Universiteit Utrecht

Opleiding Natuur- en Sterrenkunde

Finding the Effects of Jet Quenching with Machine Learning

Bachelor Thesis

David Coppens

Supervisor:

Dr. Marta Verweij
Institute for Subatomic Physics

Daily Supervisor:

Bas Hofman
Institute for Subatomic Physics

February, 2021

Abstract

In the collisions of heavy ions we can find that a quark-gluon plasma forms. A result of this quark-gluon plasma is that jetquenching can take place. There has been a lot of research to study the workings of jetquenching. When we are looking at particle collisions however, it is not yet possible to reliably tell if a jet has been quenched based on their final state. In this research we wanted to find a way to look at the final state of a particle collision and tell if jet quenching had taken place. We firstly investigated the jets and looked at which parameters are important. After finding useful variables of the jets we used a machine learning algorithm to try and find a pattern between hundreds of thousands of simulated non quenched jets. We then looked at how this algorithm treated simulated quenched jets, to see if it would recognise them as being different.

The Algorithm seemed successful at finding a difference between the normal jets and the quenched jets. It is however unsure how the algorithm found this pattern. We can therefore not say with certainty if the algorithm was successful. An important reason why the algorithm seemed to find a difference is due to the way the events are generated. Additional research is therefore needed to further study this problem, and to try to find whether machine learning can find a difference between quenched and non quenched jets. It could then also be used to look at real data of events and find out if they have been quenched and by how much.

Contents

1 Introduction

The Greek philosophers already believed there should be a particle that can not be split up in multiple particles, an atomos or "uncuttable"[1]. In the nineteenth and early twentieth century a lot of research was done to find this unsplittable particle and when they thought they found it they called it the atom. It was later found out however that this atom was not really unsplittable, when the atom was found to consist of electrons, protons and neutrons[2]. A new field of physics now known as particle physics investigates those particles that are smaller than atoms called subatomic particles. It was later found however that the protons and neutrons are themselves splittable, because they are made up out of quarks[3]. These quarks can not be measured in nature due to color confinement[4]. But their existence has been proven[5].

At the Large Hadron Collider (LHC) at CERN for the last decades there has been a lot of research into subatomic particles such as quarks. This has helped us to gain a deeper understanding of our world today, but might also help us to understand how it has been formed at the big bang. The experiments at the LHC use the scattering of particles to find out how particles interact with each other. In the process of scattering, two particles collide with each other at a high energy and form hundreds of different particles. This is known as a parton shower. In this shower we can recognize jets which consist of multiple particles that are related to each other. An interesting phenomenon can take place in the scattering of heavy-ions such as lead atoms. A Quark Gluon Plasma can be formed which can interact with the particles. This modifies the way the particles can scatter and alters the properties of jets, this process is called jet quenching. The probability of jet quenching occurring is usually small, dependent on the temperature of the collision [6], and it is not easy to determine if a jet is quenched and by how much. It is therefore hard to show if it has happened, based on the value of one observable. We are also only able to look at the end product of a scattering process and can not tell what is happening, only what it results in. It is however very useful to be able to tell if jet quenching has taken place, because it can help us deeper understand jet quenching and the implications it has for physics in general.

In this study we try to find out if it is possible to find out if a scattering process has undergone jet quenching. We will do this by training a Machine Learning algorithm on a set of normal (not quenched) jets. Then we will use this algorithm to find a difference between other normal jets and quenched jets. We want to use this to find a certain pattern in quenched jets that makes them distinct from normal jets. In the hope to eventually be able to tell from the resulting particles of a scattering whether or not it is likely that jet quenching has taken place. This will here be done with simulated events and not real ones. Eventually it would be useful to use this same method on real events if it works on simulations. For the simulated proton scatterings without jet quenching we use an event generator called Pythia [7]. For the simulated lead scatterings with jet quenching we use an event generator called Jewel [8]. The main goals of this study are: Can we find a difference between Pythia and Jewel events, based on their jet structure using Machine Learning? And if so, can we find the physical observables associated to this difference?

In this experiment we have tested four different Machine Learning algorithms to find a difference between the Pythia and Jewel jets. At least one of these algorithms seems to find a difference between the Pythia and Jewel jets. This result could tell us that Machine Learning

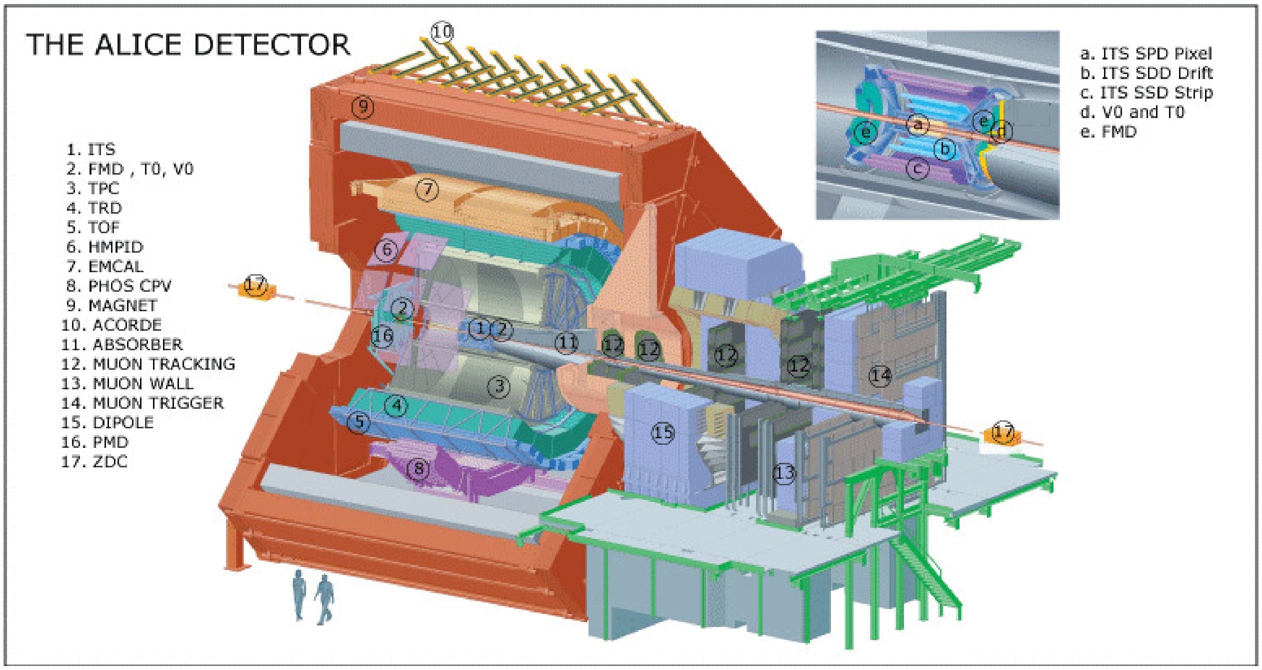

Figure 1: The Alice detector. Each of the numbers represents a part of the detector that is used for the detection of different particles. All these detectors together can show us what happened at a particle collision. Figure retrieved from [11].

can find out if jet quenching has taken place. It is not yet possible to show the physical reason why this difference is found, and if it really has anything to do with jet quenching. It is also possible that the difference exists, due to the way in which the samples have been generated. This means that further research into this topic will be important to find out if there are useful results to be found.

2 Theoretical background

2.1 The Large Hadron Collider

In particle accelerators particles are accelerated and then made to collide with each other at high energies. The largest of these particle accelerators is the Large Hadron Collider (LHC) located at CERN, The European Organization for Nuclear Research. At the LHC there are two beams of particles that are accelerated to velocities close to the speed of light, using superconducting magnets at temperatures close to 0 K [9]. In this process of accelerating the particle beams, the particles reach an energy of up to 6.5 TeV. Two particles from different beams can collide with each other with a total impact energy of 13 TeV. These collisions can happen at 4 different detectors, that are designed for different research purposes. These detectors are ALICE, ATLAS, CMS and LHCb.

In this thesis we will only look at the ALICE detector which specifically looks at the collisions of heavy-ions such as lead [10]. The ALICE detector (see figure 1) consists of various different components that all neatly work together to measure what happens when the two particles collide. After the collision all the particles that have been created first pass through the Inner Tracking Systems (ITS), the Time Projection Chamber (TPC) and Transition Radiation Detector (TRD) detector. These detectors accurately measure the trajectory of all the charged particles. In the various different detectors that follow, the identity of each of the particles is determined. The entire detector is in a magnetic field which bends the trajectories of all charged particles which helps in calculating the momenta of the different particles. There are also detectors that tell us something about the photons created at the collision, these detectors are PHOS, CPV, PMD and EMCAL.

The interaction of the colliding particles is described by all the fundamental forces, these are: the strong force, the weak force, the electromagnetic force, and gravity, although the effect of gravity on this process is very small. Each of these detectors shows us a part of what happens after a particle collision. This can help us gain a deeper understanding of all the forces that are at work in this process, and how it affects particles in an collision.

2.2 Particle Collisions

The physics of what is happening inside a particle collision is mostly ruled by strong interactions. This is captured in Quantum Chromodynamics (QCD)[12], it describes the interactions between quarks and gluons. These are the subatomic particles that make up protons and neutrons, the building blocks of atoms. QCD can not be solved algebraicly and there are therefore two strategies that are being used to find solutions. The first is to use lattice QCD, which is a numerical solution. To find an accurate solution, it is however necessary to use more memory-points in a computer than is practically feasible. The more useful strategy is perturbative QCD. This makes use of the fact that the strong coupling α_s is small, which leads to terms which contain a α_s^n with a large value for n to fall out of the equation. This means that it is possible to accurately describe simple QCD processes and also to give good approximations of more difficult processes. The electromagnetic force also has an effect on particle collisions. All charged particles and photons can interact with each other with the electromagnetic force. The theory that describes these interactions is Quantum Electrodynamics (QED). It is most important in the interactions of electrons and positrons with photons. Research in this field was started by among others Dirac[13] and Feynmann[14]. The weak interaction also has an effect on particle collisions. Together with QED it is usually used in the theory of electroweak interactions. This unifies both the theories of electromagnetic interactions and weak interactions, and considers them to be two manifestations of one force[15].

2.3 Quark-Gluon Plasma

All atoms, which constitute almost everything around us, are made of protons, neutrons and electrons. The protons and neutrons in their turn are made up of quarks. These quarks interact with each other via gluons through the strong force. An effect of the strong force is that at low density and low temperature it gets stronger when distances get larger. This means that, as the name strong force implies, the bond between quarks due to the strong force very strong and almost unbreakable is. Therefore, protons and neutrons, which are

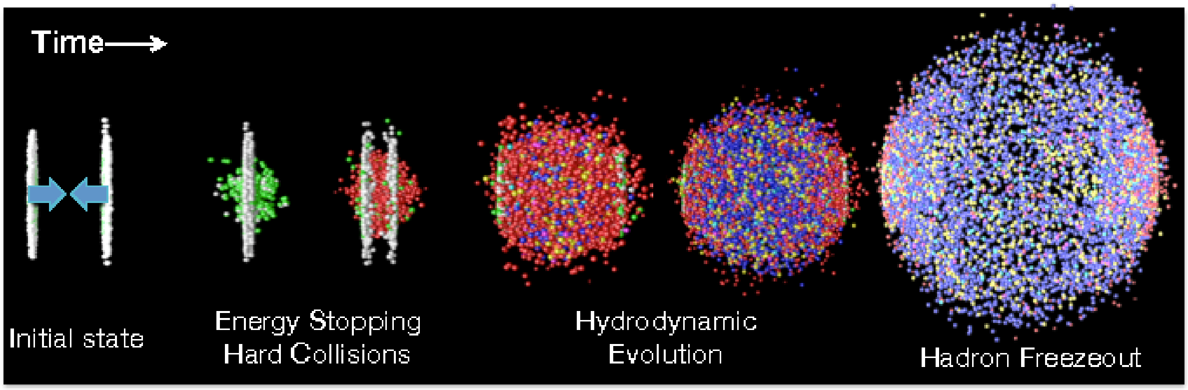

Figure 2: A depiction of the forming of Quark-Gluon Plasma (QGP) in a heavy-ion collision. The red parts of the figure is QGP. It initially expands very fast but it then cools down and falls down into hadrons. Figure retrieved from [19]

made up of quarks, are very stable. This also means that free quarks are not detectable as free particles[16]. Because of the strong force quarks are only observable in bound states together with other quarks.

The Quark-Gluon Plasma (QGP) is a state of matter that is formed in collisions of heavy-ions at sufficiently high energy. In the quark-gluon plasma the usual confinement of quarks and gluons changes such that they are free to move around. Usually they are confined inside protons or neutrons inside atoms. The quark-gluon plasma is an interesting field of study because it allows us to see the behaviour of quarks and gluons if they are not confined. It can be formed due to a phase transition from a hadron gas [17]. An interesting feature of this quark gluon plasma is that it has near zero viscosity. Meaning that the quarks are almost completly free to flow unhindered by each other [18]. In figure 2 we can see the evolution of the QGP. The red part of it is the QGP. As can be seen it grows fast initially but in the end it vanishes, because it hadronizes. Hadronization is the process where the quarks and gluons of a collisions form hadrons, after they have lost a lot of energy. Hadrons can be either baryons, which are made up of three quarks, or mesons, which are made up of a quark and an antiquark. These hadrons then enter the detector and can be detected. After these hadrons have been detected we are able to reconstruct what happened in the collision.

2.4 Jets

In the above mentioned collisions of two accelerated particles the particles scatter and form a multitude of different particles that is known as a particle shower. Here the initial particles lose their energy in lots of steps where they emit a particle, which can also emit another particle. This can result in a situation with hundreds of particles. A detector can only detect the final state of the branching and not how they were formed. This does not leave us with a good structure of the parton shower. It is therefore hard to really understand what has happened if we only look at the final particles. A strategy that is being used to make more sense of what has happened in the scattering process is to use jets. A jet is a part of all the particles that are very close to each other and are assumed to have the same parent particle.

This means that a collision can be split up into a few high energy jets instead of hundreds of single lower energy particles. There is no perfect way to find those jets because it is impossible to find which exact particles "are related" to each other. Instead there are a few different algorithms that all have their distinct advantages and disadvantages. A few of these algorithms are: Cambridge-Aachen, anti-kt, kt, Jade, JetClu and MidPoint. After using these algorithms, we get a few important properties of the jets like its momentum, direction and the amount of particles it contains. It can also tell us something of the expected structure inside the jet, and how it was formed[20]. These jets and their substructure are being used to gain a better understanding of what is happening during scatterings, because they simplify the data without losing too much valuable information, making it easier to understand. These jets are also very closely related to the QCD splitting functions that define how the collision takes place. The construction of the most used jets is collinear an infrared safe. It is therefore theoretically a very natural and useful tool to interpret particle collisions.

2.5 Jet Quenching

What has been mentioned about jets applies to both proton-proton collisions and heavy-ions collisions. An interesting phenomenon takes place in heavy-ions collisions, because of the quark-gluon plasma that is being created there. The partons inside the jets have the possibility to react with the quark-gluon plasma. This can happen in the form of a particle bending away due to an interaction with the quark-gluon plasma. It can also be due to a particle exchanging energy with the quark-gluon plasma. This then results in a change of the resulting hadrons of the scattering and thus in the way the jets look. This process is called jet quenching. In the first years of this century the first experimental evidence of this process was found [21] [22] both at the RHIC and the LHC. Jet quenching has as an effect that the jets resulting from it will have lost energy [23] [24]. This has been used to prove that interactions with the QGP take place, resulting in jet quenching.

3 Method

3.1 Dataset

We want to find a difference between the unquenched proton-proton scatterings and the quenched lead-lead scatterings. We could use real data acquired by experiments at the LHC, but the choice has been made for simulated data. The reason is because the theory behind the scatterings in proton-proton collisions is quit well understood, therefore the simulations resemble the real experiments very well. The big advantage of simulated events is that they are far more practical to access and use, because it can be done from any computer. For the proton-proton scatterings a program called Pythia[7] was used, and for the lead-lead scatterings Jewel[8] was used. It is important to know that not all Jewel jets are quenched, only a small part. This is done because in real life there also is just a small chance that a jet is being quenched. In order to make sure the Pythia and Jewel events could properly be compared, we needed to make sure that they were the same except for the jet quenching. This was done by creating both the Pythia and Jewel events at $\hat{p_T} = 150$ GeV which is the transverse momentum of the leading jet. In total there where 500.000 Pythia events generated, with a tune of 14[25]. This tune enables us to specify the values of some parameters to have results that better match real experiments. The Jewel events that were used have been downloaded from the following website[26], giving 57.913 Jewel events. Now that we had these events, we needed to find the jets and their important features, we used the JetToyHI software by Dr. Marta Verweij[27] for this. This had to be done for both the Pythia and Jewel events. After this we wanted to filter the jets based on some parameters for the jets. This was done to make sure that the jets were very similar, and to prevent that jets that are very different from most jets would make a big impact. The first condition was that we wanted each of the jets to have a p_T greater than 160 GeV. The second condition was that we wanted for the pseudo-rapidity to be in the range of $-2 < \eta < 2$. The pseudo-rapidity is the spatial coordinate for where the jet lands along the angle where 0 means that the jet direction is orthogonal to the direction of the incoming particles, and infinity corresponds to the jet direction being parallel to the direction of the incoming particles. This is useful because at larger values of $|\eta|$ we do not always have a detector and if we do the resolution is worse than for small values. It was therefore useful to remove those jets to make sure they do not affect the outcome of the experiment. This gave us 470.271 Pythia jets and 66.228 Jewel jets. From these jets we used 404.043 Pythia jets as a training set, which gave us 66.228 Pythia and Jewel jets that were being used to compare the two types of jets.

3.2 Analysing the dataset

Now that there is a dataset for both the Pythia and the Jewel sample, it is important to compare the two. It is firstly interesting to inspect some of their important features. In particular we want to take a look at the structure of the jets that were being produced, to find out interesting attributes of the jets or differences between Pythia and Jewel. There are different features of the jets that can be looked at. One of these features is the dr of a splitting of a jet. This is the angle between a particles in a jet and another particle that has split off. For most jets this happens multiple times, as will be shown later. We here use the first splitting where both of the splittings remained with at least 10% of the energy of the starting particle. This means that there was a zcut of 0.1. This angle can have any value between 0 and 0.5. It can be seen in the top half of Figure 3 how the value of this angle, dr, is distributed for both Pythia and Jewel. As can be seen in the figure the distribution of the dr of the first splitting is very similar for both Pythia and Jewel. But when you place them on top of each other you find that Jewel has an even larger peak at small angles, than Jewel. An important other feature of the jets is a parameter called z_g, which is the relative amount of energy of the lowest energy particle after a splitting. This z_g is a parameter that exists for each splitting, just like the dr's. The distribution for the z_g of the first splitting can be seen in the bottom half of Figure 3. It is important to note again that there is a z cut, which means that both particles after the splitting must have at least 10% of the energy of the particle before the splitting. This can easily been seen in the figure, since there are no splittings that have a z_g smaller than 0.1. If this z cut would not be there, there would be a lot of splittings with a z_g smaller than 0.1, which does not give a very interesting result. It can be seen for both Pythia and Jewel that small values of z_g are more likely. When the distributions for Pythia and Jewel are placed on top of each other, you can see that the two

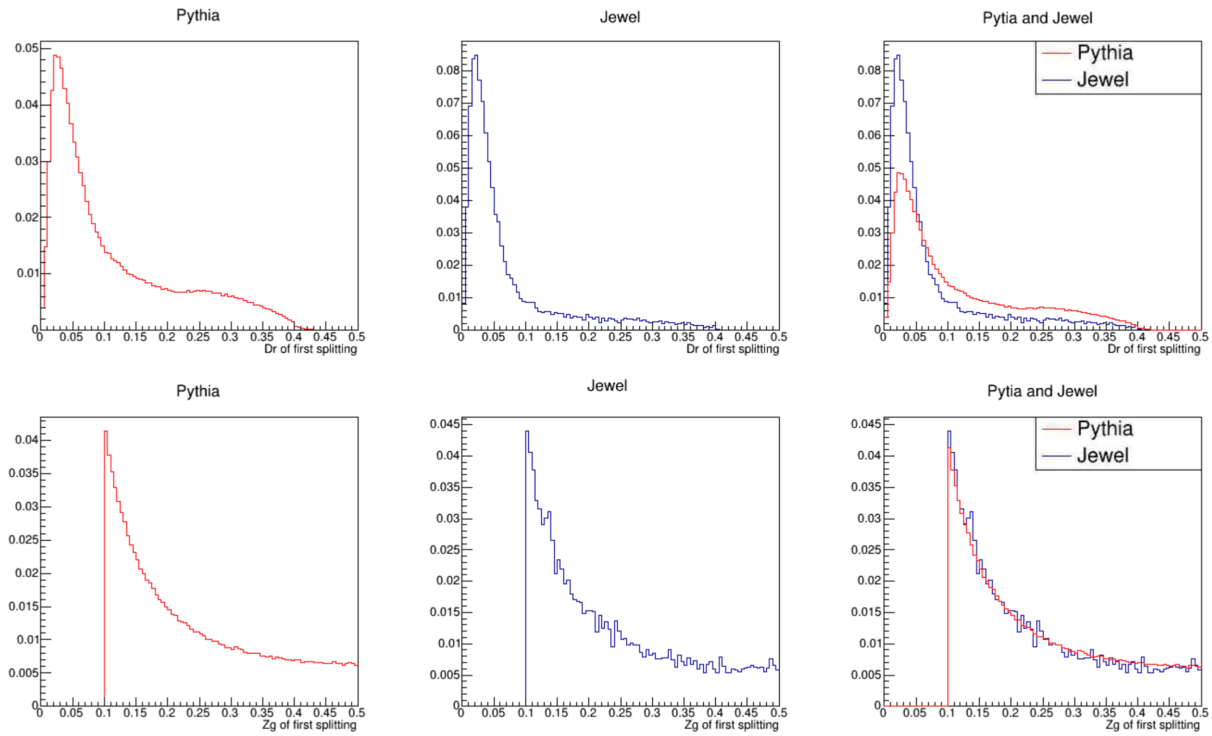

Figure 3: Above: 1 dimensional histograms of the dr of the first splitting. This is the angle between the resulting particles in a single splitting. Below: 1 dimensional histograms of the z_g of the first splitting. This z_g is the part of energy of the original particle in the resulting particle with the least energy, which has a maximum of 0.5. There is a z cut of 0.1 here, this means that both of the particles will have at least 10% of the energy of the starting particle. For both avove and below: on the left the Pythia events and in the middle the Jewel events. On the right the Pythia and Jewel events are plotted together, here Pythia is red and Jewel is blue.

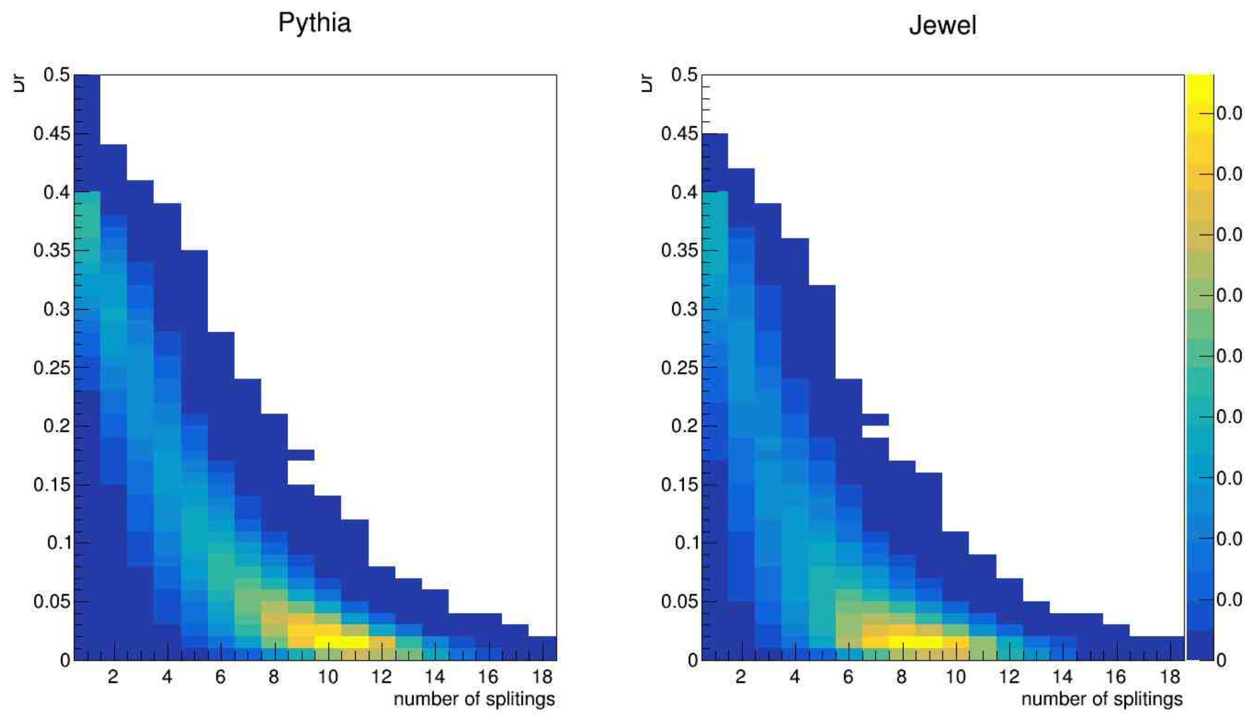

Figure 4: 2D histograms of the dr for all the different splittings for both Pythia and Jewel. On the x axis you can see the number of splittings (so the first, second, third etc.). On the y axis you can see the value for dr at this splitting. Here all the jets are added together.

are almost the same. The differences that can be seen are mostly because the dataset of Jewel was smaller which is probably the reason why it fluctuates the way it does.

A next step we wanted to look at is how the dr and the z_g are distributed for all their splittings and not just their first. The result of this can be seen in Figure 4. Here you can see for all the jets together what the value for dr for each of their splittings. And how it progresses for following splittings. On the x axis you can see which number of splittings it is you are looking at and on the y axis you can see the value for dr of this angle. You can see that for a higher number of splittings the values for dr start to decrease. It is also important to note that the amount of jets is decreasing for higher numbers, because a jet that has already had a splitting does not always have another splitting. We can see how many splittings a jet has in figure 6. It shows that Jewel jets tends to have fewer splittings than Pythia jets. In figure 4 it is important to know that there is no z cut which there was in Figure 3. We could otherwise interpret the first column in the left part of Figure 4 to be the equivalent as the top left plot of Figure 3. It is noticeable that in this figure we have for the first splitting a much higher value for the first dr because of the lack of this z cut. We can also see that for higher number of splittings the dr tends to get smaller. We expect that this angle dr is decreasing for each next splitting. Based on Figure 4 we can say that it looks like the angle is decreasing for each step, but we can not say that for sure. Another way to visualize this decrease in the angle is to look at the difference between two

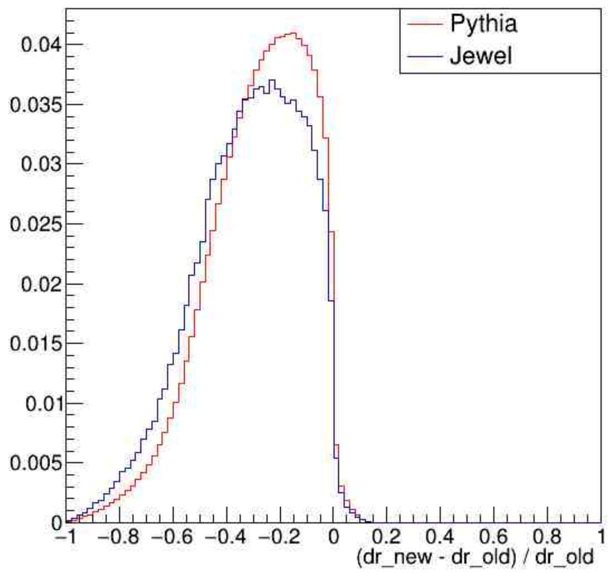

Figure 5: The difference between the angle, dr, of two preceding splittings. The angle should usually get smaller for following steps. We can see that is happening, because most results are negative. only a very small part has a value higher than 0.

each other preceding angles. This can be seen in Figure 5. It is clear that the vast majority of the splittings gets smaller for successive jets. It can be seen that there are some values that are larger than 0, which means the angle gets bigger in two following steps. This is not very common both for Pythia and Jewel. For Pythia this happens in 1.37% of cases, and for Jewel it happens in 1.10% of cases. Since this is such a small part it is not something that is investigated further here. We are however not sure why this happens, and what causes this increase in angle.

3.3 Machine Learning

3.3.1 What is Machine Learning?

Now that we have analysed and compared the two datasets. We want to find a difference between the two that will help us understand the effects of jet quenching. The way we did this is by using a Machine Learning algorithm. Machine Learning uses a lot of data to "train" an algorithm to find a pattern that we can not necessarily see as humans. It is usually given both input and the corresponding output, to make it learn to find a pattern, so it can eventually guess the desired output from new input. We here use a different kind of algorithm because we don't give it an output when training. This kind of Machine Learning algorithm is known as unsupervised. The specific type of algorithm that we used is "anomaly search" or "outlier detection". It looks at the given points and looks at common trends, and it gives a prediction value to datapoints based on how much they follow the trends. Here a higher value means it looks more like the trend. This way it is able to spot points that have

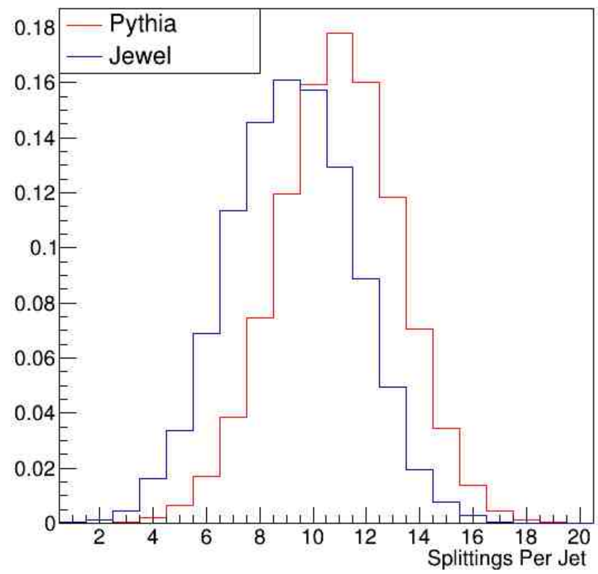

Figure 6: In this figure we can see for both Pythia and Jewel how many splittings there are per jet. We can see that the general shape of the distributions seems to be the same. However the Jewel distribution seems to be shifted to the left by roughly 2. This means that the Jewel jets tend to have fewer splittings than the Pythia jets.

a low prediction value and are different than the training set. In this experiment we trained the algorithm on a part of the Pythia dataset. The algorithms could then look at the other part of the Pythia dataset and the Jewel dataset. These could then be compared with each other, to find out what is different about the Jewel data.

3.3.2 Which Machine Learning algorithm should we use?

A choice had to be made for which Machine Learning algorithm should be used. The choice was made for the Scikit library[28], because this has a lot of different algorithms and they are very easy to use. Then from this library a specific algorithm had to be chosen. As before mentioned we needed an anomaly search algorithm. Scikit has four different possible anomaly search algorithms, these are: Elliptic Envelope, One-Class SVM, Isolation Forest and Local Outlier Factor. Each of these algorithms has the same goal, but they achieve it differently and it is therefore very situation dependent which is best to be used in a specific situation. For this experiment all four of them have been tested and compared to each other. For each of these algorithms, the results can be seen in Figure 7. Here we have plotted a histogram for each of the algorithms with how many jets fall under each prediction value for both Pythia and Jewel. It is important to note that each of the algorithms uses a different scale for the prediction values, but that doesn't make a difference in their effectiveness. We are still able to compare them to each other. We can see for all of them except the One-Class SVM, that the Jewel jets are more present at lower prediction values than the Pythia jets. The majority of Jewel jets are still situated at high prediction values. This is exactly as we

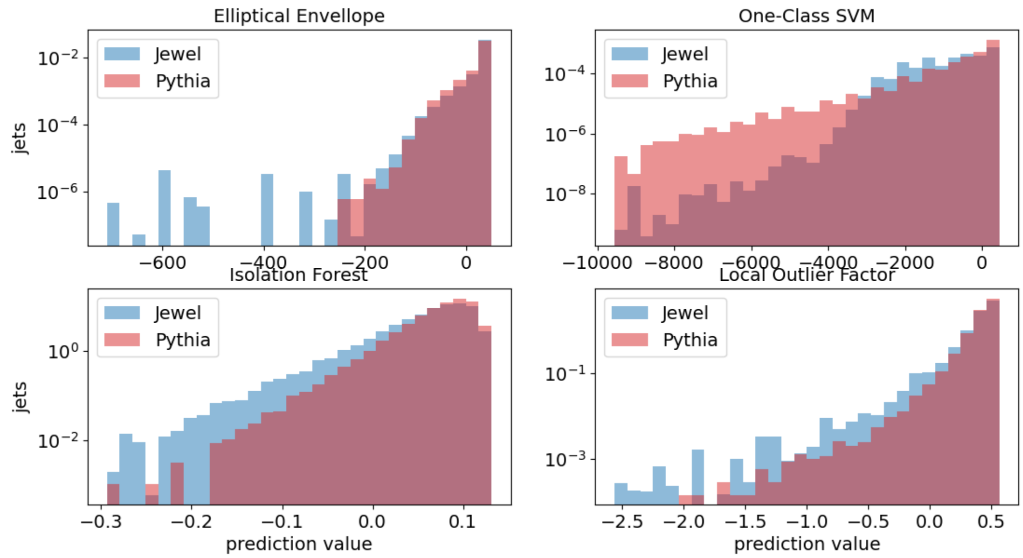

Figure 7: A comparison between the prediction values of different Machine Learning algorithms. The higher the prediction value is the more it looks like Pythia. The red bars represent the Pythia events, the blue bars represent the Jewel events. For each of these algorithms except One-Class SVM, the Pythia events tend to have a higher expectation value.

expect because most Jewel jets should be "the same" as Pythia jets, because of no or small jet quenching. A small part of Jewel jets should be different because of jet quenching. To get a better understanding of how the algorithms compare we made a Receiver Operating Characteristic (ROC) plot which can be seen in Figure 8. This plots the "efficiency of the Pythia jets" versus "1 - the efficiency of the Jewel jets". Here the efficiency is given by the part of jets that have a prediction value that is higher than a certain number. The black line represents the line you would get if the algorithm would not do anything and just gives every jet a random value. Lines that are to the right and above of it indicate that the algorithm did in fact do something and found a certain pattern which all of the algorithms seem to have done. From the graphs we can see that Isolation Forest (green) and One-Class SVM (orange) are doing the best job at finding a difference between Jewel and Pythia. One-Class SVM in particular is interesting because in Figure 7 it seemed to show no good results, but here it does. This can be understood because almost all the jets of One-Class SVM are close to 0, which heavily influences the result, and therefore make the plot in Figure 7 almost useless. This shows that there is no big difference between Pythia and Jewel for low prediction values, but there is a large difference between high prediction values. This is the opposite of what we want because we want to look at what happens at low prediction values, because that is where jet quenching happens. That is why One-Class SVM is not a very good algorithm for this experiment. Therefore we chose Isolation Forest as the algorithm to do the rest of this research with, since it shows good results in both of these plots.

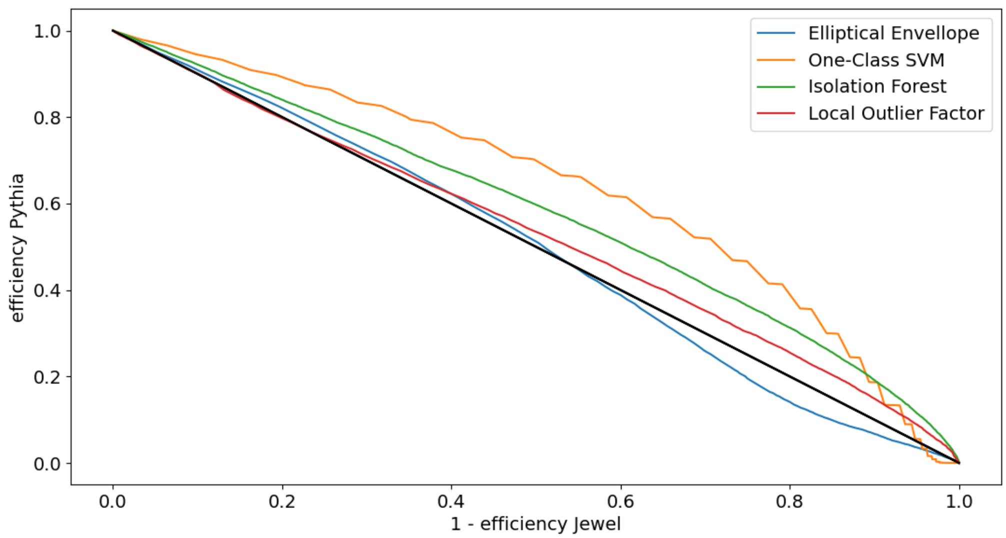

Figure 8: A comparison between ROC plots of different Machine Learning algorithms. The black line represents a algorithm that does not find a difference between Pythia and Jewel at all. The more a line is to the top and right the better it is able to find a difference between Pythia and Jewel. Isolation forest and One-Class SVM seem to show the best result.

3.4 Procedure

To obtain the goal of finding out which jets have been quenched, we have used the following methodology. The algorithm required that each datapoint, in this experiment a jet, was given as an array of numbers. For each jet we used a few values to train and test the algorithm. These values can be found in Table 1.

Table 1: All the parameters that are being used in the Machine Learning algorithm.

1	dr, the angle, of the first splitting
2	z_g, the energy fraction, of the firts splitting
3	average dr of all splittings
4	standard deviation of dr of all splittings
5	average z_g of all splittings
6	standard deviation of z_g of all splittings
7	the total amount of splittings

To understand these variables we need to know a bit more about how the jets look. An example of how a jet might look can be seen in Fig 9. Here we see one starting particle and multiple particles splitting off of it. For each of these particles it is possible that they also split multiple times, but we don't look at that here. We only look at the so called "leading branch", which is the branch with the highest energy in each splitting. Each splittings occurs under an angle, dr. It also gives both branches a part of the initial particles energy, z_g. z_g is the fraction of the total energy of the mother branch that is given to the branch with the

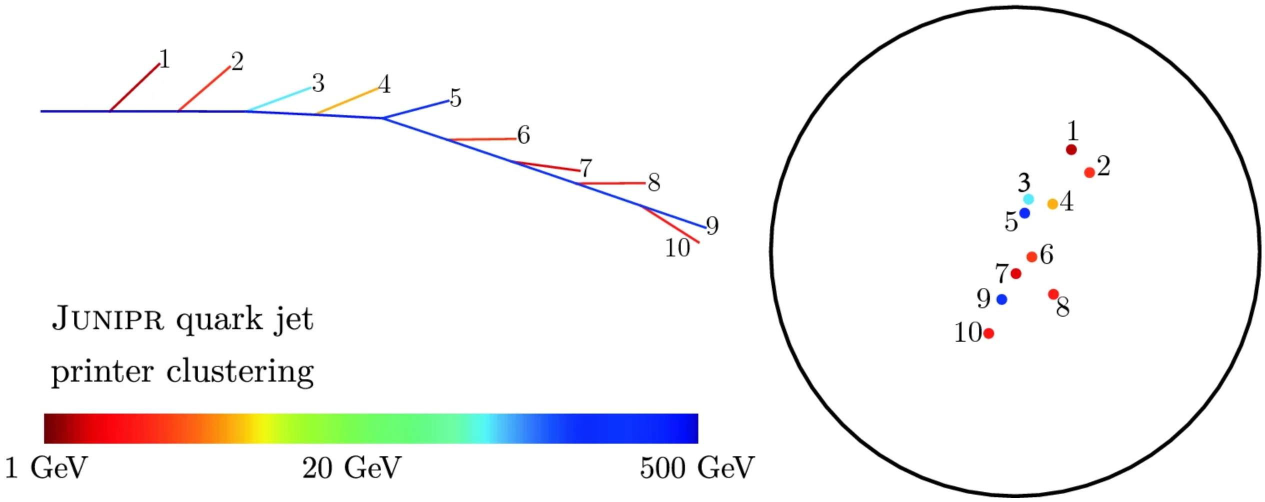

Figure 9: The shape of a random jet. We can see the splittings of the leading branch. Which is the particle after each splitting that has the highest energy. At each splitting the direction of the particle is bend to result in a constant total momentum. each of the numbers can represent a single particle but also a subjet. Figure retrieved from [29].

lowest energy after the splitting, so it is always the case that $z_g \leqslant 0.5$. For the first splitting there is a so-called zcut which means that z_g has to be at least 0.1, because otherwise we face the problem that the splitting is so small that it does not have a significant effect on the original particle, and thus does not give us a lot of relevant information. If the first splitting z_g is smaller than 0.1 then it gets ignored and we take the next splitting as the first splitting. The total number of splittings starts from this splitting, and counts all the splittings in the leading branch. There are thus these seven parameters that we use. These are not all the parameters that could be used. The p_T of the jets was initially used, but it was later chosen not to use it because of how the Jewel events were generated. They are created with an artificially enhanced probability to create a high p_T jet. This is corrected afterwards by applying event weights, but this has as a side effect that it on average leads to a higher value for p_T. This resulted in a lot of jets receiving a low prediction value simply because they have a high p_T. The idea to use the dr and z_g for the first couple of splittings (for example 5) also existed, but some jets had only one or two splittings. The algorithm however needed a fixed amount of parameters for each jets, making it hard to use that properly. Different values could also be used like the k_t of a jet, but this would be correlated to other parameters, making it less useful and difficult to comprehend afterwards. In further research it could be very interesting to look at other parameters for the Machine Learning algorithms.

4 Results and discussion

In the previous paragraph we have seen that we can use Machine Learning to find a difference between the generated Pythia and Jewel events. Here we want to look at what this difference is and if we can reliable tell if a jet has undergone jet quenching. We want to do this by

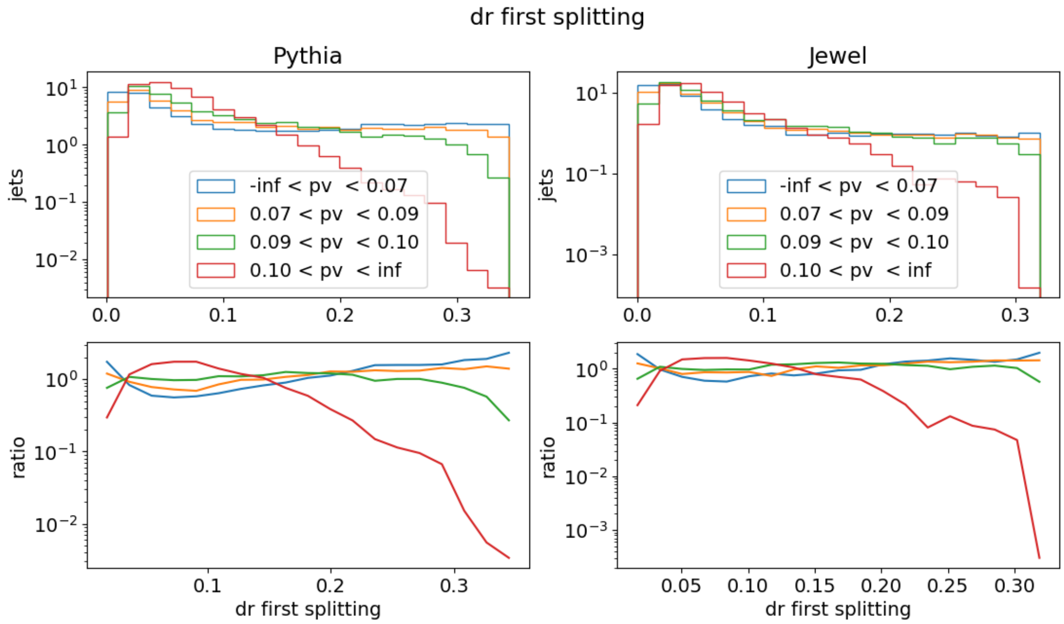

Figure 10: Above: normalized histograms of the dr of the first splitting against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

first grouping the different jets into four groups based on their prediction values. We then want to look at how these groups of jets compare to each other, by comparing their values for certain parameters. The goal of this is to find a difference between the Pythia jets and the Jewel jets when you compare the values for different prediction values. What we can get from this are plots like Figure 10. Here we see at the top a normalized histogram of the dr of the first splitting, split up between the four different values for the prediction values. Note that higher prediction values correspond to being more Pythia-like, so red corresponds to being the most Pythia-like, and blue is the least Pythia-like. In the bottom graphs we can see a ratio plot for each of the prediction values. This gives for each prediction value, the part of the total jets that are in this bin over the average part of the jets in this bin for all the prediction values. This better shows us their relative values compared to each other. We can clearly see that the high prediction value of red peaks at a dr between 0.05 and 0.10 and falls of sharply after that. For the other three values the graphs tend to look a lot the same. An interesting and somewhat problematic feature to notice is that this shape seems to be very much the same for both the Pythia and the Jewel jets. This means that from figure 10 we can conclude that the Machine Learning Algorithm has found that a high value for the dr of the first splitting to be different, but since it happens for both Pythia and Jewel we can not conclude that it has anything to do with jet quenching. This is because we know that Pythia has no jet quenching and it still shows a difference between the different prediction values. This difference for Jewel that seems to be very similar can therefore not be due to jet quenching.

We can make and have made multiple figures like Figure 10 for each of the different

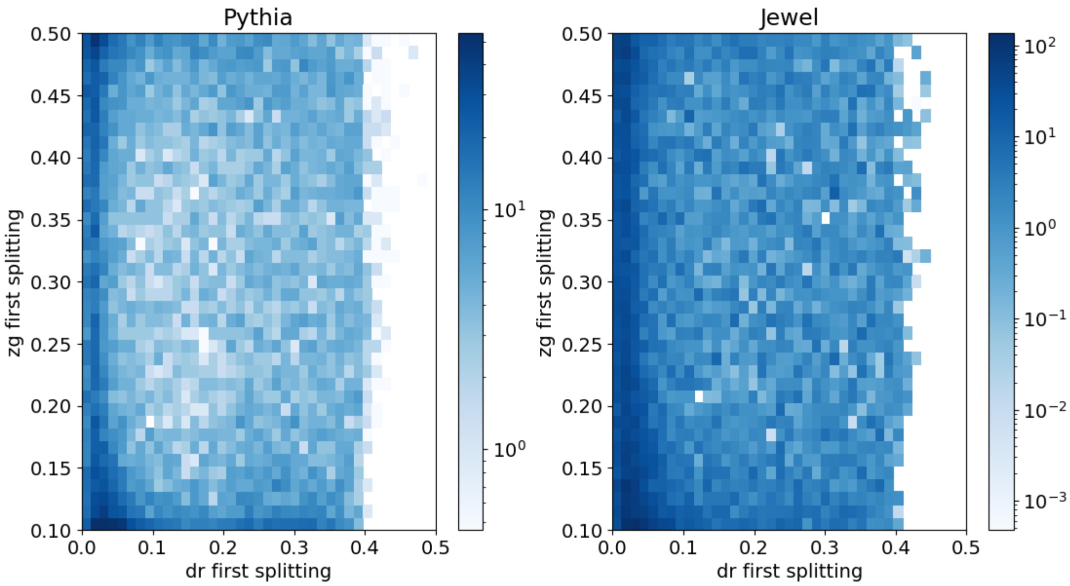

Figure 11: Left: a 2 dimensional histogram of the dr of the first splitting against the z_g of the first splitting with the lowest prediction values for the Pythia events. Right: the same variables plotted for the Jewel events.

parameters that were being used to train the algorithm and also for other parameters (see Appendix). The reason why we want to do this is to hopefully find a difference between Jewel and Pythia when it comes to the way the jets are distributed for the different prediction values. Upon doing so we have however not found any noticeable differences that would lead us to believe that they show an important difference between the Pythia and the Jewel sample.

An important next step that we could now take is to look at 2d histograms of two variables against each other. This can help us to find out if there are relations there. This however is difficult to achieve visually because it is not possible to plot multiple 2d histograms in one figure. This means we should use one figure for each prediction value, which makes it very hard to compare them to each other. That is why in Figure 11 we have only plotted the lowest prediction value in the next figure, because it should be where jet quenching takes place. We can see in this figure that both the Pythia and Jewel plots do not seem to show obvious large differences. Note that the scaling is very different between the two plots. We can of course make a similar plot to these here by plotting each pair of parameters of a jet against each other. This will give us a large amount of plots making it a lot of work to go through. It might however give us valuable insight in how to find jet quenching. In this experiment we have unfortunately not had the chance to properly look at each of these histograms and determine if it could be useful. This could however prove to be useful in a follow up experiment. So far however we have not been able to find a good way to differentiate between jets that have been quenched and jets that have not been quenched.

It is important to look at why we did find differences between the prediction values of

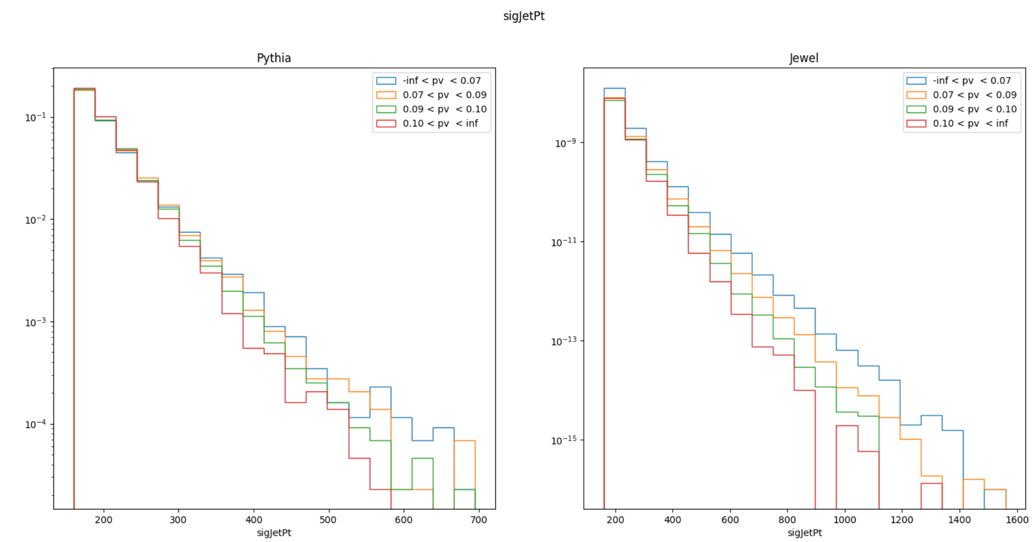

Figure 12: A histogram of the amount of jets for different values of jet p_T. On the left for Pythia, on the Right for Jewel.

the Pythia jets and the Jewel jets in Section 3.3, but we can not find a reason why this difference happens in the plots in this Section. An important reason why this could be the case is because of how the Jewel jets have been created. To increase the chances of jets being quenched the Jewel software changes the probabilities of certain things to happen. This leads to more jet quenching but it also has the effect of a larger part of the jets having for example a higher value for jet p_T. This can be seen in Figure 12 . The Jewel software makes up for this problem by giving every jet a proper associated weight, which ensures that the results are still consistent with the physics of real events when we take an average assuming the corresponding weight. This means that we can see the same exponential decrease happen for both Pythia and Jewel for the jet p_T. We see however that for Jewel there are still events for larger values of jet p_T, but we see the same logarithmic decrease happen. This leads us into a scenario where it seems like there are a more Jewel events with a low prediction value as can be seen in Figure 7, but we don't see that same result in Figure 10. This means that we might be able to find a difference between the Jewel and Pythia events using Machine Learning, but the results here do not seem to give us a decisive proof for it. It is therefore necessary to do more research on this topic to be able to really answer the research question.

5 Conclusions

In this experiment we used machine learning to try to find the effects of jet quenching in Jewel jets to compare them to Pythia jets. The results have been mixed and need to be further looked into to find out if there is a real detectable difference between the Pythia and Jewel events. When researching the possible Machine Learning algorithms we found that some of them could find a difference between the Pythia jets and the Jewel jets, like we expected.

It can however not be told with certainty whether this is caused by the jet quenching in the Jewel jets or by the way the Jewel jets are being generated. We then looked at the different parameters that were used as input of the algorithm. We did this by splitting the jets in groups, based on their prediction value which showed how much the jets resembled the training Pythia jets. This way we could investigate what jets that seem to be very different from normal Pythia look like. We could not find a noticeable difference between the Pythia and Jewel jets if we looked at these prediction value groups. Further research into this could still lead to useful results, as this experiment was not complete in its assessment of these differences.

6 Outlook

There are a few different things that could be improved in future research to find better results and be able to properly answer the research questions. The first thing that could be done is to take a better look at the data and results this experiment generated. It is definitely not impossible that there is a good and useful conclusion to be drawn from it. In this experiment there was however not enough time to do that. One can look at the before mentioned one and two dimensional histograms. But also other interesting ways of visualizing could be used like a Lund plane[30]. The second thing that could be done is doing this experiment but changing some aspects of it. One way to do this is by increasing the sample size, this should allow for a better trained algorithm and a larger amount of data to test the algorithm on. This could give a clearer picture of what is happening by giving more data which could highlight the interesting abnormalities in the data from the statistical abnormalities. Another way to do this is by using different parameters for the Machine Learning. One could for example use different features of the dr's and z_g's of the jets to train the algorithm, where for example the dr and z_g of more than just the first splitting is being used. Or perhaps by not only looking at the leading branch of a jet but also at its other branches. It can also be tested if the p_T of a jet can be used as a parameter, to find out if that might have an effect. Another option for this is to use a different Machine Learning Algorithm. The algorithm that was used here, called "Isolation Forest", seems to work the best out of the four tested algorithms, but they are definitely not the only possible algorithms that exist. The tensorflow library[31] for example could be used, since it also has different Machine Learning options. Another way to change the experiment is by trying different settings for the jets when they are being generated, such as a different $\hat{p_T}$. A third way in which we could really make this experiment easier is if there would be any way to guarantee that a jet is being quenched. If in the production of the jets there would be some kind of variable that tells us that jet quenching has taken place, it would be a lot easier to find a pattern using Machine Learning.

If there are useful results produced with this experiment, the next step would be to use this on real data. Because it would not be useful if this theory only works on simulated events it should also work on real data This would allow us to better understand the real physics behind jet quenching. It is however expected that this analysis could straightforwardly be translation to the real data because they resemble the simulations very closely in case op proton-proton collisions.

7 Laymen summary

In dit bacheloronderzoek wordt er onderzoek gedaan naar de botsingen van elementaire deeltjes. In deeltjesversnellers worden twee deeltjes versneld tot snelheden dichtbij de lichtsnelheid, waardoor ze veel energie hebben. Vervolgens worden ze met elkaar in botsing gebracht. In die botsing worden veel andere deeltjes geproduceerd wat resulteert in honderden deeltjes die in verschillende richtingen vliegen. Deze deeltjes kunnen vervolgens worden waargenomen. Aan de hand van deze deeltjes die uiteindelijk ontstaan is het mogelijk om te verklaren wat er is gebeurd. Om dit makkelijker te maken worden de deeltjes vaak beschouwd als "jets". Dit betekent dat een groep deeltjes, die in ongeveer dezelfde richting beweegt, beschouwd wordt als een ding, een jet.

In dit onderzoek wordt gekeken naar twee soorten botsingen, door middel van deze jets. Een botsing tussen twee protonen en een botsing tussen twee loodkernen. Als voor beide soorten botsingen de kernen dezelfde energie hebben, zijn de twee soorten botsingen normaliter vrijwel hetzelfde. Er is echter een quark gluon plasma dat ontstaat bij de botsing van de loodkernen. Dit quark gluon plasma kan interacties aangaan met de deeltjes die ontstaan. Dit kan een effect hebben op hoe de jets gevormd worden. Dit proces heet "jetquenching". Dit jetquenching gebeurt niet heel vaak en het is vaak moeilijk om met zekerheid te zeggen of het heeft plaatsgevonden voor een jet. Dit heeft er vooral mee te maken dat bij deze botsingen er verschillende dingen kunnen gebeuren met een bepaalde kansverdeling. Jetquenching verandert niet de mogenlijke resultaten van een botsing, het verandert alleen de kans dat een bepaald resultaat plaatsvindt. Vergelijkbaar met bijvoorbeeld een verzwaarde dobbelsteen, die een 50% kans geeft op een zes, maar kan nog steeds alle andere getallen gooien. Op basis van één worp kan er niet worden gezegd of de dobbelsteen verzwaard is, maar aan de hand van honderd worpen kan dat wel.

Het doel van dit onderzoek was om met behulp van een machine learning algoritme een patroon te vinden dat ons zou kunnen vertellen of jetquenching had plaatsgevonden. Er is geen gebruik gemaakt van echte data van een deeltjesversneller. In plaats daarvan is er gebruik gemaakt van gegenereerde data. Dit omdat de theorie van hoe de botsingen plaatsvinden goed bekend is en de data dus goed is. Daarbij is het gebruik daarvan makkelijker. Als het werkt op deze gegenereerde dataset, dan moet het ook werken op echte data.

We hebben met een machine learning algoritme kunnen vinden dat er een meetbaar verschil blijkt te zijn tussen de twee gegenereerde datasets. Dit betekent waarschijnlijk dat het het algoritme gelukt is om een verschil te vinden tussen de proton-proton botsingen en de lood-lood botsingen. Het probleem op dit moment is dat we nog niet met zekerheid kunnen zeggen waar dat verschil vandaan komt. Het zou namelijk kunnen zijn dat dit verschil niks met jetquenching te maken heeft, maar eigenlijk een andere reden heeft. Er moet daarom nog meer onderzoek gedaan worden om met zekerheid te kunnen zeggen of dit algoritme in staat is om jetquenching te detecteren. Daarbij is het heel interessant om te kijken welke preciese fysische grootheden, zoals bijvoorbeeld de energie of richting van deeltjes, het algoritme gebruikt om te bepalen of jetquenching heeft plaatsgevonden. Dit zou ons kunnen helpen om een beter begrip te krijgen van de data die we detecteren in deeltjesversnellers. Dit kan ons helpen begrijpen hoe de deeltjes, waaruit alles om ons heen is opgebouwd, werken.

References

[1] A. G. Van Melsen, *From atomos to atom: The history of the concept atom* (Courier Corporation, 2004).

[2] J. J. Thomson, *On bodies smaller than atoms* (US Government Printing Office, 1901).

[3] W. N. Cottingham and D. A. Greenwood, *An introduction to the standard model of particle physics* (Cambridge university press, 2007).

[4] J. Greensite, *An introduction to the confinement problem*, vol. 821 (Springer, 2011).

[5] *The nobel prize in physics 1990*, URL `https://www.nobelprize.org/prizes/physics/1990/summary/`.

[6] K. Zapp, G. Ingelman, J. Rathsman, J. Stachel, and U. A. Wiedemann, The European Physical Journal C **60**, 617 (2009).

[7] URL `http://home.thep.lu.se/~torbjorn/pythia81html/Welcome.html`.

[8] *Hepforge*, URL `https://jewel.hepforge.org/`.

[9] *Cern accelerating science*, URL `https://home.cern/science/accelerators/large-hadron-collider`.

[10] *Alice - a large ion collider experiment*, URL `http://aliceinfo.cern.ch/Public/en/Chapter1/results.html`.

[11] *A large ion collider experiment*, URL `http://aliceinfo.cern.ch/Public/en/Chapter2/Chap2Experiment-en.html`.

[12] R. K. Ellis, W. J. Stirling, and B. R. Webber, *QCD and collider physics* (Cambridge university press, 2003).

[13] P. A. M. Dirac, Proceedings of the Royal Society of London Series A **114**, 243 (1927).

[14] R. P. Feynman, Physical Review **76**, 769 (1949).

[15] A. Salam and J. C. Ward, Il Nuovo Cimento **11**, 568 (1959).

[16] *Quark-gluon plasma*, URL `https://www.physicscentral.com/explore/action/gluon.cfm`.

[17] R. S. Bhalerao, arXiv preprint arXiv:1404.3294 (2014).

[18] O. b. L. A. N. S. Los Alamos National Laboratory, *Quark-gluon plasma*, URL `https://www.lanl.gov/projects/dense-plasma-theory/background/quark-gluon-plasma.php`.

[19] T. K. Nayak, Pramana **79**, 719 (2012).

[20] R. Kogler, B. Nachman, A. Schmidt, L. Asquith, E. Winkels, M. Campanelli, C. Delitzsch, P. Harris, A. Hinzmann, D. Kar, et al., Reviews of Modern Physics **91**, 045003 (2019).

[21] M. Gyulassy, in *Structure and dynamics of elementary matter* (Springer, 2004), pp. 159–182.

[22] *Cern accelerating science*, URL `https://cms.cern/news/jet-quenching-observed-cms-heavy-ion-collisions`.

[23] J. Casalderrey-Solana and C. A. Salgado, arXiv preprint arXiv:0712.3443 (2007).

[24] I. Lokhtin and A. Snigirev, The European Physical Journal C-Particles and Fields **45**, 211 (2006).

[25] URL `http://home.thep.lu.se/~torbjorn/pythia81html/Tunes.html`.

[26] URL `https://jetquenchingtools.web.cern.ch/JetQuenchingTools/samples/jewel_NR_2.2_5.02_Sep18/dijets_pthat150/PU14Simple/`.

[27] Mverwe, *mverwe/jettoyhi*, URL `https://github.com/mverwe/JetToyHI/tree/forbsc`.

[28] *learn*, URL `https://scikit-learn.org/stable/index.html`.

[29] A. L. K. Datta, A. L. K. Datta, J. H. J. Gallicchio, M. K. J. Cogan, E. M. PT. Komiske, E. M. PT. Komiske, T. P. G. Kasieczka, J. C. D. Guest, B. N. EM. Metodiev, M. F. T. Cohen, et al., *Junipr: a framework for unsupervised machine learning in particle physics* (1970), URL `https://link.springer.com/article/10.1140/epjc/s10052-019-6607-9`.

[30] H. A. Andrews, L. Apolinario, R. A. Bertens, C. Bierlich, M. Cacciari, Y. Chen, Y.-T. Chien, L. C. Mendez, M. Deak, D. d'Enterria, et al., Journal of Physics G: Nuclear and Particle Physics **47**, 065102 (2020).

[31] URL `https://www.tensorflow.org/`.

A Appendix

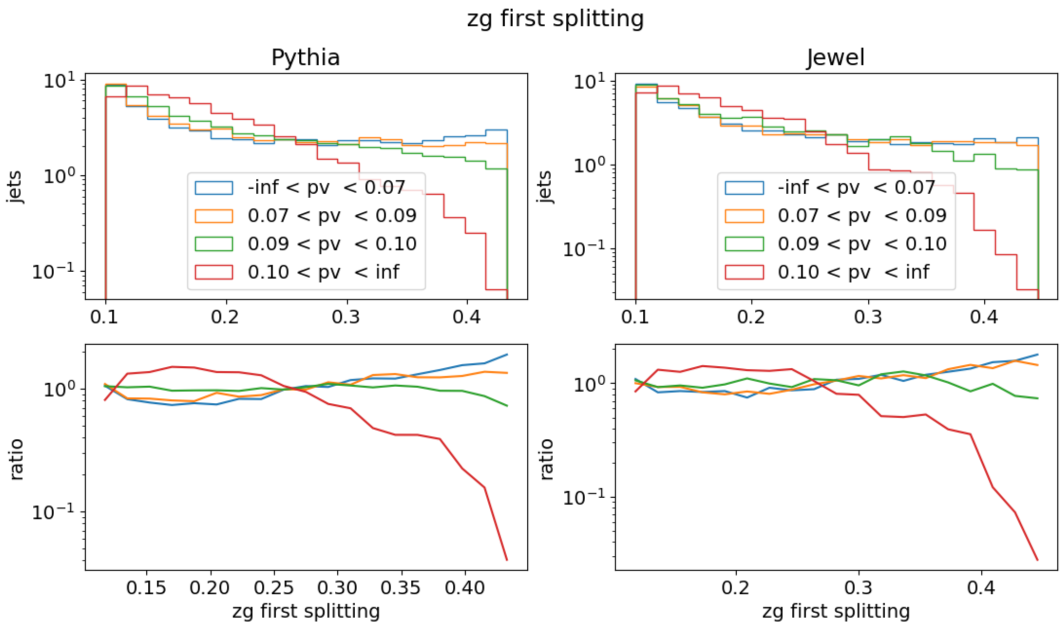

Figure 13: Above: normalized histograms of the zg of the first splitting against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

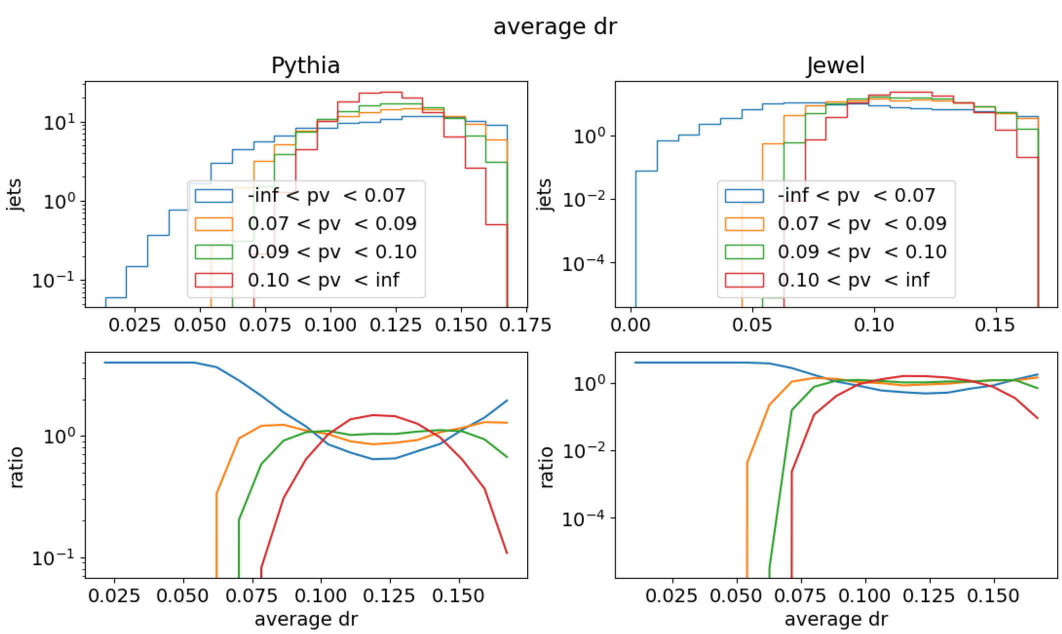

Figure 14: Above: normalized histograms of the average of the dr of all splittings against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

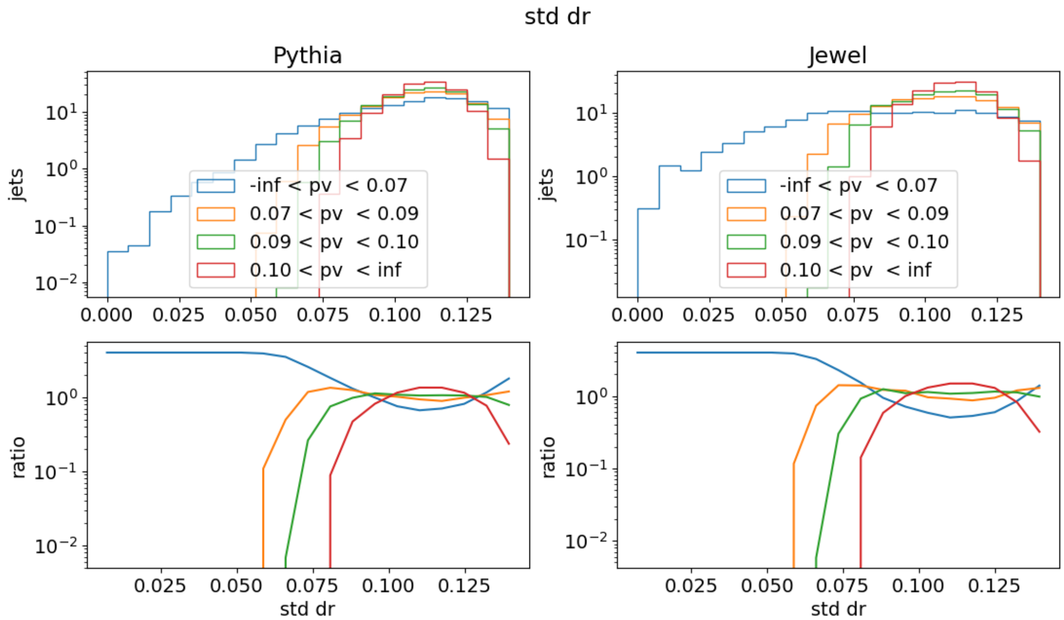

Figure 15: Above: normalized histograms of the standard deviation of the dr of all splittings against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

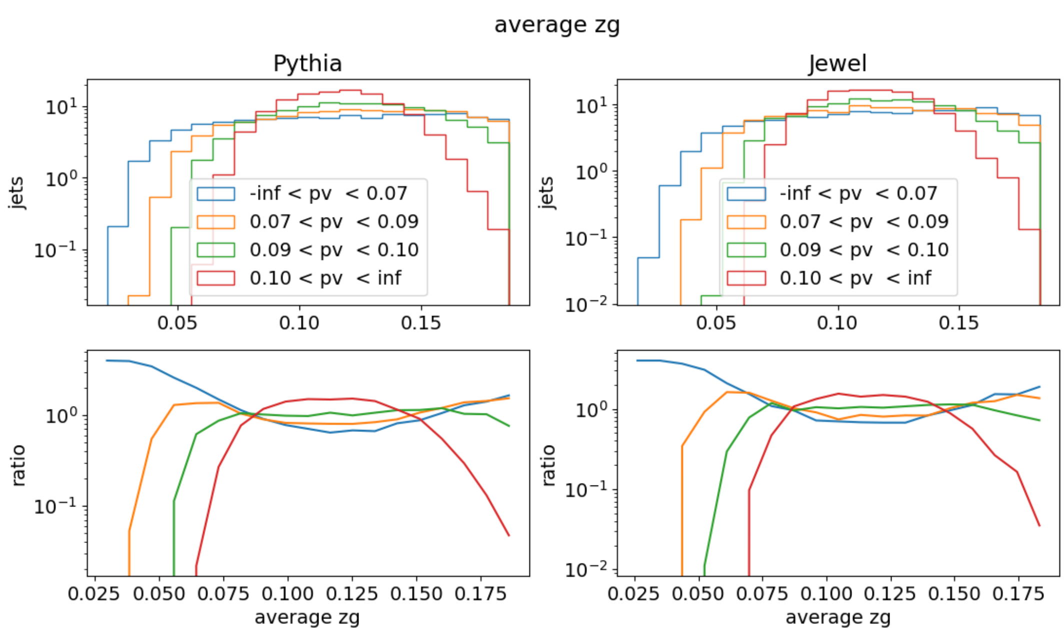

Figure 16: Above: normalized histograms of the average of the zg of all splittings against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

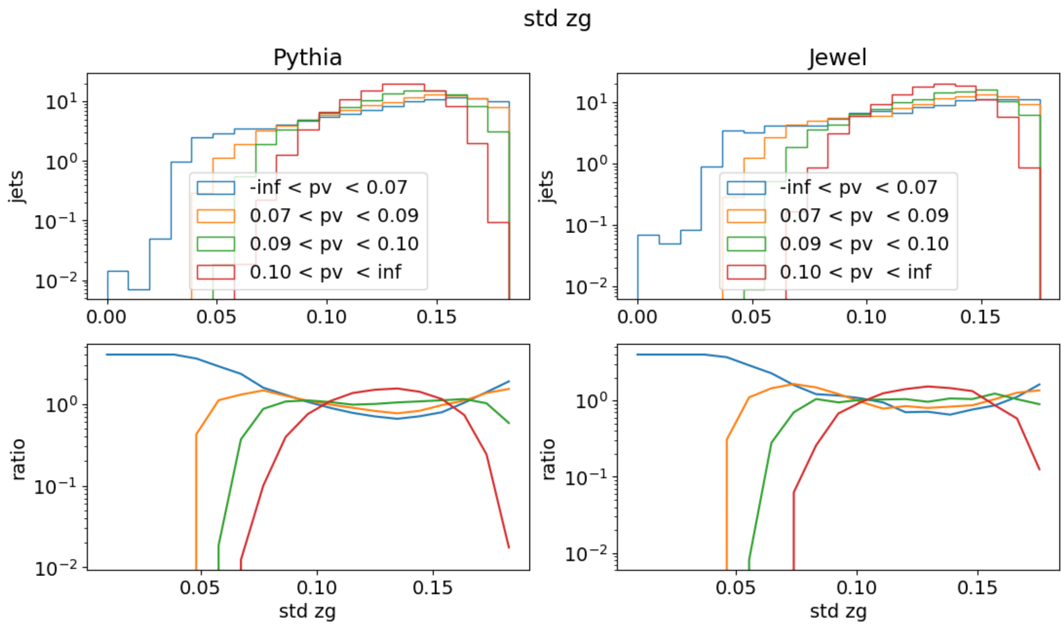

Figure 17: Above: normalized histograms of the standard deviation of the dr of all splittings against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

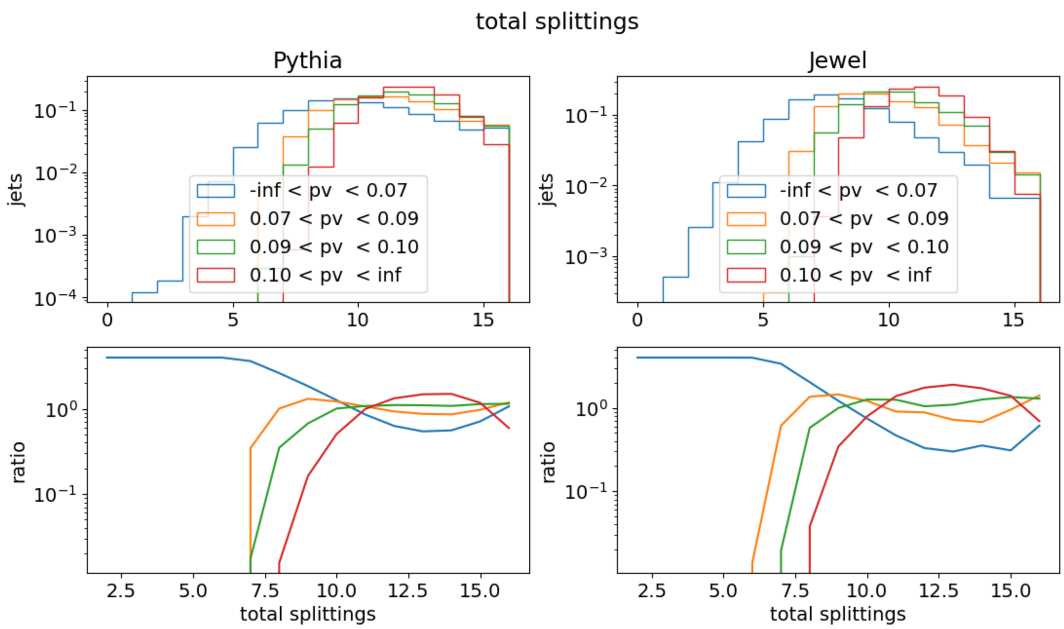

Figure 18: Above: normalized histograms of the total amount of splittings against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

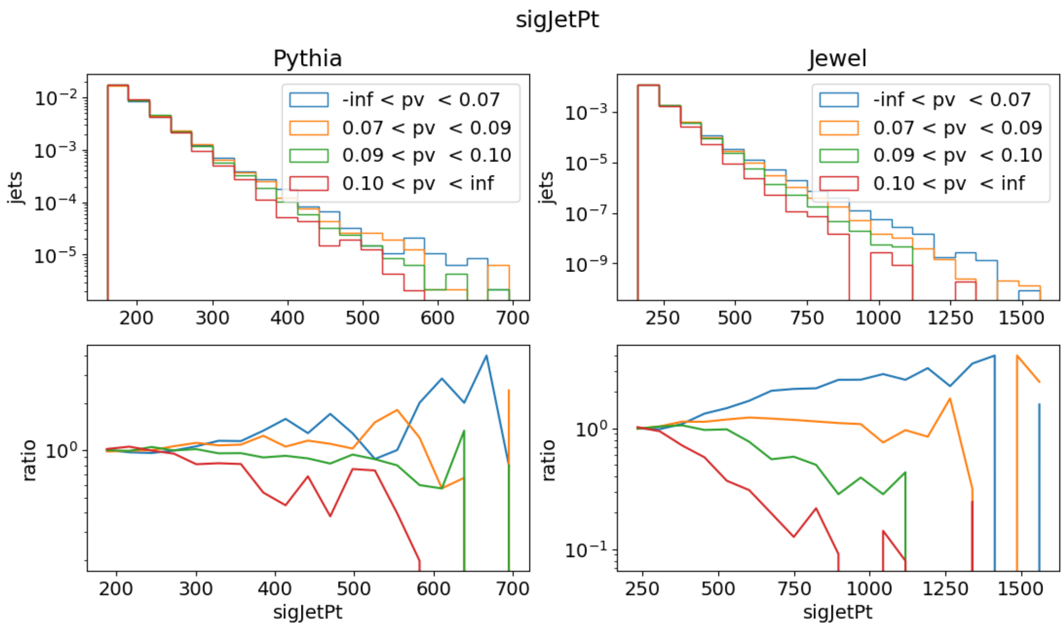

Figure 19: Above: normalized histograms of the $\hat{p_T}$ against the amount of jets, for both Pythia and jewel. Below: ratio plots of those same variables. Where the ratio is defined as the value divided by the average value of all the bins.

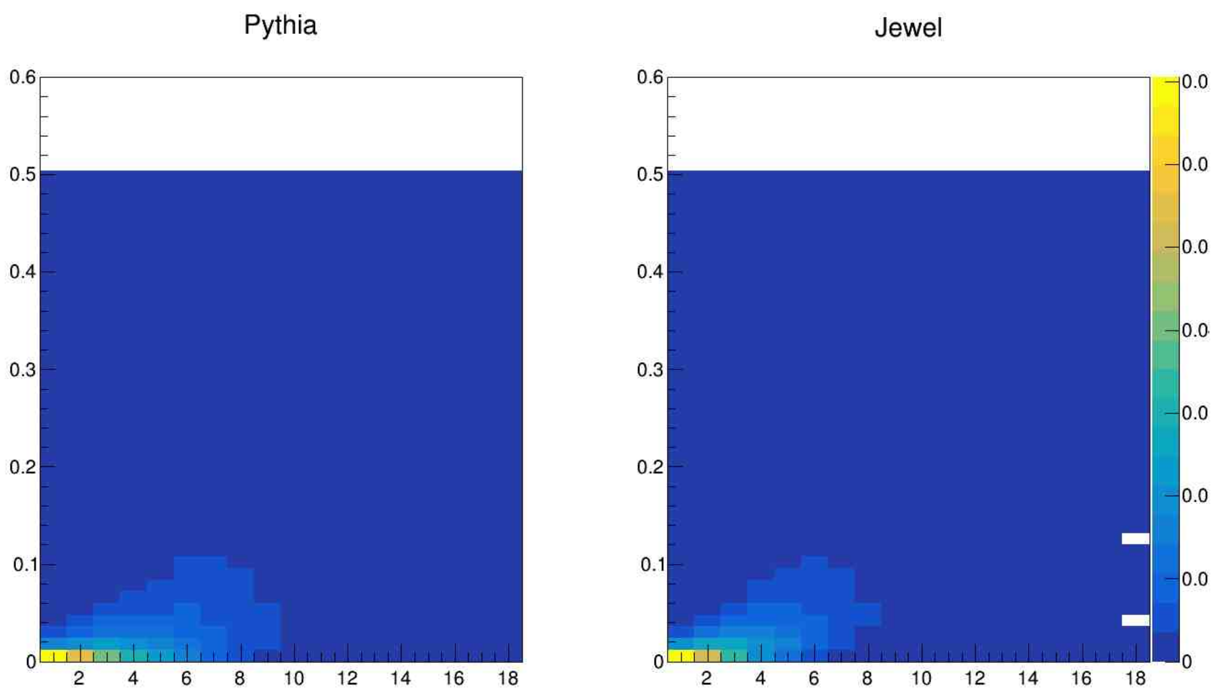

Figure 20: 2D histograms of the dr for all the different splittings for both Pythia and Jewel.On the x axis you can see the number of splittings (so the first, second, third etc.). On they axis you can see the value for dr at this splitting. Here all the jets are added together.