Search for new physics with fat jets at LHCb

CERN Summer Student Program 2021

Ida-Marie Fauske Johansson[1]
Supervisors: Lorenzo Sestini[2] and Carlos Vázquez Sierra[3]

Abstract

The search for new physics at the LHCb includes the study and reconstruction of fat $b\bar{b}$-jets. These fat jets have a large radius and a rich substructure, and taking a closer look at the internal structure of the jet may help us in uncovering new physics. In fact, with fat jet techniques, it is possible to search for low mass particles decaying in boosted $b\bar{b}$ quarks. This study applies the jet substructure technique called SoftDrop, to groom and clean up the jets by removing background and underlying events. After applying SoftDrop, its effect on the jet is studied through two analysis. First the mass distribution of the jet after applying SoftDrop is compared to the original jet, and secondly the N-subjettiness is calculated to investigate how likely it is that the jet consists of multiple subjets. In the analysis we discover that the effect of SoftDrop heavily depends on the fat jet radius and on the tuning of the SoftDrop parameters.

[1]University of Bergen, Norway
[2]Universita e INFN, Padova, Italy
[3]CERN

Contents

1 Introduction

The search for new physics at the LHCb often involves jets. Jets are collimated sprays of particles that are clustered together within a cone. The particles that initiated the jet may radiate partons and produce a shower of quark and gluons that eventually turn into hadrons that can be observed in a detector.
The LHCb is able to perform precise tracking of charged particles, it can identify hadrons and leptons and distinguish between different hadrons. LHCb can also reconstruct secondary vertices compatible with the decay of b-hadrons. Thanks to these features, LHCb can measure and identify b-jets with a good precision. For physics Beyond the Standard Model (BSM), we usually have "inclusive" signatures like $X \to b\bar{b}$ where it is necessary to reconstruct these b- and $b\bar{b}$-jets. These jets are very collimated in the forward region. For low mass objects the two jets created from the collision appear close enough to be considered as a single merged **Fat Jet**.

At the LHCb we are looking at the product of proton proton collisions. Often in these collisions the interacting partons are gluon gluon or $q\bar{q}$ pairs. The gluon and $q\bar{q}$ pairs can interact in several ways, such as $gg \to H \to b\bar{b}$ or $q\bar{q} \to Z \to b\bar{b}$.
The Higgs (H) and Z-boson (Z) are part of the Standard Model. Processes like $gg \to A \to b\bar{b}$, where A is a Higgs-like particle, are predicted by several extension of the SM. The A particle can have many different masses, and in our study we will be looking at A with mass 25 GeV, 45 GeV, 60 GeV and 80 GeV. At these masses the boosted $b\bar{b}$ - quarks that create fat jets occur when their transverse momentum exceeds their mass at $p_T \gtrsim 2m/R$.

Several methods to discriminate the source of the jets have been developed in order to classify a jet as interesting for a search [2]. These are called jet substructure techniques, as they take into consideration the internal structure of the jets. The methods generally consist of two steps. First the jet is groomed, or cleaned up, and secondly one computes observables designed to separate signal and background.
There are many different grooming techniques and observables we can use. In this studentship we will use SoftDrop grooming [3], and the mass distribution and N-subjettiness [4] as observables to separate signal and QCD background.

2 Generator Level Simulations

In preparation of the fat jet analysis, we start by looking at Generator Level Simulations, where di-jets are reconstructed with a radius of $R = 0.5$. As previously pointed out, the creation of fat jets depend on the radius of the jet as well as the transverse momentum. Therefore, before we start actually looking at fat jets, we take a quick look at the angular separation of di-jets versus their transverse momentum, normalized by the mass. With the angular separation in the (ϕ, η) space, where ϕ is the azimuthal angle and η is the pseudo rapidity, is $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$. This will help us determine what radius our study will benefit the most from.

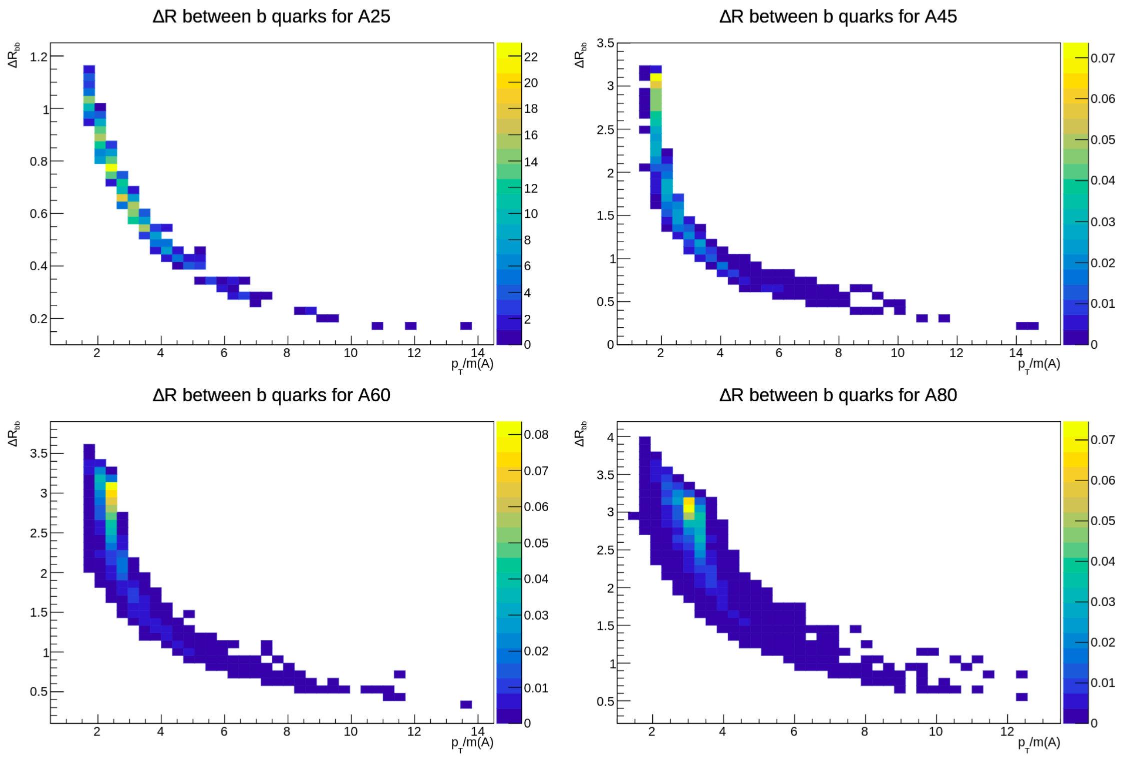

Figure 1: Angular separation as a function of the transverse momentum p_T devided by the mass of the particle.

As we can see in the plots in Figure 1, the maximum is around 3. If we look for fat jets we use the limit of $p_T \gtrsim 2m/R$, which means that the optimal radius is $R \gtrsim 2p_T/m$. This gives us a radius somewhere around 1.5 or 1.2, thus we will use both these radii in the study.

3 SoftDrop

Now we move on to the jet grooming procedure. SoftDrop is a grooming tool introduced to mitigate the impact of soft background on the fat jets. It works by removing soft, or wide angled, radiation from the jet axis. SoftDrop is a generalization of the modified Mass Drop Tagger(mMDT) [4].
The SoftDrop procedure depends on two parameters z_{cut} and β. The SoftDrop condition [3] is as follows:

$$\frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{cut}(\frac{\Delta R_{12}}{R_0})^{\beta} \tag{1}$$

where $p_{T,i}$ is the transverse momentum of the subjets with respect to the beam, ΔR_{12} is the angular separation in the (ϕ, η) space, R_0 is the jet radius, z_{cut} is the SoftDrop threshold and β is an angular exponent.

SoftDrop works by clustering the original jet with the Cambridge/Aachen (C/A) algorithm, which allows us to iteratively decluster the jet to find the subjets. Then we follow the steps of the SoftDrop procedure:

1. Declustering the jet j into two subjets by undoing the last step of the C/A algorithm. The two subjets are labeled j_1 and j_2.
2. If the subjets j_1 and j_2 pass the SoftDrop condition (1), then j is the final soft-drop jet.
3. Otherwise, define j to be the subjet with largest p_T and iterate the procedure.
4. If j can no longer be declustered, keep j as final jet.

The z_{cut} and β parameters controls the degree of jet grooming. The z_{cut} parameter keeps the hard structure and exclude soft emissions, starting from wide angles. The β parameter is as mentioned the angular exponent. When $\beta \to 0$ SoftDrop reduces to mMDT, and when $\beta \to \infty$ the

ungroomed jet is returned. For $\beta > 0$, SoftDrop acts like a groomer and removes soft radiation from the jet while still maintaining a fraction, controlled by β, of the soft-collinear radiation. For $\beta < 0$, SoftDrop acts as a tagger and removes both soft and collinear radiation.
We know that these two parameters are somehow correlated, but we do not know the nature of this relationship. Therefore we have to tune the parameters in a way that take into account this relation.

3.1 Tuning the Parameters

As the SoftDrop technique have not been used at the LHCb before we wish to tune the algorithm, to get some "optimal" values for the parameters z_{cut} and β. That way we have prepared some default values for the parameter where we can see the effect of the SoftDrop grooming properly.
For this study and in the next sections, we use the official detector-level LHCb simulation instead of the generator-level one. To locate the best values for the two parameters we use the resolution of the jet mass distribution. The resolution is obtained by taking the Most Probable Value divided by the Standard Deviation of the jet mass distribution. By considering the resolution as the bin height in our (z_{cut}, β) grid, locating the minimum provided us with some alternatives for the default values.

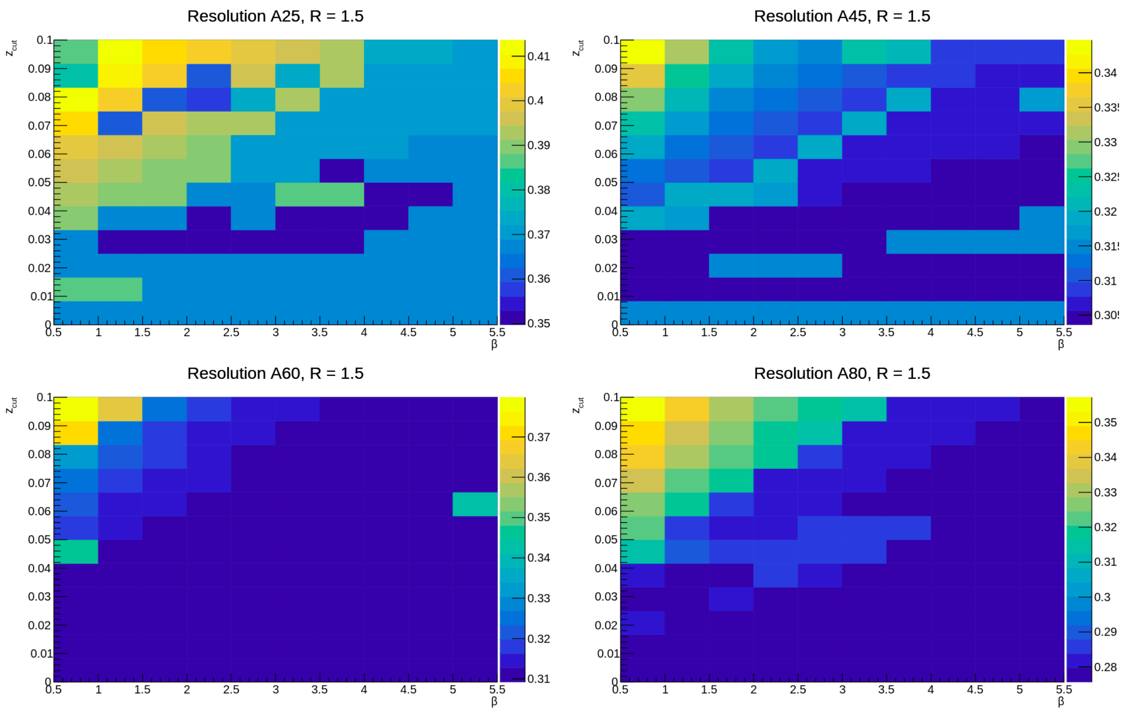

Figure 2: Mass resolution heat map for $A \to b\bar{b}$ with relevant pt cut for the different masses and with a radius of R = 1.5

In Figure 2 we can see an approximately linear correlation between the two SoftDrop parameters, as the boundaries creates diagonal lines in the plots. We are looking for a local minimum that is a minimum in all the plots for the different masses of A. We could have chosen several other combinations of β and z_{cut} that would give us minimums in all the masses, but after evaluating the N-subjettiness of the different combinations we concluded that the difference was not significant, and we therefor can choose any one of them. We choose the value of $z_{cut} = 0.025$ and $\beta = 1$ for the events clustered with $R = 1.5$.

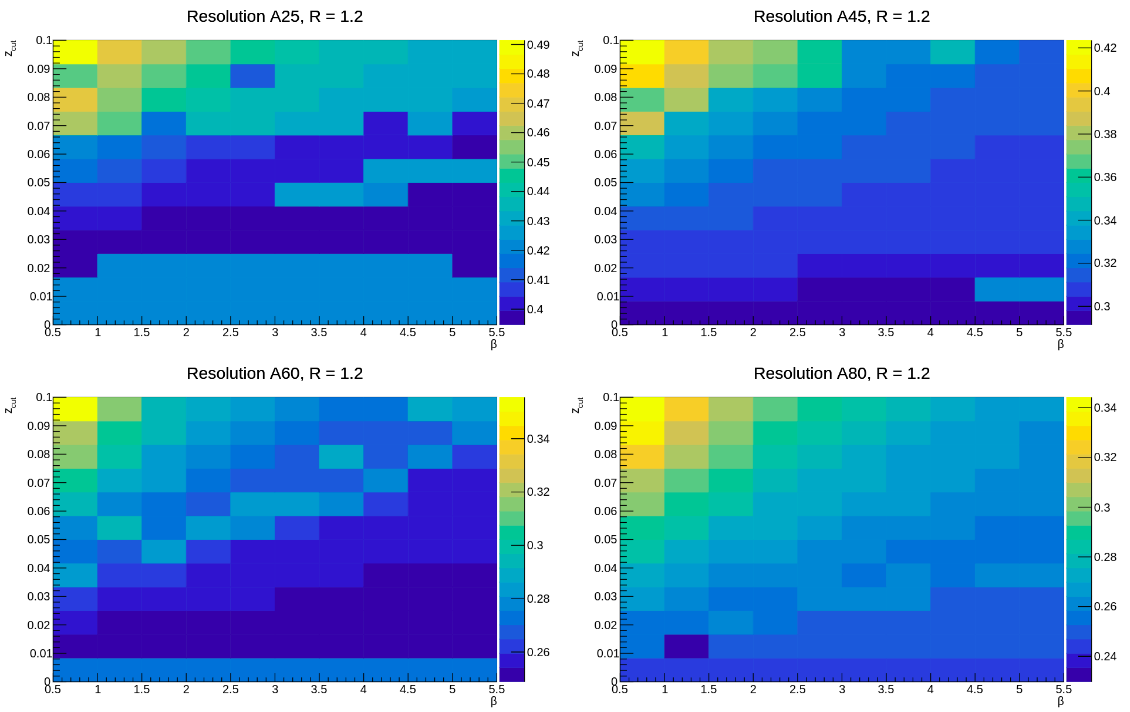

Figure 3: Mass resolution heat map for $A \to b\bar{b}$ with relevant pt cut for the different masses and with a radius of R = 1.2

For the $R = 1.2$ resolution heat maps in Figure 3, we use the same reasoning as above. However, there are no obvious common local minimum this time. The A25's high resolution in the lower z_{cut} eliminates the possibility of using the other masses minimums. It forces us to use one of the local minimums from the A25, one that is still has relatively low resolution for all the other masses. We choose the $z_{cut} = 0.0170$ and $\beta = 5.5$ for events clustered with $R = 1.2$.
We can see that the resolution for the higher masses, A60 and A80, have a lower minimum resolution for $R = 1.2$ than they have for $R = 1.5$. This is an indicator of that the $R = 1.2$ is a slightly better choice of radius for the SoftDrop analysis in this mass range. The minimum resolution for A45 is about the same, while for A25 the minimum resolution is actually higher for $R = 1.2$ than they have for $R = 1.5$. There is not much of a difference and therefore we still choose to carry on with both the radii in the analysis.

4 Analysis

Now, after preparing the SoftDrop grooming it is time to look at how well we are able to separate signal and QCD background. How the analysis is carried out depends on what model the particle is considered to be a part of. In the analysis we have particles from two different models.
The first group of particles is the particle A with different masses, which are supposed to be Higgs-like particles not predicted by the SM. These particles have an unknown p_T range until we make assumptions about the model. Therefore, when comparing to the QCD background we apply a cut to the p_T of the signal to obtain the boosted $b\bar{b}$-jets, but no cut in the QCD background. By not applying the cut to the QCD we obtain the b-jets. To illustrate this we plotted all the possible versions of cuts on our signal in one plot, in Figure 4. The signal without any cuts or b-tagging can also consist of other composite particles. The signal with b-tagging are b-jets. With the pt-cut we ensure that we get the boosted $b\bar{b}$-jets.

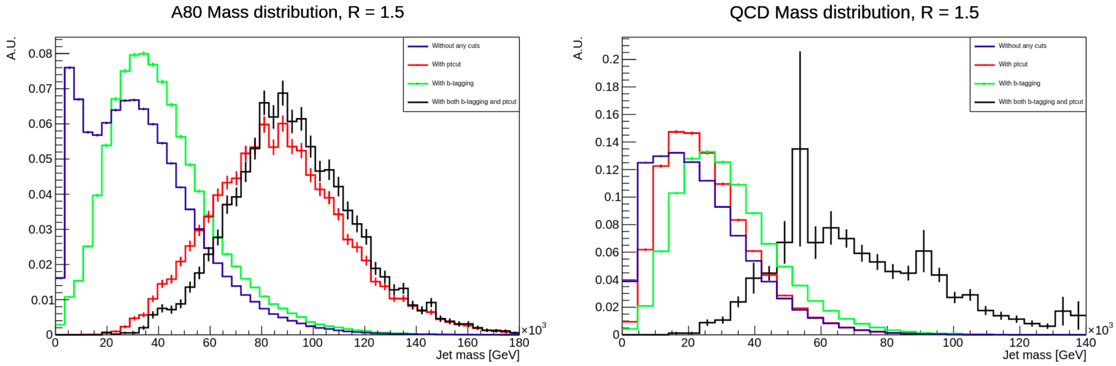

Figure 4: Mass distribution for A80 and QCD background with and withut pt-cuts and b-tagging.

In the first plot in Figure 4 we have the signal of A80 with the different cuts, and in the second we have the QCD background with the different cuts.

The second group of particles is the Higgs and Z-boson, which are a part of the SM. Because they are a part of the SM, we already know what p_T range to look at if we wish to detect them among the QCD background. Therefore it makes sense to apply the correct p_T cut to both the signal and the QCD in order to compare bb-jets. The efficiency of the signal of the SM particles is higher than the efficiency of the QCD background when applying the correct p_T cut, as we can see in Table 1.

	$p_T > 2 \cdot m(H)/R$		$p_T > 2 \cdot m(Z)/R$	
Radius	Higgs	QCD	Z	QCD
1.2	0.17%	0.000153%	0.059%	0.00119%
1.5	0.7%	0.000879%	0.286%	0.0103%

Table 1: Efficiency of Higgs, Z and QCD with relevant p_T cut

4.1 Mass Distribution

We start the analysis by looking at the mass distribution. In regards to the mass distribution it is interesting to look at the distribution of the signal and QCD before and after SoftDrop.

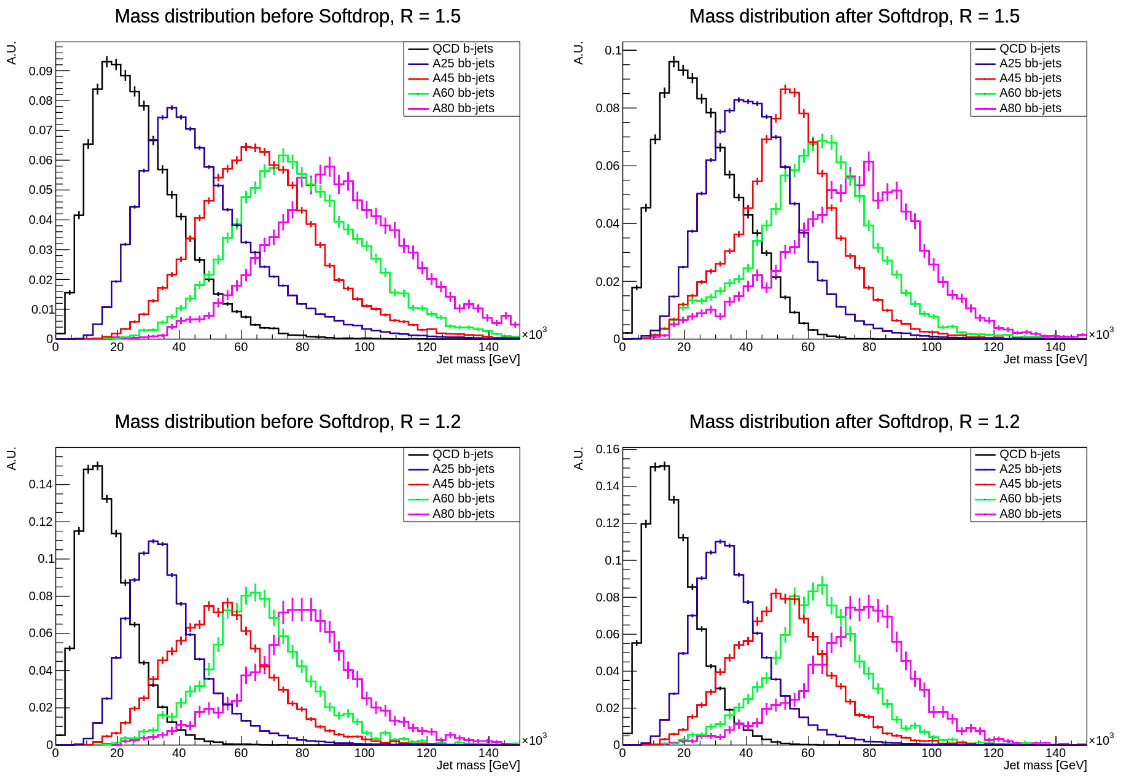

Figure 5: Mass distribution for all the masses, before and after SoftDrop. For R $=$ 1.5 the SoftDrop parameters we use are $z_{cut} = 0.025$ and $\beta = 1$ and for R $=$ 1.2 the SoftDrop parameters are $z_{cut} = 0.017$ and $\beta = 5.5$.

Figure 5 shows the mass distribution of the QCD, and all the A masses. There has not been applied a pt cut on the QCD background as previously discussed, and therefore the QCD distribution remains independent of the constraints we put on the signal. If given the QCD background and the signal, it is possible to distinguish the signal of the higher masses.
In the upper plots in Figure 5 we have the mass distribution for the $R = 1.5$, where we can see a significan change after SoftDrop. In the lower plots, where $R = 1.2$, the change is less obvious.
We will take a closer look at the difference between the mass distribution before and after SoftDrop, after we have had a quick look at the mass distribution for the Higgs- and Z-boson.

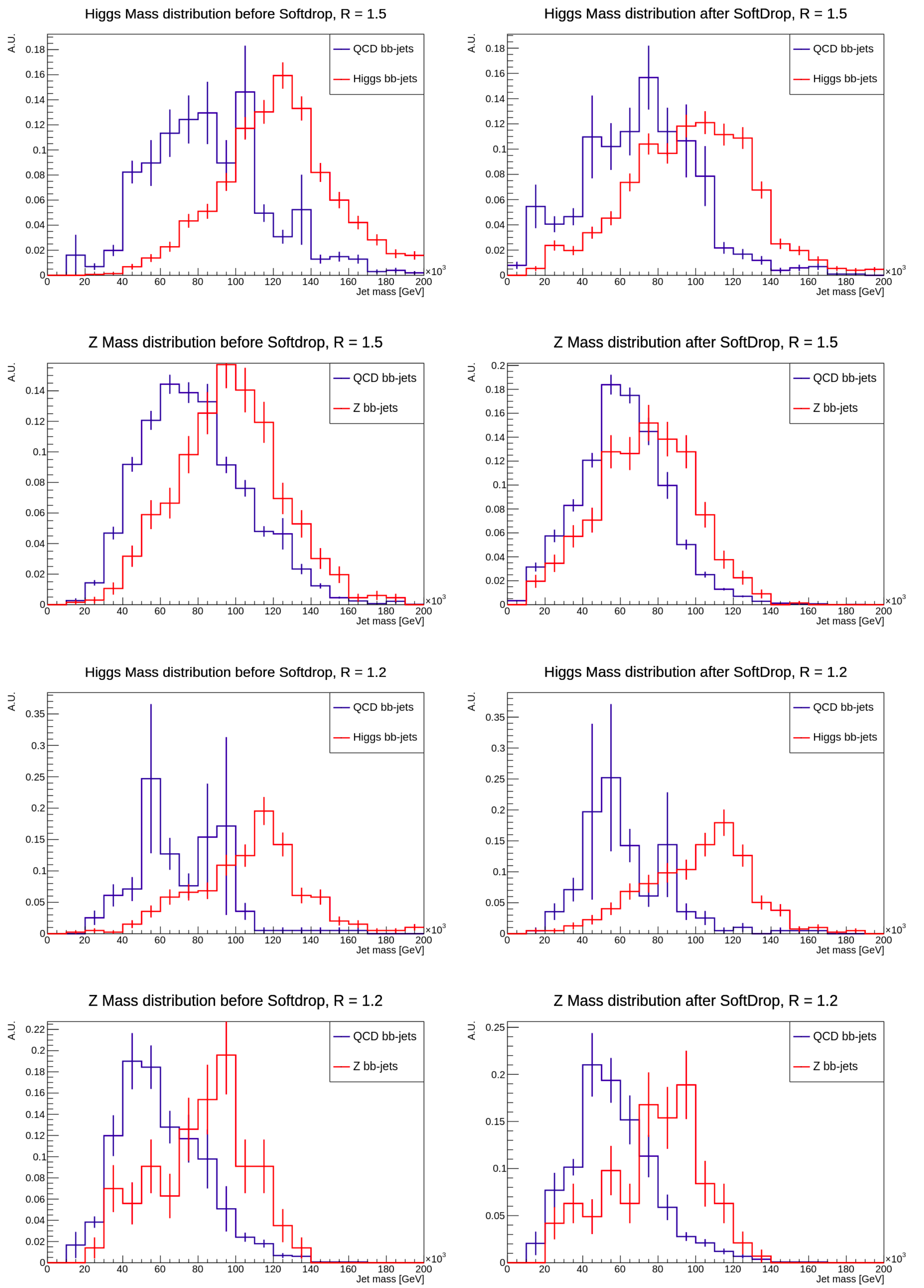

Figure 6: Mass distribution plots for the Higgs- and Z-boson, before and after SoftDrop. or R = 1.5 the SoftDrop parameters we use are $z_{cut} = 0.025$ and $\beta = 1$ and for R = 1.2 the SoftDrop parameters are $z_{cut} = 0.017$ and $\beta = 5.5$.

In Figure 6 we see the mass distribution of the $b\bar{b}$-jets belonging to the Higgs- and Z-boson compared to the $b\bar{b}$-jets of the QCD background. When looking at the Z- and Higgs, we also apply the p_T cut to the QCD to obtain the $b\bar{b}$-jets. When doing this we see that the distribution of the QCD is more similar to the signal than for the A masses. However, we are still able to observe a difference between the QCD and signal both before and after applying SoftDrop. Most of the simulated entries are removed with the p_T cut, as we can see the efficiency is very low in Table 1. Since the distributions we observe in Figure 6 have large statistical fluctuations, is not as reliable as the ones we observe for the A masses in Figure 5.

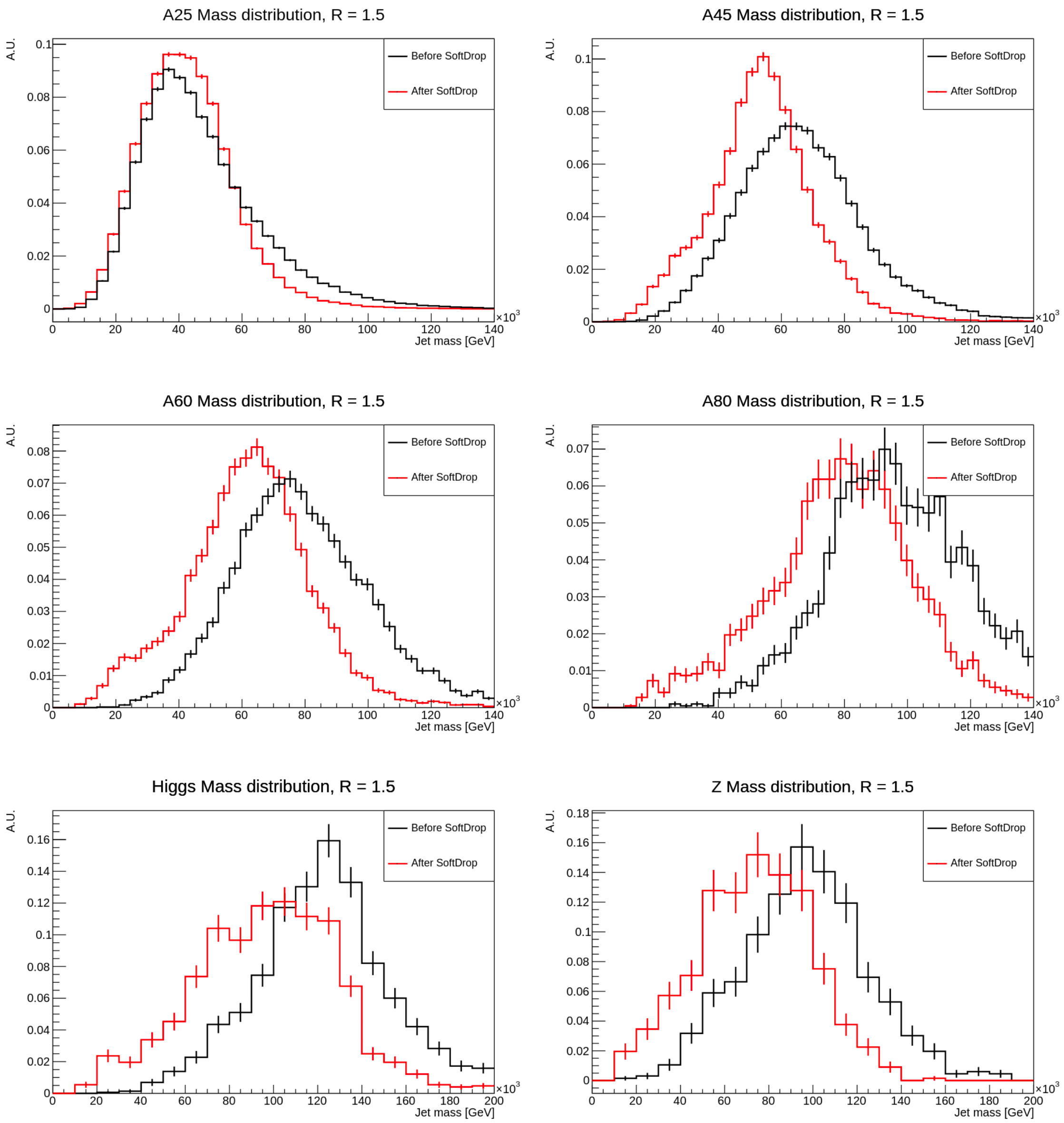

Figure 7: The mass distribution of the signal before and after SoftDrop in the same plot, for R = 1.5. With $z_{cut} = 0.025$ and $\beta = 1$.

In Figure 7, we can see that the SoftDrop is improving the distribution for $R = 1.5$. Part of the tail is removed for all the masses, and for most of them the peak is noticeably closer to the correct mass. Except for A25 which has a peak even further away from $m = 25$.

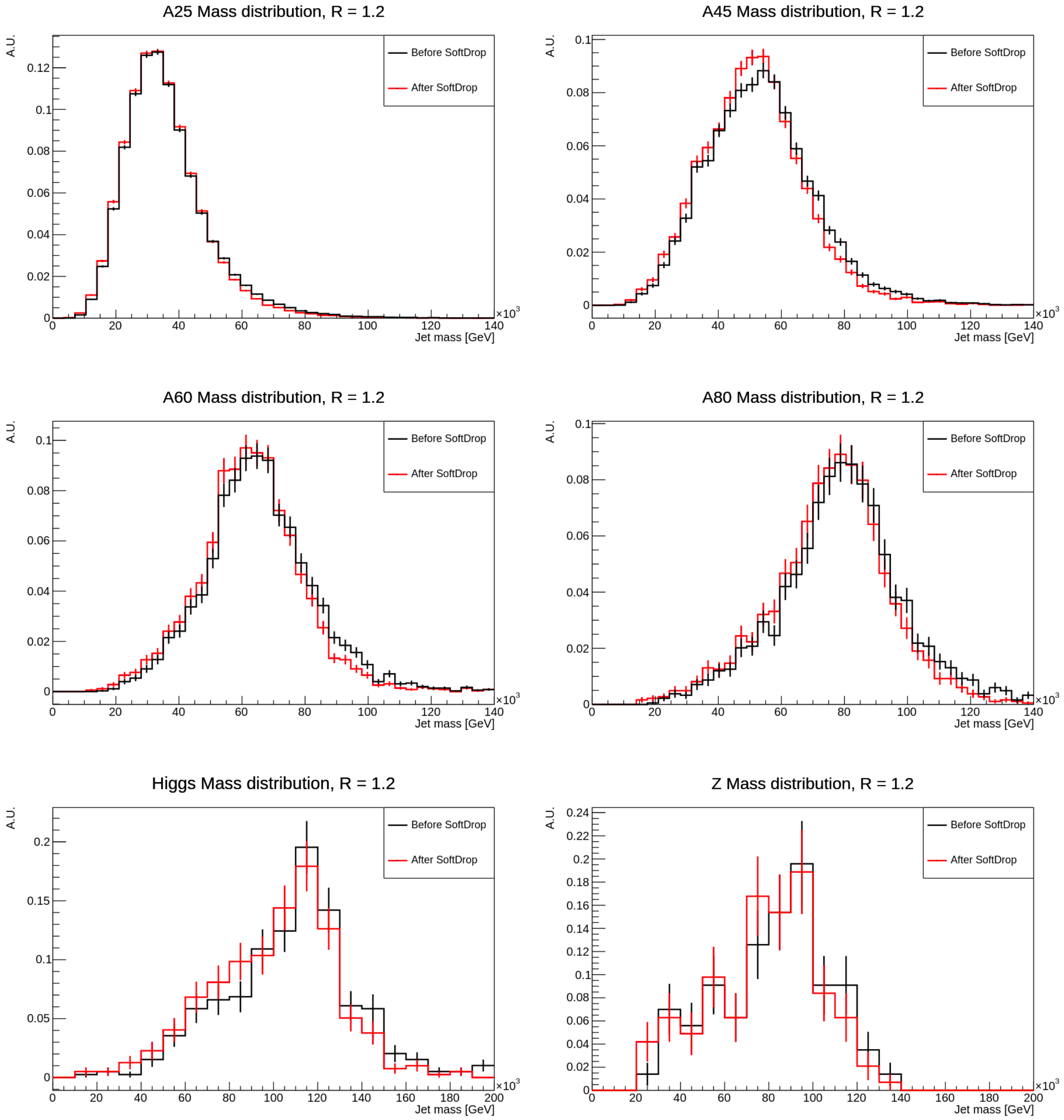

Figure 8: The mass distribution of the signal before and after SoftDrop in the same plot, for R = 1.2. With $z_{cut} = 0.017$ and $\beta = 5.5$.

In Figure 8 we can see that the difference between before and after SoftDrop is very little for $R = 1.2$. A little bit of the tail is removed, but other than that it is hard to see the effect of SoftDrop.
Considering this, it seems like $R = 1.5$ is a better radius to use in the mass distribution analysis. It is actually quite a large difference between the $R = 1.5$ and $R = 1.2$, compared to what we would expect. In the tuning part of the report we saw that the minimum mass resolution of the $R = 1.2$ signals was smaller than the minimum of the $R = 1.5$ at higher masses. We considered that to be a sign that $R = 1.2$ possibly was the better radius, but at this point in the analysis it seems to be an advantage to use the $R = 1.5$.

4.2 N-Subjettiness

N-subjettiness is a jet substructure method designed to identify boosted object, such as fat jets. We use N-subjettiness to effectively "count" the number of subjets in a given jet [4]. In a way we can say that N-subjettiness is a measure of to what degree a particular jet can be regarded as composed of N subjets. N-subjettiness is defined as:

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, ..., \Delta R_{N,k}\} \tag{2}$$

Here we have k that runs over the constituents of the jet. $p_{T,k}$ is the transverse momentum of the constituents, and $_{N,k}$ is the angular separation between the subjet N and the constituent k. It also contains the normalization factor $d_0 = \sum_k p_{T,k} R_0$, with R_0 as the clustering radius.

The τ_N is defined between 0 and 1. When $\tau_N \approx 0$ the jet has all the radiation aligned with the subjets direction, and therefore the jet has N or fewer subjets. For $\tau_N \gg 0$ a large part of the radiation is distributed away from the subjets direction, and therefore the jet has $N+1$ or more subjets. The ratio τ_2/τ_1 is considered to be a good discriminator when we consider signal versus QCD.
Therefore, we will be looking at τ_1, τ_2 and the ratio τ_2/τ_1 and compare the signal and the QCD background. We apply the N-subjettiness after SoftDrop, and send the groomed jets subjets into the N-subjettiness calculation.

In Figure 9 we see the $R = 1.5$ signal. We see that for the high masses there is more of a difference between the signal and the QCD. In the τ_2 plots, the center plots, we see that the signal is more towards 0 than the QCD. This implies that the signal jet most likely has two subjets, while it is less likely that the QCD jets do. When we look at the ratio τ_2/τ_1, it is hard to see how this is a good discriminator. For A25 and A45, the signal and QCD is almost identical, but we see for the higher masses that there is more of a difference.
As for the Higgs- and Z-boson, we know that there is a very low efficiency. This makes the results statistically limited, but the tendency of the plots is that both the τ_1 and τ_2 is closer to zero than for the A masses.

In Figure 10 we see that the τ_1 and τ_2 have the same trends for $R = 1.2$ as for the $R = 1.5$. However for the τ_2/τ_1 we see that the distributions look quite different, both in regards to the QCD and to the $R = 1.5$ signals.

5 Conclusion

In this project we have studied several fat jet techniques that can be applied in the search for new physics signals at LHCb.
SoftDrop is a useful jet substructure tool, that removes soft radiation from fat jets to mitigate the impact of background radiation. SoftDrop takes two parameters z_{cut} and β. These parameters can have many values, so we performed a test to tune the parameters for our analysis. We did this by creating a grid of z_{cut} and β, and using the mass resolution as the z-value. This way we obtained local minimums, helping us pick values for the two parameters. For $R = 1.5$ we chose $z_{cut} = 0.025$ and $\beta = 1$ and for $R = 1.2$ we chose $z_{cut} = 0.017$ and $\beta = 5.5$.

For our analysis we looked at the mass distribution and N-subjettiness.
For the mass distribution we see that its easier to distinguish between signal and QCD background when we use $R = 1.5$ instead of $R = 1.2$. The mass resolution from the tuning showed that $R = 1.2$ should have a better resolution, but it does not show in the plots we use. One reason for this might be that the mass resolution for $R = 1.2$ before SoftDrop also is lower than for $R = 1.5$. However, for $R = 1.5$ we do see a distinction between the signal and QCD, which was our main goal.
For the N-subjettiness the results show that the signal has a lower τ_2 than the QCD, which indicates that the fat jet consists of atleast two subjets.

The results from both the mass distribution and the N-subjettiness show that we are able to distinguish the signal from the QCD background. However, the study should be expanded to more masses, in particular lower masses such as 10 GeV. We believe that we would see a greater difference if we use lower masses. The simulation and use of many more jet entries would also improve the analysis, because of the low efficiency when applying p_T cuts, especially for the Higgs- and Z-boson.

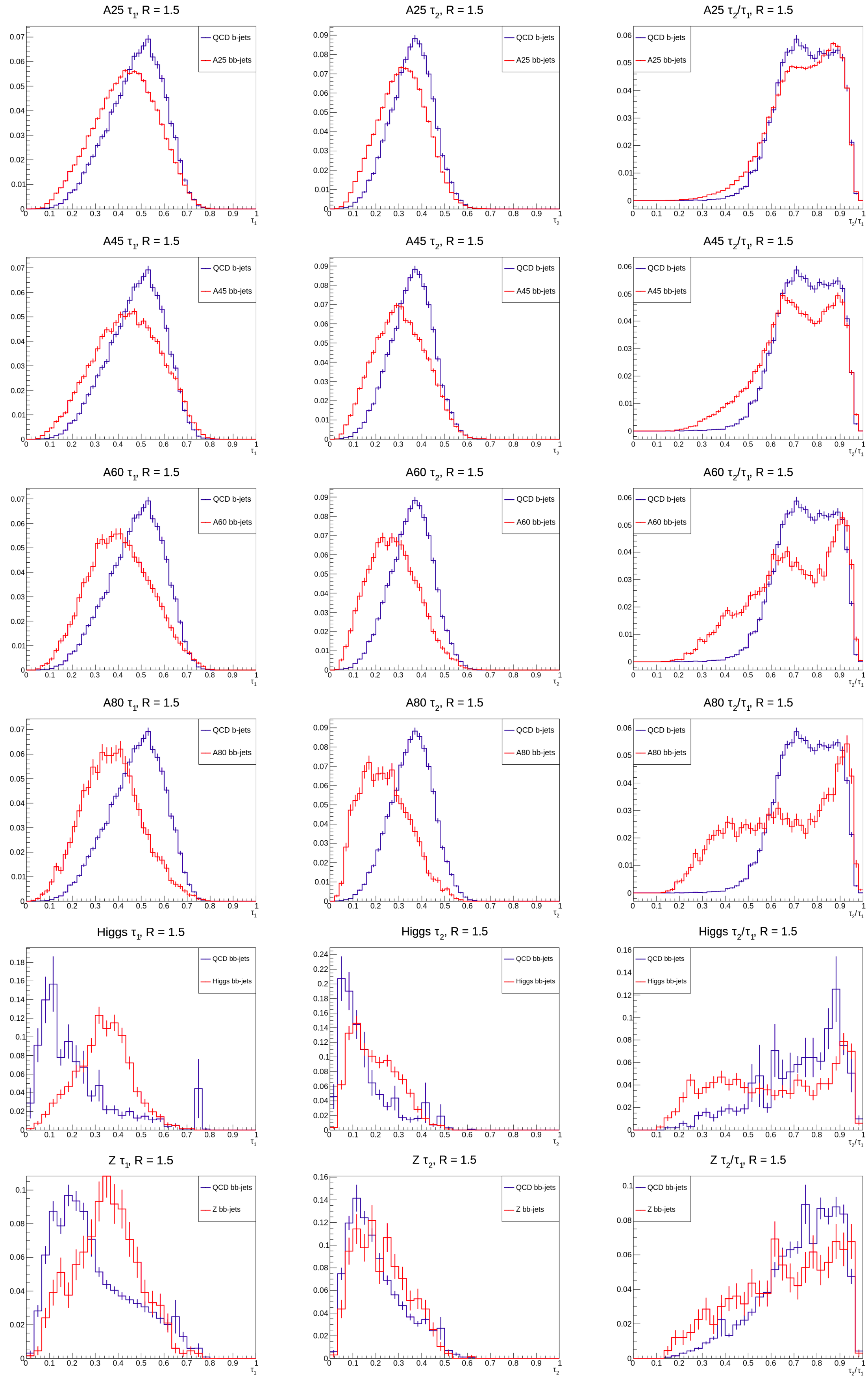

Figure 9: N-subjettiness after SoftDrop with $z_{cut} = 0.025$, $\beta = 1$ and $R = 1.5$

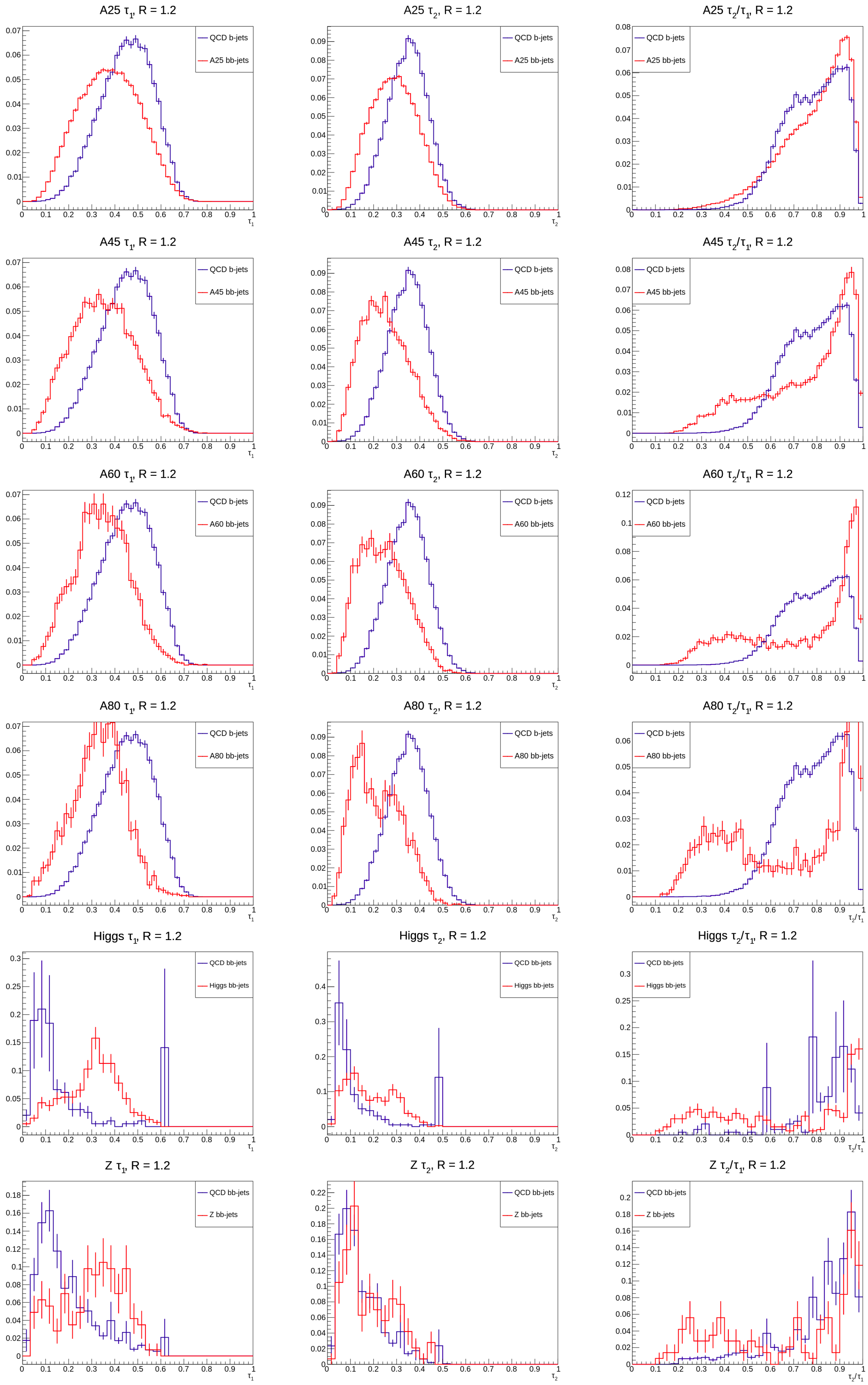

Figure 10: N-subjettiness after SoftDrop with $z_{cut} = 0.017$, $\beta = 5.5$ and $R = 1.2$

References

[1] Gavin Salam. *Theory of Fat Jets and Jet Substructure* (2012). (2012). [`https://gsalam.web.cern.ch/gsalam/talks/repo/2012-HiggsHunting-HH-boosted-review-salam.pdf`]

[2] Mrinal Dasgupta, Alessandro Fregoso, Simone Marzani and Gavin P. Salam *Towards an understanding of jet substructure* [arXiv:1307.0007 [hep-ph]]

[3] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thalera. *Soft Drop* (2014). [arXiv:1402.2657v2 [hep-ph]]

[4] Jesse Thaler and Ken Van Tilburg *Identifying Boosted Objects with N-subjettiness* [arXiv:1011.2268 [hep-ph]]