Studies in boson reconstruction for SUSY searches

Jesús Aguado López

Contents

Abstract

The aim of this study is to compare different algorithms of reconstructionf of Higgs bosons in events $hh \to 4b$. This study has been developed as part of a summer student program from Instituto de Física de Altas Energías (Barcelona), and supervised by Chiara Rizzi.

1 Introduction

This study is a lateral development to the search for higgsino pair production in the ATLAS Collaboration, withinthe SUSY-EWK research group.

Higgsinos are particles predicted by Supersymmetry (SUSY), a theory beyond the Standard Model generalizing new symmetries and predicting so called supersimetric partners to the Standard Model particles. The current search focuses on looking for the lightest supersymmetric partner (LSP) and next to lightest supersymmetric partner (NSLP), which are the particles predicted by SUSY with lowest mass of them. In this model, the LSP role would be played by the higgsino or neutralino $\tilde{\chi}_1^0$ and the NSLP would be the gravitino $\tilde{G}$.

With a simplified model, the target event to look at would be the one in Figure 1.

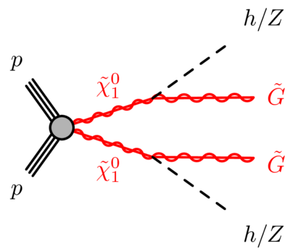


Figure 1: Decay process for the model

The main target would be the case of each of the neutralinos decaying into a Higgs boson and a gravitino. The Higgs would then decay, and with the high branching ratio of 58% for $h \to b\bar{b}$ it would be sensible to look for final states with 4 b-jets. It is also possible a decay into Z boson. This could also decay to two b quarks, but with smaller branching ratio, $\mathrm{BR}(Z \to b\bar{b}) \approx 15\%$.

The search makes use of data recorded by the ATLAS detector during the LHC Run 2 with $\sqrt{s}$ = 13 TeV and integrated luminosity of 36.1 fb^{-1}.

2 Details of the study

2.1 Samples

The Monte Carlos simulations used were produced using MadGraph5 v2.2.2 at leading order and Pythia v8.212, as specified in [1]. However, a different derivation was used: unskimmed TOPQ1 instead of SUSY 10.

2.2 Definitions

- $\mathrm{E}_T^{\mathrm{miss}}$, missing transverse momentum. It is the magnitud of the the vector sum of the transversal 4-momenta of the physical objects involved in the event. By transversal we mean perpendicular to the beam. It is the signature of particle that passes through the detectors without interacting (e.g. neutrinos or possibly new particles).
- $\Delta R(h)$ is a measure of the angular distance between two jets coming from the same candidate boson. Its definition is $\Delta R(h) = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2}$, where φ is the azimuthal angle through the beam's axis and η is the pseudorapidity of cited jet.
- m_T^{b-min}, minimum transverse mass between the $\mathrm{E}_T^{\mathrm{miss}}$ and the 3 leading b-jets.

$$m_T^{b-min} = \min_{i \geq 4}\sqrt{(\mathrm{E}_T^{\mathrm{miss}} + p_T^{j_i})^2 - (\mathrm{E}_T^{\mathrm{miss}}{}_x + p_x^{j_i})^2 - (\mathrm{E}_T^{\mathrm{miss}}{}_y + p_y^{j_i})^2}$$

2.3 Identifying

It is critical to assign the jets as coming from the bosons or as secondary products of the collision. b-tagging the jets is a important task. Previously, this had been done as assigning the jets with highest MV2c10 value as b-jets coming from the higgs. This algorithm showed good performance. However, it relies on uncalibrated information (the MV2c10 ordering), thus it can't be extended to analysis of data.

2.4 Matching

Once the selected jets are chosen, it is important to match them in pairs and associate each pair to a specific boson candidate. To do this, several approaches have been contemplated, all of them related to minimizing certain function:

- In the $hh \rightarrow 4b$ events, minimize the difference between one of the reconstructed masses and 115 GeV (value defined with signal studies), to target hh events. See [1] section 2.5.
- In events with same type of bosons, minimize the mass difference between the two boson candidates, to target events with same type of bosons.
- Minimize $\max\{\Delta R(h_1), \Delta R(h_2)\}$
- Maximize p_T of the bosons via the function $\min\{p_T(h_1), p_T(h_2)\}$

3 New algorithms

3.1 Jet choosing algorithm

A new algorithm was implemeneted for choosing the jets as coming from the boson candidates. Instead of picking the jets with highest MV2c10 value, this new algorithm uses also the information about p_T, as follows:

- If there are precisely 4 b-tagged jets, choose them.
- If there are more than 4 b-tagged jets, choose the ones with higher p_T.
- If there are less than 4 b-tagged jets, choose them and complete with non b-tagged jets with higher p_T.

The jets are considered to be b-tagged with the MV2c10 value over the 77% operating point. The performance of this new algorithm can be compared to the previous one simply by studying how much this reconstruction yields a correct match (we can do this when working with MC samples since we have the information at truth level).

3.2 Exploring new matching algorithms

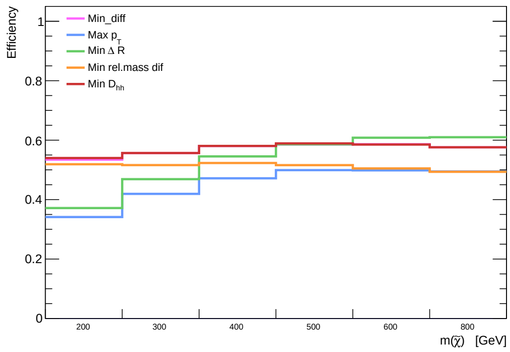
Three additional algorithms for matching the chosen jets were tested to expand the previous ones explained in section 2.4.

- Minimizing relative mass difference. Instead of simply minimizing $m(h_1) - m(h_2)$, an algorithm was tested minimizing
$$\frac{m(h_1) - m(h_2)}{m(h_1) + m(h_2)}$$
The first boson is always chosen as leading in mass.
- Minimizing the function D_{hh} defined as:
$$D_{hh} = \sqrt{m_{lead}^2 + m_{sublead}^2}|sin(\beta)|$$
where $\beta = arctan\left(\frac{m_{lead}}{m_{sublead}}\right) - arctan\left(\frac{120}{110}\right)$

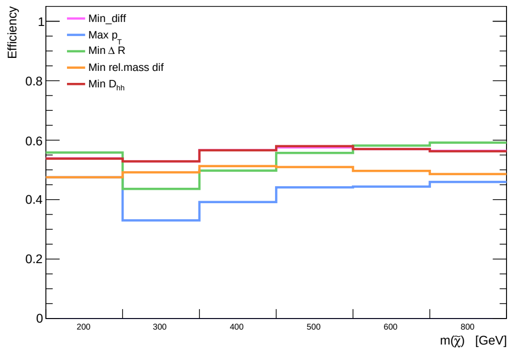
3.3 Comparative of plots: MV2c10 algorithm

Plots in Figure 2 show the fraction of events in which the different matching algorithms yield the correct reconstruction for $hh \to 4b$. This has been done selecting the four jets with the highest value of MV2c10.

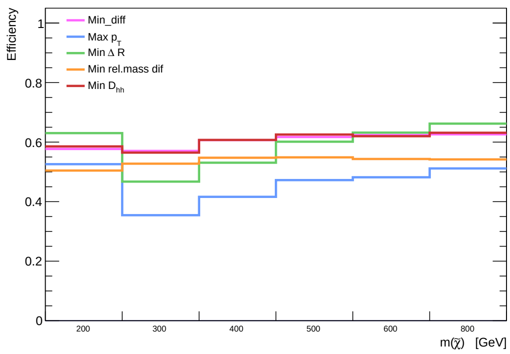
- Selection 1: at least 4 jets.
- Selection 2: $\mathrm{E}_T^{\mathrm{miss}} \geq 200$ GeV, pass the $\mathrm{E}_T^{\mathrm{miss}}$ trigger and at least 4 jets.
- Selection 3: $\mathrm{E}_T^{\mathrm{miss}} \geq 200$ GeV, pass the $\mathrm{E}_T^{\mathrm{miss}}$ trigger at least 4 jets and 3 b-tagged jets.



(a) Selection 1



(b) Selection 2



(c) Selection 3

Figure 2: Good match for different selections and matching algorithms

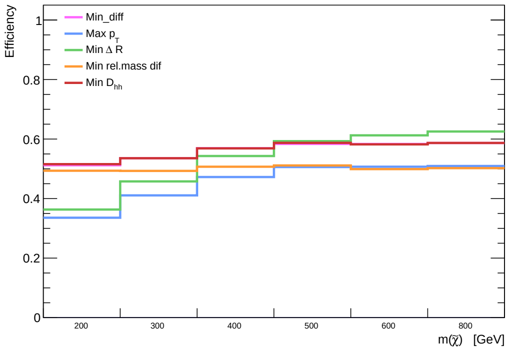
In these plots it was imposed that the match is possible, which is a truth level variable.

3.4 Comparative of plots: p_T based algorithm

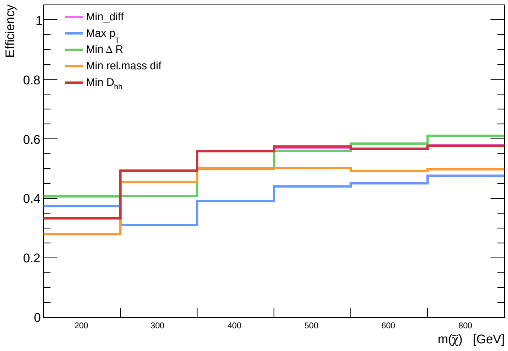
Same plots as before but for the new jet choosing algorithm.

In Figure 3 the same plots as in Figure 2 are reported, but using the jets selected with the new algorithm described in section 3.1, again for $hh \to 4b$ events.

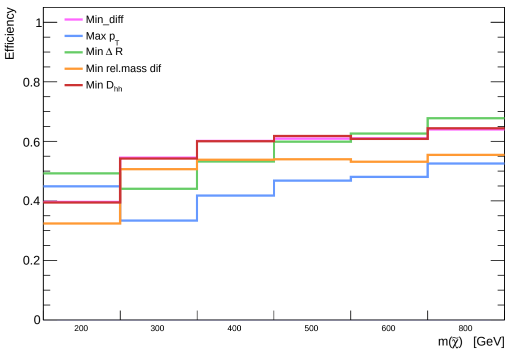
- Selection 1: at least 4 jets.
- Selection 2: $\mathrm{E}_T^{\mathrm{miss}} \geq 200$ GeV, pass the $\mathrm{E}_T^{\mathrm{miss}}$ trigger and at least 4 jets.
- Selection 3: $\mathrm{E}_T^{\mathrm{miss}} \geq 200$ GeV, pass the $\mathrm{E}_T^{\mathrm{miss}}$ trigger at least 4 jets and 3 b-tagged jets.



(a) Selection 1



(b) Selection 2



(c) Selection 3

Figure 3: Good match for different selections and matching algorithms

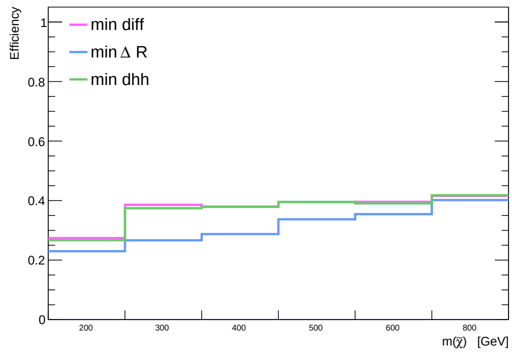
3.5 Interpretations

It is clear that minimizing ΔR works better for higher signal masses (700-800 GeV). Besides that, minimizing D_{hh} and mass difference have a very similar performance, and they perform well for intermediate mass regime. From now on we shall only consider the three best performing algorithms, D_{hh}, mass difference and ΔR.

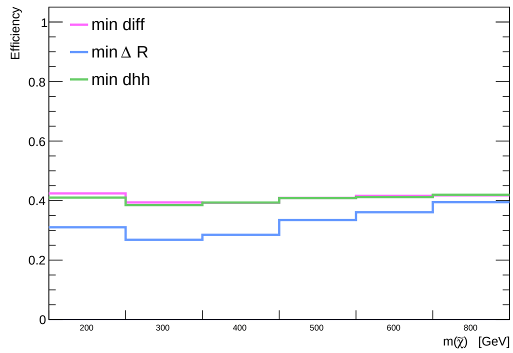
It should be noted that MV2c10 and p_T based reconstruction have a similar performance. Since it uses only calibrated information, the p_T based algorithm is preferred.

4 Variable distributions

Plots for kinematic variables are shown in Figure 4, comparing the different distributions for the two main jet choosing algorithm. All of them are produced for the third type of selection which is $hh \rightarrow 4b$, $E_T^{miss} \geq 200$ GeV, pass the E_T^{miss} trigger at least 4 jets and 3 b-tagged jets.



(a) p_T based



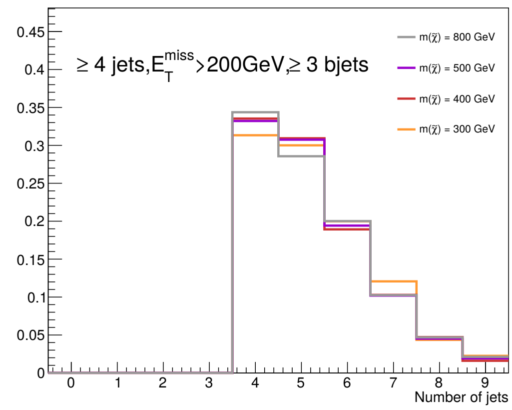
(b) MV2c10 based

Figure 4: Fraction of events with both bosons with reconstructed mass falling in 100-150 GeV (90-140 GeV for second Higgs)

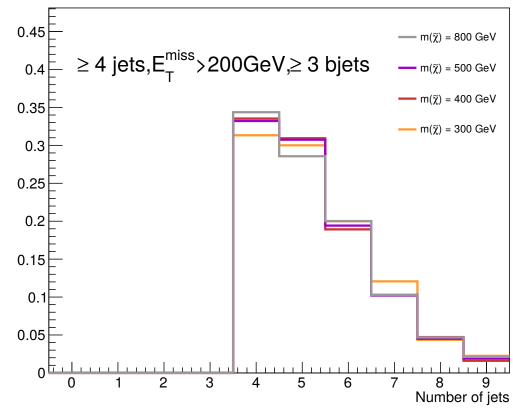
Here it can be visualized how the reconstruction works better for higher signal masses, not only in the sense of properly identifying and matching the jets but also reconstruting a mass of the candidate bosons similar to what is expected.

4.1 Jets

The number of jets and b-tagged jets are approximately equal for different signal masses as it can be seen in Figure 5 and Figure 6.

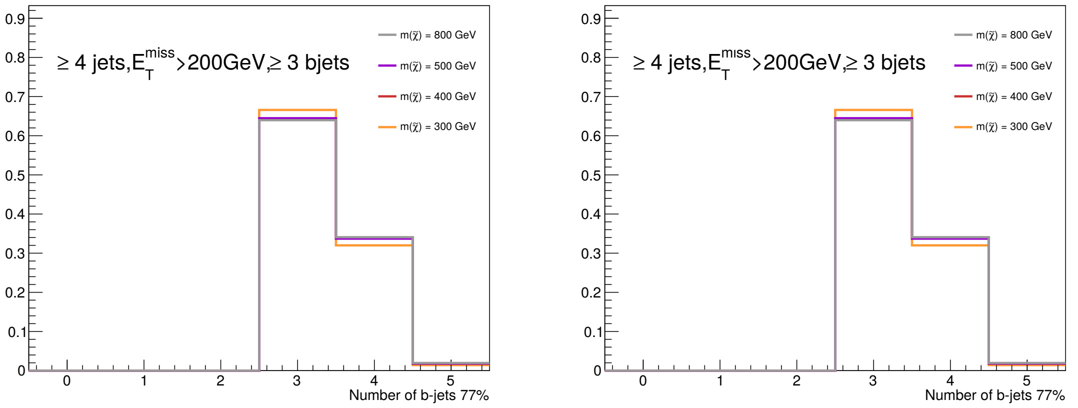


(a) p_T based



(b) MV2c10 based

Figure 5: Number of jets

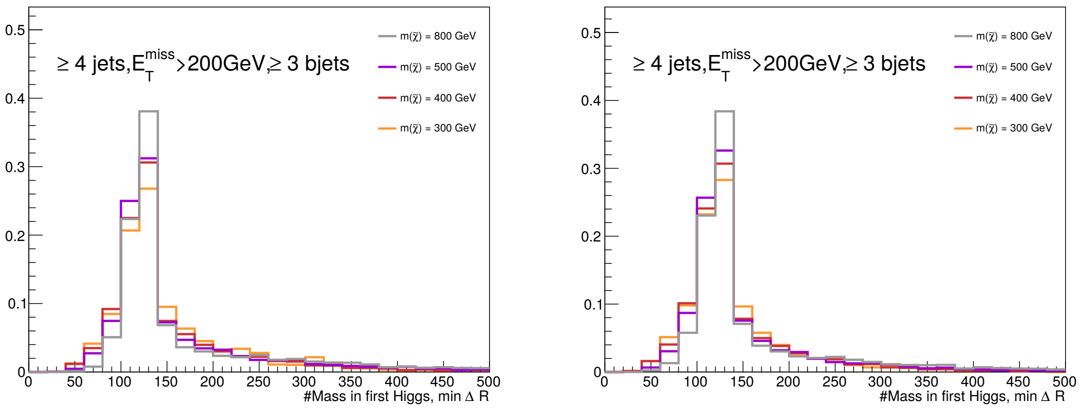


(a) p_T based

(b) MV2c10 based

Figure 6: Number of b-tagged jets

4.2 Mass distributions

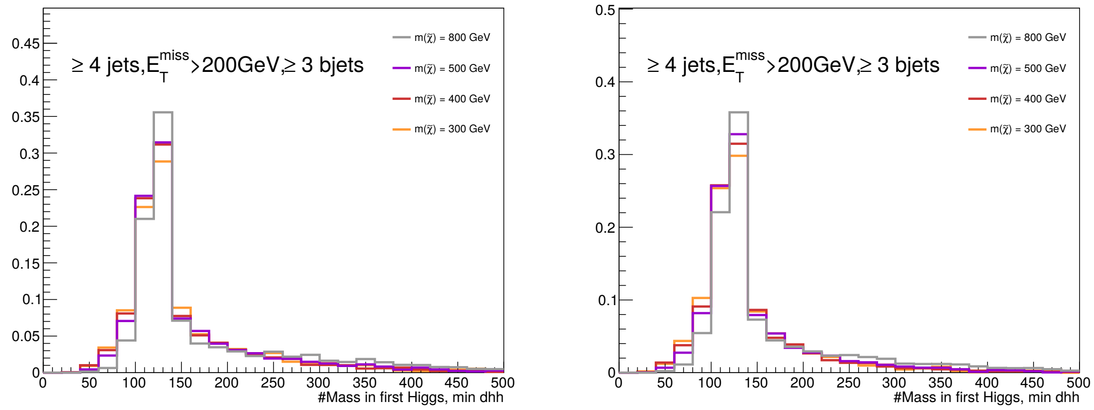


(a) p_T based

(b) MV2c10 based

Figure 7: Mass of first Higgs, min ΔR

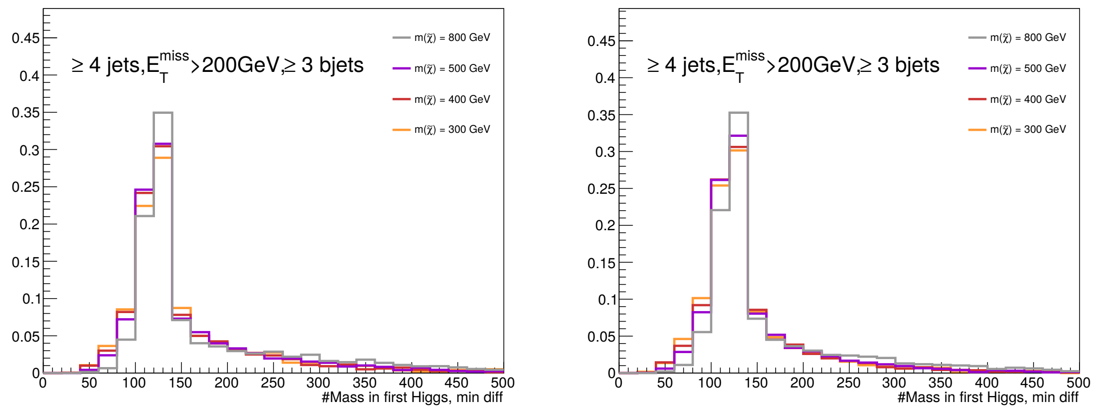
Mass distributions in Figure 7 peak around 125 GeV. Mass of the Higgs candidate of higher mass is properly reconstructed on average. Mass of the Higgs candidate with sub-leading mass tends to be understimated for all algorithms, being the distribution slightly shifted to lower values.



(a) p_T based

(b) MV2c10 based

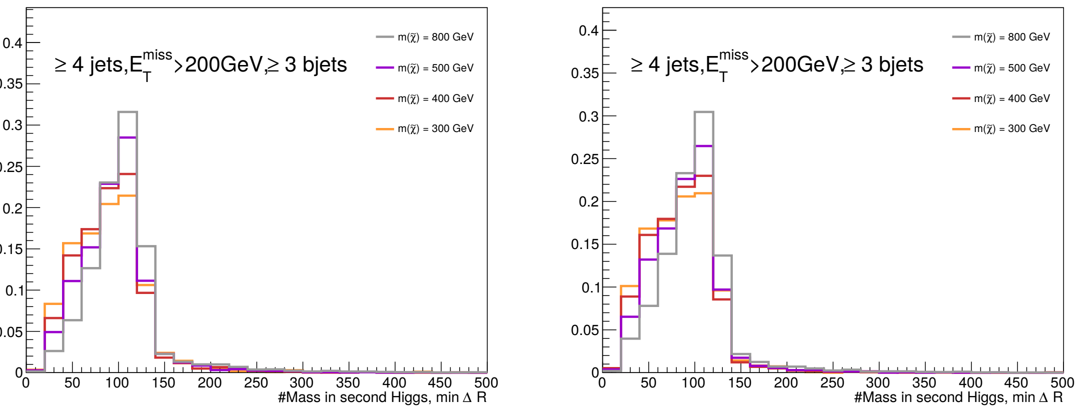
Figure 8: Mass of first Higgs, min D_{hh}



(a) p_T based

(b) MV2c10 based

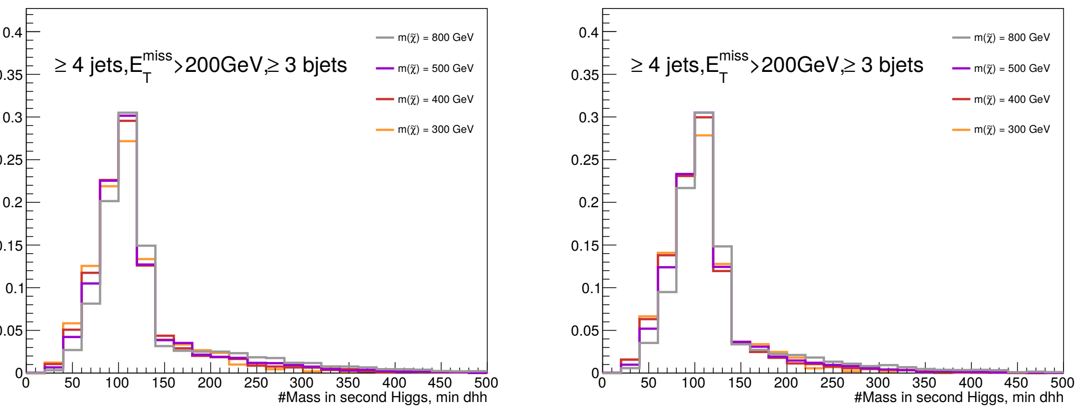
Figure 9: Mass of first Higgs, min mass difference



(a) p_T based

(b) MV2c10 based

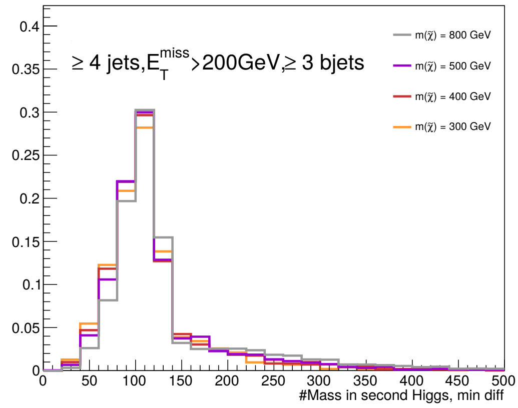
Figure 10: Mass of second Higgs, min ΔR



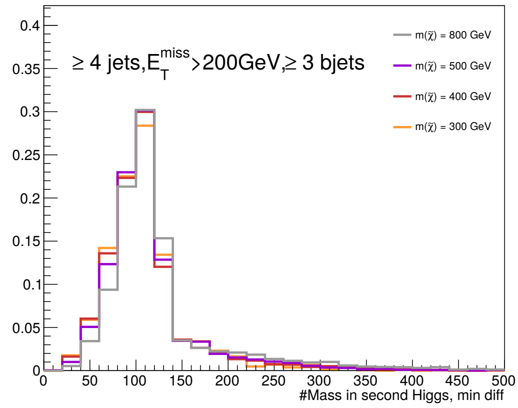
(a) p_T based

(b) MV2c10 based

Figure 11: Mass of second Higgs, min D_{hh}



(a) p_T based



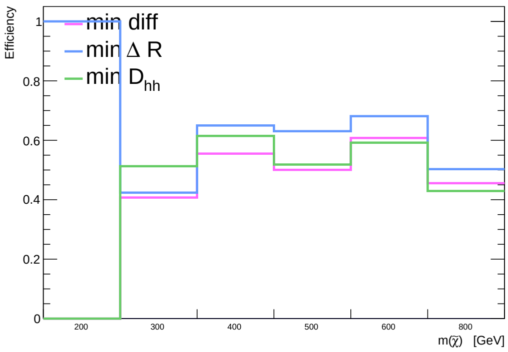
(b) MV2c10 based

Figure 12: Mass of second Higgs, min mass difference

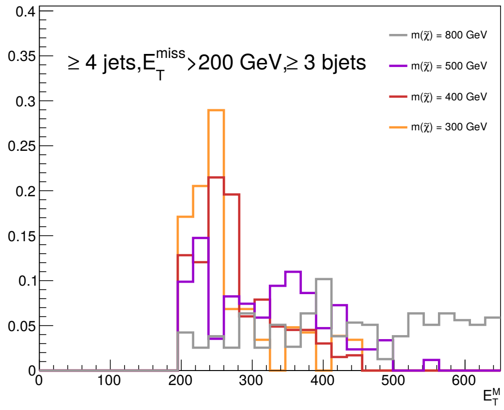
From now on we shall consider only plots for the p_T based jet-choosing algorithm.

5 ZZ case

The studies were extended to target the decays $Zh \to 4b$ and $ZZ \to 4b$. Plots in Figures 13, 14, 15 show the distribution of relevant variables for the ZZ signal.

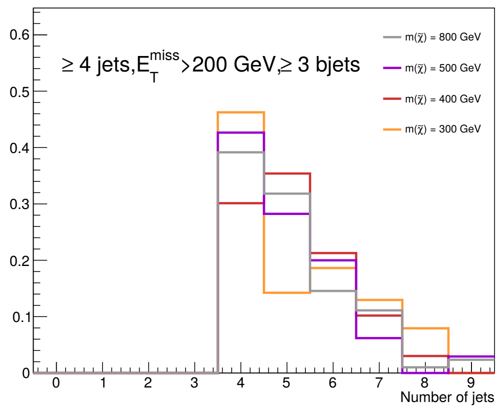


(a) Fraction of events with good match

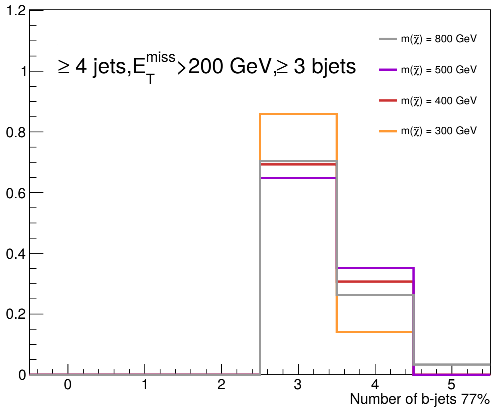


(b) E_T^{miss} distribution

Figure 13: ZZ selection 3 plots

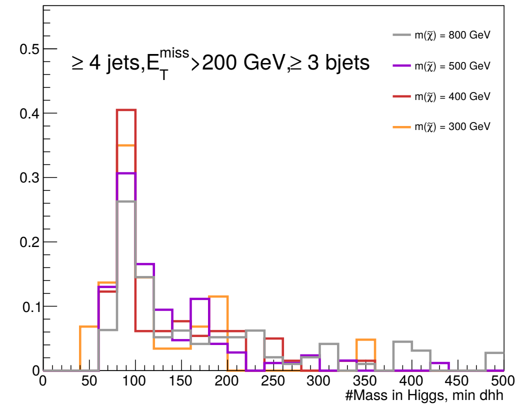


(a) Number of jets

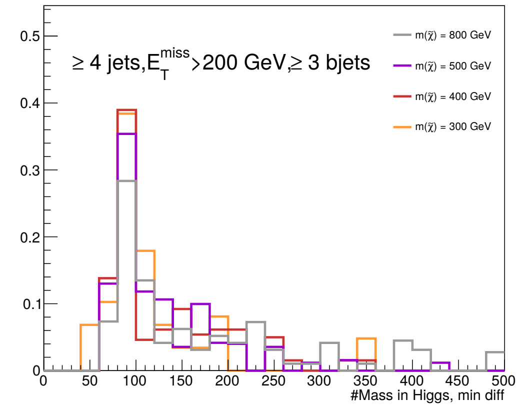


(b) Number of b-jets at 77% OP

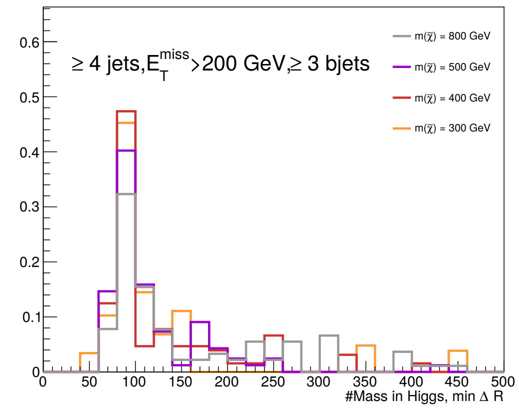
Figure 14: ZZ plots, jets



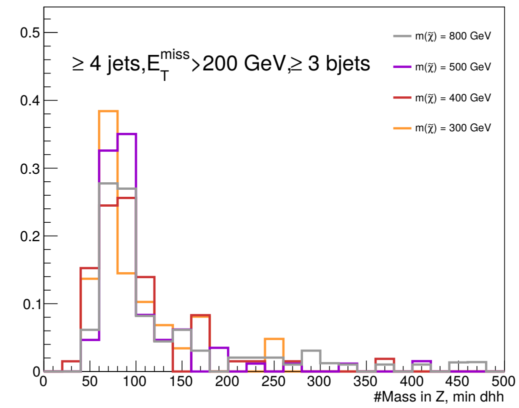
(a) Mass of Z_1 min D_{hh}



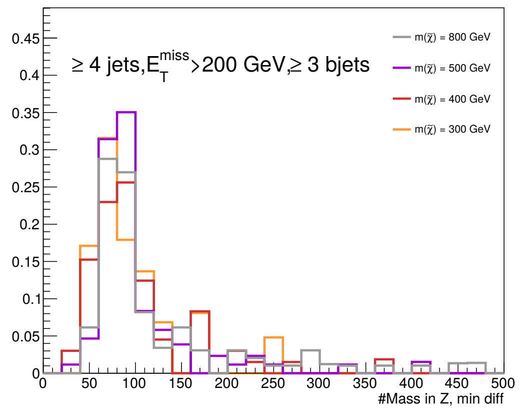
(b) Mass of Z_1, min mass difference



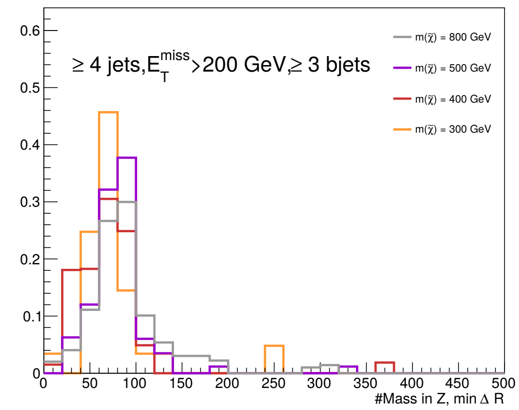
(c) Mass of Z_1 min ΔR



(d) Mass of Z_2 min D_{hh}



(e) Mass of Z_2 min mass difference

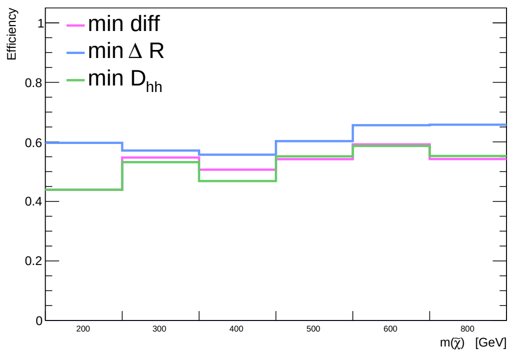


(f) Mass of Z_2 min ΔR

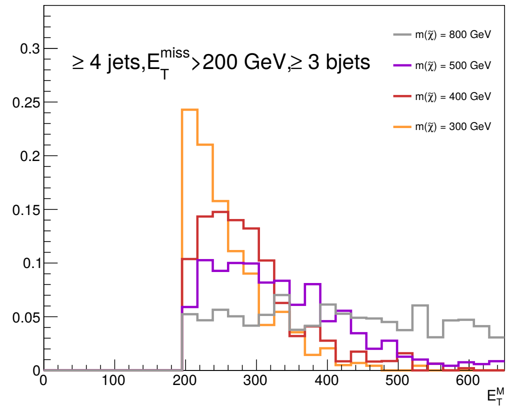
Figure 15: ZZ plots, mass distributions

6 Zh case

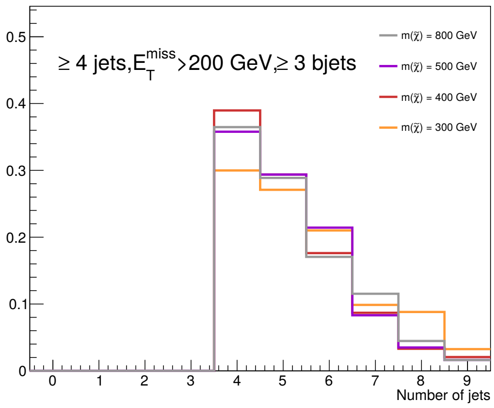
The plots in Figures 16 and 17. shwo the distribution of the relevant variables for the $Zh \rightarrow 4b$ signal. Same salection is applied in the study of ZZ case $E_T^{miss} \geq 200$ GeV, pass the E_T^{miss} trigger, at least 4 jets and 3 b-tagged jets.



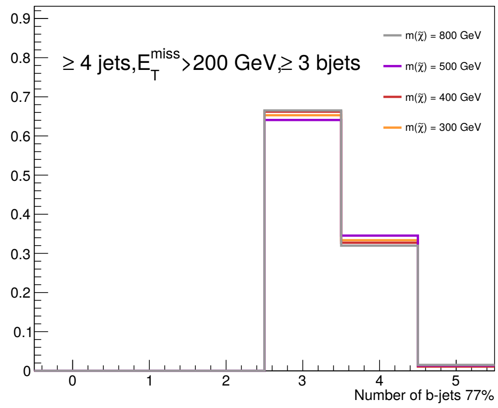
(a) Fraction of events with good match



(b) E_T^{miss} distribution

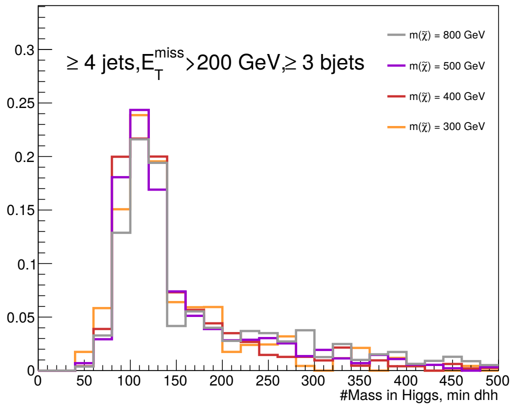


(c) Number of jets

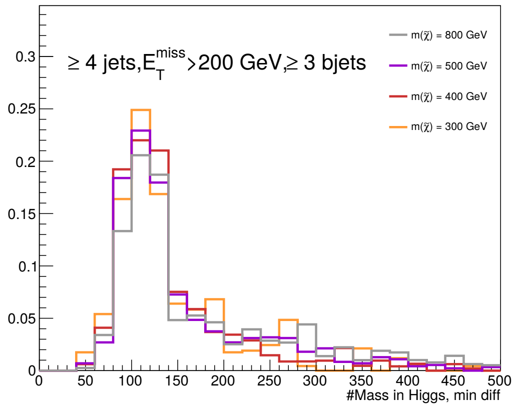


(d) Number of b-jets at 77% OP

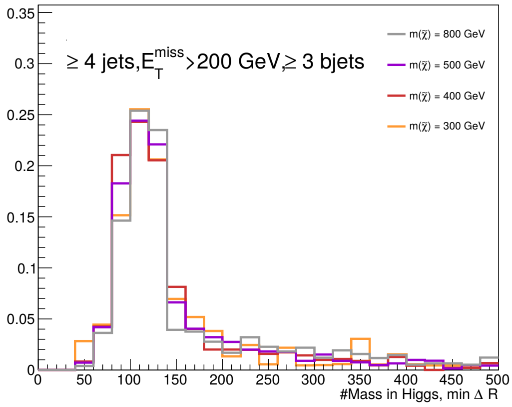
Figure 16: Zh selection 3 plots



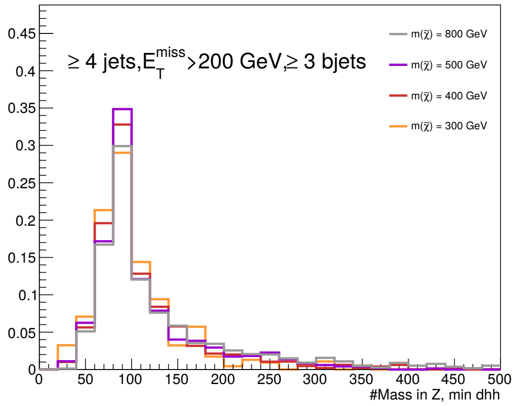
(a) Mass of Z_1 min D_{hh}



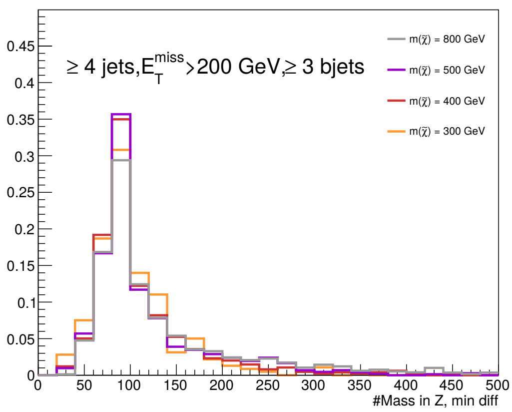
(b) Mass of Z_1, min mass difference



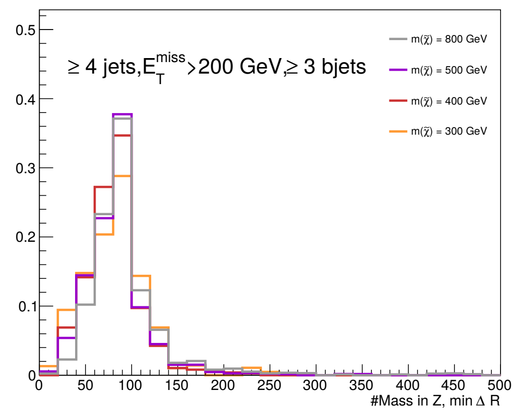
(c) Mass of Z_1 min ΔR



(d) Mass of Z_2 min $D_h h$



(e) Mass of Z_2 min mass difference

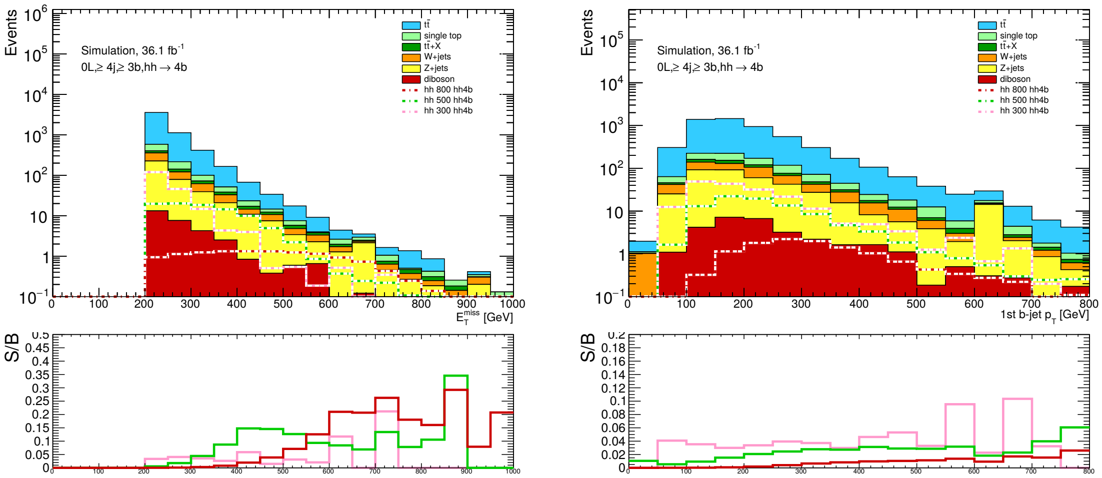


(f) Mass of Z_2 min ΔR

Figure 17: Zh plots, mass distributions

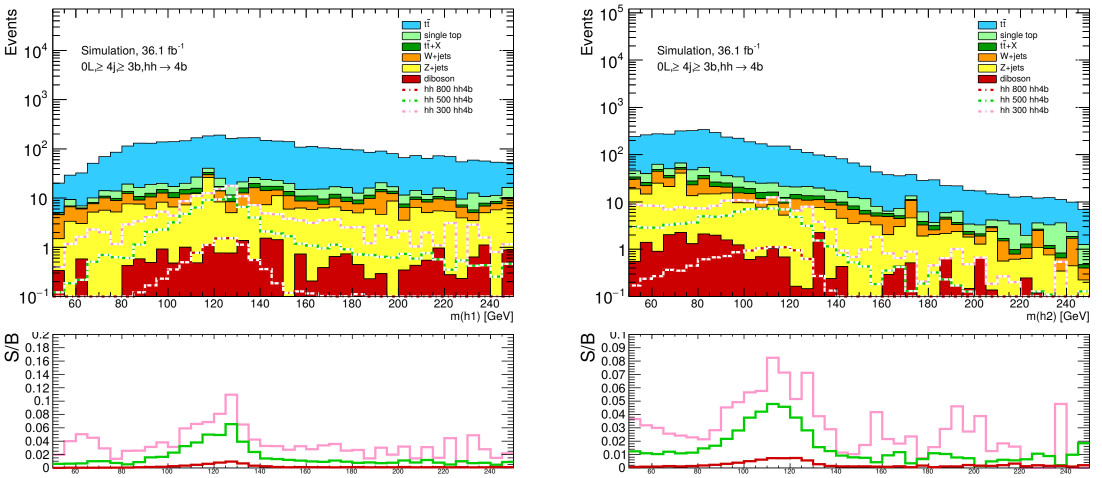
7 Signal - background estimate

The plots in Figure 18 compare signal and background events with the selection: between 4 and 6 jets, no leptons (0L), $E_T^{miss} \geq 200$ GeV, $\Delta\varphi_{min} > 0.4$, at least 3 b-tagged jets. This can be used to better understand which variables are more useful for discriminating signal and background.



(a) E_T^{miss} distribution

(b) p_T of leading jet



(c) Mass h_1, min ΔR

(d) Mass h_2, min ΔR

Figure 18: Signal-background plots

8 Conclusions

Several algorithms have been implemented with the aim of better reconstruct from events. There are two different phases.

First, a new jet choosing algorithm has been developed. This algorithm relies on missing transverse momentum to choose candidate b-jets. Its performance has been compared with the previous MV2c10 based algorithm. While having similar performance, the newly developed one has the advantage of making use of calibrated information only.

Then, several ways of matching the selected jets have been studied. This was implemented through a minimizing (or maximizing) process of different functions of variables such as ΔR, $m_T^{b-\mathrm{min}}$, E_T^{miss}, relative and absolute mass difference of reconstructed candidate bosons and D_{hh} function as explained in section 3.2. The comparison of their performance has been presented in the working analysis meetings and have been of insight for the group, yielding a better understanding of the reconstruction techniques and how different algorithms perform better depending upon the signal of interest. In the case of this study, minimizing ΔR is useful for high signal masses, while in intermediate mass regime minimizing mass difference and the function D_{hh} perform better.

All of this has been used for the proposed $hh \to 4b$ signal, but an extension to $ZZ \to 4b$ and $Zh \to 4b$ cases was relatively straightforward. In every case, different variables have been studied. The reconstructed mass of candidate mass is a variable of interest.

Finally, a simulation of signal-background was performed in order to understand which variables are better to distinguish between signal background.

References

[1] *Search for higgsino pair production in the hh topology in final states with $\geq$ 3 b-jets*
ATLAS Internal Note, July 2017. ATL-COM-PHYS-2017-824.