Determination of non-perturbative corrections in Z+jets prdouction at the LHC using iterative methods

Bachelor-Arbeit
zur Erlangung des Hochschulgrades
Bachelor of Science
im Bachelor-Studiengang Physik

vorgelegt von

Paul Moder
geboren am 24.05.1993 in Dresden

Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften
Technische Universität Dresden

Eingereicht am 14. Juni 2016

1. Gutachter: Dr. Frank Siegert
2. Gutachter: Prof. Dr. Dominik Stöckinger

Zusammenfassung

Zusammenfassung
Deutsch:

Ziel der Arbeit war es, nicht-perturbative Korrekturen für Z+jets-Produktionen am LHC mit Hilfe der Methode des iterativen Unfolding zu berechnen.
Durch Nutzung des SHERPA Event Generators wurden Z+jets-Events sowohl auf Hadron- als auch auf Parton-Level erstellt. Anschließend wurden diese Events mit Hilfe des Programms Rivet analysiert. Es wurden dabei verschiedene Observablen untersucht, die sich in ihrer Sensitivität bezüglich nicht-perturbativer Effekte unterscheiden. Durch Verwendung des Programms RooUnfold konnte schließlich ein Response-Matrix-Objekt erstellt werden, welches nicht-perturbative Korrekturen zwischen beiden Leveln enthielt.
Schlussendlich wurde dieses Objekt verwendet, um andere Parton-Level-Distributionen auf zwei unterschiedliche Methoden zu unfolden und diese dann zu vergleichen. Die verwendeten Methoden waren dabei das Bin-by-Bin-Verfahren und das iterative Unfolding.

Abstract
English:

The aim of the presented thesis was to determine non-perturbative corrections in Z+jets production at the LHC using iterative unfolding.
By using the event generator SHERPA, Z+jets events were simulated on hadron and parton level, which were analysed with the program Rivet afterwards. A lot of observables with different sensitivity for non-perturbative effects were examined. Eventually by using RooUnfold a response matrix object could be created, which contained information about the non-perturbative corrections.
Finally another parton level distribution was unfolded with two different methods by using the response matrix object. The results of the unfolding was then used to compare these methods. The first method was the bin-by-bin method and the second one was the iterative unfolding.

Contents

1 Introduction

Today the standard model of particle physics is the acknowledged theory about elementary particles and their interactions. When it was first developed there were nearly no possibilities to prove this theory. For example, the Higgs boson was already postulated by Peter Higgs in 1964, the detection however happened nearly 50 years later in 2012. Today detections of particles are possible due to the high energy colliders, which allow to investigate high energy physics. One of the most important colliders was the Large Electron-Positron Collider (LEP), which allowed to verify some assumptions of the standard model. However, it was possible to prove most parts about the existence and interactions of elementary particles, when physicists started to experiment at the Large Hadron Collider (LHC) in the year 2008. At the LHC it is possible to collide hadrons with energies higher than ever before. Currently experiments are done with an invariant mass of 13 TeV and with the help of new detectors like the ATLAS detector, it is also possible to measure the produced particles. Because of these new possibilities, researches on theories, which go beyond the standard model, can be done like theories about a super symmetrical partner for each particle (SUSY [1]).
The standard model classifies the elementary particles. First it differentiates between the spins. Particles with a spin of $\frac{1}{2}$ are called fermions. There are a bunch of different fermions, such as charged leptons (electron, muon, tau), neutral leptons (electron-neutrino, muon-neutrino, tau-neutrino), up-type quarks (up, charm, top) and down-type quarks (down, strange, bottom), all together with their antiparticles. Particles with a spin of 1 are called bosons. These bosons are further classified by the interaction, they are responsible for. For the weak interaction exist three bosons, the W^+, the W^- and the Z boson, for the electromagnetic interaction exists the photon and the gluon is needed for the strong interaction. The last particle of the standard model, which was already mentioned above, is the Higgs boson, which is the only elementary particle with a spin of 0. Its part in the standard model is to give mass to all the particles through a so called Higgs field. The achievement of the high energy colliders was not only to discover those particles, but also to measure their masses. The latest proved particle was the Higgs boson in the year 2012 with $\mathrm{m_H} = 125\mathrm{GeV}$.
With the help of these measurements and therefore the prove for the correctness of the theories in the standard model, it is now also possible to simulate high energy collisions together with the followed interactions. Simulations have the advantage, that a large amount of data points can be created, which can later be used to determine the amount of background in a future measurement, e.g. researches on unsolved questions of the standard model, like the absence of the antimatter or researches on theories, which go beyond the standard model. The simulated data points can be important for deciding, which structures in a measurement come from the researched particle and which are probably just background.
The simulation itself is just a simplification of the true process and can be divided into two different levels: the hadron and the parton level (see also section 2.1). For simulating those levels the most important approximation in particle physics is used: the perturbation theory. But, as with every

approximation, it is only as accurate as the order of made corrections. Corrections of higher order are normally ignored and those are the so called perturbative corrections. However the perturbation theory alone is not able to describe all systems, e.g. QCD interactions at lower energies behave non-perturbatively. These non-perturbative effects lead to differences between the hadron and the parton level. The focus of this thesis will lie on analysing the differences between both levels and therefore on determining the non-perturbative effects, because these are hard to predict in theory. To derive those non-perturbative effects the unfolding method is used. Normally this method is important for calculating the true distribution out of one, which was measured in the detector, because the problem is, that the measuring in the detector smears the results and therefore the measured values differ from the true ones. In this thesis, the unfolding will be used to calculate the hadron level out of a parton level simulation and therefore mimicking the non-perturbative corrections. Two methods will be compared: the bin-by-bin method and the iterative unfolding, which is using Bayes' theorem.

2 Theory

2.1 Simulating collisions

In order to simulate such complex processes like collisions between different particles and the consequential production and decay of other particles, the processes have to be simplified, they must be modelled. A typical model of how an event generator simulates a proton-proton collision is shown in figure 2.1. This model will be shortly explained in this section to understand the basics of a simulation. Detailed information can be found in the publication for event generation in SHERPA [2]. The two protons which collide are the green ellipses on the right and the left. In the simulation the collision is now divided into different parts according to their energies. First there is the big red blob in the center of the picture. This is the hard scattering process, which is the fundamental physics of the simulation and which is calculated through the use of the perturbation theory. That means, this process is calculated with matrix elements. In this picture, the simulated hard process is the production of a Higgs boson and a $t\bar{t}$ pair (the little red blobs) through the reaction of two gluons (red spirals). All of these particles have a small life span and so they decay shortly after the production. The blue spirals are also gluons, which were together with the red gluons produced through QCD radiation, but did not take part in the hard process. Furthermore the blue gluons are not calculated through the perturbation theory in the event generator, but the so called parton shower. Besides the hard process a lot of other low energy processes can also take place. Those are displayed by the purple color.

After a short time the partons hadronise, which means they become bound (light green particles), before they can decay even further (dark green particles). The simulation is stopped after these particles, because these are the ones, which would reach the detector in a real collision at the LHC and because of that this state is referred to as the truth level. Additionally, photons can be emitted by charged particles at any time (yellow waves).

For this thesis two levels are of interest: the level after the hard process and the parton shower, but before the hadronisation, which is called the parton level and the level directly after the hadronisation (the truth level), which is called hadron level.

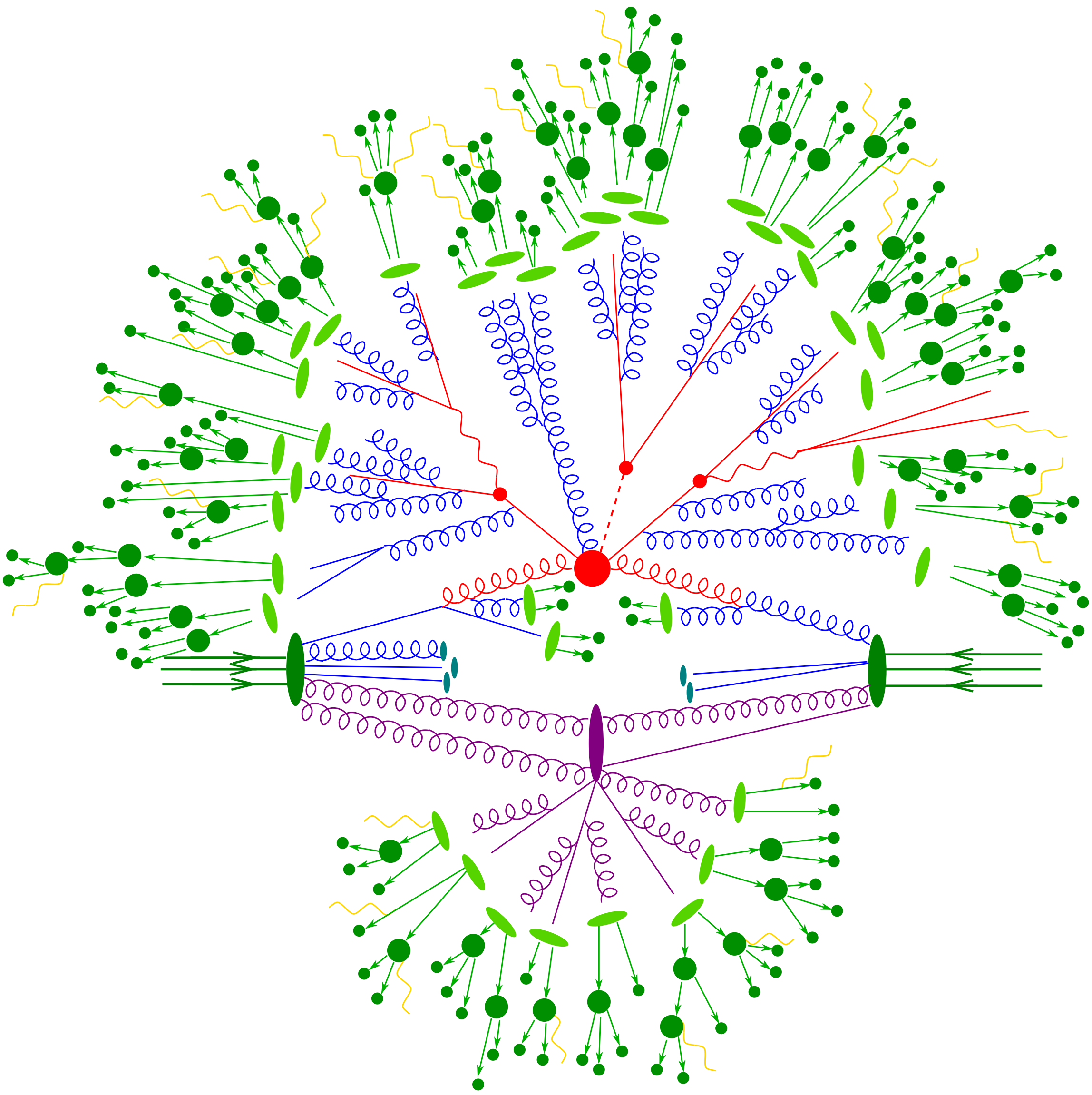


Figure 2.1: Model of a proton proton collision at the LHC [3]

2.2 Anti-k_T algorithm

Quarks and gluons can combine to composite particles, e.g. the proton is a combination of two up- and one down-quark together with gluons and virtual sea quarks[1]. But in contrast to other elementary particles quarks and gluons can not exist on their own, they need to be bound in so called hadrons. This is a result of the strong interaction, more precisely the confinement. The characteristic of the strong interaction is, that the energy between two particles increases, if the distance between them increases[2]. Quarks would be called free, if the distance between them is high enough, but then also the energy is high enough to produce a new pair of quarks, which means the quark are bound again with a different quark.

Hadrons can further be divided into two main groups: mesons and baryons. Baryons and mesons

[1]Near a hadron, sea quarks can be created out of gluons. But in general these are directly annihilated to a gluon again. This leads to constant creation and annihilation of sea quarks and gluons around the hadron, which builds the so called sea.

[2]In comparison, the energy between two electrically charged particles decreases with increasing distance.

differ in the number of quarks they are made of. Mesons are particles, which consist of two quarks, while baryons consist of three quarks. There are also particles, which are built of more than three quarks like the Z(4430) [4], which is made out of four quarks, but those are extremely rare. However, to be able to mimic the original partons, that were created, bundles of hadrons are then combined to jets. Depending on the jet clustering algorithm the jets can then have different shapes.
Now the principle of the later used jet clustering algorithm will be presented. The algorithm is called anti-k_T and was first presented in the paper of Cacciari, Salam and Soyez [5]. It is one part of a jet algorithm, which first needs the distances between entities and their distances to the proton beam. Entities can be either particles or pseudo jets. The distance between the entities i and j is called d_{ij} and the distance to the beam B d_{iB}. The clustering then works as follows. First the smallest distance is identified and if it is a distance between two entities, those two entities are recombined to a pseudo jet. If it is a distance of an entity and the beam, the entity is called a jet and is removed from the list of entities. After that the distances are calculated again and the process is repeated. This algorithm ends, if no more entities are left. The special part about this algorithm is now, how those distances are calculated:

$$d_{ij} = min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{2.1}$$

$$d_{iB} = k_{ti}^{2p}. \tag{2.2}$$

In this equations k_{ti} and k_{tj} are the transverse momentums of particle i or j and Δ_{ij} is the geometrical scale for the entities i and j

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2 \tag{2.3}$$

with the rapidities y_i and y_j and the azimuthal angles ϕ_i and ϕ_j of the particles i or j. The parameter R is the distance parameter, which is also often called jet radius. R describes the minimal angular distance[3], two reconstructed jets must at least have. It can easily be seen, that the distance between two entities decreases with a greater distance parameter, while the distances to the beam stay the same. That means that that more particles are added to one jet. The reason for that is, that the distance between entities is often smaller than the distance to the beam, when R is greater. The result is, that more particles are added to a pseudo jet entity before the entity is removed as a jet. In addition to this parameter, the parameter p was added by Cacciari et al., which describes the relative power of the energy against the geometrical scales Δ_{ij}. For different parameters p, different jet clustering algorithms are received. For $p = 1$ the algorithm is the inclusive k_T algorithm [6], if $p = 0$ the algorithm is the same as the Cambridge/Aachen algorithm [7], [8] and if p is set to -1, there is a new algorithm, called the anti-k_T algorithm.

2.3 Z+jets production

In this thesis processes will be analysed, in which a Z boson and a variable number of jets are produced, when two protons, each with an energy of 3.5 TeV, collide. As it was described in section 2.1, jets are bundles of hadrons, which can be measured in the detector. However the Z boson is an unstable particle. After its production from two partons, it decays before it can reach the detector. The Z

[3]The angular distance is calculated like the geometrical scale of two entities thorugh the rapidity distance and the azimuthal angle distance.

boson can thereby decay into hadrons, a pair of a lepton (electron, muon, tau) and its anti particle or into a neutrino and its antiparticle. In this paper, only the decay into leptons is researched, because neutrinos can not be detected in the detector and because it is hard to identify, which hadrons are a product of a Z boson decay, when other jets are produced as well. Furthermore out of the three leptons only electron-positron and muon-anti-muon pairs are analysed, because the tau-lepton itself is an unstable particle, which decays before reaching the detector. So a possible production and decay of the Z boson for this analysis could look like figure 2.2.

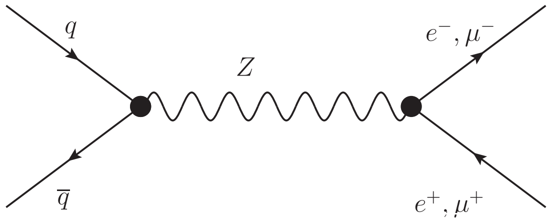


Figure 2.2: $q\overline{q} \to l^+l^-$ with a Z boson as intermediate particle

That also means, that the Z boson itself can not be measured, only its decay products. The problem is, that those leptons are not only produced in the Z decay, e.g. two leptons can also be produced due to two partons with a photon as intermediate particle instead of a Z boson, which can be seen in figure 2.3.

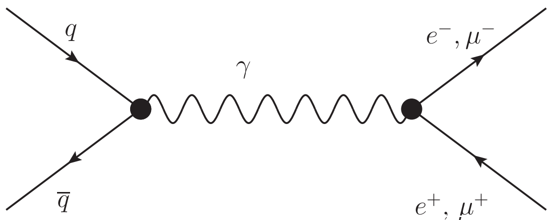


Figure 2.3: $q\overline{q} \to l^+l^-$ with a photon as intermediate particle

So the first task of the analysis is to find those leptons, which are products of the Z boson decay. This will be later discussed in section 3.3.

Besides the Z boson partons and gluons can also be produced, which later hadronise to hadrons and which are reconstructed to jets with the anti-k_T algorithm. An example for a possible jet production besides the Z boson is shown in figure 2.4.

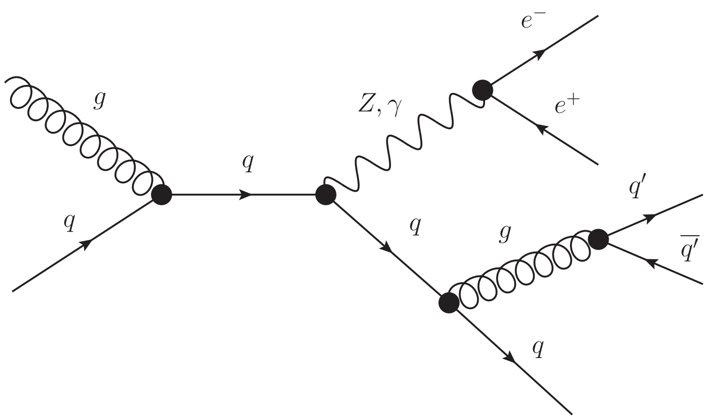


Figure 2.4: $qg \to l^+l^-qq'\overline{q'}$

2.4 The unfolding method

2.4.1 Unfolding a distribution

In every experiment with a detector, the measured values for observables differ from the true ones because of effects like limited detector performances. To calculate the true values out of the measured ones the unfolding method is used, which can be done with different approaches. Two of those approaches, which are already used in ATLAS analysis, are explained briefly in this section. More detailed information for both can be found in the publication about the unfolding at ATLAS [9].
The first approach is the bin-by-bin method, which calculates the ratio between the number of events in a bin of the true distribution and in the same bin of the measured one. The factors obtained for every bin can then be used in a different measurement to calculate the true distribution out of the measured one. So this is a more simple method and therefore easy to apply, which is why this is the most often used method, when unfolding is needed. However this method has some shortcomings, e.g. migrations of events between bins can not be estimated and it does not take into account the correlation between adjoining bins.
The second method is the iterative unfolding using the Bayes' theorem. It was developed by D'Agostini and is a lot more complex than the bin-by-bin method [10]. This section will only concentrate on the basics and the resulting algorithm, which will later be applied in the analysis to unfold the distributions. The primary advantage of the iterative unfolding is, that it is able to take migrations between different bins into account, when calculating the true distribution out of the measured one. Besides this, it has a lot of other advantages too, given in [10].
The basic for the iterative unfolding is the Bayes' theorem, which is described in equation (2.4).

$$P(C_i|E) = \frac{P(E|C_i)P(C_i)}{\sum\limits_{l=1}^{n_c} P(E|C_l)P(C_l)} \tag{2.4}$$

In this formula, the C_i ($i = 1, 2, ..., n_c$) are different causes which can produce one effect E. The probability for the cause is $P(C_i)$ and the probability of this cause to produce the effect E is $P(E|C_i)$. The calculated result $P(C_i|E)$ however is the probability, that a single event with the effect E has

originated from cause C_i. It is important to state, that the knowledge of the probabilities $P(C_i)$ can be increased by increasing the number of observations, but that does not apply to the knowledge of the probability $P(E|C_i)$.
For unfolding a distribution, the theoretical number of events, which are expected to be created due to the cause C_i, have to be calculated first.

$$\hat{n}(C_i) = n(E)P(C_i|E) \tag{2.5}$$

$P(C_i|E)$ is thereby the probability, which was estimated in equation (2.4) and $n(E)$ is the number of events with the effect E. However this equation has to be further adjusted to a real measurement, because in reality a cause C_i can create multiple effects E_j ($j = 1, 2, ..., n_E$) and not just one. If in addition the number of observed events N_{obs} and the efficiency ϵ_i to detect the cause C_i in one effect are considered as well, equation (2.5) can be rewritten.

$$\hat{n}(C_i) = \frac{1}{\epsilon_i}\sum_{j=1}^{n_E} n(E_j)P(C_i|E_j) \tag{2.6}$$

It can be seen, that ϵ_i is not allowed to be zero. If it is, then $\hat{n}(C_i)$ will also set to 0, because that would mean, that the experiment is not sensitive to the cause C_i. However equation (2.6) leads then to the true total number of events

$$\hat{N}_{\text{true}} = \sum_{i=1}^{n_C} \hat{n}(C_i) \tag{2.7}$$

together with the final probabilities of the causes

$$\hat{P}(C_i) = P(C_i|n(E)) = \frac{\hat{n}(C_i)}{\hat{N}_{\text{true}}} \tag{2.8}$$

and the overall efficiency of the observation

$$\hat{\epsilon} = \frac{N_{\text{obs}}}{\hat{N}_{\text{true.}}} \tag{2.9}$$

D'Agostini demonstrated, that the final distribution $\hat{P}(C)$ lies between the initial distribution $P_0(C)$ and the true one, which suggests to use an iterative algorithm to unfold a distribution, that is defined as follows:
1. take the initial number of events $n_0(C_i) = P_0(C)N_{\text{obs}}$
2. use the formulas above to calculate the distributions $\hat{n}(C)$ and $\hat{P}(C)$
3. use the χ^2 method to calculate the difference between the distributions $\hat{n}(C)$ and $n_0(C)$
4. replace the initial distributions $\mathrm{P_0(C)}$ and $\mathrm{n_0(C)}$ with the new calculated ones $\hat{P}(C)$ and $\hat{n}(C)$
After that, the steps 2 to 4 have to be repeated until the value for χ^2 is finally small enough. The result, one gets out of this process is finally the last distribution created in step 2.

2.4.2 Estimation of the Uncertainties according to D'Agostini

In this section the uncertainties of the iterative unfolding according to D'Agostini [10] are examined. The problem is, that it is not as simple as just calculating the square root of the number of events, because the uncertainties of the $n(C_i)$ are correlated. Due to this fact, it is also not possible to directly use the error propagation of Gauss, one needs a covariance matrix V, which also describes the correlation between the C_i. The elements of this matrix are then the sum of the uncertainties of the $n(E)$ and the unfolding matrix M with

$$M_{ij} = \frac{P(E_j|C_i)P_0(C_i)}{\left[\sum_{l=1}^{n_E} P(E_l|C_l)\right]\left[\sum_{l=1}^{n_C} P(E_j|C_l)P_0(C_l)\right]}. \tag{2.10}$$

So the uncertainties for both terms have to be calculated. The first term, the uncertainties of the $n(E)$ can be calculated very quickly.

$$V_{kl}(n(E)) = \sum_{j=1}^{n_E} M_{kj}M_{lj}n(E_j)\left(1 - \frac{n(E_j)}{\hat{N}_{\text{true}}}\right) - \sum_{\substack{i,j=1 \\ i\neq j}}^{n_E} M_{ki}M_{lj}\frac{n(E_i)n(E_j)}{\hat{N}_{\text{true}}} \tag{2.11}$$

The second term, the errors of the unfolding matrix M, are divided into statistical and systematic errors. Statistical errors occur due to the limited number of events, while systematic errors are a result of assumptions made in the simulation. The systematic uncertainties have to be calculated by methods, which are appropriate for the made assumptions. For the statistical uncertainties on the other hand, covariances between terms, which are related to the same cause-cell C_i, have to be considered. In the end, the covariance matrix elements for the unfolding matrix M can be calculated with equation (2.12).

$$V_{kl}(M) = \sum_{i,j=1}^{n_E} n(E_i)n(E_j)\text{Cov}(M_{ki}, M_{lj}) \tag{2.12}$$

$\text{Cov}(M_{ki}, M_{lj})$ describes thereby the covariances between the matrix elements for the statistical uncertainties. The detailed calculation is given in [10].

3 Implementation

In this chapter, it will be explained, how the methods described in the last chapter are used in this thesis. Especially the unfolding will have a different function, than in the normal detector analysis. As described in section 2.4 the real values of a process are not equal to the ones measured in the detector, which is why unfolding is used. In this thesis however the non-perturbative corrections are investigated, which lead to a difference in the values of the parton and the hadron level (see section 2.1). The problem is, that these corrections can not be predicted as easily as the perturbative corrections. So the idea is, that it could be possible to determine those corrections by using an unfolding method. If this is possible, higher order parton level calculations could be compared to hadron level data (simulated or measured). In this chapter it will be explained, how the different levels can be simulated, analysed and later be unfolded to get an understanding of how the results in chapter 4 are created.

3.1 Programs

To simulate and analyse the data for the given process, three different programs are needed. The first program is SHERPA [11], a Monte Carlo event generator which was developed partly at the TU Dresden and is used to simulate events for particle collisions with high energy. For example proton-proton collisions with a centre-of-mass energy of 7 TeV, which is currently also used in experiments at the LHC (3.2), can be simulated.

The second program is Rivet [12], which is needed to analyse the received events from the Monte Carlo event generator SHERPA. With Rivet it is possible to search for particles with specific characteristics and to determine their mass, transverse momentum and so on. It also shows the results in histograms and allows to compare them with measured data at the LHC, if data for that observable is existing.

The last used program is RooUnfold [13], which is implemented into Root. This program is able to unfold measured distributions or in this case, the parton level distributions to receive hadron level distributions (truth level). There are different methods implemented, in this paper the method for the bin-by-bin unfolding and Bayes' theorem (see section 2.4) are used and then compared.

3.2 Simulating events in Sherpa

To simulate a process with SHERPA, a file called Run.dat must be written. An example for such a file can be found in the appendix (A.1). In this section the most important parts of the analysis for the Z+jets production will be explained, so that the physics behind the program is clear. The file itself is split into different parts. In the first card, the so called run card, the parameters for the event generation are set. For example in this card the number and the error of the events, which generator and what mode is used and how much jets are estimated through the matrix element or rather by

using a parton shower simulation are defined.
An important part for the collision is the collider setup, where the defining of the colliding beams is done. In this example two protons (number in the PDG is 2212 [14]), each with an energy of 3500 GeV, collide. As it was explained in section 2.3, the decay of the Z boson into a tau lepton pair is not analysed. So to be later able to use the lepton container as a product of the hard process, the tau lepton has to be excluded from the container with "MASSIVE[15] 1". The number 15 indicates again the number of the tau lepton in the PDG [14]. Another important part in this card can be the parton level simulation. By default Sherpa simulates the particles on the hadron level, because this is the level, which interacts with the detector. But in this thesis the parton level is also needed. In order to simulate it, some effects for the particles have to be suppressed. These effects are the hadronization of the particles, indicated by the fragmentation, the multiple parton interaction, indicated through the MI_Handler and the intrinsic transverse momentum of the beams, indicated by the K_Perp settings.
After that comes the process card, which defines the generated hard scattering process, in this thesis the production of a Z boson together with a number of jets. So two partons of the protons collide and create a Z boson, which immediately decays into a lepton and its antiparticle. This is defined in the first line of the process card. The number 93 is a container of all the possible partons, two of them collide and produce two particles from the lepton container (90) and a number of jets, which contain partons of the parton container (93). The variable in brackets, "NJET", defines the number of particles out of this container, which are simulated through the matrix element calculation. Other particles of this container are simulated with the parton shower method [2].
As it was explained in section 2.3, lepton pairs can not only be produced due to the decay of the Z boson. A lot of other processes can have lepton pairs as products too. So the definition of the process is not enough to declare, that in this event a Z boson was created. In the analysis with Rivet, events will be removed, if particles do not have certain characteristics (section 3.3). The problem is, that decreasing the number of events results in an increasing of the statistical error. So it would be advantageous to only simulate events, which already fulfil the requirements. For this task, there is the last card, the selector card. It ensures that only events with particles, which have certain characteristics, are created. In this example the invariant mass and the transverse momentum of the outgoing leptons is already defined, but the requirements are less strict than in the later analysis, because of the statistical fluctuation in each event.
At this point an important matter for the simulation of the Z+Jets process has to be explained. When comparing the simulated data with those measured by the ATLAS corporation, it can be noticed, that the simulated cross section for this process will always lie 10-15% below the measured cross section. The reason for that is a result of the simulation, which has of course some uncertainties. The first error is, that matrix elements are calculated through the perturbation theory up until a given order, which is already an approximation. A part of this approximation is, that only tree level matrix elements are calculated and not more complex ones, which include for example a loop. These errors of approximation can be corrected with perturbative corrections. But as described in chapter 1, non-perturbative corrections are also needed, because there can also be interactions, which are not possible to calculate with the perturbation theory. In addition there are also more simplifications like the fact, that only a defined number of particles is simulated through the perturbation theory, while the rest is simulated through the parton shower (see above).
Theoretically some of these errors can be minimized, for example one could increase the number of

jets, which are calculated with a matrix element. The problem is, that this would of course need a lot more computing capacities and time and it would only minimize the error by a very small amount, because the observables measured in the analysis are those, which are not very sensitive to non-perturbative effects. So most of the differences are a product of perturbative effects and for those the simulation would have to be a lot more complex. On the other hand, the error is known to be within 10-15%, so the easier way is, to just multiply the simulated data with a factor of 1.15 after the simulation to be able to compare it to the measured data.

3.3 Analysing events with Rivet

After the simulation of the events with Sherpa, they have to be analysed. This is done with the help of a Rivet analysis[1]. An analysis in Rivet is written in C++ and normally contains three different sections. In this section, the content of each section will be described and furthermore what was implemented in this analysis.

3.3.1 init

In the first section, the init section, the requirements for the particles can be defined, in this example for the leptons to identify decay products of the Z boson. The requirements for this analysis will be the same as in the publication of Aad et al. [15]. First each lepton of a possible pair needs to have a transverse momentum of at least 20 GeV. Furthermore the charge of the two leptons inside an analysed pair must be inverse, the separation between both ($\Delta R^{ll} = \sqrt{(\Delta\Phi)^2 + (\Delta\eta)^2}$) has to be at least 0.2. Because the Z boson has a mass of around 91 GeV [14] the invariant mass of the leptons must lie between 66 GeV and 116 GeV [2]. Beyond that the absolute value of the rapidity for each lepton must also lie in a particular area, but it is different, if the decay products are electrons or muons. For a pair of an electron and positron, the absolute value of the rapidity for both particles must be smaller than 1.37 or between 1.52 and 2.47. For a pair of a muon and anti-muon, the the absolute value of the rapidity needs to be smaller than 2.4.
Besides the leptons, the jets can be also defined in the init section. Jets will later be combined to lists of jets and reconstructed with different distance parameters. For this analysis, jets will be reconstructed with three different distance parameters (0.4, 0.6 and 1.0), depending on the observable, which is analysed (see section 3.3.2). Additional to that, jets also have to fulfil certain requirements, like the leptons in this analysis. Jets, which do not fulfil those requirements are not added to the list. For this analysis, each jet has to have a transverse momentum of at least 30 GeV, the absolute value of the rapidity must be smaller than 4.4 and depending on the distance parameter, which was used, the separation between the jet and lepton must be greater than 0.5, 0.7 or 1.1.
Finally the histograms, which will later be filled, are also defined in this section. If the histograms shall be used in Root later, it is also possible to declare them as Root histograms instead of Yoda histograms.

[1] The Rivet analysis, which was the basic of the one used in this thesis can be found on this web page: https://rivet.hepforge.org/code/dev/a00605_source.html

[2] Normally the invariant mass should be the same as the one of the Z boson, but because of the fact, that the Z boson is an internal propagator in the Feynman diagram, it is allowed to be off-shell. That means, that it does not have to fulfil the Einstein energy-momentum relationship.

3.3.2 analyze and finalize

In the second section each event will be analysed and if it fulfils the requirements defined in the init section, it can be used to fill the histograms. But before that is possible, the observables have to be determined, which can be done with the help of the Rivet system. This system allows to define different jets and particles and get their characteristics like rapidity, charge and momentum, which itself contain all the remaining information like mass, transverse momentum or the absolute value of the rapidity. If different histograms require different requirements for the particles, those can also be defined again with the help of if-conditions. The observables, which are analysed in this thesis, are first of all the 28 observables of the ATLAS_2013_I1230812 analysis [12]. These observables were already measured at the ATLAS detector in the year 2013 and contain characteristics of the Z boson and even more importantly of the outgoing jets, e.g. the invariant mass of the two leading jets, the distance between them, their rapidity and also the transverse momentum of the Z boson. During this analysis the jets were reconstructed with the anti-k_T algorithm and a distance parameter of 0.4, because the non-perturbative effects do not influence the result that strong and so the non-perturbative corrections can be ignored. In this thesis those corrections shall be examined, so the analysis was supplemented with all those observables using a distance parameter of 1.0. To be able to get even more histograms, which need non-perturbative corrections, the masses of single jets using a distance parameter of 0.4, 0.6 and 1.0 were also implemented.
The finalize section, finishes the histograms by scaling them with the weight of each bin (or other weights, which were defined) and writing them into a file.

3.4 Unfolding with RooUnfold

3.4.1 The unfolding process

Before the unfolding of the parton level can begin, response matrices for every observable have to be created. In principle response matrices are 2D histograms, which are created through analysing an event on hadron and parton level. If the particles have the required properties on both levels, the observable will be analysed and the event is written into a bin of the 2D histogram with the parton level on the x axis, the hadron level on the y axis and the cross section on the z axis. To be able to analyse events on hadron and parton level at once, the simulation has to be changed. The problem is, that events of the parton and hadron level are created separately, which means, that they are not correlated in the default settings of SHERPA. So to create correlated events, the random generator seed for the simulations on hadron and parton level had to be the same for each event in both runs. To enable that the "EVENT_SEED_MODE" was set to three. Also it has to be enabled, that the events were not just simulated, but written into a hepmc-file with the command "EVENT_OUTPUT=HepMC_Short[filename]", which allows later to combine the events from the parton and the hadron level and analyse them together in one analysis. The analysis then has to be rewritten to create the 1D parton and hadron level histograms and also the response matrices for each observable.
In order to unfold, a reference run is needed to generate the response matrix object in Root for every observable. This object is generated through the 1D histograms of the hadron and parton level, as well as the 2D histogram for the observable. After the response matrix object was created, it can be

used to unfold a new parton level of a different simulation. It is possible to unfold with the bin-by-bin method or the iterative unfolding, using Bayes' theorem. In both cases, the results are histograms, which represent the hadron level of the simulation.

3.4.2 The estimation of the uncertainties for the Unfolding with RooUnfold

In section 2.4.2 the estimation of the uncertainties according to D'Agostini was presented. These uncertainties however are not enough to describe the error of the unfolded distributions, which were created with RooUnfold. Therefore a different approach to calculate the error of the used method is presented in this section.
First of all, the uncertainties are classified into statistical and systematic errors. The statistical error can be determined by repeating the same process of section 3.4.1 with slightly different conditions and then calculating the deviation. To do that, first the 2D histograms are varied 2000 times with a Gaussian distribution. That means, that for every bin in the 2D histogram a random value is created by using a Gaussian distribution. The mean value is thereby the old value of the bin and the standard deviation is the bin error. For every variation the unfolding is done and after that the standard deviation of all the 2000 times for every bin is determined, which then describes the error on the 2D histogram (err2D). The same procedure is then done with parton level histograms, which will be unfolded later (errPart). The last error, which needs to be determined is the systematic error. For this error, the unfolding of the parton level is done with the normal number of iterations and after that another time with three more iterations. The error is then the difference between both values in every bin (errIter). At the end, the error is estimated out of these three single errors:

$$err = \sqrt{err2D^2 + errPart^2 + errIter^2} \tag{3.1}$$

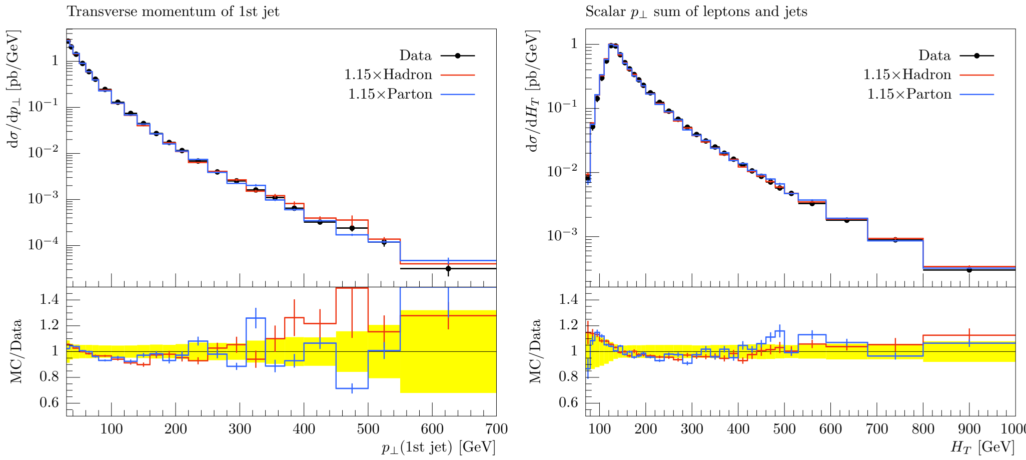
4 Results

4.1 Comparing Hadron and Parton level simulations

As written in the previous chapters, this thesis deals with the differences in simulations of the parton and hadron level. These differences can be shown after the two levels were simulated and analysed with SHERPA and Rivet. At first the analysis concentrate on the observables, which were used in the ATLAS publication [15]. Two of them are shown in figure 4.1. In these examples, it can be seen, that the difference between parton and hadron level is rather small. This is not surprising because the objective of the publication was to compare theory and experiment and the used observables were not very sensitive to non-perturbative effects. That the cross section of those levels does not differ very much is better shown in a 2D histogram (figure 4.2). It can be seen, that most of the values are arranged on a diagonal. That means, that there is practically no smearing between different bins, when hadronizing the parton level to create the hadron level. The conclusion is, that the unfolding with the bin-by-bin correction is reliable to calculate the hadron level out of the parton level.
But there are also observables, which are sensitive to non-perturbative effects, e.g. the mass of the single jets. When comparing the cross section of parton and hadron level for the mass of a single jet it can be seen in figure 4.3, that the differential cross section differs especially for small masses. The difference for small masses can be explained with the fact, that the particles are not hadronized yet, when the parton level is simulated. When the hadronization is finished the quarks are not free anymore, they are bound in hadrons. This leads to an important difference for jets. A jet consisting of only quarks is nearly collinear, while a jet made out of hadrons is not. This is important for the mass, because for example the mass of a jet made out of two particles can be determined through the following equation.

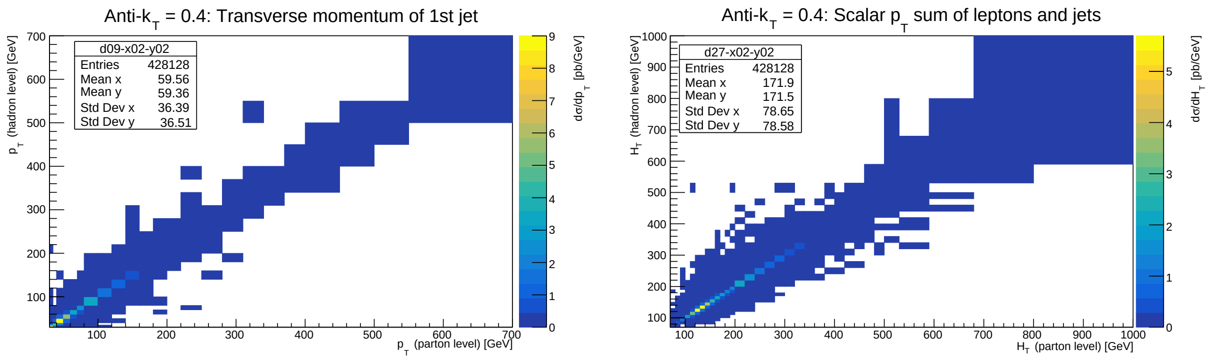
$$\begin{aligned} m^2 &= (p_1 + p_2)^2 = p_1^2 + p_2^2 + 2p_1^\mu p_{2,\mu} \\ &= E_1^2 - |\vec{p_1}|^2 + E_2^2 - |\vec{p_2}|^2 + 2E_1E_2 - 2|\vec{p_1}||\vec{p_2}| \cdot cos\Theta \end{aligned} \tag{4.1}$$

As it can be seen, the mass of the jet is smaller, if the angle between the particles Θ decreases. So if the jet of quarks is nearly collinear, its mass must be smaller than for one made out of hadrons.
This all means, that it is more unlikely to find a light jet on the hadron level resulting in a smaller cross section. This effect is better visible in the right histogram (4.3b), which uses a different distance parameter to recombine the jets, but this effect will be discussed in section 4.2. So, it can be assumed, that this observable is a lot more sensitive to non-perturbative effects than the ones before and this is also shown in the 2D histograms in figure 4.4. By observing the data in the 2D matrices it is visible, that they are partially arranged off-diagonal, what means, that there is a smearing between different bins and because of that it can be deduced, that the bin-by-bin correction should not be used to calculate the hadron level from a simulated parton level.



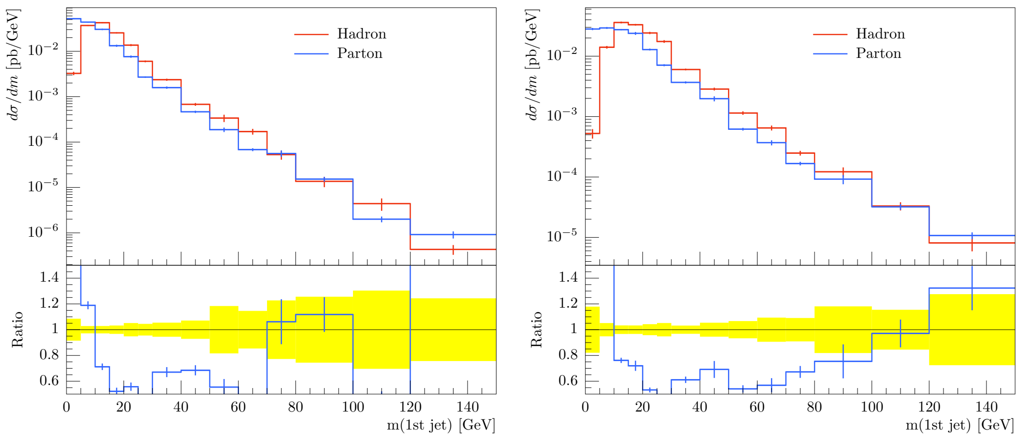
(a) **(b)**

Figure 4.1: Two examples from the used analysis: (a) Cross section over the transverse momentum of the leading jet and (b) Cross section over the scalar sum of p_T of leptons and jets



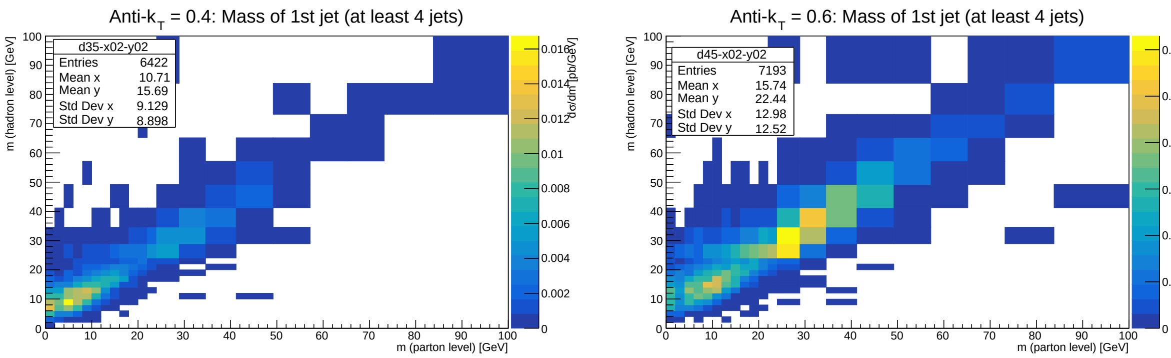
(a) **(b)**

Figure 4.2: 2D histograms of the example observables at hadron and parton level: (a) Cross section over the transverse momentum of the leading jet (b) cross section over the scalar sum of p_T of leptons and jets



(a) **(b)**

Figure 4.3: Cross section over the mass of the leading jet with at least 4 jets for a distance parameter of (a) 0.4 and (b) 0.6



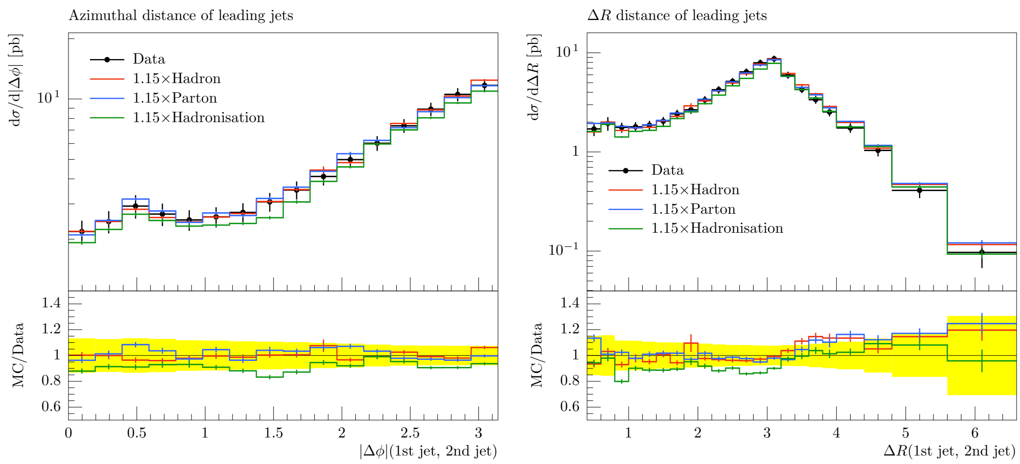
(a) **(b)**

Figure 4.4: 2D histograms for the cross section over the jet mass for the leading jet with at least 4 jets and a distance parameter of (a) 0.4 and (b) 0.6

4.2 The influence of the different hadron level calculations

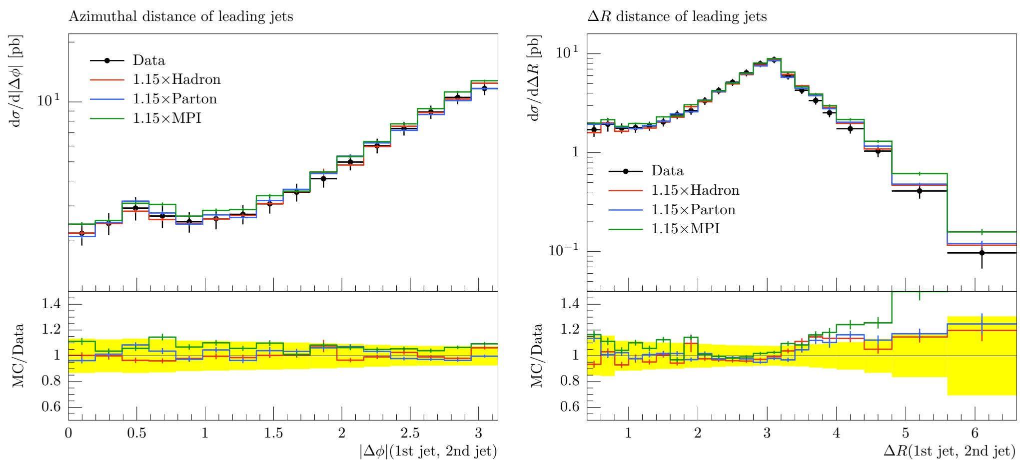
As it was explained in section 3.2 the simulation of the hadron level requires the calculation of more effects, such as the hadronization of the particles, the enabling of multiple parton interactions and the possibility for the beam to have an intrinsic transverse momentum. So in this section, the influences of the different hadron level effects are analysed and with that the produced non-perturbative effects. In a first step the effect of the hadronization will be tested. That means, that still no multiple parton interactions are allowed and that the intrinsic transverse momentum of the beams must be zero. The assumption is, that the hadronization could have a great effect on the parton level, because the quarks are then bound into mesons. That means, that the masses of the jets increase like it was explained in section 4.1, which leads to different results for jet observables. As it can be seen in figure 4.5 the hadron level simulation really differs from the parton level simulation. It is also visible, that the cross section for the hadronization is even lower than the cross section for the hadron level. That means, that the effect of the hadronization alone has a too great effect on the parton level to be enough to simulate the hadron level. This is logical, if it is considered, what was left out. In this simulation, there were no multiple parton interactions allowed, what means, that there was only the hard process of the Z+jets production. But if more processes in the simulation of the event would be allowed besides the hard process, the probability, that the event would be analysed, increases. This is a result of the requirements for the analysis. In this analysis, an event is analysed, if the leptons fulfil certain requirements, but also if at least one jet with specific characteristics exist. So if there are more processes, which are able to produce jets, the probability for fulfilling the requirements increases and with that the cross section as well. If this assumption is correct, the cross section for the simulation with only multiple parton interactions should be higher than hadron and parton level, what can be seen in figure 4.6.

The last difference to the hadron level is, that the colliding partons are allowed to have a transverse momentum, but as it can be seen in figure 4.7 this characteristic does not have a great effect on the simulation results. Furthermore the generated effect on the parton level is not as predictable as it is for the other two. Sometimes the cross section increases and sometimes decreases by a small amount. So the conclusion is, that the combination of the hadronization and the multiple parton interaction could be already very close to the simulated hadron level and as it can be seen in figure 4.8 this is



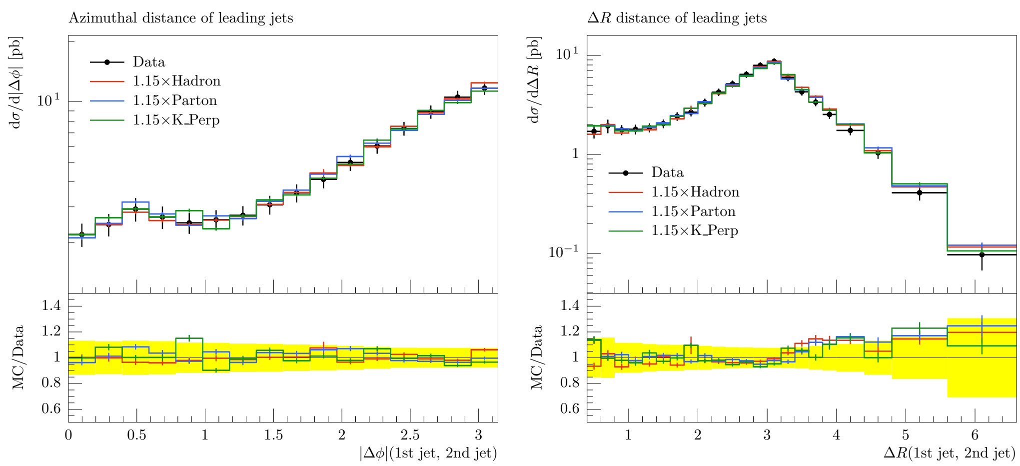
(a) **(b)**

Figure 4.5: Histograms for the parton and hadron level together with only the hadronisation for the observables (a) azimuthal distance of leading jets (b) ΔR distance of leading jets



(a) **(b)**

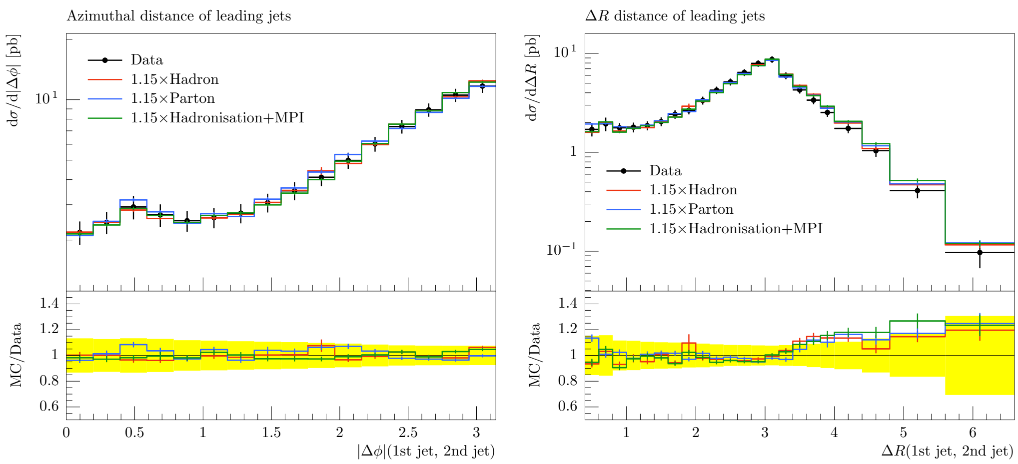
Figure 4.6: Histograms for the parton and hadron level together with only the multiple parton interactions for the observables (a) azimuthal distance of leading jets (b) ΔR distance of leading jets



(a) **(b)**

Figure 4.7: Histograms for the parton and hadron level together with only the intrinsic transverse momentum of partons for the observables (a) azimuthal distance of leading jets (b) ΔR distance of leading jets

the case. The differences between both simulations are very small and always in the error limit. That means, both effects are the main reasons for the differences between hadron and parton level. So the conclusion is, that the non-perturbative effects are mostly a product of the hadronization and the multiple parton interaction, what means, that these two effects are the most important to consider, when analyzing non-perturbative effects.



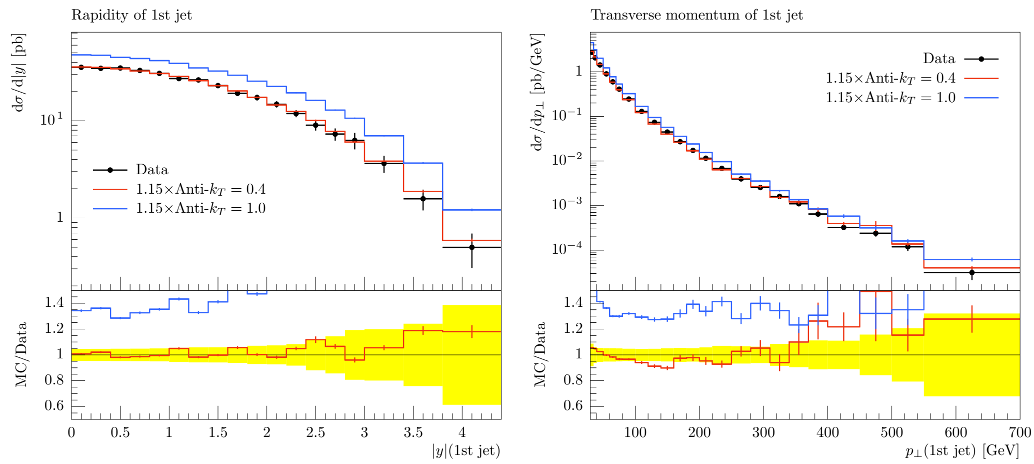
(a) **(b)**

Figure 4.8: Histograms for the parton and hadron level together with the hadronisation and multiple parton interaction for the observables (a) azimuthal distance of leading jets (b) ΔR distance of leading jets

4.3 The influence of the distance parameter R

As it was mentioned earlier the jet clustering algorithm can be used with different parameters. One of them is the distance parameter R. This section will analyse, what influence the distance parameter has on the results of the simulation. First the differences between two hadron level simulations will be analysed, which used different distance parameters for jet clustering. After that it will be researched, how the the non-perturbative effects change, when using another distance parameter.

To analyse the influence of the distance parameter, observables for jets are examined. As it can be seen in figure 4.9, a greater distance parameter leads always to a higher cross section, when also considering the possible errors. This effect can also be seen in most of the other observables too, which can be found in the appendix (B.1). The reason for this behaviour lies in the analysis. Like it was explained in section 3.3, not all events are taken into account. Only those events, which include the production of Z boson and at least one jet are analysed. Furthermore the jets also need to have a p_T of at least 30 GeV to be counted. Jets with a lower p_T are ignored. So if there are only jets with a p_T less than 30 GeV in an event, it will be ignored completely. However as it was shown in section 2.2 a greater distance parameter means, that more particles are combined to one jet. So now the question is, which particles are combined. Normally it is very unlikely, that particles with opposite p_T are combined, because if two particles have opposite p_T, the distance to the beam is in general less than the distance between the particles, which results in two different jets and not one. So in most of the cases, adding more particles to a jet leads to an increasing of the transverse momentum of that jet. The result is then, that more jets fulfil the required p_T of at least 30 GeV. That means,



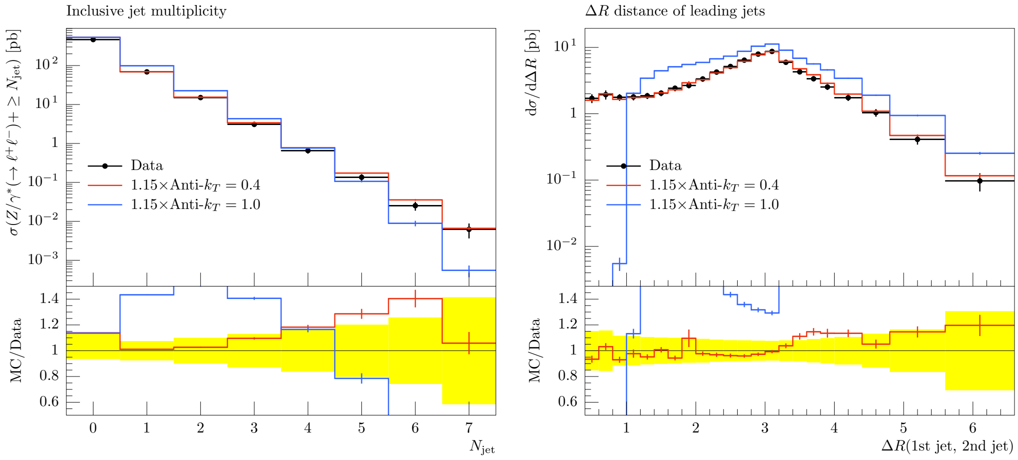
(a) **(b)**

Figure 4.9: Histograms for jet observables with the cross section over (a) the rapidity of the leading jet and (b) the transverse momentum of the leading jet

that a greater number of events is analysed leading to a higher cross section for a greater distance parameter.

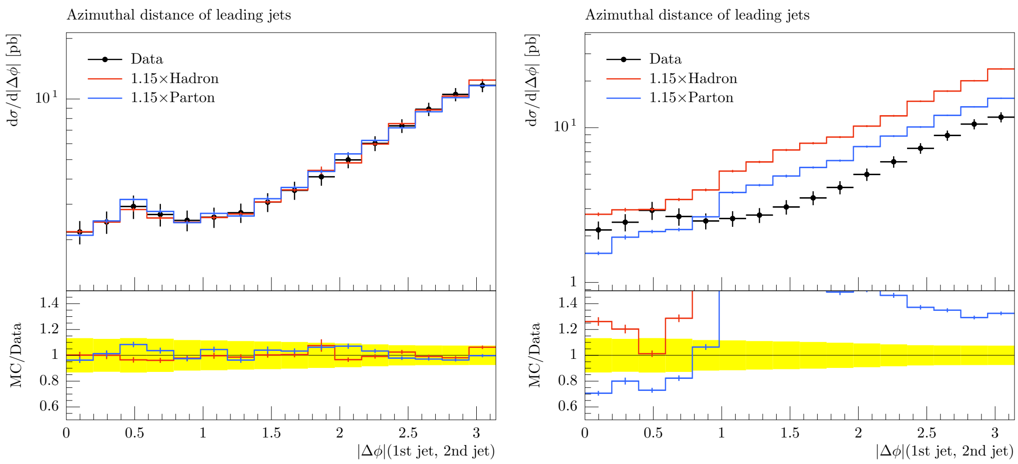
However sometimes it can be seen, that the cross section for the greater distance parameter is smaller, like in figure 4.10. But those exceptions can normally be explained quickly. In figure 4.10a for example the cross section for the number of found jets is smaller for four or more jets. The explanation is again the number of combined particles into a jet for different distance parameters. When more particles are combined to one jet and there is a finite number of particles, less jets are produced for a greater distance parameter, which leads to a smaller cross section. Figure 4.10b however can directly be explained with the distance parameter itself. The observable in this histogram is the ΔR difference, which is calculated like the geometrical scale between two entities (equation 2.3). That means, when the angular distance between two jets has to be at least 1.0, it is impossible for them to have a smaller ΔR distance than 1.0, because the calculation is the same. This is, why the cross section immediately decreases for ΔR smaller than 1.0.

Now it will be analysed, how the distance parameter influences the differences between parton and hadron level and thereby examining the influence on non-perturbative effects. As it was shown in section 4.1, the differences between hadron and parton level for the observables of the ATLAS publication [15] were very small. So the non-perturbative effects did not have a great impact on the simulation of the hadron level. This drastically changes, if the observables are analysed with a higher distance parameter, as it can be seen in figure 4.11. First of all, it can be noticed, that the simulation does not fit the measured data anymore, because in the analysis of the measuring, it was also used a distance parameter of 0.4 for reconstructing the jets. But more importantly the differences between hadron and parton level are considerably higher. That means, that the usage of a greater distance parameter increases the sensitivity to non-perturbative effects. This is better visible for observables, which were already sensitive to these effects before, like the masses of the jets. In figure 4.12 the differences between hadron and parton level are much greater for a distance parameter of 1.0 than for 0.4, especially for small masses. Furthermore the migration from one bin to another, when applying the effects for the hadron level can be seen very clearly in figure 4.12b. For a comparison, it can also be seen in the appendix (B.2), that the differences for hadron and parton level for a distance



(a) (b)

Figure 4.10: Histograms with higher cross section for R = 0.4: (a) Cross section over the number of found jets (b) Cross section over ΔR distance of leading jets

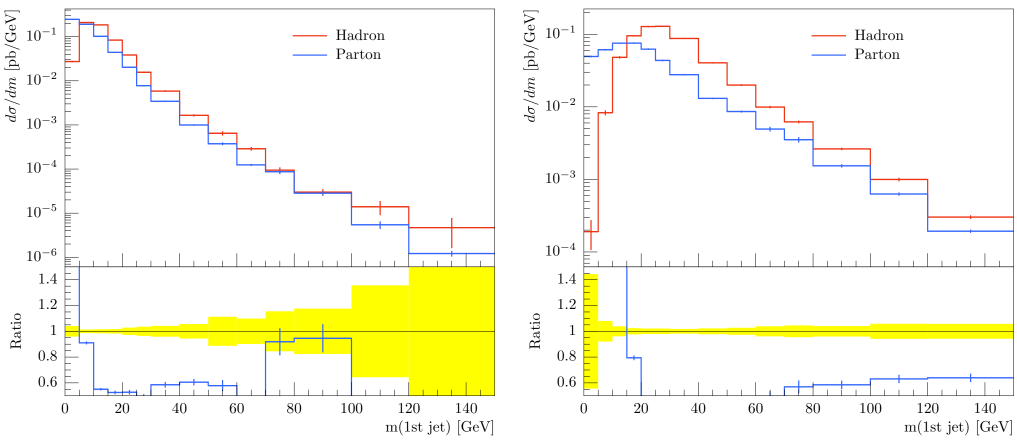


(a) (b)

Figure 4.11: Cross section over the azimuthal distance of leading jets for (a) $R = 0.4$ (b) $R = 1.0$

parameter of 0.6 are greater than for 0.4, but smaller than for 1.0. This leads to the conclusion, that observables are more important to non-perturbative-effects, if a greater distance parameter is used. To find the reason for this difference, we have to observe figure 4.11 closer. It can be noticed for a distance parameter of 1.0, that the cross section of the hadron level is higher than the cross section for the parton level for nearly every observable[1]. That means, that a greater distance parameter has a stronger influence on the hadron level. The only exception is the cross section for small masses, but that difference was already explained in section 4.1. In section 4.2 however, it was shown, that the enabling of the multiple parton interactions results in an increase of the cross section. So the conclusion is, that the distance parameter has a greater effect on the multiple parton interaction, than on the hadronization.

The reason for this lies in the simulation parameters. As it was already mentioned, the enabling of the multiple parton interaction allows the simulation of processes with lower energy besides the hard process. Furthermore it was explained in section 3.2, that only a limited number of particles is simulated through the perturbation theory and the rest with the parton shower method. So if the overall number of particles is increased, but not the number, which is calculated with the perturbation theory, more particles are simulated through the parton shower and with that the non-perturbative effects increase. This effect was already seen for $R = 0.4$ in section 4.2, but it did not have that great impact, because the particles in low energy processes have only a small p_T. So even if some of these particles were combined to jets, most of these jets did not have the required p_T of at least 30 GeV to be counted. However if the distance parameter is increased, more of these low energy particles can be combined to one jet and so the probability for a jet with a p_T of at least 30 GeV increases as well. This results in more analysed events and with that the cross section for the hadron level increases in comparison to the parton level.



(a) (b)

Figure 4.12: Cross section over the mass of the first jet with at least three jets for (a) $R = 0.4$ (b) $R = 1.0$

[1] For more observables, see the appendix (B.4).

4.4 Unfolding

4.4.1 Comparing bin-by-bin and Bayes

In this section, two unfolding methods are used to calculate a hadron level distribution. The first one is the bin-by-bin method and the second one is the iterative method, which uses Bayes' theorem. After that the two methods are compared in terms of equality to the hadron level. The aim is to analyse, if one of the unfolding methods can be used to calculate hadron level simulations out of parton level simulations and with that determining the non-perturbative corrections. There were different possibilities to compare both methods. The easiest way would be to use different simulations for the creation of the response matrix object and the one to unfold. But another intention of this thesis is to investigate, if the unfolding can be used for other parton generators as well and not just SHERPA. So it would be useful to have a simulation, which is not similar to the one used to create the response matrix object. To create such a distribution, a small mistake was built into the simulation and analysis. First the events were generated by SHERPA [11] as weighted events. That means some events will be of more importance to the result than others. However in the analysis it will be assumed, that all events have the same weight, what will then create incorrect distributions and because of that its form is changed. The assumption is, that the iterative unfolding, using Bayes' theorem, should still be applicable to generate a similar hadron level out of the parton level, but the bin-by-bin method will be problematic, because the bin-by-bin factors will now be different than when the response matrix object was created.

But before the unfolding of the incorrect distribution can begin, an important question has to be cleared. It has to be checked, if the response matrix object, which was created, can still be applied to the new incorrect distribution. In order to test this, the response matrices have to be compared. For that the bin entries of both matrices need to be normed and after that divided. If the obtained 2D matrix is nearly uniform, the effects of the integrated mistake are small enough to use the same response matrix object. This is shown for different observables in figure 4.13. It can be seen, that the 2D matrices are not completely uniform. It is sometimes even the case, that the differences in single bins are relatively great like in figure 4.13c, but the overall matrix is nearly uniform and so the response matrix object can be used for the newly obtained distribution.

After this requirement is now fulfilled, the unfolding of the weighted simulation can start. First the behaviour of observables, which had a nearly diagonal response matrix (see section 4.1), shall be examined. The assumption is, that the bin-by-bin method should be able to create a very similar hadron level for these observables. That this is actually the case can be seen in figure 4.14. However in figure 4.14b it is already visible, that the bin-by-bin method does not always provide similar results for such observables. This is better shown for the observables of the rapidity distance and the ΔR distance of the leading jets. For those observables there are more off-diagonal elements, but by far the most lie on the diagonal of the response matrix (see appendix B.3). When unfolding the parton level of these observables with the bin-by-bin method, the result differs more from the expected hadron level. In figure 4.15 it can be observed, that the bin-by-bin method collapses completely for some bins in the histogram.

This effect seems to be even stronger for observables, which have a non-diagonal response matrix, like the masses of the jets. This can be seen in figure 4.16. However one should also mention, that even

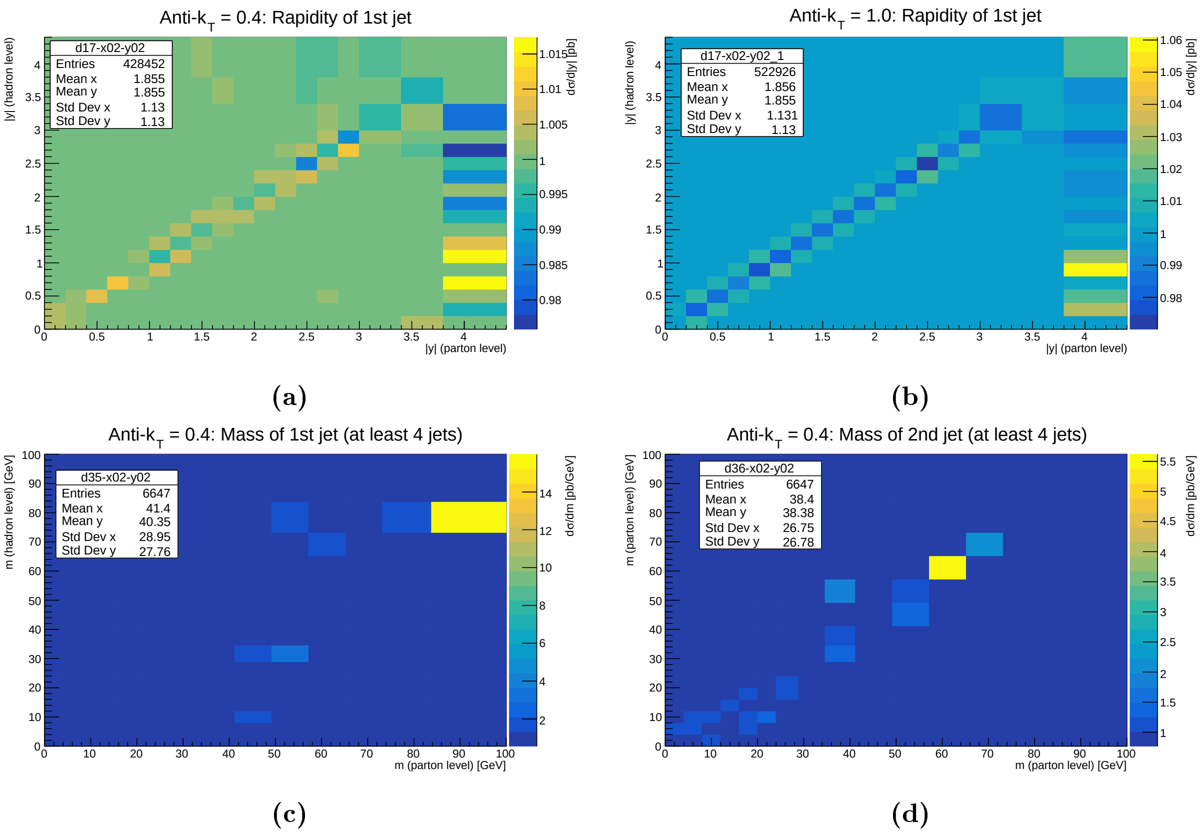


Figure 4.13: 2D ratio histograms of both simulations for different observables (a) Rapidity of first jet for distance parameter 0.4 (b) Rapidity of first jet for distance parameter 1.0 (c) Mass of first jet for at least 4 jets with a distance parameter of 0.4 (d) Mass of second jet for at least 4 jets with a distance parameter 0.4

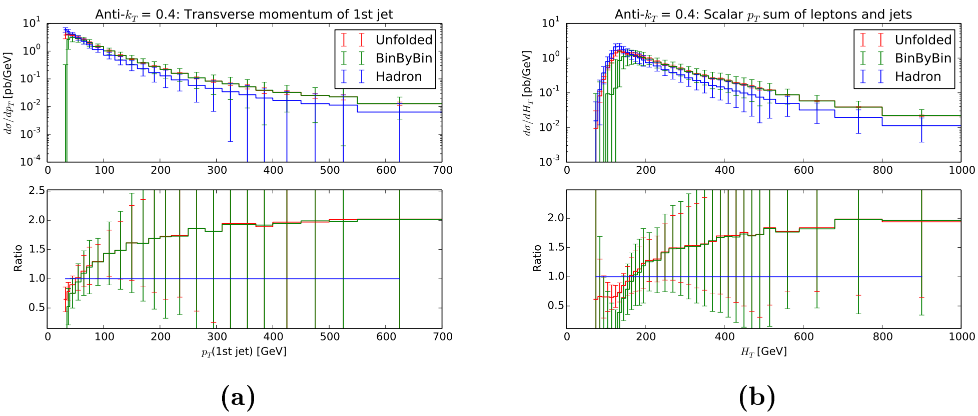
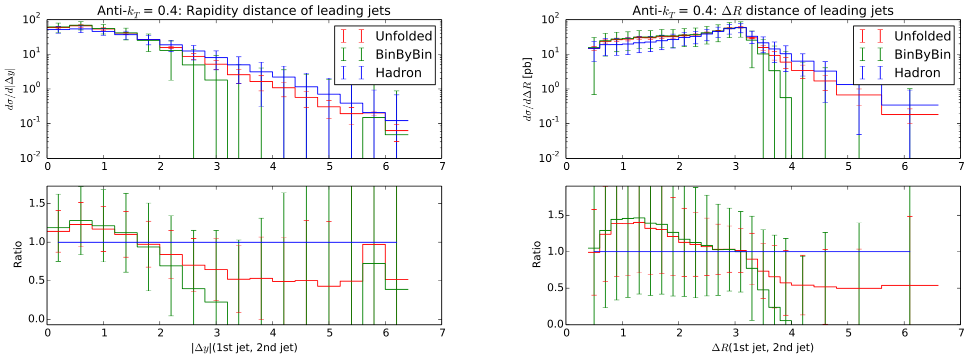


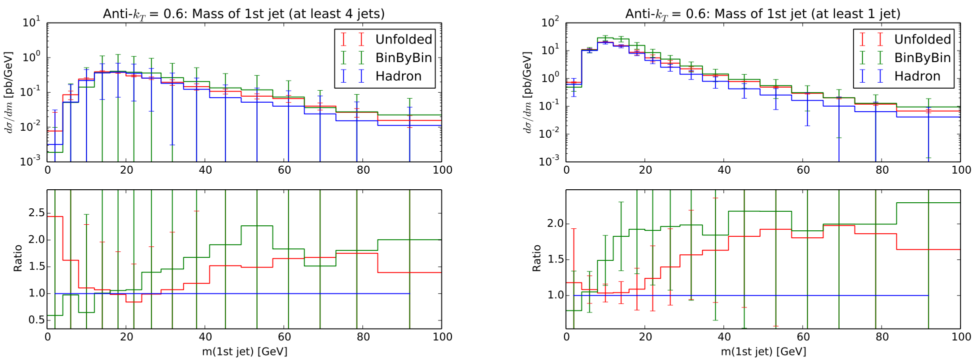
Figure 4.14: Comparing unfolded parton level by both methods with the hadron level for the observables (a) Transverse momentum of first jet and (b) scalar p_T sum of leptons and jets



(a) **(b)**

Figure 4.15: Comparing unfolded parton level by both methods with the hadron level for the observables (a) rapidity distance and (b) ΔR distance of leading jets

the iterative unfolding is not always applicable to produce the expected result for these observables, especially for small masses. This is visible in the first bins of figure 4.16a. A reason for that is the fact, that the number of events in the first bins is very low, which results in greater errors and so a greater difference in the outcome. For a better unfolding result, it is therefore important to have enough entries in each bin. This can be achieved either with more simulated events or wider bins.

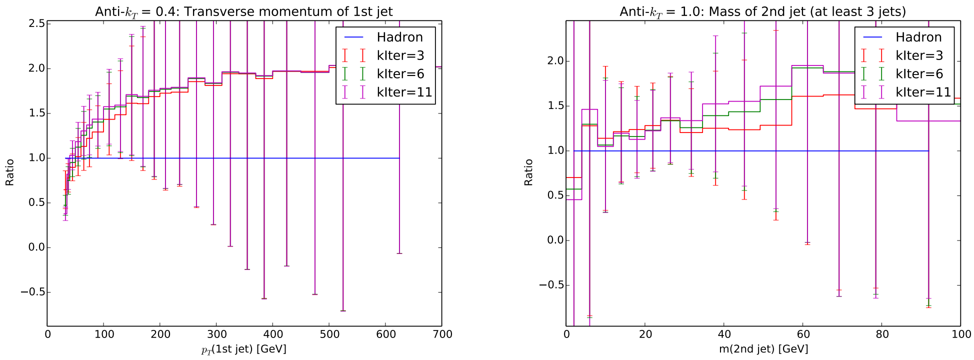


(a) **(b)**

Figure 4.16: Comparing unfolded parton level by both methods with the hadron level for the observables (a) mass of first jet with at least 4 jets for a distance parameter 1.0 (b) mass of the first jet with at least one jet for a distance parameter 0.6

4.4.2 Comparing different iteration numbers

As it was explained in section 2.4 to unfold a distribution the algorithm has to be executed more often than just once. In this section it shall be examined, what influence the number of iterations has on the result and if it makes sense for our purposes to iterate more often, because every iteration needs computing time and capacities. For this comparison the parton level was iterated three, six and eleven times and the differences to the hadron level were plotted in one histogram for each observable. Two examples are shown in figure 4.17, one for an observable, which has a nearly diagonal response matrix (figure 4.17a) and one for an observable with an off-diagonal response matrix (figure 4.17b). Interestingly one can see, that the deviation to the hadron level seems to increase, when



(a) **(b)**

Figure 4.17: Unfolded parton levels with different numbers of iteration in comparison to the hadron level for the observables (a) transverse momentum of the leading jet for a distance parameter 0.4 (b) mass of the first with at least 3 jets for a distance parameter of 1.0

more iterations are made, which is the opposite of the normal assumption[2]. However this behaviour was already shown in the article of D'Agostini [10]. He mentioned, that for an infinite number of iterations, the result starts to fluctuate because of the fact, that every bin in the true distribution is a degree of freedom and this leads to an unfolding matrix, which is the inverse of the smearing matrix. That means, that the iterative unfolding goes over to a method which uses a normal inverse matrix. But this method is not a very good approach, because it can happen, that the inverse of the smearing matrix does not exist and even if it exits, it produces a number of errors, when the uncertainties of the initial distribution are high. Apparently in this case this behaviour appears already for six iterations, which is nevertheless surprising.

[2] This is also the case for other observables, some of them can be observed in the appendix (B.5).

5 Summary and outlook

This thesis has shown, that non-perturbative effects in the simulation of the hadron level are very important to consider. It is theoretically possible to examine observables, which are not very sensitive to these effects, but already the variation of the used jet clustering algorithm can increase this sensitivity. So it is not trivial to create a hadron level distribution out of a parton level distribution, but the iterative unfolding method has proven beneficial for this task, especially in comparison with the bin-by-bin method. However it is still important to consider, that even the iterative unfolding method has its limitations. If the number of events in a bin is too small, the errors of this method can increase drastically, so it is important to always make sure, that either enough events were simulated or the bins are wide enough. Furthermore it was seen, that it is not always better to increase the number of iterations for the iterative unfolding. It is important to test the effects before using the method to find the optimal number of iterations.

It is now possible to further investigate this topic. First of all it could be interesting to examine more observables, which are sensitive to non-perturbative effects. It is also possible to use the unfolding method on processes other than the Z+jets production. If it could be shown, that this method also works for other processes and observables, the iterative unfolding would be verified as a general method for determining non-perturbative corrections.

Furthermore it could be investigated, if the response matrix object created with SHERPA can also be used to unfold a parton level distribution of another event generator (preferably a parton level event generator like BlackHat [16]). This was simulated in section 4.4, but for further proof, this has to be tested. It would be important to validate, that the response matrix object generated during the work of this thesis provides similar results for different parton level generators. The advantage is then, that it is not necessary to generate a response matrix object for every generator.

A Programs

Listing A.1: Settings file ('Run card') for the 1 million events parton level Z+jets Sherpa simulation.

```
(run){
  % general setting
  EVENTS 1M; ERROR 0.99;

  EVENT_SEED_MODE=3;
  EVENT_OUTPUT=HepMC_Short[Parton_events];

  % scales, tags for scale variations
  FSF:=1.; RSF:=1.; QSF:=1.;
  SCALES METS{FSF*MU_F2}{RSF*MU_R2}{QSF*MU_Q2};

  % tags for process setup
  NJET:=2; LJET:=0; QCUT:=20.;

  % me generator settings
  ME_SIGNAL_GENERATOR Comix Amegic LOOPGEN;
  EVENT_GENERATION_MODE=P;
  LOOPGEN:=BlackHat;

  % exclude tau from lepton container
  MASSIVE[15] 1;

  % collider setup
  BEAM_1 2212; BEAM_ENERGY_1 = 3500.;
  BEAM_2 2212; BEAM_ENERGY_2 = 3500.;

  FRAGMENTATION=Off;
  MI_HANDLER=None;
  K_PERP_MEAN_1=0.0;
  K_PERP_MEAN_2=0.0;
  K_PERP_SIGMA_1=0.0;
  K_PERP_SIGMA_2=0.0;

}(run)

(processes){
  Process 93 93 -> 90 90 93{NJET};
  Enhance_Factor 3 {3}
  Enhance_Factor 9 {4}
  Enhance_Factor 27 {5}
  Order (*,2); CKKW sqr(QCUT/E_CMS);
  Integration_Error 0.02 {4};
  Integration_Error 0.02 {5};
```

```
  Integration_Error 0.05 {6};
  Scales LOOSE_METS{FSF*MU_F2}{RSF*MU_R2}{QSF*MU_Q2} {7,8};
  Print_Graphs ./;
  End process;
}(processes)


(selector){
  Mass 11 -11 50 150     #invariante Masse einschraenken
  Mass 13 -13 50 150
  PT 11 10 E_CMS         #transversale Impulse einschraenken
  PT -11 10 E_CMS
  PT 13 10 E_CMS
  PT -13 10 E_CMS
}(selector)
```

B Histograms

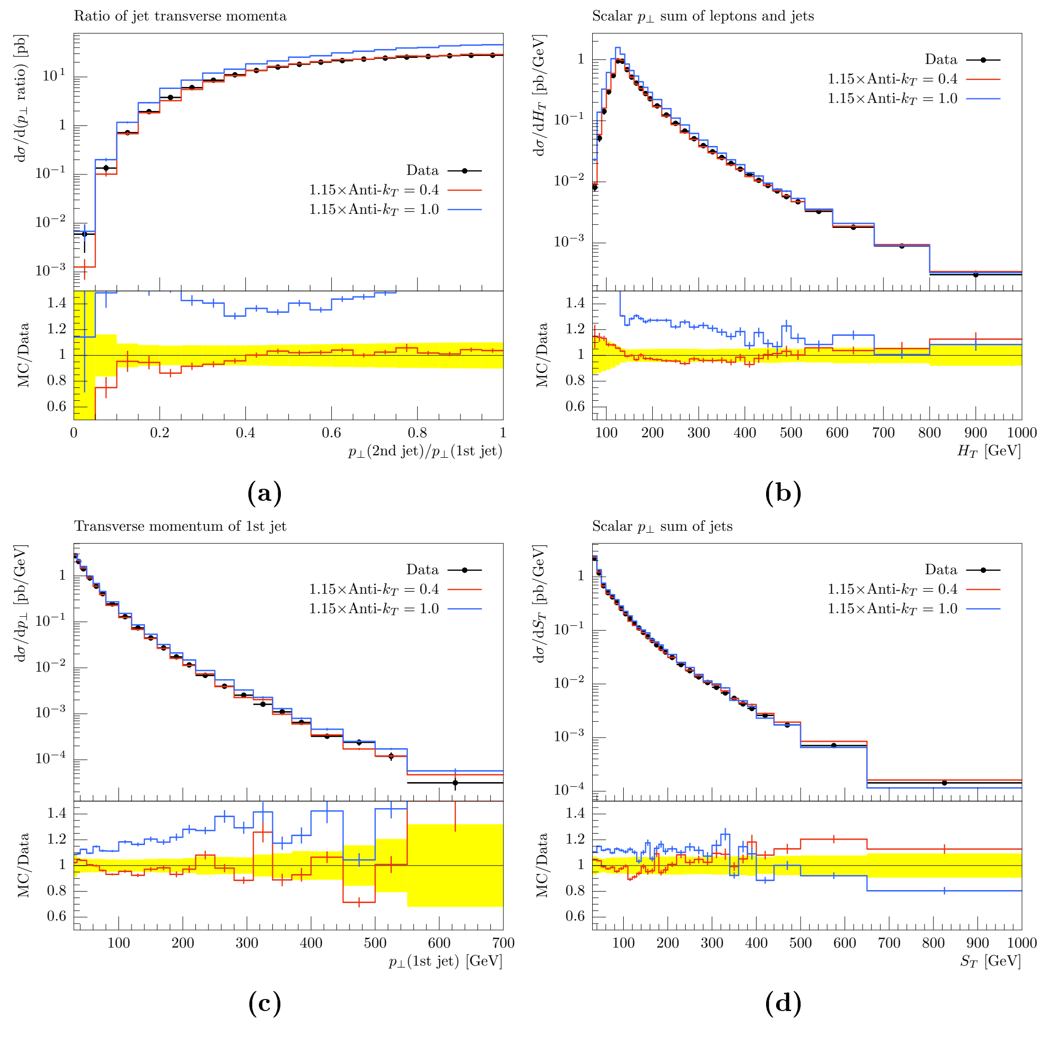
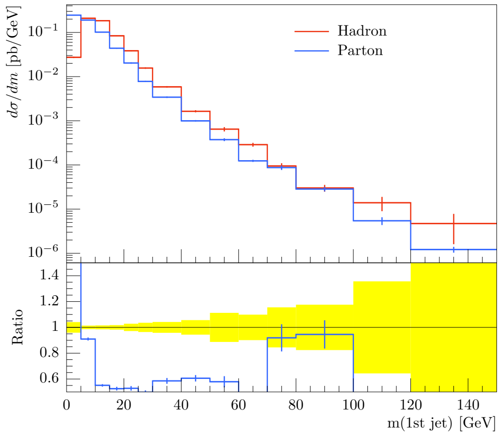
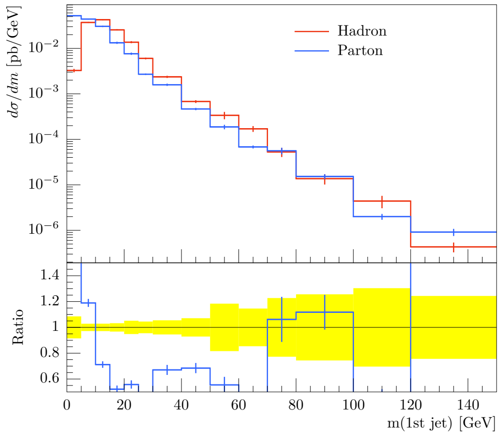


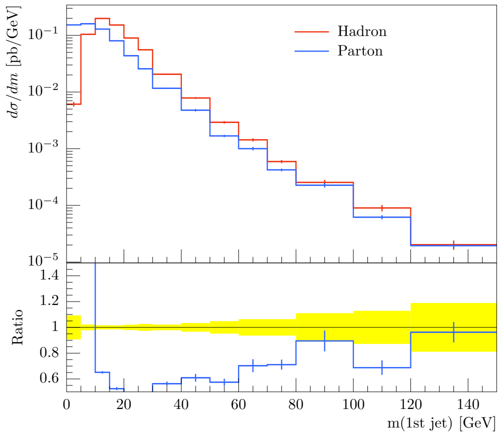
Figure B.1: Histograms on hadron level for jet observables with the cross section over (a) the ratio of the transverse momenta of the leading jets (b) scalar p_T sum of leptons and jets
Histograms on parton level for jet observables with the cross section over (c) transverse momentum of the first jet (d) scalar p_T sum of jets



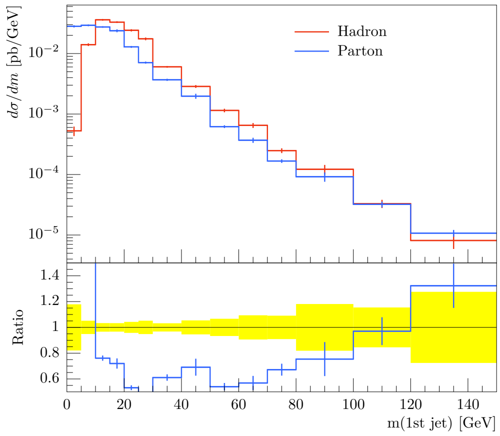
(a)



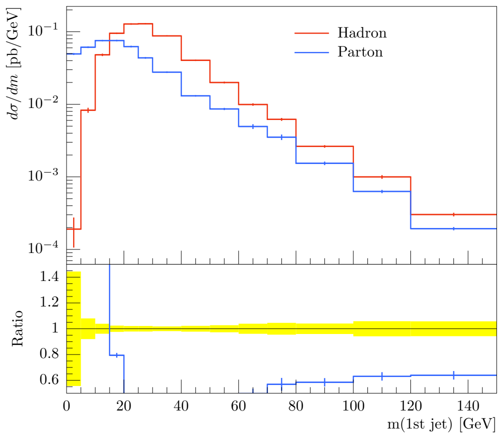
(b)



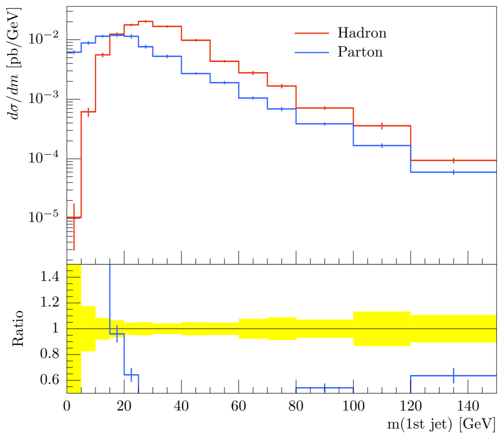
(c)



(d)

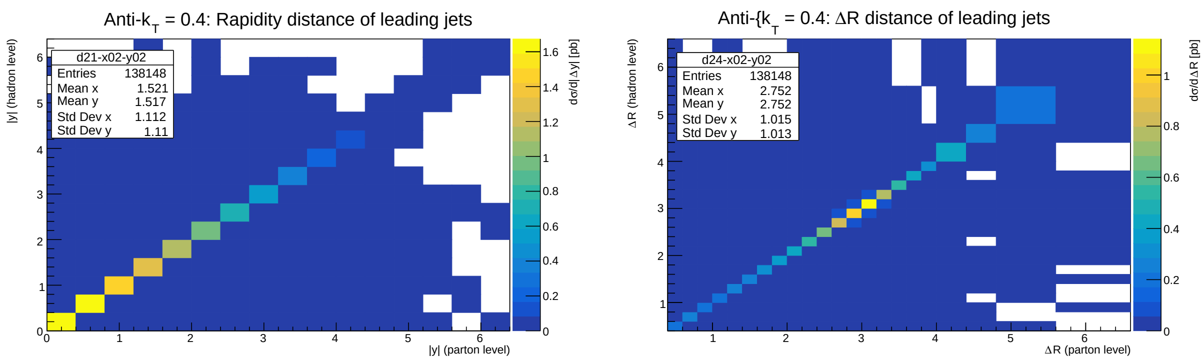


(e)



(f)

Figure B.2: Cross section over the mass of the first jet with at least three jets for (a) $R = 0.4$ (c) $R = 0.6$ (e) $R = 1.0$
Cross section over the mass of the first jet with at least four jets for (b) $R = 0.4$ (d) $R = 0.6$ (f) $R = 1.0$



(a) **(b)**

Figure B.3: 2D histograms for the cross section over (a) the rapidity distance of the leading jets (b) the ΔR distance of the leading jets

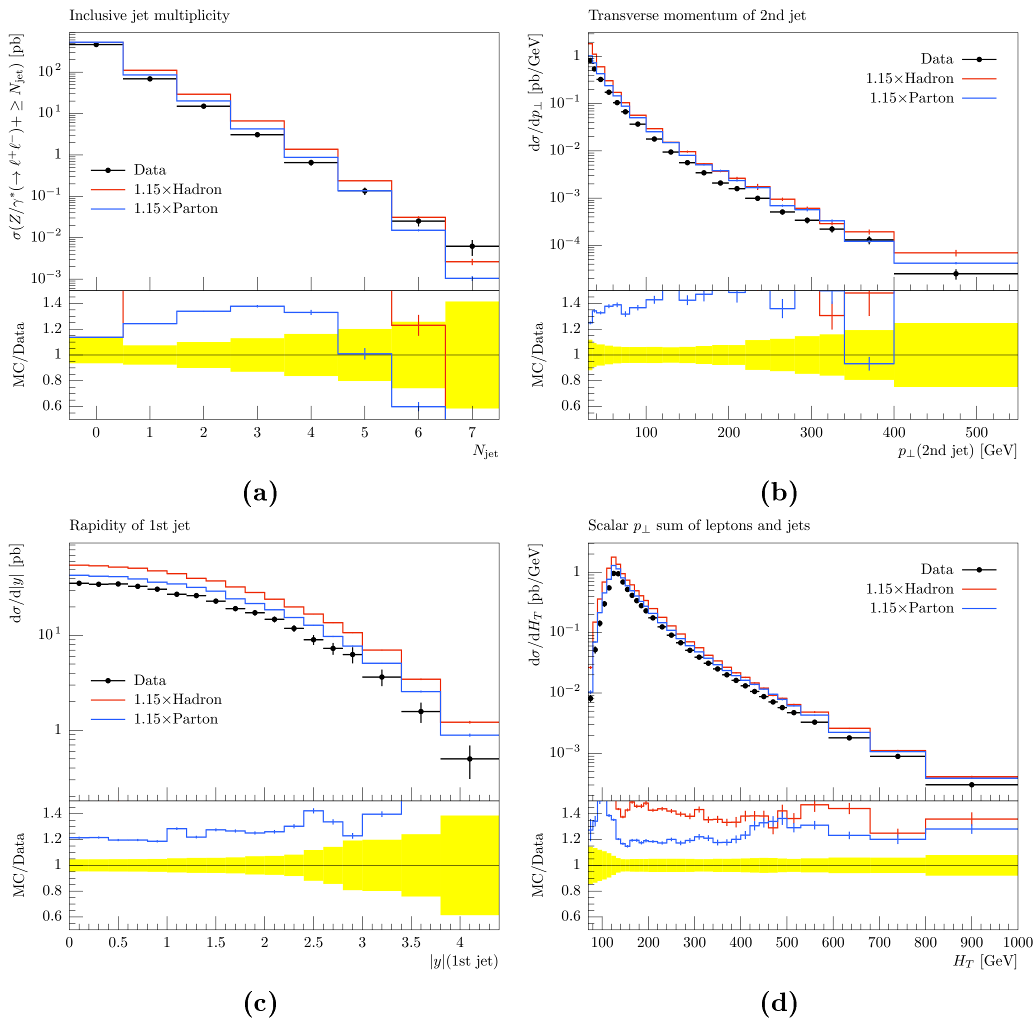
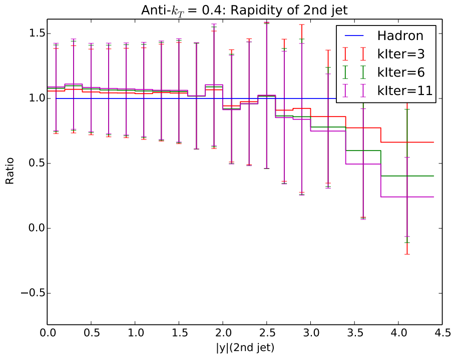
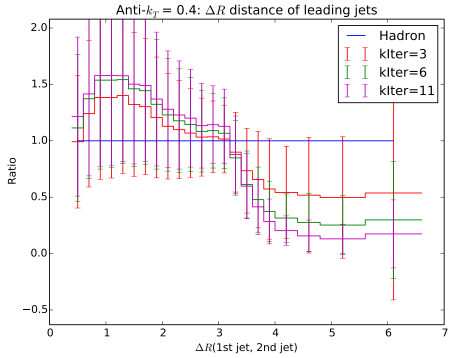


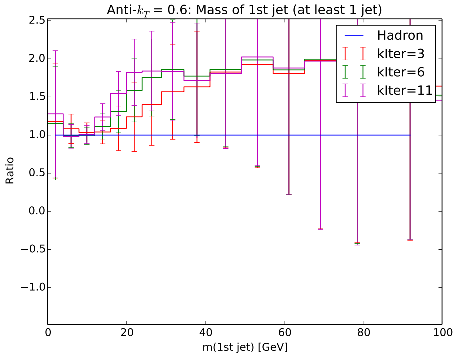
Figure B.4: Cross section for different distance parameters on hadron and parton level over (a) number of found jets (b) transverse momentum of the leading jet (c) rapidity of first jet (d) scalar p_T sum of leptons and jets



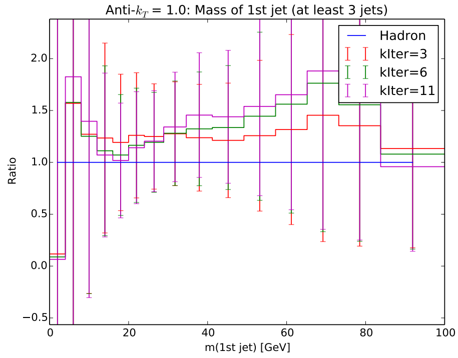
(a)



(b)



(c)



(d)

Figure B.5: Unfolded parton levels with different numbers of iteration in comparison to the hadron level for the observables (a) rapidity of second jet for a distance parameter 0.4 (b) ΔR distance of leading jets for a distance parameter of 0.4 (c) mass of the leading jet for at least 1 jet and a distance parameter of 0.6 (d) mass of the leading jet for at least 3 jets and a distance parameter of 1.0

Bibliography

[1] Stephen P. Martin. A Supersymmetry primer. 1997. [Adv. Ser. Direct. High Energy Phys.18,1(1998)].

[2] T. Gleisberg, Stefan. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.

[3] Frank Siegert. Sherpa performance, jan 2014.

[4] Roel Aaij et al. Observation of the resonant character of the $Z(4430)^-$ state. *Phys. Rev. Lett.*, 112(22):222002, 2014.

[5] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. *JHEP*, 04:063, 2008.

[6] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[7] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[8] M. Wobisch and T. Wengler. Hadronization corrections to jet cross-sections in deep inelastic scattering. In *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*, 1998.

[9] Georgios Choudalakis. Unfolding in ATLAS. 2011.

[10] G.D'Agostini. A multidimensional unfolding method based on Bayes' theorem. *Nuclear Instruments and Methods in Physics Research*, pages 487–498, March 1995.

[11] Jennifer Archibald, Timo Fischer, Tanju Gleisberg, Stefan Hoeche, Hendrik Hoeth, Frank Krauss, Ralf Kuhn, Silvan Kuttimalai, Thomas Laubrich, Radoslaw Matyszkiewicz, Andreas Schaelicke, Marek Schoenherr, Steffen Schumann, Frank Siegert, Jennifer Thompson, Jan Winter, and Korinna Zapp. *Sherpa 2.2.0 Manual*, August 2015.

[12] Andy Buckley, Jonathan Butterworth, Leif Lonnblad, David Grellscheid, Hendrik Hoeth, James Monk, Holger Schulz, and Frank Siegert. Rivet user manual. *Comput. Phys. Commun.*, 184:2803–2819, 2013.

[13] Tim Adye. Unfolding algorithms and tests using RooUnfold. In *Proceedings of the PHYSTAT 2011 Workshop, CERN, Geneva, Switzerland, January 2011, CERN-2011-006, pp 313-318*, pages 313–318, 2011.

[14] K.A. Olive et al. (Particle Data Group). Particle Data Group. *PDG*, May 2016.

[15] Georges Aad et al. Measurement of the production cross section of jets in association with a Z boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector. *JHEP*, 07:032, 2013.

[16] C. F. Berger, Z. Bern, L. J. Dixon, F. Febres Cordero, D. Forde, H. Ita, D. A. Kosower, and D. Maitre. An Automated Implementation of On-Shell Methods for One-Loop Amplitudes. *Phys. Rev.*, D78:036003, 2008.

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am Institut für Kern- und Teilchenphysik ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Paul Moder
Dresden, 14.06.2016