

# Lc TMVA selection update

02/04/2012

# New TMVA Config 1

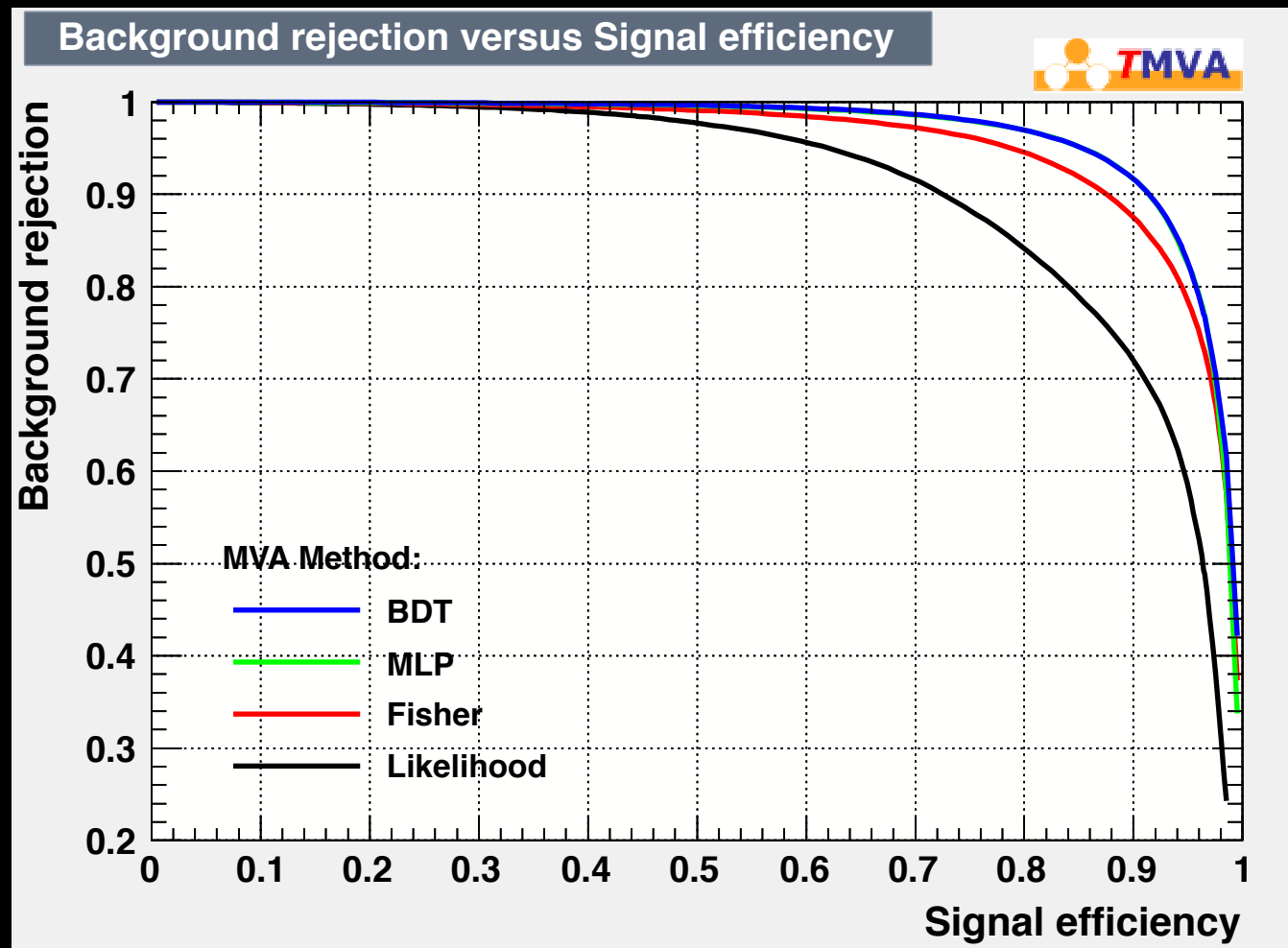
- Suspected previous attempts at MVA for Lcs did not have enough statistics to work with, hence gave gibberish.
- Have run over full Lc  $\rightarrow$  pKpi MC in bookkeeping (5m events) and moved from inclusive charm to minbias for background (10m events).
- Want to compare this to the rectangular cut based optimisation we perform manually.

# New TMVA Config 2

- TMVA Setup:
  - Global cuts:
    - Stripping 12
    - Kaon PID as  $K > 5$ , Proton PID as  $p > 5$
    - All daughters: track  $\text{Chi}^2 < 5$ ,  $P > 2\text{GeV}$
    - Truth matching with BKG CAT
  - Variables Used in Optimisation:
    - Lc: Pt, DOCA, DIRA, Vertex  $\text{Chi}^2$
    - Proton, Pion, Kaon: Pt
  - 80k signal candidates and 768k bkg candidates pass global selection. Half of each set used for testing and half for training.

# ROC Curve

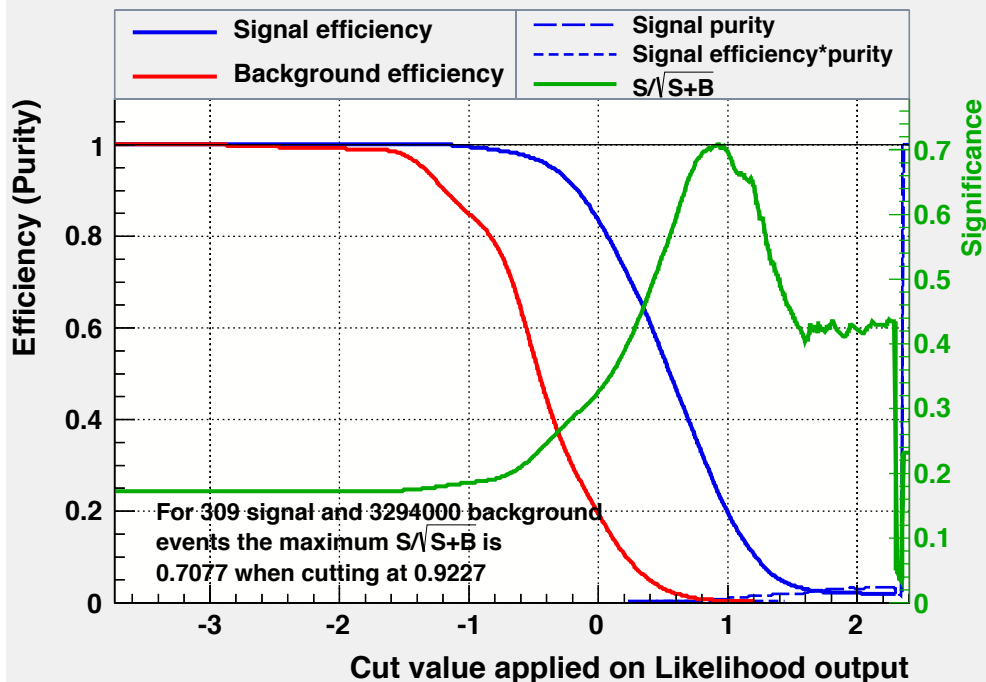
- Not much to say, except curves are smoother than before due to more statistics. BDT and MLP look best.
- Looked at MVA's cut-based methods but they spout nonsense as usual.



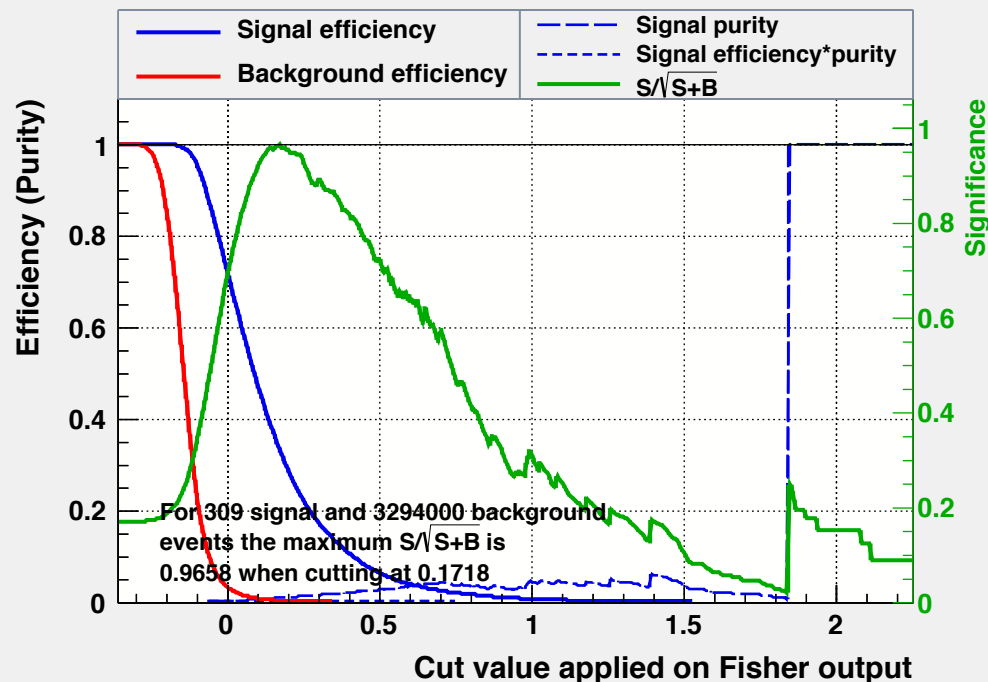
# Classifier Efficiencies

- Use the same number of truth matched Lc in the minbias sample as the signal fed in and the size of the minbias as background.
- Shown below likelihood and fisher.

Cut efficiencies and optimal cut value



Cut efficiencies and optimal cut value

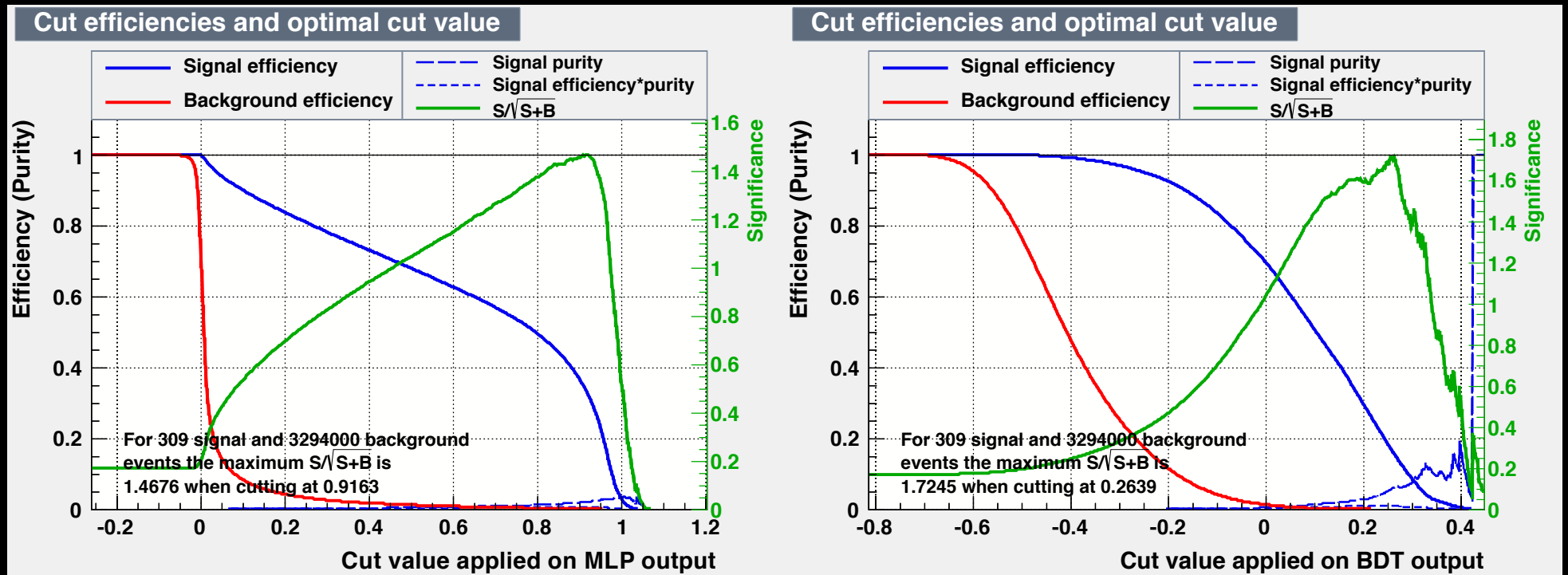


```

-----
--- Classifier ( #signal, #backgr.) Optimal-cut S/sqrt(S+B) NSig NBkg EffSig EffBkg
-----
--- Likelihood: ( 309, 3294000) 0.9227 0.707732 74.82677 11103.48 0.2422 0.003371
--- Fisher: ( 309, 3294000) 0.1718 0.965771 102.5165 11165.29 0.3318 0.00339
--- MLP: ( 309, 3294000) 0.9163 1.46757 102.5779 4782.916 0.332 0.001452
--- BDT: ( 309, 3294000) 0.2639 1.72451 47.49005 710.8696 0.1537 0.0002158
-----
    
```

# Classifier Efficiencies 2

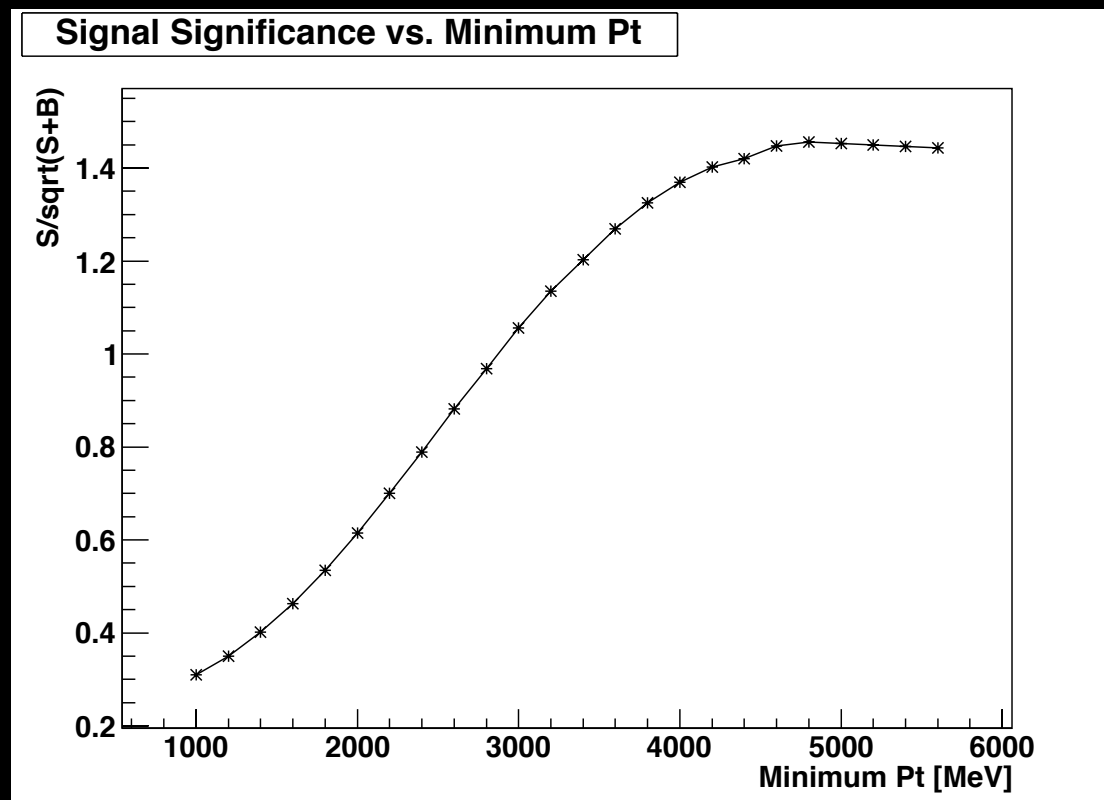
- MLP and BDT are clear winners over the less sophisticated MVA methods at these stats.



Classifier	( #signal, #backgr.)	Optimal-cut	S/sqrt(S+B)	NSig	NBkg	EffSig	EffBkg
Likelihood:	( 309, 3294000)	0.9227	0.707732	74.82677	11103.48	0.2422	0.003371
Fisher:	( 309, 3294000)	0.1718	0.965771	102.5165	11165.29	0.3318	0.00339
MLP:	( 309, 3294000)	0.9163	1.46757	102.5779	4782.916	0.332	0.001452
BDT:	( 309, 3294000)	0.2639	1.72451	47.49005	710.8696	0.1537	0.0002158

# Comparison with Rectangular cuts

- Shown below is a 1D Lc Pt optimisation using the same global cuts and dataset.
- Peak significance attained of 1.42, BDT gives 1.72, therefore ~20% better.



# Note on logfile output

- Want to draw attention to numbers of signal and background.
- Previously we looked at MVA with Lc but zero numbers of background were getting through the MVA methods.
- Now we get non-zero bkg levels through and these results can be trusted more.
- TMVA calculates statistics changes by stipulating more events just by scaling up the existing Nsig, Nbkg and significance. It doesn't move the MV cut value, which one would presumably wish to look at manually.
- So we should be getting at least 21% better performance on BDT and 3% better performance with MLP.
- Based on this feel confident recommending a BDT for offline selection.

```
-----  
--- Classifier ( #signal, #backgr.) Optimal-cut S/sqrt(S+B) NSig NBkg EffSig EffBkg  
-----  
--- Likelihood: ( 309, 3294000) 0.9227 0.707732 74.82677 11103.48 0.2422 0.003371  
--- Fisher: ( 309, 3294000) 0.1718 0.965771 102.5165 11165.29 0.3318 0.00339  
--- MLP: ( 309, 3294000) 0.9163 1.46757 102.5779 4782.916 0.332 0.001452  
--- BDT: ( 309, 3294000) 0.2639 1.72451 47.49005 710.8696 0.1537 0.0002158  
-----
```



# What next?

- Check BDT structure, and also how TMVA handles them.
- We don't blind the CF  $L_c \rightarrow pK^- \pi^+$  mode, so no harm in trying out both the simple rectangular optimisation and the BDT and MLP acquired here on the data to get the final word on significance.
- This is a priority as it will let us know roughly if we can expect to see the DCS  $L_c \rightarrow pK^+ \pi^-$  mode.