

Lambda_c Offline Selection & TMVA Update

09/02/2012

Recap

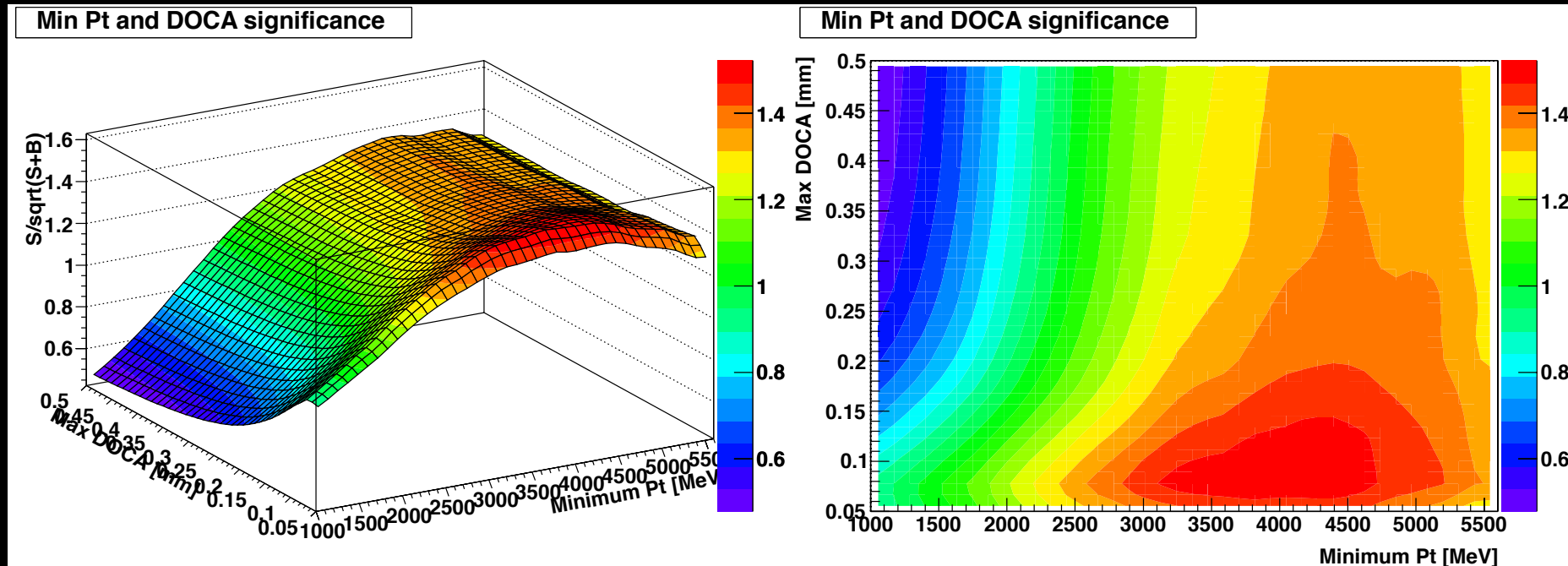
- Performed rectangular cuts using a hybrid strategy, inclusive charm MC for bkg, Lc signal for sig.
- Then moved onto TMVA optimisation. MVA methods looked promising BUT...
- TMVA's built in rectangular cut methods seem less than sensible (or at least my configuration of them, which would be funny given it's the default...)
- How do traditional cuts based methods really compare to the true MVA methods?
- Instead of using TMVA's cuts, scale the MVA methods signal and bkg to our hybrid study.

Hybrid Optimisation

- Take our 2 strongest discriminants, Lc Pt and daughter DOCA, and apply some sensible global cuts:
 - All daughters track $\chi^2 < 5$
 - proton PIDp > 5
 - Kaon PIDK > 5
- Use truth matching to get signal and bkg, plot significance...
- Inclusive charm has 98 signal events, ~150k bkg, use ratio of truth matched Lc candidates in the signal MC and inclusive charm MC as a scaling factor.

Hybrid “best” significance

- “Best” because still preliminary, could probably be improved but I doubt by a significant amount...
- Get a top significance of around 1.5σ .

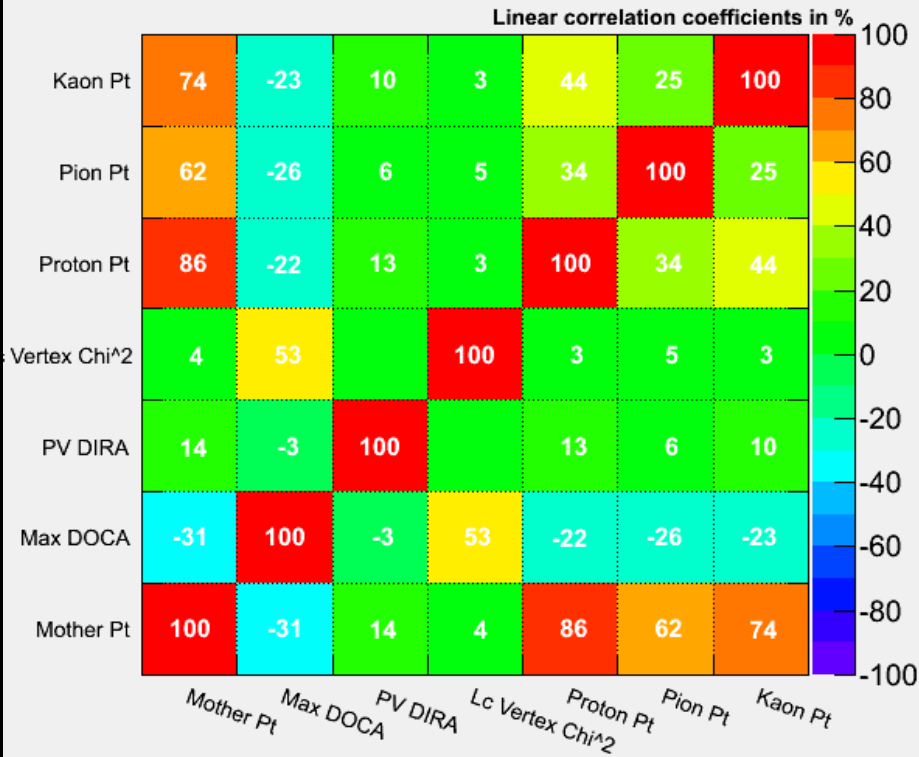


TMVA Optimisation

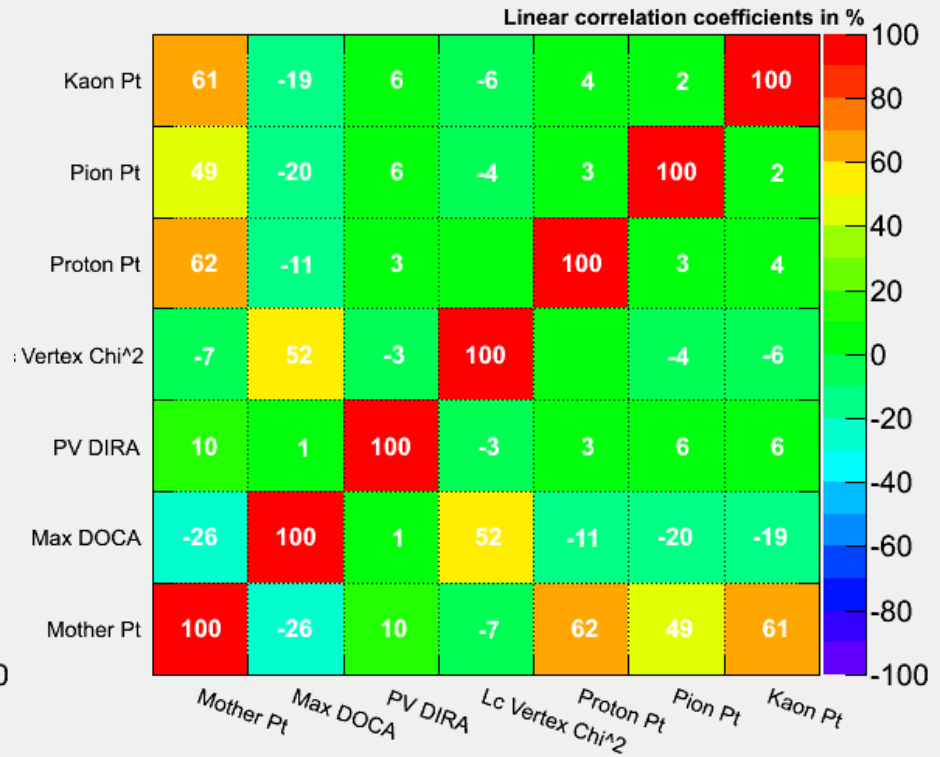
- As we have discovered, TMVA needs some thought into its setup. Ideally use variables with...
 - Strong discriminating power.
 - Low correlation (but so long as good discriminant doesn't really matter...)
 - Keep out any PID info because it can't be trusted at MC.
 - Keep out track χ^2 info to make getting efficiencies easier (Patrick feel free to call me out on any rubbish here!)
- So have implemented the same global track χ^2 and PID cuts as the hybrid and used the following variables in the optimisation...

TMVA Optimisation

Correlation Matrix (signal)

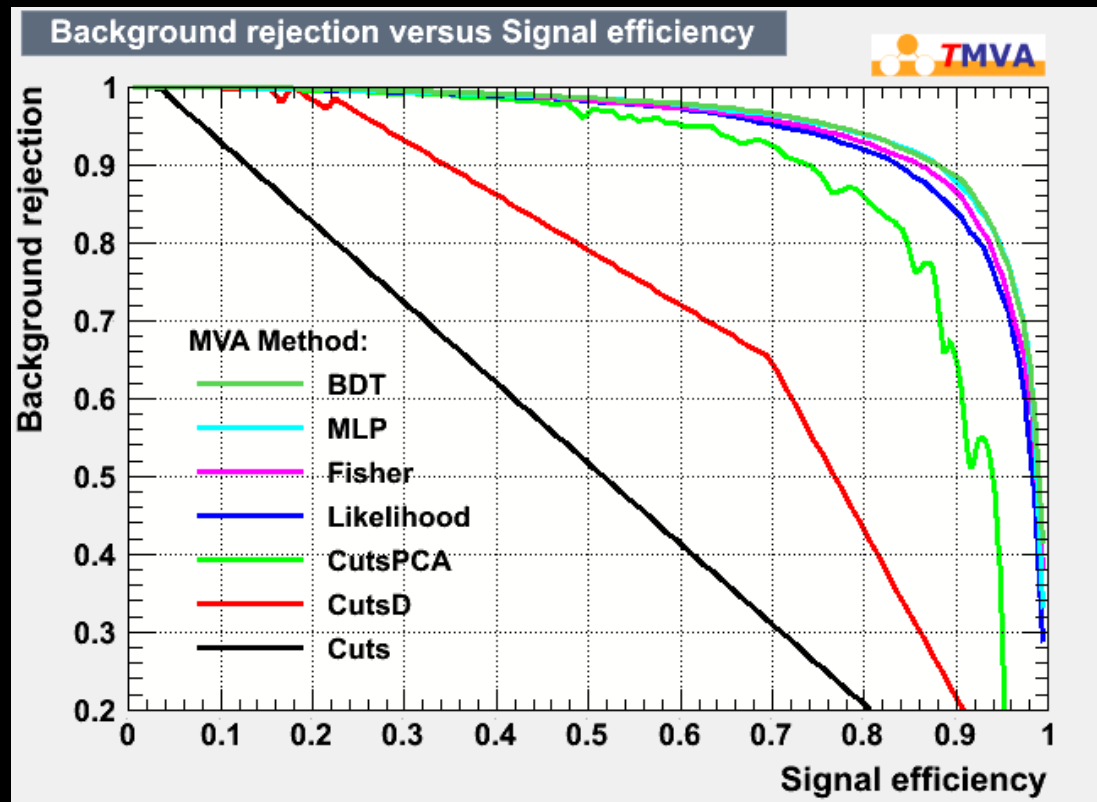


Correlation Matrix (background)



TMVA Optimisation

- ROC curve gives idea of discriminating power which is (I think!) independent of signal and background yields used.
- No surprise the cuts methods are rubbish.
- From this suspect the MVA methods are all similar.



TMVA Optimisation

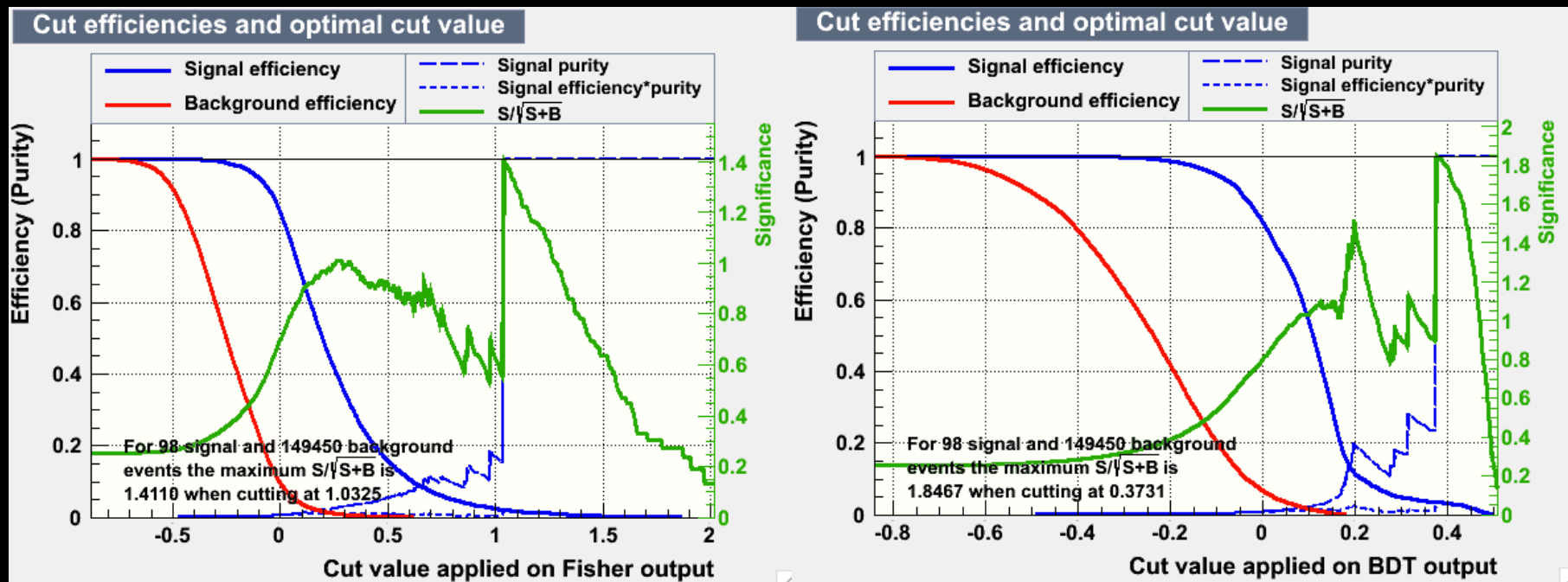
- However, signal and bkg yields not taken into account.
- Use the same signal and bkg numbers as the hybrid and check the significances (98 signal, 14950 bkg) we get...
- From the logfile:

```
=====  
Classifier ( #signal, #backgr.) Optimal-cut S/sqrt(S+B) NSig NBkg EffSig EffBkg  
-----  
Cuts: ( 98, 149450) 0.0350 0.626883 3.207675 22.97463 0.03273 0.0001537  
CutsD: ( 98, 149450) 0.0150 1.04291 1.08766 0 0.0111 0  
CutsPCA: ( 98, 149450) 0.0250 1.71744 2.949586 0 0.0301 0  
Likelihood: ( 98, 149450) 0.5330 0.981118 36.75922 1366.991 0.3751 0.009147  
Fisher: ( 98, 149450) 1.0325 1.41102 1.990971 0 0.02032 0  
MLP: ( 98, 149450) 0.8158 1.07517 51.72837 2263.002 0.5278 0.01514  
BDT: ( 98, 149450) 0.3731 1.84674 3.410459 0 0.0348 0  
-----
```

Naively expect BDT to be the clear winner. But yields of signal and bkg are very small. We should be very sensitive to the exclusion of individual events. Supported by...

TMVA Optimisation

- Looking at the classifier efficiencies:



Significance isn't exactly stable, need higher statistics to get more trustworthy results (eg on the BDT significance may be higher around 0.3 cut value with more results).

Further comments...

- So need more statistics... BUT have run over 2.5million of the 5million inclusive charm MC events on BookKeeping!
- Can double our stats, may be useful, will run on full 5million
- MLP performance looks much worse than BDT/Fischer at this stage:

- Overall outlook seems positive, we should get some improvement over rectangular cuts even at this stage with BDTs.

