Analysis of Higgs production in the VBF-VH channel at the LHC

Chaitanya Paranjape - IIT(ISM) Dhanbad, India

Abstract

This is a technical project report on the analysis of Higgs production in association with a vector boson through vector boson fusion at the LHC. The concerned process is particularly sensitive to the Higgs couplings to vector bosons (κ_W, κ_Z) and exhibits Quantum interference effects between them. In this work, we propose a analysis strategy to search for this rare process with $\mathcal{L} = 3000\ fb^{-1}$, anticipating the high luminosity LHC upgrades. We present contour plots over (κ_W, κ_Z) based upon the significance estimation of our analysis, which can help to conclusively rule out the $(\kappa_W, \kappa_Z) = \pm(1, -1)$ point with more than 95% CL.

Contents

1 Introduction

The 4 significant higgs production channels are namely gluon fusion, vector boson fusion, higgs production in association with a vector boson (higgs-strahlung) and higgs production in association with $t\bar{t}$. The Feynman diagrams for the corresponding processes are given in figure 1(left) and the total cross-section measurements for the corresponding channels are given in figure 1(right).

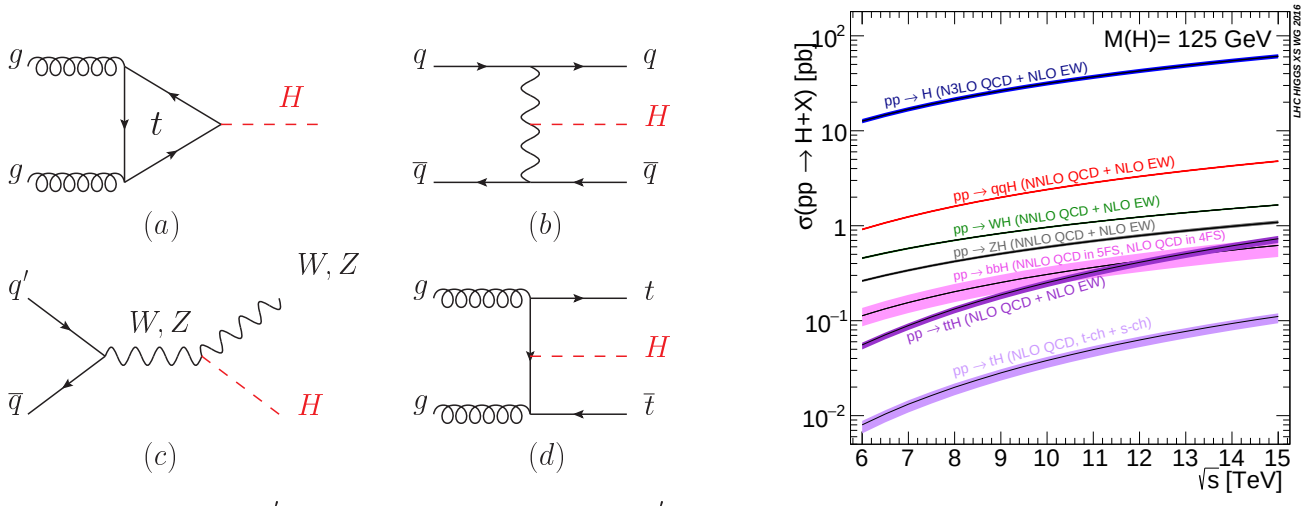

Figure 1: Feynman diagrams for main higgs production channels (left) and their corresponding cross-sections (right) [1].

Because of their dominant contribution, cross-section measurement for these prominent production channels have been analysed carefully and compared with the SM predictions for various decay modes. The recent ATLAS data for the same can be observed in figure 2.

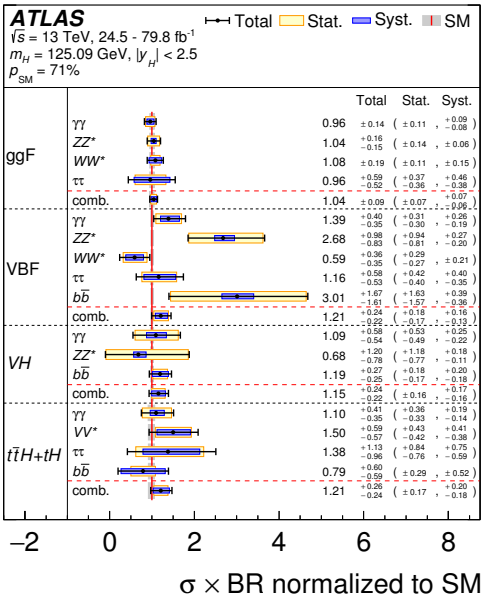

Figure 2: Net cross-sections for various production and decay modes of higgs, normalized to SM predictions [2].

Almost all of the concrete data obtained on higgs so far, has been studied through these dominant higgs production channels and some others. There are still many higgs production channels which haven't been studied at the LHC so far, owing to their small cross-sections. With the high-luminosity LHC upgrades, it would be possible to probe the much rarer higgs processes which could be key in testing the higgs properties with the SM predictions. Even though the discovery of higgs was one of greatest achievements of Standard model, there are still numerous higgs couplings which need to be tested against the SM predictions , to either confirm the SM predictions or hint towards new physics. With the data at hand, we have only been able to set some finite bounds on the observed higgs parameters. The recent ATLAS data for the Higgs couplings to fermions κ_F and to vector bosons κ_V can be viewed in 3.

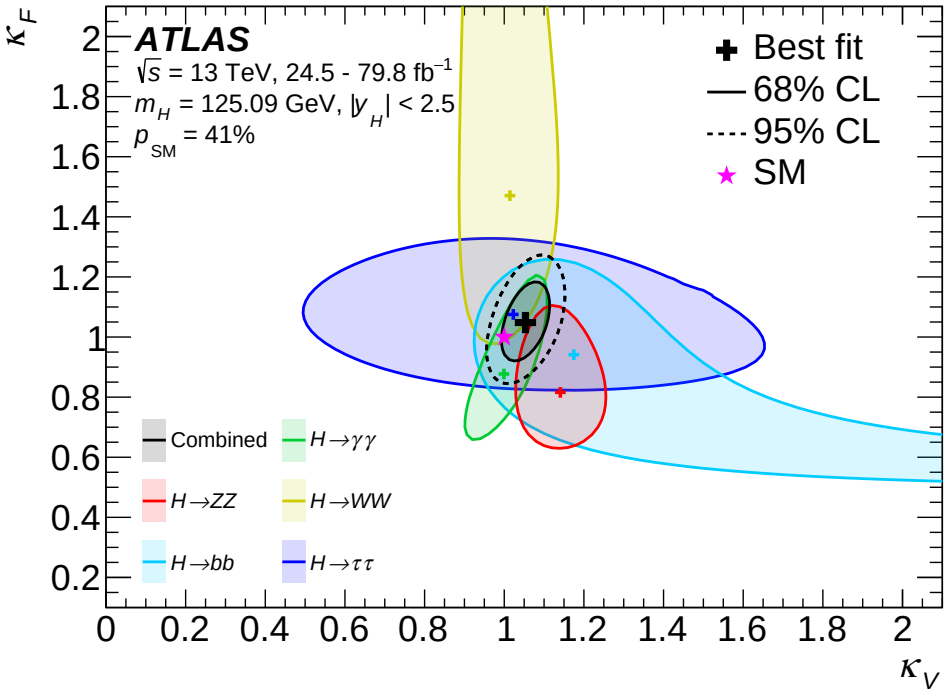

Figure 3: Negative log-likelihood contours at 68% and 95% CL in the (κ_V, κ_F) plane for the individual decay modes and their combination (in black) [2].

Therefore, anticipating the LHC upgrades, we will now focus on analyzing a rare higgs process, which could help us test the SM predictions for some of the Higgs couplings. The process we will consider is higgs production in association with Z boson thorugh Vector Boson Fusion. We believe the process to be sensitive with respect to the Higgs couplings to W and Z bosons and possess Quantum inteference effects, based upon a similar type of study previously performed in [3]. Therefore, to probe these couplings, we will try to devise an analysis strategy to search for this process at the LHC. In order to look for deviations from SM, we will parametrize the Higgs(h) couplings to W (κ_W) and Z(κ_Z) in SM as follows :

$$\mathcal{L} = h\Big(\kappa_W g m_W W_\mu^+ W_\mu^- + \kappa_Z g \frac{m_Z^2}{2m_W} Z_\mu Z_\mu\Big)$$

where g is SU(2) gauge coupling. In the standard model, $\kappa_W = \kappa_Z = 1$.

Let's investigate cross-section and detector response for such process in detail now.

2 MadGraph5-Pythia-Delphes Simulation framework

Before setting up the MG5-Pythia-Delphes pipeline, we need to install an appropriate version of ROOT [4] compatible with the Delphes version. It is advisable to have a Python version 3.7 or higher, even though version 2.7 or higher also work. Source file for a specific ROOT version can be downloaded from [5] and then one can proceed to compile ROOT from the source file. Make sure to have an appropriate version of `cmake` or `gcc` for this purpose.

The MadGraph5 homepage is [6] from where one can navigate to download the tar file for an appropriate version. We will be using the version `MG5_aMC_v2_9_4` [7]. After un-tarring the file, we can launch the MG5 interface by executing following command in the `MG5_aMC_v2_9_4` folder. (We call it simply the MG5 folder)

```
./bin/mg5_aMC
```

To download the `pythia-pgs` package, simply execute following command in the MG5 interface.

```
install pythia-pgs
```

We will be using the `Pythia6` [8] for hadronization with the `pythia-pgs 2.4.5` [9] package. Now, to install Delphes [10], simply execute following command in the MG5 interface.

```
install Delphes
```

We will be using the `Delphes 3.4.2` [11] version, which we have found to be compatible with the `ROOT 6.12.06` that we use. If any problem occurred with the installation, please make sure to have an appropriate version of `gcc` and `ROOT` installed, which could be the problem.

For a complete tutorial on getting started with this framework, please follow [12] .

To use this framework efficiently, we can script the entire simulation procedure as per our convenience and directly execute those scripts. A script with name `test_script.txt` can be executed by typing the following command in the MG5 folder :

```
./bin/mg5_aMC PATH/TO/test_script.txt
```

It is convenient to store all your scripts in a certain folder `mg5_scripts` which is inside the MG5 folder, such that the corresponding script can be executed by follwing command :

```
./bin/mg5_aMC mg5_scripts/test_script.txt
```

Scripts just contain the commands to automate the simulation process instead of entering the commands one by one into the MG5 interface.

The General structure of a script with corresponding functionalities :

```
import model model_name                #Import the model model_name from MG5/models
generate process_equation              #Generate all files for simulation of given process
output process_name                    #Store the generated files in process_name folder
launch process_name -n run_name                    #Launch simulation of the process
shower = Pythia6                                   #Use Pythia6 for Shower simulation
detector= Delphes                                  #Use Delphes for detector simulation
set nevents 10000                      #Change the number of events generated to 10000
set var_name value                     #Change values of run_card or param_card variables
...
...
```

Therefore, scripts can be used to simulate efficiently over various conditions, without having to manually enter the commands one by one. One can just start start execution of a script and wait for it to complete, to view the results.

With the MadGraph5-Pythia-Delphes framework in place, we need our κ_W and κ_Z model to analyze the processes.

3 Model with κ_W, κ_Z

First, we need to devise a model by parameterizing the Higgs couplings to vector bosons as κ_W & κ_Z. We will employ `FeynRules 2.3.32` [13] for this purpose.

Let us build up on the Standard Model description file `SM.fr` for our purposes. In order to introduce the κ_W & κ_Z as external parameters, we include the following definitions in the model parameters list `M$Parameters` .

```
   kZ == {
ParameterType -> External,
BlockName -> KAPPABLOCK,
OrderBlock -> 1,
Value -> 1,
TeX -> Subscript[[Kappa], Z],
Description -> "kappa Z"
},
kW == {
ParameterType -> External,
BlockName -> KAPPABLOCK,
OrderBlock -> 2,
Value -> 1,
TeX -> Subscript[[Kappa], W],
Description -> "kappa W"
}
```

And now, we need to modify the SM Lagrangian such that the terms originating from the Higgs to vector boson couplings are proportional to κ_W & κ_Z, and retrieve us back the SM Lagrangian at $\kappa_W = \kappa_Z = 1$. We do so by adding an extra term to the SM Lagrangian ($\mathcal{L}_{SM}$) as follows :

$$\mathcal{L} = \mathcal{L}_{SM} + h\Big\{(\kappa_W - 1)g m_W W_\mu^+ W_\mu^- + (\kappa_Z - 1)g\frac{m_Z^2}{2m_W} Z_\mu Z_\mu\Big\}$$

Where g is SU(2) gauge coupling and In the SM, $\kappa_W = \kappa_Z = 1$.

The above definition of Lagrangian ($\mathcal{L}$) is encoded into the model description file `SM.fr` and the model is exported in UFO (Universal FeynRules Output) [14] format. The UFO model files can be readily used to import into `MadGraph5` for simulation purposes.

4 VBF-ZH Process

The signal process we consider will be the production of Higgs (H) in association with a Z boson, occurring through Vector Boson Fusion. This is one of the few VBF-VH channels that we can experiment with. Concerned process is expected to be sensitive to κ_W, κ_Z and should also exhibit tree level Quantum Interference effects [3] between the two couplings. As a result, the cross-section in terms of the κ_W, κ_Z couplings can be written as follows :

$$a \cdot \kappa_W^2 \ + b \cdot \kappa_W \kappa_Z + c \cdot \kappa_Z^2$$

Where a, b, c are some constants which can be fixed empirically.

It can be noted that the cross-section for this process, in the SM, is of the order ~ 10 fb , which already makes it hard enough to detect. The Higgs to $b\bar{b}$ decay mode may give rise to large backgrounds. However, we must try to diminish their contribution, because the other decay modes have very low branching ratios leading to negligible event yields. Therefore, in order to best our chances to detect this process, we will consider the prominent decay mode $H \mapsto b\bar{b},\ Z \mapsto l^-\ l^+$ [1] .

Therefore, the signal process we simulate in the MadGraph5 is :

```
signal :  p p > z h j j QCD=0, h > b b~, z > l- l+
```

Few of the many Feynman diagrams for the VBF-ZH channel are :

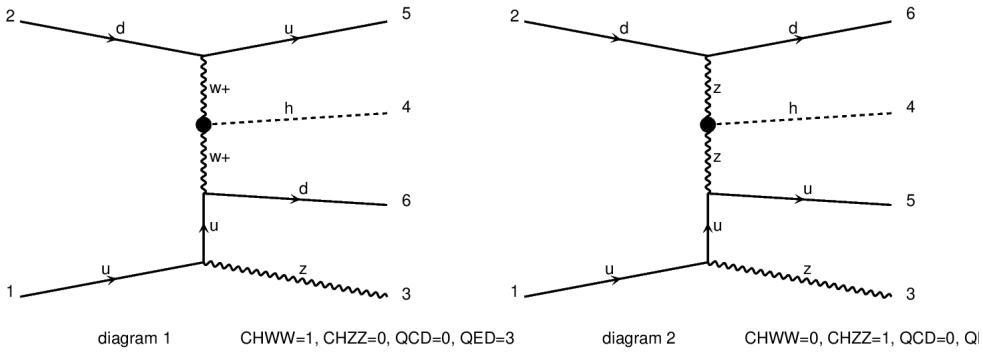

Figure 4: VBF-ZH Feynman diagrams.

The two quarks releasing the vector bosons go on to form the jets `j j` in the final state which we regard as the "VBF-Tagging Jets". These Jets exhibit a little deviation from the beam line and are detected as a forward-backward jet pair, which acts a distinctive feature to identify the VBF processes. We will deal with these distinctive characteristics in detail during the analysis part.

Notice the importance of the `QCD=0` flag in the process declaration. It instructs MG5 to generate diagrams with no gluon couplings. At higher energies, this more or less ensures that the process occurs through Vector Boson Fusion. Therefore, the same process can occur through some non-VBF channel, which we are not concerned with. Thus, we will regard such process as the first background process that we will have to deal with.

```
bcg1 :  p p > z h j j , h > b b~, z > l- l+
```

[1] $l^\pm = e^\pm$, $\mu^\pm$ and $j = g, u, c, s, d, \bar{u}, \bar{c}, \bar{s}, \bar{d}$.

Now, let us investigate how the signal $\sim$ VBF process (in blue) behaves with respect to the κ_W, κ_Z couplings. Also observe the behaviour of bcg1 process (in red).

We will plot the cross-section (in fb) for our processes by varying κ_W, κ_Z accordingly, at 13 TeV Center of mass energy.

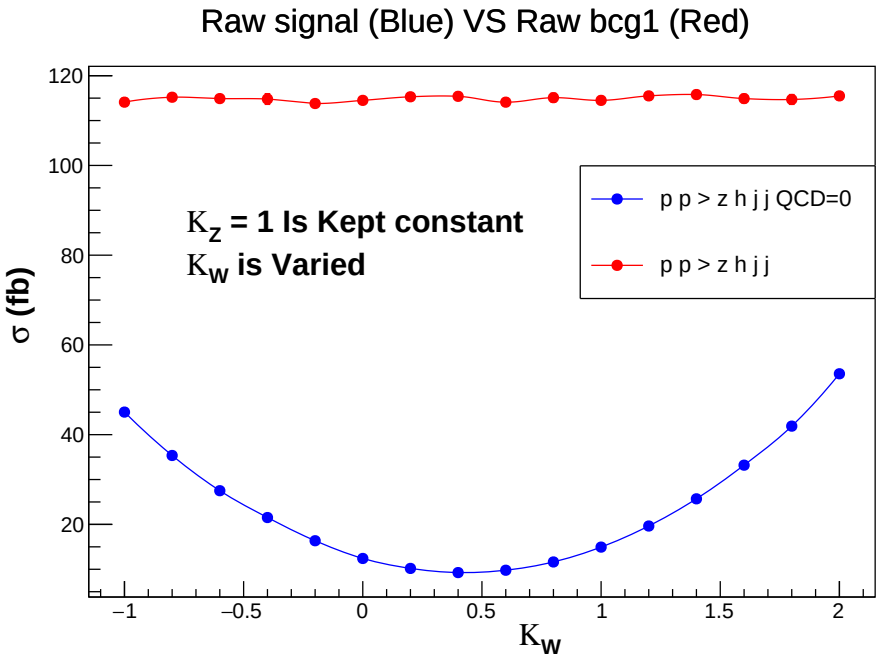

Figure 5: Signal and bcg1 dependence on κ_W.

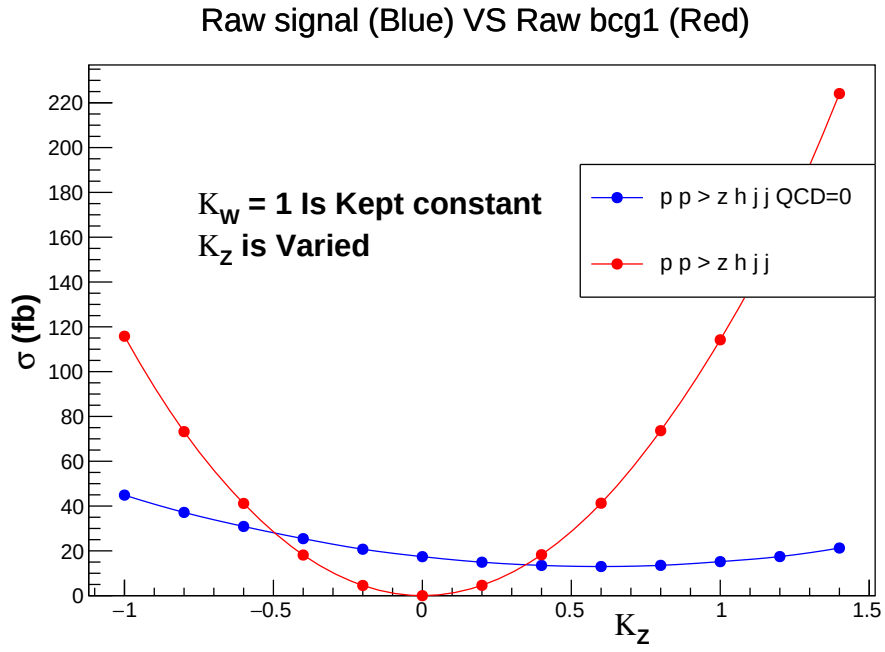

Figure 6: Signal and bcg1 dependence on κ_Z.

After a careful interpretation of the above graphs, we can corroborate our claims about the signal process possessing Quantum interference effects between the κ_W and κ_Z couplings. On the other hand, the bcg1 depends only on the κ_Z. Additionally as a rough estimate, we can model the cross-sections for the raw signal (C_1) and raw bcg1 (C_2) process as follows :

$$C_1 \sim (17.41 \text{ fb}) \cdot \kappa_W^2 - (14.755 \text{ fb}) \cdot \kappa_W \kappa_Z + (12.41 \text{ fb}) \cdot \kappa_Z^2, \qquad C_2 \sim (114.2 \text{ fb}) \cdot \kappa_Z^2$$

5 Background Processes

After fixating on the signal process to observe, we need to account for the background processes that can mimic the signal process characteristics to some extent, and possibly adulterate our observation data. First, we need to decide on the various background processes which can cause such problem. Usually, the processes with the same final state products as the signal process are the prominent backgrounds one should take care of. Let's look at the backgrounds for our case.

We already investigated the first background, which is the same process as our signal, except that it does not occur through the VBF channel. With a careful consideration, we finalized on 5 different background processes that we must account for in our analysis :

```
bcg1 :  p p > z h j j , h > b b~, z > l- l+

bcg2 :  p p > t t~ , (t > w+ b , w+ > e+ vl ) ,( t~ > w- b~ , w- > e- vl~)

  &&    p p > t t~ , (t > w+ b , w+ > mu+ vl ) ,( t~ > w- b~ , w- > mu- vl~)

bcg3 :  p p > z z j j QCD=0, z > b b~ , z > l- l+

bcg4 :  p p > z z j j , z > b b~ , z > l- l+

bcg5 :  p p > z b b~ j j , z > l- l+
```

Our task now would be to simulate all these processes in the MadGraph5-Pythia-Delphes framework and devise an analysis stratergy to discriminate between the signal and background processes. Our goal would be maximize the signal event yield (S) and minimize the net background event yield (B), as far as possible. The event yield (Y) for any process is given by,

$$Y = \mathcal{L} \cdot C_X \cdot \frac{\text{Number of Events selected}}{\text{Total number of Events Simulated}}$$

Where $\mathcal{L}$ is the total integrated luminosity and C_X is the cross-section for the corresponding process. We will be working with $\mathcal{L} = 3000\ fb^{-1}$ and therefore, will express all the cross-sections in terms of fb . The corresponding Yield error is calculated as :

$$\text{Yield-error} = \sqrt{\frac{Y^2}{\text{Number of Events selected}}}$$

We will estimate the effectiveness of the analysis with various significance (σ) measures, listed as follows :

$$\sigma = \frac{S}{\sqrt{B}}, \quad \sigma = \frac{S}{B}, \quad \sigma = \frac{S}{\sqrt{B + (\beta \cdot B)^2}}$$

Where instead of calculating the exact systematic error (β), we take $\beta = 0.1(10\%)$, which will act as a moderate benchmark.

We will be focusing on the final formula [15], which is the most general measure for the significance (σ).

We will now go ahead and simulate these processes with specific conditions, and move on to the analysis part.

6 Event generation

Let us briefly go through the event generation criteria.

First, we will only be generating events in the SM ($\kappa_W = \kappa_Z = 1$). If we look at the approximate cross-sections for all the processes, we find that almost all the processes have cross-section of the order $\sim 1\ fb$, however, the bcg2 and bcg5 have cross-sections of the order $\sim 1000\ fb$. Therefore, we will be generating larger event sample for these processes, in order to produce comparable event yields. We first produce the event samples in LHE file format [16],[17],[18] and then after Delphes detector simulation, they are stored in a Delphes Tree Root file.

For the purposes of our analysis, we used a 100k event sample for each of the signal, bcg1, bcg2, bcg3 and bcg4. A 700k event sample for the bcg5 and a 5 million event sample for bcg2.

We will also modify some of the run_card parameters of the MadGraph5, in order to generate large enough event samples for all of the signal and background processes. We use a common list of parameters and their values for all of the processes except bcg2. Following are the specific variables that we modify, along with their assigned value. (This list is for all the processes except the bcg2)

```
set ptb 20.0      # minimum pt for the b
set drbb 0.4      # min distance between b's
set mmjj 100.0    # min invariant mass of a jet pair
set xetamin 0.5   # minimum rapidity for two jets in the WBF case
set deltaeta 1.0  # minimum rapidity difference for two jets in the WBF case
set ebeam1 6500   # Energy of beamline-1
set ebeam2 6500   # Energy of beamline-2
set kW 1.0        # κ_W value
set kZ 1.0        # κ_Z value
```

Whereas, for the bcg2, we use the following extra set of variable constraints.

```
set missetmax 70.0  # maximum missing Et (sum of neutrino's momenta)
set mmll 70.0       # min invariant mass of l+l- (OSSF) lepton pair
set mmllmax 110.0   # max invariant mass of l+l- (OSSF) lepton pair
set ptb 20.0
set drbb 0.4
set mmjj 100.0
set xetamin 0.5
set deltaeta 1.0
set ebeam1 6500
set ebeam2 6500
set kW 1.0
set kZ 1.0
```

With the above constraints, we will generate the corresponding event samples in the MadGraph5-Pythia-Delphes framework. The corresponding cross-section files are also provided there. We can now devise the analysis strategy and test it on the event samples using the Delphes macros. The summary of cross-sections and corresponding processes is also prsented.

7 Summary of processes

```
signal :  p p > z h j j QCD=0, h > b b~, z > l- l+

   bcg1 :  p p > z h j j , h > b b~, z > l- l+

   bcg2 :  p p > t t~ , (t > w+ b , w+ > e+ vl ) ,( t~ > w- b~ , w- > e- vl~)

     &&    p p > t t~ , (t > w+ b , w+ > mu+ vl ) ,( t~ > w- b~ , w- > mu- vl~)

   bcg3 :  p p > z z j j QCD=0, z > b b~ , z > l- l+

   bcg4 :  p p > z z j j , z > b b~ , z > l- l+

   bcg5 :  p p > z b b~ j j , z > l- l+
```

The corresponding cross-sections reported by MadGraph5 are : (With the conditions described in the Event generation section)

```
   signal :  0.9104 fb

   bcg1 :  1.916 fb

   bcg2 :  5313.0 fb

   bcg3 :  1.214 fb

   bcg4 :  8.737 fb

   bcg5 :  1113.0 fb
```

We will actually[2] consider cross-sections reported by MadGraph5 for the raw process[3], and multiply them with Branching Ratios to obtain net cross-section. For example, cross-section reported by MadGraph5 for the raw signal process is 10.63 fb. Therefore, net cross-section for signal is $10.63 \text{ fb} \cdot BR(H \mapsto b\bar{b}) \cdot BR(Z \mapsto l^- l^+) = 0.4131$ fb. Where we use $BR(H \mapsto b\bar{b}) \sim 0.58$ [19] and $BR(Z \mapsto l^- l^+) \sim 0.067$ [20]. Cross-section files for all the concerned processes are uploaded here.

Thus, signal and bcg1 cross-sections[4] obtained with Branching Ratios (BRs) in this manner are :

```
   signal :  0.4131 fb

   bcg1 :  0.8809 fb
```

The event sample sizes are 100k each for all of the processes except for bcg2 and bcg5. For bcg5 it is 700k, while for bcg2 it is 5 million in size, as mentioned earlier.

[2] Some problem with Higgs width implements the wrong branching ratios for $H \to b\bar{b}$ in MG5.
[3] Event generation conditions in place.
[4] Later on, we will use these cross-sections.

8 Analysis Strategy

We will now go through the analysis strategy we developed, in order to diminish the background contribution as far as possible, without hurting the signal yield to a larger extent. We will be conducting our analysis on the event samples generated through the `MadGraph5-Pythia-Delphes` framework, where `Delphes` is responsible for the detector simulation, reconstructing detected particles and performing jet clustering. Later on, we will also employ the use of `FastJet` to run a separate Jet clustering algorithm. We will explain that part later. For now, we will be implementing the analysis with the help of `Delphes Macros` .

We aim to make the best use of the distinctive features of our signal process (VBF-ZH), namely the constraints associated with the VBF topology. As stated earlier, in a vector boson fusion induced process, the two quarks go on to form a pair of jets in the final state, which tend to show a little deviation from the beamline and thus are detected as a forward-backward jet pair in the detectors. Naturally, the absolute pseudo-rapidity difference between these jets $\sim |\eta_{\text{VBF}}|$ is also expected to be high. These jets are termed as the "VBF-Tagging Jet pair". There can be multiple ways to reconstruct the VBF-Tagging Jet pair. For our analysis, We will identify the forward-backward jet pair [5] with highest invariant mass as the VBF-Tagging Jet pair, based on the earlier research [21]. A number of cuts that we use, have been inspired from the same source.

First, let us introduce some of the terminology for the variables we will use throughout the project.

`VBF-Tagging Jet pair` : a forward-backward jet pair with highest invariant mass of all the forward-backward jet pairs

`MasterJet` : A jet[6] satisfying $p_T \geq 20$ GeV AND $|\eta| \leq 5$

`VBF-B-Jets` : Number of Delphes B-tagged jets, out of the two VBF-Tagging Jets

`OSSF Lepton pair` : Opposite Sign Same Flavour Lepton pair, $e^{\pm}$ or $\mu^{\pm}$

`DeltaEta` $|\eta_{\text{VBF}}|$: Absolute pseudo-rapidity difference between the two VBF-Tagging jets

`DiJetMass` : Invariant mass of the two VBF-Tagging jets

`DiBJetMass` : Invariant mass of the two Delphes B-Tagged jets

Initially, we apply a rudimentary set of cuts[7], which will help us to focus on the specialized cuts we may need to apply. The initial set of cuts that we apply are :

- Atleast one forward-backward jet pair must exist
- Number of MasterJets $\geq$ 4
- Number of Delphes B-tagged Jets $\geq$ 2
- Number of VBF-B-Jets == 0
- Invariant mass of the detected OSSF Lepton pair[8] $\in$ (81GeV, 101GeV)

[5] Forward-backward jet pair : the two jets must have opposite signs of η.
[6] All the jets under consideration from now on, must be MasterJets, including the B-tagged jets and any other jets.
[7] Cuts : all the conditions which must be satisfied in order to select an event for analysis.
[8] If no OSSF lepton pair is detected, discard that event.

With these cuts in place, we will go ahead and plot the normalized distribution for the concerned variables like `DeltaEta, DiJetMass, DiBJetMass` , which might help us discriminate signal from background. We will observe the behaviour of such variables for the concerned processes, and try to improvise cuts based upon these variables.

First let us concern ourselves with the VBF topology constraints namely the `DeltaEta & DiJetMass`. The corresponding plots are displayed in the figures 7,8.

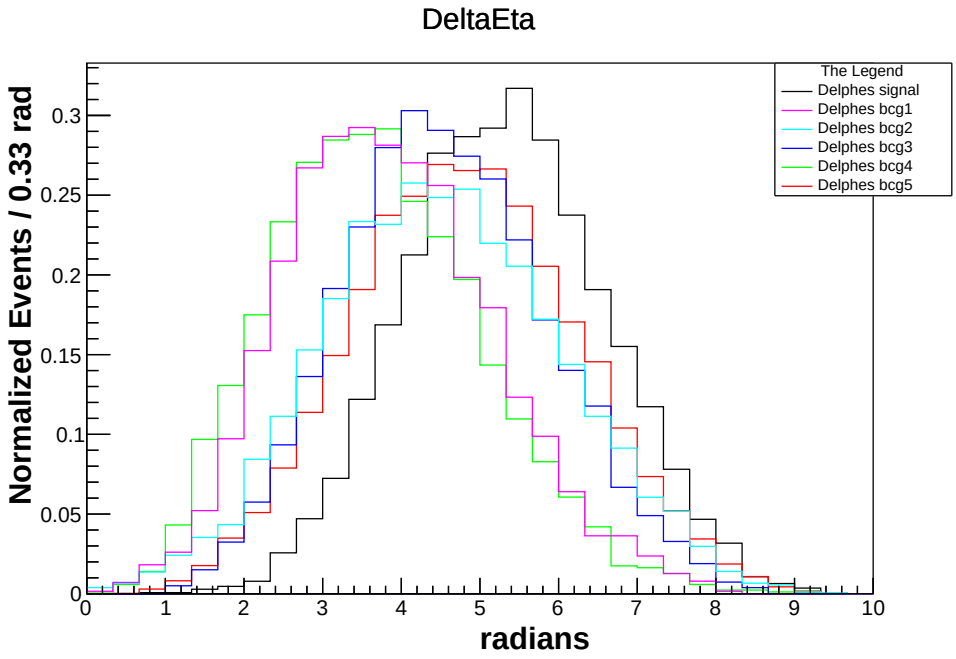

Figure 7: Normalized distribution of absolute pseudo-rapidity ($|\eta_{\text{VBF}}|$) difference for the VBF-Tagging Jets $\sim$ `DeltaEta` , for all processes.

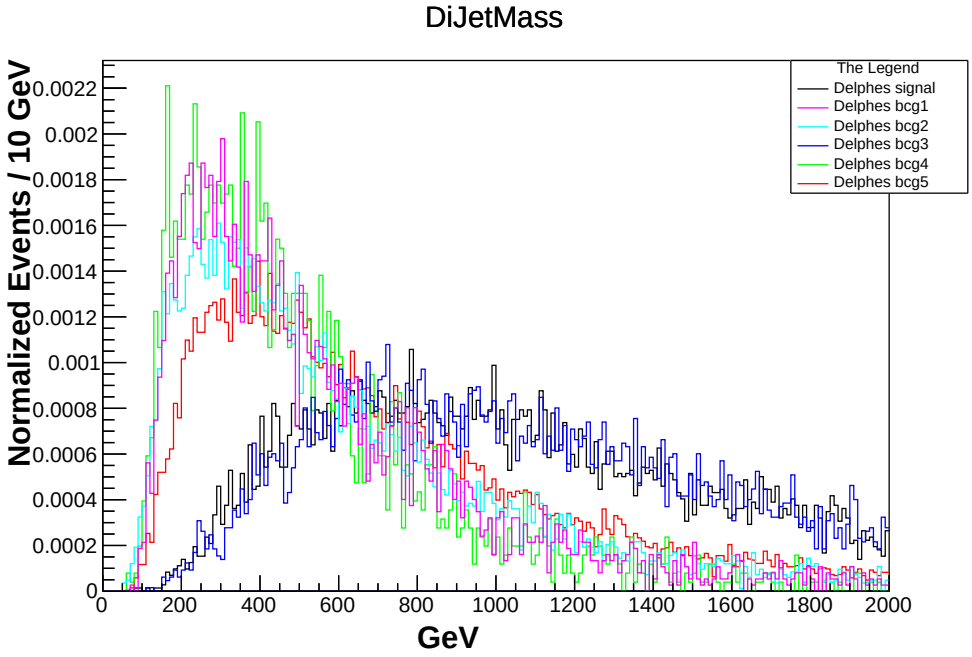

Figure 8: Normalized distribution of Invariant mass of the two VBF-Tagging Jets $\sim$ `DiJetMass` , for all processes.

Observing these distributions, we can corroborate our claims about the high $|\eta|$ value for the VBF-Tagging Jets. We can also catch the clear distinction between the distributions of `DiJetMass` for the VBF (`signal & bcg3`) and non-VBF processes. Cuts like $|\eta_{\text{VBF}}| \geq 4$ and `DiJetMass` ≥ 700 GeV can act as good starting point.

We know that, except the bcg2 and bcg5, all the other processes have cross-sections of the same order ~ 1 fb. Therefore, the individual yield ratios (S/B) for these processes are of the order $S/B \sim 1$ to begin with. As a result, with the VBF topology constraints in place, it would not be hard to even achieve $S/B \sim 10$. However, for bcg2 and bcg5, this ratio is already of the order $S/B \sim 10^{-3}$ to begin with. Implying, we need much stronger event selection criteria to reduce down these backgrounds to acceptable bounds.

Let us investigate the invariant mass of the two B-tagged jets $\sim$ `DiBJetMass` , and the angular separation between[9] them $\Delta R_{b\bar{b}} \sim$ `DRBB`, for this purpose. Refer to the figure 9 for the `DiBJetMass` plot.

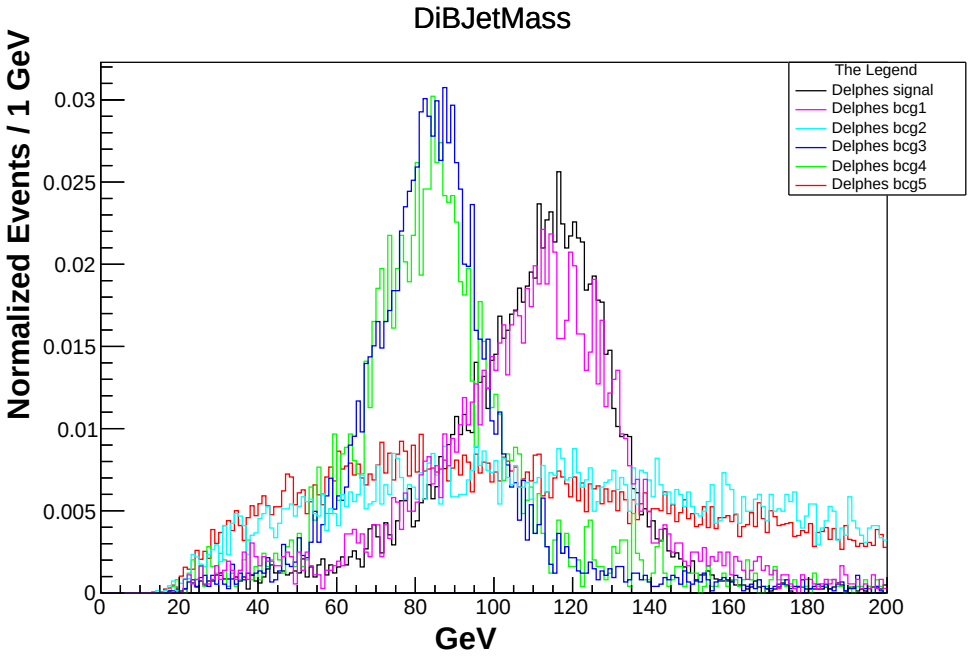

Figure 9: Normalized distribution of Invariant mass of the two B-tagged jets $\sim$ `DiBJetMass` , for all processes.

As one may predict, we observe the peaks around 125 GeV for signal and bcg1, and around 80 GeV for bcg3 and bcg4. Whereas, the bcg2 and bcg5 show continuous distribution. Thus, a window cut $\sim$ (100 GeV, 140 GeV) on the `DiBJetMass` can act as a good starting point.

Let us look at the distribution of angular separation between the two reconstructed b-jets $\Delta R_{b\bar{b}} \sim$ `DRBB` for only the bcg2 and bcg5 processes. Refer to the figure 10 for the `DRBB` plot.

One can clearly observe that a cut `DRBB` ≤ 2 will help best to get rid of the unnecessary bcg2 and bcg5 contributions. The signal peak is located around 1.5 radians, whereas the bcg2 and bcg5 have their peaks around 3 radians.

Along with these cuts, we will also place some general cuts to further reduce down the backgrounds. Observe that bcg2 contains neutrinos in final state, as a result, is likely to have large

[9] $\Delta R_{b\bar{b}} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

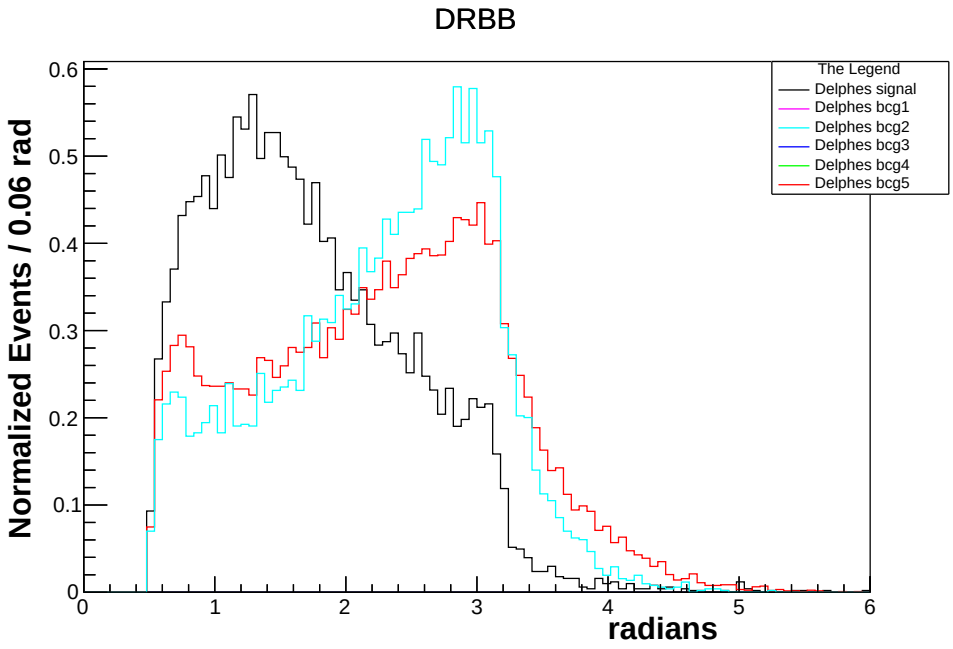

Figure 10: Normalized distribution of absolute pseudo-rapidity ($|\eta|$) difference for the VBF-Tagging Jets $\sim$ `DRBB` , for signal, bcg2 and bcg5.

missing transverse energy $\sim$ `missingET` . A cut like `missingET` < 50 GeV will help to further reduce down the bcg2.

Additionally, we will also place some constraints on the transverse momenta (p_T) of the leading jets[10]. The p_T of the first three leading jets must be greater than 100 GeV, 70 GeV, 50 GeV respectively. In a like manner, p_T of each of the two B-Jets must also be greater than 55 GeV.

Let us summarize all the cuts we have finalized, which will be implemented through the `Delphes Macros`. Initial cuts +

- `DeltaEta` ≥ 4 : ($|\eta_{\text{VBF}}| \geq 4$)
- `DiJetMass` ≥ 1000 GeV
- `DiBJetMass` $\in (110\text{ GeV}, 130\text{ GeV})$
- `DRBB` ≤ 2 : ($\Delta R_{b\bar{b}} \leq 2$)
- `missingET` < 50 GeV
- `PT-Jet1`[11] ≥ 100 GeV
- `PT-Jet2` ≥ 70 GeV
- `PT-Jet3` ≥ 50 GeV
- `PT-B-Jet1`[12] ≥ 55 GeV
- `PT-B-Jet2` ≥ 55 GeV

However, as we will explain later, we can improve upon these cuts by reconstructing the b-jets with a different jet clustering algorithm, and again placing a cut on the invariant mass of the two b jets. Let us familiarize to the FastJet framework where we will be able to do this. For convenience, the above set of cuts will be referred to as the 'Delphes cuts'.

[10]Leading jet 1 is the jet with highest p_T, leading jet 2 is the second highest p_T jet, and so on.
[11]PT-Jet1 : p_T of the first leading jet and so on.
[12]PT-B-Jet1 : p_T of the leading B-tagged Jet and so on.

9 Boosted Higgs search with FastJet

Carefully observing the `DRBB` - figure 10 graph, we can conclude the majority of the signal events have short angular separation between the two reconstructed b-jets. This implies that, the transverse momentum of Higgs must be quite high (Refer figure 11). Motivating us to employ a strategy to specifically search for the boosted Higgs decays, where the Higgs generated has a large p_T, resulting in a shorter angular separation between the b candidates. The idea behind the search is to look for fat jets, which can be formed by the b-jets emerging from the boosted Higgs. Then such fat jets can be broken down into the b-fragments which are reconstrcuted as the b-jets. This type of search has already been well tested and implemented in FastJet by J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam [22][13].

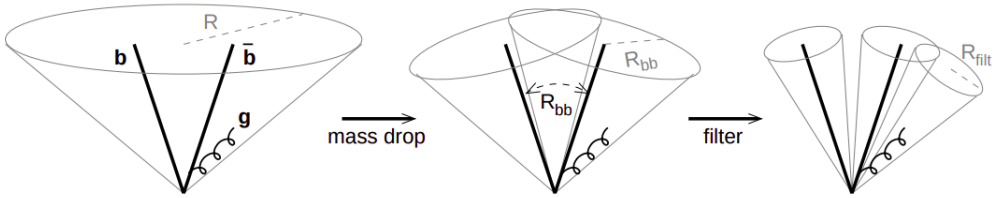

Figure 11: Pictorial representation for extracting the b-jets from the fat jet, source [22].

Therefore, our target now would be to apply the BDRS algorithm on our event data, to further improve the selectivity of the signal events with respect to the backgrounds. This particular BDRS algorithm has already been well implemented in the software package `FastJet` [23] and is distributed in one of the standard examples in the `FastJet` installation. We will be working with the `FastJet 3.4.0` . Please refer to the standard manual at [24], for installation and all other FastJet related details.

Although the algorithm is readily available to use in FastJet, the information regarding the detector simulation performed by Delphes, is encoded in the Delphes ROOT Tree file. Before we develop a way to use this data into FastJet, we first need to set up a way to use FastJet from within the ROOT. We will be using FastJet through files compiled with ACLiC, where we need to add the include path and link the necessary FastJet libraries. The exact commands we had to set in our case were as follows [14] :

```
> gSystem->AddIncludePath("-I/PATH/TO/fastjet-install/include");
> gSystem->Load("/PATH/TO/fastjet-install/lib/libfastjet.so");
> gSystem->Load("/PATH/TO/fastjet-install/lib/libfastjettools.so");
> .L fast_jet_macro.cc+
> main()
```

For more information about using FastJet with ROOT, refer to [25].

We can now move forward to develop a system in order to extract the data of detected particles from Delphes, and feed this data into FastJet to implement the BDRS algortihm.

[13] Commonly referred to as the BDRS algorithm.
[14] These commands are to be input into the ROOT interpreter before executing the fast-jet macros.

10 Final analysis framework : Delphes + FastJet

FastJet only needs access to the 4-momenta of the particles, namely the E, p_x, p_y, p_z variables of each detected particle. Additionally, we will also store the PDG-ID[15] of certain particles like leptons, to help them identify in the FastJet framework. We will define a custom Event class $\sim$ `MyEvent` , and store the data of the required particles in the vectors of this class. We will then go ahead and store this data in a ROOT tree file, making it easy to store and convenient to access from the within the FastJet framework. First let us review the definition of the `MyEvent` class.

```
   class MyEvent {
public :  MyEvent(){ px = 0; };
virtual ~MyEvent(){};

void setPx(int px_){ px = px_; };
float getPx(){return px;};

private :
float px;
};
```

Above definition is just an excerpt from the actual definition. The idea is that the each object of the class will have its own set of E, p_x, p_y, p_z variables. Therefore, we can use a vector of type `MyEvent` to store the necessary data of all the detected particles required, for each simulated event. The complete definition of this class is provided in the file `MyEvent.h` . However, to successfully implement this event class in our analysis, we first need to generate the dictionary files by compiling the header file `MyEvent.h` with ROOT, and then linking the shared library files with ROOT, before executing any of the Delphes or FastJet macros. Command to compile the header file with ROOT :

```
.L MyEvent.h+
```

and then the shared libary file can be linked via following command[16] :

```
gSystem->Load("/PATH/TO/MyEvent_h.so");
```

Additionally, one must not forget to include this header file in the start of the Delphes or FastJet macros. Following two lines must be added at the start of the macros :

```
#include "/PATH/TO/MyEvent.h"

ClassImp(MyEvent);
```

Along with the other necessary FastJet/ROOT/Delphes header files.

Now that we have the custom Event class defined, we can go ahead and access the detected particles through Delphes macros.

[15] PDG-ID : Unique ID assigned to each different particle, established by the particle data group. For more information refer : [26].

[16] Both the commands again should be enetered in the ROOT interpreter.

The detected particles are those which will be collected from the following three branches of the Delphes Tree.

`Track : HCal/eflowTracks (includes charged hadrons, electrons, muons )`

`Tower : ECal/eflowPhotons (includes photons)`

`Tower : HCal/eflowNeutralHadrons (includes neutral hadrons)`

We access all the `TLorentzVector P4()` for the above Track and Towers, and store the E, p_x, p_y, p_z for each particle detected, and push it onto the vector of the type `MyEvent` as stated earlier. We also store the PDG-ID of the leptons (will be needed later), as they are detected with a Track, which has PID as a member variable. For more details, refer to the class definitions [27]. These vectors are then saved into a ROOT Tree file, in a branch for the array of vectors. Along with these vectors, we will also save the rest of the variables used in the Delphes macros like `DeltaEta, DiJetMass` . We will use a separate branch for each individual variable. This will help us to harden the Delphes cuts later in the FastJet macros, if required.

Now, we can go ahead and load the above ROOT Tree in a FastJet macro, and process the data of all the detected particles, event-wise. Remember, our goal was to implement the BDRS algorithm on the detected particles. But before that, we will first remove the particle which are constituents of the VBF-Tagging Jets and the isolated leptons.

First, we define a UserInfo class to assign a unique ID to each particle loaded from the Tree, per event. It also stores the PDG-ID of each particle[17]. Then, we identify all the leptons with the associated PDG-ID and traverse through each one of them. Based upon the criteria described in the section 3.1.3 from [11], we select the isolated leptons and remove them from the list of particles.

Now, we perform jet-clustering on these particles with the `antikt algorithm` [28] and with the $R = 0.5$,[18] because these are the Delphes jet-clustering criteria. We again reconstruct the VBF-Tagging Jets[19] in this framework, and flag the constituent particles of these VBF jets with the help of the unique IDs. We get rid of these particles from the total list of particles loaded. Finally, on these remaining list of particles, we can go ahead and apply the procedure for BDRS algorithm.

We perform jet-clustering on remaining particles with the `Cambridge/Aachen algorithm` [29],[30] and take the Jet parameter to be $R = 2.0$ [20] along with the already defined Flavour re-combiner. We then apply a Mass-drop tagger with $\mu = 0.667$ & $y_{\text{cut}} = 0.09$ on the hardest jet i.e. jet with highest p_T. If there is no substructure found, we ignore that event, otherwise we obtain two tagged pieces which we identify as the b-jets. We can go ahead and plot invariant mass of these two tagged pieces to observe that it indeed has a peak around 125 GeV (Refer figure 12). This will prove useful in further diminishing the bcg5. Even though bcg2 has higher cross-section, bcg5 mimics the signal process to a greater extent than bcg2. Therefore, our main focus would be to reduce the bcg5, which in turn will help bcg2 to be in control. Refer to the figure 12, for a plot of the Higgs invariant mass reconstructed in the FastJet $\sim$ `Hmass`, compared with the Higgs invariant mass reconstructed in the Delphes $\sim$ `DiBJetMass` .

As we can observe, the distributions are pretty similar, whether reconstructed in Delphes or

[17] For our case, only leptons have the PDG-ID stored, other particles' PDG-ID isn't required.
[18] We also pass a Flavour recombiner, which has already been defined in the boosted-higgs example from the FastJet.
[19] Similarly as we had done in the Delphes macros.
[20] We can take $R = 1.2/1.5/2.0/2.5/3.0$ however, for our purposes, $R = 2.0$ suited to be a optimum benchmark.

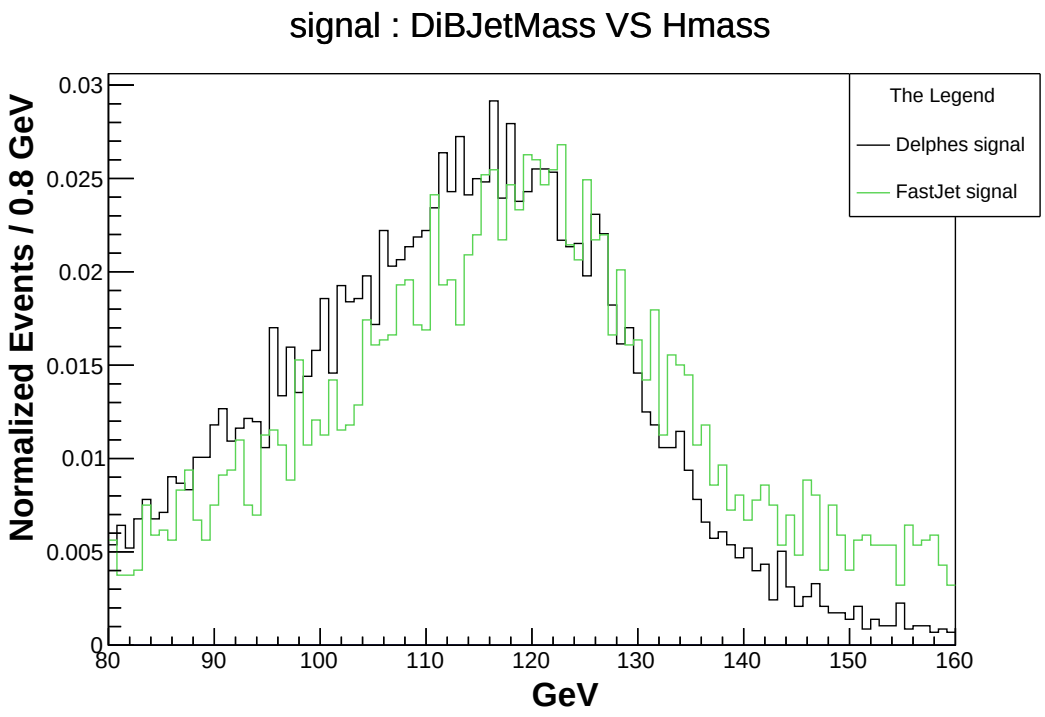

Figure 12: Normalized distribution of `DiBJetMass` Vs `Hmass`, for signal process .

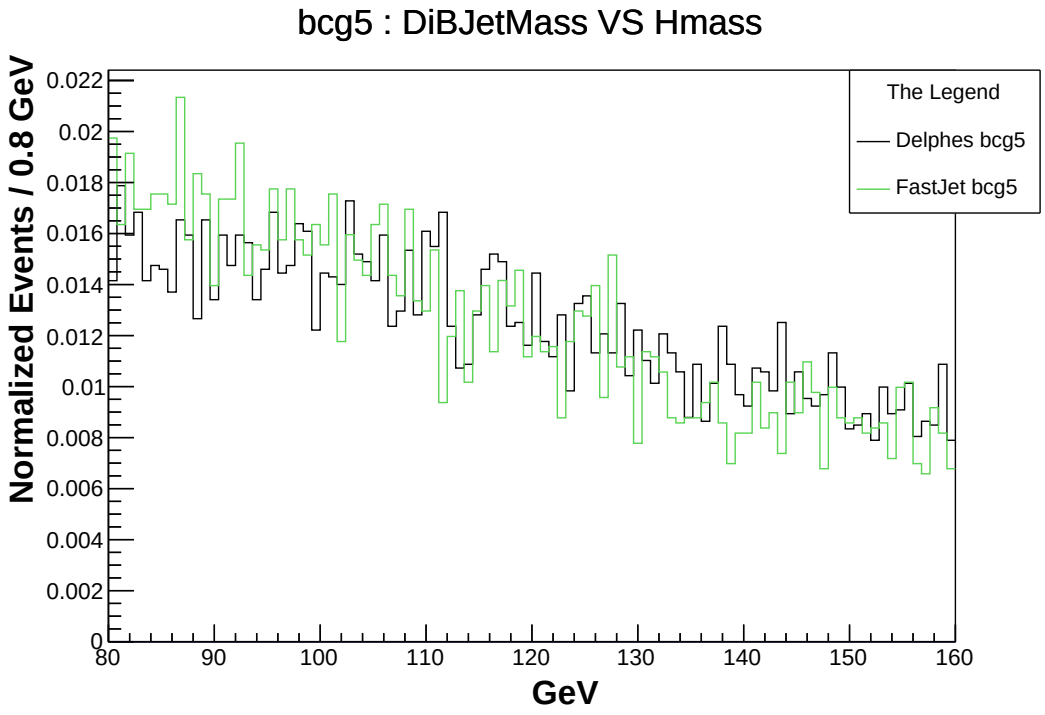

Figure 13: Normalized distribution of `DiBJetMass` Vs `Hmass`, for bcg5 process.

FastJet. However, we can still use this to our advantage by requiring the Higgs invariant mass cut to be satisfied in both frameworks, and we can see that it does help to improve the significance[21].

First, we implement all the cuts listed till now, except the conditions on `DiBJetMass` and `Hmass`, and as explained earlier, we will consider the bcg5 for now.

[21]Figure 12 and 13, both are produced with only the Initial cuts in place.

With the said cuts in place, a total of 1219 signal events[22] pass through, out of the 100k sample. For the bcg5, 350 events pass through, out of the 700k sample. Now, on these remaining events, we will apply the Higgs mass condition in both frameworks. `DiBJetMass` $\sim$ (110 GeV, 130 GeV) in the Delphes framework and `Hmass` $\sim$ (110 GeV, 130 GeV) in the FastJet framework. Tabulated results are illustrated in the table 1.

Higgs mass condition satisfied by	signal event counts	bcg5 event counts	Yield ratio S/b_5
BOTH Delphes & FastJet	503	9	0.32
AT LEAST Delphes	698	35	0.11
AT LEAST FastJet	590	25	0.13
NONE	434	308	—

Table 1: Event counts satisfying the higgs mass condition, in combination of Delphes & FastJet frameworks.

We can observe the increment in the yield ratio, when we require the Higgs mass to fall in the window of (110 GeV, 130 GeV), whether reconstructed through the Delphes or the FastJet procedure. Naturally, later on, we will also see an improvement in the final significance. Before we present the values for different significance measures, first let us briefly list down all the cuts we need to place for our analysis.

Final Cuts :

- Atleast one forward-backward jet pair must exist
- Number of MasterJets $\geq$ 4
- Number of Delphes B-tagged Jets $\geq$ 2
- Number of VBF-B-Jets == 0
- Invariant mass of the detected OSSF Lepton pair $\in$ $(81\text{GeV}, 101\text{GeV})$
- `DeltaEta` ≥ 4 : ($|\eta_{\text{VBF}}| \geq 4$)
- `DiJetMass` ≥ 1000 GeV
- `DRBB` ≤ 2 : ($\Delta R_{b\bar{b}} \leq 2$)
- `missingET` < 50 GeV
- `PT-Jet1` ≥ 100 GeV
- `PT-Jet2` ≥ 70 GeV
- `PT-Jet3` ≥ 50 GeV
- `PT-B-Jet1` ≥ 55 GeV
- `PT-B-Jet2` ≥ 55 GeV
- `DiBJetMass` $\in (110\text{ GeV}, 130\text{ GeV})$
- `Hmass` $\in (110\text{ GeV}, 130\text{ GeV})$

Where, the `DiBJetMass` is reconstructed in the Delphes with standard procedure, and `Hmass` is reconstructed in the FastJet, with the BDRS algorithm.

[22]Because the events satisfying the conditions are of the order $\sim$ 10, therefore, we would rather deal with actual counts than look at distributions.

11 Significance in the SM

With the Final cuts in place, let us review the corresponding yields[23] we obtain for different processes, refer to Table 2.

Process	Event selection	Yield
Signal (S)	503 /100k	6.23 $\pm$ 0.28
Bcg1 (b_1)	12 /100k	0.32 $\pm$ 0.09
Bcg2 (b_2)	3 /5M	9.56 $\pm$ 5.52
Bcg3 (b_3)	23 /100k	0.84 $\pm$ 0.17
Bcg4 (b_4)	1 /100k	0.26 $\pm$ 0.26
Bcg5 (b_5)	9 /700k	42.93 $\pm$ 14.31

Table 2: Event yields for all processes, final analysis.

Therefore, we have the Final yields S and B as :

$$S = 6.23, \quad \& \quad B = b_1 + b_2 + b_3 + b_4 + b_5 = 53.91$$

which gives our estimate for the analysis significance in the SM as follows :

$$\frac{S}{\sqrt{B}} = 0.85, \quad \frac{S}{B} = 0.12, \quad \frac{S}{\sqrt{B + (\beta \cdot B)^2}} = 0.68$$

If we had only used the Delphes framework for our analysis, we would have results as follows. (Table 3)

Process	Event selection	Yield
Signal (S)	698 /100k	8.65 $\pm$ 0.33
Bcg1 (b_1)	30 /100k	0.79 $\pm$ 0.14
Bcg2 (b_2)	4 /5M	12.75 $\pm$ 9.01
Bcg3 (b_3)	47 /100k	1.71 $\pm$ 0.25
Bcg4 (b_4)	2 /100k	0.52 $\pm$ 0.37
Bcg5 (b_5)	35 /700k	166.95 $\pm$ 28.22

Table 3: Event yields for all processes, if only Delphes had been employed.

The comparison between results is represented in the following table (Table 4).

Significance (σ)	Final analysis	Only Delphes analysis
$\frac{S}{\sqrt{B}}$	0.85	0.64
$\frac{S}{B}$	0.12	0.05
$\frac{S}{\sqrt{B+(\beta\cdot B)^2}}$	0.68	0.38

Table 4: Significance comparison between final analysis and only Delphes analysis.

[23] We use the cross-section obtained with Branching Ratios in case of signal and bcg1.

12 Contour Plots (κ_W, κ_Z)

We fix the analysis in the SM ($\kappa_W = \kappa_Z = 1$) with the final cuts described as before, fixing the FastJet parameter at $R = 2$.

- Case (a) : Significance with Signal yield at each point.

For the signal process, we run the analysis over a set of (κ_W, κ_Z) points and record the cross-section[24], event selection at each point by simulating 10k events. Corresponding figures are noted in the Column 4 & 5 of the following table. Using this data, yield of signal process $S(\kappa_W, \kappa_Z)$ is calculated in the column 6.

Out of the backgrounds, only background-1 will have significant dependence on κ_W, κ_Z. The cross-section for background-1 is independent of κ_W and varies $\sim \kappa_Z^2$. However, because the contribution of background-1 is already negligible, we will assume the event selection efficiency to stay approximately same over the concerned (κ_W, κ_Z) plane. Therefore, the varying net background yield can be modeled as :

$$B(\kappa_W, \kappa_Z) = b_{2,3,4,5} + b_1 \cdot \kappa_Z^2$$

where, $b_{2,3,4,5} = 53.59317$ is the constant contribution (SM contribution) from remaining backgrounds while $b_1 = 0.317144232$ is the background-1 yield in the SM. These $B(\kappa_W, \kappa_Z)$ figures are noted down in the column 8 of the corresponding table.

Now, the Significance at each (κ_W, κ_Z) point can be calculated by substituting the signal yield (S) and the background yield (B) in the following formulas. ($\beta = 0.1$ is assumed for time being)

$$\sigma(\kappa_W, \kappa_Z) = \frac{S(\kappa_W, \kappa_Z)}{\sqrt{B(\kappa_W, \kappa_Z) + \left[\beta \cdot B(\kappa_W, \kappa_Z)\right]^2}}$$

The corresponding σ figures are noted down in the column 10 of the following table.

Additionally, we can approximately model the cross-section for signal process $C_s(\kappa_W, \kappa_Z)$ as follows :

$$C_s(\kappa_W, \kappa_Z) = (14.41\ fb) \cdot \kappa_W^2 - (12.147\ fb) \cdot \kappa_W \kappa_Z + (8.367\ fb) \cdot \kappa_Z^2$$

These figures are in the column 11.

The contour plots produced in this case are represented in figure 14.

- Case (b) : Significance with SM as background

Now, we calculate the significance with a different approach, while considering the contribution from SM in background. We borrow the data from Case 1(a) and only modify the measure of significance. We employ following formula for this purpose :

$$\sigma = \frac{|A(\kappa_W, \kappa_Z) - A_{SM}|}{\sqrt{A_{SM} + (\beta \cdot A_{SM})^2}}$$

Where $A = S + B$ is the total yield at the corresponding point. Particularly, this will act as good measure of deviation from the SM, because in the SM $\sigma = 0$. The contour plots in this case are showcased in figure 15.

[24] Cross-sections are obtained from raw process cross-sections with Branching ratios.

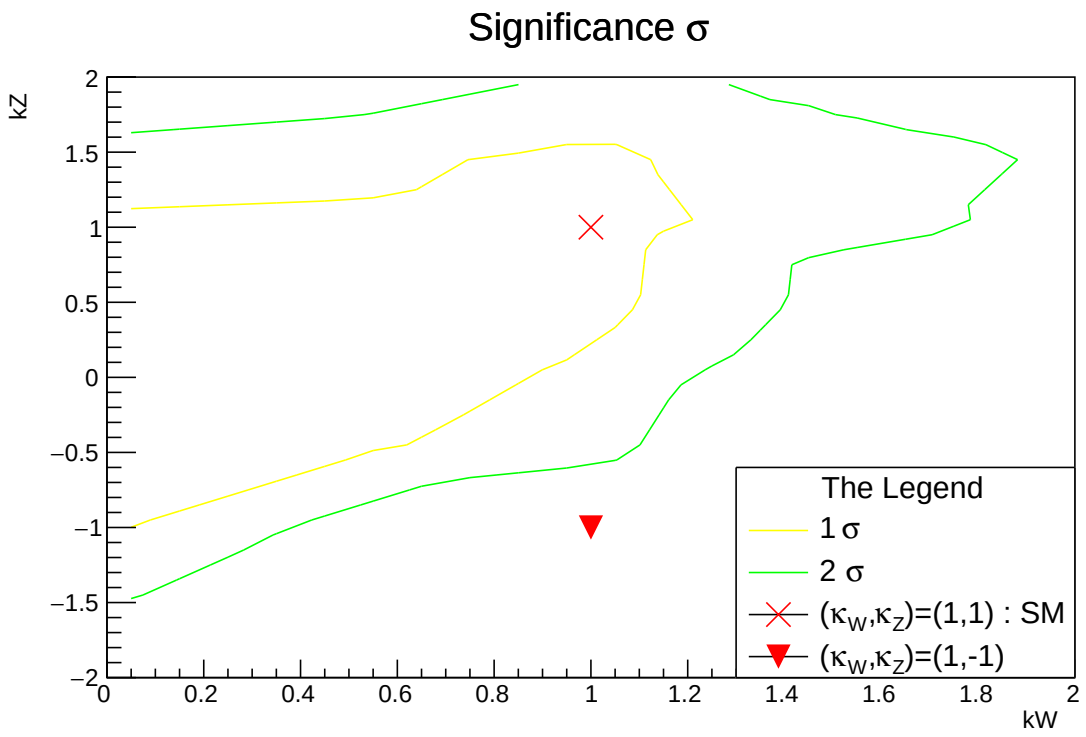

Figure 14: Contour plot - Case (a) .

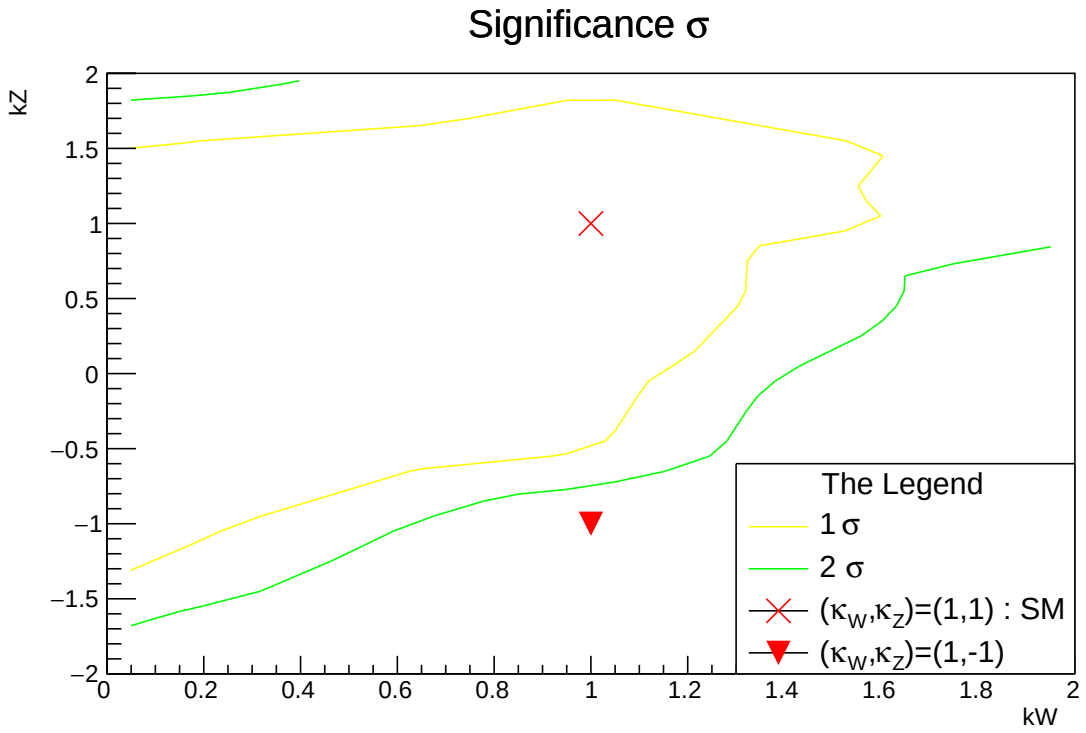

Figure 15: Contour plot - Case (b) .

The data-sets for both cases which were used to produce these plots are provided in the Appendix C.

We have successfully produced the contour plots over the (κ_W, κ_Z) plane, and ready to move forward to the conclusions.

13 FastJet-Delphes characteristics

Because we have proposed to employ two different methods to reconstruct the Higgs mass, it would be wise to carefully examine the 2D distribution of the event counts over the previously explained FastJet-Delphes strategies. We will look at the observed bare event counts per 10 GeV square in the FastJet-Delphes plane, and plot the corresponding 2D histogram in successive manner of cuts. For example, a point $(120, 160)$ in this plane means that the value of Higgs reconstructed in the Delphes framework is 120 GeV, whereas it is 160 GeV in the FastJet framework.

Ideally, according to the Final cuts we proposed, we want maximum signal density in the 110 GeV-130 GeV square, and intend to have negligible density in rest of the strip areas (marked with red lines). Correspondingly, we want minimum density of bcg5 event counts, in the same square.

With only initial cuts in place, we can look at these FastJet-Delphes characteristics in 16. Observe that signal has good strength in the required square, but the the bcg5 has a rather continuum of a distribution. This nature of the bcg5 to pertain a continuous distribution can still be observed after application of extra set of cuts.

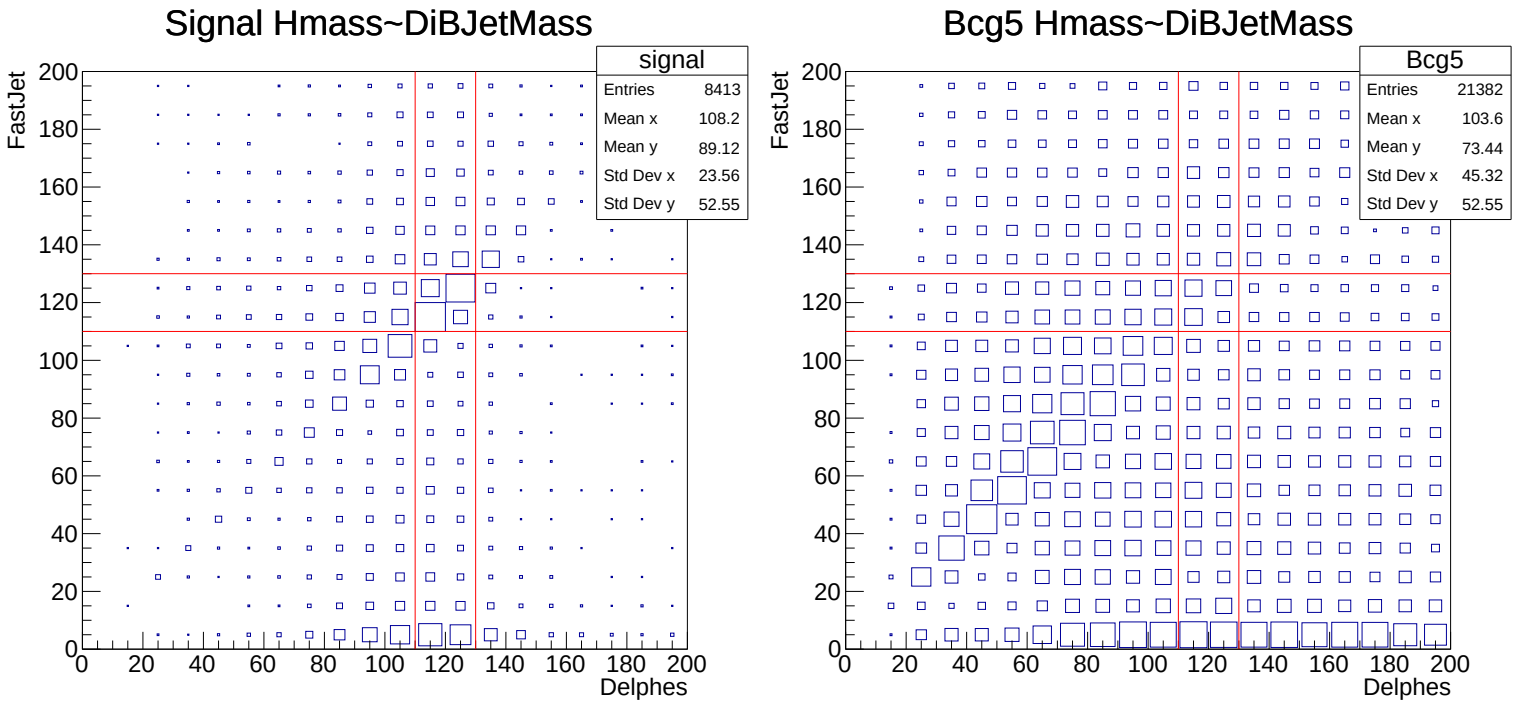

Figure 16: FastJet-Delphes characteristics for signal and bcg5 process : After Initial cuts

If we go ahead and apply the `DRBB` ≤ 2 cut, which was the motivation behind using FastJet search; we can observe for bcg5 that the distribution on the Delphes side pretty much clears out. (figure 17) But the FastJet side still exhibiting a dense enough distribution. On the contrary, the signal strength distribution is somewhat same, just the events with rest higgs decays are cleared out. This particularly shows the disagreement between the FastJet and Delphes values for bcg5, and good agreement for signal.

After the final set of cuts (except the `Hmass` and `DiBJetMass`), we can look at the final characteristics in figure 18. As required, the signal has maximum contribution in the required square, with negligible densities in the remaining strips. For bcg5 however, we still have comparable contribution

in the remaining strip areas to that of the square, pertaining to the continuum distribution nature of bcg5. This more or less motivates our argument to make effective use of both the frameworks to further reduce down the bcg5, whose every count is 1000 times stronger than signal.

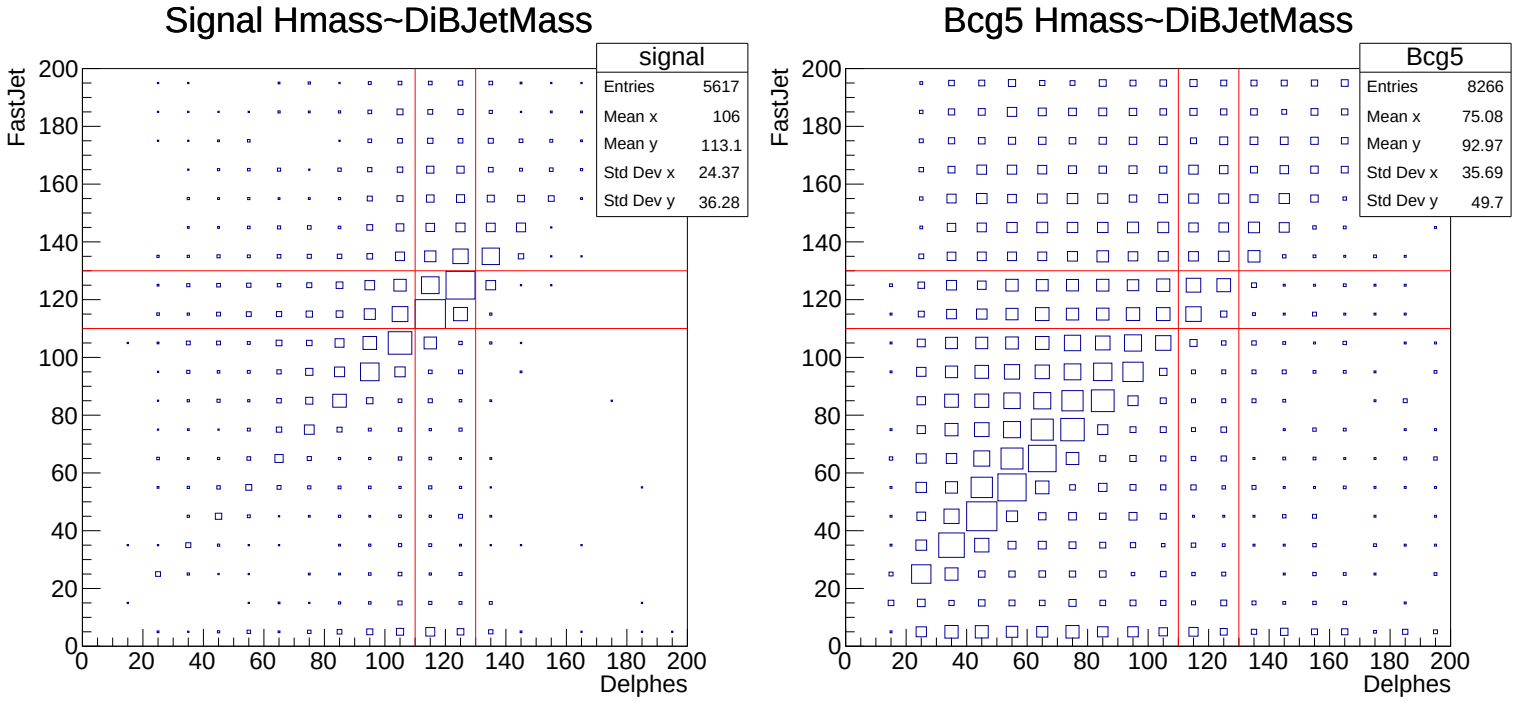

Figure 17: FastJet-Delphes characteristics for signal and bcg5 process : After Initial + `DRBB` ≤ 2 cut

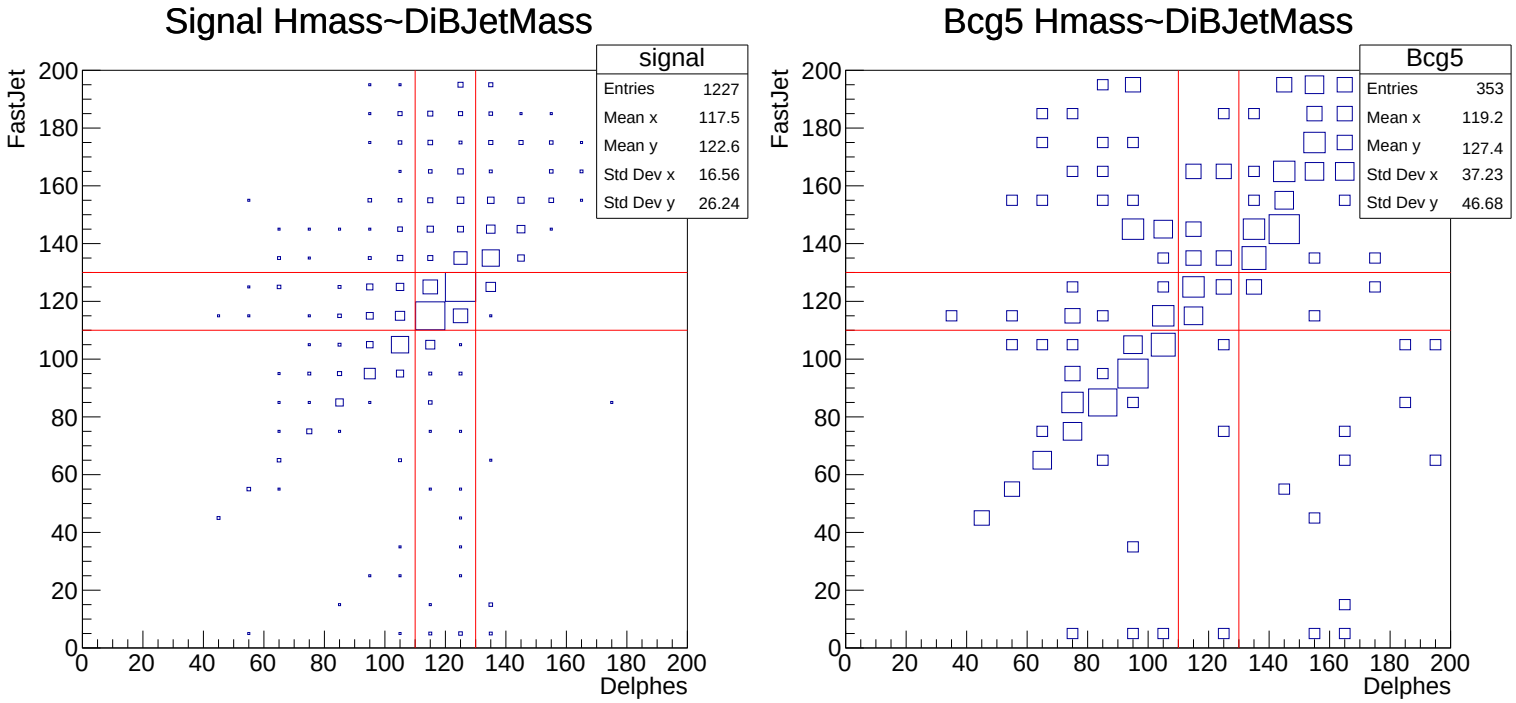

Figure 18: FastJet-Delphes characteristics for signal and bcg5 process : After all but the `Hmass-DiBJetMass` cuts

14 Signal distributions at $(\kappa_W, \kappa_Z) = (1, -1)$ Point

Observation of contour plots hints towards clear indication that the $(\kappa_W, \kappa_Z) = (1, -1)$ point can be well excluded with the expected SM observation data. The corresponding point ('-1 point') has almost 3 times larger cross-section than the SM case, which makes it distinguishable. Let us look at the distributions for signal process at this point, compared to the expected SM signal.

One can observe that for almost all of the variables, the distribution is somewhat similar, or has a growing tail on the stronger side of our cuts. Only the `DRBB` variable shows a significantly altered distribution, which increases its event selection and consequently the expected significance. (figure 22)

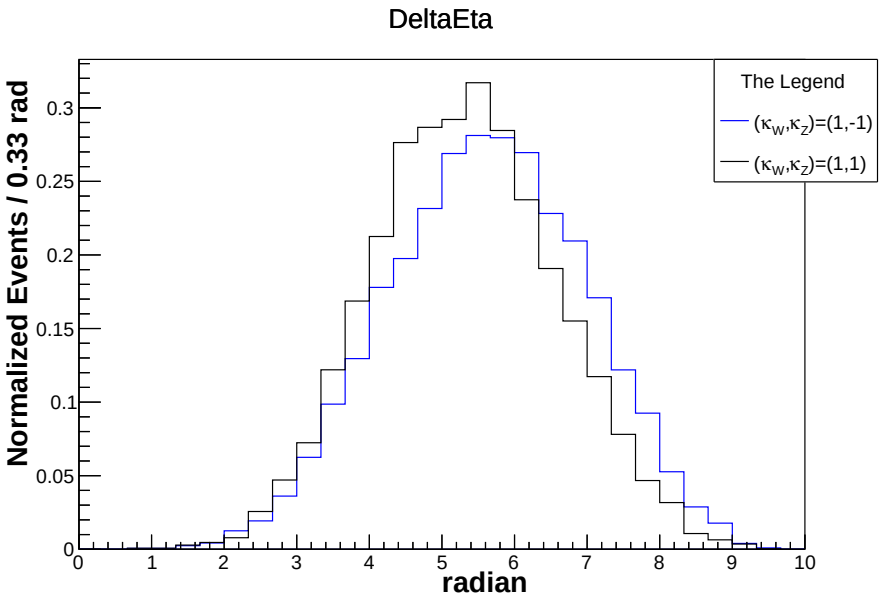

Figure 19: `DeltaEta` distribution for signal at SM Vs signal at -1 point

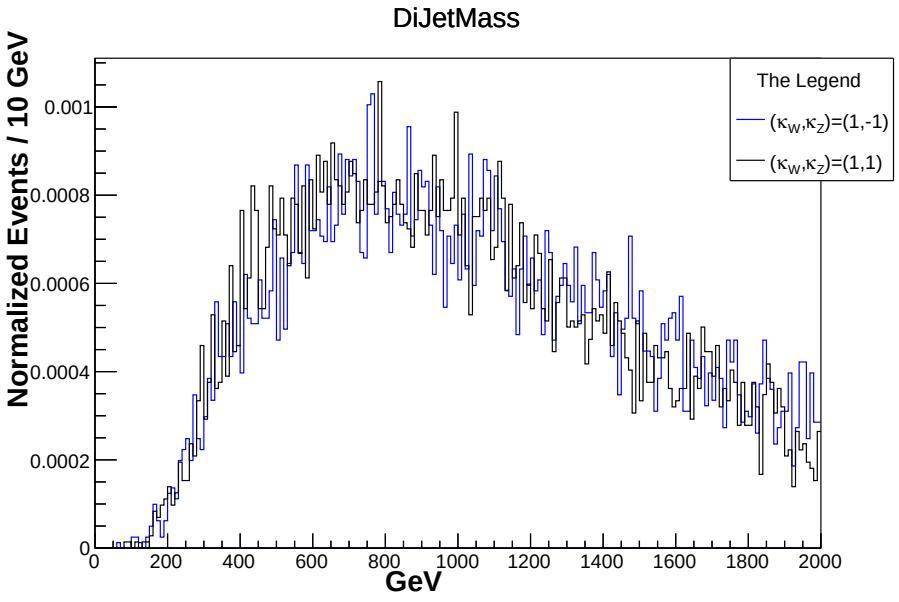

Figure 20: `DiJetMass` distribution for signal at SM Vs signal at -1 point

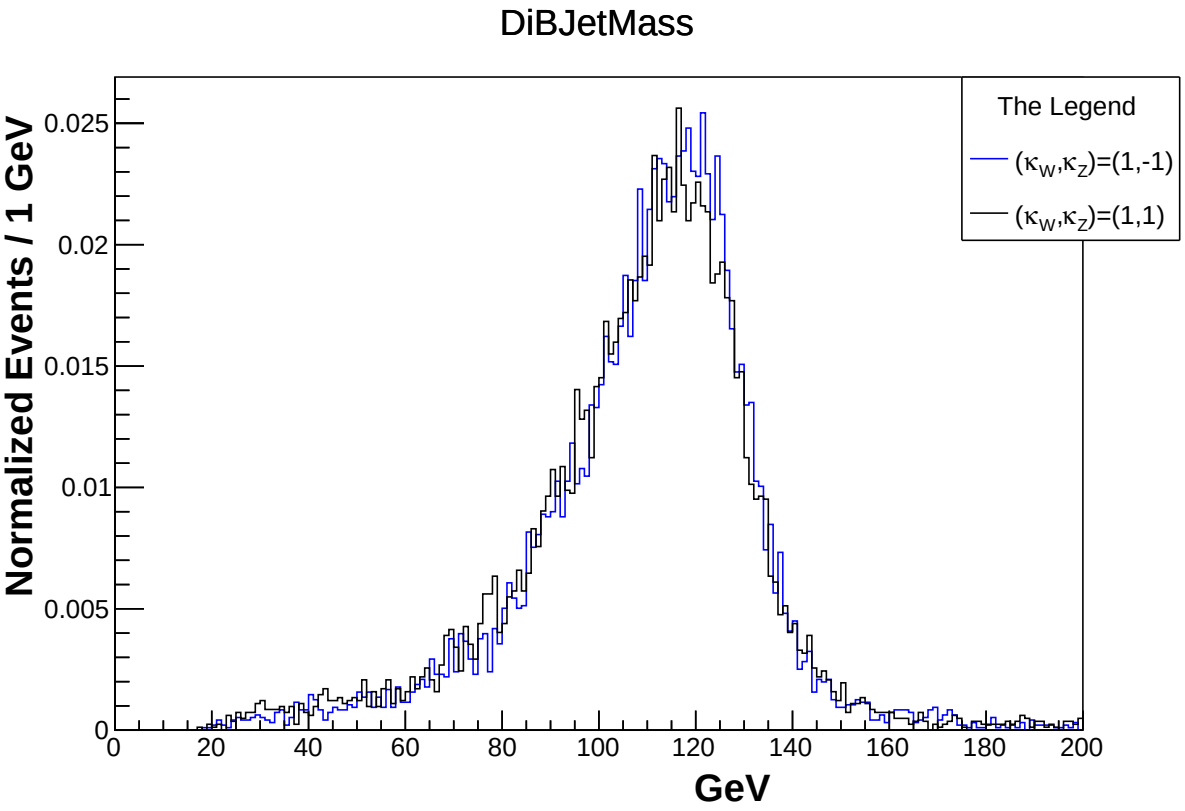

Figure 21: `DiBJetMass` distribution for signal at SM Vs signal at -1 point

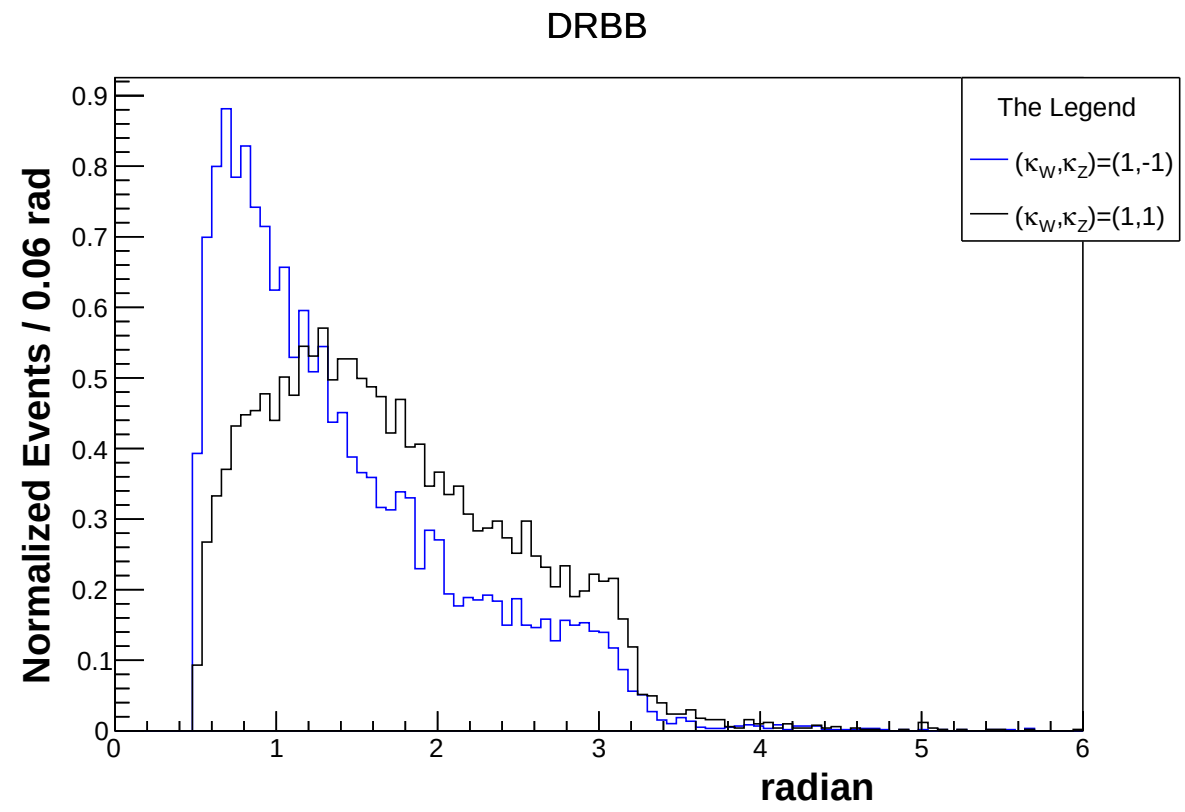

Figure 22: `DRBB` distribution for signal at SM Vs signal at -1 point

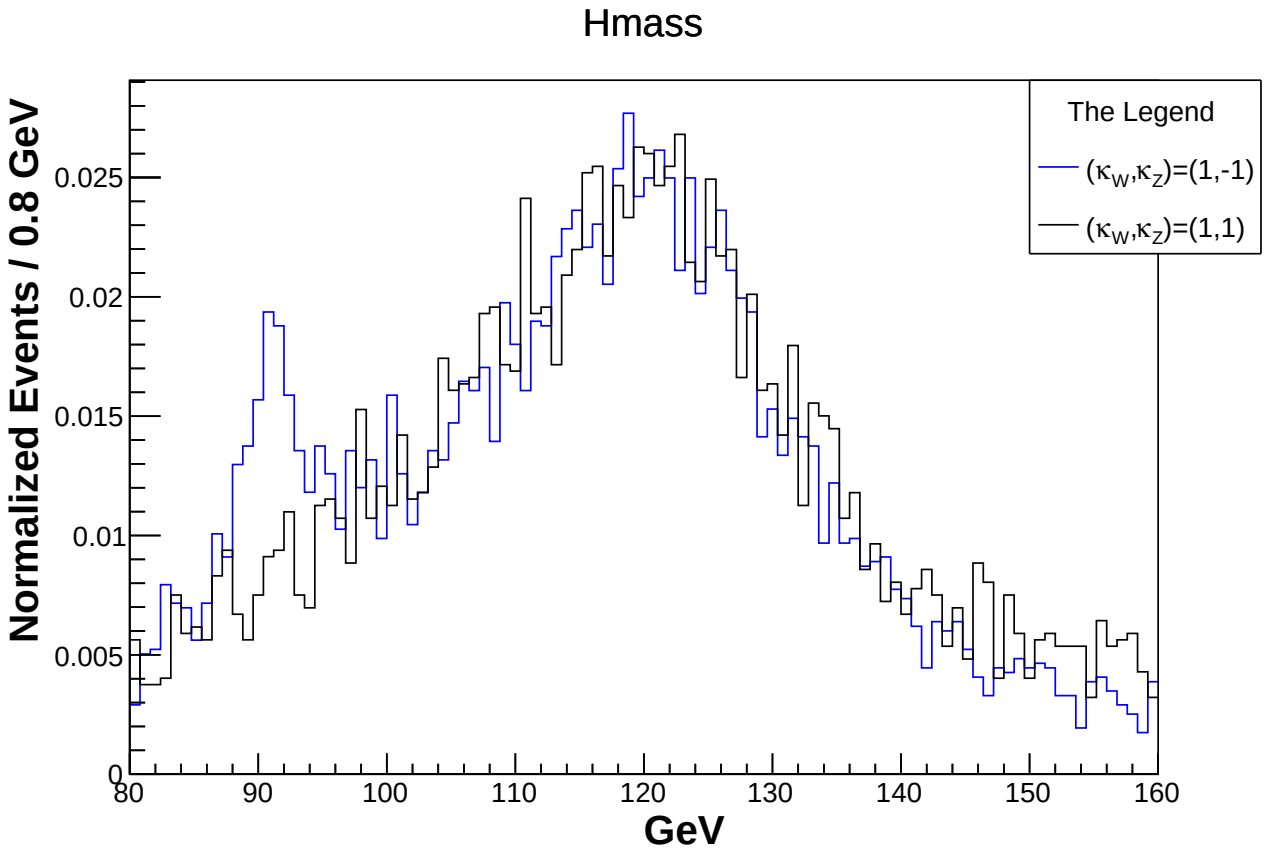

Figure 23: `Hmass` distribution for signal at SM Vs signal at -1 point

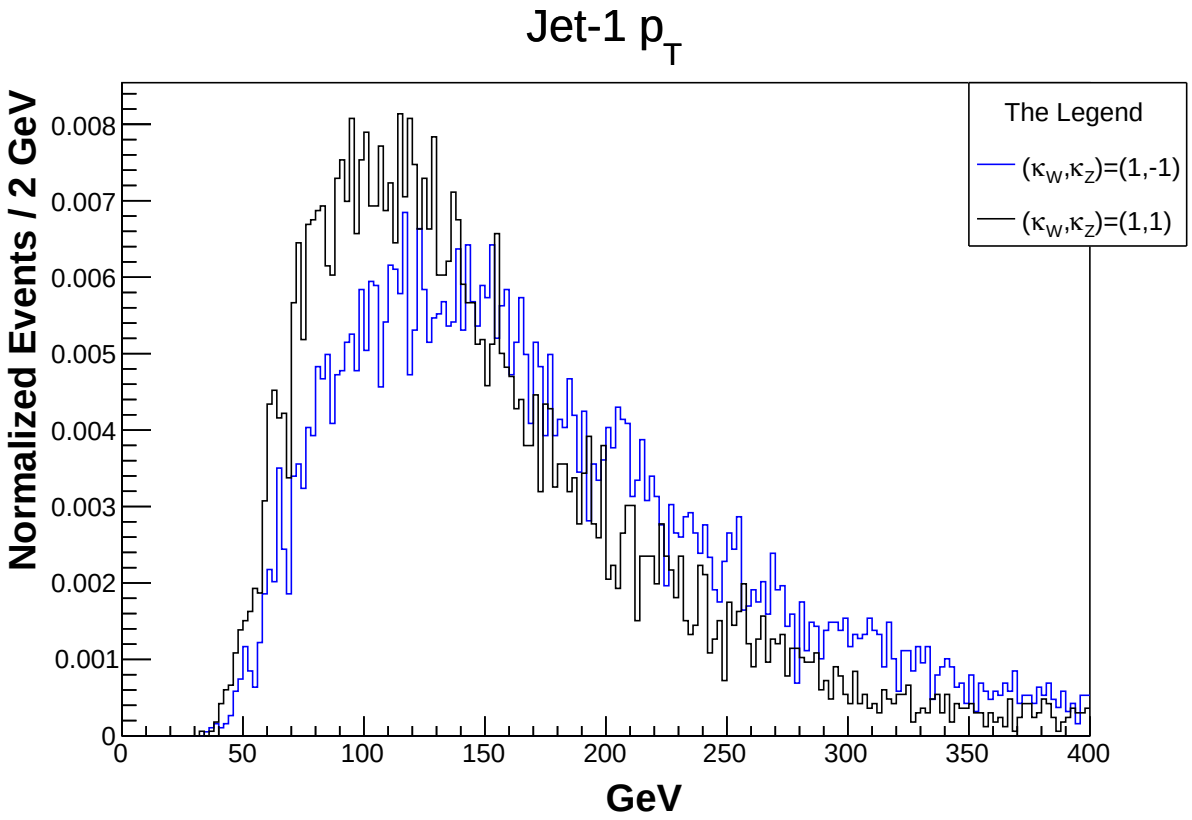

Figure 24: `Jet-1` p_T distribution for signal at SM Vs signal at -1 point

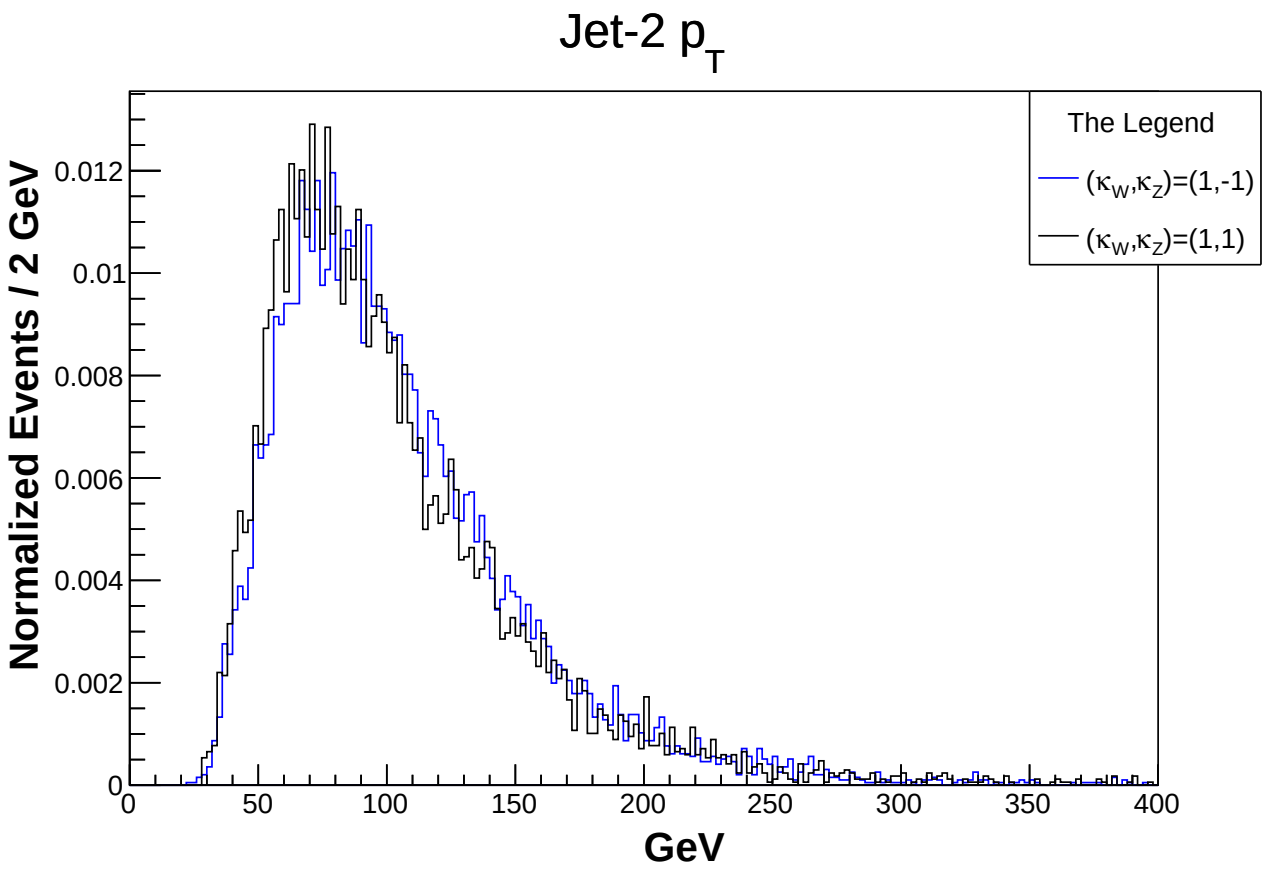

Figure 25: `Jet-2` p_T distribution for signal at SM Vs signal at -1 point

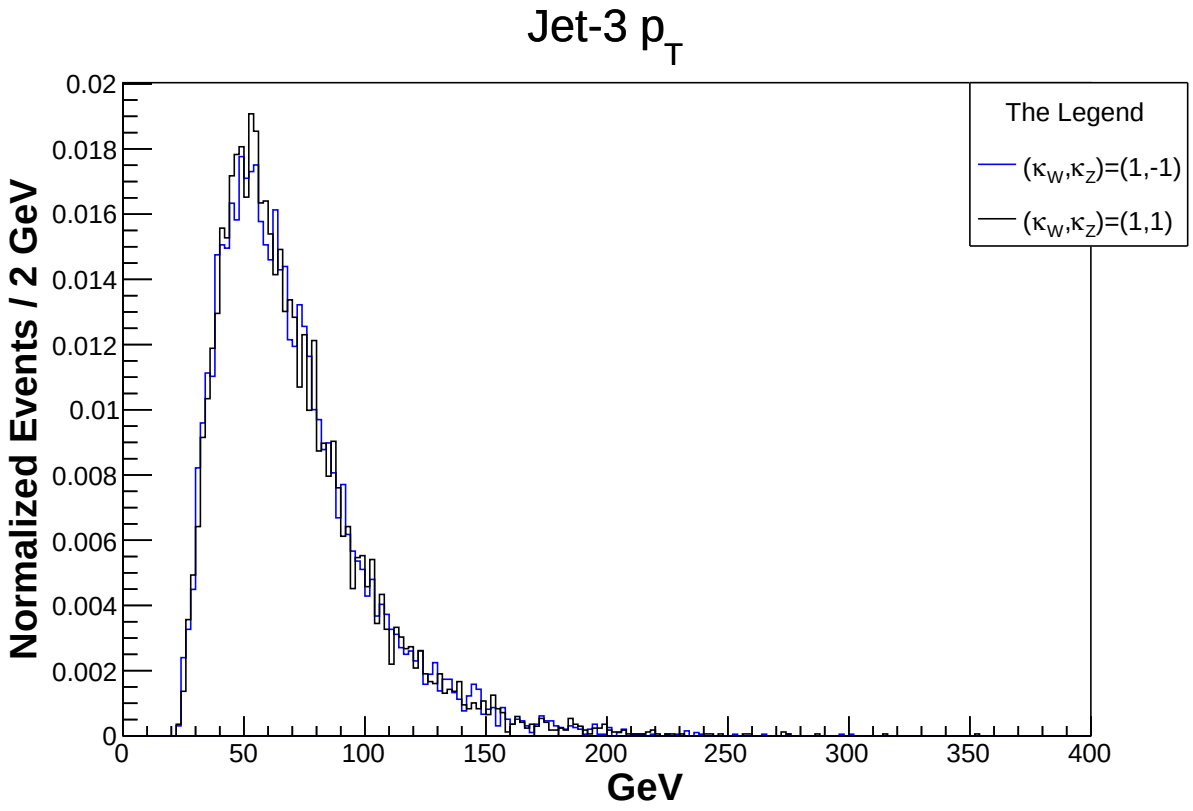

Figure 26: `Jet-3` p_T distribution for signal at SM Vs signal at -1 point

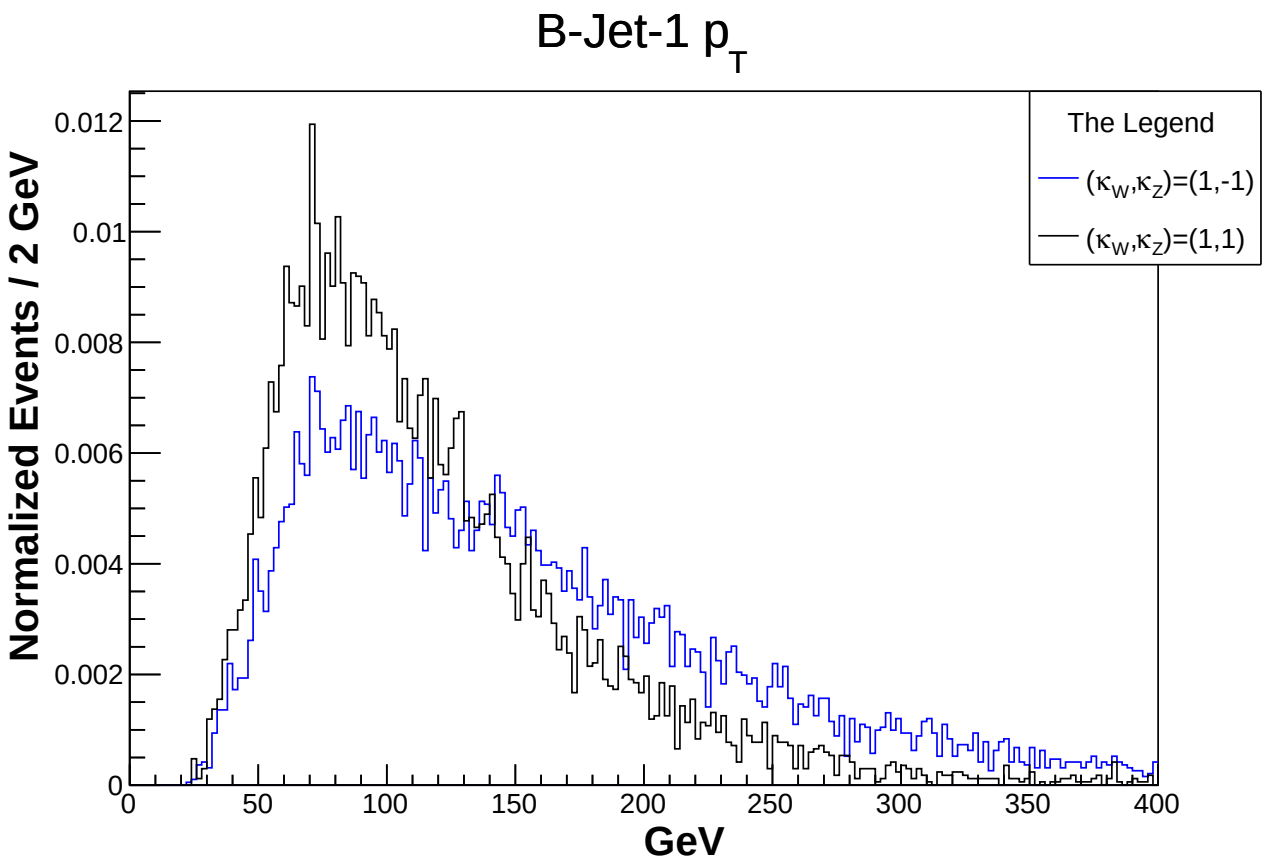

Figure 27: `B-Jet-1` p_T distribution for signal at SM Vs signal at -1 point

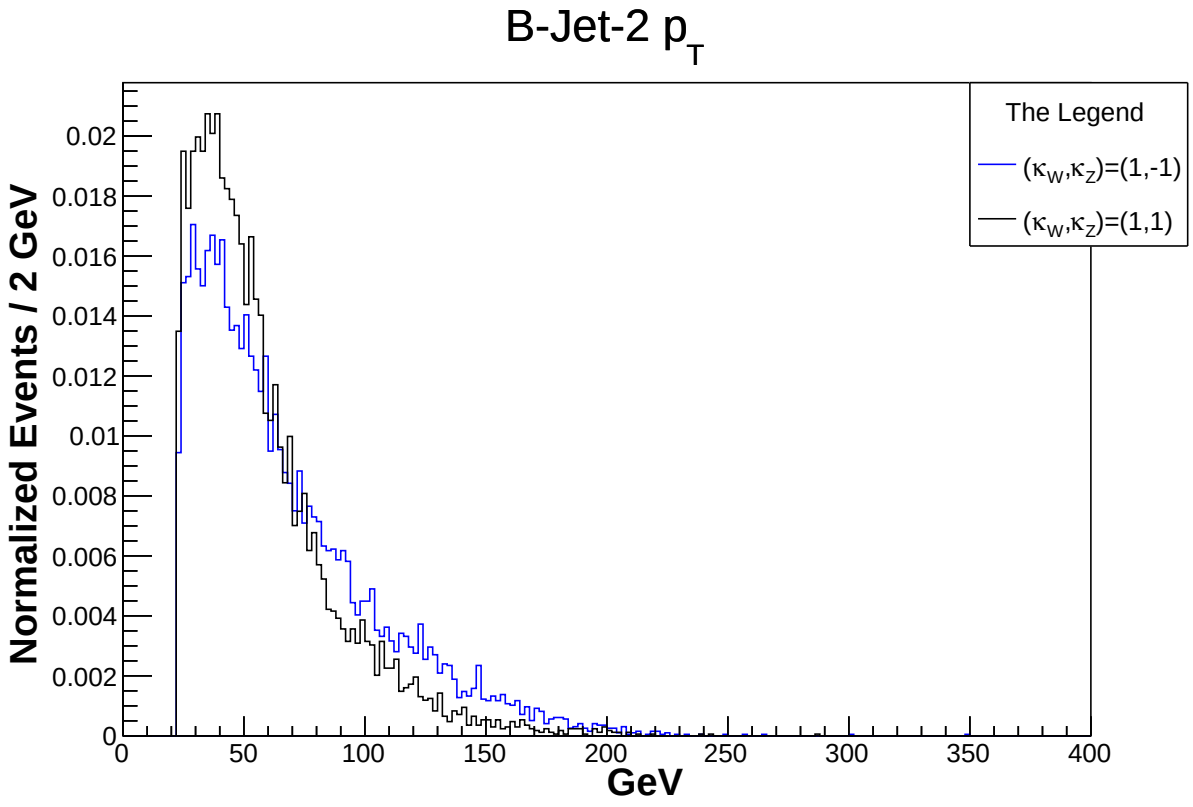

Figure 28: `B-Jet-2` p_T distribution for signal at SM Vs signal at -1 point

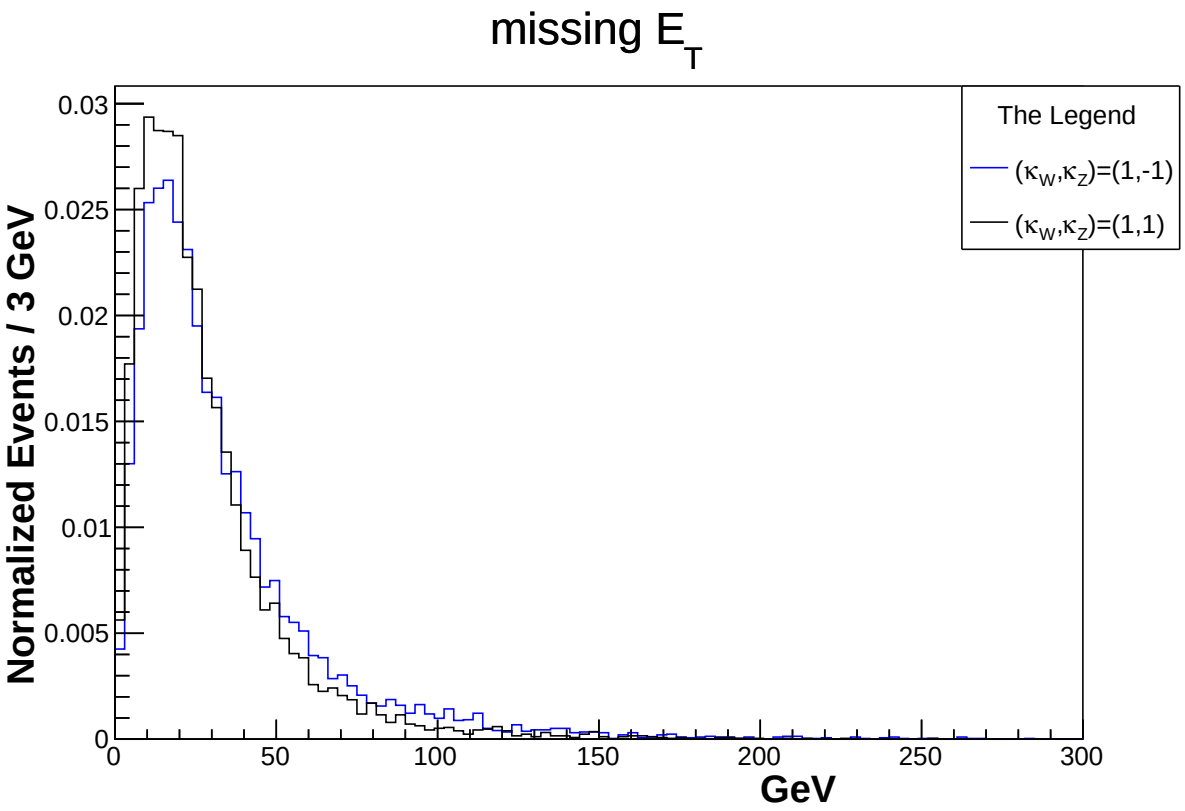

Figure 29: `missingET` distribution for signal at SM Vs signal at -1 point

15 Conclusion

We have conclusively performed a study to search for the VBF-VH process at LHC, anticipating the high luminosity upgrade. Being a much rare process, we worked with $\mathcal{L} = 3000\ fb^{-1}$ for our analysis and evaluated the proposed process yields. Particularly, we have established the results for VBF-ZH process with $H \mapsto b\bar{b}$, $Z \mapsto l^{-}l^{+}$ as the decay channel. The sensitivity of process with κ_W, κ_Z couplings and the claims about Quantum interference effects, have all been verified earlier in the section for VBF-ZH process.

The search strategy proposed involves a numerous cuts and reconstructing the invariant mass of higgs in two different ways. One where the 2 b-jets are reconstructed in the standard Delphes framework, with antikt algorithm - R=0.5. The other method where the b-jets are reconstructed according to BDRS algorithm, with Cambridge/Aachen algorithm R=2.0 . The consolidated application of two alternate methods proves to be fruitful in order to improve the significance of the analysis. In significance estimation calculations, we account for 10% systematic error as a benchmark. Effectiveness of the analysis over the (κ_W, κ_Z) plane is represented with the help of contour plots in the corresponding section, namely the figures 14,15.

Observing the associated figures for all cases, we can confidently propose that this search strategy can help to conclusively rule out the $(\kappa_W, \kappa_Z) = (1, -1)$ point, which shows at least 2σ deviation from the expected SM case. By symmetry arguments, we can also argue the same case for $(\kappa_W, \kappa_Z) = (-1, 1)$ point.

This will help in constricting the parameter space of Higgs to vector boson couplings to further bounds, and this search strategy can be adapted to look for other rare channels in light of the HL-LHC upgrades.

16 Acknowledgement

I am humbled and very grateful to have had Prof. Daniel Stolarski's Invaluable guidance throughout this project, and Dr. Yongcheng Wu's vital assistance. This project wouldn't have been possible without the valuable discussions with them and their constant support. I am also thankful to the Carleton University for allowing me to work as a virtual intern, which provided the administrative and technical support to make this project possible. I would also like to formally thank the Mitacs organization for allowing undergraduates an opportunity to work with Canada's leading research frontier. I worked on this project as an awardee of the Mitacs Globalink Research Internship program.

References

[1] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG .

[2] G. Aad *et al.* [ATLAS], Phys. Rev. D **101**, no.1, 012002 (2020) doi:10.1103/PhysRevD.101.012002 [arXiv:1909.02845 [hep-ex]].

[3] D. Stolarski and Y. Wu, Phys. Rev. D **102**, no.3, 033006 (2020) doi:10.1103/PhysRevD.102.033006 [arXiv:2006.09374 [hep-ph]].

[4] Rene Brun and Fons Rademakers, ROOT - An Object Oriented Data Analysis Framework, Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389 (1997) 81-86.

[5] https://root.cern/install/all_releases/ .

[6] http://madgraph.phys.ucl.ac.be/index.html .

[7] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli and M. Zaro, JHEP **07**, 079 (2014) doi:10.1007/JHEP07(2014)079 [arXiv:1405.0301 [hep-ph]].

[8] T. Sjostrand, S. Mrenna and P. Z. Skands, JHEP **05**, 026 (2006) doi:10.1088/1126-6708/2006/05/026 [arXiv:hep-ph/0603175 [hep-ph]].

[9] https://github.com/HEPcodes/pythia-pgs .

[10] S. Ovyn, X. Rouby and V. Lemaitre, [arXiv:0903.2225 [hep-ph]].

[11] J. de Favereau *et al.* [DELPHES 3], JHEP **02**, 057 (2014) doi:10.1007/JHEP02(2014)057 [arXiv:1307.6346 [hep-ex]].

[12] https://twiki.cern.ch/twiki/bin/view/CMSPublic/MadgraphTutorial .

[13] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, Comput. Phys. Commun. **185**, 2250-2300 (2014) doi:10.1016/j.cpc.2014.04.012 [arXiv:1310.1921 [hep-ph]].

[14] C. Degrande, C. Duhr, B. Fuks, D. Grellscheid, O. Mattelaer and T. Reiter, Comput. Phys. Commun. **183**, 1201-1214 (2012) doi:10.1016/j.cpc.2012.01.022 [arXiv:1108.2040 [hep-ph]].

[15] P. Schwaller, D. Stolarski and A. Weiler, JHEP **05**, 059 (2015) doi:10.1007/JHEP05(2015)059 [arXiv:1502.05409 [hep-ph]].

[16] J. Alwall, A. Ballestrero, P. Bartalini, S. Belov, E. Boos, A. Buckley, J. M. Butterworth, L. Dudko, S. Frixione and L. Garren, *et al.* Comput. Phys. Commun. **176**, 300-304 (2007) doi:10.1016/j.cpc.2006.11.010 [arXiv:hep-ph/0609017 [hep-ph]].

[17] J. M. Butterworth, A. Arbey, L. Basso, S. Belov, A. Bharucha, F. Braam, A. Buckley, M. Campanelli, R. Chierici and A. Djouadi, *et al.* [arXiv:1003.1643 [hep-ph]].

[18] J. R. Andersen, S. Badger, L. Barze, J. Bellm, F. U. Bernlochner, A. Buckley, J. Butterworth, N. Chanon, M. Chiesa and A. Cooper-Sarkar, *et al.* [arXiv:1405.1067 [hep-ph]].

[19] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR .

[20] https://pdg.lbl.gov/2018/listings/rpp2018-list-z-boson.pdf .

[21] V. Khachatryan *et al.* [CMS], Phys. Rev. D **92**, no.3, 032008 (2015) doi:10.1103/PhysRevD.92.032008 [arXiv:1506.01010 [hep-ex]].

[22] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) doi:10.1103/PhysRevLett.100.242001 [arXiv:0802.2470 [hep-ph]].

[23] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) doi:10.1140/epjc/s10052-012-1896-2 [arXiv:1111.6097 [hep-ph]].

[24] `http://fastjet.fr/repo/fastjet-doc-3.4.0.pdf` .

[25] `http://fastjet.fr/FAQ.html` .

[26] `https://pdg.lbl.gov/2019/reviews/rpp2019-rev-monte-carlo-numbering.pdf` .

[27] `https://github.com/delphes/delphes/blob/master/classes/DelphesClasses.h` .

[28] M. Cacciari, G. P. Salam and G. Soyez, JHEP **04**, 063 (2008) doi:10.1088/1126-6708/2008/04/063 [arXiv:0802.1189 [hep-ph]].

[29] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **08**, 001 (1997) doi:10.1088/1126-6708/1997/08/001 [arXiv:hep-ph/9707323 [hep-ph]].

[30] M. Wobisch and T. Wengler, [arXiv:hep-ph/9907280 [hep-ph]].

17 Appendix A : $H \mapsto \tau\bar{\tau}$ Decay mode

The final analysis was performed with the $H \mapsto b\bar{b}$ decay mode, however, we did consider the search for VBF-ZH through a different decay mode[25], namely $H \mapsto \tau\bar{\tau}$ mode. We would like to present excerpts of our findings with this particular decay mode, emphasizing the vitality of higher cross-section for a complete analysis. Therefore, how this study led us back to the $b\bar{b}$ decay mode.

Let us directly present the associated signal and background processes :

```
signal :  p p > z h j j QCD=0, z > l- l+, h > ta- ta+    ( 0.04366 fb )

bcg1 :  p p > z h j j QCD=0, z > l- l+, h > ta- ta+    ( 0.09199 fb )

bcg2 :  p p > z z j j QCD=0, z > ta- ta+ , z > l- l+    ( 0.2778 fb )

bcg3 :  p p > z z j j , z > ta- ta+ , z > l- l+    ( 2.002 fb )

bcg4 :  p p > w+ z j j j, w+ > vt ta+ , z > l- l+

  &&    p p > w- z j j j, w- > vt~ ta- , z > l- l+    ( 10.33 fb )

bcg5 :  p p > w+ w- z j j , z > l- l+ , w+ > vt ta+, w- > vt~ ta-    ( 0.01187 fb)
```

The corresponding cross-sections reported by MadGraph5 are also provided, where the processes were simulated with following conditions to be able to generate sizeable event samples. :

For signal, bcg1, bcg2, bcg3 :

```
set mmjj 100.0    # min invariant mass of a jet pair
set xetamin 0.5   # minimum rapidity for two jets in the WBF case
set deltaeta 1.0  # minimum rapidity difference for two jets in the WBF case
set ebeam1 6500   # Energy of beamline-1
set ebeam2 6500   # Energy of beamline-2
set kW 1.0        # κ_W value
set kZ 1.0        # κ_Z value
```

For bcg4,

```
set misset 10.0   # minimum missing Et (sum of neutrino's momenta)
set mmjj 50.0     # min invariant mass of a jet pair
set xetamin 0.5   # minimum rapidity for two jets in the WBF case
set deltaeta 1.0  # minimum rapidity difference for two jets in the WBF case
set ebeam1 6500   # Energy of beamline-1
set ebeam2 6500   # Energy of beamline-2
set kW 1.0        # κ_W value
set kZ 1.0        # κ_Z value
```

[25] Because of the huge background contributions in the $H \mapsto b\bar{b}$ decay mode.

For bcg5,

```
set misset 30.0   # minimum missing Et (sum of neutrino's momenta)
set mmjj 100.0    # min invariant mass of a jet pair
set xetamin 0.5   # minimum rapidity for two jets in the WBF case
set deltaeta 1.0  # minimum rapidity difference for two jets in the WBF case
set ebeam1 6500   # Energy of beamline-1
set ebeam2 6500   # Energy of beamline-2
set kW 1.0        # κ_W value
set kZ 1.0        # κ_Z value
```

With these conditions, we produce 100k event samples for all processes, except for bcg5 we have 78166[26] events. If we assume selection efficiency of 1, the event yields we are starting with are :

$S = 130.98, \quad b_1 = 275.97, \quad b_2 = 833.4 \quad b_3 = 6006, \quad b_4 = 30990, \quad b_5 = 35.61$

Now, if we apply a rudimentary set of cuts as follows :

- There must be at least 1 VBF-Tagging Jet pair candidate
- Number of MasterJets ≥ 2
- There must be at least 2 Delphes Tau-Tagged Jets, Both must have TauWeight[27] > 0.5

And perform the event selection with these, we will get event yields as follows :

$S = 9.72, \quad b_1 = 17.45, \quad b_2 = 50.97 \quad b_3 = 275.6754, \quad b_4 = 22.6227, \quad b_5 = 1.89$

Even though the background can be reduced significantly, observe that signal yield S falls down below 10 in almost just the basic cuts. This is because of extremely low cross-section of signal process, which can't be compensated even by the $\mathcal{L} = 3000\ fb^{-1}$. As a result, we start with just $S \sim 100$ to work with. In simple terms, because of the low cross-section, we don't have enough space to play around the signal process, and diminish the backgrounds within acceptable bounds.

This is because of the low branching ratio of $H \mapsto \tau\bar{\tau}$ decay mode, which is approximately 6%. On the other hand, $H \mapsto b\bar{b}$ dominates the decay channel of Higgs with branching fraction of order $\sim 58\%$. This allows us to work around the signal process and control the background contribution within acceptable bounds. Therefore, we ended up working with $b\bar{b}$ decay mode for our final analysis, with following process as the final signal process.

```
signal :  p p > z h j j QCD=0, h > b b~, z > l- l+
```

[26] We just fail to generate 100k, so work with what we have, because bcg5 is negligible anyways.
[27] Although TauWeight is a Truth level variable, we are using it just for investigation.

18 Appendix B : Additional plots

In the analysis strategy section, we directly presented few of the cuts namely the Jet-p_T cuts and missing E_T cut. Here we provide the normalized distributions for the corresponding variables, which motivated the final cuts. All the plots are produced with the Initial cuts in place, as described in the earlier section. We consider only the prominent processes namely signal,bcg2 and bcg5.

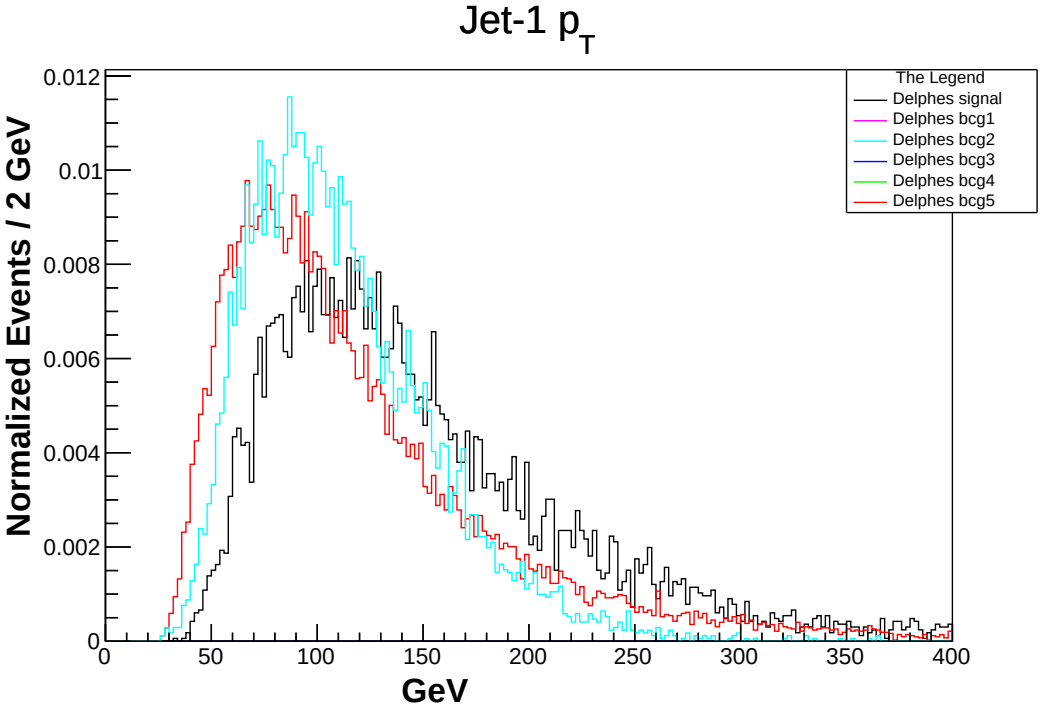

Figure 30: Normalized distribution for Jet-1 p_T .

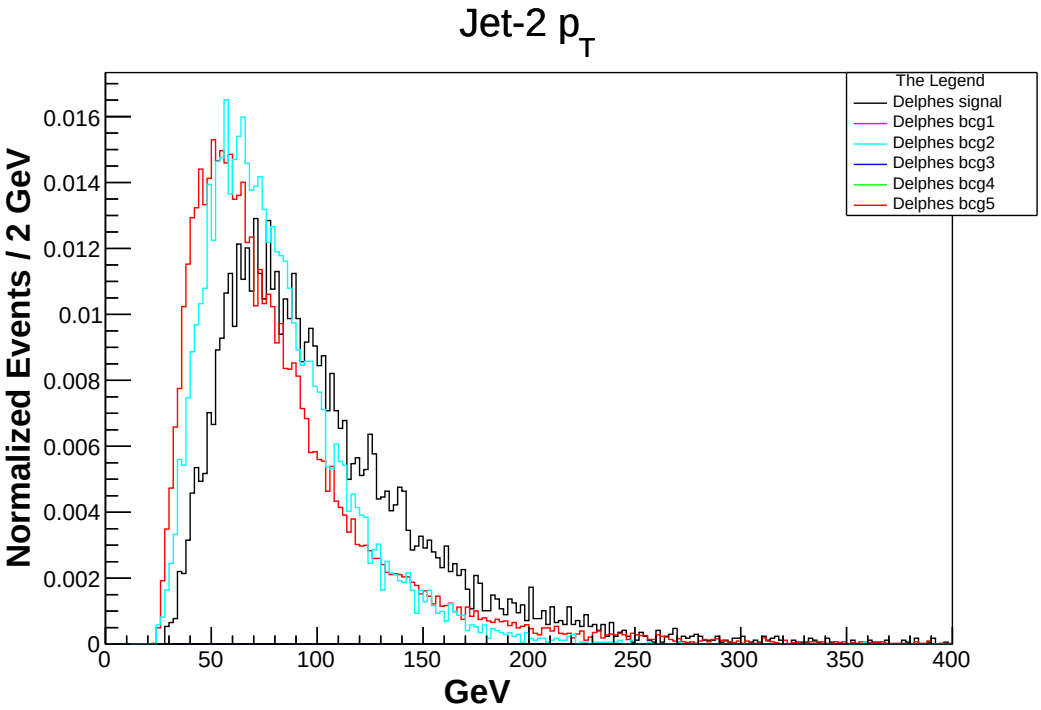

Figure 31: Normalized distribution for Jet-2 p_T .

The normal Jet-p_T distributions are represented in figures 30,31 and 32. Whereas the B-Jet-p_T distributions are in the figures 33 and 34. We prioritize the cuts on normal Jets over the B-Jets, because of their reliability over B-Jets which have may have B-Tagging inefficiencies associated with it. Therefore, harder cuts are placed over normal jets rather B-Jets. The missing E_T distribution is displayed in figure 35.

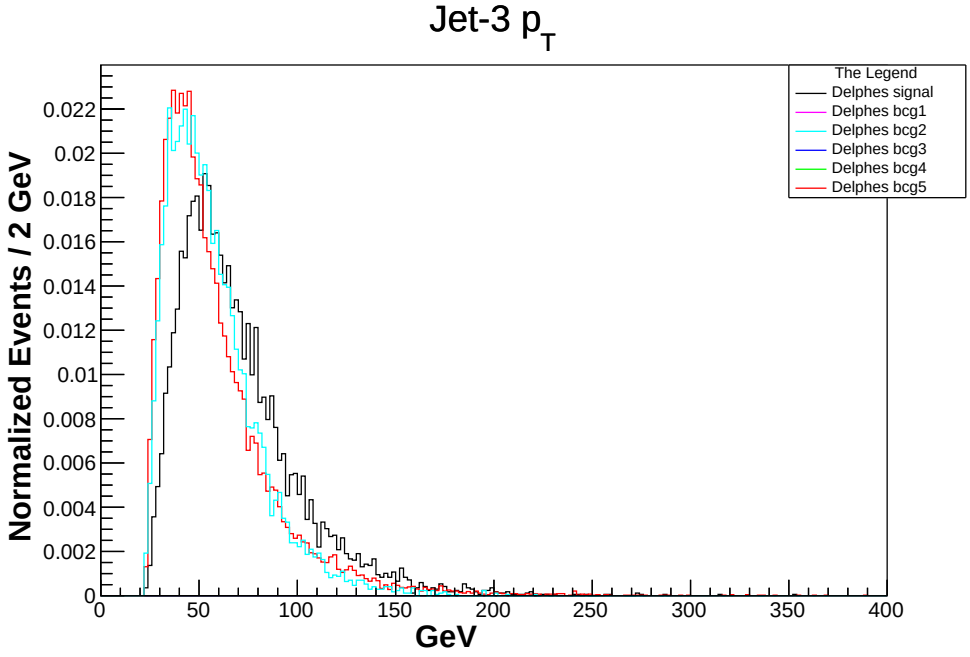

Figure 32: Normalized distribution for Jet-3 p_T .

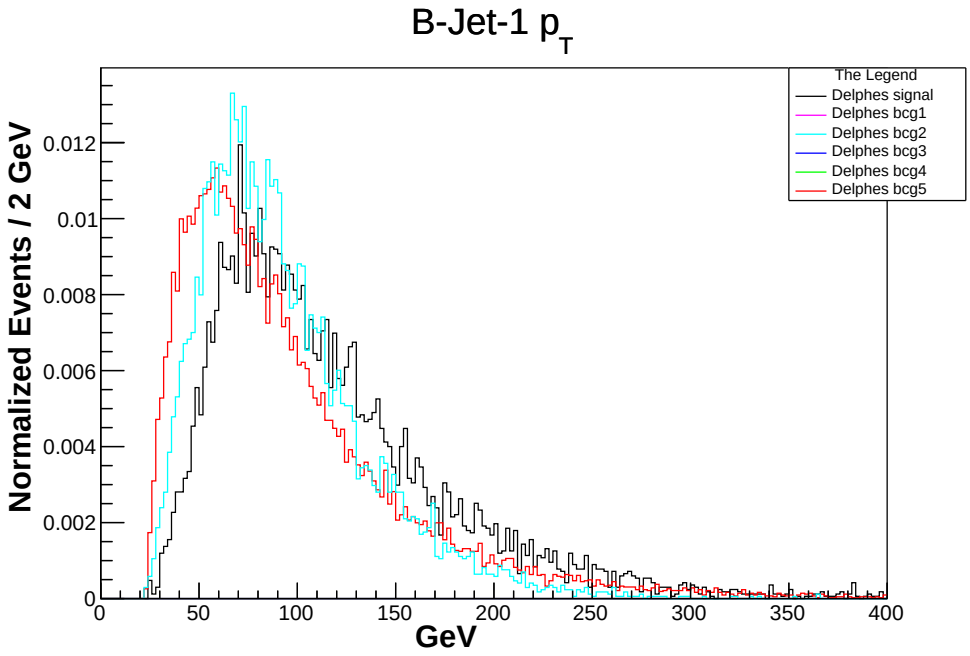

Figure 33: Normalized distribution for B-Jet-1 p_T .

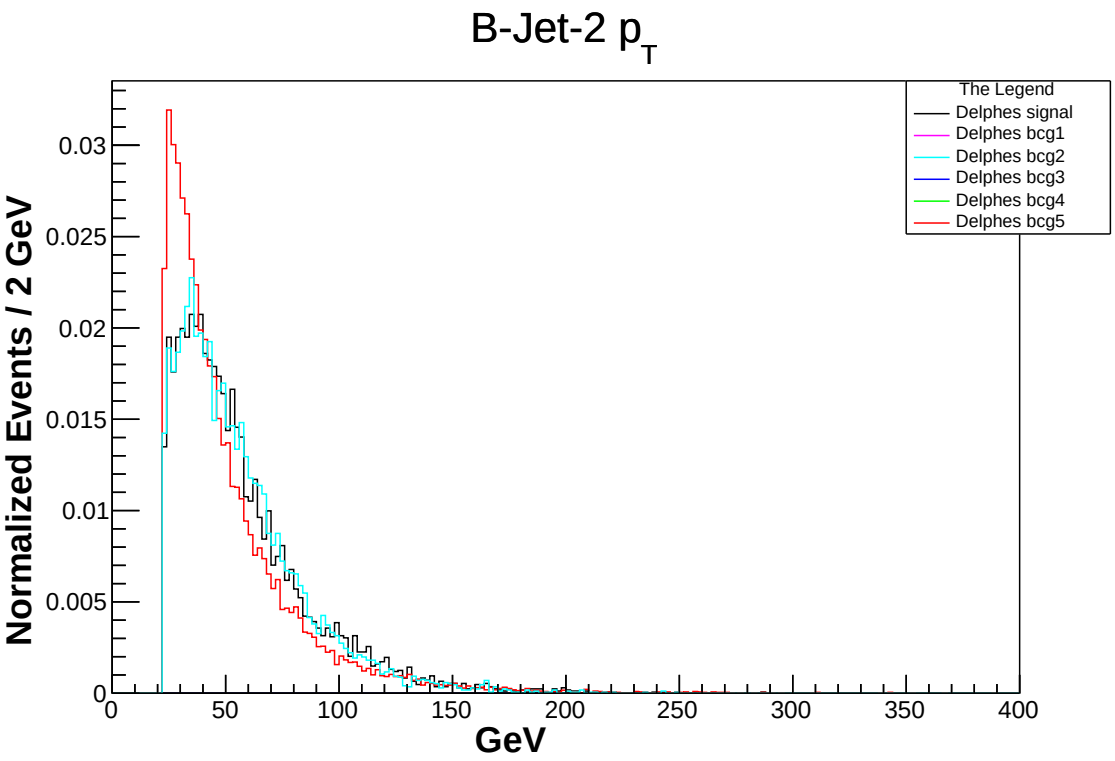

Figure 34: Normalized distribution for B-Jet-2 p_T .

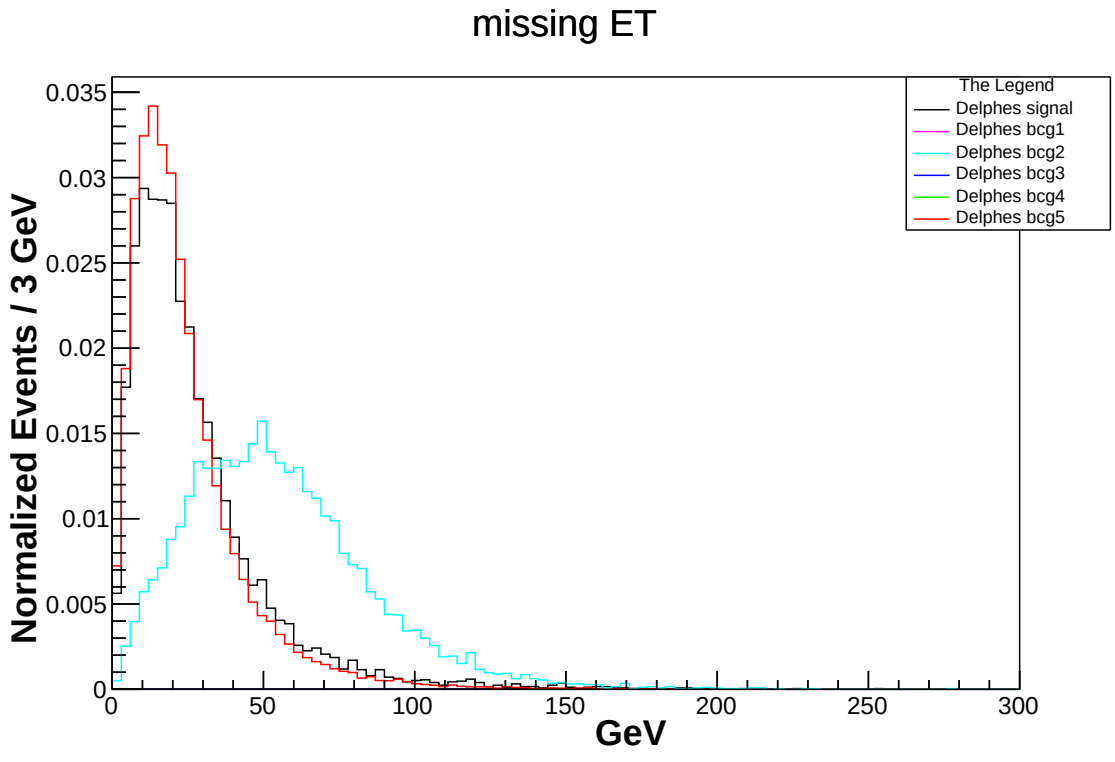

Figure 35: Normalized distribution for missing E_T .

19 Appendix C : (κ_W, κ_Z) Data-sets

There are total 2 data-sets PDF, one for each case. They are arranged in a chronological order as follows :

- Case (a) :
- Case (b) :

kW	kZ	Cross-section for Raw process (fb)	Cross-section with BR values fb (C)	Events selected out of 10k (d)	Signal Yield (S) = 3000*C*d/10k	Signal Yield Error = sqrt(S*S/d)	Net Background Yield (B)	Background Yield Error	Significance (σ)	Raw Cross-section Approximation
1	1	10.63	0.4130818	45	5.5766043	0.831311087	53.91031423	20.3597574	0.612209459	10.63
-2	2	138.9	5.397654	86	139.2594732	15.01673519	54.86174693	20.45130906	15.10837844	139.696
-2	1.5	113.3	4.402838	60	79.251084	10.23127095	54.30674452	20.40553323	8.657367873	112.90675
-2	1	90.23	3.5063378	59	62.06217906	8.079807505	53.91031423	20.3597574	6.813295512	90.301
-2	0.5	71.48	2.7777128	70	58.3319688	6.972003795	53.67245606	20.31398158	6.422925492	71.87875
-2	0	57.51	2.2348386	72	48.27251376	5.688970304	53.59317	20.26820575	5.320583215	57.64
-2	-0.5	47.64	1.8512904	56	31.10167872	4.15613664	53.67245606	20.31398158	3.424601796	47.58475
-2	-1	41.77	1.6231822	60	29.2172796	3.771934577	53.91031423	20.3597574	3.207524502	41.713
-2	-1.5	39.94	1.5520684	58	27.00599016	3.546060234	54.30674452	20.40553323	2.950127365	40.02475
-2	-2	42.43	1.6488298	44	21.76455336	3.281129874	54.86174693	20.45130906	2.361254866	42.52
-1	2	72.6	2.821236	68	57.5532144	6.979352414	54.86174693	20.45130906	6.243997078	72.172
-1	1.5	51.63	2.0063418	68	40.92937272	4.963415497	54.30674452	20.40553323	4.471114066	51.45625
-1	1	34.97	1.3589342	71	28.94529846	3.43517493	53.91031423	20.3597574	3.177665932	34.924
-1	0.5	22.59	0.8778474	77	20.27827494	2.310923902	53.67245606	20.31398158	2.232838214	22.57525
-1	0	14.3	0.555698	59	9.8358546	1.280519199	53.59317	20.26820575	1.084105194	14.41
-1	-0.5	10.39	0.4037554	56	6.78309072	0.906428625	53.67245606	20.31398158	0.746885236	10.42825
-1	-1	10.61	0.4123046	50	6.184569	0.874630136	53.91031423	20.3597574	0.678952896	10.63
-1	-1.5	14.93	0.5801798	51	8.87675094	1.242993755	54.30674452	20.40553323	0.969693972	15.01525
-1	-2	23.67	0.9198162	60	16.5566916	2.137459695	54.86174693	20.45130906	1.796249524	23.584
0	2	33.2	1.290152	78	30.1895568	3.418294184	54.86174693	20.45130906	3.27529064	33.468
0	1.5	18.83	0.7317338	67	14.70784938	1.796849786	54.30674452	20.40553323	1.606681654	18.82575
0	1	8.367	0.32514162	76	7.413228936	0.850355678	53.91031423	20.3597574	0.813837352	8.367
0	0.5	2.086	0.08106196	75	1.8238941	0.21060515	53.67245606	20.31398158	0.200828742	2.09175
0	-0.5	2.073	0.08055678	91	2.199200094	0.230538811	53.67245606	20.31398158	0.242153636	2.09175
0	-1	8.406	0.32665716	82	8.035766136	0.887401918	53.91031423	20.3597574	0.882180584	8.367
0	-1.5	18.75	0.728625	79	17.2684125	1.942848197	54.30674452	20.40553323	1.886396905	18.82575
0	-2	33.43	1.2990898	84	32.73706296	3.571906406	54.86174693	20.45130906	3.551671746	33.468
1	2	23.51	0.9135986	64	17.54109312	2.19263664	54.86174693	20.45130906	1.90304808	23.584
1	1.5	15	0.5829	43	7.51941	1.146699275	54.30674452	20.40553323	0.821418399	15.01525
1	1	10.63	0.4130818	45	5.5766043	0.831311087	53.91031423	20.3597574	0.612209459	10.63
1	0.5	10.43	0.4053098	50	6.079647	0.859791924	53.67245606	20.31398158	0.66942914	10.42825
1	0	14.41	0.5599726	63	10.58348214	1.333393417	53.59317	20.26820575	1.166508496	14.41
1	-0.5	22.65	0.880179	55	14.5229535	1.95827465	53.67245606	20.31398158	1.59912052	22.57525
1	-1	34.85	1.354271	87	35.3464731	3.789539687	53.91031423	20.3597574	3.880398178	34.924
1	-1.5	51.3	1.993518	73	43.6580442	5.109787578	54.30674452	20.40553323	4.769193431	51.45625
1	-2	71.88	2.7932568	78	65.36220912	7.400812829	54.86174693	20.45130906	7.091201543	72.172
2	2	42.34	1.6453324	43	21.22478796	3.23674982	54.86174693	20.45130906	2.30269526	42.52
2	1.5	39.88	1.5497368	42	19.52668368	3.013032705	54.30674452	20.40553323	2.133089864	40.02475
2	1	41.83	1.6255138	44	21.45678216	3.23473162	53.91031423	20.3597574	2.355563401	41.713
2	0.5	47.26	1.8365236	65	35.8122102	4.441957987	53.67245606	20.31398158	3.943277804	47.58475
2	0	57.58	2.2375588	71	47.66000244	5.656201673	53.59317	20.26820575	5.253072386	57.64
2	-0.5	71.68	2.7854848	65	54.3169536	6.737188923	53.67245606	20.31398158	5.980832691	71.87875
2	-1	90.35	3.511001	61	64.2513183	8.226538327	53.91031423	20.3597574	7.05362308	90.301
2	-1.5	113.2	4.398952	75	98.97642	11.42881255	54.30674452	20.40553323	10.81215846	112.90675
2	-2	139.7	5.428742	57	92.8314882	12.29583099	54.86174693	20.45130906	10.07136694	139.696
-0.5	0.5	8.724	0.33901464	76	7.729533792	0.886638334	53.67245606	20.31398158	0.85109796	8.731
-0.5	-0.5	2.65	0.102979	38	1.1739606	0.190441557	53.67245606	20.31398158	0.129264649	2.6575
0.5	0.5	2.664	0.10352304	47	1.459674864	0.212915462	53.67245606	20.31398158	0.160724609	2.6575
0.5	-0.5	8.676	0.33714936	78	7.889295024	0.893286757	53.67245606	20.31398158	0.868689248	8.731

kW	kZ	Cross-section for Raw process (fb)	Cross-section with BR values fb (C)	Events selected out of 10k (d)	Signal Yield (S) = 3000*C*d/10k	Signal Yield Error = sqrt(S*S/d)	Net Background Yield (B)	Background Yield Error	Significance (σ)	Raw Cross-section Approximation
1	1	10.63	0.4130818	45	5.5766043	0.831311087	53.91031423	20.3597574	0	10.63
-2	2	138.9	5.397654	86	139.2594732	15.01673519	54.86174693	20.45130906	13.8223680639204	139.696
-2	1.5	113.3	4.402838	60	79.251084	10.23127095	54.30674452	20.40553323	7.60456559465448	112.90675
-2	1	90.23	3.5063378	59	62.06217906	8.079807505	53.91031423	20.3597574	5.79914922388931	90.301
-2	0.5	71.48	2.7777128	70	58.3319688	6.972003795	53.67245606	20.31398158	5.39176336897522	71.87875
-2	0	57.51	2.2348386	72	48.27251376	5.688970304	53.59317	20.26820575	4.35085917349938	57.64
-2	-0.5	47.64	1.8512904	56	31.10167872	4.15613664	53.67245606	20.31398158	2.59613784653376	47.58475
-2	-1	41.77	1.6231822	60	29.2172796	3.771934577	53.91031423	20.3597574	2.42709403380367	41.713
-2	-1.5	39.94	1.5520684	58	27.00599016	3.546060234	54.30674452	20.40553323	2.24076968176656	40.02475
-2	-2	42.43	1.6488298	44	21.76455336	3.281129874	54.86174693	20.45130906	1.75963215074366	42.52
-1	2	72.6	2.821236	68	57.5532144	6.979352414	54.86174693	20.45130906	5.43391156920028	72.172
-1	1.5	51.63	2.0063418	68	40.92937272	4.963415497	54.30674452	20.40553323	3.67022801193959	51.45625
-1	1	34.97	1.3589342	71	28.94529846	3.43517493	53.91031423	20.3597574	2.39917082677138	34.924
-1	0.5	22.59	0.8778474	77	20.27827494	2.310923902	53.67245606	20.31398158	1.48494202518574	22.57525
-1	0	14.3	0.555698	59	9.8358546	1.280519199	53.59317	20.26820575	0.640720337013289	14.41
-1	-0.5	10.39	0.4037554	56	6.78309072	0.906428625	53.67245606	20.31398158	0.0994451824990266	10.42825
-1	-1	10.61	0.4123046	50	6.184569	0.874630136	53.91031423	20.3597574	0.0624173474006747	10.63
-1	-1.5	14.93	0.5801798	51	8.87675094	1.242993755	54.30674452	20.40553323	0.379512851566034	15.01525
-1	-2	23.67	0.9198162	60	16.5566916	2.137459695	54.86174693	20.45130906	1.22496168177132	23.584
0	2	33.2	1.290152	78	30.1895568	3.418294184	54.86174693	20.45130906	2.62459368488251	33.468
0	1.5	18.83	0.7317338	67	14.70784938	1.796849786	54.30674452	20.40553323	0.978168480744324	18.82575
0	1	8.367	0.32514162	76	7.413228936	0.850355678	53.91031423	20.3597574	0.188558980648169	8.367
0	0.5	2.086	0.08106196	75	1.8238941	0.21060515	53.67245606	20.31398158	0.40969573771848	2.09175
0	-0.5	2.073	0.08055678	91	2.199200094	0.230538811	53.67245606	20.31398158	0.371164564075988	2.09175
0	-1	8.406	0.32665716	82	8.035766136	0.887401918	53.91031423	20.3597574	0.252472377305499	8.367
0	-1.5	18.75	0.728625	79	17.2684125	1.942848197	54.30674452	20.40553323	1.24105129523581	18.82575
0	-2	33.43	1.2990898	84	32.73706296	3.571906406	54.86174693	20.45130906	2.88613598913901	33.468
1	2	23.51	0.9135986	64	17.54109312	2.19263664	54.86174693	20.45130906	1.32602631808413	23.584
1	1.5	15	0.5829	43	7.51941	1.146699275	54.30674452	20.40553323	0.240160052082216	15.01525
1	1	10.63	0.4130818	45	5.5766043	0.831311087	53.91031423	20.3597574	0	10.63
1	0.5	10.43	0.4053098	50	6.079647	0.859791924	53.67245606	20.31398158	0.0272254887542012	10.42825
1	0	14.41	0.5599726	63	10.58348214	1.333393417	53.59317	20.26820575	0.481476328891886	14.41
1	-0.5	22.65	0.880179	55	14.5229535	1.95827465	53.67245606	20.31398158	0.894066119073711	22.57525
1	-1	34.85	1.354271	87	35.3464731	3.789539687	53.91031423	20.3597574	3.05635399315022	34.924
1	-1.5	51.3	1.993518	73	43.6580442	5.109787578	54.30674452	20.40553323	3.95036980587369	51.45625
1	-2	71.88	2.7932568	78	65.36220912	7.400812829	54.86174693	20.45130906	6.2356300131323	72.172
2	2	42.34	1.6453324	43	21.22478796	3.23674982	54.86174693	20.45130906	1.70421652330282	42.52
2	1.5	39.88	1.5497368	42	19.52668368	3.013032705	54.30674452	20.40553323	1.47289909560131	40.02475
2	1	41.83	1.6255138	44	21.45678216	3.23473162	53.91031423	20.3597574	1.63035468655359	41.713
2	0.5	47.26	1.8365236	65	35.8122102	4.441957987	53.67245606	20.31398158	3.07974942332847	47.58475
2	0	57.58	2.2375588	71	47.66000244	5.656201673	53.59317	20.26820575	4.28797510636767	57.64
2	-0.5	71.68	2.7854848	65	54.3169536	6.737188923	53.67245606	20.31398158	4.97955775504362	71.87875
2	-1	90.35	3.511001	61	64.2513183	8.226538327	53.91031423	20.3597574	6.02389941423146	90.301
2	-1.5	113.2	4.398952	75	98.97642	11.42881255	54.30674452	20.40553323	9.62968735282647	112.90675
2	-2	139.7	5.428742	57	92.8314882	12.29583099	54.86174693	20.45130906	9.05579158524428	139.696
-0.5	0.5	8.724	0.33901464	76	7.729533792	0.886638334	53.67245606	20.31398158	0.196612719324233	8.731
-0.5	-0.5	2.65	0.102979	38	1.1739606	0.190441557	53.67245606	20.31398158	0.476421875350597	2.6575
0.5	0.5	2.664	0.10352304	47	1.459674864	0.212915462	53.67245606	20.31398158	0.447088755853198	2.6575
0.5	-0.5	8.676	0.33714936	78	7.889295024	0.893286757	53.67245606	20.31398158	0.213014848156364	8.731