Department of Physics and Astronomy
Heidelberg University

Bachelor Thesis in Physics
submitted by

Lisa-Katrin Kümmerer

born in Stuttgart (Germany)

September 2023

Optimization of CBM physics performance for Ξ^- baryon yield measurements with machine learning techniques

This Bachelor Thesis has been carried out by Lisa-Katrin Kümmerer at the
Physikalisches Institute in Heidelberg
under the supervision of
Prof. Dr. Slivia Masciocchi

Abstract

The Compressed Baryonic Matter (CBM) experiment at FAIR will investigate the QCD phase diagram in the region of high net-baryon densities ($\mu_B > 500$ MeV) and low temperatures ($T < 120$ MeV). The SIS100 accelerator will provide the ion beams for the fixed target heavy-ion collisions in an energy range of $\sqrt{s_{NN}} = 2.86 - 4.93$ GeV. With an unprecedented interaction rate of up to 10 MHz, the CBM detector will be built to operate in the low energy regime, enabling feasible studies of the state of matter. The (multi-)strange baryons are crucial in determining the equation of state at high baryon density and the hadronization process from deconfined QCD matter.

In this thesis, the performance for the Ξ^- selection in minimum bias Au+Au collisions at $\sqrt{s_{NN}} = 4.93$ GeV will be presented. Based on a Monte Carlo sample simulating the CBM detector response, the Ξ^- hyperon is reconstructed via the weak decay channel
$\Xi^- \rightarrow \pi^-(\Lambda^0 \rightarrow \pi^- p)$.

For the reduction of the data size, which is driven by the large combinatorial background, specific data skimming criteria are optimized in this work. In the future, this filtering procedure might be applied previous to data storage itself, working as a first kind of trigger. To then obtain an optimal and stable separation between signal and background candidates, the machine learning tool XGBoost is used. Machine learning allows for efficient, non-linear and multidimensional selection criteria to be implemented in an heavy-ion collision environment.

An iterative approach on machine learning selection is introduced, involving two consecutive models. This procedure allows to improve the selection accuracy by removing background in each step (data skimming and ML selection) before zooming deeper in data characteristics to include finer differences in feature correlations in the next step. The strategy proves to be more efficient than a single model selection for the signal separation and enables to extract the Ξ^- raw yield.

Zusammenfassung

Das Compressed Baryonic Matter (CBM) Experiment wird als Teil von FAIR das QCD Phasendiagramm im Bereich hoher Netto-Baryondichten ($\mu_B > 500\,\text{MeV}$) und tiefer Temperaturen ($T < 120\,\text{MeV}$) untersuchen. Der SIS100 Ringbeschleuniger wird Ionenstrahlen für die Schwerionenkollisionen mit festem Target (engl. für Ziel) im Energiebereich von $\sqrt{s_{NN}} = 2.86 - 4.93\,\text{GeV}$ liefern. Mit einer beispiellosen Interaktionsrate von 10 MHz wird der CBM Detektor entwickelt, um im Tiefenergieregime geeignete Studien des Aggregatzustandes zu ermöglichen. Dabei sind (multi-)strange Baryonen ausschlaggebend, um die Zustandsgleichung bei hoher baryonischer Dichte und den Hadronisierungsprozess von unbeschränkter (engl. deconfined) QCD Materie zu ermitteln.

In dieser Arbeit wird die Performance einer Ξ^- Selektion in minimum bias (minimale Triggerbedingung) Au+Au Kollisionen bei $\sqrt{s_{NN}} = 4.93\,\text{GeV}$ präsentiert. Aus einem Monte Carlo Sample, welches die CBM Detektorreaktion simuliert, wird das Ξ^- Hyperon über den schwachen Zerfallskanal $\Xi^- \rightarrow \pi^-(\Lambda^0 \rightarrow \pi^- p)$ rekonstruiert.

Aufgrund der großen Menge an kombinatorischem Hintergrund werden in dieser Arbeit spezifische Filterkriterien optimisiert, um die Größe des Datensets zu reduzieren. In Zukunft kann dieser Filterprozess schon vor der Datenspeicherung als erste Art von Trigger angewandt werden. Um danach eine optimale und stabile Trennung von Signal und Hintergrund zu erlangen, wird das Machine Learning Werkzeug XGBoost verwendet. Machine Learning erlaubt eine Implementierung von effizienten, nicht-linearen und multidimensionalen Selektionskriterien in der Umgebung von Schwerionenkollisionen.

Es wird ein iterativer Ansatz zur Machine Learning Selektion vorgestellt, welcher zwei aufeinanderfolgende Modelle enthält. Diese Prozedur erlaubt eine verbesserte Selektionsgenauigkeit, indem in jedem Schritt (Filter, sowie zwei Machine Learning Modelle) Hintergrundkandidaten verworfen werden, bevor die Eigenschaften im nächsten Schritt genauer betrachtet werden, wobei feinere Korrelationen in den Variablen miteinbezogen werden. Die Strategie zeigt sich effizienter als eine Selektion mit nur einem Modell zur Trennung von Signalkandidaten und ermöglicht die Extraktion des reinen Ξ^- Ertrags.

Contents

1. Introduction

In the physics of fundamental interactions and elementary particles, the search for the basic constituents of matter seems to have come to a stop. While the previously assumed indivisibility of components proved to be wrong for both atoms and protons successively, the description of elementary particles in the Standard Model does not allow a further partition. Even though the theory lacks in terms of completeness, predictions are found to be sufficiently accurate in experiments to allow the model to be widely accepted as factual. The following short introduction to the Standard Model is based on information given in [1].

The elementary particles include fermions, spin $1/2$ particles that make up the actual matter, as well as gauge bosons, spin 1 particles that are carriers of one of the fundamental forces each. Quantum field theories are used to describe these forces and interactions based on the underlying symmetries.

Quantum Chromodynamics (QCD), which strictly follows a SU(3) symmetry, describes the strong interaction coupling to color charge. The carriers of the color charges (red, green or blue) are fermions called quarks (and antiquarks with anticolors respectively). The messenger bosons (or gauge bosons) of the strong interaction, the gluons, are colored themselves as well. Each holds both color and anticolor, resulting in an octet of truly colored states. Coupling to color while carrying the same charge, gluons exhibit self-interaction. Together with a very large strong coupling constant α_s, this causes the so called color-confinement. The constantly self-interacting gluons are pulled into a tube like connection between color charged particles. The force connecting them is constant, which cause the energy to rise linearly with the distance. A separation would therefore require infinite energy and is not practicable.

$$V(\mathbf{r}) = -\frac{3}{4}\frac{\alpha_s}{r} + \kappa\mathbf{r} \tag{1.1}$$

The QCD potential between a quark and antiquark pair of the same color given in Eq. (1.1) demonstrates the effect of confinement. While the first term induces a Coulomb-like interaction on short distances $\mathbf{r}$, the linear term with $\kappa \sim 1\,\mathrm{GeV}$ causes a divergence for high distances, which makes a separation impossible. Carriers of color charge do therefore not appear isolated, but form color neutral states by hadronization, including baryons and mesons. Baryons are effectively colorless due to their composition of three (valence) quarks with a different color each. Mesons on the other hand consist of two (valence) quarks (or rather quark and antiquark) carrying color and according anticolor, which cancel each other, resulting in a non-colored state. Regular baryonic matter (as for example atomic nuclei) is formed by only a residual effect of the strong force acting between the colorless states.

One of the unique properties of QCD is the strong dependence of the interaction strength on the energy scale of the underlying process. The running of the coupling α_s, which strongly depends on the squared 4-momentum transfer Q^2, is depicted in Fig. 1.1. In the low energy regime, the coupling is very large (of order $O(1)$), resulting in the previously described confinement of color charged particles. This prohibits any perturbative calculation like used in Quantum Electrodynamics (QED), which is based on the assumption of a small coupling. Another approach, lattice QCD, allows for a precise description for certain properties of QCD, but is limited due to its computational expensive calculations. It can therefore not completely compensate the lack of predictive power in a non-perturbative theory.

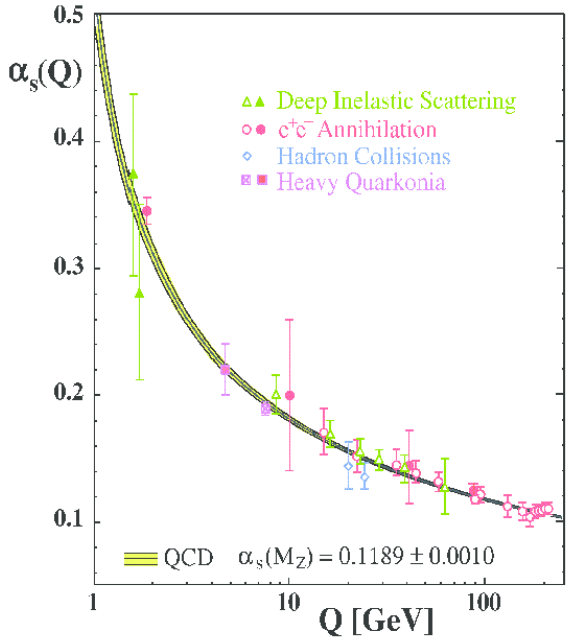


Figure 1.1.: Running of the strong coupling $\alpha_s(Q^2)$, taken from [2]

On small distances (or for interactions with a high momentum transfer Q^2), the coupling decreases, which leads to an effect known as asymptotic freedom. At energies above 100 GeV, the coupling has dropped to $O(0.1)$ already. For processes with momentum transfer much greater than the QCD scale $\Lambda_{QCD} \sim 200$ MeV, the coupling is sufficiently small to allow for an application of perturbative calculation. With radii of the order of 1 fm, this also applies for the interactions inside of hadrons.

Due to the reduced interaction strength based on the decreased coupling, the color anti-screening effects causing confinement can be neglected. Quarks and gluons are deconfined and can be treated as free in this regime.

1.1. The QCD phase diagram

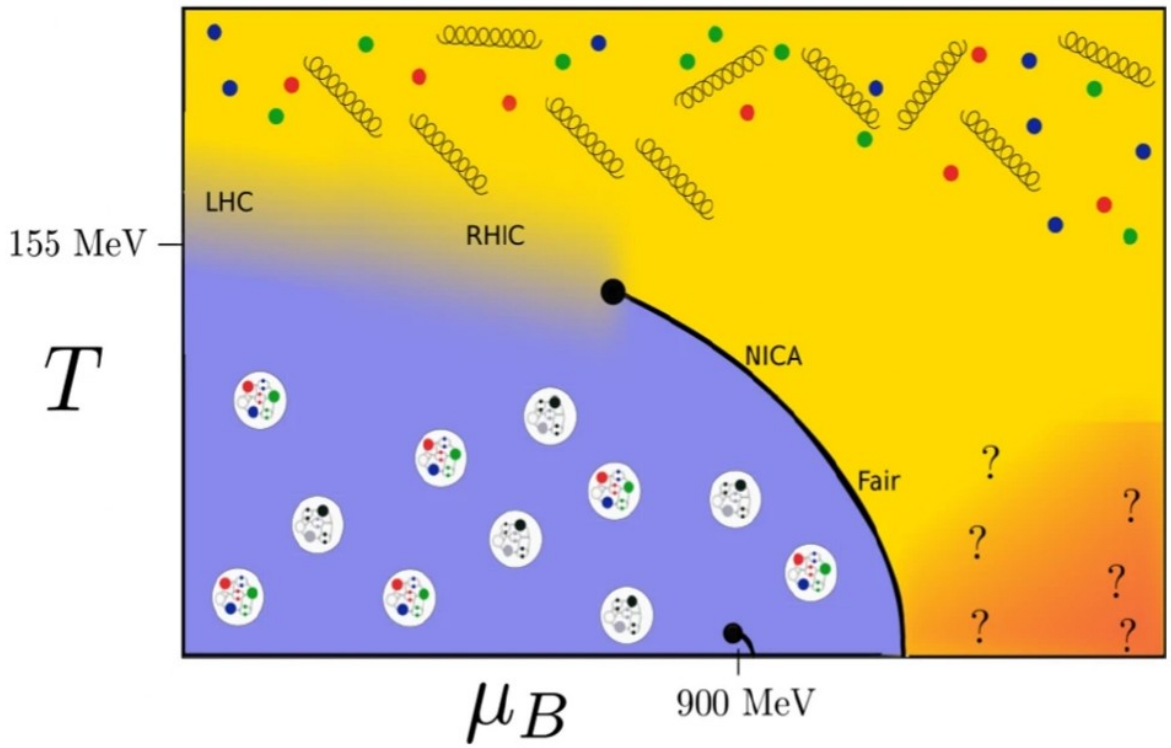


Figure 1.2.: QCD phase diagram, taken from [3]

Hadronic matter only dominates the small part at low temperatures and densities of the QCD phase diagram shown in Figure 1.2. At high temperature or density, another state of matter, the quark-gluon plasma (QGP) is formed, where the quarks and gluons are deconfined. For negligible baryo-chemical potential $\mu_B \sim 0$, a smooth transition without encountering an actual phase transition is expected. Lattice QCD calculations predict this cross over at the critical temperature $T_C = 156 \pm 1.5$ MeV [4]. For higher baryo-chemical potential, a first order transition is expected between hadronic matter and QGP. Beyond that, for even lower temperature and higher density, completely new phases might exist.

The equation of state (EoS) relates the pressure p, the volume V, and the absolute temperature T of a system. High-energy heavy-ion collisions offer the unique possibility to study the fundamental properties of nuclear matter in the laboratory. While the EoS for vanishing net-baryon density (LHC energies) is now well constrained by lattice QCD calculations, the high density and intermediate- to low-temperature EoS is still not well known.

The Compressed Baryonic Matter experiment (CBM), with a collision center of mass energy range of $\sqrt{s_{NN}} = 2.86 - 4.93$ GeV, will explore the phase diagram for temperature $T < 120$ MeV

and $\mu_B > 500\,\text{MeV}$. In this region, the first order transition is expected to be found. According to recent calculations, the critical endpoint, where the first order transition changes to the cross over, might be located in the CBM range as well. Contrary to the outdated depiction in Figure 1.2, the critical endpoint is expected at $(T_{CEP}, \mu_{B\,CEP}) = (107, 635)$ MeV [5].

With high center-of-mass energies, heavy-ion collider experiments at RHIC or LHC explore the high temperature region. This allows to conclude on the state of matter that existed in the early universe, microseconds after the Big-Bang, and the following freeze-out. Up to now, the low temperature, high baryo-chemical potential regime is poorly understood on the other hand. With three times nuclear density ρ_0 and a temperature of only a few keV, the state of the inner core of compact stars is placed in this region of the QCD phase diagram. Observables like mass, radius and electromagnetic radiation can deliver restrictions, but the actual EoS at high densities can not be derived without a direct investigation of the phase diagram at high baryo-chemical potential. Particularly, more knowledge is needed to describe states with high density, but non-negligible temperature like existent in supernovae and neutron star mergers. With densities of up to $8\rho_0$, comparable conditions can be expected in CBM collisions, allowing to characterize QCD phases at high densities. More information about the QCD phase diagram and the physics goals of CBM can be found in [6] and [7], which the short introduction in this thesis is based on.

1.2. Motivation for multistrange hyperon studies

Due to their light masses, the three lightest quark flavours up, down, and strange can be assumed to be produced near chemical equilibrium at high energies. Contrary to the low energy regime, where the heavier strange quark production is suppressed towards the up/down quark production, this enhanced strangeness gives information on the modified degrees of freedom (DOF). This allows to directly conclude on the characteristics of the state existing in the fireball emerging in a heavy-ion collision. Apart from this, the formation of multistrange baryons is favoured due to subsequential multi-collisions of the fragments while the fireball cools down. This process directly depends on the density and dynamics of the medium. Measuring strangeness, and particularly multistrange baryons therefore allows to conclude directly on the emerging state. Being the lightest charged multistrange hyperon, a full analysis of the Ξ^- yield in the experimental data and the study of the physics performance that can be obtained with the CBM detector are essential. To facilitate this, an optimized process for the selection of signal candidates in unbiased Monte Carlo data samples will be presented in this thesis.

2. The CBM experiment at FAIR

The Facility for Antiproton and Ion Research (FAIR) is a ion accelerator complex planned to research on hadron, nuclear, atomic and plasma physics. Located at the GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt, Germany, it will expand the existing accelerator SIS18. This will act as injector to the planned Schwerionensynchrotron SIS100, which will then provide the high intensity ion beams for the CBM experiment. With a rigidity of 100 Tm, the superconducting ring accelerator will provide gold ion beams with a kinetic energy between 2 and 11 AGeV. In the fixed target Au+Au collisions this equals an energy range of $\sqrt{s_{NN}} = 2.86 - 4.93$ GeV. The extracted beams supplied for the CBM experiment will reach up to 10^9 gold ions per second, resulting in collision rates of up to 10 MHz. To resolve the collisions with such a high rate, an elaborated detector system is necessary (Fig. 2.1).

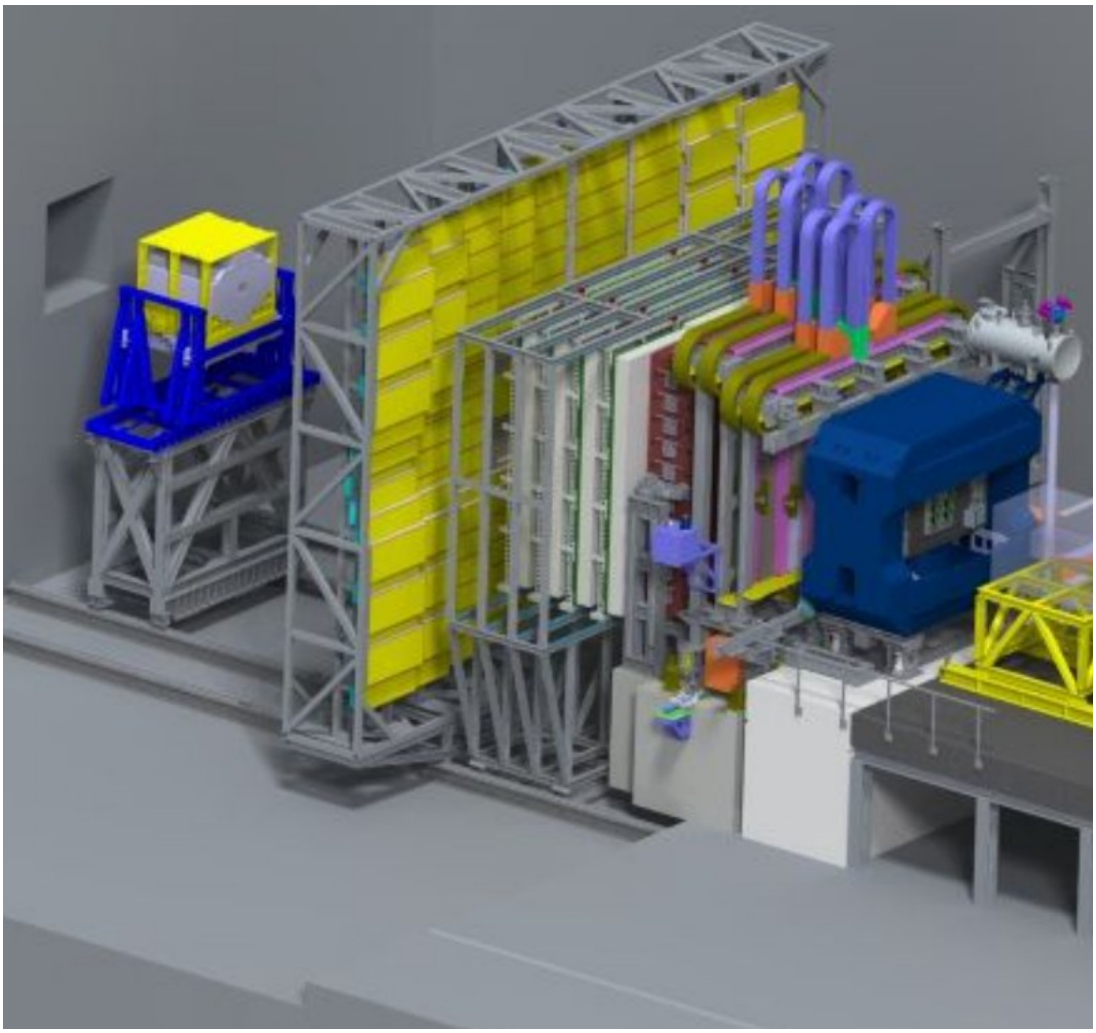


Figure 2.1.: The CBM detector system (internal material from the CBM collaboration)

The detectors include

- **MVD & STS** the Micro Vertex Detector and the Silicon Tracking System will be installed inside a superconducting magnet right behind the target position to track charged particles
- **RICH & TRD / MuCH** a Ring Imaging CHerenkov detector and a Transition Radiation Detector will be used for electron identification and pion surpression. In a muon setup, the RICH detector can be replaced by a Muon CHamber
- **TOF** a Time of Flight detector will measure the velocity of charged particles
- **PSD** the Projectile Spectator Detector will give information on the centrality and reaction plane of events by measuring the non-interacting nucleons (spectators)

For the reconstruction of short lived hyperons, no direct measurements are practicable. With a decay length of $c\tau = 4.9\,\text{cm}$, the Ξ^- does not reach the detector system itself, but is reconstructed via its decay products. To measure those daughters, the tracking of charged particles and particle identification (PID) are necessary. The function of the respective detectors is briefly explained in the following. While the measurements of other detectors can be used to improve the results, they are not used directly for this analysis and are therefore not elaborated further.

2.1. Tracking system

Together with the MVD, the STS will track charged particles emerging from the target position. Being placed inside a magnet, the bending of a charged particle track in the magnetic field then allows to measure its momentum. The the STS will act as the main tracking detector in the CBM apparatus, while the MVD will allow to improve the resolution for tracking and (particularly) the reconstruction of vertices.

2.1.1. Magnet

The superconducting dipole magnet will be placed around the target position and will cover both MVD and STS. In the gap of $140 \times 300\,\text{cm}$ it will generate a magnetic field throughout both detectors with 1 Tm over a 1 m length, resulting in a maximal field strength of 1 T [8].

2.1.2. Micro Vertex Detector

Located right behind the target position (relative to the incoming beam), the MVD will allow to improve the tracking resolution drastically. It will consist of four stations with Monolithic Active Pixel Sensors (MAPS) chips sitting $5 - 20\,\text{cm}$ downstream of the collision point. With a

thickness of 50 μm for each of the Complementary Metal-Oxide-Semiconductor (CMOS) pixel sensors, it will measure charged particles with a spatial precision of 5 μm.

2.1.3. Silicon Tracking System

The STS will act as the main detector of the CBM experiment. Composed by 8 tracking stations, located between 30 and 100 cm downstream of the target position, the STS will cover an opening angle of $2.5° < \Theta < 25°$ and a total area of 4 m^2. The layers of silicon detectors will consist of altogether 896 double-sided microstrip sensors (each with a thickness of 300 μm, a length between 20 and 60 mm, placed with a strip pitch of 58 μm and stereo angle of 7.5 μm). They enable a spatial resolution of 25 μm and a time stamp resolution of 5 ns, resulting in a momentum resolution of $\Delta p/p = 1.8\,\%$ for up to 700 charged particles per event, 10^7 events per second [9].

2.2. Particle identification

To be able to assign a particle identity in form of a mass measurement, a TOF detector will be installed. Measuring the time t that a charged particle needs to pass through the distance l, the velocity $\beta = \frac{l}{t}$ is calculated. With the STS information on the particles momentum p this enables to conclude on the mass via

$$m^2 = p^2 \left(\frac{1}{\beta^2} - 1 \right). \tag{2.1}$$

The TOF-wall of 120 m^2 will be placed 6 – 10 m downstream of the target, which again results in an opening angle of $2.5° < \Theta < 25°$. It will consist of Multigap Resistive Plate Chambers (MRPCs), which measure the ionization caused by the charged particle while passing through gaps filled with gas in between resistive plates. To be able to distinguish high energy particles a time resolution of < 80 ps will be necessary [10].

3. Data sample and reconstruction

3.1. Data sample

The data sample utilized in this analysis consists of a Monte Carlo (MC) simulation of minimum bias Au+Au collisions with a beam momentum of 12 AGeV. It is obtained with the Dubna Cascade Model and the Quark Gluon String Model (DCM-QGSM); the Statistical Multifragmentation Model (SMM) is used to reproduce the subsequent fragments formation [11]. The particles weak decay, propagation through the CBM detector material and the related response is modeled with the GEANT4 [12] toolkit and reconstructed with algorithms implemented in the CBMRoot framework [13].

3.2. Reconstruction of short-lived particles

Particle tracks throughout the detector system are formed by combining hits in STS and MVD detectors according to the Cellular Automaton (CA) algorithm. Combining reconstructed tracks by assigning them to events (collisions) [14] gives information on the collision point, as well as event related information like particle multiplicity.

Decay candidates are reconstructed using the PFSimple Package [15], which - being a simplified version of the Kalman Filter Particle (KFParticle) Package - implements the Kalman Filter mathematics [16]. All particle tracks, both primary (produced in the collision) and secondary (produced in subsequent decay or interaction with detector material) are described equivalently by a state vector

$$r = (x, y, z, p_x, p_y, p_z, E, s) \tag{3.1}$$

with the particles coordinates (x, y, z), momentum (p_x, p_y, p_z), energy E and (if production and decay vertex are known) $s = \frac{l}{p}$, the decay length l normalized to the momentum p [17]. The algorithm then optimizes the state vector to describe a trajectory by iteratively adding measurements, each updating the state vector according to the residual of prediction and measurement. Here, 'measurements' implies hits on STS and MVD detectors for directly measured particles, but also the state vectors of daughter particles for reconstructed mothers. The uncertainties of measurements are included as weight for the computation of the state vector, resulting in an optimal description of each particle track by the state vector and its covariance matrix C.

This equivalent description allows to add or subtract tracks together in order to form short lived, decayed particles. In contrast to other approaches (i.e. simply adding daughter particle tracks at their point of closest approach), the complete tracks of all particles including their

covariance are reconstructed. This allows to extrapolate particles trajectories to any chosen point in the system and to reconstruct primary (collision point, PV) as well as secondary (decay points, SV) vertices [18, 19]. Knowing both tracks and vertices, topological features can be used to identify a potential decay. With the covariance matrix C of each track, distances $\Delta\vec{r}$ can be given in relation to their uncertainties in the form of χ^2 values:

$$\chi^2 = \Delta\vec{r}^T C^{-1} \Delta\vec{r} = \sum_{i,j} C_{ij}^{-1} \Delta r_i \Delta r_j \tag{3.2}$$

The main features used in this analysis are listed below. For a simple two-body decay $a \rightarrow bc$, the same features are visualized in Figure 3.1.

- DCA: distance of closest approach between the daughter particles
- L: decay length of the mother particle
- $L/\Delta L$: decay length L normalized to its uncertainty ΔL
- χ^2_{prim}: χ^2 between daughter particle track and PV
- χ^2_{geo}: χ^2 between daughter particle tracks
- χ^2_{topo}: χ^2 between mother particle track and PV + χ^2 between daughter particle tracks
- $cos(\alpha)$: cosine of angle between mother and daughter particle tracks
- $cos(\alpha_{topo})$: cosine of angle between mother momentum and connection PV-SV

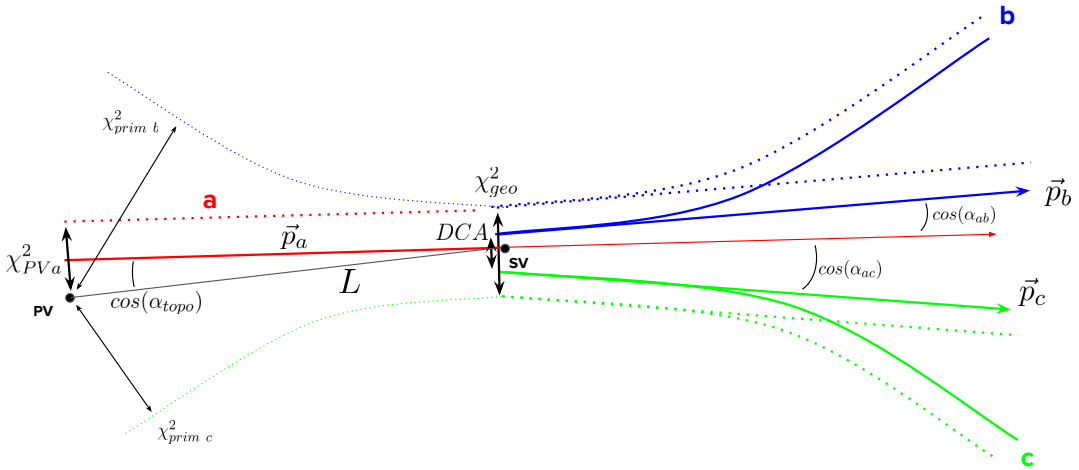


Figure 3.1.: Topological features of a two-body decay $a \rightarrow bc$ as reconstructed with KFParticle Package

3.3. Reconstruction of Ξ^- candidates

The Ξ^-, being a short lived particle, is reconstructed via its weak decay to $\Lambda^0\pi^-$ with a branching ratio of 99.887 ± 0.035 %. The Λ^0 itself is reconstructed from its decay products π^- and proton (branching ratio of 64.1 ± 0.5 %) [20]. The decay chain and some of the most important topological variables are shown in Fig. 3.2.

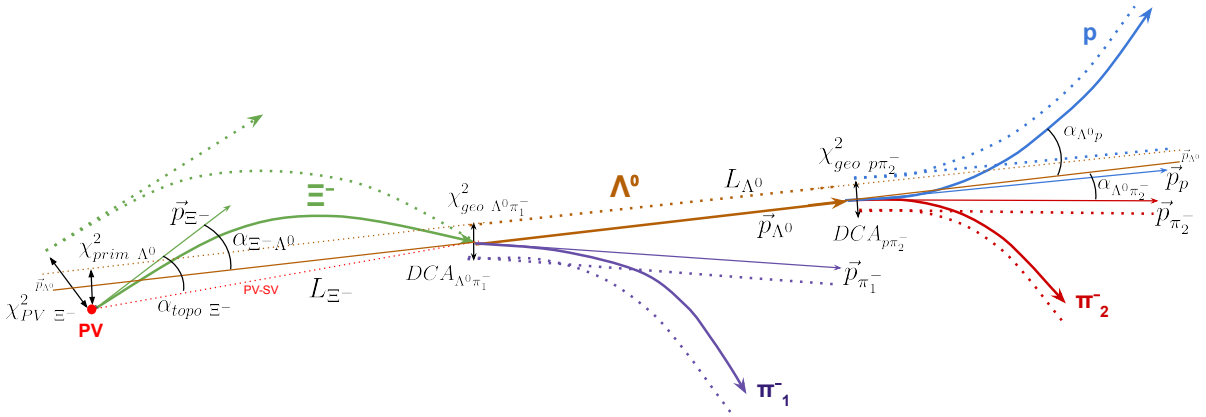


Figure 3.2.: Ξ^- decay chain with some of the most important reconstructable variables

Only the charged daughter particle tracks, (proton and pions) can be reconstructed directly from MVD and STS measurements. A Λ^0 candidate is reconstructed by combining a negative track with a positive track. Adding another negatively charged track, a Ξ^- candidate is formed [21].

Aside from the charge, no constraints are imposed on the measured daughters at the time of the reconstruction. Particularly, no PID selection is applied. To reduce the data size nonetheless, combinatorial background is rejected by constraining the reconstructed Λ^0 mass to the 5σ region of the expected mass $m_{\Lambda^0} = 1.1157 \pm 0.0015\,\mathrm{GeV/c^2}$. Additionally, a non-linear mass constraint is applied on the reconstructed energy and momentum to match the real Λ^0 mass $m^2_{\Lambda^0} = E^2 - p^2$. This recalculation does not act as selection, but updates the Λ^0 state vector. Constraining the invariant mass of the (reconstructed) daughter previous to the reconstruction of mother particles improves the mass resolution, providing a better separation of signal and background.

4. Analysis strategy

The invariant mass distribution of $\Lambda^0\pi^-$ pairs does not show a visible Ξ^- signal peak in data samples without any selection due to the dominant combinatorial background (compare Figures A.1, A.2 in the Appendix). In order to reduce the background and extract signal candidates, a selection process including manual data skimming as well as machine learning techniques is developed and presented.

Reconstructing short lived particles by combining tracks based on their charge allows the reconstruction without the need of PID for the daughters, but constructs a lot of combinatorics at the same time. Aside from misidentified particles, this includes particles (proton, pions, as well as Λ^0) that do not come from a Ξ^- decay chain, as well as actual Ξ^- daughters that are wrongly combined with other particles. Based on the different origin, the properties of background and signal candidates vary in both, the one-dimensional distribution of some features, as well as multidimensional correlation of features. While the first can be suppressed by data skimming, machine learning (ML) allows to extract signal in a non-linear way.

Using a Monte Carlo sample allows to optimize the selection process by separating signal (MC-true) and background (MC-false) candidates based on their label. In this way, the differing properties can be made visible and used in the filtering procedure. Additionally, supervised ML is practicable, which implies availability and usage of the real label for each candidate. Finally, the selection quality can be verified using the MC labels of the remaining candidates after selection.

4.1. Data skimming

A set of constraints on the reconstructed decay variables is chosen to exploit differences in the one-dimensional signal and background distributions. With a signal to background ratio of about 3.5×10^{-7} in samples without selection, the preselection is necessary. In order to save time and computational resources during the model building, this needs to be done previous to further selection with ML techniques.

Even data storage itself is challenging due to the size of the samples. With the high CBM interaction rate, background combinatorics results in about 5×10^{11} reconstructed Ξ^- candidates per second. Because of this, the filtering procedure might even be applied ahead of data storage, working as a first kind of trigger. Saving only candidates in the chosen limits for further Ξ^- analysis reduces the needed storage drastically.

4.2. Multidimensional selection with machine learning

After having skimmed the data, ML is exploited in order to apply non-linear selection. To do so, a model is optimized and trained on the MC sample before applying a selection based on its prediction score. In this analysis, the Heavy Ion Physics Environment for Machine Learning (hipe4ml) package [22] is used to train the model on learning the differences in the signal and background sample (distinguishable by their MC label). To mimic the approach on real data, where no MC label is available, the sets are additionally separated in the invariant Ξ^- mass distribution. For the training, signal is only included in the Ξ^- mass peak region, while background candidates in the same region are excluded.

4.2.1. Classification

The model output consists of an assigned Boosted Decision Tree (BDT) score between 0 and 1 for each candidate. This can be understood as signal probability, where a score close to 1 indicates a signal candidate, opposed to a background candidate which should achieve a BDT score close to 0.

The used Extreme Gradient Boost (XGBoost) algorithm [23] combines weak learners to a boosted decision tree. Weak learners are in this case binary decision trees, that - taken on their own - manage to classify data just marginally better than random classification. The combination is done in an order such that a second order Taylor expansion of a Loss Function is minimized for each step. The approximation includes both gradient and hessian of the loss function, hence the name gradient boosting. In the following analysis, this loss function is calculated as the mean squared error (MSE) between prediction and true label, the second order Taylor expansion therefore gives the exact solution. By adding a regularization term, the complexity of each weak learner is minimized as well, which is meant to compensate for missing statistics due to a non-infinitely large data sample.

The XGBoost algorithm depends on hyperparameters that directly modify the learning process. A tuning is necessary to adapt to the properties of the data sample under analysis. While a small sample with easily distinguishable characteristics might be accurately classifiable by a conservative, simple model, higher statistics with evenly distributed points require a deeper, more complex description. One of the challenges in adjusting the hyperparameters is to avoid overtraining. This happens if the ML model is allowed to learn local fluctuations of data features and results in a classification based on random, non generalizable characteristics. Since these are peculiar to the sample used for training, the model will perform worse on unknown data. To ensure the model recognizes only statistical, general differences, the performance needs to be tested on unknown data.

For the optimization, the bayesian based Optuna framework [24] is used in this analysis, which implements the k-fold cross validation method [25]. In this approach, the sample is split in k parts, with each being a validation set for a model trained on the remaining $k-1$ parts. Minimizing the mean error between model output and true label on the k-folds, this gives the optimal performance on an unknown data sample.

Permittable ranges for the hyperperameters need to be adjusted manually before optimization. For practical reasons, no infinite volumes of hyperparameters can be scanned. Additionally, the k-fold algorithm tends to deliver a slightly too complex ML model. While its aim is to optimize the prediction a model achieves on unknown data, it does not consider potential differences to the model performance on the known training set. However, overtraining does not necessarily cause a bad prediction score itself. The problem can only be certainly eliminated by comparing the performance on train and test sets, making sure they do not differ much.

The following hyperparameters are being optimized for this analysis:

- **number of estimators:** the number of combined weak learners
- **maximum tree depth:** the allowed depth of a weak learner (number of decisions it is allowed to make). This directly specifies the variable correlation the model will know - allowing only 2 decisions only includes 2D correlations
- **learning rate:** the weight on each weak learner before it is added to the BDT
- **gamma:** minimum loss reduction required to allow further splitting in a weak learner
- **alpha:** weight on regularization term
- **scaling positive weight:** control the weighting of positive/negative entries to compensate for non balanced signal/background set size

4.2.2. Selection

After having optimized the training for the present data sample, the model is applied on a new sample (constructed in the same way, but not used for training). A selection is applied by accepting only candidates above a certain BDT probability. This method allows to separate signal and background based on the ML classification, which considers multidimensional feature correlation. An optimal BDT threshold is chosen by maximizing the significance

$$sign = \frac{S}{\sqrt{S+B}} \tag{4.1}$$

with the MC-true signal count S and the MC-false background count B after selection. With an approximation of a negligible deviation on the background count, this achieves a minimal relative deviance on the signal count (for the signal deviance $\sigma(S)$).

$$\frac{1}{sign} \sim \frac{\sigma(S)}{S} \tag{4.2}$$

To verify the selection quality for a chosen BDT threshold, the remaining signal to background ratio is computed in a 5σ range of the peak in the reconstructed Ξ^- invariant mass. With $\sigma = 0.002\,\text{GeV}$, this region contains 95% of the signal distribution as explained in more detail in Chapter 5.3.4. Additionally, a high signal efficiency (i.e. the percentage of remaining signal after selection) is an indicator for an accurate and beneficial selection.

4.3. Iterative selection approach

In this analysis, an iterative ML approach is used. Two models are trained and applied in sequence in order to improve the selection applied by a single algorithm. A moderate BDT threshold is chosen on the first model score in order to reduce background without a high signal loss, similarly to the previous data skimming procedure. Then, a second ML model is trained on the remaining data sample according to the strategy described in 4.2. Using a ML based preselection a more precise classification is achieved by allowing to zoom into multidimensional correlation of features. While only differences in the one-dimensional distribution of signal and background were used in the data skimming, correlations between features are considered here in the ML selection. Allowing a too complex learning process results in overtraining very quickly though. Using a first model to remove the differences in low dimensional correlations, before training a second model on deeper correlations will allow to exploit these data characteristics without including random connections.

4.4. Extraction of signal

After having optimized the selection strategy, the shape of the signal and background candidates needs to be modelled by a fitting function. This will allow to extract the Ξ^- yield as it would be done in real data, where a MC label will not be present.

5. Results

5.1. Data skimming

The data samples contain candidates that evidently do not origin from Ξ^- decay, differing very much from the signal shape in various features as shown in Figure 5.1. This allows to manually optimize a set of one-dimensional constraints on those features to reject background candidates. The chosen selection criteria are listed in Table 5.1.

With all criteria applied, the background is reduced by 85% of the original size, at the same time only 4% of the signal is rejected. In spite of the unavoidable signal loss, this selection is necessary to reduce the background. With about 50×10^3 candidates per event, the background dominates the sample previous to the selection, while a true Ξ^- only comes to pass every sixtieth event on average. Due to this proportion, each reduction (even with small percentages) removes a large amount of background candidates, while only a few signal candidates are affected.

invariant mass range	$m_{\Lambda^0\pi_1^-}$	ϵ (1.28, 1.37) [GeV/c^2]
on Ξ^- decay	$\chi^2_{topo\ \Xi^-}$	< 3000
	$\chi^2_{geo\ \Lambda^0\pi_1^-}$	< 5000
	$DCA_{\Lambda^0\pi_1^-}$	$< 10\,\mathrm{cm}$
	L_{Ξ^-}	$> -10\,\mathrm{cm}$
on Λ^0 decay	$\chi^2_{geo\ p\pi_2^-}$	< 1000
	$DCA_{p\pi_2^-}$	$< 5\,\mathrm{cm}$
	L_{Λ^0}	$> -10\,\mathrm{cm}$
positively defined values	all χ^2	> 0

Table 5.1.: Selection criteria applied for data skimming

To choose the appropriate selection criteria, signal and background shapes are compared for variables where differences are expected. The most striking difference is visible in the reconstructed invariant Ξ^- mass (Figure 5.1, top left panel), which shows a clear signal peak at the expected Ξ^- mass value of 1.32 GeV, but a near flat background level. In terms of background reduction, constraining the allowed mass range is the most effective. By rejecting all candidates outside the accepted area (non-dashed area, equals more than 20σ), 81 % of the background are removed with a 2 % signal loss.

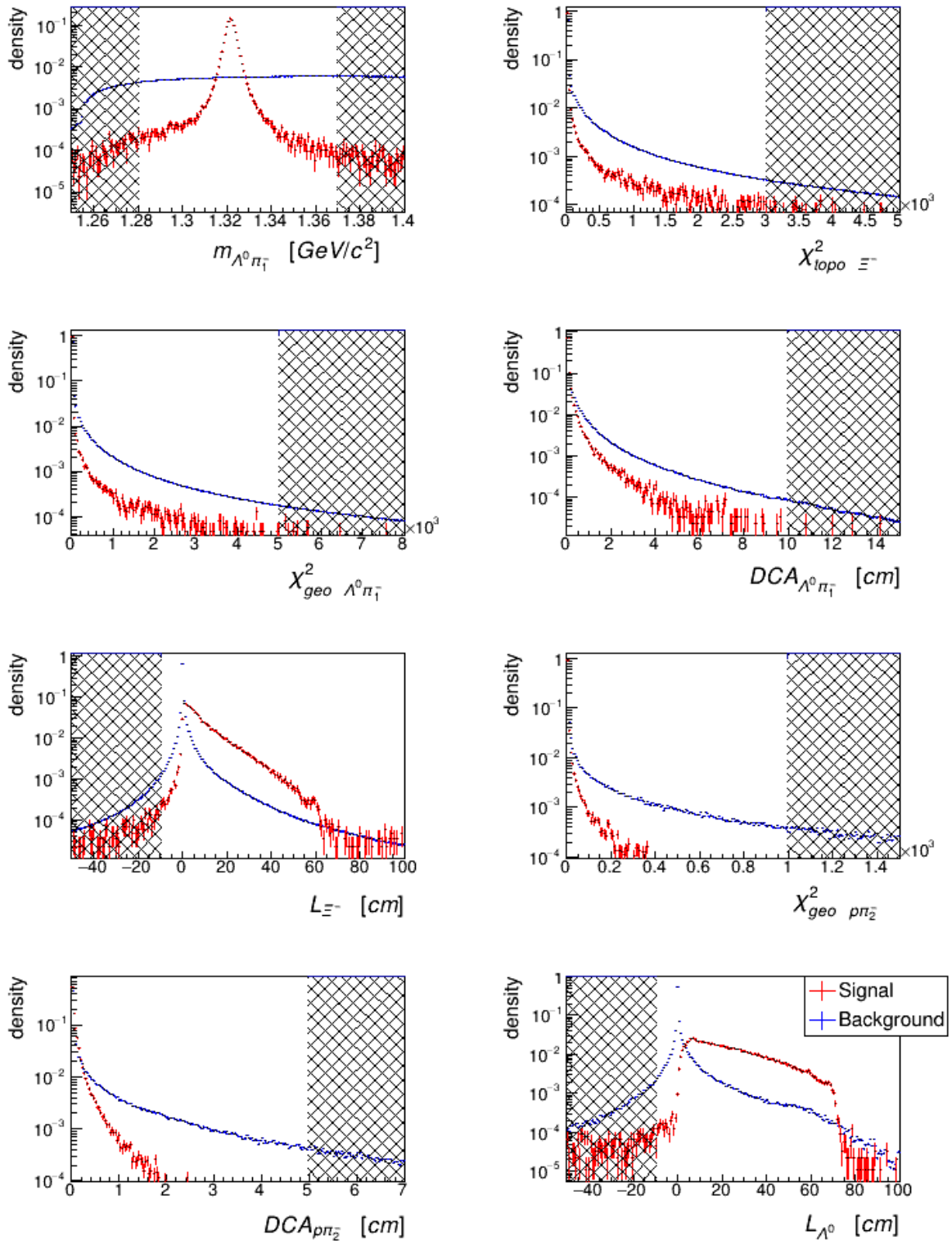


Figure 5.1.: Signal and background distribution in the features used for data skimming. The dashed area is rejected

To further select the reconstructed primary particle, the Ξ^- track is forced to originate close to the primary vertex (target position) by putting an upper limit on the $\chi^2_{topo\ \Xi^-}$ variable. This selects only candidates consistent with particles produced in the collision vertex, possibly removing secondary Ξ^- particles, but mostly combinatorial background. The reconstructed Λ^0 is a secondary particle in the Ξ^- decay and is therefore not necessarily expected to point back to the primary vertex. On the other hand, no proximity can be excluded due to the short Ξ^- lifetime, which is why $\chi^2_{topo\ \Lambda^0}$ remains unrestricted. Excluding candidates with large distances between the Ξ^- track to the primary vertex, 5 % of the background are rejected while 98 % of the signal are preserved.

Investigating variables that describe the decay geometry for both Ξ^- and Λ^0 decay, constraints can be applied in order to remove candidates with wrongly assigned daughter particles that do not originate in the same decay vertex. Quite intuitively, this includes candidates where the daughter tracks simply do not come close to each other. By restricting the DCA variables, those candidates are removed. An upper limit on the χ^2_{geo} variables removes candidates where the distance is too big in consideration of the uncertainties of the tracks. Rejecting the dashed area for higher values in all four variables plotted in Figure 5.1, combined 'daughter' particles that actually do not originate from the same decay vertex are removed. These geometrical constraints on the Ξ^- decay and on the Λ^0 decay remove 12 % and 3 % of the background respectively, while preserving 99 % of the signal candidates.

An additional constraint is applied on the decay length L of both Ξ^- and Λ^0. Being a scalar length, L is positively defined. In numerical calculations however, negative values can appear due to the finite precision in the reconstruction process. For actual signal candidates, this is highly unlikely, resulting in a strongly asymmetrical distribution (see Figure 5.1, L variables). For randomly combined pairs on the other hand, the effect takes place regularly, resulting in an enhanced distribution of background candidates with values below zero. To avoid a high signal loss, a selection rejecting only large negative values is chosen. The constraint on the Ξ^- decay length removes 2 %, the constraint on the Λ^0 decay length 3 % of the background with a 1 % signal loss.

Finally, candidates with values that cannot be mathematically justified are removed. Being a product of a squared value and an uncertainty, χ^2 variables cannot be negative by construction as in Eq. (3.2). In the reconstruction process, negative entries can appear due to numerical effects. For instance, the numerical calculation of the inverted covariance matrix can lead to negative entries. When rejecting all values below 0 for all χ^2 variables, about 1 % of all (signal and background) candidates are removed.

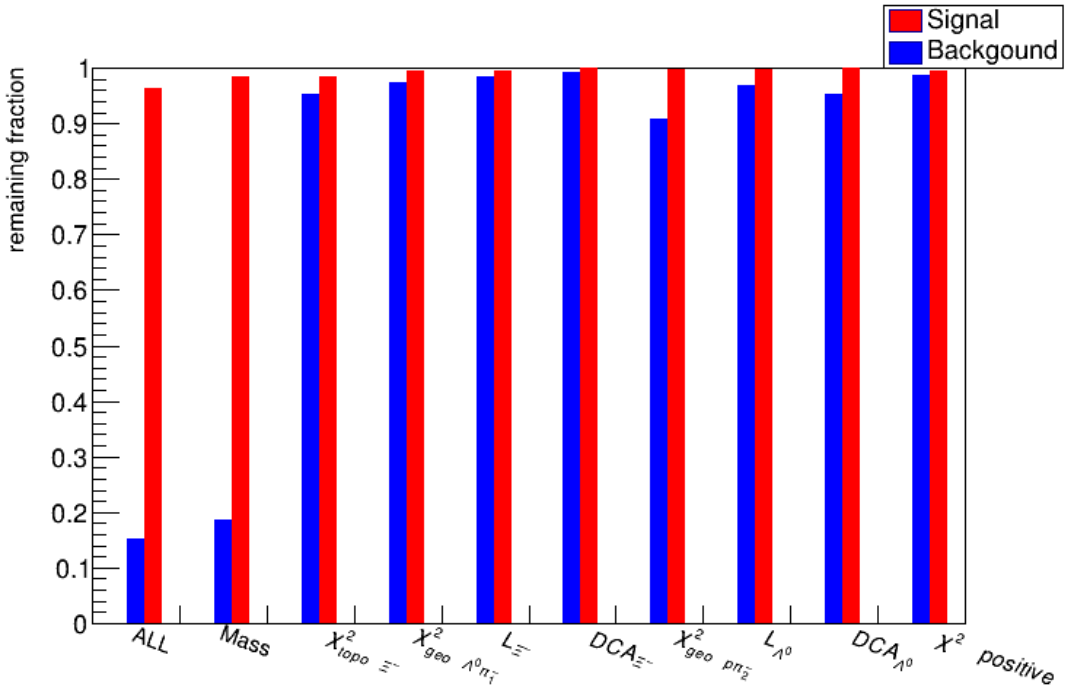


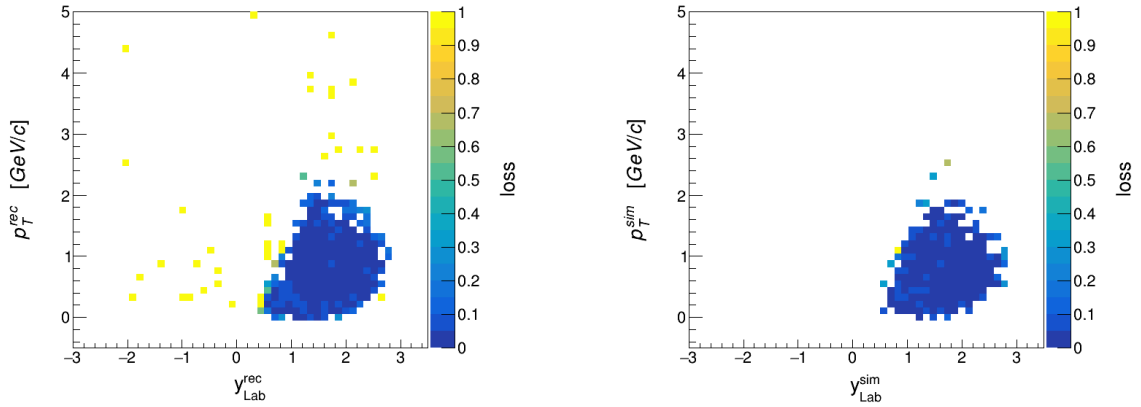
Figure 5.2.: Remaining data fraction after each selection criteria listed in Table 5.1 individually. The first bin shows the impact of all combined constraints

For each single constraint, the impact on both signal and background rejection is quantified in Figure 5.2. The remaining fractions may show small variations depending on the number of candidates in the used sample, however no significant deviations are expected. Seeing that only a small total count of signal is rejected overall, systematic effects are prevented. The remaining variations can be assigned to statistical fluctuations when the data sample was randomly sampled.

To verify that the signal loss does not preferentially happens on a specific region in phase space, the loss in rapidity y_{Lab} and transverse momentum p_T correlation is quantified in Figure 5.3(a). The loss is calculated as

$$signal\ loss = \frac{\#_{\text{signal before selection}} - \#_{\text{signal after selection}}}{\#_{\text{signal before selection}}} \tag{5.1}$$

Ξ^- with low rapidity, as well as high p_T seem to be particularly affected by the preselection, this originates in the reconstruction process however. Reconstructable signal candidates (simulated candidates with charged daughters that are reconstructed with the tracking system) do not necessarily coincide with the reconstructed candidate. Looking at the distribution for reconstructable Ξ^- in Figure 5.3(b), the loss is indeed uniform. Furthermore, the high p_T and low y_{Lab} points do not even exist for simulated data, but originate due to the finite detector and track resolution. The reinforced loss in this region is not caused by an imbalance in the phase space, but by the problems in reconstruction of the affected candidates.



(a) reconstructed Ξ^- (b) simulated and reconstructable Ξ^-

Figure 5.3.: Signal loss in transverse momentum, rapidity phase space with selection listed in Table 5.1. The yellow points in (a) indicate a high loss of signal for marginal candidates, which do not exist in the simulated candidates (b)

5.2. Selection with machine learning

Having reached a 2×10^{-6} signal to background ratio after data skimming, machine learning is used for a further optimization of the background and signal separation. To do so, a ML model is trained as described in section 4.2. The signal of five million events is used for the training, which equals about 85×10^3 MC-true candidates. Three times as many background candidates ($\sim 255 \times 10^3$ candidates) are added in order to reflect the overrepresentation of background in real data. To verify the necessity of this, the same training is done on a $1:1$ set. The resulting model shows a less precise classification, lacking knowledge on the data characteristics due to missing statistics of the widely distributed background.

For the training, signal and background samples are additionally separated in the reconstructed invariant Ξ^- mass (see Figure 5.4) in order imitate the approach used on real data without available MC labels.

The samples are then split into two parts: 60 % are taken to train the model, while the other 40% are used to test the model performance. In this way, any signs of overtraining can be identified by comparing the performance in train and test samples directly. In order to test the stability of the model, the sample sizes are varied, see section 5.2.2.

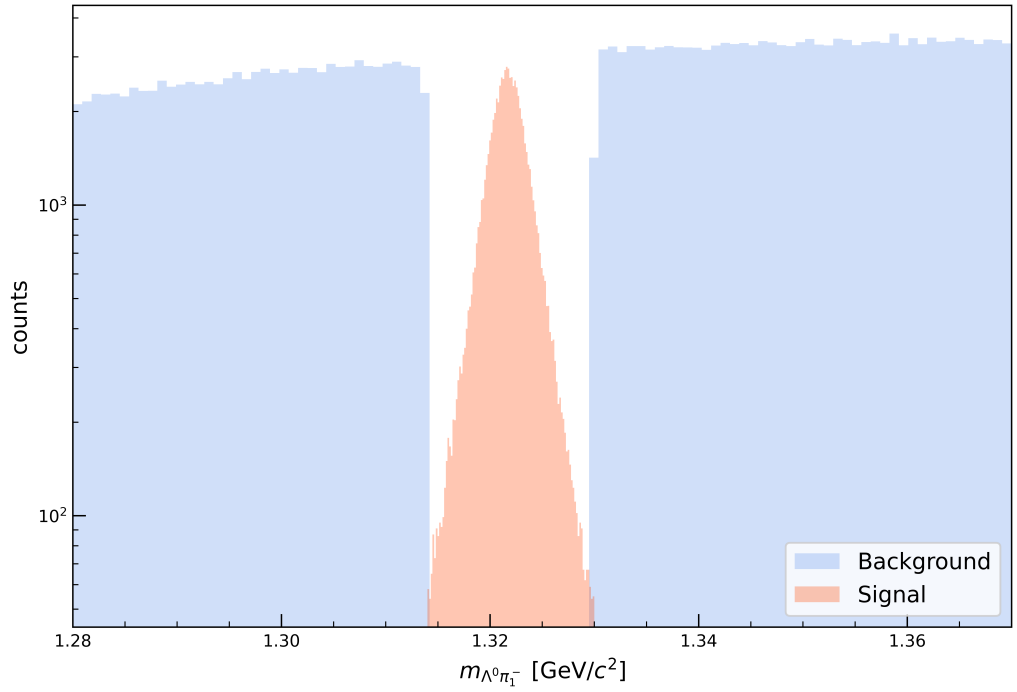


Figure 5.4.: Invariant mass distribution of the MC samples used for training and testing of the model (signal to background 1 : 3)

The training variables are chosen to include all topological features of both Ξ^- and Λ^0 decays (see section 3.2). Only the decay length L is removed to exclude double information also contained in the $L/\Delta L$ variable. Additionally, PID of the daughters in form of the TOF measurement m^2 is included for all three charged daughters if available, but not required (daughters that have not reached the TOF detector are not rejected). For the reconstructed Λ^0, a particle identification is added in form of the reconstructed invariant Λ^0 mass previous to the state vectors recalculation to match the real Λ^0 mass (see section 3.3). The shape of signal, as well as background distributions in the training features is visualized in Figure 5.5. It is verified that none of the chosen features show direct correlation to the Ξ^- mass in order to ensure an unbiased selection. The covariance matrices for signal and background distributions in the training features and Ξ^- mass can be found in the Appendix, Figures A.3, A.4.

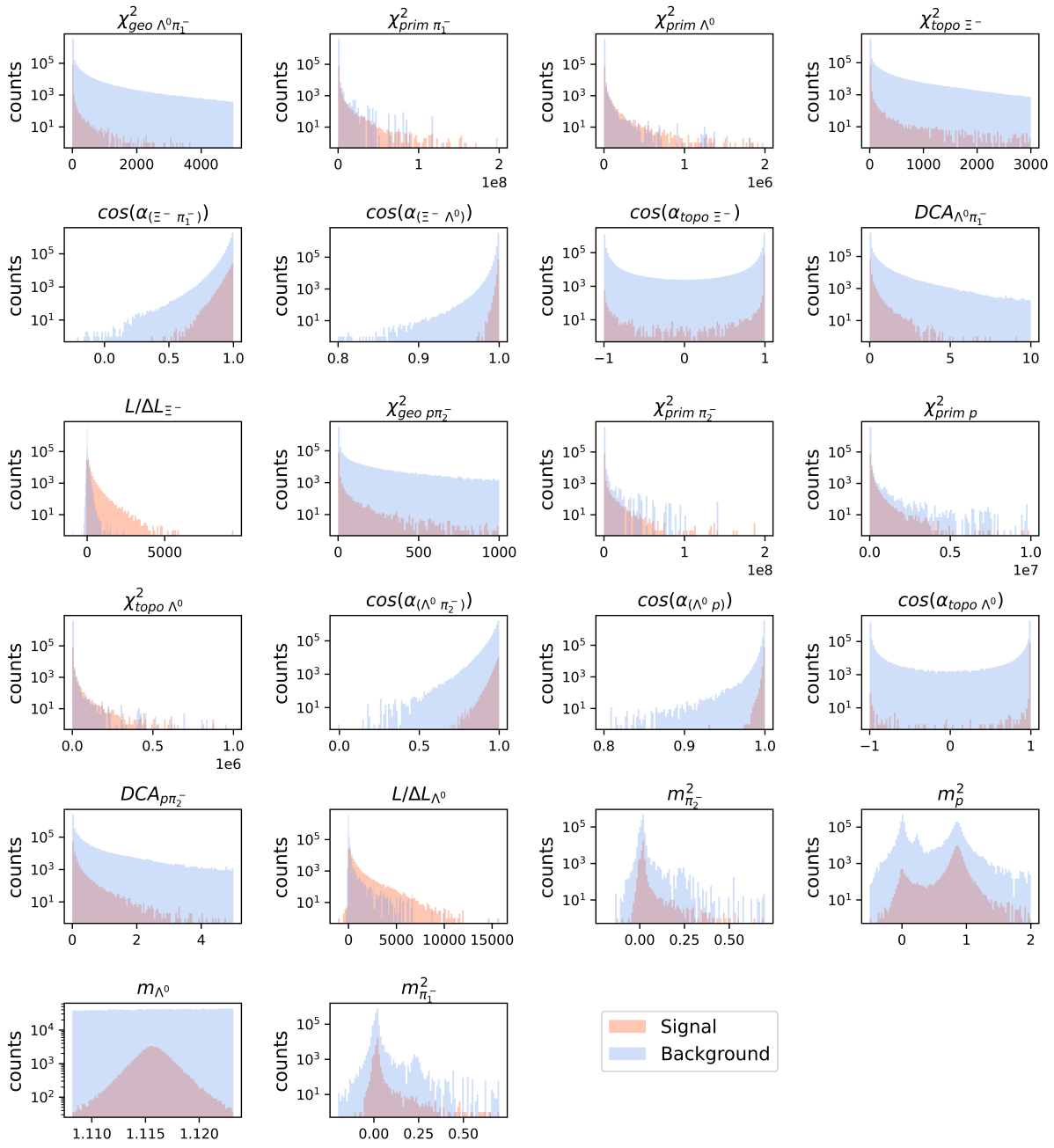


Figure 5.5.: Signal and background distributions for features used for model training

5.2.1. Hyperparameter optimization

In a first attempt, the hyperparameter ranges are chosen based on previously optimized selection strategies for short lived-particles without further grounds in order to identify the needed adjustments for the present sample in general. The resulting model shows proper classification, but the performance differs on test and train sample. Trying to prevent this overfitting without loosing efficiency, different restrictions on the hyperparameters are considered and tested.

- The maximum depth is reduced. Allowing only 3 decisions for each of the combined binary trees instead of 4 conceals deeper variable correlation. This forces the model decision to act on more general characteristics. Indeed, the output on train and test sets are more similar, even if not satisfyingly alike. Restricting the depth to 2 decisions eliminates the overfitting problem, but shows a lack of performance due to the shallowness.

- The allowed number of estimators is restricted. With fewer combined weak learners, the boosted decision tree is hoped to become more general by removing some of the less important, too specific, binary trees. Again, the output shows less, but not little enough signs of overfitting.

- The learning rate is halved. Making the influence of each single decision less important, the general features should outweigh the few random coincidences that are included in the classification. This can prevent overfitting without actually removing the wrongly included decisions, but only by weakening their effect on the output. In the present case, the reduction does not show the desired effect of preventing overfitting.

- Trying to reduce the model complexity without having to restrict hyperparameters, the allowed ranges of gamma, alpha and the positive weight scaling are increased. Contrary to the learning rate weighting each decision, alpha acts as weight on the regularization term, which (if increased) impedes a too complex model. Strengthening the requirements before allowing a new decision in each weak learners (increase in gamma) replaces a general limit on the maximum depth. An increased scaling reinforces the identification of signal over background, which should remove decisions based on statistical fluctuations in the background set.
  A more conservative model impedes the study of data characteristics with low impact on the general output. With the lacking statistical relevance of random properties, this can exclude the decisions that cause overfitting. However, the changes do not show any effect in reality. In the optimization with Optuna, the expanded ranges are not exploited. Forcing an increase of the three hyperparameters anyway does not keep the model from overfitting.

Since a more conservative model cannot compensate for the complexity sufficiently, limits are chosen for the number of estimators, maximum depth and learning rate. In this way, each restriction can be placed less tight individually, but prevent overfitting jointly. Both optimal hyperparameters and allowed ranges are given in Table 5.2.

hyperparameter	allowed range	optimized value
number of estimators	(100, 300)	206
maximum depth	(2, 3)	3
learning rate	(0.01, 0.05)	0.044
gamma	(0, 1)	1
alpha	(2, 20)	12
scaling positive weight	(1, 10)	3

Table 5.2.: Chosen ranges and optimized hyperparameters for the first XGBoost model

5.2.2. Stability

The learning curve describes the dependency of the model performance to the available data statistics. With increasing training set size, the model gains experience of the data characteristics, resulting in a more realistic classification. The root mean square error (RMSE) between model prediction and true label (0 for background, 1 for signal) decreases and converges to a minimum value as shown in Figure 5.6. The curves of train and test set do not differ much, indicating little to no overtraining. Containing a total count of $\sim 2 \times 10^5$ candidates, the used training sample is located in the stable region with a minimal error. The model training is expected to not depend on statistics and to show no deviation for a slightly modified training sample size.

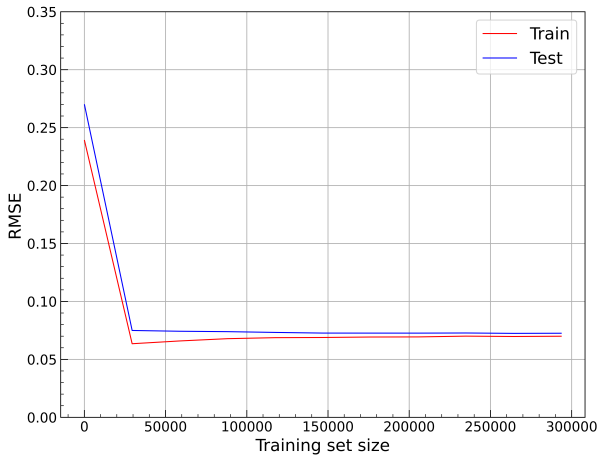


Figure 5.6.: Learning curve of the first XGBoost model.
The RMSE is calculated in dependency to the total count of candidates (signal + background) contained in the sample used for training

Making sure this is not a 'lucky size', resulting in a maximal performance due to an exceptional data set, the training sample size is varied from 40 % to 80 % of the total set (before, 60 % were used for the training, with the remainder available as independent test sample). The variation equals about 1.4 to 2.6×10^5 total candidates (signal+background) respectively. To compare the

training process, the resulting optimized hyperparameters for the different samples are listed in Table 5.3. Seeing no difference in the training as well as the in model output, the model is taken as stable and not dependent on the train and test sample.

hyperparameter	small set (40 %)	default set (60 %)	large set (80 %)
number of estimators	219	206	208
maximum depth	3	3	3
learning rate	0.042	0.044	0.039
gamma	1	1	1
alpha	11	12	8
scaling positive weight	4	3	2

Table 5.3.: Optimal hyperparameters for different training set sizes when allowing the same ranges as given in Table 5.2. The similarity is an indicator for the stability of the model

5.2.3. Feature importance

Investigating the importance of each feature for the model training, the shapley additive explanation (shap) values are calculated (Figure 5.7). Shap values quantify the importance of each feature in the classification output. To understand the process, the influence is also visualized depending on the feature value (Figure 5.7, right part). The first entry for example shows that a low distance of the proton track to the primary vertex might strongly decrease the output value (indicating a background candidate), while a higher signal probability is only achieved for a large distances. This is in accordance with the assumption that the daughter particle track does not point back to the collision point.

Overall, it only a few features have a significant impact on the model decision, while showing interesting correlations. The three most important features can be used to remove wrongly combined daughter that are actually produced in the collision directly (see section 5.3.2). The models objective does not seem to remove random combinations as expected, but to separate those primary (non-)daughters. Particularly, primary protons seem to be singled out. Only some of the less important features distinguish false combinations rather than primary particles. Contrary to the three most important features that are used to reject secondary particles, $\chi^2_{topo\ \Xi^-}$ qualifies the Ξ^- as primary particle. High values indicate a high distance between Ξ^- track and primary vertex, which allows to reject the candidate as background. A preferably forward motion of both Ξ^- and Λ^0 causes larger values for both $cos(\alpha_{topo})$ variables as favored for a signal classification, again identifying tracks that cannot be associated to particles produced in the collision. This can be understood as a rejection of false combinations which do not point

back to the collision rather than a separation of secondary Ξ^-, which are not included in the data sample used for training. Only the Λ^0 decay length $L/\Delta L_{\Lambda^0}$ indicates an importance of the daughters common origin. The favoured high value for a signal classification happens for a high resolution of the decay vertex, which in turn is achieved for matching daughter tracks. Features concerning the decay structure itself, verifying the common origin of daughters like DCA, χ^2_{geo} or connecting daughters with mother tracks like $cos(\alpha)$ as well as the PID of daughters (mass information m^2) seem to be insignificant in the models decision making.

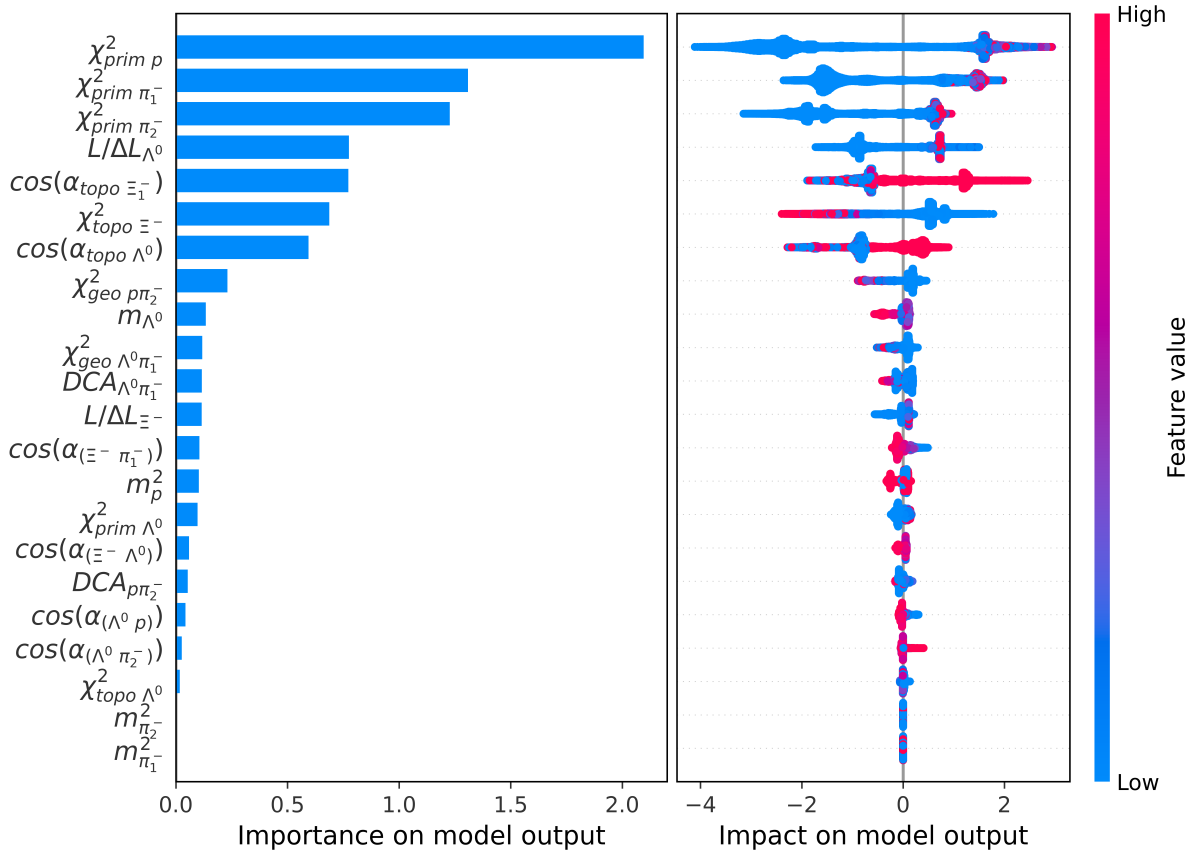


Figure 5.7.: Shap values for features used in training

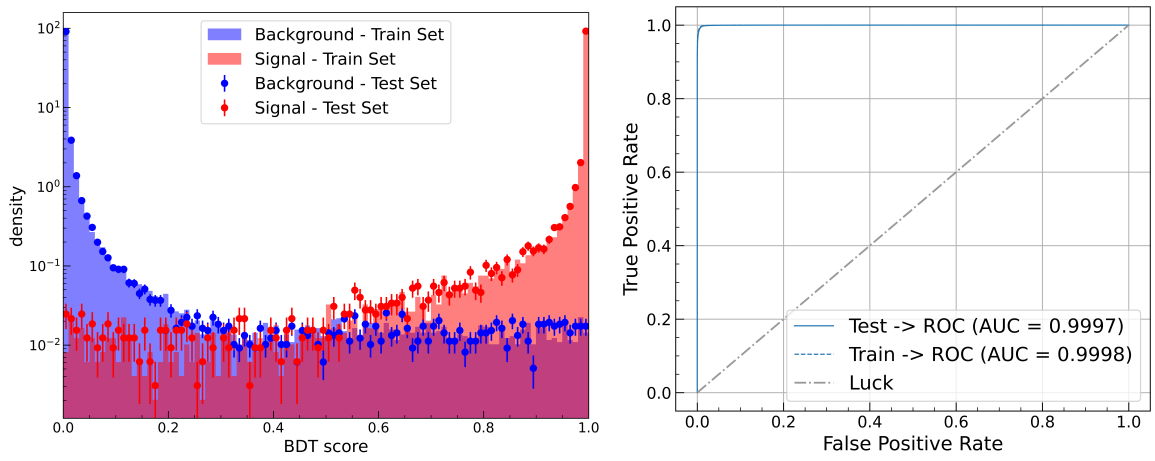
To test the stability of the model with respect to the excessive number of not important variables, the eight least important variables are excluded from the training as a test. The resulting model does not show any less signs of overfitting, however the proven stability (see section 5.2.2) is lost. Increasing the sample size used for training changes the optimal training process. Limited to the same ranges, the optimal hyperparameters vary strongly for different sample sizes, indicating an unintended dependency on statistics. Even though the features seem to be unimportant for the classification itself, their availability is significant for the models stability. The features therefore remain included in the training.

5.2.4. Model performance

In order to visualize the model output, the boosted decision tree (BDT) score is plotted separately for signal and background in both test and training sample in Figure 5.8(a). The shape of MC-true signal differs much from the background candidates, which makes a selection possible. Understanding the BDT score as the probability of a candidate being signal, the accumulation of signal candidates at high scores (and background at low scores respectively) proves the successful classification. A machine leaning based selection by accepting only candidates above a certain BDT score will allow to separate signal candidates.

Looking at the receiver operating characteristic (ROC) curve in Figure 5.8(b) confirms the prediction accuracy. Depending on the applied BDT threshold, this compares true positive (MC-true signal and selected) with false positive counts (selected data that is actually background). For a complete random classification this would result in a identity curve with an integral of 0.5; a perfect classification would provide a Heaviside step function with an area under the curve (ROC-AUC) of 1. The achieved score of both train and test set of above 0.999 again shows the efficient model performance.

Both plots verify that in the optimization process there is no overtraining. A differing performance in train and test samples would result in a different shape in both BDT output as well as ROC curve. Seeing that both train and test shapes are nearly indistinguishable proves the independency of the output to the used sample.



(a) Model output

(b) ROC curve

Figure 5.8.: Output and ROC curve quantify the model performance with optimized set and hyperparameters

5.2.5. Selection optimization

The optimal selection (maximizing the significance as explained in section 4.2) is achieved for a selection at BDT > 0.9994. While this gives a high signal to background ratio of 24.9 in the 5σ mass range, with only 49% the signal efficiency is not quite satisfactory. A selection at a lower BDT threshold would improve the efficiency, but lack in background rejection. The course of all, significance, efficiency and signal to background ratio for different thresholds is shown in Figure 5.9.

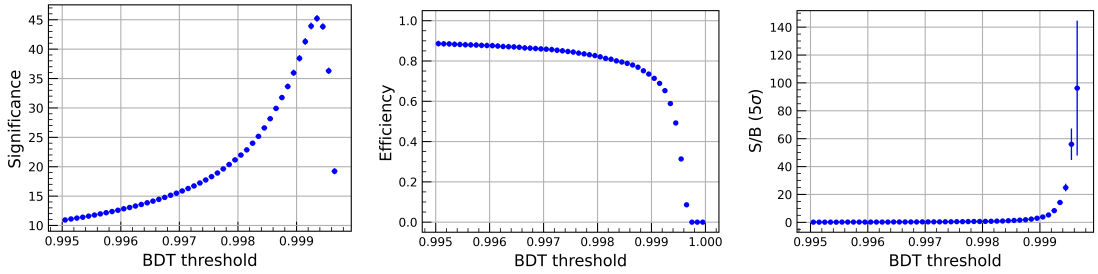


Figure 5.9.: Significance, efficiency and signal to background ratio in dependence to the applied BDT threshold

Trying to understand the selection process, the five most important features in training (compare Figure 5.7(a)) are visualized. To get an idea of the models classification process, the distribution of both signal and background candidates is compared before and after a moderate selection in Figure 5.10. A selection at BDT > 0.6 achieves both an efficiency and a background rejection of more than 99%, allowing to demonstrate the differences without a substantial loss of signal.

While looking at the single variable distribution (diagonal), only small differences strike in signal and background distributions. However, in the two dimensional correlation of features (below-diagonal), the background and the signal distributions show significant differences, underlining the importance of the multidimensional selection for a signal/background separation. After applying a BDT selection, those differences are reduced drastically (above-diagonal).

As expected, background candidates that differ visibly from the signal distribution are removed in the selection. This happens similarly to the data skimming done before (see section 5.1). While this was done in one dimension only, machine learning considers more dimensional correlations for the classification. Trying to improve the selection, the resemblance of the models classification process to the preselection brings up the idea to use it as such, and to train (yet another) BDT model on the (now twice) preselected sample.

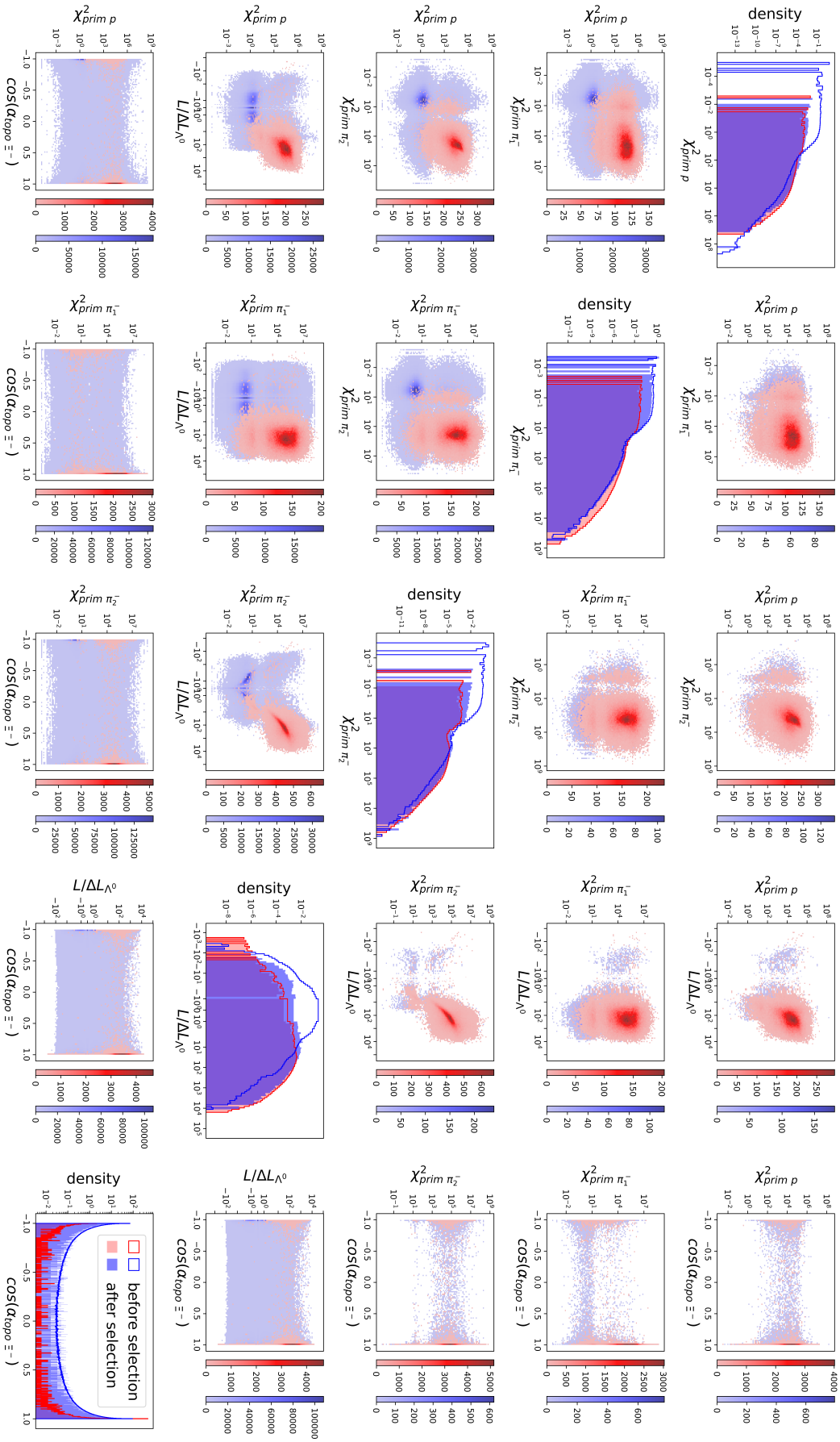


Figure 5.10.: 2D distribution in most important training features
Red:Signal Blue:Background
unselected data below diagonal, BDT > 0.6 selected data above diagonal

5.2.6. Iterative selection

A moderate threshold on the first BDT (> 0.6) is applied in order to reduce background without actually rejecting signal. A second model is trained on this selected sample. To be able to compare the two strategies, the training is carried out identical for both models. Parallel to the training of the first model, a $60 : 40$ train-test set is formed from signal of five million events (without the candidates rejected in the first model selection) and a $3 : 1$ background proportion, which enables a more efficient classification than same sized sets. The same features are made available in the training, containing topological features of each decay (section 3.2) and mass information of all daughters. With this equality in the training process, any differences can only origin due to the prior selection of the samples, indicating if a further background rejection previous to ML selection improves the output as assumed.

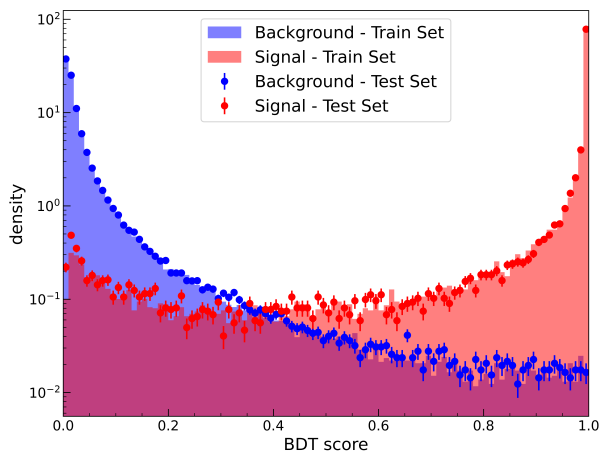
In the optimization of the hyperparameters with the Optuna framework, estimators and learning rate need to be limited in order to avoid overtraining (again, increasing ranges for gamma, alpha and positive weighting does not improve this). Still, in contrast to the first model the limits can be chosen more relaxed without causing an overfit, particularly a maximum tree depth of 4 can be allowed, indicating a change in the training already. The optimized hyperparameters as well as the allowed ranges are given in Table 5.4.

hyperparameter	allowed range	optimized value
number of estimators	(100, 450)	327
maximum depth	(2, 4)	4
learning rate	(0.01, 0.085)	0.066
gamma	(0, 1)	1
alpha	(2, 20)	17
scaling positive weight	(1, 10)	2

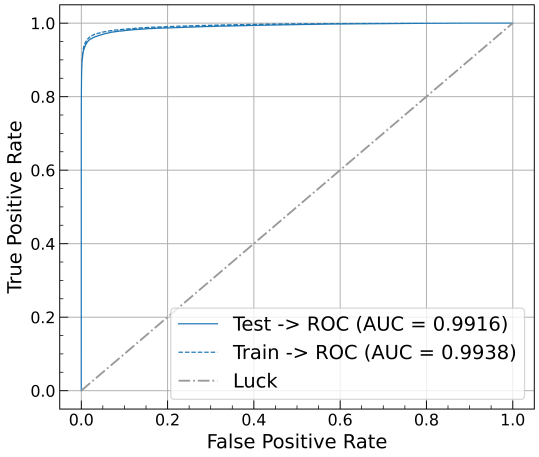
Table 5.4.: Optimal hyperparameters for second XGBoost model

Again, BDT output and ROC curve (Figure 5.11(a), (b)) show that the performance of the second model does not differ in test and train set, verifying a non-overtrained model. While the overall ROC-AUC is slightly worse than in the first model, with over 0.99 it is still satisfying. This proves that even though the data sample has been skimmed twice, manually and with machine learning, a second model is able to discover enough new properties for a further classification. A second selection based on the second BDT score is therefore expected to provide an improved signal/background separation.

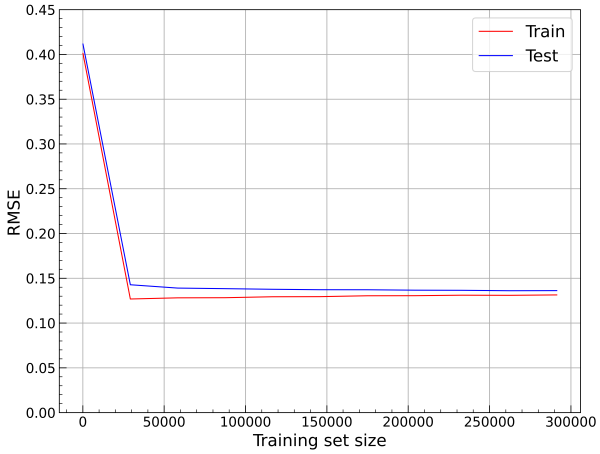
Looking at the learning curve in Figure 5.11(c), the statistics again prove to be sufficient. With a training set size of $\sim 2 \times 10^5$ candidates in total, the sample lies in the stable minimal error range. The curves for train and test set do not differ here, indicating no overfitting.



(a) Model output



(b) ROC curve



(c) Learning curve

Figure 5.11.: Model output, ROC curve and learning curve quantify the performance for the second model

As expected, comparing the shap-values proves that more of the features are of importance in the second model training (compare Figure 5.7, 5.12). Even though there are still features available that do not seem to be used at all, the classification does not rely on one single feature mostly. The prominent features of the first model appear as important again (compare section 5.2.3), indicating that these properties have not been exploited yet. Still, previously unused features are now included in the training as well, proving that the second model follows a different process. With $\cos(\alpha_{topo\ \Xi^-})$ and $\chi^2_{topo\ \Xi^-}$, the classification of Ξ^- as primary particle seems to be prioritized. In contrast to the first model, DCA and χ^2_{geo} variables are significant, meaning that the model rejects candidates with wrongly combined daughter particles that actually do not originate at a common vertex. This separation happens based on the new important decay features mostly, but since correlations are included on a deeper level, the old, already used features reappear in the importance and can be exploited further.

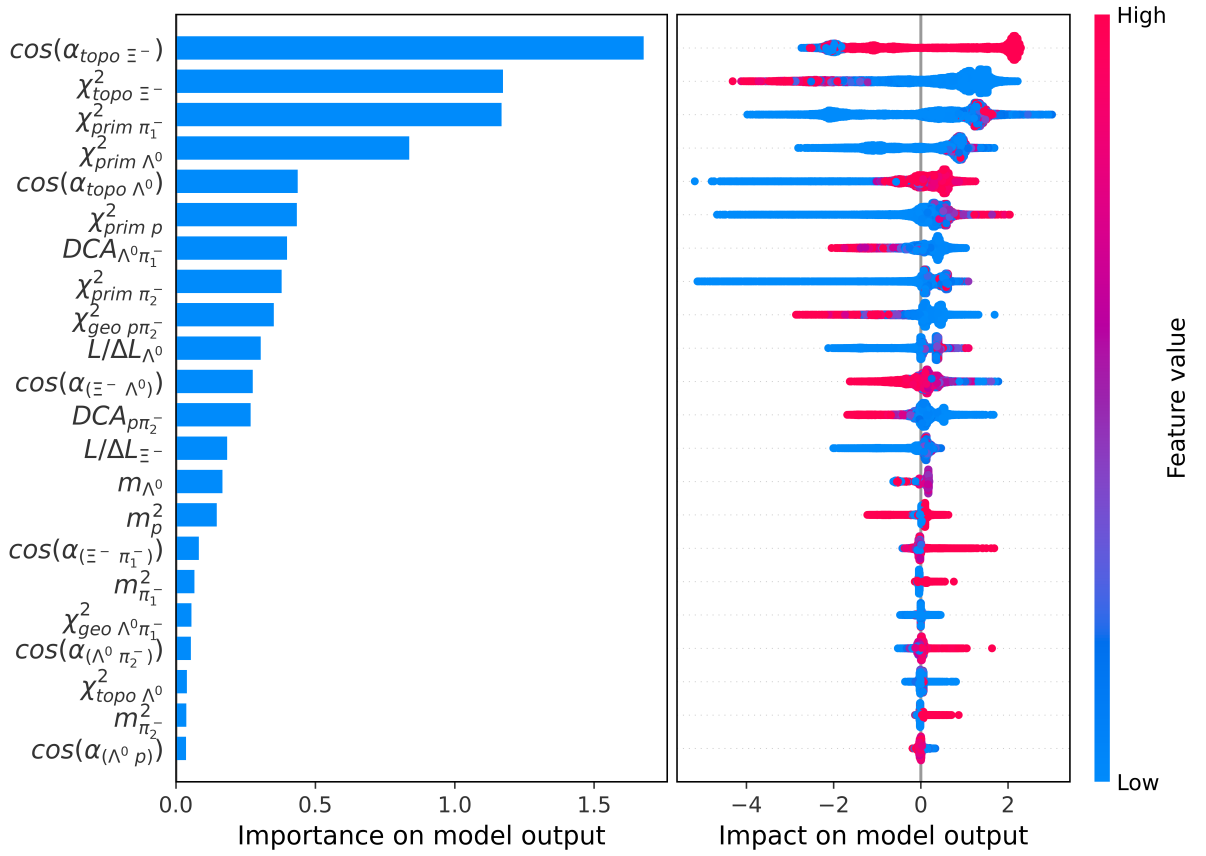


Figure 5.12.: Shap values for features used in second model training

The increased depth of the second model allows to find these correlations and use them for classification. This validates the motivation for an iterative model selection. A possible explanation for the improvement lies in the features used in the first model training. As described in section 5.2.3, the three main features can be used to reject the three measured daughter particles as primary particles produced in the collision itself. Allowing a single model to zoom deeper into correlations, their special and similar characteristics are mixed with random appearances, leading to an overfitting model. Rejecting the candidates before training a second model on the other hand, those properties are removed already, allowing to exploit finer differences in the feature correlation of general background candidates that do not show similarities due to their common origin. These newly used traits now allow to improve the classification.

5.3. Selection output

An optimal selection based on the models classification is now found again by maximizing the significance (Eq. (4.1)). This is done in dependence to the BDT score of the second model, with a fixed threshold of 0.6 on the first models score. The maximum is found for a threshold at BDT > 0.997, leading to a selection with 68% signal efficiency. The mass shape of the remaining data sample after selection in Figure 5.13 shows a significant peak at the expected position. A signal to background ratio of 39.0 is achieved in the 5σ range around the peak (1.312 – 1.332 GeV).

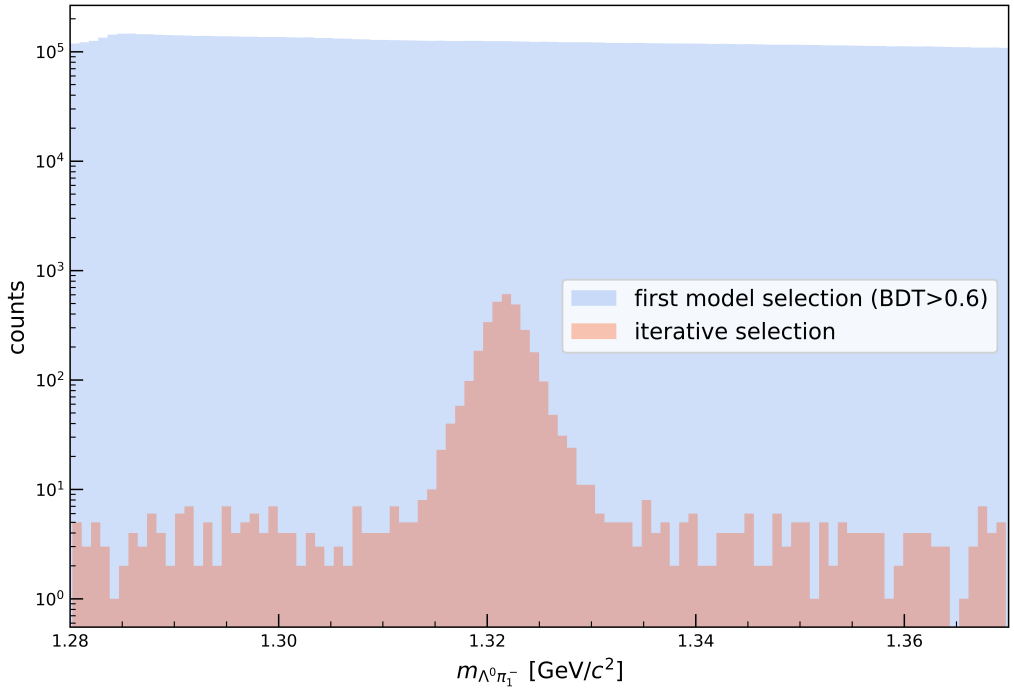


Figure 5.13.: Invariant mass distribution after iterative selection compared to single model selection at BDT > 0.6

5.3.1. Performance comparison between single model vs. iterative selection

To compare the single model selection (section 5.2.5) to the iterative approach, they are applied on the same sample. The performance of both is visualized in Figure 5.14 as a function of the applied BDT constraint. While the overall efficiency of a dual model selection is reduced, the significance follows a constant, higher level. Looking at the ratio of signal to background in the selected data samples, the iterative selection improves the output drastically for a higher BDT threshold.

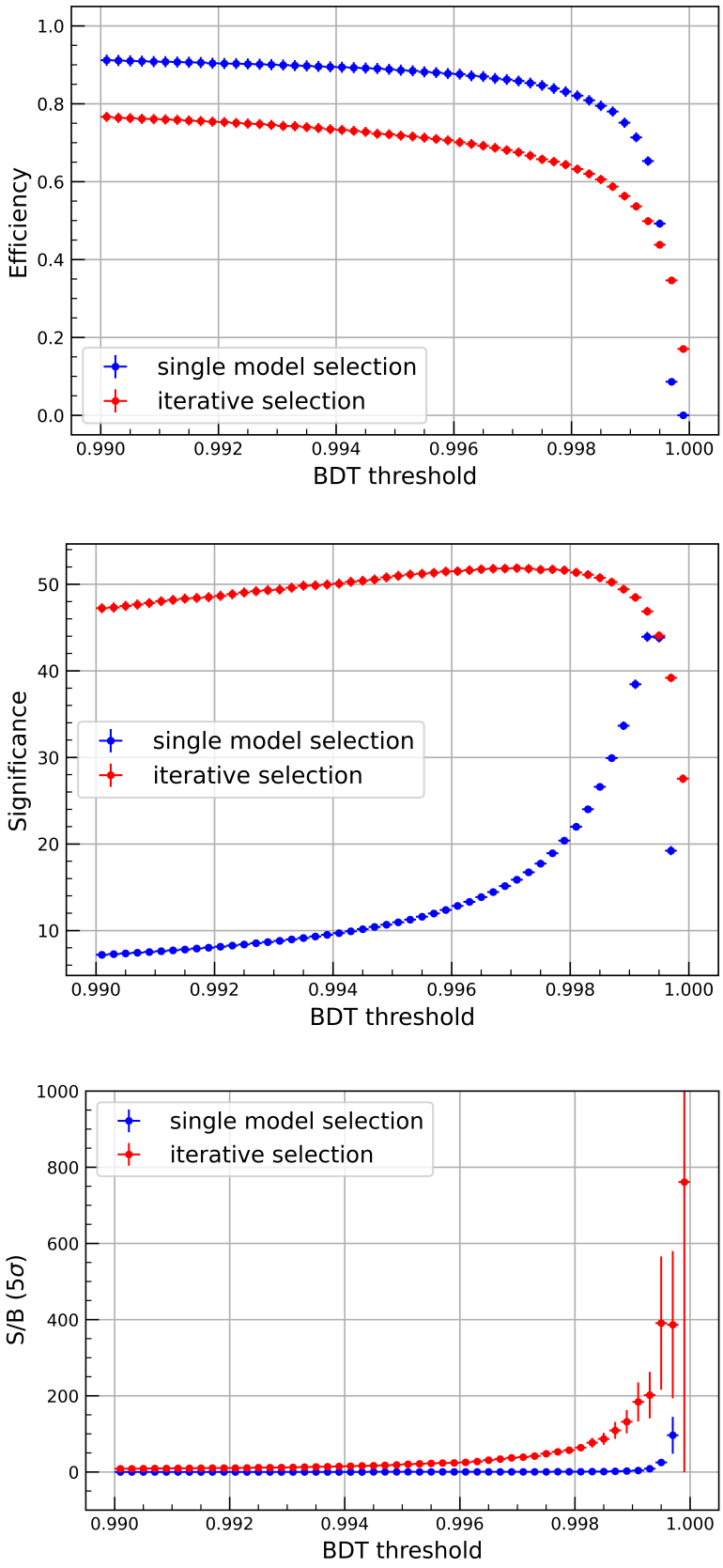


Figure 5.14.: Efficiency, Significance and signal to background ratio in dependence to applied BDT threshold for both single model and iterative selection

Considering that a maximized significance is desired, the optimal selection is achieved at a lower BDT constraint in the iterative approach. This reduced threshold allows for an equally strong background rejection with a smaller loss of signal candidates at the same time. At the optimal selection threshold, the two-model selection achieves both a higher signal efficiency and a higher signal to background ratio as given in Table 5.5.

	single model	iterative selection
optimal BDT	0.9994	0.997 (on second BDT model)
significance	44 ± 1	52 ± 1
signal efficiency (%)	49 ± 1	68 ± 2
S/B (in 5σ range)	24.9 ± 2.7	39.0 ± 4.5

Table 5.5.: Performance at optimal BDT threshold

Visualizing the different selection performances, the invariant mass distribution of the selected candidates is compared in Figure 5.15. While the distributions do not differ much for small masses, less candidates remain after a single model selection for higher masses, resulting in an asymmetric level at the edges. The iterative selection on the other hand allows for a symmetric selection with a higher mass peak, resulting in a increase when comparing the proportion of both selections. This asymmetry emerging in a single model selection indicates that an iterative selection is not only more precise, but also less dependent on the reconstructed mass.

Overall, the iterative selection approach allows for a more precise, more efficient and less biased selection. Allowing to exploit different features after having removed a large part of similar background candidates based on (mostly) only one feature improves the classification and enables a better signal extraction.

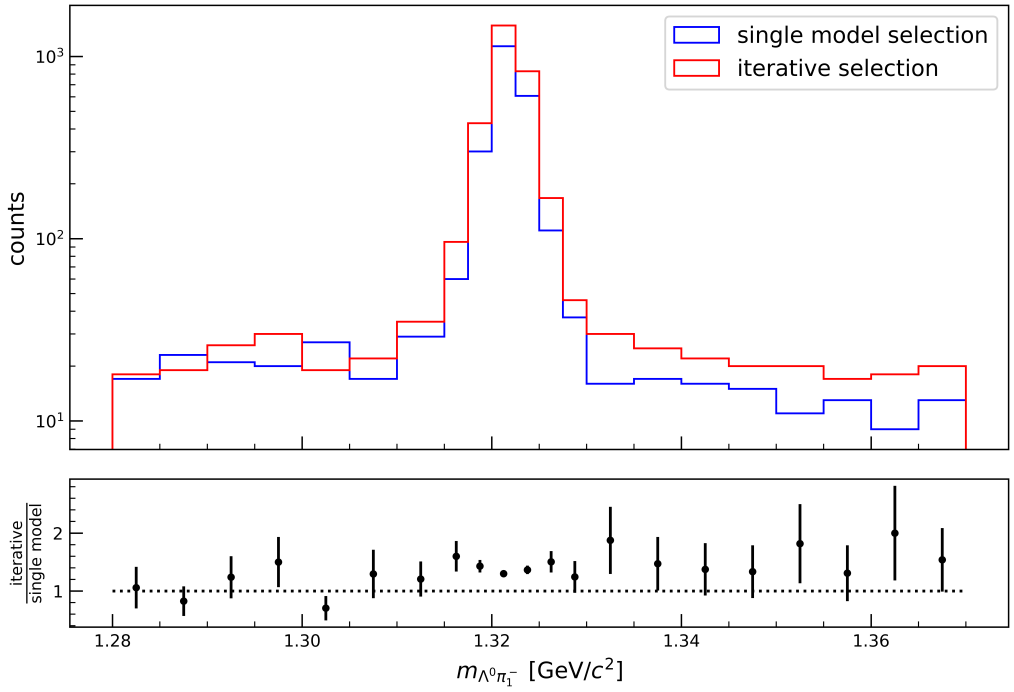


Figure 5.15.: Invariant mass distribution after selection, comparing single model vs. iterative selection

5.3.2. Performance comparison between iterative vs. box selection

While a selection with machine learning techniques is expected to achieve precise results, a lot of time and computational resources are needed for the training and selection optimization, particularly to process large amounts of statistics. As a faster and resource-saving way of signal extraction without the need of ML, a standard selection is implemented in the Kalman Filter Particle Finder for particles in over fifty decay channels [26]. Based on the respective decay geometry, it contains rectangular selection criteria to remove background candidates. More details to this selection strategy can be found in [21].

The default criteria for the Ξ^- selection include lower limits for the χ^2_{prim} variables of the three measured secondary tracks (both pions and proton) as well as on $\chi^2_{topo\ \Lambda^0}$. This verifies that the tracks of all secondary particles do not lead back to the primary vertex, rejecting assigned daughter particles that actually origin in the collision directly. The Ξ^- on the other hand, being a primary particle, is required to track back to the PV by putting an upper limit on the $\chi^2_{topo\ \Xi^-}$ variable. To ensure the daughters share a common origin, both DCA and χ^2_{geo} variables for both decays are constrained. In addition, the decay length relative to its error $L/\Delta L$ is restricted for both Ξ^- and Λ^0, verifying a high resolution of both decay vertices. By restricting the Λ^0 invariant mass to a 3σ range of the expected mass peak, the particles identity is verified, resulting in a better resolution of the reconstructed Ξ^- decay.

selection on Ξ^- decay		selection on Λ^0 decay	
$\chi^2_{topo\ \Xi^-}$	< 5	$\chi^2_{topo\ \Lambda^0}$	> 5
$\chi^2_{geo\ \Lambda^0\pi^-_1}$	< 6	$\chi^2_{geo\ p\pi^-_2}$	< 3
$\chi^2_{prim\ \pi^-_1}$	> 18.42	$\chi^2_{prim\ \pi^-_2}$	> 18.42
		$\chi^2_{prim\ p}$	> 18.42
$DCA_{\Lambda^0\pi^-_1}$	< 1	$DCA_{p\pi^-_2}$	< 1
$L/\Delta L_{\Xi^-}$	> 5	$L/\Delta L_{\Lambda^0}$	> 10
		m_{Λ^0}	in 3σ range

Table 5.6.: Default box selection criteria for the Ξ^- particle

The box selection criteria for the Ξ^- particle are given given in Table 5.6. In contrast to the machine learning selection, no PID information is used for the three measured daughter particles. Including the TOF-mass information for both pions and the proton in the ML training improves the classification, but makes the two selection strategies nonequivalent. In order to enable an appropriate comparison, a PID restriction is added to the box selection. To do so, a fourth order polynomial is used to constrain the relation of squared mass and momentum of the three measured particles as shown in Figure 5.16.

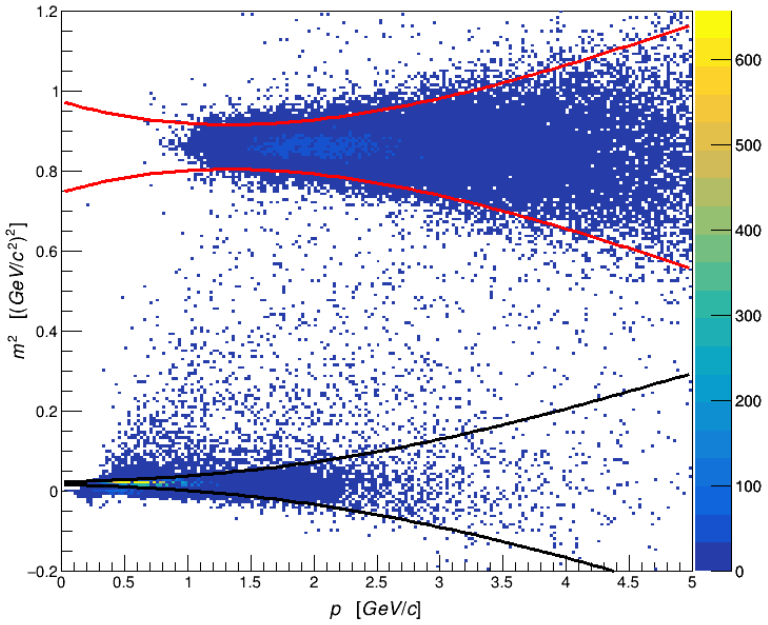


Figure 5.16.: Squared mass - momentum relation of the charged daughter particles (pions, proton) for a signal only sample with the selection criteria described in section 5.1 applied. The additional PID constraint added to box selection excludes protons are excluded outside the red band and pions outside the black bands respectively

Applied on the data sample used in this analysis, the box selection succeeds in selecting a significant signal peak. Consisting almost only of linear constraints, this proves the precision of the rectangular selection strategy. Still, comparing to the previously optimized machine learning selection, it does show a deficit concerning efficiency. While a signal loss is inevitable in order to reject enough background with linear constraints, machine learning allows to select candidates in a multi dimensional parameter space. This leads to a more efficient and precise selection as displayed in Table 5.7.

	box selection	iterative selection
significance	36 ± 1	52 ± 1
signal efficiency (%)	44 ± 1	68 ± 2
S/B (in 5σ range)	9.6 ± 0.7	39.0 ± 4.5

Table 5.7.: Performance of box selection vs. iterative selection

Comparing the selection strategies in the mass shape of the remaining data sample in Figure 5.17, the iterative approach shows a higher signal peak as well as a lower, flat background level outside of the Ξ^- invariant mass peak region. This results in a clear peak when considering the proportion of the remainder, again proving the precision of a selection with machine learning techniques.

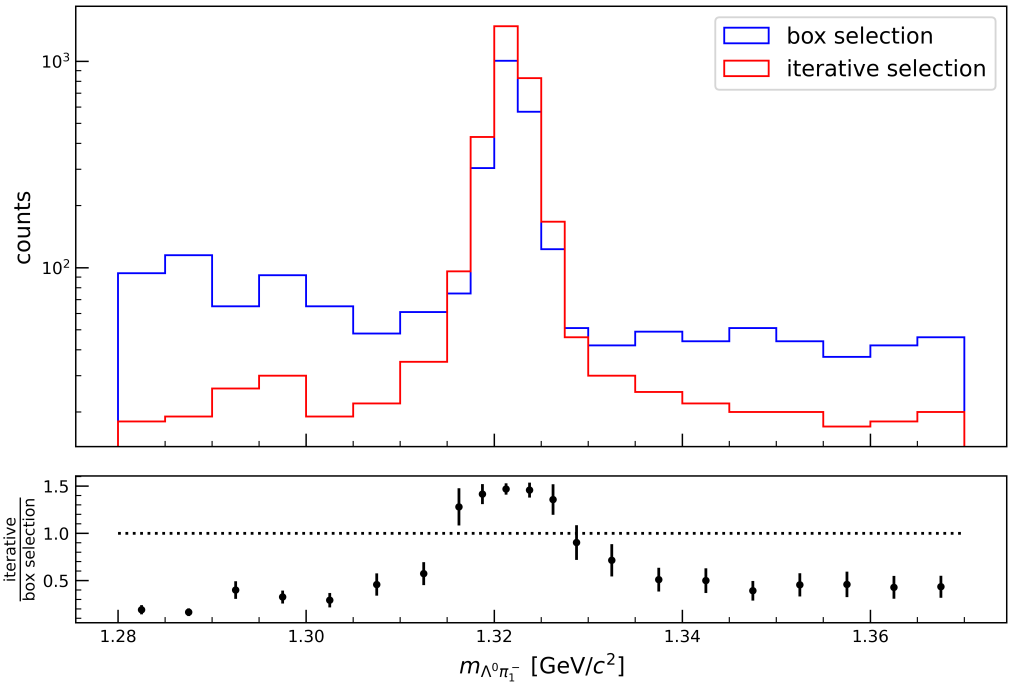


Figure 5.17.: Invariant mass distribution after selection, comparing box and iterative selection

5.3.3. Efficiency study in (p_T, y_{Lab}) phase space

Applying the complete presented selection strategy including data skimming and iterative ML selection on a pure sample achieves a signal efficiency of 67 ± 2 %. While this is sufficient overall, the efficiency can vary severely in different phase space intervals, i.e. in the transverse momentum p_T, rapidity y_{Lab} correlation. To make sure the selection does not have a 'blind spot' with lacking accuracy, the proportion of remaining signal after the selection process is displayed in (p_T, y_{Lab}) dependency in Figure 5.18.

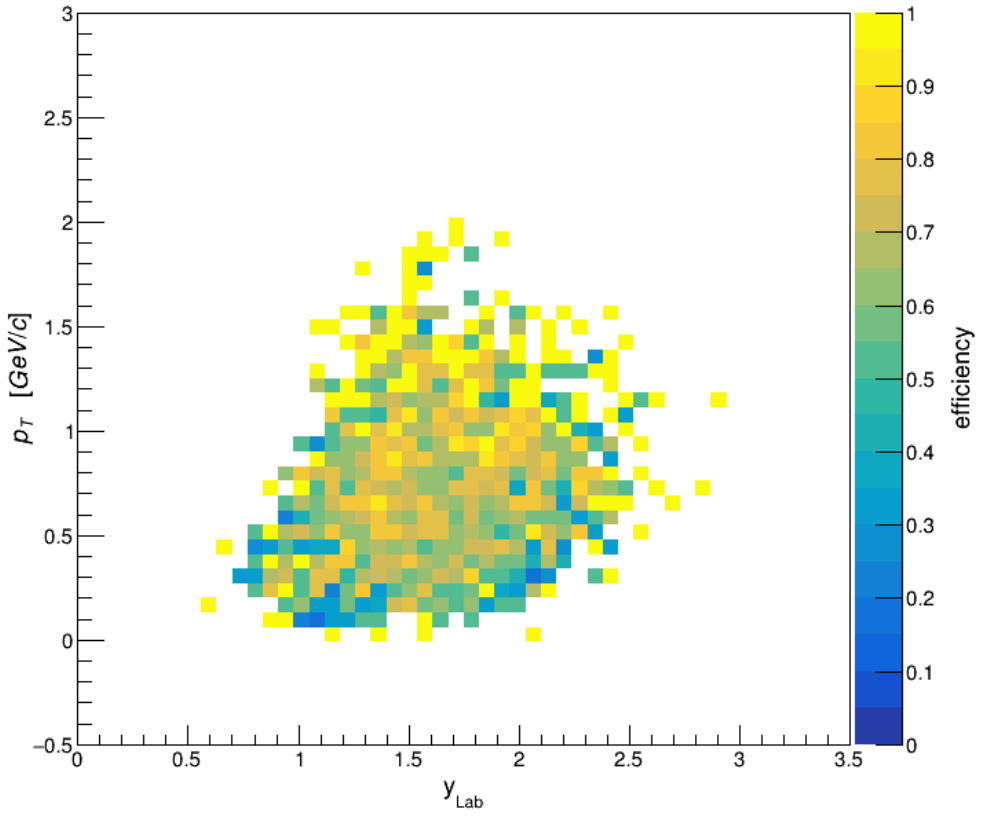


Figure 5.18.: Efficiency of the presented selection strategy (data skimming and iterative ML selection) in rapidity, transverse momentum space

Indeed, the efficiency shows a trend indicating a poor selection for low p_T and low y_{Lab} especially. In this region, the characteristics seem to differ too much from the common trend to enable a precise general description. This does not origin in the selection, but is explained by the experimental setup however. Candidates with extreme momentum values (very high, as well as very low) are placed on the corners of the CBM acceptance. Not all daughter particles might hit all detectors and might therefore not be reliably reconstructable, which falsifies the description of the reconstructed Ξ^- candidate. Those differences might then be targeted in the selection, which in turn shows a strong impact in the corner region.

The imbalance of the signal efficiency is therefore not taken as flaw in the selection itself, but as experimental disadvantage. Still, a multi-differential selection optimization might compensate for this. For this, independent ML models are trained for different (p_T, y_{Lab}) bins, allowing each to concentrate on the most prominent features in the respective sector, which in turn is expected to improve the efficiency in the separate regions.

5.3.4. Raw signal yield extraction

After optimizing the selection strategy, the remaining data sample will need to be characterized by a fitting function. This will allow to extract reconstructed signal yield without the the need of MC labels. While the presented selection strategy in this thesis is not fully optimized, the general output should already resemble the overall expectation. To enable a suitable description, the expected shape of the invariant Ξ^- mass distribution is modelled. For this, four different fit functions are tested on a MC-true data sample containing only signal candidates with no selection applied. In Figure 5.19, the resulting best fits are visualized.

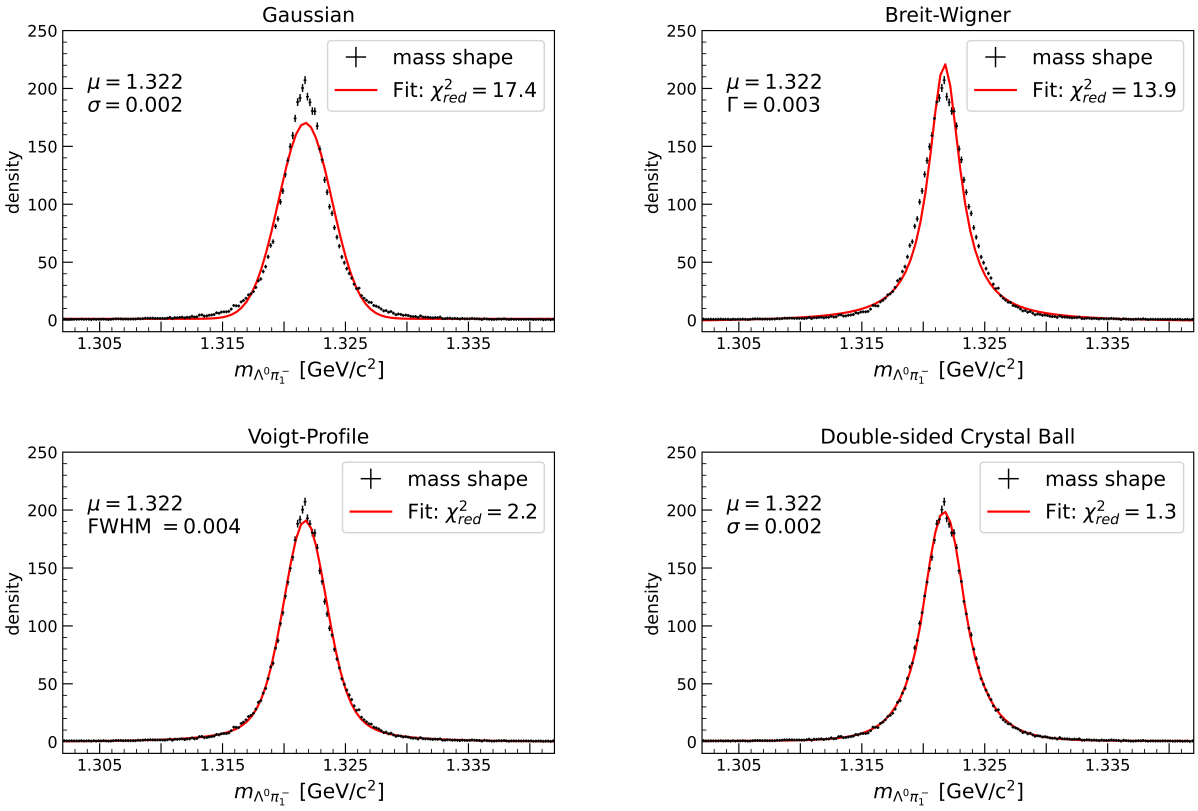


Figure 5.19.: different fit function approaches for invariant mass distribution of a signal-only sample with no selection

While a pure Gaussian distribution shows to be too broad, a Breit-Wigner distribution turns out too narrow. The shape shows good accordance with the Voigt-Profile, being the Convolution of a Breit-Wigner and a Gaussian. Theoretically, the resonance curve per se is expected to follow a Breit-Wigner distribution, the detector resolution causes an widening in Gaussian form. The Convolution of both should therefore be a fitting description.

Actually, the Double-sided Crystal Ball (DSCB) function shows the best description of the shape, achieving the lowest χ^2_{red} score of 1.3. The DSCB function describes a continuous distribution following a Gaussian for the peak region, while the tails on both sides follow asymmetric power laws.

$$f(x) = N \begin{cases} \frac{n_l}{\alpha_l}\, exp\left(-\frac{\alpha_l^2}{2}\right) \times \left(\frac{n_l}{\alpha_l} - \alpha_l - \frac{x-\mu}{\sigma}\right)^{-n_l} & \text{if } \frac{x-\mu}{\sigma} \leq -\alpha_l \\ exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) & \text{if } -\alpha_l < \frac{x-\mu}{\sigma} < \alpha_h \\ \frac{n_h}{\alpha_h}\, exp\left(-\frac{\alpha_h^2}{2}\right) \times \left(\frac{n_h}{\alpha_h} - \alpha_h + \frac{x-\mu}{\sigma}\right)^{-n_h} & \text{if } \frac{x-\mu}{\sigma} \geq \alpha_h \end{cases} \tag{5.2}$$

In Eq. (5.2), the DSCB is gives with the Gaussian parameters mean μ and standard deviation σ, asymmetric tail parameters cross over α and power law n for both low-end and high-end tail (indices l (low) and h (high) respectively). As extra parameter, N allows for a non-unity normalization. Additionally, an offset parameter is allowed in the fit.

The optimized DSCB Function locates the mass peak at $m_{\Lambda^0\pi_1^-} = 1.322 \pm 0.002\,\mathrm{GeV/c^2}$. This is in accordance with the input PDG value [20] used in GEANT transport. For the Ξ^- resonance, a Breit-Wigner distribution with a Full-Maximum-Half-Width of only 4 μeV would be expected, corresponding to the decay width $\Gamma = \frac{\hbar}{\tau}$ for the mean Ξ^- lifetime of $\tau = 1.639 \times 10^{-10}$s [20]. The deviance in the fit is of course not understood as decay width. The shape of the distribution purely originates from the experimental setup and the reconstruction process. A finite detector resolution causes Gaussian distributed measurements. With a non-uniform resolution throughout the system, the distribution is no longer random according to the central theorem, but a convolution of various functions describing these resolution alterations. While the effect does not matter much for the core, the tails are enhanced on both sides, resulting in a shape best described by a DSCB.

Being wider than a standard distribution, the chosen mass integration corresponds to $1.312 < m_{\Lambda^0\pi_1^-} < 1.332\,\mathrm{GeV}$, which is equal to the 5σ range of the peak.

Using this description of the mass shape, a multi-step fit routine is used for the remaining data fraction after the optimized iterative selection process. First, the 5σ mass peak region is excluded to allow an unbiased background characterization. Seeing that any curvature would overestimate the complexity of the distribution, a zeroth-order polynomial is chosen to model

the background. This constant level is then added to a DSCB function before optimizing it to describe the whole mass shape. The resulting shape and optimized fit functions are visualized in Figure 5.20.

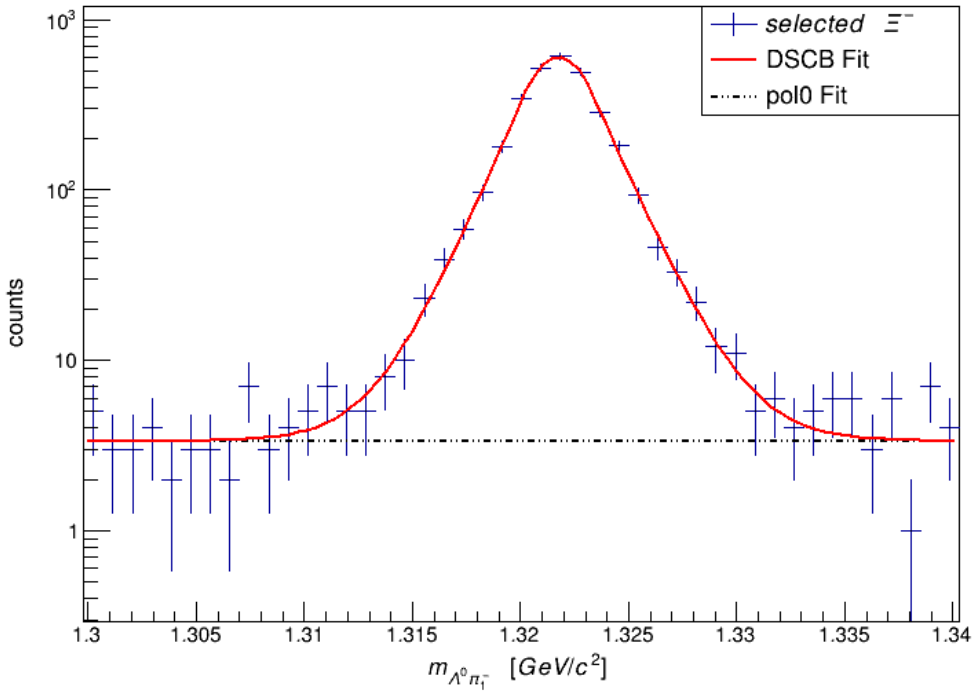


Figure 5.20.: Invariant mass shape after selection described by the optimized DSCB function and flat background level. The sample contains candidates in $p_T \in (0, 3)$ GeV/c and $y_{Lab} \in (0.5, 3)$

The exact fit parameters and the respective fit uncertainties are given in Table 5.8.

Gaussian parameters	mean μ	$1.32175 \pm 0.000\,05$ GeV/c^2
	standard deviation σ	$0.00157 \pm 0.000\,08$ GeV/c^2
	normalization N	599 ± 18
low-end tail parameters	power law n_l	48 ± 112
	cutoff α_l	1.1 ± 0.1
high-end tail parameters	power law n_h	44 ± 161
	cutoff α_h	1.0 ± 0.2
Background (pol0 parameter)		3.3 ± 0.4

Table 5.8.: optimal fitparameters for the DSCB function and flat offset on the selected mass shape

The good fit of the DSCB for both, before and after selection, proves that the general invariant mass shape does not change and is not distorted. With a χ^2_{red} of 0.70, the good description of the data is confirmed. The total count of candidates in each class is be approximated as integral over the functions. Integrating over the 5σ range of the mass peak gives a signal to background ratio of 40.5 ± 5.1, which is in accordance with the previously counted (Monte Carlo assigned) ratio of 39.0 ± 4.5. This agreement verifies that no fake peak is selected in the process. A total count of 3004 ± 169 selected candidates are assigned as signal in the peak fit.

6. Conclusion and outlook

In this thesis, a selection strategy for the Ξ^- baryon in Au+Au collisions at a center of mass energy of $\sqrt{s_{NN}} = 4.93\,\text{GeV}$ has been presented. A MC sample produced to simulate the conditions of the future CBM detector was used, where the Ξ^- candidates were reconstructed via the decay channel $\Xi^- \rightarrow \pi^-(\Lambda^0 \rightarrow p\pi^-)$ in the same way as it will be done for real data. To extract signal in the background dominated sample, a selection strategy including manual data skimming and machine learning techniques was optimized.

A set of selection criteria was chosen to separate signal and background candidates in visibly differing one-dimensional feature distributions. Subsequently, ML was exploited to single out signal candidates in a multidimensional way. An iterative approach was implemented, consisting in the optimization and training of two consecutive BDT models. Overall, the presented selection strategy achieves a signal efficiency with 67±2 %, allowing to extract 3004±169 signal candidates from 250×10^3 collision events with a signal to background ratio of 40.5 ± 5.1 in the 5σ range of the mass peak. This strongly indicates that Ξ^- yield measurements will prove feasible with the CBM apparatus.

Furthermore, the quality of an iterative selection strategy has been shown. Using an ML based preselection (similarly to the one-dimensional data skimming before) allows to remove background candidates based on low-level feature correlations, before using a second ML model to find only finer differences hidden deeper in the data characteristics. Compared to a single ML model selection, this allows to decrease the signal loss while simultaneously increasing the background reduction.

In the future, further optimization of the selection strategy can be expected to improve the Ξ^- yield. This includes multi-differential selection in different (p_T, y_{Lab}) bins, but also another approach of dual ML selection. Having discovered in this analysis that the background candidates show strong similarities depending on their origin, this can be exploited further. A selection of secondary Λ^0 candidates preceding the Ξ^- selection can replace the presented iterative approach. In this analysis, the first model mainly ensures the rejection of primary particles, however requiring secondary Λ^0s for the Ξ^- reconstruction itself would render this process redundant. A Ξ^- selection based on a previous non-prompt Λ^0 selection might therefore be an equivalent - if not more effective - strategy.

List of Acronyms

BDT Boosted Decision Tree (or its output score respectively)

CBM Compressed Baryonic Matter

DSCB Double-Sided Crystal Ball Function

EoS Equation of State

FAIR Facility for Antiproton and Ion Research

GSI Gesellschaft für Schwerionenforschung

MC Monte Carlo

ML Machine Learning

MVD Micro Vertex Detector

PID Particle Identification

PV Primary Vertex, collision point

QCD Quantum Chomodynamics

QED Quantum Electrodynamics

QGP Quark-Gluon Plasma

ROC Reciever Operating Characteristic

ROC-AUC ROC - area under the curve

SIS Schwerionensynchrotron

STS Silicon Tracking System

SV Secondary Vertex, decay point

TOF Time of Flight detector

List of Features

χ^2 distance $\Delta\vec{r}$ normalized to Covariance matrix C, see equation 3.2

χ^2_{topo} χ^2 between mother track and PV + χ^2 between daughter tracks

χ^2_{prim} χ^2 between daughter track and PV

χ^2_{geo} χ^2 between daughter tracks

DCA distance of closest approach between daughter tracks

L decay length of mother particle

$L/\Delta L$ decay length normalized to its uncertainty

$cos(\alpha)$ cosine of angle between mother and daughter tack

$cos(\alpha_{topo})$ cosine of angle between mother track and connection PV-SV

m^2 squared mass of measured particles, used as PID, TOF information

Bibliography

[1] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013. DOI: 10.1017/CBO9781139525367.

[2] A. Chaudhuri. „A Short Course on Relativistic Heavy Ion Collisions“. In: (July 2012). DOI: 10.1088/978-0-750-31060-4.

[3] Jana N. Guenther. *Overview of the QCD phase diagram – Recent progress from the lattice*. 2021. arXiv: 2010.15503 [hep-lat].

[4] A. Bazavov et al. „Chiral crossover in QCD at zero and non-zero chemical potentials“. In: *Physics Letters B* 795 (2019), pp. 15–21. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2019.05.013. URL: https://www.sciencedirect.com/science/article/pii/S0370269319303223.

[5] Wei-jie Fu, Jan M. Pawlowski, and Fabian Rennecke. „QCD phase structure at finite temperature and density“. In: *Phys. Rev. D* 101.5 (2020), p. 054032. DOI: 10.1103/PhysRevD.101.054032. arXiv: 1909.02991 [hep-ph].

[6] Bengt Friman et al., eds. *The CBM Physics Book: Compressed Baryonic Matter in Laboratory Experiments*. 1st ed. Vol. 814. Lecture Notes in Physics. Springer Berlin/Heidelberg, 2011. URL: https://doi.org/10.1007/978-3-642-13293-3.

[7] T. Ablyazimov et al. „Challenges in QCD matter physics –The scientific programme of the Compressed Baryonic Matter experiment at FAIR“. In: *Eur. Phys. J. A* 53.3 (2017), p. 60. DOI: 10.1140/epja/i2017-12248-y. arXiv: 1607.01487 [nucl-ex].

[8] Alexander Malakhov and Alexey Shabunov, eds. *Technical Design Report for the CBM Superconducting Dipole Magnet*. Darmstadt: GSI, 2013, 80 S. URL: https://repository.gsi.de/record/109025.

[9] Johann Heuser et al., eds. *[GSI Report 2013-4] Technical Design Report for the CBM Silicon Tracking System (STS)*. Darmstadt: GSI, 2013, 167 p. URL: https://repository.gsi.de/record/54798.

[10] Norbert Herrmann, ed. *Technical Design Report for the CBM Time-of-Flight System (TOF)*. Darmstadt: GSI, 2014, 182 S. URL: https://repository.gsi.de/record/109024.

[11] M. Baznat et al. „Monte-Carlo Generator of Heavy Ion Collisions DCM-SMM“. In: *Physics of Particles and Nuclei Letters* 17.3 (May 2020), pp. 303–324. DOI: 10.1134/s1547477120030024. URL: https://doi.org/10.1134/S1547477120030024.

[12] S. Agostinelli, J. Allison, and K. Amako. „GEANT4: A simulation toolkit“. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: https://www.sciencedirect.com/science/article/pii/S0168900203013688.

[13] The CBM Collaboration. *CBMRoot*. URL: https://git.cbm.gsi.de/computing/cbmroot (visited on 07/08/2023).

[14] Valentina Akishina and Ivan Kisel. „Online event reconstruction in the CBM experiment at FAIR“. In: *EPJ Web Conf.* 173 (2018), p. 01002.

[15] O. Lubynets et al. *PFSimple*. URL: https://git.cbm.gsi.de/pwg-c2f/analysis/pf_simple (visited on 08/07/2023).

[16] Sergey Gorbunov. „On-line reconstruction algorithms for the CBM and ALICE experiments“. PhD thesis. Johann Wolfgang Goethe University, Frankfurt am Main, Germany, 2013.

[17] S. Gorbunov and I. Kisel. *Reconstruction of decayed particles based on the Kalman filter*. CBM-SOFT-note-2007-003. 2007.

[18] S. Gorbunov and I. Kisel. *Primary vertex fit based on the Kalman filter*. CBM-SOFT-note-2006-001. 2006.

[19] S. Gorbunov and I. Kisel. *Secondary vertex fit based on the Kalman filter*. CBM-SOFT-note-2006-002. 2006.

[20] R. L. Workman et al. „Review of Particle Physics“. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097. URL: https://pdglive.lbl.gov/Particle.action?init=0&node=S022&home=BXXX030. last accessed on 10.07.23.

[21] Maksym Zyzak. „Online selection of short-lived particles on many-core computer architectures in the CBM experiment at FAIR“. PhD thesis. Johann Wolfgang Goethe University, Frankfurt am Main, Germany, 2015.

[22] Luca Barioglio et al. *Minimal heavy-ion physics environment for Machine Learning (hipe4ml)*. en. Apr. 2022. DOI: 10.5281/ZENODO.7014886. URL: https://zenodo.org/record/7014886.

[23] Tianqi Chen and Carlos Guestrin. „XGBoost: A Scalable Tree Boosting System“. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, Aug. 2016. DOI: 10.1145/2939672.2939785. URL: https://doi.org/10.1145/2939672.2939785.

[24] Takuya Akiba et al. „Optuna: A Next-generation Hyperparameter Optimization Framework“. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, July 2019. DOI: 10.1145/3292500.3330701. URL: https://doi.org/10.1145/3292500.3330701.

[25] Gareth James et al. „Resampling Methods“. In: *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer US, 2021, pp. 197–223. ISBN: 978-1-0716-1418-1. DOI: 10.1007/978-1-0716-1418-1_5. URL: https://doi.org/10.1007/978-1-0716-1418-1_5.

[26] M. Zyzak et al. „The KF Particle Finder package for short-lived particles reconstruction for CBM“. In: *Scientific Report 2012*. Vol. 2013-1. GSI Report. Darmstadt: GSI Helmholtzzentrum für Schwerionenforschung, 2013, 79 p. URL: https://repository.gsi.de/record/51983.

A. Appendix

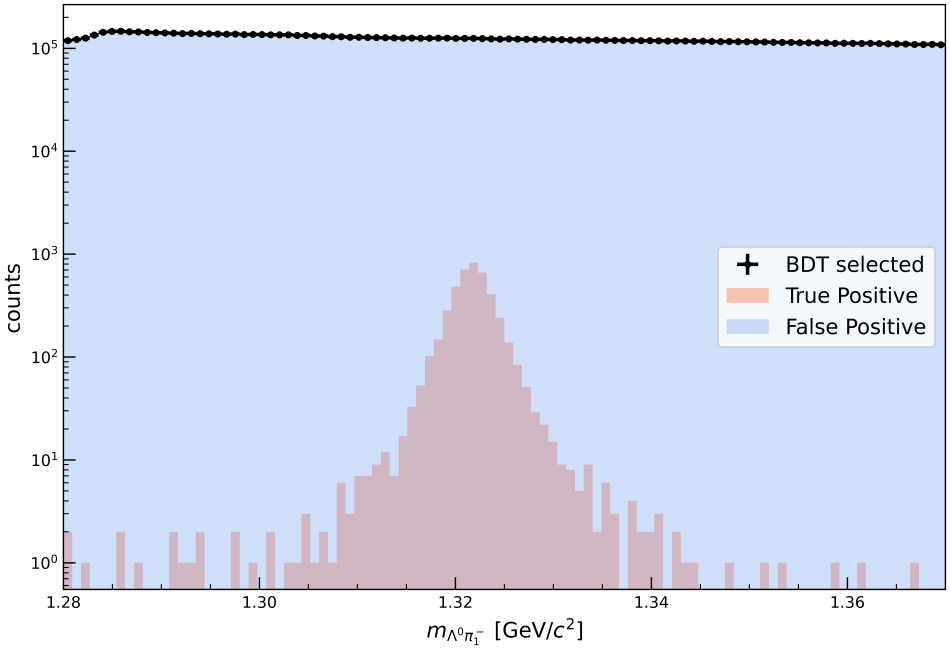


Figure A.1.: Invariant mass distribution after single model selection with BDT > 0.6. Even with applied data-skimming and machine learning selection, the signal peak is not visible in the distribution (black pointers)

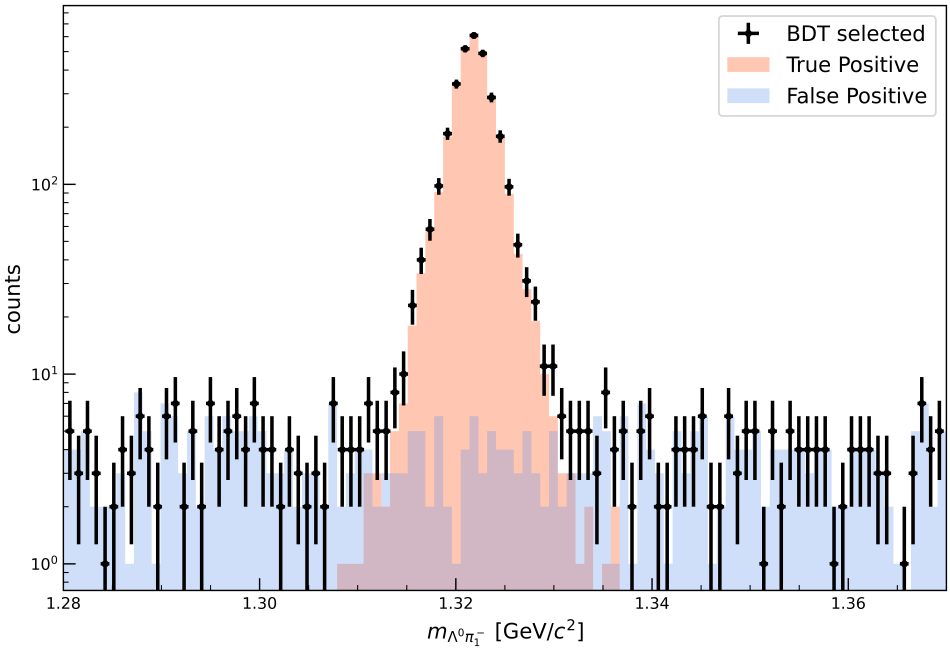


Figure A.2.: Invariant mass distribution after iterative model selection

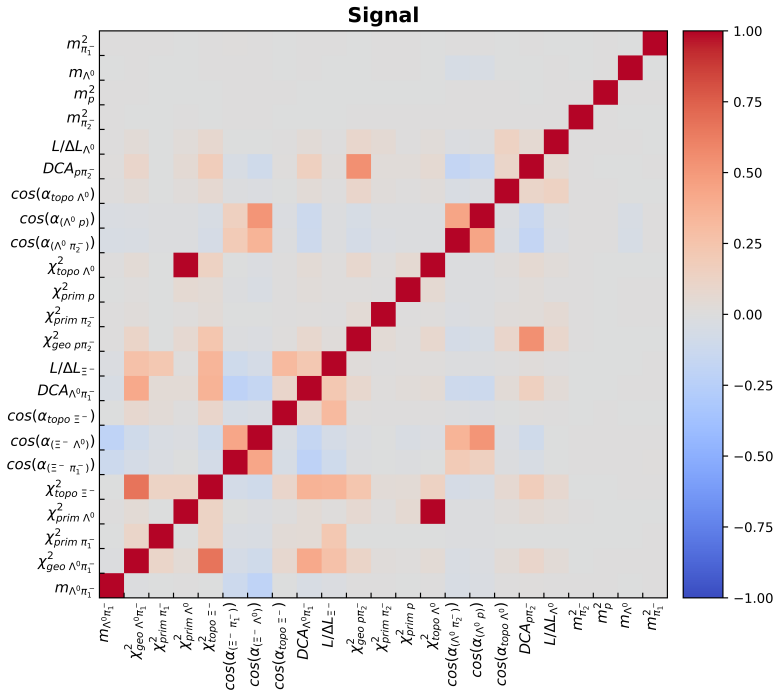


Figure A.3.: Covariance matrix of invariant Ξ^- mass and features used for training (signal candidates)

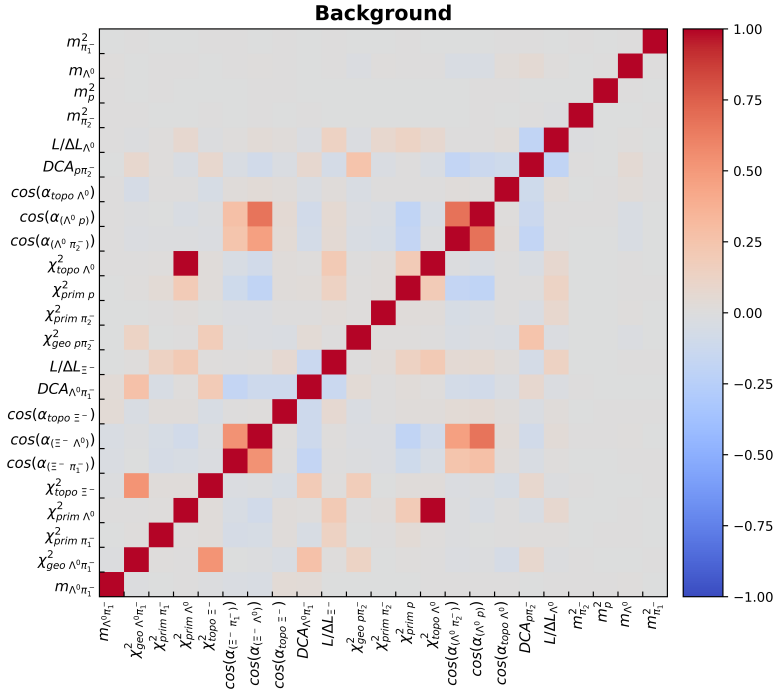


Figure A.4.: Covariance matrix of invariant Ξ^- mass and features used for training (background candidates)

List of exact variables used in selection

$\chi^2_{geo\ \Lambda^0\pi^-_1}$

χ^2 between Λ^0 and π^-_1 (Ξ^- daughter particles) tracks

$\chi^2_{prim\ \pi^-_1}$

χ^2 between π^-_1 (Ξ^- daughter) track and PV

$\chi^2_{prim\ \Lambda^0}$

χ^2 between Λ^0 (Ξ^- daughter) track and PV

$\chi^2_{topo\ \Xi^-}$

χ^2 between Ξ^- track and PV + χ^2 between Ξ^- daughter particle tracks

$\chi^2_{geo\ p\pi^-_2}$

χ^2 between p and π^-_2 (Λ^0 daughter particles) tracks

$\chi^2_{prim\ \pi^-_2}$

χ^2 between π^-_2 (Λ^0 daughter) track and PV

$\chi^2_{prim\ p}$

χ^2 between p (Λ^0 daughter) track and PV

$\chi^2_{topo\ \Lambda^0}$

χ^2 between Λ^0 track and PV + χ^2 between Λ^0 daughter particle tracks

$\cos(\alpha_{(\Xi^-\pi^-_1)})$

cosine of angle between Ξ^- and π^-_1 (Ξ^- daughter) tracks

$\cos(\alpha_{(\Xi^-\Lambda^0)})$

cosine of angle between Ξ^- and Λ^0 (Ξ^- daughter) tracks

$\cos(\alpha_{topo\ \Xi^-})$

cosine of angle between Ξ^- track and connection PV-SV

$\cos(\alpha_{(\Lambda^0\pi^-_2)})$

cosine of angle between Λ^0 and π^-_2 (Λ^0 daughter) tracks

$\cos(\alpha_{(\Lambda^0 p)})$

cosine of angle between Λ^0 and p (Λ^0 daughter) tracks

$\cos(\alpha_{topo\ \Lambda^0})$

cosine of angle between Λ^0 track and connection PV-SV

$DCA_{\Lambda^0\pi_1^-}$

distance of closest approach between Ξ^- daughters

$DCA_{p\pi_2^-}$

distance of closest approach between Λ^0 daughters

L_{Ξ^-}

decay length of Ξ^- (distance between PV and Ξ^- decay vertex)

L_{Λ^0}

decay length of Λ^0 (distance between PV and Λ^0 decay vertex)

$L/\Delta L_{\Xi^-}$

decay length of Ξ^- normalized to uncertainty

$L/\Delta L_{\Lambda^0}$

decay length of Λ^0 normalized to uncertainty

$m_{\Lambda^0\pi_1^-}$

invariant Ξ^- mass reconstructed from Λ^0 and π_1^-

m_{Λ^0}

reconstructed invariant Λ^0 mass previous to non-linear mass constraint (see Chapter 3.3)

$m^2_{\pi_1^-}$

squared mass of π_1^- (Ξ^- daughter), TOF information

$m^2_{\pi_2^-}$

squared mass of π_2^- (Λ^0 daughter), TOF information

m^2_p

squared mass of p (Λ^0 daughter), TOF information

p_T

transverse Ξ^- momentum

y_{Lab}

Ξ^- rapidity

Declaration of Authorship

Ich versichere, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den 7. September 2023