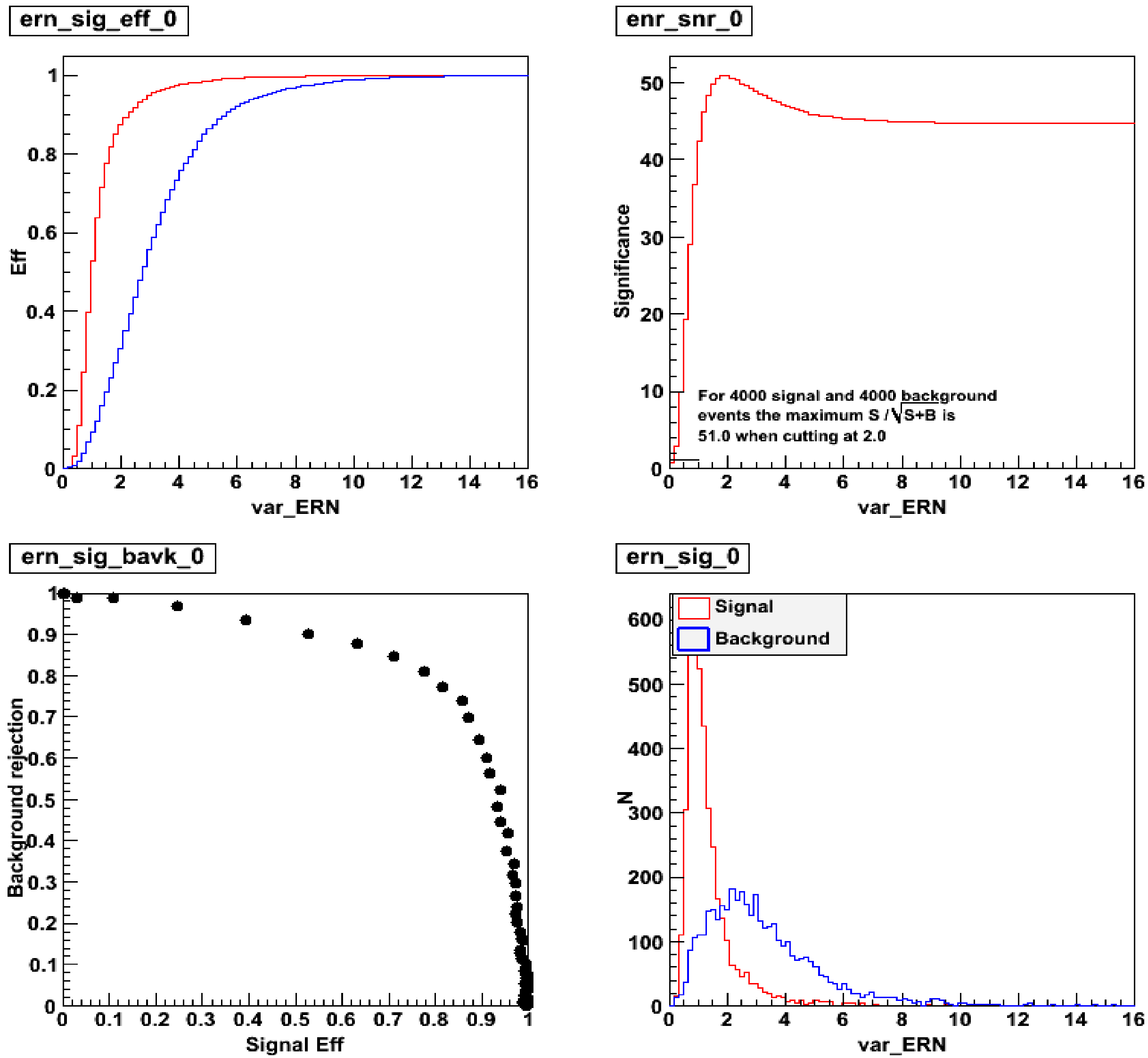
Distributions of the variables
trigger > 30, 50-10000GeV; core position is inside an area: 500x500m2



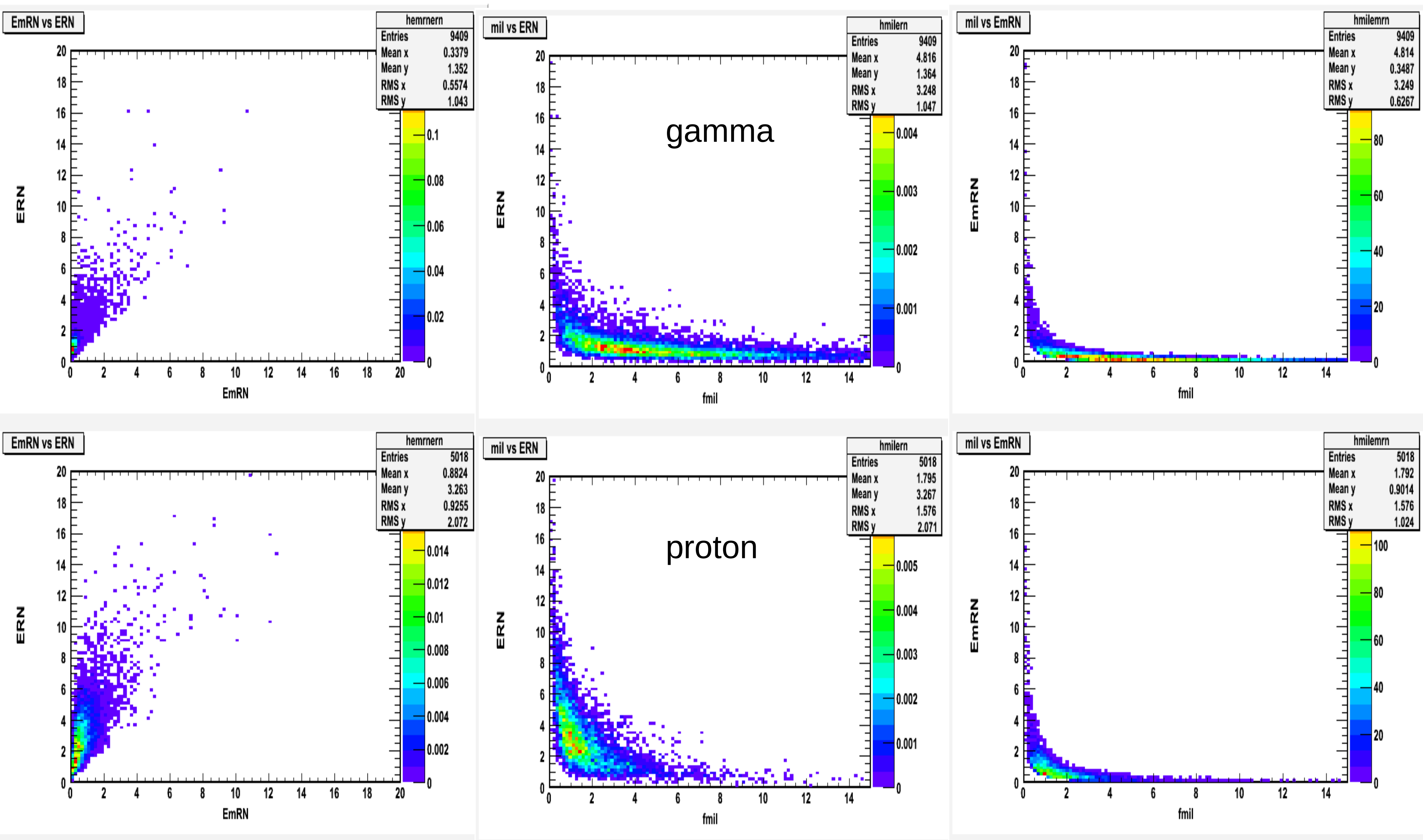
Distributions of the variables
trigger > 30, 50-10000GeV; core position is inside the detector area



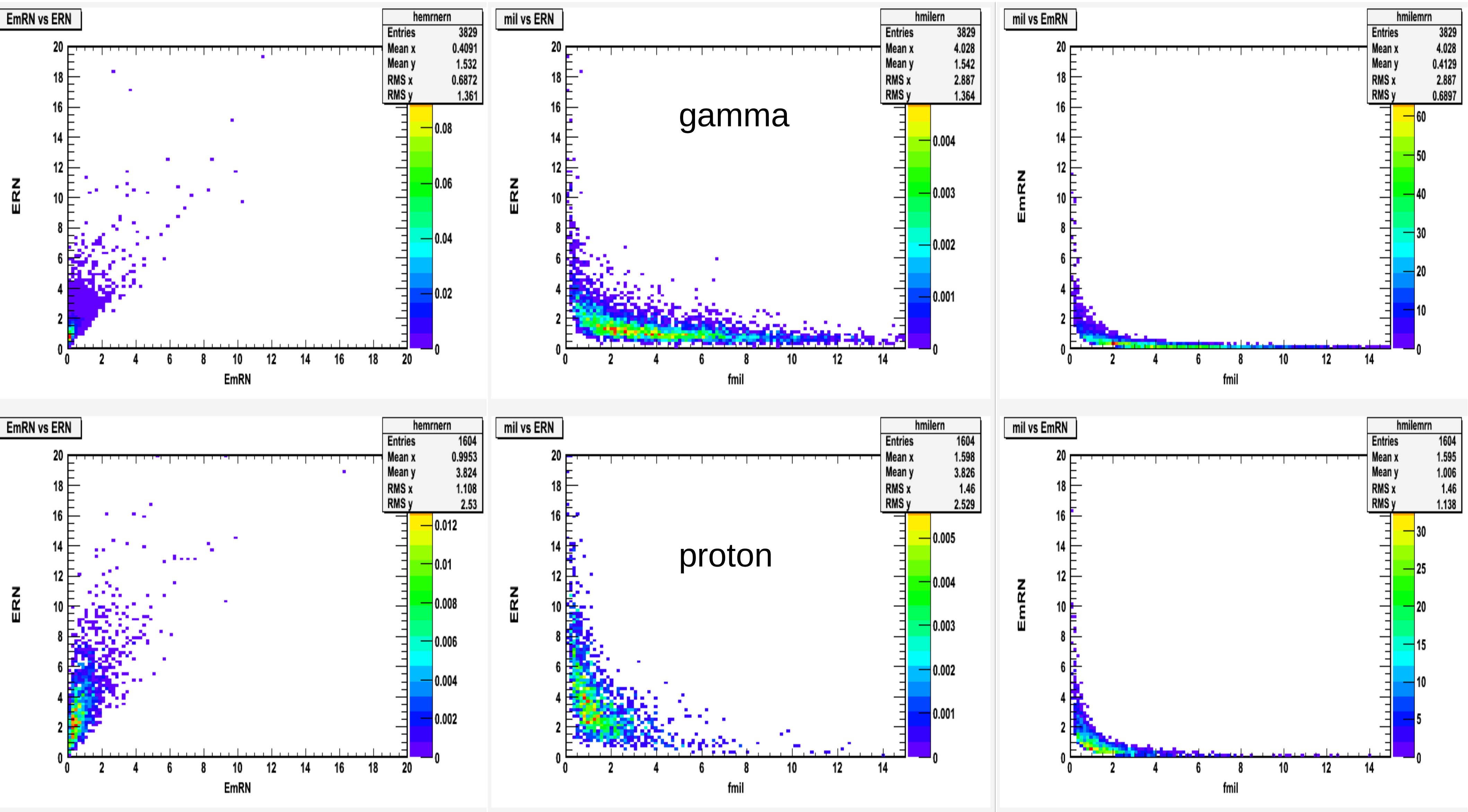
Single parameter (feature) analysis - simple cut method
later on is used data in large area for statistical reason only



Correlations between the features (large area)



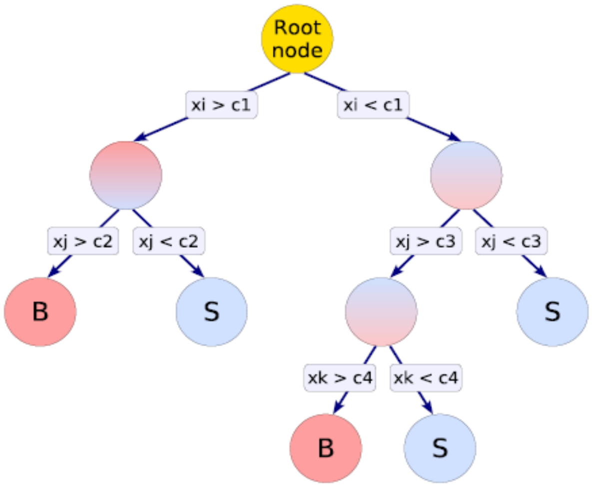
Correlations between the features (detector area)



Many of TMVA non parametric algorithms have been used and 5 of them have better performance for this task

- **Rectangular cuts** -The RC based on the usage of an ensemble of rectangular cuts on discriminating variables.
- **Artificial Neural Networks** (ANN nonlinear discriminant analysis)- two different realizations of ANN have been used (ANI-ANN from ANI program package and TMLP-ANN from TMVA ROOT program package).
- **Support Vector Machine** (SVM)-The main idea of the SVM approach to classification problems is to build a hyperplane that separates signal and background vectors (events) using only a minimal subset of all training vectors (support vectors). The position of the hyperplane is obtained by maximizing the distance between the vector to be classified and the support vectors.

- **Boosted Decision Trees** (BDT)-A decision tree (BDT) is a binary tree structured classifier similar to the one sketched in Fig. Repeated left/right (yes/no) decisions are taken on one single variable at a time until a stop criterion is fulfilled.



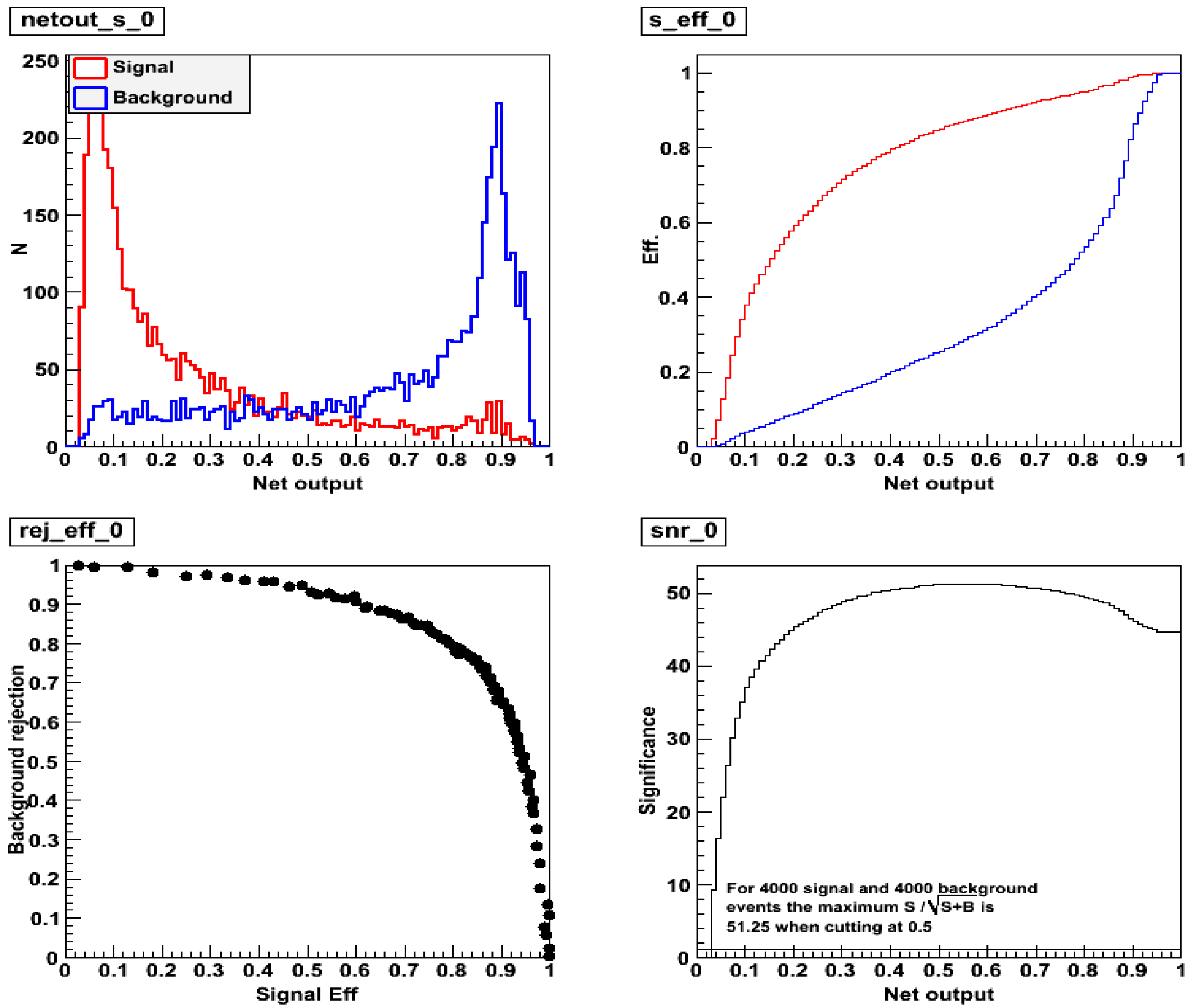
- **Predictive learning via rule ensembles** (RuleFit)-Friedman-Popescu's RuleFit method uses an ensemble of so-called rules create a scoring function with good classification power. Each rule defined by a sequence of cuts, such as:

$r1(x) = I(x_1>100.0) .I(x_3 < 40.0)$

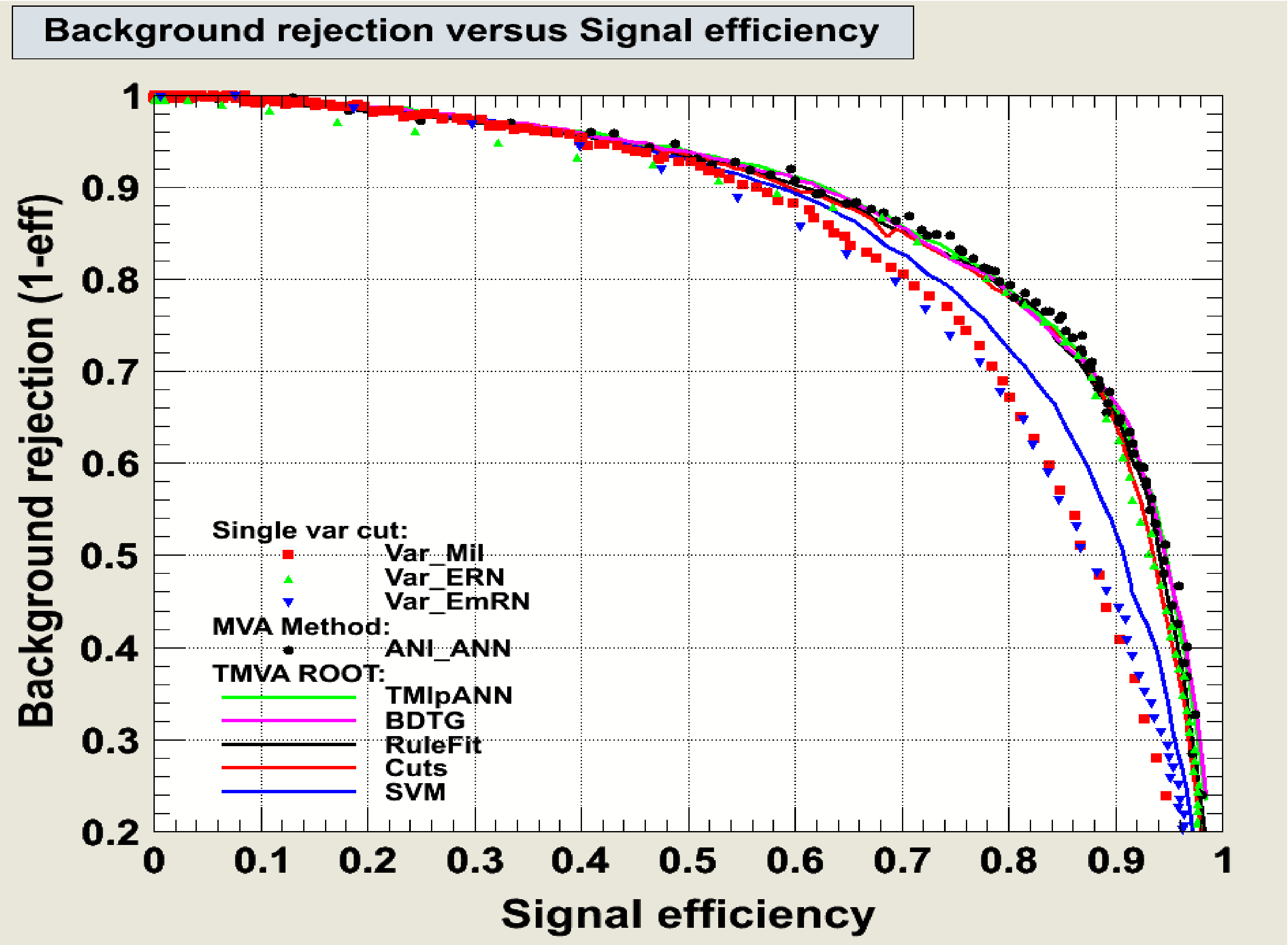
$r2(x)= I(0.45<x_4<1.0).I(x_1>150.0)$

r3(x)=................, where the x_i are discriminating input variables, and I(: : :) returns the truth of its argument. A rule applied on a given event is non-zero only if all of its cuts are satisfied, in which case the rule returns 1.

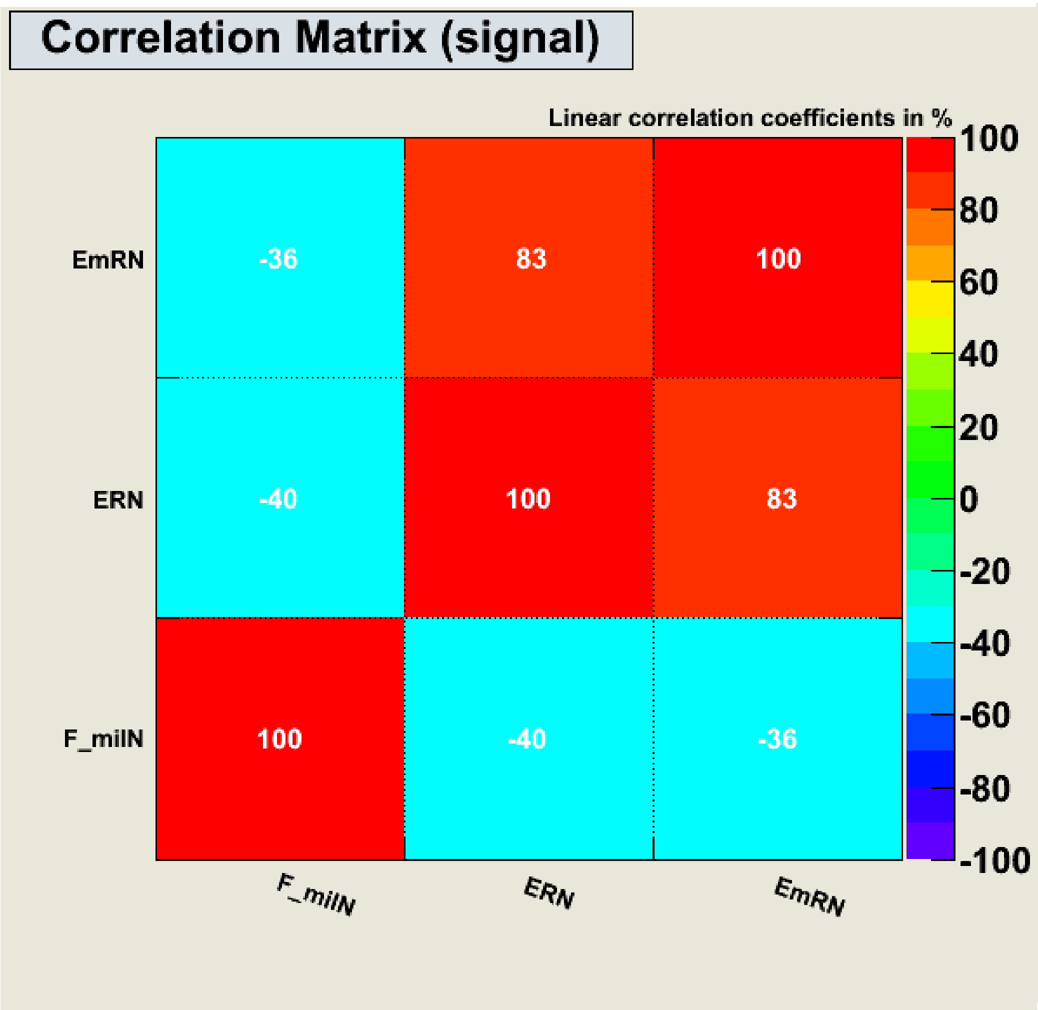
Results:ANI_ANN

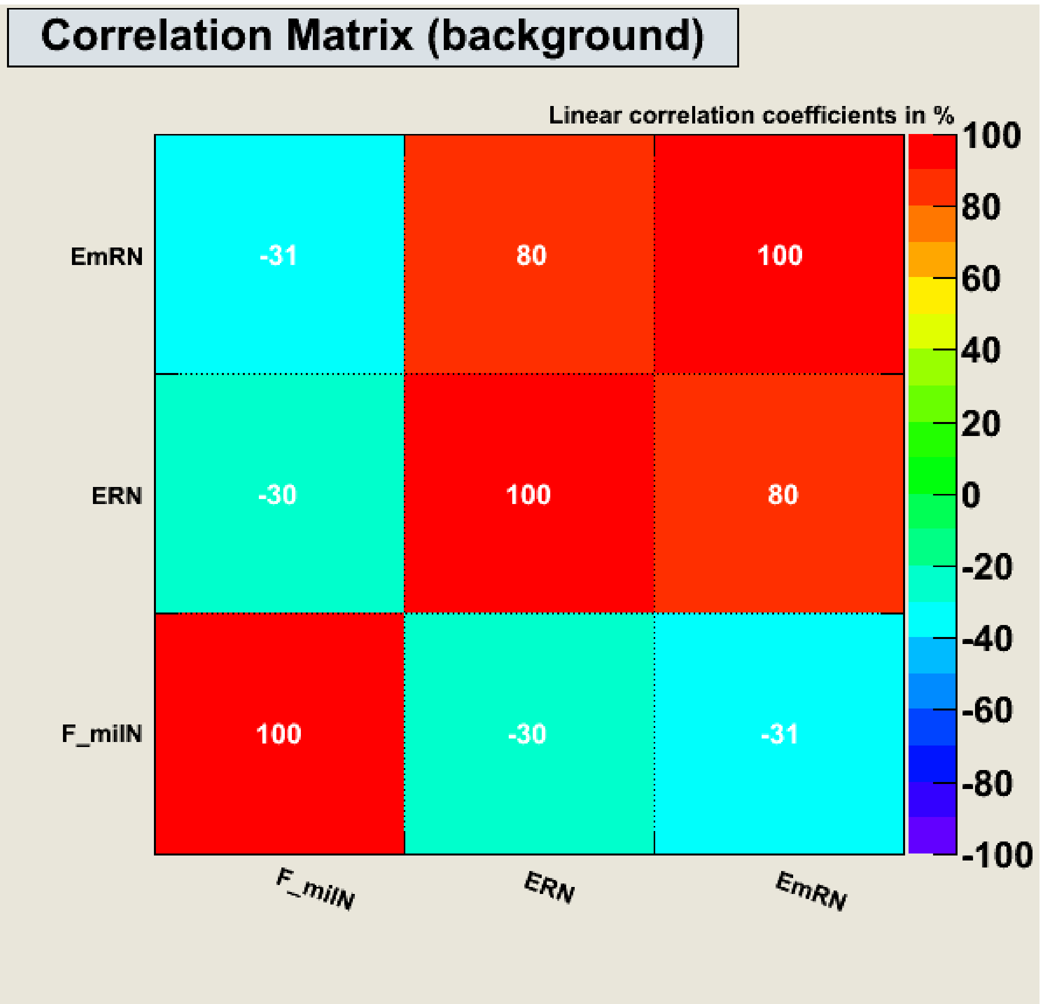


All together



Correlations between variables





Conclusions

- Obtained features are good enough for the simple cut usage. There are correlations between the features, so non linear algorithms should be used.
- The use of all features together demonstrate some improvement in discrimination performance.
- For better performance it should be found a feature that has good performance for the discrimination and is not correlated with one of the used features or the correlation is significantly different for two classes.

Acknowledgments
The author is grateful to Yerevan physics institute to spent a sabbatical year in YerPhi and PASPA DGAPA of UNAM for a partial financial support.