Universität Zürich[UZH]

June 21, 2017

Studies of $B^+ \to K^+\pi^+\pi^-e^+e^-$ decays at LHCb

Bachelor thesis of
Jonas Eschle

Supervised by
Prof. Nicola Serra
Dr. Rafael Silva Coutinho

Abstract

Studies of the rare decay $B^+ \to K^+\pi^+\pi^-e^+e^-$ are performed using proton-proton collision data, corresponding to an integrated luminosity of about 1 fb^{-1}, collected by the LHCb experiment at the centre-of-mass energy of 7 TeV. A novel reweighting procedure based on the boosting technique and decision trees is applied to reduce the simulation to data differences using the control channels $B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-$ and $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$. A MVA algorithm is designed to reduce the combinatorial background. Finally, a preliminary blind fit of the $K^+\pi^+\pi^-e^+e^-$ invariant mass is performed. To proceed with the analysis, further studies on the background contamination are proposed.

Contents

1 Introduction

1.1 Standard Model

The Standard Model (SM) of particle physics is a collection of theories describing the most fundamental laws of nature except of gravity. It is embedded in the theoretical framework of the quantum field theory and had great success with the prediction and explanation of particles and their behaviour.

According to the SM, there are three fundamental forces apart from gravity, which is several orders of magnitude weaker. The forces result from the exchange of mediators. The weak force interacts with nearly all particles and despite its name, which comes from the low frequency it appears in our energy scales, it is responsible for particle interactions with the most non-conserved quantities. It is mediated via the heavy Z and $W^{\pm}$ bosons. A prominent example is the β decay which occurs via the weak force. The strong force keeps the nuclei together and is responsible for nuclei-nuclei interactions. Its mediators are gluons which interact via three different types of colour charges. The electromagnetic force is responsible for particle-particle interactions as well as all macroscopic electrodynamic effects. It acts via the exchange of photons between electrically charged particles.

All of the matter is made up from the following twelve particles. For every particle, there exists a corresponding anti-particle with the same mass but opposite quantum numbers. The subatomic particles, which make up the nuclei of atoms, are quarks. There are three generations of quarks. In each generation, two kind of quarks exist, one with the electric charge of $+2/3$, the other one with $-1/3$. They can interact via the electromagnetic, weak and the strong force and exist in nature as groups of either two (quark anti-quark pair) or three (three quarks with different colour charges) quarks[1]. Analogous to the quarks, there are three families of leptons. Two different kind per family exist, a lepton and a corresponding lepton-neutrino, whereas the former carries electric charge and the latter does not. As leptons do not carry colour charge, the neutrino is left to interact via the weak force only. The Higgs boson is an excitation in the Higgs-field which, simplified, is responsible for the fact that the other particles actually have mass.

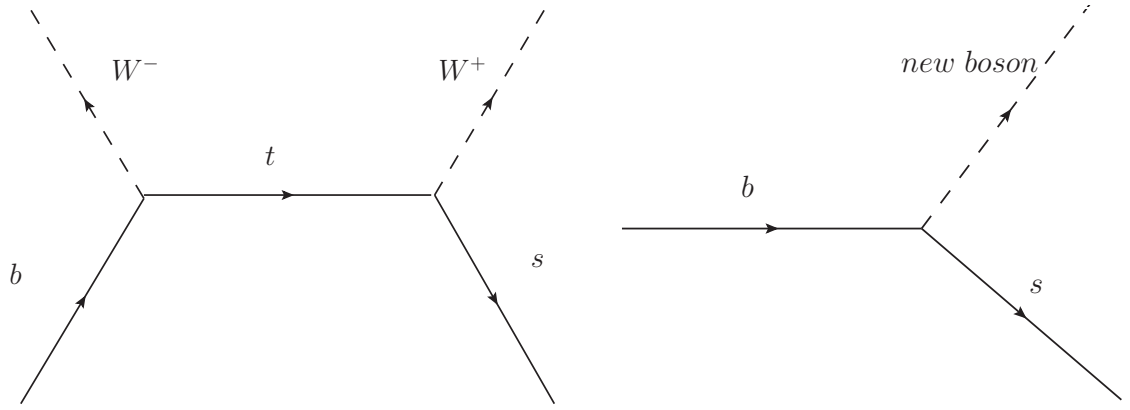
Heavier quarks decay into lighter ones under the restriction to decay into oppositely charged quarks. Their transition is described by the GIM-mechanism, which suppresses such decays naturally [1]. This allows flavour changing neutral currents (FCNC), decays where a quark changes its family but not its charge, to only occur in higher order diagrams.

1.2 New Physics

Despite the SM being a remarkable theory in its predictive power, it fails to accommodate elements such as gravity, dark matter, dark energy and neutrino oscillation. Hence, there is a general agreement on the fact that the SM is not the final description of nature and the searches for rare or SM forbidden processes are of great interest. Those are usually summarized as the search for New Physics (NP).

Rare B decays through the $b \to s$ transition via FCNC can only occur at loop-level, as seen before, via electroweak loop (penguin) and box diagrams. As the decay is strongly suppressed, it provides an effective way to check the predictions of the SM as they are sensitive to small corrections contributed from NP. Such could enter the quantum loops as shown in Fig. 2a and change the branching fraction significantly.

[1] There exist exotic states as well which are not mentioned here.

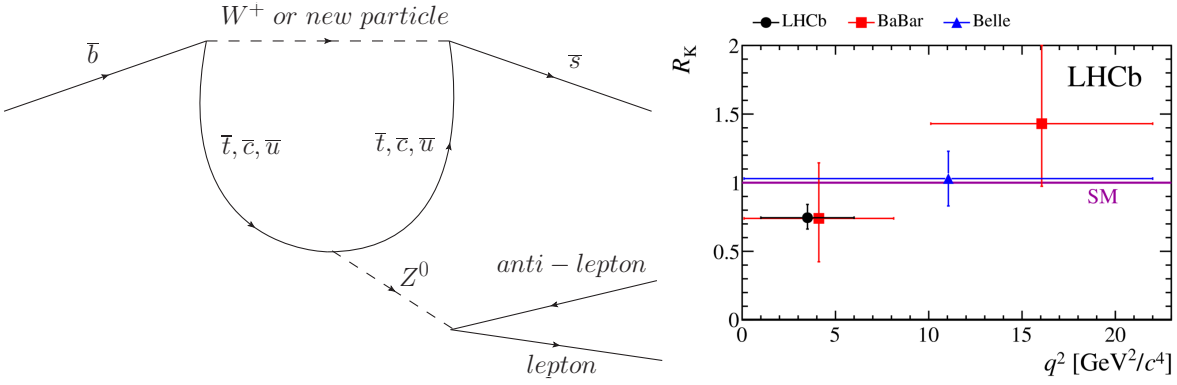


(a) standard model (b) new physics

Figure 1: FCNC as in 1a are naturally suppressed by the GIM-mechanism and occur in higher order diagrams only. Currently forbidden in the SM are transitions as shown in the tree level diagram 1b.

According to the SM, leptons carry the same weak charge and particles couple equally to different flavours of leptons – hereafter referred to as lepton flavour universality (LFU). Measurements of the branching ratio of $B^+ \to K^+\ell^+\ell^-$ with ℓ either equal e or μ hint a possible deviation from the SM with a significance of 2.6σ [2], recent results from the measurement of the branching ratio of $B^0 \to K^{*0}\ell\ell$ seem to confirm this deviation [3].

Therefore, additional studies of semileptonic b-hadron decays are of great interest. An interesting final state to consider is the decay channel $B^+ \to K^+\pi^+\pi^-\ell^+\ell^-$, where only the muonic channel $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ has been observed so far [4]. This measurement can be used for branching ratio tests later on. In this thesis the yet unobserved $B^+ \to K^+\pi^+\pi^-e^+e^-$ is examined. The current branching fraction of $B^+ \to K^+\pi^+\pi^-e^+e^-$ is predicted to be $\mathcal{B} = (2.7^{+1.5+0.0}_{-1.2-0.3}) \times 10^{-6}$ whereas the first error originates from the uncertainty of the form-factor and the second one from the error of the kaon mixing angle



(a) (b)

Figure 2: A penguin diagram of $b \to s$ transition, as it also occurs in the decay under study, is shown in 2a with possible NP contribution through the *new particle* as it occurs in the decay under study. In 2b, the currently measured $\mathcal{B}$ of $B^+ \to K^+\ell^+\ell^-$ with ℓ either e or μ at the LHCb is consistent with the SM prediction at the 2.6 σ level.

θ_{K_1} (see below) [5].

The decay under study can be accessed via different resonances. The main contribution comes from the $B^+ \to K_1(1270)(\to K^+\pi^+\pi^-)e^+e^-$ decay. The K_1 is a mixture of the two orbital angular momentum states $K_{1A}(1^3P_1)$ and $K_{1B}(1^1P_1)$ with two physical states, $K_1(1270)$ and $K_1(1400)$. The number in brackets refer to their mass in MeV[2]. Their mixing is given by

$$\begin{pmatrix} |K_1(1270)\rangle \\ |K_1(1400)\rangle \end{pmatrix} = \begin{pmatrix} \cos\theta_{K_1} & \sin\theta_{K_1} \\ -\sin\theta_{K_1} & \cos\theta_{K_1} \end{pmatrix} \begin{pmatrix} |K_{1A}\rangle \\ |K_{1B}\rangle \end{pmatrix} \tag{1}$$

where θ_{K_1} is the mixing angle. Although there is no common agreement on an angle, the mixing seems to be maximal. Both K_1 decay into the same final state of $K^+\pi^+\pi^-$.

The resonant decays $B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-$ as well as $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$ have already been observed and the latter is used as normalisation channel.

[2]Natural units with $\hbar = c = 1$ are used throughout.

2 LHCb Experiment

2.1 LHC

The Large Hadron Collider (LHC) is a proton-proton synchrotron situated nearly 200 m below the surface in a tunnel. The ring of superconducting magnets has a total length of 27 km containing two beam pipes filled with protons that are brought to collision at several points. Those collisions occurred at a total centre-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$ in 2011 and 8 TeV in 2012. After an upgrade, the energy has been increased to the centre-of-mass energy of 13 TeV in 2015 and 2016. The proton beams interact simultaneously in four detector points in the LHC ring which experiments built around, ATLAS, CMS, LHCb and ALICE. Two of them, CMS and ATLAS, are more general-purpose experiments with a toroidal structure covering the whole space around the interaction point and operating at the full collision rate. With another goal in mind, there is also ALICE, an experiment designed to study gluon-plasma and high-density events. For a fraction of the running time, the LHC is filled with lead-ions in order to create lead-proton or lead-lead interactions.

2.2 Detector

The Large Hadron Collider beauty (LHCb) is one out of four experiments situated at the LHC at CERN [6]. The LHCb is designed to perform high-precision measurements of particles containing b and c quarks to study rare decays and CP violation. In contrast to the other experiments located at the LHC, the LHCb is a single-arm forward spectrometer. This allows for measurements in the region of the pseudorapidity range $2 < \eta < 5$, the predominant flight direction of $b\bar{b}$-production.

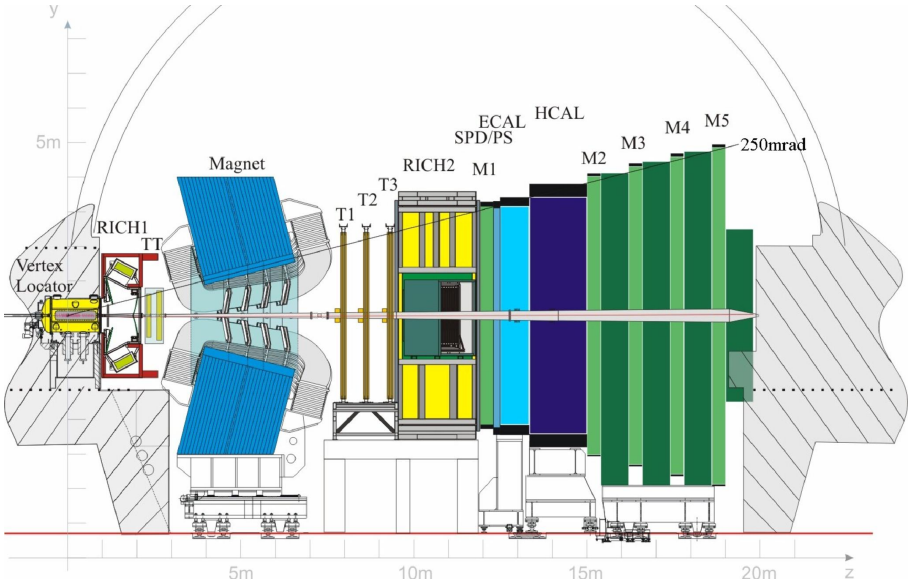


Figure 3: A schematic view of the non-bending plane of the LHCb detector. Particles are produced in the collision point on the left side inside the vertex locator and are bent by the magnet afterwards.

2.2.1 Vertex locator

An important aspect of the b and c physics is the tracking of the particles close to the interaction point in order to precisely determine the primary and secondary vertices of heavy mesons. At the LHCb this is achieved with the vertex locator (VELO), a series of modules with sensors made up of lightweight, radiation-hard silicon-strips. Each sensor is able to either measure the azimuthal coordinate or the radial distance to the beam axis, a single module contains both complementary sensors. The tracker is located about 8 mm from the aligned beam and is placed inside a beam-pipe independent vacuum system.

2.2.2 Tracking system

In addition to the VELO, several other tracking stations measure the tracks and bending of the particles. In front of the 4 Tm dipole magnets, the Tracker Turicensis is installed. It consists of four layers of silicon-strip detectors and allows for the detection of low-momenta particles which will be bent away in the magnetic field.

After the dipole magnet, the three tracking stations T1, T2 and T3 are placed. Each of them consists of an inner tracker situated close to the beam pipe and an outer tracker, covering the largest area of the tracker plane. The inner tracker is a silicon-strip detector covering the area with a high density of tracks. Another detector technique is used in the outer tracker as it covers a greater area without the need for the same precision as required in the inner tracker. In this case four layers of straw tubes filled with gas are used as drift chambers.

2.2.3 RICH

In b physics it is important to have a good discrimination between charged particles, *e.g.* K and π. In order to achieve a good particle identification, there is a ring imaging Cherenkov detector (RICH) on each side of the magnet, which measure the Cherenkov emission angle θ_c. The Cherenkov radiation is detected by pixel hybrid photon detectors. As the angle of the radiation relative to the particles flight direction depends on the velocity of the passing particle only, using additionally the information about the momentum from the tracker allows to determine the mass of the particle and therefore its identity. The Cherenkov angle also depends on the materials refractive index the charged particle is passing through. In order to cover a large momentum range with a good angle resolution, the RICH detectors are filled with materials of different refractive indices.

2.2.4 Calorimeter

A calorimeter measures the total as well as the differential energy loss by completely absorbing it through interactions with the material. For the LHCb, a classical architecture of an electromagnetic calorimeter (ECAL) in front of a hadronic calorimeter (HCAL) was chosen. Both are optimized for particle identification, mostly for e/π and π^0/γ discrimination, as well as for a fast readout. The information will be used, among others, in the first trigger stage (see Sec. 2.3).

In front of the ECAL, a scintillator pad detector is placed to detect the pass-through of charged particles followed by a pre-shower detector. The ECAL itself is built of a sampling scintillator/lead structure (shashlik technology) and has a total depth of $25X_0$. As the hit density rapidly drops with increasing distance from the beam pipe, the ECAL is split into three different sections with appropriate cell sizes.

The HCAL of the LHCb is a sampling calorimeter with a special structure. It consists of lead/scintillator tiles directed *parallel* to the beam-pipe. Each thin row consisting of several tiles has a neighbour-row with inverted lead/scintillator tiles. The scintillation light is detected by photomultiplier tubes and collected by fibres. The total length equals to 5.6 hadronic interaction lengths.

2.2.5 Muon system

The muon system is responsible for the identification of muons and provides a standalone, fast signal to the trigger in case of muons with high transverse momentum (p_{T}) passing it. The whole system is composed of the five stations M1–M5. M1 is placed in front of the calorimeters to improve the p_{T} resolution whereas the others are located downstream. The stations are separated by iron absorbers to prevent any non-muons from passing through the detectors. All systems provide spatial resolved hit information with decreasing segmentation scale for increasing distance to the beam pipe. M4 and M5 are mostly used for penetration testing and offer only sparsely location informations.

2.3 Trigger

At the nominal LHC conditions, the bunch crossing frequency can reach up to 40 MHz which leaves 25 ns in between two crossings. This high frequency has to be reduced down to 1 kHz in order to be able to store the data for offline analysis. Two trigger-systems, a low-level trigger (L0) and a high-level trigger (HLT) consisting of two stages, HLT1 and HLT2 select which events to keep.

The L0 stage is a hardware implemented trigger and consists of a custom electronics set-up built with FPGA. It takes information from three different sources into account. The first is a pile-up system inside the VELO, estimating the number of events that occurred during the collision. Information from the calorimeter is used to estimates the transverse energy (E_{T}) of certain particles and decides to keep the event in case of high E_{T}. The muon system feeds the trigger with information about the p_{T} of muons in order to trigger on a certain threshold.

The next stage is the HLT1. It reconstructs some parts of the tracks to confirm the L0 decision as well as to further reduce the event rate. This is now low enough to allow the HLT2 to reconstruct b events and make more refined decisions. The events which pass HLT2 with a frequency of around 1 kHz are then stored for offline analysis.

A general distinction is made on whether an event passed the trigger because of the events signature itself (trigger on signal, `TOS`) or because of some other particles signature (trigger independent of signal, `TIS`).

2.4 Software

Once the events are stored, offline tools are used to reconstruct and fit tracks and apply sets of exclusive selections prior to the data manipulation.

2.4.1 Track reconstruction and fit

For the event reconstruction, information from the tracking system (including the VELO) is used. First of all, a clustering algorithm determines *track seeds* by searching for candidates in a low magnetic field region of the spectrometer. A Kalman filter algorithm is then

fitted to the data using the track seeds as initialisation. An advantage of reconstructing and fitting with this algorithm is that the result is equivalent to a least square fit of the tracks to the hits. For the particle propagation with the Kalman filter, the inhomogeneous magnetic field as well as multiple scattering occurring from detector material is taken into account.

3 Dataset

The data used in the analysis was collected at the LHCb experiment in the year 2011, which corresponds to an integrated luminosity of $1\,\text{fb}^{-1}$ recorded at a centre-of-mass energy $\sqrt{s} = 7\,\text{TeV}$.

Three different decay channels are considered in this analysis.

- $B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-$: a control channel that is used to correct for the largest discrepancies between data and simulation.
- $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$: the normalisation channel which is used for some minor kinematic corrections.
- $B^+ \to K^+\pi^+\pi^-e^+e^-$: the signal mode under study.

3.1 Signal simulation

The Monte Carlo simulated samples (MC) are generated using Pythia 8 with a specific LHCb configuration [7]. An average number of pp interactions per bunch crossing[3] of 2 is used and a constraint is applied on the generator level to limit the $K^+\pi^+\pi^-$ invariant mass. In order to reduce possible systematic effects from the detector and surrounding, the magnetic polarity is switched during the data taking. Therefore also both magnetic polarities are generated and then merged.

3.2 Stripping

For the J/ψ resonant decays Stripping21r0p1 with the Bu2LLK line was used whereas for the rare decay Stripping20r1 was used. The cuts which are applied to the sample are listed in Table 1.

3.3 Preselection

For this study, only the central q^2 region is analysed corresponding to $1 < m^2_{e^+e^-} < 6\,\text{GeV}$. This is chosen in order to reduce the contribution from the resonant mode.

The signal candidates in this analysis are triggered by three different trigger categories and merged into one sample. The categories are exclusive and are evaluated in the following order:

- `L0 Electron` requires events to be `TOS` with respect to the `L0 Electron` trigger line.
- `L0 Hadron` requires events to be `TOS` with respect to the `L0 Hadron` trigger line.
- `L0 TIS` requires events to be `TIS` with respect to the `L0 Global` trigger line and therefore to be triggered by other particles.

[3] This also includes the number of not visible (for the detector) interactions and is therefore only used in the context of simulated events.

Table 1: Stripping requirements.

Object	Requirement
Event	$N_{\mathrm{PV}} > 1$
	$n_{SPD} < 600$
K	`hasRICH`
	$\mathrm{DLL}_{K\pi} > -5$
	χ^2_{IP} PV > 9
	$\chi^2_{track} < 3$
	$GhostProba < 0.4$
π	`hasRICH`
	χ^2_{IP} PV > 9
	$\chi^2_{track} < 3$
	$GhostProba < 0.4$
e	`hasCalo`
	$\mathrm{DLL}_{e\pi} > 0$
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}$
	χ^2_{IP} PV > 9
μ	`isMuon`
	$p_{\mathrm{T}} > 300\,\mathrm{MeV}$
	χ^2_{IP} PV > 9
$\ell\ell$	$m < 5500\,\mathrm{MeV}$
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf} < 9$
	origin vertex χ^2 separation > 16
B	$\lvert m - m^{\mathrm{PDG}}_{B^0}\rvert < 1000\,\mathrm{MeV}$
	`DIRA` > 0.9995
	χ^2_{IP} PV < 25
	$\chi^2_{\mathrm{vtx}}/\mathrm{ndf} < 9$
	PV χ^2 separation > 100
K_1	$0 < m < 6000\,\mathrm{MeV}$
	$\chi^2_{\mathrm{vtx}} < 12$
	sum hadron $p_{\mathrm{T}} > 800\,\mathrm{MeV}$
	sum hadron $\chi^2_{\mathrm{IP}} > 48$

It is additionally required that the particle triggering the `L0 Electron` (`L0 Hadron`) trigger line is a lepton (hadron). So events triggered by the `L0 Electron` and `L0 Hadron` trigger line that have been hadrons and electrons, respectively, are removed from the samples.

Finally, only events passing the HLT1 as well as the HLT2 trigger decision for TOS are kept.

With those requirements applied, still a sizeable amount of background is in our sample. To further remove that, a multivariate analysis (MVA) to discriminate between our signal and the combinatorial background is applied as described in Sect. 5.1. In order to be able to perform the MVA as well as to get an unbiased yield estimation later on, any physical background reaching into our region of interest has to be removed. Therefore strong preselection cuts are proposed as listed in Table 2.

Table 2: Preselection cuts

Particle	Variable	Cut	Explanation
B^+	χ^2_{vtx}/ndf	< 6	Reconstruction quality of the vertex per number of degrees of freedom
	AMAXDOCA	< 4	Maximum distance of closest approach with all tracks
	p_{T}	$> 3\,\text{GeV}$	Transverse momentum
	$DIRA$	> 0.9998	DIRection Angle; the cosine of the angle between the reconstructed momentum of the B^+ and its direction of flight.
	${\chi^2}_{FD}$	> 150	Significance of the flight distance with respect to the PV
J/ψ	χ^2_{IP} PV	> 1	Difference in the vertex-fit χ^2 of a given PV reconstructed with and without the current track
	χ^2_{vtx}/ndf	< 6	Reconstruction quality of the vertex per number of degrees of freedom
K_1	χ^2_{IP} PV	> 3	Difference in the vertex-fit χ^2 of a given PV reconstructed with and without the current track
K^+	ProbNNk	> 0.02	Neural network based particle identification probability to be a K
	GhostProb	< 0.3	Probability obtained from a MVA algorithm that track is a ghost
π^+ and π^-	probNNpi	> 0.02	Neural network based particle identification probability to be a π
	GhostProb	< 0.3	Probability obtained from a MVA algorithm that track is a ghost
π^+, π^- and K^+	sum of χ^2_{IP}	> 200	Difference in the vertex-fit χ^2 of a given PV reconstructed with and without the current track
e^+ and e^-	sum of χ^2_{IP}	> 200	Difference in the vertex-fit χ^2 of a given PV reconstructed with and without the current track
	GhostProb	< 0.3	Probability obtained from a MVA algorithm that track is a ghost

4 Simulation corrections

Physical intrinsic reasons such as non-converging QCD-calculations, free parameters in the SM and limited computation resources lead to differences between simulated signals and events observed in the experiment. One of the problems related to this misalignment is the possible bias in the attempt to increase the signal-to-noise ratio with a MVA background rejection. Therefore, the difference is estimated and reduced by adding event-weights to the simulated events in order to improve the agreement between the two distributions.

4.1 Reweighting techniques

The general concept of the re-weighting examined in this thesis is given as follows:

1. Compare the signal distributions of the data of the normalisation channel in specific variables with the corresponding MC.

2. Understand their differences and learn which events are likely to occur more often in the data sample.

3. Correct the generated signal events by applying weights to each event in order to compensate the differences learnt. So events occurring more often in the real sample then in the generated sample receive higher weights and vice versa.

To be able to compare the generated signal events and the data signal events, the *sPlot* technique is used, which statistically subtracts the combinatorial background from the sample [8]. Thereby weights are calculated, the sWeights, which requires to perform a fit to the B^+ mass in the data sample as shown in Fig. 4. Throughout this section, data refers to signal sWeighted data which is handled as the equivalent of generated signal events.

To learn from the differences, generalise this knowledge and correct the target distribution, several different techniques are available.

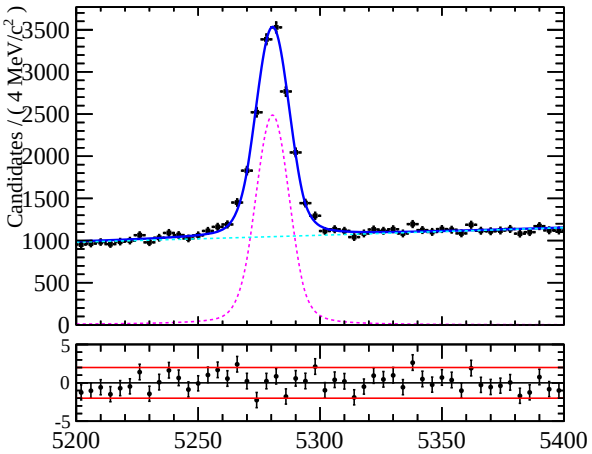


Figure 4: Fit to the B mass of $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+\pi^+\pi^-$ to obtain the sWeights.

4.1.1 Binned reweighting

A simple but widely used approach is to bin the two different samples, data and MC, in the variable which needs to be corrected. Then every bin of one sample is divided by the corresponding bin of the other sample, which results in a step-function containing the ratios.

This approach is very easy and fast, but has its limitations and disadvantages. If a bin has only a few events, the ratio fluctuates greatly and does not provide reliable weights. This is especially a problem if one considers higher dimension. As the simple approach only reweights a single variable, this is often not sufficient. Variables are one dimensional projections of a multi-dimensional distribution and therefore the reweighting does not properly account for higher order correlations. Binning however can be done in multi-dimension as well, but without a significant amount of data, the curse of dimensionality creates sparse bins with only a few events in each, leading to the fluctuations mentioned above.

4.1.2 Gradient boosted reweighting

An algorithm that tries to overcome those limitations is the gradient boosted reweighting [9]. The main characteristic of this approach is to split both samples using a decision tree (DT). The optimal split is determined by maximising a binned χ^2 fit. The ratio between the number of events of both samples in each bin is calculated and applied as corrections to the MC. The same procedure is iteratively repeated by taking the weights from the previous splits into consideration. From this procedure – discriminate samples, update data weights, repeat – comes the "boosting" in the name. Although this allows for good corrections in higher dimensional spaces due to the DT and low event regions, the algorithm is sensitive to its hyper-parameters and can often overfit. When using this approach it is a crucial part to make sure that the latter does not occur.

4.2 Performance

To find the optimal reweighting hyper-parameters and to be able to compare the different approaches, a metric for the reweighting quality should be established. Unfortunately, the comparison of two multi-dimensional distributions[4] is not so simple. In contrast to one dimension, no order of events is defined in multidimensional distributions. An order of events is often used in non-parametric tests like Kolmogorov-Smirnoff, Anderson-Darling *etc.* Although certain approaches like density kernels exist for multidimensional distributions, they are infeasible for our case due to the lack of events and/or high dimensionality. On the other hand, the question that arises is not whether a certain statistics test can distinguish our samples, but if the MVA algorithm, which will be used in the MVA afterwards, can. Therefore it seems natural to rely on its predictions to find a reliable metric. In the following, three different approaches to investigate this problem are described.

[4] A naive approach would be to compare the one dimensional projections, which correspond to the physical variables. But whereas two different projections imply different distributions, two similar projections do not imply similar distributions. An illustrative explanation is shown in the Appendix in Fig. 14.

4.2.1 Simple discrimination

A classifier[5] is trained and tested on the reweighted MC sample and on the real data using stratified k-folding and the variables that will be used later on in the MVA shown in Table 7. To test the performance of the classifier, a single-valued metric is needed. Therefore, the receiver operation characteristic (ROC) curve is drawn and the area under the curve (AUC) is calculated [10]. Notice that the same metric will be used later in the MVA. Here the idea is that the lower this score is the less the classifier is able to discriminate the two distributions. Less discrimination power means that the two distributions are more similar under the assumption that an optimised MVA algorithm is used.

Even tough this approach yields a good idea of the similarity of the two samples, it can be blind to some kind of overfitting. The problem arises with the event weights and the randomised training- respectively test-sample drawing. If an event a with an event weight w_a is drawn, what actually is draw is not one event but w_a times the event a. This is then not a randomised, uncorrelated sampling any more as drawing the event a implies also drawing the event a again, namely $w_a - 1$ times (for $w_a > 1$). So the prerequisition to make statements, namely the randomised splitting, is not given any more. The effect from this sample biasing is that the classifier makes incorrect predictions with wrongly gained strong confidence which in turn lowers the ROC AUC more then we expect it to be. This effect is further illustrated in the Appendix in Fig. 15. Although the effect decreases for large samples and is not expected to be too large for our case, it *can* even lead to ROC AUC values well below the 0.5 mark, which is usually assumed to be the lowest possible score.

4.2.2 Label the data

Another possible approach is to train a classifier on the original MC sample without corrections as well as on the real data. This trained algorithm can be used to make predictions on three distinct samples and can be used to get hints for possible overfitting. The number of events that are predicted as real data from the following events are counted and interpreted:

- MC: The lowest count is expected as most of the events will be predicted as MC.
- reweighted MC: A count as high as possible (but not higher then the real) is aimed for as higher values mean more events in the sample look like a real event to the classifier.
- real data: The highest count is expected.

Ideally, the count of the reweighted MC sample lies between the other two counts as close as possible to the real data. However, this score system should be used only as a guideline indication, since it only provides information about single events and not the distribution itself. So a real-like MC event with an extra large weight will wrongly dominate the score.

[5]Classifier refers to a MVA algorithm which predicts the class-label (in comparison to regression).

4.2.3 Count the data

In order to compare the distribution and not single events, a simple but robust approach is to train a classifier to discriminate between generated and real data, which is basically the same as described in Sec. 4.2.1 but instead of making predictions on both the generated and the real sample, only the latter is considered. Then the number of real events predicted as real is counted. The more the classifier was able to learn from the distributions, the more real events will be predict correctly[6]. So the goal is to minimize that score. Compared to the approach described in sec. 4.2.1, the bias due to weights is constant and originates only from the weights of the real sample. Therefore, changing the weights of the generated samples, as a reweighting algorithm does, does not change the bias. Although this score does not offer informations at the percent level of optimisation, it is a good indication of overfitting and complementary to the other scores.

4.3 Corrections

To find the optimal reweighting algorithm, the scores described in sec. 4.2 as well as visual comparisons of the variable distribution are used to estimate a good configuration. The values obtained for the different parameters are shown in table 3.

Two stages of corrections are applied in order to gain the best results without biasing the data. The first stage considers the $B^+ \to J/\psi(\to \mu^+\mu^-)K^+\pi^+\pi^-$ decay and is

Table 3: Hyper-parameter configurations for the gradient boosted reweighting. Two separated values means the first one was used for the first reweighting stage and accordingly for the second one. A single value means the same parameter was used in both stages.

Parameter	Value	Explanation
n_estimators	240/140	Number of boosting rounds to be performed (see comment *learning_rate*)
learning_rate	0.05	A factor by which the weights of each boosting stage are multiplied by. There is a trade-off between the learning_rate and n_estimators and the ratio determines (basically) how complex our model is.
max_depth	3	Maximum depth of the DT. Higher values create more complex models and are able to get higher order correlations but tend to over-fit.
min_samples_leaf	100	Determines the minimum number of events in a leaf in order to split. Larger values create more conservative models and can help to avoid overfitting.
loss_regularization	8/10	Adds a regularisation term to the weight inside the logarithm of the loss-function.
gb_args: subsample	0.8	The fraction of the data that is used to train each DT. Reduces overfitting.

[6]It has to be noted here that predictions are just a cut on the classifier output. To use the output for our purpose, equalized class-weights are required. Furthermore, the classifier itself has to generate probability-like predictions which XGBoost does. This is not per se the case for most algorithms.

responsible for the largest corrections. All variables are listed in table 4. Mostly $nTracks$ as well as $nSPDHits$ seem to differ largely between MC and data. The second stage of corrections uses the $B^+ \to J/\psi\,(\to e^+e^-)K^+\pi^+\pi^-$ sample and is less significant. It corrects the kinematics of the decay products. The variables are listed in table 5.

Table 4: First stage reweighting variables in $B^+ \to J/\psi\,(\to \mu^+\mu^-)K^+\pi^+\pi^-$

Variable	Explanation
nTracks	Track multiplicity of the event.
nSPDhits	Number of hits in the scintillation pad detector.
B p_{T}	Transverse momentum of the B.
B χ^2_{vtx}	Quality of the vertex reconstruction.

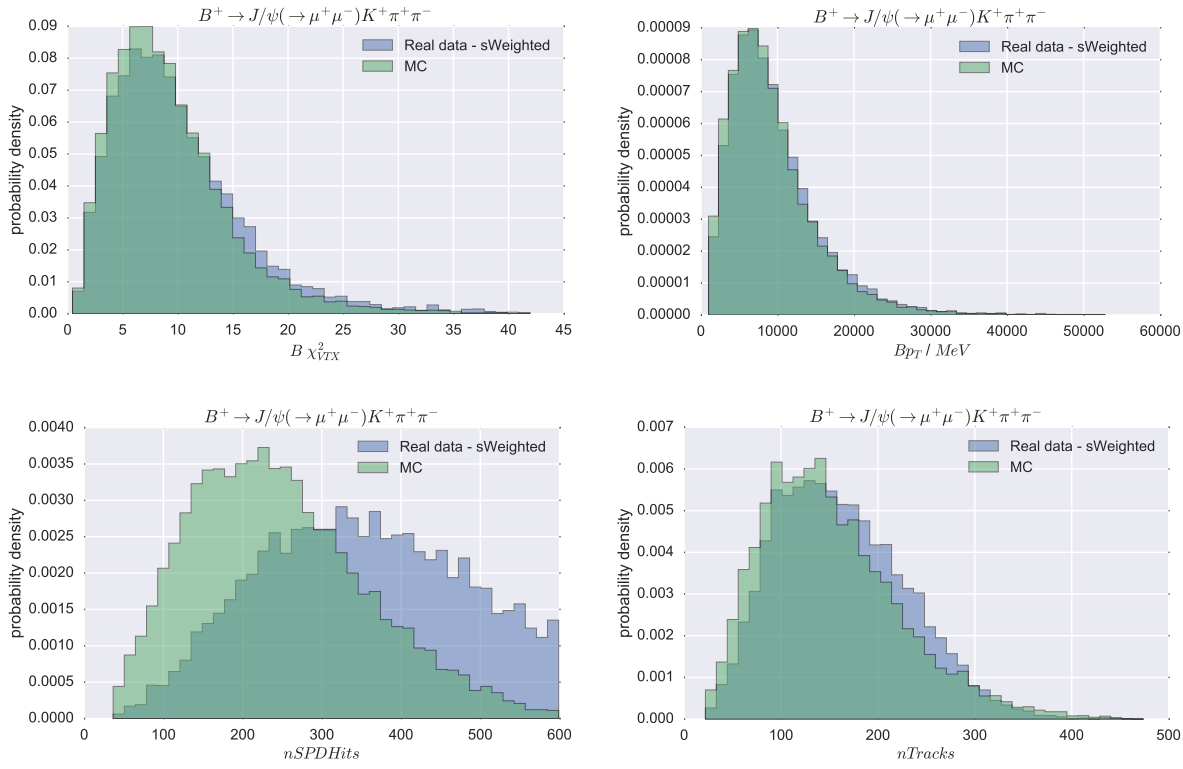


Figure 5: Variables used in the first stage of the reweighting procedure.

Table 5: Second stage reweighting variables in $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$

Variable	Explanation
$\min(h_{p_\mathrm{T}})$	Minimum p_T of the hadronic decay products
$\max(h_{p_\mathrm{T}})$	Maximum p_T of the hadronic decay products
$\min(\ell_{p_\mathrm{T}})$	Minimum p_T of the leptonic decay products
$\max(\ell_{p_\mathrm{T}})$	Maximum p_T of the leptonic decay products

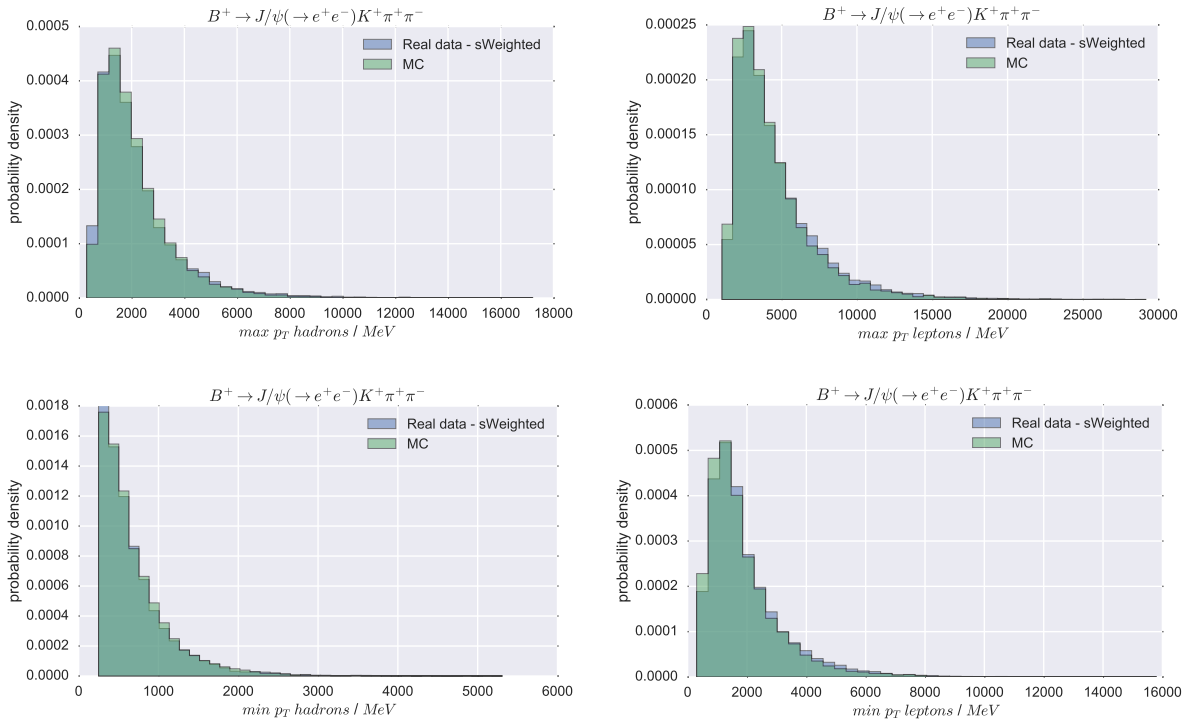


Figure 6: Variables used in the second stage of the reweighting procedure.

5 Selection

5.1 MVA

In order to further reduce the background contribution, a multivariate analysis is performed. The combinatorial background sample used in the training has been selected to be the upper sideband above 5600 MeV of the vertex constrained B^+ invariant mass, whereas the signal sample is defined as the reweighted MC described above.

5.1.1 Optimize algorithm

Several classifiers are examined. These are trained and tested on the samples using a cross-validation technique due to the limited amount of events available. A stratified k-folding strategy is applied. This works as follows:

1. The data (both signal and background) is split into k sub-samples, each containing the same fraction of a certain class[7].
2. A training set consisting of $k-1$ sub-samples and a testing set consisting of one sub-sample are created.
3. The algorithm is trained on the training set and tested on the testing set.
4. Predictions made by the algorithm on the test set are collected.
5. This is done k times, every time with a different sub-sample as testing set, so that in the end a prediction for every event is made.

For the evaluation and comparison of the performance, the ROC AUC is used with the goal to maximise it. Before the classifiers are compared against each other, a hyper-parameter optimisation is performed for each[8] classifier.

The best performing algorithm for our case is a boosted decision tree (BDT) implementation, the extreme gradient boosting (XGB) algorithm with DT as base classifiers [11–13]. Similar performance is obtained by other algorithms such as random forests and deep neural networks(DNN). The random forest averages the predictions of several DT but uses for our application critically more memory while not outperforming the XGB. The DNN is on one hand intrinsically hard and time-consuming to train and to optimise its architecture. On the other hand they are in general able to outperform most of the other classifiers but usually only with low-level features and a lot of data available in order to get the extra correlations and not just pick up noise. As there are only high-level features and a limited amount of data available, seeing no superior performance from the DNN was expected.

The final configuration used for the XGB in the selection can be seen in Table 6.

[7] A class in this context refers to the "label" or the "y"; here we have two classes, signal and background.

[8] For DNN, several well-performing architectures from other analyses were used as inspiration for the set-up, then tested and varied.

Parameter	Value	Explanation
n_estimators	500	Number of base classifiers (DT); equals to number of boosting rounds to be performed.
eta	0.1	A factor the weights of each boosting stage is multiplied by. There is a trade-off between eta and n_estimators and the ratio determines mostly how complex our model is. In other boosting algorithms, this is usually called the "learning rate".
max_depth	6	Maximum depth of the DT. Higher values create more complex models which are able to get higher order correlations.
gamma	0	Determines the minimum gain required to perform a split. Larger values create more conservative models
subsample	0.8	The fraction of the data that is used to train each DT. Reduces over-fitting.

Table 6: Hyper-parameter configuration of the XGB classifier used for the selection.

5.1.2 Feature selection

To achieve a good discrimination power, it is crucial to use the appropriate input variables. Badly simulated features are not used in order to avoid training of MC against real data instead of signal against background. In addition, any direct correlation of the features with the B^+ mass has to be avoided in order to perform an unbiased yield estimation later on.

Particle	Variable	Explanation
B^+, J/ψ, K_1	$log(p_{\mathrm{T}})$	Transverse momentum
	$log(\chi^2_{\mathrm{vtx}})$	Quality of the vertex reconstruction
	$log(\chi^2_{\mathrm{IP}})$	Difference in the vertex-fit χ^2 of a given PV reconstructed with and without the current track
	$log(\chi^2_{FD})$	Significance of the flight distance with respect to the PV
B^+	$log(DIRA)$	DIRection Angle; the cosine of the angle between the reconstructed momentum of the B^+ and its direction of flight.
	$log(AMAXDOCA)$	Maximum distance of closest approach with all tracks
	$log(\chi^2_{VTXiso}\ one\ track)$	Measure for the isolation of the reconstructed track by removing the track under consideration and repeat the reconstruction.
	$log(\chi^2_{VTXiso}\ two\ track)$	
K_1, J/ψ	$log(1-\cos(\theta))$	The angle θ is between the particles flight direction and the beam axis
π^+, π^-	$\log(p_{\mathrm{T}})$	Transverse momentum

Table 7: Variables used in the training of the XGB classifier.

5.1.3 Prediction and performance

In order to get reliable predictions from the classifier, a k-folding strategy is used on the data sample. Therefore, the data is split into k folds and the right sideband of the training sample is trained against the MC signal. The algorithm makes predictions on the full data, not just on the right sideband, of the fold not used in the training.

To apply an optimal cut based on the predictions, there are several Figure of Merits (FoM) available. Depending on the goal of the analysis, a different one may be chosen. As this study aims for the first detection, the Punzi FoM

$$FoM_{Punzi} = \frac{S}{\sqrt{B} + \sigma/2} \tag{2}$$

with $\sigma = 5$ is selected and maximized [14]. This yields the highest sensitivity for an observation with a significance of 5 σ.

5.2 Efficiencies

Not all particles produced in a collision are captured by the detector. There is a limited, geometrically determined detector acceptance range. To determine how many events are lost outside this range, the efficiency of the remaining events is calculated using generated events. For the $B^+ \to K^+\pi^+\pi^-e^+e^-$ sample used in this thesis the efficiency is $0.148 + \mathcal{O}(10^{-4})$.

An overview over the cuts applied so far as well as their respective efficiency is given in Table 8. Applying all cuts yields a total efficiency of $\varepsilon_{\text{tot}} = 0.0557\%$.

Table 8: Efficiencies of the different cuts. For every cut, the above ones are applied as well. The relative efficiency refers to the loss because of this cut with respect to the previous cut.

Number of events	Cut added	Relative efficiency
$2,065,330$	Geometrical	14.8%
$66,185$	Stripping	3.20%
$25,622$	HLT	38.7%
$14,192$	q^2 region	55.4%
$11,254$	Preselection	79.3%
7789	MVA	69.2%

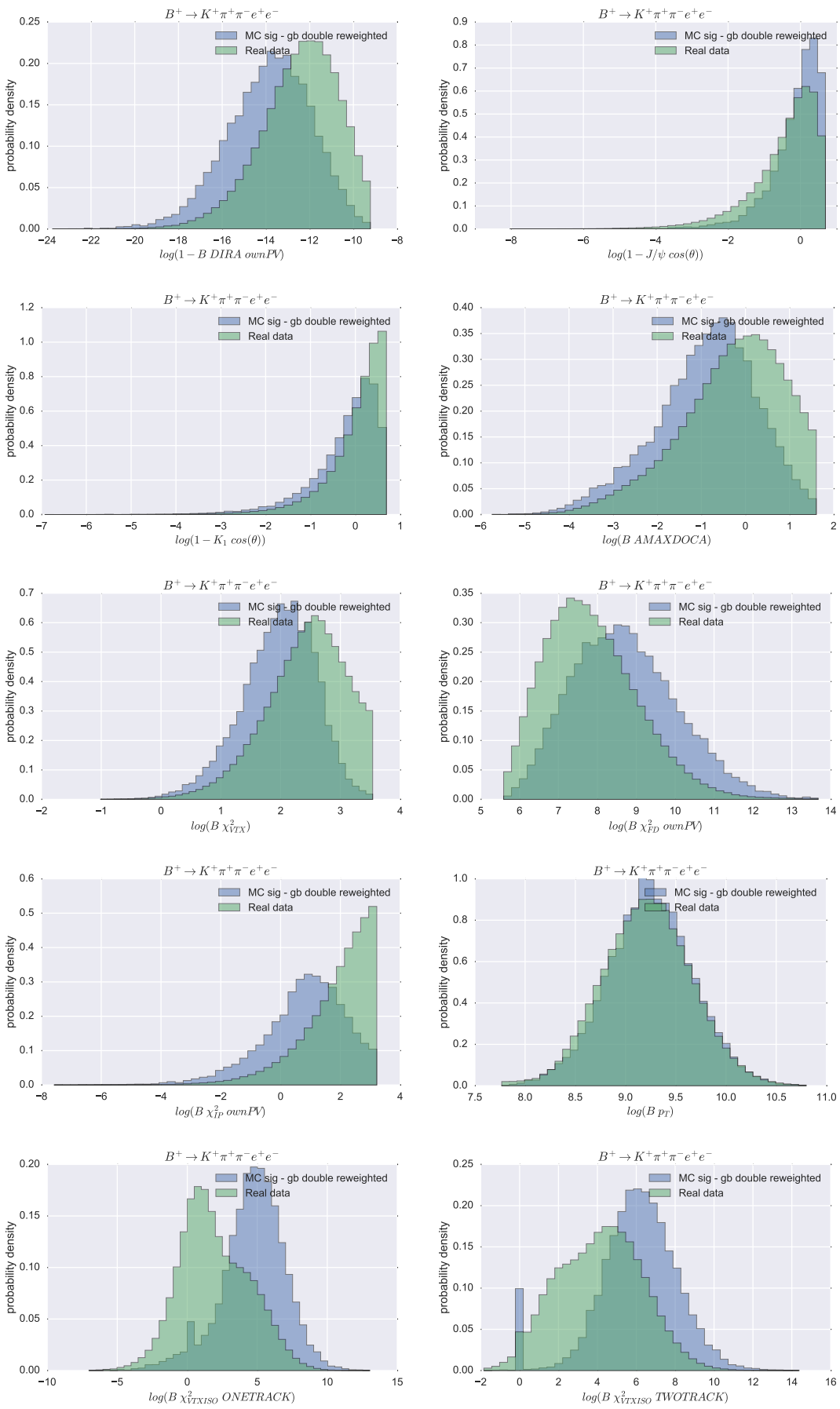


Figure 7: Features used in the training of the BDT.

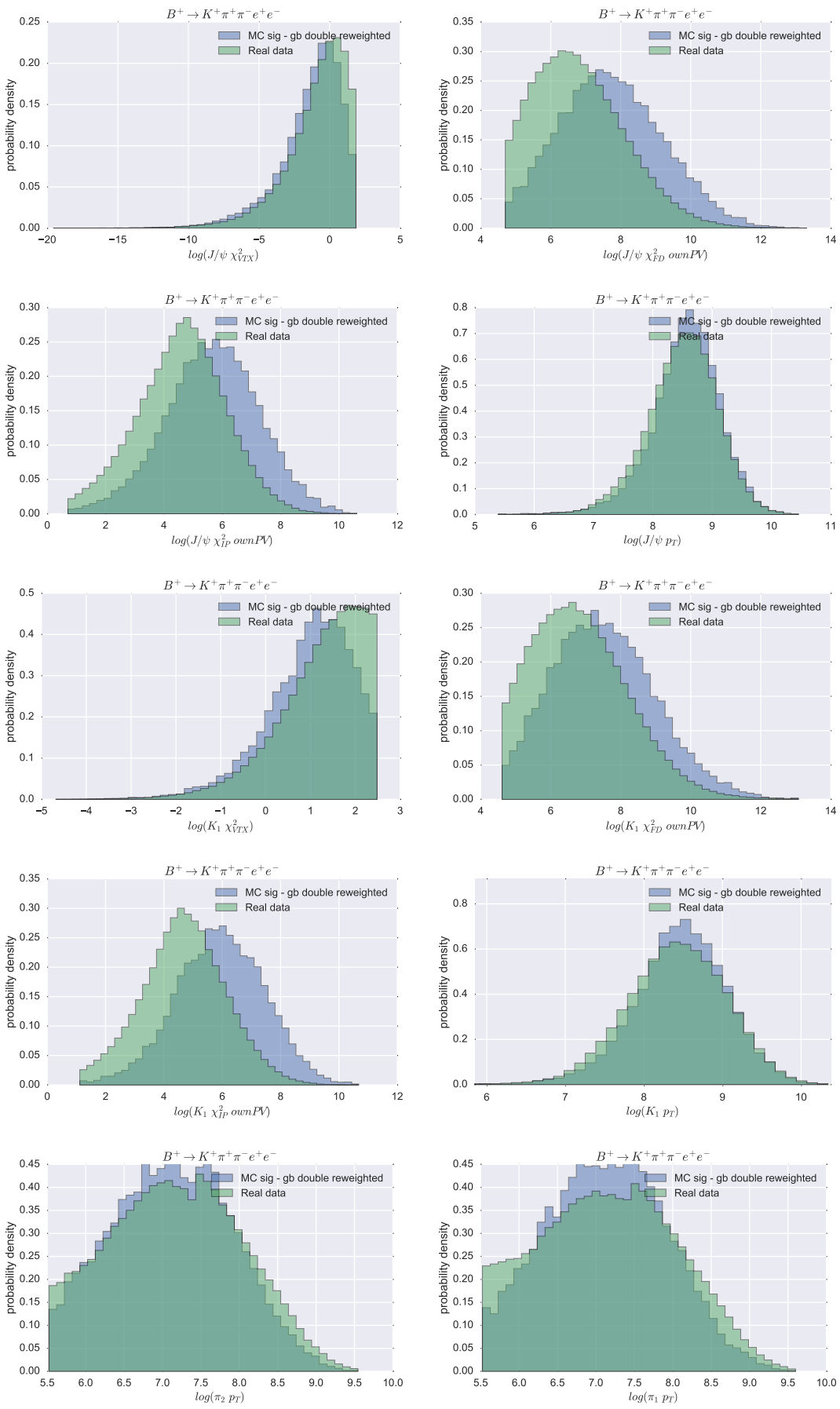
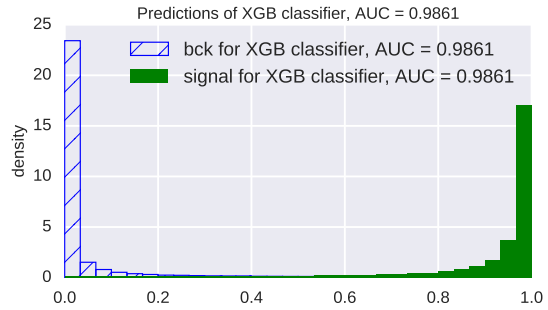
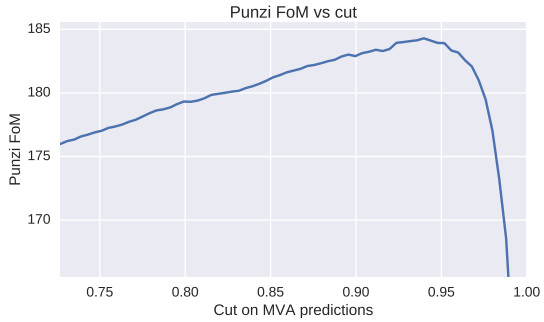


Figure 7: Features used in the training of the BDT.



(a)



(b)

Figure 8: The output of the XGB for the performance evaluation is shown in 8a resulting in a ROC AUC of 0.986. Background (bck) refers to the right sideband. In 8b, the optimisation of the cut is determined. The Punzi FoM is plotted against the cut applied on the predictions. The optimal cut is at 94%.

6 Mass fit

6.1 Yield estimation

To know the order of magnitude of how many events are expected to be found in the fit, two different estimations are calculated. The first one is a standalone estimation, which involves the theoretical predictions of the branching fraction $\mathcal{B}\ (B^+ \to K^+\pi^+\pi^- e^+ e^-)$. The number of events that are in our sample can be estimated

$$n_{events} = \int \mathcal{L}\ dt \cdot \sigma_{b\bar{b}} \cdot \varepsilon_{\text{tot}}/\epsilon_{geo} \cdot f(B^+) \cdot 2 \cdot \mathcal{B}(B \to K_1 e^+ e^-), \tag{3}$$

with an integrated luminosity $\int \mathcal{L} = 1.11\,\text{fb}^{-1}$, a $b\bar{b}$ production cross section in the accepted η region of $\sigma_{b\bar{b}} = 72.0 \pm 0.3\,(\text{stat}) \pm 6.8\,(\text{syst})\ \mu\text{b}$ [15], an efficiency of $\varepsilon_{\text{tot}}/\epsilon_{geo} = 0.377\%$ as the geometric efficiency is already taken into account in the production cross section $\sigma_{b\bar{b}}$, the hadronisation factor of $f(B^+) = 0.377 \pm 0.005\%$, which is obtained using the methods described in [16] with the f_s/f_d ratio from [17] under the assumption that $f_u \approx f_d$, and the branching ratio $\mathcal{B}(B \to K_1 e^+ e^-) = (2.7^{+1.5+0.0}_{-1.2-0.3}) \times 10^{-6}$. This estimation yields $n_{events} \approx 602 \pm 341$ where the total uncertainty is dominated by the statistical uncertainty on $\mathcal{B}(B \to K_1 e^+ e^-)$.

The second estimation uses the already measured ratio of the $B^0 \to K^{*0}\ell^+\ell^-$ decays with ℓ equal to either e or μ. Together with the $\mathcal{B}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-)$, we can estimate the yield for our mode. As several factors are the same between our mode and the $\mu^+\mu^-$ final state, only the yield, the different integrated luminosities and parts of the efficiency have to be taken into account. The estimated number of events is given by

$$n_{events} = \frac{\mathcal{B}(B^0 \to K^{*0} e^+ e^-)}{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)} \cdot n_{events}(B^+ \to K^+\pi^+\pi^-\mu^+\mu^-) \cdot \frac{{\varepsilon_{\text{tot}}}^{e^+e^-}}{{\varepsilon_{\text{tot}}}^{\mu^+\mu^-}} \cdot \frac{\mathcal{L}^{e^+e^-}}{\mathcal{L}^{\mu^+\mu^-}} \tag{4}$$

with the ratio $R_{K^{*0}} = \dfrac{\mathcal{B}(B^0 \to K^{*0}\mu^+\mu^-)}{\mathcal{B}(B^0 \to K^{*0} e^+ e^-)} = 0.69^{+0.11}_{-0.07}\,(\text{stat}) \pm 0.05\,(\text{syst})$ [3], the number of events obtained from the fit to the $\mu^+\mu^-$ final state $n_{events}^{\mu^+\mu^-} = 144.80^{+14.89}_{-14.31}$, the efficiency of the $\mu^+\mu^-$ final state ${\varepsilon_{\text{tot}}}^{\mu^+\mu^-}/\epsilon_{geo} = 1.062$ [4] and the efficiency for our mode obtained in Sec. 5.2 ${\varepsilon_{\text{tot}}}^{e^+e^-}/\epsilon_{geo} = 0.377$, both without the geometric acceptance, the integrated luminosity for the e^+e^- mode $\mathcal{L}^{e^+e^-} = 1.11\,\text{fb}^{-1}$ and for the $\mu^+\mu^-$ mode $\mathcal{L}^{\mu^+\mu^-} = 3.19\,\text{fb}^{-1}$. This yields an estimated number of events of $n_{events} \approx 19.0^{+3.6}_{-2.8}$.

Both estimations do not agree well with each other. Comparing with the measured branch ratio and the predictions of $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$, the measured one is lower by a factor of about six, a similar deviation is expected here. Also the uncertainties on the predicted branching ratio of $B^+ \to K^+\pi^+\pi^- e^+ e^-$ is comparably large. This considerations favour the second estimation, which still would yield enough events for an observation, at least if the data taken in 2012 is used as well.

6.2 Fits

To determine the number of events, a fit to the vertex constrained B invariant mass is performed. From the ROOT software package, the RooFit library with python bindings is used.

The probability density function (pdf) for the fit is constructed using a linear combination of an exponential pdf as background shape and a double crystal-ball (CB) function[9] for the signal shape [18]. A double CB function is a linear combination of two CB functions, the ratio of the two normalisations is a fit parameter. An extended unbinned maximum likelihood fit is performed, leaving the number of the background and signal events as free parameters to the fit.

Four fits are performed in total to fix certain parameters and to correct for simulation differences. First, the fit is performed on the J/ψ samples and the ratio between the MC and data mean is taken to correct the mean obtained in the non-resonant mode with that factor.

1. Fit to the B invariant mass with vertex and J/ψ mass constrained of $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$ MC as shown in Fig. 9
   - Fit without background.
   - All parameters are floating freely including the ration between the two CB functions.

2. Fit to the B invariant mass with vertex and J/ψ mass constrained of $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$ data as shown in Fig. 10
   - Fit with background.
   - Exponential tail parameters and fraction are fixed from MC fit.
   - Free parameters are the mean, width and the scaling.

3. Fit to the B invariant mass with vertex constrained of $B^+ \to K^+\pi^+\pi^-e^+e^-$ MC as shown in Fig. 11
   - Fit without background.

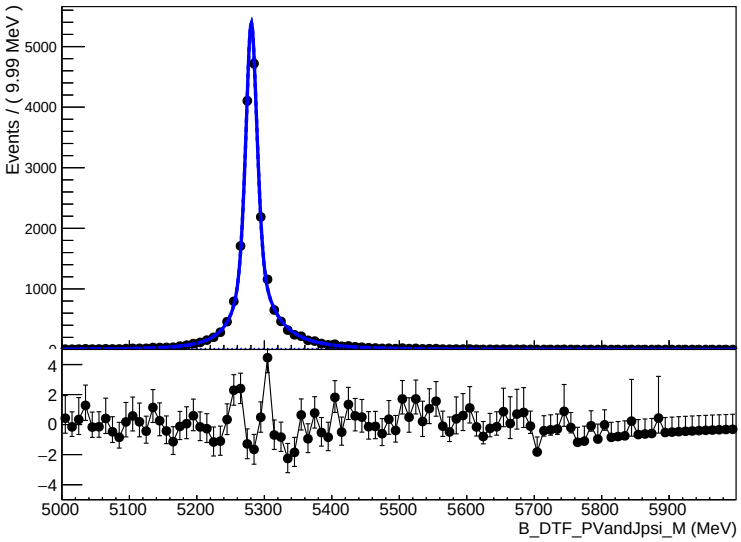


Figure 9: Fit to $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$ MC

[9] A CB function is a Gaussian distribution with exponential tails.

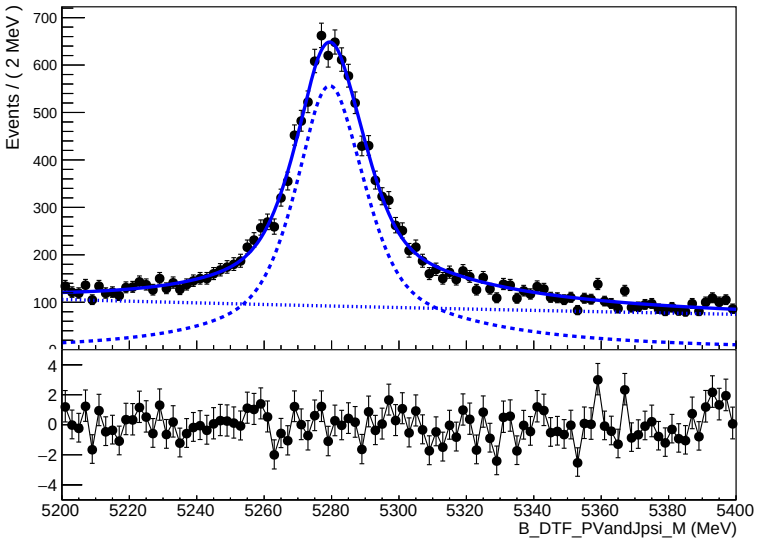


Figure 10: Fit to $B^+ \to J/\psi(\to e^+e^-)K^+\pi^+\pi^-$ data

- All parameters are floating freely.

4. Blind-fit to the B invariant mass with vertex constrained of $B^+ \to K^+\pi^+\pi^-e^+e^-$ data blinding the region $5100 - 5380\,\text{MeV}$ around the B mass of $5279\,\text{MeV}$ as shown in Fig. 12.

   - Fit with background
   - All signal parameters are fixed from the previous MC fit. The mean is corrected by the ratio of the mean between the fits to the MC and data of the J/ψ.

For an unblinding of the fit, a clean signal region is required. As can be seen in Fig. 12 at the lower bound of the blinded region, there seems to be a peak, most probably

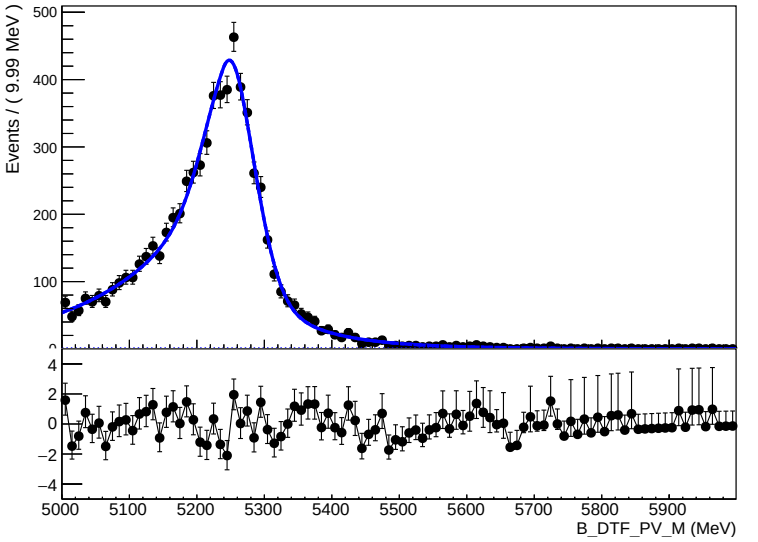


Figure 11: Fit to $B^+ \to K^+\pi^+\pi^-e^+e^-$ MC

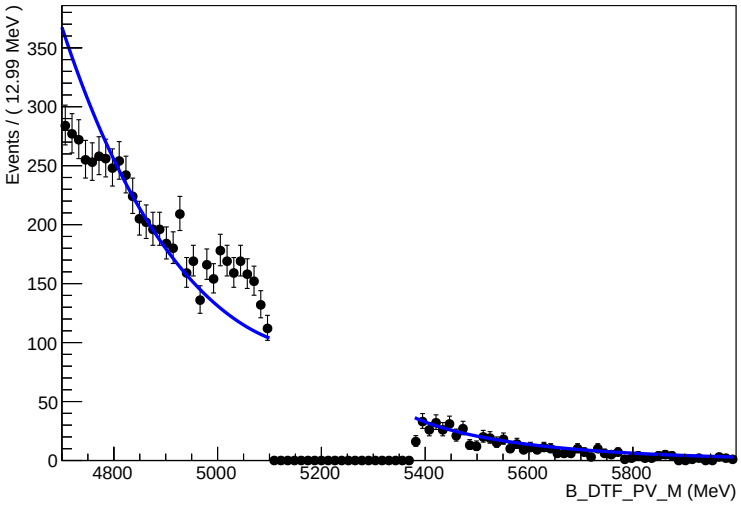


Figure 12: Fit to $B^+ \to K^+\pi^+\pi^-e^+e^-$ data with the region $5100 - 5380\,\text{MeV}$ blinded.

originating from physical background and reaching into our signal region. This background would bias our yield and has to be further investigated before an unblinding of the fit is possible.

7 Discussion

In this thesis the yet unobserved rare B-decay $B^+ \to K^+\pi^+\pi^- e^+ e^-$ at LHCb was studied using a total amount of $1\,\mathrm{fb}^{-1}$ of data. The same mode with $\mu^+\mu^-$ in the final state has already been observed and could be used for ratio tests of the lepton flavour universality.

First, the stripping lines have been applied to the sample. A strong preselection is then applied in order to remove physical contributions to the background. This is necessary to further reduce the combinatorial background in our signal region with a MVA and to have an unbiased yield. The yield is estimated and a blind fit to the vertex constrained B invariant mass is performed successfully.

It turned out that the cuts were insufficient as peaks are occurring in the data right next to the blinded region, which most probably come from physical background and reaches into our signal region. It is further to note that the classifier used in the MVA showed an unexpected high performance. This can happen if the classifier actually trains on the signature of physical background which differs much more from the signal then combinatorial background does. Another reason for this could be the differences between generated and real events which leads to the classifier being trained to distinguish those two instead of signal versus background. On one hand the generated and real sample do not seem to differ too large for this decay and on the other hand a multidimensional reweighting procedure has been applied in order to further reduce the differences.

To continue with this analysis and to perform a first detection, it would be a possible step to perform a more in-deep study of the background and to remove any remaining physical contributions from it.

Acknowledgements

I would like to to thank the Physics Department of the University of Zurich, the CERN collaboration and especially the LHCb collaboration for providing such an impressive infrastructure of computing power, accelerators and detectors.

I would like to express my gratitude to Prof. Nicola Serra from the University of Zurich for letting me do my bachelor thesis in his research group and who managed to give me a very positive impression on the world of data analysis. I like to especially thank my supervisor Dr. Rafael Silva Coutinho for the support and explanations for all kind of problems and questions I had as well as for the encouragement if things did not go as well as expected. I also like to thank Dr. Albert Piug for his support, advices and discussions. I further want to thank my colleague Alexander Daetwyler for explanations and discussions as well as for providing me several, useful code snippets.

A Appendix

A.1 Preselection

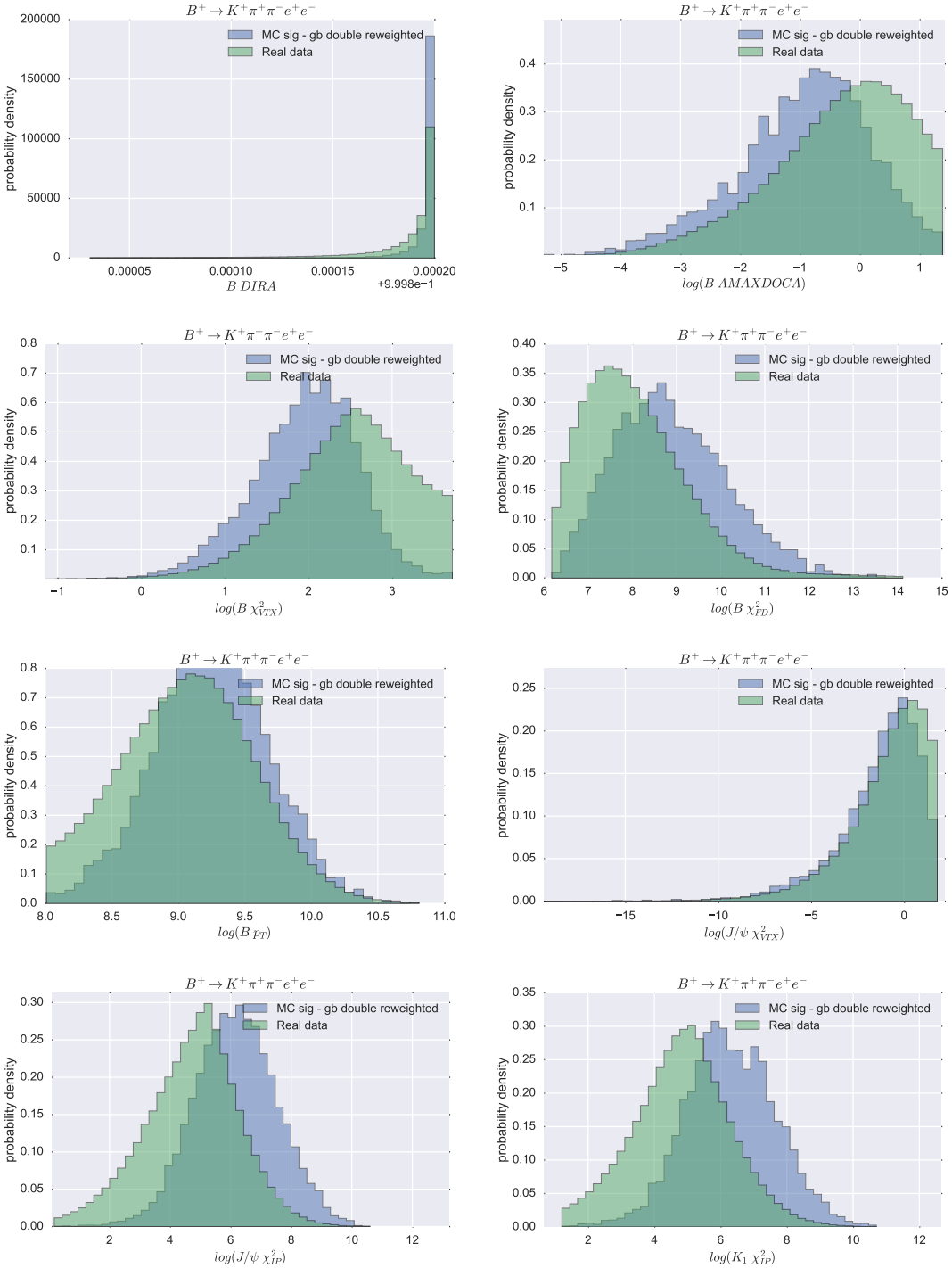


Figure 13: Variables used in the preselection as described in Sect. 3.3

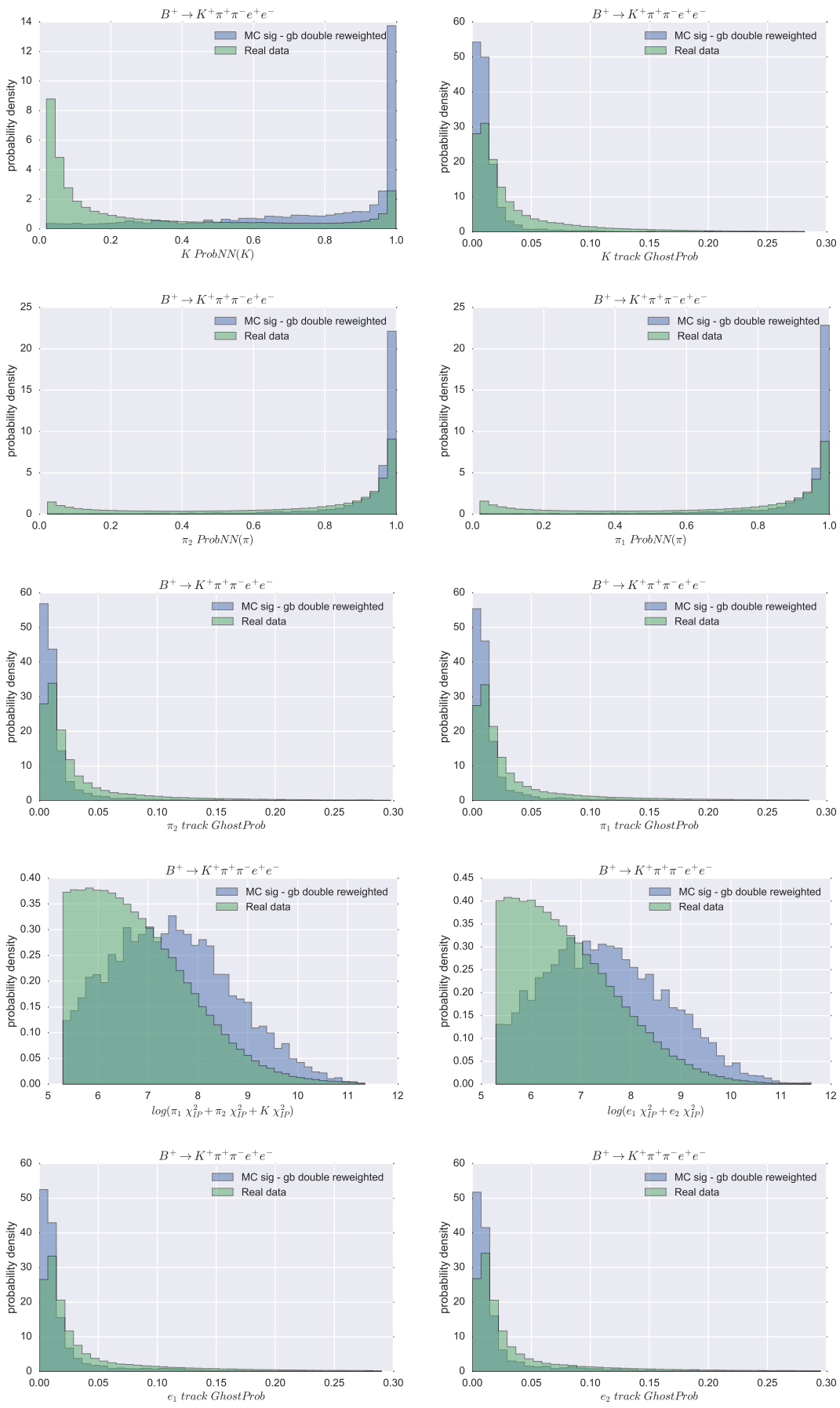


Figure 13: Variables used in the preselection as described in Sect. 3.3

A.2 Reweighting

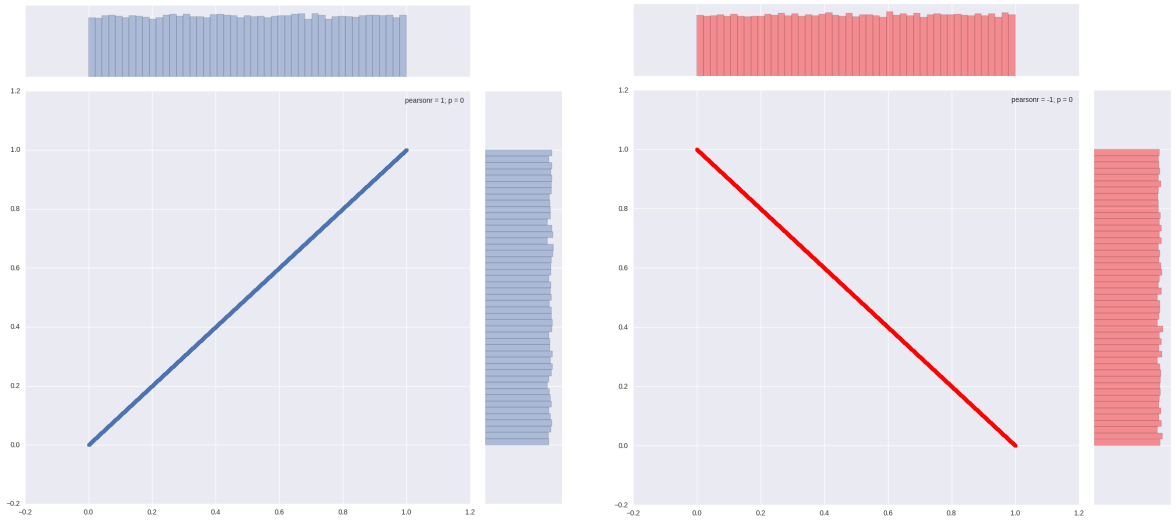


Figure 14: A two dimensional distribution and its projections. Even tough the distributions can be easily discriminated by looking at their higher order – second order here – correlations, there projections do not reveal that.

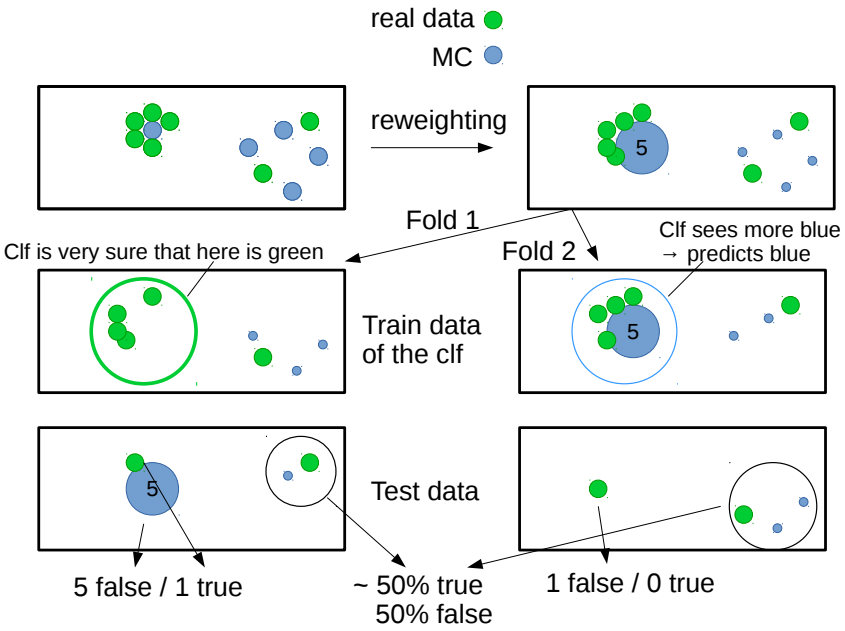


Figure 15: ROC AUC bias with weights visualized. The reweighter works quite well for this example and assigns a weight of 5 to the single blue point. Then the data is split in two different ways (Fold 1 and 2) into training and test data in order to compare two possible outcomes. The total outcome can be thought as an average of both cases.

A.3 Selection

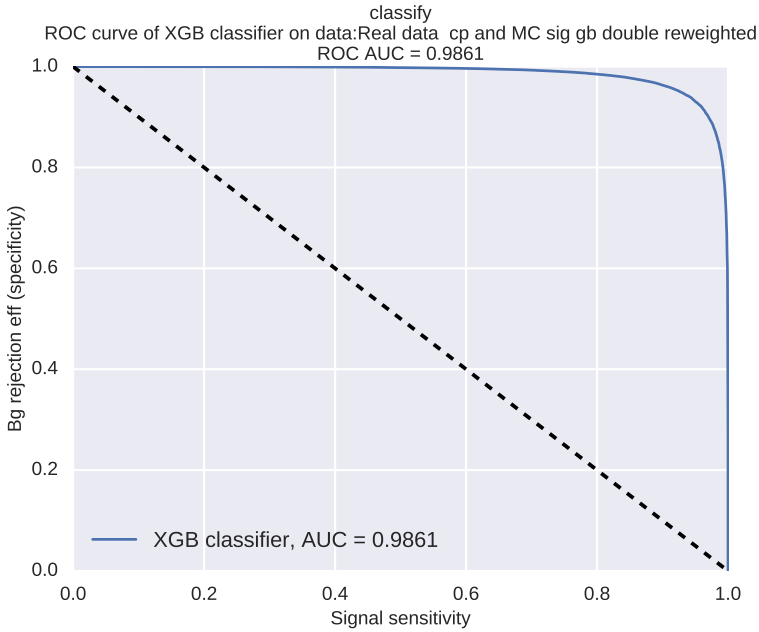


Figure 16: ROC curve of the XGB trained on the MC against the right side band (B mass vertex constrained $> 5600\,\mathrm{MeV}$) of $B^+ \to K^+\pi^+\pi^-e^+e^-$.

References

[1] S. L. Glashow, J. Iliopoulos, and L. Maiani, *Weak interactions with lepton-hadron symmetry*, Phys. Rev. D **2** (1970) 1285.

[2] LHCb, R. Aaij *et al.*, *Test of lepton universality using $B^+ \to K^+\ell^+\ell^-$ decays*, Phys. Rev. Lett. **113** (2014) 151601, arXiv:1406.6482.

[3] LHCb, R. Aaij *et al.*, *Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays*, arXiv:1705.05802.

[4] LHCb, R. Aaij *et al.*, *First observations of the rare decays $B^+ \to K^+\pi^+\pi^-\mu^+\mu^-$ and $B^+ \to \phi K^+\mu^+\mu^-$*, JHEP **10** (2014) 064, arXiv:1408.1137.

[5] H. Hatanaka and K.-C. Yang, $K_1(1270) - K_1(1400)$, Phys. Rev. D **78** (2008) 074007.

[6] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[7] I. Belyaev *et al.*, *Handling of the generation of primary events in Gauss, the LHCb simulation framework*, J. Phys. Conf. Ser. **331** (2011) 032047.

[8] M. Pivk and F. R. Le Diberder, *sPlot: A statistical tool to unfold data distributions*, Nucl. Instrum. Meth. **A555** (2005) 356, arXiv:physics/0402083.

[9] LHCb collaboration, A. A. Alves Jr. *et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.

[10] A. P. Bradley, *The use of the area under the roc curve in the evaluation of machine learning algorithms*, Pattern Recogn. **30** (1997) 1145.

[11] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, Wadsworth international group, Belmont, California, USA, 1984.

[12] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, J. Comput. Syst. Sci. **55** (1997) 119.

[13] T. Chen and C. Guestrin, *Xgboost: A scalable tree boosting system*, CoRR **abs/1603.02754** (2016).

[14] G. Punzi, *Sensitivity of searches for new signals and its optimization*, in *Statistical Problems in Particle Physics, Astrophysics, and Cosmology* (L. Lyons, R. Mount, and R. Reitmeyer, eds.), p. 79, 2003. arXiv:physics/0308063.

[15] LHCb, R. Aaij *et al.*, *Measurement of the b-quark production cross-section in 7 and 13 TeV pp collisions*, Phys. Rev. Lett. **118** (2017), no. 5 052002, arXiv:1612.05140.

[16] LHCb collaboration, R. Aaij *et al.*, *Search for the $\Lambda_b^0 \to \Lambda\eta$ and $\Lambda_b^0 \to \Lambda\eta'$ decays with the LHCb detector*, JHEP **09** (2015) 006, arXiv:1505.03295.

[17] LHCb collaboration, R. Aaij *et al.*, *Measurement of the fragmentation fraction ratio f_s/f_d and its dependence on B meson kinematics*, JHEP **04** (2013) 001, arXiv:1301.5286.

[18] T. Skwarnicki, *A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances*, PhD thesis, Institute of Nuclear Physics, Krakow, 1986, DESY-F31-86-02.