ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Reconstruction of semileptonic decays and search for $B_s^0 \to K^{*-}\mu^+\nu_\mu$ at the LHCb experiment

Author:
Surapat EK-IN

Supervisor:
Dr. Michel DE CIAN

Director:
Prof. Tatsuya NAKADA

Abstract

A search for the semileptonic decay of $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$ is performed on the data collected by the LHCb experiment in proton-proton collisions at the centre-of-mass energy of 8 TeV corresponding to the integrated luminosity of 2.08 fb^{-1}. This decay is induced by a $b \to u$ transition at quark level which allows access to a measurement of the Cabibbo-Kobayashi-Maskawa matrix element $|V_{ub}|$. A novel technique using neural networks for the reconstruction of B_s^0 mesons is presented. By applying this technique, the invariant mass of the $\mu^+\nu_\mu$ system can be determined. This has profound applications in studies with non-fully reconstructed semileptonic decays.

Contents

1 Introduction

The Standard Model (SM) of particle physics describes fundamental particles and three forces of nature: the electromagnetic (EM), strong, and weak force. The SM consists of fermions and bosons. The fermions have spin $^1/_2$ and respect the Fermi-Dirac statistics. They can be classified into two types depending on their interactions with fundamental forces. Leptons include particles that couple with the weak force. They are grouped into charged leptons (electron (e), muon (μ), and tau (τ)), which additionally interact with the EM force, and their corresponded neutrinos (ν_e, ν_μ, and ν_τ). The other types are quarks, which can couple with all of three forces. They are categorised into up-type (up (u), charm (c), and top (t)) and down-type (down (d), strange (s), and beauty (b)) which have $+^2/_3$ and $-^1/_3$ of electric charge, respectively. The interaction between these particles can be considered as an exchange of a force carrier - a gauge boson. This gauge boson has spin 1 and follows Bose-Einstein statistics. The SM includes the photon (EM force), gluon (strong force), and $W^\pm, Z^0$ (weak force) as spin 1 gauge bosons. Massive particles in the SM acquire their masses due to the Higgs mechanism. This mechanism excites a sole scalar Higgs boson (spin 0) in the SM and this boson was discovered at the Large Hadron Collider in 2012 at CMS and ATLAS detectors [1, 2]. These fundamental particles are summarised in Figure 1.

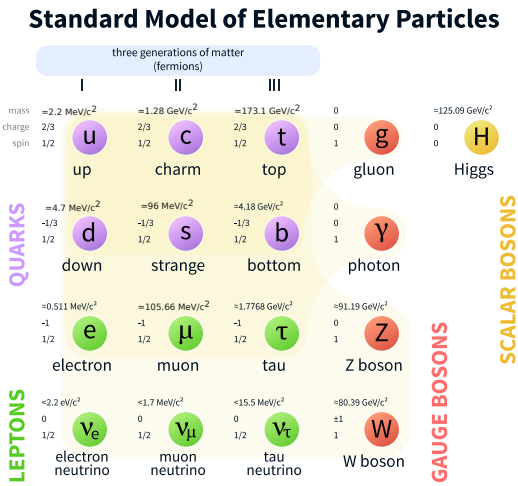


Figure 1: The Standard Model of particle physics [3].

The Cabibbo-Kobayashi-Maskawa (CKM) matrix represents the transition amplitude of quark flavour-changing through the exchange of a $W^\pm$ boson. This mechanism arises from the Yukawa interaction between spinor massless quark fields with the scalar Higgs field [4]. The interaction results in transitions between up-type and down-type left-handed chiral quark states. The matrix can be parametrised with 4 free parameters, but these values are not predicted in the SM. Measurements of these parameters in different processes must be consistent in the SM.

The quark flavour mixing parameter of u and b quarks ($|V_{ub}^L|$) in the CKM matrix is one of the least well-known parameters. Measurements of this quantity show a difference between inclusive and exclusive semileptonic decays of b-mesons, as reviewed in Section 1.1. In the SM, the W-boson only interacts with a left-handed chiral fermion - called a left-handed current. For simplicity's sake, $|V_{ub}|$ throughout this report refers to only left-handed coupling. A measurement of $|V_{ub}|$ using $\Lambda_b^0 \to p\mu^-\overline{\nu}_\mu$

from the LHCb collaboration seems to confirm this discrepancy [5]. Figure 2 represents the combined results of left-handed coupling $|V_{ub}|$ in inclusive and exclusive decays as a function of the fractional right handed coupling (ϵ_R). The overlap of the 68% confidence level bands of exclusive modes suggest no right-handed coupling as predicted by the SM, whilst the overlap between exclusive and inclusive does. This provides a strong motivation for a study of other exclusive semileptonic decays of the b-hadron to support the measurement of $|V_{ub}|$.

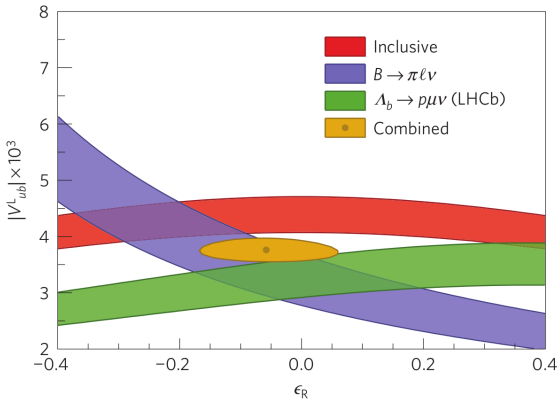


Figure 2: Experimental constraints on the left-handed coupling, $|V^L_{ub}|$ and the fractional right-handed coupling, ϵ_R [5].

In this report, we present a search for $B^0_s \to K^{*-}\mu^+\nu_\nu$ at the LHCb experiment. This decay provides access to a measurement of the quark coupling strength $|V_{ub}|$. Since there is missing information from the neutrino in the decay, the invariant mass of the B^0_s meson cannot be reconstructed, making the discrimination against background difficult. Regression techniques are implemented, allowing for a prediction of the B^0_s momentum as well as the invariant mass of $\mu^+\nu_\nu$ system, which measurements of physics quantities, *i.e.* branching fraction, left-right handed weak currents, depend on.

The outline of this report is arranged as follows. In this section, the physics motivations are given. Several techniques for a reconstruction of B^0_s mesons are evaluated in Section 3, using Monte Carlo simulation at the LHCb detector.We summarise an optimisation study on the LHCb data and a measurement of the decay's branching fraction in Section 4. We note that the analysis described in this report is performed on both $B^0_s \to K^{*-}(\to K^0_S\pi^-)\mu^+\nu_\mu$ and $\overline{B}^0_s \to K^{*+}(\to \overline{K}^0_S\pi^+)\mu^-\overline{\nu}_\mu$; for simplicity, only $B^0_s \to K^{*-}(\to K^0_S\pi^-)\mu^+\nu_\mu$ will be explained unless stated otherwise. As the decay chain in this study is complicated, a convention $(A|B)$ is introduced to represent a daughter particle A decaying from a resonance B, *i.e.* $(\pi^-|K^{*-})$ implies a π^- that comes from a K^{*-}.

1.1 CKM matrix and measurement of $|V_{ub}|$

The Cabibbo-Kobayashi-Maskawa (CKM) matrix describes the transitions between quark states [6, 7]. Each element V_{ij} represents a transition probability between state i and j. Since there are 6 flavours of quarks in the Standard Model, the matrix can be written as:

$$V_{CKM} \equiv \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1}$$

which is required to be 3×3 unitary matrix, e.g. $V^*_{CKM} V_{CKM} = 1$, if there exists no other quark flavours.

The matrix can be parametrised in terms of three real angles and one physical phase which is known as Kobayashi-Maskawa phase (KM phase). There are many possible parametrisations. It is useful to show the Wolfenstein parametrisation [8, 9]. Then, V_{CKM} can be rewritten as follows (to $\mathcal{O}(\lambda^3)$):

$$V_{CKM} = \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3[1 - (\rho + i\eta)] & -A\lambda^2 & 1 \end{pmatrix}, \tag{2}$$

where A, λ, ρ, and η are the free parameters which rely on experimental measurements. The KM phase can be measured by the complex term $\rho + i\eta$. This is the only Charge-Parity violation term in the SM. Elements in the matrix are fundamental parameters of the SM, therefore their precise determination is crucial and their values have to be consistent with each other.

The determination of $|V_{ub}|$ is complicated due to large background from $b \to c$ transitions. The most precise measurement of $|V_{ub}|$ is quoted from the inclusive decay $B \to X_u e \nu_e$ by the CLEO, BABAR, and Belle experiments. The combined value of $|V_{ub}|$ is found to be [4]

$$|V_{ub}| = (4.41 \pm 0.15^{+0.15}_{-0.19}) \times 10^{-3} \tag{3}$$

However, study of an exclusive mode can be pursued if the form factors are known. This factor describes the internal structure of a hadron and is computed theoretically by using lattice quantum chromodynamics calculations (LQCD) or light-cone sum rules (LCSR). The BABAR experiment studied the exclusive mode $B \to \pi \ell \nu$. With the available and theoretical calculations of the form factor at different ranges of invariant mass square of the $\ell\nu$ system (q^2), the combined value of $|V_{ub}|$ is found to be inconsistent with the inclusive mode as [10, 11]

$$|V_{ub}| = (3.28 \pm 0.29) \times 10^{-3} \tag{4}$$

The most recent measurement of $|V_{ub}|$ is from the LHCb collaboration in a b-baryon decay. A neutrino is not detected in the LHCb detector, causing missing information in the decay. Semileptonic decays of a b-hadron are not well-reconstructed because 4-momentums of interacting particles in pp collisions, gluons and quarks inside protons, are unknown. An accurate measurement of $|V_{ub}|$ can be obtained by measuring the ratio of branching fractions with respect to $|V_{cb}|$.

$$\frac{|V_{ub}|^2}{|V_{cb}|^2} = \frac{\mathcal{B}(\Lambda_b^0 \to p\mu^-\overline{\nu}_\mu)}{\mathcal{B}(\Lambda_b^0 \to \Lambda_c^+\mu^-\overline{\nu}_\mu)} R_{FF} \tag{5}$$

R_{FF} is a ratio between the from factors of those decays. By using the world average value of $|V_{cb}|$ and LQCD calculation for the considered decay mode, $|V_{ub}|$ is found to be [5]

$$|V_{ub}| = (3.27 \pm 0.15 \pm 0.16 \pm 0.06) \times 10^{-3} \tag{6}$$

This seems to agree with the previously measured result in exclusive decays. Therefore, more studies of $b \to u$ transitions should be undertaken.

1.2 Physics motivation for studying $B_s^0 \to K^{*-}$ transitions

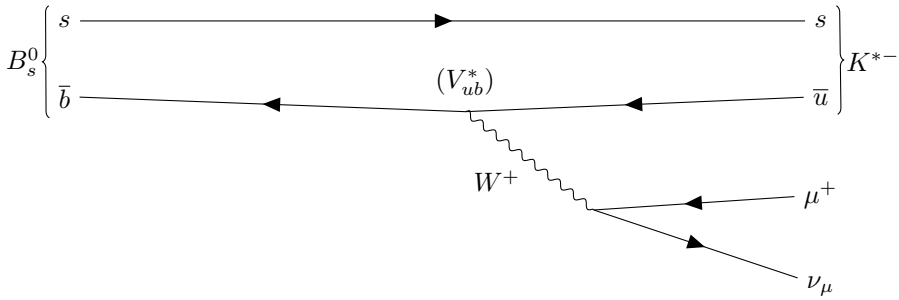


Figure 3: Feynman diagram for the process $B_s^0 \to K^{*-}\mu^+\nu_\mu$.

B_s^0 mesons decaying to $K^-\mu^+\nu_\mu$ provide a promising benchmark for the $|V_{ub}|$ measurement, since the LQCD calculation for this decay is available [13]. However, the B_s^0 decaying to $K^{*-}\mu^+\nu_\nu$ is more interesting because the structure of the weak current in the W-boson exchange as shown in Figure 3 is accessible by performing an angular analysis as the K^{*-} decays particularly to $K_S^0\pi^-$. The left-right coupling of the weak interaction can thus be measured directly in this decay, to further support (or weaken) the contradiction of $|V_{ub}|$ measurements. The form factors from LCSR calculations have as well been determined [14]. We show a schematic of the decay chain together with the angular observables in Figure 4.

Furthermore, the combination study of this decay with the recent angular analysis of $B^0 \to K^{*0}\mu^+\mu^-$ [15] is favourable. The secondary decay $K^* \to K\pi$ is identical and leads to an one-to-one correspondence between angular observables [16].

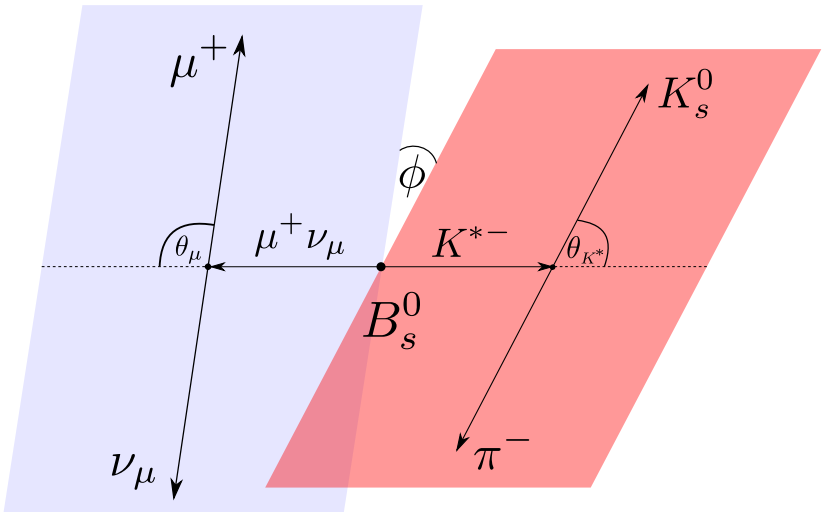


Figure 4: Schematic of the decay $B_s^0 \to K^{*-}\mu^+\nu_\nu$ in B_s^0's reference frame, defined according to the Jacob-Wick helicity amplitude decomposition formalism - Jackson's Convention [17, 18]. Note that the arrows in the figure refer to the momentum of the particle in its mother's rest frame. The measurement of the angular observables θ_μ, θ_{K^*}, and ϕ can be used to test the consistency with the SM scenario.

1.3 Challenges in a search for semileptonic decays of b-mesons

The quark flavour-changing is associated with the charged weak interaction, involving an exchange of a W-boson. As shown in the Feynman diagram of the process $B_s^0 \to K^{*-}\mu^+\nu_\mu$ (Figure 3), the W-boson is emitted from the transition of $\bar{b}$ quark state to $\bar{u}$ quark state. This W-boson can couple to the lepton and its corresponding neutrino.

A semileptonic decay is a type of decay in which a lepton is produced in association with other particles. There are 3 charged leptons including electron (e), muon (μ), and tau (τ), in the SM. Due to the electron's mass, which is the smallest among charged leptons, the electron loses its energy via Bremsstrahlung process more than others. On the other hand, the tau lepton is the heaviest, but it decays to other hadrons after travelling a few millimetres in the LHCb detector. This results in a difficult reconstruction.

Unlike electron and tau, when a muon is produced at high energy, it is able to propagate through the whole detector without decaying to other particles. The energy loss of the muon in the material in the detector is mostly via the ionisation process as it does not interact hadronically. Since the muon produced at the LHC experiments are dominantly minimum ionizing particles, the muon only deposits little energy in the detector. Thus, a muon can penetrate through several metres of material, and this is the reason why the muon detector is placed at the very end of the detector. This makes the muon identification sufficiently simple. Therefore, the semileptonic decay with the muon and muon neutrino final state is chosen for the search in this study.

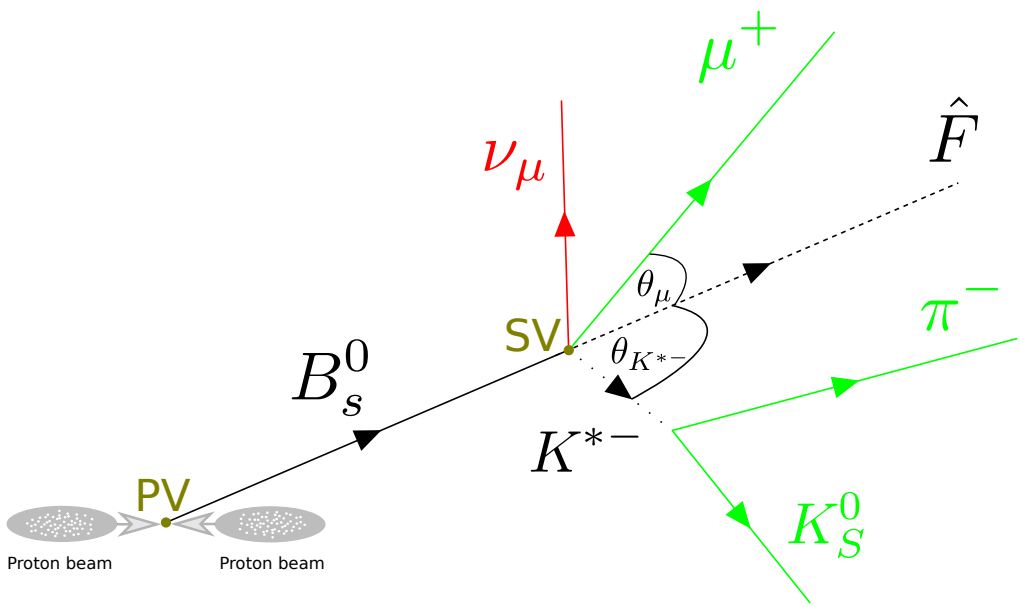


Figure 5: Schematic decay of $B_s^0 \to K^{*-}\mu^+\nu_\mu$. The point where the B_s^0 is produced is denoted as the primary vertex (PV). The B_s^0 meson propagates until it decays at the scondary vertex (SV). Arrows indicate momentum vectors of each particle of a long-lived particle (green), undetected particle (red), and unstable particle (black). The dotted line represents a prompt decaying particle whose track is not detected in a detector. Note that the K_S^0 decays further specifically to $\pi^-\pi^+$. The unit flight vector ($\vec{F}$) is shown as a dashed black line.

Figure 5 represents the decay chain of $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\nu$, where μ^+, π^-, and K_S^0 travel and can be detected. A neutrino interacts only weakly with matter and cannot be detected. This causes an unbalanced momentum of the B_s^0 decay products in a reconstructed event. To reconstruct the B_s^0 meson, the momentum of the neutrino ν_μ is necessary to be known. The transverse component of the neutrino with respect to the B_s^0's flight vector $\vec{F}$ ($P_{\nu_\mu}^{\perp}$) is obtained from the vector sum of the transverse components of K^{*-} and μ^+. The parallel component ($P_{\nu_\mu}^{\perp}$) can be computed using the Energy-Momentum conservation by projecting the decay products' vector ($\vec{P}_A$; $A = K^{*-}, \mu^+$) to $\vec{F}$ as

$$P_A^{\perp} = |\vec{P}_A \times \hat{F}| \tag{7}$$

$$P_A^{||} = \vec{P}_A \cdot \hat{F} \tag{8}$$

for perpendicular and parallel components. Then, the momentum of B_s^0 can be written as

$$P_{B_s} = P_{vis}^{||} + P_{miss}^{||} \tag{9}$$

$$P_{miss}^{||} = -a \pm \sqrt{r} \tag{10}$$

where the $P_{miss}^{||}$ is the parallel momentum of the neutrino. This is obtained from solving a quadratic equation for $P_{miss}^{||}$ which results in two solutions. The subscript vis refers to the visible components where all detected particles are considered, e.g. E_{vis} (P_{vis}) is the energy (momentum) calculated from the 4-vector sum of the μ^+ and K^{*-} energy (momentum). The parameters a and $\sqrt{r}$ are

$$\begin{aligned} a &= \frac{P_{vis}^{||}\left(m_{B_s}^2 - M_{vis}^2 - 2(P_{vis}^{\perp})^2\right)}{2\left((P_{vis}^{||})^2 - E_{vis}^2\right)} \\ r &= \frac{E_{vis}^2\left(m_{B_s}^2 - M_{vis}^2 - 2(P_{vis}^{\perp})^2\right)^2}{4\left((P_{vis}^{||})^2 - E_{vis}^2\right)^2} + \frac{\left(E_{vis}P_{vis}^{\perp}\right)^2}{(P_{vis}^{||})^2 - E_{vis}^2} \end{aligned} \tag{11}$$

Since the decay width of the B_s^0 meson is relatively small compared to its mass ($m_{B_s} = 5366.77 \pm 0.24\,\mathrm{MeV}/c^2$), m_{B_s} in Equation 11 can be fixed. The two solutions of the B_s^0's absolute momentum are denoted as

$$\begin{aligned} P_+ &= P_{vis}^{||} - a + \sqrt{r} \\ P_- &= P_{vis}^{||} - a - \sqrt{r} \end{aligned} \tag{12}$$

These two solutions arise when we look at this problem in the B_s^0's rest frame. The daughter particles, including K^{*-}, μ^+, and ν_μ, are boosted to the B_s^0's rest frame. There are two topologies of the ν_μ momentum that have the same transverse momentum with respect to the B_s^0 momentum vector. It is unknown how often which solution occurs. This causes the study of general semileptonic decay difficult in the signal reconstruction and, hence, background discrimination.

A recent study made an effort to predict the correct solution of quadratic equation by considering the flight vector of the B_s^0 meson. It provides a possible way to study semileptonic decay as well as background optimisation [20]. However, the resolution of the predicted P_{B_s} is broad. We extend this work in Section 3 by using advance multivariate analysis methods, deep neural networks, to further improve the B_s^0 resolution and the rate of methods for predicting the corrected solution.

Since interesting physics properties depend mostly on the invariant mass square of the $\mu\nu_\mu$ system ($q^2(\mu\nu_\mu)$), *e.g.* branching fraction, or hadronic form factors [21]; one needs a way to calculate it. A

study from [22] with $B^0 \to D^- \mu^+ \nu_\nu$ suggests that the $q^2(\mu\nu_\mu)$ strongly correlates with the momentum, perpendicularly to the b-meson flight vector, ($P^{\perp}_{K^*}$) of the hadron produced in association with $\mu\nu_\mu$ which is the D^- in the study. The correlation between $P^{\perp}_{K^*}$ and $q^2(\mu\nu_\mu)$ is investigated, and found to be correlated with -87.3 percent using the Pearson's correlation coefficient as shown in Figure 6. This component may be used to represent the invariant mass square of $\mu\nu_\mu$ system.

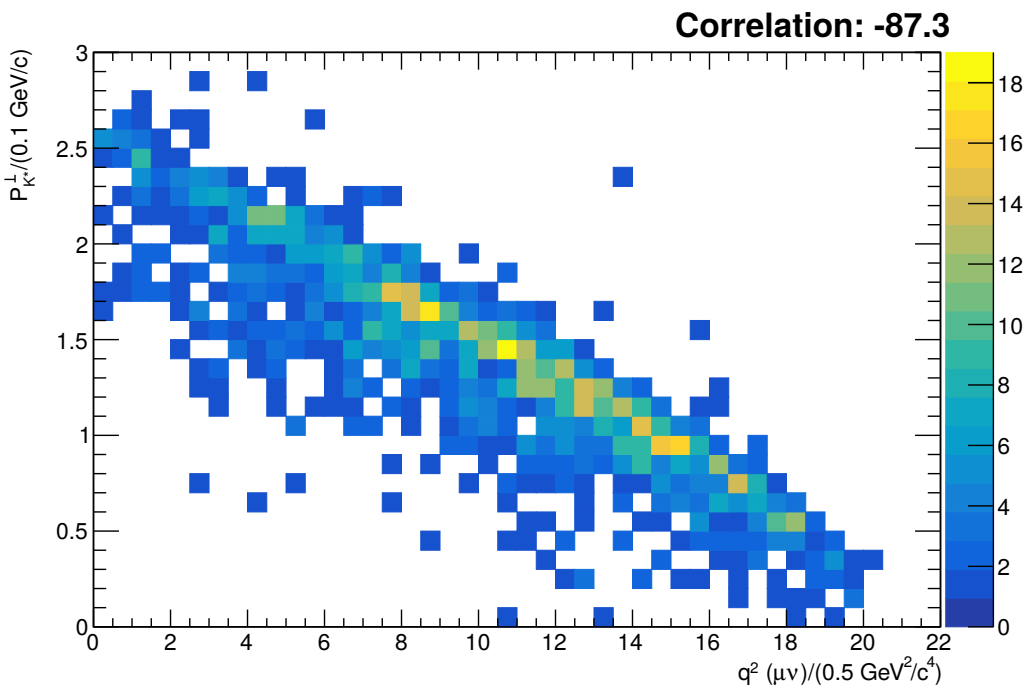


Figure 6: Correlation between the perpendicular momentum of the K^{*-} ($P^{\perp}_{K^*}$) and invariant mass square of the $\mu\nu_\mu$ system ($q^2_{\mu\nu_\mu}$). The sample is acquired from Monte Carlo simulation and is the same as used in Section 3.

Another possibility to study the semileptonic decay of the B^0_s meson is to predict accurately the momentum of the B^0_s. Then, the $q^2(\mu\nu_\mu)$ can be determined from the square of an invariant quantity between the projected 4-momentum vector of the B^0_s and K^{*-} on the B^0_s's flight vector with the fixed mass m_{B_s}.

$$q^2(\mu\nu_\mu) = \left(\sqrt{P^2_{B_s} + m^2_{B_s}} - E_{K^*}\right)^2 - (P_{B_s} - P^{||}_{K^*})^2 - (P^{\perp}_{K^*})^2 \tag{13}$$

where P_{B_s} and E_{K^*} are the predicted momentum of B_s and the energy of K^{*-}, respectively. The reconstructed q^2 throughout this study will be according to the above equation where P_{B_s} is resolved depending on methods as will be described in Section 3.

2 The LHCb detector

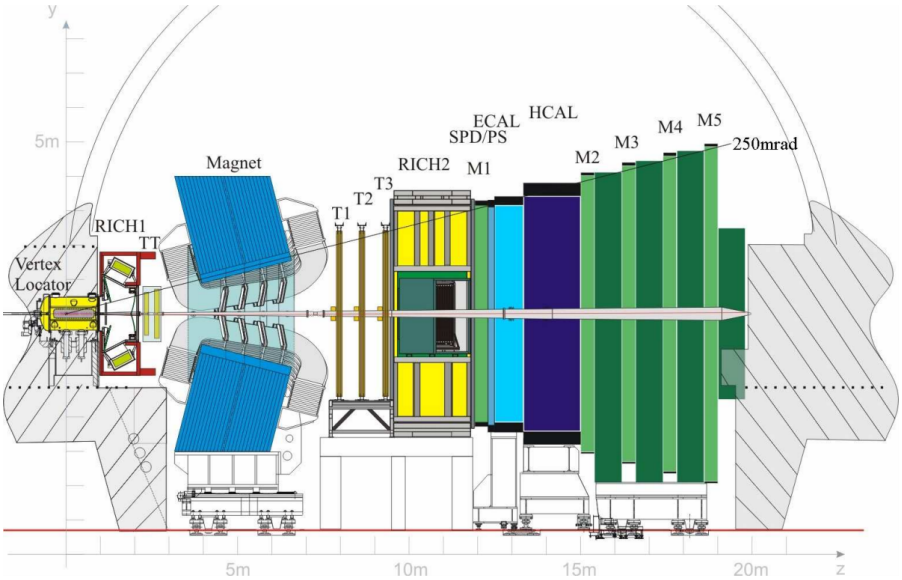


Figure 7: The LHCb detector

The LHCb detector is a single-arm forward spectrometer dedicated to search for indirect evidence of new physics in the decay of particles comprised of a beauty or charm quark. The coverage of the detector is in the range $2 < \eta < 5$ of pseudorapidity (η) which is optimised for b-hadron production at the Large Hadron Collider (LHC). The $b\bar{b}$ cross section in this region is higher than in the range $0 < \eta < 2$. After they are produced, they hadronise to form b-hadrons, e.g. B^0, $B^{\pm}$, B_s^0, Λ_b^0 and etc. [23]. The distinguishing feature of the collision point at this detector is the choice of low pile-up events. The pile-up is a quantity of simultaneous collisions at the collision point which is around ~ 3 pp collisions at the LHCb detector (depending on year and run conditions). This leads to clean a event reconstruction and efficient computing performance. The detector is mainly composed of a vertex locator, a tracking system, Cherenkov detectors, calorimeters, and muon chambers as shown in Figure 7 [24–26].

2.1 Vertex locator (VELO)

The VErtex LOcator (VELO) is a silicon sensor strip detector placed around the collision point covering a geometrical acceptance of $2 < \eta < 5$. It aims to specify precisely the primary vertex (PV) and secondary vertex (SV) in interactions. This is important in a study of b and c hadrons where they are formed at the PV and decay at the SV. The VELO is designed as two semicircular sensor shapes. During the data acquisition period, these two parts are moved toward each other and leave an aperture at its centre for the LHC proton beam propagation. To avoid radiation damage and prolong the lifetime of the detector, these halves are separated outside the data acquisition period. The VELO has resolutions of impact parameters in the transverse to the beam axis $x - y$ plane (IP_x and IP_y)

depending on a particle transverse momentum (p_T) as shown in Figure 8.

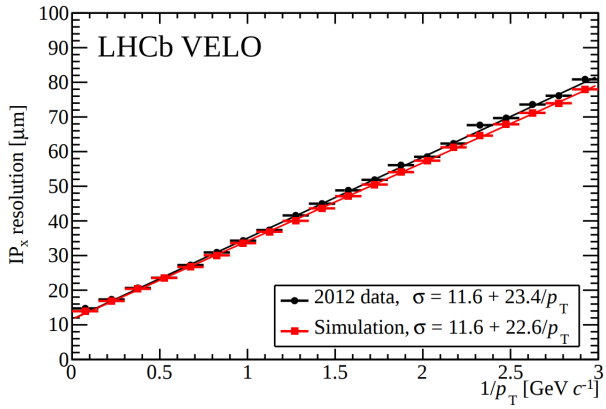


Figure 8: Resolution of impact parameters as a function of an inverse of transverse momentum of a particle [28].

2.2 Tracking system

A large-area silicon-strip detector, Tracker Turicencis (TT), located upstream of a dipole magnet with a bending power of about 4 Tm, and three stations of silicon-strip detectors and straw drift tubes (T1, T2, and T3) placed downstream of the magnet. It provides a measurement of the momentum $\vec{p}$ of charged particles using the particles' curvature. The resolution of the measured particle's momentum depends on hit and track reconstruction as well as the magnetic field. The tracking system can provide a resolution around $0.4\% - 0.6\%$ in a range 10 GeV $< |\vec{p}| <$ 60 GeV of momentum.

Types of tracks detected in the LHCb detector are categorised depending on how the tracks are found in different subdetectors as shown in Figure 9. The Velo track is a track detected only in the $VELO$, while the long track is detected additionally in the whole tracking system. The upstream track is similar to the long track but the track is bended out of the detector before the three tracker stations. The tracks which are not shown in the $VELO$ are downstream and T tracks. The T track is found only in the three stations where the downstream is also found in the TT.

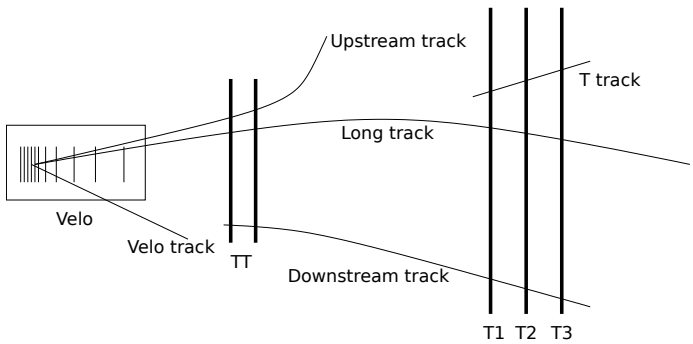


Figure 9: Track types available in the LHCb detector [29].

2.3 Ring Imaging Cherenkov Detectors (RICH)

Two stations of RICH detectors are located after VELO (RICH1) and the downstream trackers (RICH2) as shown in Figure 7. They are a photodetector that captures Cherenkov radiation from high energy particles propagating with higher velocity than the speed of light in a medium. The relation between particle's velocity in a natural unit, β, and the angle of Cherenkov light with respect to a particle's velocity vector θ is

$$\beta = \frac{1}{n \cos\theta} \ , \tag{14}$$

where n is the refractive index of the medium. The information obtained from the RICH detectors plays an important role in particle identification at the LHCb detector, since particles have different masses leading to different signatures in the relation between Cherenkov angle and particle momentum as shown in Figure 10.

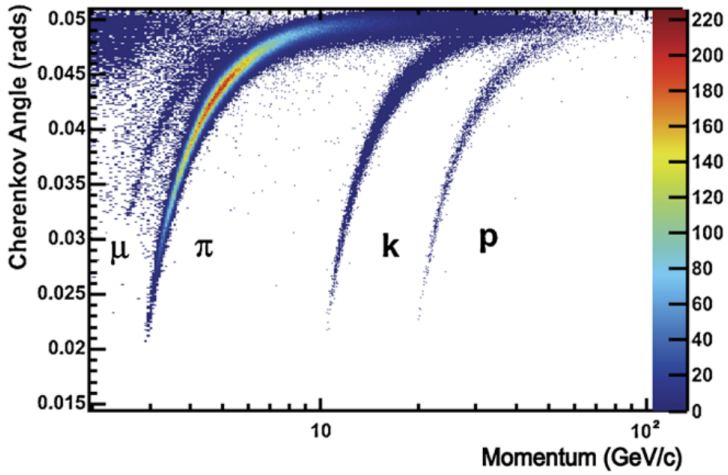


Figure 10: Reconstructed Cherenkov angle as a function of track momentum in C_4F_{10} radiator [30].

2.4 Calorimeter system

The calorimeter system is designed to absorb and measure a particle's energy when it travels through the detector. In addition to the particle identification, photons, electrons and hadrons are identified by a calorimeter system consisting of a scintillating pad detector and preshower detector (SPD and PS), an electromagnetic calorimeter (ECAL) and a hadronic calorimeter (HCAL). The SPD and PS are subdetectors made of scintillators. They provide fast particle identification and background rejection during an online data acquisition. Particles, interacting via the EM force, will lose their energies by inducing electromagnetic showers in the ECAL. Hadronic particles which can interact through strong interaction deposit their energies in a hadronic calorimeter (HCAL). The signature of each long-lived particle (electron e, hadron h, and photon) in energy deposition in the calorimeter system is shown in Figure 11.

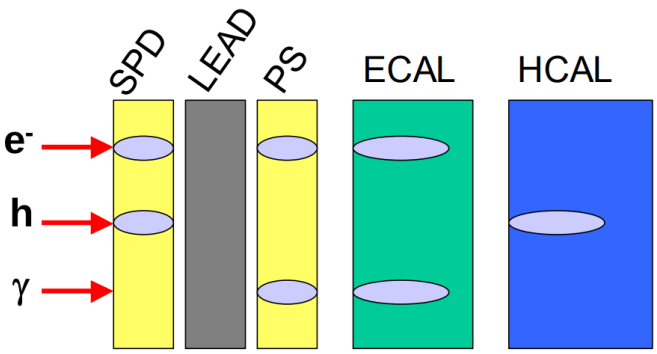


Figure 11: Signal deposited on the different parts of the calorimeter by an electron, a hadron, and a photon [31].

2.5 Muon chambers

Bremsstrahlung radiation is EM radiation emitted when a charged particle is accelerated. The emission power of this process is inversely proportional to the mass of the charged particle. This is the main process for an electron to produce the electromagnetic shower the calorimeters. Since the muon mass is 200 times the electron mass, a muon deposits less energy in the calorimeters. It is also a lepton which does not interact strongly. To detect the muons, muon chambers are placed at the very end of the detector. The system is composed of alternating layers of iron and multiwire proportional chambers to detect ionisation of the gas inside the chambers by the muon. The chambers are composed of five stations (M1-M5) located after the calorimeters except M1. This station is for the precise position measurement before the multiple scattering in the calorimeters. The particle rates in the M1 station are much higher than the rest. The triple gas electron multiplier (GEM) is required for the M1 detector close to the beam pipe to deal with the harsh environment.

2.6 Trigger system

The trigger system is a system that decides whether events occurring in a detector are worth to be kept or not. High energy physics experiments are generally searching for events, which have a relatively low probability to be produced. This system will optimise the selection collecting signal events and rejecting background events. Typically in the LHC experiment, the system is divided into hardware and software triggers.

The hardware-based trigger, as known as Level-0 (L0) trigger, rapidly selects events that have muons or high transverse energy particles from collisions. The events are classified further in the HLT which is a software-based trigger system. It will determine physical signatures of signal events and store them. The HLT is divided into 2 parts denoted as HLT1 and HLT2. The HLT1 reconstructs a partial event where a few tracks are chosen based on their transverse momentum and impact parameter (IP) with respect to the PV. It uses this information to confirm or reject a L0 candidate particle. Since the amount of events passing the L0 and HLT1 are comparable with the amount of events used in an

offline-analysis, the fast full detector reconstruction is performed to select exclusive or inclusive decays. Figure 12 illustrates a flow of an event acceptance beginning from the collisions to event collections in a storage.

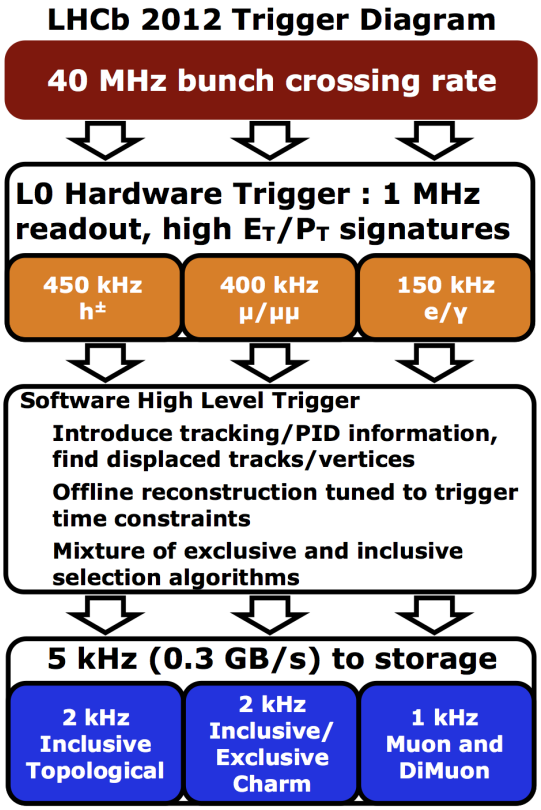


Figure 12: Trigger schemes used in the LHCb detector for data collected in 2011 and 2012 (Run I) [32].

3 Reconstruction of B_s^0 decays with one missing particle

The study of the semileptonic decay poses a challenge due to the presence of the undetectable neutrino. In a hadron collider such as the LHC, a b-hadron tends to have a large Lorentz boost in forward pseudorapidity covered by the LHCb detector. The flight vector of the b-hadron can be constructed using the PV and the SV. This can be exploited as explained in Section 1.2. The momentum of a neutrino produced in association with a lepton and a hadron can be computed with respect to its mother flight vector.

In this section, various methods for the prediction of the b-hadron momentum are presented by exploiting the decay topology. The decay $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$ is used for comparing the methods. We explore and compare analytic methods, *i.e.* corrected mass and proportional mass, as well as multivariate regression methods. These are proposed for the accurate reconstruction of the invariant mass of the $\mu\nu_\mu$ system ($q^2(\mu\nu_\mu)$) as well as background rejection. The sample is acquired from Monte Carlo (MC) simulation and LHCb detector simulation. In the simulation, pp collisions are generated using PYTHIA with the specific LHCb configuration [33, 34]. Decays of particles are calculated using EVTGEN [35]. The detector responses are implemented using GEANT4 [36, 37].

3.1 Preprocessing data

The sample is processed to correspond to the data collected by the LHCb detector. It is required to pass stripping and trigger selections as will be explained in Section 4. Final state particles and the B_s^0, K^{*-}, and K_S^0, are selected further to meet criteria of typical trigger and analysis selections of LHCb, as presented in Table 1.

Particle	Variable	Accepted region	Description
B_s^0	$\lvert\vec{F}\rvert$	> 3 mm	Flight distance of B_s^0 .
μ^+	$\lvert P\rvert$	> 5 GeV/c	Absolute momentum of a muon.
	p_T	$> 1{,}000$ MeV/c	Transverse momentum of a muon.
	η	$1.9 > \eta > 4.9$	Pseudorapidity of a muon.
K_S^0	$\lvert P\rvert$	> 5 GeV/c	Absolute momentum of a K_S^0.
	p_T	> 700 MeV/c	Transverse momentum of a K_S^0.
	η	$1.9 > \eta > 4.9$	Pseudorapidity momentum of a K_S^0.

Table 1: Basic selections for B_s^0, μ^+, and K_S^0.

3.2 Analytic methods

There are several techniques to reconstruct the semileptonic decay of a B_s^0. One of the widely used is known as "corrected mass". This property is a reconstructed mass of the transverse component of the

decay by considering the mass and the transverse momentum of detectable tracks [38]. It is defined as

$$m_{corr} = \sqrt{m_{vis}^2 + (P_{vis}^{\perp})^2} + P_{vis}^{\perp} \tag{15}$$

m_{corr} has been used in many semileptonic decay studies at the LHCb detector.

The B_s^0 momentum (P_{B_s}) can be estimated by using a scale factor with the ratio between the B_s^0 and visible mass as [39]

$$P_{B_s} = \frac{m_{B_s}}{m_{vis}} P_{vis}^{||} \tag{16}$$

Estimation of P_{B_s} by the above equation will be denoted as "proportional" mass or "proportional" momentum depending on the context throughout this report.

Since these techniques are analytic, they are simple to implement. Uncertainties of these are also well-determined. However, the mass resolution is large, leaving a difficulty in background discrimination. Novel techniques using multivariate analysis are explored in the next section to further improve these techniques.

3.3 Multivariate regression

Kinematic variables correlated with the b-meson momentum are implemented to resolve the ambiguity of the quadratic solutions. In the study of [20], information of a flight vector including flight distance $|\vec{F}|$ and flight angle θ_{flight} are used in a simple least squares linear regression algorithm to determine P_{B_s}. They are related through

$$\begin{aligned} P_{B_s} &= \frac{P_{B_s}^{\perp}}{\sin\theta_{flight}} \\ P_{B_s} &= \frac{m_{B_s}|\vec{F}|}{t} \ , \end{aligned}$$

where $P_{B_s}^{\perp}$ and t are the momentum of the B_s^0 perpendicular to the beam axis and decay time, respectively. Figure 13 represents these two variables. $B_s^0 \to K^-\mu\nu_\mu$ is used as a benchmark in this study.

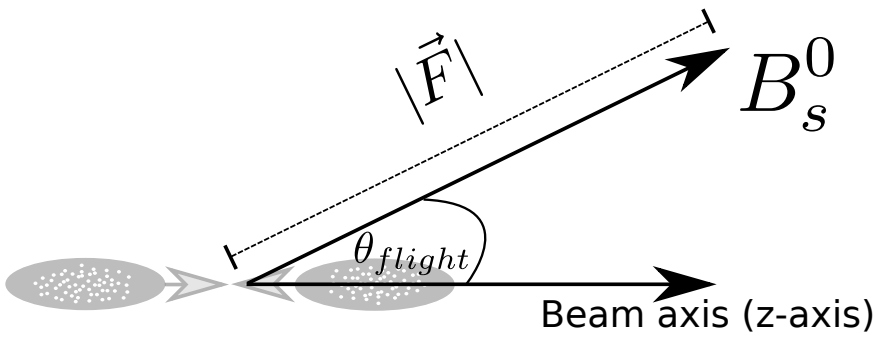


Figure 13: Schematic of determination of flight vector information: flight distance $|\vec{F}|$ and flight angle θ_{flight}.

However, the momentum resolution of the reconstructed P_{B_s} using this method is broad. We apply this method in the $B_s^0 \to K^{*-}\mu^+\nu_\mu$ decay and implement it as a standard to further improve the momentum resolution of the reconstructed P_{B_s}. The package "Scikit-learn" will be used in a similar way as in their study [40].

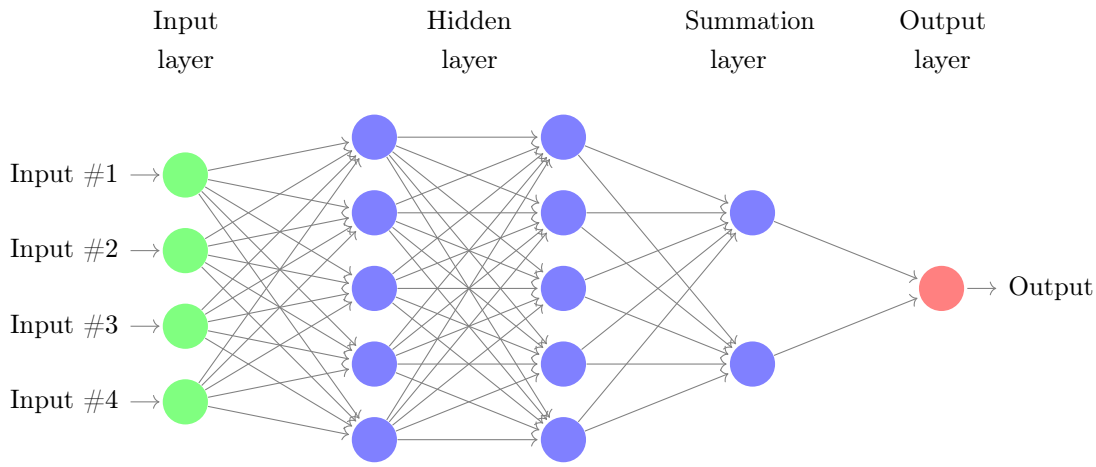


Figure 14: General Regression Neural Network structure.

The Artificial Neural Network (ANN) using the Keras package with the Tensorflow backend [41, 42] is introduced for enhancing the performance of the P_{B_s} prediction. This ANN mimics the function of neurons in a biological brain in which each training node carries an activating function of various strength to train the data. The ANN is composed of input, hidden, and output layers as shown in Figure 14. The input layer takes an array of input variables which are related to the desired value in the output layer. The hidden layers are situated in-between to combine the variables in several ways. Arrows in this figure represent connections from one layer to the next which form a large number of massively interconnected processors (neurons). The network makes a prediction of the output value by adjusting weights on each connection.

Neural networks (NN) used in this study are designed by combining the General Regression Neural Networks (GRNN) structure [43] with the backpropagation algorithm in the Keras package. The GRNN has an additional hidden layer with 2 nodes on top of a general hidden layer before the output layer, as shown in Figure 14. This is known as the summation layer. In the original GRNN, these two nodes are numerator and denominator of a division that provides a highly parallel structure for the output calculation. We adapt this to be corresponded with the quadratic solution parameter $(a, \sqrt{r})$ as in Equation 11 and train the network using the correct solution of P_{B_s} as an output.

The main benefit of the GRNN relative to other techniques is that the network can process without iterations. On the other hand, this requires substantial computation to evaluate the weights on each connection [43]. The backpropagation algorithm is applied to overcome this. It modifies a weight of each connection iteratively depending on an error of predicted and true output values. Then, an error function is defined to guide how the weights will be adjusted. This procedure often leads to optimising this error function using a stochastic gradient decent method. Since there are many input variables in the neural network, the normal method, gradient decent method, to find the optimised point often

converge to a local minima of the error function. The stochastic method takes randomly part of data to compute a gradient and, then, combine with other parts which have different gradients but share the same global minima. In this way, the optimised point of the error function can be achieved.

We intend to resolve the degeneracy of the solutions of the B_s^0 momentum. These solutions may be related to kinematic and topological variables of a b-hadron and its decay products. Therefore, those variables will be input to the neural networks. Both solutions of P_{B_s} are also put in. These might seem to be redundant variables, but it helps calibrating the output value in this study case. The trained events are weighted depending on the uncertainty of the SV to avoid training on high uncertainty events. The error function which is required to be optimised is the "root mean square error (RMSE)":

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}(P_{PRED,i}^2 - P_{TRUE,i}^2)}{N}}, \tag{17}$$

where $P_{PRED,i}$ and $P_{TRUE,i}$ are the B_s^0 predicted and true momentum at event i, respectively. N is the number of events. This function is optimised by using a stochastic gradient descent algorithm with "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION"[44, 45]. Table 2 summarises the input variables as well as the error function to train neural networks.

Feature	Variables	Description
Input	B_s^0 flight	Two variables related to the B_s^0 flight vector, namely $1/\sin\theta_{flight}$ and $\lvert\vec{F}\rvert$.
	Kinematic variables	Perpendicular and parallel momentum components of the B_s^0 decay products - μ and K^{*-}.
	Quadratic solution	P_+ and P_- from Equation 12.
Output	$P_{\pm}^{corrected}$	Corrected solution of P_{B_s} as in Equation 12
Additional	Weight	Inverse of error propagation of SV uncertainties. $((\sigma_x^{SV})^2 + (\sigma_y^{SV})^2 + (\sigma_z^{SV})^2)^{-1}$
	Error function	Root mean square error (RMSE).
	Activation function (Hidden Layer)	tanh function to activate each connection.
	Activation function (Summation Layer)	Linear function to activate each connection.

Table 2: Variables and settings for training neural networks.

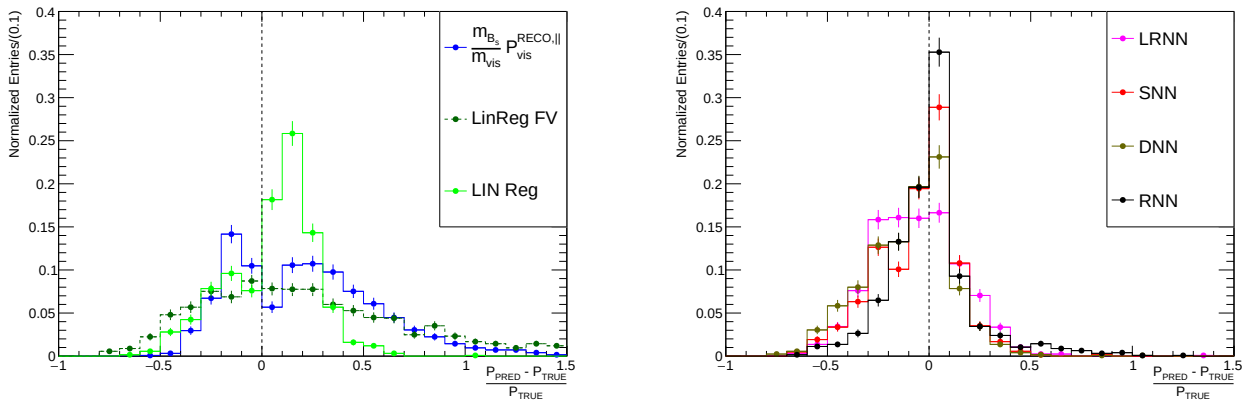
For the number of hidden layers in a neural network, we test on three models. The linear regression neural network is deployed where the input layers are connected directly with the output layer. This is to represent how the neural network can process without the hidden layer. Then, 1 (3) hidden layers are added to the shallow (deep) NN model together with a summation layer. In addition to the feed forward connection line of the NN, a recurrent neural network (RNN) is also introduced. This type of NN has a special hidden layer which can transfer information to the next trained event. Meaning that if an event is trained in this layer, the next event also considers the weight computed from the event before. This results in robustness to predict the value of the output layer. Even if applications of RNN are generally to solve time-dependent problems [46]. We found that this NN is applicable for choosing a corrected solution of B_s^0 as well. All methods to predict P_{B_s} are summarised in Table 3.

Method	Name	Abbreviation	No. Hidden layers
Deterministic	Proportional momentum	$\frac{m_{B_s}}{m_{vis}} P_{vis}^{RECO,\|\|}$	-
	LinReg with only 2 flight vector variables	LinReg FV	-
	LinReg with variables the same as NN	LinReg	-
Neural networks	Linear Regression NN	LRNN	0
	Shallow NN	SNN	1
	Deep NN	DNN	3
	Recurrent NN	RNN	3

Table 3: Method for $P_{B_s^0}$ prediction using proportional mass, linear regression (LinReg), and neural network with different architectures. Each hidden layer is composed of 32 nodes.

Two-third of the $B_s^0 \to K^{*-}\mu^+\nu_\mu$ sample ($\sim$ 2420 events) is used for training the models. They are required to pass the basic selections as shown in Table 11 in Appendix C. The rest of the events ($\sim$ 1250 events) are for controlling model overfitting. Note that the events used in this section have different stripping selections as will be explained in Section 4.1.

Overfitting is avoided by deploying regularisation methods. In each iteration of neural network trainings, the weights on each connection are updated by optimising the error function together with the sum of the square of the weights (L_2 regularisation). The later is added to penalise model complexity for not perfectly fitted to the training data [47]. Another regularisation is called Dropout regularisation. Since the nodes in each hidden layer may represent similar features, the nodes are randomly dropped out in each iteration [48]. We train the networks until the error functions are unchanged and compatible with the testing set. From this point, the testing set will be implemented to evaluate between methods.



(a) Deterministic methods

(b) Neural networks

Figure 15: Residual distribution of predicted B_s^0 momentum from various methods.

Figure 15 represents residual distributions of the predicted values P_{PRED} of each model and the true value P_{TRUE} in the simulation. The proportional momentum and linear regressions methods have

larger widths than the neural networks models. This implies that the neural networks provide the predicted $P_{B_s^0}$ more precisely. The RNN model has the best resolution among other neural network models. We present the distribution of the B_s^0 momentum using this model (RNN) and the true value in Figure 16. The correlation coefficient can reach upto almost 90% whereas around 70% was achieved by [20].

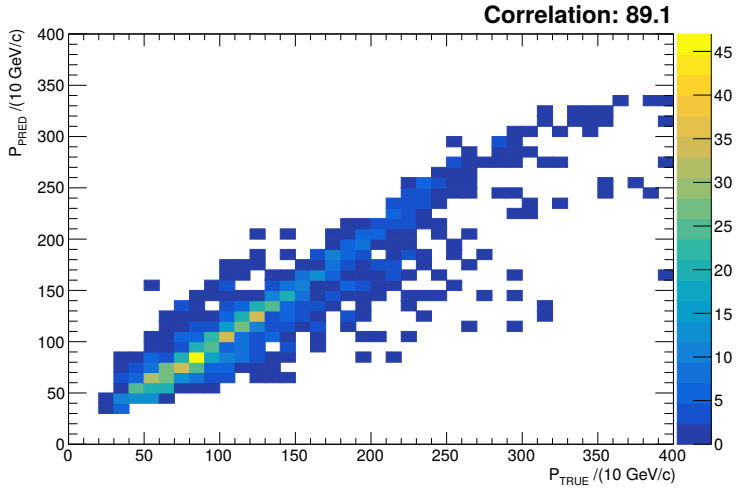


Figure 16: Distribution of predicted B_s^0 momentum using recurrent neural networks (RNN) P_{PRED} versus the true momentum at a generator level.

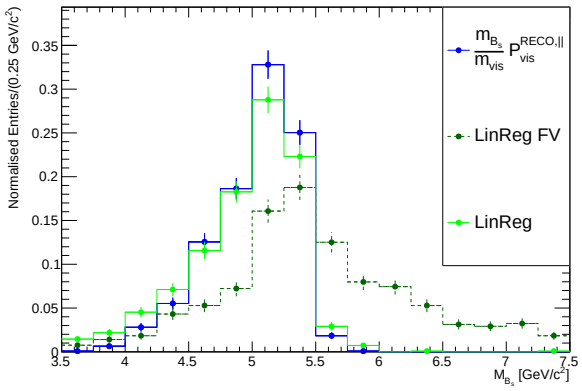
3.4 Application to study the semileptonic decy

We show the B_s^0 invariant mass as well as missing mass square M^2_{miss} of the system. The M^2_{miss} is an alternative quantity to specify undetected particles' masses which is supposed to be near zero because of the massless neutrino. It is defined as

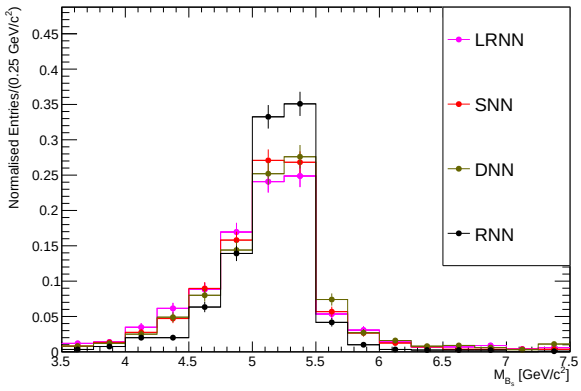
$$M^2_{miss} = \left(\sqrt{P^2_{B_s} + m^2_{B_s}} - E_{vis}\right)^2 - (P_{B_s^0} - P^{||}_{vis})^2 - (P^{\perp}_{vis})^2, \tag{18}$$

where P_{B_s} is acquired depending on the specific methods used. Figure 17 shows that the reconstructed quantities with the linear regression by using only the flight vector has a larger width than the other methods. By comparing the deterministic methods and neural networks, the latter provides distributions that are closer to expected values than the former; mass distributions near the B_s^0 mass and missing mass square near zero. The RNN is the best method to reconstruct both mass distribution and missing mass square.

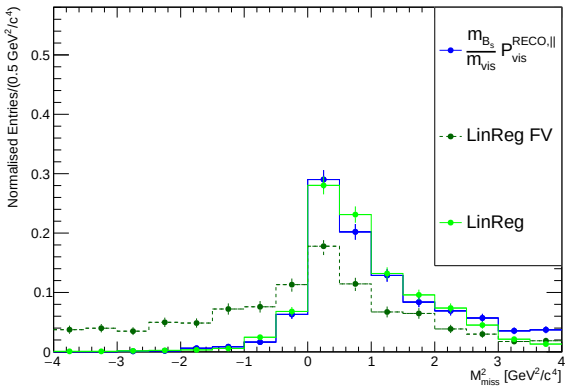
These models are also tested on how well they predict the correct P_{B_s} solution as shown in Figure 18 as a function of η and q^2 (the true value). The proportional mass is clearly separated from the others by having the lowest fraction of the correct solution. The average fractions of neural networks linear regression methods have similar trends and are distinct from the proportional mass method. The ratios of correct predictions from the neural networks increase as a function of $q^2(\mu\nu_\mu)$ and fluctuate at high q^2 value, but this is not the case for the RNN method which still rises and can reach up to around 80% in $14 < q^2 < 20$ GeV^2/c^4.



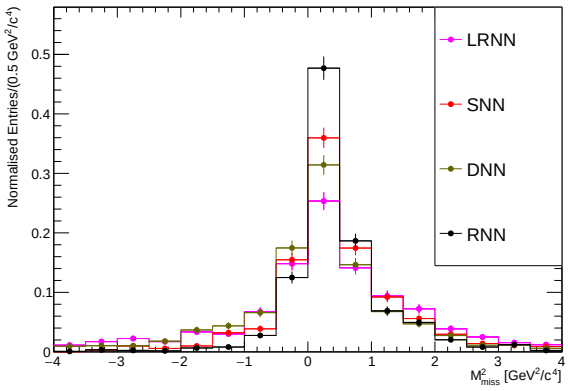
(a) Reconstructed B_s^0 mass using deterministic methods



(b) Reconstructed B_s^0 mass using neural networks

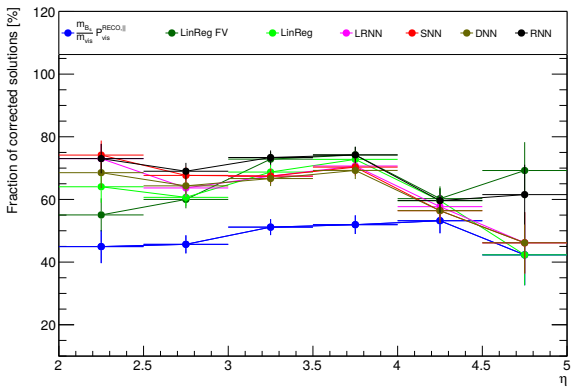


(c) Reconstructed B_s^0 missing mass square using deterministic methods

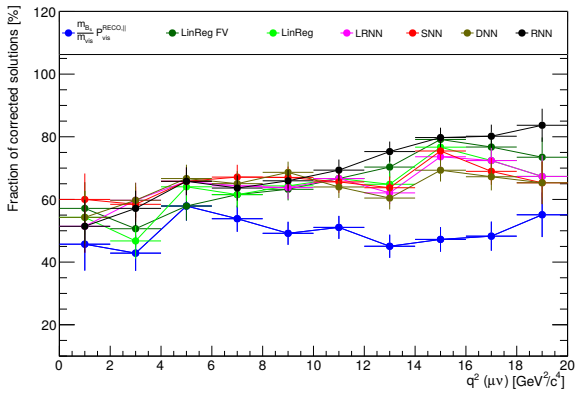


(d) Reconstructed B_s^0 missing mass square using neural networks

Figure 17: Distribution of reconstructed B_s^0 mass and missing mass square with various methods.



(a) Pseudorapidity η



(b) $q^2(\mu\nu_\mu)$

Figure 18: Fraction of which the correct B_s^0 momentum solution is chosen as a function of pseudorapidity and invariant mass of $\mu\nu_\mu$ system.

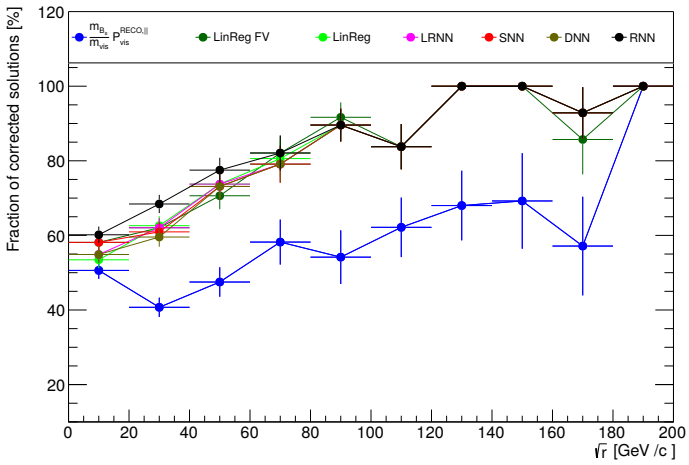


Figure 19: Fraction of which the correct B_s^0 momentum solution is chosen as a function of a half of difference between two solutions of P_{B_s} ($\sqrt{r}$).

Furthermore, the dependence of the ratio on half of difference between the two solutions of P_{B_s} ($\sqrt{r}$) is investigated as shown in Figure 19. The proportional mass method is still isolated, while the others have the same trend. They can go over 80% when $\sqrt{r} > 60$ GeV. This implies that these regression methods become more distinct when both solutions are well separated. The study of the resolution of q^2 as a function of $\sqrt{r}$ should be mentioned for the accurate prediction because if the difference between these two solutions ($2\sqrt{r}$) is smaller than the resolution of q^2, either of solution can be chosen.

Figure 20 illustrates the correlation of the $q^2(\mu\nu_\mu)$ between the true value and the value reconstructed using P_{B_s} from the RNN method. This RNN can reconstruct the $q^2(\mu\nu_\mu)$ correlated with the true value at over 90%. This is highly-correlated as expected from the precise prediction of P_{B_s} using this method.

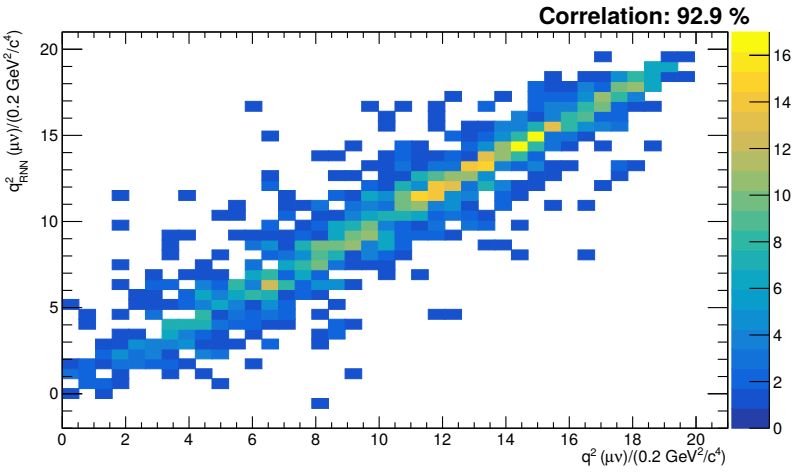


Figure 20: Correlation between invariant mass square of $\mu\nu$ system (q^2) between the predicted value using the RNN and the true value.

4 Analysis of $B_s^0 \to K^{*-}\mu^+\nu_\mu$ on the LHCb Run I data at 8 TeV

In this part, the search for $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$ is performed on the data collected by the LHCb detector in 2012 from proton-proton collisions at a centre-of-mass energy at 8 TeV, corresponding to a luminosity of 2.08 fb^{-1}. The reduction of background events is described. The analysis emphasises on optimising signal events distributed in the mass distribution of the K_S^0 and the K^{*-}.

Preliminary selections are studied to optimise signal over background events. Stripping and trigger selections are implemented to select events that have a topology specific to this decay chain. Background events are vetoed by implementing a selection of trigger lines as well as a multivariate analysis for classification. When a Monte Carlo sample is used, it must be scaled to correspond to the luminosity in data. This computation and referenced quantities are described in Appendix A.1. Note that since this is the first search for this decay channel, a qualitative analysis is performed to explore an efficient selection of the decay.

4.1 Stripping selection

The stripping line **StrippingB2XuMuNuBs2KstarLine** is used for selecting candidate events from the LHCb data [49]. This stripping line extracts the events with a topology typical to the decay $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$. The selection is shown in Table 4 and Table 5.

	Cut type	Accepted region	Description
B_s^0	$\chi^2_{vtx}/ndof$	< 4	Quality of the vertex fit.
	$\cos(DIRA)$	> 0.99	Cosine of the angle between the particle's momentum and the vector pointing from the primary vertex to the secondary vertex.
	M_{corr}	> 2500 MeV/c < 7000 MeV/c	Corrected mass of B_s.
μ	p_T	> 1000 MeV/c	Transverse momentum.
	P	> 3000 MeV/c	Absolute momentum.
	Ghost. Prob.	< 0.35	Probability to be a ghost track; a track not corresponding to a real particle.
	χ^2_{IP}	> 12	Quality difference between primary vertex fits when the μ candidate is added.

Table 4: Stripping requirements **StrippingB2XuMuNuBs2KstarLine** for the candidate $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$.

Note that the variables shown in red colour in Table 5 cause the exclusion of about half of the signal events as seen in the simulation sample [1]. Because if the $K_S^0 \to \pi^+\pi^-$ is reconstructed from

[1] This is a faulty implementation of the K_S^0 selection in the stripping selections.

downstream tracks, the χ^2_{IP} and FD χ^2 are small due to the large uncertainty of the track extrapolation to PV. These selections were not applied in Section 3 because a substantial amount of events is needed to train the neural networks. Then, we apply these as shown as "Imp. Stripping [MC]" in Appendix C to make the simulation corresponding with the data.

	Cut type	Accepted region	Description
$(\pi^-\|K^{*-})$	ADAMASS	< 200 MeV/c	Acceptance region of reconstructed K^{*-} mass ($M_{K^{*-}} \in [m_{K^{*-}} - 200, m_{K^{*-}} + 200]$ MeV/c^2). The $M_{K^{*-}}$ and $m_{K^{*-}}$ are reconstructed and mean K^{*-} mass from [4], respectively.
	p_T	> 100 MeV/c	Transverse momentum.
	P	> 2000 MeV/c	Absolute momentum.
	χ^2_{IP}	> 9	Quality difference between primary vertex fits when the π^- from K^{*-} candidate is added.
	PIDK	< 10	Particle identification likelihood to be a kaon.
$(K^0_S\|K^{*-})$	p_T	> 700 MeV/c	Transverse momentum.
	FD	< 20 mm	Flight distance.
	Mass	> 456 MeV/c < 536 MeV/c	Reconstructed mass.
	FD χ^2	> 100	Significance of K^0_S's flight distance.
	$\chi^2_{vtx}/ndof$	< 10	Quality of the vertex fit.
	χ^2_{IP}	> 8	Quality difference between primary vertex fits when the K^0_S candidate is added.
$(\pi^-\|K^0_S)$	p_T	> 250 MeV/c	Transverse momentum.
	p	> 2000 MeV/c	Absolute momentum.
	χ^2_{IP}	> 50	Quality difference between primary vertex fits when the π^- from K^0_S candidate is added.

Table 5: Stripping requirements of **StrippingB2XuMuNuBs2KstarLine** for the candidate $B^0_s \to K^{*-}(\to K^0_S\pi^-)\mu^+\nu_\mu$. The convention $(A|B)$ represents a daughter particle A decaying from B, *i.e.* $(\pi^-|K^{*-})$ implies a π^- that comes from a resonance K^{*-}.

4.2 Trigger selection

In the LHCb experiment, the TISTOS method is used to classify the candidates. In this method, events are classified depending on their information in the trigger system as Triggered On Signal (TOS), Triggered Independent of Signal (TIS), and Triggered On Both (TOB) [50]. A particle candidate is TOS when a corresponding trigger line is fired. If the event is still triggered when the candidate is removed, the particle candidate is considered to be TIS of that trigger selection. Because there might be other particles in the event which can be accepted by that trigger [51]. Since there are some events in which a particle candidate can be both TIS and TOS, we classify them as TOB. To acquire candidates for $B^0_s \to K^{*-}(\to K^0_s\pi^-)\mu^+\nu_\nu$, we apply preliminary selections to extract events with a muon using L0

and HLT1 triggers.

Trigger Level	Trigger line	Description
L0	Muon Decision TOS on B_s^0	Trigger on muon
HLT1	Track Muon Decision TOS on B_s^0	Multivariate analysis on one track of a particle

Table 6: Requirements for the trigger selection.

4.3 Background optimisation

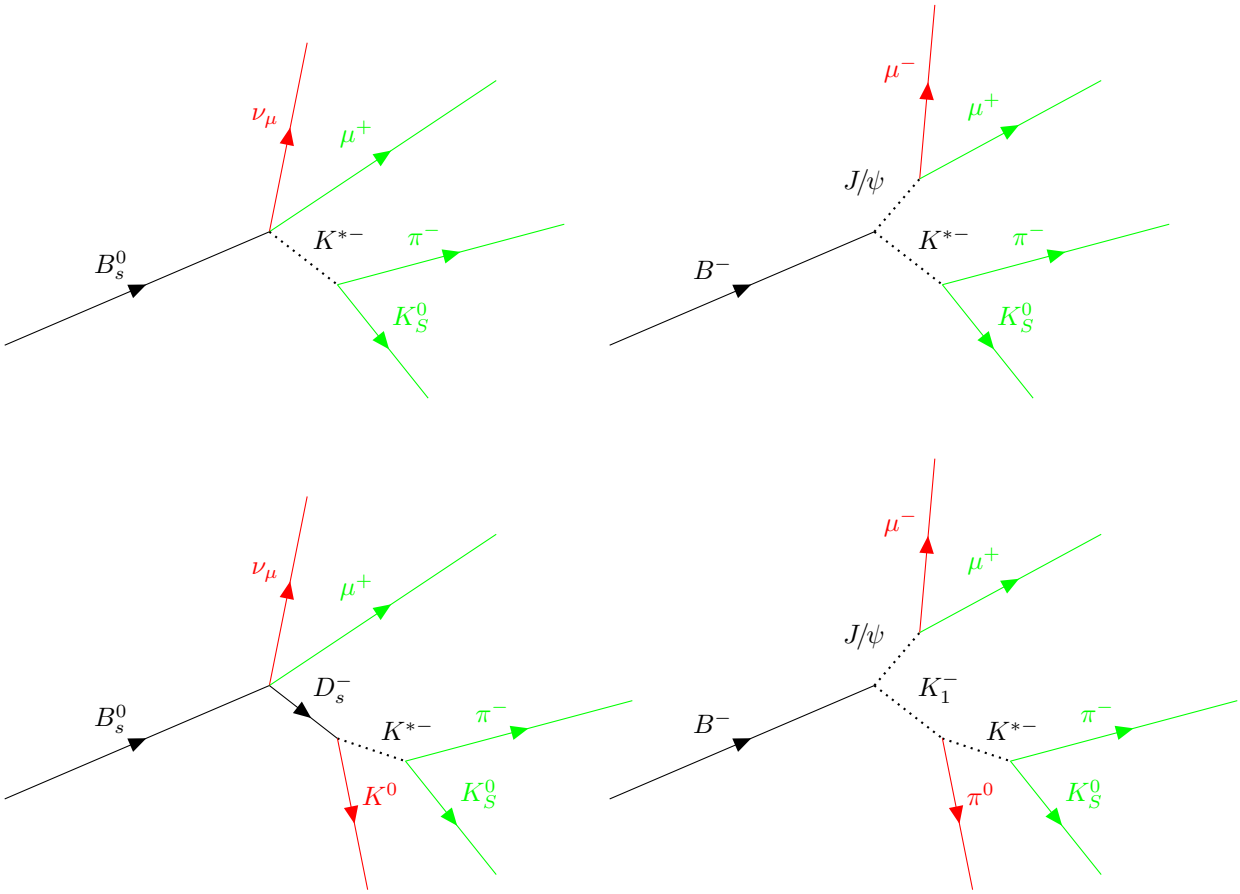


Figure 21: Schematic decay of signal $B_s^0 \to K^{*-}\mu^+\nu_\mu$ and background $B_s^0 \to D_s^-(\to K^{*-}K^0)\mu^+\nu_\mu$ $B^- \to K^{*-}J/\psi^0(\to \mu^+\mu^-)$, $B^- \to K_1^-(\to K^{*-}\pi^0)J/\psi^0(\to \mu^+\mu^-)$. The decay of $K^{*-} \to K_S^0\pi^-$ is the same for signal and all backgrounds. Colour indications are similar to Figure 5. The dotted line represents a prompt decaying particle whose track is not detected in the detector. This figure is just for illustration.

Physical background events, which could mimic the signal's topology, include

- $B_s^0 \to D_s^+(\to K^{*-}K^0)\mu^+\nu_\mu$
- $B^- \to K^{*-}J/\psi(\to \mu^+\mu^-)$
- $B^- \to K_1^-(\to K^{*-}\pi^0)J/\psi(\to \mu^+\mu^-)$

We show the signal and background topologies in Figure 21. The K^{*-}, J/ψ, and K_1^- are prompt decaying particles. The intermediate state J/ψ can be vetoed by excluding events with the HLT2 trigger line "HLT2DiMuonDetachedJPsi". This trigger is on when two muons found in the detector have an invariant mass in the J/ψ mass window. In addition, invariant mass of μ and K^{*-} together with a particle track that is compatible with the fitted SV of B_s^0 (SmallestDeltaChi2MassOneTrack) is computed by assuming a mass of that track to be a pion. We exclude events for this variable in a range from 5200 MeV/c^2 to 5400 MeV/c^2. This significantly reduces background with $J/\psi \to \mu^+\mu^-$ when one of these muons is unreconstructed. This can be seen in simulation between this the signal and background from $B^- \to K^{*-}J/\psi(\to \mu^+\mu^-)$.

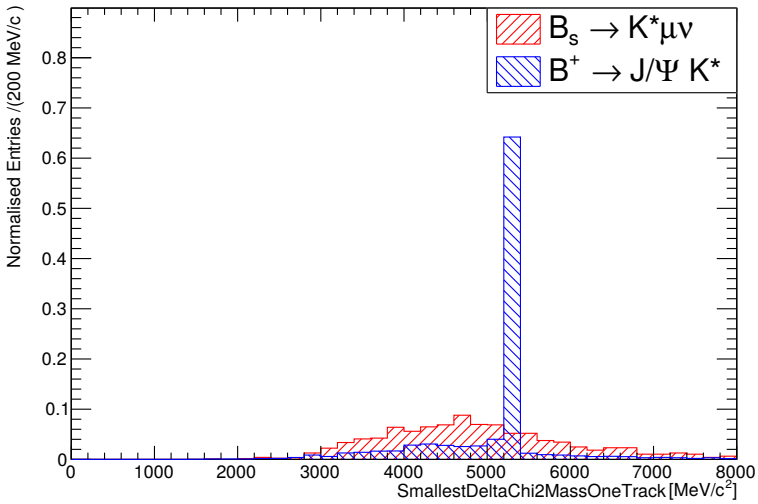


Figure 22: SmallestDeltaChi2MassOneTrack mass distribution of signal and $B^- \to K^{*-}J/\psi(\to \mu^+\mu^-)$ background from Monte Carlo sample

Particles should be selected strictly depending on the probability to be a specific type. A cut on a probability of the particle identification variable $ProbNN$ is applied. This $ProbNN_X$ is the probability of a detected track to be a particle X obtained from neural networks using information from various parts in the LHCb detector [25]. The $ProbNN$ is preliminarily required to be larger than 0.2 for μ, $(\pi^-|K^{*-})$, and $(\pi^-|K_S^0)$.

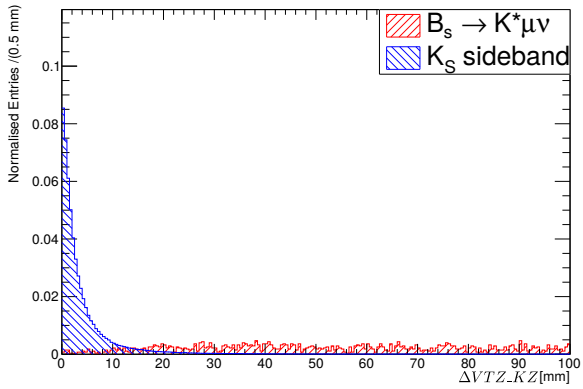
Moreover, a preliminary cut on the probability of a track to be a ghost track with $ProbNNghost$ > 0.2 on all particles' tracks, except two pions from the decay $K_S^0 \to \pi^-\pi^+$, is applied. The ghost track is defined as the track not corresponding to a real particle. To be able to implement the regression methods explained in the previous section, the kinematic variables in the events must imply a physical solution of $P_{B_s^0}$, *i.e.* $r > 0$ as in Equation 11.

K^0_S should propagate through the detector with $c\tau = 2.6844\,\text{cm}$ where c is the speed of light and τ is a lifetime of K^0_S in its restframe [4]. Its lifetime in the lab frame is longer than this $c\tau$ and its propagation should be detected and isolated from other combinatorial background in the reconstructed K^0_S mass distribution. The event candidates are optimised using a difference between decay vertices of K^{*-} and K^0_S in the proton beam axis (z-axis), namely ΔVTX_KZ. The optimisation is done using the signal MC sample scaled to be corresponding to the luminosity of 2.08 fb^{-1} as of the LHCb data in 2012. Events which have a $\pi^-\pi^+$ invariant mass of [465, 480] and [515, 530] MeV/c^2 are obtained to represent the background distribution. The cut is chosen by optimising the significance (Signi).

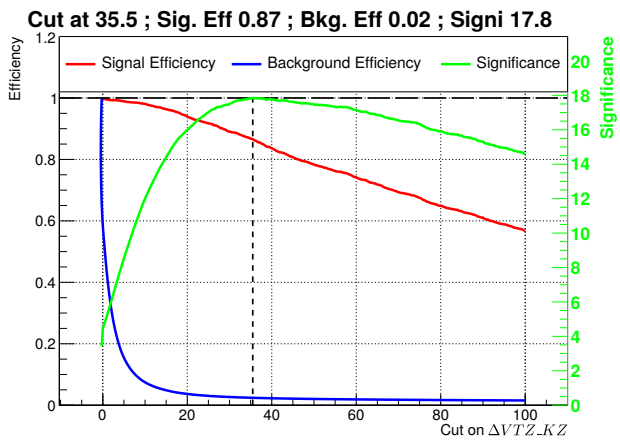
Signal efficiency ϵ_s and background efficiency ϵ_B are defined as

$$\begin{aligned} \text{Signi} &= \frac{S_i}{\sqrt{S_i + B_i}} \\ \epsilon_s &= \frac{S_i}{S_0} \\ \epsilon_B &= \frac{B_i}{B_0} \end{aligned} \tag{19}$$

where S_0, B_0 are total amount of signal and background yields before this cut. S_i and B_i are the amount of background and signal where the cut is applied at a point i. Figure 23 represents the distribution of this variable between signal and background together with the significance optimisation. The optimal point is found to be at $\Delta VTX\_KZ = 3.55$ cm where about the 87% of signal is selected and almost all of the background is excluded. The distribution of the $\pi^+\pi^-$ invariant mass before and after this selection is shown in Figure 24.

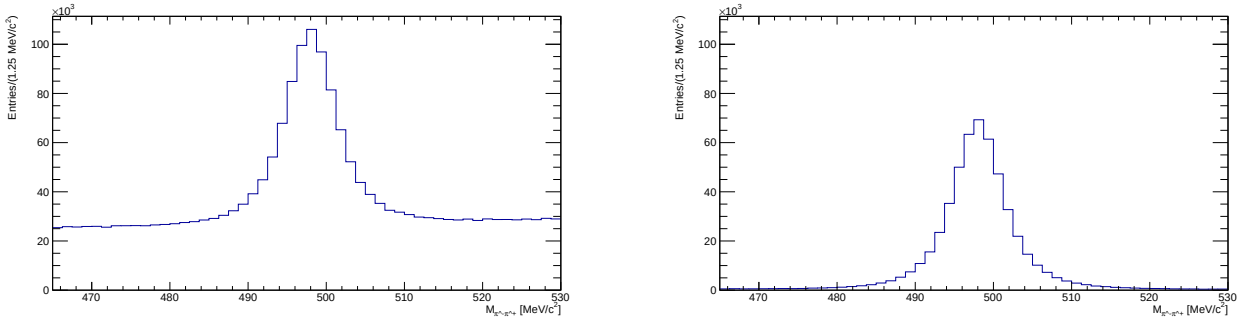


(a) Distribution of difference between decay vertices of K^{*-} and K^0_s



(b) Significance optimisation

Figure 23: Optimisation of a cut on the difference between z-coordinate of the decay vertices of K^{*-} and K^0_s (ΔVTX_KZ). The red distribution in (a) is obtained from MC sample of the signal event. The blue distribution is extracted from the reconstructed K_S mass distribution $m_{K_s} \in$ [465, 480] and [515, 530] GeV/c^2 (K_S sideband).



(a) Before the selection

(b) After the selection

Figure 24: Invariant mass of $\pi^+\pi^-$ distribution.

4.4 Gradient Boosting Decision Tree

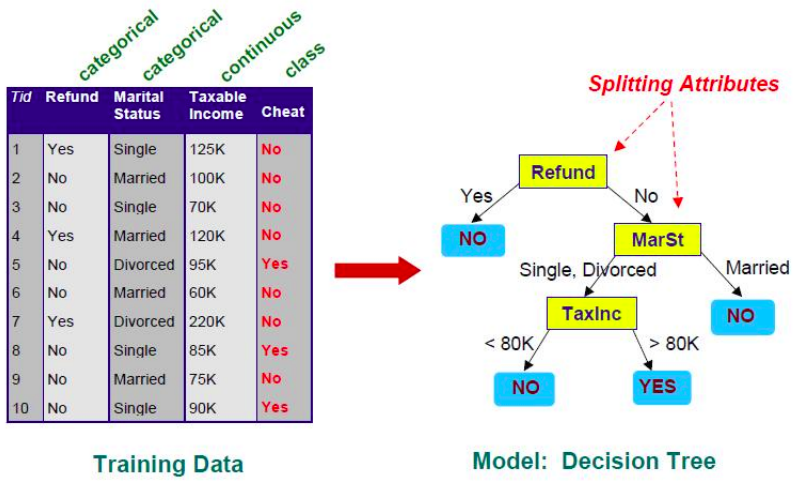


Figure 25: An example decision tree diagram from which data is projected into a flowchart-like diagram [52].

A decision tree classifier is an non-linear algorithm to distinguish between different data. It uses a flowchart-like diagram of decisions with possibilities of outcome. As an example in Figure 25, it contains several test nodes on input variables. However, the classifier alone is not powerful enough to distinguish data because it randomly defines rules on each variables without considering how much weight should be put on each test node [53]. To promote this weak classifier, the boosting algorithm is taken into account. This algorithm is different from the adaptive boosting algorithm which is used in most of LHCb analysis on how it deals with the weak learner. The adaptive boosting takes a base classifier to find prediction errors of that classifier and the algorithm will improve the algorithm based on those errors. The improvement is iterated with modification of a weight of incorrect prediction until higher accuracy is achieved [54]. The combination of this algorithm with the tree diagram is called "Boosted Decision Tree (BDT)".

The gradient boosting algorithm considers remaining errors of the sample and build up a second classifier in such a way to minimize the overall error of the classifier [55]. This is the same procedure as the gradient decent method for specifying an minimum point of the error function explained in Section 3. The combination of this algorithm with the decision tree is called "Gradient Boosting Decision tree (GBDT)".

We implement the GBDT classifier using the extreme gradient boosting decision tree (XGBDT) in the "XGBoost" package [56]. The usage of those boosting algorithms with the decision tree often leads to the overtraining. This is expected to be under control in this package because it has additional options to perform the L_2 regularisation to the tree model.

This XGBDT is used to classify signal event from combinatorial background distributed in the invariant mass distribution of $K_S^0\pi^-$ ($m_{K_S^0\pi^-}$). The background sample is obtained from events which have mass $m_{K_S^0\pi^-} \in [690, 800] \cup [1000, 1100]$ MeV/c^2, denoted as K^{*-} sideband background. Prior to the classification, events with visible mass lower than 2500 MeV/c^2 are excluded to reduce number of background events. This does not affect much the amount of signal as shown in Figure 26. We also present the $m_{K_S^0\pi^-}$ distribution for both signal from MC sample and background in Figure 27.

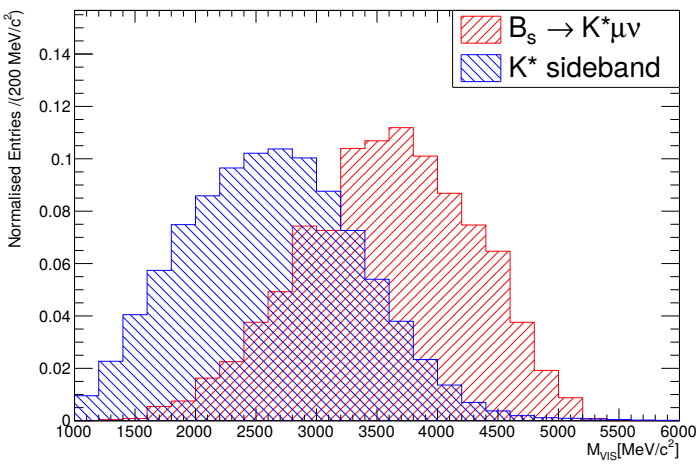


Figure 26: Visible mass distribution of signal events from Monte Carlo sample and background sample taken from events which have invariant mass of $K_s^0\pi^-$ between [690, 800] and [1000, 1100] MeV/c^2.

The kinematic variables, which characterise the topology of the events that will be employed in the multivariate analysis to discriminate between signal and background, are chosen and described in Table 7 for B_s^0 candidates, K^{*-}, and μ in each event. We also present the distributions of these variables from all data in Figure 29.

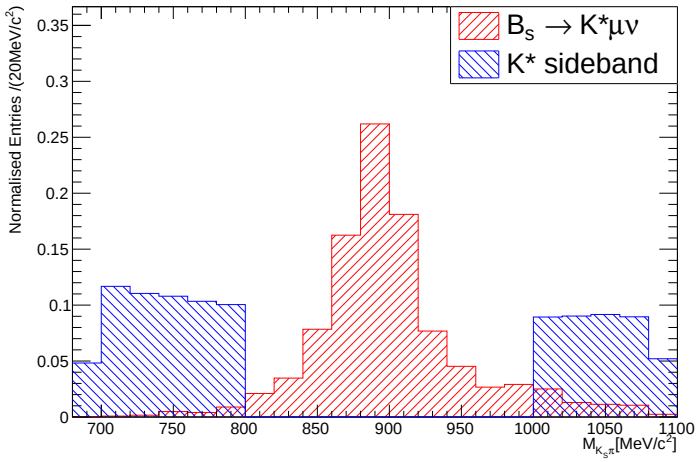


Figure 27: $M_{K_s^0\pi^-}$ distribution of signal events from the Monte Carlo sample and the background sample taken from events which have invariant mass of $K_s^0\pi^-$ between [690, 800] and [1000, 1100] MeV/c^2.

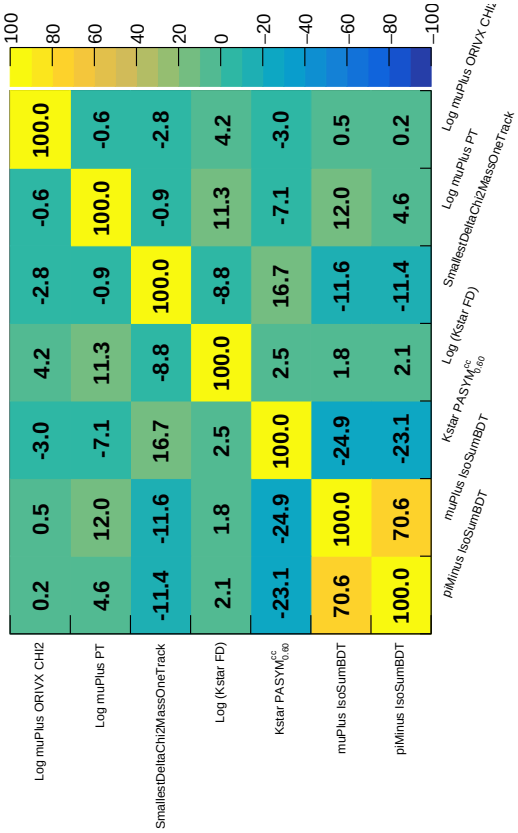
Particle	Variable	Description
B_s^0	Max(BDT$_\mu^{Iso}$ and BDT$_{\pi^-}^{Iso}$)	Maximum values between BDT$_\mu^{Iso}$ and BDT$_{\pi^-}^{Iso}$. The BDT$_X^{Iso}$ is a decision from the BDT training against background with a particle X isolated from the signal decay chain.
K^{*-}	Kstar PTASYM$_{0.60}^{nc}$	$\frac{p_{T,K^{*-}}-\sum p_T}{p_{T,K^{*-}}+\sum p_T}$ where $p_{T,K^{*-}}$ is the transverse momentum of K^{*-} and $\sum p_T$ is the sum of the transverse momentum of all charged particle track in a cone around K^{*-} candidate with a radius $R = \sqrt{(\Delta\phi)^2 + (\eta)^2} = 0.60$ (η is pseudorapidity and ϕ is azimuthal angle.)
	Log(Kstar FD)	Flight distance of K^{*-}.
	SmallestDeltaChi2MassOneTrack	Invariant mass of a reconstructed K^{*-} and an additional track which has the smallest difference between χ^2 of vertex fit with the K^{*-} track.
μ	muPlus PT	Transverse momentum of a muon.
	Log muPlus ORIVX CHI2	Quality of an original vertex of a muon. Note that this is the same vertex as the B_s^0 decay vertex.

Table 7: Discriminating variables used in the XGBDT.

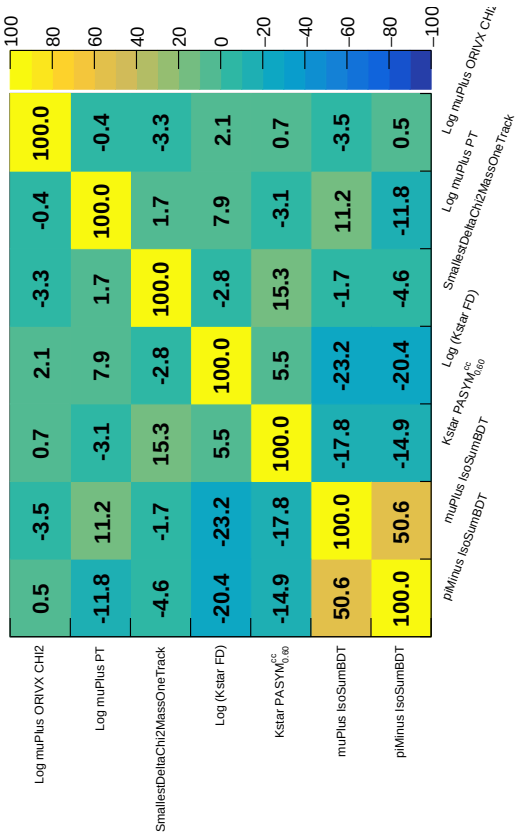
To minimize the bias of choosing specific variables for the multivariate analysis, the variables should be independent from each other. The discriminating variables are checked for their correlations using the Pearson's correlation coefficient which is explained in Appendix A.2. The results of all events

are presented in percent in Figure 28 for both signal and background.

Probabilities for a particle track to be isolated from other tracks in an event are denoted as BDT_{μ}^{Iso} and $\mathrm{BDT}_{\pi^-}^{Iso}$ for μ and π^-. These variables are from training a classifier using a BDT to isolate against background which have additional particle apart from the signal decay chain. The BDT_{μ}^{Iso} and $\mathrm{BDT}_{\pi^-}^{Iso}$ are correlated with around 70% and 50% for signal and background, respectively. So higher value of these variables is chosen to be put in the XGBDT. We found no significant correlation between each pair of the other discriminating variables.



(a) Background



(b) Signal

Figure 28: Correlation matrices between the variables described in Table 7 of all data: (a) Signal events from Monte Carlo sample, (b) background sample taken from events which have invariant mass of $K_s^0\pi^-$ between [690, 800] and [1000, 1100] MeV/c^2 as shown in Figure 27.

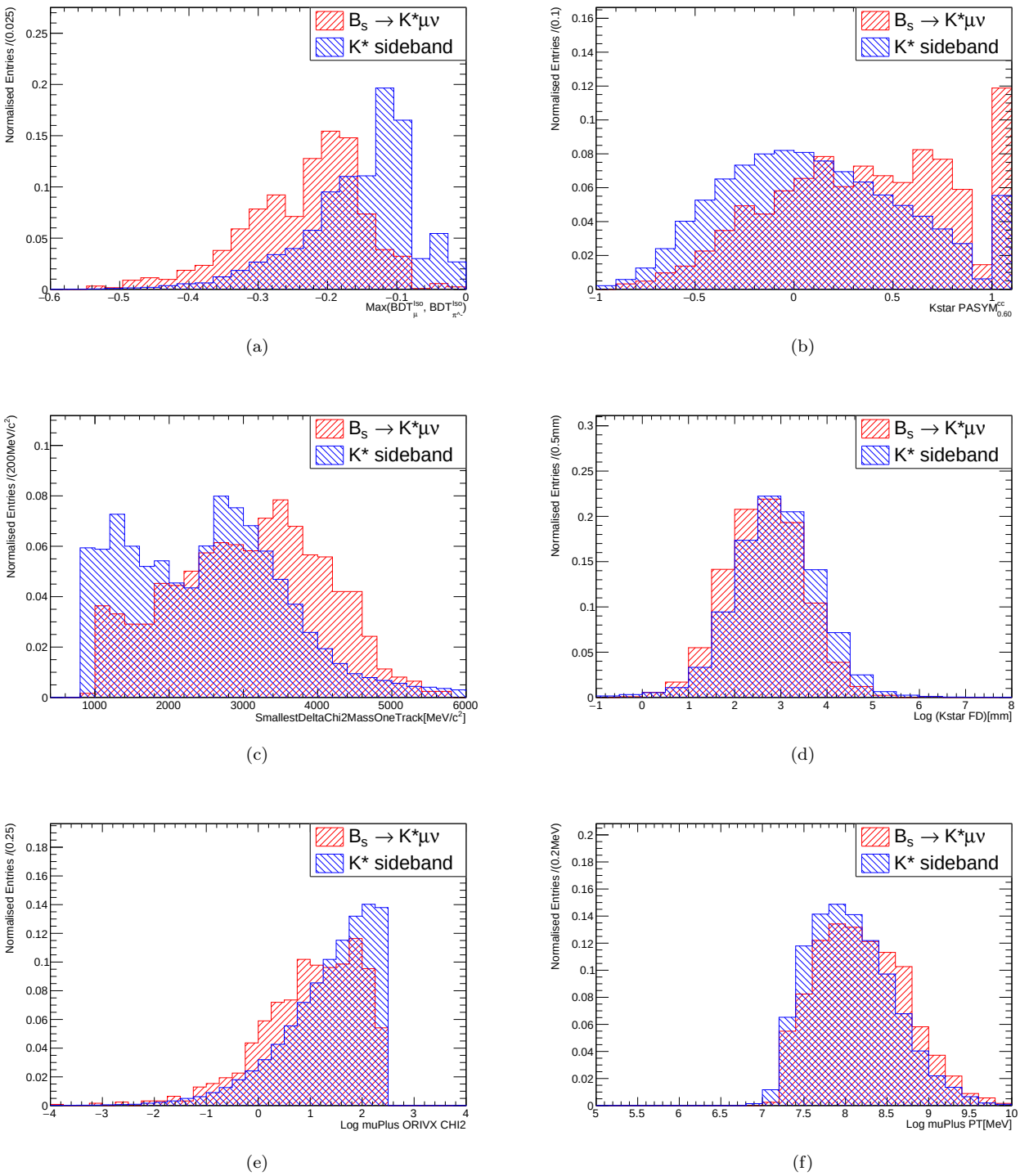


Figure 29: Distribution of the variables described in Table 7

Before the XGBDT is trained on the sample, the discriminating variables are decorrelated using the Principle Component Analysis (PCA) method. This method will transform the variables by multiplying them with the inverse of the square root of its covariance matrix. The related covariance matrix is as well diagonalised, meaning that each pair of variables is uncorrelated. This transformation

also enhances the performance of the BDT in distinguishing between signal and background events.

The XGBDT is trained on two-third of the sample after the cut on the visible mass and, then, tested on the rest. The relevance parameter, which ranks the discriminating variables depending on their discriminating power, is extracted and presented in Table 8. This implies that the two variables in the μ-track are the two most important parameters contributing in the classifier.

Particle	Variable	Relevance
B_s^0	Max(BDT$_\mu^{Iso}$ and BDT$_{\pi^-}^{Iso}$)	0.2068
K^{*-}	Kstar PTASYM$_{0.60}^{nc}$	0.1780
	Log(Kstar FD)	0.0798
	SmallestDeltaChi2MassOneTrack	0.0415
μ	muPlus PT	0.2073
	Log muPlus ORIVX CHI2	0.2866

Table 8: Relevance of the variables used in the XGBDT classifier. The higher relevance of a variable, the more important that variable is in the classifier.

The classifier is evaluated using the testing sample. The area under the receiver operating characteristic curve, known as AUROC, is found to be 0.816 in the testing sample as shown in Figure 30. The AUROC is an area under the signal acceptance rate as a function of background rejection for different cut-off points. A classifier with no overlap in the signal and background distribution has the AUROC equals to 1. This tells how good the classifier is. In this case, the classifier has a good performance for specifying the samples.

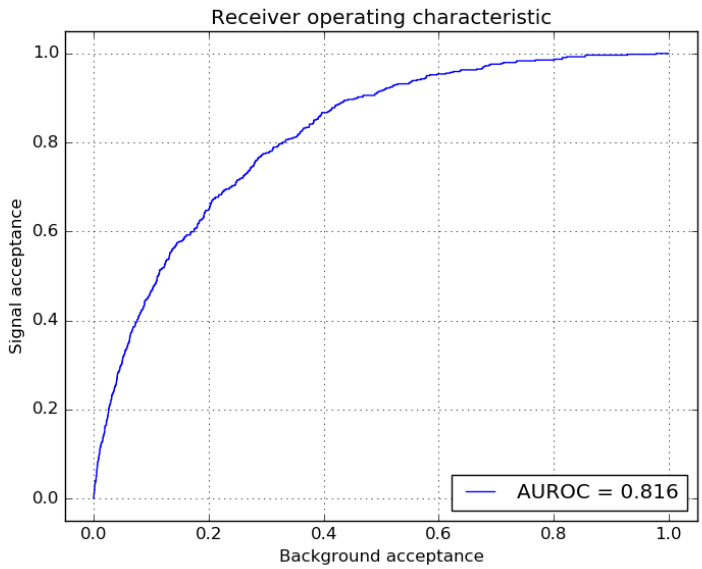
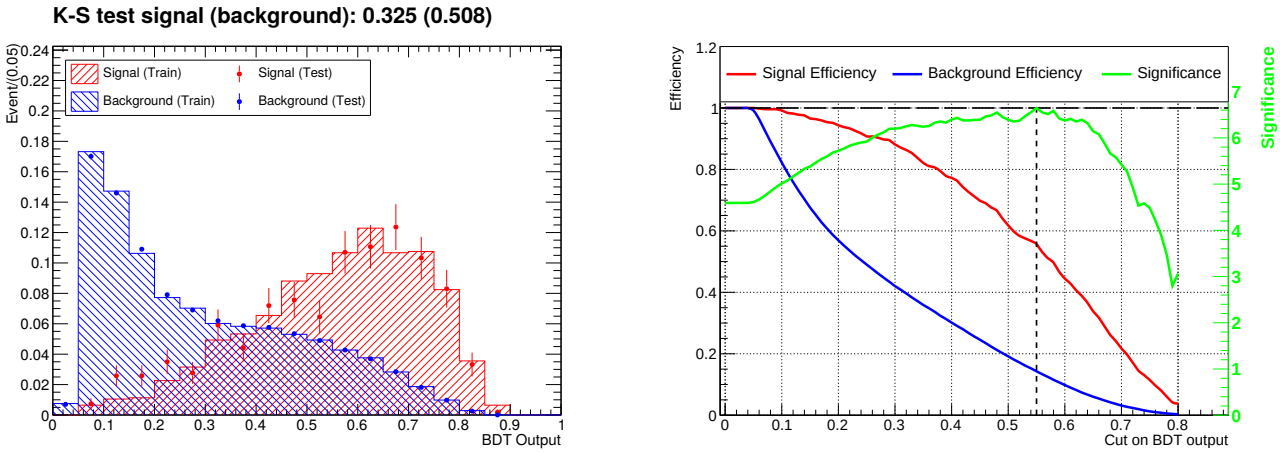


Figure 30: Receiver Operating Characteristic (ROC) curve of the testing set.

The precision, which is defined as a fraction between correct prediction and amount of data, is found to be 78% in determining the background distribution and 70% for the signal distribution in the testing sample. The precisions are also tested on the training sample. They are found to be the same as from the testing sample. This ensures that the classifier is not overtrained. To further search for a sign of overtraining, we use the Kolmogorov-Smirnov (K-S) test in the signal and background distributions of the output from the XGBDT classifier between training and testing samples as shown in Figure 31 (a). The test is a non-parametric test that compares 2 samples to investigate whether they come from the same distribution. We found no significant sign of overtraining.



(a) BDT output from the training and testing samples.

(b) Significance optimisation.

Figure 31: Optimisation of the BDT cut.

The cut on this XGBDT output is chosen by optimising the significance using Equation 19. We use the same definition of signal, scaled to the 2012 LHCb data luminosity, to represent S_i. For B_i, it is extracted from a fit of the $m_{K_S^0\pi^-}$ mass distribution within the range $m_{KS\pi^-} \in [800, 1000]$ MeV/c^2 which is the signal region. The K^* mass distribution can be described by the Breit-Wigner distribution fitted with an unbinned maximum likelihood. The parameters of this distribution are obtained and fixed from the signal Monte Carlo sample. The distribution is shown in Figure 32. The background is described by an exponential distribution. The B_i is then the number of background events in the K^* mass fit at the XGBDT cut point i. The amount of signal S_i does not come from the fit because the K^* peak may be composed of other physical backgrounds than signal which need to be vetoed. The signal S_i is, then, estimated using the Monte Carlo sample.

An optimal point is obtained with the BDT output cut at > 0.55 as shown in Figure 31 (b). The mass distribution (M_{K*}) at this point in data is presented in Figure 33. Yields are extracted with 2422 ± 173 of the K^{*-} candidates and 17919 ± 213 of the background.

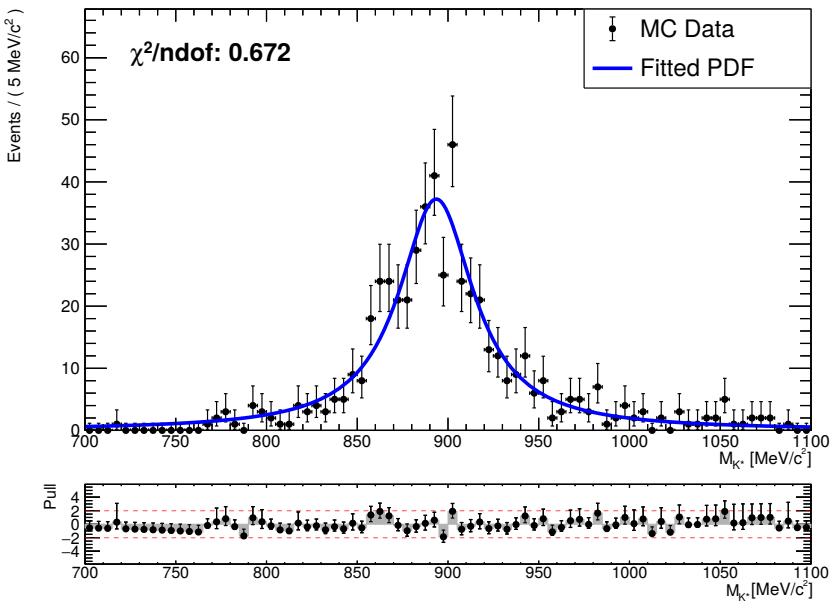


Figure 32: Mass distribution fitting (M_{K*}) of $K^{*-} \rightarrow K_s^0\pi^-$ from MC data with Breit-Wigner distribution function along with the pull plot.

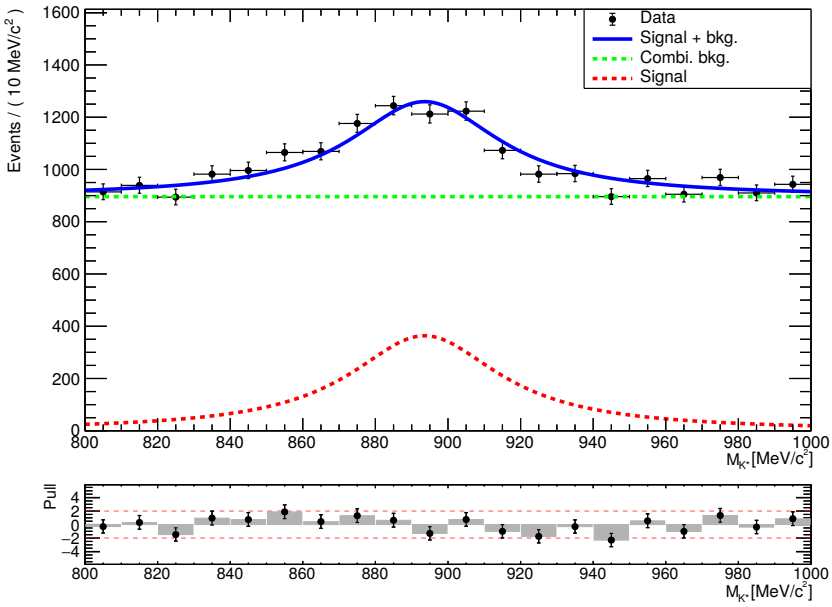


Figure 33: Fitting of $K^{*-} \rightarrow K_s^0\pi^-$ in the signal region - the invariant mass of K^{*-} in range [800, 1000] MeV/c along with the corresponding pull plot.

4.5 Kinematic distribution

After the XGBDT cut, the fit, as shown in Figure 34, is performed by maximising the binned likelihood on the corrected mass (M_{Corr}) distribution to extract the amount of signal. Physical backgrounds, including $B_s^0 \to D_s^+(\to K^{*-}K^0)\mu^+\nu_\mu$ and $B^- \to J/\Psi K^{*-}$, are modelled from MC samples. The background distribution is estimated from the remaining background in the sideband background testing set ($m_{K_S^0\pi^-} \in [690, 800] \cup [1000, 1100]$ MeV/c^2). However, the fit can identify only the background distributions with 19389 ± 537 events and $B_s^0 \to D_s^+(\to K^{*-}K^0)\mu^+\nu_\mu$ with 941 ± 480 in data due to its predominance.

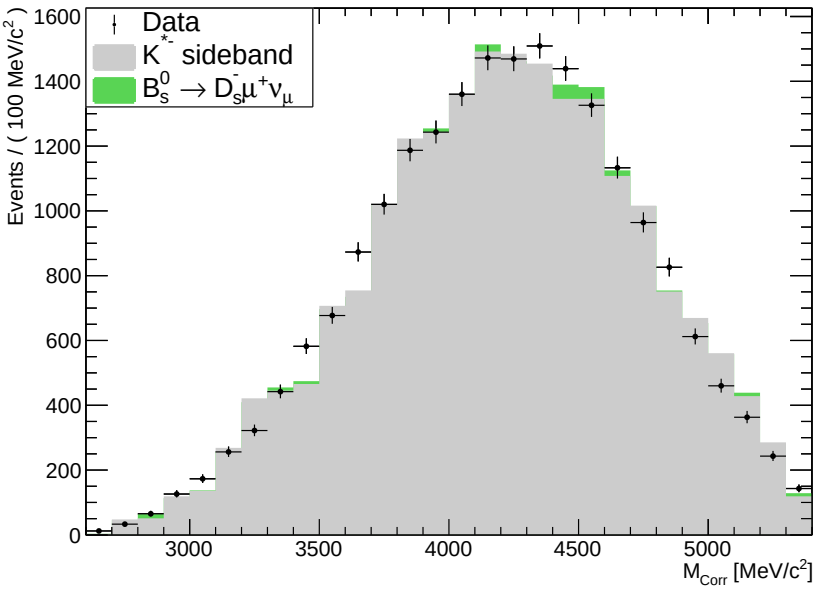


Figure 34: Binned maximum likelihood fit of the corrected mass of $K^{*-}\mu^+$.

Alternatively, the sideband background distribution is scaled to the number of what is found in the M_{K^*} distribution in the fit of Figure 33. The M_{Corr} distribution of the scaled sideband background is plotted together with the corrected mass of the data in Figure 35. This plot implies that the K^{*-} found in the M_{K^*} distribution occupies mostly around $3600 < M_{Corr} < 4900$ MeV/c^2. However, the signal is mostly distributed in the higher corrected mass region.

Distributions of reconstructed quantities, including M_{corr}, M_{vis}, reconstructed mass, missing mass square, and q^2, of signal and physical backgrounds from Monte Carlo samples are presented in Figure 36. We also include the scaled sideband background in these plots to see how these distributions are expected in the data. Only the reconstructed values from the regression methods using RNN and the proportional momentum are shown because the RNN is the best to specify the P_{B_s} and the proportional mass is the worst but its distribution depends on the visible mass which is distinguishable between signal and background.

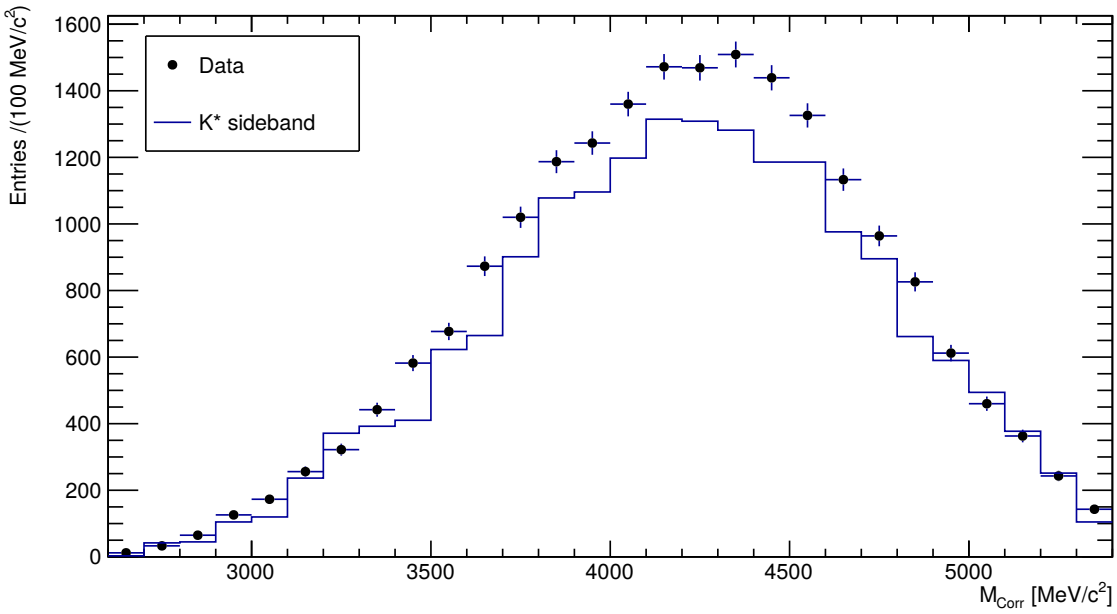


Figure 35: Corrected mass distribution from data with estimated sideband background.

As described before, the K^{*-} candidates found in the data have a M_{Corr} distribution between 3600 and 4900 MeV/c^2. Figure 36 (a) suggests that these K^{*-} may come mostly from $B_s^0 \to D_s^- \mu^+ \nu_\mu$ background. It is also possible that the background (K^{*-} sideband in Figure 36) is not pure. There might be other physical backgrounds from $b \to c$ transitions which have not been considered and excluded in the data.

By comparing reconstructed quantities between using RNN and proportional mass methods in Figure 36 ((c), (e), (g) for RNN; (d), (f), (h) for proportional mass), the later seems to have more discrimination power than the former. Even if the RNN is powerful in specifying the P_{B_s}, the mass-related quantities (M_{B_s} and M^2_{Miss}) reconstructed from this method have a similar distribution in both signal and background. This may be because of the training process. All of the neural network models are trained using the corrected solution of P_{B_s}, which the mass of a B_s^0 meson is constrained, as the output. The reconstructions of the mass-related quantities may converge into the constrain value as the M_{B_s} to $m_{B_s}(\sim 5366\,\text{MeV}/c^2)$ and M^2_{Miss} to 0 MeV^2/c^4. According to the input variables in the NN models calculated from quantities in detector level, the corrected solution is also computed using variables in the detector level with Equation 11 to be consistent. The true value of P_{B_s} in the generator level could have been implemented instead. However, the RNN is still useful in the determination of $q^2(\mu\nu_\mu)$ because M_{B_s} has been fixed as shown in Equation 13.

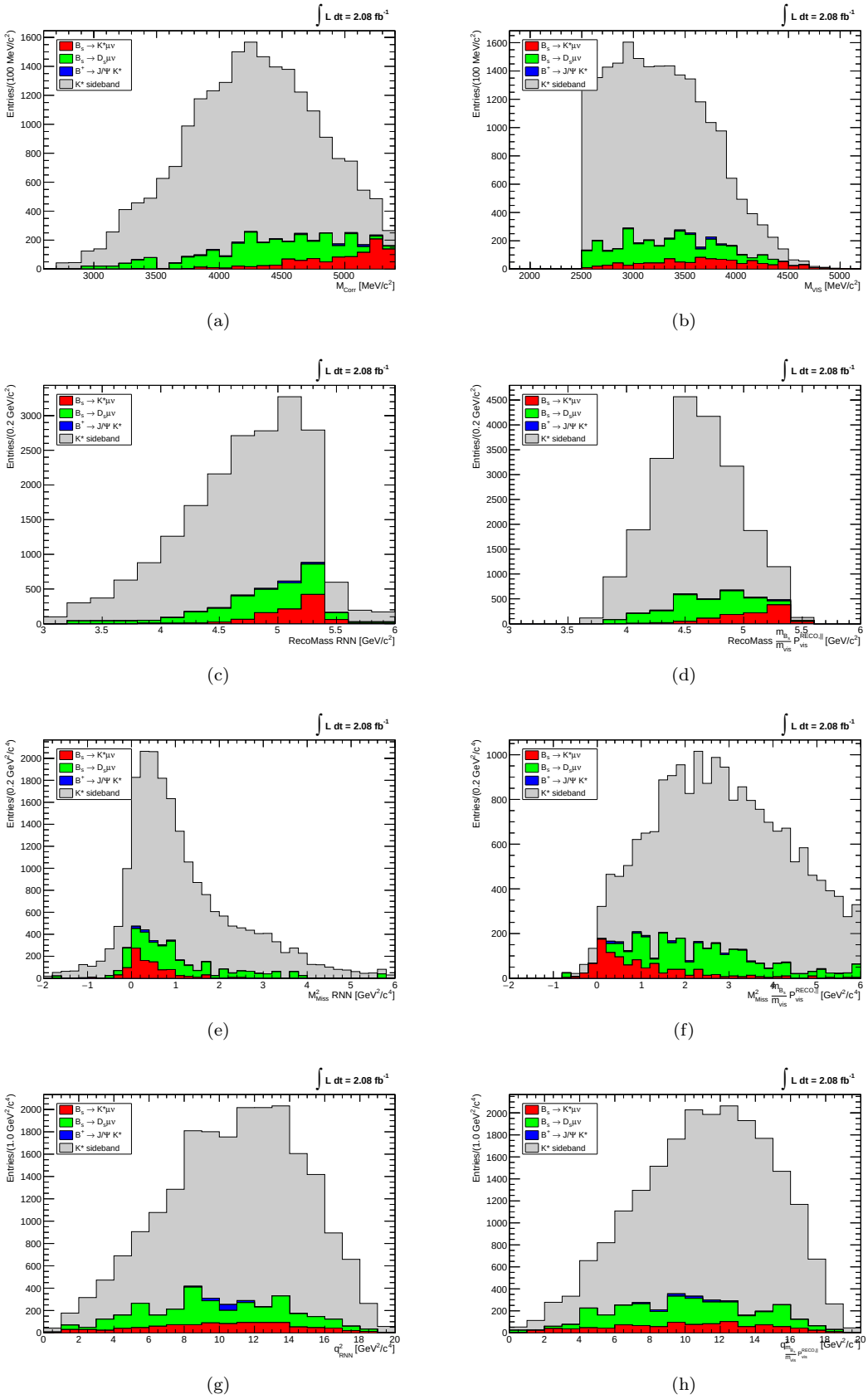


Figure 36: Distribution of various kinematic variables of signal and background after applying the optimised BDT cut. The background $B^- \to K_1^-(\to K^{*-}\pi^0)J/\psi^0(\to \mu^+\mu^-)$ is excluded due to low statistic after the BDT cut.

The $_s$Plot technique can be used to extract only the signal distribution from the fits in Figure 33. In this method, a distribution of interest is extracted by considering 2 sets of variables: discriminating variables and control variables. A discriminating variable is a variable for which the distributions of all the sources of events are known while a control variable is a variable for which the distributions of some sources of events are either unknown. This technique performs a maximum Likelihood fit with the aim to reproduce the signal distribution of a control variable independently from the known properties of the control variable itself. The technique uses fitting parameters in the K^* signal distribution as the discriminating variables to compute a correlation between signal and background in the merged distribution [57].

The weight obtained from the $_s$Plot technique, so called *s-weight*, is computed for the signal distribution from Figure 33 and applied to all of events in data. Figure 37 illustrates the M_{Corr} distribution after applying the *s-weight*. It is uncertain that the signal can be seen in this plot. There might be a distribution from $B_s^0 \to D_s^- \mu^+ \nu_\mu$ and large statistical fluctuation from the $_s$Plot technique.

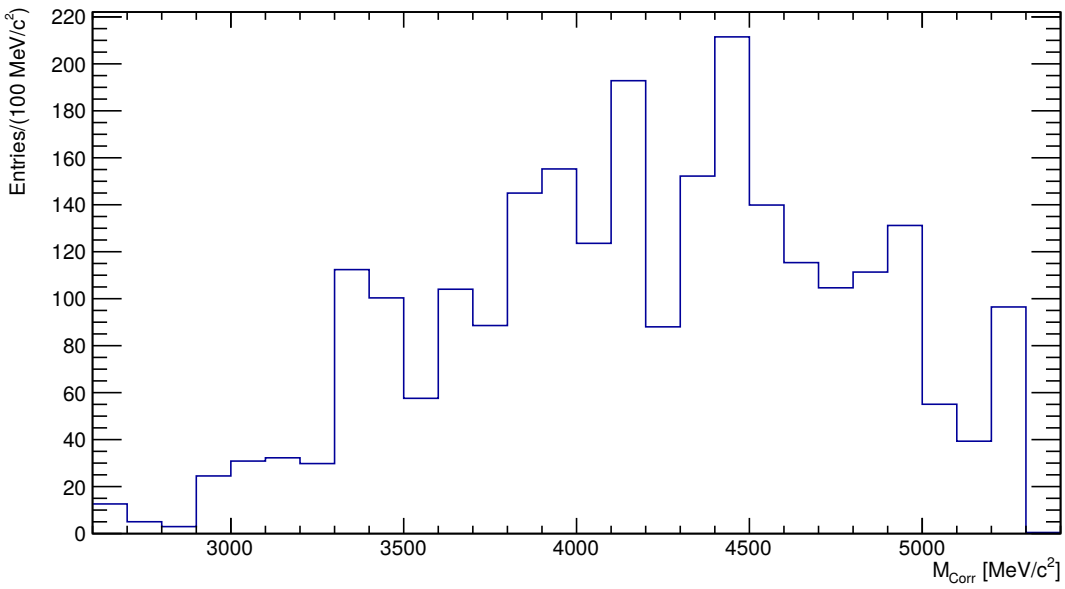


Figure 37: M_{Corr} distribution after applying the $_s$Plot to extract only K^{*-} distribution.

5 Conclusion and outlook

In conclusion, we propose and compare several methods to estimate the momentum of a B_s^0 meson decaying semileptonically. The semileptonic decay of a b-hadron poses a challenge due to the unreconstructed neutrino which causes the B_s^0 momentum to be not well-defined. Even if the neutrino momentum can be computed analytically, this leads to a quadratic equation which has two solutions for the neutrino momentum. Regression techniques are trained using kinematic properties of the decay products. Various models of neural networks are implemented as well as linear regression methods to resolve the ambiguity of two solutions from the quadratic equation.

Although the neural networks predict the momentum of B_s^0 quite well, they provide similar reconstructed mass distributions on signal and background. This results in poor discrimination between those distributions. The benefit of using these models is that the momentum resolution is good, meaning that the B_s^0 momentum is well-defined. The invariant mass of $\mu\nu_\mu$ system, q^2, can be computed from this predicted momentum by constraining on the B_s^0 mass. This is beneficial in studies of semileptonic decays where hadronic form factors and branching fractions are dependent on this q^2. To distinguish between signal and background distributions, it is better to use analytic methods, particularly the proportional mass, in which the momentum of the B_s^0 is reconstructed from the ratio between the fixed B_s^0 mass and the visible mass. Here the visible mass of the background events is distributed such that it can be distinguished from the signal.

Further development of other techniques is encouraged to provide both better B_s^0 momentum resolution and background discrimination. One possibility may be to train the regressor with the reconstructed B_s^0 momentum using the truth 4-momentum of the neutrino from the generator level. The training will then be independent from the B_s^0 mass assumption. More Monte Carlo samples are also required because it is more useful for the network to be applicable on any semileptonic decay.

The first search for $B_s^0 \to K^{*-}\mu^+\nu_\mu$ is performed on data collected by the LHCb experiment in 2012. Preliminary selections are chosen to reduce physical background including $B_s^0 \to D_s^-\mu^+\nu_\mu$, $B^- \to J/\psi K^{*-}$, and $B^- \to J/\psi K_1^-$.

The selections are not sufficiently optimised for a good signal over background discrimination. The main problem is the background distributed in the K^{*-} mass distribution. This could come from other physical backgrounds rather than combinatorial background which should be further investigated. Since the sideband of the K^{*-} mass distribution has the lower visible mass than the signal, this implies that the background may have other unreconstructed massive particles, such as π^0 and K^0. So further study to isolate unreconstructed massive neutral particles are needed. One may look at the decay $B^- \to D^-(\to K_S^0\pi^+\pi^0)\mu^+\nu_\mu$ and conduct a study to exclude this decay from data. Another problem is the faulty implementation in the stripping selection on K_S^0. This causes the reduction of signal candidates by a factor of two. The analysis should be re-performed after the stripping selection will be improved in data.

A Relevant statistical computations

A.1 Luminosity scaling

When an optimisation on each cut is performed, the number of events in the Monte Carlo sample is scaled to be corresponding to the expected number ($N_{expected}$) at the referred luminosity ($\mathcal{L}_{ref}$).

$$N_{expected} = \frac{\mathcal{L}_{ref}}{\mathcal{L}_{MC}} N_{MC}, \tag{20}$$

where $\mathcal{L}_{MC}$ and N_{MC} are luminosity and number of events in a MC sample. In general, the relation between luminosity ($\mathcal{L}$) and number of event (N) can be written as

$$N = \mathcal{L}\sigma_{pp\to b\bar{b}X,fw} f_{s(d)} \mathcal{B}(Process)\epsilon, \tag{21}$$

where ϵ is an efficiency which includes reconstruction and selection efficiencies. $\sigma_{pp\to b\bar{b}X,fw}$ is the cross section of $b\bar{b}$ quarks pair production in association with any particles X in the forward region covered by the LHCb detector from pp-collisions. $f_{s(d)}$ is the fragmentation fraction for the b-quark to hadronise to produce a B_s^0 (B^-)-meson . The branching fraction of a process ($\mathcal{B}(Process)$) depends on the decay chain of that process. Decay chains and the associated branching fractions used in this study are summarised in Table 9.

	Process	Percent
Fractions of decay modes	$\mathcal{B}(B_s^0 \to K^{*-}\mu^+\nu_\mu)$	$\sim 3 \times 10^{-4}$
	$\mathcal{B}(B_s^0 \to D_s^-\mu^+\nu_\mu)$	0.079 ± 0.024
	$\mathcal{B}(B^- \to J/\psi^0 K^{*-})$	$(1.44 \pm 0.08) \times 10^{-3}$
	$\mathcal{B}(B^- \to J/\psi^0 K_1^-)$	$(1.8 \pm 0.5) \times 10^{-3}$
	$\mathcal{B}(K^{*-} \to K_S^0\pi^-)$	$\frac{2}{3} \times \frac{1}{2} \times 1$
	$\mathcal{B}(K_S^0 \to \pi^+\pi^-)$	0.6920 ± 0.0005
	$\mathcal{B}(J/\psi^0 \to \mu^+\mu^-)$	0.05961 ± 0.00033
	$\mathcal{B}(D_s^- \to K^{*-}K^0)$	0.054 ± 0.012
	$\mathcal{B}(K_1^- \to K^{*-}\pi^0)$	$\frac{1}{3} \times (0.16 \pm 0.05)$
Production constants	$\sigma_{pp\to b\bar{b}X,fw7TeV}$	$72.0 \pm 0.3 \pm 6.8\ \mu b$
	$\sigma_{pp\to b\bar{b}X,fw8TeV}$	$\sim 75.0\ \mu b$ [Approximated value]
	$\frac{\sigma_{13TeV}}{\sigma_{7TeV}}_{pp\to b\bar{b}X,fw}$	$2.14 \pm 0.02 \pm 0.13$
	f_s/f_d	0.256 ± 0.020

Table 9: Branching fraction of decay processes as well as constants used in this study. The $\mathcal{B}(B_s^0 \to K^{*-}\mu^+\nu_\mu)$ is estimated from $\mathcal{B}(B_s^0 \to K^-\mu^+\nu_\mu)$ [58] multiplied by a factor of 3. This factor is from the fact that K^{*-} is a vector meson while K^- is a scalar meson. The $\sigma_{pp\to b\bar{b}X,fw8TeV}$ is roughly extrapolated from $\sigma_{pp\to b\bar{b}X,fw7TeV}$. Fractional numbers are calculated using Clebsch–Gordan coefficients. These values are obtained from [4, 59–61].

A.2 Pearson's correlation

A correlation between two variables is determined by Pearson's correlation coefficient P_r. It is defined as

$$P_r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{22}$$

where x_i and y_i are two variables with their average $\bar{x}_i$ and $\bar{y}_i$. The above formula is used to calculate correlation coefficients between each pair of the discriminating variables.

A.3 Pull plot

The deviation between a fit function and data points can be extracted by a pull plot. This method takes each bin with data X_i and uncertainty σ_i which is usually Poissonian ($\sigma_i \equiv \sqrt{X_i}$). The pull ($p$) is defined comparing with the value of the fit function Y_i at the bin's centre as

$$p = \frac{X_i - Y_i}{\sigma_i} \tag{23}$$

The summation over all bins (N_{bins}) refers to a quality of the fit (χ^2)

$$\chi^2 = \sum_{i=1}^{N_{bins}} p_i^2 \tag{24}$$

A.4 Breit-Wigner distribution

The resonance of K^{*-} is modelled with the non-relativistic Breit-Wigner distribution with the mean (μ) and width at half maximum (Γ). It is defined as a function of an invariant mass of the reconstructed K^{*-} (x) as follows

$$f(x : \mu, \Gamma) = N\frac{1}{(x-\mu)^2 + (\Gamma/2)^2}, \tag{25}$$

where N is a normalisation factor. Table 10 presents a fit result of the K^{*-} mass distribution as shown in Figure 32

Variable	Value	Description
`AMOUNT of DATA`	542	Amount of data used for fitting.
`BW_mean`	893.54 ± 1.48	Mean value μ of the Breit-Wigner distribution.
`BW_width`	50.34 ± 3.19	Width at half maximum of the Breit-Wigner distribution.

Table 10: Fitting result of K^{*-} invariant mass distribution from Monte Carlo samples of $B_s^0 \to K^{*-}(\to K_S^0\pi^-)\mu^+\nu_\mu$.

B Flows of neural network stacks

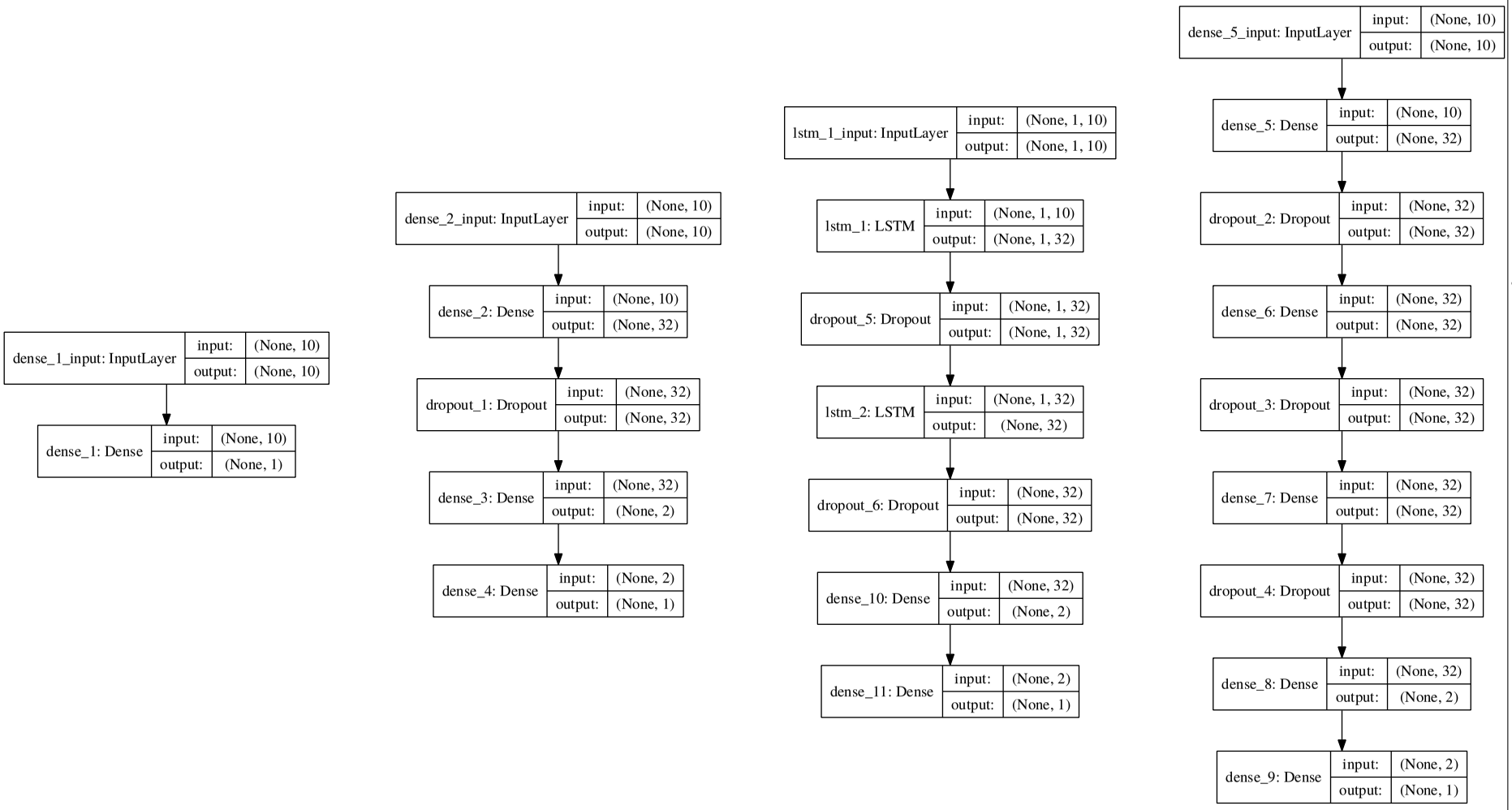


Figure 38: Stacks of several neural network models visualised using Keras package. From left to right, they are denoted as LRNN, SNN, RNN, and DNN as explained in Table 3. Each box marked as "Dense" and "LSTM" refers to one layer with a number of nodes indicated in output channel. The "Dropout" box means half of the nodes in the previous layer will be vanished at the end of training.

C Cut Summary

Process Cut	$B_s \to K^*\mu\nu$ MC	$B_s \to D_s\mu\nu$ MC	$B^+ \to J/\Psi K^*$ MC	$B^+ \to J/\Psi K_1$ MC	Data
Generated Event [MC]	1,928,339	1,387,479	2,023,855	2,717,356	-
Selected Events [MC]	6,939	1,923	12,792	7,016	-
Imp. Stripping [MC]	4,572	829	6,464	2,956	-
Truth Matching [MC]/ Stripping [Data]	4,181	604	4,056	120	5,526,393
Trigger selections	3,406	516	3,205	83	4,428,170
Vetoed Trigger	3,381	511	862	18	4,352,149
Basic Selections	3,285	439	790	13	4,336,747
$ProbNN$ Selections	2,962	392	718	11	2,113,060
5200 MeV $<$ SmallestDeltaChi2MassOneTrack $<$ 5400 MeV	2,776	372	487	9	2,089,451
$\Delta VTX\_KZ < 35.5$ mm	2,403	326	356	9	524,574
$M_{vis} < 2.5$ GeV	2,241	248	342	7	289,883
Detector valid solution	1,779	237	312	7	270,876
XGBDT $<$ 0.55 (SignalRegion | Sideband)	1,008	116	105	-	(20,341 | 18,461)

Table 11: Summary of applied cuts to choose candidate events. Cuts which are marked as [MC] are applied only for Monte Carlo samples and [Data] for the data.

D Prospect on combined search using Run I and Run II

To achieve a statistical power of the measurement, a combined search for this decay should be conducted. Table 12 presents the amount of signal events expected for data collected by the LHCb experiment in 2016 and 2017. We project the amounts of signal by extrapolating from the expected amount in this study with an improvement of the stripping selection to retrieve events with K^0_{S} downstream tracks.

With more data as well as more optimised selections, it could lead to the first observation of the decay $B^0_s \to K^{*-}\mu^+\nu_\mu$ as well as the measurement of $|V_{ub}|$ from this exclusive decay.

Data	Year	Energy (TeV)	Integrated luminosity (fb^{-1})	Expected events
Run I	2011	7	1.11	[1124]
	2012	8	2.08	1011 [2022]
Run II	2015	13	0.33	[642]
	2016	13	1.67	[3247]
	2017	13	1.71	[3325]

Table 12: Prospect amount of signal in different Run of the LHC. Values in square brackets are expected values when the downstream track is recovered from the stripping selections as explained in Section 4.1.

References

[1] G. Aad and Others, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett.* **B716** (2012) 1–29, `arXiv:1207.7214 [hep-ex]`.

[2] S. Chatrchyan and Others, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett.* **B716** (2012) 30–61, `arXiv:1207.7235 [hep-ex]`.

[3] Wikipedia, "The Standard Model of elementary particles." `https://en.wikipedia.org/wiki/Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg`.

[4] J. Beringer and Others, "Review of Particle Physics (RPP)," *Chinese Physics C* **40** (2016) .

[5] The LHCb collaboration, "Determination of the quark coupling strength $|V_{ub}|$ using baryonic decays," *Nature Phys.* **11** (2015) 743–747, `arXiv:1504.01568 [hep-ex]`.

[6] N. Cabibbo, "Unitary Symmetry and Nonleptonic Decays," *Physical Review Letters* (1964) .

[7] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics* (1973) .

[8] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa Matrix," *Phys. Rev. Lett.* **51** (1983) 1945.

[9] A. J. Buras, M. E. Lautenbacher, and G. Ostermaier, "Waiting for the top quark mass, $K^+ \to \pi^+$ neutrino anti-neutrino, B_s^0 - $\bar{B}_s^0$ mixing and CP asymmetries in B decays," *Phys. Rev.* **D50** (1994) 3433–3446, `arXiv:hep-ph/9403384 [hep-ph]`.

[10] B. Aubert, R. Barate, D. Boutigny, *et al.*, "Measurement of the Cabibbo-Kobayashi-Maskawa Matrix Element $|V_{ub}|$ with $B \to \pi\ell\nu$ Decays,".

[11] J. A. Bailey, C. Bernard, C. Detar, *et al.*, "$B \to \pi\ell\nu$ semileptonic form factor from three-flavor lattice QCD: A model-independent determination of $|V_{ub}|$," *Physical Review D - Particles, Fields, Gravitation and Cosmology* **79** no. 5, (2009) 1–27, `arXiv:0811.3640v3`.

[12] J. Ellis, "TikZ-Feynman: Feynman diagrams with TikZ," *Computer Physics Communications* **210** no. January, (2016) 31, `arXiv:1601.05437`. `http://arxiv.org/abs/1601.05437`.

[13] J. M. Flynn, T. Izubuchi, T. Kawanai, *et al.*, "$B \to \pi\ell\nu$ and $B_s \to K\ell\nu$ form factors and $|V_{ub}|$ from 2+1-flavor lattice QCD with domain-wall light quarks and relativistic heavy quarks," *Phys. Rev. D* **91** no. 7, (2015) 74510, `arXiv:1501.05373 [hep-lat]`.

[14] A. Bharucha, D. M. Straub, and R. Zwicky, "$B \to V\ell^+\ell^-$ in the Standard Model from light-cone sum rules," *Journal of High Energy Physics* **08** (2016) 98. `http://link.springer.com/10.1007/JHEP08(2016)098`.

[15] The LHCb collaboration, "Angular analysis of the $B^0 \to K^{*0}\mu^+\mu^-$ decay using 3 fb^{-1} of integrated luminosity," *Journal of High Energy Physics* **2016** no. 2, (2016) 1–79.

[16] T. Feldmann, B. Müller, and D. Van Dyk, "Analyzing $b \to u$ transitions in semileptonic $\bar{B}_s \to K^{*+}(\to K\pi)\ell^-\bar{\nu}_\ell$ decays," *Physical Review D* **92** no. 3, (2015) 1–14.

[17] J. D. Jackson, "Particle and polarization angular distribution for two and three-body decays," *in Lectures delivered at the Les Houches summer school of theoretical physics* (1965) 325–366.

[18] M. Jacob and G. Wick, "On the General Theory with of Collisions Spin for Particles," *Annals Phys.* **7** (1959) 404–428.

[19] L. Ferreira, "Study of $B_s \to D_s\mu\nu_\mu$ decays at the LHCb experiment," tech. rep., EPFL, Lausanne, Switzerland, 2017.

[20] G. Ciezarek, A. Lupato, M. Rotondo, and M. Vesterinen, "Reconstruction of semileptonically decaying beauty hadrons produced in high energy pp collisions," *JHEP* **02** (2017) 21, `arXiv:1611.08522 [hep-ex]`.

[21] M. Neubert, "Heavy-quark symmetry," *Physics Reports* **245** no. 5-6, (1994) 259–395, `arXiv:9306320 [hep-ph]`.

[22] V. S. Kirsebom, "Comparison of $B_s^0 \to D_s^-\mu^+\nu_\mu$ and $B^0 \to D_s^-\mu^+\nu_\mu$ to test SU(3) symmetry," tech. rep., CERN Summer Student Project, 2017. `https://lhcb.web.cern.ch/lhcb/Collaboration_prizes/2017/Veronica.pdf`.

[23] The LHCb collaboration, "LHC-B: A Dedicated LHC Collider Beauty Experiment for Precision Measurements of CP Violation," tech. rep., CERN-LHCC-95-5, 1995. `http://cds.cern.ch/record/290868`.

[24] The LHCb collaboration, "The LHCb Detector at the LHC," *Journal of Instrumentation* **3** no. 08, (2008) S08005. `http://stacks.iop.org/1748-0221/3/i=08/a=S08005?key=crossref.358ac80e1a6b6ba36f68c89dc0c4bed4`.

[25] The LHCb collaboration, "LHCb Detector Performance," *International Journal of Modern Physics A* **30** no. 7, (2015) 1530022, `arXiv:1412.6352`. `http://arxiv.org/abs/1412.6352%0Ahttp://dx.doi.org/10.1142/S0217751X15300227`.

[26] The LHCb collaboration, "Template for writing LHCb papers," *LHCb Internal note* (2015) 1–6. `papers2://publication/uuid/4FB75CCA-A8DC-489C-B623-08C80CF83EE9`.

[27] The LHCb collaboration, "VErtex LOcator (VELO)." `https://lhcb-public.web.cern.ch/lhcb-public/en/detector/VELO-en.html`.

[28] R. Aaij and Others, "Performance of the LHCb Vertex Locator," *JINST* **9** (2014) P09007, `arXiv:1405.7808 [physics.ins-det]`.

[29] The LHCb collaboration, "Tracking strategies used in LHCb." `https://twiki.cern.ch/twiki/bin/view/LHCb/LHCbTrackingStrategies`.

[30] M. Adinolfi, G. Aglieri Rinella, E. Albrecht, *et al.*, "Performance of the LHCb RICH detector at the LHC," *European Physical Journal C* **73** no. 5, (2013) 1–17, `arXiv:1211.6759`.

[31] E. P. Olloqui, "LHCb Preshower(PS) and Scintillating Pad Detector (SPD): Commissioning, calibration, and monitoring," *Journal of Physics: Conference Series* **160** (2009) .

[32] The LHCb collaboration, "Trigger Schemes." https://lhcb.web.cern.ch/lhcb/speakersbureau/html/TriggerScheme.html.

[33] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.* **178** (2008) 852–867, arXiv:0710.3820 [hep-ph].

[34] I. Belyaev and Others, "Handling of the generation of primary events in Gauss, the LHCb simulation framework," *J. Phys. Conf. Ser.* **331** (2011) 32047.

[35] D. J. Lange, "The EvtGen particle decay simulation package," *Nucl. Instrum. Meth.* **A462** (2001) 152–155.

[36] S. Agostinelli and Others, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth.* **A506** (2003) 250–303.

[37] M. Clemencic, G. Corti, S. Easo, *et al.*, "The LHCb simulation application, Gauss: Design, evolution and experience," *J. Phys. Conf. Ser.* **331** (2011) 32023.

[38] K. Abe, K. Abe, T. Akagi, *et al.*, "Measurement of R_b Using a Vertex Mass Tag," *Physical Review Letters* **80** no. 4, (1998) 2–7.

[39] The LHCb collaboration, "Measurement of the Ratio of Branching Fractions $\mathcal{B}(\bar{B}^0 \to D^{*+}\tau^-\bar{(}\nu)_\tau)/\mathcal{B}(\bar{B}^0 \to D^{*+}\mu^-\bar{(}\nu)_\mu)$," *Physical Review Letters* **115** no. 11, (2015) 1–10.

[40] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research* **12** (2011) 2825–2830.

[41] F. and others Chollet, "Keras," 2015. https://keras.io.

[42] Martin Abadi, Ashish Agarwal, Paul Barham, *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. http://tensorflow.org/.

[43] D. F. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks* **2** no. 6, (1991) 568–576.

[44] D. P. Kingma and J. L. Ba, "ADAM : A METHOD FOR STOCHASTIC OPTIMIZATION," in *International Conference on Learning Representations*, pp. 1–15. 2015. arXiv:arXiv:1412.6980v8.

[45] S. J. Reddi, S. Kale, and S. Kumar, "ON THE CONVERGENCE OF ADAM AND BEYOND," in *International Conference on Learning Representations*, pp. 1–23. 2018.

[46] L. C. Jain and L. R. Medsker, *Recurrent Neural Networks: Design and Applications.* CRC Press, Inc., Boca Raton, FL, USA, 1st ed., 1999.

[47] A. Krogh and J. A. Hertz, "A Simple Weight Decay Can Improve Generalization," in *Advances in Neural Information Processing Systems 4*, J. E. Moody, S. J. Hanson, and R. P. Lippmann, eds., pp. 950–957. Morgan-Kaufmann, 1992. http://papers.nips.cc/paper/563-a-simple-weight-decay-can-improve-generalization.pdf.

[48] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research* **15** (2014) 1929–1958. http://jmlr.org/papers/v15/srivastava14a.html.

[49] LHCb collaboration: Computing Group, "LHCb Experiment: STRIPPING project," 2017. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/stripping/.

[50] S. Tolk, J. Albrecht, F. Dettori, and A. Pellegrino, "Data driven trigger efficiency determination at LHCb," *LHCb-PUB-2014-039* (2015) .

[51] The LHCb collaboration, "Second analysis steps: TisTos DIY." https://lhcb.github.io/second-analysis-steps/18-tistos-diy.html.

[52] L. Guo, "Classification and Decision Tree Classifier Introduction." http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/lguo/decisionTree.html.

[53] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and Regression Trees.* 1984.

[54] J. Zhu, H. Zou, S. Rosset, and T. Hastie, "Multi-class Adaboost," *Statistics and Its Interface* **2** (2009) 349–360, arXiv:arXiv:1011.1669v3. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.158.4221.

[55] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics* **29** no. 5, (2001) 1189–1232, arXiv:arXiv:1011.1669v3.

[56] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pp. 785–794. ACM, New York, NY, USA, 2016. http://doi.acm.org/10.1145/2939672.2939785.

[57] M. Pivk and F. R. Le Diberder, "SPlot: A Statistical tool to unfold data distributions," *Nucl. Instrum. Meth.* **A555** (2005) 356–369, arXiv:physics/0402083 [physics.data-an].

[58] S. Braun, "Update on $|V_{ub}|$ from $B_s^0 \to K^-\mu^+\nu_\mu$," 2017. https://indico.cern.ch/event/603961/contributions/2532455/attachments/1440166/2216790/ASweek_Vub.pdf.

[59] The LHCb collaboration, "Measurement of the b -Quark Production Cross Section in 7 and 13 TeV pp Collisions," *Physical Review Letters* **118** no. 5, (2017) 1–11.

[60] The LHCb collaboration, "Measurement of the fragmentation fraction ratio fs/fdand its dependence on B meson kinematics," *Journal of High Energy Physics* **2013** no. 4, (2013) 1–16, arXiv:1301.5286 [hep-ex].

[61] The LHCb collaboration, "Measurement of b hadron production fractions in 7 TeV pp collisions," *Physical Review D* **85** no. 3, (2012) 1–16, arXiv:1111.2357.

Acknowledgements

This work would not have been possible without the help from people who supported and encouraged me during my study at EPFL.

First of all, I would like to express my deep gratitude to Dr. Michel De Cian, my research supervisor, for his patient guidance, enthusiastic encouragement and useful critiques of this research work. He was always available to answer my questions unhesitantly and passionately.

I also appreciate that Prof. Tatsuya Nakada accepted to be my thesis director. I am glad to have an opportunity to conduct a thesis with the first spokeperson of the LHCb collaboration.

I would like to thank Prof. Aurelio Bay and Prof. Olivier Schneider who allowed me to do research with the Laboratory for High Energy Physics (LPHE) since I started my Master study at EPFL. I thank the members of the LPHE, particularly Maria Elena Stramaglia who gave me very useful advice about future careers in physics.

My grateful thanks are also extended to my fellow Master students, for constructive physics and statistics discussions during the time we spent in BSP 612.

Finally, without encouragement and support from my family, this research would have not been possible.

S. Ek-In