Search for CP violation in top quark pair at $\sqrt{s} = 13\,\text{TeV}$ in CMS

Chen-Yu Chuang*
Department of physics,
National Taiwan University, Taipei, Taiwan

Advisor
Prof. Kai-Feng Chen
Department of physics,
National Taiwan University, Taipei, Taiwan

October 3, 2020

*jeff0118252718@gmail.com

Abstract

This is a template abstract for my LaTeX notes.

Contents

List of Figures

List of Tables

I Introduction

Particle physics has been started from

I.1 Ingredients of the materials

The standard language of

II Experimental Apparatus

III Physical Objects Reconstruction and Selection

III.1 Vertex

III.2 Pileup Issue

Because of the high luminosity of pp collision in CMS, there are more than one pp collision occuring in every time when the the bunches cross one another. The multiple collision in one crossing event is called **pile-up(pileup)**. In the present, the LHC is operating at an instantaneous luminosity of $0.7 \times 10^{34} cm^{-2} s^{-1}$, and the proton bunches cross inside CMS every 50 ns, so there are more proton in one bunch and showing high pileup. Also, the high granularity and efficiency of CMS Tracker make them distinguish the many tracks in an event. The pileup image is shown below(Fig.1).

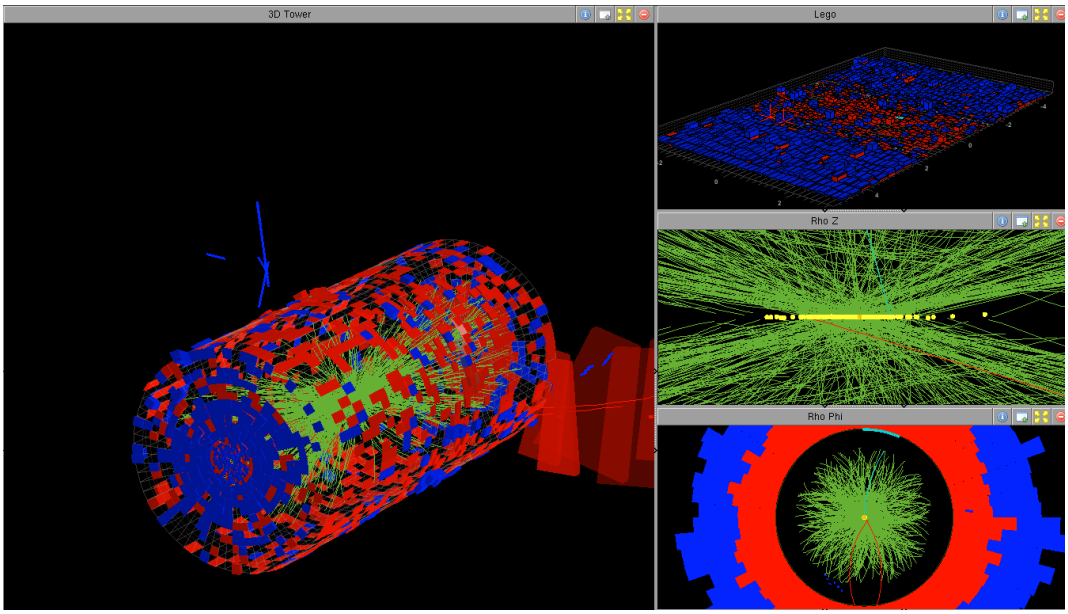


Figure 1: 78 reconstructed vertices in one crossing from high pileup run (3D, lego, on Pho-z plane, on Rho-Phi plane)

III.3 Lepton

The selected lepton in the analysis is required to obey the criteria of one passed *selected lepton* and zero *veto lepton* passed. The veto criteria means that there would be no lepton passing the veto criteria except the selected one. In other words, the veto criteria can filter the physical objects which are lepton-like but not really like after reconstructed from particle level to detector level. The selected criteria corresponds to tight lepton's criteria, and veto criteria follows loose lepton's criteria:

III.3.1 Muon

III.3.2 Electron

III.4 Jet

III.4.1 B-tagged jet

III.5 Trigger

IV Data and Simulation samples

IV.1 Data sample

IV.2 Simulation sample

IV.3 Correction on simulation sample

IV.3.1 Pile-up Reweighing

As mention in section.III.2, pileup is the Although Monte Carlo sample may roughly cover the pileup interactions, the final distribution is also sensitive to the subtle content of primary vertex. Furthermore, the distribution of reconstructed vertices can be differently affected by the offline event selection cut and some high level trigger between real data and Monte Carlo sample. To fix the descrepancy between data and simulation,

IV.3.2 Jet Energy correction, smearing and resolution

IV.3.3 Efficiency Scale Factor

IV.3.4 Weigh to Luminosity

Process sample	Cross Section (pb)	k-factor	Events Number	Generator
$t\bar{t} \to b\bar{b}jjl\nu$	XXX	1	CCCCCC	AAAAAA
cell7	cell8	XXXXXX	cell7	cell8

V Events Selection and Reconstruction

To calculate the asymmetry from top quark in $t\bar{t}$, there must be some selections to extract the $t\bar{t}$ sample from other background samples. Also, there would be several analysis strategies to recognize physical objects and pick them out. The chapter will do these discussion and make some comparison and organization.

In this analysis, doing reasearch by the channel of lepton + jets, which means 2 tops have different decay modes. One decays leptonically, and another decays hadronically(they will be called *hadronic top* and *leptonic top* in following context). It is expected that the hadronic top can be constructed with 1 b-tagged jet and 2 non b-tagged jets and the other top can be constructed imcompletely with 1 b-tagged jet, 1 lepton with missing neutrino 4-momentum by detector issue.

V.1 Trigger

In addition to L1 online trigger mentioned in III.5, there are also offline trigger here to filter specific events. It's called **High Level Trigger(HLT)**. It's a software trigger usually filter events with kinematics like energy(E), invariant mass(M), transverse momentum(p_T), etc. With final states including one selected isolated lepton in this analysis, if they are unprescaled during the entire data collecting, single lepton triggers are chose. For both 2016 data and MC, there are the HLT we use on SingleMuon and SingleElectron sample individually:

- **Single Muon**: *HLT IsoMu24* ***or*** *HLT IsoTkMu24*, which pick out events whose muon with reconstructed $p_T > 24$ GeV, or track reconstructed muons with reconstructed $p_T > 24$ GeV.
- **Single Electron**: *HLT Ele27 WPTight*, which pick out events whose electron with reconstructed $p_T > 27$ GeV.

V.2 Events Selection : Signal Region

For the **Signal Region(SR)**, the region with selection cut to extract signal in an analysis, there are the selection cuts below, with the physical objects' criteria menntioned in Chapter.III:

- 1 selected lepton which are a tight muon or a tight electron
- 0 lepton pass veto criteria which are loose muon and loose electron criterion
- $\geq$ 4 selected jets with passing medium jet criteria
- exact 2 btagged jets(deepCSV Medium criteria) in these selected jets
- each selected jets are isolated from the selected lepton with $\Delta R > 0.4$

The ΔR criteria here means the angle distribution between selected jets and selected lepton in $\phi - \eta$ phase space with $\Delta R > 0.4$. This would avoid the confusing cases that the lepton is really coming from jets or that the jets have some correlation with lepton in reconstruction process... etc.

$$\Delta R = \sqrt{(\eta_{jet} - \eta_{lep})^2 - (\phi_{jet} - \phi_{lep})^2} > 0.4 \tag{1}$$

Under SR selection to extract the semi-leptonic $t\bar{t}$(signal), and also following the high level trigger, they are shown in muon and electron channels to classify the selected sample in the following analysis.

V.3 Events Reconstruction

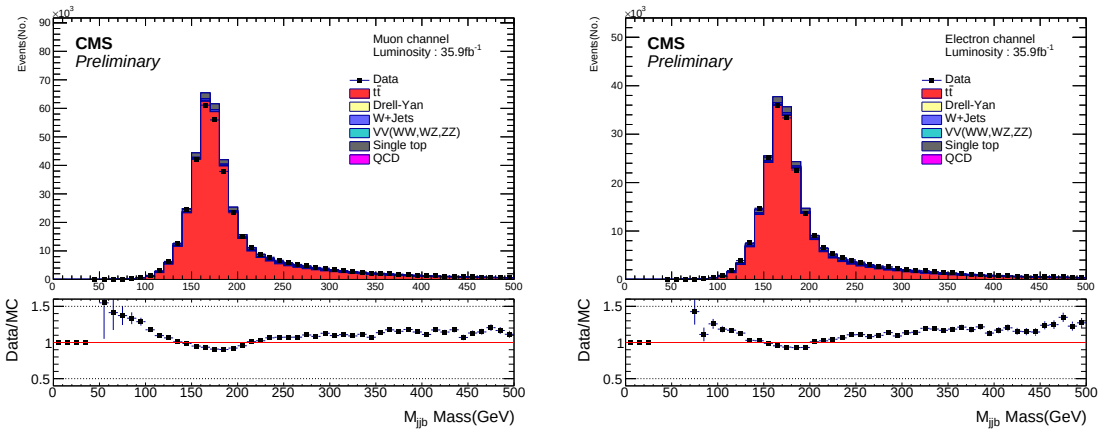
To reconstruct the semi-leptonic $t\bar{t}$ system, it is suggested that to reconstruct the hadronic top quark. It's an advantage that we can avoid dealing with missing 4-momentum from neutrino decays from leptonic top. And under the SR selection, if we can reconstruct the top which decays hadronically, the decay objects of leptonic decay top would be picked up simultaneously. This is a direct and common way to correctly identify selected candidates.

V.3.1 χ^2_{min} Method

There are multiple combinations of 1 b-tagged jet and 2 non b-tagged jet in an event. And how do we reconstruct hadronic top? In known and published analysis, based on the reco-level invariant mass of top quark itself and the intermediate particle in decay - W boson, they are used as constraints to invariant mass which is reconstructed from each combination. With the defined χ^2_{min} value, which shows below:

$$\chi^2 = (\frac{m_{jjb} - m_t}{\sigma_t})^2 + (\frac{m_{jj} - m_W}{\sigma_W})^2 \tag{2}$$

where m_t, m_W, σ_t and σ_W are the mean and width of top quark and W boson, coming from artificial fitting with jets corresponding to real decay quark with generator information in simulation sample(in appendix****), which are 168.15GeV, 81.25GeV, 20.6GeV and 12.1GeV seperately. Back to the part of χ^2, for each combination, m_{jjb} is the invariant mass of 1 b-tagged jet and 2 non-btagged jets; m_{jj} is the invariat mass of 2 non b-tagged jets which are same 2 jets in m_{jjb}. the combination who have the minimum χ^2 value in all of them is chosen as physical obsects coming from hadronically decay top.



(a) muon channel

(b) electron channel

Figure 2: Reconstructed M_{jjb} with χ^2_{min} algorithm (w/o cut)

V.3.2 MVA Method

However, to improve the reconstruction performance in my analysis, there is another method - **Multi-Variate Analysis (MVA)** can be adopted to do well in this reconstruction part. It has been not a common way to used in this step yet compared to signal and background samples' classification. In order to check the improvement between usual and new method, in the following rest analysis, the comparison of analysis results of χ^2_{min} method and MVA method will be shown simultaneously.

The concept of MVA is to use basic machine learning method to classify signal and background, we'll take advantage of MVA discriminating ability to improve the correctness rate of selection compared with χ^2_{min} method.

As the usage of classification, one should define the "signal" and "background" in MVA configuration. In most particle physics analysis with MVA, "signal" means the physics sample which is expected to be analyzed and as a target sample and the "background" means samples from the other known physics. Different as usual, in this analysis,

- **Sample** : signal simulation sample($t\bar{t}$ MC) with full event selection(V.2)
- Randomly half for training sample, another half for testing sample
- In each event,
    1. **Signal** is recognized as the correct combination of objects hadronically decaying from $t(\bar{t})$-quark in an event
    2. **Background** are all the other incorrect combinations in an event

To identify whether one combination is the 2 jets and the correct detector-level's objects from generator-level's particles, the ΔR method is used to match them. The matching method is to compare the angle distribution between detector-level objects and generator-level particles with $\Delta R < 0.4$.

$$\Delta R = \sqrt{(\eta_{det} - \eta_{gen})^2 - (\phi_{det} - \phi_{gen})^2} < 0.4 \tag{3}$$

With concept of machine learning, we need to train with informations of classes("signal" and "backgroud" in this case) and get out an algorithm to make distinguishment. Following the original χ^2_{min} method's variables, we started MVA with inputting 2 variables : m_{jj}, m_{jjb}, as informations to be used for distinguishment. There are three machine learning methods I used for testing: **MLP**(ANN, Artificial Neuro-Network), **BDT**(Boosted Decision Tree), **BDTG**(Boosted Decision Tree Gradient).

With respect to the input variables for distinguishment training, we can check the seperation shows on input variables originally in advance:

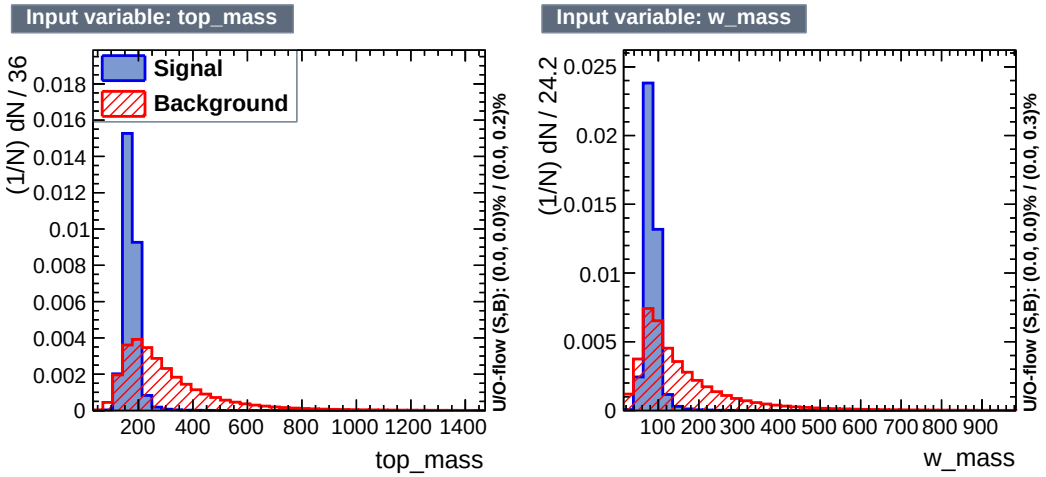
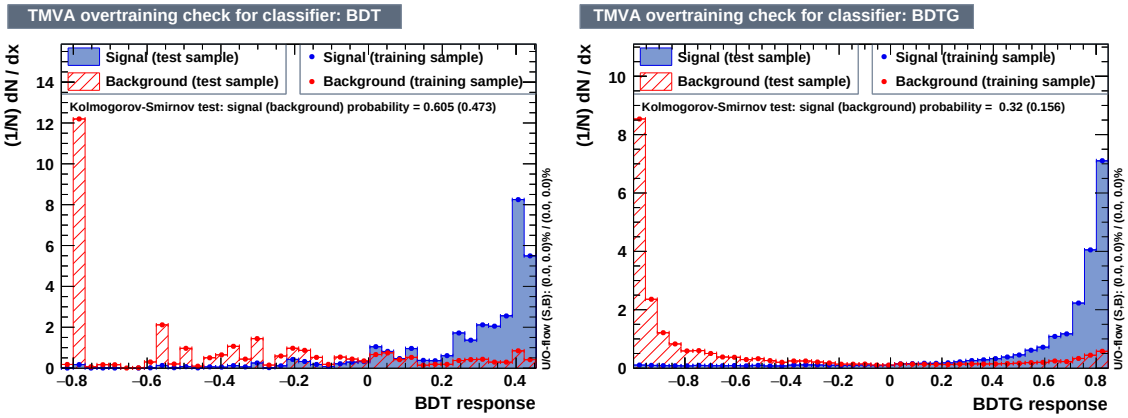


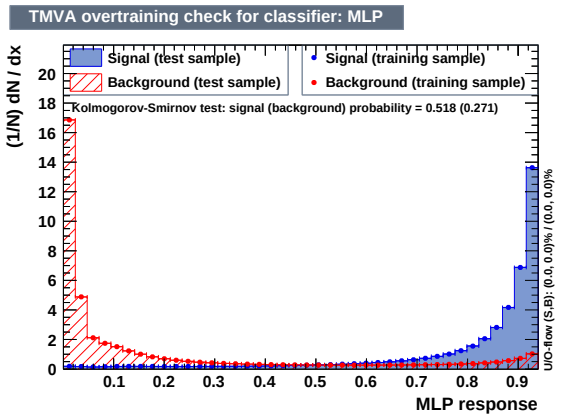
Figure 3: Input training variables separation between "signal" and "background".(2 variables)

And the training result shows below:



(a) BDT's separation result

(b) BDTG's separation result



(c) MLP's separation result

Figure 4: The seperating distribution on signal and background

The separation plots Fig.4 shows the machine learning methods' separating ability through the input training variables. The distributions in these plots are the final separation performance between "signal" and "background"(right and wrong objects combination) with the training values which is propagated from separations of all input variables(Fig.3). It is also the separation performance under training sample and testing sample with given machine learning method. As we can see that if we want to pick out "signal"(right combination), just choose the jjb combination who has the *highist MVA score* as the right one.

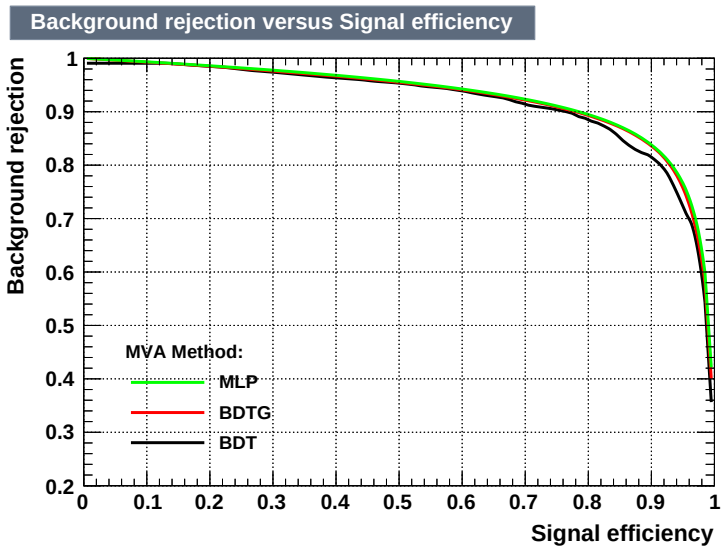
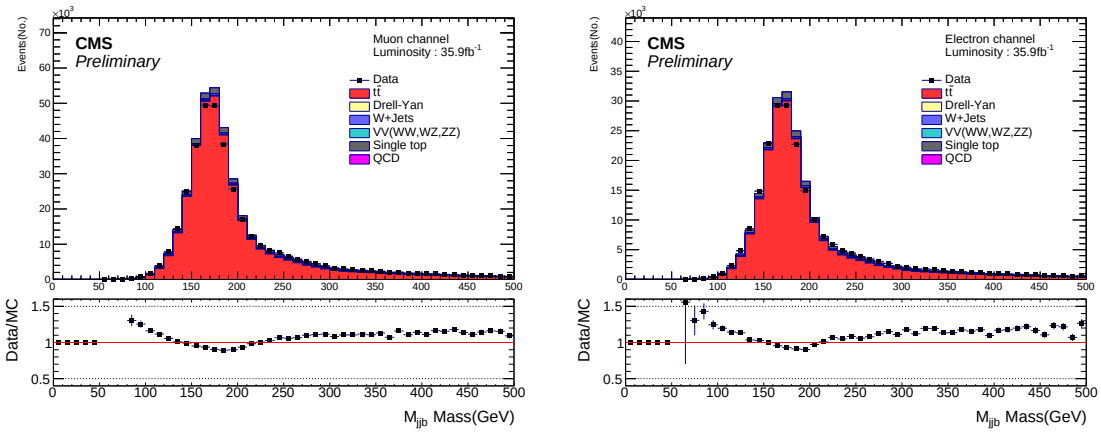


Figure 5: Receiver Operating Characteristic(ROC) curve of various machine learning method

The **Receiver Operating Characteristic(ROC)** curve plot Fig.5 can tell if it's given a cut on the output MVA score in separation distribution Fig.4 to extract the "signal", how the background rejection ratio vary when signal efficiency change. It can be infered that, the good performance is that when one rejects more ratio of background and at the same time reserves more ratio of signal(high signal efficiency). In other words, the bigger area under the ROC curve, the better the method is. However, the previous ROC criterion are just available and meaningful for the common case - one use MVA to separate the 2 different physics sample which are independent to each other, for example, use MVA to separate sigle t and $t\bar{t}$ sample. In this analysis case, the signal and background are not separate like that kind in common. Being not independent 2 samples, signal and background are at the same time in one event instead. There are couple of complicated correlation between them. Therefore, the separation and ROC plots are not really fair anymore. As they going to not to be relative directly, there must be a standard to tell how MVA perform(That is the $b\bar{b}$ separation in V.4).

There are also the reconstructed M_{jjb} with MVA algorithm.(m_{jjb}, m_{jj}) There is just MLP (instead of BDT/BDTG)results shown here.



(a) muon channel (b) electron channel

Figure 6: Reconstructed M_{jjb} with 2 variables MLP algorithm (w/o cut)

And also, it is better to use MVA with more variables than just m_{jj} and m_{jjb}. There are 3 variables sets after a bunch of trials here to input and train. It's a remindful item that they are the variables of each combination in any event.

1. The first set: (2 variables)
   - m_{jjb}, m_{jj}
2. The second set: (8 variables)
   - m_{jjb}, m_{jj}
   - 2 jets'(jj) *sum of* P_T, $\Delta\phi$, $\Delta\eta$
   - selected lepton and leptonic b-jet's *sum of* P_T, $\Delta\phi$, $\Delta\eta$
3. The third set: (20 variables)
   - m_{jjb}, m_{jj}
   - 2 jets'(jj) *sum of* P_T, $|\Delta P_T|$, ΔR
   - hadronic W(j+j) and hadronic b-jet's *sum of* P_T, $|\Delta P_T|$, ΔR
   - selected lepton and hadronic b-jet's *sum of* P_T, $|\Delta P_T|$, ΔR
   - selected lepton and hadronic W(j+j)'s *sum of* P_T, $|\Delta P_T|$, ΔR
   - hadronic W(j+j) and MET's *sum of* P_T, $|\Delta P_T|$, $\Delta\phi$
   - hadronic b-jet and MET's *sum of* P_T, $|\Delta P_T|$, $\Delta\phi$

Besides the training result of the 2 variables' set have been shown(Fig.3, Fig.4, Fig.5), there are also the training results of 8 variables and 20 variables' cases:

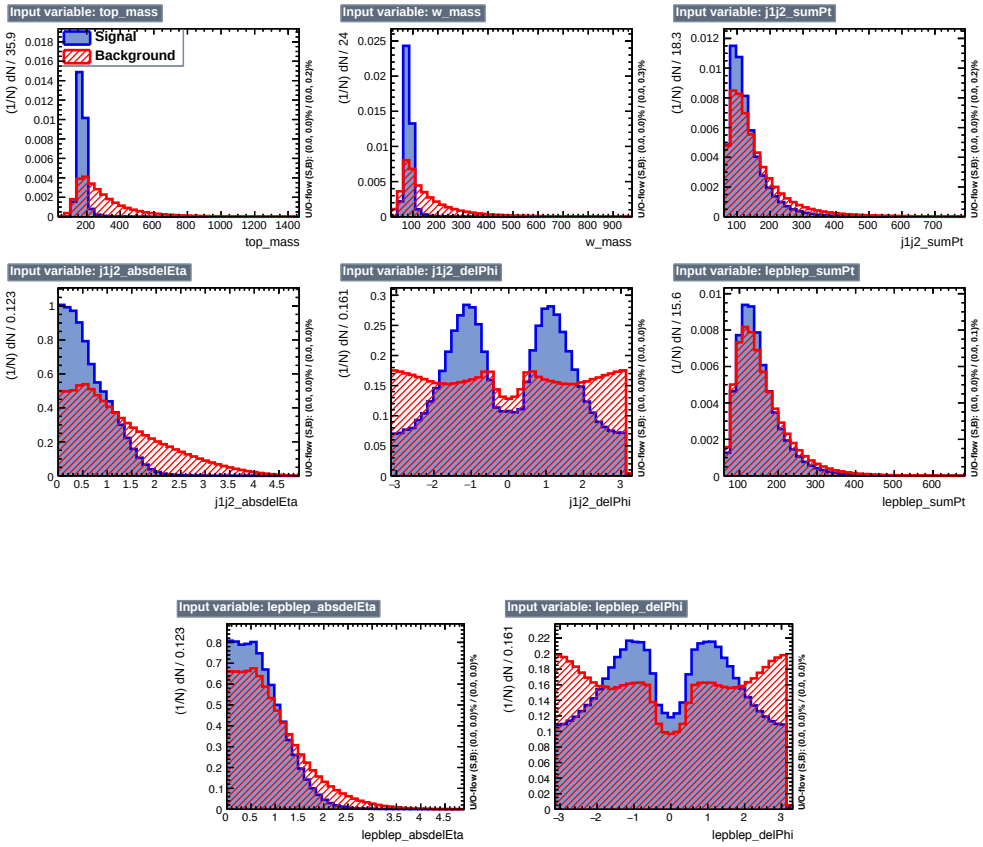
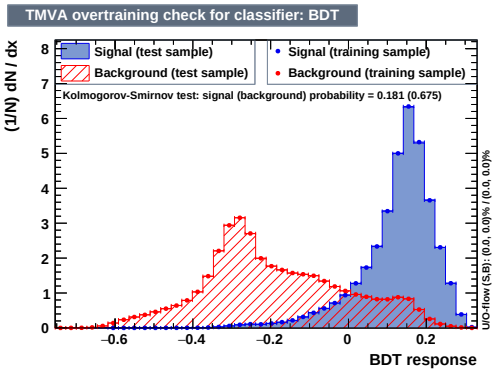
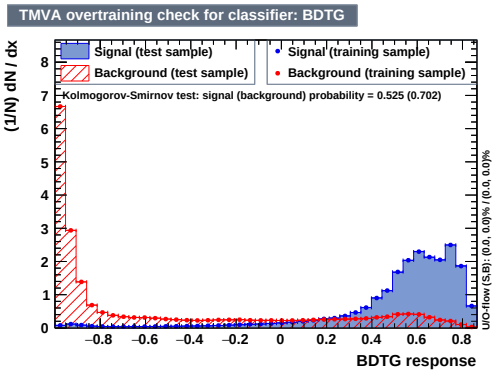


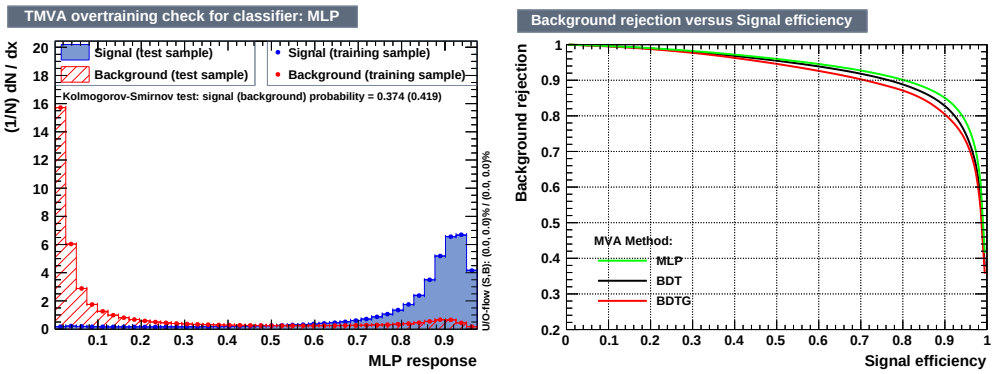
Figure 7: Input training variables separation between "signal" and "background".(8 variables)



(a) BDT's separation result (8 vars)

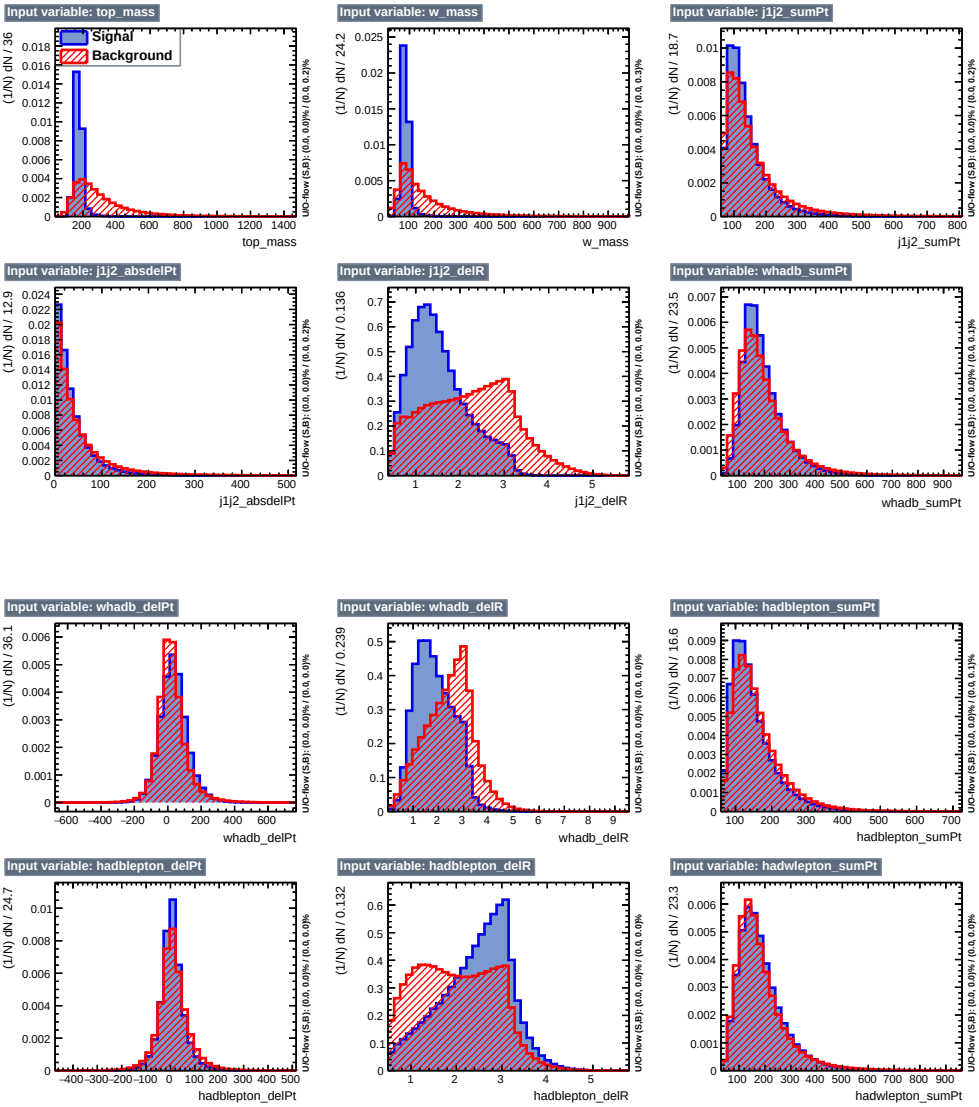


(b) BDTG's separation result (8 vars)



(c) MLP's separation result (8 vars) (d) ROC curve (8 vars)

Figure 8: The training result of 8 variables set



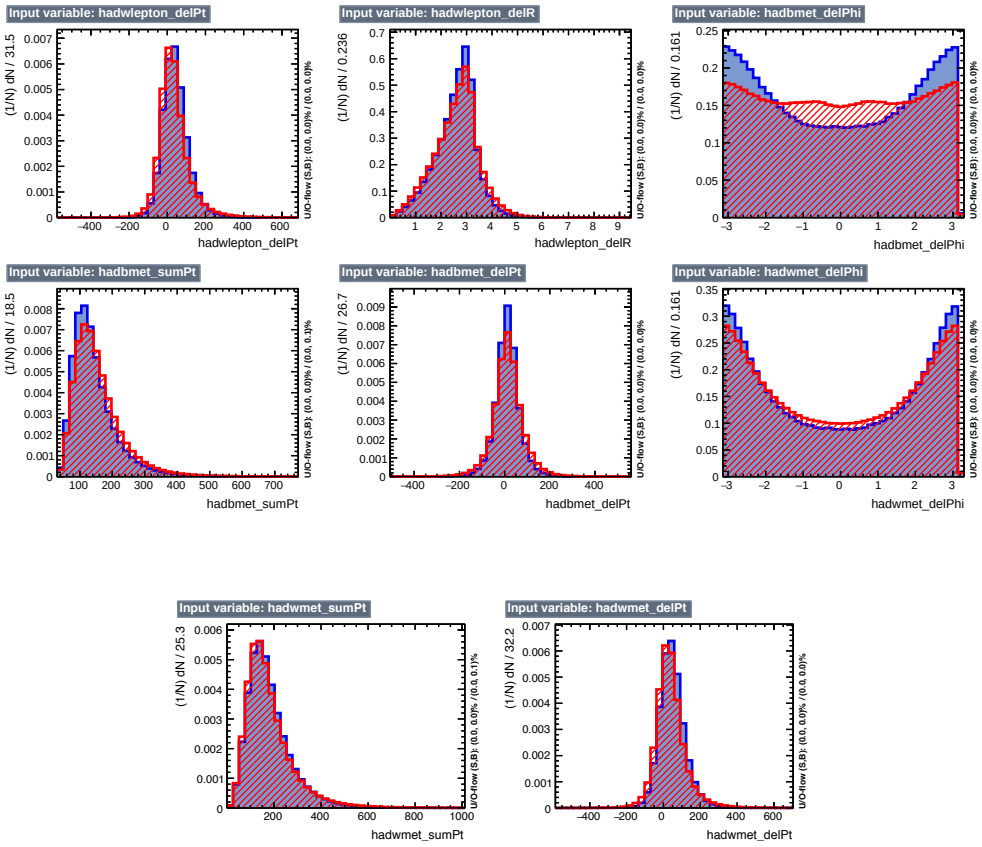
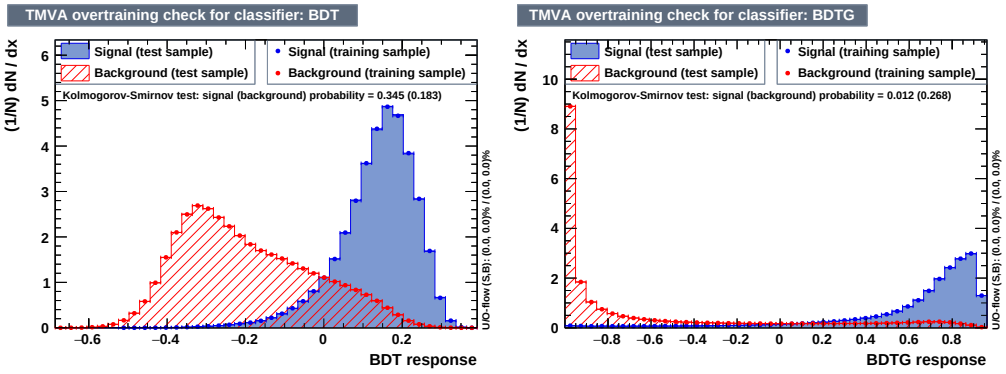
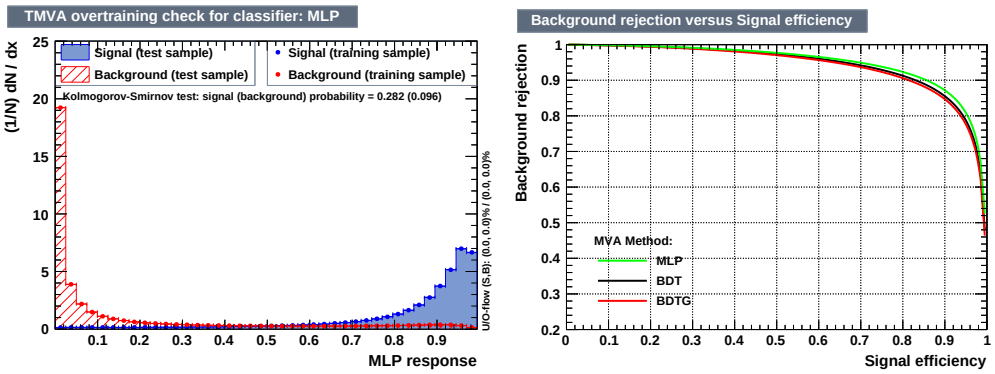


Figure 9: Input training variables separation between "signal" and "background".(20 variables)



(a) BDT's separation result (20 vars) (b) BDTG's separation result (20 vars)

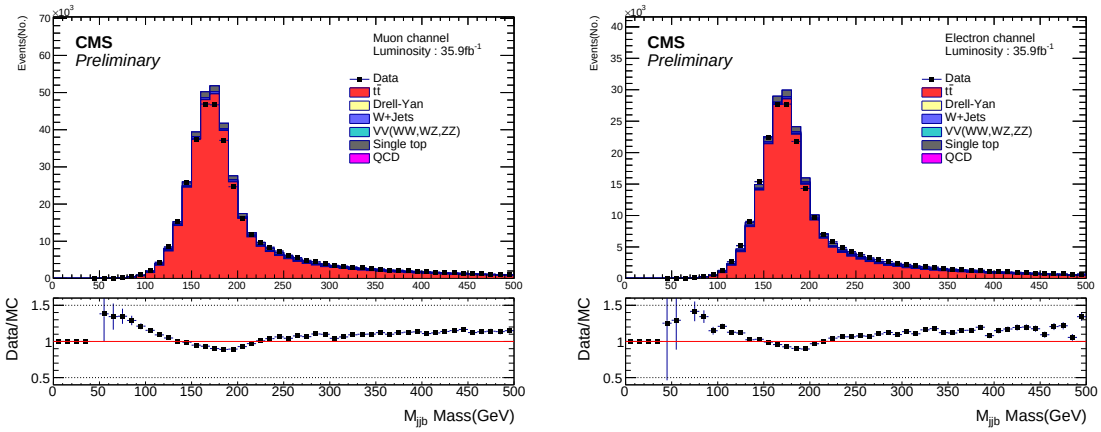


(c) MLP's separation result (20 vars) (d) ROC curve (20 vars)

Figure 10: The training result of 20 variables set

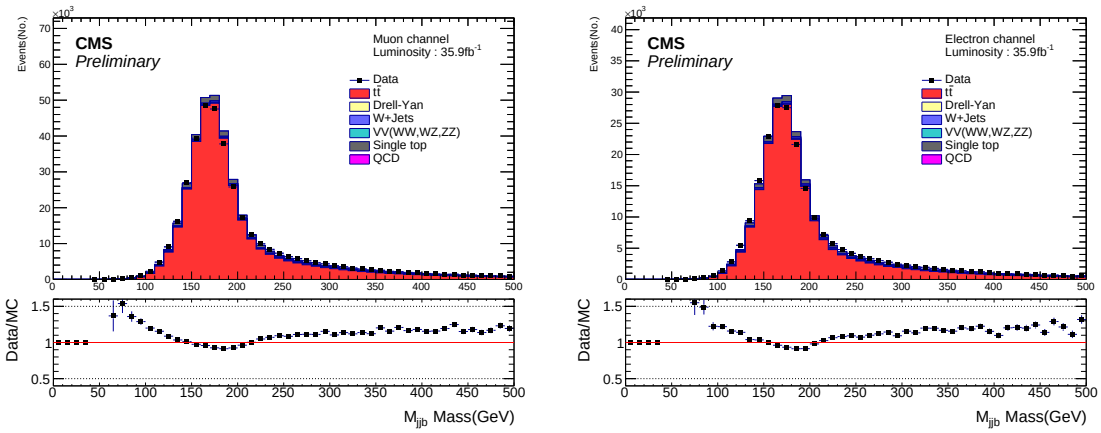
This is shown that the MLP(ANN) training algorithm has the best performance under ROC criteria from training results. But as previous mentioning, the ROC performance cannot totally represent the training performance completely in these analysis case.

There are results of reconstructed M_{jjb} by 8 variables' and 20 variables' MLP algorithm.



(a) muon channel (b) electron channel

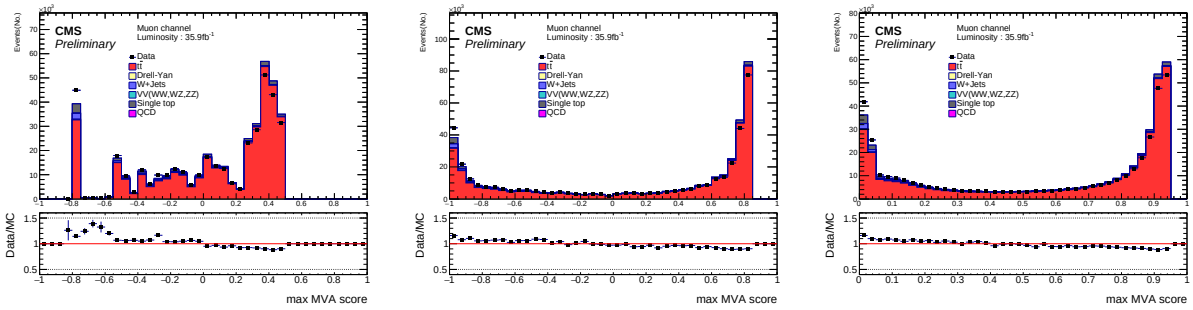
Figure 11: Reconstructed M_{jjb} with 8 variables MLP algorithm (w/o cut)



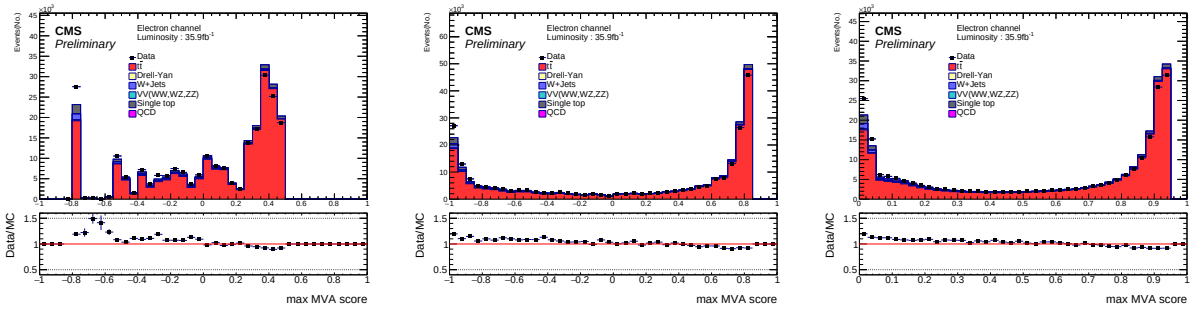
(a) muon channel (b) electron channel

Figure 12: Reconstructed M_{jjb} with 20 variables MLP algorithm (w/o cut)

There are the max MVA score's distribution with 2/8/20 variables and BDT/BDTG/MLP results.

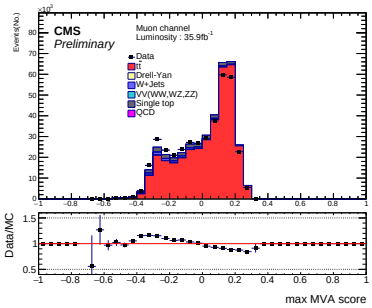


(a) BDT (2 vars, mu-ch) (b) BDTG (2 vars, mu-ch) (c) MLP (2 vars, mu-ch)

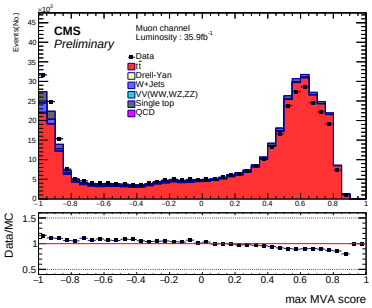


(d) BDT (2 vars, el-ch) (e) BDTG (2 vars, el-ch) (f) MLP (2 vars, el-ch)

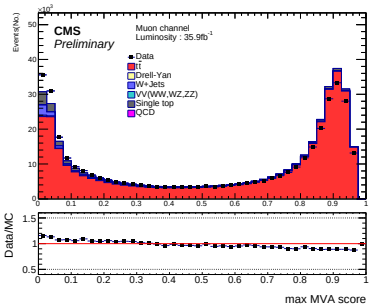
Figure 13: max MVA score in each event, comparing Data and MC.(2 variables training)



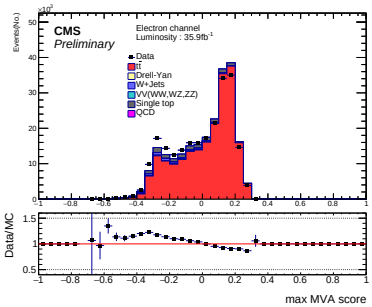
(a) BDT (8 vars, mu-ch)



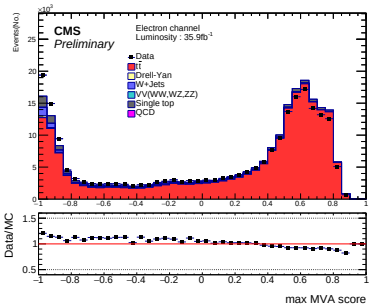
(b) BDTG (8 vars, mu-ch)



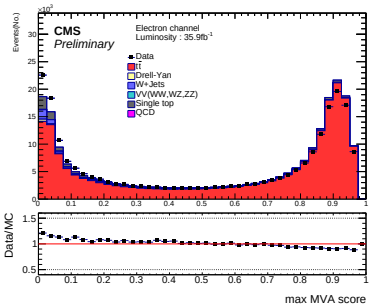
(c) MLP (8 vars, mu-ch)



(d) BDT (8 vars, el-ch)

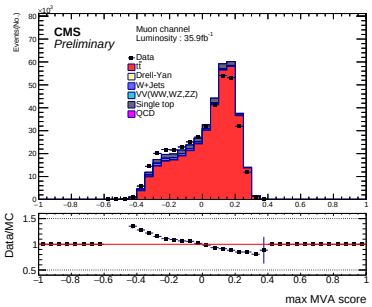


(e) BDTG (8 vars, el-ch)

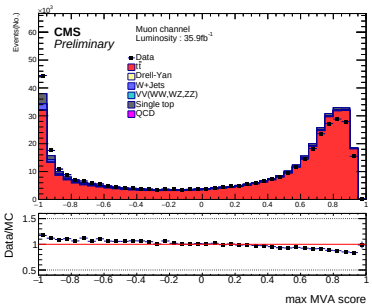


(f) MLP (8 vars, el-ch)

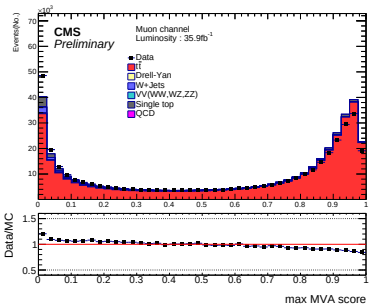
Figure 14: max MVA score in each event, comparing Data and MC.(8 variables training)



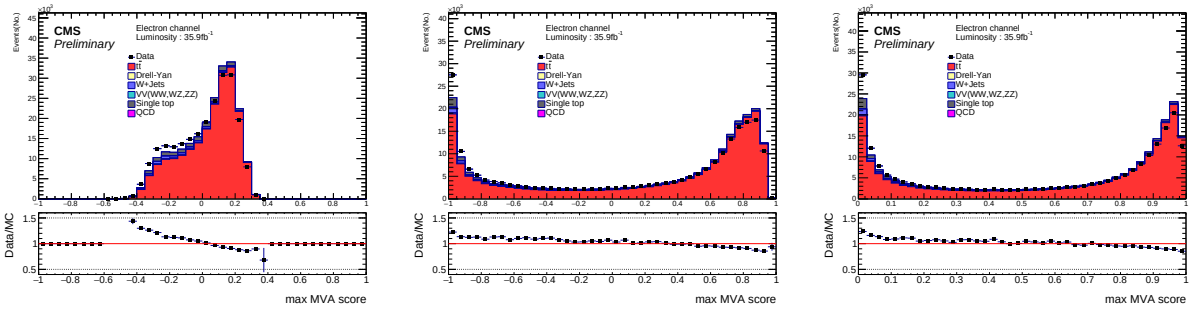
(a) BDT (20 vars, mu-ch)



(b) BDTG (20 vars, mu-ch)



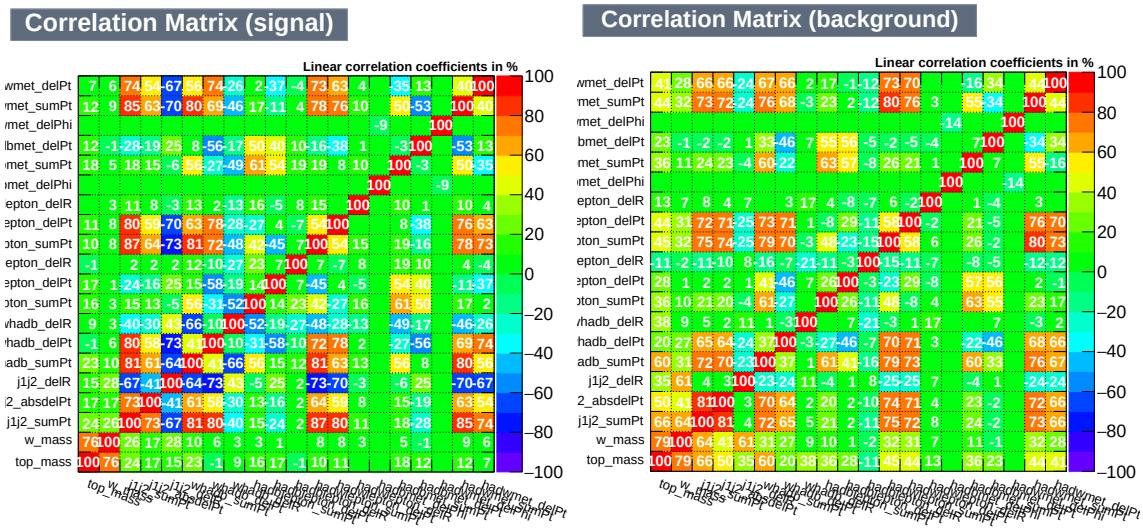
(c) MLP (20 vars, mu-ch)



(d) BDT (20 vars, el-ch) (e) BDTG (20 vars, el-ch) (f) MLP (20 vars, el-ch)

Figure 15: max MVA score in each event, comparing Data and MC.(20 variables training)

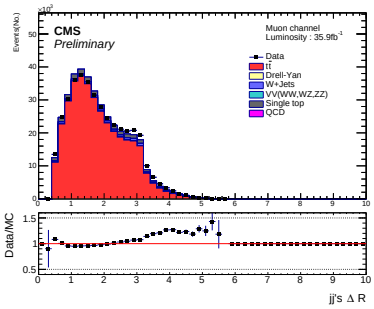
The training variables may have correlation with each other. If there are too much characteristic to be trained alike between variables, the most of input informations would duplicate. To check the variables' correlation score is necessary: (20 variables' case is shown)

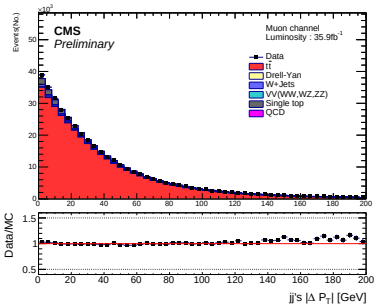


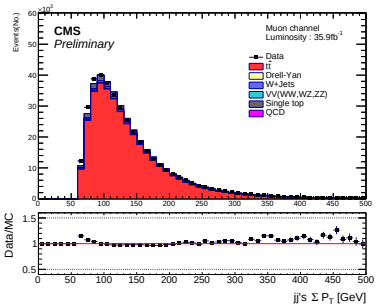
(a) Variables in signal (b) Variables in background

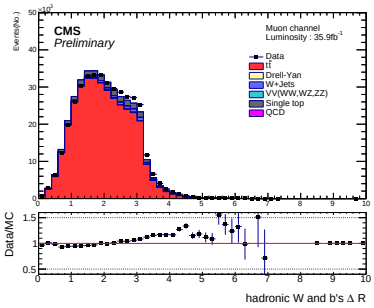
Figure 16: Variables correlation with each other

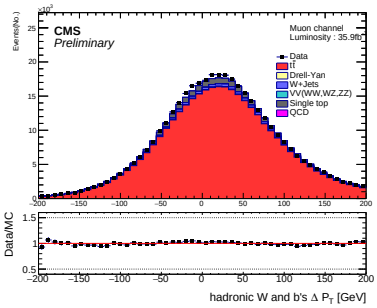
No excess of correlation appear in training variables. The training variables would be also validated match between Data and MC:(20 variables MLP, muon channel is shown)

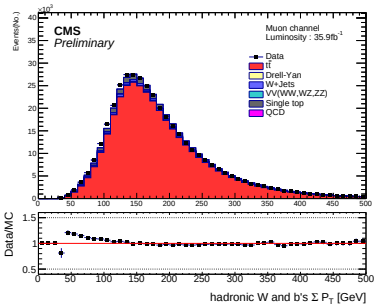

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
jj's Δ R

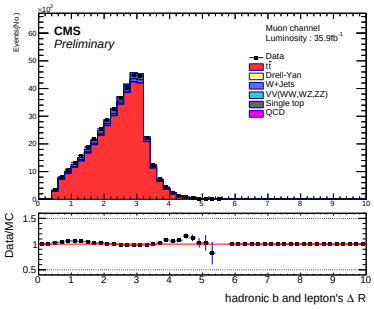

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
jj's |Δ P$_T$| [GeV]

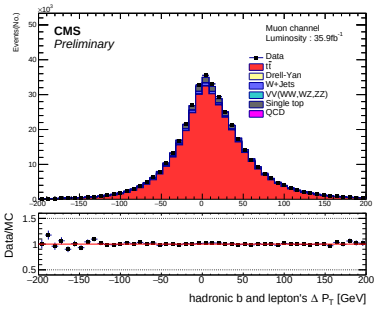

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
jj's Σ P$_T$ [GeV]

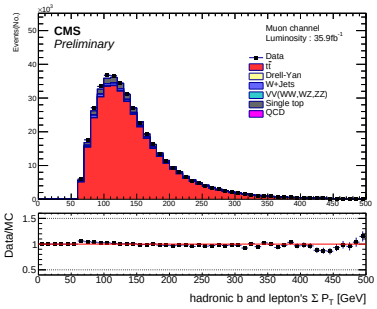

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and b's Δ R

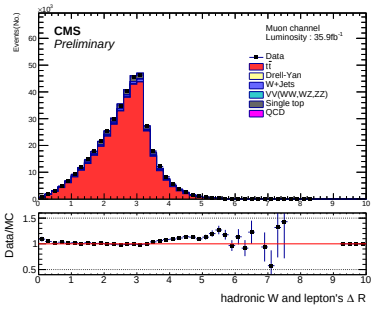

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and b's Δ P$_T$ [GeV]

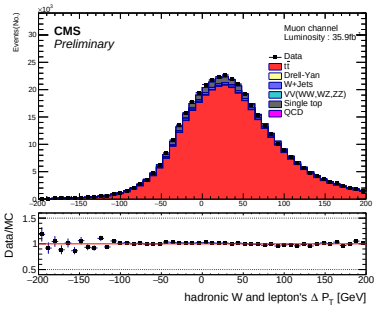

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and b's Σ P$_T$ [GeV]

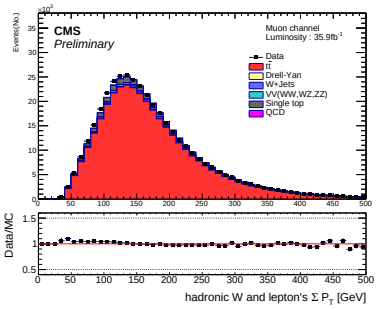

CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic b and lepton's Δ R


CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic b and lepton's Δ P$_T$ [GeV]


CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic b and lepton's Σ P$_T$ [GeV]


CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and lepton's Δ R


CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and lepton's Δ P$_T$ [GeV]


CMS
Preliminary
Muon channel
Luminosity : 35.9fb^{-1}
Data
tt̄
Drell-Yan
W+Jets
VV(WW,WZ,ZZ)
Single top
QCD
Events(No.)
Data/MC
hadronic W and lepton's Σ P$_T$ [GeV]

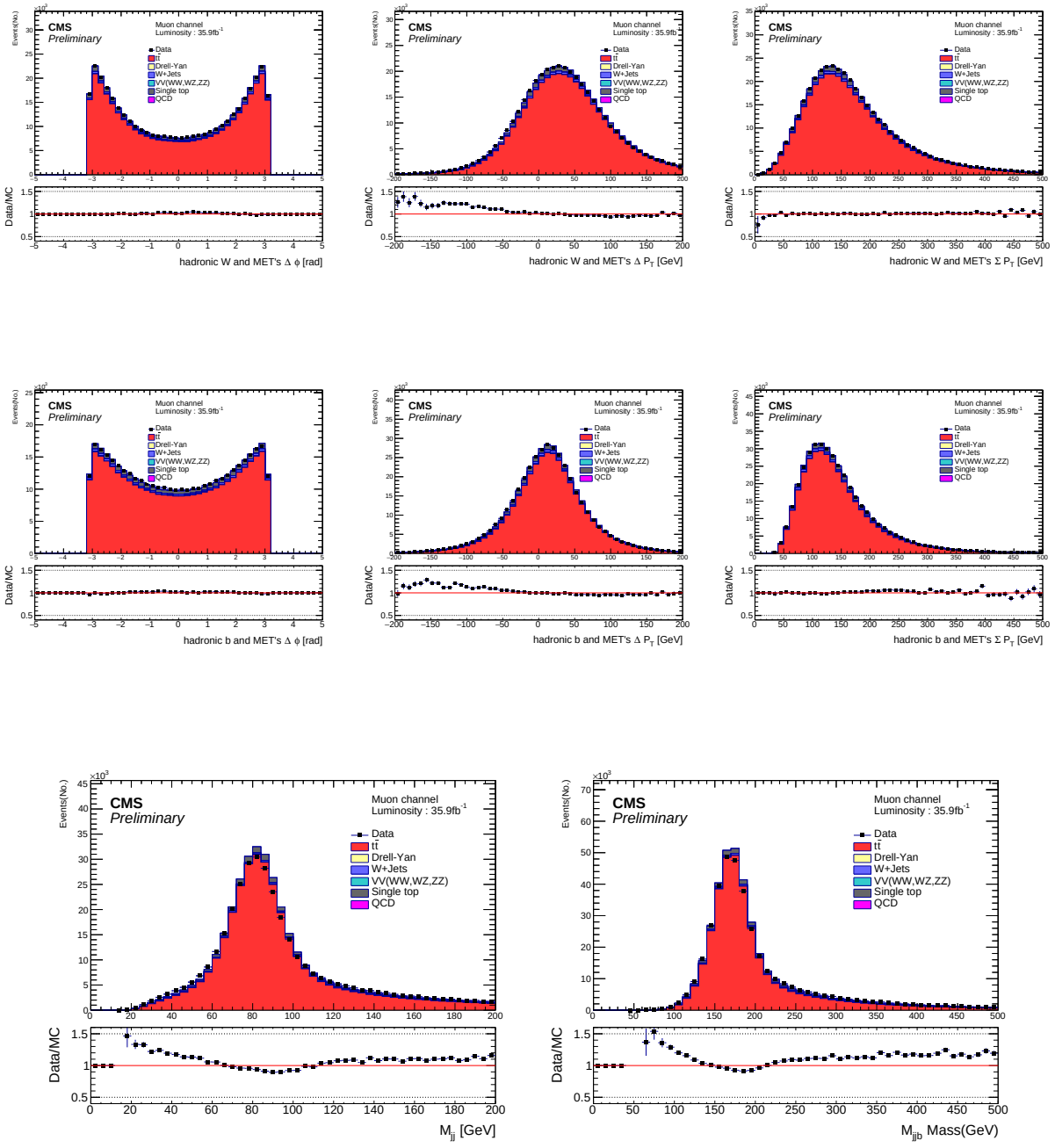


Figure 17: 20 variables' Data/MC comparison(MLP, mu-ch)

V.4 $b\bar{b}$ separation and distinguishment

The good correctness of chosen physical objects is necessary in this analysis. By the basis for discovering new physics, the observables require precise identification of particles. And the important is, in our selected observables, to distinguish b- between $\bar{b}$ -quark which are both b-tagged jets detected from smashing and hadronizatoin of b-flavor quarks. For example, the

identification is highly correlated with observable O_6 which is $q_l(\vec{p_b}-\vec{p\bar{b}})\cdot(\vec{p_l}\times\vec{p_{j_1}})$, the mis-ordered $b/(\bar{b})$ will cause the wrong sign on this observable. On the other hand, it is not requirement for us to correctly pick up the 2 jets from hadronic top quark in this analysis. Also, the long tail in reconstructed hadronic top invariant mass(M_{jjb}, Fig.6, Fig.11, Fig.12) deviated too much from known top mass 172.5GeV. And the source of this tail is infered coming from $b/\bar{b}$ wrong identified cases. In this way, the identification between b and $\bar{b}$ is the most critical implication. And that is also what we need to test the performance of reconstruction algorithm(χ^2_{min},MVA).

All the following study in V.4 is worked only under the signal simulation sample($t\bar{t}$ Monte Carlo). All the selected physical objects include jets and leptons will be matched to particle with the generator-level information in MC sample with $\Delta R < 0.4$ method.(Eq.3)

To standardlize the $b/\bar{b}$ distinguishment, there are 3 classifications listed:

- **Correct**: b-quark is identified as b-jets and $\bar{b}$-quark is identified as $\bar{b}$-jets
- **b/$\bar{\text{b}}$ mis-identified**: b-quark is identified as $\bar{b}$-jets and $\bar{b}$-quark is identified as b-jets
- **Mistag**: non $b(\bar{b})$-quark is identified as $b(\bar{b})$-jet

With applying the various algorithm on signal simulation sample($t\bar{t}$ MC), there are the $b/\bar{b}$-distinguishment results after reconstructed the hadronic top and make a identification of physical objects via χ^2_{min} and MVA methods:

Table 1: $b/\bar{b}$ disdinguishment under different algorithm (w/ o cut)

[%]		Correct	$b\bar{b}$ mis-identified	Mistag
χ^2_{min}		61.18	25.21	13.38
2 variables	MLP	63.18	23.22	13.60
	BDT	60.89	25.49	13.60
	BDTG	63.02	23.37	13.61
8 variables	MLP	72.36	14.02	13.62
	BDT	72.02	14.37	13.60
	BDTG	71.00	15.39	13.61
20 variables	MLP	71.27	15.13	13.61
	BDT	71.07	15.32	13.61
	BDTG	69.59	16.79	13.61

This is shown that the MLP algorithm have the best performance under these 3 training variables sets. As our expected, the 2 variables training result is close to χ^2_{min} method's because

of their same variables input to be deduced. And the 8 variables set and 20 variables set are both good at discreminate $b/\bar{b}$, which can lead us to choose these to be optimized one. It is also known that *Mistag*'s ratio in Table.1 are almost the same since it is primarily pre-decided by the b-tagged selection.

In addition to applying algorithm to reconstruct M_{jjb} and identify the physical objects, there are also some optimization cut done on this step. In the concept of χ^2_{min} method, the smaller the χ^2 value(Eq.2), the less possibility that this combination is the one we want. So we can apply cut on χ^2 value. If in an event the chosen combination whose χ^2_{min} is still bigger than the cut, the event will be threw away. These thrown away events are identified as the events which come from the mis-tag objects from detector or some events come from other physics decay(background) not $t\bar{t}$. The most important is that, it can cut out the events which are not $b/\bar{b}$-correct distinguished. So the cut is set to be at $\chi^2_{min} < 20$, and to standardize the optimization, there are also use the $b/\bar{b}$ distinguishment results to see the performance.

There is the cross-check plot(Fig.18) which shows that the high χ^2 value correspond to the m_{jjb} distant from the expected m_{jjb} in Eq.2.

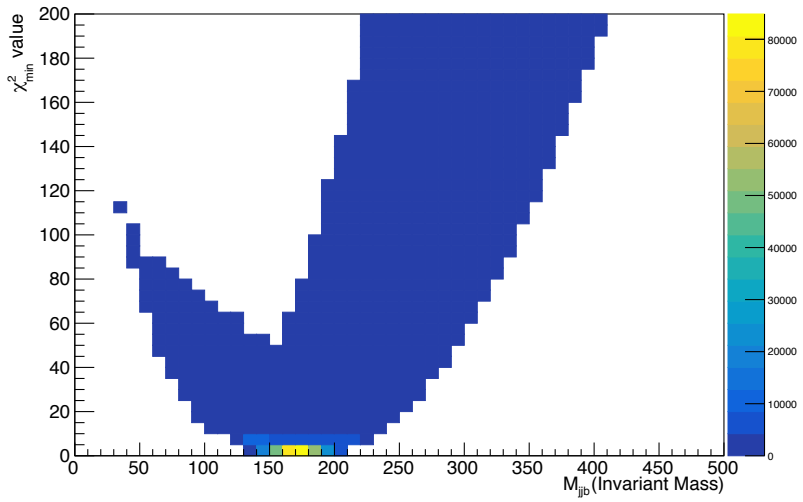


Figure 18: χ^2_{min}-reco m_{jjb}

To check that the cut on the χ^2_{min} really rule out incorrect $b/\bar{b}$-identified more effeciently than correct case. This plot(Fig.19) shows that the incorrect cases(exchange-identified and mistag) do have more ratio at the higher χ^2_{min} score than the correct case. It make the χ^2_{min} cut more persuasive.

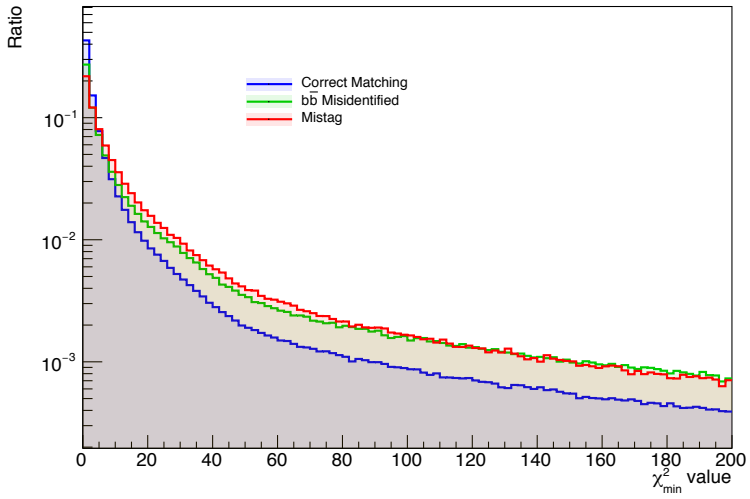


Figure 19: $b/\bar{b}$ identified result related to χ^2_{min}(pdf)

There is the result that if we cut on the χ^2_{min} value and how the 3 $b/\bar{b}$-identified classifications' ratios vary. It is also shown that the events efficiency which is the ratio of events survived under this optimized cut.

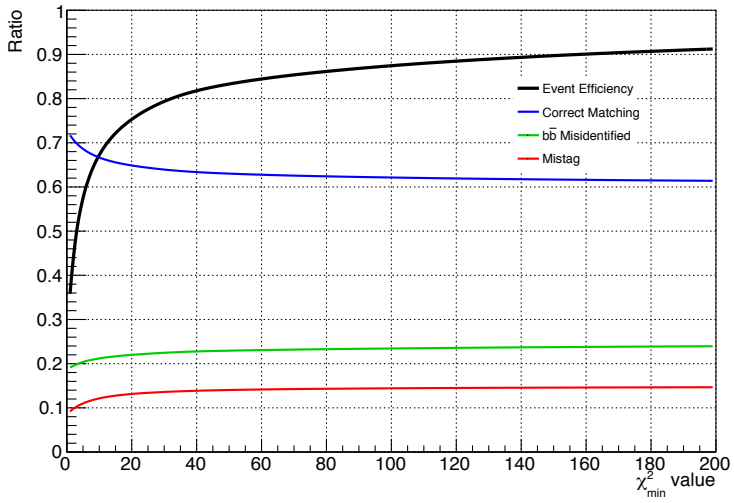


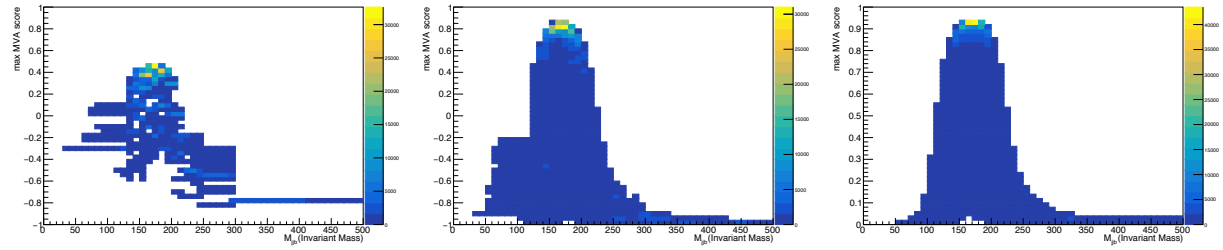
Figure 20: cut on χ^2_{min} and ratio of 3 classification

Table 2: $b/\bar{b}$ disdinguishment under χ^2_{min} method (w/ $\chi^2_{min} < 20$ cut)

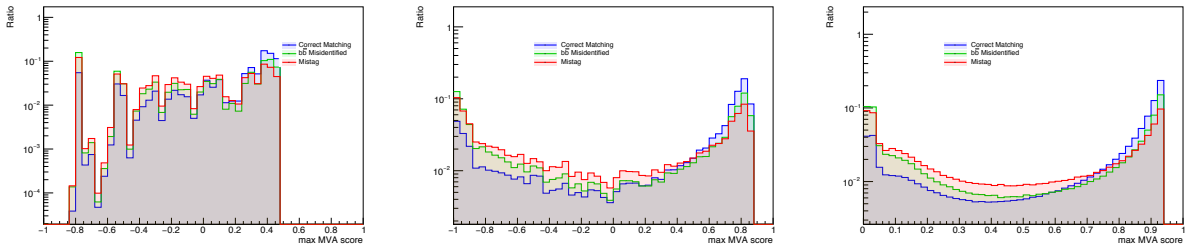
[%] Channel	Correct	$b\bar{b}$ mis-identified	Mistag

χ^2_{min}	Electron Channel	65.47	22.29	12.24
	Muon Channel	65.65	22.29	12.06

And as same concept as the optimized cut on χ^2_{min} value in χ^2_{min} method, there would be also optimization cut applied on MVA method. In comparison with that low χ^2 value is better choice, combination with high MVA score is the better one in MVA method. It also can have a cut on MVA score to subtract incorrect $b/\bar{b}$-identified selection. Same scenario as χ^2_{min} method, if in an event the maximum MVA score(which belongs to the chosen combination) is still lower than the cut, the event will be threw away. We still need to check the relation between MVA score and M_{jjb}, and also check the ratio of 3 $b/\bar{b}$-classifications that the incorrect cases has more ratio at lower MVA score. There are the plots of 2 variables/8 variables/20 variables sets with MLP/BDT/BDTG training method:

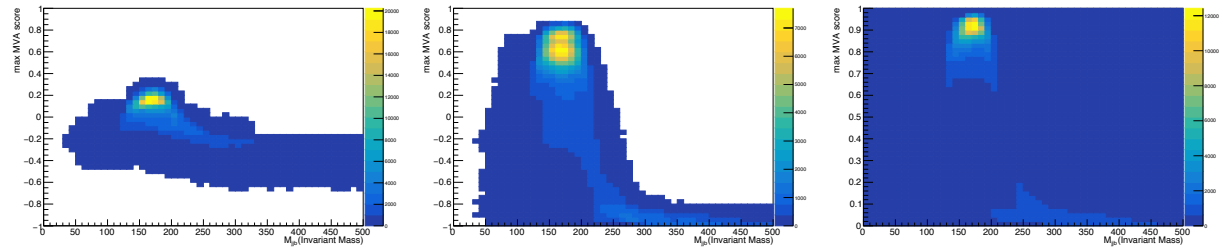


(a) BDT (b) BDTG (c) MLP

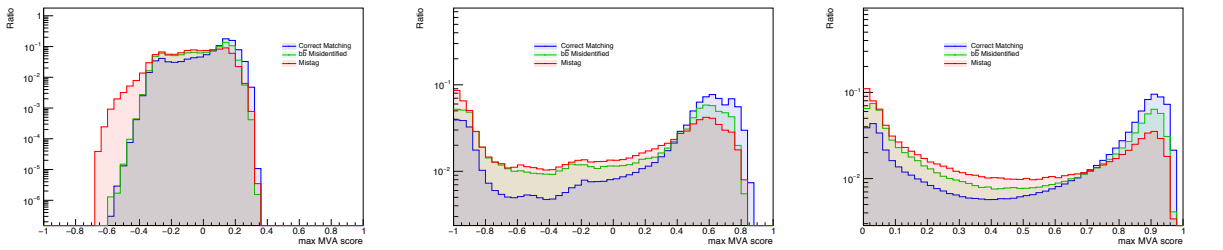


(d) BDT (e) BDTG (f) MLP

Figure 21: MVA score-reco m_{jjb} plots and $b/\bar{b}$ identified result related to MVA value (2 variables training)

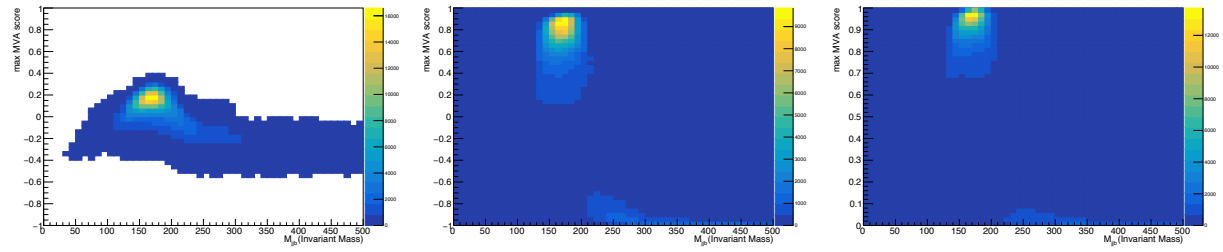


(a) BDT (b) BDTG (c) MLP

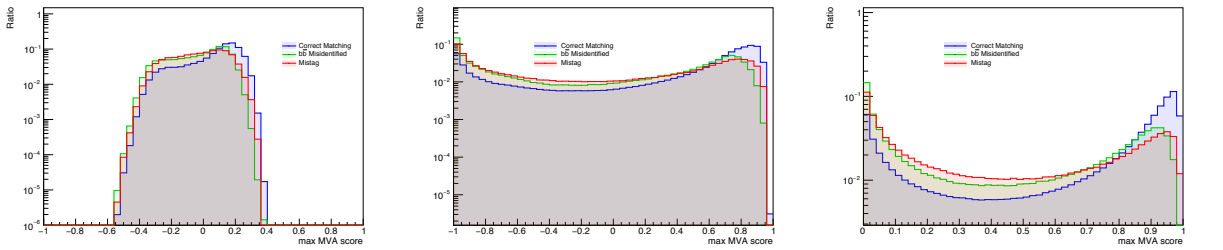


(d) BDT (e) BDTG (f) MLP

Figure 22: MVA score-reco m_{jjb} plots and $b/\bar{b}$ identified result related to MVA value (8 variables training)



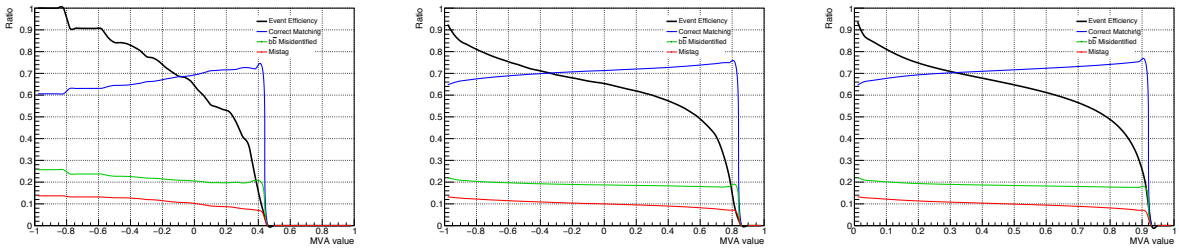
(a) BDT (b) BDTG (c) MLP



(d) BDT (e) BDTG (f) MLP

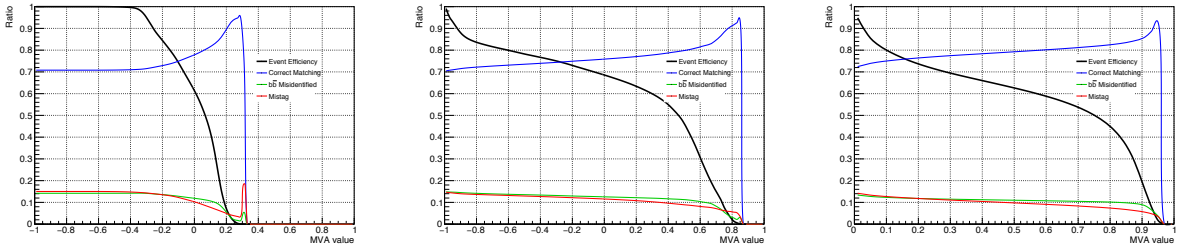
Figure 23: MVA score-reco m_{jjb} plots and $b/\bar{b}$ identified result related to MVA value (20 variables training)

There are also the effeciency plots shows that if it's given a cut on MVA score, how these 3 $b/\bar{b}$-identified cases' ratio vary.



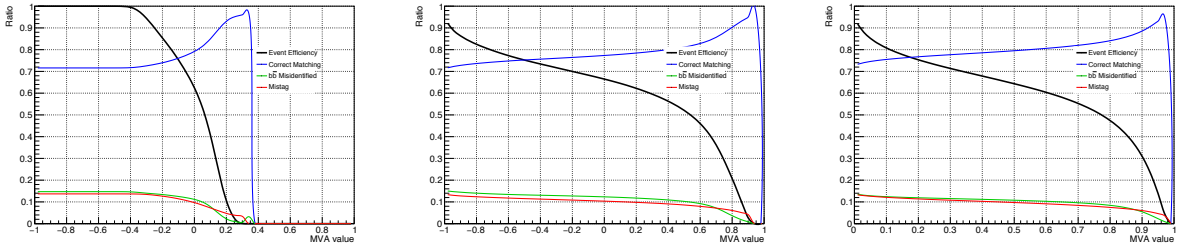
(a) BDT (b) BDTG (c) MLP

Figure 24: cut on MVA score and ratio of 3 classification (2 variables training)



(a) BDT (b) BDTG (c) MLP

Figure 25: cut on MVA score and ratio of 3 classification (8 variables training)



(a) BDT (b) BDTG (c) MLP

Figure 26: cut on MVA score and ratio of 3 classification (20 variables training)

The correct ratio increase when the MVA score-cut is applied on higher value. Based on the target to optimized from χ^2_{min} method, we have cut on the same events efficiency as χ^2_{min} method when cutting on $\chi^2_{min} < 20$, which is $\sim$75%, to get the better ratio of correct-$b/\bar{b}$ case. There are the MVA-score cut at events efficiency $\sim$75%:

Table 3: MVA-score cut at events efficiency ∼75%

MVA score cut	BDT	BDTG	MLP(ANN)
2 variables	-0.21	-0.57	0.2
8 variables	-0.1	-0.28	0.16
20 variables	-0.1	-0.5	0.22

There are the $b/\bar{b}$-identified classifications' ratio after cut on events efficiency at $\sim 75\%$(cut value in Table.3), test by signal simulation sample($t\bar{t}$ MC):

Table 4: $b/\bar{b}$ disdinguishment under different algorithm (w/ MVA cut)

[%]		Correct	$b\bar{b}$ mis-identified	Mistag
χ^2_{min}		65.58	22.33	12.09
2 variables	MLP	69.22	19.43	11.35
	BDT	68.26	20.95	10.79
	BDTG	68.95	19.69	11.36
8 variables	MLP	76.88	12.35	10.77
	BDT	75.98	13.08	10.94
	BDTG	75.37	13.68	10.95
20 variables	MLP	76.41	12.34	11.25
	BDT	75.13	13.20	11.67
	BDTG	74.72	13.76	11.52

Comparing Table.1 to Table.4, it really make progress on the $b/\bar{b}$ separation with cut on algorithm value.

Besides the hadronic top's mass spectrum, there would be checked that the leptonic top's invariant mass(M_{lb}) shows. Even though the missing neutrino's 4-momentum cause the leptonic top reconstructed incompletely, it's also valuable of M_{lb} to be an independent information for other analysis strategy. Exactly, the M_{lb} is the primary variable to be analyzed for the relation between signal/background and between data/simulation in this analysis! For instatnce, the following background estimation(Chapter.VI).

And for the $b/\bar{b}$ part, $b/\bar{b}$-identified types can also be shown under M_{lb}, they are also separated obviously with χ^2_{min} method:

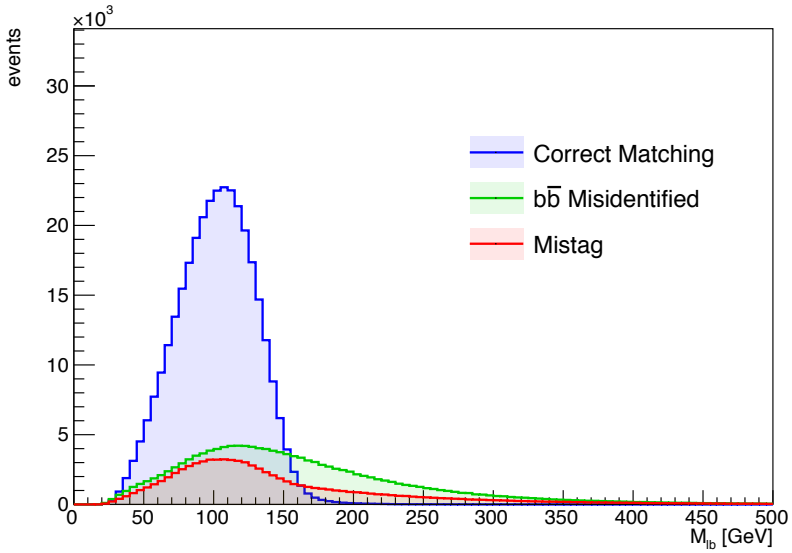
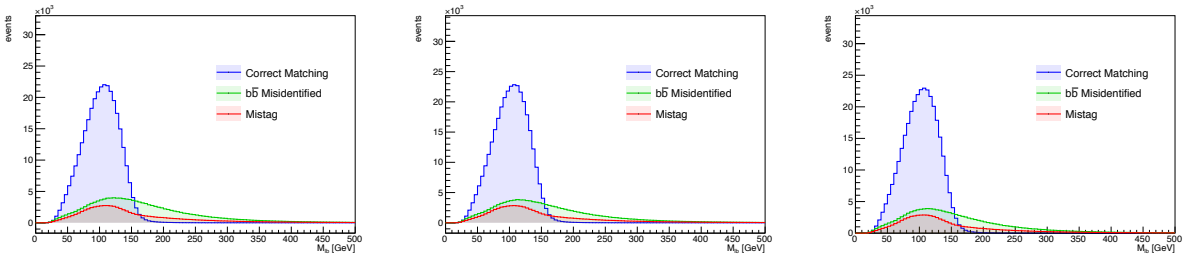


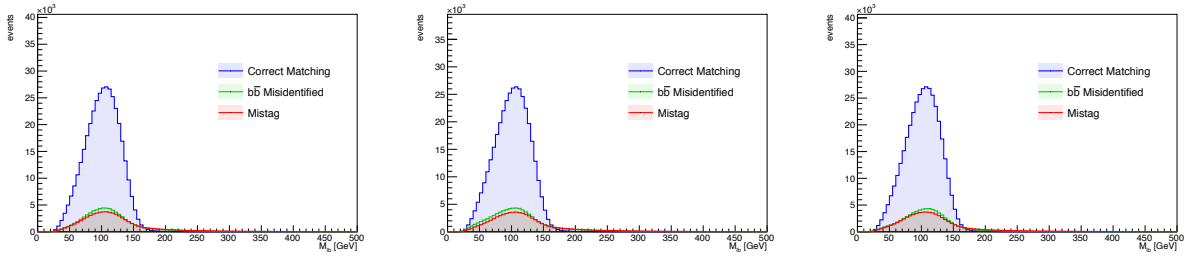
Figure 27: The M_{lb} and corresponding $b/\bar{b}$-seperation under χ^2_{min} method.

It's decide to cut on $M_{lb} = 150$ GeV and leave the events below the value. It can effectively eliminate the high ratio of events whose decay objects are wrong-identified. And follow the same reason as χ^2_{min} method, the MVA methods are also check with M_{lb} spectrum:



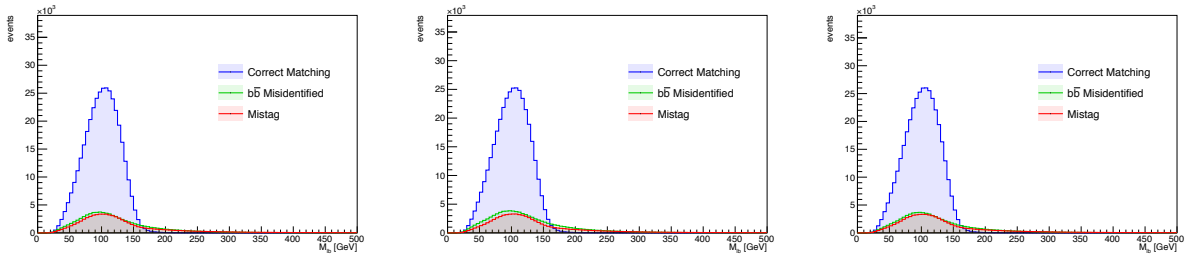
(a) BDT (b) BDTG (c) MLP

Figure 28: Relation between M_{lb} and $b/\bar{b}$ with MVA reconstruction result.(2 variables)



(a) BDT (b) BDTG (c) MLP

Figure 29: Relation between M_{lb} and $b/\bar{b}$ with MVA reconstruction result.(8 variables)



(a) BDT (b) BDTG (c) MLP

Figure 30: Relation between M_{lb} and $b/\bar{b}$ with MVA reconstruction result.(20 variables)

In 8 variables and 20 variables cases, the M_{lb} looks not to be seperated by $b/\bar{b}$-identification obviously, that is, it may not improve when applying M_{lb} cut at 150 GeV. However, there are the $b/\bar{b}$-identified ratio after MVA(or χ^2) cut and M_{lb} cut.

Table 5: $b/\bar{b}$ disdinguishment under different algorithm (w/ MVA cut and M_{lb} cut)

[%]		Correct	$b\bar{b}$ mis-identified	Mistag	Events efficiency
χ^2_{min}		75.56	14.70	9.74	65.39
2 variables	MLP	78.03	12.92	9.00	64.84
	BDT	77.66	13.69	8.65	64.92
	BDTG	77.91	13.01	9.08	64.57
8 variables	MLP	78.77	11.50	9.73	71.55
	BDT	78.12	12.05	9.83	71.44
	BDTG	77.05	12.94	10.01	71.29
20 variables	MLP	80.51	10.24	9.25	69.26
	BDT	79.34	11.01	9.65	69.09
	BDTG	78.67	11.76	9.57	69.99

- Final χ^2_{min} strategy is decided to be with χ^2_{min} value cut at 20 and M_{lb} cut at 150GeV. It is called $\chi^2_{\mathbf{min}}$ **strategy** in the following content.

The improvement is still shown with M_{lb} cut in all MVA method by comparing Table.4 and Table.5. The MLP is the most suitable algorithm after these test. The 8 variables case will be eliminated for the reason explained at Fig.41 in V.5, so the finally decided method which is the most optimized is 20 variables treaining with MLP algorithm. There are 2 choice of analysis strategy with this 20 variables MLP training result:

1. **The first** is with MVA(MLP) cut at 0.22 (events efficiency $\sim 75\%$) and also cut on M_{lb} at 150GeV.
   The correct $b/\bar{b}$ ratio is $\sim 5\%(80.51\%/75.56\%)$ better and the events efficiency is $\sim 4\%(69.26\%/65.39\%)$ more than χ^2_{min} strategy.
   This MVA result would be called **MVA-A strategy** in the following content;

2. **The second** is only with MVA(MLP) cut at 0.22 but M_{lb} cut.
   Though the correct $b/\bar{b}$ ratio is just $\sim 1\%(76.88\%/75.56\%)$ better, the events efficiency would be reserve $\sim 9.5\%(74.97\%/65.39\%)$ more than χ^2_{min} strategy. The advantage of this is that the more statistics we retain, the precise the calculation of asymmetry eventually because the statistical and systematic uncertainty would be smaller.(Explained at Chapter****)
   This MVA result would be called **MVA-B strategy** in the following content.

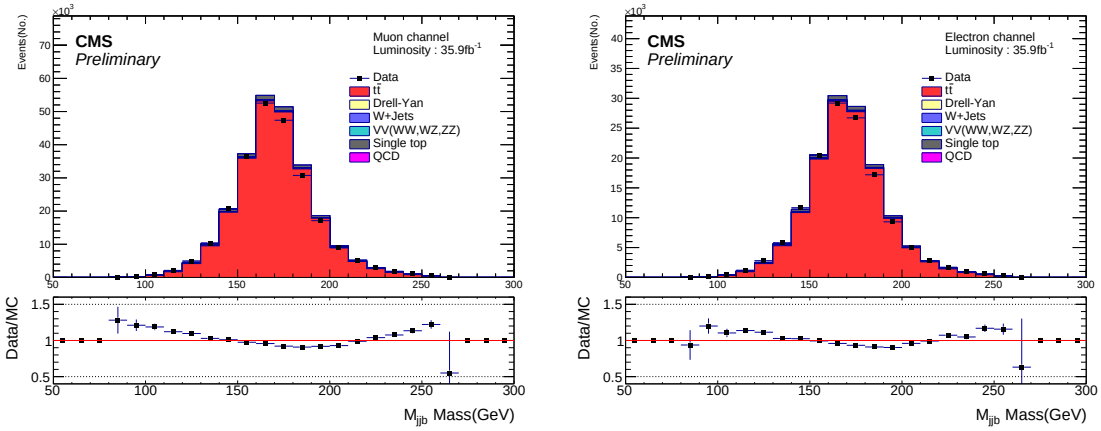
There are the data events number passed full-selection(V.2) also with $\chi^2_{\mathbf{min}}$ **strategy** and the expected signal($t\bar{t}$) and background events number of simulation sample in Table.6. The Table.7 present the expected process ratio after full-selection cut and analysis strategy; The Data and MC comparison plots of hadronic top's invariant mass(M_{jjb}) and leptonic top's invariant mass(M_{lb}) are also shown() with this criteria.

Table 6: Data and MC events number passing the full selection(w/ χ^2_{min}, M_{lb} cut)

	Muon Channel	Electron Channel
Data	243790	136151
Expected $t\bar{t}$	244680	135600
Expected background	9877.63	5537.45

Table 7: Expected process ratio passing the full selection(w/ χ^2_{min}, M_{lb} cut)

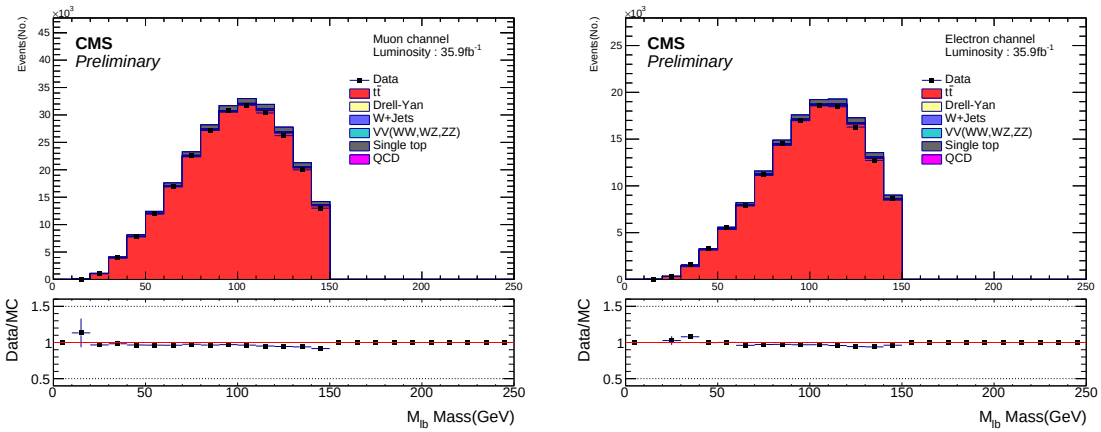
Process	Muon Channel (%)	Electron Channel (%)
$t\bar{t} + jets$	96.12	96.08
$Z/\gamma^* + jets$	0.17	0.31
$W + jets$	0.77	0.70
$ZZ/WW/WZ$	0.04	0.05
$Single\ top$	2.91	2.86



(a) Muon Channel

(b) Electron Channe

Figure 31: Data and MC comparison plots of hadronic top's invariant mass(M_{jjb} w/ χ^2_{min}, M_{lb} cut)



(a) Muon Channel (b) Electron Channel

Figure 32: Data and MC comparison plots of leptonic top's invariant mass(M_{lb} w/ χ^2_{min}, M_{lb} cut)

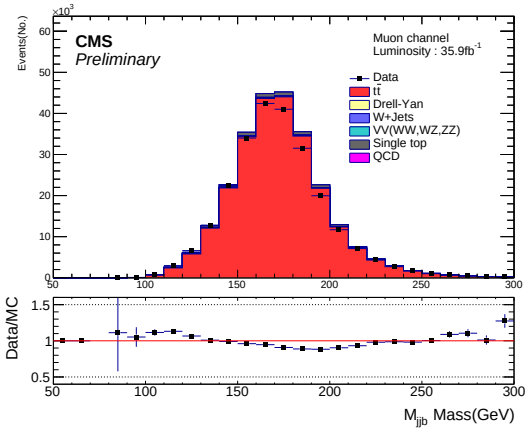
There are also the full selection results with **MVA-A strategy**(Table.8, Table.9, Fig.33, Fig.34) and **MVA-B strategy**(Table.10, Table.11, Fig.35, Fig.36):

Table 8: Data and MC events number passing the full selection in SR(w/ MVA, M_{lb} cut)

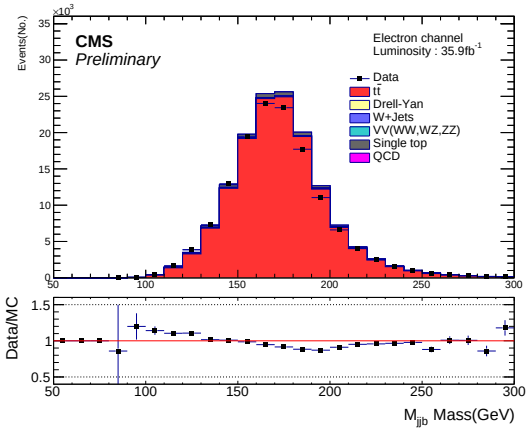
	Muon Channel	Electron Channel
Data	245436	139432
Expected $t\bar{t}$	251862	142211
Expected background	9164.09	5238.32

Table 9: Expected process ratio passing the full selection in SR(w/ MVA, M_{lb} cut)

Process	Muon Channel (%)	Electron Channel (%)
$t\bar{t}+jets$	96.49	96.45
Z/γ^*+jets	0.14	0.26
$W+jets$	0.67	0.62
$ZZ/WW/WZ$	0.03	0.05
$Single\ top$	2.67	2.64

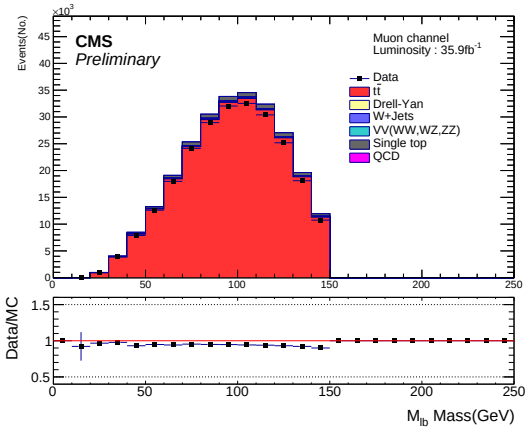


(a) Muon Channel

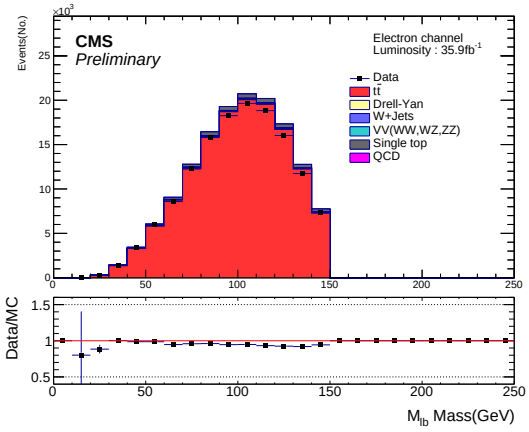


(b) Electron Channe

Figure 33: Data and MC comparison plots of hadronic top's invariant mass(M_{jjb} w/ MVA, M_{lb} cut)



(a) Muon Channel



(b) Electron Channel

Figure 34: Data and MC comparison plots of leptonic top's invariant mass(M_{lb} w/ MVA, M_{lb} cut)

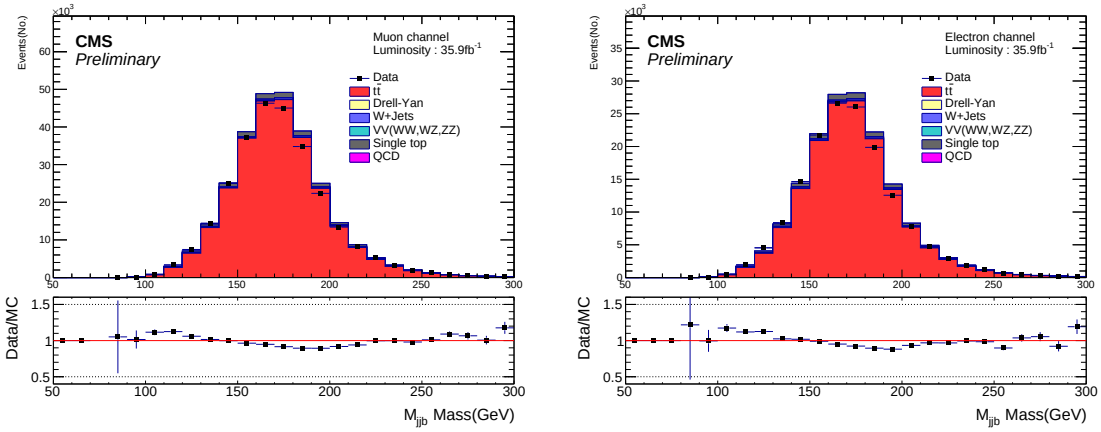
Table 10: Data and MC events number passing the full selection in SR(w/ MVA cut)

	Muon Channel	Electron Channel
Data	272583	157510

Expected $t\bar{t}$	273154	155766
Expected background	14715.9	8957.06

Table 11: Expected process ratio passing the full selection in SR(w/ MVA cut)

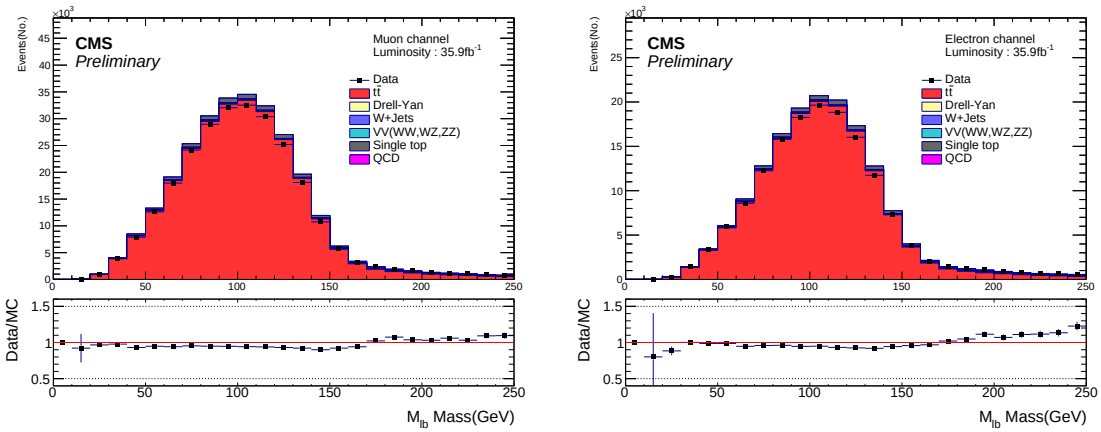
Process	Muon Channel (%)	Electron Channel (%)
$t\bar{t} + jets$	94.89	94.56
$Z/\gamma^* + jets$	0.21	0.39
$W + jets$	1.22	1.17
$ZZ/WW/WZ$	0.05	0.06
$Single\ top$	3.63	3.82



(a) Muon Channel

(b) Electron Channe

Figure 35: Data and MC comparison plots of hadronic top's invariant mass(M_{jjb} w/ MVA cut)



(a) Muon Channel

(b) Electron Channel

Figure 36: Data and MC comparison plots of leptonic top's invariant mass(M_{lb} w/ MVA cut)

As previously mentioned, the M_{lb} instead of M_{jjb} is the variable used to do background estimation and calculate the target asymmetry. This is because the χ^2_{min} method and MVA method are directly use the M_{jjb} variables to distingush physical object, it is decided by us and being artificial. The M_{jjb} is just used to validate the reconstruction and objects identification, relatively, the information-isolated variable M_{lb} is appropriate to used in the following analysis.

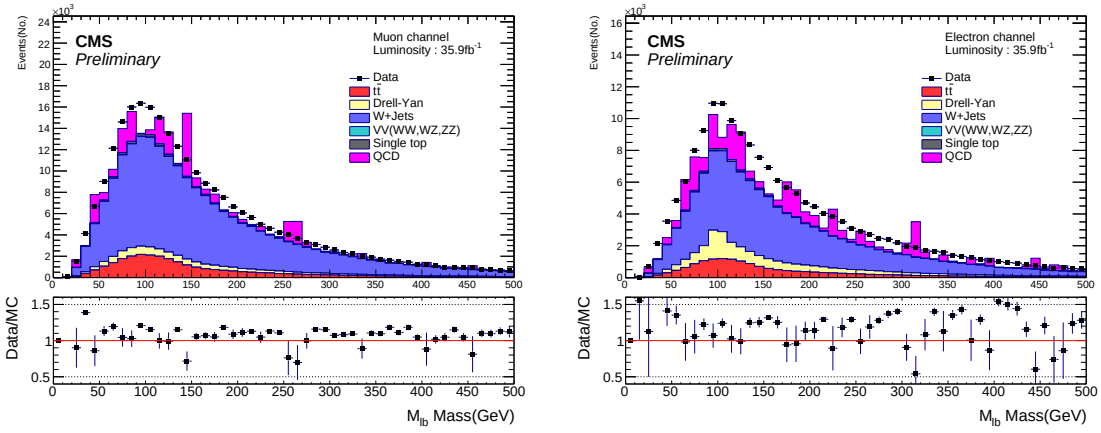
V.5 Control Region

In the calculation of asymmetry of $t\bar{t}$, there are some effect from non-$t\bar{t}$ background diving in and causing spurious result. To study the background asymmetry effects from detector and reconstruction, and to subtract these effects from data's asymmetry calculation, there are 2 **Control Region(CR)** samples introduced. Control region usually means a region which is orthogonal to signal region and also usually used to study the background in an analysis. The background estimation and study part is set in Chapter.VI. In the analysis, the first CR is a **W+jets-dominant CR** samples:

- 1 selected lepton which are a tight muon or a tight electron
- 0 lepton pass veto criteria which are loose lepton criterion **with NO isolation restriction***
- $\geq$ 4 selected jets with passing medium jet criteria
- **NO** btagged jets(deepCSV Loose) in these selected jets*
- each selected jets are isolated from the selected lepton with $\Delta R > 0.4$

The selection of b-tagged jets and veto lepton criteria is set to have difference between SR(V.2) and W+jets-dominant CR(V.5). Zero jet passes deepCSV-Loose btagged criteria may enhance the fraction of background and diminish $t\bar{t}$ contribution, which is based on difference of efficiency of $t\bar{t}$ and background passing btagged algorithm(deepCSV); The modification of veto lepton criteria to release veto isolation restriction means tighten the leptons selection. It may eliminate the possibility of some lepton-like objects are the secondary lepton comes from b-quark jet's propagating, in other words, lessen the possibility of existence of b-jets.

Besides, the reconstruction method of M_{jjb} is same with SR by χ^2_{min} method or MVA method(20 variables, MLP). However the M_{lb} is the critical variable for this analysis to do background estimation, the reconstruction of M_{lb} is necessary. In the W+jets-dominant CR sample, there are not 2 selected b jets in the beginning of reconstruction, it is 2 different criteria for χ^2_{min} method and MVA method to reconstruct M_{jjb} and M_{lb} in CR. For the χ^2_{min} method, we put in all the jets in χ^2 equation Eq.2 to pick out the combination with minimum χ^2 value as the ingredients of $M_{jjb(j)}$(one jet is seen as "b"). And for the $M_{lb(j)}$ part, there are $>= 2$ jets rest. The jet which is the most close to selected lepton by ΔR is the jet seen as the b in M_{lb}; For the MVA method, there is already leptonic b's information input to be trained for the algorithm which reconstruct M_{jjb}, which means the product algorithm from training would distinguish leptonic b automatically not just hadronic jjb, so we just input all the selected jets in the training algorithm to see when the highest MVA score occur, which one is the role of hadronic b, which one is the role of leptonic b...,etc. It is used to study background with M_{lb} variable with CR sample, so there is not M_{lb} cut for CR selection strategy. There are the M_{lb} with χ^2_{min} reconstruction of W+jets-dominant CR:



(a) Muon Channel

(b) Electron Channel

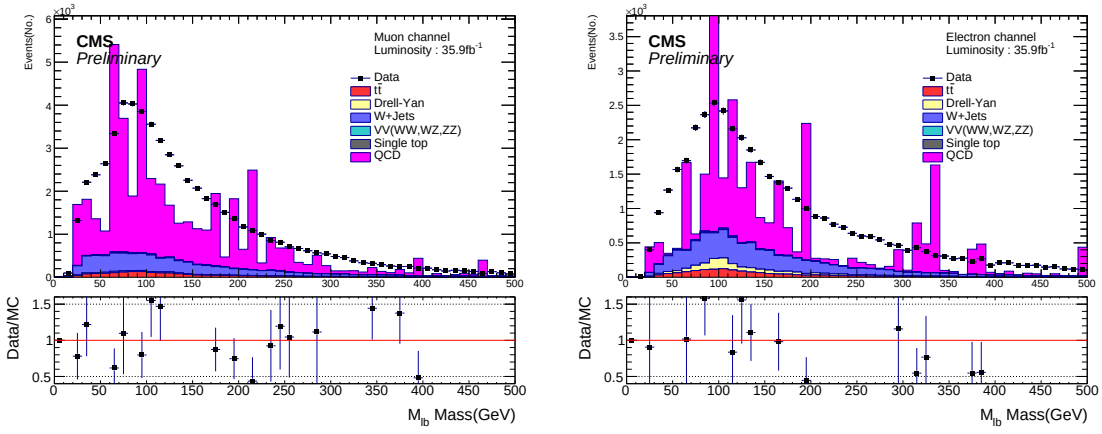
Figure 37: Data and MC comparison plots of leptonic top's invariant mass(w/ χ^2_{min}-reco) in W+jets-dominant CR

There are more spike-shape happening in QCD MC samples shown in Fig.**??**. They come from the reason that QCD has much larger cross section than other simulation samples. In this

case, if one QCD simulation event pass or fail the selection, it would be reweighed by a large value. The necessity to reweigh to luminosity leads to that one QCD simulation event's effect would cause an enormous variation showing on plots after selection. To make up this demonstration type, there is the second CR which is enriched with QCD samples:

- 1 selected lepton with tight lepton with **inverse ISOlation criteria***
- 0 lepton pass veto criteria which are loose lepton criterion **with NO isolation restriction**
- $\geq$ 4 selected jets with passing medium jet criteria
- **NO** btagged jets(deepCSV Loose) in these selected jets
- each selected jets are isolated from the selected lepton with $\Delta R > 0.4$

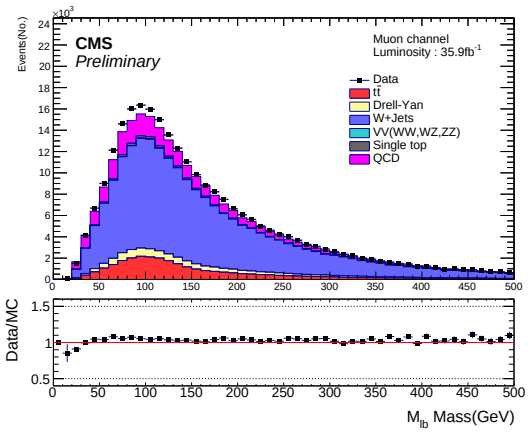
We can see that between W+jets-dominant CR and QCD-dominant CR, there is just a deviation that the selected lepton is chosed as the non-isolated one because of the property of QCD sample. And below is the Data/MC comparison of M_{lb} with χ^2_{min} reconstruction in QCD-dominant CR:



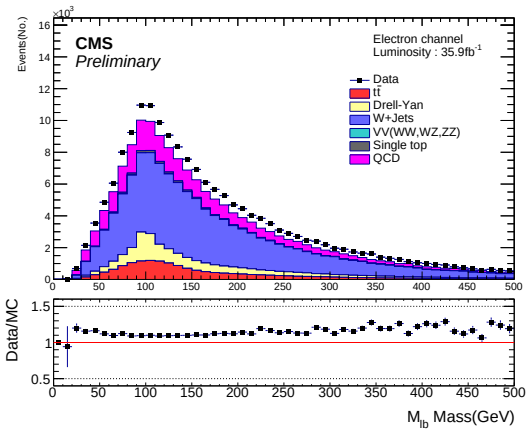
(a) Muon Channel (b) Electron Channel

Figure 38: Data and MC comparison plots of leptonic top's invariant mass(w/ χ^2_{min}-reco) in QCD-dominant CR

This QCD-dominant CR which is exactly orthoganal to W+jets-dominant CR could be studied as one of the "background" of W+jets-dominant CR. Therefore the shape of data in QCD-dominant CR may represent the shape of QCD in W+jets-dominant CR and also the number of events follows the events number of QCD MC. This method is called **Data-Driven** method for using data's characteristic to supplant simulation's. There are the Data/MC comparison of M_{lb} with χ^2_{min} reconstruction in W+jets-dominant CR after applying QCD's data-driven.



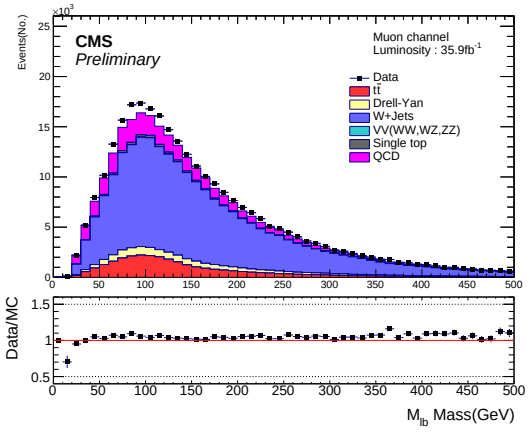
(a) Muon Channel



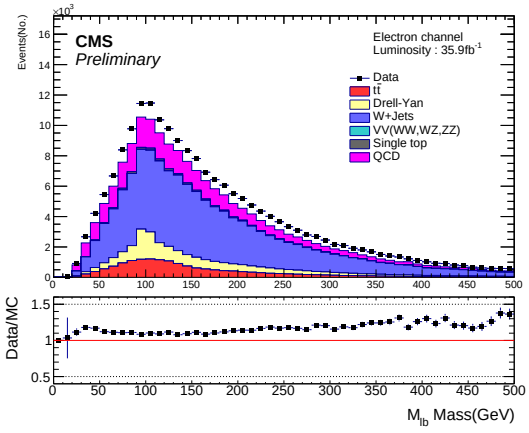
(b) Electron Channel

Figure 39: Data and MC comparison plots of leptonic top's invariant mass(w/ χ^2_{min}-reco) in W+jets-dominant CR(w/ data-driven QCD)

And the CR comparison's plots of MVA method(20 variables, MLP) are also shown here(after data-driven of QCD):



(a) Muon Channel



(b) Electron Channel

Figure 40: Data and MC comparison plots of leptonic top's invariant mass(w/ MVA-reco) in W+jets-dominant CR(w/ data-driven QCD)

Furthermore, the process ratio and composition of W+jets-dominant CR are calculated:

Table 12: Expected process ratio passing the full selection and χ^2_{min}-reconstruction in W+jets-dominant CR(w/ χ^2_{min} cut)

Process	Muon Channel (%)	Electron Channel (%)
$t\bar{t}+jets$	10.70	9.60
Z/γ^*+jets	5.27	10.86
$W+jets$	70.16	58.32
$Single\ top$	1.33	1.22
QCD	12.54	20.0

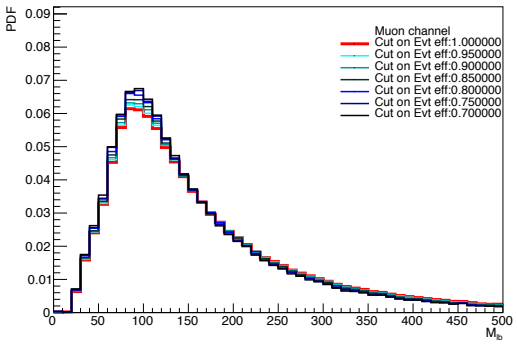
Table 13: Expected process ratio passing the full selection and MVA-reconstruction in W+jets-dominant CR(w/ MVA cut)

Process	Muon Channel (%)	Electron Channel (%)
$t\bar{t}+jets$	11.17	10.03
Z/γ^*+jets	5.26	10.82
$W+jets$	69.74	58.03
$Single\ top$	1.35	1.24
QCD	12.48	19.87

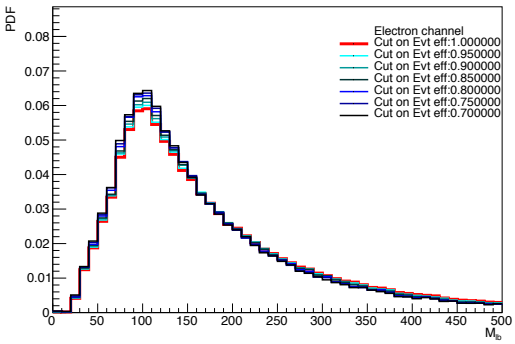
Though the non-$t\bar{t}$ background in SR is minority relative to $t\bar{t}$, there must be eliminated for the precision measurement. Because of little ratio of non-$t\bar{t}$ background, there may be much fluctuation and bias to study low statistics samples. For improvement, there is also a data-driven method to make W+jets-dominant CR's data as the role of background simulation sample in SR. It can smooth the issue of low MC statistics. And for the analysis, the W+jets-dominant CR's data is used to do template fit to estimate signal and background yields by its M_{lb} shape. At the same time, the effect of asymmetry from background is also studied with W+jets-dominant CR's data.

For the following template fit(VI.1), it is necessary to retain the isolation of information about M_{lb} when reconstructing the hadronic top mass M_{jjb}. There is a check to see if a cut on mva score is given, whether the M_{lb}'s pdf(probability density function) shape change. If the cut on mva score which is designed and trained for reconstructing M_{jjb} would vary the M_{lb} shape, there are some obvious interference from reconstructing M_{jjb} to information of M_{lb}. That is what we

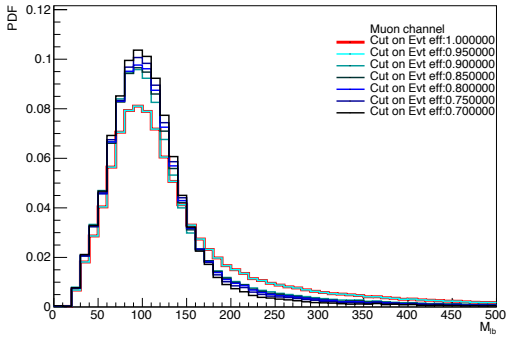
need to circumvent for following analysis strategy. There are the W+jets-dominant CR's data's M_{lb} shape with MVA score cuts below(Fig.41). The 8 variables training case obviosly vary when mva-cut applied. It is because for the 8 training variables(3), there are leptonic b and lepton's information(selected lepton and leptonic b-jet's *sum of* P_T, $\Delta\phi$, $\Delta\eta$) input for training. These informations can directly use the M_{lb}'s information to distinguish objects, making the product algorithm is M_{lb}-related. So this is the reason why the 8 variables training case was abandoned.



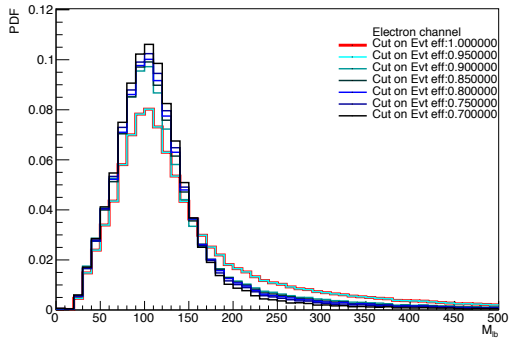
(a) 2 variables, MLP, mu-ch



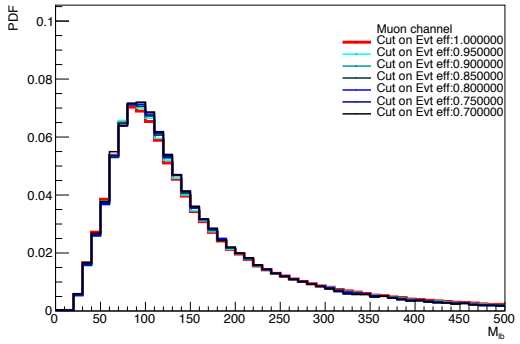
(b) 2 variables, MLP, el-ch



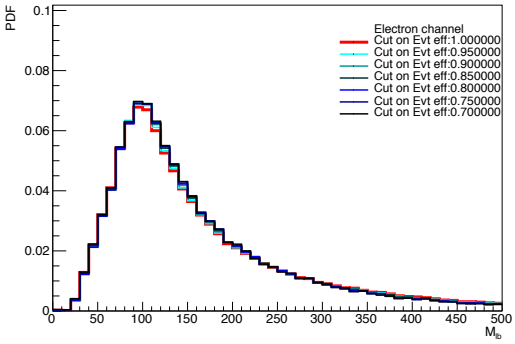
(c) 8 variables, MLP, mu-ch



(d) 8 variables, MLP, el-ch



(e) 20 variables, MLP, mu-ch



(f) 20 variables, MLP, el-ch

Figure 41: Relation between M_{lb} shape(W+jets-dominant CR's data) when cut on MVA score which leave different events efficiency

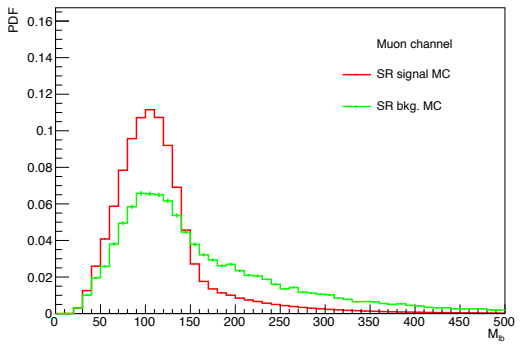
Here is an arrangement of signal region and 2 control region as a table:

VI Background Estimation

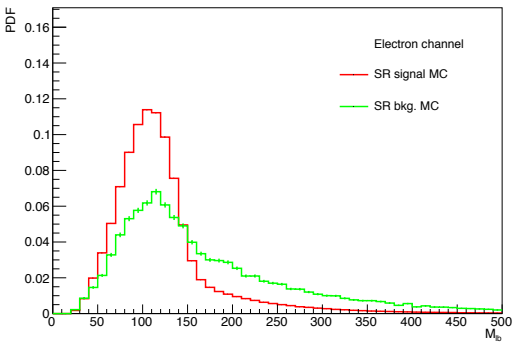
According to the simulation, the events of interests occupy diminant ratio under signal region selection(includes χ^2_{min}, MVA-A, MVA-B strategy), but also about 4∼6 % of events which came from non-$t\bar{t}$ background are collected. Even though the selected events are under high signal purity, the randomly combined background events would contaminate the final measurement of asymmetry. In fact, the minority of background may still dilute the calculating results of A'_{cp} critically in this precision measurement analysis. Thus, the background study should be attached importance to. The A'_{cp} value of $t\bar{t}$ would be extracted, at the same time, backgrounds' contribution would be cast aside.

It is a common and usual way to estimate background in data by using simulation samples. It depends more on the theoretical modeling and detector simulation with simulation sample. The background in simulation cannot be quite close to real data perfectly. This situation would disorder the asymmetry measurement eventually. It is suggested that in many high energy experiments' analysis to directly acquire background from real data. Therefore, the data-driven way is applied.

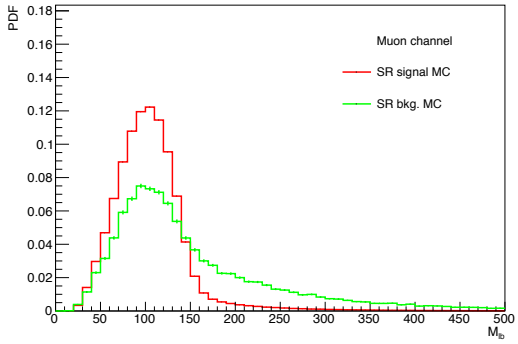
In this analysis, it is found that the M_{lb} variable shows really difference between signal and background in shape(Fig.42). Thus, with the diverse shape, there is an advantage to discern signal and background and then seperate them with template fit.



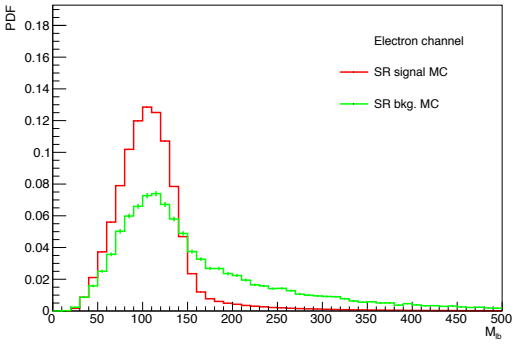
(a) χ^2_{min}, mu-ch



(b) χ^2_{min}, el-ch



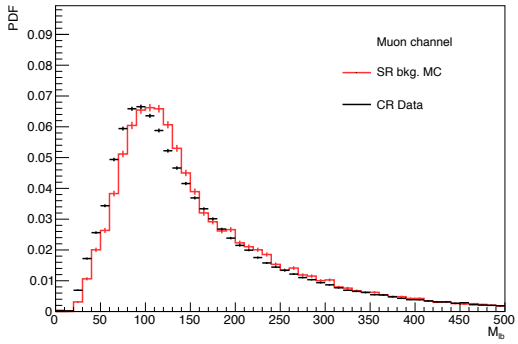
(c) MVA(20 variables, MLP), mu-ch



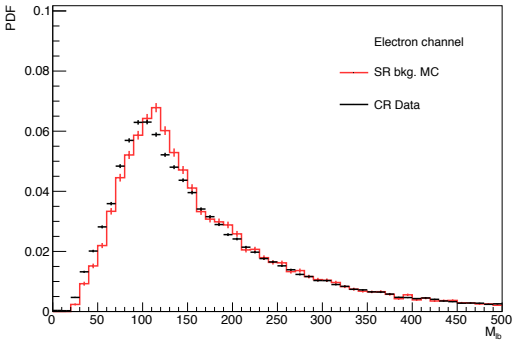
(d) MVA(20 variables, MLP), el-ch

Figure 42: Comparison of M_{lb} shape between SR's signal and background

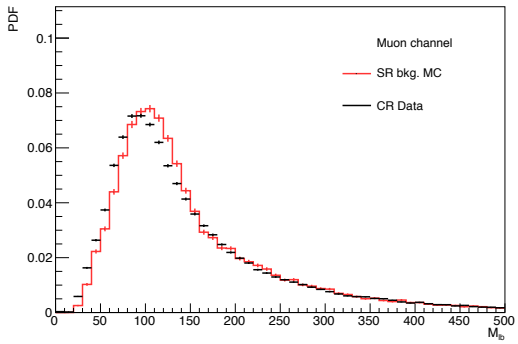
As mentioned that the W+jets-dominant CR's data is used to data-driven background in SR, it must be checked that how the similarity of M_{lb}'s pdf shape they have:



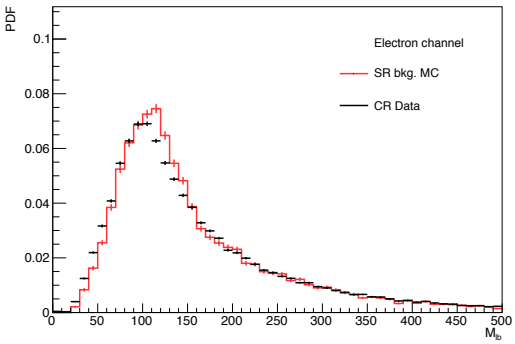
(a) χ^2_{min}, mu-ch



(b) χ^2_{min}, el-ch



(c) MVA(20 variables, MLP), mu-ch



(d) MVA(20 variables, MLP), el-ch

Figure 43: Comparison of M_{lb} shape between SR's background and CR's data

VI.1 Template Fit

The method **Template Fit** is used to estimate background yields and subtract them in this analysis. The template in this method means a pdf(probability density function) with unvaried shape under specific variable's spectrum. Also following pdf's characteristic, the template have to be normalized(integrated area of this shape is unity). This template fit is one of general fitting with **Maximum-Likelihood Method**.

With respect to Maximum-Likelihood Method, it uses the likelihood function L and minimizes the $-2\ln L$ by computing resource to get the more appropriate parameters in function we want to predict. The signal and background yields in this analysis are both the target parameters in fitting. The likelihood function comes from joint pdf of N independent observation with $\boldsymbol{X}$ which is the set of observables, the joint pdf is Eq.4.

$$P(X|\theta) = \prod_{i=1}^{N} f(X^i|\theta) \tag{4}$$

Eq.4 shows the product of all i-th observation's function and the θ is the set of parameters in the pdf. When we substitute the observed data $\boldsymbol{X^0}$ for variables $\boldsymbol{X}$, the PDF becomes the **Likelihood function L(θ)**:

$$L(\theta) = P(X^0|\theta) = p(N|\theta)\prod_{i=1}^{N} f(X^i|\theta) \tag{5}$$

There are N times measurements in likelihood function. Also, the $p(N|\theta)$ is usually a poisson distribution in most common case because of the common counting condition. In order to get the most appropriate parameters to match the observed data, the larger value of likelihood function $L(\theta)$, the more matched the parameters demonstrate. Therefore, the maximum $L(\theta)$ case should be the target for final decision of parameters. However, the modern technics are usually used to minimize some values, so it is turned to negative sign and being have to be minimized. On the other hand, it is convenient to calculate with addition instead of product, so doing the natural logarithm on likelihood function make it easier(Eq.6). In conclusion, Maximum-Likelihood method minimizes the $-2\ln L(\theta)$ and get the fitting parameters.

$$L(\theta) \implies \ln L(\theta) \quad , \prod_{i=1}^{N} f(X^i|\theta) \implies \sum_{i=1}^{N} \ln L(\theta) \tag{6}$$

In the cases that the number of observation is also the parameters and as random variables, the **extended-likelihood function** is often adopted. The extended-likelihood function includes the poisson distribution $p(N|\theta)$ of Eq.5 in. So the extended-likelihood function shows as the

pattern in Eq.7.

$$L(\theta)_{ext} = p(N|\theta)\prod_{i=1}^{N} f(X^i|\theta) = \frac{e^{-\mu(\theta)}\mu(\theta)^N}{N!}\prod_{i=1}^{N} f(X^i|\theta) \tag{7}$$

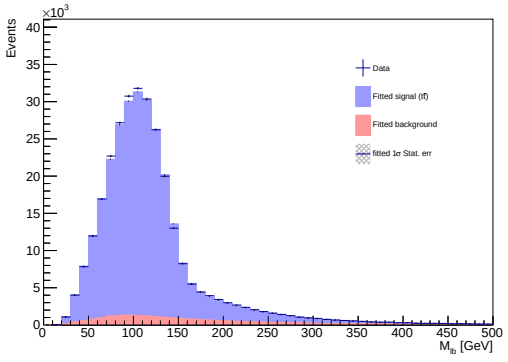
In this analysis, there are signal and background templates being pdf shapes, and the only parameters of θ set are the event numbers of signal and background(n_S, n_B). Thus the extended-likelihood function of this analysis is demonstrated as Eq.8. The f_S, f_B are the fraction of signal and background. That is to say, $f_S = n_S/(n_S + n_B)$, $f_B = n_B/(n_S + n_B)$, and the P_S and P_B are the templates of signal and background. Also, we just fit under the M_{lb} variable(oboservable), so the X^i in Eq.7 is M_{lb}.

$$\begin{aligned} L(n_S, n_B)_{ext} &= \frac{e^{-(n_S+n_B)}(n_S+n_B)^N}{N!}\prod_{i=1}^{N} f(M_{lb}^i|n_S, n_B) \\ &= \frac{e^{-(n_S+n_B)}(n_S+n_B)^N}{N!}\prod_{i=1}^{N}[f_S P_S(M_{lb}^i) + f_B P_B(M_{lb}^i)] \\ &= \frac{e^{-(n_S+n_B)}}{N!}\prod_{i=1}^{N}[n_S P_S(M_{lb}^i) + n_B P_B(M_{lb}^i)] \end{aligned} \tag{8}$$

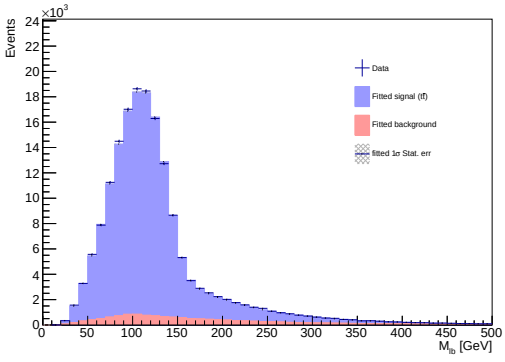
Eventually, in our analysis we use binned-likelihood function as the observed mode. In other words, the $i = 1, 2, \ldots, N$ means that input observations are the results of N bins under the observable – M_{lb} spectrum. By contrast, the unbinned-likelihood method uses event-by-event as the observation, costing too much resource and time, so it has not been used.

VI.1.1 Fitting Result

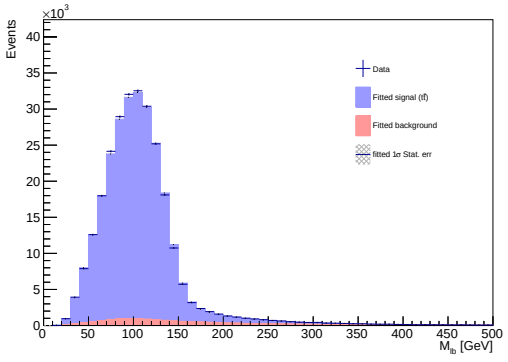
With the Maximum-Likelihood Method, there is the fitting result(Fig.44). The signal template is $t\bar{t}$ MC in SR and the background template is real data in CR, both are under M_{lb} spectrum.



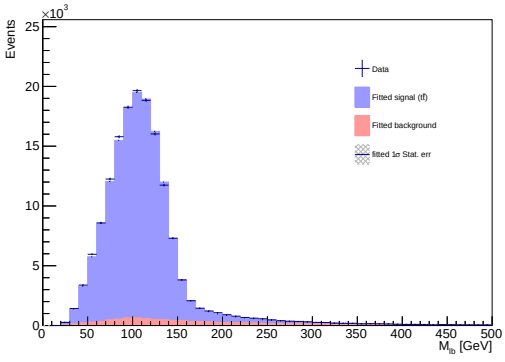
(a) χ^2_{min}, mu-ch



(b) χ^2_{min}, el-ch



(c) MVA(20 variables, MLP), mu-ch



(d) MVA(20 variables, MLP), el-ch

Figure 44: Results of template fit, with fitted yields ccomparing data and MC.(χ^2, MVA methods are both shown)

Futhermore, the fitted yields of signal($t\bar{t}$) and background are shown under χ^2 and MVA cases(Table.14, 15, 16).

Table 14: Data and MC estimated events number after template fit(w/ χ^2_{min}, M_{lb} cut)

	Muon Channel	Electron Channel
Data	243790	136151
Estimated $t\bar{t}$	230766	135600
Estimated background	12146.9	5537.45

Table 15: Data and MC estimated events number after template fit(w/ MVA-A strategy)

	Muon Channel	Electron Channel
Data	245436	139432
Estimated $t\bar{t}$	236226	133894
Estimated background	8520.95	5102.1

Table 16: Data and MC estimated events number after template fit(w/ MVA-B strategy)

	Muon Channel	Electron Channel
Data	272583	157510
Estimated $t\bar{t}$	256196	146656
Estimated background	15479	10217.6

VI.1.2 Goodness of Fit

It is necessary to check whether it is fine to use the fitting result. There is a common method to test the goodness of fit – $\boldsymbol{\chi}^2$ **test**. In a set of 2 data histogram, we use the χ^2 value between them to see the degree they match to each other. In general case, χ^2 value are the sum of N independent normal variables distribution, shown below(Eq.9). The $\chi^2(N)$ and $P(x)$ in Eq.9 is also a distribution called χ^2 distribution. The only parameter of χ^2 distribution is the N which also means *degree of freedom*. Degree of freedom means the number of independent terms in one distribution which affect the variation of this distribution. The basic concept to apply χ^2 to test goodness of fit is – to see all the N independent variables in histogram as normal distribution and operate them to be standard, and then sum all of them to compare with the corresponding $\chi^2(N)$ distribution. Commonly, we see where the derived χ^2 value drop on the corresponding χ^2 distribution and take the p-value as the criteria. If p-value is large, and this means one data distribution is not so match to the other one, and vice versa. For simlified one criteria, if 2 set of data points are well match to each, the derived χ^2 value should be close to expectation value

of corresponding $\chi^2(N)$ distribution – N.

$$
\begin{aligned}
Q = \sum_{i=1}^{N} X_i^2 \implies \chi^2(N) \\
\chi^2(N) \implies P(x) = \frac{\frac{1}{2}(\frac{x}{2})^{\frac{N}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{N}{2})} \\
Exp < P(x) >= N \ \ (D.o.F)
\end{aligned}
\tag{9}
$$

For this analysis, to compare the 2 histogram under 1 variable's spectrum, it is usually taken the events number in any bins as independent variables. There are 1 set of real data points and 1 set of MC points which follows the fitting result. And then, to standardize all the bins' variables to be normal distribution variables, we operate them as Eq.10's first equation. After standardization, we just sum up all the bins' operated variables to be the χ^2 value of these 2 set of data points (Eq.10-2). The x_i^d means the events number of real data in i-th bin, x_i^s is the events number of MC in i-th bin. Also, the standard error for standardization is poisson error of counting events number of MC in i-th bin – $\sqrt{x_i^s}$:

$$
\begin{aligned}
X_i = \frac{x_i^d - x_i^s}{\sqrt{x_i^s}} \\
Q = \sum_{i=1}^{N} X_i^2 = \sum_{i=1}^{N} \frac{(x_i^d - x_i^s)^2}{(\sqrt{x_i^s})^2}
\end{aligned}
\tag{10}
$$

In this case, the bins range under M_{lb} is from 0GeV to 500GeV with 50 bins. The parameters n_S, n_B in fitting give these data points 2 constraint, so the degree of freedom of these data points are (50-2)=48. So with the simple method, there are the comparison between derived χ^2 value(Q) to the expectation value 48 $\longrightarrow \frac{Q}{48}$, there are the goodness of fit under χ^2_{min} and MVA reconstruction strategy.

Table 17: $\frac{Q}{D.o.F}$ of fitting results (χ^2_{min} and MVA strategy)

	Muon Channel	Electron Channel
χ^2_{min}	2.54258	1.39517
MVA	2.52261	2.15768

We can see that in electron channel we have better fitting demonstration than in muon

channel. However, it is fine to have these goodness-of-fit results under all cases.

VI.2 Background subtraction

The final target of background subtraction is using the fitted yields results to subtract the background contribution in A'_{cp}. The process are shown below as a illustration diagram(Fig.45). First of all, we use signal and background templates to do the template fit and get fitted yields N_{data} and N_{bkg} which both include positive and negative value of specific observable. Secondly, subtract both the positive and negative observable's yields of background from data's part, and there are the effective positive and negative observable's yields $N(O>0)_eff.$ and $N(O<0)_eff.$. Therefore, the final A'_{cp} can be calculated by effective yields which are expected purely contributed by signal($t\bar{t}$).

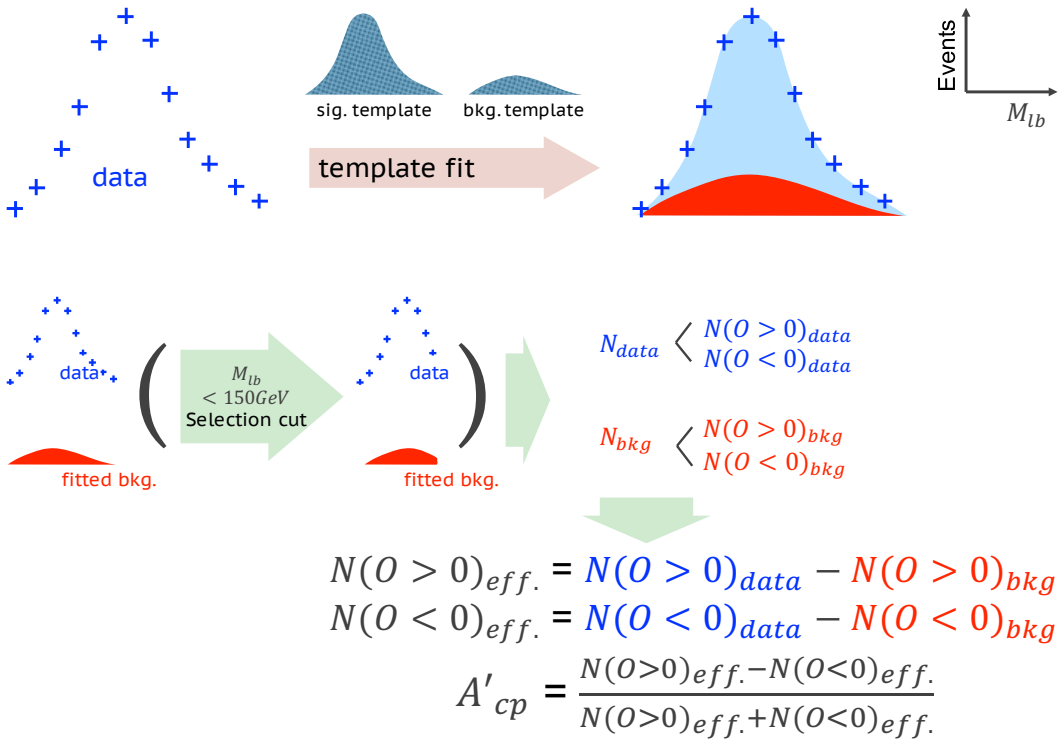


Figure 45: Process of background subtraction

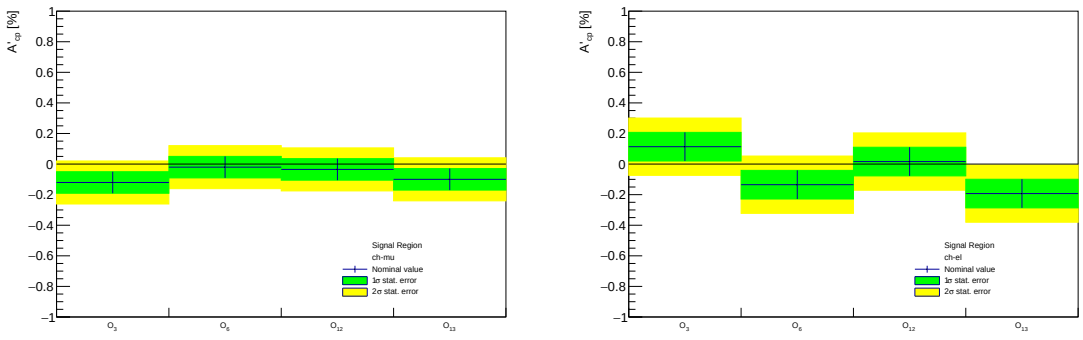
VII Asymmetry Bias

In the standard model(SM), the A_{cp} of $t\bar{t}$ and other background are expected to be zero, that is to say, there is no asymmetry. However, the A'_{cp} value may be affected by the reconstructing and detector demonstrating, and also the particles' characteristic may interact with detector, so the A'_{cp} might be different from the real A_{cp}. And the reconstruction effects from the detector can be studied by simulation sample. Also, the detector and reconstruction bias is expected to be the major systematic uncertainty of the analysis.

The best way to estimate the detector and reconstruction issue is to see the A'_{cp} bias of a sample which supposes to be zero in A_{cp}. And then, the standard deviation(error) of A'_{cp} is taken as the detector and reconstruction bias. The sample needs to be irrelavent with the target signal of this analysis – $t\bar{t}$ because we expect the possible source of asymmetry is in $t\bar{t}$. Accordingly, the SR background sample is appropriate to be studied for the detector and reconstruction bias since it is supposed to be without A_{cp} and it is nothing relative to the signal in the analysis($t\bar{t}$). However, the SR background sample has low statistics. Therefore, the method in 8TeV analysis is using the W+jets-dominant CR data sample as the sample supposed to be without A_{cp}, because the W+jets CR data sample is used to data-driven the SR background sample. Yet, in this analysis, we use the analyzed data itself and a fake data method to doing this study. It is advatage of using SR data itself that the statistics is more than CR data and that the W+jets-dominant CR smaple still includes over 10% $t\bar{t}$. And the chapter is to introduce the fake data method to calculate the detector and reconstruction bias.

VII.1 Asymmetry in simulation sample

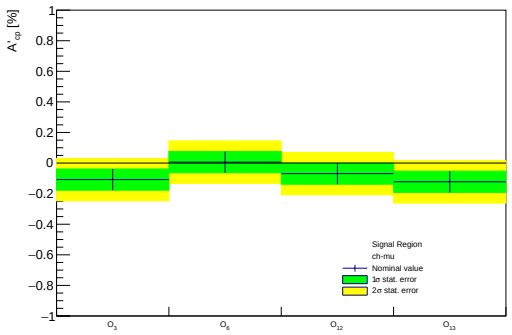
There are the A'_{cp} of simulated $t\bar{t}$ sample in SR with 3 different reconstruction results(χ^2_{min}, MVA-A, MVA-B). It is expected to have no asymmetry because of the simulation sample is generated based on standard model(SM).



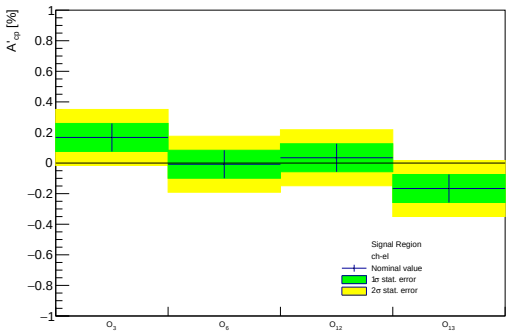
(a) Muon channel

(b) Electron channel

Figure 46: A'_{cp} of simulated signal sample($t\bar{t}$) by χ^2_{min} reconstruction strategy

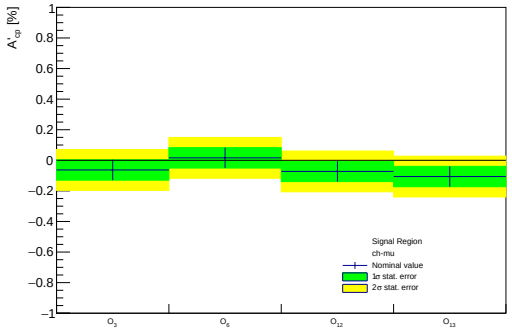


(a) Muon channel

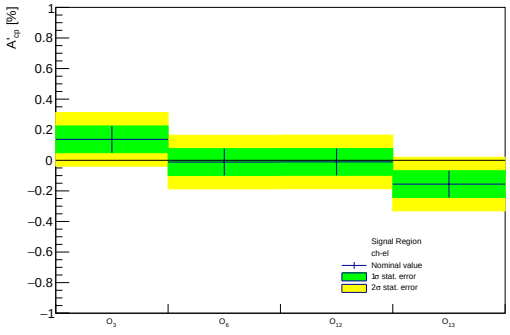


(b) Electron channel

Figure 47: A'_{cp} of simulated signal sample($t\bar{t}$) by MVA-A reconstruction strategy



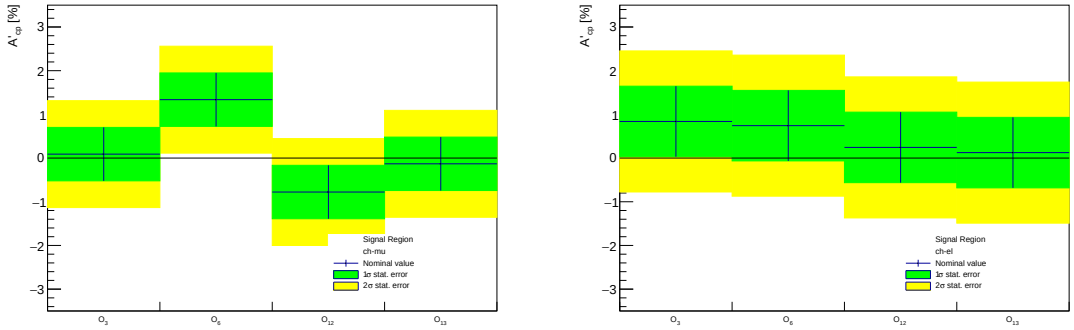
(a) Muon channel



(b) Electron channel

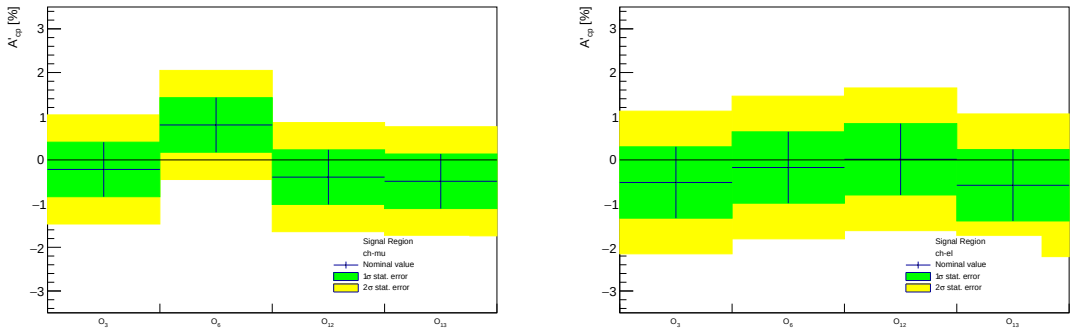
Figure 48: A'_{cp} of simulated signal sample($t\bar{t}$) by MVA-B reconstruction strategy

Futhermore, as previous mentioning, the A'_{cp} of simulated background sample in SR is participated as a fine sample to check the detector and reconstruction bias. There are the checks whether they are has no asymmetry.



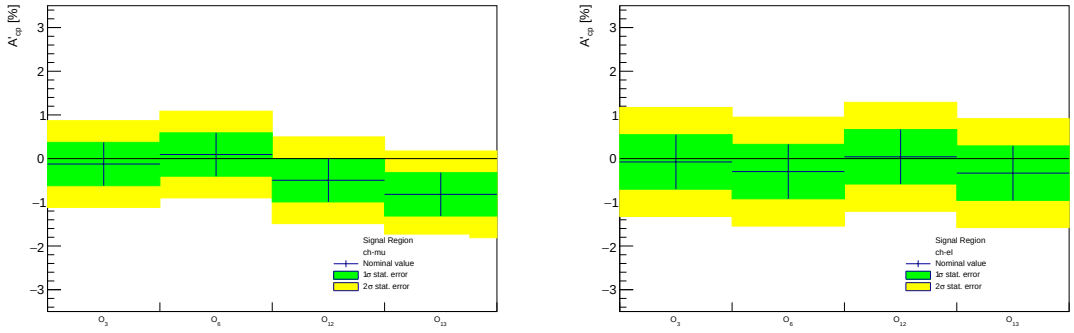
(a) Muon channel (b) Electron channel

Figure 49: A'_{cp} of simulated background sample by χ^2_{min} reconstruction strategy



(a) Muon channel (b) Electron channel

Figure 50: A'_{cp} of simulated background sample by MVA-A reconstruction strategy



(a) Muon channel (b) Electron channel

Figure 51: A'_{cp} of simulated background sample by MVA-B reconstruction strategy

The asymmetry of simulated signal and background sample in three reconstruction strategies are all in 2 statistical uncertainty and are regarded confirming without asymmetry.

VII.2 Events mixing method

Using the real data of SR itself to make some fake-data sets can improve the measurerment of detector and reconstruction bias as systematic uncertainty. The method to make fake-data sets is called **events mixing method**. Events mixing method may discharge the effect of asymmetry of observables according to the structure of these triple-product observables.

For the detail, the events mixing method is to pass both b-jet's and the leading light jet(j_1)'s four momentum to the b-jet's and j_1's four momentum of the next i-th event. Therefore, a chosen i gave a set of fake-data. For example, the 2nd set of fake-data is that, the first event's informations are passed to the third event, the second event's informations are passed to the fourth event, and the third event's informations are passed to the fifth event, etc. There is a reference graph of event mixing method(Fig.52).

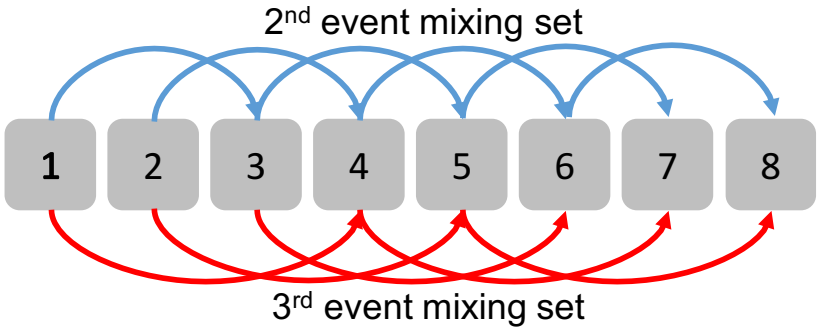
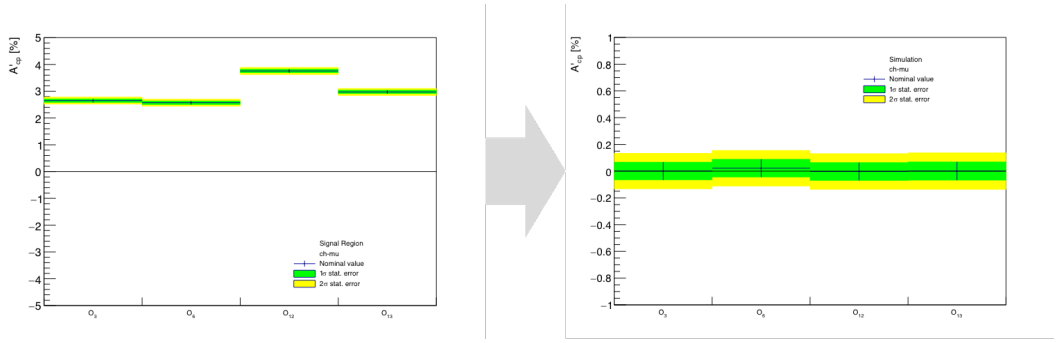
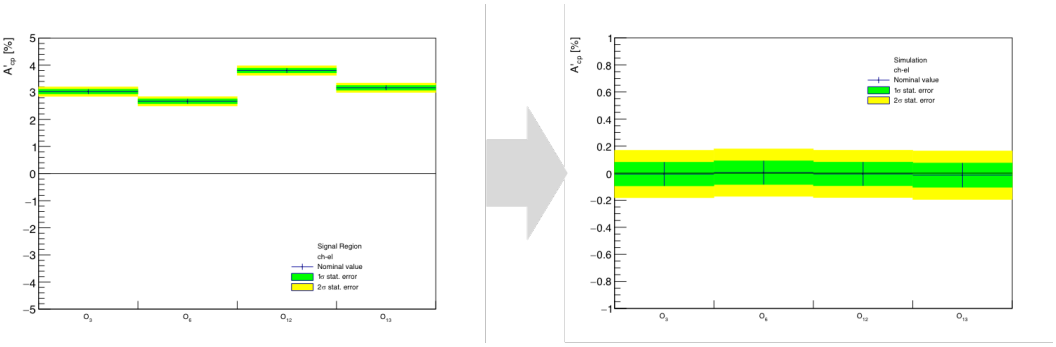


Figure 52: Illustration graph of event mixing method

Before applying this method on real data to make fake data, it is needed to check whether the events mixing method really mix the information of particle kinematics and really cancel out the intrinsic asymmetry factor. To check the canceling out effect, there is an artificial generated A'_{cp} sample prepared by duplicating random positive events with given probability. Then apply the event mixing result on this sample, the results are shown below(Fig.53), and we can confirm that the method would really cancel out the A'_{cp}.



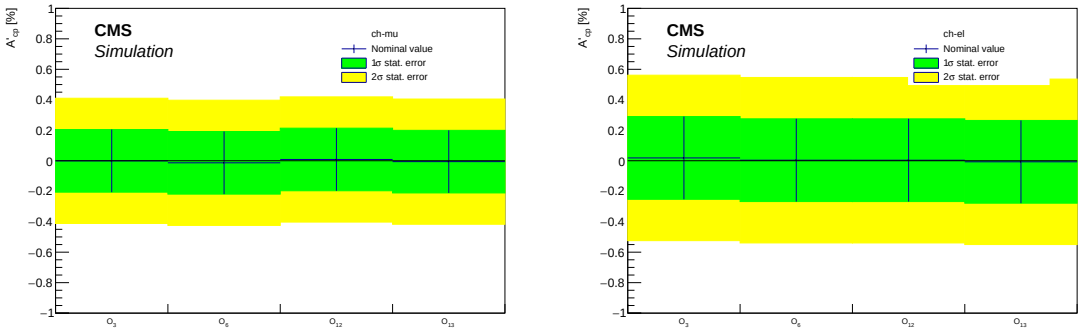
(a) Muon channel



(b) Electron channel

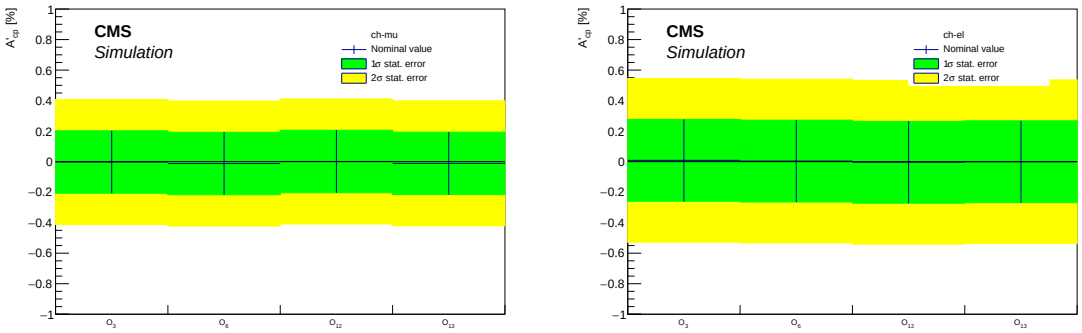
Figure 53:

And then to get the reconstruction and detector bias, we generate 1000 sets of fake-data samples.Then, we measure their A'_{cp}, and average the 1000 results mean values and standard deviations. The derived standard deviations of 4 observables are considered as the detector and reconstruction bias also the major systematic uncertainty of the 4 observables' final results. There are the plots of reconstruction and detector bias with event mixing method under three reconstruction strategies(χ^2_{min}, MVA-A, MVA-B):



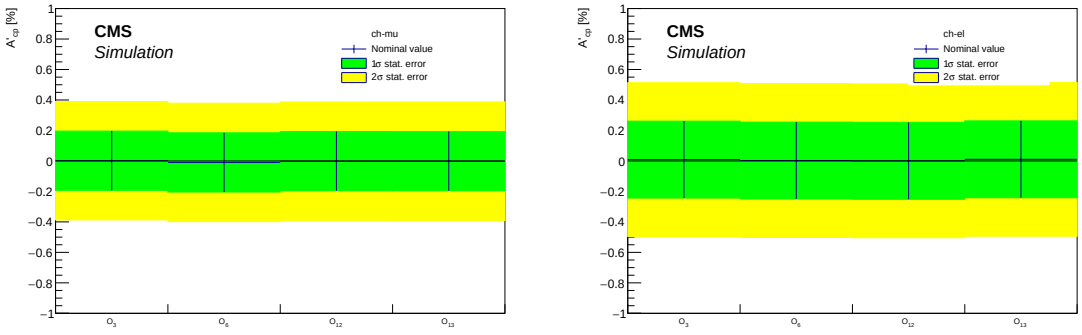
(a) Muon channel (b) Electron channel

Figure 54: average A'_{cp} of 1000 sets fake-data by χ^2_{min} reconstruction strategy



(a) Muon channel (b) Electron channel

Figure 55: average A'_{cp} of 1000 sets fake-data by MVA-A reconstruction strategy



(a) Muon channel (b) Electron channel

Figure 56: average A'_{cp} of 1000 sets fake-data by MVA-B reconstruction strategy

Table 18: Reconstruction and detector bias of χ^2_{min} strategy

[%]	O_3	O_6	O_{12}	O_{14}
Muon channel	± 0.205	± 0.205	± 0.205	± 0.205
Electron channel	± 0.271	± 0.271	± 0.271	± 0.271

Table 19: Reconstruction and detector bias of MVA-A strategy

[%]	O_3	O_6	O_{12}	O_{14}
Muon channel	± 0.204	± 0.204	± 0.204	± 0.204
Electron channel	± 0.268	± 0.268	± 0.268	± 0.268

Table 20: Reconstruction and detector bias of MVA-B strategy

[%]	O_3	O_6	O_{12}	O_{14}
Muon channel	± 0.194	± 0.194	± 0.194	± 0.194
Electron channel	± 0.252	± 0.252	± 0.252	± 0.252

VIII Systematic Uncertainty

Besides the statistical uncertainty, there are also some systematic uncertainty from detector's issue, reconstruction procedure, simulating process...etc. In this snalysis, the primary systematic uncertainties are from the simulation correction.

VIII.1 Introduction of systematic uncertainty

VIII.1.1 Pileup re-weighting

Since the pileup performance under data and simulation(MC) have discrepancy, there is reweighing of pileup in any events which is mentioned in section.IV.3.1. However, there is also an uncertainty of pileup reweight. The pp inelastic scattering cross section is used to calculate the pileup distribution and then used to implement the pileup reweight, so this cross section has its mean value and uncertainty $69.2mb^{-1}$ $\pm 5\%$. Then the pileup reweighing factor may be affected by this cross section as the pileup reweighting systematic uncertainty.

VIII.1.2 Jet energy correction and resolution

VIII.1.3 b-tagged scale factor

VIII.1.4 Lepton identification, isolation, reconstruction and trigger efficiencies

eee

VIII.2 Implementation and results

The approaches to apply the systematic uncertainty are shown below. First, we

IX CP Asymmetry in Top quarks pair and corresponding observables/models

IX.1 Models

IX.1.1 CEDM model

IX.1.2 2HDM model

IX.2 Observables

IX.2.1 Classification

IX.2.2 Observable property

X Results and Conclusion

References