

VBF $H \rightarrow b\bar{b}$ weekly

Elisa Guido

23 January 2014

General random news (I)

- ❖ high stat. mc12 signal samples (VBF and gg filtered) AVAILABLE :-)
- ❖ high stat. mc11 signal samples (VBF and gg filtered) submitted... priorities could be an issue now since data challenge has started: trying to check with the person in charge
http://panda.cern.ch:25880/server/pandamon/query/?mode=taskquery&qTaskName=mc11_7TeV%25181768%VBFH%25&qsubmit=QuerySubmit
http://panda.cern.ch:25880/server/pandamon/query/?mode=taskquery&qTaskName=mc11_7TeV%25181769%ggH%25&qsubmit=QuerySubmit
- ❖ ntuples with the new jets' ordering (eta-based) are now available for 2011 and 2012
/mnt/project_mnt/atlas/atlas_gen_fs/atlas/data/Hbb/2012/Final_140120/
/mnt/project_mnt/atlas/atlas_gen_fs/atlas/data/Hbb/2011/Final_140120/
- ❖ **internal note** template available here (temporarily, going to ask for space on svn):
<http://www.ge.infn.it/~guido/Atlas/Hbb/internalNote/>
- ❖ Run2 L1 items study has been shown at the Jet Trigger meeting —> work on-going
<https://indico.cern.ch/conferenceDisplay.py?confId=295877>

General random news (II)

❖ talks and papers to have a look at:

- <http://cds.cern.ch/record/1643500> (ATLAS q/g tagger on 7TeV data)
- <https://indico.cern.ch/conferenceDisplay.py?confId=296010> (talk by Jia Jian Teoh on gluon-splitting identification for boosted topologies, study just started)
- <https://indico.cern.ch/getFile.py/access?contribId=0&resId=0&materialId=2&confId=293325> (talk on FTK triggers by John Alison, starting from page 24 VBF Hbb triggers. Pag.27 adding muon triggers, pag.28 adding electron triggers. L1 eff. studies without selection. A senior student will spend a month working on this —> we need to coordinate efforts)

Change in the jets labeling

- ❖ What we did:

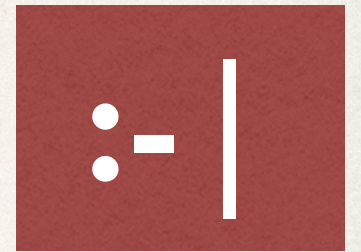
- ❖ the 4 jets are ordered on the basis of their MV1 variable: the 2 most b-like being the b's, the 2 less b-like being the VBF's
- ❖ in the case a jet is out of the tracker acceptance (i.e. no MV1 associated), it is automatically a VBF jet candidate. If more than one, they are ordered with eta, i.e. the more forward are VBF's
- ❖ **Problems:** apply the b-jet calibrations; biasing anti-b tagged samples (data control sample going to be used to determine the Z fraction, and the m_{bb} shape)

- ❖ What we do now:

- ❖ the 4 jets are ordered on the basis of their eta: the 2 most central being the b's, the 2 more forward being the VBF's
- ❖ the b's are matched to the L2 and EF objects
- ❖ We have in any case to live with this choice. Can we gain something, at the same time?

Do things look better now? (I)

$S/\sqrt{S+N}$ [here done for 2012]



- cuts for data only
- cuts for MC only

10X more signal!

wrt previous analysis with the same cuts, but different jets' labeling



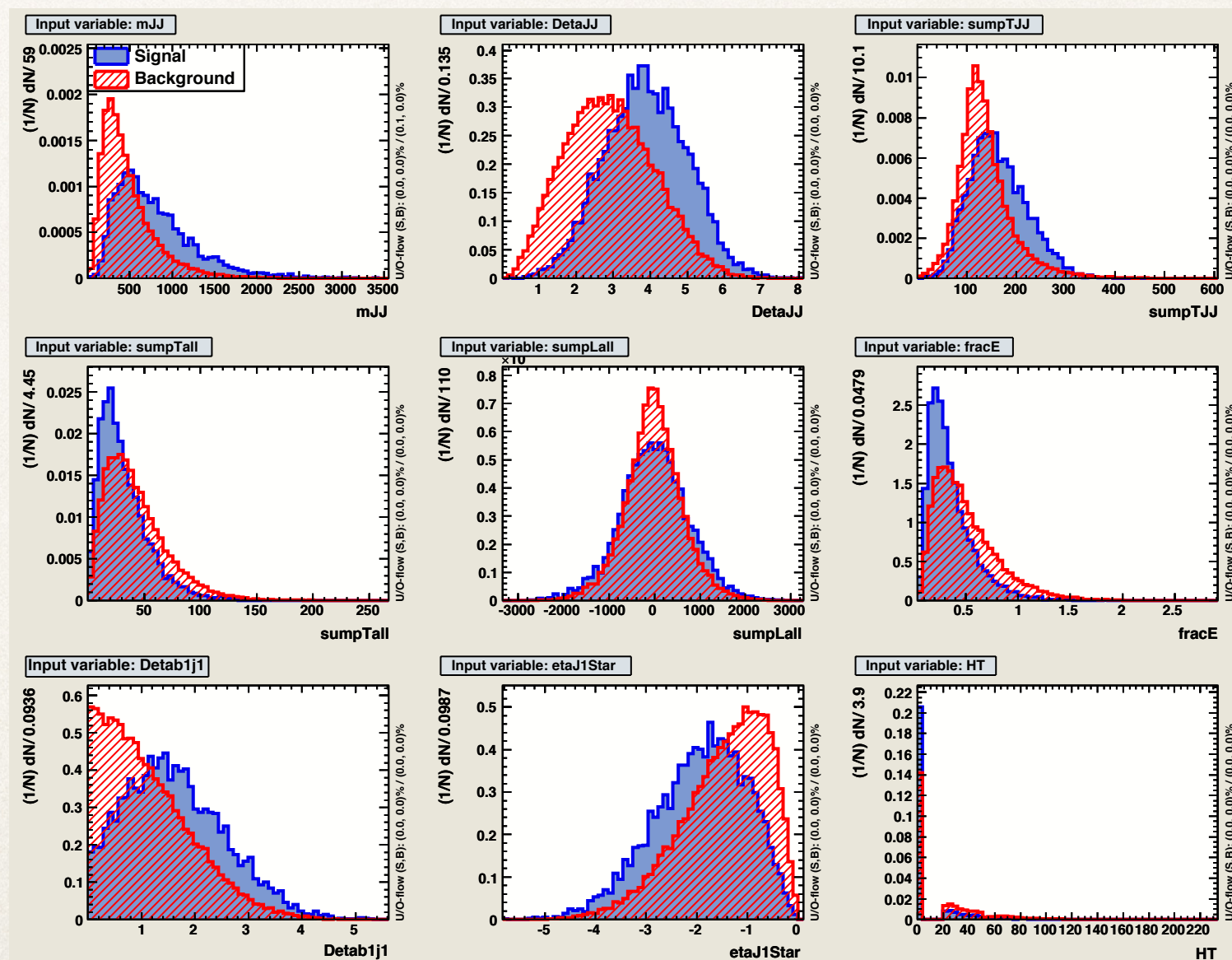
2012	N _{bkg} (20fb ⁻¹)	cut eff	tot eff		N _{sig} (20fb ⁻¹)	cut eff	tot eff	S/sqrt(S+N)	(new-old)/old
slimmed/generated	3,51E+07			1498880	18.483				
GRL	3,29E+07	0,94	0,937						
PV	3,29E+07	1,00	0,937	1498472	18.478	1,00	1,000		
LAr	3,28E+07	1,00	0,935	1498472	18.478	1,00	1,000		
trigger	3,28E+07	1,00	0,935	87769	1.082	0,06	0,059		
4J50 (p _T >50,η<4.5,IJVFI>0.75)	5,58E+06	0,17	0,159	23064	284	0,26	0,015	0,120	
classified events				23064	284	1,00	0,015	0,120	
truth matched events				19362	239	0,84	0,013	0,101	
eta sorted events				14878	183	0,77	0,010	0,078	
b's matched to L2&EF	1,06E+06	0,19	0,030	12296	152	0,83	0,008	0,147	
b's central	1,06E+06	1,00	0,030	12268	151	1,00	0,008	0,147	
b's tagged offline (80%)	7,77E+05	0,73	0,022	11886	147	0,97	0,008	0,166	
0.8<ΔR _{bb} <2.2	3,02E+05	0,39	0,009	9840	121	0,83	0,007	0,221	
VBF cut [m _{JJ} >500, Δη _{JJ} >3]	9,27E+04	0,31	0,003	6529	81	0,66	0,004	0,264	
SR [100<m _{bb} <140 GeV]	2,96E+04	0,32	0,001	5320	66	0,81	0,004	0,381	-0,03

Remember that for data the slimming already takes in the trigger requests → trigger efficiency is 1 in the table

Do things look better now? (II)

MVA [here done for 2012]

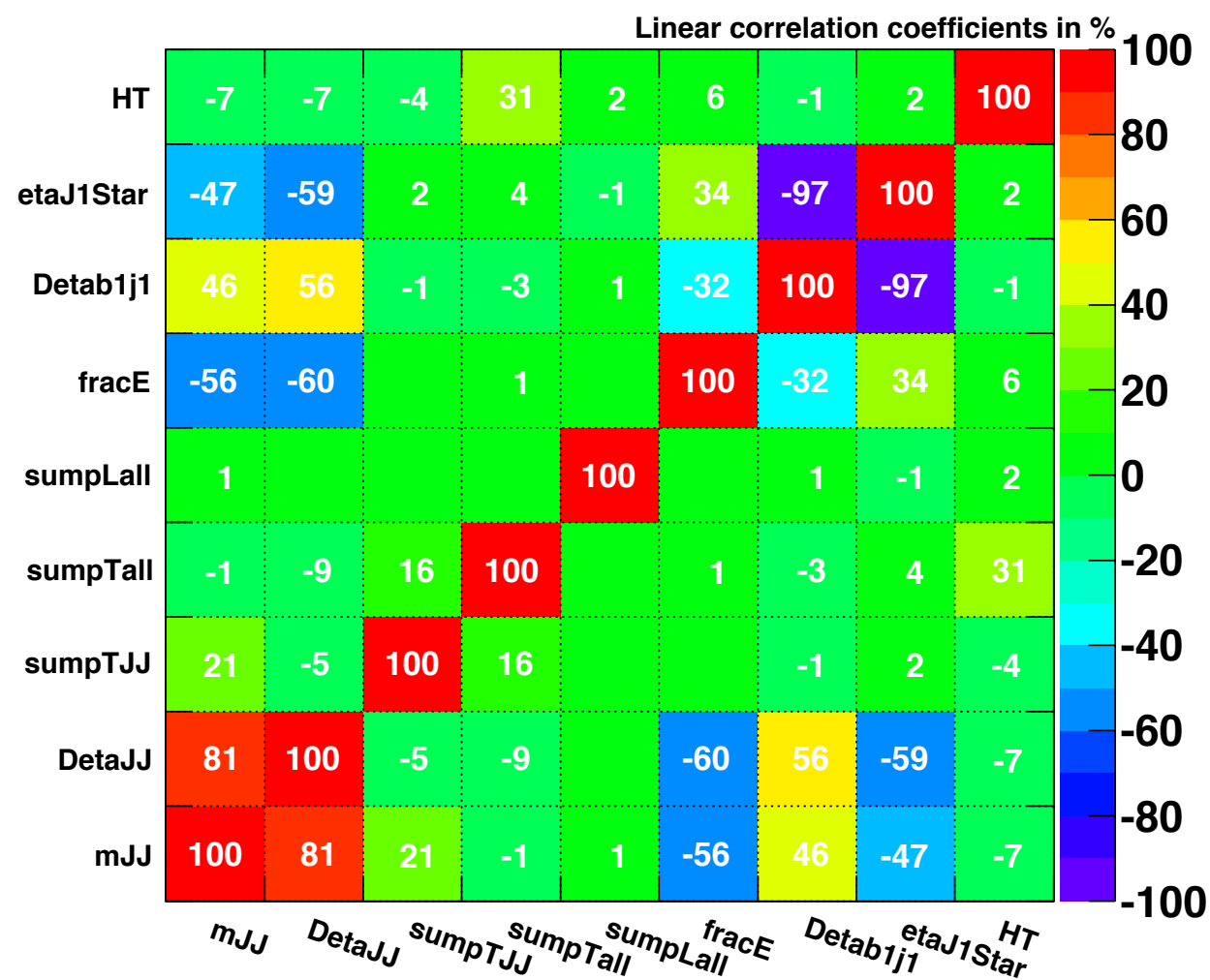
- ❖ Of course, changing the labeling of the jets results in changing some of the input variables of the TMVA - mostly the kinematical and angular ones for the VBF jets
- ❖ At present, I simply re-used the previous set of input variables —> understanding and optimization here



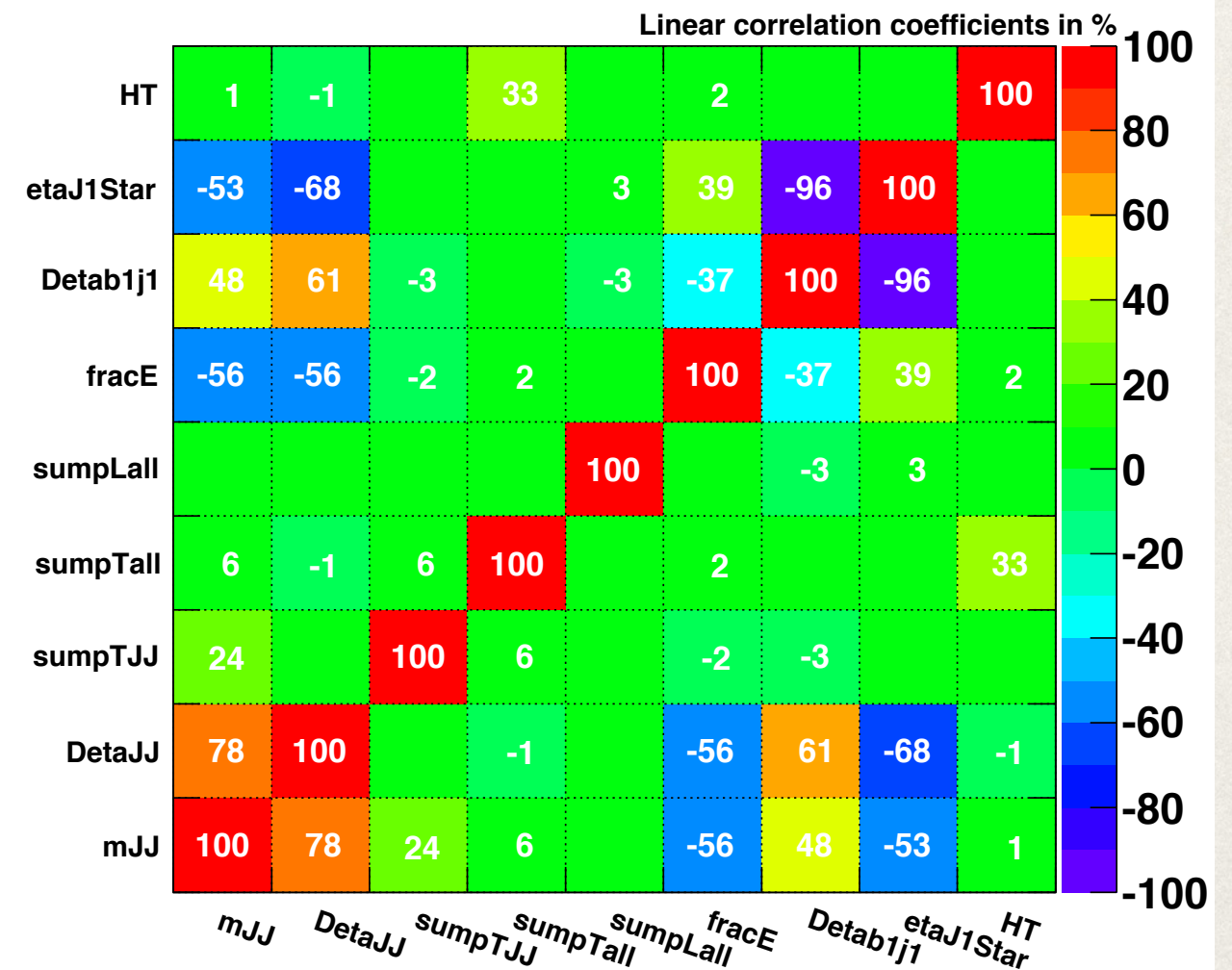
Rank : Variable : Separation

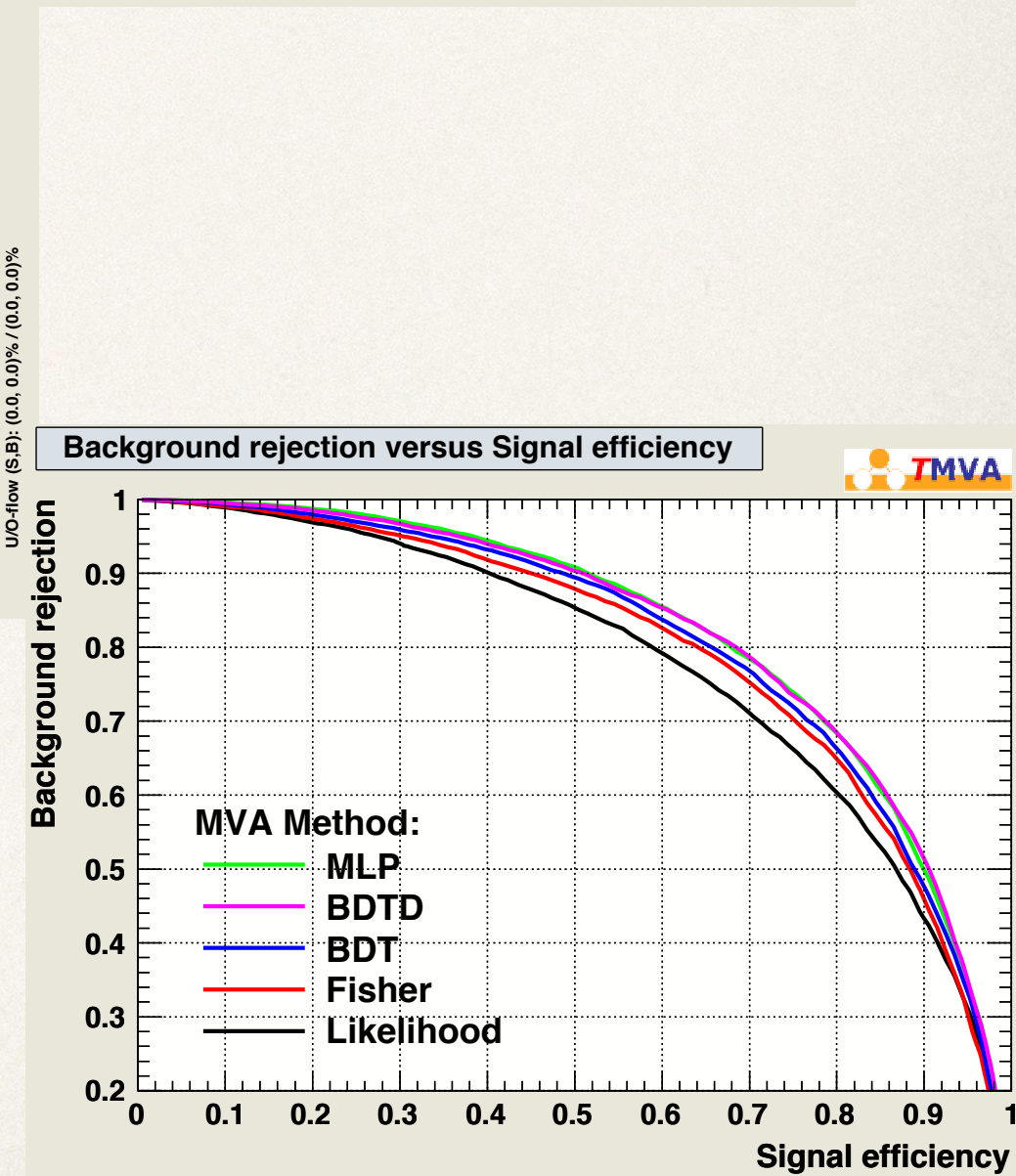
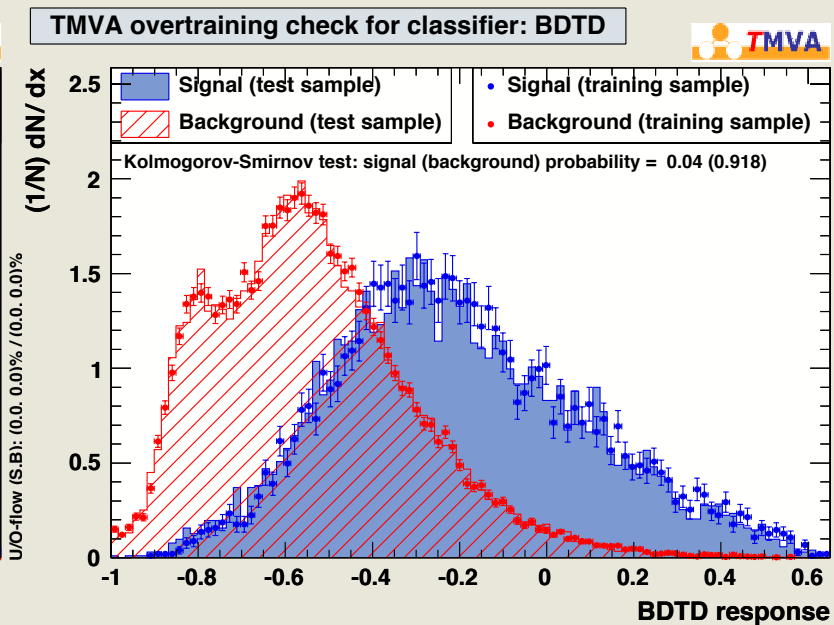
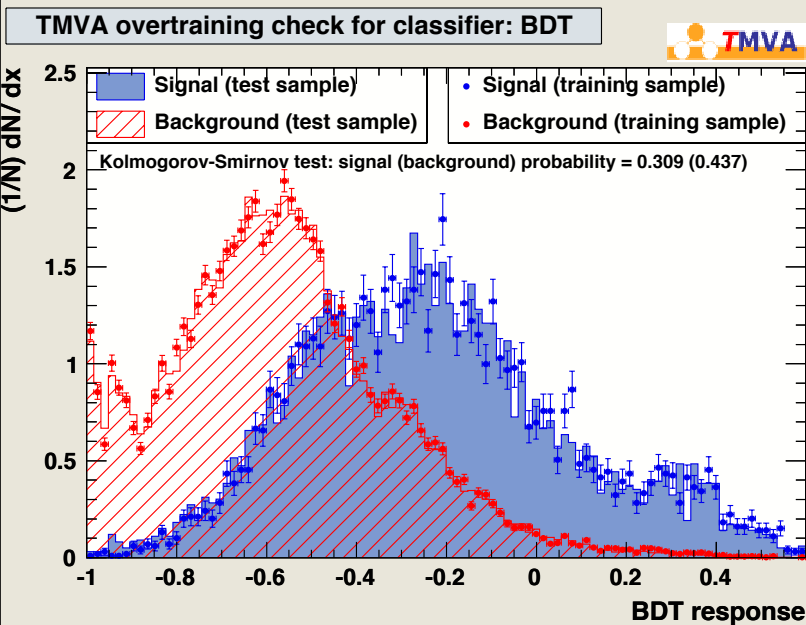
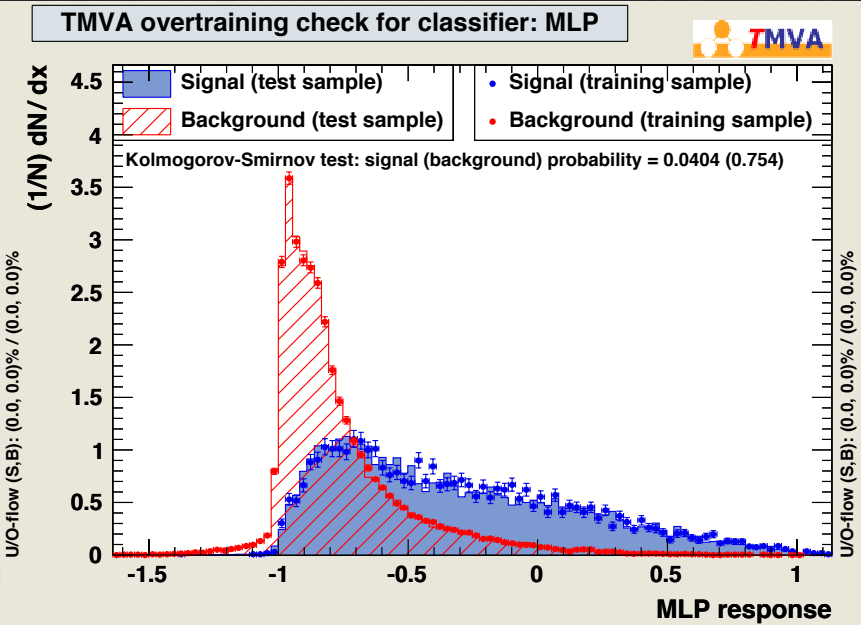
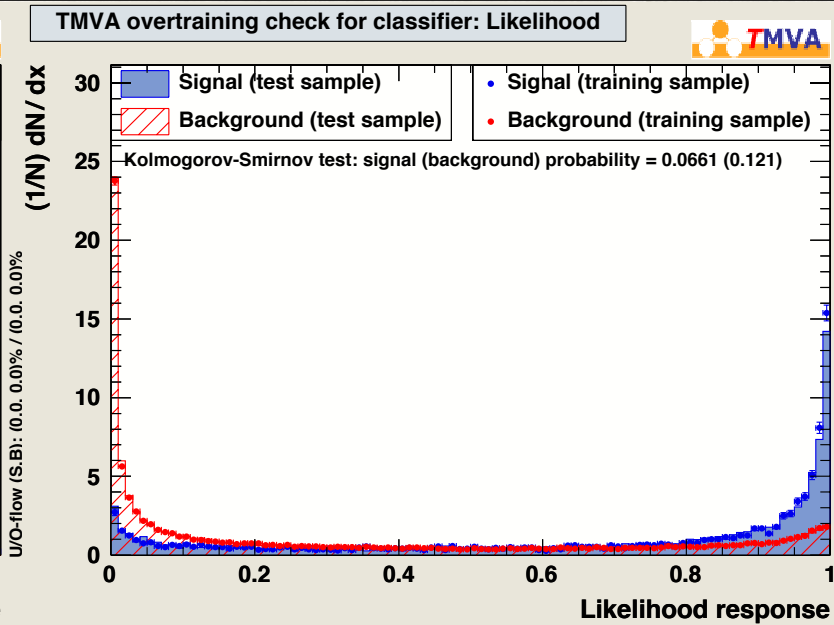
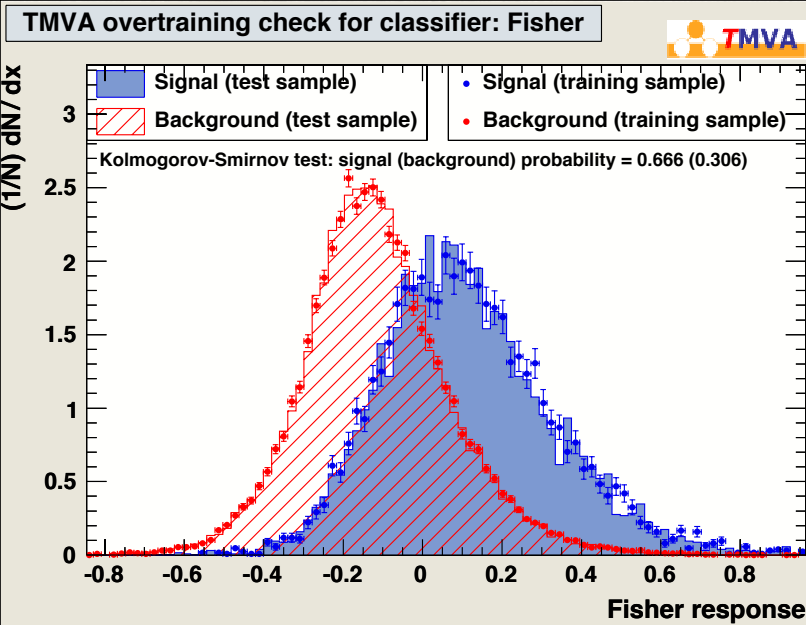
 1 : mJJ : 1.730e-01
 2 : DetaJJ : 1.560e-01
 3 : Detab1j1 : 8.301e-02
 4 : fracE : 7.992e-02
 5 : sumpTJJ : 7.960e-02
 6 : HT : 7.918e-02
 7 : etaJ1Star : 7.621e-02
 8 : sumpTall : 4.886e-02
 9 : sumpLall : 1.833e-02

Correlation Matrix (signal)

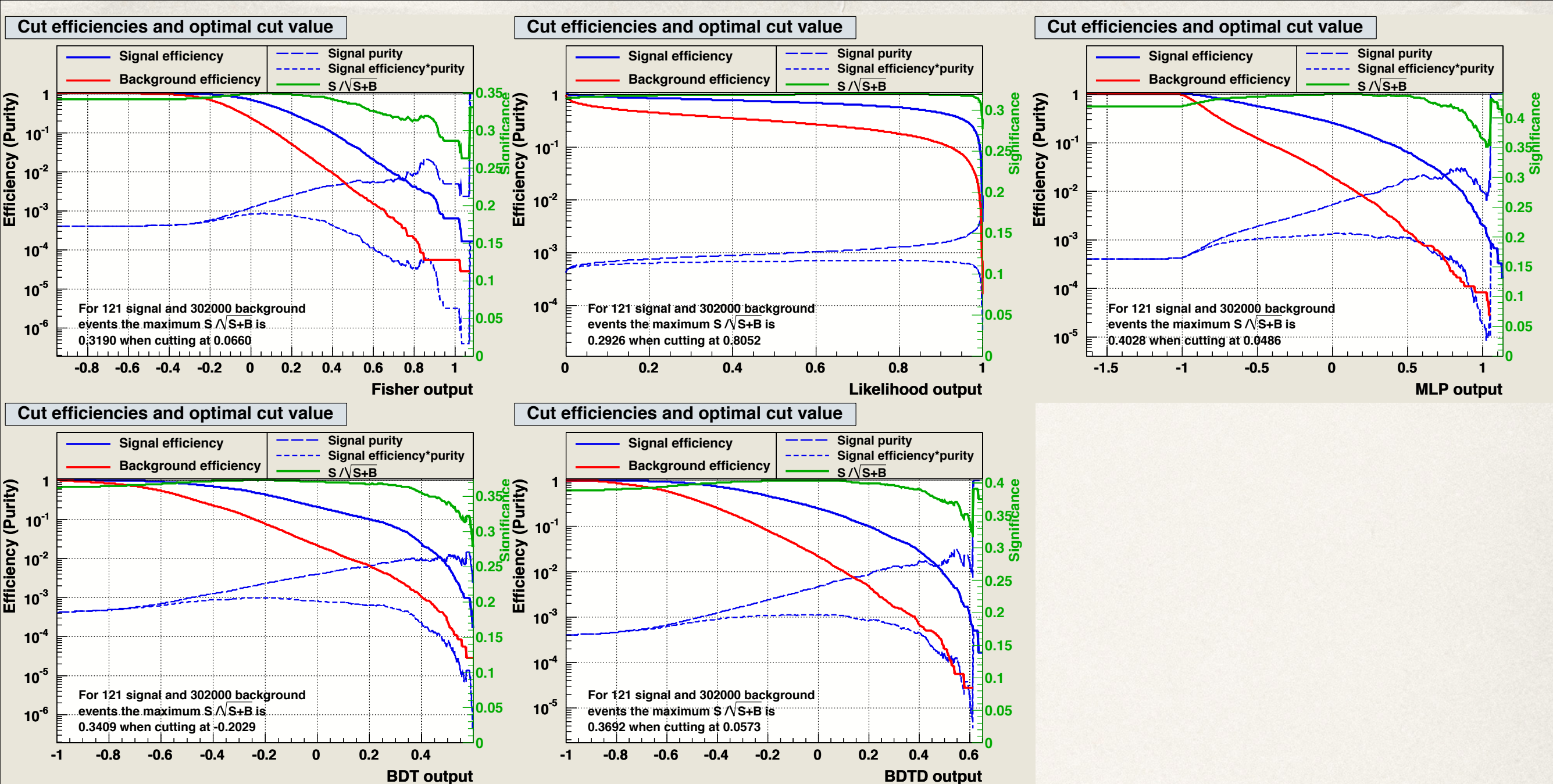


Correlation Matrix (background)





Signal training entries: 6202
Signal testing entries: 6203
Background training entries: 36751
Background testing entries: 36752



Classifier	(#signal, #backgr.)	Optimal-cut	$S/\sqrt{S+B}$	NSig	NBkg	EffSig	EffBkg
Likelihood:	(121, 302000)	0.8052	0.292594	66.86894	52163.04	0.5526	0.1727
Fisher:	(121, 302000)	0.0660	0.318994	66.79091	43773.24	0.552	0.1449
MLP:	(121, 302000)	0.0486	0.402824	27.50443	4634.523	0.2273	0.01535
BDT:	(121, 302000)	-0.2029	0.340871	51.96582	23189.05	0.4295	0.07678
BDTD:	(121, 302000)	0.0573	0.369165	23.34951	3977.144	0.193	0.01317

improvement factor wrt cut-based (page 5) ~1.5

Possible improvements (I)

$|JVF| > 0.75 \longrightarrow \text{no JVF cut}$



- Now that a high stat. signal sample is available: re-visit all the cuts in the event selection in order to understand if there are possible improvements in terms of $S/\sqrt{S+N}$

wrt page 4



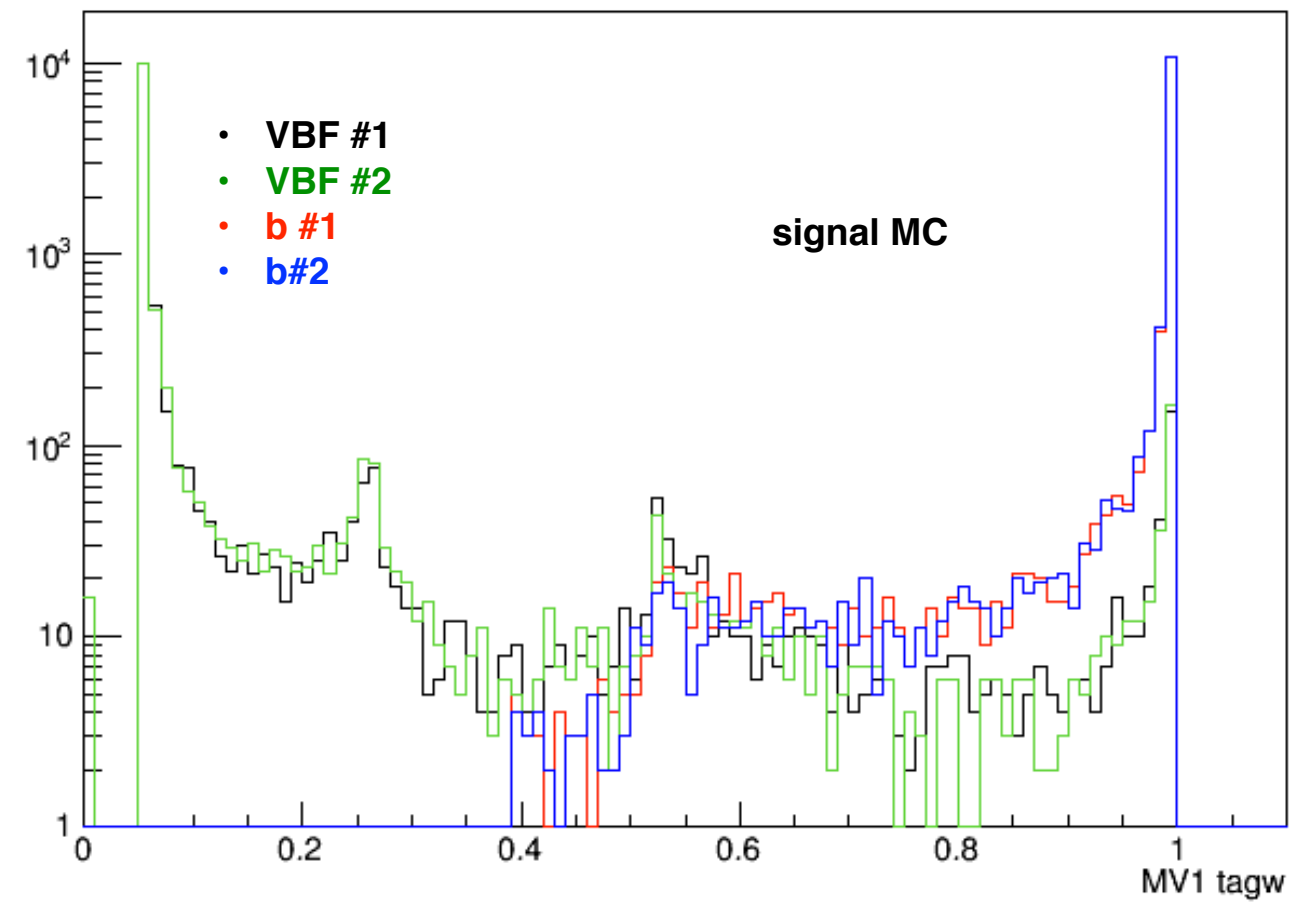
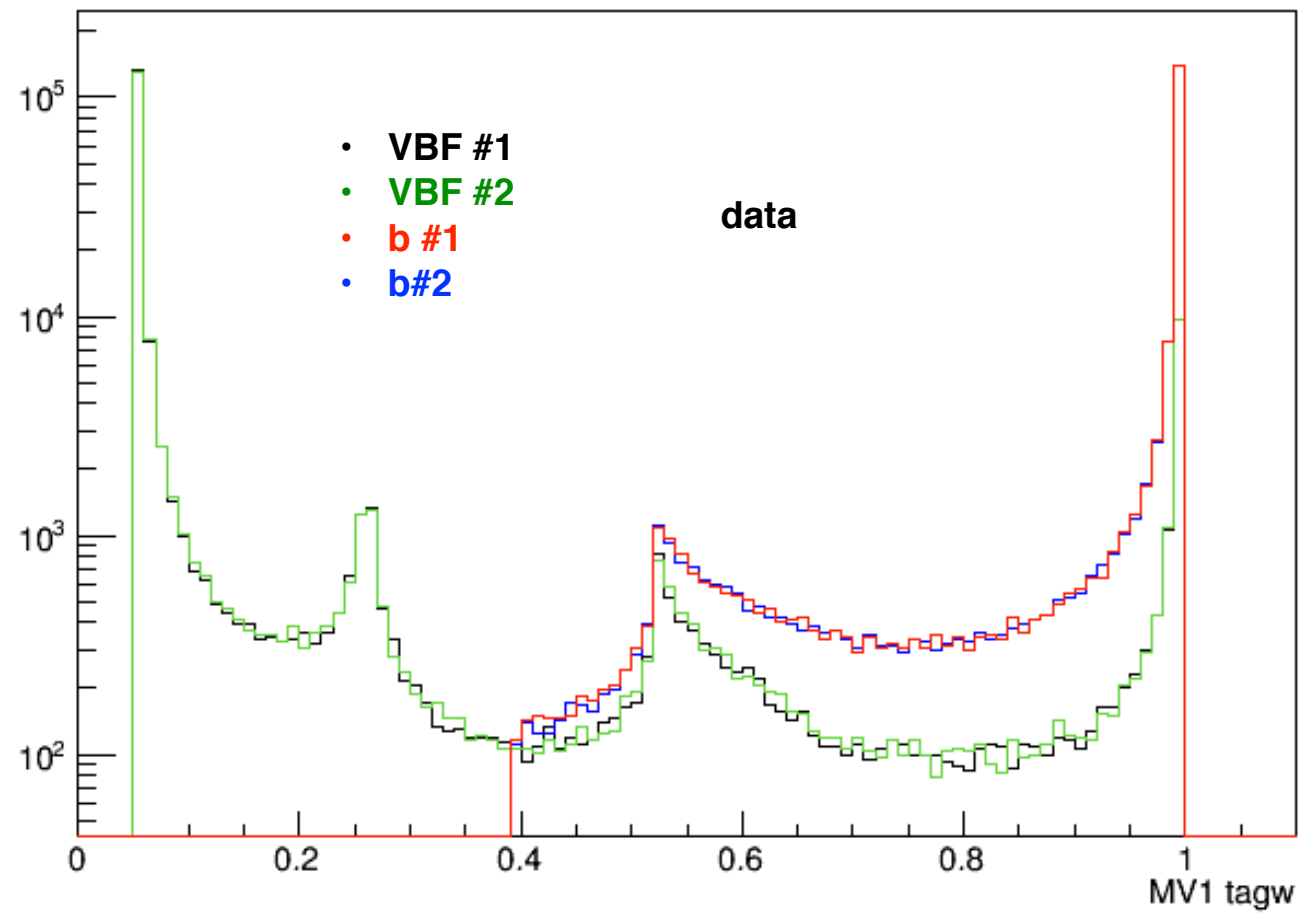
default	N_{bkg} (20fb ⁻¹)	cut eff	tot eff		N_{sig} (20fb ⁻¹)	cut eff	tot eff	$S/\sqrt{S+N}$	(new-old)/old
slimmed/generated	3,51E+07			1498880	18.483				
GRL	3,29E+07	0,94	0,937						
PV	3,29E+07	1,00	0,937	1498472	18.478	1,00	1,000		
LAr	3,28E+07	1,00	0,935	1498472	18.478	1,00	1,000		
trigger	3,28E+07	1,00	0,935	87769	1.082	0,06	0,059		
4J50 ($p_T > 50, \eta < 4.5$)	6,73E+06	0,21	0,192	28704	354	0,33	0,019	0,136	
classified events				28704	354	1,00	0,019	0,136	
truth matched events				24349	300	0,85	0,016	0,116	
eta sorted events				18994	234	0,78	0,013	0,090	
b's matched to L2&EF	1,29E+06	0,19	0,037	15687	193	0,83	0,010	0,171	
b's central	1,28E+06	1,00	0,037	15653	193	1,00	0,010	0,170	
b's tagged offline (80%)	9,40E+05	0,73	0,027	15082	186	0,96	0,010	0,192	
$0.8 < \Delta R_{bb} < 2.2$	3,65E+05	0,39	0,010	12405	153	0,82	0,008	0,253	
VBF cut [$m_{JJ} > 500, \Delta\eta_{JJ} > 3$]	1,13E+05	0,31	0,003	8258	102	0,67	0,006	0,303	
SR [$100 < m_{bb} < 140 \text{ GeV}$]	3,63E+04	0,32	0,001	6708	83	0,81	0,004	0,433	0,14

- Checking other config. (JVF for b's only?) and all other cuts

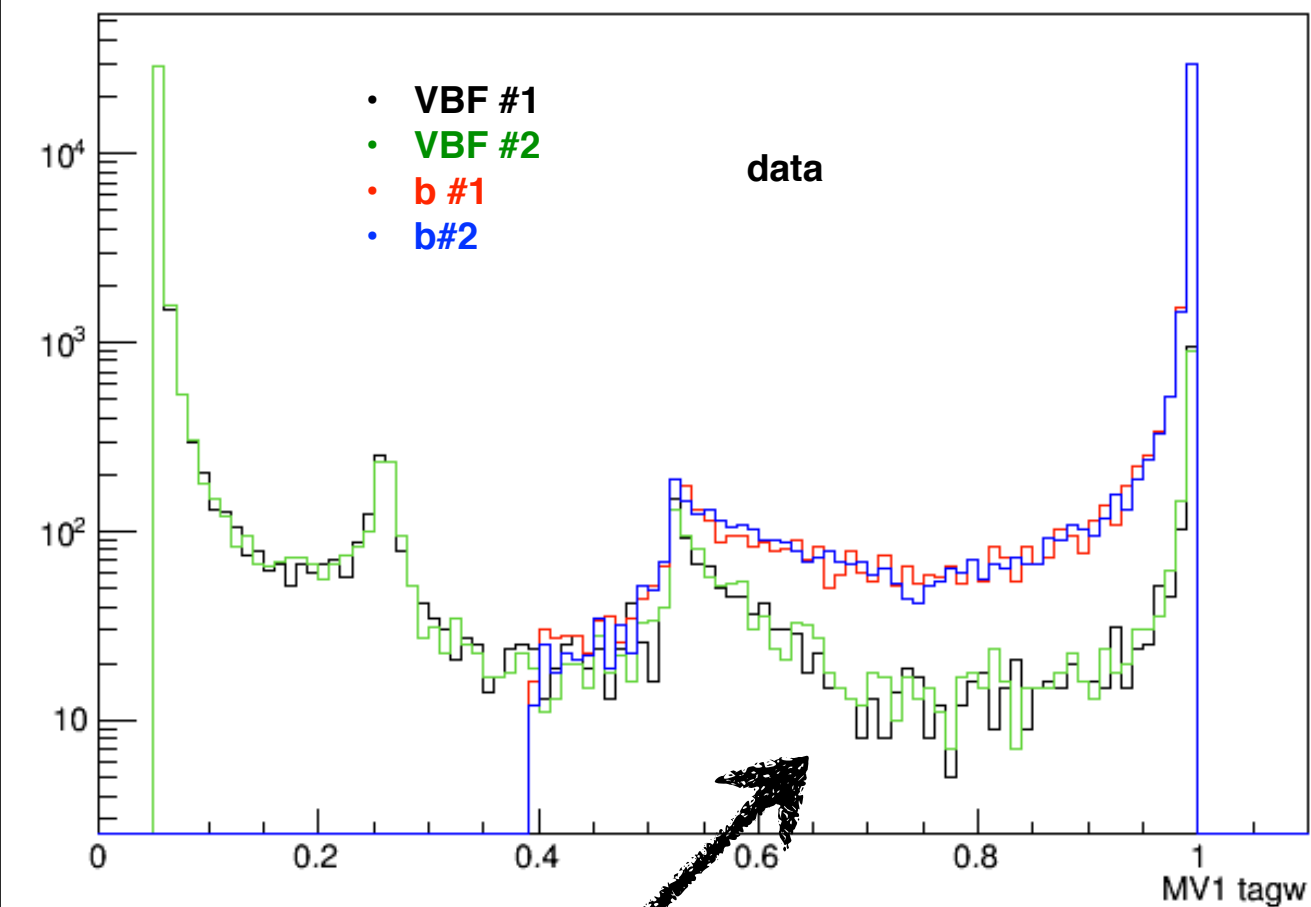


~10% improvement wrt old approach... not negligible!

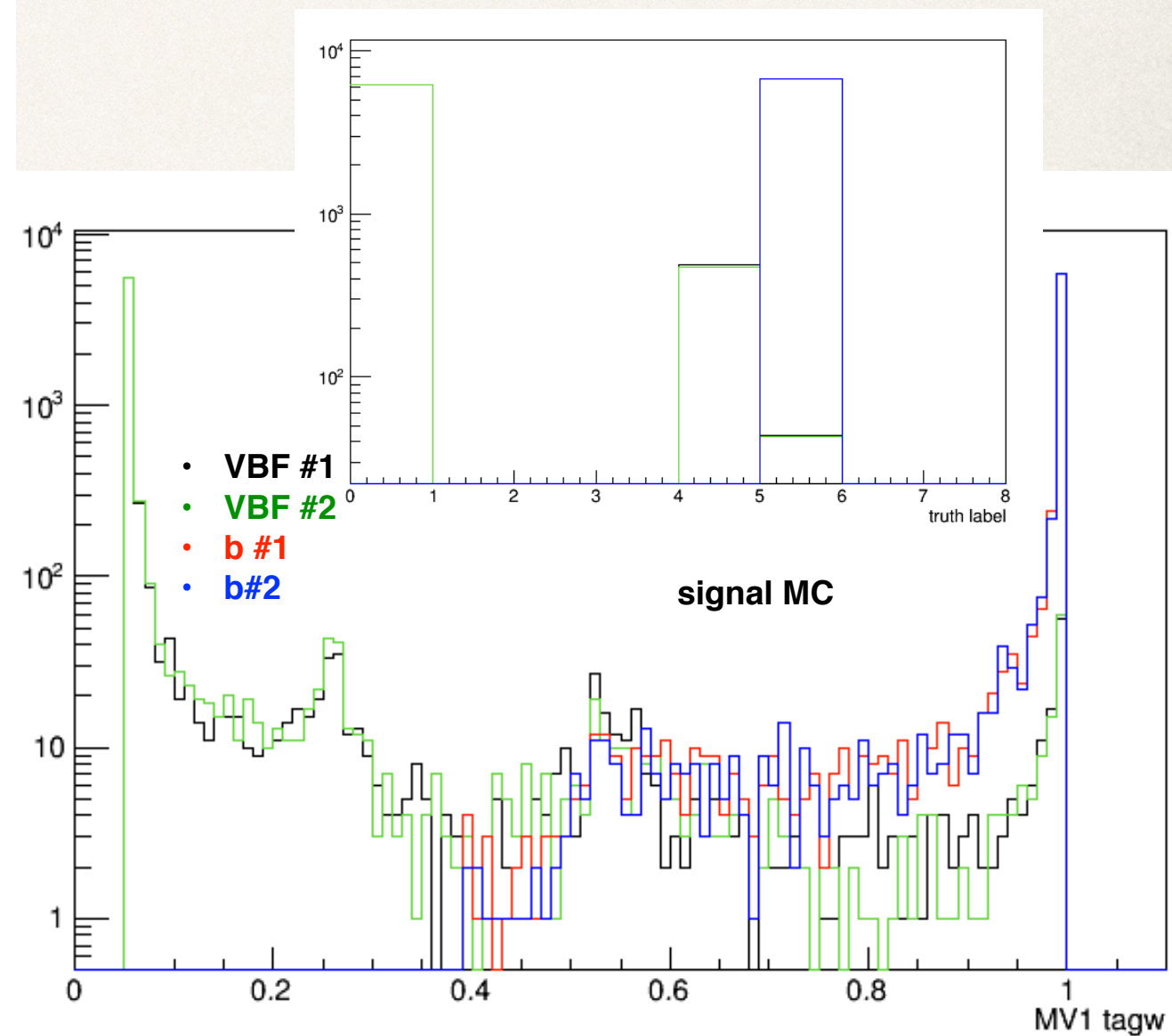
MV1 tagw



MV1 tagw at the end of the selection



anti b-tagging VBF jets?



Possible improvements (II)

tighter b-tagging requirements

% eff. b #1 - % eff. b #2

at the b-tag stage
(before VBF cuts or MVA)
in the SR

at the end of the selection
in the SR

2012 b-tag	N _{bkg} (20fb ⁻¹)		N _{sig} (20fb ⁻¹)	S/sqrt(S+N)	improv.
80-80	1,77E+05	12197	150	0,357	
70-80	1,61E+05	11748	145	0,361	0,010
60-80	1,44E+05	11011	136	0,357	-0,000
80-70	1,61E+05	11792	145	0,362	0,013
80-60	1,44E+05	11037	136	0,358	0,001
70-60	1,33E+05	10630	131	0,359	0,005
70-70	1,48E+05	11362	140	0,365	0,020
60-70	1,33E+05	10650	131	0,360	0,007
60-60	1,20E+05	9966	123	0,354	-0,008
80-80	3,63E+04	6708	83	0,433	
70-80	3,35E+04	6460	80	0,435	0,003
60-80	3,03E+04	6026	74	0,427	-0,016
80-70	3,35E+04	6477	80	0,436	0,005
80-60	3,03E+04	6045	75	0,428	-0,013
70-60	2,82E+04	5824	72	0,427	-0,015
70-70	3,00E+04	6106	75	0,434	0,002
60-70	2,82E+04	5819	72	0,427	-0,015
60-60	2,56E+04	5437	67	0,418	-0,035

wrt page 10

To-do-list

- ❖ Check the event selection cut-by-cut, looking for configurations maximizing $S/\sqrt{S+N}$
 - ❖ re-evaluate the gain by moving the pT thresholds
 - ❖ anti b-tagging for VBF jets?
 - ❖ JVF for b jets?
- ❖ Work on the MVA: define the input variables, define the best discriminator, and the optimized cut on the output variable
- ❖ Write...