Performance study of future Λ_b^0 production with ALICE3

Author
Maurice Vincent Jongerhuis

Supervisor
Dr. Panos Christakoglou

Second supervisor
Dr. Marta Verweij

July 12, 2022

Abstract

The potential for the performance of the measurement of the Λ_b^0 of the proposed experimental setup ALICE3 during run 5 at the Large Hadron Collider has been studied and is reported in this thesis. The focus was on the optimization of the statistical significance in both proton-proton and Pb-Pb collisions by utilizing selection criteria and machine learning algorithms. This study shows that these measurements are very much feasible in both collision systems reaching values for the significance above 100 in multiple fine bins between $3.0 < p_\mathrm{T} < 10.0$ GeV/c. The optimal analysis strategy has been achieved by the inclusion of particle identification with the time-of-flight and ring cherenkov detector, and parameters of the daughters and granddaughters of the Λ_b^0 candidates. Further improvements in the low- and high-p_T regions are suggested to cater to the physics performance of the detector.

Contents

1 Introduction

In the most recent century, advancements and innovations in technology opened up the route to modern day subatomic physics. On the experimental front, particle accelerators and colliders have played a major role in verifying and measuring properties of fundamental and composite particles. Together with developments of quantum field theory (QFT), the Standard Model (SM) of particle physics has been created. This model attempts to describe the behaviour of a set fundamental particles by means of excitations of quantum fields. This set makes the distinction between 2 types of particles. One group contains the elementary bosons, the force carriers, that are used as intermediate fields that transfer information between particles and hence give rise to forces. The other group are the elementary fermions, the particles on which the forces act. The latter contains quarks and leptons, which are the ingredients for the atoms we are familiar with in our daily lives. For example, three quarks bound together with gluons can form a proton.

Although we only see them bound as hadrons, when meeting certain criteria, quarks and gluons undergo a phase transition to a state of matter called a quark gluon plasma (QGP), where they are no longer bound by color charge. This was initially considered to be very similar to the plasma made of ions and electrons, where nuclei and electrons are no longer bound. One consequence of this state is that quarks can freely move on their own at distances larger than hadronic radii, as opposed to the three quarks locked together in a proton. As one may have expected, this state only occurs at very high temperatures, higher than the mass of light hadrons. At the very early stages of the universe, all matter was so densely packed that everything was a quark gluon plasma. As the universe expanded, its pressure and temperature dropped and the QGP 'freezes' into the bound hadron states. This hadronization of the QGP is a complex process with many open questions and consequences that could potentially explain the structure and constituents of the universe today. As beautiful as the early universe is, it is not very practical as an experimental stage.

The standard model has been used and tested in many cases. Although it can describe many different areas of particle physics, it also has its difficulties. The strong force is highly nonlinear and has a very strong coupling constant, due to which it is very difficult to find solutions at low energies. A tool that is used to still solve the equations is lattice quantum chromodynamics (lQCD). The idea is that quarks are placed on a lattice with an intermediate spacing, which induces a lower limit on the momentum and therefore avoid the complexity at low energies. By reducing the spacing, a limit can be found which is an approximation of the spacetime continuum. This way, lQCD has successfully predicted experimental data before[1]. One of these predictions is the formation of a quark-gluon plasma at temperatures above the Hagedorn temperature ($150 MeV$ or $1.7 \cdot 10^{12} K$)[2]. This makes the investigation of the QGP as a means to test and expand the standard model a very interesting field.
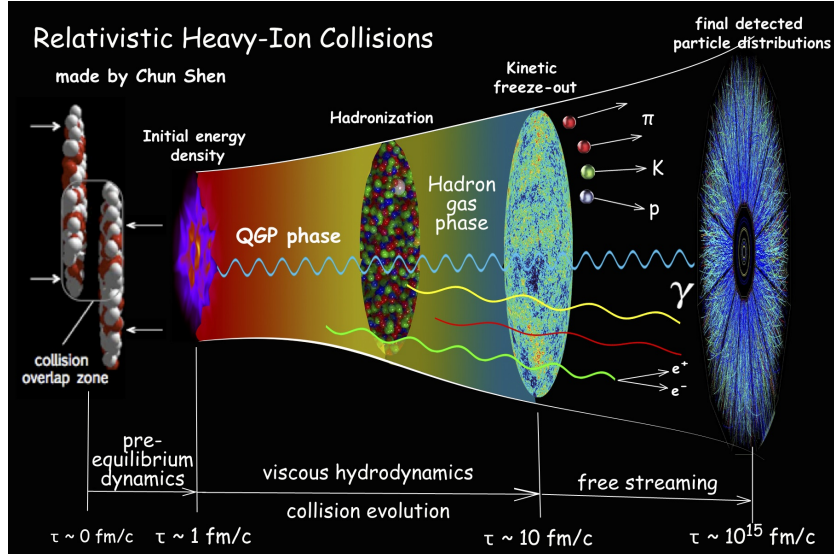
Figure 1: Evolution of a heavy ion collision. Source:[3]

1.1 Physics motivation

In order to obtain exotic states of hadrons and a QGP, a very high energy density is required. This can be acquired by accelerating particles and colliding them in a controlled manner. An example of this is the acceleration of lead nuclei in multiple stages to velocities very close to the speed of light. This is exactly the experimental setup of the Large Hadron Collider (LHC), where lead nuclei obtain energies of the order of TeV. Such collisions provide the necessary conditions for the creation of a QGP which, due to the symmetry of the collision, appears as a stationary ball of plasma in the laboratory frame. This plasma then expands and hadronizes and momentarily can be described as a hadronic gas before the kinetic freeze-out, from which point forward they barely interact with other hadrons. A schematic overview of this evolution can be seen in Figure 1.

In particular, quarks from the heavier generations are very rarely seen in the universe because they decay very quickly and require a lot of energy to be produced in the first place, making it very difficult to measure their interactions. As a comparison, the mass of the beauty quark is $4.24GeV$, a factor 30 larger than the Hagedorn temperature[4]. Hence, even during the lifetime of the QGP, this third generation quark can typically not be produced thermally in the medium. As a consequence, the heavy charm and beauty quarks are almost exclusively produced during the very early stages of the event. During the initial collision of the nuclei, the available energy is still high enough to overcome the production threshold[5]. This also means that these exotic quarks are a good probe to study the QGP, since they experience most, if not all, of the stages of the medium throughout their lifetime. Even though they may experience the QGP in its entirety, due to their mass they may not be fully thermalized during the expansion of the medium. Comparisons with statistical hadronization models seem to suggest that the beauty quark is not fully thermalized during the plasma phase[6]. First signs for this effect can be found in measurements described by references [7][8]. However, further measurements on beauty-baryons and -mesons are required to provide concise evidence. If it would be the case, the heavy flavour quarks keep some information from the initial state of the collision,

4

which can give further insight in the coupling strength between the quarks and the medium [5]. Additionally, the process of hadronization, when the plasma cools down and returns to a color confined structure, is not very well understood either. Since it is such a fundamental process, it could shine light on the properties of the medium as well as individual interactions of elementary particles.

One measurement that can be done makes use of the geometric properties of the event. The initial colliding particles will not always line up centre-to-centre, hence, the overlapping region is typically not circular. The amount of overlap is often indicated by the centrality, the distance between the centres of the ions as a percentage of the radius of the ion. An event with a centrality of 0% therefore means the ions collided perfectly centre-to-centre and 90% means they only touched peripherally. The resulting elliptic-shaped QGP, as shown in Figure 2, can actually help to study the hadronization process. The density in the centre of the plasma is

Figure 2: Simplified image of the elliptic shape of the QGP right after the collision

higher than the edges, resulting in an anisotropic pressure gradient. The pressure is higher in the reaction plane[1], which causes partons to receive a momentum boost in-plane as compared to out-of-plane when the system evolves as a bulk [9]. Of course the initial ions are not perfect spheres and include parton configurations described by parton distribution functions (PDF), which in turn leads to fluctuations in the initial conditions of the event. As a result, the plasma can be described as the sum of all higher harmonics[10]. The anisotropic parameters, which from now on will be called flow parameters, can be characterized by the Fourier coefficients
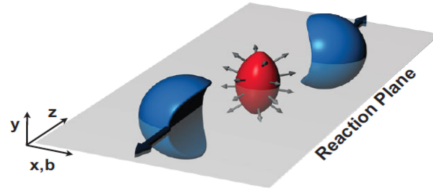
$$v_n = \langle \cos\left[n(\phi - \Psi_n)\right]\rangle \tag{1}$$

Where n is the order of the flow harmonic, ϕ is the azimuthal angle of particles with respect to the reaction plane, and Ψ_n is the angle of the spatial plane of symmetry of harmonic n. Measuring these parameters in both the spatial space and the momentum space raises the ability to investigate properties of the medium like the ratio of the sheer viscosity to the entropy density (η/s) and the thermalization of heavy quarks[11]. Figure 3 shows the measured elliptic flow of several particles during run1 of the ALICE experiment. The low transverse momentum region ($p_T < 2$ GeV/c) can be described very well by hydrodynamical models [12]. The pressure induces a radial flow that pushes partons away from the interaction site, hence increasing their transverse momentum. Consequentially the amount of low p_T particles reduces. This effect has a dependence on the mass of the parton, where the heavier generations experience an increased boost. Combined with the anisotropic gradient this causes an increased depletion of low-p_T particles in-plane based on the mass, reducing the ellipticity of the flow in that p_T range. This effect is clearly observed as shown in Figure 3, where the elliptic flow is ordered solely by the mass of the hadrons.

On the other end of the spectrum, the $p_T > 8$ GeV/c region, different processes are dominant. As the momentum component perpendicular to the beam axis is large, the path of motion will tend to pass through more of the plasma. As these particles traverse they may interact with the medium resulting in an energy loss. This effect is illustrated in Figure 4 and shows a comparison with proton-proton collision. This image indicates two different relevant measurements that can

---

[1]The reaction plane is defined by the plane spanned by the beam axis and the vector pointing from the centre of one nuclei to that of the other at the moment of collision.
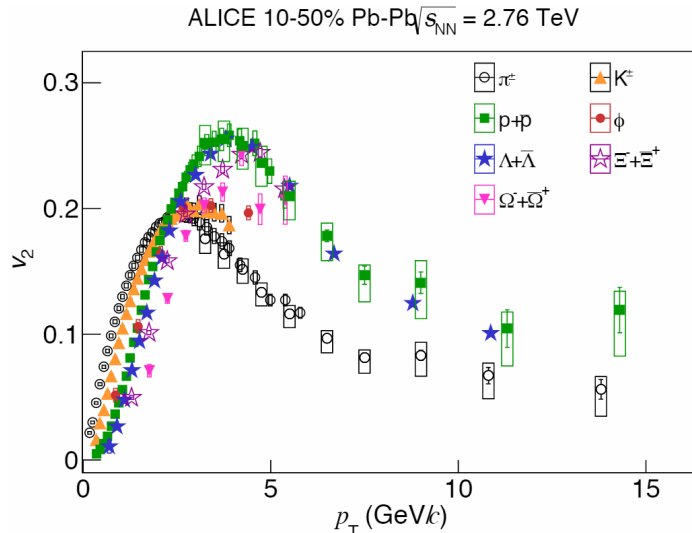
Figure 3: The measurements of v_2 for several particles as a function of transverse momentum in the centrality bin 10-50% [9].

be done that may indicate the presence of the QGP. The first being that when these high-p_T particles (jets) are created, they are oriented back-to-back, with similar momentum pointing in opposite directions. If their origin is not in the centre of the QGP, it is expected that one of these particles travels through the medium for a longer distance as compared to their twin on the other end. The former will therefore have more interactions and lose more energy, a process also known as jet quenching. One could measure an asymmetry of the high-p_T particles to observe this effect.

The second measurement that can be done is a comparison with the momentum distribution of particles in proton-proton collisions. As there is no, or very little, creation of a QGP, the jets keep more of their original energy as compared to the heavy-ion case. The ratio between the yield in any collision with ion A and the equivalent yield in proton-proton collisions is called the nuclear modification factor R_{AA} and has been measured by, for instance, CMS [13] and is shown in Figure 5. In all centrality intervals the factor is below unity, indicating that the high-p_T particles are suppressed by the medium. Additionally, in the more central collisions, the suppression is stronger, which coincides with the idea that there is more plasma to transverse through.

The effects on the momentum of these particles is only based on the electric- and color-charge, and not the mass of the partons. Hence in the high-p_T region, the elliptic flow can be described by the path-length dependence resulting in no significant dependence on the particle species[13, 14].
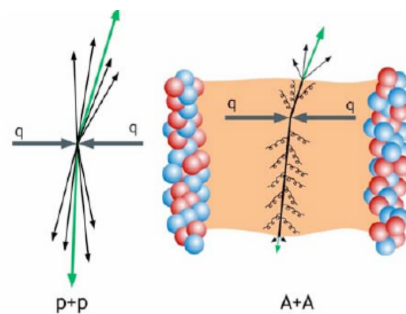

Figure 4: Simplified illustration of high-p_T particles in proton-proton (left) and heavy ion (right) collisions, where the orange blob represents the created QGP.
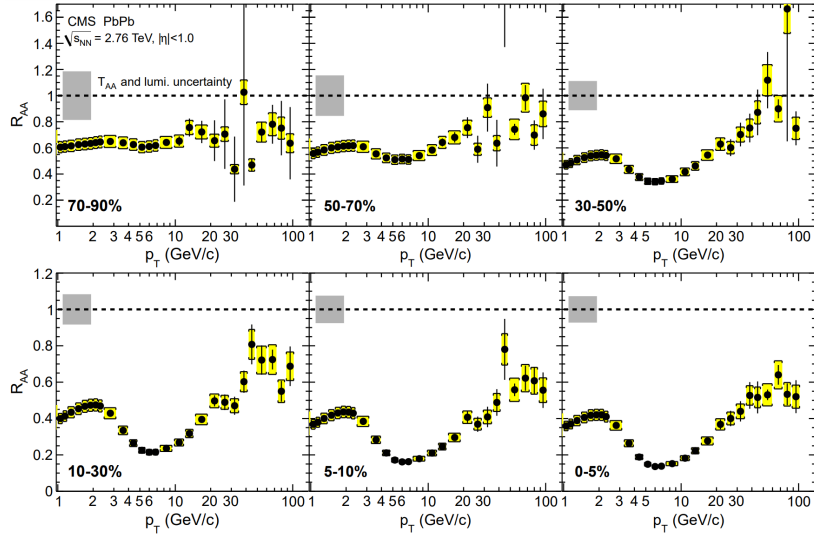
Figure 5: Nuclear modification factor R_{AA} as a function of transverse momentum in different centrality regions, measured by CMS [13].

The first experimental setup of ALICE was used for the first two runs of the LHC and collected data between 2009 and 2018. The Λ_c^+ has been measured in p-p, p-Pb and Pb-Pb collisions from which a nuclear modification factor could be determined. The result, as shown in Figure 6, is very limited and was only determined in a single coarse bin[15]. The limiting factors being the reconstruction efficiency, due to the spatial resolution, and amount of data taken in heavy ion collisions. These limitation are even more profound for the third generation of quarks, where no meaningful measurement for, for instance, the Λ_b^0 could be made. Naturally, this calls for upgraded detectors to increase the precision and a higher luminosity of collisions.
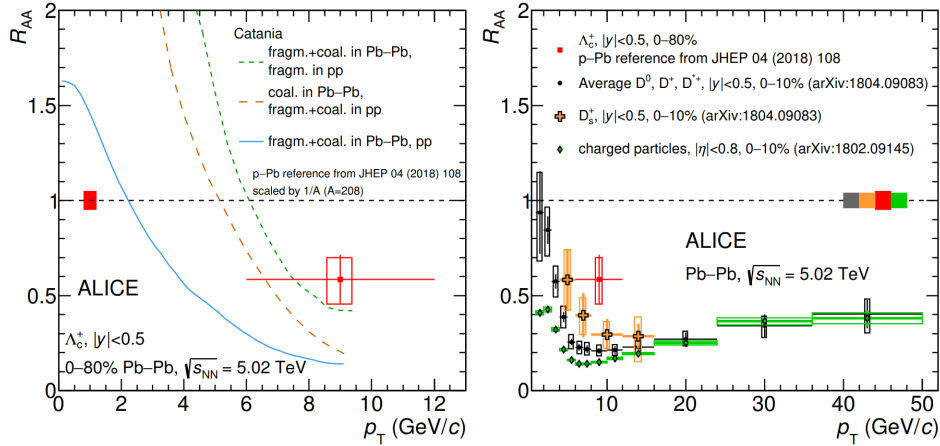

Figure 6: Nuclear modification factor of the Λ_c^+ as measured by ALICE. On the left a comparison with different models and on the right a comparison with other particles.[15]

7

2 Experimental setup of ALICE at the LHC

During the recent long shutdown (LS2) of the LHC, each of the experiments upgraded their setups with new and better subdetectors and readout systems. This version of ALICE aims to better study the QGP and heavier generation of quarks. Previously, the flow of light hadrons (containing up and down quarks) and some hyperons containing strange quarks have been measured very precisely[9]. To expand this research, to for instance the Λ_c^+, a new Inner Tracking System (ITS) has been installed, which from now on will be called ITS2 [16].

This system has three layers of very compact pixel sensor chips close to the interaction point, forming what is called the inner barrel, and 4 layers further away[17]. The pseudo-rapidity range for the new inner barrel is $|\eta| < 1.98$. The first layer surrounding the beampipe has a radius of 2.3 cm as compared to 3.9 cm in the previous version. As a consequence, the accuracy with which the impact parameter can be determined greatly increases. In an environment with a high multiplicity, as in the case of lead-lead collisions, it is increasingly important to be able to distinguish which track came from which vertex. Figure 7 shows the geometry of the ITS2 and a performance comparison of the impact parameter between the old data points in blue, and the upgraded version in red. The improvement is most noticeable at high-$p_{\rm T}$, where a factor 20 is realised. Even at low-$p_{\rm T}$ the resolution is reduced by a factor 4.

Another factor in the equation of the resolution of the detector is the material budget. This is measured by describing the thickness of a layer as a percentage of the radiation length (X_0), such that it includes the density and electromagnetic properties of the material. A lower value in this case means that less energy is extracted from the particles that pass through, and therefore the tracks are altered less, increasing the accuracy of the spatial resolution. The pixels used are Monolithic Active Pixel Sensors (MAPS) and are a factor 7 thinner compared to those in ITS1. Combined with improvements in the circuitry of the power distribution and data lines, the material budget has been reduced by a factor 5, from $1.5\%X_0$ to $0.3\%X_0$ per layer. These smaller pixels have a much faster readout as well, such that the number of collisions that can be processed improves from 1kHz to 100kHz in Pb-Pb collisions, and 400kHz in p-p collisions, allowing compatibility with the high-luminosity plans of the LHC. As a side-effect of this design, the charge signal amplitude will no longer be measured. Instead, a binary readout of the pixels is used. This loss in PID information was studied and concluded to be marginal[18]. The three layers combined cover a surface of $10m^2$ and count $12.5 \cdot 10^9$ individual pixels.

Additionally, the readout of the Time Projection Chamber (TPC) has been upgraded to accommodate for the increased number of interactions per second that are planned for the upcoming years. To detect particles closer to the beampipe a Muon Forward Tracker (MFT) and a Fast Interaction Trigger Detector (FIT) have been employed to remove unwanted signal and increase the spatial resolution of the detector as a whole.

Although the ITS2 was recently installed, an extra upgrade is planned for run4, expected to start in 2029. The new subdetector is called ITS3 and will have the same middle and outer barrel as is the current setup. However, the inner barrel may be replaced by flexible printed circuits, bend around the beamaxis. Not only is the first layer closer to the interaction point, $r = 1.8$ mm, options to place the first two layers inside the beampipe were initially considered. With recent technological advances in the fabrication process of sensors, it is possible to print pixel sensors of arbitrary size. Previously they were limited to a few centimeters in both directions. This allows the opportunity to print an entire strip in one go, to cover the length of the entire barrel. As a result, much of the readout electronics and power distribution can be moved outside the acceptance region of the detector and less power is required to run the detector.
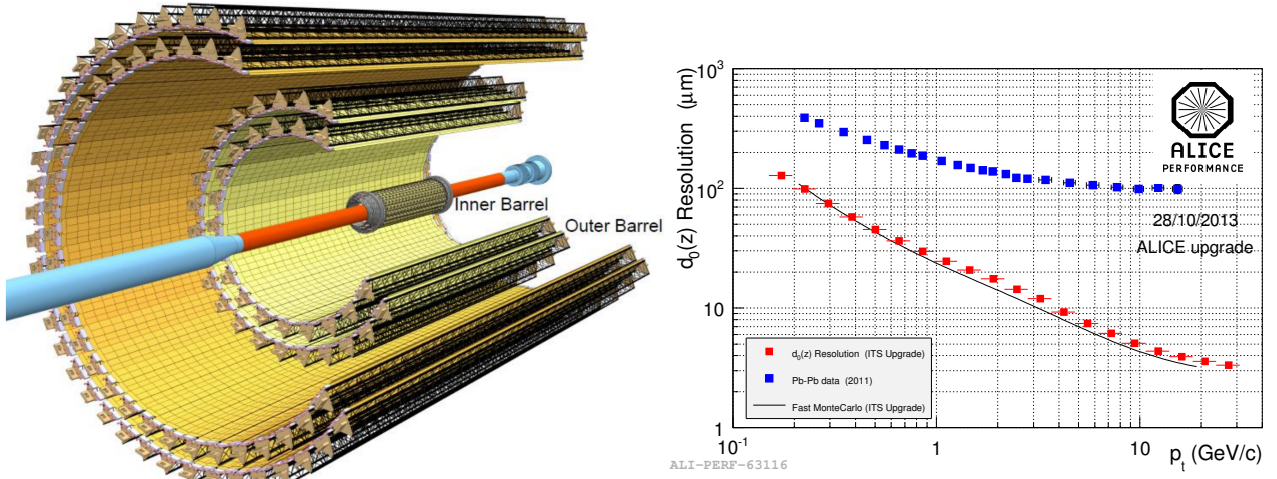
Figure 7: On the left a schematic of the geometry of ITS2 and on the right a comparison with ITS regarding the performance in a measurement of the impact parameter (d_0) in the longitudinal direction as function of transverse momentum [17]

In turn, this allows the option to cool the system solely by air, further reducing the amount of material used. Overall, this reduces the material budget significantly to $0.05\% X_0$ per inner layer. Simulation studies have been performed to obtain an expectation of the improvement in the physics performance. Figure 8 shows a comparison of the Λ_c^+ production for ITS2 (blue markers) and ITS3 (red markers). The left plot presents the significance while the right plot the signal over background ratio both drawn as a function of transverse momentum for the 10% most central Pb–Pb collisions at $\sqrt{s_{NN}} = 5.5$ TeV. The combined efforts result in a much better ability to distinguish signal from background, especially in the low-p_T region, where the signal to background ratio has been increased by a factor 10. Consequently, the significance improves by a factor up to 4, which allows a much more refined measured of, for instance, the elliptic flow of this charmed baryon.
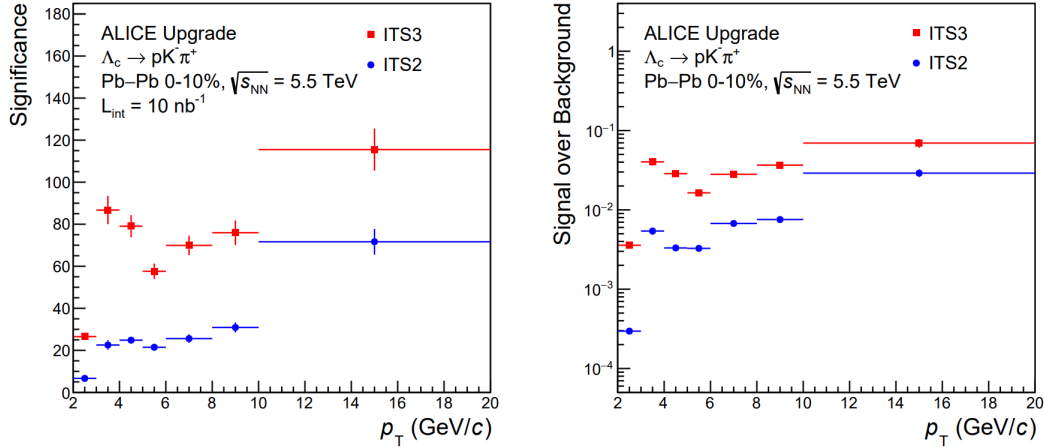
Figure 8: Comparison of ITS2 and ITS3 for $\Lambda_c^+ \to pK^-\pi$ production in very central(0-10%) Pb-Pb collisions at centre of mass energies $\sqrt{s_{NN}} = 5.5 TeV$ ($L_{int} = 10 nb^{-1}$): statistical significance (left) and S/B ratio (right) as a function of transverse momentum[19].

2.1 ALICE3 detector

The exotic hadrons formed from the heavy quarks created in the QGP typically decay very fast, often before they are able to reach any detector. Hence many of the exotic states can only be reconstructed by combining information of the decay products. To extract as much information as possible, a new detector concept, currently called ALICE3, is proposed [20]. Figure 9 shows a schematic overview of the current concept of the upgraded detector setup[21].

A high resolution vertex detector is placed inside the beam-pipe, surrounded by layers of silicon pixel detectors for particle tracking. The outer layers of the barrel segment cover a range up to $|\eta| \approx 1.5$, where the inner layers in the beampipe go up to $|\eta| \approx 4.0$. In addition, 9 layers of endcap disks increase the coverage in the region $1.75 < |\eta| < 4.0$. To improve the efficiency in the direction close to the beampipe (eg. pseudorapidity $3 < |\eta| < 5$), a forward conversion tracker (FCT) based on the same principles is placed in rings around the beampipe a few meters away from the interaction site. Additionally the Time-Of-Flight (TOF) detector measures the timing of charged particles extremely precisely in the barrel segment $|\eta| < 1.9$, such that the velocity can be determined and two collisions very close in time can be distinguished[22]. This is measured with Multigap Resistive Plate Chamber (MRPC) strip detectors and reaches a timing resolution on the order of tens of picoseconds. For supplementary particle identification, a Ring-Imaging Cherenkov (RICH) detector is used. The barrel segment covers up to $|\eta| \approx 1.75$ and the two endcaps extend this to $|\eta| < 4.0$. As the charged particles travel through a refractive medium, cherenkov radiation is emitted as long as the particles have velocities higher than the local speed of light. The angle between the momentum vector and the emitted cherenkov radiation is correlated to the nature of the particle, especially in certain momentum ranges. This allows for a probabilistic identification label, proven to be very useful for the identification of protons, kaons and pions.

The electromagnetic calorimeter (ECal) consists of a high energy resolution barrel segment of lead tungstate ($PbWO_4$) scintillators in the forward and midrapidity region ($|\eta| < 1.75$). Together with a large acceptance endcap segment, $1.75 < |\eta| < 4$, that is made of layers of lead and plastic scintillators, the energy of charged particles can be measured by completely capturing
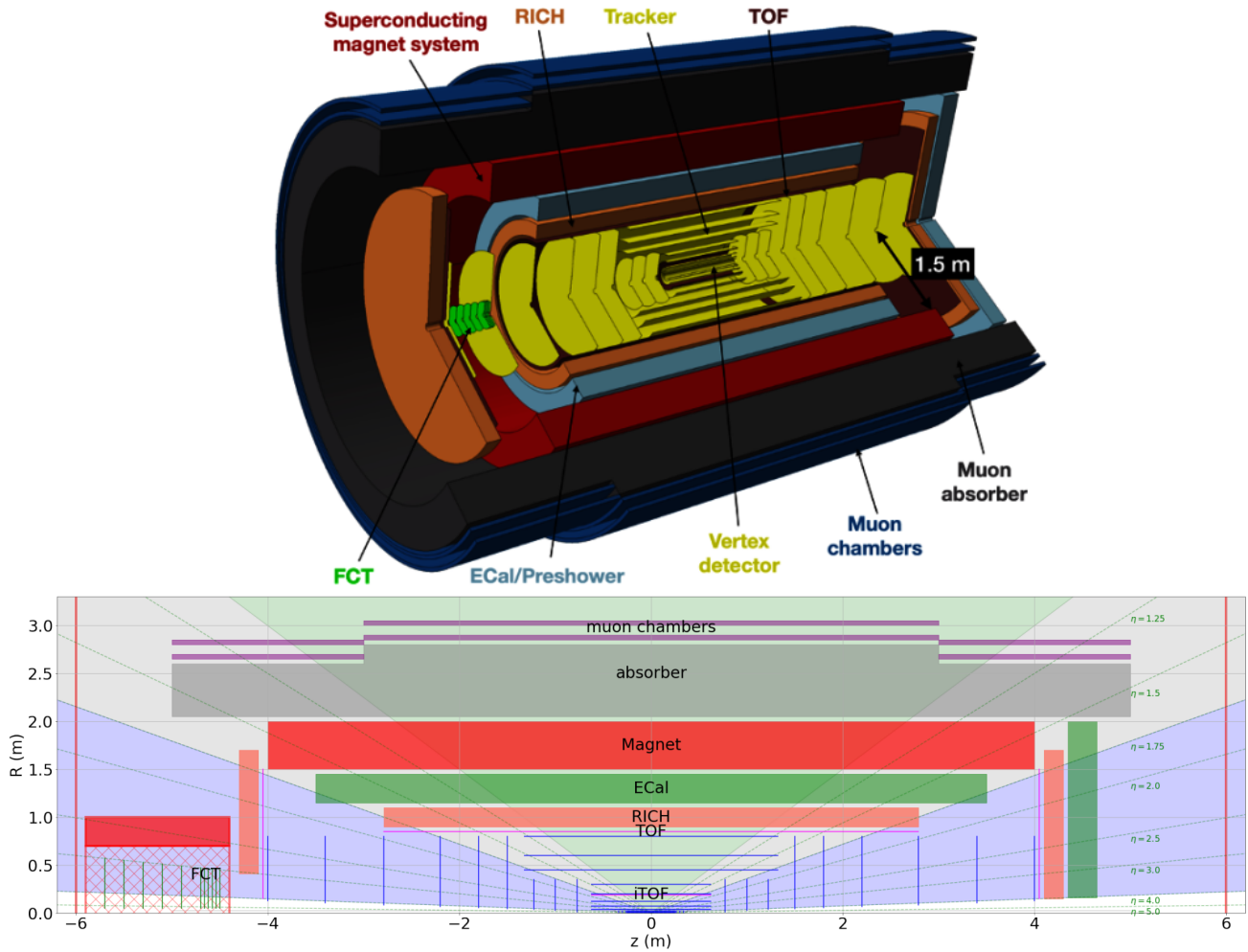
10

Figure 9: At the top a schematic overview of the concept of the ALICE-3 experimental setup. At the bottom a longitudinal cross-section detailing the coverage of each subdetector. Source:[21]

their kinetic energy. The lead layer's function is to interact with (neutral)hadrons in order to 'force' it to fragment into a shower of particles that can be measured with the scintillator layers. The muon absorber is a thick layer of non-magnetic steel that aims to capture any leftover hadrons, such that the only remaining particles after this layer are most likely muons. The muon chambers in the outer layer then measure the position of these charged particles.

The combination of all the data then allows one to reconstruct particle tracks that reached the detectors of the event. Since one of the most interesting parts of the event is that which does not reach the detectors, the primary vertex and exotic short-lived hadrons need to be reverse-engineered by tracing the daughter's tracks back in time and reconstructing the hadrons of interest. This is where most of the analysis is performed, after the events happened, and is also where this thesis is located in the research chain.

With this proposed setup, ALICE3 will have a much improved accuracy compared to previous

iterations. With detectors even closer to the interaction point, which benefits the measurements of heavy quarks greatly, all measurements are much more refined. Figure 10 shows the expected resolution for the impact parameter in the longitudinal direction for ALICE3 in p-p collisions. Not only does it improve by a factor 5 in the acceptance region that was previously present, the extension to higher pseudorapidity provides the opportunity to explore the low-p_T regions with greater accuracy.
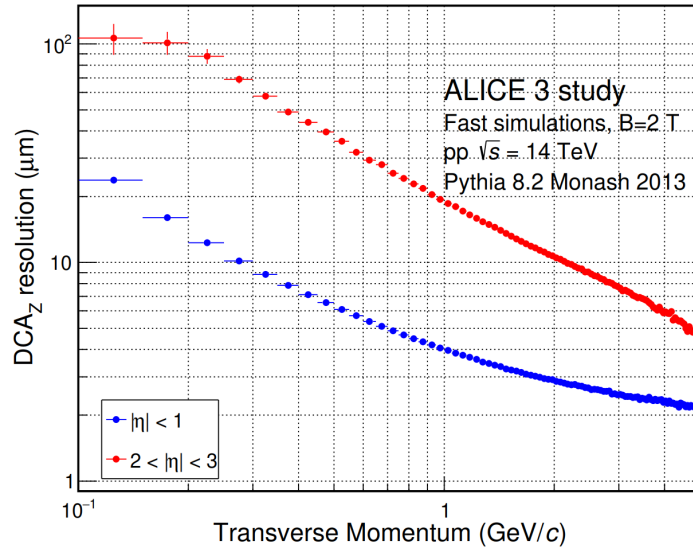

Figure 10: Expected resolution for the impact parameter ($DCA_z = d_0$) in the longitudinal direction for ALICE3. Although it shows the case of p-p collisions, it can be loosely compared to Figure 7. It shows an improvement of about a factor 5 as compared to ITS3 for low pseudo-rapidity tracks($|\eta| < 1$) and the additional acceptance region performs similarly to the previous low-η region[20].

Overall, the current data sample is very limited, which makes it difficult to test certain models. The future of ALICE aims to provide precise measurements of the heavy beauty quarks and shine light on the hadronization and quark-medium interactions in very exotic states of matter. One of the goals of ALICE3 is to measure the elliptic flow of beauty baryons and mesons. This upgraded version of the experiment has the capability to obtain the accuracy and amount of statistics required to perform measurements on these exotic states. In particular, the beauty sector is expected to be enabled by these upgrades. A critical step in this measurement is the analysis of the production of these hadrons. This thesis will provide an analysis on the performance of ALICE3 regarding the production of the Λ_b^0 baryon over a wide range of transverse momentum, namely from 1.0 up to 24.0 GeV/c. It provides the accuracy of the measurements of the flow parameters, based on which claims can or can not be made about the physics.

3 Traditional analysis with Λ_c^+ as example

As the research expands to heavier quarks, it becomes more and more unlikely to come across a relevant interaction in a collision. If, for example, the particle of interest only appears once in a million collisions, there needs to be a method of removing as many irrelevant events and events that appear to be containing the relevant particle, but in fact do not. Only limited information can be reconstructed from an event, from which a candidate can be selected. Many other processes or particles would give a very similar response in the detector and are traditionally separated by so-called 'cuts', requirements for certain measurable parameters to be within a certain range. Additionally, machine learning algorithms can be used to optimize this process, which will be one of the main focus points of this thesis. Mainly a boosted decision tree (BDT) is used to evaluate and optimize the performance of the measurement of Λ_b^0 hadrons in the ALICE-3 experiment. These measurements can be used as a probe for properties of the QGP.

3.1 Raw data from ALICE/simulation

The underlying machinery to simulate high-energy particle collisions is PYTHIA [23]. This general purpose event generator contains tune-able physics models to provide simulations tailored to ones needs. This research is interested in heavy-flavour physics, so in order to study the properties of these hadrons $c\bar{c}$- and $b\bar{b}$-enriched proton-proton collision events are used for the analysis. As compared to a minimum bias sample, these events contain much more information to construct the parameter distributions of the particle of interest by forcing an increased amount of signal events. In either case, the number of background events is abundant for the purpose of this analysis. Later on, a correction is made to achieve the correct physical results. Additionally Delphes is used, which adds functionalities to propagate tracks to the outer regions of the detector and include the magnetic field that is present in the experiment[24]. To emulate the detector's response to the event, a fast parametrization of the detector's properties is used. The output of this combined simulation contains the tracks of all particles involved in the collision, including all their decay products. A more accurate and complete method would be to simulate each part of the detector individually. However, since there are hundreds of thousands of output channels, this would cost a lot computational power. For the purpose of this study, such levels of accuracy are not required and therefore not used.

Most of the analysis is performed in the PWGHF (Physics Working Group: Heavy Flavour) section of the Online-Offline (O^2) software environment [25]. An elaborate schematic of the process can be found in Figure 11, however, the relevant steps will be discussed here.

A track skimmer goes over the list of all reconstructed tracks and creates combinations of the tracks that could be originating from a 2-point or 3-point secondary vertex. Some basic information about each track is added as well, eg. charge, transverse momenta (p_T) and statistical values for particle identification (PID) where applicable.

3.2 Candidate creator

This part of the software aims to construct a list of potential candidate tracks. Depending on the lifetime of these particles, they may only be detected by the inner parts of the detector, or do not even reach the detector at all. Even if they are measured directly, it is extremely hard to correctly determine their nature due to the limited amount of data. One approach is to combine reconstructed tracks of the decay products into a potential track of the mother particle, the candidate. In order to reduce the amount of possible combinations, one could pick a specific decay channel and only take combinations of those daughter particles. Let us take the
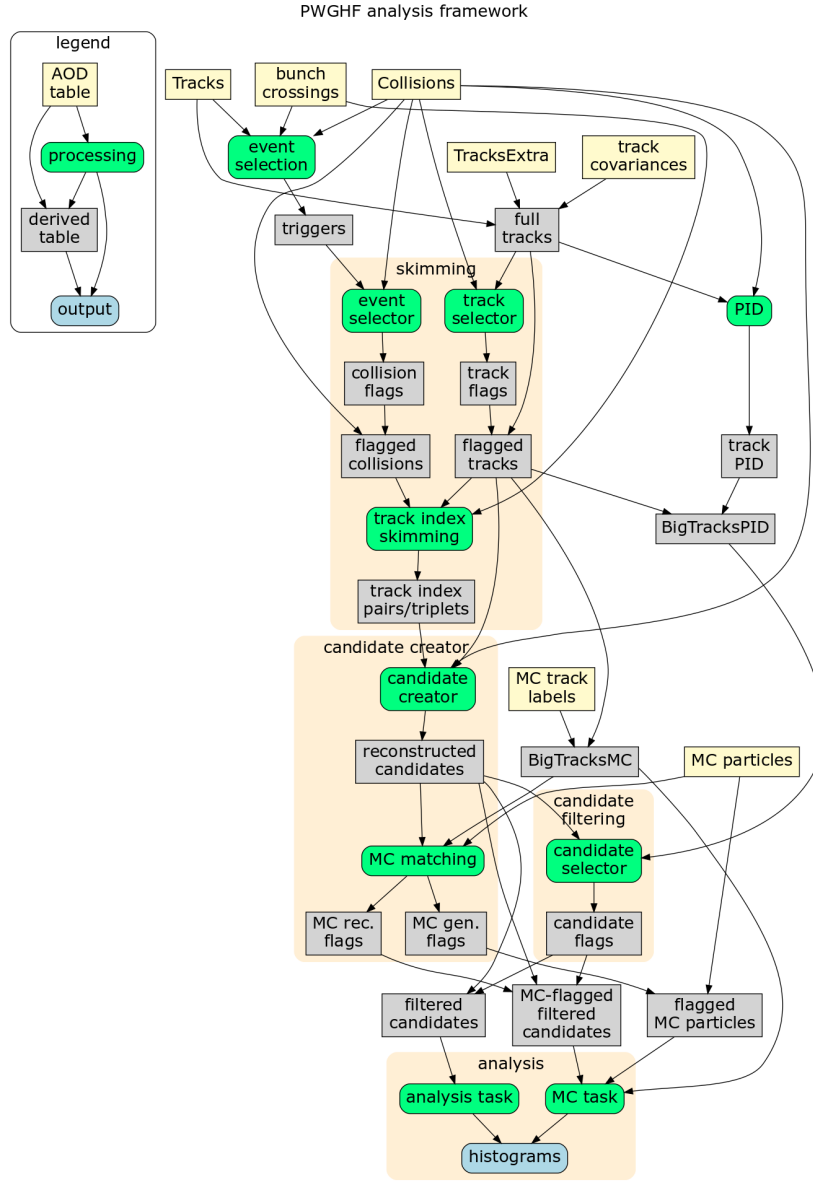
Figure 11: Schematic of the PWGHF analysis workflow in the O^2 software. Source:[25]

Λ_c^+ as example. A relatively large fraction of these charmed baryons decay into a proton, kaon and a pion, with a branching ratio of $BR(\Lambda_c^+ \rightarrow p + K^- + \pi^+) = 6.28 \pm 0.32\%$[4]. Since all these daughter particles have typical decay lengths larger than the size of the detector, they can be measured directly, making it a perfect choice for reconstruction of the Λ_c^+. The candidate creator takes the list of tracks potentially involved in a 3-point vertex and determines associated parameters of this secondary vertex. For instance, the impact parameters of each of the daughter

particles, the invariant mass of the candidate mother particle and the pointing angle[2]. Due to the presence of the magnetic field in the experiment, the charge of the measured particles can be determined by their curvature relative to their momentum vector. This way many of the 3-prong vertex candidates can be disregarded with great certainty because of an electric charge mismatch.

3.3 Candidate selector and traditional cuts

Since protons, kaons and pions are common products in the decay of many other particles, it is very likely that a set of three can be reconstructed without them originating from the particle of interest. These candidates are considered to be 'background events'. The goal of the analysis is to make a selection of the candidates, such that as much background events as possible are cut out, while maintaining most, if not all, of the signal events. Since the reconstructed and physically-measured data do not contain labels of the true nature of the tracks, simulations are very important. The reconstructed tracks can be compared with those in the monte-carlo simulation. This process called monte-carlo matching, provides a way to evaluate the performance of the analysis by a comparison with the generated particles (eg. determine a false-positive rate, reconstruction efficiency etc.).
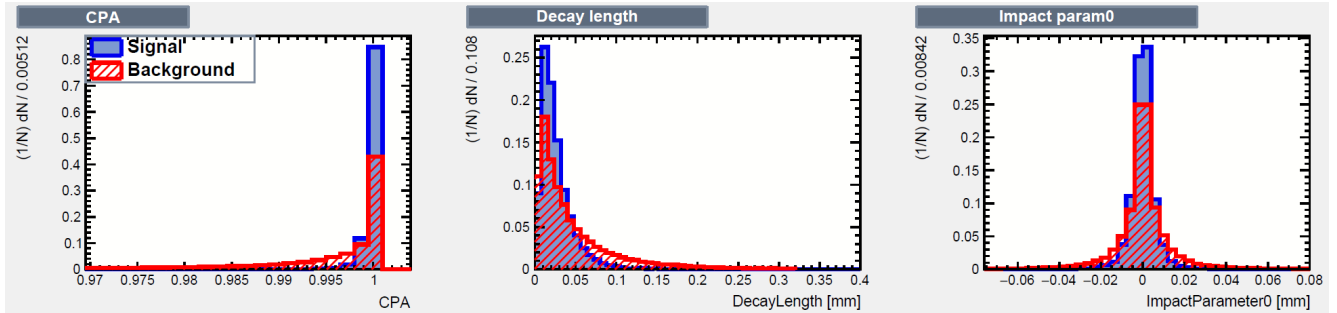

Figure 12: Some examples of normalized distributions of parameters of the $\Lambda_c^+ \rightarrow p^+ + K^- + \pi^+$ decay candidates, categorized in signal and background. Data is obtained from p-p collisions at $\sqrt{s} = 14.0 TeV$ with transverse momentum of the mother particle $4.0 < p_{\mathrm{T}} < 6.0$ GeV/c. (left) cosine of the pointing angle, (middle) decay length of the mother particle and (right) impact parameter of the proton.

Figure 12 shows a few examples of parameter distributions of reconstructed Λ_c^+ candidates with a transverse momentum between $4.0 < p_{\mathrm{T}} < 6.0$ GeV/c. The candidates originate from a $c\bar{c}$-enriched sample of proton-proton collisions at $\sqrt{s} = 14.0 TeV$. It is assumed that the parameters measured are independent on the type of collision (eg. heavy ion or proton collisions). On the vertical axis is a normalized value of the number of occurrences as function of the parameter value. The signal (blue) and background (red) have slightly different distributions. Hence, limitations on these event parameters can be used to create an effective selection cut. In this example, the cosine of the pointing angle (CPA) of the signal events is, on average, closer to unity. Therefore, a requirement of $CPA > 0.995$ would cut away some background while leaving the signal untouched. This traditional method allows to tune the selection criteria to optimize the analysis. The exact selection criteria used can be found in Appendix A and the resulting

---

[2]The pointing angle is defined as the angle between the momentum vector of the particle and the vector from the primary to the secondary vertex. It is a useful parameter as it indicates the origin of the candidate, where heavy quarks mostly originate from the primary vertex.
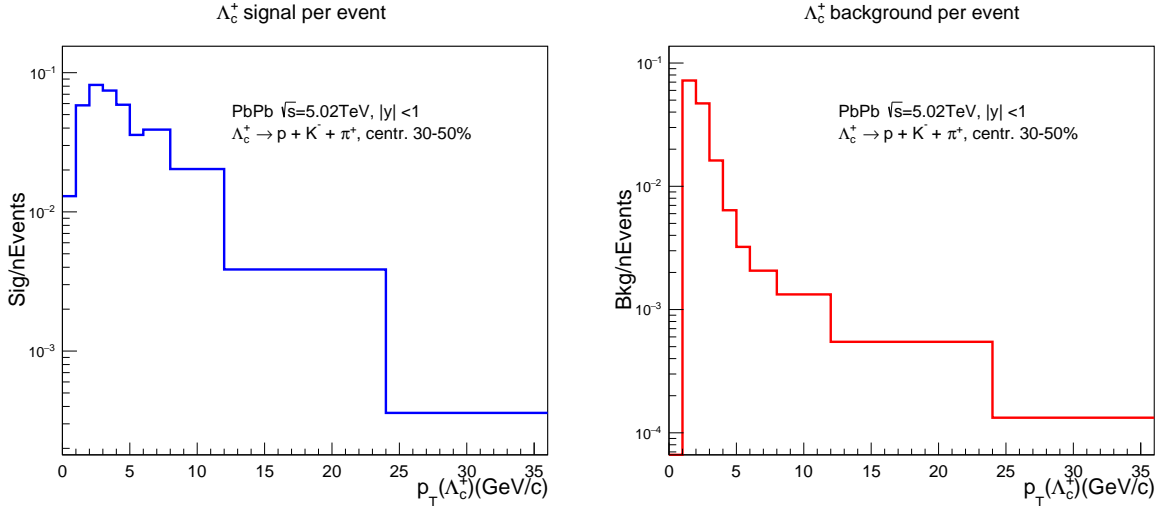
Figure 13: The number of expected signal (left) and background (right) candidates per event in Pb-Pb collisions at $\sqrt{s} = 5.02 TeV$ after applying selection criteria.

number of expected signal events after applying the cuts is shown in Figure 13. These histograms describe the number of accepted events per collision as a function of transverse momentum in lead-lead collisions at $\sqrt{s_{NN}} = 5.02 TeV$. In both cases an exponential decrease at mid- and high-p_{T} can be observed, which is partially a signature from the overall p_{T} distributions of particles in these types of collisions. At very low p_{T} the reduced sensitivity of the detector takes over, drastically reducing the number of candidates.

One potential value to optimize the selection criteria for is the signal to background ratio. This may be useful for determining branching fractions and ratios etc. Another value is the significance, defined as

$$significance = \frac{N_s}{\sqrt{N_s + N_b}} \tag{2}$$

where N_s and N_b are the number of accepted signal and background events respectively. The maximization of this latter measure will be the goal on this research as it serves the purpose of the examination of properties of the quark-guon plasma well, as discussed in subsection 1.1.

3.4 Yield from theory (PYTHIA)

The events used in the analysis originate from simulations and are altered to enhance their purpose in the analysis. In the case of Λ_c^+, the sample is enriched with $c\bar{c}$ pairs, which in turn cause an increase in the number of Λ_c^+ particles. This is useful to study the effectiveness of the selection criteria as discussed in previous sections. However, it does not give a correct physical representation of what results can be expected from the experiment. To obtain the proper yield, models can be used to estimate the amount of true signal events that can be expected. The number of obtained signal candidates per event is

$$\frac{N_{s,pp}}{N_{evt}} = 2 \cdot \Delta p_{\mathrm{T}} \cdot \Delta y \cdot BR(\Lambda_c^+ \rightarrow p + K^- + \pi^+) \cdot (Acc \times \epsilon)_{\Lambda_c^+} \cdot (d\sigma/dp_{\mathrm{T}})_{|y|<1.44}^{\Lambda_c^+} \quad (3)$$

$$\frac{N_{s,PbPb}}{N_{evt}} = \frac{N_{s,pp}}{N_{evt}} \cdot \langle N_{coll} \rangle \cdot R_{AA}^{\Lambda_c^+} \quad (4)$$

In the previous equation the factor 2 is there to include anti-particles, Δp_{T} is the binwidth in the transverse momentum range, Δy is the binwidth in the pseudorapidity range, $BR(...)$ is the respective branching ratio, N_{evt} is the expected number of events during the experiment, $(Acc \times \epsilon)_{\Lambda_c^+}$ is acceptance times efficiency obtained from the analysis of the simulation data, $(d\sigma/dp_{\mathrm{T}})_{|y|<1.44}^{\Lambda_c^+}$ is the expected cross-section of Λ_c^+ obtained from PYTHIA[23], $\langle N_{coll} \rangle$ is the mean number of nucleon collisions in the centrality region being investigated[26], and $R_{AA}^{\Lambda_c^+}$ is the nuclear modification factor obtained from TAMU [15, 27].
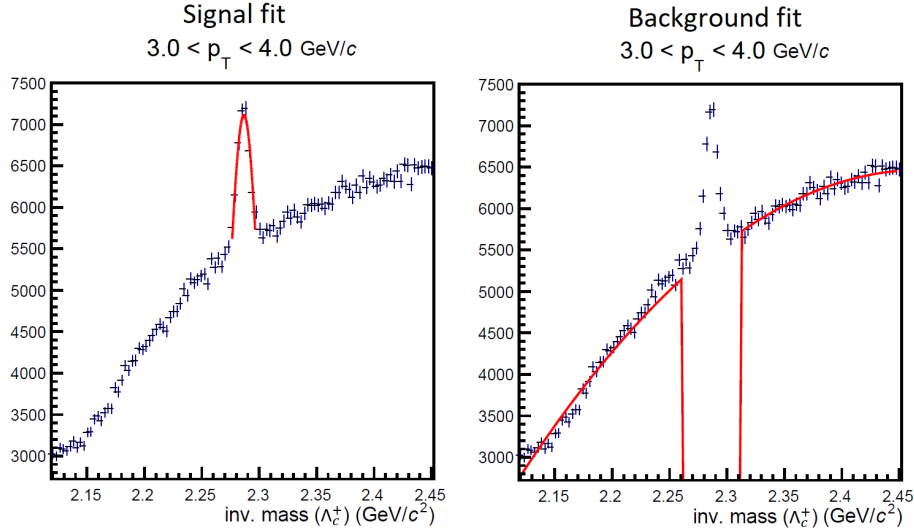
Figure 14: Mass distribution of Λ_c^+ candidates in the region $3.0 < p_{\mathrm{T}} < 4.0$ GeV/c from $c\bar{c}$-enriched p-p collisions at $\sqrt{s} = 14.0 TeV$. Left: example of fitting the signal to the peak around the mass hypothesis of the Λ_c^+, where the fit is shown for the region $\pm 3\sigma$. Right: example of the fit of the background, excluding the region around the signal peak $\pm 5\sigma$.

Similarly, the background is altered by these modifications as well. A different sample is used that contains the least of amount of artificial modifications, a minimum-bias (MB) sample.

By convention, for the calculation of the significance, a combination of three particle-tracks is considered a background event if the obtained invariant mass of the candidate falls within 3 standard deviations of the signal peak. Hence, from the enriched sample, the peak in the invariant mass distribution is fitted by a gaussian to determine the $\pm 3\sigma$-region. Figure 14 shows an example of how the signal is fit of Λ_c^+ candidates in the region $3.0 < p_\mathrm{T} < 4.0$ GeV/c from $c\bar{c}$-enriched p-p collisions at $\sqrt{s} = 14.0 TeV$. Background events are combinations of other random distributions and therefore have no particular reason to have a narrow peak at a specific value. The sidebands surrounding the signal peak are fitted with a polynomial excluding the signal region to avoid bias. In this example the same distribution is shown solely for indicative purposes. However, to get the most realistic estimation, the mass distribution of the MB sample is fit to a polynomial excluding the signal region as determined from the enriched sample. By interpolation and integrating of the fit in the $\pm 3\sigma$-region, the number of background events considered for the significance can be determined. Once divided by the number of events used in the sample, this provides the number of background events per collision and can be scaled according to the number of events N_{evt} later on.

3.5   Results, efficiencies and significance

From the monte-carlo matching procedure, an efficiency for the reconstruction of signal events as function of transverse momentum can be determined of which the result is shown in Figure 15. At very high momentum the reconstruction efficiency is very high because the particle density is relatively low and these particles pass through a large part of the detector. However, not many candidates obtain such large values of transverse momentum and therefore the significance is not that high in this region, as shown in Figure 16 for the case of Pb-Pb collisions as a function of transverse momentum. On the other end of the spectrum, a particle with very low transverse momentum will have large curvature and might not cross a significant part of the sensitive areas of the detector and thus not getting fully reconstructed. Although the number of signal events is much higher at low-p_T, due to the reconstruction efficiency, the significance is still fairly low. The intermediate region is where the performance peaks. The number of signal events per collisions is still high (Figure 13), the reconstruction efficiency is decent, and the background contamination is much lower as compared to the low-p_T region. The total number of events used ($N_{evt} = 3.8 \cdot 10^{10}$) is the amount expected during run5 at the time of writing. Note, however, that selection criteria used are not perfectly optimized as that was not the goal for now.
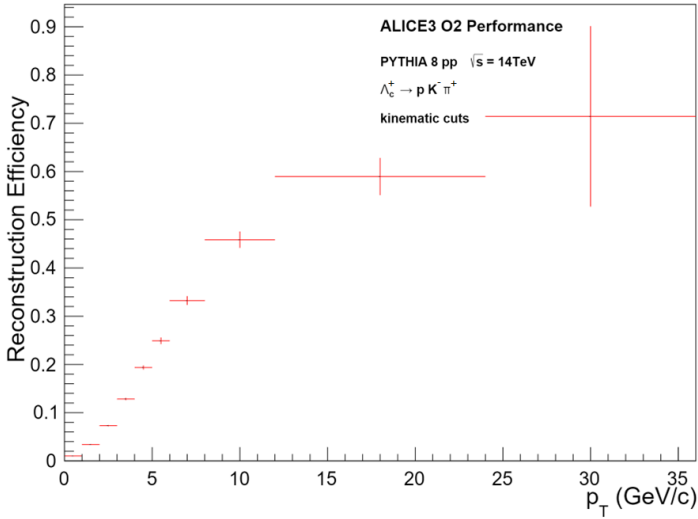

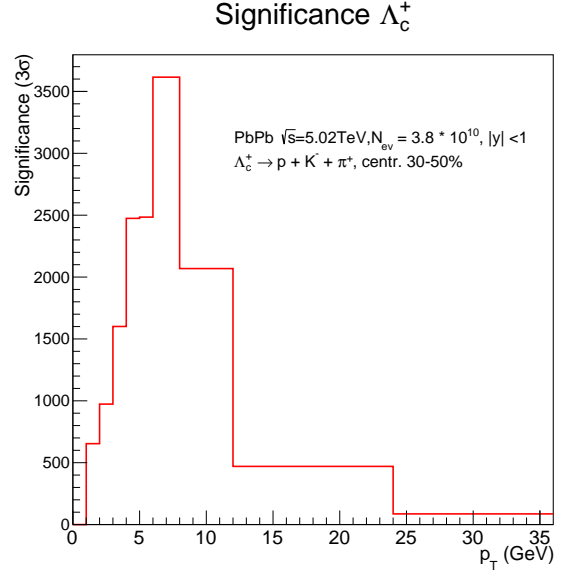Figure 15: Reconstruction efficiency of Λ_c^+ after applying selection criteria.

Figure 16: The obtained significance of Λ_c^+ in Pb-Pb collisions at $\sqrt{s} = 5.02 TeV$, after applying selection criteria.

4   Machine learning algorithms in particle physics

The process of finding the optimal cuts can be very tedious by hand. However, since there is a single value (significance) that needs to maximized, the optimization is very well fit for machine learning algorithms. Naively one could use a monte-carlo scheme with a set of random cuts that can be varied in each iteration with a resulting significance as measure of performance. This should work decently well to solve the problem to first order. However, there are underlying secrets that may not be exposed by the parameter distributions themselves. One of these is the correlation between measurable parameters. As the particles from a signal event are coming from the same mother particle, these correlations should be present for certain parameters, like the momentum and angles between the momentum vectors.

4.1   Boosted decision tree

The next logical step could be to include certain correlations, eg. if parameter $x_1 > a$ and $x_2 < b$, it may be flagged as a signal event. This is exactly the principle of a Decision Tree algorithm. One can create several layers of such decisions and let the result be a classification, the event is either labeled as signal of background. Figure 17 and Figure 18 show an example of a simple decision tree and the resulting map of the classification. The automated algorithm used in this analysis (from the Toolkit for MultiVariate data Analysis, TMVA, as part of the ROOT software package [28]) works as follows: the first node scans the sample's parameter distribution and chooses a parameter and cut value that yields the highest separation power. This decision then splits the sample in two parts, each used in the second-level nodes. These repeat the process of finding the optimal decision which continues until some maximum depth or minimal statistically significant separation is obtained. With p being the purity of the sample, the separation power can be defined in multiple different ways[28, 29], with no noticeable difference in performance:

- Gini Index, defined by $p \cdot (1 - p)$;
- Cross entropy, defined by $-p \cdot ln(p) - (1 - p) \cdot ln(1 - p)$;
- Misclassification error, defined by $1 - max(p, 1 - p)$;
- Statistical significance, defined by Equation 2

Although it adds the opportunity to label certain regions within the parameter space, it is only limited to placing fences perpendicular to the parameter axis. For example, if some optimal region is circular, eg $x_1^2 + x_2^2 = r^2$, it can only be approximated by a rectangle. If, a priori, this pattern is expected one could add a new parameter $x_3 = x_1^2 + x_2^2$ and the decision tree can find a single cut value on the x_3-axis to include this property into the selection criteria. The obvious downsides are that a priori knowledge is required, or all combinations of parameters should be added and see which one may work, or make the tree infinitely deep to approximate the circle by an infinite number of squares. The former solution may not be feasible as it is reliant on theory and the latter two would lose to the tremendous cost in computational power.

A different approach is to instead use a forest of decision trees and let each grow on their own, create and learn their own set of decisions. In the end one can take the average of the opinions of all the trees and assign some number to how convinced the forest is about the classification of the event. Such an ensemble model is called a Boosted Decision Tree (BDT), where the 'boosting' is a reference to how the algorithm learns from misclassified events. Each tree is generated sequentially, where the first tree uses the training sample as-is. The decisions are determined based on the separation between the distributions of the signal and background
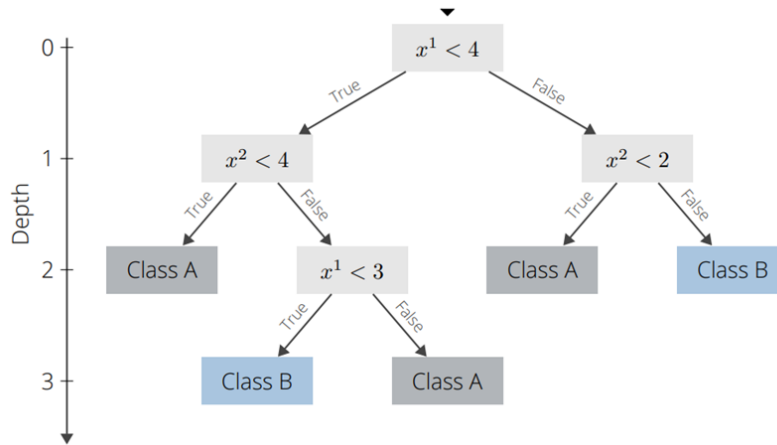
Figure 17: schematic of a very simple scheme of a decision tree. Each layer contains an if-statement that determines how an event is progressed through the tree. The result is a classification of the event, typically either 'signal' or 'background'. Source: [30]
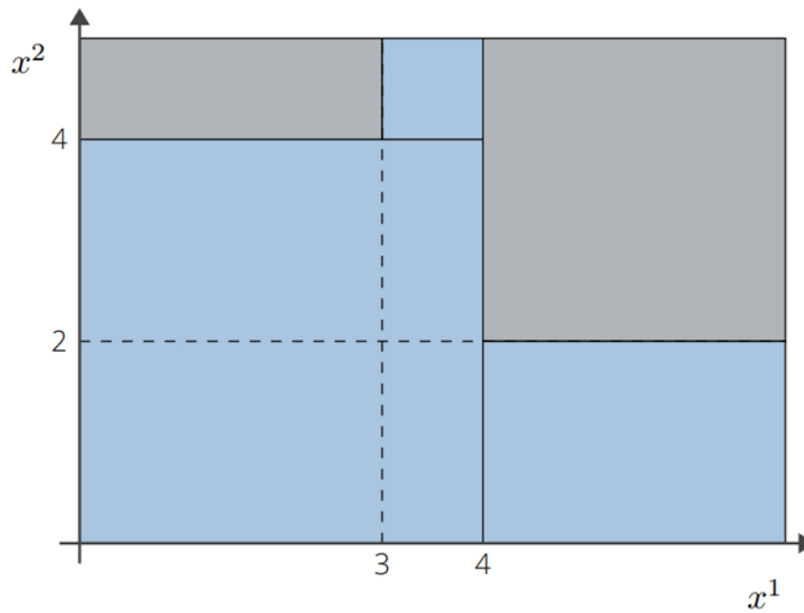
Figure 18: The resulting map of the classification by the decision tree shown in Figure 17. Such a simple model is only able to make distinctions between classes based on cuts perpendicular to the parameter axis. This yields the rectangular classifications in parameter space as shown in this image. Source: [30]

event parameters. Afterwards, the entire training sample is evaluated with this tree and wrongly classified events receive a higher weighting, hence the name boosting. This will slightly change the parameter distributions of the sample for the next tree. When the second tree trains on this

sample, it tends to focus more towards the complicated area where signal and background are very much alike. With each iteration the sample is reweighted and therefore each tree makes different decisions. Naively one could think the last tree of the forest will be the 'strongest' since it contains the most information. However, for the same reason, it uses a heavily modified training sample that is designed to classify a very specific region in parameter space. Hence, the entire forest should still be used, to preserve the classification of the entire parameter space. A welcoming side-effect is that boosting increases the statistical stability of the sample. When a single tree is used, a slight fluctuation in the event parameters can mean the difference between two completely different decisions, whereas the boosts yields the exploration of many different decision structures.

4.1.1 Types of boosting

The finesse of the algorithm now lies in how the boosting or reweighting is actually implemented. The default BDT used makes use of adaptive boosting, where the weights of the events are changed based on the error rate $R_{error} = N_{misclassification}/N_{events}$

$$\alpha = \frac{1 - R_{error}}{R_{error}} \tag{5}$$

where α is the multiplier for the weight of all misclassified events. The general idea is to loop over all previous trees, determine their classification and error rate, calculate the relevant boost and then the new boosted classifier is given by

$$y_{Boost}(\mathbf{x}) = \frac{1}{N_{trees}} \sum_{i}^{N_{trees}} \beta \ln{(\alpha_i)} h_i(\mathbf{x}) \tag{6}$$

where $\mathbf{x}$ respresents the tuple of event parameters, the first factor on the right hand side accounts for the renormalization to obtain the same number of events after boosting, β is the learning rate, α is the multiplier for the weight of an individual event, and $h(\mathbf{x})$ is the classification of an event ($+1$ for signal and -1 for background). The learning rate is a modifiable parameter that regulates the effect of the boost per iteration. A slower (lower) learning rate typically avoids overtraining and can increase performance when combined with an increased number of trees in the forest to compensate for the lower progress per iteration. This, of course, comes at a cost of computational time.

A different type of boosting is a gradient boost, where the forest can be seen as a function and the algorithm attempts minimize the difference between the response and a true value y of the sample. Each iteration adds to the function and a set of weights P is generated to fit the function. The function can be described as

$$F(\mathbf{x}; P) = \sum_{i}^{N_{trees}} \gamma_i f(x; a_i) \tag{7}$$

where $P \in \{\gamma_i; a_i\}_0^N$ is a set of weights, γ is a weight for a tree, and a is the weight for an event. To find the parameters a deviation is measured by a Loss-function, which in theory can be any function. An exponential loss function leads to the adaptive boosting previously described. The implementation used for the gradient boost uses a binomial log-likelihood loss function

$$L(F; y) = \ln\left(1 + e^{-2F(\mathbf{x})y}\right) \tag{8}$$

P is chosen such that $L(F; y)$ is minimal. The set of P is appended in each iteration as the forest grows. Similarly to before, a learning rate can used with the same consequences. The intention behind this implementation is that the adaptive boost does not deal well with outliers, mislabeled events and noisy parameters, which should be improved upon with this loss-function since the loss is much smaller for extreme cases relative to the case of adaptive boosting.

A completely different approach uses Fisher discriminants, where the parameter space is transformed to separate classes. The idea is to define a new axis in parameter hyperspace along which the signal and background are as much separated as possible, while keeping the events within a class close together[31]. The separation can then be described as the ratio of the variance between the classes to the variance within the classes

$$S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(\vec{a} \cdot \vec{x}_S - \vec{a} \cdot \vec{x}_B)^2}{\vec{a}^T \Sigma_S \vec{a} + \vec{a}^T \Sigma_B \vec{a}} \quad (9)$$

where $\vec{x}_C$ is the sample mean of class C, Σ_C is the covariance of a class, and $\vec{a}$ is the vector of the new axis, hence, the separation is perpendicular to $\vec{a}$. The separation S is maximum when

$$\vec{a} \propto \frac{\vec{x}_S - \vec{x}_B}{\Sigma_S + \Sigma_B} \quad (10)$$

Figure 19: Example of axis transformation to maximize class separation. Source: wikipedia

Figure 19 shows visually how a new axis can separate two very simple distributions better than the original axes. Since the variance of both classes is identical, this essentially means that the new axis passes through the two sample means.

The classification can then be constructed from the following information: the sample mean $\bar{x}_i$ for each parameter i, the sample mean of each class for every parameter $\bar{x}_{C,i}$, and the covariance matrix of the sample *within* a class W_{ij} and *between* the classes B_{ij}, where j denotes the event number. The Fisher coefficients are then given by

$$F_i = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{j=1}^{N_{evt}} W_{ij}^{-1} (\bar{X}_{Sij} - \bar{X}_{Bij}) \quad (11)$$

where $\bar{X}_{Cij} = x_j - \bar{x}_{C,i}$ is the difference between the value of a parameter in event j and the sample mean of class C. Typically one reweights the events for training such that $N_S = N_B$, in which case the fraction in front of the sum becomes $\frac{1}{2}$. These coefficients are indicative of the discrimination(separation) power and are used to choose which axis $\vec{a}$ and cut value is to be used in a node of the BDT.

In the end, an average of the three implementations was used, where the difference between them is used as contribution for the systematic uncertainty in the significance. The specific settings for each implementation in the TMVA framework can be found in Table 1. *MinNodeSize* depicts the minimum percentage of the training events required in a leaf node, setting this too low may cause overtraining. *MaxDepth* is the maximum number of sequential choices in a tree(see Figure 17). *Shrinkage* and *AdaBoostBeta* are the parameters to modify the learning rate in the gradient boost and adaptive boost algorithms respectively. *BaggedSampleFraction* is the fraction of the original sample that determines the size of the bagged sample when it is used.
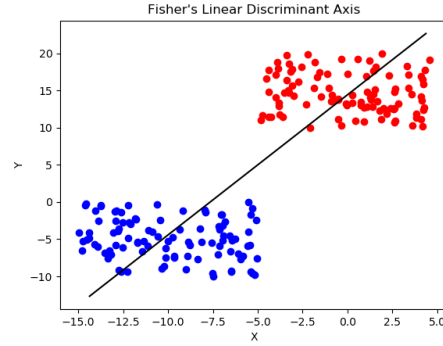
SeparationType states the way the separation power is determined as described in subsection 4.1. *nCuts* is the number of different selection cuts evaluated when determining a new decision in a node. Some slight variations of these values have been tested, however, no noticeable difference was observed and therefore it was not further investigated.

Setting	Adaptive Boost (BDTA)	Gradient Boost (BDTG)	Fisher Discriminant (BDTF)
N_{Trees}	850	1000	50
MinNodeSize	2.5%	2.5%	2.5%
MaxDepth	3	3	3
BoostType	AdaBoost	Grad	AdaBoost
Shrinkage	-	0.10	-
AdaBoostBeta	0.5	-	0.5
UseBaggedBoost	✓	✓	×
BaggedSampleFraction	0.5	0.5	-
UseFisherCuts	×	×	✓
SeparationType	GiniIndex	-	GiniIndex
nCuts	20	20	20

Table 1: The specific settings used for each variation of BDT in the TMVA framework.

4.1.2 Output of a BDT

The output value, which in the rest of the text will be called the BDT response, is a new parameter that can be cut on in the same way how the other parameters were used traditionally. In this specific framework (TMVA) a value between -1 and $+1$ is assigned to each candidate, where the negative extreme indicates that the entire forest agrees it is a background event, and the positive end a signal event respectively. Hence, the goal of training the BDT is to grow trees such that most of the forest classifies the candidates as signal or background correctly. Figure 20 shows an example of the distributions created by such an algorithm. One can remove all the events that have a BDT response lower than a certain value, such that the remaining sample is better suited for one's needs. Depending on the loss function or other boosting method used, these distributions will show different behaviour. The example shown used adaptive boosting, which 'punishes' misclassified candidates way harder than it rewards correct classifications. Therefore, the algorithm does not attempt to push as much signal to $+1$ as possible or vice versa, but instead focuses on the overlapping region in the middle. Since a cut will keep everything above that value anyway it does not matter how much higher the BDT response of an event is compared to the cut. As a result, the computational power is mostly spend on distinguishing the difficult to separate background events from the signal. The distributions shown have a clear peak for both signal and background, and especially the background contains difficult-to-classify tail between -0.2 and 0.1.

Since the (simulated) events used to train and test the BDT often do not correspond with the amount of signal and background events in a physical experiment, one can use the signal and background efficiency as a function of BDT response instead. This allows the calculation of the significance for any amount of expected signal and background events as a function of the cut value applied on the BDT response. Figure 21 shows the efficiencies, signal purity, and resulting significance superimposed. Starting at the left hand side, with a BDT cut of -0.5, nothing would be removed from the sample and therefore the original significance will be obtained. As the cut value increases, it first starts removing mostly background events from the peak between

−0.4 and −0.2, and the significance slowly increases. Above values of −0.05, a rapidly increasing number of signal events are being lost, which has a strong impact on the significance. In this example, a maximum significance is obtained when cutting away all events with a response lower than 0.0344, at which point about 85% of the background is removed and the signal suffered 20% loss from collateral damage.

An important assumption made in this analysis procedure is that sample used to train the BDT is very much representative of expected data from the experiment, to the extend that scaling the number of events by orders of magnitude is still valid. Since the events are from generated p-p collisions, it is also assumed that the reconstructed candidates have similar parameter distributions and correlations to the candidates measured in Pb-Pb collisions. Therefore the same BDTs can be used for both collision types.
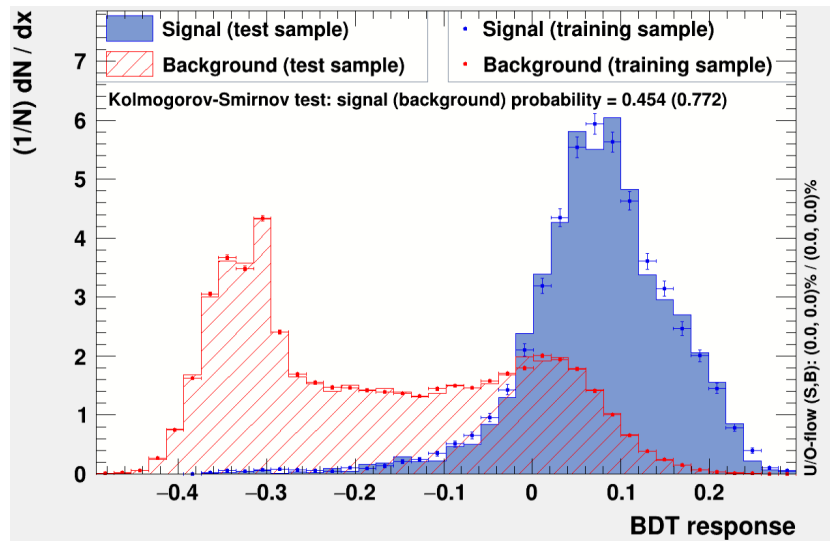
Figure 20: Example of normalized signal and background distributions as a function of BDT response obtained from an adaptive boost algorithm.

4.1.3 Parameter ranking

The importance of each of the variables used in the algorithm can be described by multiple factors. Primarily by the number of occurrences of that variable in the forest, reweighted by the separation gain (ratio of separation index before and after the split) squared of the decision in that occurrence and the amount of events used in that node. For the method that uses the Fisher discriminants, the coefficients (Equation 11) can be used to create a ranking for the overall forest. The discrimination power of parameter i is

$$\lambda_i = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{j=1}^{N_{evt}} F_i(\bar{X}_{Sij} - \bar{X}_{Bij}) \tag{12}$$

If, for any reason, one needs to reduce the dimensionality of the BDT either for future generations or for faster application (eg. during the triggering of events), the ranking provides a good list of importance which indicates which variables can be removed or ignored with the least amount of impact on the performance.

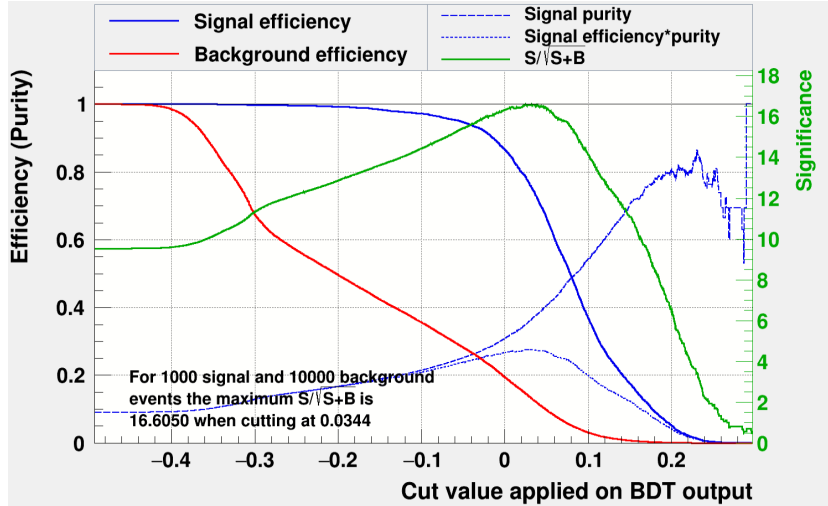Figure 21: The resulting efficiencies, signal purity, and significance as a function of the cut value applied on the BDT response of the sample distributions of Figure 20 scaled to a case where 1000 signal and 10 000 background events are expected.

4.2 Improvement of the analysis

This tool can be applied in many different ways. One way is to completely replace the selection cuts procedure by a BDT. One advantage of this is that one has plenty of events that can be used to train the BDT, since all possible candidates that passed the candidate creator stage are kept. The obvious downside is, since many of these candidates can be easily disregarded as background by a topological cut, a lot of the computational power is 'wasted' on the easily achievable decisions. One could compensate for this by increasing the depth of the trees, at an exponential increase in the computational cost. The other extreme would be to first optimize the selection criteria to obtain the maximum significance with strict topological cuts and then use the selected events to generate a BDT on top of that. This way the algorithm's power is only used to distinguish the most difficult background events from the signal. These events will have extremely similar parameter distributions, but perhaps not the correlations between them as one may expect from signal events. As a result, this will give a boost on top of the analysis procedure that was already decent. The downsides are that the amount of events can become very scarce, especially once the sample is split in p_T-bins, and the cuts may have removed a part of the spectrum that would have worked really well in the BDT. Hence, the optimal strategy probably lies somewhere in between. To find this point, one could try many different levels of strictness in the selection criteria, given one has plenty of computational power. Another possibility to make it a bit more efficient could be to construct some parameter space that includes the candidate parameters and the BDT response, test different locations in this space sampled from a Latin hypercube algorithm, use this set to train Gaussian process emulators and finally use a Markov chain Monte Carlo random walk through the parameter space to find the best configuration [32, 33]. As the purpose of this thesis is find a first estimate for what to expect from ALICE3, such an elaborate optimization was not performed. Instead, a semi-optimized set of selection criteria (see Appendix A) was used with the BDT as final addition. The resulting significance and the effect of the BDT cut in the Λ_c^+ analysis can be seen in Figure 22. With the selection criteria used, the BDT really adds its value in the low-p_T region, where an increase by a factor 5 can

be noticed. In the intermediate($5 < p_T < 12$ GeV/c) and high-p_T regions it generally doubles the resulting significance.
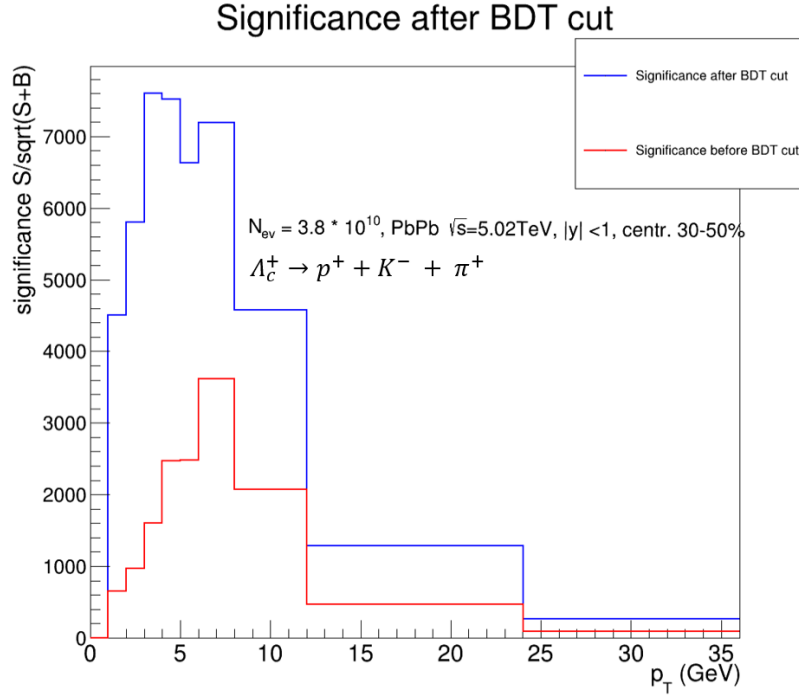
Significance after BDT cut

Figure 22: The significance of Λ_c^+ measured in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02TeV$. Both cases use the same set of selection criteria. One has an additional BDT-response cut as final step in the optimization process.

5 Λ_b^0 performance studies

5.1 Decay channel

To extend the research of the QGP to the even heavier beauty-quark sector, one can look at the production of Λ_b^0 baryons. As the label suggest, it has very similar quark contents as the Λ_c^+, namely *udb*. Another added benefit is that a major hadronic decay channel is $\Lambda_b^0 \rightarrow \Lambda_c^+ + \pi^-$. Hence, one can use the selected Λ_c^+ candidates and match an additional pion to form a Λ_b^0 vertex. The Λ_c^+ candidates from the previous sections before the BDT cut are used as a starting point for the Λ_b^0 analysis. This has the advantage of having decently selected tracks from the selection criteria, but not too strictly selected like the BDT would do. Additionally, one can include the Λ_c^+ candidate parameters in the BDT for the Λ_b^0 where it is used to optimize the analysis for this specific case.

Both the signal and background events used for the analysis originate from a $b\bar{b}$-enriched sample generated by PYTHIA and simulated by DELPHES[23, 24]. These events and their parameter distributions are assumed to be representative for both p-p and Pb-Pb collisions, up to the scaling factors obtained from theoretical yield for the signal and the minimum-bias sample for the background.

5.2 Selection criteria

Since the analysis is built upon the Λ_c^+ candidates from the previous sections, the selection criteria for those particles are inherited. Although rather simple, they induce the necessary requirements on the Λ_c^+ (from now on considered a daughter particle) and its decay products (from now on considered granddaughters). One such limitation is the reconstruction efficiency. Only selected Λ_c^+ candidates are considered in the construction of Λ_b^0 candidates. Hence, in low-p_T regions many true Λ_b^0 tracks do not have the opportunity to be reconstructed. For example, for momenta between 1.0 and 2.0 Gev/c, only one in two thousand was reconstructed. The full plot of the reconstruction efficiency as function of transverse momentum can be found in Appendix C. For the Λ_b^0 candidates, the selection criteria have been more advanced. In particular, more topological parameters have been included and cut on rather strictly. Figure 23 shows some of the initial parameter distribution of Λ_b^0 candidates in the transverse momentum range $5.0 < p_T < 7.0$ GeV/c. The distributions for the other bins are shown in Appendix B.

Most notable additions are the p_T dependant cuts on the pointing angle. Not only does the spatial resolution increase at higher values of momenta, as described in section 2, the tracks are also straighter. Hence, at high p_T the cosine of the pointing angle (Figure 23a) tends to be closer to unity for signal events, shown in red, compared to the background, depicted by the blue distribution. A progressively tighter cut on the CPA at higher momentum accommodates this topological property. Due to the nature of how vertices are found, it is likely that fake tracks or vertices are reconstructed from particles that were spatially close to each other, but do not originate from a true vertex. Hence, many background events are constructed from the abundance of particles close to the primary vertex. This is represented in the distribution of the decay length (Figure 23b), where the background contains a very sharp peak very close to zero. These candidates may come from the artifacts described which would yield extremely short tracks. By requiring the decay length of the candidate to be a minimum value, this source of background can be minimized.

The impact parameters (Figure 23e&f) of the daughters are now cut on as well, where both signal and background peak around zero. The relative fraction of candidates closer to zero is

higher for background, hence a minimum value is required for both daughters to remove the large peak in the background. The sum of the squared distances of closest approach of the daughters to the reconstructed secondary vertex at the optimal χ^2-fit, labeled as *Chi2PCA* (Figure 23h), also proved to be a successful addition to the selection criteria. Since the daughters of a true Λ_b^0 tracks in fact came from the decay vertex of said track, it is expected that the distance between the daughter tracks and the vertex is very small. Practically, this is only limited by the resolution of the detector. Although a very large portion of background events inherit the same property, it also contains a tail at larger values, which the true signal does not have. The full list of selection criteria used, including exact values, can be found in Appendix D.
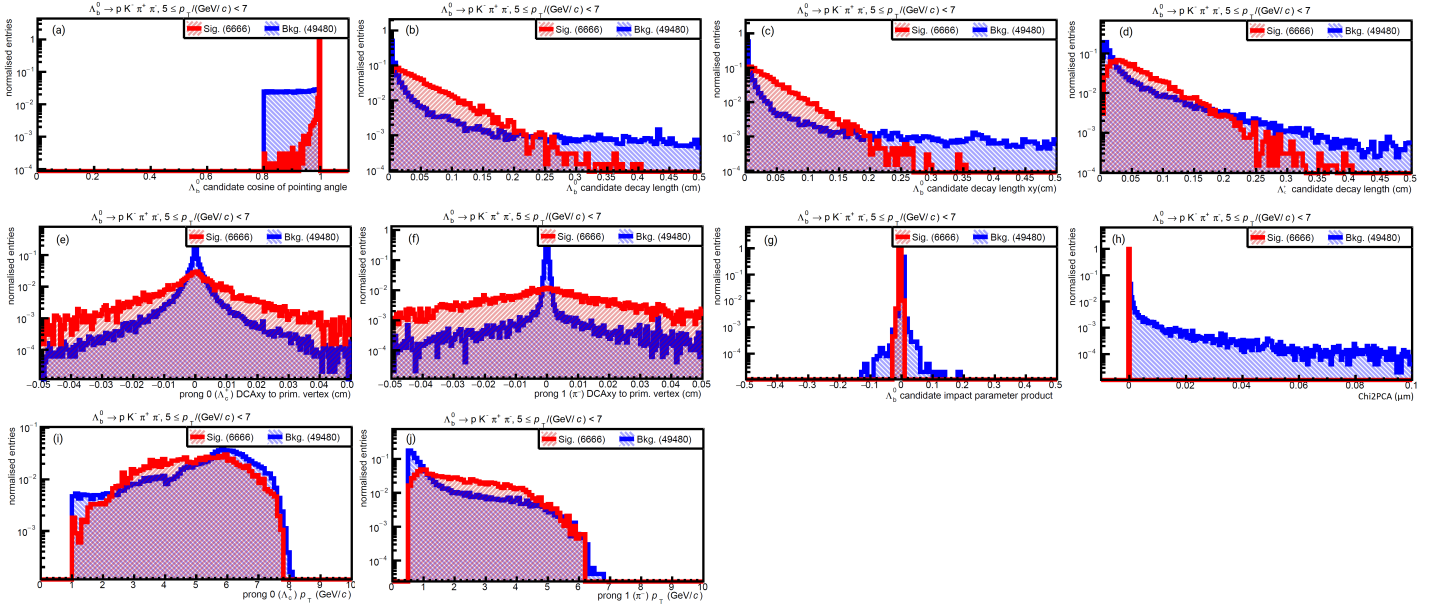
Figure 23: Some parameter distributions for the created Λ_b^0 candidates in the transverse momentum range $5.0 < p_T < 7.0$ GeV/c. a) cosine of the pointing angle, b) Λ_b^0 decay length, c) Λ_b^0 decay length in transverse plane, d) Λ_c^+ decay length, e) impact parameter of Λ_c^+, f) impact parameter of π^- daughter, g) product of the daughter's impact parameter, h) sum of the squared distances of closest approach of the daughters to the reconstructed secondary vertex at the optimal χ^2-fit, i) transverse momentum of Λ_c^+, j) transverse momentum of π^- daughter.

5.3 Yield from theory (PYTHIA/FONLL)

The yield can be computed similarly to the case of a charmed baryon, as in Equation 3 & Equation 4, where the cross-section is replaced by that of the beauty version. For a more thorough analysis and systematic uncertainty estimation it was decided to include an estimation for the cross-section from FONLL as well[34]. Since this provides the cross-section of the b-quark individually, additional factors are required. The number of obtained signal candidates per event in this case can be described by

$$\frac{N_{s,pp}}{N_{evt}} = 2 \cdot \Delta p_\mathrm{T} \cdot \Delta y \cdot BR(\Lambda_b^0 \to \Lambda_c^+ + \pi^-) \cdot BR(\Lambda_c^+ \to p + K^- + \pi^+) \cdot (Acc \times \epsilon)_{\Lambda_b^0} \cdot$$
$$(d\sigma/dp_\mathrm{T})_{|y|<1.44}^b \cdot f(b \to B) \cdot F(\Lambda_b^0/B) \quad (13)$$

$$\frac{N_{s,PbPb}}{N_{evt}} = \frac{N_{s,pp}}{N_{evt}} \cdot \langle N_{coll} \rangle \cdot R_{AA}^{\Lambda_b^0} \quad (14)$$

where N_s is the amount of expected accepted signal events, the factor 2 is there to include anti-particles, Δp_T is the binwidth in the transverse momentum range, Δy is the binwidth in the pseudorapidity range, $BR(...)$ are the respective branching ratios, $(d\sigma/dp_\mathrm{T})_{|y|<1.44}^b$ is the expected cross-section of b-quarks obtained from FONLL[34], $f(b \to B)$ is the fraction of b-quarks that end up in B-mesons[35], $F(\Lambda_b^0/B)$ is the ratio between Λ_b^0 baryons and B-mesons[36], $\langle N_{coll} \rangle$ is the mean number of nucleon collisions in the centrality region being investigated[26], and $R_{AA}^{\Lambda_b^0}$ is the nuclear modification factor for the respective baryon, for this analysis it is approximated by the R_{AA} of the Λ_c^+ from TAMU [15, 27].

Many of these factors have different values depending on the source (theoretical models, experiments etc.). In order to estimate the systematic uncertainty, variations of different sources have been considered. The final uncertainty was obtained as the quadratic sum of the corresponding individual contribution of each source. The list below summarizes the sources that have been considered

- For the cross-section $(d\sigma/dp_\mathrm{T})$, PYTHIA[23] and FONLL[34] are included, where their individually provided uncertainties are considered statistical uncertainties and the difference between them systematic,

- the branching ratios contain the different entries in the PDG listing, both the statistical and systematic uncertainties can be found in the reference [4],

- the uncertainties in the nuclear overlap function, $\left\langle T_{AA}^{30-50\%} \right\rangle$, are inherited from reference [26],

- the nuclear modification factor, R_{AA}, was assumed to be identical to that of the Λ_c^+, with uncertainties containing contributions from TAMU[27] and Catania as described in reference [15].

The resulting signal and background yield as a function of p_T can be seen in Figure 24 for p-p collisions and in Figure 25 for Pb-Pb collisions. By far the most dominant source of uncertainty in the signal yield is the difference between the cross-section predictions in PYTHIA and FONLL. Namely, PYTHIA consistently predicted value almost twice as high across the momentum range. Similar to the case of Λ_c^+, the signal yield is limited by the reconstruction efficiency in the low-p_T region and the exponentially decreasing number of particles at high-p_T. Surprisingly, the background yield is very similar for both collision systems. Naively one could expect there

Figure 24: The number of expected signal (left) and background (right) candidates per event in p-p collisions at $\sqrt{s} = 14.0 TeV$ after applying selection criteria.
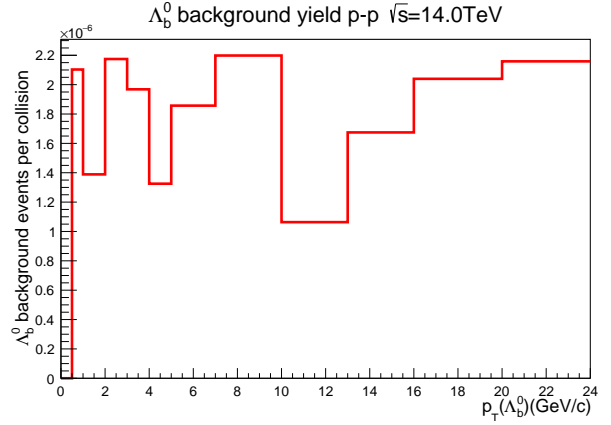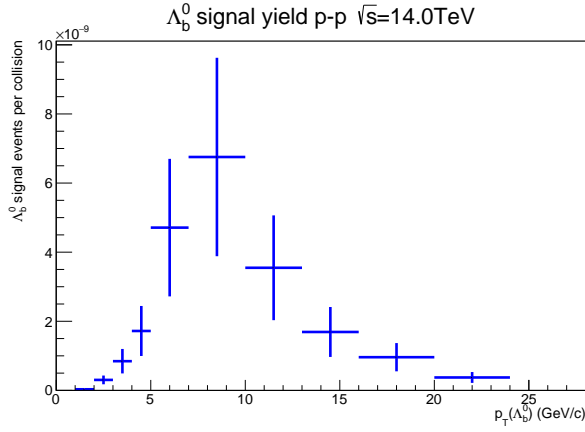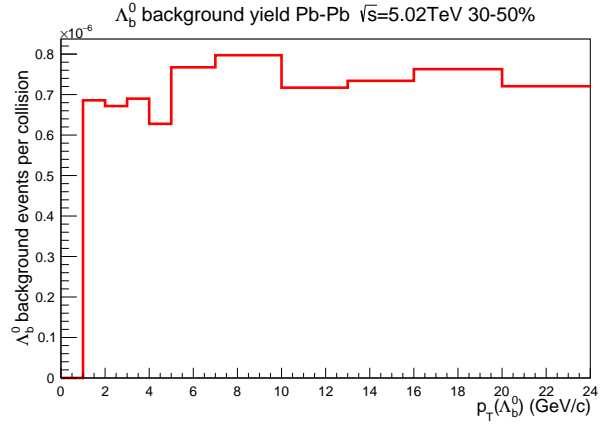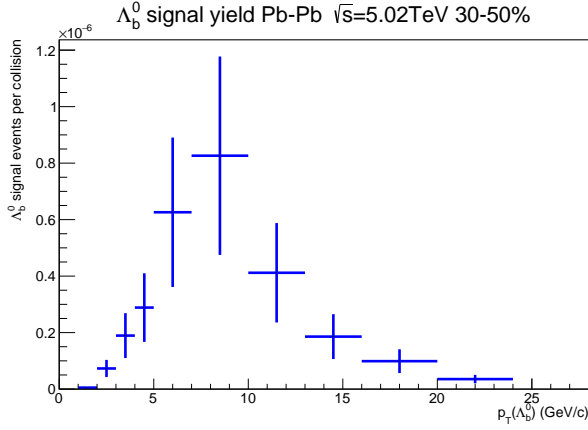

Figure 25: The number of expected signal (left) and background (right) candidates per event in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02 TeV$ after applying selection criteria.

to be more background in heavy ion collisions since the overall number particles in the final state is significantly higher. No satisfying explanation has been found yet and is worth further investigation.

5.4 BDT

The following parameters of the Λ_b^0 candidate itself are used to train the BDT

- CPA: Cosine of the pointing angle
- CPAXY: Cosine of the pointing angle in the transverse plane
- Chi2PCA: The sum of the squared distances of closest approach of the daughters to the reconstructed secondary vertex at the optimal χ^2-fit
- DecayLength: decay length of the Λ_b^0 candidate
- DecayLengthXY: decay length of the Λ_b^0 candidate in the transverse plane
- $d_0(\Lambda_c^+)$: impact parameter of the Λ_c^+
- $d_0(\pi^-)$: the impact parameter of the pion daughter

All these parameters were also part of the selection criteria and serves as a simple first estimate of what a BDT can add to the performance of the analysis. The parameter ranking for each BDT can be seen in Table 2. A higher ranking indicates the parameter was used more often and/or had more discriminating power. The adaptive and gradient boost seem to rank each parameter very similarly. Since they are based on the same principle, this may have been expected. The implementation that makes use of Fisher discriminants seems to pick up different patterns and ranks the pointing angle above the impact parameters. In all cases however, *Chi2PCA* ranks very low, a factor 10 less important than the second lowest. In BDTF it is not even used at all. This could indicate that the selective cut placed on this parameter earlier was either perfect or perhaps too strict. If the latter is the case, it may have removed too many candidates resulting in a loss of information, or if the former is true, all the relevant events are exactly within the accepted window. To test which of these is a valid argument, one can loosen the cut and re-train the BDTs with the new sample.

	BDTA		BDTG		BDTF	
Rank	Variable	Importance	Variable	Importance	Variable	Importance
1	$d_0(\Lambda_c^+)$	$2.88 \cdot 10^{-1}$	$d_0(\Lambda_c^+)$	$3.07 \cdot 10^{-1}$	CPA	$3.16 \cdot 10^{-1}$
2	$d_0(\pi^-)$	$2.08 \cdot 10^{-1}$	$d_0(\pi^-)$	$2.27 \cdot 10^{-1}$	CPAXY	$2.42 \cdot 10^{-1}$
3	DecayLength	$1.73 \cdot 10^{-1}$	DecayLength	$2.03 \cdot 10^{-1}$	$d_0(\Lambda_c^+)$	$1.43 \cdot 10^{-1}$
4	DecayLengthXY	$1.37 \cdot 10^{-1}$	DecayLengthXY	$1.55 \cdot 10^{-1}$	DecayLengthXY	$1.12 \cdot 10^{-1}$
5	CPA	$1.33 \cdot 10^{-1}$	CPA	$7.30 \cdot 10^{-2}$	DecayLength	$1.10 \cdot 10^{-1}$
6	CPAXY	$5.80 \cdot 10^{-2}$	CPAXY	$3.02 \cdot 10^{-2}$	$d_0(\pi^-)$	$7.68 \cdot 10^{-2}$
7	Chi2PCA	$3.77 \cdot 10^{-3}$	Chi2PCA	$4.88 \cdot 10^{-3}$	Chi2PCA	$0.00 \cdot 10^{+0}$

Table 2: Ranking of the Λ_b^0 candidate parameters used in the various BDT algorithms in the region $5.0 < p_T < 7.0$ GeV/c. The importance is a normalized value derived from the discrimination power of the parameter. A higher ranking indicates the parameter was used more often and/or had more discriminating power.

After training the BDTs, they are tested on a random subset of the data that was excluded from the training dataset. In order to check if overtraining did not occur, one can plot the BDT response distribution for both signal and background and compare the testing and training subsets. If the BDT was focused on a specific property that was randomly present in the training, but not in the testing sample, these distribution do not overlap. For instance, a large part of the distribution for the testing set could be incorrectly labeled as background, because it
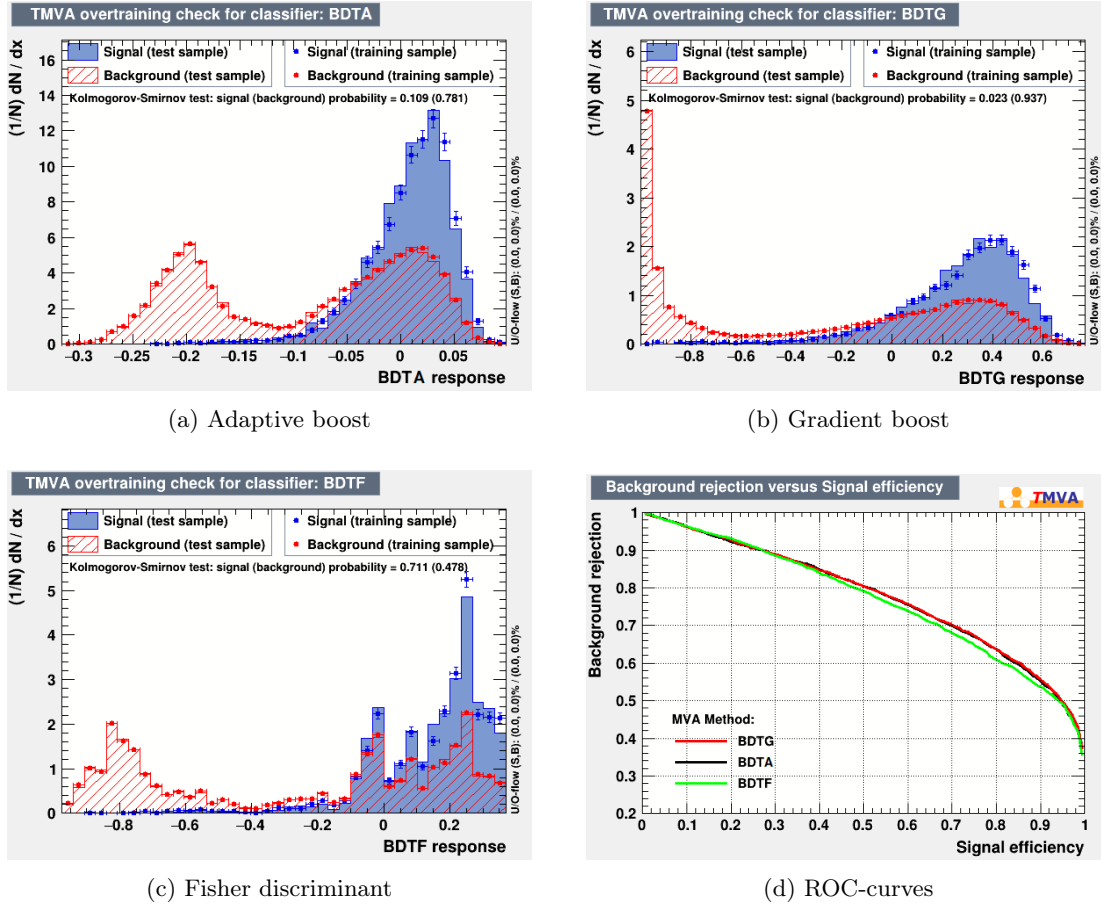
(a) Adaptive boost

(b) Gradient boost

(c) Fisher discriminant

(d) ROC-curves

Figure 26: a,b,c) BDT response distributions for training and testing samples for the three different implementations used. These are the results of training and testing on Λ_b^0 candidates in the region $5.0 < p_T < 7.0$ GeV/c using only 7 parameters as listed in Table 2. Figure (d) shows the resulting ROC-curves for the different variations

did not contain this specific feature, resulting in a lower BDT response. Figure 26 shows these distribution for the BDTs discussed previously. For BDTA and BDTG the peaks for signal events for the testing sample seems to be slightly lower than the training set, which draws some attention. The error bars (not visually included for test sample) are relatively large due to the limited number of events in the overall data sample, hence, this could be an unfortunate randomization of the two subsets. Either way, all three algorithms clearly have two peaks in the background distribution. This may suggest that there is a subset of the background that is easily separated by a single parameter cut, or correlation between two parameters. The other subset is very much signal-like to the point where it has the same BDT response distribution.

A metric to evaluate the performance of a BDT is the ROC-curve, also shown in Figure 26. This plots the background rejection (1−efficiency) as a function of the signal efficiency. This curve is obtained from evaluating the efficiencies at each possible BDT cut. The resulting curve describes all possible outcomes the BDT can obtain, independent on the number of events or scaling requirements. In a utopian case this curve would pass through the coordinates (1.0 ;

1.0), which would mean the BDT can reject all background while keeping all signal candidates as well. Realistically this will never be the case. Therefore, one way to evaluate the performance is by integrating these curves, which would yield a value between 0.5 and 1.0. A higher value approximately means better performance. Depending on the signal to background ratio and shape of the curve an optimal significance may be achieved at different BDT response cuts. Therefore, different ROC-curves can yield different results even though their integral would be the same. As the plot suggest, the integral of these three cases are fairly similar, namely 0.778 for BDTA, 0.779 for BDTG and 0.769 for BDTF, indicating their performance is on a similar level.

As described in section 4 as well, the different variations of the BDTs provide different results when scaling to the proper obtained yield from subsection 5.3. In this analysis, each BDT is evaluated on their own, with a respective BDT cut value and resulting significance. The average of the latter is used and the differences are used as uncorrelated systematic uncertainties in the significance. This way the final uncertainty in the significance can be described as the quadratic sum of propagated standard deviations.

$$
\begin{aligned}
\sigma^2_{significance} &= \frac{d^2}{dN_s^2} \frac{N_s}{\sqrt{N_s + N_B}} \cdot \sigma^2_{signalyield} + \frac{1}{N_{BDT}} \sum_i^{N_{BDT}} (\mu_{sign} - \mu_i)^2 \\
&= \left[\frac{2N_B + N_S}{2(N_B + N_S)^{3/2}} \right]^2 \cdot \sigma^2_{signalyield} + \frac{1}{N_{BDT}} \sum_i^{N_{BDT}} (\mu_{sign} - \mu_i)^2
\end{aligned}
\tag{15}
$$

where μ_{sign} is the mean significance, μ_i is the significance produced by BDT version i, and N_{BDT} is the number of BDT's used, in this case 3, and the first term on the right hand side is the derivative of the significance with respect to the signal yield. For the case of p-p collisions in the region $5.0 < p_T < 7.0$ GeV/c, the obtained significance as a function of BDT cut value is shown in Figure 27. The efficiency curves in these plots show the same behaviour as the BDT response distributions and the two peaks in the background can be observed as the two areas where the background efficiency decreases rapidly. Although the different algorithms ranked the parameters differently and underlying principles can be very much different, the resulting optimal significance is still within 2% of each other. At low- and high-p_T this difference increased, potentially due to the lower number of events present in those bins. Apart from one outlier, the spread was typically 5 to 13%.

A comparison with the baseline of just applying the selection criteria can be made to check to impact of the BDTs on the performance. This is illustrated in Figure 28 by plotting the obtained significance (data points) as a function of transverse momentum. The red line indicates the state after the selection criteria. In both collision types the best improvement is in the low- and high-p_T regions, where the significance generally increases by a factor 3. Typically the signal to background ratio is lower in these regions and the BDTs have higher values for the integral of the ROC-curve in low-p_T regions (plots of the integral as function of p_T, including comments on the interpretation, can be found in Appendix F). Therefore, there is relatively more background to get rid of and they are better at doing so. In the intermediate region, although the significance peaks here, the signal to background ratio and the absolute number of signal candidates accepted result in a reduced improvement by adding the BDT cut. The systematic uncertainty, represented by blue boxes, is fairly large, especially in proton-proton collisions. The dominant factor in that is the uncertainty in the Λ_b^0 cross-section, which, due to the signal to background ratio in these collisions, results in a larger relative uncertainty in p-p as compared to Pb-Pb in the significance.
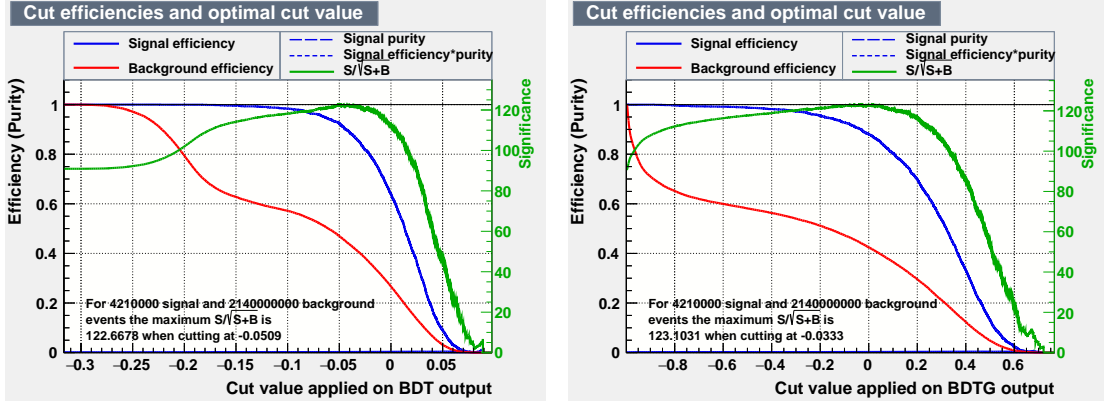
(a) Adaptive boost

(b) Gradient boost
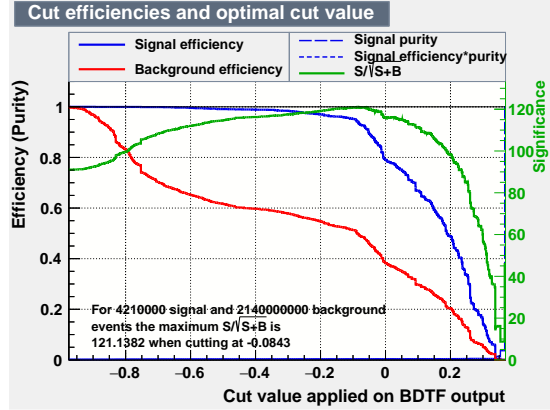
(c) Fisher discriminant

Figure 27: Significance as a function of the BDT cut for the three different implementations used. These are the results of training and testing on Λ_b^0 candidates in the region $5.0 < p_{\mathrm{T}} < 7.0$ GeV/c using only 7 parameters as listed in Table 2. The number of signal and background events is scaled to the expected number in p-p collisions as shown in Figure 24.
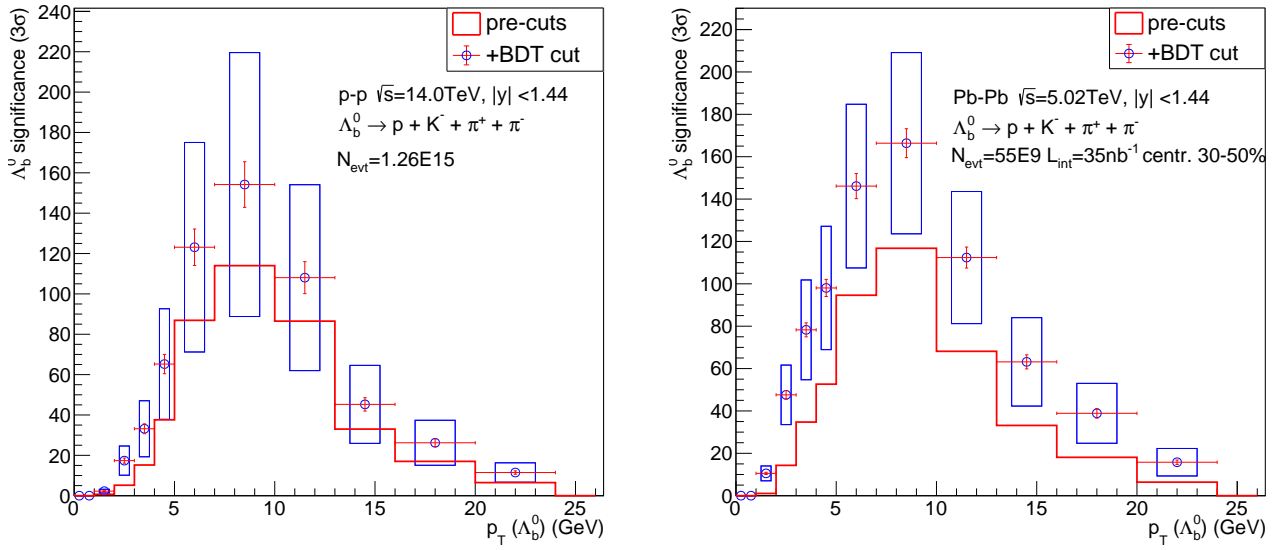
Figure 28: The expected significance for Λ_b^0 obtained from applying the selection criteria (red line) and after applying the additional BDT cut(data points) for p-p collision (left) and Pb-Pb collisions (right). The blue boxes represent the systematic uncertainties while the red cross represents the statistical uncertainty. The error bars for the case of only pre-cuts have been omitted for visual clarity, an inclusive plot can be seen in Appendix E.

5.5 Inclusion of Λ_c^+ parameters and PID information

In order to further improve the analysis, one can add the granddaughter's info and particle identification information (PID). As discussed earlier, this includes the parameters that were used for the BDT of the Λ_c^+, to be used for optimizing the Λ_b^0 analysis. Additionally, the invariant mass of the Λ_c^+ is included. For the sake of the analysis, the previous cut was very loose such that the background distribution can be established in the region surrounding the signal peak. For the purpose of the Λ_b^0 analysis, this is no longer required, hence it is possible to more tightly cut on the charmed daughter's mass. It was left completely up to the BDT how to use this freedom instead of manually narrowing the mass window. As a consequence, the separation power of this parameter in the BDT is very high, however, since an equivalent cut could perhaps be done by hand as well, this should be taken with a grain of salt.

The four final state particles (the pion from the initial decay and the three decay products of the Λ_c^+) are all suitable particles for the PID systems present in ALICE, namely the Time Of Flight (TOF) and Ring-Imaging Cherenkov (RICH) information is added as parameters for the BDT. Both give a number of standard deviations that the track is away from the ideal PID fit for the respective detectors. In a sense it depict how convinced each system is what the true nature of these tracks are. Not every track has an assigned value from each subdetector and in particular, many tracks were unable to obtain measurements from the RICH. Instead some meaningless number is used instead to fill an otherwise empty gap. For some reason about 30% of the granddaughters and about 75% of the π^- daughters did not register a hit and got the meaningless number assigned. It meant that the Fisher discriminant's implementation (BDTF), was unable to perform proper training due to reaching many situations where there is zero variance in the parameter distributions. This is very much an issue caused by the implementation that most likely can be improved upon. Hence, two situations are considered, the first containing all three BDTs and no RICH information, and the second only using BDTA and BDTG with RICH information included.

Correlations between the parameters of candidates in the region $5.0 < p_T < 7.0$ GeV/c are shown in Figure 29 for signal and Figure 30 for background candidates. PID parameters are labeled with either TOF() or RICH() with the particle inside the brackets. Since the decay products of the Λ_c^+ contain two positively charged particles, the numbering of the tracks may be different for each candidate when they are produced in the candidate creator (eg. $pK^-\pi^+$ or π^+K^-p). These tracks will be evaluated both with respect to a proton and a pion and are labeled as Trk1 and Trk2. The issue with missing RICH hits cascades into the correlations. If multiple tracks in one event end up without a hit in the RICH, they get the same value, inducing a correlation that has no physical meaning. This potentially masks a meaningful correlation between parameters that could otherwise be useful for the analysis. The two different evaluations of the positive granddaughter's tracks obviously have a very strong correlation, since they originate from the same source, just different representations. Similarly, some parameters are based on 3 dimensional vectors, like the decay length, and contain an additional projection in the transverse plane, DecayLengthXY. As expected, the correlation is very strong between these parameters as one is just the other parameter plus a free component in the z-direction.

A more interesting correlation is between the impact parameters of the daughters with that of the granddaughters. A common factor in the determination of these values is the vertex fitting. Hence, a different location for the vertex would directly affect all impact parameters. This correlation is not as strong in the background. If a background candidate originate from a true vertex as well, just a different particle species, the correlation should still exist. However, random combination of tracks that reconstruct a non-existing vertex would not have this correlation.
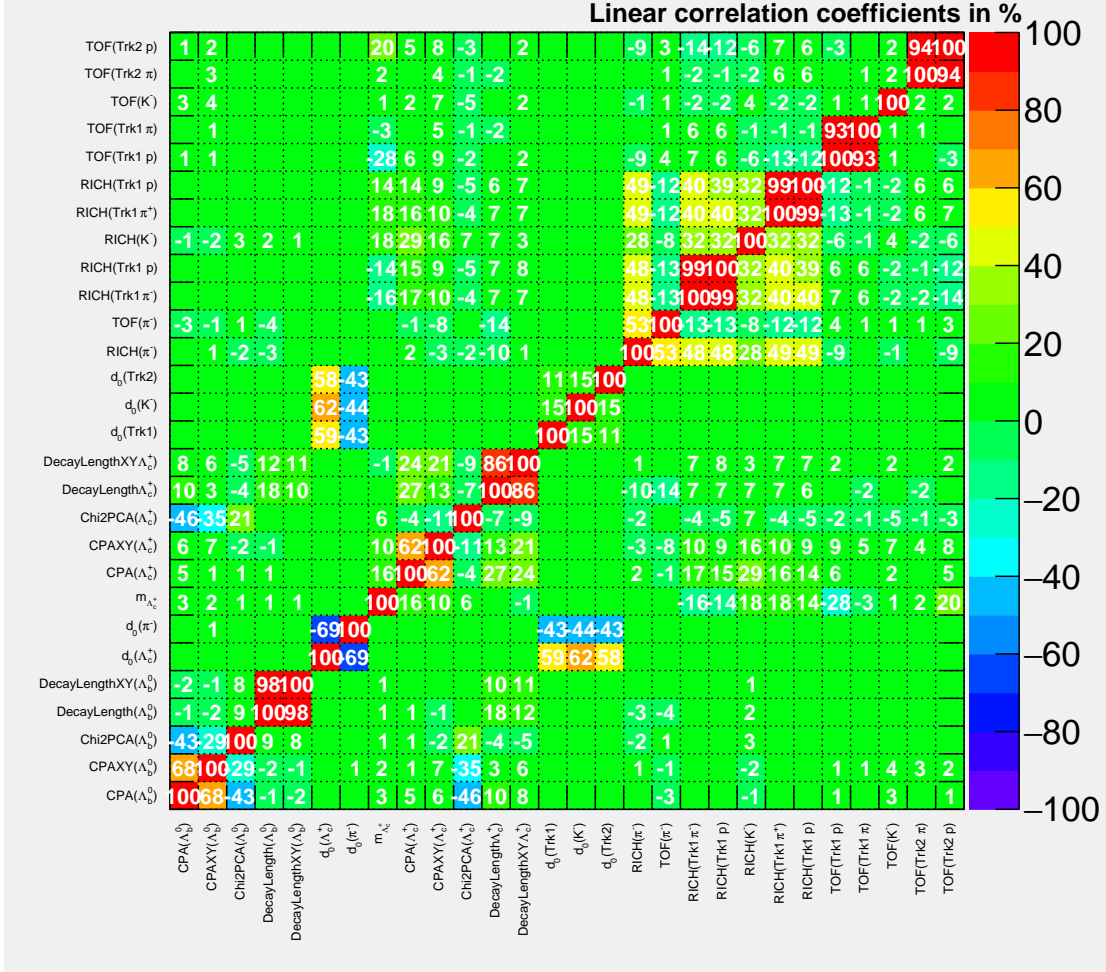
Perhaps the difference between the correlation in signal and background could indicate the number of combinatorial background candidates. The sum of distances of closest approach (chi2PCA), however, shows a decently strong correlation with the pointing angle. Surprising the Chi2PCA of the Λ_c^+ also contains this correlation with the pointing angle of the Λ_b^0 for background candidates. This could potentially be explained by some source of background and is a useful property that can be exploited in the BDT algorithms.

Figure 29: Correlation matrix of the input parameters of true signal Λ_b^0 candidates in the region $5.0 < p_T < 7.0$ GeV/c.

Unfortunately for the list of events with these additional set of parameters, a limited number of candidates were available for training and testing of the BDTs. Especially in the region $1.0 < p_T < 2.0$ GeV/c, containing only 20 signal events, this strongly limited the potential of the analysis. The resulting parameter ranking after training on this longer list of parameters for the case without RICH and with RICH included can be found in Table 3 and Table 4 respectively. The patterns from the case with only the topological parameters of the Λ_b^0 candidate can be observed here as well. BDTF highly ranks the pointing angle, even above the Λ_c^+ mass, and the impact parameters are more important for the other two algorithms. BDTA and BDTG

Figure 30: Correlation matrix of the input parameters of background Λ_b^0 candidates in the region $5.0 < p_{\mathrm{T}} < 7.0$ GeV/c.

however, seem to benefit more from adding the Λ_c^+ parameters as they are ranked very high in all 4 cases. General consensus between all the BDTs discussed here is that the time of flight information of the proton should not be disregarded, ranking at least in the top 4 in all cases. When adding the RICH information, the proton information ranks relatively high again. Overall the RICH information seems to add good separation power to the systems, even though not all tracks have a hit in this detector.

Rank	BDTA Variable	Importance	BDTG Variable	Importance	BDTF Variable	Importance
1	$m_{\Lambda_c^+}$	$1.06 \cdot 10^{-1}$	$m_{\Lambda_c^+}$	$1.40 \cdot 10^{-1}$	CPA(Λ_b^0)	$1.93 \cdot 10^{-1}$
2	TOF(Trk2 p)	$9.12 \cdot 10^{-2}$	CPA(Λ_c^+)	$7.30 \cdot 10^{-2}$	$m_{\Lambda_c^+}$	$1.17 \cdot 10^{-1}$
3	CPA(Λ_c^+)	$7.64 \cdot 10^{-2}$	TOF(Trk2 p)	$6.94 \cdot 10^{-2}$	DecayLengthXY(Λ_c^+)	$9.93 \cdot 10^{-2}$
4	DecayLengthXY(Λ_c^+)	$6.24 \cdot 10^{-2}$	DecayLength(Λ_c^+)	$6.34 \cdot 10^{-2}$	TOF(Trk2 p)	$8.89 \cdot 10^{-2}$
5	DecayLength(Λ_c^+)	$5.72 \cdot 10^{-2}$	DecayLengthXY(Λ_c^+)	$6.03 \cdot 10^{-2}$	DecayLengthXY(Λ_b^0)	$7.36 \cdot 10^{-2}$
6	$d_0(K^-)$	$5.69 \cdot 10^{-2}$	$d_0(K^-)$	$5.89 \cdot 10^{-2}$	DecayLength(Λ_b^0)	$6.69 \cdot 10^{-2}$
7	d_0(Trk2)	$5.24 \cdot 10^{-2}$	TOF(Trk1 p)	$5.49 \cdot 10^{-2}$	Chi2PCA(Λ_c^+)	$6.18 \cdot 10^{-2}$
8	$d_0(\Lambda_c^+)$	$5.21 \cdot 10^{-2}$	d_0(Trk1)	$5.46 \cdot 10^{-2}$	DecayLength(Λ_c^+)	$6.13 \cdot 10^{-2}$
9	d_0(Trk1)	$5.16 \cdot 10^{-2}$	$d_0(\Lambda_c^+)$	$5.06 \cdot 10^{-2}$	TOF(Trk2 π^+)	$5.19 \cdot 10^{-2}$
10	TOF(Trk1 p)	$5.08 \cdot 10^{-2}$	d_0(Trk2)	$4.90 \cdot 10^{-2}$	CPAXY(Λ_b^0)	$4.68 \cdot 10^{-2}$
11	CPA(Λ_b^0)	$4.88 \cdot 10^{-2}$	TOF(Trk2 π^+)	$4.54 \cdot 10^{-2}$	TOF(Trk1 p)	$2.83 \cdot 10^{-2}$
12	CPAXY(Λ_c^+)	$4.79 \cdot 10^{-2}$	$d_0(\pi^-)$	$4.47 \cdot 10^{-2}$	$d_0(\Lambda_c^+)$	$2.46 \cdot 10^{-2}$
13	TOF(Trk2 π^+)	$4.63 \cdot 10^{-2}$	CPAXY(Λ_c^+)	$4.47 \cdot 10^{-2}$	CPA(Λ_c^+)	$2.19 \cdot 10^{-2}$
14	$d_0(\pi^-)$	$3.72 \cdot 10^{-2}$	DecayLength(Λ_b^0)	$3.30 \cdot 10^{-2}$	TOF(Trk1 π^+)	$1.73 \cdot 10^{-2}$
15	TOF(Trk1 π^+)	$3.25 \cdot 10^{-2}$	CPA(Λ_b^0)	$2.97 \cdot 10^{-2}$	d_0(Trk1)	$1.62 \cdot 10^{-2}$
16	DecayLength(Λ_b^0)	$2.78 \cdot 10^{-2}$	DecayLengthXY(Λ_b^0)	$2.69 \cdot 10^{-2}$	TOF(π^-)	$9.84 \cdot 10^{-3}$
17	TOF(π^-)	$2.60 \cdot 10^{-2}$	TOF(Trk1 π^+)	$2.61 \cdot 10^{-2}$	d_0(Trk2)	$8.31 \cdot 10^{-3}$
18	DecayLengthXY(Λ_b^0)	$2.55 \cdot 10^{-2}$	Chi2PCA(Λ_c^+)	$2.39 \cdot 10^{-2}$	$d_0(K^-)$	$7.90 \cdot 10^{-3}$
19	CPAXY(Λ_b^0)	$1.95 \cdot 10^{-2}$	TOF(π^-)	$2.26 \cdot 10^{-2}$	$d_0(\pi^-)$	$4.50 \cdot 10^{-3}$
20	Chi2PCA(Λ_c^+)	$1.95 \cdot 10^{-2}$	CPAXY(Λ_b^0)	$1.77 \cdot 10^{-2}$	Chi2PCA(Λ_b^0)	$9.80 \cdot 10^{-4}$
21	TOF(K^-)	$1.16 \cdot 10^{-2}$	TOF(K^-)	$1.18 \cdot 10^{-2}$	CPAXY(Λ_c^+)	$3.87 \cdot 10^{-6}$
22	Chi2PCA(Λ_b^0)	$0.00 \cdot 10^{+0}$	Chi2PCA(Λ_b^0)	$0.00 \cdot 10^{+0}$	TOF(K^-)	$5.71 \cdot 10^{-11}$

Table 3: Ranking of the Λ_b^0 candidate parameters used in the various BDT algorithms in the region $5.0 < p_\mathrm{T} < 7.0$ GeV/c. This version contains the topological parameters of the Λ_b^0 candidate itself, TOF information of the final state particles, and the Λ_c^+ candidate parameters. The importance is a normalized value derived from the discrimination power of the parameter. A higher ranking indicates the parameter was used more often and/or had more discriminating power.

	BDTA		BDTG	
Rank	Variable	Importance	Variable	Importance
1	$m_{\Lambda_c^+}$	$9.22 \cdot 10^{-2}$	$m_{\Lambda_c^+}$	$1.08 \cdot 10^{-1}$
2	TOF(Trk2 p)	$6.91 \cdot 10^{-2}$	CPA(Λ_c^+)	$5.62 \cdot 10^{-2}$
3	CPA(Λ_c^+)	$6.39 \cdot 10^{-2}$	TOF(Trk2 p)	$5.19 \cdot 10^{-2}$
4	DecayLengthXY(Λ_c^+)	$4.95 \cdot 10^{-2}$	$d_0(K^-)$	$4.64 \cdot 10^{-2}$
5	$d_0(K^-)$	$4.78 \cdot 10^{-2}$	DecayLength(Λ_c^+)	$4.57 \cdot 10^{-2}$
6	DecayLength(Λ_c^+)	$4.47 \cdot 10^{-2}$	DecayLengthXY(Λ_c^+)	$4.56 \cdot 10^{-2}$
7	d_0(Trk1)	$4.25 \cdot 10^{-2}$	d_0(Trk1)	$4.48 \cdot 10^{-2}$
8	CPA(Λ_b^0)	$4.23 \cdot 10^{-2}$	RICH(Trk2 p)	$4.30 \cdot 10^{-2}$
9	$d_0(\Lambda_c^+)$	$4.14 \cdot 10^{-2}$	RICH(Trk2 π^+)	$4.12 \cdot 10^{-2}$
10	d_0(Trk2)	$3.89 \cdot 10^{-2}$	TOF(Trk1 p)	$4.12 \cdot 10^{-2}$
11	TOF(Trk1 p)	$3.77 \cdot 10^{-2}$	$d_0(\Lambda_c^+)$	$4.12 \cdot 10^{-2}$
12	RICH(Trk2 p)	$3.74 \cdot 10^{-2}$	TOF(Trk2 π^+)	$4.06 \cdot 10^{-2}$
13	RICH(Trk2 π^+)	$3.64 \cdot 10^{-2}$	d_0(Trk2)	$3.89 \cdot 10^{-2}$
14	CPAXY(Λ_c^+)	$3.61 \cdot 10^{-2}$	RICH(π^-)	$3.74 \cdot 10^{-2}$
15	TOF(Trk1 π^+)	$3.38 \cdot 10^{-2}$	RICH(Trk1 p)	$3.69 \cdot 10^{-2}$
16	$d_0(\pi^-)$	$3.29 \cdot 10^{-2}$	RICH(Trk1 π^+)	$3.41 \cdot 10^{-2}$
17	RICH(Trk1 p)	$3.26 \cdot 10^{-2}$	CPAXY(Λ_c^+)	$3.40 \cdot 10^{-2}$
18	TOF(Trk2 π^+)	$3.24 \cdot 10^{-2}$	$d_0(\pi^-)$	$3.06 \cdot 10^{-2}$
19	RICH(Trk1 π^+)	$3.23 \cdot 10^{-2}$	RICH(K^-)	$2.74 \cdot 10^{-2}$
20	RICH(K^-)	$2.83 \cdot 10^{-2}$	TOF(Trk1 π^+)	$2.72 \cdot 10^{-2}$
21	RICH(π^-)	$2.56 \cdot 10^{-2}$	DecayLength(Λ_b^0)	$2.52 \cdot 10^{-2}$
22	CPAXY(Λ_b^0)	$2.09 \cdot 10^{-2}$	CPA(Λ_b^0)	$2.11 \cdot 10^{-2}$
23	TOF(π^-)	$2.02 \cdot 10^{-2}$	DecayLengthXY(Λ_b^0)	$2.03 \cdot 10^{-2}$
24	DecayLength(Λ_b^0)	$1.97 \cdot 10^{-2}$	Chi2PCA(Λ_c^+)	$1.92 \cdot 10^{-2}$
25	Chi2PCA(Λ_c^+)	$1.83 \cdot 10^{-2}$	TOF(π^-)	$1.86 \cdot 10^{-2}$
26	DecayLengthXY(Λ_b^0)	$1.28 \cdot 10^{-2}$	TOF(K^-)	$1.53 \cdot 10^{-2}$
27	TOF(K^-)	$1.05 \cdot 10^{-2}$	CPAXY(Λ_b^0)	$8.53 \cdot 10^{-3}$
28	Chi2PCA(Λ_b^0)	$0.00 \cdot 10^{+0}$	Chi2PCA(Λ_b^0)	$0.00 \cdot 10^{+0}$

Table 4: Ranking of the Λ_b^0 candidate parameters used in the various BDT algorithms in the region $5.0 < p_{\mathrm{T}} < 7.0$ GeV/c. This version contains the topological parameters of the Λ_b^0 candidate itself, TOF and RICH information of the final state particles, and the Λ_c^+ candidate parameters. The importance is a normalized value derived from the discrimination power of the parameter. A higher ranking indicates the parameter was used more often and/or had more discriminating power.

Although not individually included here, each p_{T} bin has their own set of rankings. Due to the sequential evolutionary nature of BDT algorithms, a slight change in parameter distributions caused by random selection of training subsamples can lead to different boosts in the first few iteration, which in turn impacts all future generations. Hence, in the end, the ranking of parameters varies. This uncertainty is especially evident for parameters in the middle of the list, where the values for the importance are very close. Similarly, this effect spans across p_{T} bins, such that the same algorithm can produce significantly different rankings for neighbouring bins. In order to still obtain a first guess at the p_{T} dependence of the importance of certain parameters, a comparison was made by eye. At high p_{T} the PID information loses some importance while the decay length and impact parameters of the Λ_b^0 increase in this region. In the lower p_{T} regions, the variations seem to be more intense because of which no significantly consistent pattern could
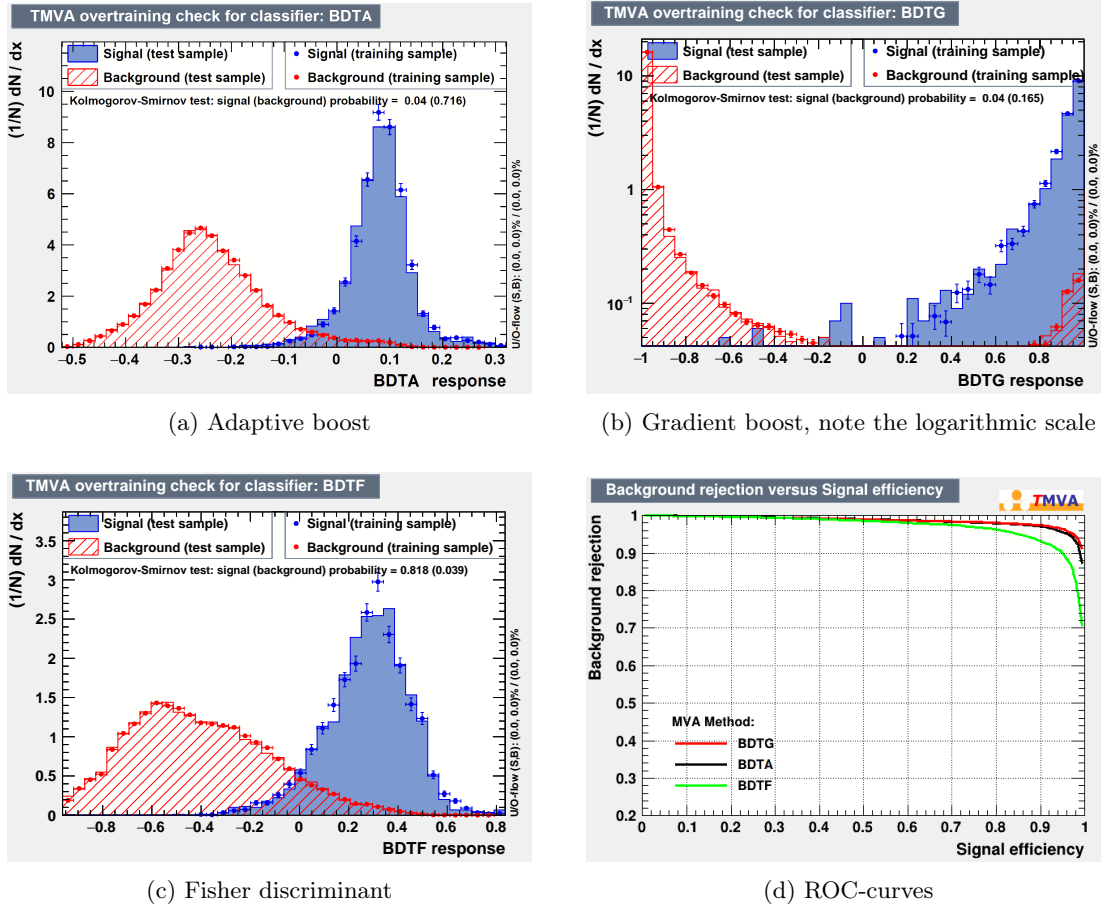
(a) Adaptive boost

(b) Gradient boost, note the logarithmic scale

(c) Fisher discriminant

(d) ROC-curves

Figure 31: a,b,c) BDT response distributions for training and testing samples for the three different implementations used. These are the results of training and testing on Λ_b^0 candidates in the region $5.0 < p_T < 7.0$ GeV/c using the 22 parameters as listed in Table 3. Figure (d) shows the resulting ROC-curves for the different variations.

be observed. Overall, the rankings seem fairly similar to the mid-p_T list shown in the tables here.

The improved BDT response distributions can be seen in Figure 31 for the case with no RICH information included. The most notable difference when comparing with the previous iteration is the background distributions. The second peak that was overlapping greatly with the signal has been reduced significantly to a point where it is more like a tail of the first peak. Only in the BDTG distribution is it distinctly noticeable as a separate peak, however, logarithmic scaling on the vertical axis was required to notice this. Adding the RICH information did not change the profile of these plots, it only slightly increased the separation.

These BDTs were scaled accordingly and their resulting combined significance has been plotted as a function of transverse momentum in Figure 32 (without RICH) and Figure 33 (with RICH included). In both these cases the significance in proton-proton collisions was improved by up to a factor 4. The increased power to correctly classify background is a major improvement in this case due to the low signal to background ratio ($\approx 1/1000$). In the Pb-Pb collisions, the signal
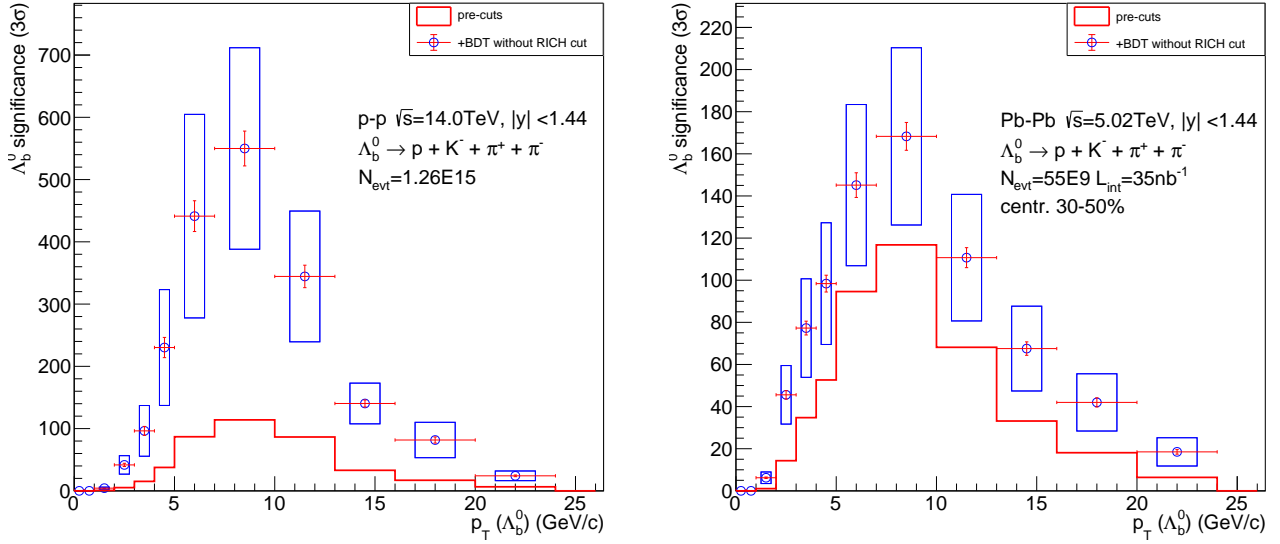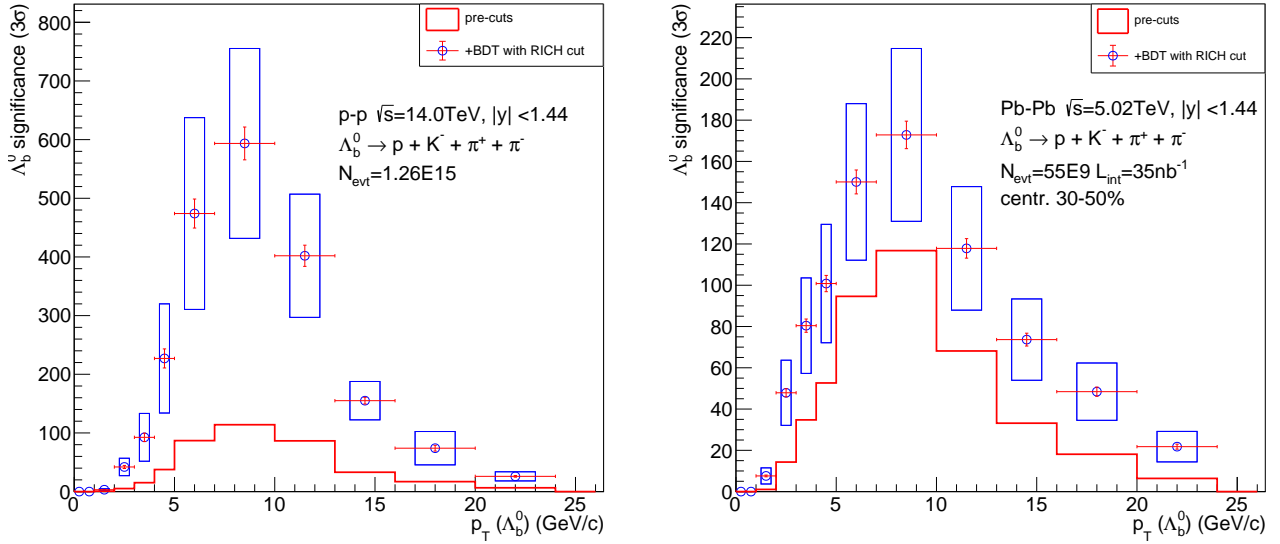
Figure 32: The expected significance for Λ_b^0 obtained from applying the selection criteria (red line) and after applying the additional BDT cut (data points) for p-p collision (left) and Pb-Pb collisions (right). The BDTs used the parameters as listed in Table 3. The blue boxes represent the systematic uncertainties while the red cross represents the statistical uncertainty. The error bars for the case of only pre-cuts have been omitted for visual clarity, an inclusive plot can be seen in Appendix E.

to background ratio is much higher, close to unity in many p_T regions. Combined with the fact that Run5 expects much fewer heavy ion collisions, the significance was mostly limited by the number of signal events, even after applying just the selection criteria. After applying the cuts of these new BDTs in Pb-Pb events, the signal to background ratio is close to one in the lowest and highest p_T bins and increased up to nearly 20 between 7.0 and 10.0 GeV/c. A comparison between the different analyses is shown in Figure 34. In many cases, the significance in Pb-Pb in only slightly higher than the short list of input parameters. In the region $1.0 < p_T < 2.0$ GeV/c, the inclusion of additional parameters resulted in a lower significance. However, this can be attributed to reduced number of signal event that could be used for training and testing.

Figure 33: The expected significance for Λ_b^0 obtained from applying the selection criteria (red line) and after applying the additional BDT cut (data points) for p-p collision (left) and Pb-Pb collisions (right). The BDTs used the parameters as listed in Table 4. The blue boxes represent the systematic uncertainties while the red cross represents the statistical uncertainty. The error bars for the case of only pre-cuts have been omitted for visual clarity, an inclusive plot can be seen in Appendix E.
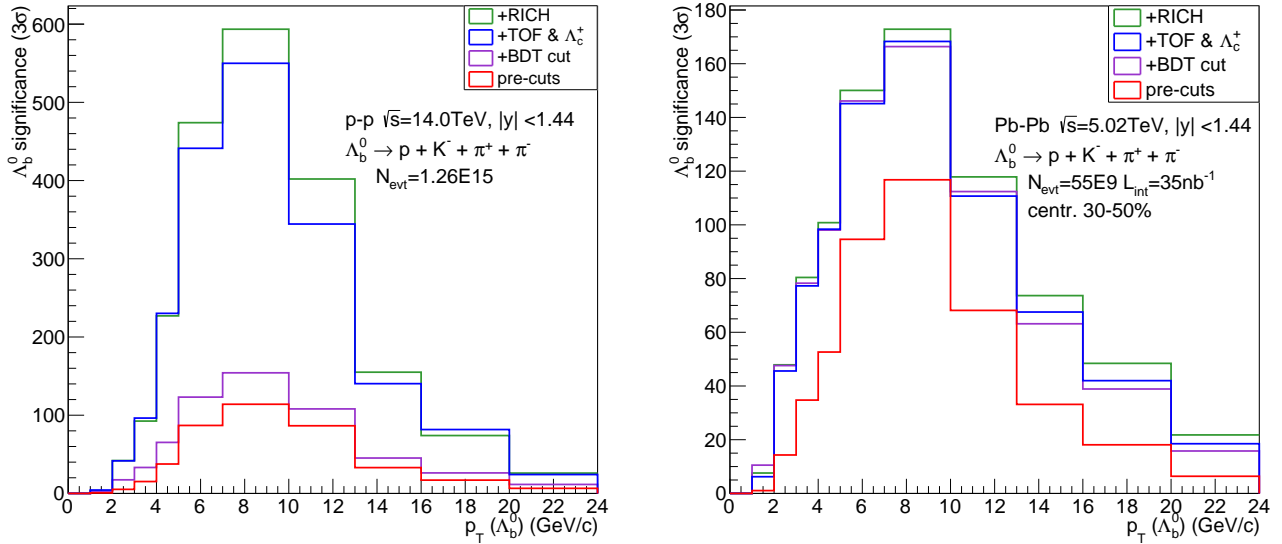
Figure 34: Comparison of the resulting significance of the different analysis methods. Error bars have been omitted for visual clarity.

6 Conclusion

In summary, this thesis presented a feasibility study on the Λ_b^0 analysis with the proposed ALICE3 experimental setup. From $b\bar{b}-$enriched proton-proton collision samples generated by PYTHIA and simulated by DELPHES[23, 24], candidates are reconstructed based on the following decay channel: $\Lambda_b^0 \to \pi^- + \Lambda_c^+ \to \pi^- + p + K^- + \pi^+$. Based on topological and kinematic selection criteria, candidates were selected with the goal to reduce the number of background events. Selected candidates were used to train and test multiple boosted decision trees in order to further separate background from signal. The number of events was scaled according to theoretical yields for the signal, and minimum bias samples for background candidates. The machine learning algorithms are applied as a final step to optimize for the statistical significance of the signal.

Different sets of input parameters for the BDTs have been compared to study the impact of an increased parameter space. For the measurement in proton-proton collisions the addition of Λ_c^+ candidate parameters and PID information of the granddaughters was shown to have a great impact on the power to correctly classify background events. Due to the signal to background ratio in this type of collision, this greatly improved to obtained significance. In lead-lead collisions, this ratio was much higher, by several order of magnitude, hence, suppression of the background was not the limiting factor in the potential significance. Instead, the reconstruction efficiency of signal events was the main bottleneck. This could potentially be improved upon by loosening the selection criteria, to recover some of the lost signal events. The BDTs have shown to be capable in rejecting background events when the amount of information provided is sufficient. Therefore, the drastic increase of background candidates obtained from these new criteria may not be an issue and the resulting significance may increase as a consequence.

Despite the relatively low hitcount, the RICH information tends to lead to an increase in performance as addition to the TOF regarding particle identification of the final state particles. If the cause of this effect could be identified and perhaps improved upon, the RICH can surpass the TOF in importance for this specific decay. However, from the results obtained here, the TOF seems more reliable and nearly equivalent in performance for the Λ_b^0 analysis.

The potential of the machine learning algorithms is most likely not fully revealed. Especially in low-p_T bins, the number of available signal events for training was very low. Not only does this reduce the opportunity for the algorithms to find unique and useful patterns, it raises concern about the validity of this datasample being representative of the full spectrum. Additionally, the assumption was made that the parameter distributions and correlations for the candidates generated in p-p collisions are also valid for those in Pb-Pb collisions. Intuitively this seems valid since the Λ_b^0 has a typical decay time considerably larger than the kinematic freeze-out time of the QGP. All the parameters used in this analysis are based on the decay products of the Λ_b^0, which originate from the free streaming phase and therefore should not be affected much by the type of initial collision. However, this was not investigated and may be of relevance.

Overall, it can be concluded that in a large momentum range the Λ_b^0 can be measured pretty well in fine p_T bins. At low p_T, several improvements can still be made and different analysis strategies are to be explored if one so desires.

To further study the physics that ALICE3 can explore, it is very interesting to look at, for instance, the B_s^0 and B_c^+ mesons. The mass ordering of the elliptic flow currently measured for light particles is yet be tested for third generation quarks. Comparison of the Λ_b^0 and the B mesons can give further insight in the validity of current hydrodynamical models or give rise to undiscovered mechanisms. To facilitate this, one could perform a similar analysis for

the B mesons as is performed for the Λ_b^0 here. Based on these analyses, the elliptic flow can be estimated, as shown in Figure 35 for the Λ_b^0. Additionally, in this plot a comparison is made between the expected measurements with ALICE3, based on the result in this thesis, and the expected results during run3 (ITS2) and run4 (ITS3). ALICE3 extends the momentum range where a measurement is considered to 1 GeV/c on the lower end and up to 24 GeV/c on the higher end. Despite the finer binning, as compared to ITS2&3, the uncertainties can be drastically reduced. Consequentially, the physics of heavy flavour quarks in a QGP can be explored in great detail.
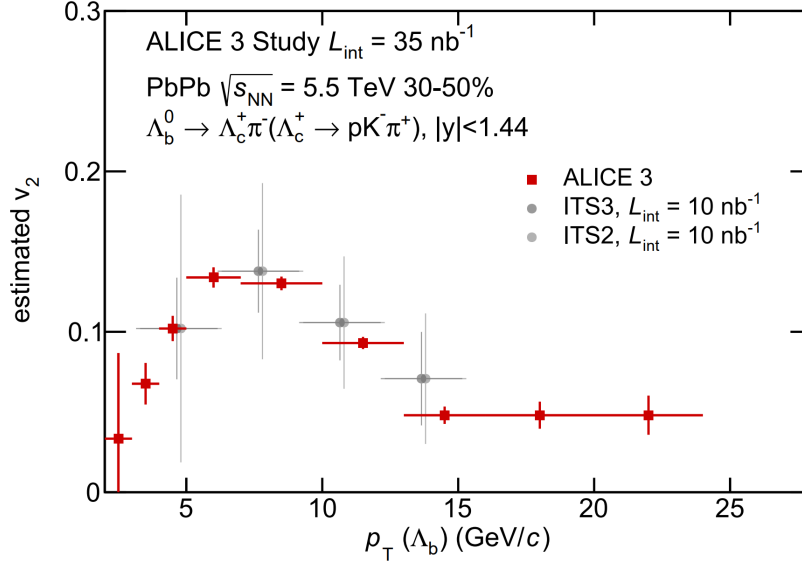
Figure 35: Estimated elliptic flow (v_2) of Λ_b^0 in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02 TeV$.

7 Acknowledgements

References

[1] C. T. H. Davies et al. "High precision lattice QCD confronts experiment". In: *Phys. Rev. Lett.* 92 (2004), p. 022001. DOI: 10.1103/PhysRevLett.92.022001. arXiv: hep-lat/0304004.

[2] Johann Rafelski. "Melting Hadrons, Boiling Quarks". In: *Eur. Phys. J. A* 51.9 (2015), p. 114. DOI: 10.1140/epja/i2015-15114-0. arXiv: 1508.03260 [nucl-th].

[3] Chun Shen. *Heavy ion collision evolution*. URL: https://u.osu.edu/vishnu/ (visited on 02/17/2022).

[4] P.A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[5] Francesco Prino and Ralf Rapp. "Open Heavy Flavor in QCD Matter and in Nuclear Collisions". In: *J. Phys. G* 43.9 (2016), p. 093002. DOI: 10.1088/0954-3899/43/9/093002. arXiv: 1603.00529 [nucl-ex].

[6] Shreyasi Acharya et al. "Inclusive quarkonium production in pp collisions at $\sqrt{s} = 5.02$ TeV". In: (Sept. 2021). arXiv: 2109.15240 [nucl-ex].

[7] Albert M Sirunyan et al. "Suppression of Excited Υ States Relative to the Ground State in Pb-Pb Collisions at $\sqrt{s_{NN}}$=5.02 TeV". In: *Phys. Rev. Lett.* 120.14 (2018), p. 142301. DOI: 10.1103/PhysRevLett.120.142301. arXiv: 1706.05984 [hep-ex].

[8] Shreyasi Acharya et al. "Υ production and nuclear modification at forward rapidity in Pb–Pb collisions at sNN=5.02TeV". In: *Phys. Lett. B* 822 (2021), p. 136579. DOI: 10.1016/j.physletb.2021.136579. arXiv: 2011.05758 [nucl-ex].

[9] Betty Bezverkhny Abelev et al. "Elliptic flow of identified hadrons in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV". In: *JHEP* 06 (2015), p. 190. DOI: 10.1007/JHEP06(2015)190. arXiv: 1405.4632 [nucl-ex].

[10] S. Voloshin and Y. Zhang. "Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions". In: *Z. Phys. C* 70 (1996), pp. 665–672. DOI: 10.1007/s002880050141. arXiv: hep-ph/9407282.

[11] John Adams et al. "Experimental and theoretical challenges in the search for the quark gluon plasma: The STAR Collaboration's critical assessment of the evidence from RHIC collisions". In: *Nucl. Phys. A* 757 (2005), pp. 102–183. DOI: 10.1016/j.nuclphysa.2005.03.085. arXiv: nucl-ex/0501009.

[12] Huichao Song. "Hydrodynamic Modeling and the QGP Shear Viscosity". In: *Eur. Phys. J. A* 48 (2012), p. 163. DOI: 10.1140/epja/i2012-12163-9. arXiv: 1207.2396 [nucl-th].

[13] Serguei Chatrchyan et al. "Study of high-pT charged particle suppression in PbPb compared to *pp* collisions at $\sqrt{s_{NN}} = 2.76$ TeV". In: *Eur. Phys. J. C* 72 (2012), p. 1945. DOI: 10.1140/epjc/s10052-012-1945-x. arXiv: 1202.2554 [nucl-ex].

[14] Nestor Armesto et al. "Testing the color charge and mass dependence of parton energy loss with heavy-to-light ratios at RHIC and CERN LHC". In: *Phys. Rev. D* 71 (2005), p. 054027. DOI: 10.1103/PhysRevD.71.054027. arXiv: hep-ph/0501225.

[15] Shreyasi Acharya et al. "Λ_c^+ production in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV". In: *Phys. Lett. B* 793 (2019), pp. 212–223. DOI: 10.1016/j.physletb.2019.04.046. arXiv: 1809.10922 [nucl-ex].

[16] Letizia Diamante. *Upgrading Alice: What's in store for the next two years?* Feb. 2019. URL: https://home.cern/news/news/experiments/upgrading-alice-whats-store-next-two-years.

[17] B Abelev et al. "Technical Design Report for the Upgrade of the ALICE Inner Tracking System". In: *J. Phys. G* 41 (2014), p. 087002. DOI: 10.1088/0954-3899/41/8/087002.

[18] Giacomo Contin. "The MAPS-based ITS Upgrade for ALICE". In: *PoS* Vertex2019 (2020), p. 003. DOI: 10.22323/1.373.0003. arXiv: 2001.03042 [physics.ins-det].

[19] Luciano Musa. *Letter of Intent for an ALICE ITS Upgrade in LS3*. Tech. rep. Geneva: CERN, Dec. 2019. URL: https://cds.cern.ch/record/2703140.

[20] Collaboration ALICE. *Letter of intent for ALICE 3: A next generation heavy-ion experiment at the LHC*. Tech. rep. Geneva: CERN, Mar. 2022. URL: http://cds.cern.ch/record/2803563.

[21] M. van Leeuwen J. Klein. *Alice 3 experiment concept*. URL: https://home.cern/news/news/physics/alice-3-workshop-towards-next-generation-heavy-ion-experiment-2030s (visited on 02/17/2022).

[22] ALICE collaboration. *Alice TOF*. URL: https://alice-collaboration.web.cern.ch/node/34978 (visited on 02/17/2022).

[23] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Computer Physics Communications* 191 (June 2015), pp. 159–177. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2015.01.024. URL: http://dx.doi.org/10.1016/j.cpc.2015.01.024.

[24] ALICE collaboration. *ALICE O2 analysis framework documentation*. URL: https://github.com/AliceO2Group/DelphesO2 (visited on 05/18/2022).

[25] ALICE collaboration. *ALICE O2 analysis framework documentation*. URL: https://aliceo2group.github.io/analysis-framework/ (visited on 03/15/2022).

[26] ALICE collaboration. "Centrality determination in heavy ion collisions". In: (Aug. 2018). URL: https://cds.cern.ch/record/2636623.

[27] Xingbo Zhao and Ralf Rapp. "Medium Modifications and Production of Charmonia at LHC". In: *Nucl. Phys. A* 859 (2011), pp. 114–125. DOI: 10.1016/j.nuclphysa.2011.05.001. arXiv: 1102.2194 [hep-ph].

[28] Andreas Hocker et al. "TMVA - Toolkit for Multivariate Data Analysis". In: (Mar. 2007). arXiv: physics/0703039.

[29] L Breiman. "Classification and Regression Trees". In: *Wadsworth Statistics/probability Series* (1984).

[30] Bob Stienen. "Working in High-Dimensional Parameter Spaces". Applications of machine learning in particle physics phenomenology. PhD thesis. Radboud Universiteit Nijmegen, July 2021.

[31] R. A. Fisher. "The use of multiple measurements in taxonomic problems". In: *Annals of Eugenics* 7.2 (1936), pp. 179–188. DOI: https://doi.org/10.1111/j.1469-1809.1936.tb02137.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1936.tb02137.x.

[32] Jonah E. Bernhard et al. "Applying Bayesian parameter estimation to relativistic heavy-ion collisions: Simultaneous characterization of the initial state and quark-gluon plasma medium". In: *Physical Review C* 94.2 (Aug. 2016). ISSN: 2469-9993. DOI: 10.1103/physrevc.94.024907. arXiv: 1605.03954 [nucl-th]. URL: http://dx.doi.org/10.1103/PhysRevC.94.024907.

[33] Yingru Xu et al. "Data-driven analysis for the temperature and momentum dependence of the heavy-quark diffusion coefficient in relativistic heavy-ion collisions". In: *Physical Review C* 97.1 (Jan. 2018). DOI: `10.1103/physrevc.97.014907`. arXiv: `1710.00807` `[nucl-th]`. URL: `https://doi.org/10.1103%2Fphysrevc.97.014907`.

[34] Matteo Cacciari, Mario Greco, and Paolo Nason. "The pT spectrum in heavy-flavour hadroproduction". In: *Journal of High Energy Physics* 1998.05 (May 1998), pp. 007–007. ISSN: 1029-8479. DOI: `10.1088/1126-6708/1998/05/007`. URL: `http://dx.doi.org/10.1088/1126-6708/1998/05/007`.

[35] Heavy Flavor Averaging Group (HFLAV) et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of summer 2016". In: *The European Physical Journal C* 77.895 (Dec. 2017). DOI: `s10052-017-5058-4`. URL: `https://doi.org/10.1140/epjc/s10052-017-5058-4`.

[36] R. Aaij et al. "Measurement of b hadron fractions in 13 TeV pp collisions". In: *Phys. Rev. D* 100 (3 Aug. 2019), p. 031102. DOI: `10.1103/PhysRevD.100.031102`. URL: `https://link.aps.org/doi/10.1103/PhysRevD.100.031102`.

Appendices

A Λ_c^+ selection criteria

These are the selection criteria used for the initial selection of Λ_c^+ candidates, eg. before the training and testing of a BDT.

Variable	selection criterion
$\lvert \Delta m_{\Lambda_c^+} \rvert$ GeV/c^2	<0.4
p_{T} proton (GeV/c)	>0.4
p_{T} kaon (GeV/c)	>0.4
p_{T} pion (GeV/c)	>0.4
Decay length (mm)	>0.5

Table 5: The selection criteria used for the first reconstruction of Λ_c^+ candidates. Where the mass difference is with respect to the hypothesis of the PDG [4].

B Λ_b^0 candidate parameter distributions

Following are the parameter distribution for the created Λ_b^0 candidates.
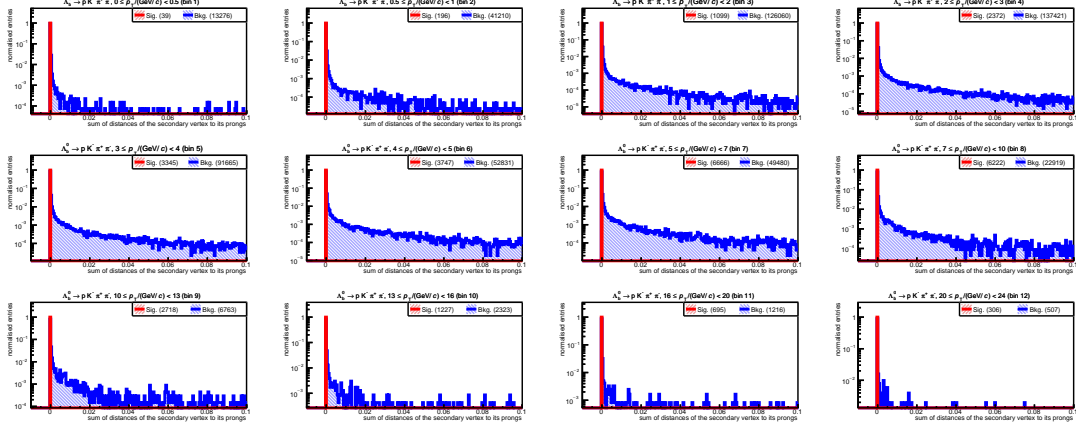
Figure 36: The distribution of the sum of the distances of closest approach of the daughter particles to the reconstructed secondary vertex at the optimal χ^2-fit for Λ_b^0 signal and background.
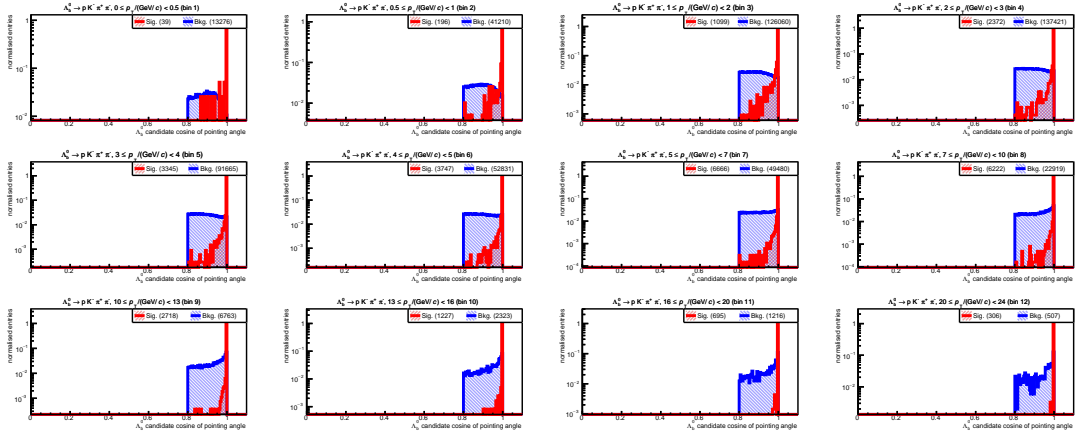
Figure 37: The distribution of the cosine of the pointing angle for Λ_b^0 signal and background.
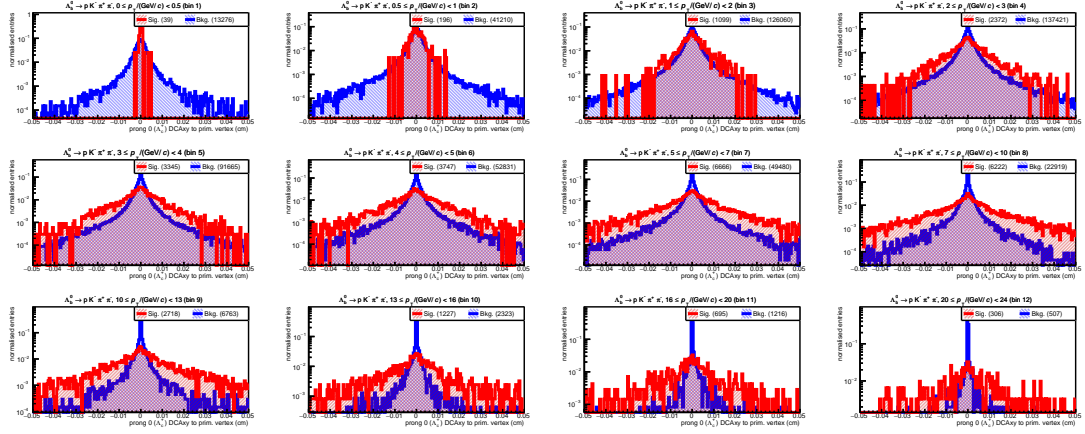
Figure 38: The distribution of the impact parameter of the first prong of the decay (i.e. Λ_c^+) for signal and background.
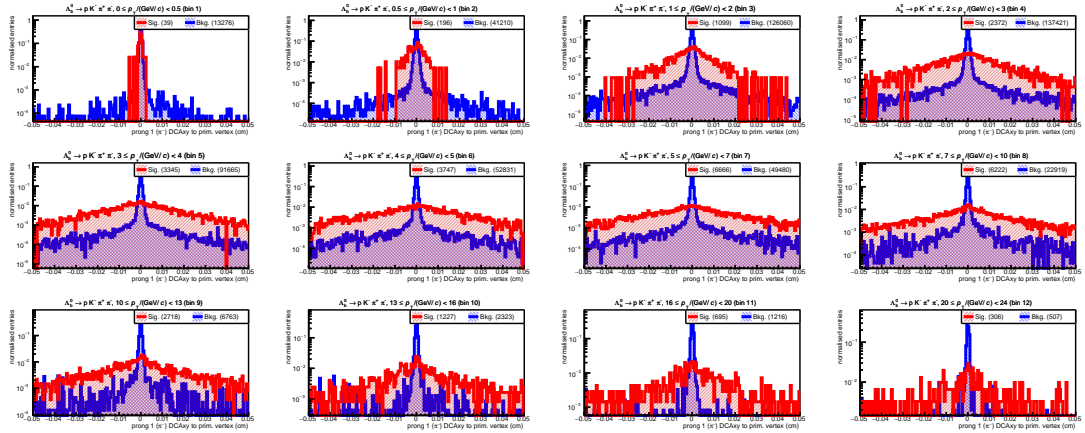
Figure 39: The distribution of the impact parameter of the second prong of the decay (i.e. π^-) for signal and background.
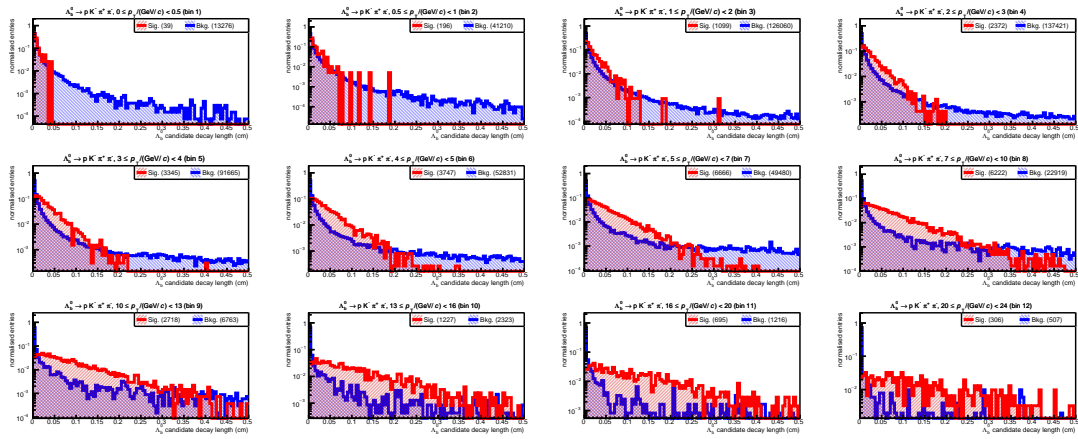
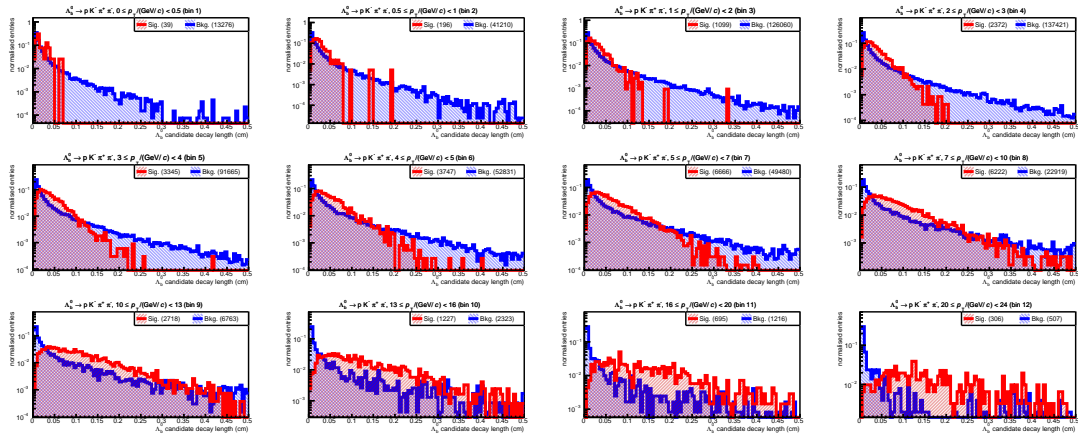Figure 40: The distribution of the decay length of Λ_b^0 signal and background.

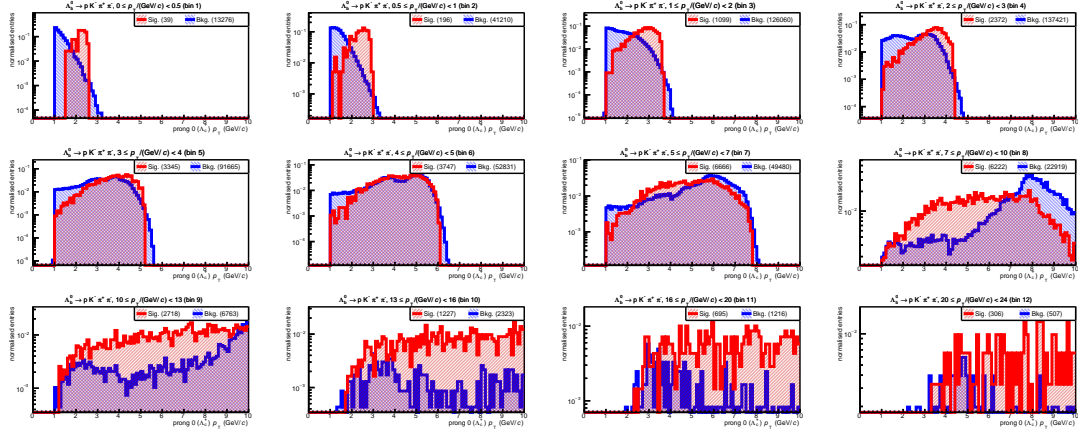Figure 41: The distribution of the decay length of Λ_c^+ signal and background.

Figure 42: The distribution of the transverse momentum of the first prong of the decay (i.e. Λ_c^+) for signal and background.
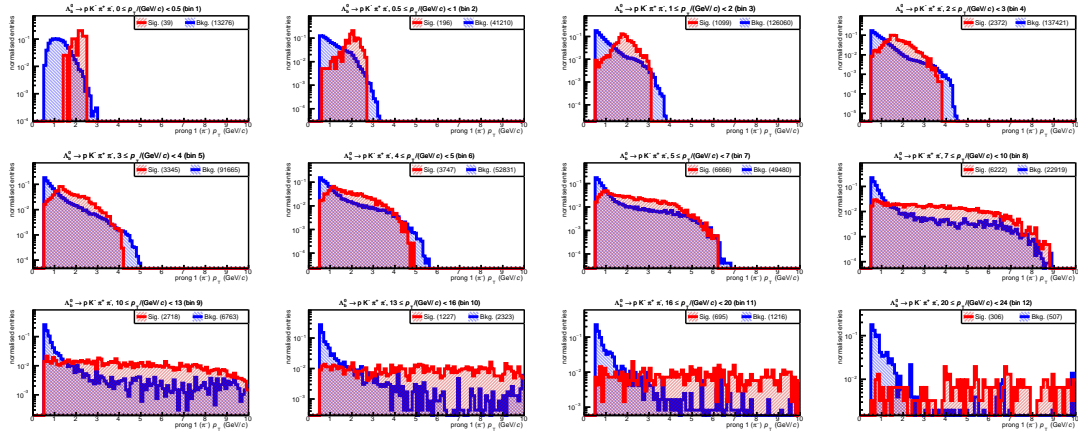

Figure 43: The distribution of the transverse momentum of the second prong of the decay (i.e. π^-) for signal and background.

C Λ_b^0 reconstruction efficiency

The following plots show the reconstruction efficiency as a function of transverse momentum. The plot on the right has logarithmic scaling on the vertical axis for clarity at low p_T. Due to the cascading decay, the reconstruction efficiency of the Λ_c^+ is also a factor in determining that of the Λ_b^0. As a consequence, the reconstruction efficiency is very low at low transverse momentum.
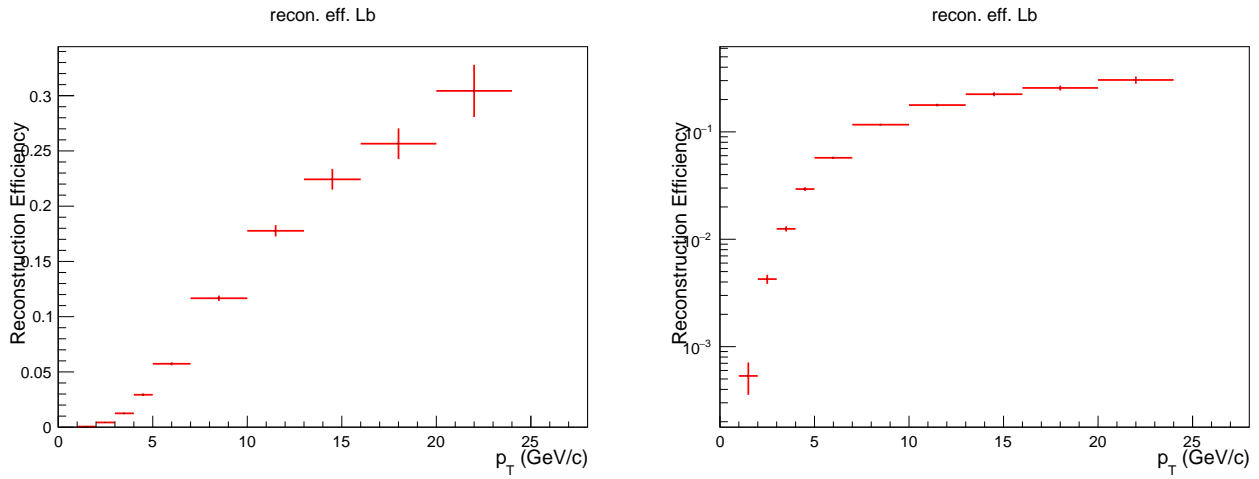

Figure 44: Reconstruction efficiency as determined by the monte-carlo matching procedure after applying the selection criteria. Both plots contain the same information, the version of the right just contains a logarithmic vertical axis for clarity at low p_T.

D Λ_b^0 selection criteria

These are the selection criteria used for the initial selection of Λ_b^0 candidates. The remaining candidates are used for the training and testing of the BDT's.

Variable	$0.5 < p_{\rm T} < 1$ GeV/c	$1 < p_{\rm T} < 2$ GeV/c	$2 < p_{\rm T} < 3$ GeV/c	$3 < p_{\rm T} < 4$ GeV/c
$|\Delta m_{\Lambda_b^0}|$ GeV/c^2	<1.0	<1.0	<1.0	<1.0
chi2PCA (μm^2)	<2	<2	<2	<2
CPA	>0.98	>0.98	>0.98	>0.98
CPAxy	>0.98	>0.98	>0.98	>0.98
D0Prong0 (μm)	>30	>30	>30	>30
D0Prong1 (μm)	>15	>15	>15	>15
Decay length (μm)	>50	>50	>50	>50
$p_{\rm T}$ pion (GeV/c)	>0.5	>0.5	>0.5	>0.5
$p_{\rm T}$ λ_c^+ (GeV/c)	>1.0	>1.0	>1.0	>1.0

Table 6: The selection criteria used for the reconstruction of Λ_b^0 candidates. Where the mass difference is with respect to the hypothesis of the PDG [4].

Variable	$4 < p_{\rm T} < 5$ GeV/c	$5 < p_{\rm T} < 7$ GeV/c	$7 < p_{\rm T} < 10$ GeV/c	$10 < p_{\rm T} < 13$ GeV/c
$|\Delta m_{\Lambda_b^0}|$ GeV/c^2	<1.0	<1.0	<1.0	<1.0
chi2PCA (μm^2)	<2	<2	<2	<2
CPA	>0.985	>0.985	>0.99	>0.99
CPAxy	>0.985	>0.985	>0.99	>0.99
D0Prong0 (μm)	>30	>30	>30	>20
D0Prong1 (μm)	>15	>15	>15	>15
Decay length (μm)	>50	>50	>50	>50
$p_{\rm T}$ pion (GeV/c)	>0.5	>0.5	>0.5	>0.5
$p_{\rm T}$ λ_c^+ (GeV/c)	>1.0	>1.0	>1.0	>1.0

Table 7: The selection criteria used for the reconstruction of Λ_b^0 candidates. Where the mass difference is with respect to the hypothesis of the PDG [4].

Variable	$13 < p_{\mathrm{T}} < 16$ GeV/c	$16 < p_{\mathrm{T}} < 20$ GeV/c	$20 < p_{\mathrm{T}} < 24$ GeV/c
$|\Delta m_{\Lambda_b^0}|$ GeV/c^2	<1.0	<1.0	<1.0
chi2PCA (μm^2)	<2	<2	<2
CPA	>0.994	>0.994	>0.994
CPAxy	>0.994	>0.994	>0.994
D0Prong0 (μm)	>20	>20	>20
D0Prong1 (μm)	>15	>15	>15
Decay length (μm)	>50	>50	>50
p_{T} pion (GeV/c)	>0.5	>0.5	>0.5
p_{T} λ_c^+ (GeV/c)	>1.0	>1.0	>1.0

Table 8: The selection criteria used for the reconstruction of Λ_b^0 candidates. Where the mass difference is with respect to the hypothesis of the PDG [4].

E Λ_b^0 significance

Following are the significance obtained by only applying selection criteria, without any form of machine learning involved.
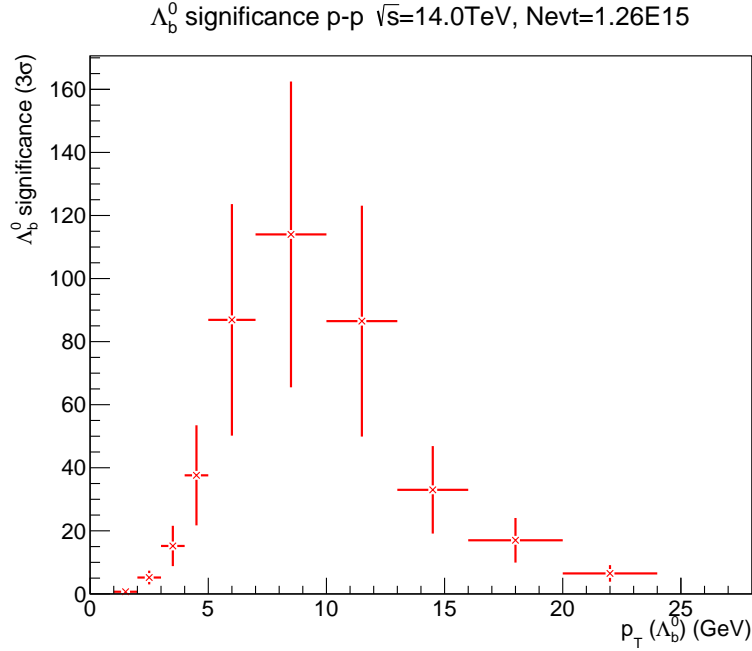

Figure 45: Significance of Λ_b^0 in p-p collisions at $\sqrt{s} = 14.0 TeV$ with only the selection criteria as listed in Appendix D.
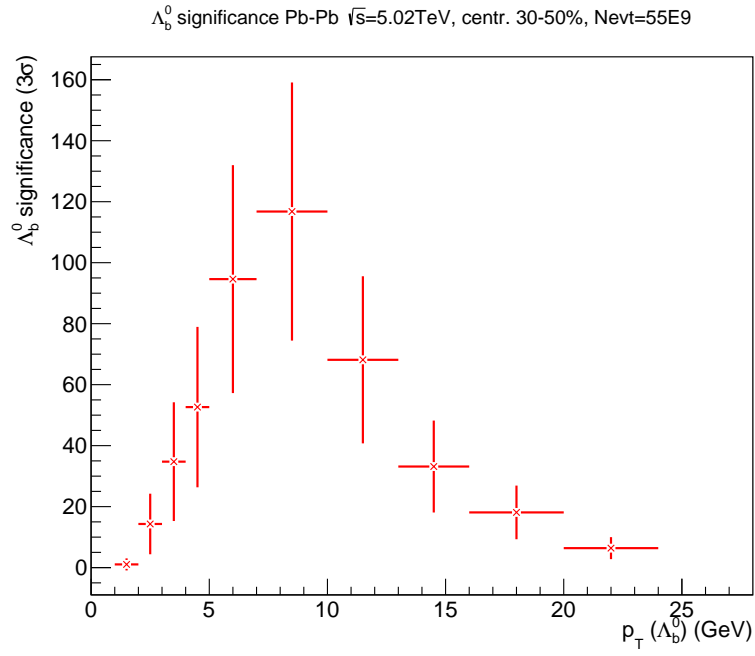
Figure 46: Significance of Λ_b^0 in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02TeV$ with only the selection criteria as listed in Appendix D.

F   integrals of ROC-curve

Following are the integrals of the ROC-curves as a function of transverse momentum of the Λ_b^0 candidate for the three different analysis strategies. They do not contains any uncertainties, however, they certainly depend on the randomized subsample of training events. Hence, one could determine some form of uncertainty by training on different subsamples. Information not directly obvious in these plots are some of the BDT artifacts. Especially in regions with very few number of testing events, the ROC-curves become discontinuous causing an error in the integral. As a consequence, these plots should not be taken too literal as they are solely included to sketch the negative relation with transverse momentum.
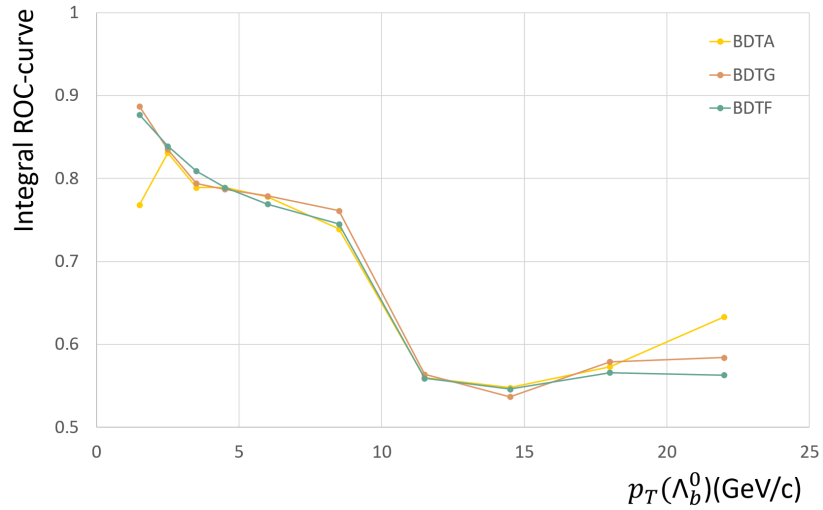

Figure 47: Integral of the ROC-Curves as a function of transverse momentum of the Λ_b^0 candidate of the three different BDTs using the simple list of input parameters as per Table 2.
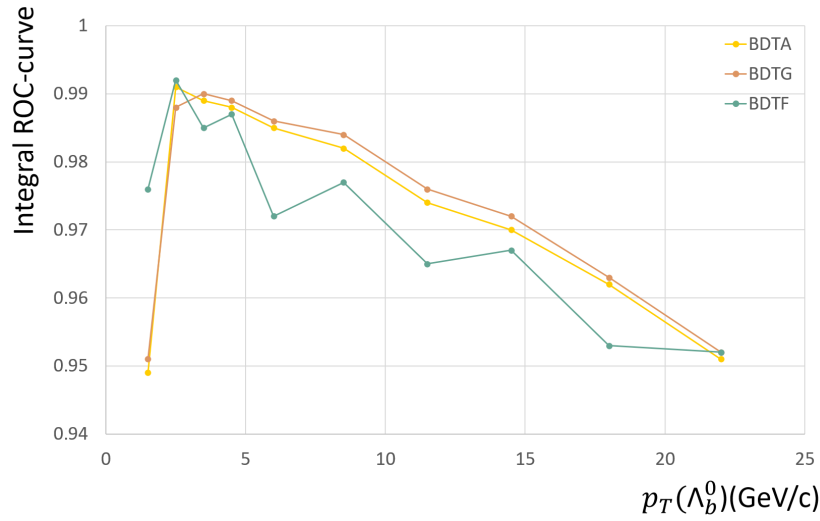
Figure 48: Integral of the ROC-Curves as a function of transverse momentum of the Λ_b^0 candidate of the three different BDTs using the simple list of input parameters as per Table 3.
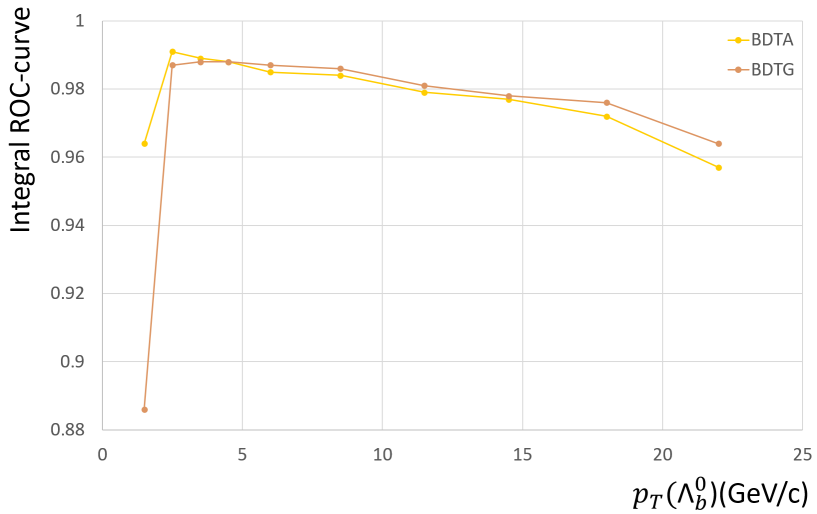

Figure 49: Integral of the ROC-Curves as a function of transverse momentum of the Λ_b^0 candidate of the two different BDTs using the simple list of input parameters as per Table 4.