UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Corso di Laurea Magistrale in Fisica

Tesi di Laurea

Measurement of $b\bar{b}$ production in s = 13 TeV pp collisions with the CMS experiment

Relatore
Prof. Franco Simonetto

Laureando
Enrico Lusiani

Anno Accademico 2018/2019

Contents

Chapter 1

Introduction

A precise measurement of the $\sigma\,\mathrm{pp} \to \mathrm{b\bar{b}}X$ is of extreme importance for testing next to leading order Quantum Chromodynamics calculations. Previous analysis at lower energies have been made by the UA1 collaboration at the cern $S\bar{p}ps$ collider, with a $\sqrt{s} = 0.63\,\mathrm{TeV}$ ([1]) and by the the CDF and D0 experiments at Tevatron, at $\sqrt{s} = 1.8\,\mathrm{TeV}$ and $\sqrt{s} = 1.96\,\mathrm{TeV}$ ([2], [3], [4] and [5]). More recently there have been measurements in LHC at $\sqrt{s} = 7\,\mathrm{TeV}$ from the LHCb experiment, using semiinclusive decays ([6]), and from CMS, using fully reconstructed $\mathrm{B}^+ \to \mathrm{J/\psi K} \to \mu^+\mu^-\mathrm{K}$ decays ([7]).

In our thesis we present a novel method for the measurement of this cross-section at 13 TeV using semileptonic decays, that does not depend on measure of the luminosity, the branching ratio of the processes or the muon reconstruction efficiencies. The measurement uses the ratio the numbers of semileptonic decays on pileup and on a vertex with an already fully reconstructed B-meson. All the usual normalizations will be simplified by virtue of using the same process on the same events. We expect instead our uncertainty to be dominated by the inaccuracies of the models from the Monte Carlo and the similarities between their parameters. We will develop new tags based on Machine Learning techniques to help the fit converge and minimize fit uncertainties and correlations. In the end the result presented will be ratio between the integral cross-section of the $\mathrm{pp} \to \mathrm{b\bar{b}}X$ and the "minimum-bias cross-section" in the CMS experiment, an already well measured value. We will also discuss on the feasability of this new method and whether it is possible to achieve better accuracies than the previously used ones.

Chapter 2

The CMS detector

CMS (Compact Muon Solenoid) is one of the two general purpose detectors active at the Large Hadron Collider at CERN, the other being ATLAS. It's optimized for the search of new physics, in particular:

- Higgs Physics: photons and leptons detection, decay vertex reconstruction
- Jet Physics: track and jet reconstruction with fine granularity
- Dark Matter, SUSY and other new physics research
- Exploration of physics at the TeV scale

In collider physics we often work in an environment with an obvious simmetry: the beam axis. When choosing a coordinate system one would be tempted to use spherical or cilindrical coordinates. Unfortunately these corrdinates are not invariant with respect to lorentz transformations, making it unsuitable to use.

A common base used is instead the (p_T, η, φ). This three variables are related to the spherical coordinates (p, θ, φ) with the formulas

$$p_T = p \cdot \sin\theta \tag{2.1}$$

$$\eta = -\ln(\tan\frac{\theta}{2}) = \operatorname{arctanh}\left(\frac{p_L}{|p|}\right) \tag{2.2}$$

$$\varphi = \varphi \tag{2.3}$$

The advantage of this coordinate system is that difference in η, unlike difference in θ are Lorentz invariants for boosts along the beam axis.

CMS is structured as a cilindrical detector, composed by various layers with different responsabilities.

The innermost layer is the tracker. The tracker is designed to provide high resolution track reconstruction, as well as primary and secondary vertex (primary particle decays) reconstruction. Inside the tracker is split in two different subregions according to the distance from the beam spot. The innermost one is a silicon pixel tracker, while externally there is a silicon microstrips tracker. The separation is both to increase the resolution when close to the event and to account for the higher flux of particles at lower radii. The tracker provides a momentum resolution of $0.5\,\% \oplus 15\,\% \cdot p_T[\text{TeV}]$ in the central region with $|\eta| < 1.6$ degrading to $0.5\,\% \oplus 60\,\% \cdot p_T[\text{TeV}]$ as $|\eta|$ approaches 2.5. The trasversal impact parameter resolutions is between 10 and 100 m.

After the tracker is the ElectroMagnetic calorimeter (ECAL). The ECAL is made with 80000 PbWO4 scintillating crystals distributed between the barrel and the endcap. The main function of the ECAL is intercepting electrons and photons and collecting and thus estimating their energy. CMS' ECAL can reach a resolution of $\frac{2.8\,\%}{\sqrt{E}} \oplus \frac{0.12}{E} \oplus 0.30\,\%$.

The next layer is the Hadronic Calorimeter (HCAL). The HCAL is needed to measure the energy of hadrons, which due to the differents interactions produce a signal deeper in the detector. To avoid losing energy due to particles traversing the whole volume of the detector, the HCAL is made of alternating layers of brass (or steel), which slows down particles, and liquid scintillator material to read the energy. The HCAL resolution is $\frac{70\,\%}{\sqrt{E}} \oplus 7\,\%$.

All the previous layers were placed inside the solenoid that gives the name to the experiment. The solenoid can produce a magnetic field of up to 4 T. The particles from the interaction are curved by the action of the magnetic field, with a radius depending on their trasversal momentum. Using the signal that the particles leave in the tracker, we can measure the radius and thus the p_T.

Outside the solenoid (but embedded in the return yoke) are the muon chambers. The muon chambers are of three different types: Drift Tubes (in the barrel), Cathode Strip Chambers (in the endcaps) and Resistive Plate

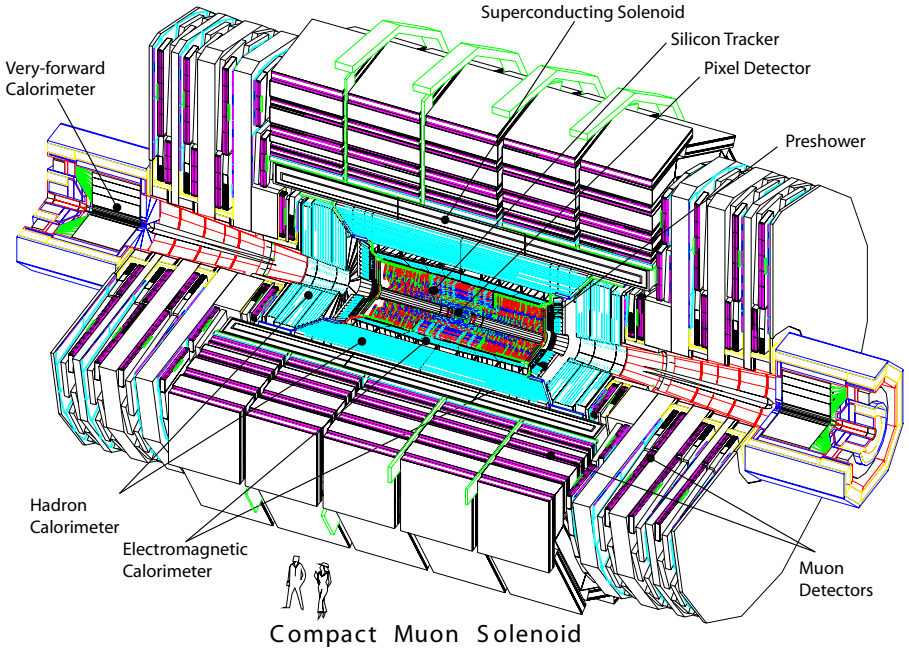


Figure 2.1: *cms*

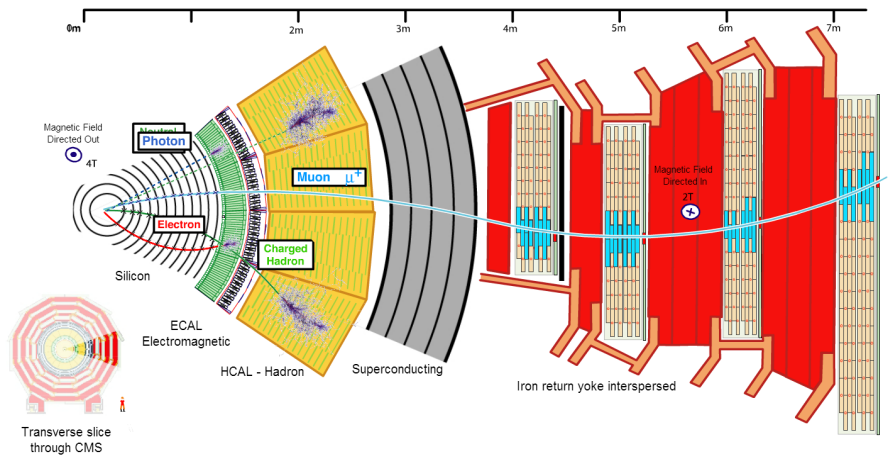


Figure 2.2: *cms trasv*

Chambers. The specialized muon detectors are the outermost layer because we expect that the only particles able to pass through both calorimeters and the iron return yoke are the muons, giving us a clean signal.

Chapter 3

Analysis methods

As stated in the previous chapters, the objective of this thesis is the measure of the $pp \to b\bar{b}X$ cross-section in the collision in the Large Hadron Collider. The analysis relies on having a fully reconstructed secondary B-meson vertex in the selected event and uses the decay of a b quark into a μ.

For a normal LHC collision, the probability of finding a μ (μ) from a b quark decay is a sum of the probabilities of many processes, as the μ can be produced by the virtual W from the b quark decay ("direct" μ), coming from the decay of a particle coupled to the W (such as a τ or a charmed meson) or from the decay of the quark the b decayed into (usually a c). Since not all the processes are mutually exclusive, to compute the probability we would be forced to use combinatorial terms. In our case we are interested in the total number of such μ collected after enough events, and this allows us to ignore the combinatorial factors, which are integrated in a term called B_μ that represents the average number of μ produced in a decay of a b quark.

Ignoring for the moment non-b background, the average number of μ where the B^+ and the μ come from two different collisions recored in the same bunch crossing ($N_{\mu,pil}$, "pileup" μ) is:

$$\begin{aligned} \left\langle N_{\mu,pil} | N_{PU} \right\rangle &= N_{PU} \cdot \frac{\sigma(pp \to b\bar{b}X)}{\sigma(pp \to X)} \cdot 2B_\mu \cdot \varepsilon_{detection} \\ &= N_{PU} \cdot R_b \cdot 2B_\mu \cdot \varepsilon_{detection} \end{aligned} \tag{3.1}$$

where N_{PU} is the number of reconstructed pp collisions ("primary vertices") excluding the one with the B^+ and is equal to $N_{PV} - 1$, $\varepsilon_{detection}$ is the detection probability of a μ from that specific decay and $\sigma(pp \to b\bar{b}X)$ is the probability of a $b\bar{b}$ pair being produced in a pp collision. We also used in the last passage the factor R_b, which relates the cross-section we want to the so called "minimum bias cross-section", a well known parameter of LHC and the CMS detector that includes the actual total inelastic cross-section of the proton scattering and the efficiency in the reconstruction of the vertex. This parameter is what we will actually measure in the analysis.

If we restrict the analysis to the μ that come from the same vertex as a reconstructed B-meson ("signal" μ), the number (called here $N_{\mu,osb}$, for opposite side b) is instead

$$\left\langle N_{\mu,osb} \right\rangle = B_\mu \cdot \varepsilon_{detection} \tag{3.2}$$

As we can see, we lost the factors pertaining to the b cross-section, as b quarks are mostly produced in opposite flavor pairs at the LHC ($t\bar{t}$ couples where one of the t does not decay into a b are negligible from what we see in our data) and we are already requiring the presence of a first one by selecting vertices with a B-meson.

The ratio of these two quantities contains our R_b factor and is independent on B_μ and $\varepsilon_{detection}$. For each N_{PU}, we have

$$\frac{\left\langle N_{\mu,pil} | N_{PU} \right\rangle}{\left\langle N_{\mu,osb} \right\rangle} = 2N_{PU} \cdot R_b \tag{3.3}$$

Equation 3.3 is valid for each N, but uses the population average of $N_{\mu,pil}$ and $N_{\mu,osb}$ which is not known. If we replace that with the sample average, the equation becomes an approximation. To improve our estimate we can average the values for each N_{PU}. The equation becomes

$$\begin{aligned}\widehat{R_{\mathrm{b}}} &= \frac{1}{2}\sum_n p_n \frac{\overline{N_{\mu,pil}|N_{PU}=n}}{n\cdot\overline{N_{\mu,osb}}} \\ &= \frac{1}{2\overline{N_{\mu,osb}}}\sum_n p_n \frac{\overline{N_{\mu,pil}|N_{PU}=n}}{n} \\ &= \frac{N_0}{2\sum_k N^k_{\mu,osb}}\sum_n p_n \frac{\sum_{k,N^k_{PU}=n} N^k_{\mu,pil}}{n\cdot\#(N_{PU}=n)}\end{aligned} \tag{3.4}$$

In the last passage we simply expanded the average number of muons for pileup and signal, with k as an index running over the number of bunch crossings, and used $\#(N_{PU}=n)$ as the number of bunch crossing that resulted in n interactions between protons. $\widehat{R_{\mathrm{b}}}$ here is an estimator of the true value of R_{b}. We can make a further approximation on this formula, if we consider that $\frac{\#(N_{PU}=n)}{N_0} \simeq p_n$ in the limit of an infinitely large sample. We can thus write

$$\begin{aligned}\widehat{R_{\mathrm{b}}} &= \frac{N_0}{2\sum_k N^k_{\mu,osb}}\sum_n p_n \frac{\sum_{k,N^k_{PU}=n} N^k_{\mu,pil}}{n\cdot\#(N_{PU}=n)} \\ &= \frac{1}{2\sum_k N^k_{\mu,osb}}\sum_n \frac{\sum_{k,N^k_{PU}=n} N^k_{\mu,pil}}{n} \\ &= \frac{1}{2}\frac{\sum_k \frac{N^k_{\mu,pil}}{N^k_{PU}}}{\sum_k N^k_{\mu,osb}}\end{aligned} \tag{3.5}$$

Unlike the previous forms, this equation contains quantities that we can estimate, the total number of b muons from signal and the total number of b muons from pileup weighted by number of pileup vertices.

The approximations used in the process may actually make $\widehat{R_{\mathrm{b}}}$ a biased estimator, but we verified trough a quick Monte Carlo simulation that this is not the case.

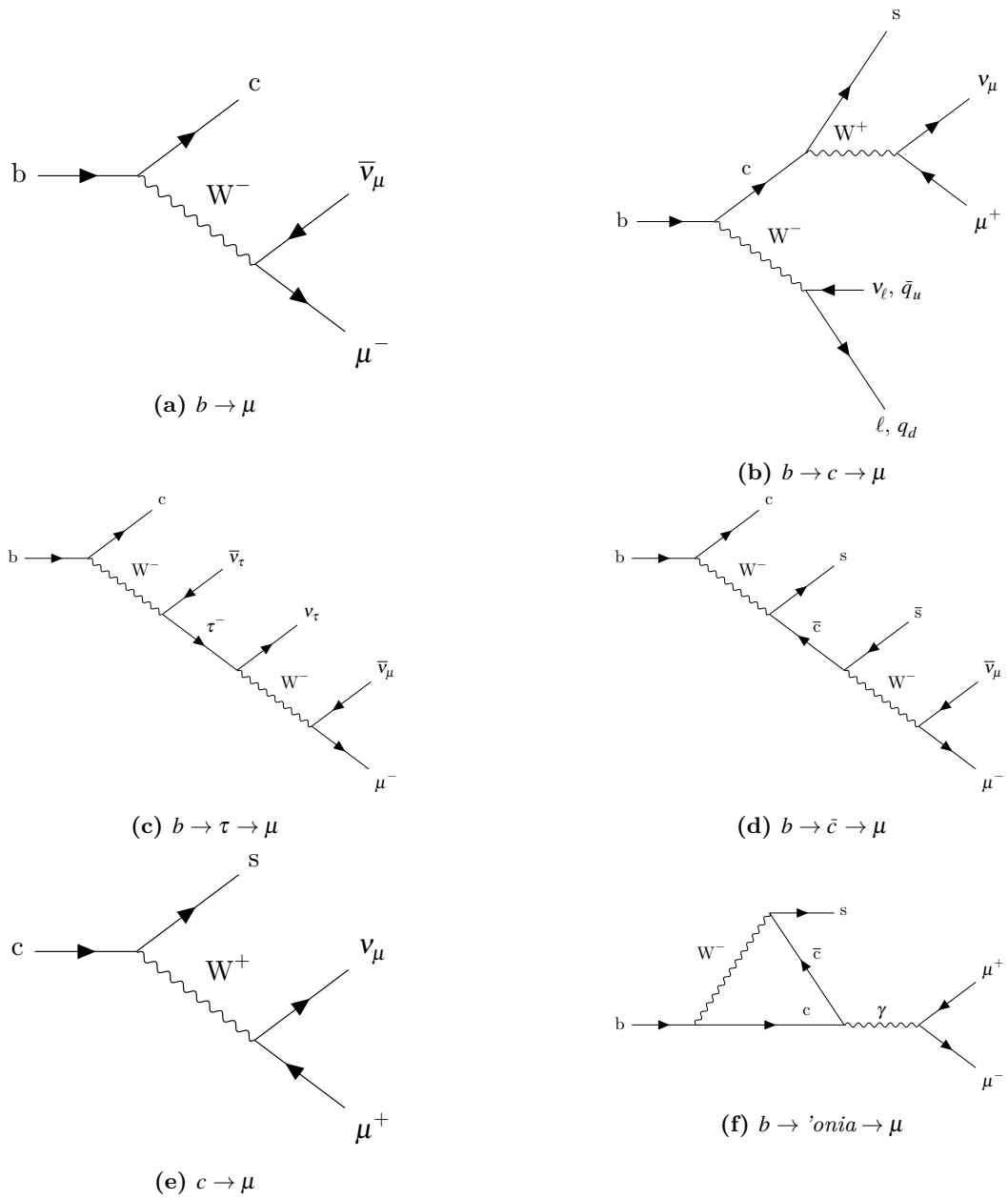
3.1. Signal and background sources

All the above equations are ignoring background sources, as we are assuming that all the μ in the event are derived from a b decay, which obviously doesn't hold for a real event. We need to use various tagging variables to be able to count only μ produced in a b decay.

We group our events in four categories (two signal and two background):

- $\mathrm{b} \to \mu$, direct μ, as shown in Figure 3.1a
- $\mathrm{b} \to X \to \mu$, non-prompt μ, as shown in Fig 3.1c and 3.1d
- $\mathrm{c} \to \mu$, prompt charm hadron decay, as shown in Figure 3.1e
- "fake" μ, hadrons identified as μ inside the CMS track reconstruction algorithms. This includes charged hadrons crossing all the detector material and reaching the μ chambers ("punch through, sail through"), and μ from the leptonic or semileptonic decays of pions and kaons. The background from Υ, $\mathrm{W}^{\pm}$ or Z^0 decays is negligible.

The fake category itself can be split in 3 additional categories: fake muons appearingly coming from the pp interaction (as they were produced as a result of the decay of only very short lived particles), fake muons with only one long lived ancestor and fake muons with two long lived ancestors.



(a) $b \to \mu$ **(b)** $b \to c \to \mu$ **(c)** $b \to \tau \to \mu$ **(d)** $b \to \bar{c} \to \mu$ **(e)** $c \to \mu$ **(f)** $b \to$ *'onia* $\to \mu$

Figure 3.1: *Main Feynman diagrams involved in the muon production for the considered categories*

Decay category	Composition of MC sample
$b \to \mu$	0.689
$b \to X \to \mu$	0.176
$c \to \mu$	0.0511
fake muons	0.0833

Table 3.1: *Fractions of muons in each category in the MC samples. These values refer to the whole sample, ignoring differences in* η*, or* p_T

3.2. Discrimination of signal from background

For the separation of signal from background we mainly use three different tags. The tags are only outlined here but will be explained more throughly in chapter 5.

The first one is the output of an MVA, trained to reject fake μ using variables linked to the quality of the

track reconstruction (a Deep Neural Network to be more precise). Usually these kinds of MVA are used as a selection for the tracks, to reject as many fake μ as possible while minimizing the number of real μ rejected. This means that the output distributions for fake and real μ have to be as different as possible, with the bulk of the distribution as close as possible to the opposite sides of the codomain of the output. This property makes the variable a good fit for our use case. Using the MC directly as a model, without trying to describe analitically the shape of the distribution, we superimpose the distributions on top of the data and from the result of the fit we extract the number of fakes in the sample.

A second variable used is another DNN, trained on isolation based variables (isolation is a measure of how many other tracks are around the track we are reconstructing as μ). We notice that isolation seemed to be a good tag to select direct μ from other categories. The use of this variable is very similar to the fake rejection MVA, fitting the output to MC models.

The last MVA variable is a mix of the previous two. This tag uses informations from the fake MVA, from the isolation MVA and from the impact parameter to build an output ideally ablt to separe all the categories. It's used as the last stop before the impact parameter fit, summarizing the information from previous tags.

The main tag used in the thesis is the impact parameter, already defined before. Since B-hadrons typically have a longer mean life, the distribution of the impact parameter for the μ coming from their decay is expected to have an higher mean. Using generated Monte Carlo samples, we extract the distributions separed by decay and fit them with an appropriate analytical model. The data is then fitted with a linear combination of these models.

3.3. Reconstructed vertices

We said in the previous section that we need a fully reconstructed B-meson decay vertex for our analysis. This is because we need to be certain that the mother particle is actually a B-meson and also because we need to be able to exclude the tracks coming from this decay from the count of μ for the analysis.

The chosen decays to be reconstructed are:

- $\mathrm{B}^+ \to \mathrm{J}/\psi\mathrm{K}^+ \to \mu^+\mu^-\mathrm{K}^+$
- $\mathrm{B}^0 \to \mathrm{J}/\psi\mathrm{K}^* \to \mu^+\mu^-\mathrm{K}^\pm\pi^\mp$
- $\mathrm{B}_\mathrm{s}^0 \to \mathrm{J}/\psi\phi \to \mu^+\mu^-\mathrm{K}^+\mathrm{K}^-$

This decay were chosen because they leave a very distinctive track in the CMS tracker (2 μ with invariant mass close to that of the J/ψ, one extra track for the B^+ or two extra tracks for the B^0 and B_s^0, all forming a common vertex). They also the decays that are often used in CP violation measures, allowing us to use large Monte Carlo samples that are produced for those analysis.

The μ from the J/ψ and the other tracks from the decay are of course removed from the count of the μ for the analysis.

Chapter 4

Samples used

During the analysis we used 5 different datasets, 3 Monte Carlo samples as a model for the fit, one MC sample for training the MVAs and one data sample, all from the year 2016.

The Monte Carlo samples are dataset generated for other parallel analysis. Each of them is generated so that every event contains at least one secondary vertex of a specific decay, which is used as a label for the dataset. The used datasets are

- $B_s^0 \to J/\psi\phi \to \mu^+\mu^- K^+K^-$, $\sim 98 \cdot 10^6$ events, $2.825 \cdot 10^7$ after secondary vertex reconstruction
- $B^+ \to J/\psi K^+ \to \mu^+\mu^- K^+$, $\sim 19 \cdot 10^6$ events, $7.986 \cdot 10^6$ after SV reconstruction
- $B^0 \to J/\psi K^* \to \mu^+\mu^- K^\pm\pi^\mp$, $\sim 73 \cdot 10^6$ events, $1.9734 \cdot 10^7$ after SV reconstruction
- $B^0 \to \mu^+\mu^- K^* \to \mu^+\mu^- K^\pm\pi^\mp$, $\sim 121 \cdot 10^6$ events, $3.103 \cdot 10^6$ muons

The last sample does not require a reconstructed vertex, as it is only used to train the muon MVAs. For that we only need to be sure to exclude the muons from the B-decay, a process that has a much higher efficiency than a full reconstruction.

The data sample is the union of all the so called "Charmonium" datasets collected in the year 2016. This means that they contain only events that passed a specific set of High Level Triggers (HLT). Every high level trigger is a set of selections that can be computed extremely fast in a computer farm, as these triggers are executed online during the acquisition and are used to decide if a specific event is worth being stored or not. In the case of the Charmonium dataset the triggers are designed to ensure the presence of at least one charmonium meson decayin ito two μ inside the event.

In each dataset we had to chose one of the three possible decays (as explained in the previous section) to reconstruct. In the MC samples we simply reconstructed the required decay, to maximize the number of events not rejected. In the data sample we instead chose to reconstruct the $B^+ \to J/\psi K^+$ decay, as of the three decays it's one of the most likely to occur.

After a first skim, that only selected events with the required decay, some additional selections are applied on the events.

The first one is a selection on the HLT, to reduce background. We require that each event passed a trigger that is internally called HLT_DoubleMu4_JpsiTrk_Displaced. The trigger requires that in the event there are:

- Two μ with total invariant mass close to that of the J/ψ (in a window of $150\,\mathrm{MeV}$ around it), forming a common vertex fitted with a probability $P > 10\%$
- The vertex has a distance from the beam spot L_{xy} in the transversal plane greater than 3 times its error.
- The cosine of the angle between L_{xy} and the total momentum of the μ is greater than 0.9 (we don't want this cut to be strict, since it's considering only the J/ψ momentum instead of the full B-meson)
- An extra track, with $p_T > 0.8$, compatible with the μ's vertex with a $\frac{\chi^2}{\mathrm{NDF}} > 10$. No extra requirements are made on this vertex

The second set of cuts applied are quality cuts for the main secondary vertex at the offline analysis stage. These cuts include

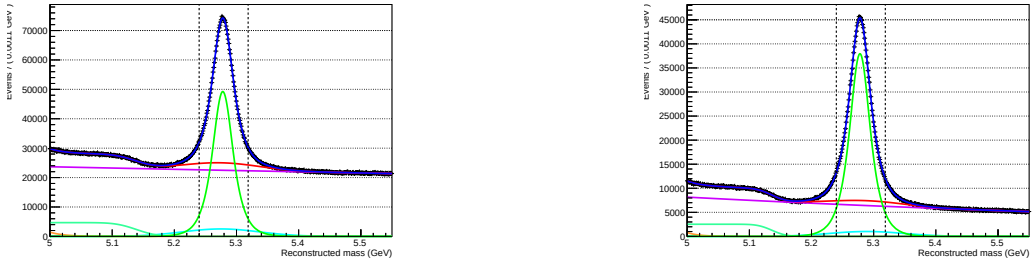
- $\cos\alpha < 0.99$, with α angle between the secondary vertex direction (computed as the 3d distance between pp interaction point and B decay vertex) and the total momentum of the tracks from the decay. This is a measurement of how much momentum we are missing from the decay, to notice potential missing tracks.

- Vertex fit probability $P > 10\,\%$.
- Distance from the beam axis greater than 3 times the error on it. Since B-meson have a relatively long mean life, with this cut we can reject combinatorial background from prompt J/ψ production.
- Total p_T of the meson greater than 10 GeV.
- Total p_T of the J/ψ from the decay greater than 8 GeV.
- Each μ from the J/ψ decay with p_T greater than 4 GeV and η smaller than 2.2.

Table 4.1 shows a summary of the efficiency of the cuts applied on the secondary vertex. The strict mass window in the last row is to reduce the bias due to the inclusion of events that do not contain the required meson but are accepted by our selections. The cut was applied using a window wide twice the sigma of the second largest gaussian in the Figure 4.1b fit.

Cut	Events	
SV reconstruction	$(6.2323 \pm 0.0007) \cdot 10^7$	
HLT DoubleMu4_JpsiTrk_Displaced	$(1.4290 \pm 0.0040) \cdot 10^7$	Figure 4.1a
secondary vertex quality	$(5.346 \pm 0.002) \cdot 10^6$	Figure 4.1b
strict mass window (0.04 GeV)	$(1.999 \pm 0.001) \cdot 10^6$	

Table 4.1: *Summary of selections applied to the single events in the datasets*



(a) *Before cuts* **(b)** *After cuts*

Figure 4.1: *Fully reconstructed B-meson mass, before and after the quality cuts applied. The vertical lines are the limits of the region used for the strict mass cut*

Further cuts are applied on the additional μ collected in the events. We require the μ to have $p_T > 4\,\text{GeV}$ and error on the impact parameter $E_{xy} < 0.005\,\text{cm}$. We can see in Table 4.2 the effect these selection make on signal or pileup μ. Additionally, the last rows show the abundance of signal and pileup μ, after selections, in the p_T and η windows we use during the analysis.

Cut	μ Signal	μ Pileup
$p_T > 4\,\text{GeV}$	$(6.39 \pm 0.02) \cdot 10^4$	$(1.66 \pm 0.02) \cdot 10^4$
After all previous cuts:		
barrel	$(3.84 \pm 0.06) \cdot 10^4$	$(8.24 \pm 0.09) \cdot 10^3$
endcap	$(2.55 \pm 0.02) \cdot 10^4$	$(8.36 \pm 0.09) \cdot 10^3$
$4\,\text{GeV} < p_T < 5\,\text{GeV}$	$(1.82 \pm 0.01) \cdot 10^4$	$(8.05 \pm 0.09) \cdot 10^3$
$5\,\text{GeV} < p_T < 7\,\text{GeV}$	$(2.03 \pm 0.01) \cdot 10^4$	$(5.46 \pm 0.07) \cdot 10^3$
$7\,\text{GeV} < p_T < 10\,\text{GeV}$	$(1.34 \pm 0.01) \cdot 10^4$	$(1.95 \pm 0.04) \cdot 10^3$
$p_T > 10\,\text{GeV}$	$(1.20 \pm 0.01) \cdot 10^4$	$(1.13 \pm 0.03) \cdot 10^3$

Table 4.2: *Effect of selections on signal data (center column) and pile up (right column)*

4.1. Preliminary checks on Monte Carlo sample

Before starting with the analysis it is necessary to check if the data from the Monte Carlo samples is a faithful representation for the data in the Charmonium dataset. It's particularly important in our case as we completely rely on the Monte Carlo models for the fit.

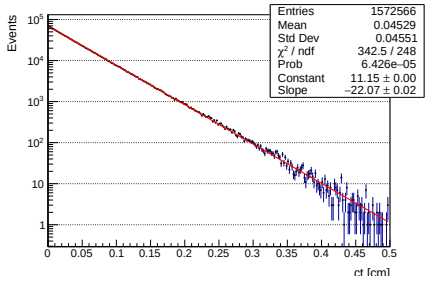
4.1.1 B-hadrons *ct*

We first check that the B meson are generated with the correct lifetime as reported in the Particle Data Book (PDG), by fitting the proper decay time for each B hadron, which can be easily computed from the generation information.

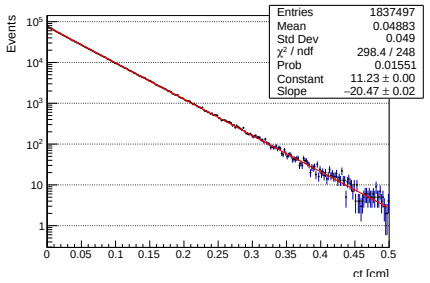
The algorithm iterated on all the B-hadrons in each event of the datasets and collected each decay time in histograms separed by particle type (including charge). The histograms were then fitted with an exponential distribution, and the parameters from the fit were compared to the tabulated mean life of the particle.

We exclude from each event the meson that would trigger our reconstruction, as its decay time would be biased.

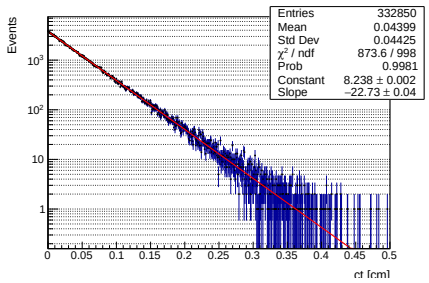
The fit results are nicely compatible with the world average for each B hadron considered, as shown in Figure 4.2, Figure 4.3 and Table 4.3, for both particles and antiparticles.



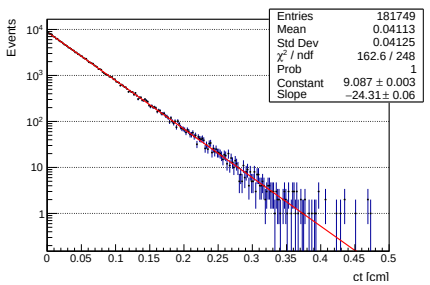
(a) B^0



(b) B^+

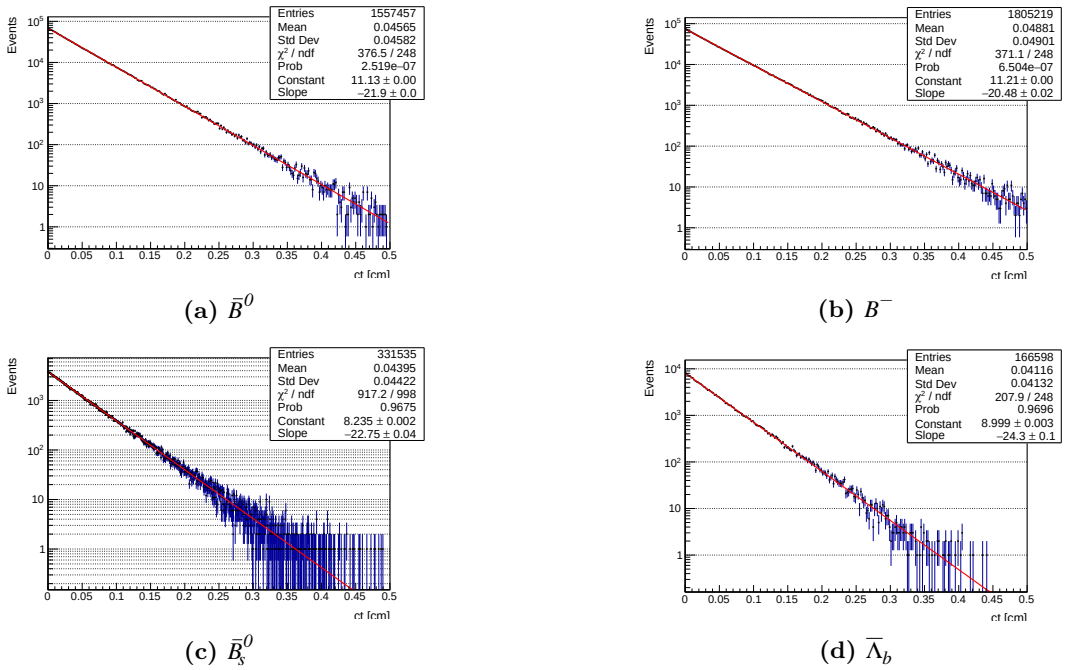


(c) B_s^0



(d) Λ_b

Figure 4.2: *Generated ct from Monte Carlo information for different B-hadrons*



(a) $\bar{B}^0$ **(b)** B^-

(c) $\bar{B}_s^0$ **(d)** $\overline{\Lambda}_b$

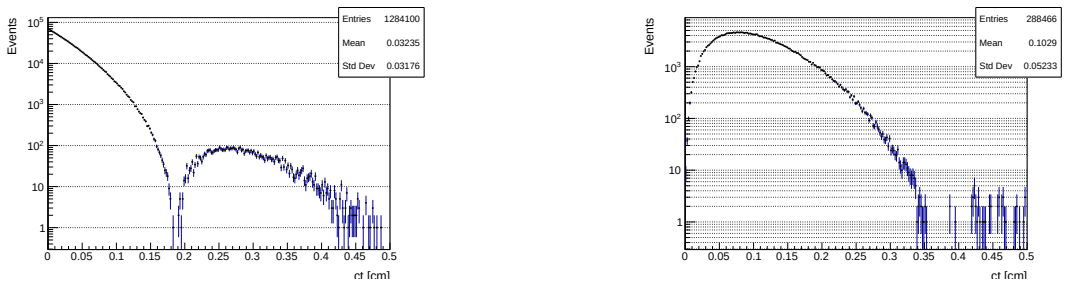
Figure 4.3: *Generated ct from Monte Carlo information for different anti B-hadrons*

Hadron	PDG [cm]	From fit [cm]
B^0	0.04557	0.0453 ± 0.00004
$\overline{B}^0$	0.04557	0.04566 ± 0.00004
B^+	0.04911	0.04885 ± 0.00004
B^-	0.04911	0.04883 ± 0.00004
B_s^0	0.04527	0.04399 ± 0.00008
$\overline{B}_s^0$	0.04527	0.04395 ± 0.00008
Λ_b	0.04395	0.04113 ± 0.0001
$\overline{\Lambda}_b$	0.04395	0.0412 ± 0.0001

Table 4.3: *Fitted ct versus the tabulated one, for different flavours of B-hadron*

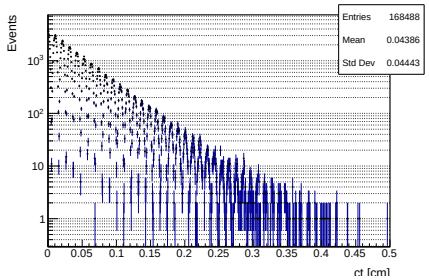
4.1.2 Flavour oscillations

We then test the oscillation frequency for B^0 and B_s^0 mesons, by comparing the proper decay time distributions for mixed and unmixed mesons, see Figure 4.4 for Bd and Figure 4.5 for Bs mesons.

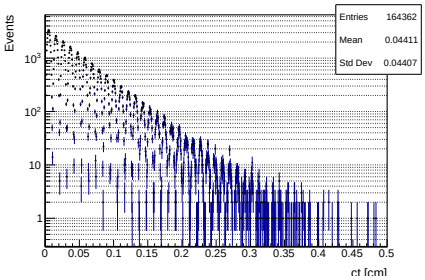


(a) *Unoscillated mesons* **(b)** *Oscillated mesons*

Figure 4.4: *Generated ct from Monte Carlo information, for meson that oscillated an even or odd number of times. In this example the* B^0 *mesons*



(a) *Unoscillated mesons*



(b) *Oscillated mesons*

Figure 4.5: *Generated ct from Monte Carlo information, for meson that oscillated an even or odd number of times. In this example the* B_s^0 *mesons*

We then compute the asimmetry between the two histograms. The asimmetry is defined as

$$H_{asymm} = \frac{H_1 - H_2}{H_1 + H_2} \tag{4.1}$$

We expect the histogram of unoscillated meson to b proportional to

$$H_{unosc} \propto \mathrm{e}^{-\frac{t}{\tau}} \cdot \frac{(1+\cos\omega t)}{2} \tag{4.2}$$

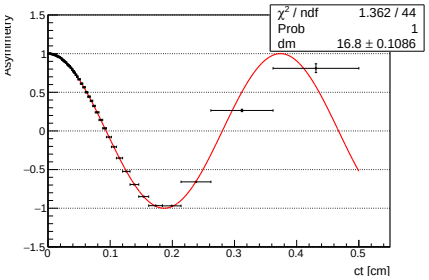
where τ is the mean life of the meson and ω is its oscillation rate, and the histogram of oscillated meson to be proportional to

$$H_{osc} \propto \mathrm{e}^{-\frac{t}{\tau}} \cdot \frac{(1-\cos\omega t)}{2} \tag{4.3}$$

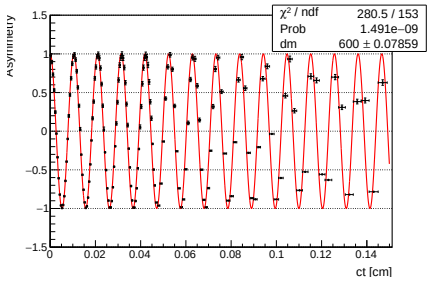
The asimmetry between the two is then

$$H_{asymm} = \frac{H_{unosc} - H_{osc}}{H_{unosc} + H_{osc}} = \frac{\frac{(1+\cos\omega t)}{2} - \frac{(1-\cos\omega t)}{2}}{\frac{(1+\cos\omega t)}{2} + \frac{(1-\cos\omega t)}{2}} = \cos\omega t \tag{4.4}$$

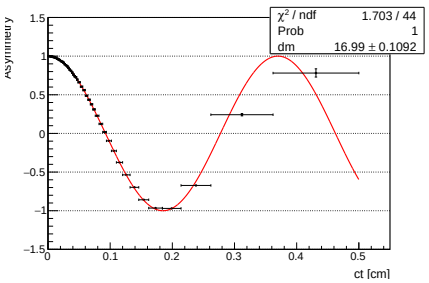
The resulting asimmetry histograms are showed in Figure 4.6, where a fit with a cosine is overlayed, showing the extremely good fit between data and model.



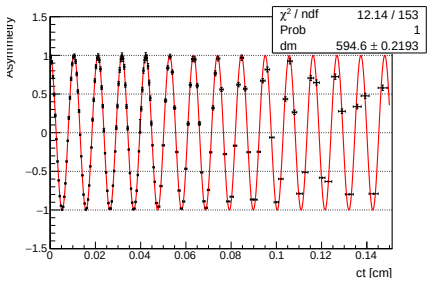
(a) B^0



(b) B_s^0



(c) $\bar{B}^0$



(d) $\bar{B}_s^0$

Figure 4.6: *Ratios between mixed and unmixed mesons as function of generated ct*

4.1.3 Separation of signal from pileup μ

For the analysis it is necessary to separe μ coming from pileup vertices from μ coming from the main one. This is accomplished using the μ track's longitudinal impact parameter D_z with respect to the main primary vertex. The main primary vertex was selected by choosing the one that minimized the difference between the decay total momentum and the vector from the candidate to the secondary vertex. During the analysis it was found that this method gave the best results in term of pileup rejection.

On the Monte Carlo samples, where we can indentify a track as being from pileup, we plotted the dz distribution in the two cases. The results can be seen in Figure 4.7.

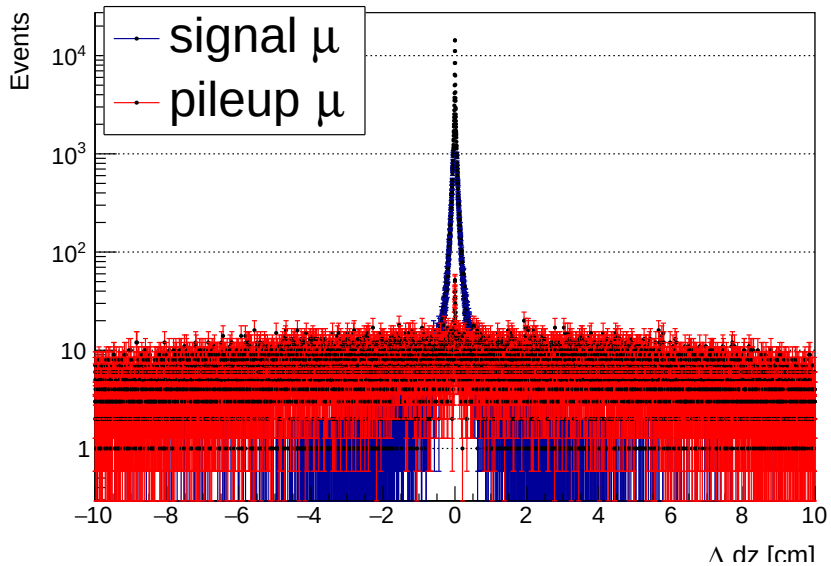


Figure 4.7: *Longitudinal impact parameter of the tracks with respect to the main primary vertex. In red pileup muon tracks and in blue signal*

As we can see, pileup tracks have a largely different distributions, with a much higher variance as we expected. Signal tracks instead have a much narrower distribution and are concentrated in an area of size $\sim$0.5 cm.

To select a cut to apply to the impact parameters of μ when working with real data we have to consider the signal efficiency, that is how much signal we are throwing away on average by narrowing the cut, and the false positive efficiency, how much background we are instead pulling in by enlarging it. A good way to display these values is the ROC curve, shown in Figure 4.8. The ROC curve plots the signal efficiency (on y) vs the pileup efficiency (on x). The steep rise at the beginning shows that by widening a little the cut window we can pull in most of the signal while leaving the pileup efficiency low.

The main efficiency values as function of the cut size are also shown in Table 4.4.

	Dz				
cut [cm]	Signal	Bkg	Sig/Bkg	Total Signal	Total Bkg
0.5	96.6%	6.2%	48.7	230660	74028
0.4	96.2%	5.1%	58.9	230660	74028
0.3	95.4%	4.0%	74.4	230660	74028
0.2	93.6%	2.9%	99.5	230660	74028
0.15	91.5%	2.4%	121	230660	74028
0.1	87.1%	1.8%	154	230660	74028
0.075	82.9%	1.4%	178	230660	74028
0.05	75.5%	1.1%	205	230660	74028

Table 4.4: *Summary of the results of various cut widths in*

The selected size for the cut was in the end chosen to be 0.4, which implies a signal efficiency of 96.2 % and a false positive efficiency of 5.1 %.

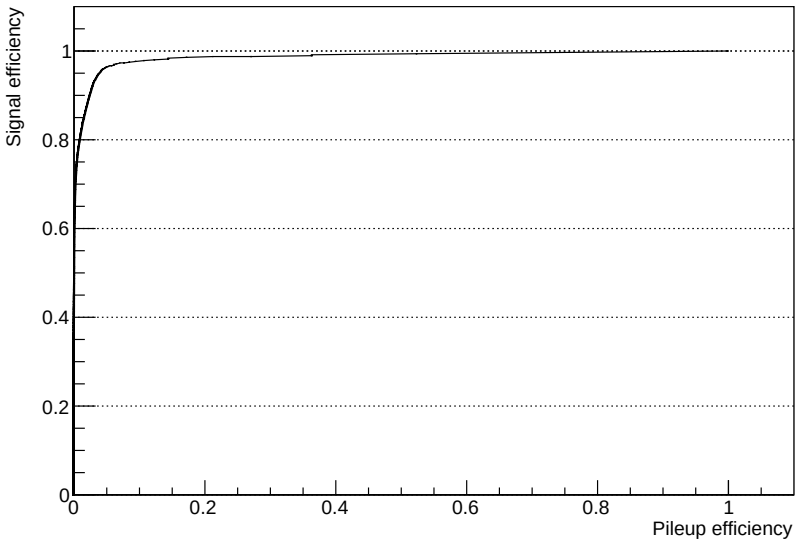


Figure 4.8: *Pileup vs signal efficiency, variying the cut width*

Chapter 5

Tagging variables

We present now some details on the variables that will be used as tag to select the μ that come from a B-meson decay.

5.1. Tag: impact parameter

The first tag used is the impact parameter (D_{xy} from now on) of the μ. We shall here explain that the μ impact parameter is correlated with the proper decay time of its ancestor particle. This makes it a good handle to recognize μ coming from a b decay, as B-mesons have a greater mean life than other lighter mesons.

5.1.1 Definition and modelization

Most of the analysis is based on a quantity called impact parameter. This is a reconstructed variable linked to tracks, commonly used in collider physics.

The impact parameter of a track is the distance of closest approach to some reference point, usually the point of interaction of the two protons that generated the particle or the beam spot (the center of the region where the two beams collide) of the event. It allows distinguishing, on a statistical basis, particles produced from the primary interaction vertex from those originated by the decay of a long-lived hadron, as explained below.

Choosing the reference point where the proton interacted is theoretically a better choice, as the equations that we will show in the next sections are simpler (they won't need to account for the vertex position relative to the beamspot). The problems with that approach is that the our best approximation of the interaction point is the primary vertex. Due to the algorithm used in CMS, the tracks used to fit the primary vertex tend to have a biased D_{xy} distribution, with an overabundance of tracks close to 0 and diminution of tracks with medium D_{xy} (Figure 5.1).

The formula for the impact parameter changes depending on how many long lived particles are present in the decay chain that led to the production of the particle associated to the track. We will first study the case where there is only one previous particle, coming directly from the pp interaction. This is the case for both the $b \to \mu$ and $c \to \mu$.

Single-particle decay

If a particle created in the primary vertex lives for a time t, its decay length in the lab frame is given by

$$l = \gamma\beta ct \tag{5.1}$$

where β and γ are the relativistic parameters (respectively $\beta = \frac{v}{c}$ e $\gamma = \frac{1}{\sqrt{1-\beta^2}} = \frac{E}{mc^2}$). If this particles decays, we define the trasversal impact parameter of any daughter particle with respect to the primary vertex as

$$D_{xy} = l \sin\theta \sin\psi = \gamma\beta ct \sin\theta \sin\psi = \frac{\gamma\beta ct \sin\psi}{\cosh\eta} \tag{5.2}$$

using η as defined before and ψ as the angle between the direction of the momentum of the mother (also called ancestor) particle and that of the daughter. The sign of the impact parameter is determined here by the sign of

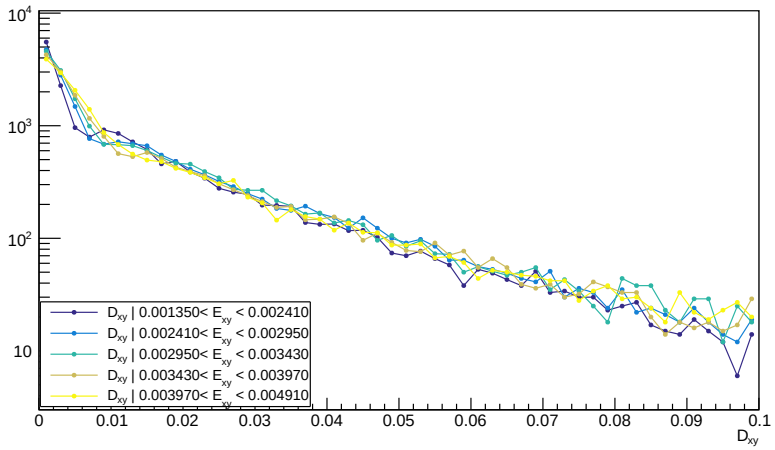


Figure 5.1: *Full range of D_{xy} showing the biased behaviour at low values. The slices are at different windows of E_{xy}. All the graphs are normalized to 1 to highlight the common behaviour*

ψ. This choice is different from what is commonly used in high energy physics, which usually relates the sign to apparent start point of the track, compared to the primary vertex. The advantage of using our version will be clear later, when analyzing the impact parameter of tracks coming from a longer decay chain.

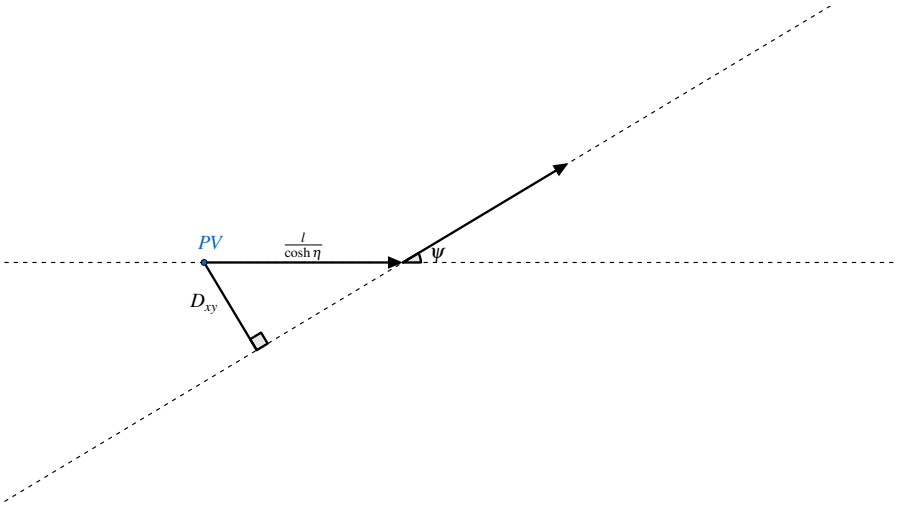


Figure 5.2: *Definition of the impact parameter of a track*

An important feature of the impact parameter, and the main motivation for our use ot it in the analysis, is that as the parent particle becomes highly relativistic, the daughter's impact parameter becomes insensitive to the parent particle momentum. This is caused by the cancellation between the decay length, which increases with higher momentum (the factor γ in Equation 5.2) and the decay angle, which decreases with higher momentum due to the lorentz boost.

Let's assume that the daughter particle's mass is small compared to that of the parent, a good approximation in our case, when the daughter particle is a μ and the parent is usually an heavy quark such as c or b. This assumption fails when treating the fake category, which, as we will see, shows a different distribution. Applying a boost to move from the parent particle (trasversal, since ψ lies on that plane) rest frame to the laboratory, the $\sin\psi$ factor transform as

$$\sin\psi = \frac{\sin\psi_{CM}}{\gamma_T\left(1+\beta_T\cos\psi_{CM}\right)}; \quad -\pi < \psi_{CM} < \pi \tag{5.3}$$

Here γ_T and β_T refer to the relativistic parameters in the trasversal plane ($\beta_T = \frac{v_T}{c}$ and $\gamma_T = \frac{1}{\sqrt{1-\beta_T^2}}$).

We can insert this inside Equation Equation 5.2, yelding

$$\begin{aligned} D_{xy} &= \frac{\gamma\beta ct \sin\psi_{CM}}{\cosh\eta\,\gamma_T\left(1+\beta_T\cos\psi_{CM}\right)} \\ &= \frac{\beta_T ct \sin\psi_{CM}}{1+\beta_T\cos\psi_{CM}} \end{aligned} \tag{5.4}$$

where in the last passage we used the fact that $\beta\gamma = \frac{p}{m}$ and $p_T \cdot \cosh\eta = p$.

As we can see, this expression depends very weakly on the momentum of the parent particle. Additionally we can see that the only dependence on the pseudorapidity is trough p_T and thus through β_T.

Let us now study the shape of the resulting distribution.

We can split Equation 5.4 in two factors, the mother particle ct and the p_T/angular part.

The distribution of the product z of two random variables x and y can be written as:

$$P_{z=xy}(z) = \int_{Dom(x)} P_x(x)\cdot P_y(z/x)\frac{1}{|x|}\,\mathrm{d}x \tag{5.5}$$

In our case, if we take $y = ct$ (exponentially distributed) and $x = \beta_T\gamma_T\sin\psi$ we can write

$$P_{D_{xy}}(D_{xy}) = \int_{Dom(x)} P_x(x)\cdot\frac{e^{-\frac{D_{xy}}{\tau\cdot x}}}{\tau\cdot|x|}\,\mathrm{d}x = \int_{Dom(x)} P_x(x)\cdot Expo(D_{xy};\,\tau\cdot x)\,\mathrm{d}x \tag{5.6}$$

where $Expo(x;\,\tau)$ means an exponential distribution. Here x may be negative and the mean life of the final exponential term shares the same property. This simply means that the exponential will span negative values of D_{xy}.

We need now to find and expression for P_x. From an analysis of $\frac{D_{xy}}{ct}$ in the MC samples we can see that the actual distribution has a gaussian-like distribution and that its shape does not depend on the decay considered (see Figure 5.3), with the only exception in the fake category. This last property allows us to consider the angular part of the distribution as a invariant in the fit, and instead use only the decay mean life as a parameter. For obtaining the proper parameters we fitted the shape with a sum of 2 gaussians centered in 0. It's important to note that the actual σ of the gaussian can be reabsorbed inside the τ parameter of the exponential. In the models we chose to leave it at 0.5, which is close to the fitted parameter.

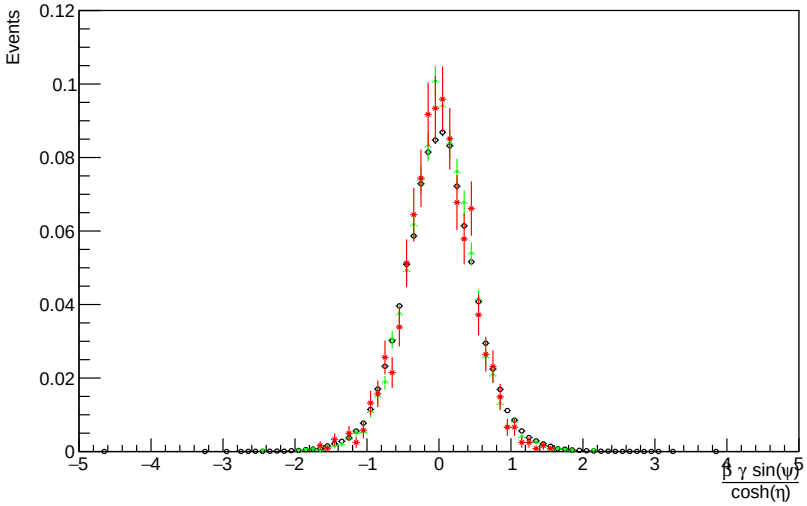


Figure 5.3: $\frac{\beta\gamma\sin\psi}{\cosh\eta}$ *for decays with only one intermediate particle. As we can see the distribution is extremely similar*

Because of simmetry, x's distribution has to be simmetrical with respect to 0. Since from Equation 5.6 the sign of x determines the sign of D_{xy}, the impact parameter distribution will be simmetrical too. Furthermore, since $|x|$ has no lower bound (and thus neither does $\tau\cdot x$), we expect $P_{D_{xy}}(D_{xy})$ to have some kind of cusp in $D_{xy} = 0$, a fact that is confirmed in the MC samples. An example of the resulting shape is shown in Figure 5.4, obtained from Monte Carlo.

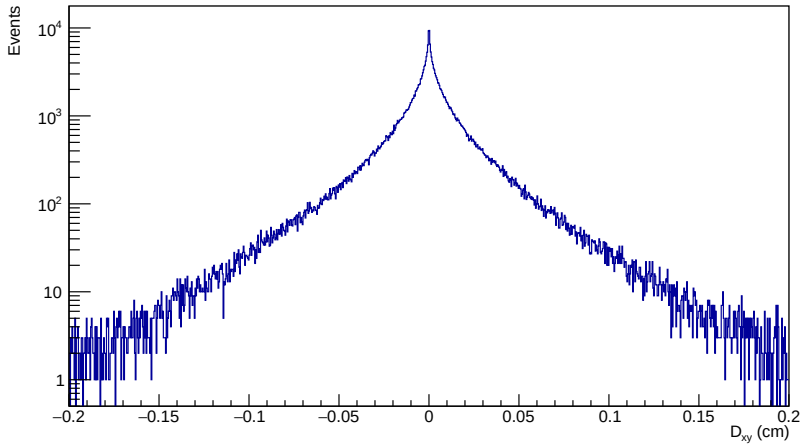


Figure 5.4: *Full generated* D_{xy} *distribution for bl muons*

Two-particle decay

The case with more than one particle in the decay chain is more complex. From a geometrical analysis of the event (Figure 5.5) we can see that each step of the decay contributes to the total D_{xy} with a term proportional to the length of its life and to the sinus of the angle between itself and the final track.

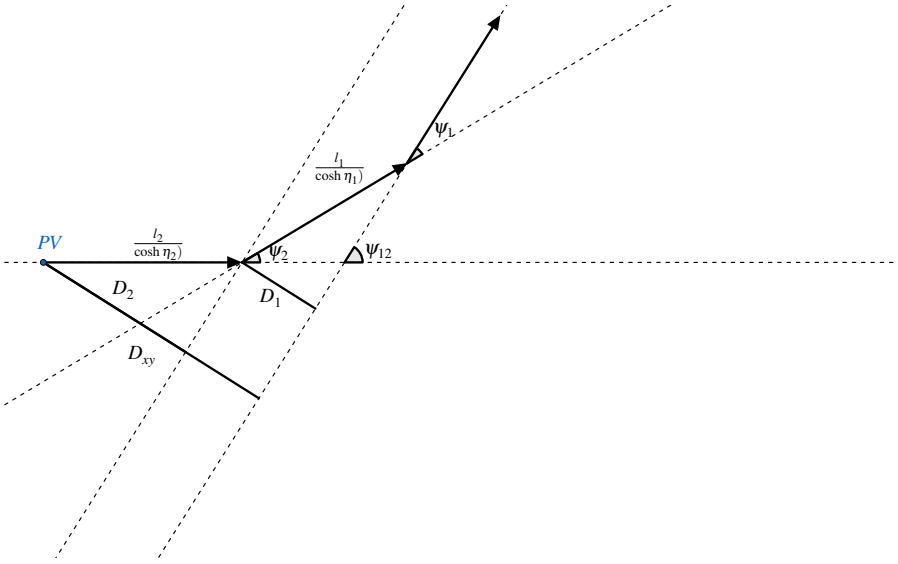


Figure 5.5: *dxy multi*

The total D_{xy} is then:

$$D_{xy} = \sum_i \frac{\beta_i \gamma_i c t_i \sin\psi_{if}}{\cosh\eta_i} \tag{5.7}$$

where i is the index over the various steps of the decay and ψ_{if} is the angle between the momentum of the step i of the decay and the track.

To understand the distribution of this version we will first analyze the distribution of the sum of two exponentially distributed variables.

$$P_{z=x+y}(z) = \left[P_x \otimes P_y\right](z) = \int_{-\infty}^{\infty} Expo(z-t;\,\tau_1)Expo(t;\,\tau_2)\,\mathrm{d}t \tag{5.8}$$

Here we used again *Expo* meaning an exponential distribution and $\tau_{1,2}$ are the parameters of the two distributions. For the result of the convolution we will consider only the case where $z > 0$, but an identically shaped tail is present on the other half of the axis. The equation for the $z < 0$ tail can be obtained simply flipping the sign of z, τ_1 and τ_2.

$$P_{z=x+y}(z) = \begin{cases} 0 & \tau_1 < 0 \wedge \tau_2 < 0 \\ \frac{e^{-\frac{z}{\tau_1}}}{\tau_1 - \tau_2} & \tau_1 > 0 \wedge \tau_2 < 0 \\ \frac{e^{-\frac{z}{\tau_2}}}{\tau_2 - \tau_1} & \tau_1 < 0 \wedge \tau_2 > 0 \\ \frac{e^{-\frac{z}{\tau_1}} - e^{-\frac{z}{\tau_2}}}{\tau_1 - \tau_2} & \tau_1 > 0 \wedge \tau_2 > 0 \end{cases} \tag{5.9}$$

We will call this distribution $ExpSum(x;\, \tau_1, \tau_2)$.

D_{xy} in this case is the sum of the two terms D_1 and D_2, thus the distribution is the convolution of the two. Each of the two terms is distributed as in Equation 5.6 (with P_x yet to be determined in this case), so the full equation is

$$\begin{aligned} P_{D_{xy}}(D_{xy}) &= \left[P_{D_1} \otimes P_{D_2}\right](D_{xy}) \\ &= \int_{Dom(x_1)} \int_{Dom(x_2)} P_{x_1}(x_1) \cdot P_{x_2}(x_2) \cdot \left[Expo(\tau_1 \cdot x_1) \otimes Expo(\tau_2 \cdot x_2)\right](D_{xy})\ \mathrm{d}x_1\ \mathrm{d}x_2 \\ &= \int_{Dom(x_1)} \int_{Dom(x_2)} P_{x_1}(x_1) \cdot P_{x_2}(x_2) \cdot ExpSum(D_{xy};\, \tau_1 \cdot x_1, \tau_2 \cdot x_2)\ \mathrm{d}x_1\ \mathrm{d}x_2 \end{aligned} \tag{5.10}$$

where x_1 and x_2 are the angular terms in Equation 5.7.

In both steps the distributions of x are similar to the ones found in Figure 5.1.1. This time, this property is shared by the fake category too. A comparison can be found in Figure 5.6 for the last step and Figure 5.7 for the other.

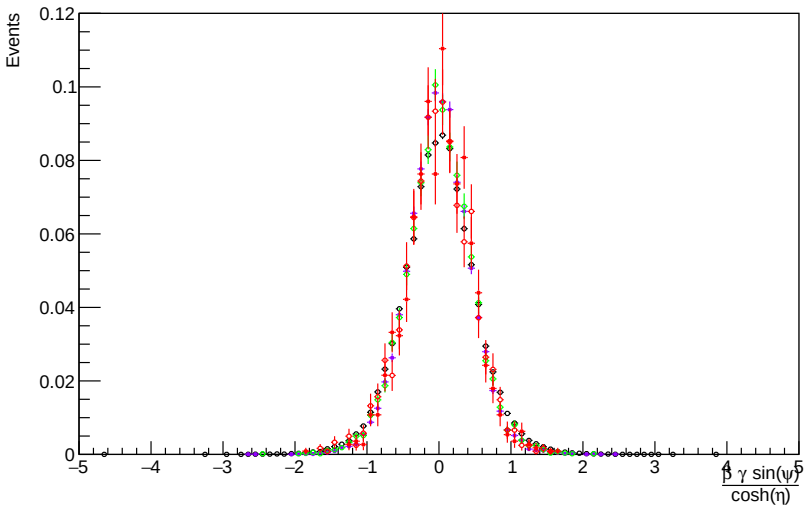


Figure 5.6: $\frac{\beta\gamma\sin\psi}{\cosh\eta}$ *for decays with two intermediate particles (full circles) compared to the ones with only one (empty circles). The colors follow the conventions explained before. Only the last step of the decay is considered (the one closest to the track)*

A full distribution of D_{xy} can be seen in Figure 5.8, obtained from data from MC samples.

Other kind of decays

Almost none of our reconstructed muons show a 3 or 4 steps decay, to the point of sheer negligibility.

A good share of the fake category (about 75 %) comes instead directly from the primary vertex. These "muons"' impact parameter is distributed as a Dirac's delta and its only parameter is thus the resolution.

5.1.2 Impact parameter resolution

After modeling the distribution of D_{xy} in our categories, we want now to account for the effects of resolution.

First we want to check if the track D_{xy} coming from reconstruction is well described by the formulas shown in subsection 5.1.1. For this we have two ways:

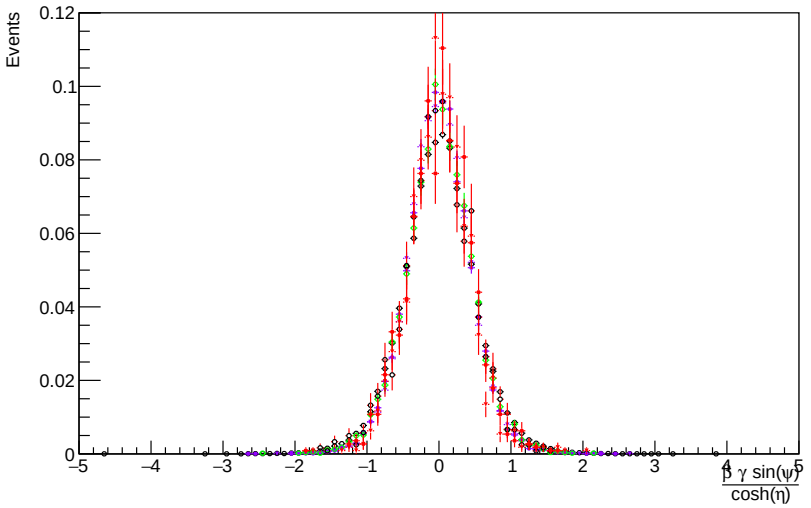


Figure 5.7: $\frac{\beta\gamma\sin\psi}{\cosh\eta}$ *for decays with two intermediate particles (full circles and full triangles) compared to the ones with only one (empty circles). The colors follow the conventions explained before. In case of multiple intermediate particles all steps are considered*

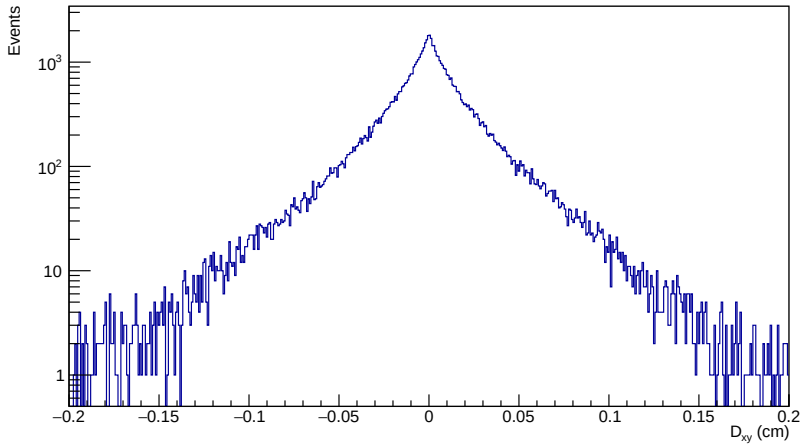


Figure 5.8: *Full generated* D_{xy} *distribution for bxl muons*

- a 2D plot of the B hadron proper decay time versus the μ impact parameter, to evidence the correlation between the two.

- 1D plots of the difference between the two variables, optionally normalized over the precision in the impact parameter measurement (provided by the reconstruction). The resulting distribution should be centered around 0 and if normalized it should be a standard normal distribution.

The plots are made iterating over each μ in the events and extracting its ancestor using the generation information.

The 2D plot is shown in Figure 5.9, in double logaritmic axis.

The plots for the difference of the two variables are shown in Figure 5.10, and normalization is added in Figure 5.11. The normalized plots show a good accord with the standard normal distribution model, but with longer tails, demonstrating that there is part of the error that is not accounted for. Our hypothesis is that the extra error comes from the distribution of the interaction points around the beam spot, but this is not easily verified, as we have no information on its distribution other that the PV distribution, which is affected by the track reconstruction precision.

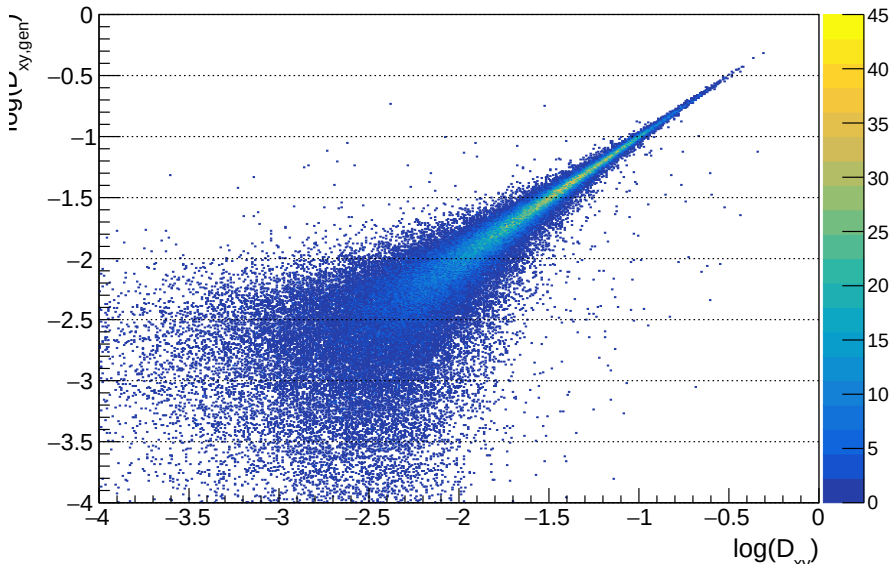
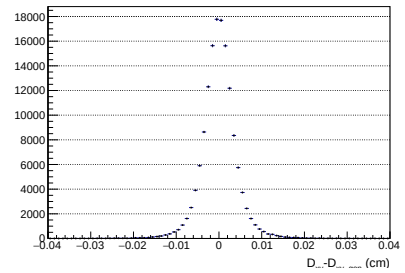
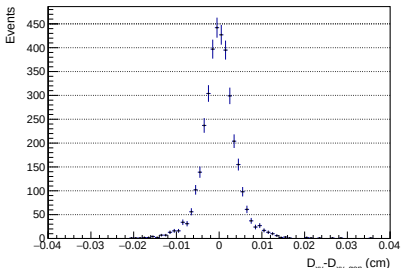


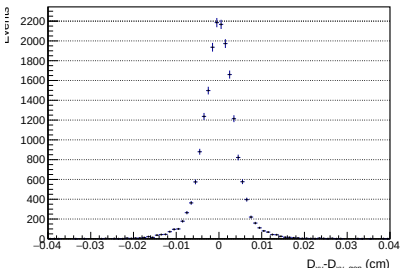
Figure 5.9: *2D plot of* $\ln(D_{xy})$ *vs* $\ln(ct)$



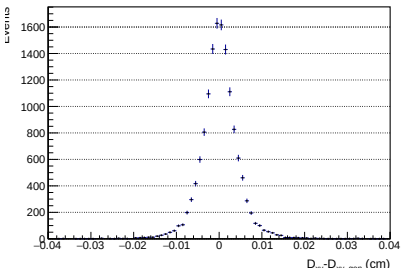
(a) $b \to \mu$



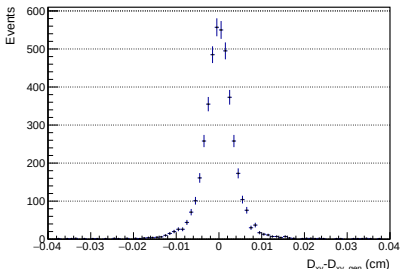
(b) $b \to \tau \to \mu$



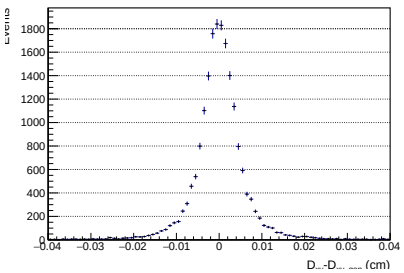
(c) $b \to \bar{c} \to \mu$



(d) $b \to c \to \mu$

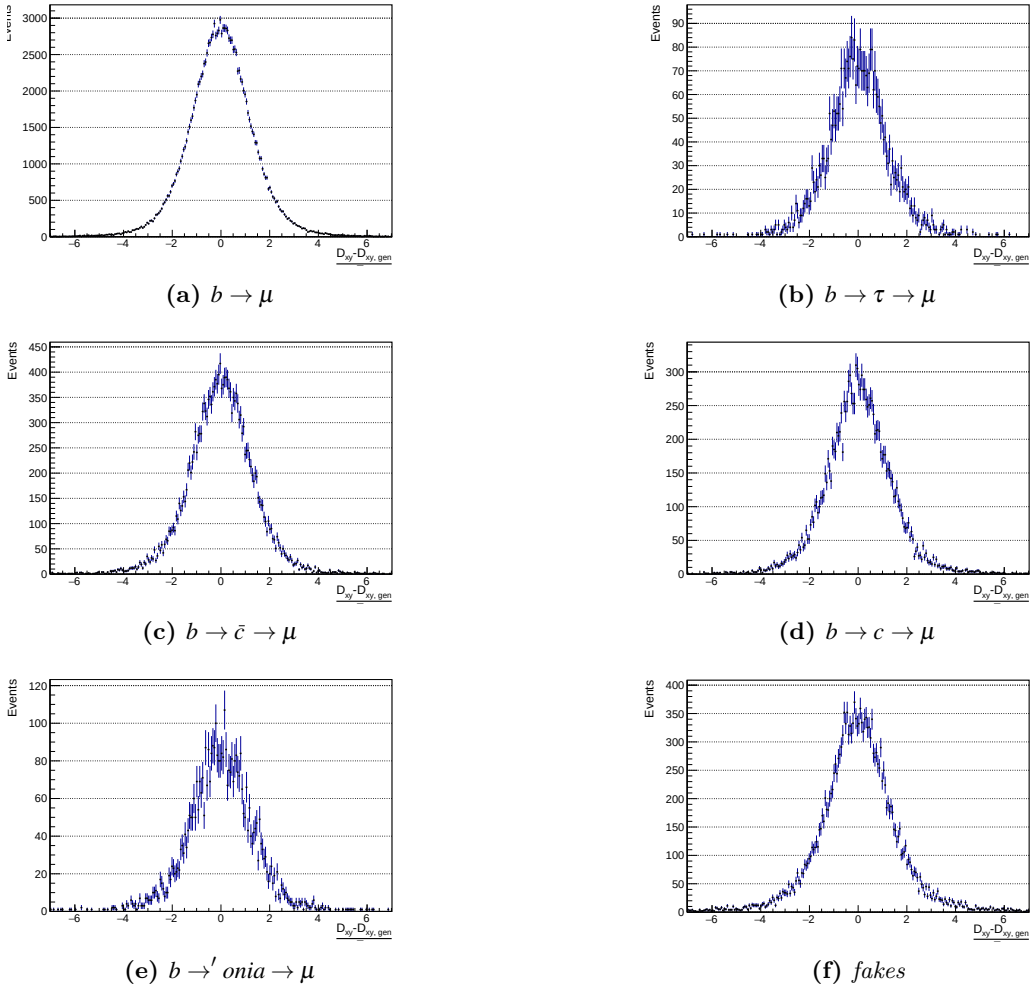


(e) $b \to ' onia \to \mu$



(f) *fakes*

Figure 5.10: *Difference between the muons impact parameter and its ancestor's decay time, projected on the xy (trasversal) plane*



(a) $b \to \mu$ **(b)** $b \to \tau \to \mu$

(c) $b \to \bar{c} \to \mu$ **(d)** $b \to c \to \mu$

(e) $b \to' onia \to \mu$ **(f)** *fakes*

Figure 5.11: *Difference between the muons impact parameter and its ancestor's decay time, projected on the xy (trasversal) plane, normalized to the impact parameter error*

To find the dependency between the error on D_{xy} and the actual resolution we make a plot of the latter as a function of the former. This plot is obtained using the standard deviation in Figure 5.10 divided in E_{xy} slices. The slices can be seen in Figure 5.12, while the σ can be seen in Figure 5.13.

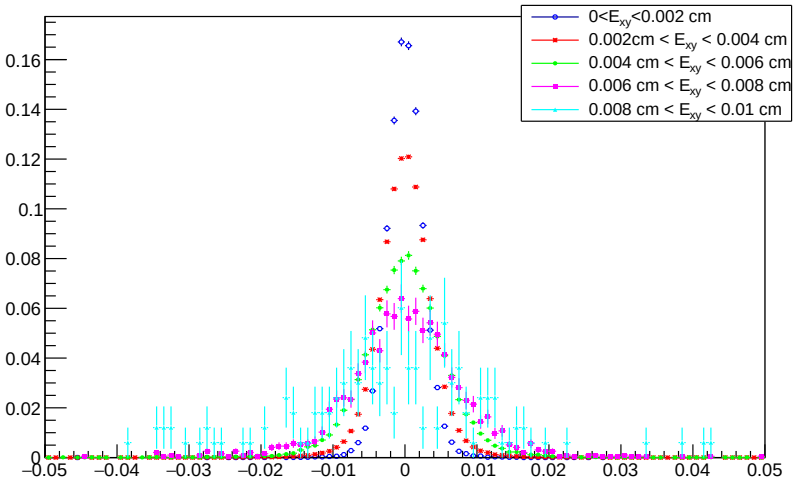


Figure 5.12: *Distribution of* $D_{xy} - D_{xy,gen}$ *in slices of* E_{xy}

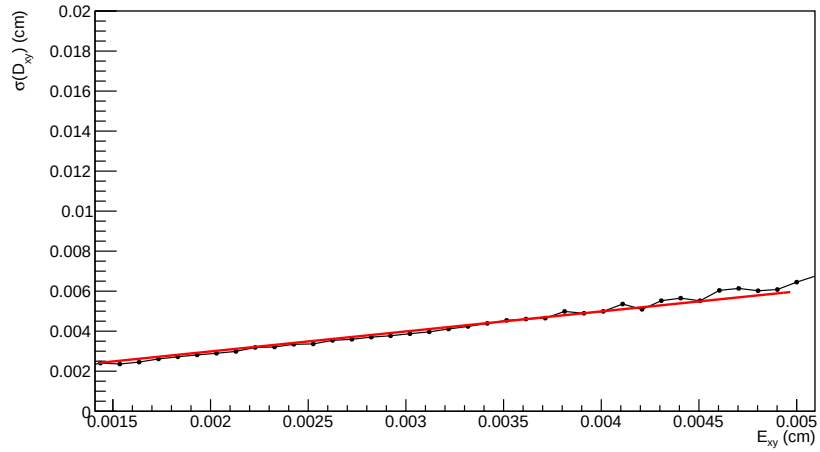


Figure 5.13: *Standard deviation of* $D_{xy} - D_{xy,gen}$ *as a function of* E_{xy}

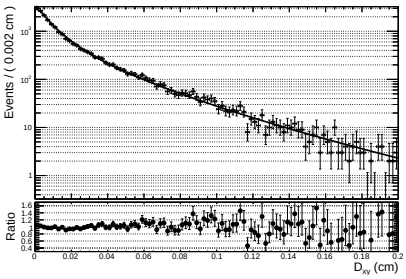
5.1.3 Use as a tagging variable

We want to fit the data with a distribution that is the sum of 4 mutually exclusive processes, corresponding to the decay categories we have talked about in the previous sections. Each category has a particular shape that allow us to recognize it among the others. Ideally we would like the D_{xy} distributions to be as different as possible among the categories, to reduce the correlations between the linear combination parameters in the total model.

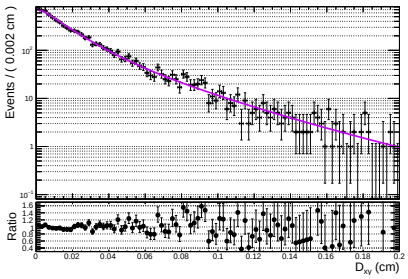
The models are obtained from the Monte Carlo samples, with a binned fit over the D_{xy} distributions. Every distribution is fitted with the proper model from subsection 5.1.1. For the bl and cl categories we used a single intermediate particle decay, for bxl we used a two intermediate particles decay and for fakes we used a sum of single particle, double particle and 0 particles.

In the models we are forced to assume that the distribution is the same both for signal and pileup μ as our model samples does not contain generation informations for the pileup μ.

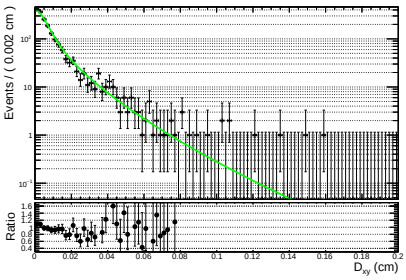
Figure 5.14 shows the final fitted models in the merged dataset, with the fit superimposed.



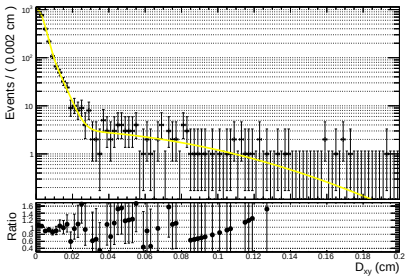
(a) *Direct b muons*



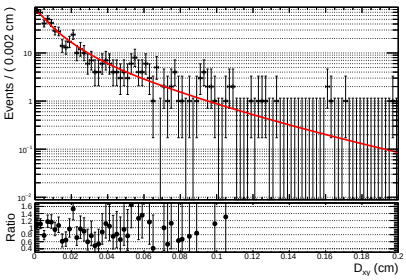
(b) *Indirect b muons*



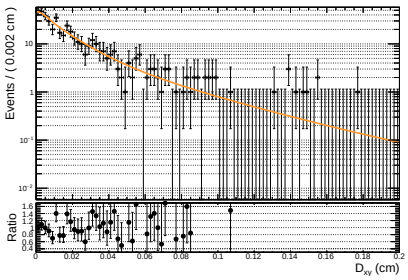
(c) *Muons from c decays*



(d) *Fake muons with no long lived ancestors*



(e) *Fake muons with one long lived ancestor*



(f) *Fake muons with two long lived ancestors*

Figure 5.14: *Impact parameter models for the 4 categories*

Parameter	Sample			
	All	B^0	B^+	B_s^0
Direct: τ	(0.0481 ± 0.0005) cm	(0.0476 ± 0.0008) cm	(0.048 ± 0.001) cm	(0.0487 ± 0.0007) cm
Indirect: τ_1	(0.018 ± 0.002) cm	(0.016 ± 0.003) cm	(0.011 ± 0.005) cm	(0.022 ± 0.004) cm
Indirect: τ_1	(0.049 ± 0.002) cm	(0.051 ± 0.003) cm	(0.054 ± 0.004) cm	(0.046 ± 0.003) cm
From charm: τ	(0.0218 ± 0.0007) cm	(0.021 ± 0.001) cm	(0.023 ± 0.002) cm	(0.022 ± 0.001) cm
Fake: $\tau_{chain=1}$	(0.054 ± 0.003) cm	(0.038 ± 0.004) cm	(0.10 ± 0.02) cm	(0.040 ± 0.004) cm
Fake: $\tau_{1,chain=2}$	(0.007 ± 0.007) cm	(0.00 ± 0.06) cm	(0.00 ± 0.06) cm	(0.02 ± 0.01) cm
Fake: $\tau_{2,chain=2}$	(0.058 ± 0.005) cm	(0.055 ± 0.006) cm	(0.04 ± 0.01) cm	(0.06 ± 0.01) cm
PV resolution, comp1	0.00039 ± 0.00010	0.0004 ± 0.0001	0.0006 ± 0.0004	0.0003 ± 0.0001
PV resolution, comp2	0.0091 ± 0.0007	0.010 ± 0.001	0.008 ± 0.002	0.0080 ± 0.0008
PV resolution, comp3	0.071 ± 0.006	0.08 ± 0.01	0.09 ± 0.02	0.059 ± 0.006
PV model, comp1 fraction	0.75 ± 0.03	0.79 ± 0.04	0.7 ± 0.1	0.70 ± 0.05
PV model, comp2 vs comp3	0.20 ± 0.03	0.19 ± 0.04	0.2 ± 0.1	0.19 ± 0.04

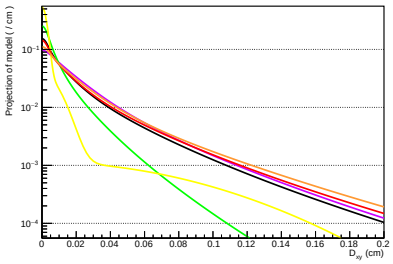
Table 5.1: *Comparison of the parameter from the fit of the three used Monte Carlo samples*

The fake category with no long lived ancestors is particularly important. As this category has the impact parameter equal to 0, the distribution we are fitting is actually only the resolution model.

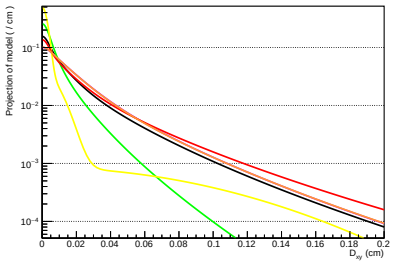
Other things to note is the similarity between the bl and bxl categories and the similarity between fakes and muons from b. The former is caused by the shorter mean life of the c mesons: the distribution of sum of two exponentials converge to the longest of the two after the variable is more than some mean-lives of the shortest. The latter is instead caused by the composition of the fake category: a large part of this sample is composed

by mesons coming from decay with means lives similar to those of bl or bxl muons and the impact parameter is not sensitive to the flavour of the particle that creates the track. While the impact parameter is similar, we can't say the same for the rest of the event: this "muons" have very different spectrums in p_T, η, quality of the track and isolation, allowing us to treat them as a separate category.

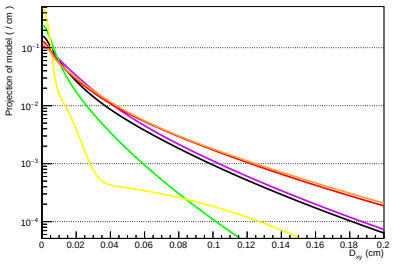
For reasons that will be clear later (section 5.2), we will perform the final fit on windows of p_T and η. We cannot be sure that the distributions stay the same in each window, so the models are interpolated separately. The windows used are the one already introduced in Table 4.2. Figure 5.15 and Figure 5.16 show all of the models, with each canvas representing one particular selection window.



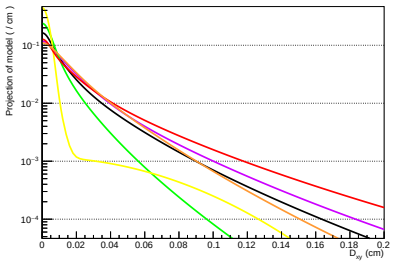
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *muons in barrel*



(b) $5\,\text{GeV} < p_T < 7\,\text{GeV}$

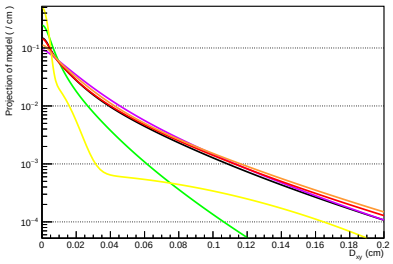


(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$

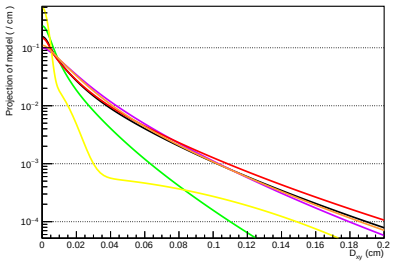


(d) $p_T > 10\,\text{GeV}$

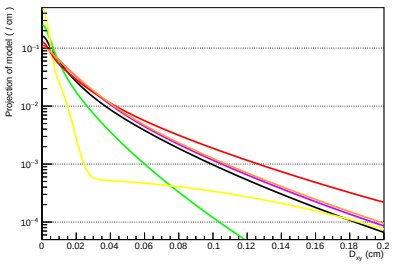
Figure 5.15: *Models of the* D_{xy} *distribution, from Monte Carlo samples, barrel region*



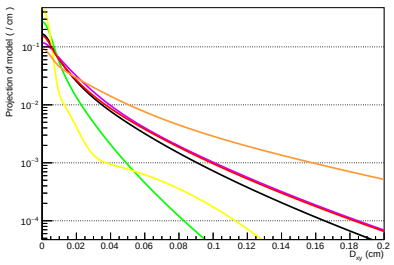
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$



(b) $5\,\text{GeV} < p_T < 7\,\text{GeV}$



(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$



(d) $p_T > 10\,\text{GeV}$

Figure 5.16: *Models of the* D_{xy} *distribution, from Monte Carlo samples, endcap region*

It is important to note that while the distributions are different enough for a fit to converge (aside for the aforementioned signal categories), the correlations are still big enough to generate large uncertainties in the results of the fit. This problem will be solved with the help of the additional tagging variables.

5.2. Tag: fake rejection MVA

The second tag used is what we will refer here as "fake rejection MVA". This MVA was developed in the Padova CMS group to reduce the occurrence of fakes in a separate analysis, but showed great potential for us to help our main fit converge.

This variable represents the output of a Deep Neural Network (DNN) trained to reject fake μ based on the quality of the track fit. In particular, this MVA uses as input variables:

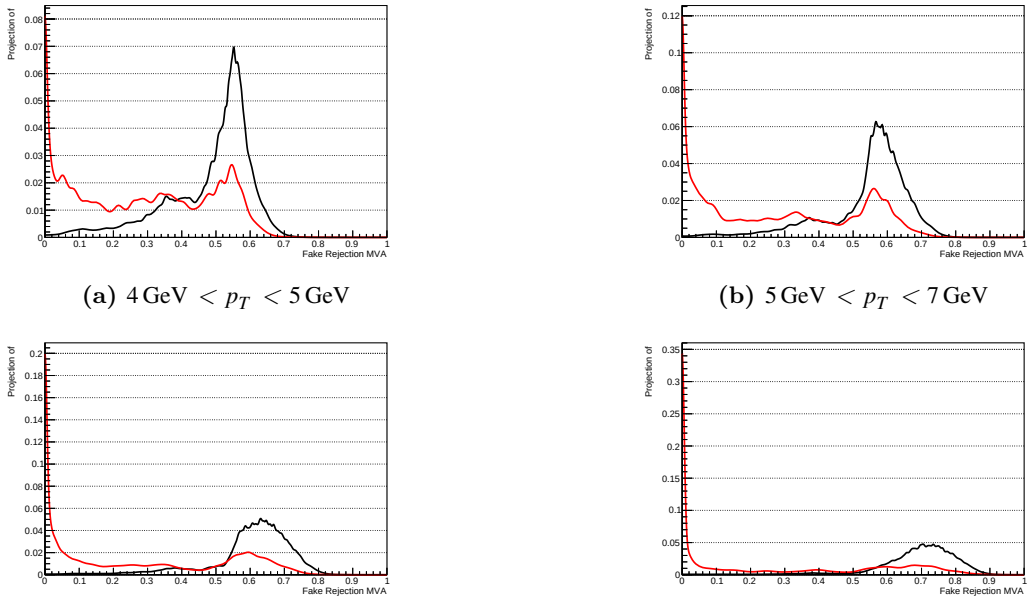
- Quality of track fit
  - in the μ chambers
  - in the tracker
  - matching of the two
- muon p_T
- muon η

More detailed informations, as well as the distributions of the input variables for fake and real μ, can be found in the Appendix A.

Since the CMS detector reacts quite differently wheter the μ interacts with the barrel or endcap sections, the MVA was trained separating the μ based on η and trained independently. Furthermore, since p_T is an input and we can't assume the same p_T distribution for data and models, to be able to get a good match in the distributions we are forced to fit in p_T windows, as it was already shown in the section 5.1. This leaves us with 8 different distributions.

For this variable there was no obvious analytical representation, so we opted to use directly the MC distributions as a model to be fitted on the data. Again, we searched the MC samples for difference in the distributions, and in the end merged the samples for the final model. We applied a variable kernel density estimation ([8]) to distribution from the Monte Carlo sample.

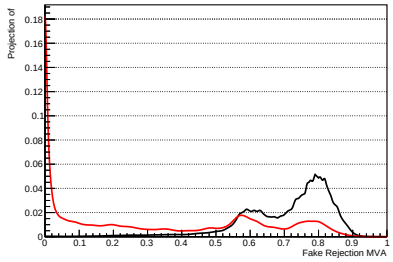
Figure 5.17 and Figure 5.18 show all the models for this variable.



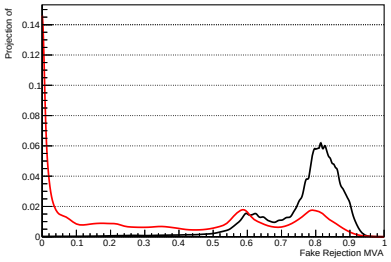
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$ **(b)** $5\,\text{GeV} < p_T < 7\,\text{GeV}$

(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$ **(d)** $p_T > 10\,\text{GeV}$

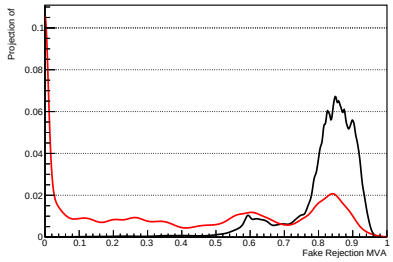
Figure 5.17: *Models of the Fake Mva distribution, from Monte Carlo samples, barrel region*



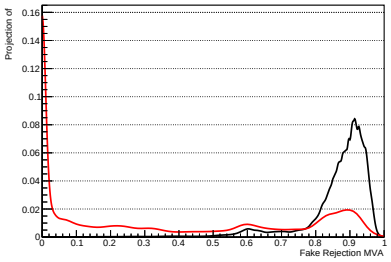
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$



(b) $5\,\text{GeV} < p_T < 7\,\text{GeV}$



(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$



(d) $p_T > 10\,\text{GeV}$

Figure 5.18: *Models of the Fake Mva distribution, from Monte Carlo samples, endcap region*

From the fit of this variable it's possible to provide an estimation of the number of the fake μ in the sample which is more precise than what we can get from the impact parameter alone. This number is used to place a constraint on both the fake and real μ weight

5.3. Tag: isolation

In our analysis of the MC samples we noticed that the isolation appeared to be a good tag to select direct μ from other μ categories.

Isolation is a variable quantifying how many separate tracks are found around the μ from its same PV, weighted by their momentum. It is defined as

$$I = \frac{p_T}{\sum_{i \in dR < 0.4, i \neq \mu} p_{T,i}} \tag{5.11}$$

where the summation is performed over the tracks that are inside what we call a $\Delta\eta\Delta\varphi$ cone of 0.4 width from the μ, meaning that

$$\Delta R(i,\mu) = \sqrt{(\eta_i - \eta_\mu)^2 + (\varphi_i - \varphi_\mu)^2} < 0.4$$

.

When there are no tracks inside the cone, ths isolation diverges and the μ is said to be fully isolated.

As we can see in Table 5.2, isolated μ are naturally enriched in signal. S/N ratio can be improved in the non isolated data by considering further tags. We need now a way to give the same estimation for non isolated μ.

Decay cat	Iso fraction	Comp. of iso sample	Comp. of MC sample
$b \to \mu$	0.133	0.871	0.689
$b \to X \to \mu$	0.0373	0.0627	0.176
$c \to \mu$	0.0577	0.0282	0.0511
fake	0.0485	0.0385	0.0833

Table 5.2: *Informations about fully isolated muons. First column is the fraction of muons in that category which are isolated, second is the composition of the isolated sample and the third is the composition of the full sample, for comparison.*

The chosen method was to use two additional variables, the cone charge Q

$$Q = \frac{\sum_{i \in dR<0.4} q_i \cdot p_{T,i}^{1.5}}{\sum_{i \in dR<0.4} p_{T,i}^{1.5}} \tag{5.12}$$

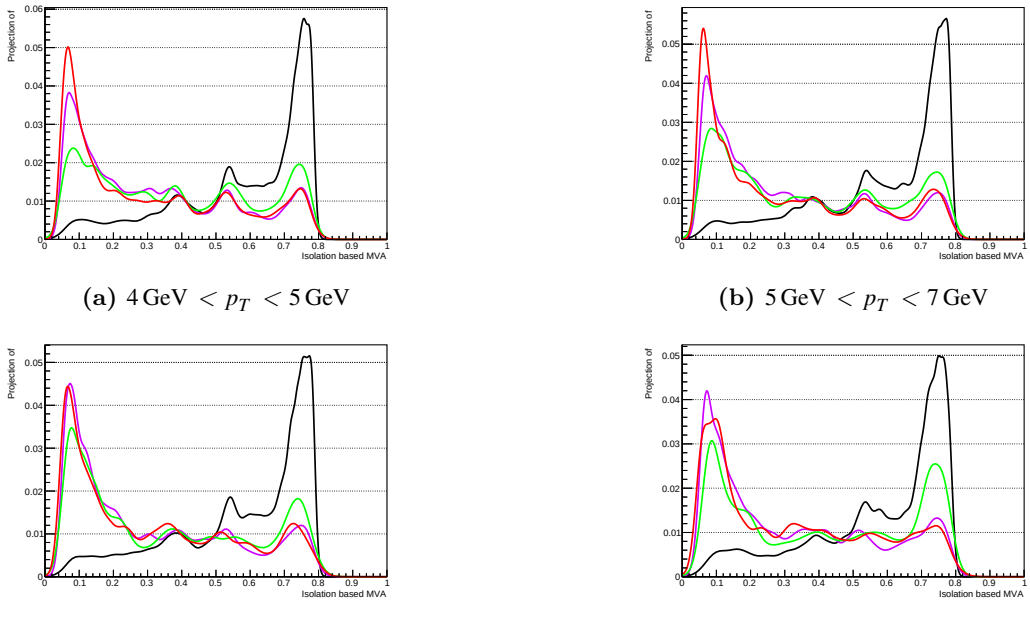
and the cone ΔR

$$\Delta R = \Delta R \left(p_\mu, \sum_{i \in dR<0.4, i \neq \mu} \vec{p}_i \right) \tag{5.13}$$

These variables, together with the isolation defined in Equation 5.11 are used as input for a second MVA method, another DNN. The DNN was trained on the entirety of the $B^0 \to \mu^- \mu^+ K^*$ sample, selecting only muons not coming from the decay B^0 decay, non-fully isolated, with $p_T > 4\,\text{GeV}$. After all the cuts, we used half of the remaining μ for the training sample and half of them for the test sample, where we verified tha absence of overtraining. The training used the direct μ as signal and the other categories as background, as this was the most important distinction that we were able to find in the input variables (Q and ΔR show extremely similar distributions for indirect μ, charm μ and fakes).

The distribution of the input variables is show in section A.1.

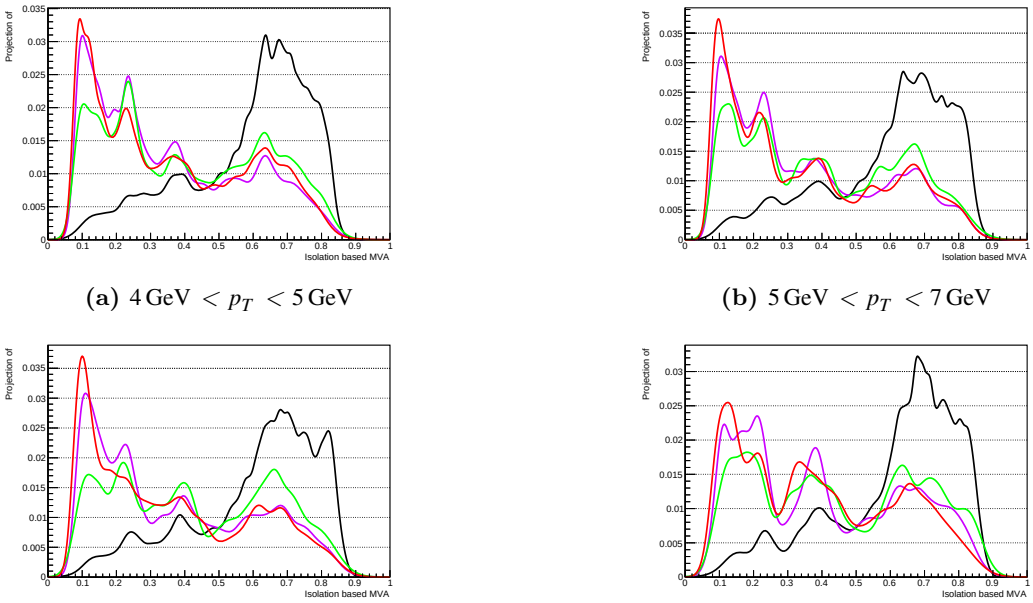
The models are obtained from data using the same process defined in section 5.2, and are shown in Figure 5.19 and Figure 5.19.



(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$ **(b)** $5\,\text{GeV} < p_T < 7\,\text{GeV}$

(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$ **(d)** $p_T > 10\,\text{GeV}$

Figure 5.19: *Models of the Isolation Mva distribution, from Monte Carlo samples, barrel region*



(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$ **(b)** $5\,\text{GeV} < p_T < 7\,\text{GeV}$

(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$ **(d)** $p_T > 10\,\text{GeV}$

Figure 5.20: *Models of the Isolation Mva distribution, from Monte Carlo samples, endcap region*

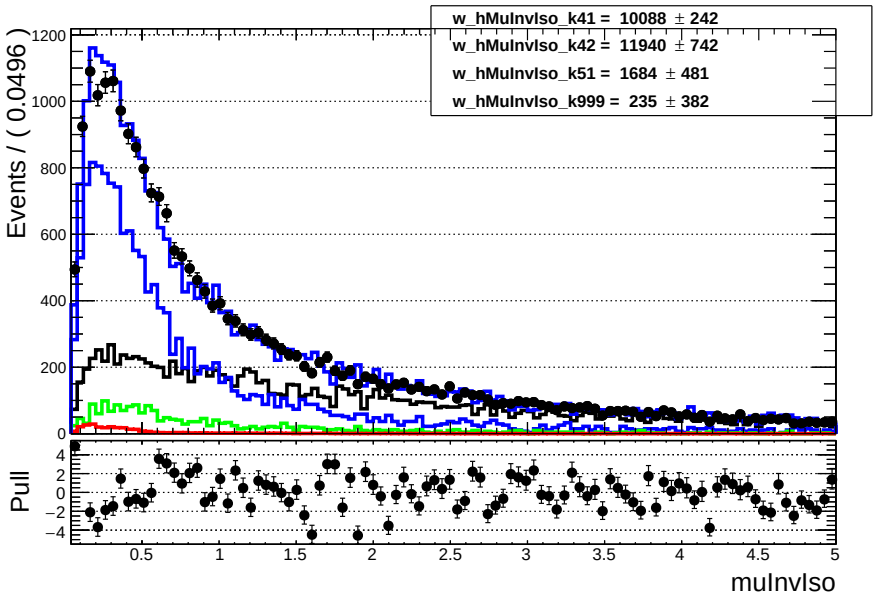


Figure 5.21: *Example of fit using the isolation variable I defined before.*

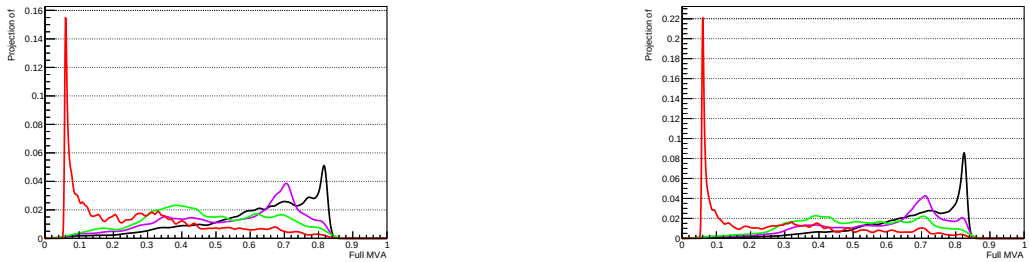
An important fact to note is that the isolation from Equation 5.11 has separate distributions in each of the 4 categories (see Figure 5.21 as an example), a fact that is propagated inside the MVA, even if the DNN was trained with only two super-categories, direct μ and the rest. This distinction was found to be necessary when fitting, as the composition of what in this MVA is called background (indirect b μ, c μ and fakes) is not the same in the data or Monte Carlo, leading to a different distribution for the super-category. The output of the MVA can be thus used similarly to the impact parameter, returning an estimante of every single component of the distribution. On real data though, the high correlation between the background components stopped us from relying on that information, using the fit, together with the full isolated μ information, only as a starting point for the subsequent one and as a constraint for weights of the direct component and the rest of the μ.

5.4. Tag: Full variables MVA

To summarize the two previous tags we developed a third MVA model, containing all the variablesused before, with the addition of D_{xy} and E_{xy}. This MVA has in theory enough informations to discriminate all categories, thus it was trained for a multiclass classification. The output of such a training is a set of 4 (our number of classes) variables, which add up to 1, each able to select the corresponding category. The tag we used was a linear combination of these 4 outputs, also trained to have a 4-modal distribution.

Since the output often presented a singular point when D_{xy} was low (in particular, lower than E_{xy}), due to the maximum in 0, we replaced this variable with the more well behaved $\log D_{xy}$. This managed to remove the singularity, leading to a smoother distribution.

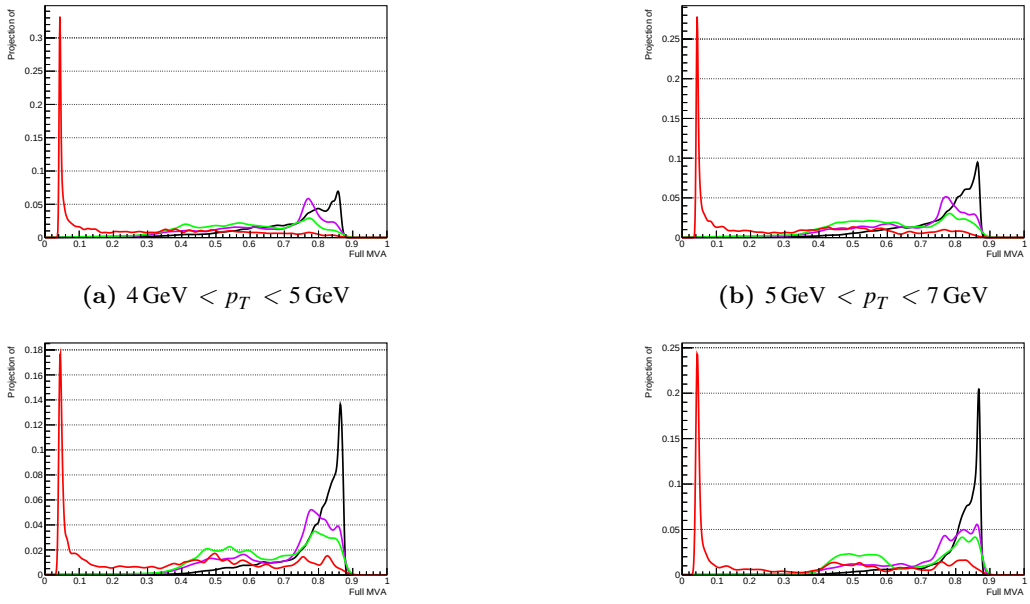
The model distributions from MC can be found in Figure 5.22 for the barrel and Figure 5.23 for the endcap.



(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$ **(b)** $5\,\text{GeV} < p_T < 7\,\text{GeV}$

(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$ **(d)** $p_T > 10\,\text{GeV}$

Figure 5.22: *Models of the Isolation Mva distribution, from Monte Carlo samples, barrel region*



(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$ **(b)** $5\,\text{GeV} < p_T < 7\,\text{GeV}$

(c) $7\,\text{GeV} < p_T < 10\,\text{GeV}$ **(d)** $p_T > 10\,\text{GeV}$

Figure 5.23: *Models of the Isolation Mva distribution, from Monte Carlo samples, endcap region*

Chapter 6

Analysis

We will now explain the various steps of the analysis that were executed.

6.1. Fit method

All the fits presented in the next sessions are done minimizing a binned likelyhood, using the models obtained in chapter 5. The possibility of an unbinned fit was considered but in the end it was discarded due to the high computational time that the fit was taking. The main free parameters used in the fits are the weights of the various components of the distribution.

For signal μ it is enough to add the parameter as coefficient in a linear combinations of the model. The result of the fit outputs the weight of our distribution in the sample.

For pileup μ we need to account for another problem: the parameter we need as output for the fit is not the total number of μ, but the total number of μ normalized by the number of primary vertices in the event (see Equation 3.5). The first method that was considered was weighting each μ with a factor of $w = \frac{1}{2\cdot(N_{PV}-1)}$, but while this worked for the MVA fits, it was causing problems when fitting the impact parameter, where the reduced weight of the pileup sample interfered with the simultaneous estimation of the common parameters. In the end we choose to fit twice in each window: the first time, with no weights added is meant to find the appropriate parameters for the models, while the second is executed after fixing the models and weighting the dataset and is meant to find the weights.

Regardless of the method used for this calculation, it was important to verify that the added weight did not introduce biases. Comparing the normalized distributions before and after the weighting showed no difference, allowing us to continue. This is a purely qualitative comparison, but it was the best available, as without the exact weight we are unableto compare the distributions with the known models.

6.1.1 Conditional observables

A Probability Density Function (PDF) G is called a conditional PDF if it describes the distribution of one or more observables X *given* the value of an observable y. We denote such a PDF with $G(X|y)$. While mathematically G is still a function of X and y, the normalization is different in the two cases:

- G(x,y) is normalized as $\int\int G(x,y)\ \mathrm{d}x\ \mathrm{d}y = 1$
- G(x|y) is normalized as $\int G(x|y)\ \mathrm{d}x = 1 \quad \forall y$

The observable y is called a conditional observable for the PDF G.

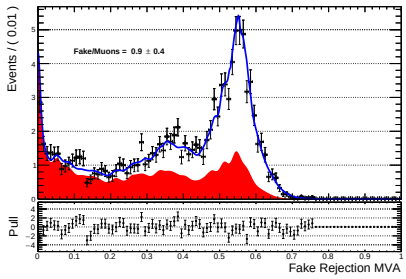
During a fit we usually input the distribution as a simple function of X and y and it is necessary to specify wether some of the variables are to be used as conditionals. This means that we are not interested in estimating the distribution of this particular observables, but they act as spectators, integrated numerically from the data to normalize the rest of the distribution.

6.2. Fake MVA fit

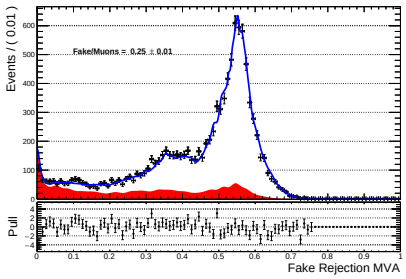
The first step for the analysis was to fit the models of the fake rejection MVA output over the data. The fit was performed separately in the already mentioned p_T and η windows, to accomodate the difference in

the distributions of the MVA. Additionally, the data was fitted indipendently for signal ($abs(\mathrm{dz}) < 0.4\,\mathrm{cm}$) and pileup ($abs(\mathrm{dz}) > 0.4\,\mathrm{cm}$) μ.

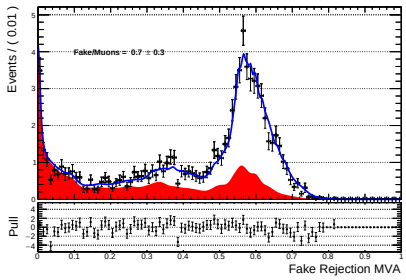
The assumption that we made, that the MVA keeps the same distribution for signal and pileup μ was verified to hold, since no particular problems were found when fitting the two samples. The interpolation converged properly in all of the windows.



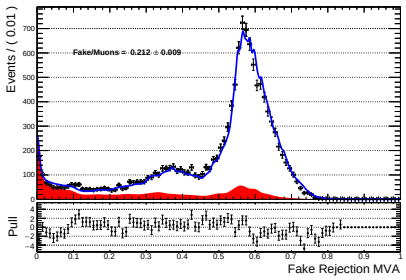
(a) $4\,\mathrm{GeV} < p_T < 5\,\mathrm{GeV}$, *pileup muons*



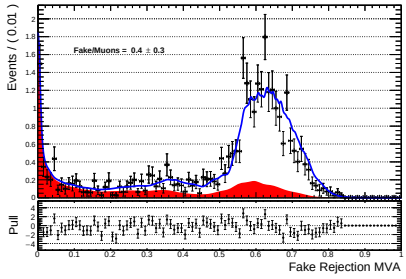
(b) $4\,\mathrm{GeV} < p_T < 5\,\mathrm{GeV}$, *signal muons*



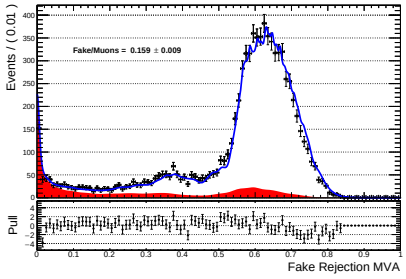
(c) $5\,\mathrm{GeV} < p_T < 7\,\mathrm{GeV}$, *pileup muons*



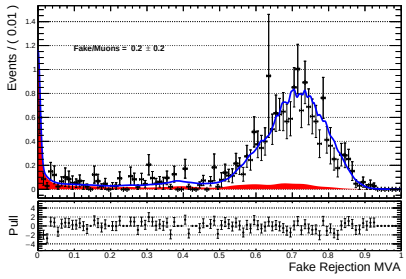
(d) $5\,\mathrm{GeV} < p_T < 7\,\mathrm{GeV}$, *signal muons*



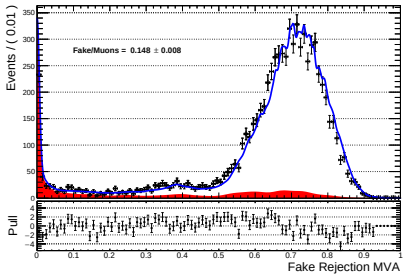
(e) $7\,\mathrm{GeV} < p_T < 10\,\mathrm{GeV}$, *pileup muons*



(f) $7\,\mathrm{GeV} < p_T < 10\,\mathrm{GeV}$, *signal muons*

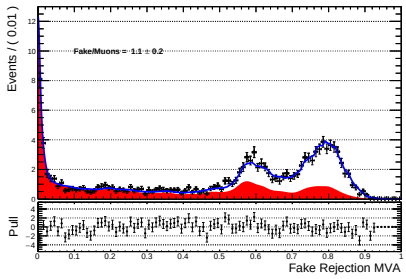


(g) $p_T > 10\,\mathrm{GeV}$, *pileup muons*

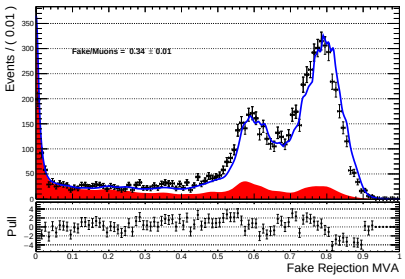


(h) $p_T > 10\,\mathrm{GeV}$, *signal muons*

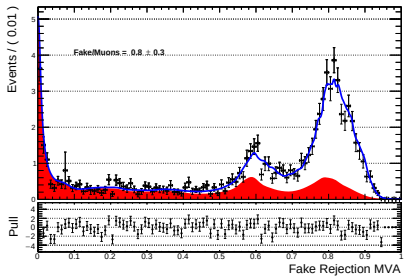
Figure 6.1: *Fitted distributions of the fake rejection MVA, for muons in barrel*



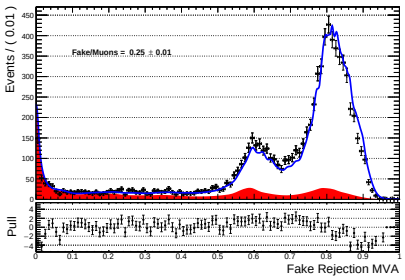
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



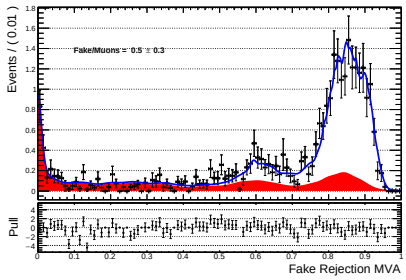
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



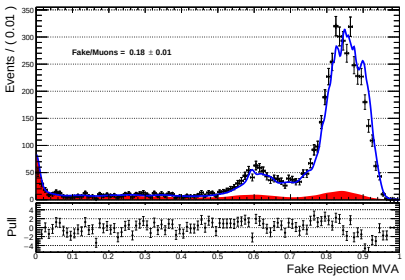
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



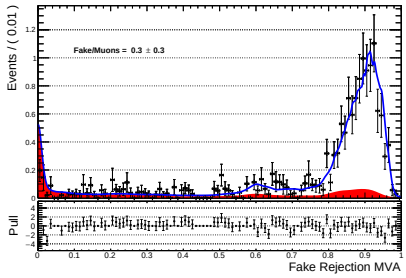
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



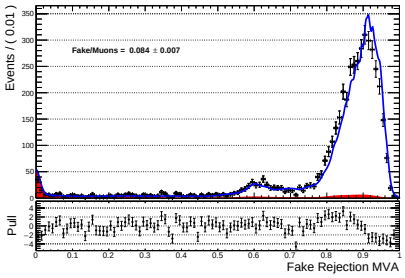
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*



(g) $p_T > 10\,\text{GeV}$, *pileup muons*



(h) $p_T > 10\,\text{GeV}$, *signal muons*

Figure 6.2: *Fitted distributions of the fake rejection MVA, for muons in endcap*

6.3. Isolation MVA fit

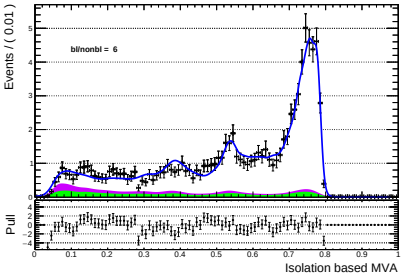
We now present the results of the fit of the isolation based MVA. The plots shown in Figure 6.3 and Figure 6.4 are valid only for non fully isolated μ, as these are the only ones where the input variables are defined.

As an input of the fit we have the weight of the fake category computed before, as a constraint to help the fit converge, especially in the non direct components that are too correlated to give meaningful results alone.

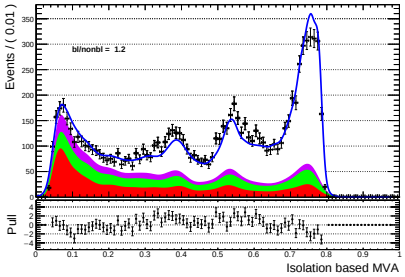
After the fit we use the isolation ratio (first column) from Table 5.2 to estimate the amount of isolated μ in each category and we add these to the count.

After this the numbers are treated like those from the previous MVA, as an hint and constraint for the final impact parameter fit.

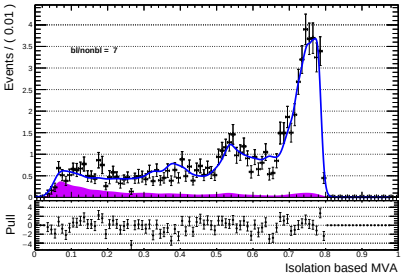
The results of the fit are shown in Figure 6.3 and Figure 6.4. As we can see, the model does not have a perfect agreement with the data.



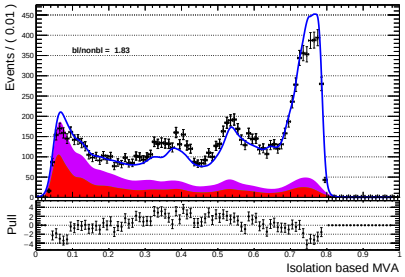
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



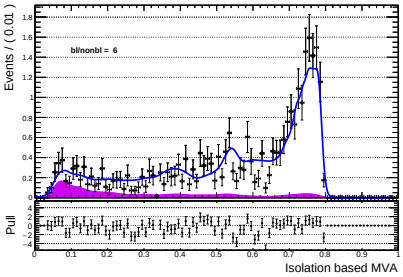
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



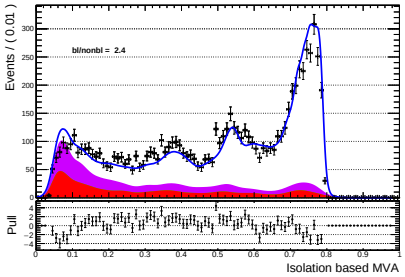
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



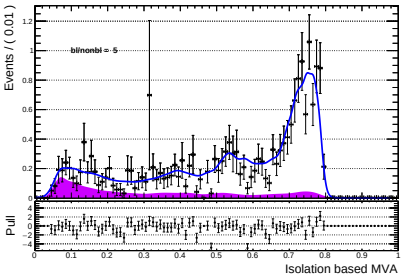
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



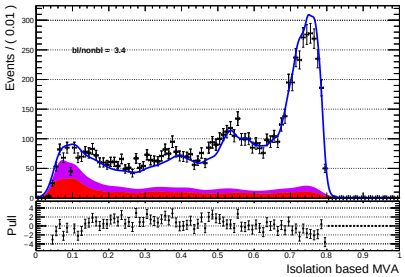
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*

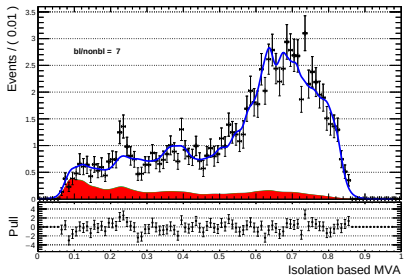


(g) $p_T > 10\,\text{GeV}$, *pileup muons*

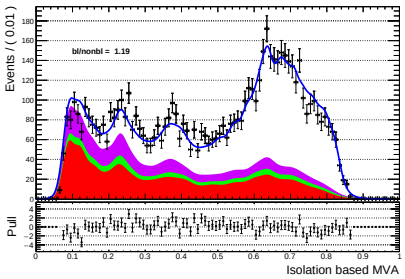


(h) $p_T > 10\,\text{GeV}$, *signal muons*

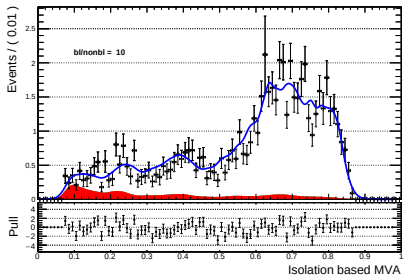
Figure 6.3: *Fitted distributions of the isolation based MVA, for muons in barrel*



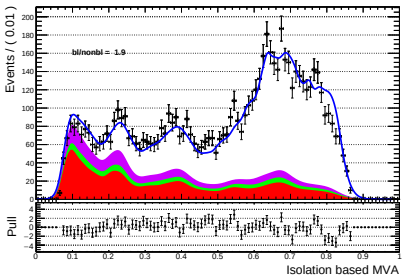
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



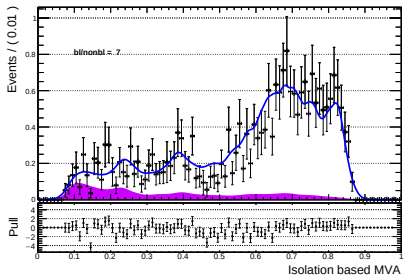
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



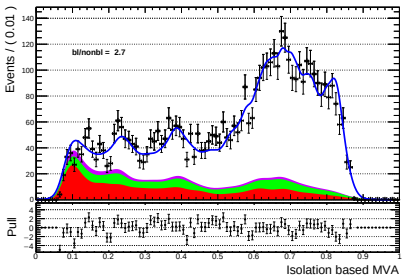
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



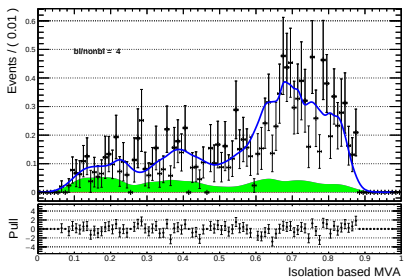
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



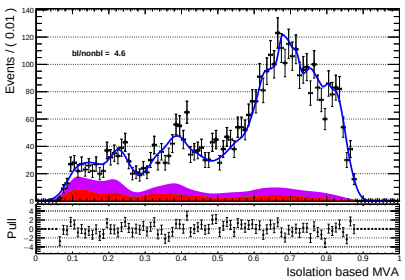
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*



(g) $p_T > 10\,\text{GeV}$, *pileup muons*



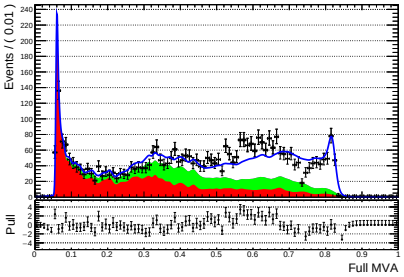
(h) $p_T > 10\,\text{GeV}$, *signal muons*

Figure 6.4: *Fitted distributions of the isolation based MVA, for muons in endcap*

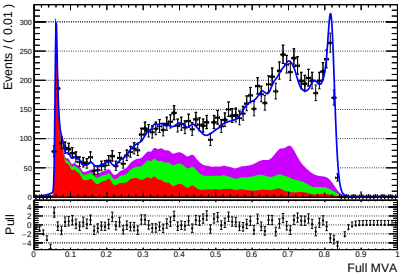
6.4. Full MVA fit

The third MVA is used as a final preprocessing step before the D_{xy} fit, estimating all categories and giving improved constraints on it. The fit proceeds as the previous ones, additionally using the informations fron the isolation fit.

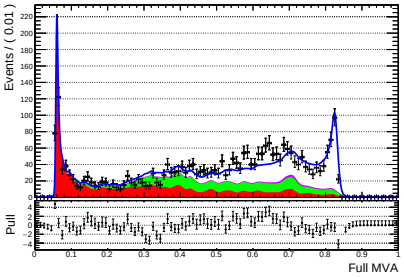
After this fit we can already produce results for the value of R_b, summarised in Table 6.1.



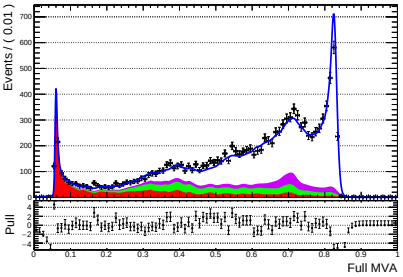
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



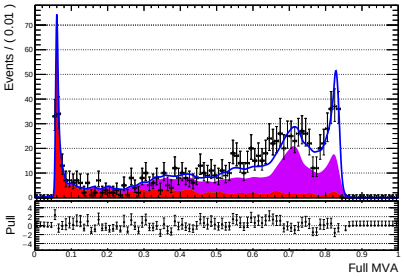
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



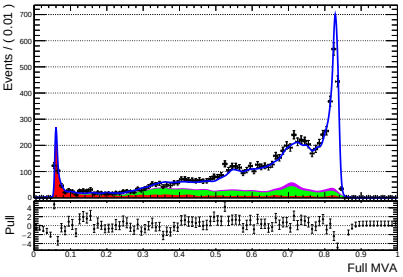
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



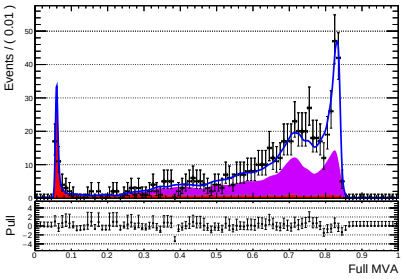
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



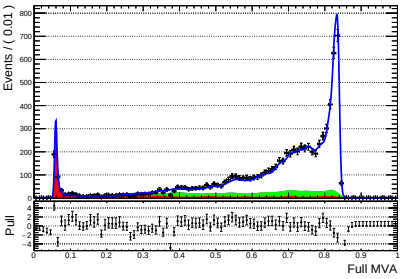
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*

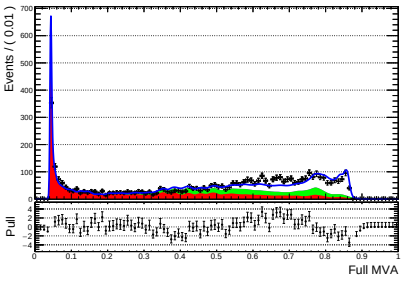


(g) $p_T > 10\,\text{GeV}$, *pileup muons*

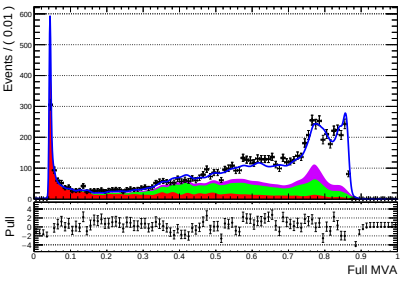


(h) $p_T > 10\,\text{GeV}$, *signal muons*

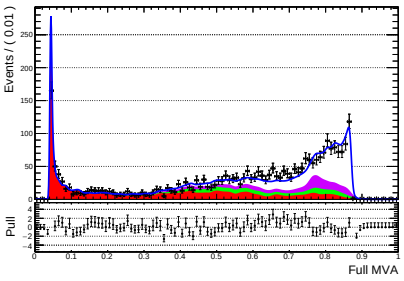
Figure 6.5: *Fitted distributions of the isolation based MVA, for muons in barrel*



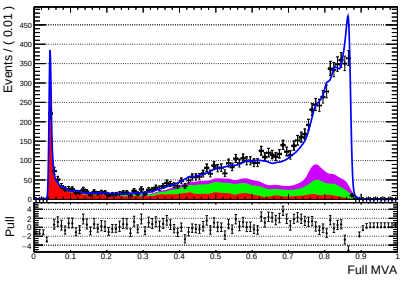
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



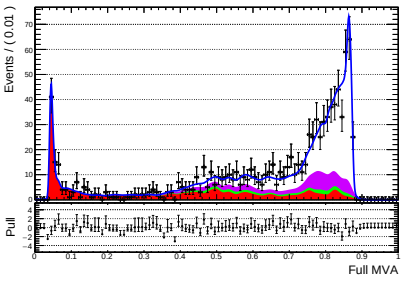
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



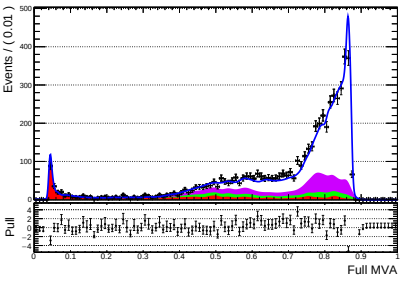
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



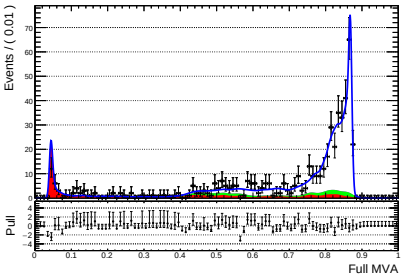
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



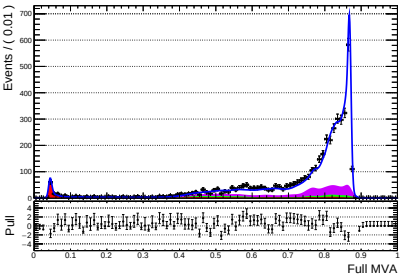
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*



(g) $p_T > 10\,\text{GeV}$, *pileup muons*



(h) $p_T > 10\,\text{GeV}$, *signal muons*

Figure 6.6: *Fitted distributions of the isolation based MVA, for muons in endcap*

6.5. Final fit

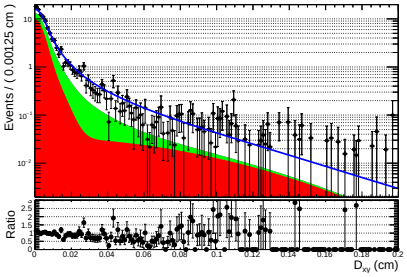
The main analysis of the thesis is the impact parameter fit.

Unlike the MVA, here our models are analytical, with parameters that we estimated from MC. To avoid sistematic errors due to an imprecise estimation of the parameter of the models, we allow the model parameters to float around their computed values. We add instead gaussian constraints in each parameter, to force it to stay close to its initial value, with the constraint sigma taken from the estimated error in the models fit. This removes a source sistematic errors at the cost of increasing the statistical one and the complexity of the fit.

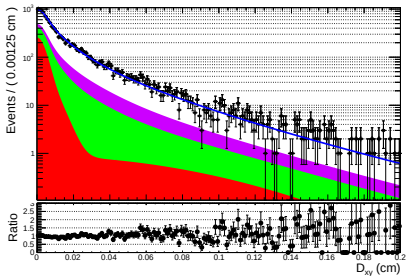
Since we are forced to assume the same model for signal and pileup, we cannot allow the model parameters to float differently in the two subsamples. The two models used in the fit have thus shared parameters and the fit is executed simultaneously for both of them. This is the reason why the second method in section 6.1 was discarder: if we lowered the weight of the pileup μ, the shared parameters would converge to values accomodating the signal much more, leaving the pileup inappropriate parameters that do not follow its shape.

Bin	Number of b		
	Signal	Pileup	Ratio(total)
barrel/pt40	7132 ± 212	43 ± 14	0.006 ± 0.002
barrel/pt50	8884 ± 242	34 ± 11	0.004 ± 0.001
barrel/pt70	5808 ± 187	22 ± 11	0.004 ± 0.002
barrel/pt100	5889 ± 123	17 ± 4	0.0029 ± 0.0007
barrel	27712 ± 392	117 ± 22	0.0042 ± 0.0008
endcap/pt40	4432 ± 191	44 ± 13	0.010 ± 0.003
endcap/pt50	5742 ± 226	44 ± 37	0.008 ± 0.006
endcap/pt70	4107 ± 168	19 ± 20	0.005 ± 0.005
endcap/pt100	3869 ± 87	12 ± 3	0.0031 ± 0.0008
endcap	18150 ± 351	118 ± 44	0.007 ± 0.002
All	45862 ± 526	235 ± 49	0.005 ± 0.001

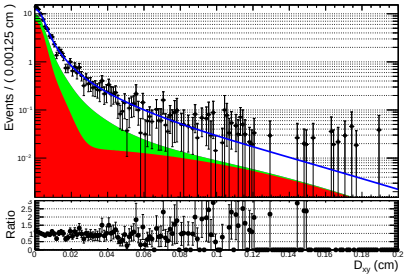
Table 6.1: *Summary of the results from the full-variables mva fit*



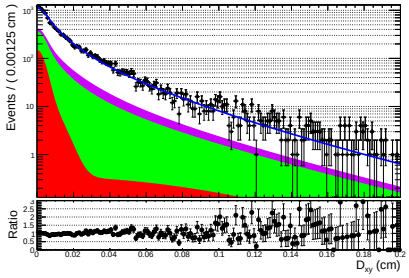
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



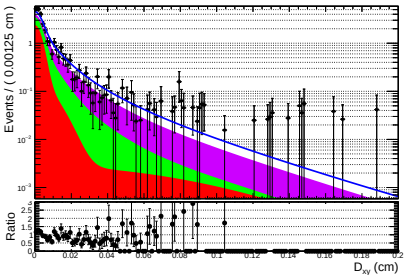
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



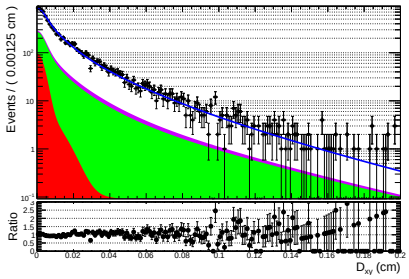
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



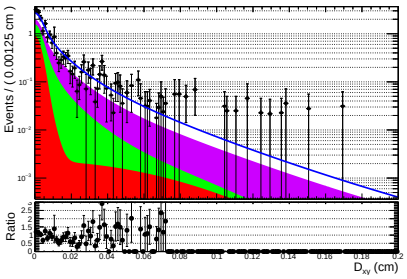
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



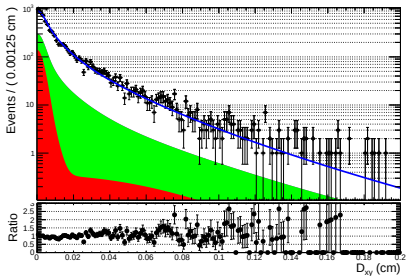
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*

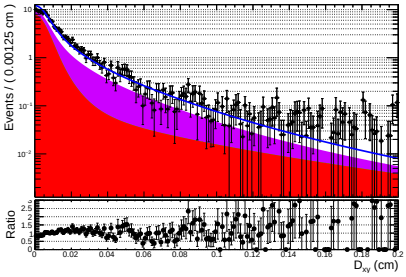


(g) $p_T > 10\,\text{GeV}$, *pileup muons*

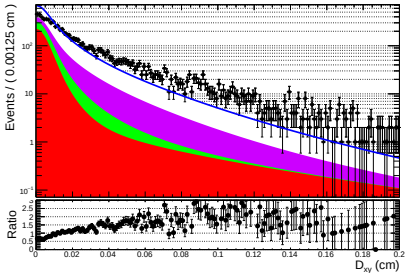


(h) $p_T > 10\,\text{GeV}$, *signal muons*

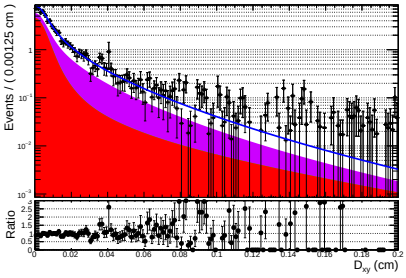
Figure 6.7: *Fitted distributions of the impact parameter, for muons in barrel*



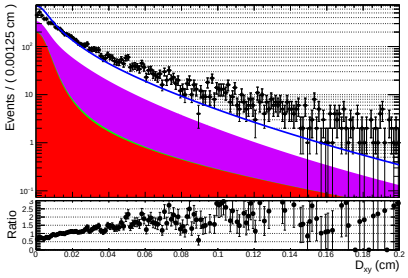
(a) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *pileup muons*



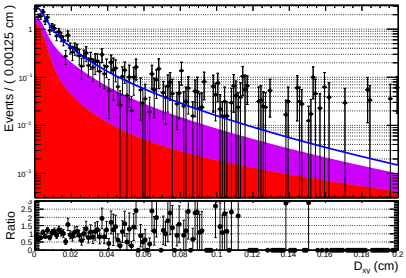
(b) $4\,\text{GeV} < p_T < 5\,\text{GeV}$, *signal muons*



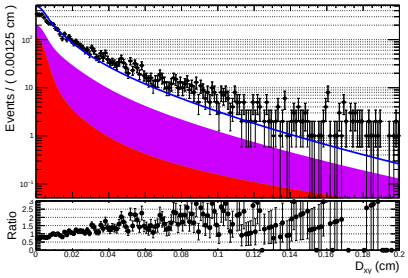
(c) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *pileup muons*



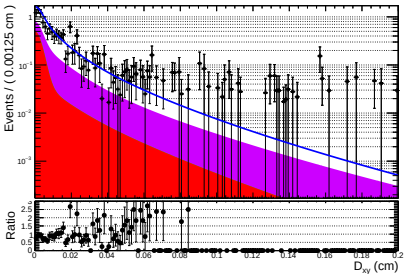
(d) $5\,\text{GeV} < p_T < 7\,\text{GeV}$, *signal muons*



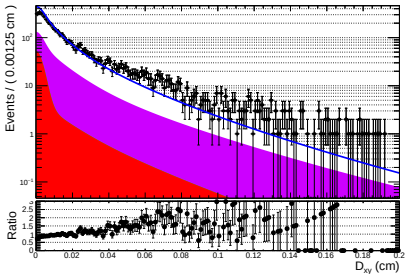
(e) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *pileup muons*



(f) $7\,\text{GeV} < p_T < 10\,\text{GeV}$, *signal muons*



(g) $p_T > 10\,\text{GeV}$, *pileup muons*



(h) $p_T > 10\,\text{GeV}$, *signal muons*

Figure 6.8: *Fitted distributions of the impact parameter, for muons in endcap*

The fits of each window of p_T, η and D_z are shown in Figure 6.7 and Figure 6.8. In the figures the read area represents fake μ, the green area μ from c quarks and the white area the μ from b quarks.

Chapter 7

Results

From the fits in chapter 6 we can finally compute R_{b}.

We first retrieve the ratio in each window of η and p_T. We use here the correlation from the fit to propagate the errors. The correlation here is due to the shared parameters (the resolution for example) in the fit.

When merging the result to the η window and finally in the whole range, the result is calculated adding separately the b-quark counts for signal and pileup. Since the windows are disjoint we assume no correlation between the parameters in different range. The correlation between the signal and pileup count are instead correctly propagated.

Table 7.1 is a summary of the found values.

	Number of b		
Bin	Signal	Pileup	Ratio(total)
barrel/pt40	7468 ± 190	35 ± 10	0.005 ± 0.001
barrel/pt50	9335 ± 180	30 ± 10	0.003 ± 0.001
barrel/pt70	6169 ± 156	13 ± 7	0.002 ± 0.001
barrel/pt100	5835 ± 77	12 ± 7	0.002 ± 0.001
barrel	28807 ± 314	90 ± 18	0.0031 ± 0.0006
All	28807 ± 314	90 ± 18	0.0031 ± 0.0006

Table 7.1: *Final results*

Appendix A

Input variables for fake rejection MVA

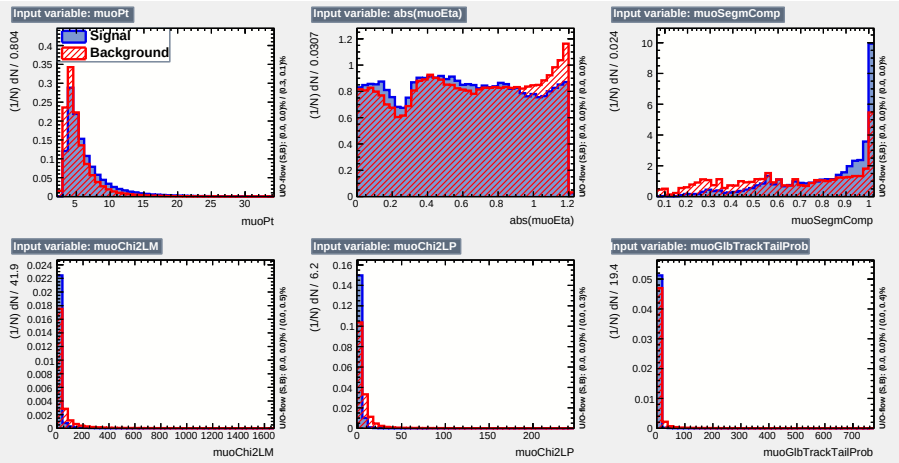


Figure A.1: *Fake rejection MVA input variables, first group*

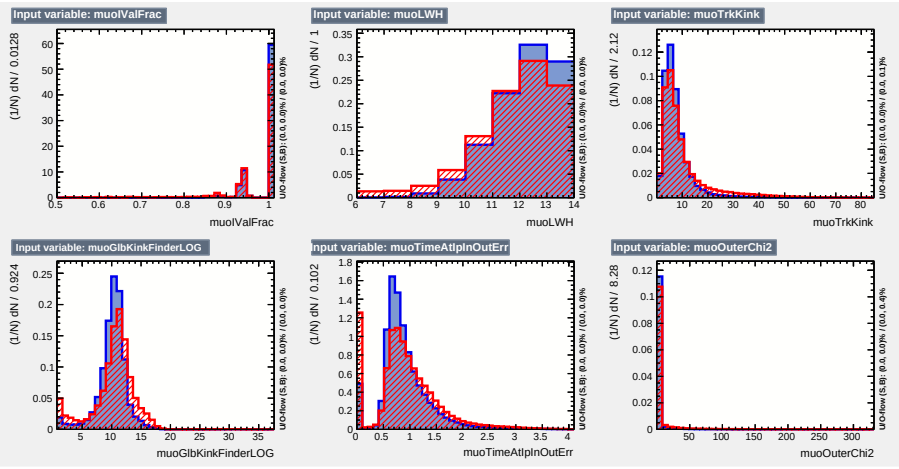


Figure A.2: *Fake rejection MVA input variables, second group*

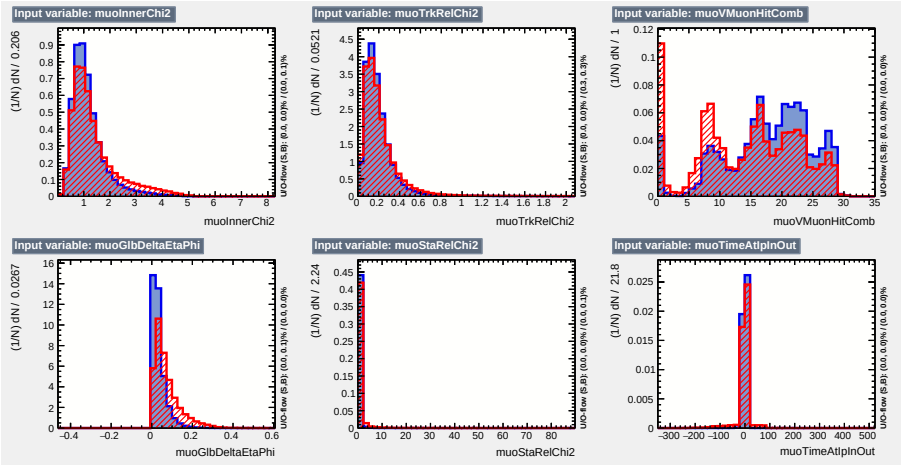


Figure A.3: *Fake rejection MVA input variables, third group*

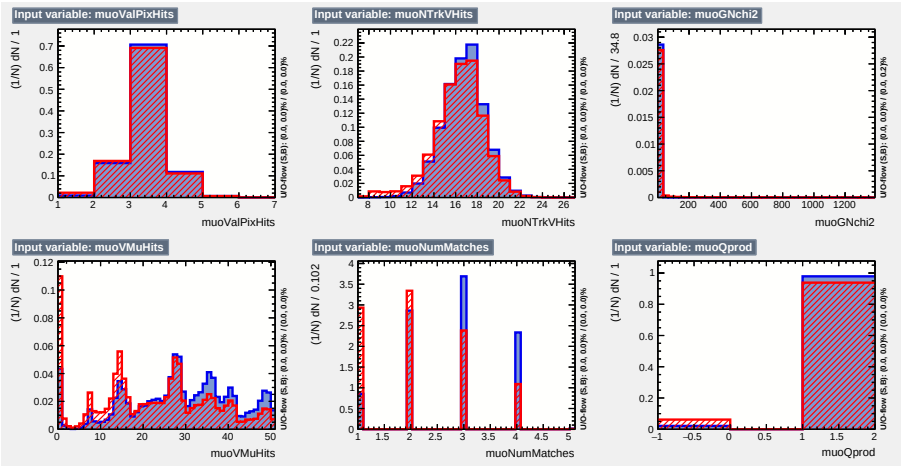


Figure A.4: *Fake rejection MVA input variables, fourth group*

A.1. Isolation variables

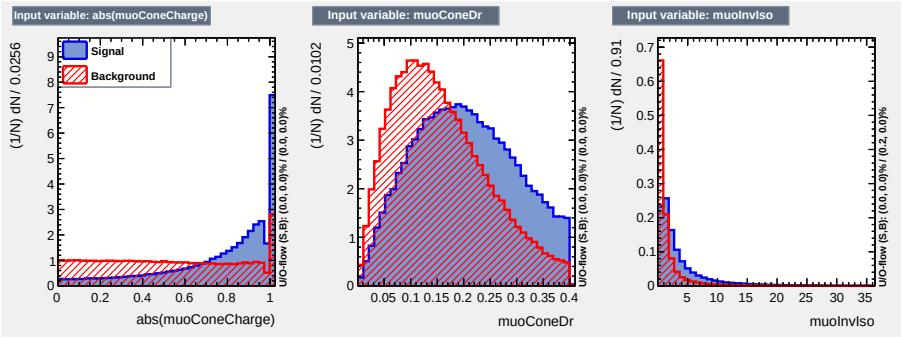


Figure A.5: *iso mva input/iso mva input*

Bibliography

[1] UA1 collaboration. "Beauty production at the CERN proton-antiproton collider". In: *Physics Letters B* 186.2 (1987), pp. 237–246. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(87)90287-5. URL: http://www.sciencedirect.com/science/article/pii/0370269387902875.

[2] F. Abe. "Measurement of the bottom quark production cross section using semileptonic decay electrons in pp^- collisions at $\sqrt{s}$ =1.8 TeV". In: *Phys. Rev. Lett.* 71 (4 July 1993), pp. 500–504. DOI: 10.1103/PhysRevLett.71.500. URL: https://link.aps.org/doi/10.1103/PhysRevLett.71.500.

[3] CDF Collaboration. "Measurement of the b-hadron production cross section using decays to $\mu^- D^0 X$ final states in $p\overline{p}$ collisions at $\sqrt{s} = 1.96$ TeV". In: *Phys. Rev. D* 79 (9 May 2009), p. 092003. DOI: 10.1103/PhysRevD.79.092003. URL: https://link.aps.org/doi/10.1103/PhysRevD.79.092003.

[4] D0 Collaboration. "Inclusive μ and b-Quark Production Cross Sections in $p\overline{p}$ Collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 74 (18 May 1995), pp. 3548–3552. DOI: 10.1103/PhysRevLett.74.3548. URL: https://link.aps.org/doi/10.1103/PhysRevLett.74.3548.

[5] D0 Collaboration. "The bb production cross section and angular correlations in pp collisions at s=1.8 TeV". In: *Physics Letters B* 487.3 (2000), pp. 264–272. ISSN: 0370-2693. DOI: https://doi.org/10.1016/S0370-2693(00)00844-3. URL: http://www.sciencedirect.com/science/article/pii/S0370269300008443.

[6] LHCb collaboration. "Measurement of (pp→bb$^-$X) at s=7 TeV in the forward region". In: *Physics Letters B* 694.3 (2010), pp. 209–216. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2010.10.010. URL: http://www.sciencedirect.com/science/article/pii/S0370269310012074.

[7] CMS Collaboration. "Measurement of the B^+ Production Cross Section in pp Collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Rev. Lett.* 106 (11 Mar. 2011), p. 112001. DOI: 10.1103/PhysRevLett.106.112001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.112001.

[8] K. Cranmer. "Kernel estimation in high-energy physics". In: *Computer Physics Communications* 136.3 (2001), pp. 198–207. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(00)00243-5. URL: http://www.sciencedirect.com/science/article/pii/S0010465500002435.