Update on $\tau \to \mu\mu\mu$ searches

J.Albrecht[1], M.Calvi[2], M.Chrzaszcz[3,4], L. Gavardi[2], J.Harrison[5], B. Khanji[2],G. Lafferty[5], E. Rodrigues[5], N. Serra[3], P. Seyfert[6]

September 19, 2013

[1] Dortmund, [2] Milano, [3] Zurich, [4] Krakow,

[5] Manchester, [6] Heidelberg

Status

1fb^{-1} analysis of $\tau \to \mu\mu\mu$ and $\tau \to p\mu\mu$ appeared in PLB.

Contents lists available at SciVerse ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

Searches for violation of lepton flavour and baryon number in tau lepton decays at LHCb ☆

LHCb Collaboration

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 27 May 2013
Accepted 29 May 2013
Available online xxxx
Editor: L. Rolandi

ABSTRACT

Searches for the lepton flavour violating decay $\tau^- \to \mu^-\mu^+\mu^-$ and the lepton flavour and baryon number violating decays $\tau^- \to \bar{p}\mu^+\mu^-$ and $\tau^- \to p\mu^-\mu^-$ have been carried out using proton–proton collision data, corresponding to an integrated luminosity of 1.0 fb^{-1}, taken by the LHCb experiment at $\sqrt{s} = 7$ TeV. No evidence has been found for any signal, and limits have been set at 90% confidence level on the branching fractions: $\mathcal{B}(\tau^- \to \mu^-\mu^+\mu^-) < 8.0 \times 10^{-8}$, $\mathcal{B}(\tau^- \to \bar{p}\mu^+\mu^-) < 3.3 \times 10^{-7}$ and $\mathcal{B}(\tau^- \to p\mu^-\mu^-) < 4.4 \times 10^{-7}$. The results for the $\tau^- \to \bar{p}\mu^+\mu^-$ and $\tau^- \to p\mu^-\mu^-$ decay modes represent the first direct experimental limits on these channels.

2011 results:

1. Obtained limit for $\tau \to \mu\mu\mu$: 8.0×10^{-8}.
2. Belle(BaBar) results: $2.1(3.2) \times 10^{-8}$ at 90% CL.
3. For 2012 + 2011 planned to implement several improvements.

Status

For now we use:

1. Stripping 20.
2. Signal sample: official+Krakow produced sample ($1M + 1M$).
3. *bb* and *cc* samples: official+Krakow. In total 30M events.
4. General strategy stays the same as 2011.

Cross section update

Analysis uses the knowledge of $c\bar{c}$ and $b\bar{b}$ cross sections. In 2011 both were measured by LHCb. For 2012 for the moment we assume:

- $\sigma_{b\bar{b}}^{8TeV} = 298 \pm 36\mu b$ from LHCB-PAPER-2013-016
- $\sigma_{c\bar{c}}^{8TeV} = \sigma_{c\bar{c}}^{7TeV} \times \frac{8}{7} = 6950 \pm 1100\mu b$

Cross checks on $c\bar{c}$

1. Pythia cross section calculation.
2. Comparing D_s yields in data.

Generated MC samples

1. In the 2011 analysis one of the complications from MC was the wrong mixture of tau sources.
2. For 2012 we solved this problem by simulating signal in 5 parts. One for each production channel:

$$\tau \to \mu\mu\mu = \begin{cases} B \to \tau \to \mu\mu\mu & 11.6\% \\ B \to D_s \to \tau \to \mu\mu\mu & 8.7\% \\ B \to D \to \tau \to \mu\mu\mu & 0.2\% \\ D_s \to \tau \to \mu\mu\mu & 75.0\% \\ D \to \tau \to \mu\mu\mu & 4.4\% \end{cases}$$

MC Generator Cuts

In order to use computing resources in more efficient way we introduced generator level cuts.

Signal sample[1]		Background sample(Dimuon)[2]	
$p_{t\mu}$	$> 250 MeV$	$p_{t\mu}$	$> 280 MeV$
p_{μ}	$> 2.5 GeV$	p_{μ}	$> 2.9 GeV$
		$m(\mu\mu)$	$< 4.5 GeV$
		$\mathrm{DOCA}(\mu\mu)$	$< 0.35 mm$

Gain a factor of $\sim 2 - 3$ in signal statistics compared to 2011 and factor of 8 in background.

[1] $X \to \tau \to 3\mu$, $\mathrm{D_s} \to \eta(\mu\mu\gamma)\mu\nu$, $\mathrm{D_s} \to \phi(\mu\mu)\pi$
[2] $c\bar{c}$, $b\bar{b}$

Trigger lines

In 2011 we took all trigger lines into account. Studies shown we can gain on limiting ourselves to specific lines (2011 data sample).

Line Name	ϵ[%]	ϵ'[%]	β[%]	β'[%]
Hlt2CharmSemilepD2HMuMu	81.7	81.7	56.8	56.8
Hlt2DiMuonDetached	75.0	12.5	54.1	17.6
Hlt2TriMuonTau	66.3	2.9	60.0	12.2
Others	–	2.2	–	11.6

, where ϵ is the signal efficiency (any Hlt2physics), ϵ' is the gain of the efficiency.
β is the efficiency of background and β' is the gain of the bck efficiency
Rule of thumb (using $\frac{s}{\sqrt{b}}$ FOM) tells us that we can gain $\mathcal{O}(5\%)$.

Normalization channel

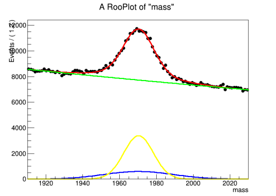
As last year we will use $D_s \to \phi(\mu\mu)\pi$.Similarly to signal channels we produced them with correct proportion:

1. $cc \to D_s \to \phi(\mu\mu)\pi$ 89.7%
2. $bb \to D_s \to \phi(\mu\mu)\pi$ 10.3%

We avoid reweighting of the samples as in 2011.

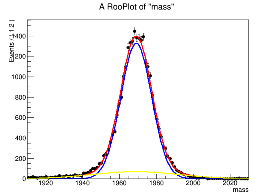
Mass correction

$D_s \to \phi(\mu\mu)\pi$ in data.



- mean $= 1970.3 \pm 0.9 MeV$

$D_s \to \phi(\mu\mu)\pi$ in MC.

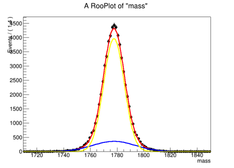


- mean $= 1969.1 \pm 0.60 MeV$

- $m_{\tau\to 3\mu} = \dfrac{1970.3}{1969.1} \times 1777.7 = 1778.8 \pm 1.1 MeV$

  In agreement with 2011.

Fit $\tau \to \mu\mu\mu$ in MC.



Background samples normalization

For the normalization of background samples($c\bar{c}$ and $b\bar{b}$) we used generator cuts efficiencies and corrected the nominal cross section accordingly:

$$\mathcal{L} = \frac{N_{MC}}{\varepsilon_{acc} \times \varepsilon_{gen} \times \sigma_{LHCb}}$$

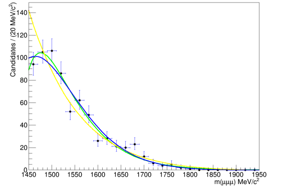
The obtained luminosities(per 1M events):

1. $\mathcal{L}_{cc} = 0.25 \pm 0.04 pb^{-1}$
2. $\mathcal{L}_{bb} = 1.20 \pm 0.15 pb^{-1}$

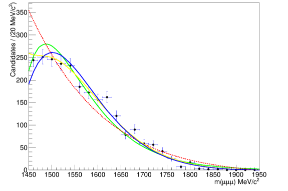
Dominant uncertainty from the cross section.

$D_s \to \eta(\mu\mu\gamma)\mu\nu$

1. The dominant background source of peaking background in this analysis is $D_s \to \eta(\mu\mu\gamma)\mu\nu$
2. In 2011 we suffered from lack of MC statistics.
3. Thanks to generator cuts our pdfs became more stable.
4. Pdf used: $\mathcal{P} = exp(m) \times Pol^n(m)$



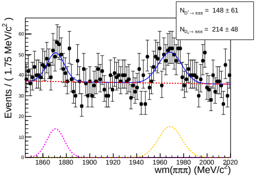
PID:**0.65**; **0.725**,GEO:**−0.48**; **0.05**

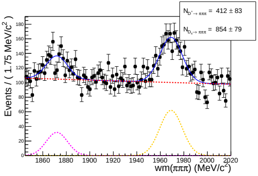


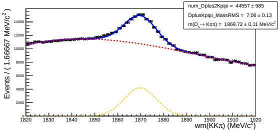
PID:**0.725**; **0.0.86**,GEO:**0.35**; **0.65**

$D \to$ hhh

In 2011 we saw a triple miss-ID background: $D^+ \to K\pi\pi$. This background was in trash-bins that were not used in the analysis. Also new sources of bck($D_x \to 3\pi$) are well under control.







- 2011 data
- 2012 data
- 2012 data

In 2012 there is still no significant amount of triple mis-ID background in the bins important to the analysis.

Isolating parameters

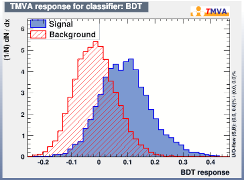
Inputs for isolating parameter(based on Giampiero work):

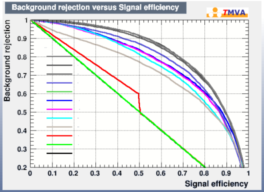
Variable	Description
IP χ^2	Impact parameter χ^2 wrt any PV
IP	Impact parameter wrt any PV
angle	angle between μ and track
doca	doca between the μ and the track
PVdis	$\lvert\overrightarrow{TV} - \overrightarrow{PV}\rvert$, signed according to $z_{TV} - z_{PV}$.
SVdis	$\lvert\overrightarrow{TV} - \overrightarrow{SV}\rvert$, signed according to $z_{STV} - z_{PV}$.
fc	$\frac{\lvert\overrightarrow{P_\mu} + \overrightarrow{P_{tr}} \times \alpha}{\lvert\overrightarrow{P_\mu} + \overrightarrow{P_{tr}} \times \alpha + P_{T_\mu} + P_{T_{tr}}}$ [3]

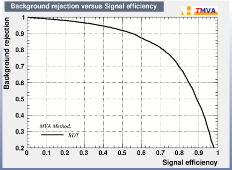
[3] α is the angle between $\overrightarrow{P}_\mu + \overrightarrow{P}_{tr}$ and $\overrightarrow{PV} - \overrightarrow{TV}$

Isolating parameters

1. In 2011 we used the isolation parameter developed for $B_s^0 \to \mu\mu$. For 2012 data we optimised the isolation parameter for our channel based on MVA(BDT).
2. We follow two approaches: train a MVA on signal vs. bkg tracks, and the isolating vs. non-isolating tracks.
3. We see a big improvement compared to old isolation.







Ensemble Selection

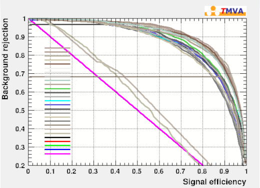
1. In the last few years people winning leading machine learning contests started to combine their classifiers to squeeze the best out of them.
2. This technique/method is know as Ensemble Selection or Blending.
3. The plan for $\tau \to \mu\mu\mu$ is to take it to the next level.
4. Combine not only different signal classifiers, but also different τ sources(slide 4).
5. Allows for usage different isolating parameters for each channel.

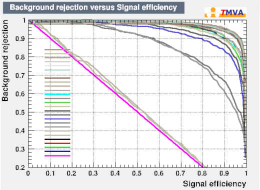
Ensemble Selection – How to

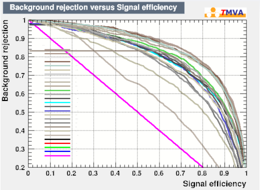
How to make an Ensemble Selection

1. Construct a reduced training set.
2. Train you different models on the reduced training set.
3. Combine/Blend all the models on the rest of the data set.
4. The output is a function that mixes the individual model predictions into a blended prediction, hopefully better than any individual result.

Ensemble Selection

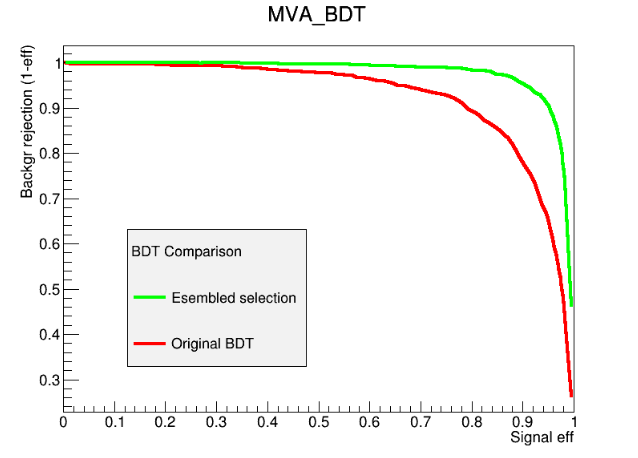






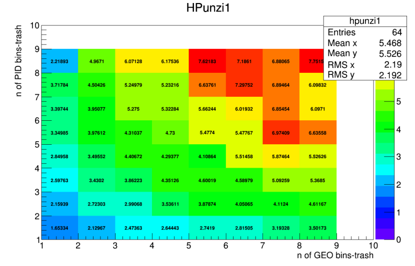
- $B \to D \to \tau$
- $D \to \tau$
- $B \to D_s \to \tau$

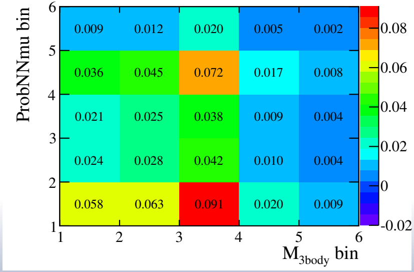
Ensemble Selection



Binning optimisation

For the 2011 analysis we had two classifiers: *PIDNN* and M_{GEO}. Each of them we optimised separately. For the 2012 analysis we are performing a simultaneous 2D optimisation.





- FOM as a function of N. of bins.
- Signal efficiency in 2011 binning.

Model dependence

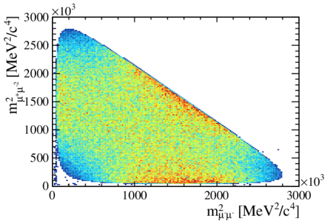
Minimal Lepton Flavour Violation Model[a]

[a] arXiv:0707.0988

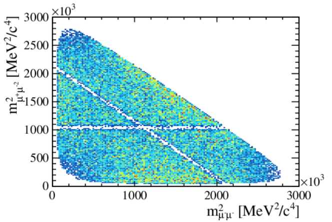
- In effective-field-theory we introduce new operators that at electro-weak scale are compatible with $SU(2)_L \times U(1)$.
- Left handed lepton doublets add right handed lepton singlets follow the group symmetry: $G_{LF} = SU(3)_L \times SU(3)_E$.
- LFV arises from breaking this group.
- We focus on three operators that have dominant contribution to NP:
  1. Purely left handed iterations: $(\overline{L}\gamma_\mu L)(\overline{L}\gamma^\mu L)$
  2. Mix term: $(\overline{R}\gamma_\mu R)(\overline{L}\gamma^\mu L)$
  3. Radiative operator: $g'(\overline{L}H\sigma_{\mu\nu}R)B^{\mu\nu}$

Reweighting MC samples

Reconstruction:



Offline:

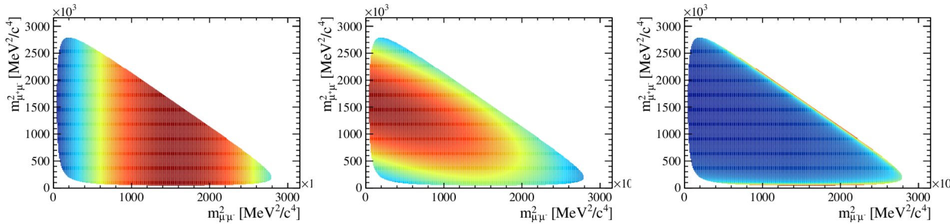


$$\epsilon_{gen\&rec} = C\epsilon_{gen\&rec}^{LHCbMC} \sum \rho^{model}(m_{12}, m_{23}) \quad (1)$$

- Simulated signal events with PHSP
- Take into account reconstruction and selection.
- Reweight accordingly to a given distribution.

Reweighting MC samples

$(\overline{L}\gamma_\mu L)(\overline{L}\gamma^\mu L)$ $(\overline{R}\gamma_\mu R)(\overline{L}\gamma^\mu L)$ $g'(\overline{L}H\sigma_{\mu\nu}R)B^{\mu\nu}$

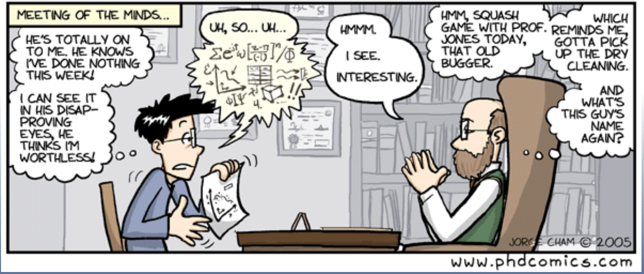


$$\epsilon_{gen\&rec} = C\epsilon_{gen\&rec}^{LHCbMC} \sum \rho^{model}(m_{12}, m_{23}) \qquad (1)$$

- Simulated signal events with PHSP
- Take into account reconstruction and selection.
- Reweight accordingly to a given distribution.

Conclusions

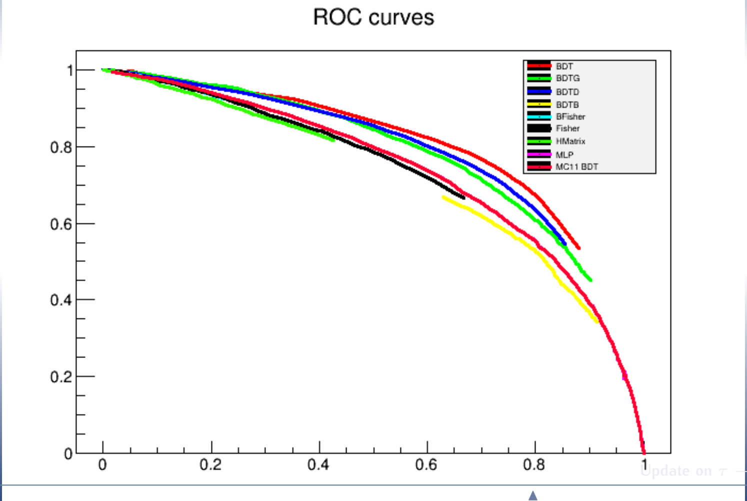
1. Analysis is well underway.
2. More efficient use of computing resources and increased MC statistics helps at all ends
3. Hope to improve the MVA/binning.



BACKUP

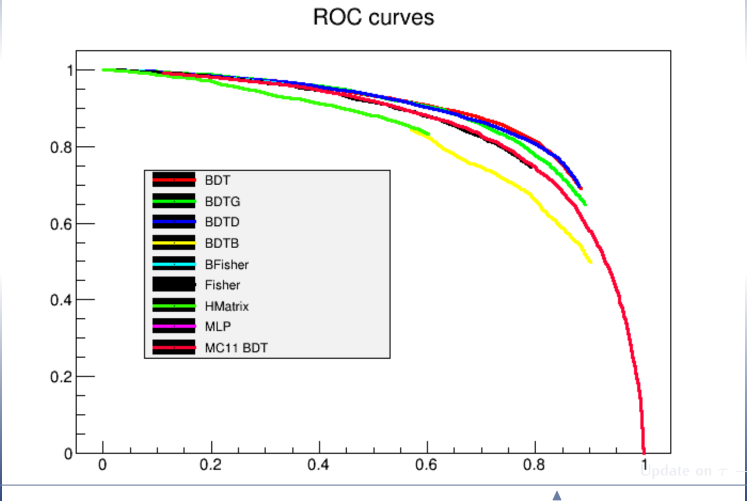
$B \to \tau$

We really suck in selecting this channel.



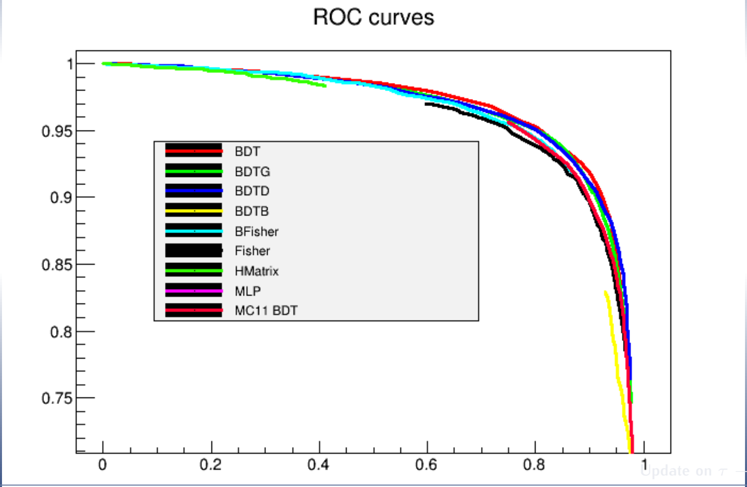
$B \to D_s \to \tau$

On the biggest contributing channel we are quite optimal.



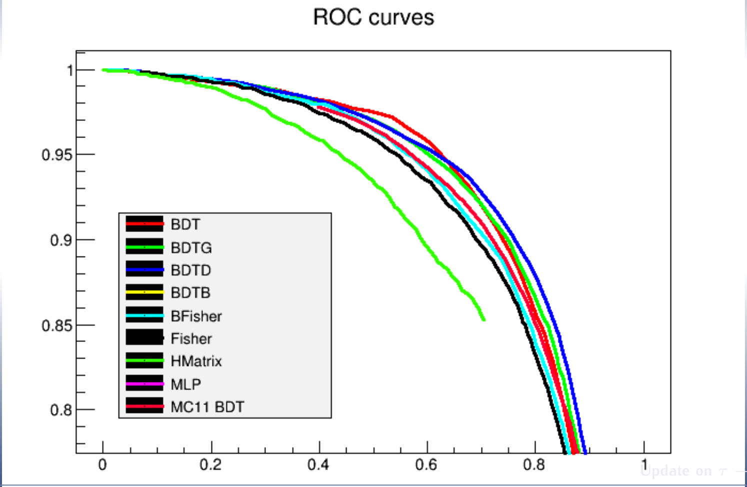
$D_s \to \tau$

On the biggest contributing channel we are quite optimal.



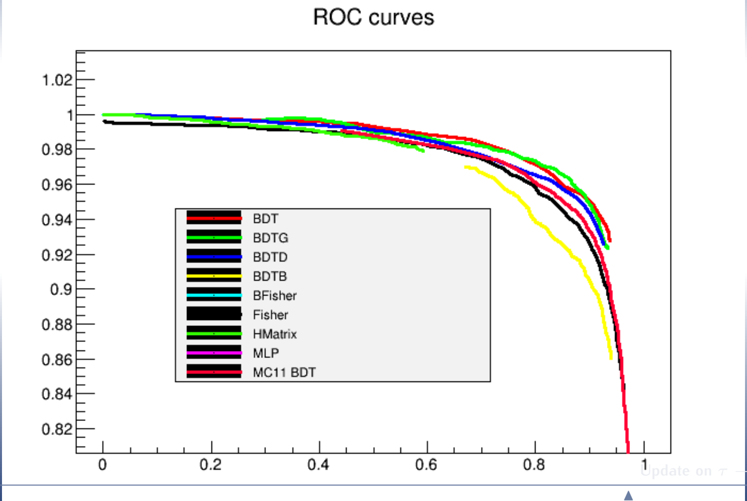
$B \to D^+ \to \tau$

On the biggest contributing channel we are quite optimal.



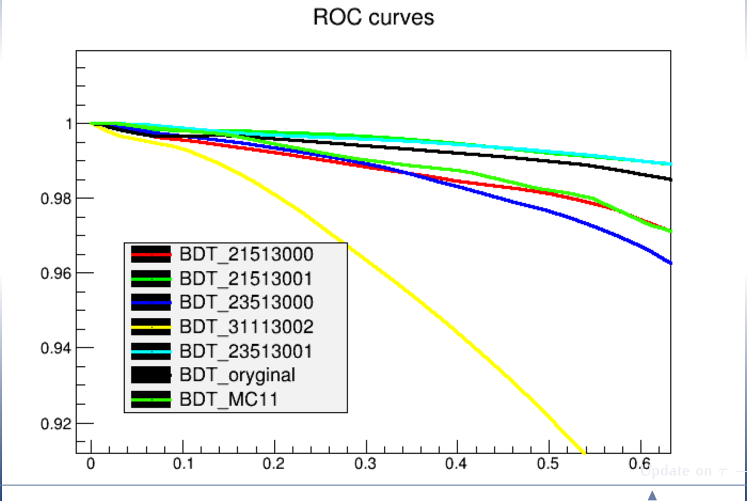
$D^+ \to \tau$

On the biggest contributing channel we are quite optimal.



Comparison on mix sample

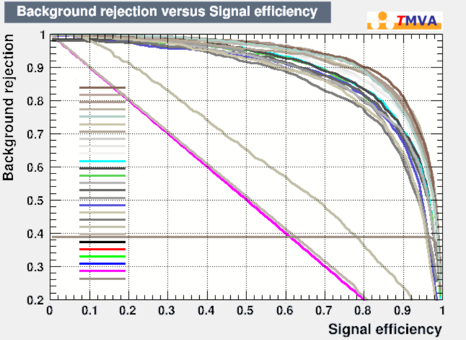
On the biggest contributing channel we are quite optimal.

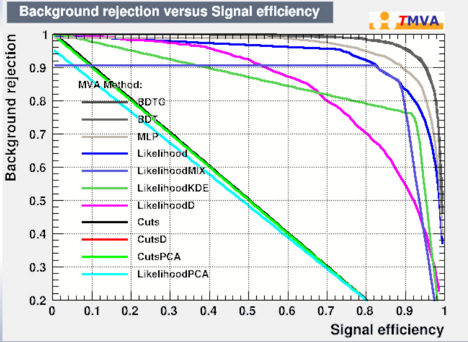


Conclusions on TMVA

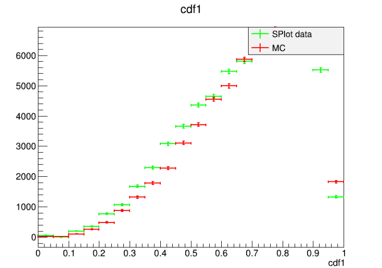
- Each of the signal components is enormously larger than MVA trained on mix.
- Method looks very promising if we can find a nice blending method(work for next week).
- Mayby discusion on TMVA/MatrixNet/Neurobayes is next to leading order effect compared to this method?

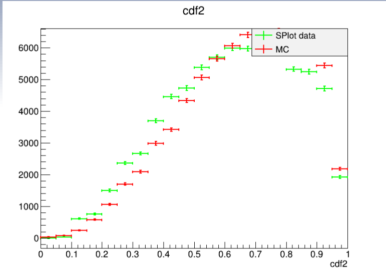
Comparison on mix sample

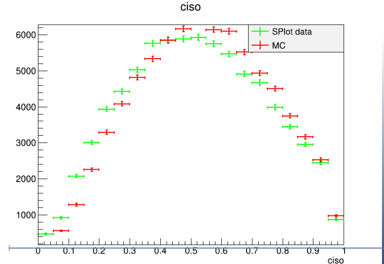


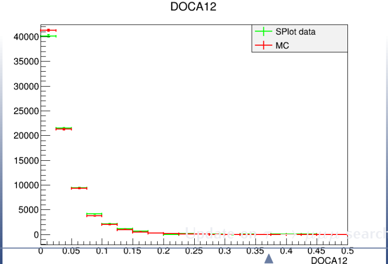


D_s correction

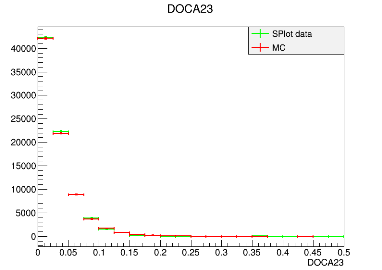


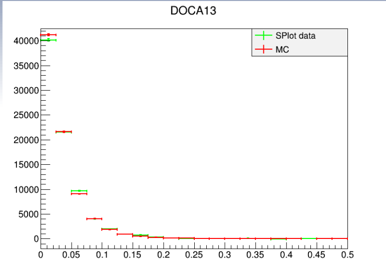


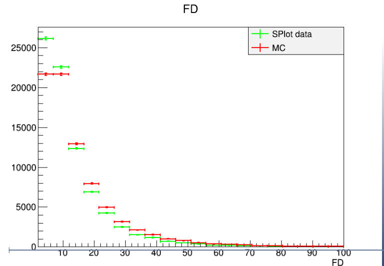


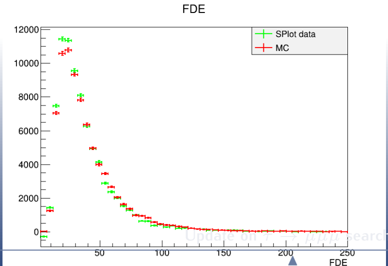


D_s correction

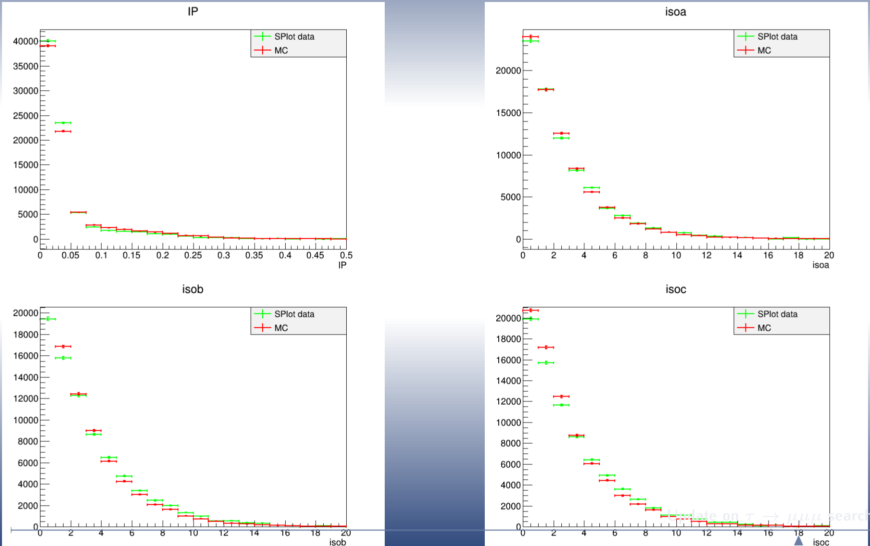




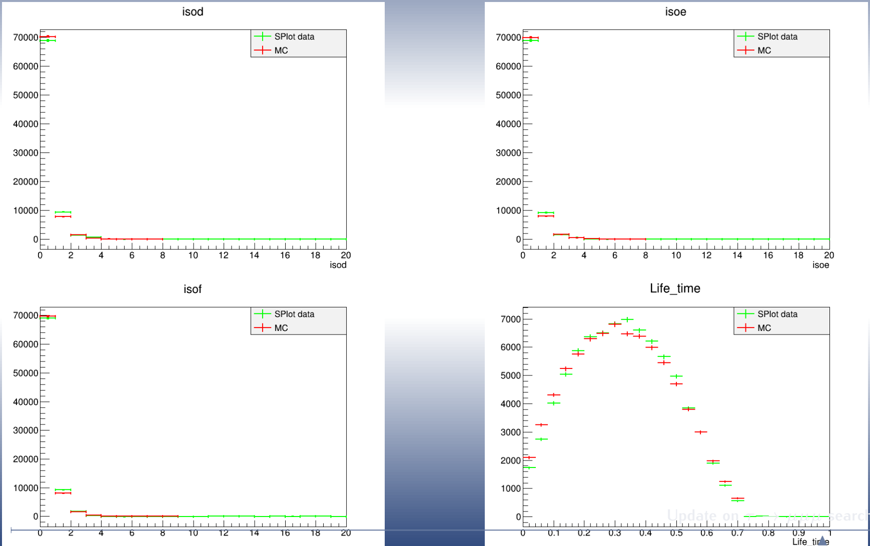




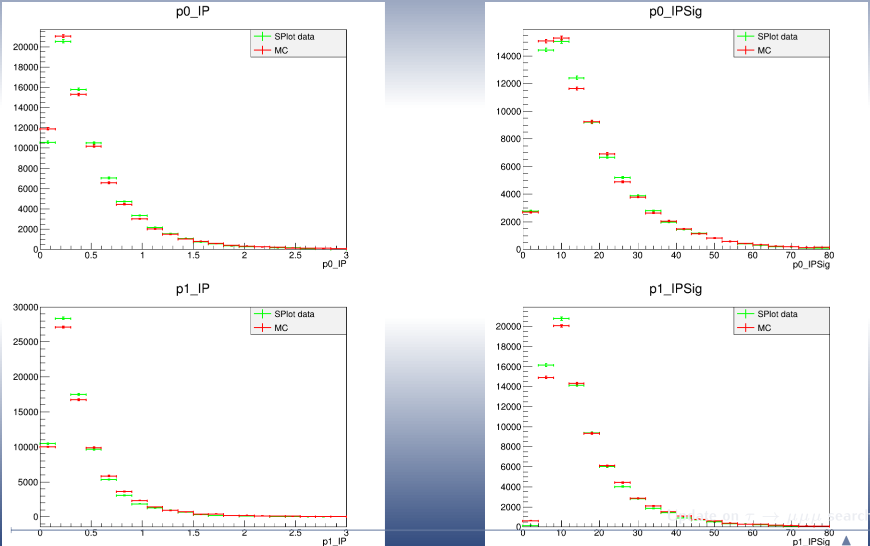
D_s correction



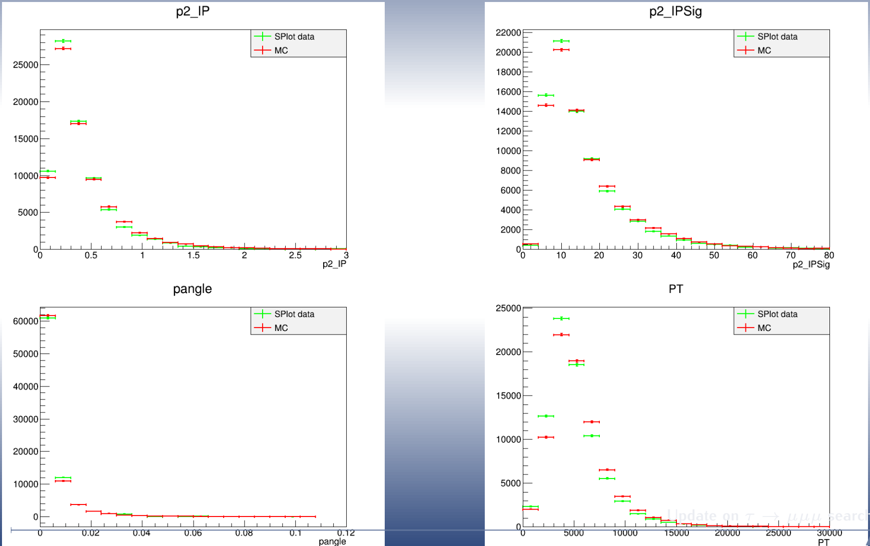
D_s correction



D_s correction



D_s correction



D_s correction