

Università degli Studi di Bari "Aldo Moro"

DIPARTIMENTO DI FISICA INTERATENEO "MICHELANGELO MERLIN"

Dottorato di Ricerca in Fisica
Ciclo XXXVI

Settore Scientifico Disciplinare: FIS/01

B_c physics with LHC Run-3 data at the CMS experiment

Dottorando:
Dott. Ümit Sözbilir

Supervisore:
Ch.mo Prof. Alexis Pompili

Coordinatore:
Ch.mo Prof. Giuseppe Gonnella

Esame Finale 2024

Contents

Introduction **iii**

1 The B_c meson and the $\bar{b}c$ system **3**
1.1 The Standard Model of Particle Physics 3
1.2 Quantum Chromodynamics 7
1.2.1 The Constituent Quark Model 7
1.2.2 The Quantum Chromodynamics theory 8
1.2.3 Hadrons with unconventional quarks content 10
1.3 Introduction to the B_c meson 12
1.4 B_c meson production at the LHC 13
1.4.1 The simulation of events with a B_c meson: the BCVEGPY generator 16
1.5 The $(c\bar{b})$ mass spectrum 17
1.6 B_c decays 19
1.6.1 The B_c decays experimentally observed 21
1.6.2 The B_c lifetime 22
1.7 Prospects in the B_c physics: excited states 23
1.7.1 Recently observed excited states in the $\bar{b}c$ spectrum . . . 23
1.7.2 Search for further excited states in the $\bar{b}c$ spectrum . . . 27
1.8 Prospects in the B_c physics: observation of doubly-beauty baryons and tetraquarks 29
1.8.1 Search for signatures of double beauty hadrons 29
1.8.2 Estimation of the experimental sensitivity for the search of double beauty hadrons 32

2 The CMS experiment at the LHC **33**
2.1 The Large Hadron Collider 34
2.1.1 Proton-proton interactions 34
2.1.2 Instantaneous and integrated luminosity 36
2.2 The CMS detector 39
2.2.1 Geometry and reference system 39
2.2.2 Magnet 41

2.2.3 The tracking system . . . 42
2.2.4 Muon chambers system . . . 44
2.2.5 The CMS trigger system . . . 45
2.3 Track reconstruction in CMS . . . 47
2.3.1 Primary Vertex and beamspot reconstruction . . . 49
2.4 Muon reconstruction at CMS . . . 53
2.4.1 Muon identification . . . 55
2.5 A new dedicated trigger for the B_c meson in Run-3 . . . 57

3 ML techniques for HEP data analysis: an `XGBoost` study 59
3.1 Introduction to `XGBoost` . . . 60
3.1.1 Performance Metrics and ROC curve . . . 62
3.1.2 Kolmogorov-Smirnov test . . . 64
3.2 Data and Monte Carlo samples as inputs for training . . . 65
3.2.1 Definition of Signal Region and Control Region . . . 66
3.3 Performances of cut-based selections . . . 69
3.3.1 Displaced cut-based selection as reference . . . 71
3.3.2 Inclusive cut-based selection as reference . . . 73
3.4 `XGBoost` for analysis selection . . . 75
3.4.1 Performance of `XGBoost` displaced model . . . 76
3.4.2 Performance of `XGBoost` inclusive model . . . 83

4 Optimisation of the `XGBoost` selection 89
4.1 `Optuna`'s Timing Performance . . . 90
4.2 AUC estimate and its uncertainty . . . 92
4.2.1 AUC uncertanity in the displaced set-up . . . 93
4.2.2 AUC uncertanity in the inclusive set-up . . . 94
4.3 `XGBoost` optimization and signal extraction . . . 96
4.3.1 The Optuna Figure of Merit . . . 96
4.3.2 The Signal Extraction Figure of Merit . . . 96
4.3.3 Optimization workflow . . . 98
4.4 Hyperparameter Optimisation for Displaced Approach . . . 101
4.4.1 BDT optimization with Optuna . . . 101
4.4.2 Signal extraction optimization . . . 106
4.4.3 Comparison between different SE FoMs . . . 108
4.4.4 Results from the optimized model . . . 111
4.5 Hyperparameter Optimisation for Inclusive Approach . . . 115
4.5.1 BDT optimization with Optuna . . . 115
4.5.2 Signal extraction optimization . . . 121
4.5.3 Comparison of different SE FoMs . . . 125
4.5.4 Results with the Punzi FoM optimization . . . 127

4.5.5 Results with the PxSS FoM optimization 129

5 Preliminary B_c lifetime measurement 135
5.1 Selection Efficiency 136
5.1.1 Inclusive cut-based approach 138
5.1.2 Inclusive `XGBoost`-based approach 142
5.1.3 Displaced cut-based approach 145
5.1.4 Displaced `XGBoost`-based approach 147
5.1.5 Summary table 149
5.2 Reconstruction efficiency 151
5.2.1 Inclusive cut-based approach 153
5.2.2 Inclusive XGB-based approach 154
5.2.3 Displaced cut-based approach 156
5.2.4 Displaced XGB-based approach 157
5.2.5 Summary table 159
5.3 Correction of the measured $c\tau$ values 160
5.3.1 Additional test on corrections' factorization 162
5.4 Lifetime Measurement for Optimized `XGBoost inclusive` Approach 164
5.4.1 Flight length distribution in control region 164
5.4.2 Flight length distribution in signal region 168
5.4.3 Preliminary extraction of the B_c lifetime 170

A Selection of the physics variables for `XGBoost` 177
A.1 Variables' selection for Displaced Approach 178
A.2 Variables' selection for Inclusive Approach 188

B Strategy for the determination of the B_c primary vertex 201
B.1 Definition of the proper decay length 202
B.1.1 Comparison between 2D and 3D lifetime definitions . . . 203
B.2 Determination of the Primary Vertex choice 206

C Cabibbo suppressed $B_c \to J/\psi K$ signal in CMS Run-3 data 211
C.1 Estimation of the $B_c \to J/\psi K$ contribution with the `XGBoost` inclusive approach 212
C.2 Estimation of the $B_c \to J/\psi K$ contribution 215

D Glossary of Variables 219

Bibliography 223

Introduction

The work described in this thesis deals with a detailed study focusing on a performant extraction of the signal of the B_c meson in its decay to $J/\psi\,\pi$ that is the most suitable decay mode to reconstruct B_c mesons at the Compact Muon Solenoid (CMS) experiment, exploiting the Run-3 data collected in the years 2022-23 at the CERN Large Hadron Collider (LHC). Specifically, large part of the study focuses on the performance comparison between a standard cut-based selection of the B_c signal and one based on Machine Learning (ML), showing the advantage of using `XGBoost`, namely a parallel Gradient Boosted Decision Trees based library, against a "classical" (non AI-powered) approach. It is also investigated and developed the optimization of the `XGBoost` hyperparameters exploiting a novel software tool called `Optuna`.

Chapter 1 introduces the production of B_c mesons at proton-proton collisions at the LHC and its decay modes. This introduction does not aim to be exhaustive but functional to the rest of the thesis by providing concepts and contexts useful later. The prospects in B_c physics at CMS are explored and discussed in detail, focusing on two main topics that seem to be promising when the full Run-3 dataset will be available. One topic is the investigation of the excited states of the B_c spectrum, and in particular of the 2P, 3S and 3P states, unobserved so far; their transitions, either radiative or with a dipion emission, are expected to bring to the B_c ground state in the final state. The second topic deals with the search for B_c-based signatures of double beauty hadrons (either baryons which are expected to exist but never observed and possible double-heavy tetraquarks). In both cases these searches rely on an efficient and effective extraction of a sample of B_c mesons, as much large as possible.

Chapter 2 describes the experimental setup and the data collection process at the CMS experiment, in particular during the Run-3. It includes an overview of the CMS detector, its sub-systems relevant to B_c meson studies, and the specific challenges associated with tracking, muon reconstruction and identification, and the trigger system and strategy to allow, at CMS, Heavy Flavour physics even in an extremely high-luminosity environment. This chapter ends up describing the High Level Trigger expressly configured for the Run-3

data taking to inclusively trigger events possibly having a B_c meson decaying to $J/\psi\,\pi$, with the "inclusive" feature meaning that there is no requirement - at HLT level - in the displacement of the B_c from its production vertex, as it was for Run-2. This choice was motivated by the aim to reconstruct as many B_c mesons as possible, for the reason explained above, and possibly with a minimal bias on the proper decay time distribution.

Chapter 3 introduces the ML models used, their training on both real and simulated datasets, and the performance metrics. Afterwards, it illustrates, first, the extraction of the signal of the physics case under exam (the $B_c \to J/\psi\,\pi$ decay), with a standard cut-based selection guided by analysis experience, in two cases corresponding to two different reconstruction topologies: displaced and inclusive, depending on whether the displacement observables (related to the B_c meson decay length) are used or not, respectively. Secondly it presents the application of a trained `XGBoost`-based selection in the same two cases, and its better performance - in comparison to the conventional selection technique - is discussed in terms of standard figures-of-merit such as signal significance and purity.

Chapter 4 discusses the implementation of the hyperparameters' optimization of the BDT-based signal extraction discussed in Chapter 3. This optimization is carried out, for both displaced and inclusive reconstructed B_c decay topologies, within the `Optuna` cutting-edge software framework, by implementing, on the top of a user defined `Optuna` FoM, a further optimized choice of the final best Working Point in terms of additional figures-of-merit such as the Punzi FoM or the product of purity times signal significance. The performances' gain of this optimization approach is compared with the unoptimized XGBoost results obtained in the previous chapter.

Chapter 5 provides a preliminary discussion of the biases that can be introduced by the selections developed in the previous chapters and the reconstruction of the proper time distribution of the B_c selected candidates.

The Appendices, following the Conclusions, contain further detailed information to support the studies presented or even add a further study, as in the case of the estimation of the possible Cabibbo-suppressed $B_c \to J/\psi K$ decay's contribution to the analyzed $J/\psi\,\pi$ final state.

Chapter 1

The B_c meson and the $\bar{b}c$ system

1.1 The Standard Model of Particle Physics

The Standard Model of particle physics (SM) currently provide the best understanding of elementary particles and their interactions (the electromagnetic, weak and strong nuclear interactions). The fourth interaction in nature, the gravitational force, is not described by the SM; its strength is measured to be many orders of magnitude lower than the other three fundamental forces, thus it is expected to have negligible impact onto the fundamental particles' processes at accessible energies and their phenomenology. The SM is a renormalizable field theory based on a local symmetry that generalizes the structure of the Quantum Electrodynamics (QED) to a larger set of conserved currents and charges (eight strong charges and four electroweak charges). It is based on the gauge group $SU(3)_C \otimes SU(2)_L \otimes U(1)_\gamma$, where $SU(3)_C$ gauge group is the symmetry group of strong interactions, and $SU(2)_L \otimes U(1)_\gamma$ is the unified gauge group of weak and electromagnetic (electroweak) interactions.

QED describes the interaction between two electrically charged particles with the exchange of photons. The fundamental particles that build up matter in the universe exist as either quarks (which have color and electroweak charge) and leptons (which only have electroweak charge). These are known as fermions and have a half-integer spin. For each elementary fermion there exists its corresponding antiparticle (with same mass and opposite electric charge).

There are three generations of quarks and leptons, each containing a weakly charged doublet of quarks (in 3 color replicas) and a colorless weakly charged doublet with a charged lepton and its corresponding neutrino:

- Generation I : quarks u and d, leptons e^- and $\bar{\nu}_e$;
- Generation II : quarks c and s, leptons μ^- and $\bar{\nu}_\mu$;
- Generation III : quarks t and b, leptons τ^- and $\bar{\nu}_\tau$.

Name	Spin	Charge	Force	Mass [GeV]
γ (photon)	1	0	Electromagnetic	0
Z^0	1	0	Weak	91.1876 ± 0.0021
$W^{\pm}$	1	± 1	Weak	80.377 ± 0.012
g (gluon)	1	0	Strong	0
H (Higgs)	0	0	-	125.25 ± 0.17

Table 1.1: Selected properties of the gauge bosons in the SM. The electric charge Q is shown in units of absolute electron charge. Reported values are from Ref. [1]

Leptons are free to exist individually whereas quarks are confined to bound states (called *hadrons*): conventional *mesons* are made up of a quark (q) and an anti-quark ($\bar{q}$ forming a ($q\bar{q}$) system and conventional *baryons* are three-quarks systems (qqq). However also possible unconventional quarks' structure (called *exotic hadrons*), such as tetraquark and pentaquark states, are currently under investigation, both experimentally and theoretically.

The interaction between the fermions is governed by fundamental forces mediated by gauge bosons, which have an integer spin. Each of the three fundamental interactions in the SM (the electromagnetic, the strong, and weak nuclear forces) is mediated by a certain spin 1 boson; specifically, the photon (γ) mediates the electromagnetic interaction, the gauge bosons $W^{\pm}$ ans Z^0 the weak one, whereas the eight gluons (gs) are the mediators of the strong force. The photon and the gluons are massless as a consequence of the exact conservation of the corresponding symmetry generators (the electric charge and the 8 color charges). On the countrary, the weak bosons have large masses ($m_W \approx 80.4$ GeV, $m_Z \approx 91.2$ GeV) because the corresponding symmetries are broken. Table 1.1 summarizes the properties of gauge bosons.

The quarks belonging to the first generation are the *up* and *down*; *charm* and *strange* quarks belong to the second one, *top* and *beauty* to the third. The former of each family is characterized by an electric charge $Q_q = +2/3$ (in units of the absolute electron charge), the latter by $Q_q = -1/3$. The corresponding anti-quarks are characterized by an opposite-sign charge. The quark masses increase with the generations [1]: if in the first generation $m(u)$ and $m(d)$ are estimated to be of the order of few MeV, the other masses span several order of magnitudes, from $m(s) = 93.4^{+8.6}_{-3.4}$ MeV to $m(t) = 172.69 \pm 0.30$ GeV. Together with the top quark, the quarks charm ($m_c = 1.27 \pm 0.02$ GeV) and beauty ($m_b = 4.18^{+0.03}_{-0.02}$ GeV) are considered heavy quarks. When we speak about *Heavy Flavour Physics* we mean the Physics of mesons, baryons and unconventional structures with at least one heavy quark. Therefore the heavy flavour realm includes; 1) charmed mesons (D^0, D^+, D_s^+, ...) and baryons (Λ_c, Ξ_c, ...), beauty mesons (B^0, B^+, B_s^0, ...) and baryons (Λ_b, Ξ_b, ...), having

one heavy quark, 2) mesons with two heavy quarks such as *charmonium* ($\bar{c}c$ system), *bottomonium* ($\bar{b}b$ system) and 3) the $\bar{b}c$ system (being B_c^+ its ground state) and baryons with two heavy quarks (of which only the Ξ_{cc} has been observed so far) [2].

Mesons and baryons containing a top quark are not expected to be observed; same consideration holds for the hypothesized *toponium* ($\bar{t}t$ system). This is because the Standard Model predicts the mean lifetime of top quarks to be roughly $5 \cdot 10^{-25}$ s which is about a twentieth of the timescale for strong interactions, and therefore they do not form hadrons; in other words, the top quark "decays" before hadronization. Nevertheless, the gluon-gluon production mechanism of the toponium at the LHC is currently investigated through proposed experimental signatures (see, for instance [3].

How particles acquire mass in the SM is by spontaneous symmetry breaking through the Brout-Englert-Higgs mechanism [4, 5, 6], which requires the existence of a Higgs boson with zero spin, which was discovered in 2012 jointly by the CMS and ATLAS experiments at the LHC [7, 8]. The Higgs field interact with the fermionic sector through *Yukawa terms*; fermions acquire a mass proportional to the vacuum expectation value of the Higgs field and the strength of the coupling. Particles that do not interact with the Higgs field (photon and gluons) remain massless. Fermion couplings to the Higgs field not only determines their masses but they induce a misalignment of quark mass eigenstates with respect to the eigenstates of the weak charges, thus allowing all fermions of heavy families to decay to lighter ones. These couplings provide the only mechanism, within the SM, that can account for the observed violation of the CP invariance, the latter representing the invariance of the laws of nature under mirror reflection (parity P) and the interchange of particles with their anti-particles (charge conjugation C).

As already mentioned, the SM is described by a non-Abelian gauge theory containing the unitary composite symmetry group $SU(3)_c \otimes SU(2)_L \otimes U(1)_\gamma$. The $SU(2) \otimes U(1)$ group describes the electroweak interaction, whereas $SU(3)$ describes the strong interaction between quarks and gluons as it will be better explained in the next section. The SM Lagrangian is commonly divided in two parts:

$$\mathcal{L}_{\mathcal{SM}} = \mathcal{L}_{\mathcal{QED}} + \mathcal{L}_{\mathcal{QCD}}$$

where $\mathcal{L}_{\mathcal{QED}}$ holds the QED interactions and its gauge bosons, whereas $\mathcal{L}_{\mathcal{QCD}}$ holds the gluon fields and their interactions.

In the SM, the masses and mixings of quarks originate from the Yukawa interactions with the Higgs condensate. The *quark mixing*, by convention, is restricted to the down quarks:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_L = \mathbf{V} \cdot \begin{pmatrix} d \\ s \\ b \end{pmatrix}_L \tag{1.1}$$

where the primed quarks are the mass eigenstaes and the others are the flavour eigenstates. $\mathbf{V}$ is the Cabibbo-Kobayashi-Maskawa (CKM) matrix which contains one free imaginary phase, being the source of the CP violation, and the three free real angles. The CKM matrix is usually written as

$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.2}$$

so that the entries are labelled by the quark flavours. The probability of the transition from a quark i to a quark j is therefore proportional to $|V_{ij}|^2$. The CKM elements are fundamental parameters of the SM that need to be accurately determined from experiments. The unitarity of the CKM matrix imposes 6 equations that can be interpreted as unitarity triangles in the complex plane. Measuring the parameters of the unitarity triangle serves two purposes:

- to determine some fundamental parameters of the Standard Model;
- to test whether the CKM mechanism of flavor and CP violation is indeed correct, or whether there are hints of deviations from the Standard Model.

Weak decays of hadrons containing heavy quarks are employed for tests of the Standard Model and measurements of its parameters. In particular, they offer the most direct way to determine the real angles of the CKM matrix, to test the unitarity of the CKM matrix and explore the physics of CP violation. Until now no significant deviations from the CKM mechanism have been established.

1.2 Quantum Chromodynamics

1.2.1 The Constituent Quark Model

Our understanding of the strong interaction began in 1963, when Gell-Man and Zweig postulated that the strongly interacting particles observed by experiments (hadrons such as the proton, neutron and pion) are not fundamental particles, but instead bound states composed of more basic constituents [9, 10]. These constituents, that are now thought to be fundamental particles, subsequently became known as the quarks. Even at the time of the proposal, the properties of the known spectrum of hadrons put tight constraints on the properties of the quarks. The π and ρ mesons were known to have $J^P = 0^-$ and 1^- respectively (where J and P denote the spin and parity quantum numbers, respectively). These mesons could naturally be interpreted as bound states of a quark and anti-quark with zero orbital angular momentum. The known hadrons could be classified within an $SU(3)$ group of three spin 1/2 particles, each with a unique flavor; the up (u), down (d), and strange (s) quarks. In order to allow the model to also predict the baryon spectrum, the quarks were constrained to be fractionally charged: $+2/3$ e for u and $-1/3$ e for d and s quarks. At first glance, this model naturally explained the known spectra of the mesons and baryons.

The Gell-Mann–Zweig picture was remarkably successful; it accounted for all of the known hadrons at that time and predicted the existence of additional hadrons that were subsequently discovered. More than 40 years later, the Gell-Mann–Zweig idea (also known as the Constituent Quark Model, or CQM) remains an effective scheme for classifying all of the known hadrons, although the number of quarks has expanded to include the charm (c), bottom (b) and top (t) quarks.

However, this model could not explain the existence of baryons consisting of three quarks of the same flavor with spin $J = 3/2$ (e.g. Δ^{++}), for which there was strong experimental evidence. The quarks were thought to be spin-1/2 particles, and so the Fermi-Dirac statistics required the overall baryon wave function to be completely anti-symmetric under the exchange of any quark pair. To reconcile this issue, a further quantum number was introduced, unique only to quarks: the color quantum number. If the baryon wave function were totally symmetric under the exchange of spin and flavor, then requiring the wave function to be anti-symmetric under the exchange of color quantum numbers would have allowed the overall wave function to be totally anti-symmetric and thus consistent with the expectation for fermions.

While the quark model was phenomenologically successful, and the two most important concepts in our modern understanding of strong interaction,

quarks and color, had been introduced, the model raised just as many questions as it answered. For example, the new color degrees of freedom could lead to a spectrum of hadronic states far richer than that observed at the time. This led to the ad-hoc requirement that only color singlet states can exist in nature. However, it would not be until the development of Quantum Chromodynamics (QCD) that a satisfactory explanation of this feature could be provided, along with answers to many other remaining questions.

1.2.2 The Quantum Chromodynamics theory

Our current understanding is that the forces that bind quarks into hadrons are described by the non-Abelian field theory called Quantum Chromodynamics (QCD) and based upon the $SU(3)$ symmetry group. The QCD theory was born within the context of a mathematical investigation of isotopic spin as a gauge theory by Yang and Mills [11].

In contrast to electrodynamics, for which the interaction is mediated by a photon, which corresponds to the single generator of the $U(1)$ symmetry group, the quanta of a gauged isospin theory carry charges. One consequence of this circumstance is a different scale dependence of the interaction strength. The electrodynamic force becomes stronger at short distances (large momentum-transfer scales), whereas, in a Yang-Mills type of theory, the force becomes weaker at short distances (*asymptotic freedom*). The relevant calculation, not initially as a signal of asymptotic freedom, first appeared in a theory based on gauged SU(2) symmetry. But a gauged SU(3), as the theory of the strong interactions, contains the "color" ingredient necessary to understand why the ground states of baryons have quark wave functions that are symmetric in space $\times$ spin $\times$ flavor, where the last term denotes the quark label $(u, d, s, ...)$. Each quark comes in one of three colors, and a wave function totally anti-symmetric in color can be constructed by taking one of each color. The strong interaction is mediated by eight gluons corresponding to the eight generators of the $SU(3)$ symmetry group. The behavior of the strong interaction at long distances is also different from that of the electrodynamic interaction. The gluons, quanta of the strong interactions, interact with one another in such a way that lines of force between a quark and an antiquark bunch up into a tube of essentially constant cross-section area, leading by Gauss' Law to a constant force at large distances, or a linearly rising potential. When this potential becomes strong enough, a new quark-antiquark pair is created, shielding the color charges of the original pair. Thus, quark confinement is an essential consequence of QCD. On the other hand leptons and the electroweak gauge bosons (γ, $W^{\pm}$, and Z^0) have no color charge. There are three color charges in QCD: red (r), green (g) and blue (b), which are the orthogonal states in $SU(3)$ color space. The $SU(3)$

color symmetry is exact, thus the strength of the QCD interaction does not depend on the color charge. Similarly to Quantum Electrodynamics (QED), where anti-particles carry opposite electric charge, in QCD anti-quarks carry opposite color charge: $\bar{r}, \bar{g}$ and $\bar{b}$.

In QED, the existence of virtual e^+e^- pairs surrounding the interacting fermions has the effect of screening their charge. At higher energies the effect of screening is reduced, naively because the length scale of the process is smaller, being inversely proportional to energy,thus the electromagnetic interaction is stronger at higher energies. For QCD, the cloud of virtual particles surrounding the quark is composed mainly of gluons, interacting with each other and with any probe particle in the neighbourhood. The effect is an anti-screening which makes the strong nuclear interaction weaker at higher energy. In the limit of infinite energy, asymptotic freedom is another important feature of QCD for quarks and this feature has important consequences as the existence of a high-energy high-pressure state called quark-gluon plasma.

Asymptotic freedom means that the binding between quarks becomes asymptotically weaker at high energy scales or small distances such that they can be considered as free particles, which subsequently allows the use of perturbation theory for calculations and predictions of production in high-energy collisions. This follows from the behavior of the *QCD effective coupling constant* $\alpha_s(Q)$ as a function of the momentum transfer Q (which sets the energy scale) [1] [12] given in Fig. 1.1. At high Q (short distances) $\alpha_s(Q)$ vanishes: quarks and gluons interact very weakly and a perturbative expansion in α_s can be performed.

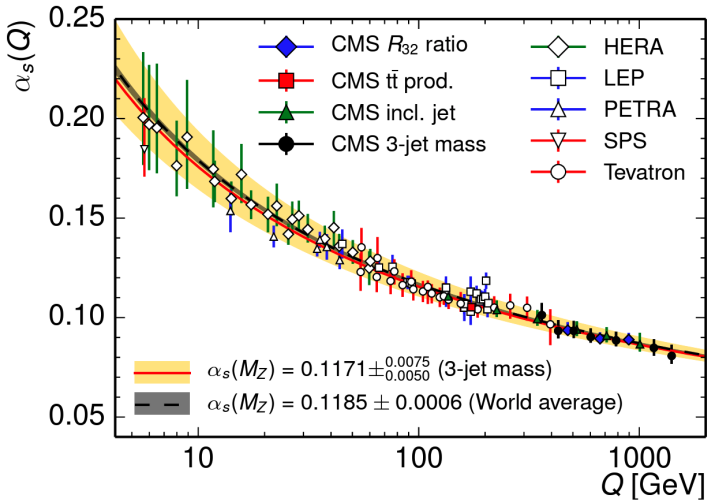


Figure 1.1: Summary of measurements of α_s as a function of the energy scale Q. Taken from [13].

Instead, at values of Q comparable to the QCD scale $\Lambda_{QCD} \approx 200$ MeV (high distances), the strong interaction amplitude becomes so large that quarks cannot behave as free particle ("confinement") and perturbation theory is no longer available and QCD is said to be in its non-perturbative regime. The non-perturbative regime requires effective theories that span from the simple bag models (describing the non-perturbative states and quark confinement by dividing the space in two regions inside and outside a non-perturbative bag), to the Lattice QCD, involving the usage of large computing clusters to simulate the QCD dynamics in a discrete space-time. It's the QCD scale that conventionally divides quarks into light and heavy. Strong interactions of hadrons containing heavy quarks are easier to understand than those of hadrons containing only light quarks. For heavy quarks the effective coupling constant $\alpha_s(m_Q)$ is small, implying that on length scales comparable to $1/m_Q$ the strong interactions are perturbative and much like the electromagnetic interactions, therefore easier to handle.

1.2.3 Hadrons with unconventional quarks content

At distance scales corresponding to the separations between quarks inside hadrons, QCD is a strongly coupled theory and perturbation theory is of limited applicability. Ultimately it is expected that numerical lattice QCD computations will generate predictions for QCD observables such as masses, transitions, and decays that can be tested by experiments. To date, models that incorporate general features of QCD have proven to be the most useful approach for describing the spectra and properties of hadrons. One prediction of these QCD-motivated models, supported by lattice QCD, is the existence of hadrons with more complex substructures than those of the simple $q\bar{q}$ mesons and the qqq baryons. Already in the seventies of last century it became clear that there was no fundamental reason why only three-quarks and quark-antiquark combinations should exist [14, 15] excluding other colorless bound quarks states or even *glueballs* (bound gluons states) and *hybrids* (bound states of gluons and quarks).

Despite considerable experimental effort, however, no clear evidence for hadrons with a non-CQM-like structure was found until the first years of the new millennium, when studies of the spectrum of charmonium mesons uncovered a number of meson candidates that do not seem to conform to CQM expectations. As already stressed, model calculations for heavy hadrons are easier and more reliable than for light ones; moreover spectra of heavy hadrons are much "cleaner" with respect to the dense spectra of light states. Therefore possible exotic states containing for example $c\bar{c}$ or $b\bar{b}$ are, in principle, expected to be handled in an easier formal framework.

Although the Higgs mechanism seems to be at the origin of the mass of all the Standard Model elementary particles like quarks, the bulk of the mass of protons and neutrons (which accounts for almost the entire mass of the visible universe) derives from the binding energy of innumerable gluons (the carriers of the strong nuclear force) that confine together the much lighter qaurks. How they are assembled remains still obscure. Knowing which hadrons exist, the pattern of their masses and structure, and what reactions they can undergo (the hadron spectroscopy) is directly related to the most fundamental issues of physics. In spite of 80 years of experimental and theoretical efforts, our understanding of hadrons remains fragmentary, as revealed by a large number of unexpected discoveries of "exotic" hadrons made in the last 20 years. Experimental spectroscopy not only is filling out "excited" multiplets of hadrons foreseen by the quark-model ("mesons" as $q\bar{q}$ bound states and "baryons" as qqq that is 3 quarks aggregates) but also lies at the forefront of new discoveries in High Energy Physics (HEP) and in this case we talk about *exotic spectroscopy*. The experimental assessment of the conventional states is a crucial aspect of the identification of possible exotic states of the matter like tetraquarks and pentaquarks.

The hadron spectroscopy program at CMS has already included some investigation of the B_c system which is experimentally challenging but also promising as proved by the observations of the two excited states $B_c(2S)^+$ and $B_c^*(2S)^+$. Indeed CMS can be very competitive (and complementary to LHCb) in final states with a B_c^+, as it will be better discussed later in this chapter.

1.3 Introduction to the B_c meson

The B_c meson is the ground state of the $c\bar{b}$ system and thus is the lightest meson consisting of two heavy quarks of different flavours. Being a double heavy-quark-antiquark bound state, a $c\bar{b}$ state is similar to the charmonium ($c\bar{c}$) and bottomonium ($b\bar{b}$) systems. In fact, the B_c meson bound-state dynamics can be treated in the non-relativistic expansion by QCD-inspired effective models worked out for the charmonium and bottomonium states. However, since it carries (open) flavour, the B_c meson has some specific production and decay peculiarities.

Fistly, the production mechanism for the $c\bar{b}$ system is different from that for $c\bar{c}$ and $b\bar{b}$ systems, because it requires the production of two heavy quark-antiquark pairs. The $c\bar{b}$ production cross section is lower than that for hidden-flavour quarkonia, because the leading-order diagrams are of higher order in the coupling constants and the phase-space is suppressed owing to the presence of the additional heavy quarks.

Moreover, being an open flavor state, the B_c meson decays via the weak interaction (i.e. it has no strong or electromagnetic decay channels). Thus, differently from the other heavy quarkonia, the B_c meson is a relatively long-living. On the other hand, its lifetime is shorter than that of the other beauty mesons since both the constituent quarks compete in the decay. Therefore, the study of B_c decays is of particular interest to improve the understanding of QCD dynamics and the measurement of the parameters of the weak Lagrangian. An accurate measurement of the B_c lifetime can also provide information on the masses of charm and beauty quarks.

The first observation of the B_c meson was at the end of the last century - at the Tevatron proton-antiproton collider - through its decay to the $J/\psi\ell^+\nu_\ell$ final state [16]. Since the uncertainties in the mass measurements are essential in the semileptonic channel, since the neutrino momentum is not detected directly, thus a first precise mass measurement of B_c meson came later with the first observation of the its much rarer decay to the $J/\psi\pi$ final state [17].

In the non-relativistic potential models the c and $\bar{b}$ quarks are tightly bound in a very compact system and have a rich spectroscopy of excited states. These spectroscopic aspects as well as the several decays characterizing the B_c meson will be discussed, in detail, in the following sections of this chapter, starting by some production aspects that will be introduced in the next section.

1.4 B_c meson production at the LHC

At the LHC design centre-of-mass energy (14 TeV), the total $b\bar{b}$ production cross section from the proton-proton collisions is assumed to be ~ 500 μb. This means the production of 10^{12} $b\bar{b}$ pairs in a standard (10^7 s) year of running at the operational luminosity of 10^{34} cm^{-2} s^{-1}. The large centre of mass energy implies that a complete spectrum of b hadrons will be available, including B_s, B_c and Λ_b mesons. However, less than 1% of all inelastic events contain b quarks, hence triggering is a critical issue. Moreover multiple pp collisions within the same bunch crossing (so-called *pile-up*) significantly complicate the reconstruction of the beauty hadron production vertex and decay vertex as happens for CMS and ATLAS. In contrast, at LHCb, the luminosity is locally controlled by appropriately focusing the beam to yield a mean value of instantaneous luminosity of 2×10^{33} cm^{-2} s^{-1}, where most events have a single pp interaction with an expected pile-up of around 5-6 in the LHCb Upgrade.

Among the LHC experiments, both the ATLAS and CMS detectors have the tracking systems covering the central rapidity region, $|\eta| < 2.5$, whereas the LHCb detector is a forward single-arm spectrometer covering $1.9 < \eta < 4.9$, and designed to perform B-physics throughout its lifetime. In view of the increasing pile-up and trigger rates (higher p_T threshold need to be applied at trigger level to keep the rates under control because of the limited bandwidth), it was imagined that the initial running period would have been most feasible for the B-physics in the CMS and ATLAS experiments, but, afterwords, the possibility to provide relevant results in the Heavy Flavour Physics would have been very limited (to channels like $B^0_{s,d} \to \mu^+\mu^-$ and not much more). In CMS this limitation has been overcome by a few concurrent development features, including the usage of GPUs in relevant part of the reconstruction of the events in the High Level Trigger and the development of the Tier0 capabilities, the development of trigger strategies, the adoption of partially delaying of the prompt reconstruction (in CMS this is called *B-parking*).

Figure 1.2 shows the η and p_T coverages of the detector tracking system for B physics measurements in the ATLAS, CMS and LHCb experiments.

The direct production mechanism of B_c is completely different from that of hidden flavour quarkonia since it requires the joint production of the two additional heavy quarks ($\bar{c}$ and b), which results in the low value of the B_c production cross section in comparison with the J/ψ and Υ families. On the

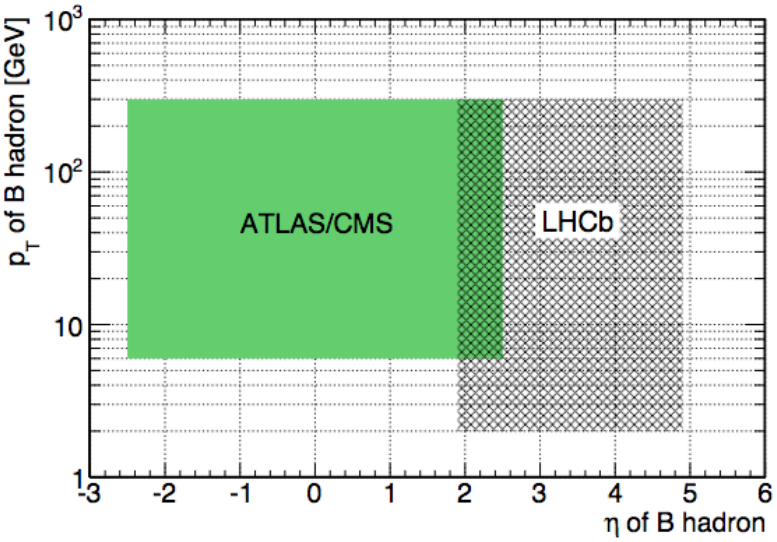


Figure 1.2: Beauty hadron transverse momentum p_T as a function of the pseudorapidity η, showing the (η, p_T) regions covered by ATLAS and CMS, compared to that covered by LHCb.

other hand, having no strong or electromagnetic annihilation decay channel, the excited higher-mass bound $c\bar{b}$ states, below the BD thresholds, cascade down into the ground pseudoscalar state through hadronic (dipion emission) and electromagnetic (photon emission) transitions into the lowest state with a probability close to unity, whereas the annihilation channels are essential for hidden flavour quarkonia.

The direct (+indirect) production of B_c mesons can be thought as a process made of two (or three) steps:

- a) the perturbative hard production of $c\bar{c}$ and $b\bar{b}$ pairs created by a parton collision;
- b) a soft non-perturbative binding of quarks in the final state, which can be estimated in the framework of non-relativistic potential models. As a result $\bar{b}$ quark and a c quark bind to form the B_c meson (direct production) in its ground state or in one of its excited states.
- c) The excited state cascades down to the ground state (indirect production).

As a consequence the yield of B_c mesons is estimated to be of the order of 10^{-3} with respect to the other beauty hadrons. The description of the direct production of the B_c meson is based on the factorization of hard parton production of heavy quarks ($b\bar{b}$, $c\bar{c}$) and the soft coupling of $c\bar{b}$ bound state. In the first stage, the hard subprocess can be reliably calculated in the framework

of QCD perturbation theory, while in the second stage, the quark binding in the heavy quarkonium can be described in the non-relativistic potential model approach. The short-distance cross section for creating the $c\bar{b}$ bound state in the color and a generic angular momentum state can be calculated as a perturbative expansion in α_s at scales of order m_c or larger. The lowest order mechanisms for creating the $(c\bar{b})$ are processes of the order-α_s^4 such as $q\bar{q}$ and $gg \to (c\bar{b}) + b + \bar{c}$. At LHC energies, the hadronic B_c production is dominated gluon-gluon fusion because of the high gluon density. To the leading approximation of QCD perturbation theory, the calculation requires contributions of 36 diagrams in the fourth order of the α_s. The typical Feynman diagrams are plotted in Figure 1.3 for illustration purposes.

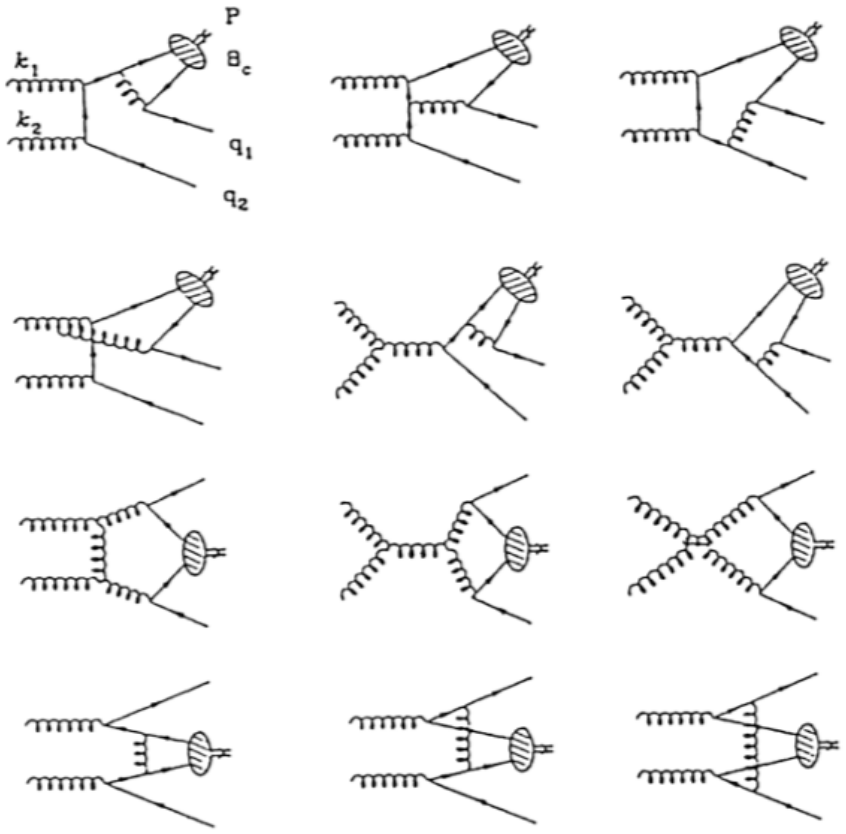


Figure 1.3: Typical Feynman diagrams for the B_c production in gluon-gluon fusion.

It should be mentioned that the production of B_c mesons can be calculated in perturbative QCD but only at transverse momenta (p_T) large compared to the B_c mass, because, in this case, the dominant process is that in which a b quark is produced by gluon fusion in the hard collision while the fragmentation provides the c quark. At lower transverse momenta of the B_c meson, a full α_s^4 calculation [18] is needed and, as already discussed, the dominant process is the one in which both the b and c quarks are produced in the hard scattering. The total uncertainty from combining all of the uncertainties in the direct cross-section for B_c production is less than an order of magnitude.

Ref. [18] and other earlier calculations provide inclusive production cross sections along with distributions in p_T and other kinematic variables. Inclusive production cross-sections should take into consideration the feed-down from all of the higher states of the $c\bar{b}$ system. Feed-down from directly produced S-wave excited states were first estimated in Ref. [19] and turned out to be larger than the cross-section for direct production by a factor ~ 5.4. The feed-down from directly produced B_c^* is relevant: its direct production is 2.4 times larger than that of the ground state (Ref. [20]) and the B_c^* decays to the ground state almost the 100% of the times. Quantitative predictions for the contribution to the inclusive B_c production cross-section from the feed-down from P-wave states came later (see, for instance, Ref. [21]).

1.4.1 The simulation of events with a B_c meson: the BCVEGPY generator

As earlier mentioned, the B_c production rate, requiring both a $\bar{b}b$ and a $\bar{c}c$ couple, is a factor $\sim 10^{-3}$ smaller compared to the single $\bar{b}b$ production. Thus, in the PYTHIA simulation program [22], only one B_c meson is produced out of 10^6 pp interactions. Therefore, a dedicated event generator for the B_c meson, BCVEGPY [23] [24], is used to enhance the simulation efficiency of B_c events, and overcome the low production rate due to the small B_c production cross section. The BCVEGPY package emulates the hadronic production of B_c^* mesons through the partonic subprocess $gg \to B_c^* + b + \bar{c}$ and is based on a complete calculation approach, namely the full pQCD complete calculation at the lowest order α_s^4. Contributions from the quark pair annihilation production mechanism are neglected because they are highly suppressed in pp collision environment.

The BCVEGPY package is integrated in the CMS software and interfaced with PYTHIA which hadronizes the whole event. PYTHIA tunes differ in the treatment of parton radiation and hadronization, and in the choice of underlying event parameters including those regulating parton showers, color reconnections, and cutoff values for the multiple parton interaction mechanism.

Values of these parameters were chosen to provide a reasonable description of existing LHC *pp* differential data measured in minimum-bias and hard QCD processes. Unstable particle decays are simulated with EVTGEN [25] and the detector response with GEANT4 [26].

1.5 The $(c\bar{b})$ mass spectrum

As an important family of hadrons, there have been few theoretical calculations of the B_c spectrum which is predicted to be very populated. Experimental progress towards establishing this spectrum has not been trivial. The poor situation on the observation and measurements of this spectrum is due to the expected low production yields, significantly smaller than those of the charmonium and bottomonium ($b\bar{b}$ and $c\bar{c}$) states.

The $c\bar{b}$ system is expected to have a rich spectroscopy or orbital and angular-momentum excitations. The $\bar{b}c$ mass spectrum is shown in Fig. 1.4. As already mentioned, having no strong or electromagnetic annihilation decay channel (cannot annihilate into gluons), the excited states below the BD thresholds (the lowest is at ≈ 7144 MeV) can only decay through radiative and hadronic (dipion) transitions to the B_c ground state, which decays only through the weak interaction. This results in a total width of $c\bar{b}$ excited states that is two orders of magnitude less than the total width of the charmonium and bottomonium excited levels, for which the annihilation channels are important. Consequently the $c\bar{b}$ states are more stable with a narrow width less than a few hundred keV.

The B_c spectrum can be estimated using non-relativistic potential models [27] and also perturbative QCD and lattice calculations. Measurements of the ground and excited states through fully reconstructed channels will allow to test the predictions of these models and to extract information on the strong interaction potential.

The unprecedented collision energies and integrated luminosities of the proton-proton (pp) data samples collected at the LHC allow, for the first time, detailed studies regarding the production and properties of B_c mesons and the exploration of the $(c\bar{b})$mass spectrum.

As an example, the two excited B_c states reported by CMS [28] were observed in the $B_c^+\pi^+\pi^-$ invariant mass spectrum exploiting a large sample of reconstructed $B_c^+ \to J/\psi\pi^+$ candidates collected during the LHC Run 2 data taking. While the $B_c(2S)^+$ decays directly to $B_c^+\pi^+\pi^-$, the $B_c^*(2S)^+$ is predicted to decay predominantly to $B_c^{*+}\pi^+\pi^-$, followed by the $B_c^{*+} \to B_c^+\gamma$ decay (see Figure 1.5). The emitted photon, due to its very low energy, is not reconstructed and therefore the observed $B_c(2S)^+$ signal appears to have a mass lower than the effective value.

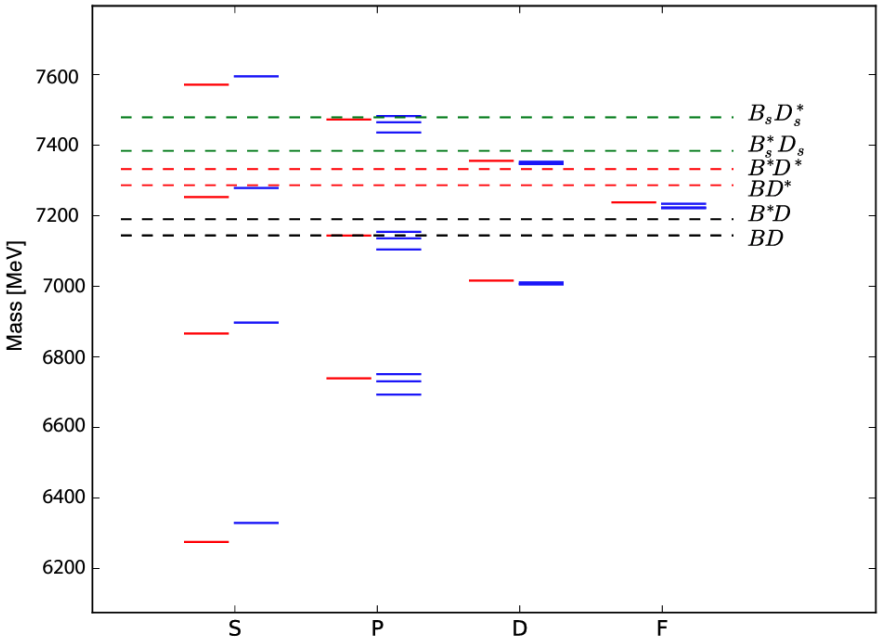


Figure 1.4: The $\bar{b}c$ spectrum shown here is borrowed from Fig.3 of [21]. Spin-singlet (-triplet) states are shown in red (blue) for each orbital-momentum family S, P, D, F (in standard notation), while the dashed lines indicate the possible BD thresholds.

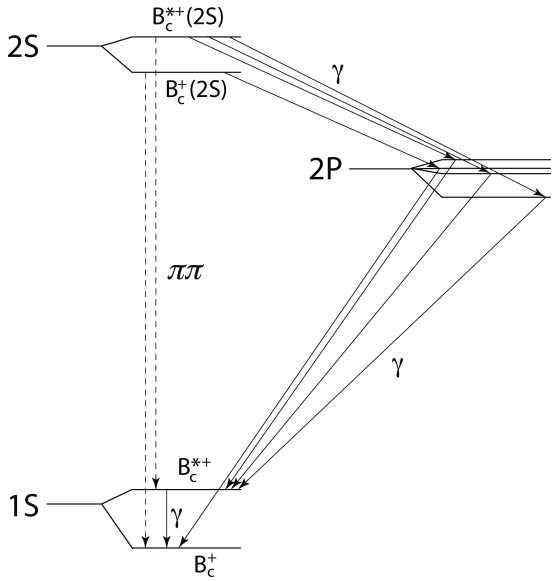


Figure 1.5: Scheme of the spectroscopic transitions among the excited $n = 2$ and $n = 1$ states of the $\bar{b}c$ system. Dashed lines represent dipion transitions.

1.6 B_c decays

As already mentioned, on one side, unlike the case of B^+, B^0 and B_s^0, the bound state dynamics of $\bar{b}c$ system can be systematically treated in a non relativistic expansion (which has proven to be very successful for the description of the quarkonia), on the other side, in contrast with the quarkonia ground states, the B_c^+ meson, laying below BD threshold, is stable for both the strong and electromagnetic annihilation interactions and thus can decay only via the weak interaction. With these properties B_c is in fact a unique example. Moreover, the $B_c^+(\bar{b}c)$ meson is also a particular state, because both heavy quarks compete in the decay. Indeed, the following three classes of transitions can be identified:

type-**I**: the decay of the $\bar{b}$-quark with the spectator c-quark ($b \to cW^{*+}$ where the asterisk highlights that it is a virtual W boson), leading to final states such as $J/\psi\pi$, $J/\psi K$, $J/\psi\,\ell\nu_\ell$ ($\ell = e, \mu, \tau$);

type-II: the decay of the c-quark ($c \to sW^{*+}$) with the spectator $\bar{b}$-quark, leading to final states like $B_s\pi$ and $B_s\ell\nu_\ell$;

type-III: the annihilation of $\bar{b}$ with c-quarks ($cb \to W^{*+}$), where the virtual W materializes into a pair lepton+neutrino ($B_c^+ \to \ell^+\nu_\ell$ where $\ell = e, \mu, \tau$) leading to final states like $\tau\nu_\tau$, or into a pair of quarks ($c\bar{s}$ favoured with respect to $u\bar{s}$ because $\|V_{cs}\| > \|V_{us}\|$, see Fig. 1.7) which then hadronizes, leading to final states like DK or multiple pions.

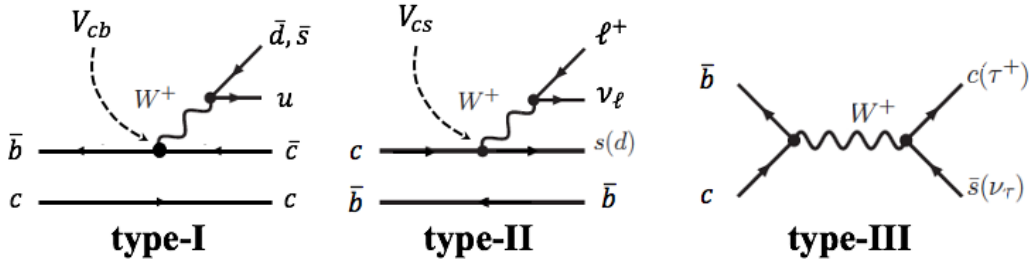


Figure 1.6: Quark diagrams for the B_c meson decays; as explained in the text they are: the c-spectator decay (type-I), the b-spectator decay (type-II) and the annihilation diagram (type-III).

The corresponding quark diagrams of these three expected contributions are shown in Fig. 1.6. Since these processes lead to different final states, their amplitudes do not interfere.

Both the $\bar{b}$-quark and c-quark can decay individually, thus overall resulting in a B_c lifetime which is about three times smaller than other ground

pseudoscalar b-flavoured mesons ($B_(u, d, s)$). Considering the hierarchy of the Cabibbo-Kobayashi-Maskawa (CKM) quark-mixing matrix elements for which $\|V_{cs}\| \gg \|V_{cb}\|$ (since $|V_{cb}\| \simeq 0.05$ and $\|V_{cs}\| \simeq 0.96$) (see Fig. 1.7), the second type of decays contributes for $\approx 70\%$ of the decay width, while the b-quark decays add about the $\approx 20\%$ and the rest $\approx 10\%$ is the contribution coming from the third class.

In the $\bar{b} \to \bar{c}c\bar{s}$ decays it is possible to estimate and separate also the Pauli Interference (PI) [29] with the c quark from the initial state, which is expected to contribute at the level of $\sim 1.5\%$, that has been already included in the b-quark decay fraction [30].

$$\begin{bmatrix} |V_{\mathrm{ud}}| & |V_{\mathrm{us}}| & |V_{\mathrm{ub}}| \\ |V_{\mathrm{cd}}| & |V_{\mathrm{cs}}| & |V_{\mathrm{cb}}| \\ |V_{\mathrm{td}}| & |V_{\mathrm{ts}}| & |V_{\mathrm{tb}}| \end{bmatrix} \approx \begin{bmatrix} 0.974 & 0.225 & 0.003 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{bmatrix}$$

Figure 1.7: CKM matrix and the numerical approximation of its elements.

Since these processes lead to different final states, their amplitudes do not interfere. When phase space and other effects are included, the predicted lifetime is in the range $0.4 - 1.4$ ps. Because of the wide range of predictions, a B_c lifetime measurement represents a test of the different assumptions made in the various calculations.

The tree-level Feynman diagrams for the Cabibbo-favoured and the Cabibbo-suppressed of the Type-I decays with a J/ψ in the final state are shown in Fig. 1.8; suppression is accounted by the fact that $\|V_{ud}\| > \|V_{us}\|$ (since $|V_{us}\| \simeq 0.225$ and $\|V_{ud}\| \simeq 0.974$) (see Fig. 1.7).

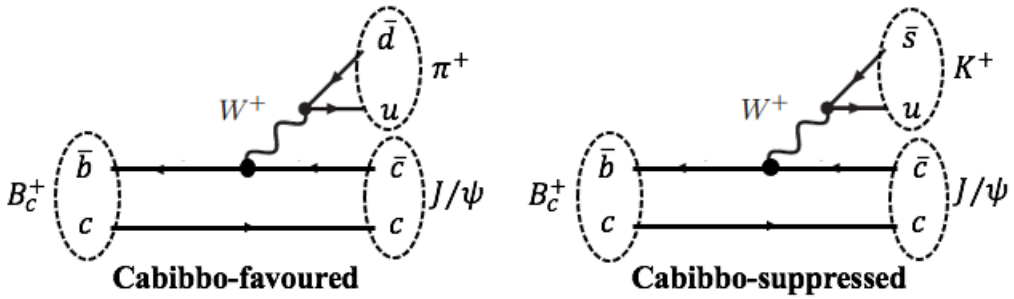


Figure 1.8: Type-I decay with a J/ψ in the final state. Left: the diagram for the $J/\psi\pi$ final state; right: the diagram for the $J/\psi K$ final state.

1.6.1 The B_c decays experimentally observed

Before the advent of the LHC, experimental measurements were limited to semileptonic channels and the fully reconstructed decays $B_c^+ \to J/\psi\pi^+$. The former has the advantage of a larger branching fraction and statistics, but it is not kinematically close (because of the escaping undetected neutrino); this explains why the former was the first decay observed but the latter has provided a precise determination of the mass (both at the CDF experiment).

The advent of the LHC has opened a new era for the experimental investigation of the B_c meson decays. The LHCb experiment has driven the field, enriching the list of the observed decay modes that are reported in Table 1.2. The $B_c^+ \to J/\psi\pi^+$ decay channel is often used as normalization channel to measure Branching Fraction ratios of B_c rare decays.

Decay channel	Type	Observation/measurements (Ref.)
$B_c^+ \to J/\psi\pi^+\pi^-\pi^+$	I	LHCb PRL 108 (2012) 251802 - CMS JHEP 01 (2015) 063
$B_c^+ \to J/\psi K^+$	I	LHCb JHEP 09 (2013) 075
$B_c^+ \to J/\psi K^+K^-\pi^+$	I	LHCb JHEP 11 (2013) 094
$B_c^+ \to J/\psi 3\pi^+2\pi^-$	I	LHCb JHEP 05 (2014) 148
$B_c^+ \to J/\psi D_s^{(*)+}$	I	LHCb PRD 87 (2013) 112012 - Atlas JHEP08 (2022) 087
$B_c^+ \to \psi(2S)\pi^+$	I	LHCb PRD 87 (2013) 071103 - LHCb EPJC 84 (2024) 468
$B_c^+ \to B_s^0\pi^+$	II	LHCb PRL 111 (2013) 181801 - LHCb JHEP 07 (2023) 066
$B_c^+ \to J/\psi p\bar{p}\pi^+$	I	LHCb PRL 113 (2014) 152003
$B_c^+ \to J/\psi\pi^+\pi^0$	I	LHCb arXiv:2402.05523
$B_c^+ \to J/\psi D^{(*)}K^{(*)}$	I	LHCb Phys. Rev. D95 (2017) 032005
$B_c^+ \to D^0K^+$	III	LHCb PRL 118 (2017) 111803

Table 1.2: Non exhausive list of the B_c meson decays observed so far. The typology for each decay mode is also indicated.

1.6.2 The B_c lifetime

The PDG average for the B_c lifetime [31] is (0.510 ± 0.009)ps and is dominated by two LHCb measurements, one exploting an high-statistics sample of semileptonic decays $B_c^+ \to J/\psi\,\mu^+\nu_\mu X$ [32],and one exploting the $B_c^+ \to J/\psi\,\pi$ decays. [33]

The PDG average incorporates also the CMS measurement (carried out with the 8 TeV dataset), $(0.541 \pm 0.026 \pm 0.014)$ps [33], obtained by measuring the B_c^+ lifetime with respect to the B^+ one, in order to reduce the systematic uncertainties that is possible since both decays have a dimuon plus one track (J/ψ + track with the track being either a π or a K) in the final state and a displaced decay vertex; this measurement is on the "high side" but still compatible with the world average guided mostly by the LHCb measurements.

1.7 Prospects in the B_c physics: excited states

In this thesis the extraction of a signal associated to the $B_c \to J/\psi\pi$ decay in the Run-3 CMS data is discussed in depth, also illustrating with a great level of details the used methods and techniques. In this section the special role of this decay as the preliminary step for ongoing or future data analyses at the LHC experiments is discussed and well highlighted.

1.7.1 Recently observed excited states in the $\bar{b}c$ spectrum

As discussed earlier, the $\bar{b}c$ spectrum below the BD thresholds (the lowest at ~ 7144 MeV) is characterized by excited states that cascade into the B_c ground state through one or more transitions (radiative or with a dipion emission). Only the two $2S$ radially excited B_c states have been observed so far, namely about all the spectrum has still to be experimentally investigated. This is currently possible only at some of the LHC experiments (LHCb, CMS and possibly ATLAS). CMS was the first experimental collaboration to observe them by disentangling this spin doublet, namely the two hyperfine excited partners $B_c(2S)^+$ and the $B_c^*(2S)^+$ [28].

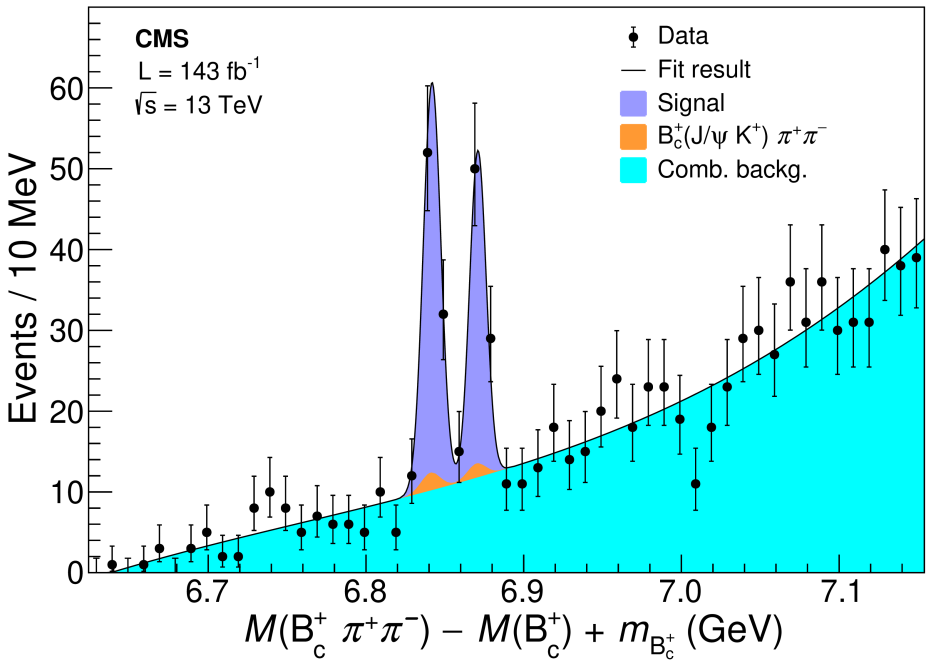


Figure 1.9: Invariant mass distribution of the $B_c\pi^+\pi^-$ candidates with the two disentangled signals; the $B_c^*(2S)^+$ peak appears at lower mass then the $B_c(2S)^+$ peak, as explained in the text.

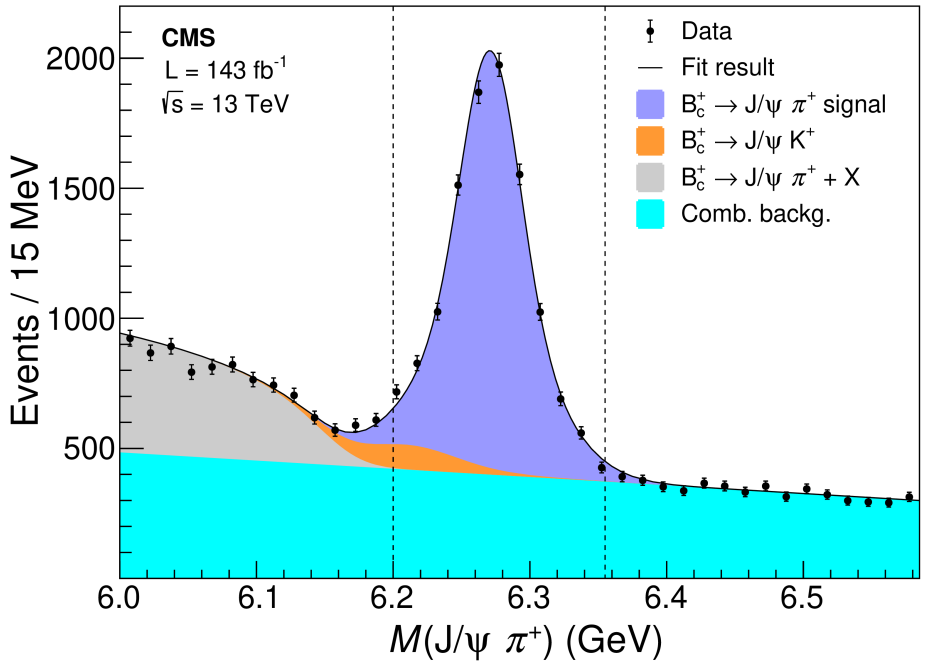


Figure 1.10: Reconstructed $B_c \to J/\psi\pi$ signal used to observe the $2S$ radially excited doublet (see next figure) in Ref. [28]. The peak half-width, driven by the mass resolution, is $\sigma = (33.5 \pm 2.5)$ MeV.

The two states were observed in the $B_c^+\pi^+\pi^-$ invariant mass spectrum, as shown in Fig. 1.9, by exploiting a large sample of reconstructed $B_c \to J/\psi\,\pi$ candidates, collected during the LHC Run 2 data taking and shown in Fig. 1.10. The reported B_c^+ yield is 7629 ± 225 extracted through a selection establishing the kinematical fiducial region defined by $p_T(B_c) > 15$ GeV and $|y(B_c)| < 2.4$.

How to understand which peak represents which excitation? Having in mind Fig. 1.5 for visual guide, the short explanation is the following. While the $B_c(2S)^+$ decays directly to the $B_c^+\pi^+\pi^-$ final state, the $B_c^*(2S)^+$ is predicted to decay predominantly to $B_c^{*+}\pi^+\pi^-$, followed by the $B_c^{*+} \to B_c^+\gamma$ decay; the emitted photon, due to its very low energy, cannot be reconstructed and therefore the observed $B_c^*(2S)^+$ signal appears to have a reconstructed invariant mass lower than the effective value, as it can be appreciated in Fig. 1.9.

In a subsequent paper [28] the CMS Collaboration also published the measurement of the ratios (R^+, R^{*+}) of production cross sections (times branching fractions) of these two radial excitations. Specifically for each of the two excited states the ratio with respect to the B_c^+ ground state was measured, as well as the production ratio between themselves:

$$\begin{aligned} \mathcal{R} = \frac{R^{*+}}{R^+} &= \frac{\sigma(B_c^*(2S)^+)}{\sigma(B_c(2S)^+)} \cdot \frac{\mathcal{B}(B_c^*(2S)^+ \to B_c^{*+}\pi^+\pi^-)}{\mathcal{B}(B_c(2S)^+ \to B_c^+\pi^+\pi^-)} = \\ &= \frac{\mathcal{N}(B_c^*(2S)^+)}{\mathcal{N}(B_c(2S)^+)} \cdot \frac{\epsilon(B_c(2S)^+)}{\epsilon(B_c^*(2S)^+)} \end{aligned} \tag{1.3}$$

Experimentally the ratio $\mathcal{R}$ is obtained from the extracted yields associated to the two excitations corrected for the combination of detection efficiencies and acceptances. CMS measured [28] $\mathcal{R} = 1.35 \pm 0.32(stat) \pm 0.09(syst)$.

Decay Mode	k_γ [keV]	Branching Fraction (%)
	1^1S_0 (6275) : weak decays	
	1^3S_1 (6329) : $\Gamma = 0.144$ keV	
$1^1S_0 + \gamma$	54	100
	2^3P_0 (6692) : $\Gamma = 53.1$ keV	
1^3S_1 (6329)	354	100
	$2P_1$ (6730) : $\Gamma = 72.5$ keV	
1^3S_1 (6329)	389	86.2
1^1S_0 (6275)	440	13.7
	$2P_1'$ (6738) : $\Gamma = 99.9$ keV	
1^1S_0 (6275)	448	92.4
1^3S_1 (6329)	397	7.51
	2^3P_2 (6750) : $\Gamma = 79.7$ keV	
1^3S_1 (6329)	409	100
	2^1S_0 (6866) : $\Gamma = 73.1$ keV	
$1^1S_0 + \pi\pi$		81.1
$2P_1'$ (6738)	126	16.5
$2P_1$ (6730)	134	2.24
	2^3S_1 (6897) : $\Gamma = 76.8$ keV	
$1^3S_1 + \pi\pi$		65.0
2^3P_0 (6692)	201	7.66
$2P_1$ (6730)	165	11.5
$2P_1'$ (6738)	157	1.13
2^3P_2 (6750)	145	14.6

Figure 1.11: Table VI of Ref. [21] is reported here. Branching fractions are estimated to be 1.0 for $B_c^{*+} \to B_c^+\gamma$, 0.81 for $B_c(2S)^+ \to B_c^+\pi^+\pi^-$ and 0.65 for $B_c^*(2S)^+ \to B_c^{*+}\pi^+\pi^-$.

It is important to experimentally determine this ratio since different models [34] [21] can bring to relevantly different predictions for the two $2S$ excitations. Considering the branching fractions in Table VI of Ref. [21], reported here in Fig. 1.11, and the production cross sections in Table VIII of Ref. [21] given here in Fig. 1.12, the authors in Ref. [21] can estimate the following projected ratio:

$$R_{EQ} = \frac{\sigma(B_c^*(2S)^+)}{\sigma(B_c(2S)^+)} \cdot \frac{\mathcal{B}(B_c^*(2S)^+ \to B_c^{*+}\pi^+\pi^-)}{\mathcal{B}(B_c(2S)^+ \to B_c^+\pi^+\pi^-)} \approx \frac{41.72}{16.94} \cdot \frac{0.65}{0.81} \approx 2.46 \cdot 0.8 \approx 1.97 \tag{1.4}$$

Of course, if the two $2S$ states were produced with equal rate, an $R_{EQ} \approx 0.8$ would be predicted. The estimated R_{EQ} is compatible with the CMS measurement reported above, $\mathcal{R} \simeq 1.35 \pm 0.33$, within two standard deviations.

TABLE VIII. Production rates (in nb) for $(c\bar{b})$ states in pp collisions at the LHC. The production rates were calculated using the BCVEGPY2.2 generator of Ref. [45], extended to include the production of $3P$ states. Color-octet contributions to s-wave production are small; we show them (following |) only for the $1S$ states.

$(c\bar{b})$ level	$\sigma(\sqrt{s} = 8$ TeV)	$\sigma(\sqrt{s} = 13$ TeV)	$\sigma(\sqrt{s} = 14$ TeV)
1^1S_0	46.8 | 1.01	80.3 | 1.75	88.0 | 1.90
1^3S_1	123.0 | 4.08	219.1 | 6.97	237.0 | 7.55
2^3P_0	1.113	1.959	2.108
2^3P_1	2.676	4.783	5.214
2^1P_1	3.185	5.702	6.166
2^3P_2	6.570	11.57	12.64
2^1S_0	9.58	16.94	18.45
2^3S_1	23.46	41.72	45.53
3^3P_0	0.915	1.642	1.806
3^3P_1	2.263	4.082	4.478
3^1P_1	2.695	4.817	5.287
3^3P_2	5.53	9.98	10.90
3^1S_0	4.23	7.53	8.08
3^3S_1	10.16	18.21	19.83

Figure 1.12: Table VIII of Ref. [21] is reported here. Production cross sections of $B_c^*(2S)^+$ and $B_c(2S)^+$ are estimated to be 16.94 nb and 41.72 nb respectively, for pp collisions at center of mass energy $\sqrt{s}$=13 TeV.

1.7.2 Search for further excited states in the $\bar{b}c$ spectrum

With reference to the mass spectrum shown in Fig. 1.4 a crucial role is played by the several thresholds (BD, B^*D, BD^*, B^*D^*, $B_s^*D_s$, $B_sD_s^*$) since the dissociation (strong decay), into a pair of one beauty and one charmed mesons will dominate the decays of the states above the relevant thresholds. Beyond the states shown in Fig. 1.5, the $1D$ doublet state is calculated to be well below the lowest threshold whereas the $2P$ multiplet is either slighlty below or around the BD thresholds. The $3S$ doublet is approximately located above B^*D and below BD^* thresholds. Let us discuss in detail what are the perspectives for the seach of the mentioned excited states.

1.7.2.1 $3S$ states

The calculations of the authors in Ref. [21] indicate that the $3S$ levels will lie closely above the relevant open flavour thresholds: the 3^1S_0 and 3^3S_1 masses have been calculated to be 7253 MeV and 7279 MeV respectively, with an indicative uncertainty of about $\pm$20 MeV, whereas the BD, B^*D thresholds are at $\approx$7150 MeV and $\approx$7195 MeV respectively. However the authors comment that it is conceivable that coupled-channel effects might push one or both states at lower mass values, and that, for this reason, it is experimentally worthy examining the $B_c\pi^+\pi^-$ mass spectrum beyond 7200 MeV for indications of both $3^1S_0 \to B_c\,\pi^+\pi^-$ and $3^3S_1 \to B_c^*\,\pi^+\pi^-$ signals. As far as the $3S$ production cross sections estimates are concerned, according to the table in Fig. 1.12, the two $3S$ states are produced at about 44% of the rate for their $2S$ counterparts. Thus, if their decay rates are not strongly affected by the thresholds, the chances to observe them adding the data currently available from Run-3 to the full Run-2 used in Ref. [28]. This will be clearer through the chapters of this thesis that illustrate how enhanced yields of reconstructed $B_c^+ \to J/\psi\pi^+$ candidates can be obtained in Run-3 data by also exploting a dedicated B_c high level trigger.

The search for the $3S$ states would follow the same procedure as in the observation of the $2S$ states [28] that can be briefly summarized as follows. The $B_c\,\pi^+\pi^-$ candidates are reconstructed by performing a kinematic fit, combining a selected B_c candidate with two opposite-sign charged tracks (assumed to be pions) and imposing a common vertex geometrical constraint. The two daughter pions used in the vertex fit are taken from the associated Primary Vertex as the lifetime of an excited B_c state is assumed to be negligible with respect to the experimental decay time resolution. Only B_c candidates in the so-called signal mass window will be used; this window is defined by the measured B_c mass and the mass resolution and its range is chosen to minimize the contamination from partially reconstructed decays (see Fig. 1.10).

1.7.2.2 $2P$ and $3P$ states

From the point of view of their masses with respect to the BD thresholds, the $2P$ and $3P$ states look more promising for their observation since they are predicted below the BD thresholds for strong decays. As can be seen in Fig. 1.5 the 2P multiplet states are expected to decay radiatively towards the B_c^* excited $1S$ state. The transition would be to $B_c^*\gamma$ followed by the undetectable decay $B_c^* \to B_c\gamma$; indeed, whereas the former photon can be detected exploiting the photon conversion mechanism, the latter is too soft and remains undetected. For instance, let us consider the $2^3P_2(6750)$ state which has the most favourable (calculated) production cross section: about $\approx 80\%$ of that calculated for the 2^1S_0, according to the table in Fig. 1.12. The transition $2^3P_2(6750) \to B_c^*\gamma$ is particularly attractive and the expected photon energy is ≈ 410 MeV. These photons could be experimentally detected by conversion with a not negligible efficiency at CMS. The photon candidates are obtained from the reconstruction of the momentum of the electron–positron pair originating from the photon conversion in the beam pipe or in the inner layers of the CMS silicon tracker. The standard algorithm used to reconstruct the converted photons uses only information from the silicon tracker. Specifically, the photon candidates are formed from two oppositely charged tracks with small angular separation, a small distance of closest approach, a conversion vertex at least 1.5 cm away from the beam axis, and a kinematic fit imposing zero mass and a common vertex. Additional calibrations studies are needed, e.g, the Photon Energy Scale calibration (PES), as in the CMS analysis that brought to the observation of the $\chi_{b1,b2}(3P)$ states via their radiative decay to $\Upsilon(3S)$ [35].

In the phenomenological models the $3P$ states lie just below, or near enough, the BD threshold. From the point of view the efficiency in reconstructing converted photons the radiative decays of the $3P$ states can be even more interesting because photons are energetically harder! However, looking at the production cross sections in the table of Fig. 1.12, the $3P$ states are characterized by a bit lower values than those characterizing the $2P$ ones, that in turn are suppressed with respect to the production rates of $2S$ states. Thus it seems feasible to look for them, in the CMS experiment, only once the full Run-3 dataset will be available.

1.7.2.3 $3D$ states

Finally, an observation of signals associated to D-wave states in the $B_c\pi^+\pi^-$ final state has been estimated in Ref. [34] and appears to be rather experimentally challenging even when adding Run-2 and Run-3 datasets because of the relatively low expected yields. Of course, in the analysis presented in Ref. [28], there is no hint of a $3D$ state in the reconstructed $B_c\,\pi^+\pi^-$ invariant mass at

the full Run-2 available statistics.

1.7.2.4 Possible exotic states

Some "complications" with the B_c spectroscopy may arise in the energy region of strong-decay meson–meson thresholds in the $\bar{b}c$ spectrum. In fact, in this energetic region, "B_c-like" structures might appear, similarly to the charmonium-like states found while investigating the charmonium spectrum. Ref. [36] is an example of theoretical paper in which the authors explore potential analogs of the $X(3872)$ state in the B_c spectrum.

1.8 Prospects in the B_c physics: observation of doubly-beauty baryons and tetraquarks

In the following part the signature of displaced B_c meson as the key to search for doubly-beauty baryons and tetraquarks will be discussed with some detail and projections.

1.8.1 Search for signatures of double beauty hadrons

The discovery of the doubly-charmed baryon Ξ_{cc}^{++} by LHCb [37] has triggered a lot of theoretical and phenomenological interest in the not-observed-so-far doubly-beauty baryons (such as $\Xi_{bb}^{0}(bbu)$, $\Xi_{bb}^{-}(bbd)$ and $\Omega_{bb}^{-}(bbs)$).

As it will be discussed soon, they share the same specific signatures that would have the hypothetical doubly-beauty tetraquarks. In fact, based on the first experimental observation of the doubly-charmed baryon as well as on Heavy Quark Symmetry (HQS) considerations, recent theoretical insights have brought the possibility that doubly-heavy tetraquarks (DHTQ), $T(Q_iQ_j\bar{q}_k\bar{q}_\ell)$, should exist in the HQS limit. Here Q_i, Q_j are either b or c quarks and the light antiquarks are $\bar{u}$, $\bar{d}$ and $\bar{s}$. In particular several theoretical papers (among them Ref. [38]) indicate that some of the ground states of the DHTQs (with quark content such as $bb\bar{u}\bar{d}$, $bb\bar{u}\bar{s}$ and $bb\bar{d}\bar{s}$) should be stable against strong and electromagnetic decays, because they are characterized by mass values below the $B\bar{B}^*$ thresholds, and thus they should be able to decay only weakly.

Since these tetraquarks below threshold should only decay weakly, they are expected to have comparable lifetimes to the ground state of doubly-beauty baryons, namely in the range $0.4-0.8$ ps. In other words, having a measurable lifetime is a characteristic signature either for the doubly-beauty baryons and for the hypothetical tetraquark states with two beauty quarks [39].

The typical expected decay of the ground state of doubly-beauty baryons is depicted in the diagram of Fig. 1.13. Here one of the b quarks in a bbq hadron decays via a $b \to \bar{c}$ transition and the produced antiquark hadronizes with the remaining b quark producing a B_c^+ meson. It is worthy to remind here that the transition $b \to c\bar{c}s$ is the dominant one according to the CKM matrix.

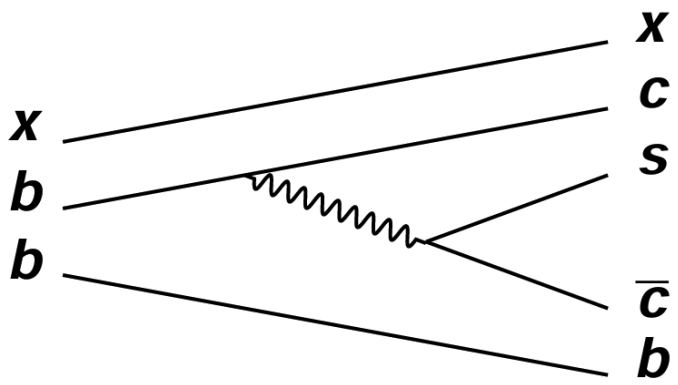


Figure 1.13: Diagram for the decay of a bbq baryon (q is the "spectator" light quark) into a B_c^- meson (with $b\bar{c}$ quark content) plus a charmed baryon. Diagram borrowed by Ref. [39].

At this point is crucial to notice that the weak decays of doubly-beauty hadrons (either baryons and stable DHTQs) are the **only** possible source of B_c^+ mesons that do not originate directly from the Primary Vertices of the LHC collisions. This is a direct consequence of the following facts: 1) all excited states of particles heavier than the b quark decay instantaneously, hence do not produce a displaced vertex, hence do not produce a displaced vertex; 2) any hadron containing both b and $\bar{b}$ constituent quarks will decay via strong or electromagnetic interactions, and hence does not have a measurable lifetime (bottomonium states are an obvious example).

From the experimental point of view one would expect a *signature of displaced B_c^+ mesons*; this specific signature should allow an inclusive approach as explained better later, since the exclusive approach appears to be not viable. In fact the exclusive reconstruction implies to be able to reconstruct the full decay chain of few decays, each of them with a typical branching fraction of the order of $10^{-2} - 10^{-3}$; furthermore, the final state would involve many particles and the overall reconstruction efficiency can be lower than $10^{-2} - 10^{-3}$.

The key point for an inclusive approach is that the experimental signature would manifest itself as a tail in the distribution of the B_c meson proper time or of the impact parameter (IP) with respect to the PV, as can be imagined with the aid of Figure 1.14 that sketches the topology of the decay chain originated by the doubly-beauty baryon. The DHTQs and the ground state of double beauty baryons are expected to have comparable lifetimes in the range

$0.4 - 0.8$ ps [39] (for reference, B_c lifetime is 0.5 ps) and this explains why the experimentally reconstructed B_c candidates would be characterized by a tail in their proper time distribution associated to the "displaced" component. As explained in Chapter 2, a well-designed inclusive J/ψ+track(s) "B_c-trigger" has been put in production in the CMS Run-3 data taking to allow the kind of searches just discussed; here "inclusive trigger" means "not-displaced" and allows to reconstruct the full proper decay time distribution.

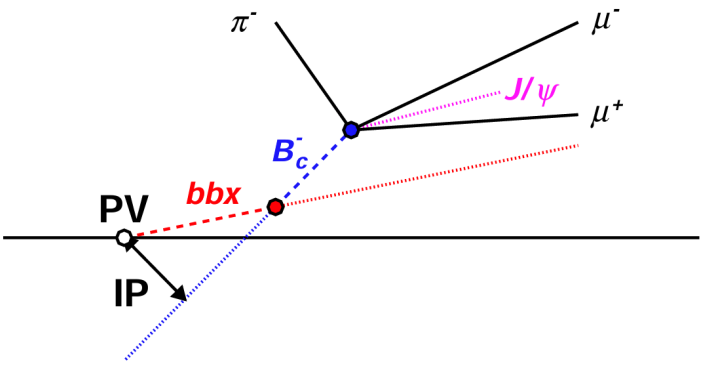


Figure 1.14: Sketch, borrowed by Ref. [39], of the *signature of displaced B_c^+ mesons*. The doubly-beauty baryon is promptly produced in the Primary Vertex (PV) and "travels" away decaying in the - well detached - secondary vertex (SV)(the red small circle). The B_c^+ meson will also "fly" away ending up in its decay to $J/\psi\pi^+$ in the - enough well detached - tertiary vertex (TV) (the blue small circle). The SV (TV) represents the B_c production (decay) vertex.

1.8.2 Estimation of the experimental sensitivity for the search of double beauty hadrons

The signature of displaced B_c^+ mesons seems to be quite attractive experimentally not only for the LHCb experiment [39] but also for CMS.

It is possible to roughly estimate the yields that could be expected for displaced B_c mesons as signature of weakly decaying doubly-beauty hadrons. The potential yields can be roughly estimated following the logic in Ref. [39] as done in the out-of-the-box calculation (characterized by relevant theoretical and experimental uncertainties and assumptions) adapted to the CMS experiment [40]. The estimation procedure and the involved assumptions are summarized in the following Table 1.3:

Numeric input	**LHCb** [39]	**CMS**
Run2+3 Integrated Luminosity	$30 fb^{-1}$	$300 fb^{-1}$
bbx hadron production cross section [38]	$\sim\mathcal{O}(1nb)$	$\sim\mathcal{O}(1nb)$
branching fraction for inclusive $bbx \to B_c^+ X$ [41]	$\sim$10%	$\sim$10%
branching fraction: $B_c^+ \to J/\psi\pi^-$ [1]	$\sim$2%	$\sim$2%
further branching fractions: $J/\psi \to \mu^+\mu^-$ [1]	$\sim$6%	$\sim$6%
overall detection efficiency	$\sim$ 10%	$\sim$ 0.5%

Table 1.3: Assumed numerical inputs to estimate the number of B_c^+ displaced mesons, following Refs. [39] [40].

The authors in Ref. [39] estimate 10 displaced B_c mesons per fb^{-1} for the LHCb experiment, thus an integrated amount of $\sim$300 displaced B_c candidates. For CMS, with respect to LHCb, there is factor 10 due to the higher integrated luminosity, but also a factor 1/20 because of the lower overall reconstruction efficiency; therefore one would approximately expect, for CMS, half of the displaced B_c mesons of LHCb, namely $\sim$ 150 candidates after adding Run-2 and Run-3. Clearly LHCb is still favoured in the searches for doubly-heavy candidates, but CMS can be competitive. For both experiments, but especially for CMS, the careful design of a dedicated trigger in Run-2 is crucial to be competitive. As discussed later an inclusive B_c trigger configured for the Run-3 data taking of CMS was configured and has been put in production to deal with this need.

Chapter 2

The CMS experiment at the LHC

The Compact Muon Solenoid (CMS) experiment [42] is a general-purpose particle physics detector at the Large Hadron Collider (LHC) [43] at the CERN laboratories. This chapter introduces LHC, presents briefly the general design of CMS, provides the most relevant details for the sub-detectors involved in the present analysis, namely the silicon Tracker and the Muon Chambers' System and describes the trigger strategy used to collect the sample for the $B_c^+ \to J/\psi\pi^+$ study case.

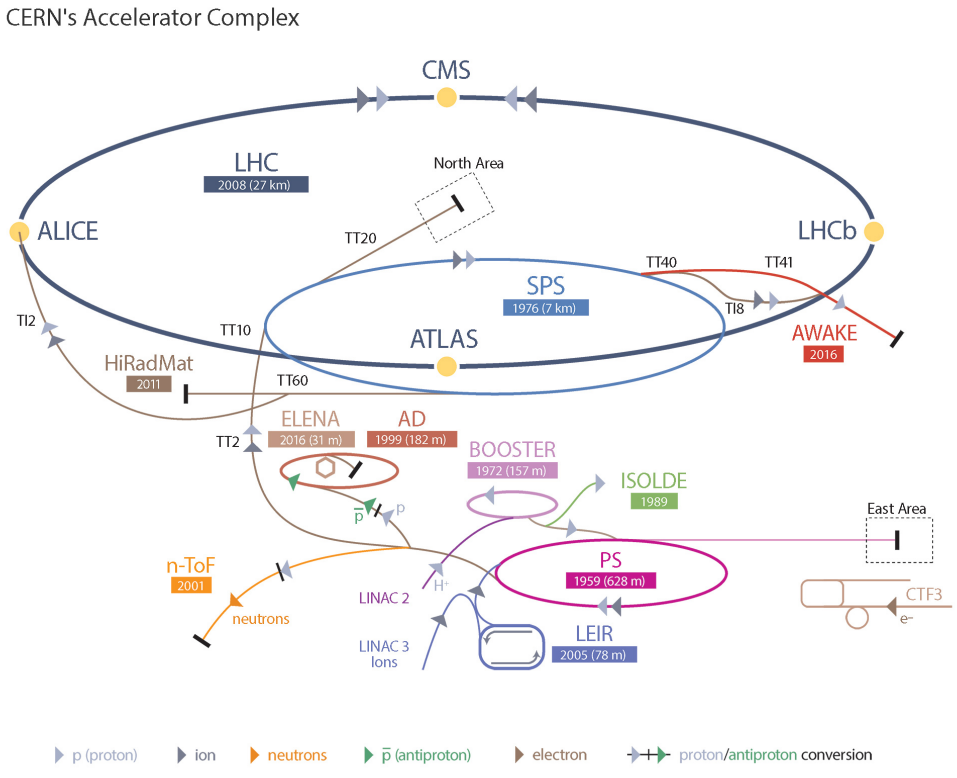


Figure 2.1: The LHC accelerator complex.

2.1 The Large Hadron Collider

The LHC is the last of a chain of accelerators as depicted in Figure 2.1. The LHC hosts four major experiments, CMS, ATLAS, ALICE and LHCb, and has been operating proton-proton (pp) collisions at high energies since 2010. The first three years are referred to as Run-1 and the center-of-mass energy ($\sqrt{s}$) was limited to 7 TeV in 2010 and 2011, and 8 TeV in 2012. During the Long Shutdown 1 (LS1) from 2013 to 2015, the machine was upgraded. Then the LHC Run-2 started from 2015 and took data till the end of 2018 at the center-of-mass energy of 13 TeV. Further upgrades came during the following Long Shutdown 2 (LS2), from 2019 to 2021, then the machine was restarted in 2022 and is operating at $\sqrt{s} = 13.6$ TeV until 2025. This four-year period is called Run-3. The LS2 upgrades included the development of new triggers for Heavy Flavour analyses; this thesis presents a preliminary analysis on data collected in 2022 and 2023 by the new trigger dedicated to the B_c meson.

2.1.1 Proton-proton interactions

The amplitude of the pp interaction is quantified by the corresponding cross-section, which consists of an elastic and an inelastic part. The elastic cross-section corresponds to all processes where the protons exchange energy but are not destroyed. In inelastic collisions instead the protons are broken to produce new hadrons. Figure 2.2 shows an overview of different measurements of the total, elastic and inelastic pp cross-section as a function of the center-of-mass energy.

In addition to their small cross-section, elastic collisions lead to a very small deflection of the resulting protons with respect to the beam axis. The final state protons are often not even entering the detector volume. What is detected by the CMS experiment is the final state of *pp* inelastic collisions for which the deflection of the final state particles with respect to the z-axis is large enough. As for any inelastic QCD process, the nature of those interactions can be split into two categories: soft and hard interactions. A hard QCD interaction implies a large momentum transfer between the interacting partons. These are the events which high-energy physicists are mostly interested in, because a large energy is available for new physics signatures such as the production of new heavy particles. In these interactions, two partons of the incoming protons interact in a way that can be described by perturbation theory. The rest of the protons will hadronize in soft processes: this is the underlying event described by non-perturbative QCD models.

When proton bunches collide inside the central part of the beam pipe of the CMS detector, only a very small fraction of the $\sim 10^{11}$ protons interact. The

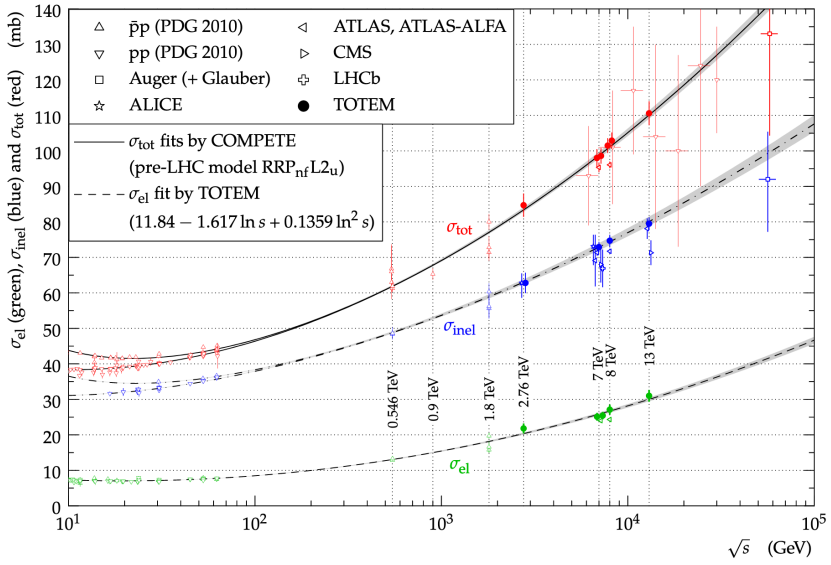


Figure 2.2: Overview of elastic (σ_{el}), inelastic (σ_{inel}), total (σ_{tot}) cross section for pp and $p\bar{p}$ collisions as a function of $\sqrt{s}$, including TOTEM measurements over the whole energy range explored by the LHC. Uncertainty band on theoretical models and/or fits are as described in the legend. The continuous black lines (lower for pp, upper for $p\bar{p}$) represent the best fits of the total cross section data by the COMPETE collaboration. The dashed line results from a fit of the elastic scattering data. The dash-dotted lines refer to the inelastic cross section and are obtained as the difference between the continuous and dashed fit curves [44].

number of extra pp interactions per bunch-crossing is called *pileup*. However, in the LHC/CMS jargon we typically refer to the *pileup* (denoted as PU) as the total number of pp interactions per bunch-crossing. Unfortunately, the pp inelastic cross-section is dominated by the soft cross-section [45], namely most of the pileup interactions are soft QCD processes (sometimes these are called *minimum bias* events). In these interactions, low momentum particles (mainly light hadrons) are produced [46].

If a hard QCD interaction occurs during a bunch-crossing, all the particles produced in the soft QCD pileup interactions will superimpose in the event. This has to be disentangled by the data analysis since this adds to the event description energy that does not come from the hard interaction. Not all PU interactions are reconstructed by the detector. If only very low energy particles are produced, it may not be possible to reconstruct the original spatial position of the interaction. What is called *primary vertex* (PV) of the interaction represents the spatial position of the hard QCD interaction and has typically the largest number of associated produced particles with respect to the vertices associated to PU interactions. Every year LHC improves its performances and pushes towards higher PU: the evolution of the PU profile throughout the years is shown in Figure 2.3.

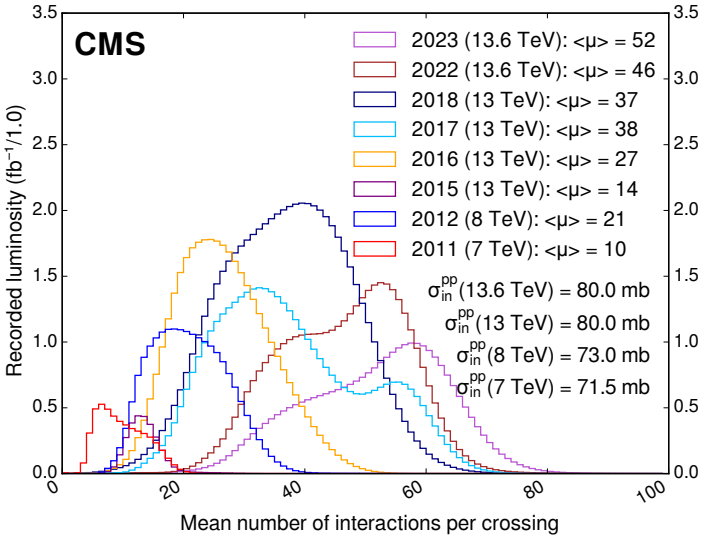


Figure 2.3: Evolution of the recorded pileup profile at CMS since the start of LHC. [47]

2.1.2 Instantaneous and integrated luminosity

The instantaneous luminosity $\mathcal{L}_{inst}$ is a measurement of the number of collisions the LHC produces per second. It is a purely machine-dependent variable. Given a process $pp \to X$, it links the rate of observed events $R \equiv dN/dt$ to its cross-section $\sigma(pp \to X)$:

$$\frac{dN}{dt} \equiv R = \mathcal{L}_{inst} \cdot \sigma(pp \to X) \tag{2.1}$$

It is measured in *inverse femtobarns per second* ($\mathrm{fb}^{-1}\mathrm{s}^{-1}$) (typically used at LHC) or in inverse squared centimeters per second cm^{-2} s^{-1} (with 1 barn being equal to 10^{-24} cm^2). The integrated luminosity L_{int} is the integral of the instantaneous luminosity over a time interval of data taking ($L_{int} = \int \mathcal{L}_{inst}\, dt$), and it is measured in fb^{-1}. The instantaneous luminosity depends on a number of variables as given by the equation:

$$\mathcal{L}_{inst} = \frac{N_b^2 n_b f_{rev} \gamma}{4\pi \epsilon_n \beta^*} F \tag{2.2}$$

where N_b is the number of protons per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ is the Lorentz factor and ϵ_n is the normalized transverse beam emittance. The latter describes the beam spread in the position-momentum phase space. β^* is the value of the function β, which describes the transverse size of the beam, at the interaction point; β^* is strictly related to the pileup. Finally F is a luminosity reduction factor to cope for the non-zero crossing angle between the beams. Beams must cross at a non-zero

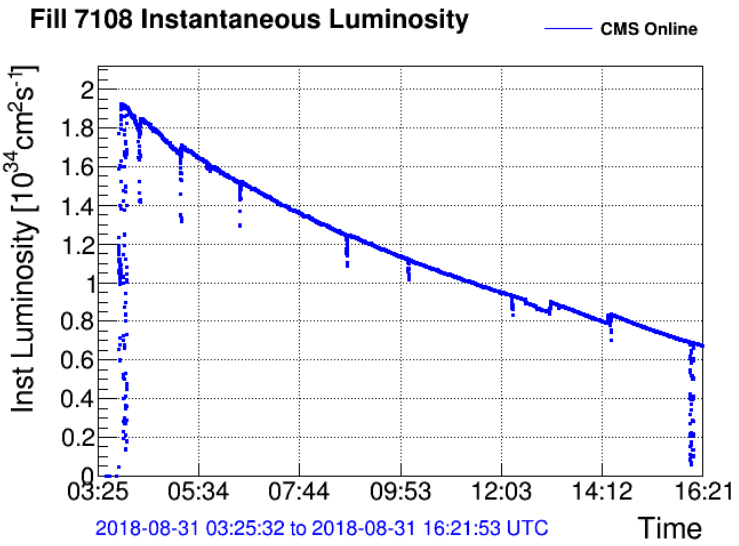


Figure 2.4: Instantaneous luminosity profile for a typical Run-2/2018 fill with the characteristic shape of a falling exponential. [48]

angle to avoid long-distance (typically electromagnetic) interactions between the beams before and after the interaction point.

Each injection of the pp is called *fill* in jargon. In a typical Run-1 or Run-2 fill, the instantaneous luminosity decreases exponentially from the initial peak luminosity along a beam fill. After some hours of data taking for the same fill (about 10 hours in the last year of Run-2), the instantaneous luminosity has decreased so much that it is more efficient to abort the fill and refill the machine with new beams. The luminosity profile in Run-3 is different than in Run-2: in fact it includes the novelty of the *lumi-levelling*, which consists in maintaining the luminosity at its peak value at the beginning of the fill for some hours (and till about half of the fill itself) by continuously adjusting the beam focusing. Examples of the instantaneous luminosity profile in 2018 and 2022 are shown in Figures 2.4 and 2.5 respectively.

The luminosity integrated by CMS since the start of LHC is shown in Figure 2.6. The total integrated luminosity recorded amounts to 38.0 fb^{-1} in 2022 and 29.4 fb^{-1} in 2023, which sum up to luminosity integrated in Run-1 and Run-2 respectively of 29.4 fb^{-1} and 163.3 fb^{-1}. A central system certifies the collected data, ensuring that all the sub-system are properly working during the data-taking: the amount of certified data is almost close to the total.

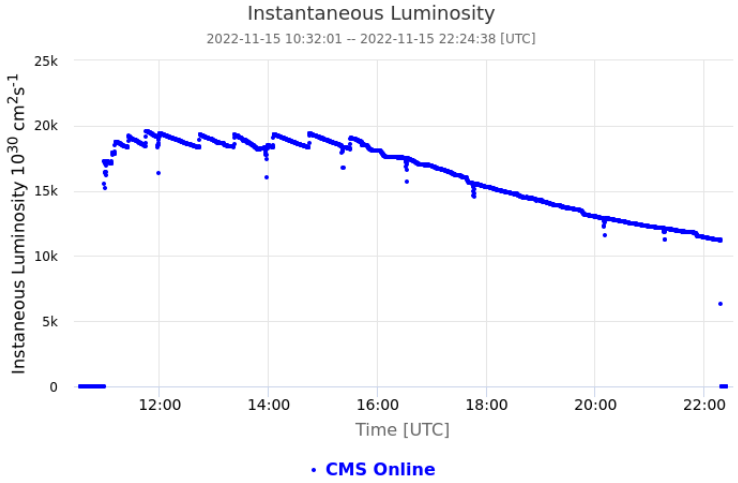


Figure 2.5: Instantaneous luminosity profile for a typical Run-3/2022 fill. The luminosity is levelled for 5 hours at a value of approximately $1.9 \cdot 10^{34}$ cm^{2}s^{-1}, corresponding to almost half of the fill duration time. [48]

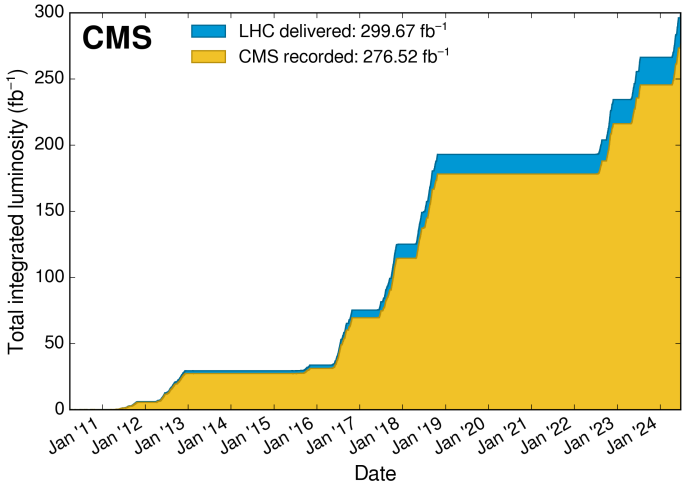


Figure 2.6: Luminosity integrated by CMS since the start of LHC. [47]

2.2 The CMS detector

The CMS design [42] was driven by the goals of the LHC physics program. The detector (shown in Figure 2.7) has a cylindrical shape with an overall length of 28.7 m, of which 21.6 m make the main cylinder with a diameter of 15 m, and the rest of the length comes from the forward calorimeter. The total mass is 14000 t. The main detector is made of a central barrel section that is closed with an endcap section on both ends to cover most of the 4π solid angle.

Peculiar feature of the CMS detector is the solenoidal high magnetic field ($\simeq$ 3.8 T) in the barrel part to bend the trajectories of charged particles. The silicon tracking system, the Electromagnetic CALorimeter (ECAL), and the Hadronic CALorimeter (HCAL) are located inside the 6 m diameter bore of the magnet. Outside of the solenoid, the muon tracking system is sandwiched in between the layers of the steel return yoke for the magnetic field. The high magnetic field not only provides a large bending power within a compact spectrometer, but also avoids stringent demands on muon-chambers resolution and alignment. The return field saturates 1.5 m of iron, allowing 4 muon stations to be integrated to ensure robustness and full geometric coverage. The muon spectrometer is composed by four stations of Drift Tube (DT) detectors in the barrel region (MB) and four stations of Cathode Strip Chambers (CSCs) in the endcaps (ME). Both the barrel and the endcaps muon chambers are coupled to Resistive Plate Chambers (RPCs) to ensure redundancy and robustness to the muon trigger and reconstruction. During LS2 each endcap has been upgraded with the addition of one station of Gaseous Electron Multiplier (GEM) detectors to provide redundancy at large pseudorapidity.

In this analysis, only tracks and muons reconstructed candidates are used. Therefore, in this thesis, the description of the CMS detector is focused only on tracker and muons systems.

2.2.1 Geometry and reference system

CMS has a standardized reference system, sketched in Figure 2.8. The origin of the cartesian coordinate system of the detector is centered at the nominal collision point. The x-axis points horizontally and radially inward to the center of the LHC ring and the y-axis points vertically upward. The coordinate system is conventionally right-handed and thus the z-axis points horizontally along the counter clockwise beam direction. The radius is defined in the xy (transverse) plane as $r = \sqrt{x^2 + y^2}$.

Since the products of the collisions will fly outward from the collision point, it is reasonable to use cylindrical coordinates for the description (used by reconstruction algorithms) based on the azimuthal angle ϕ, defined as the angle

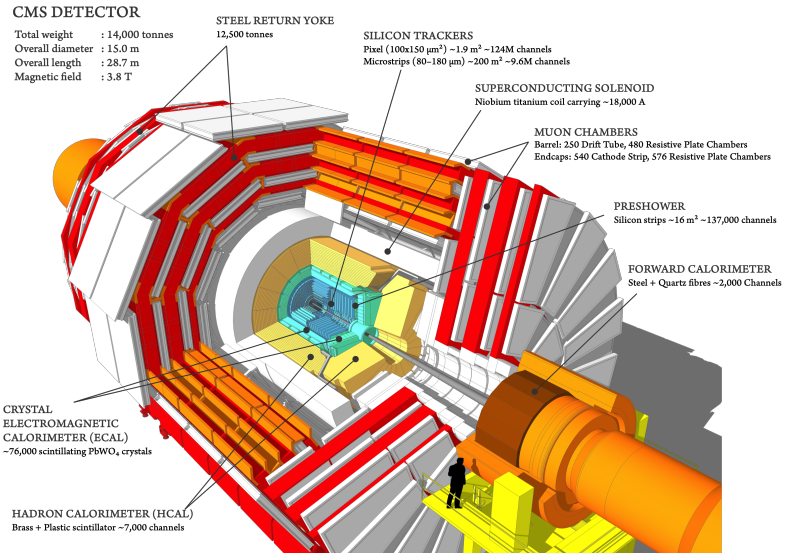


Figure 2.7: Sectional view of the CMS detector. The LHC beams travel in opposite directions along the central axis of the CMS cylinder colliding in the middle of the CMS detector [49, 50].

measured from the x-axis in the xy plane, the radial coordinate r also measured in the xy plane, and the polar angle θ, measured from the z-axis. Instead of the polar angle the pseudorapidity η is used, which is zero in the xy plane and goes to positive (negative) infinity, respectively, towards the positive (negative) z-axis; the forward regions of the detector mean are therefore characterized by $|\eta| > 3$. The η is built from the 3-dimensional angle θ computed from the z-axis:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.3}$$

Another variable commonly used is the ΔR variable, that is defined as $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. It is used to evaluate how spatially close two objects (defined by their 4-vectors) are to each other.

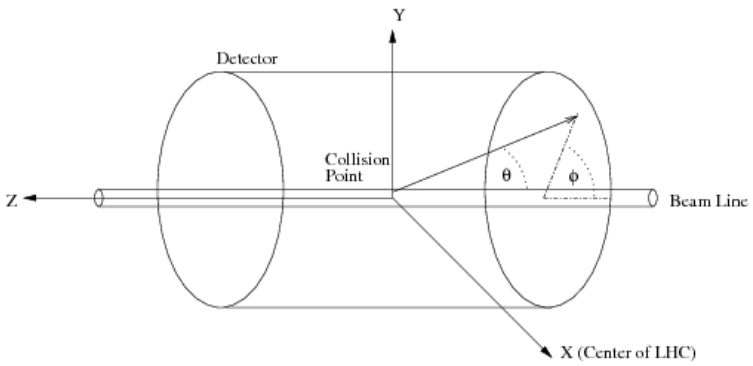


Figure 2.8: CMS coordinate system [51].

Definition of kinematic variables

The composite structure of a proton causes the two partons involved in the collision to interact without fixed energies. In an hadron collider the total energy of an event is unknown. For this reason it is not possible to define the total and missing energy of the event, but only the total and missing transverse energies, namely those in the xy plane transverse to the longitudinal directions of the beams (z-axis).

Moreover, the centre of mass may be boosted along the beam direction therefore is useful to use experimental quantities that are invariant under such boosts. According to these definitions, the momentum of a particle can be divided in two components: the longitudinal momentum p_z and the transverse momentum p_T, defined as

$$p_T = \sqrt{p_x^2 + p_y^2} \, . \tag{2.4}$$

The rapidity is defined as

$$y = \frac{1}{2} \cdot \ln \left(\frac{E + p_z}{E - p_z} \right) \, , \tag{2.5}$$

and has the property of being additive under Lorentz boosts along the z direction, i.e. it is simply shifted by a constant when subjected to such transformations. For ultra-relativistic particles ($p \gg m$) the rapidity can be approximated by the pseudorapidity. As already pointed out the latter is reconstructed from the measurement of the θ angle by means of Eq. 2.3, and thus it can be calculated also for particles whose mass and momentum are not measured.

2.2.2 Magnet

The solenoid of the CMS detector produces uniform magnetic field in the axial direction, while the flux return is assured by an external iron yoke organized in 3 layers and equipped with the muon system chambers installed in between. The momentum of the charged particles is estimated by measuring the particles trajectories inside the solenoid and the transverse momentum resolution is given by:

$$\frac{\Delta p_T}{p_T} = \Delta s \cdot \frac{8 \, p_T}{0.3 \, B \, R^2} \, , \tag{2.6}$$

where $p = m\beta\gamma$ is the particle relativistic momentum, B is the magnetic field, s is the sagitta and R is the solenoid radius. Therefore strong field and large radius are the ingredients to reach an optimal momentum resolution; the CMS design preferred a high field within a compact space.

The superconducting magnet of the CMS detector (its position can be appreciated later in Figure 2.12) has a length of 12.5 m and a diameter of the

cold bore of 6.3 m. It is made from a 4-layer winding of Nb-Ti cable reinforced with aluminium, weighting a total of 220 t, and kept at a temperature of 4.5 K with liquid helium. It was designed to produce a field of 4 T but it operates at a lower field of 3.8 T. The magnetic field is generated by a 18 kA current circulation in the cables. The magnet system stores an energy of 2.5 GJ.

2.2.3 The tracking system

The CMS tracking system is a sub-detector that carries out a series of non-destructive measurements, called *tracker hits*, along the path of a crossing ionizing particle. Those measurements are done through the ionization of a medium, solid or gas, by the charged particle. The CMS tracking system, or simply *silicon tracker*, is responsible for the tracking of the charged particles' path in the innermost part of the CMS detector. The CMS Tracker covers an η range of $[-2.5, 2.5]$ and it is located at the very center of the CMS detector in order to determine precisely the position of the primary and secondary vertices, i.e. where the particles emerge from the collision region. Being embedded in a cylindrical solenoid, with its magnetic field curving the particles' trajectory, it provides information on the transverse momentum and the charge of the particles. It also measures the stopping power dE/dx for each hit, that is the energy loss per unit length. Unfortunately this cannot be exploited in an hadron identification at the needed momentum ranges. Together with the calorimeters, the Tracker participates to the reconstruction of electrons and charged hadrons. Combined with the signals collected in the muon system, it is used to reconstruct the muons.

The crucial feature of the Tracker is that it is the closest sub-detector to the collision point. As a consequence, it must meet the following requirements:

- high granularity ensuring high spatial resolution;
- fast response, i.e. relatively small readout time;
- the ability to handle high radiation doses ();
- the ability to handle high track density, i.e. being characterized by a low track occupancy (i.e. the number of tracks per number of pixels/strips).

These requirements led to the need of two dedicated technologies to handle the high particle flux from the pp collisions: the *pixel* detector, closer to the interaction point, and the *strip* detector, with the latter surrounding the former. The pixel detector is made of relatively smaller sensors that guarantee the needed high granularity in the innermost part of the tracking system. To meet such requirements, the detector electronics needs high powers, hence a cooling

system must be put in place. Cooling also reduces leakage current in the silicon sensors. The full tracker system is cooled at $-20\,^oC$ by means of liquid C_6F_{14}. We will refer to *pixel tracker* and *strip tracker* for the innermost and outermost tracker sub-systems, respectively. In Figure 2.9 the arrangement of the sub-parts of the tracker (as in 2016) and the corresponding η ranges are sketched. The pixel detector, initially made of three layers, was upgraded in 2017 with the addition of a fourth one. Figure 2.10 provides a way to illustrate the impact of the addition of the fourth pixel layer on a physical observable, namely the proper decay length uncertainty for the $B_s^0 \to J/\psi\phi$ which is a crucial ingredient in the procedure of the extraction of the CP-violating phase ϕ_s: the mode of 2018 distribution is about $5\mu m$ lower than the 2012 one.

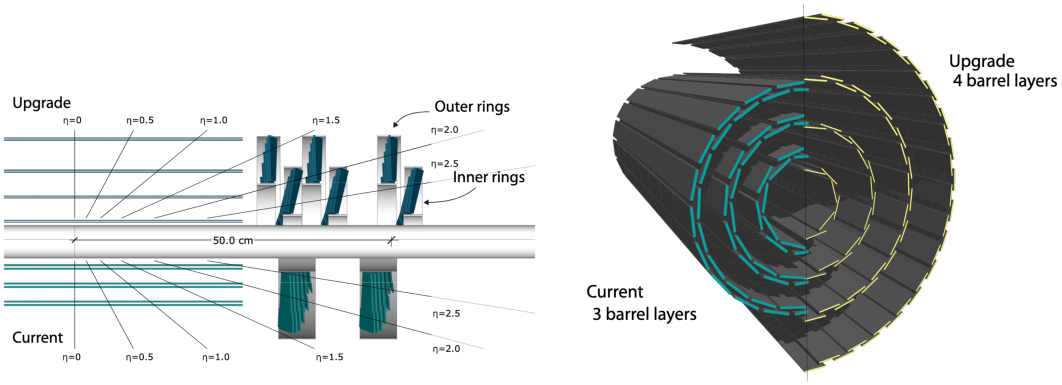


Figure 2.9: Left: Conceptual layout comparing the different layers and disks in the current and upgrade pixel detectors. Right: Transverse-oblique view comparing the pixel barrel layers in the current and upgrade detectors [52].

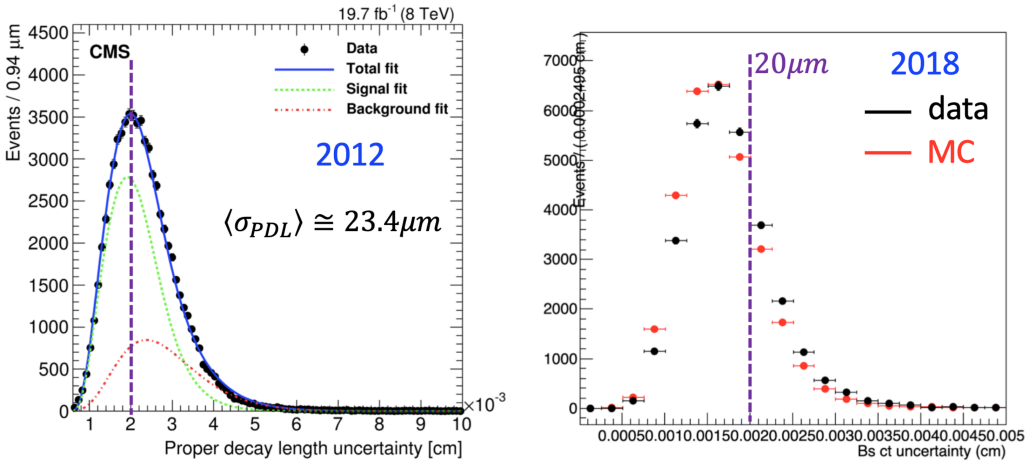


Figure 2.10: Left: proper decay length uncertainty obtained for $B_s^0 \to J/\psi\phi$ candidates from 2012 $\sqrt{s} = 8$ TeV data [53]. Right: same observable for 2018 $\sqrt{s} = 13$ TeV data [54].

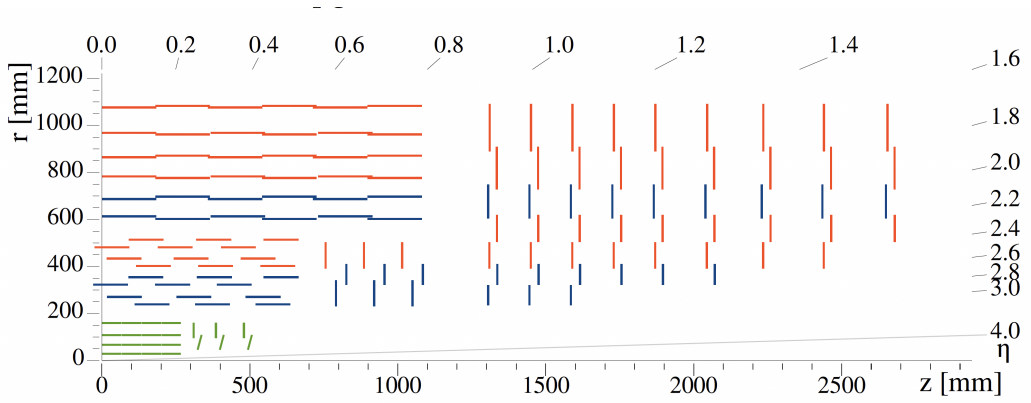


Figure 2.11: Sketch of one quarter of the Phase-1 CMS tracking systemin $r - z$ view. The pixel detector is shown in green, while single-sided and double-sided strip modules are depicted as red and blue segments, respectively. [55].

The basic idea of a sensor in a tracking system is to have an active volume immersed in an electric field. When a charged particle traverses the medium, it ionizes the atoms of the medium thus creating free charges that drift towards electrodes to be collected and an induced signal is read out. In silicon, however, the number of free carriers created by thermal excitations at room temperature is much higher (typically 4 orders of magnitude) than the number of free carriers created by the passage of a minimum ionizing particle. For a given material, a minimum ionizing particle (MIP) is a particle that has a momentum corresponding to the minimum plateau of the Bethe-Bloch distribution: it will deposit a minimal amount of energy by ionization. The solution to this apparent issue is to dope the material with donors (elements with an extra electron with respect to silicon) or acceptors (elements with one electron less, or an extra hole with respect to silicon). Typical silicon sensors are made of p-n junctions, combining a p-doped silicon bulk (with extra acceptors) with a n-doped silicon bulk (with extra donors). In such a device, electrons migrate to the p-doped region and holes move to the n-doped region, resulting in a depleted region, in the middle, characterized by the absence of free carriers. This depleted region represents the active detection volume of the silicon device.

2.2.4 Muon chambers system

The muon spectrometer is placed outside the magnet and can provide an independent muon identification, robust trigger, and enough accurate momentum measurement and charge for muons with momenta from a few GeV to a few TeV. It consists of four stations of detectors integrated into the iron flux return yoke so that both the 3.8 T magnetic field inside the solenoid and the 1.8 T average return field can be used as bending field.

The muon spectrometer is composed of four independent sub-detectors. In the barrel ($|\eta| < 1.2$), where the particle rate is lower and the residual magnetic field are low, drift tubes (DTs) are installed. In the endcaps, where the particle rate is higher and a large residual magnetic field is present, CSCs are used. The coverage of the DT and the CSC systems goes up to $|\eta| < 2.4$. In the region $|\eta| < 1.9$ RPCs are present as well. GEMs are installed in front of the first station of CSCs. All the sub-detectors are used for tracking the muons and for trigger on muon candidates in order to guarantee robustness and redundancy. The layout of the system is presented in Figure 2.12.

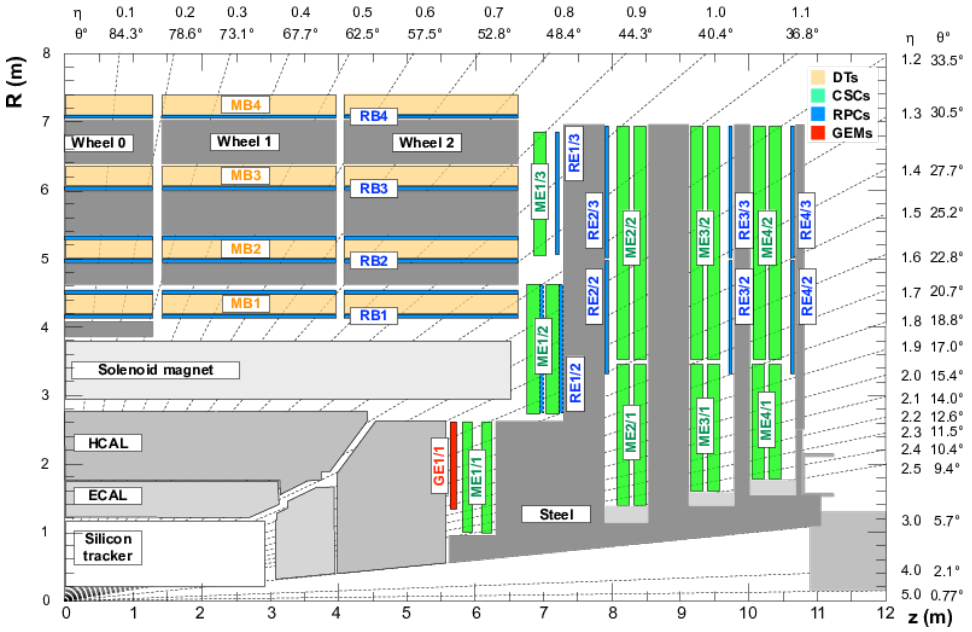


Figure 2.12: A quadrant of the CMS detector with the different muon sub-detectors highlighted [56].

2.2.5 The CMS trigger system

At the LHC protons collide every 25 ns and a trigger system to select events of interest is needed, as otherwise saving any event would lead to an unaffordable average throughtput of approximately 40 TB/s. CMS uses a tiered two-level trigger system to select physics events and reduce the data volume: the first level is the Level-1 Trigger (L1T), while the second is the High Level Trigger (HLT) (Figure 2.13).

The L1 Trigger is an hardware based trigger that receives data from some of the CMS sub-detectors, namely the Calorimeters and the Muon system. Some physics objects are already available at the L1 stage with coarse granularity thanks to a fast reconstruction; no information from the Tracker is included.

Relatively simple variables and algorithms are implemented in the L1 trigger firmware (e.g. muon p_T rough estimate, opposite sign requirement on two muons, calculation of their invariant mass or their angular separation). The L1 Trigger is capable of accepting or rejecting an event within a fixed time window of approximately 4 μs and reduces the input rate of 40 MHz to an output rate of approximately 100 kHz, limited by the hardware capabilities. A single L1 trigger is usually called in jargon *L1 seed*.

The second tier of the trigger system, the HLT, is instead a software-based trigger that can rely on event information from all the sub-detectors. A HLT trigger algorithm, usually known as *HLT path*, includes the reconstruction of physics objects with a relatively good precision, rather similar to the precision provided by the offline reconstruction.

The HLT menu at Run-2 was run on CPUs and on average, during 2018, it took around 350 ms to elaborate a decision. The HLT menu in 2018 was tuned so that the average HLT trigger rate was about 1.3 kHz, limited by the passing bandwidth for the storage of reconstructed events. Since 2018 the standard trigger strategy is complemented with a so-called *parking* trigger strategy: the HLT bandwidth limit is by-passed by storing the triggered data without running any further step, and instead running the reconstruction workflow only when the computing resources available. This approach - pursued by CMS in different manners across the years - has granted since 2018 few extra kHz of HLT rate.

Moreover in Run-3 part of the reconstruction (mainly tracking and jets in hadronic calorimeter) has been rewritten to run on heterogeneous (CPU+GPU) system: this is a crucial upgrade towards the LHC Phase-2. More complex and scalable algorithms have been implemented, thus improving the CMS reconstruction performance while fitting in the same resource budget.

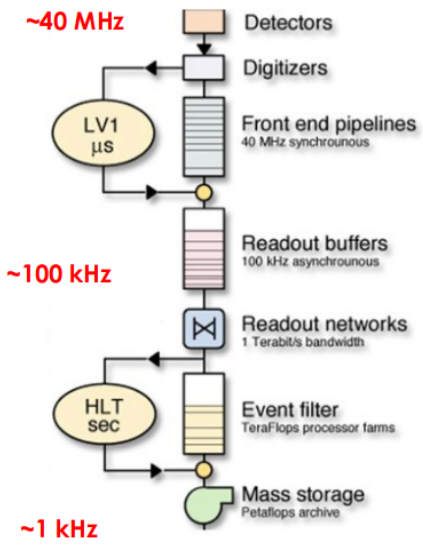


Figure 2.13: Scheme for the CMS Trigger system [57]. Protons collide every 25 ns at the LHC and the particles produced in the collision interact with the detectors. The detectors produce an electric signal which is quickly digitized and passed to the L1 system. A response whether the event must be saved or discarded is produced within a fixed latency of 3.8 μs. The L1 system is tuned so that its expected output event rate is approximately 100 kHz. The events are then further processed and reconstructed by the HLT farm in hundreds of milliseconds. The HLT system applies selection and brings down to few kHz the event rate to be stored.

2.3 Track reconstruction in CMS

The trajectory of a charged particle in the rather uniform magnetic field present in the tracker volume is an helix and thus is described by 5 parameters. Track reconstruction algorithms rely on a good estimate of the pp interaction region called *beamspot*; the exact definition and features of the beamspot will be given later, in subsection 2.3.1. After the beamspot reconstruction, track reconstruction starts using the pixel hits. The CMS standard track reconstruction uses an adaptation of the combinatorial Kalman filter named Combinatorial Track Finder (CTF) [58]. Multiple iterations of the CTF track reconstruction are sequentially performed (iterative tracking) starting firstly to search for those tracks that are easiest to find (large p_T, produced close to or within the beamspot); once the hits associated with these tracks are removed the subsequent iterations are performed in a cleaner environment, in which more difficult classes of tracks can be reconstructed. The final set of tracks is obtained by merging the previous classes in order to form the collection of reconstructed tracks.

On the basis of the whole set of hits reconstructed in the tracker within an

event, each CTF iteration typically proceeds in the following four steps: (a) seed generation, (b) track finding, (c) track fitting, (d) track selection.

The seeds define the initial estimate of the trajectory parameters and their corresponding uncertainties. Seeds are constructed in the inner part of the tracker on the basis of few inner hits. In Run-3 track seeds are built from a triplet or quadruplet (with different quality criteria) of pixel hits. During the 2023 data taking an additional track seeds collection based on pixel doublets has been included in order to mitigate the tracking inefficiency due to an issue in some pixel modules of the third and fourth barrel layer in a region around $\phi = -1$.

Track candidates are built on the basis of the seeds and the trajectory is initially defined from them. An iterative algorithm based on the Kalman filter method starts from a coarse estimate of the track parameters provided by the trajectory seed and builds track candidates by adding hits from successive detector layers and updating the track parameters at each step iteratively.

The trajectory is refitted using a Kalman filter operator starting from the innermost hit and proceeding outwards. After a smoothing stage, a second filter, initialized with the results of the first one, is run backwards towards the beam line. The optimal track parameters are then obtained from the weighted average of the output of these two filters. In Run 2, the CMS track reconstruction algorithm used an iterative approach based on combinatorial Kalman Filter (CKF), consisting of twelve main iterations targeting different track topologies and seeded with different seed tracks. The name assigned to each iteration stems either from the order of execution, or from the seeding topology. For Run 3, a new algorithm has been developed for track pattern recognition (or track building), named mkFit [59], that maximally exploits parallelization and vectorization in multi-core CPU architectures. This algorithm has been deployed in the CMS software for a subset of six tracking iterations [60]:

- *InitialPreSplitting*, initial iteration before splitting merged pixel clusters in dense jet environments;
- *Initial*, initial iteration;
- *HighPtTriplet*, high-p_{T} triplet iteration;
- *DetachedQuad*, detached quadruplet iteration;
- *DetachedTriplet*, detached triplet iteration;
- *PixelLess*, pixel-less iteration.

The mkFit algorithm allows to retain a similar physics performance with respect to the traditional CKF-based pattern recognition, while improving the computational performance of the CMS track reconstruction.

Quality requirements are applied to the reconstructed track candidates in order to discard fake tracks (tracks not associated to any real charged particle). Track selection criteria have different threshold values for each iteration and are defined as *loose*, *tight* and *high purity*, thus providing progressively more stringent requirements that reduce the fake rate but also the track reconstruction efficiency.

Results for tracking efficiency are shown in Fig. 2.14. The efficiency is measured by using simulated particles generated within $|\eta| < 2.5$, with a distance between the extrapolated production point on the beam line and the center of the beamspot < 3 cm and < 30 cm for r and $|z|$ respectively.

2.3.1 Primary Vertex and beamspot reconstruction

The goal of the Primary Vertex (PV) reconstruction is to measure the position and the associated uncertainty of all pp interaction vertices in each event (this includes also those coming from pileup collisions). This is performed in three steps: track selection, track clustering and vertex fitting.

The first step consists in choosing tracks consistent with being produced in the primary interaction region. This is accomplished by imposing requirements on the transverse IP significance, number of hits in the pixel and strip detectors and quality of the track fit. The selected tracks are then grouped into clusters on the basis of their z coordinates at their point of closest approach to the centre of the beamspot. The clustering is performed using a deterministic annealing (DA) algorithm, described in Ref. [62]. The candidate vertices identified via the DA clustering are finally fitted to find the best estimation of the vertex parameters. The fit is performed using an adaptive vertex fitter [63], which assigns to each track a weight based on the spatial compatibility of the track with the vertex.

The efficiency of reconstructing a PV and the resolution of the reconstructed PV position depends strongly on the number of tracks used to fit the vertex and their p_T, as shown in Figure 2.16. The resolution in x and z for primary vertices reconstructed using at least 50 tracks are less than 20 μm and 25 μm, respectively, for *minimum bias* events. It significantly improves (up to 10 μm and 12 μm) for jet-enriched events, where the average track p_T is higher.

The beamspot is defined as a 3D profile of the luminous region where the LHC beams collide in the central region of the CMS detector. As distinct from the PV reconstruction that is performed for each collision, the beamspot parameters are measured from an average over many events (precisely, it is

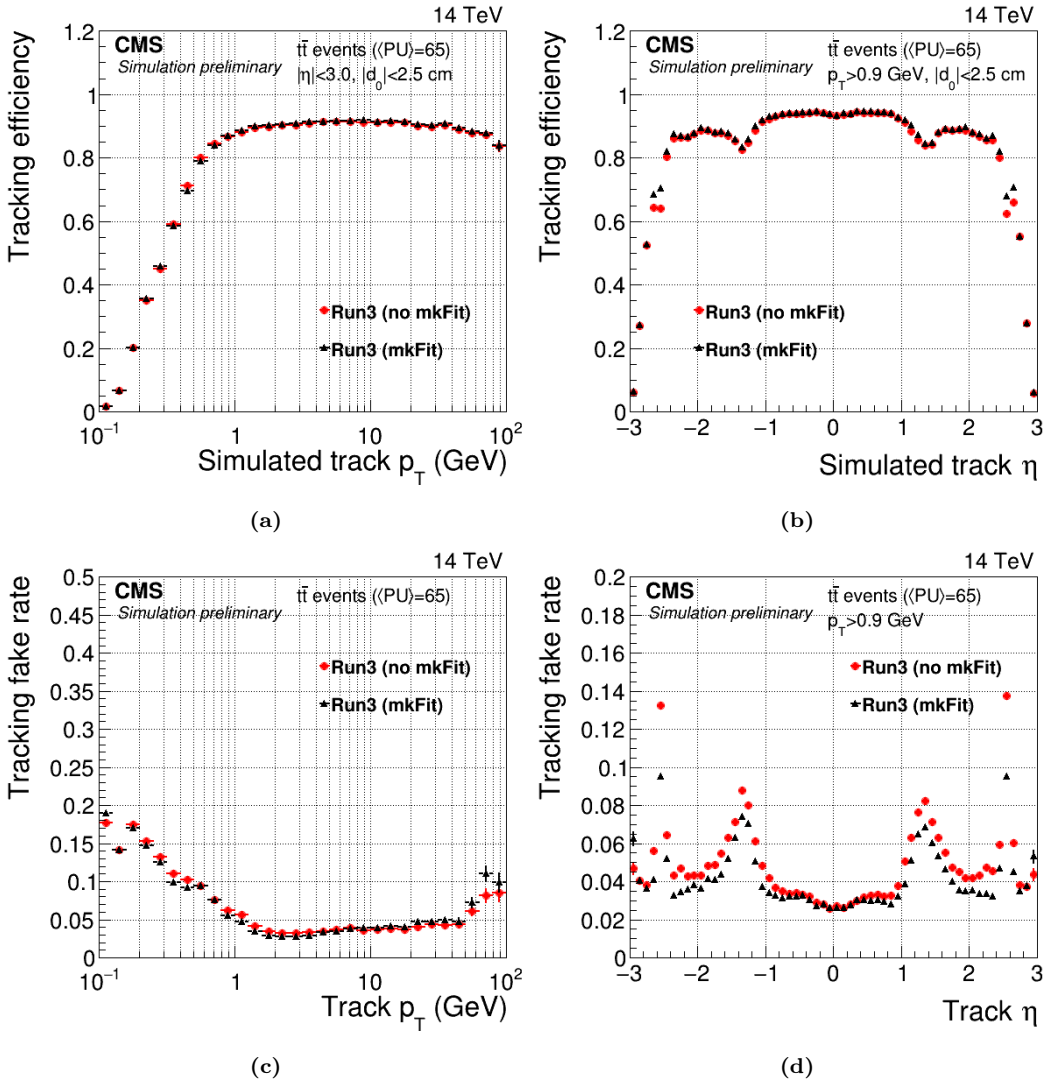


Figure 2.14: Top row: Tracking efficiency as a function of the simulated track p_{T} (left) and pseudorapidity η (right) for CKF tracking (red) and Run 3 tracking using mkFit (black). The efficiency with mkFit is consistent with traditional CKF tracking for both variables. Bottom row: Tracking fake rate as a function of reconstructed track p_{T} (left) and pseudorapidity η (right) for CKF tracking (red) and Run 3 tracking using mkFit (black). The fake rate with mkFit is generally lower than that with CKF tracking. All plots correspond to simulated $\mathrm{t\bar{t}}$ events at $\sqrt{s} = 14$ TeV with an average pileup of 65. [60]

determined for every *luminosity section*, corresponding to a period of ≈ 23 s of data taking). The position of the centre of the beamspot is determined through the combination of two methods. The first, used to measure the z coordinate and the RMS widths of the luminous region, consists in a fit to the 3D distribution of primary vertex positions. The second, used to determine the transverse coordinates, only needs track information and exploits the correlation between d_0 (minimum transverse distance from the beam axis) and ϕ

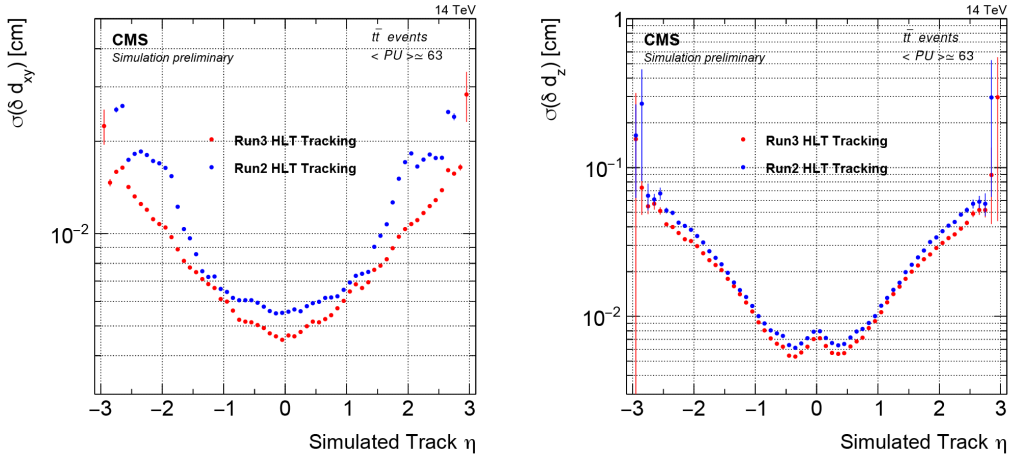


Figure 2.15: Comparison of tracking resolution as a function of pseudorapidity η between Run-2 and Run-3 HLT tracking. Left: The d_{xy} resolution for Run-2 HLT tracking (blue) and Run-3 HLT single-iteration tracking (red). Right: The d_z resolution for Run-2 HLT tracking (blue) and Run-3 HLT single-iteration tracking (red). Both plots are based on $t\bar{t}$ events with an average pileup of 63 at 14 TeV. [61]

track parameters that appears when the beamspot is displaced with respect to its expected position. The beamspot position is used to estimate the position of the interaction point prior to the reconstruction of the PV and to provide an additional constraint in the reconstruction of all the PVs in an event.

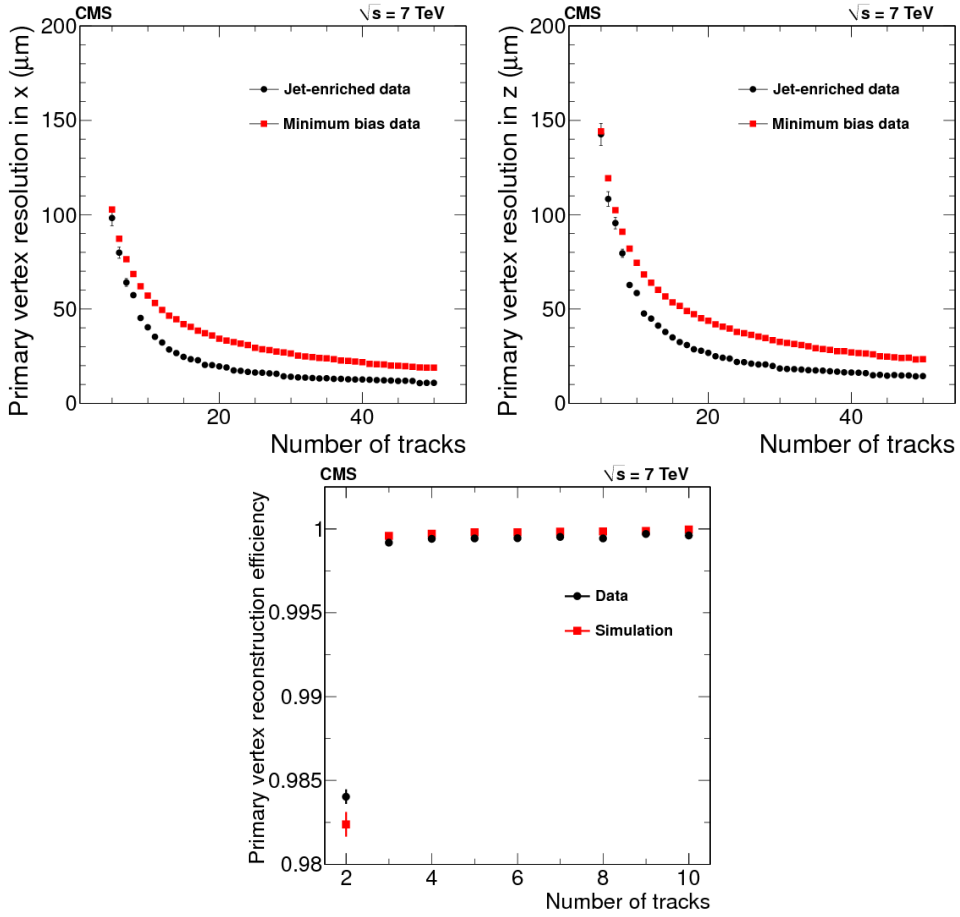


Figure 2.16: Upper plots show the x and z resolution as a function of the number of tracks at the fitted vertex, for two kinds of events with different average track p_T values [64]. The tracks in jet-enriched events have significantly higher mean p_T than *minimum bias* data, resulting in a better vertex resolution. Bottom plot shows the primary vertex reconstruction efficiency as a function of the number of tracks in a cluster, measured in *minimum bias* data and in Monte Carlo simulation. All plots are based on events at $\sqrt{s} = 7$ TeV during the 2010–2011 data-taking period with appropriate pileup conditions for jet-enriched and minimum bias events. [64]

2.4 Muon reconstruction at CMS

The CMS ability to trigger on, and accurately reconstruct, muons with high efficiency relies on muon reconstruction algorithms that have been designed to achieve these goals. The muon reconstruction is performed with the information registered by the Muon System and the Tracker. [65]

Several muon reconstruction strategies are available in CMS, in order to fulfill the specific needs of different analyses. In general, the reconstruction of muons consists of three main stages:

- Local reconstruction. In each muon chamber, the raw data from the detector read-out are reconstructed as reconstructed hits (*recHits*); in CSC and DT chambers, such recHits are then fitted to track stubs (*segments*).
- StandAlone reconstruction. RecHits and segments in the muon spectrometer are collected, using infromation from all the muon subdetectors, then they are fitted to form the so called *StandAlone* muon tracks.
- Global reconstruction. StandAlone tracks are matched to compatible tracks in the inner tracker and a global fit is performed using the whole set of available measurements (recHits or segments): the resulting tracks are called *Global* muon tracks, as sketched in Figure 2.17.

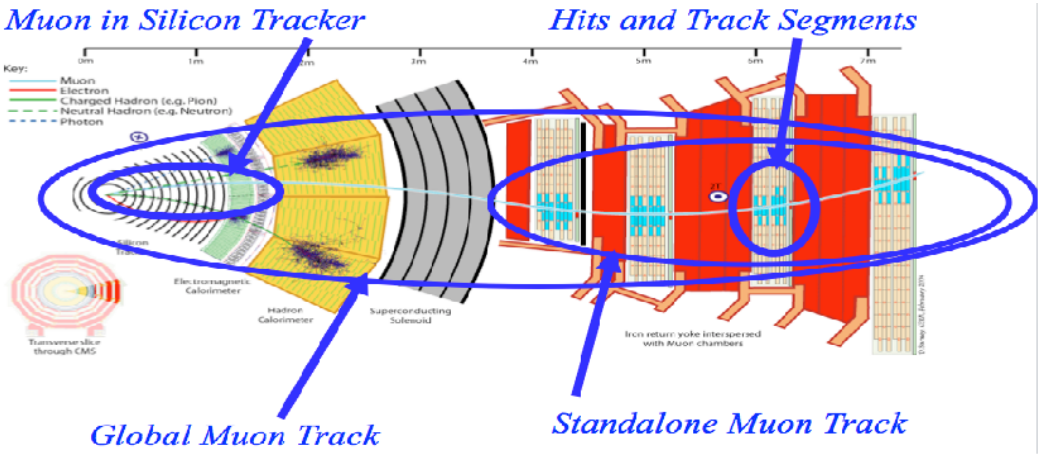


Figure 2.17: *StandAlone* muon tracks, once merged with matching tracks in the tracker, allow to create the *Global* muon tracks, as sketched in the transverse section of CMS.

There is an additional complementary approach with respect to global reconstruction: the reconstruction which starts from the inner tracker tracks and searches for matching segments in the muon spectrometer. If at least one muon segment matches the extrapolated track, the corresponding tracker track qualifies as a *Tracker* muon. The muon candidates produced with this method are named *Tracker* muons.

Indeed, a relevant fraction of muons with transverse momentum below $4 - 5$ GeV does not "release" enough hits in the muon spectrometer to be reconstructed as StandAlone muons (see Figure 2.18) and moreover some muons can be partially undetected by escaping in the gap between the wheels: this complementary approach allows to improve the CMS muon tracking efficiency.

After the completion of both (outside-in and inside-out) algorithms, the reconstructed StandAlone, Global, and Tracker muons are merged into a single software object, with the addition of further information, like isolation and energy loss eventually associated to calorimeter towers. All the available information can be used for further identification algorithms that differ in the balance between efficiency and purity (namely fake rate level) of the muon sample.

The efficiency of the different muons reconstruction algorithms is shown in Figure 2.18 for illustration purposes.

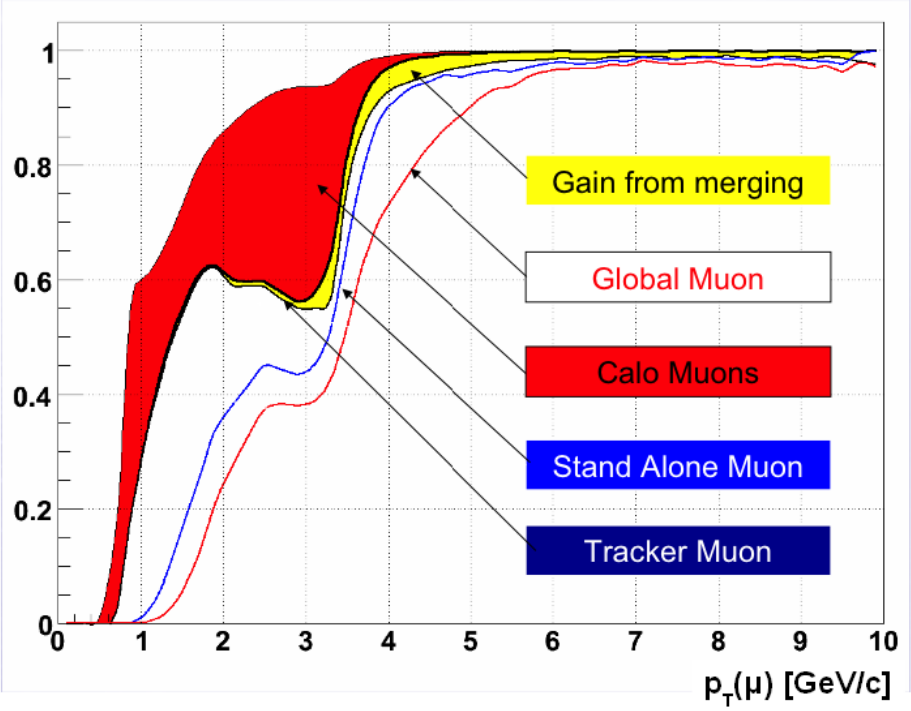


Figure 2.18: Performance of the muon reconstruction efficiency for the different types of muons (*Tracker* muons, *Global* muons and *StandAlone* muons) as a function of the transverse momentum. The efficiency for the muon considered as a complex of the three reconstructions is also provided. For historical reasons, in this relatively old study, the efficiency of the calorimetric muons (*Calo* muons) is also reported but this kind of physics object is not used anymore within CMS. The yellow band represents the difference between the efficiency of *Tracker* muons only and the efficiency of muons from the three collections merged. [Courtesy of D. Kovalskyi]

2.4.1 Muon identification

The Tracker muon reconstruction is more efficient than the Global muon reconstruction at low momenta ($p_T > 0.5$ GeV and $p > 2.5$ GeV) because it requires only a single muon segment in the muon system, whereas Global muon reconstruction is designed to have high efficiency for muons penetrating through more than one muon station and requires segments in at least two muon stations. Owing to the high efficiency of the tracker-track reconstruction and the very high efficiency of reconstructing segments in the muon system, about 99% of muons produced in pp collisions within the geometrical acceptance of the muon system and having sufficiently high momentum are reconstructed either as a Global Muon or a Tracker Muon, and very often as both of them. The majority of the reconstructed muons are either Global or/and Tracker. Candidates found both by the Global Muon and the Tracker Muon approaches that share the same tracker track are merged into a single candidate. Only a very small fraction of times (about 1%) a StandAlone muon cannot be used to build a Global or Tracker muon. Muons reconstructed only as StandAlone muon tracks have worse momentum resolution and higher admixture of cosmic-ray muons than the Global and Tracker muons.

The combination of different algorithms provides robust and efficient muon reconstruction. Physics analyses can set the desired balance between identification efficiency and purity by applying a selection based on various muon identification variables. Nevertheless the muon identification algorithms provide few identified muon lists by applying further criteria to the Global or Tracker muons. The most common lists are the *Soft* and the *Tight* muons. The former (latter) are defined starting from Tracker (Global) muons and adding further requirements. Specifically:

- Tight muons are Global muons with two additional requirements (called *GMPromptTight*):
  - the normalized chi-square of the global fit must satisfy the condition $\tilde{\chi}^2_N < 10$,
  - there must be at least one *valid* hit (hits are declared *invalid* as a penalty for not participating to the seed of an effectively reconstructed StandAlone muon).
- Soft muons [66] are Arbitrated Tracker muons of the type *TMOneStationTight* with additional quality requirements summarized below:
  - `highPurity`: the reconstructed tracks are selected with the "high purity" flag [64];

– `trackerLayersWithMeasurement` > 5: a minimal number of tracker layers involved in the track measurement guarantees a good p_T measurement and helps to suppress decays-in-flight;

– `pixelLayersWithMeasurement` > 0: there should be at least one pixel hit in the tracker part of the muon track. The innermost part of the tracker is an important handle to discard muons from decays-in-flight of light hadrons.

– $|d_{xy}| < 0.3$ cm and $|d_z| < 20$ cm: these are applied to the transverse and longitudinal impact parameters defined - respectively - as the transverse and longitudinal distance of closest approach of the extrapolated muon track with respect to the beamspot. They are effective in the suppression of cosmic muons, muons from decays-in-flight and tracks from pileup.

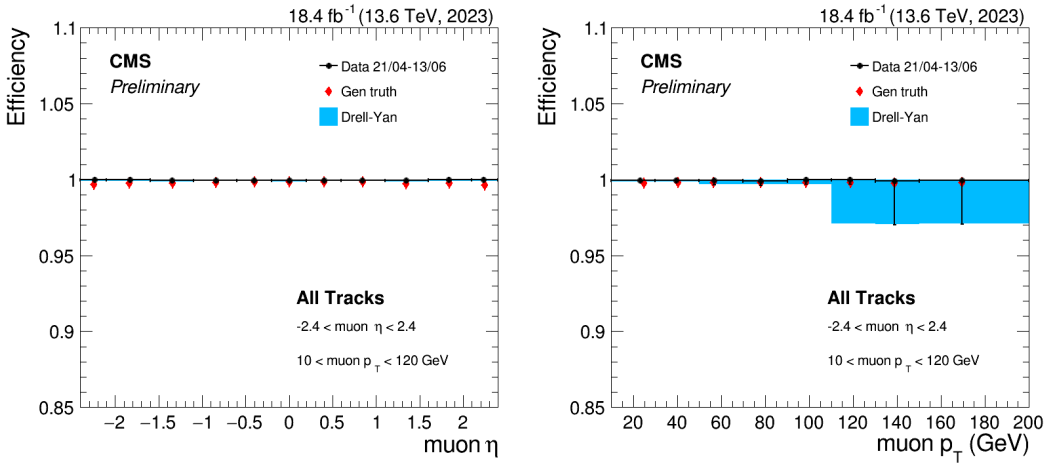


Figure 2.19: Tracking efficiency in the All Tracks collection for 2023 data (black dots), Gen truth study (red diamonds), and Madgraph DY (light blue rectangles) as a function of η (left) and p_{T} (right) of probe standalone muons. The uncertainties include statistical and matching probability contributions. [67]

The recommendation for CMS data analyses is to use the standard lists of identified muons. For Heavy Flavour Physics analyses, typically, the *Soft* muons are used to exploit the better efficiency of Tracker muons at low p_T. The *Tight* muons selection, instead, is used in many physics analyses in CMS, and among them for instance in the Standard Model and Higgs physics. Soft muons are used in this analysis.

2.5 A new dedicated trigger for the B_c meson in Run-3

As mentioned in Section 1.7.1, CMS published results on the B_c meson spectrum [28, 68] obtained using the Run-2 data. More than 7000 B_c candidates in the final state of two opposite sign muons and a charged pion were selected using a trigger requiring two muons coming from a J/ψ whose production vertex is displaced from the PV, and an additional track, also significantly detached from the PV, compatible with same di-muon vertex (`displaced JpsiTrk` trigger). This trigger is generally serving many analyses related to heavy flavour spectroscopy.

A new trigger strategy dedicated to the B_c physics program has been deployed for Run-3: it consists of removing the displacement requirement and to compensate the consequent trigger rate increase by selecting only J/ψ-plus-track candidates with an invariant mass compatible with the B_c meson, therefore designing a dedicated trigger path. For this reason this HLT path is denoted as `inclusive Bc` trigger.

The `inclusive Bc` trigger is based on two L1 seeds commonly used to trigger on soft J/ψ mesons in logical OR combination. The first targets high-quality opposite-sign muons with a pseudorapidity restriction ($|\eta(\mu^{\pm})| < 1.5$) and limited angular separation between muons ($\Delta R < 1.4$); the second, instead, targets high-quality opposite-sign muons in the full pseudorapidity acceptance, minimum transverse momentum $p_T(\mu^{\pm}) > 4$ GeV, and again a limited angular separation between muons ($\Delta R < 1.2$).

At HLT level the standard muon reconstruction is carried out and then a set of selection criteria is first applied on the dimuon candidate: muon transverse momentum $p_T(\mu^{\pm}) > 4$ GeV, two oppositely charged muons, dimuon transverse momentum $p_T(\mu^+\mu^-) > 6.8$ GeV, dimuon invariant mass $2.9 < m(\mu^+\mu^-) < 3.3$ GeV, dimuon vertex probability $P_{vtx}(\mu^+\mu^-) > 0.1$, and cosine of dimuon pointing angle[1] $\cos(\alpha_{\mu\mu}) > 0.9$.

If a dimuon candidate passes the selection, the tracking sequence is run locally in a cone-like region around the dimuon candidate. Afterwards the following selection is applied on the dimuon-plus-track candidate: track transverse momentum $p_T(trk) > 1.2$ GeV, dimuon-plus-track vertex probability $\chi^2_{vtx}(\mu\mu trk) < 10$, cosine of dimuon-plus-track pointing angle $\cos(\alpha_{\mu\mu trk}) > 0.9$, pion mass from PDG assigned to the track [1], and invariant mass of the dimuon-plus-track candidate $5.95 < m(\mu^+\mu^-\pi^{\pm}) < 6.55$ GeV (the track is as-

[1] Angle between the particle momentum at the decay vertex and the particle flight direction determined by the production vertex and the decay vertex, projected on the transverse plane.

signed the mass of the pion). Most of the HLT selection, summarized in Table 2.1, is inherited from the `displaced JpsiTrk` trigger (already active in Run-2 and kept in Run-3), but modified in order to remove the selection related to displacement requirements.

Variables	**Units**	`inclusive` B_c	`displaced JpsiTrk`
$p_T(\mu^{\pm})$	GeV	> 4	> 4
$\lvert\eta(\mu^{\pm})\rvert$	-	< 2.5	< 2.5
OS muons	-	✓	✓
$p_T(\mu\mu)$	GeV	> 6.8	> 6.8
$m(\mu\mu)$	GeV	[2.9, 3.3]	[2.9, 3.3]
$P_{vtx}(\mu\mu)$	-	> 0.1	> 0.1
$\cos(\alpha_{\mu\mu})$	-	> 0.9	> 0.9
$L_{xy}/\sigma_{L_{xy}}$	-	-	> 3
$p_T(trk)$	GeV	> 1.2	1.2
$\chi^2_{vtx}(\mu\mu trk)$	-	< 10	< 10
$\cos(\alpha_{\mu\mu trk})$	-	> 0.9	> 0.9
trk-ID	-	π	π
$d_0(trk)/\sigma_{d_0}$	-	-	> 2
$m(\mu\mu trk)$	GeV	[5.95, 6.55]	-

Table 2.1: Comparison between the selections applied in the inclusive HLT path dedicated to the B_c meson and in the `displaced JpsiTrk` trigger.

Chapter 3

ML techniques for HEP data analysis: an XGBoost study

The development of the non-displaced trigger for the B_c meson (`inclusive Bc` discussed in Sec. 2.5) is mainly supported by the need to collect an unbiased proper lifetime distribution B_c candidates. In fact previous CMS lifetime measurements had to deal with data collected by the `displaced JpsiTrk` trigger [69]: efficiency functions to model the bias introduced by such a selection were extracted by MC studies. This correction could bias the contribution coming from doubly beauty hadrons, which are expected to have a signature of displaced B_c (discussed in Chapter 1). On the contrary, thanks to the new trigger, the Run-3 data sample contains an unbiased collection of the B_c mesons: nevertheless this comes with the price of a large contribution from the combinatorial background, thus affecting those B_c mesons which decay rather promptly and are not significantly detached from the PV.

The demanding task of discriminating the signal candidates from an overwhelming background has been explored with different approaches: on one hand a "standard" cut-based selection has been designed, on the other hand novel Machine Learning (ML) techniques have been investigated to perform the selection. In fact, the effort of candidates selection in order to extract a sample of B_c candidates can be straightforwardly interpreted as a binary classification problem, where the B_c candidates (signal) need to be discriminated from other background sources.

In particular for each of the two approaches, two different sets of input variables have been considered: the `inclusive` selections do not consider displacement-related variables as input, in order to keep the advantage coming from a nearly unbiased proper lifetime distribution; on the contrary the complementary approach of `displaced` selections, namely including information related to the candidate's displacement from the primary vertex, is also

pursued.

This chapter will compare the capabilities of a standard cut-based approach with those obtained by applying machine learning techniques. For this purpose, XGBoost (XGB) [70], which is an open-source software based on Boosted Decisional Trees (BDTs), will be used. Thanks to its tree structure, it can effectively learn from large datasets - available thanks to the large statistics collected by CMS - and capture complex relationships. It also offers high performance and flexibility, which improve the accuracy of the analysis. Computations for the machine learning training were performed using a server of the HPC cluster hosted at the ReCaS-Bari Data Center. This server is equipped with 2 AMD EPYC 7H12 64-Core Processor at 2.60 GHz, about 2 TB of RAM memory, and an NVIDIA A100-PCIE-40GB GPU.

3.1 Introduction to XGBoost

Machine Learning is a subset of Artificial Intelligence which focuses on developing algorithms that allow computing tasks to learn from data and make predictions or decisions based on data. One of the powerful techniques within ML is boosting, specifically Gradient Boosting, which is used to create a strong predictive model by combining multiple weak models, usually decision trees.

Gradient Boosting works by iteratively adding decision trees to a model, where each new tree corrects the errors made by the previous ones. This is done by fitting the new tree to the gradient of the loss function, which measures the difference between the predicted and actual values. The decision trees in gradient boosting are composed of nodes (splits) and leaves (predicted outcomes). The maximum depth of the tree, which refers to the maximum number of splits from the root to the furthest leaf, is a crucial hyperparameter influencing the model's complexity and performance.

Extreme Gradient Boosting (XGBoost) [70] is an advanced implementation of the gradient boosting algorithm. It introduces several innovations that enhance performance and computational efficiency. XGBoost uses a more regularized model formalization to control overfitting, and it incorporates an efficient handling of sparse data and a weighted quantile sketch for approximate tree learning. Additionally, XGBoost supports parallel processing in a transparent way, which significantly speeds up the training process. These enhancements make XGBoost particularly effective for large-scale datasets and complex prediction problems.

In ML, two key concepts are the *objective functions* and the *evaluation metrics*. An objective function is used by the learning algorithm to guide the training process by defining what constitutes a good model. In XGBoost, one

commonly used objective function for binary classification is the binary logistic function (`binary:logistic`), which outputs probabilities that an observation belongs to the positive class, ranging from 0 to 1. This is particularly useful in distinguishing signal (nominally classified as 1) from background (nominally classified as 0), which aligns with the goal of enhancing the B_c signal visibility by increasing signal candidate counts and reducing background candidates.

On the other hand, evaluation metrics are used to assess the performance of the trained model. One such metric is the *logarithmic loss* (also known as *cross entropy loss* or *log loss*), which measures how well the predicted probabilities match the true class labels. *Log loss* is particularly useful in binary and multi-class classification problems as it evaluates the accuracy of probability estimates of models. Besides being an objective function, *log loss* serves as a critical performance metric that allows to assess the model's performance in probabilistic terms.

For binary classification, *log loss* is calculated using the following formula:

$$log\ loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \tag{3.1}$$

where N is the number of candidates to be classified, y_i is the true class label of the i-th sample (0 or 1), and $\hat{y}_i$ is the predicted probability that the i-th candidate belongs to class 1. Since the considered problem is a binary classification, it is trivial that $1 - \hat{y}_i$ is instead the probability that the i-th candidate belongs to class 0.

In this study, the `binary:logistic` objective function is chosen to train the model, and *log loss* is used to evaluate its performance. A lower *log loss* value indicates that the model's predicted probabilities are close to the actual class labels, which reflecting better model performance. This approach allows us to fine-tune the model effectively and assess its performance accurately, especially in scenarios involving uncertainties.

Many parameters are involved in the setting of `XGBoost`: they are called, in jargon, *hyperparameters*. They play a vital role in defining the structure of a machine learning model and controlling its learning process. They are set always prior to the training and they can greatly influence the model performance. For the `XGBoost` model applied in this study, the focus was set on tuning the following key hyperparameters to enhance the classification performance:

- *learning_rate*: a factor (usually far below 1) that slows down the gradient descent for the *log loss* minimization, thus preventing to pick local minima
- *gamma*: a value that specifies the minimum loss reduction required to split a node

- *max_depth*: maximum allowed depth of the decision trees
- *num_boost_round*: number of boosting rounds, namely the number of trees concurring to the final XGB decision
- *scale_pos_weight*: the ratio of the number of negative class samples to the number of positive class samples, used to balance the dataset.

All the other hyperparameters were maintained at their default settings as provided by the XGBoost library.

3.1.1 Performance Metrics and ROC curve

XGBoost is an example of *supervised* machine learning technique: the tool is first trained on a data sample (*training sample)* where each candidate has been already classified *a priori* (either as signal (S) or background (B), in a binary classification problem), then its performance are validated on another independent data sample (*test sample*) with *a priori* classification.

In the presented binary classification problem the background is digitized as 0, while the signal as 1. The prediction on the candidate output from XGBoost is a number between 0 and 1: an output threshold must be defined in order to convert the prediction into a binary classification. Candidates with output above (below) the threshold will be classified as signal (background). Since misclassification can occur, the results of a binary classifier on a candidate in the test sample can be grouped into four different cases mutually exclusive:

- True Positive (TP): Correct positive predictions (true signal classified as signal)
- True Negative (TN): Correct negative predictions (true background classified as background)
- False Positive (FP): Incorrect positive predictions (true background classified as signal)
- False Negative (FN): Incorrect negative predictions (true signal classified as background)

The Receiver Operating Characteristic (ROC) curve is a performance metric that displays the properties of the binary classifier as a function of the output threshold. The ROC curve shows in particular the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) as the output threshold is varied. TPR and FPR are defined as:

$$TPR = \frac{TP}{TP + FN} \tag{3.2}$$

$$FPR = \frac{FP}{TN + FP} \tag{3.3}$$

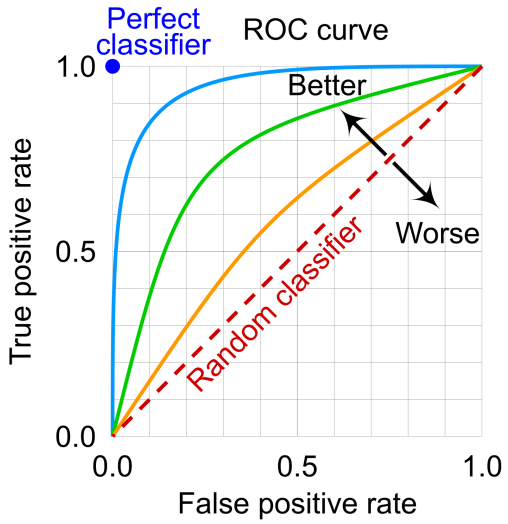


Figure 3.1: The ROC curve is a binary classifier performance metric. A random classifier is characterized by a line bisecting the TPR vs FPR plot (red dashed line), while the ROC curve for a perfect classifier is a point corresponding to TPR = 1 and FPR = 0. The blue, green, and yellow solid lines above the red line describe performances better than a random coin toss: performances are better as Area Under the Curve (AUC) is larger.

In the context of data selection obtained by a set of selection criteria TPR, represents the signal efficiency (ϵ_S) whereas FPR represents the background contamination (ϵ_B) i.e. the probability survival of the residual background. An optimal classifier would perfectly separate the signal from the background, resulting in $TPR = 1$ and $FPR = 0$, corresponding to the top-left blue point in Figure 3.1. Conversely, a random classifier will output a prediction uniformly between 0 and 1 either the candidate is true signal or true background, thus leading to the ROC Curve illustrated by the red dashed line in Fig. 3.1. A well-behaving trained classifier is expected to have much better performances than a random one.

The Area Under the ROC Curve (AUC) represents the classifier's overall performance across all the thresholds. A perfect model scores $AUC = 1$, while a random classifier scores $AUC = 0.5$. For multiclass classification, a combined AUC for all classes can be produced by treating each class as the signal and the others as the background during the ROC curve computation, repeating this for all classes. The uncertainty on the AUC measurement depends on the size of the training sample. In this study, the AUC uncertainty, given the available

size of the training sample, is estimated to be less than one permill; this is calculated as the RMS error of the AUC from five independent repetitions of the same training.

3.1.2 Kolmogorov-Smirnov test

A well-trained algorithm is expected to extract the general features of the training dataset in order to identify them on a target sample, thus discriminating between different classes. Nevertheless, an algorithm that learns also features peculiar to the specific training sample, which are not representative of the underlying process, is considered to be overtrained.

The Kolmogorov-Smirnov (KS) two-sample test [71] is a statistical method used to determine whether two samples come from the same continuous distribution. It is particularly useful in binary classification problems to understand if the trained algorithm is *overtrained*. The basic principle of the KS test is to calculate the largest vertical difference between two Cumulative Distribution Functions (CDFs) of the two distributions to be compared. The KS test can be used in fact to compare the XGBoutput (also called *score*) distributions for true signal (or background) candidates on the train sample and on the test sample. If the distributions from the train and test samples are not compatible with each other (which means that the KS test is failed), it is reasonable to assume that the algorithm was overtrained and has learned patterns specific to the training set that do not generalize well to the test set. On the contrary, a successful KS test is interpreted as a motivation for good training.

Its uses the maximum difference between the two CDFs as the test statistic. The KS test formulation can be given as follows:

- $f_n(x)$ and $g_m(x)$ are the two normalized distributions which are to be compared, each represents the distribution of the variable x for a different sample; n and m denote the sizes of each sample,
- $F_n(x)$ and $G_m(x)$ are respectively the CDFs of $f_n(x)$ and $g_m(x)$
- $D_{n,m}$ is the KS test statistic for a two-sample comparison and is defined as the largest absolute difference between the two CDFs, namely:

$$D_{n,m} = sup_x |F_n(x) - G_m(x)| \tag{3.4}$$

  where sup (*supremum*) represents the largest upper bound of a given set: it practically means finding the largest difference between two functions.

The distribution for the $D_{n,m}$ test statistics is well known but rather complex: tables are available to compute the significance level α corresponding to the

observed value, but still it is easier to compare it with known *critical values* which represent a good approximation of the exact value in the limit of large samples. The critical value D_α for a level of significance α in the large samples limit is given by:

$$D_\alpha = c(\alpha)\sqrt{\frac{n+m}{nm}} \tag{3.5}$$

where n and m are the sizes of the two samples, and $c(\alpha)$ is a tabulated factor for a given significance level (see Table 3.1). If $D_{n,m}$ exceeds this critical value, the null hypothesis that the two samples come from the same distribution is rejected and the KS test is failed.

α	0.10	0.05	0.025	0.01	0.005	0.001
$c(\alpha)$	1.22	1.36	1.48	1.63	1.73	1.95

Table 3.1: Factors $c(\alpha)$ in the calculation of critical values for the KS test statistic corresponding to a level of significance α [72].

It is equivalently possible to extract a p-value of the KS test statistic, that expresses the probability of obtaining a value below the observed one. A p-value below the level of significance (typically 0.05 or less) implies the rejection of the null hypothesis, namely the KS test is failed. Instead, a p-value above the level of significance is interpreted as a successful KS test.

3.2 Data and Monte Carlo samples as inputs for training

In order to train an `XGBoost` model for the extraction of the B_c signal, a Monte Carlo (MC) sample of approximately 500 thousands of $B_c^\pm \to J/\psi\pi^\pm$ events, in an equal mixture of B_c^+ and B_c^- candidates, was used. The sample was produced by using `PYTHIA8` [73] to generate high-energy pp collision events, coupled to the dedicated B_c generator (`BCVEGPY` [23]) discussed in Chapter 1 and `EvtGen` [25] to handle the heavy flavour decay chain under study. The MC sample is used as signal sample for the `XGBoost` training, and also to define a Signal Region (SR), that is a region of phase-space where the signal is expected to exist.

The background sample is instead obtained with a data-driven approach. In the real data sample a control region (CR), namely a region of the phase-space where no signal is expected (as complementary of the SR), is defined. This control region is called as *sidebands* and it is defined later. The amount of data available at the moment of the finalization of this thesis (2022 and 2023

datasets) corresponds to an equivalent integrated luminosity of 68.1 fb^{-1} of pp collisions collected by CMS during the LHC Run-3.

Both the data and MC events are processed by a dedicated `analyzer` module within the CMS `C++` reconstruction software framework, `CMSSW`. The `analyzer` implements candidates' reconstruction, a minimal preliminary selection and records all measured observables for each candidate in a ROOT file. The ROOT file is structured to be tabular: each entry corresponds to a B_c candidate and carries all the relevant physical information. It can happen that more than a B_c candidate is reconstructed in an event (namely a pp collision), and multiple candidate information is accommodated in the ROOT file as well. However, this contribution is negligible after selection (0.2% or 0.3% according to the selection), as also discussed later.

The B_c candidate is reconstructed, first, by looking for two opposite sign muons coming from a common vertex and compatible with a J/ψ meson. Subsequently, the analyzer looks for a charged track to be attached to the J/ψ candidate and form a B_c candidate: the pion mass is assigned to the charged track by default, then the two muons and the track are fitted to a common vertex with the kinematical constraint that the invariant mass of the di-muon is fixed to the J/ψ mass value from the PDG. Physical quantities derived from the kinematically constrained vertex fit will be denoted by indicating "J/ψ" instead of "$\mu\mu$" (e.g. the B_c candidate invariant mass is $m(J/\psi\pi)$). The common vertex from the fit is the Secondary Vertex (SV), namely the decay vertex of the B_c. The B_c production vertex, essentially coinciding with the "best" Primary Vertex (PV) of the event, is identified through a dedicated procedure. Among the many pileup vertices reconstructed by CMS (as described in Sec. 2.3.1), the one chosen as PV is the one that has minimum distance from the line where the momentum of the B_c candidate lies. The motivation for this particular choice is discussed in detail in Appendix B.

Even though both the data and MC samples go through the `analyzer` step and its selection, some residual differences are still present due to the filtering of rather soft B_c mesons in the MC production chain. Therefore, a further step, called `preprocess`, is applied on top of the analyzer selection to both the data and MC samples. The details of the two selections (`analyzer` and `preprocess`) are reported in Table 3.2. The full list and definitions of the variables are provided in the glossary in Appendix D.

3.2.1 Definition of Signal Region and Control Region

In the present work the Control Region is designated by defining a window for the invariant mass of the B_c candidate. For a Gaussian signal the 99.7% of the signal lies within a window centered in the Gaussian mean μ and with

Variables	**Units**	`analyzer`	`preprocess`
$p_T(\mu^{\pm})$	GeV	> 4	> 4
$\lvert\eta(\mu^{\pm})\rvert$	-	< 2.5	< 2.5
OS muons	-	✓	✓
μ HLT matched	-	✓	✓
Soft muons ID	-	✓	✓
$p_T(\mu\mu)$	GeV	-	10
$m(\mu\mu)$	GeV	[2.9, 3.3]	[2.9, 3.3]
d.c.a.$(\mu\mu)$	cm	< 0.5	< 0.5
$P_{vtx}(\mu\mu)$	-	> 0.01	> 0.01 (0.1)
trk HLT matched	-	×	× (✓)
$\lvert\eta(trk)\rvert$	-	< 2.5	< 2.5
$p_T(trk)$	GeV	> 0.5	> 1.2 (2.4)
$p_T(J/\psi\pi)$	GeV	> 8	> 10
$P_{vtx}(J/\psi\pi)$	-	> 0.01	> 0.01 (0.04)
trk ID	-	π	π
$m(J/\psi\pi)$	GeV	[6.05, 6.5]	[6.05, 6.5]

Table 3.2: Summary table comparing the preliminary selections applied at `analyzer`, and later offline to prepare the signal and background samples for `XGBoost` (`preprocess`). A tighter version of the `preprocess` is considered in parentheses: it will be used for the only purpose of `XGBoost` validation in Sec. 3.3.

an half-width of 3σ. In our case, the σ quantity is extracted by fitting the signal MC sample after the `preprocess` selection is applied; specifically the B_c invariant mass distribution fit is interpolated by using a Johnson p.d.f. model [74] (Figure 3.2), which is a generalization of the Gaussian p.d.f. that takes into account possible shape asymmetries by means of two parameters (δ and γ); the Johnson p.d.f. is a Gaussian when $\delta = 1$ and $\gamma = 0$. The extracted σ is ≈ 31 MeV and represents the mass resolution characterizing the CMS reconstruction. Thus, the signal window, centered on the B_c mass value estimated from the fit, is $6.1839 < m(J/\psi\pi) < 6.3695$ GeV.

The sidebands, namely the regions outside the signal window, define the control region for the real data, as almost no signal is expected to be there. To summarize, the signal and background samples used as an input for the `XGBoost` training and taken by the signal MC and real data samples respectively. The sigma is used to define the signal and control regions. With the available simulated and real data samples, the training data are characterized by about 10 millions of background candidates in the real data sidebands, and half million of simulated signal candidates respectively.

The designed selection is then applied to the whole available data sample,

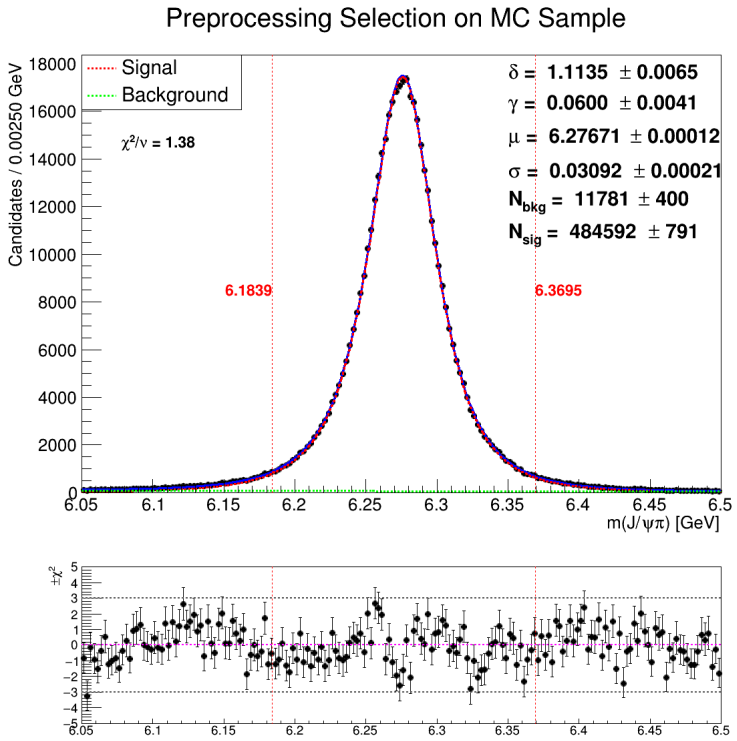


Figure 3.2: Top plot: Distribution of the invariant mass of the B_c candidate on MC sample after the `preprocess` selection. The distribution is modeled with a Johnson p.d.f. [74] and a binned ML fit is performed. The 3σ window is delimited by the vertical red dashed lines. Bottom plot: Bin-by-bin pull distribution.

and the B_c signal is extracted by performing a Maximum Likelihood (ML) fit to the mass spectrum. Different selections are compared across this thesis work by means of two figures of merit: the *signal significance* (SS) and the *purity*(p).

The signal significance is defined as

$$SS = \frac{S}{\sqrt{S+B}} \tag{3.6}$$

where S is the signal yield in the 3σ signal window, while B is the estimated background yield in the same window. For different purpose a similar figure-of-merit will be introduced and used later.

The purity is, instead, defined considering the signal and background yields (S and B) in a smaller window, with an half-width of 2σ, according the following formula:

$$p = \frac{S}{S+B} = \frac{1}{\sqrt{S+B}} \cdot SS \tag{3.7}$$

The purity is a number between 0 and 1, where 0 corresponds to the absence of signal in the considered 2σ window, while 1 corresponds to an ideal pure sample without any background contamination.

3.3 Performances of cut-based selections

The purpose of this chapter is to compare an XGBoost-based approach with a standard "cut-based" selections. As previously discussed (Sec. 3.1.1), the performance of an `XGBoost` BDT can be quantified by means of a ROC curve, which provides a comprehensive measure of the classifier's ability to distinguish between signal and background across various thresholds. However, cut-based selections is characterized by a unique specific working point (WP) to be compared with the `XGBoost` ROC curve. In this context, this WP will be denoted as ϵ_S and ϵ_B.

The overwhelming background still present after the `preprocess` step is not suitable for performing a preliminary fit before the selection, which is necessary to establish a reference. Therefore, in the context of this chapter, a `tighter preprocess` selection is considered to better define the signal and background regions. This `tighter preprocess` will be dropped in the following chapters once the comparison between standard and ML-based approaches is established.

The evaluation of the cut-based WP is based on the following strategy that has been used to estimate the performance of the cut-based selections:

1. A fit (signal + background model) is performed after the cut-based selection to determine the composition in terms of signal (S) and background (B) within the signal region.

2. Another fit (signal+background model) is performed before the cut-based selection and after the `tighter preprocess` selection. The signal (S') and background (B') yields in the signal region after this preprocess step are used as reference in the denominator of the following quantities:

$$\epsilon_S \equiv \mathrm{WP}_{\mathrm{TPR}} = \frac{S}{S'} \tag{3.8}$$

$$\epsilon_B \equiv \mathrm{WP}_{\mathrm{FPR}} = \frac{B}{B'} \tag{3.9}$$

where, again, the definition of the quantities used is as follows:

- S: Signal yield in the signal region (SR) after the cut-based selection;
- S': Signal yield in the signal region (SR) after the `tighter preprocess` selection and before the cut-based selection;
- B: Background yield in the signal region (SR) after the cut-based selection.

- B': Background yield in the signal region (SR) after the `tighter preprocess` selection and before the cut-based selection.

A scheme for the considered selections (better described later in Sec. 3.3.1 and 3.3.2) is presented in Figure 3.3.

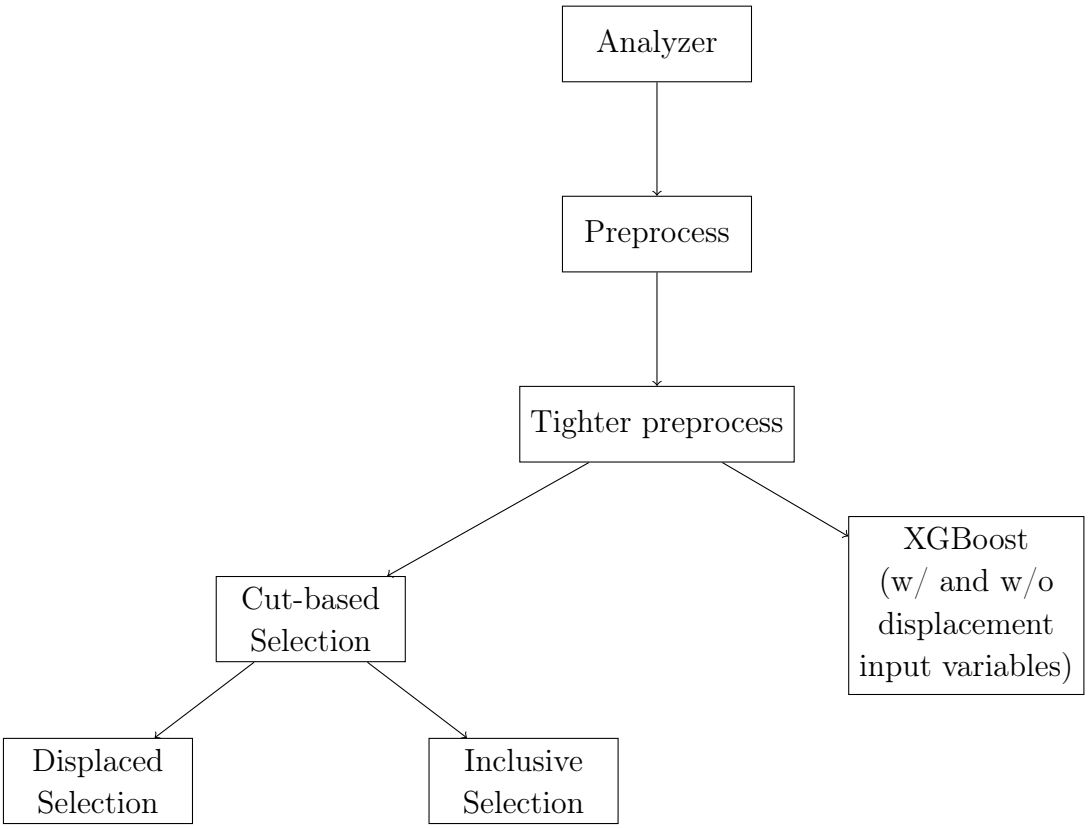


Figure 3.3: Diagram of the cut-based selections used in the present chapter to compare the performance of `XGBoost` (in two different settings) with standard cut-based approaches (denoted as `displaced`, described in Sec. 3.3.1, and `inclusive`, described in Sec. 3.3.2).

A set of few selections has been defined to be as loose as possible, yet still allowing a stable signal extraction through a fitting procedure that is relatively insensitive to small variations in the applied selection criteria.

The following further selection is applied:

- $P_{vtx}(\mu^+\mu^- trk) > 0.04$
- $p_T(trk) > 2.4$ GeV

The tighter `preprocess` selection (summarized in Table 3.2) enhances the signal peak, as shown in Figure 3.4, despite the large contribution from the combinatorial background. The signal model is a Gaussian, while the background model is a first-order polynomial: the obtained signal resolution is compatible with the resolution of approximately 31 MeV extracted from a fit on the signal MC sample (Fig. 3.2).

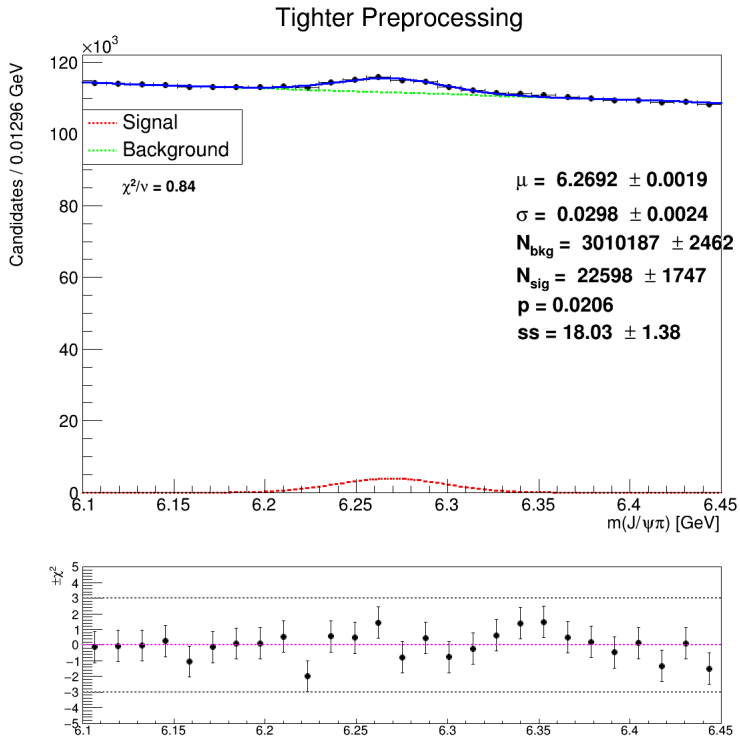


Figure 3.4: Top plot: Distribution of the invariant mass of the B_c candidates after the tighter `preprocess` selection. A clear signal shape compatible with the B_c meson in terms of mass and resolution is observable on top of an overwhelming background. The signal and background yields in the SR are extracted by means of a Binned Maximum Likelihood fit. The signal model is Gaussian (red dashed line), while the background model is a first-order polynomial (green dashed line). Bottom plot: Bin-by-bin pulls distribution.

3.3.1 Displaced cut-based selection as reference

As anticipated in Chapter 1 the $B_c^+ \to J/\psi\pi^+$ decay channel has been already reconstructed in CMS with the LHC Run-2 data thanks to the `displaced JpsiTrk trigger` (discussed in Sec. 2.5). The selection designed in the framework of the Run-2 analysis is considered as a reference and will be denoted from now on as `cut-based displaced reference`. The details of the selection are reported in Table 3.3.

The cut-based `displaced` selection is applied to the Run-3 data collected in years 2022-2023 and the obtained invariant mass distribution of the B_c candidates is shown in Figure 3.5. The distribution is fit with a sum of a Gaussian p.d.f., that accounts for the B_c signal, and an exponential p.d.f., describing the combinatorial background. A possible improvement of the background description can be obtained by including a contribution from partially reconstructed candidates (as in the Run-2 published analysis) which is not included in the present comparison as it would have a negligible impact on the fit result. The

Variables	**Units**	`preprocess`	`inclusive`	`displaced`
$p_T(\mu^{\pm})$	GeV	> 4	> 4	> 4
$\lvert\eta(\mu^{\pm})\rvert$	-	< 2.5	< 2.5	< 2.5
OS muons	-	✓	✓	✓
μ HLT matched	-	✓	✓	✓
Soft muons	-	✓	✓	✓
$p_T(\mu\mu)$	GeV	10	> 15	> 10
$m(\mu\mu)$	GeV	[2.9, 3.3]	[2.9, 3.3]	[2.9, 3.3]
d.c.a($\mu\mu$)	cm	-	< 0.5	< 0.5
$P_{vtx}(\mu\mu)$	-	> 0.01	> 0.1	> 0.1
$\cos(\alpha_{\mu\mu})$	-	-	-	> 0.9
$L_{xy}/\sigma_{L_{xy}}$	cm	-	-	> 0.01
trk HLT matched	-	× (✓)	×	×
$\lvert\eta(trk)\rvert$	-	< 2.5	< 2.5	< 2.5
$p_T(trk)$	GeV	> 1.2 (2.4)	> 5.0	> 3.5
$p_T(J/\psi\pi)$	GeV	> 10	> 15	> 15
$P_{vtx}(J/\psi\pi)$	-	> 0.01 (0.04)	> 0.1	> 0.1
$\cos(\alpha^{3D}_{B_c})$	-	-	> 0.997	> 0.98
trk ID	-	π	π	π
$d_0(trk)/\sigma_{d_0}$	-	-	-	> 2
$m(J/\psi\pi)$	GeV	[6.05, 6.5]	[6.05, 6.5]	[6.05, 6.5]
ct	cm	-	-	> 0.01

Table 3.3: Summary table of the `preprocess` selection (tighter version in parentheses), the newly-developed cut-based selection (`inclusive`), and the (`displaced`) selection discussed earlier.

word "partially reconstructed background" refers to decays into a $J/\psi+\pi+X$ final state, where X can be one or more particles that are not reconstructed (typically pions). Thus these decays can contribute to the background at lower mass values than the B_c^+ nominal mass and are labelled as "partially reconstructed background". A signal region of 3σ is defined after the fit to extract the signal (background) yield in the SR. According to Eq. 3.8 and 3.9, the working point for the cut-based `displaced` selection corresponds to a couple of values, $\epsilon_S \equiv \mathrm{WP_{TPR}} \approx 0.2124$ and $\epsilon_B \equiv \mathrm{WP_{FPR}} \approx 0.0023$ (cut-based WP).

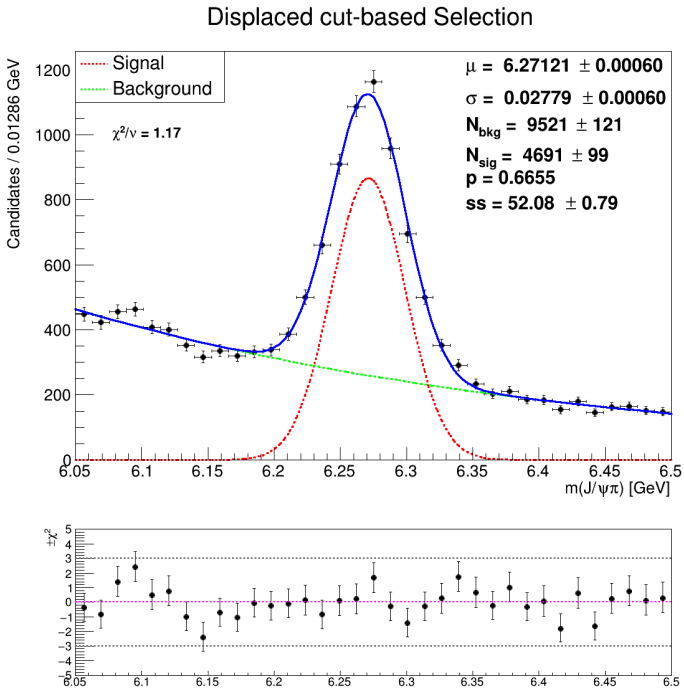


Figure 3.5: Top plot: Distribution of the invariant mass of the B_c candidates after the `displaced` selection (inherited from the Run-2 analysis). The signal model (red dashed line) is a Gaussian, while the background model (green dashed line) is a falling exponential. The result of the UML fit are superimposed (solid blue line). Bottom plot: Bin-by-bin pull distribution.

3.3.2 Inclusive cut-based selection as reference

With the start of the LHC Run-3 the new dedicated `inclusive Bc` trigger (discussed in Sec. 2.5) started taking data. A study using a cut-based approach was preliminarily studied in order to assess and monitor the trigger performances: this selection will be considered as a reference for the "inclusive" study and will be denoted from now on as cut-based `inclusive reference`. The details of the selection are reported in Table 3.3 and compared with those of the cut-based `displaced reference` introduced earlier.

The invariant mass distribution of the B_c candidates selected by the *inclusive reference* selection is shown in Figure 3.6. The distribution is fit with a sum of a Gaussian p.d.f., that accounts for the signal, and an exponential p.d.f., describing the combinatorial background. A clear signal is observed in spite of the large background contribution, with respect to the displaced selection (Figure 3.5). The large background can be easily explained by the fact that the displaced selection benefits from cutting on displacement variables like $\cos(\alpha_{\mu\mu})$, $L_{xy}/\sigma_{L_{xy}}$, and others as shown in Table 3.3.

A signal region of 3σ is defined after the fit to extract the signal (back-

ground) yield in the SR. The working point for the cut-based `inclusive` selection corresponds to a couple of values $\mathrm{WP_{TPR}} = \approx 0.2406$ and $\mathrm{WP_{FPR}} = \approx 0.0209$.

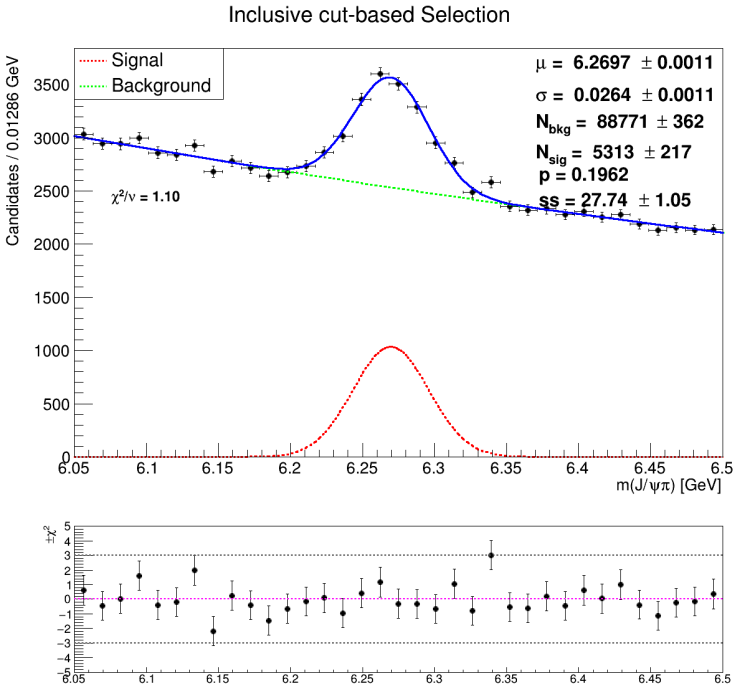


Figure 3.6: Top plot: Distribution of the invariant mass of the B_c candidates collected after the `Inclusive Selection`. The results of the UML fit are superimposed (solid blue line). The signal model (red dashed line) is a Gaussian function, while the background model (green dashed line) is a an exponential. Bottom plot: Bin-by-bin pull distribution.

Quantity	cut-based `inclusive`	cut-based `displaced`
S	5298 ± 217	4677 ± 99
B	31174 ± 127	3391 ± 43
B'	1490888 ± 867	1490888 ± 867
S'	22025 ± 1747	22025 ± 1747
ϵ_S	0.2406 ± 0.0215	0.2124 ± 0.0174
ϵ_B	0.0209 ± 0.0001	0.0023 ± 0.00003
Purity	0.1962	0.6655

Table 3.4: Summary table of signal and background candidates obtained according to `inclusive` and `displaced` selections and calculated $\epsilon_S \equiv \mathrm{WP_{TPR}}$ and $\epsilon_B \equiv \mathrm{WP_{FPR}}$ values. The S' and B' values are obtained from the tighter `preprocess` selection and are common for both approaches.

3.4 XGBoost for analysis selection

The ROOT data files (called in old HEP jargon *ntuples*) in tabular format are converted into `pandas` *dataframes* in order to be compatible with the `XGBoost` library. The data are prepared by creating a *dataframe* for each signal and background, then adding a tag-variable to respectively label the signal (tag 1) and the background (tag 0). Afterwards, the two dataframes are merged and then randomly shuffled in order to avoid any selection bias in the subsequent steps. The new merged dataframe is ready to be split in two parts: the larger (70%) is used as training dataset, whereas the smaller (30%) for test purposes.

In order to evaluate the performance of `XGBoost` for the analysis selection, two different settings are considered, labelled as `XGBoost inclusive` and `displaced` approach in analogy to the cut-based approaches. In the `inclusive` approach no quantities carrying information related to the displacement of the B_c candidate from the Primary Vertex are used, and this can be fairly compared with the cut-based inclusive selection discussed earlier in Sec. 3.3. In both cases a minimal set of input variables is defined by following an iterative procedure to reduce the input list while keeping a good description of the physics process. The iterative procedure of input reduction, based on the reference ranking, is described with some details in Appendix A: the resulting chosen variables for each approach are reported in Table 3.5.

Variables	Displaced	Inclusive
$\cos(\alpha^{3D}_{B_c})$	✓	✓
$L_{xy}/\sigma_{L_{xy}}$	✓	×
$P_{vtx}(J/\psi\pi)$	✓	✓
$p_T(J/\psi\pi)$	✓	✓
$\Delta R(J/\psi, track)$	✓	-
$\frac{d_0(trk)}{\sigma_{d_0}}$	✓	×
$\Delta R(\mu^+\mu^-)$	✓	-
$\cos(\alpha^{2D}_{\mu^+\mu^-})$	✓	✓
$\lvert\eta(trk)\rvert$	✓	✓
$p_T(trk)$	✓	✓

Table 3.5: Variables used in `displaced` and `inclusive` `XGBoost` approaches. The descriptions of the variables are explained in Appendix D. Variables marked with × are correlated to B_c displacement and are not used in the `inclusive` approach.

Given the high statistical significance of the data and the relatively low number of boosting rounds used, the `early_stopping_round` mechanism was not used in this study. Later, in Chapter 4, the optimization of the hyperparameters for each model will be investigated and discussed in detail. Just for

Hyperparameter	Value
learning_rate	0.05
max_depth	2
gamma	0
num_boost_round	500
scale_pos_weight	6.636

Table 3.6: Hyperparameters used in XGBoost for training after tighter preprocessing for comparison with reference studies.

reference, Table 3.6 shows the baseline values used in the XGBoost performance study in this chapter.

3.4.1 Performance of XGBoost displaced model

In this section the results of an XGBoost model trained on the displaced set of variables (as listed in Table 3.5) is presented. The workflow for the data preparation is discussed in Sec. 3.4.

The *log loss* curve for the displaced approach is shown in Figure 3.7; it can be appreciated that the curves for the training and test sets are not significantly separated, indicating that there is no overtraining. Moreover the trend of the *log loss* function is considered safe as it steadily decreases. No important gain in the *log loss* function was observed for the same model when more than 500 trees are considered.

The XGB output (*score*) on the signal and background samples is shown in Figure 3.8. Separate (normalized) distributions are shown for the train and test sample: the signal (background) distribution is peaking at 1 (0) as expected. The signal (background) train and test distributions are well compatible with each other: this is quantified by applying the Kolmogorov-Smirnov test to these distributions. The results for the signal and background samples are respectively shown in Figures 3.9 and Figure 3.10. In both cases a large p-value for the compatibility of the distributions is observed ($p = 0.78$ on the signal sample, $p = 0.98$ on the background sample): it can be calculated that the considered model is properly trained.

The clear separation between the signal and background distributions in Figure 3.8 demonstrates the classification power of the XGB displaced model. This can be quantified by the resulting ROC curve, shown in Figure 3.11. The ROC AUC for the XGB displaced model is found to be $\simeq 0.966$ (with an uncertainty less then permill), and it is compared with the results from cut-based displaced reference (discussed in Sec. 3.3.1). The large AUC indicates the effectiveness of the model; the reference working point (green dot in Figure

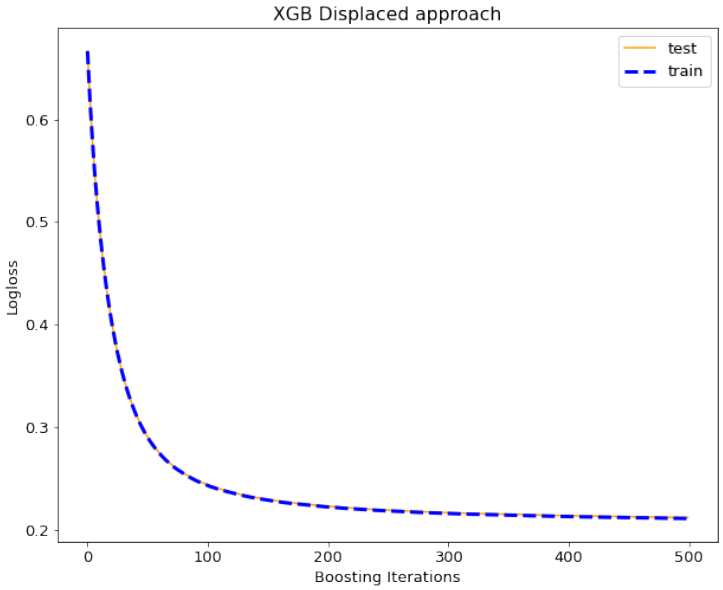


Figure 3.7: *Log loss* curve for the `displaced XGB` model. The *log loss* function is evaluated on the train sample (blue curve) and the test sample (orange curve). The two curves are very close to each other (practically superimposed), indicating no overtraining. The *log loss* value stabilizes when 500 (or more) trees are considered.

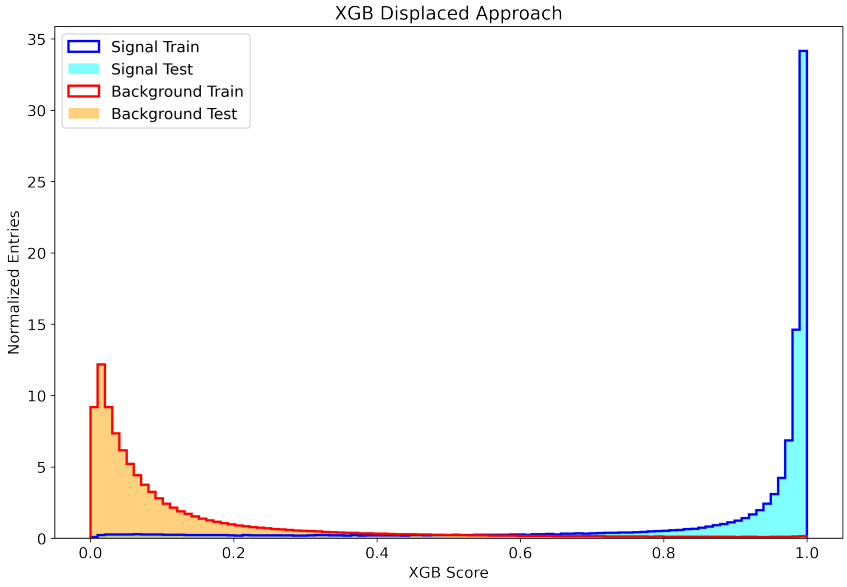


Figure 3.8: Score distribution curves for the `XGB` displaced model output on the train and test samples, separately for signal and background candidates. A clear separation between the background and signal distributions indicating the effectiveness of the classification model. The compatibility between signal (background) train and test histograms - later evaluated with the KS test - indicates that the model is properly trained (no overfitting).

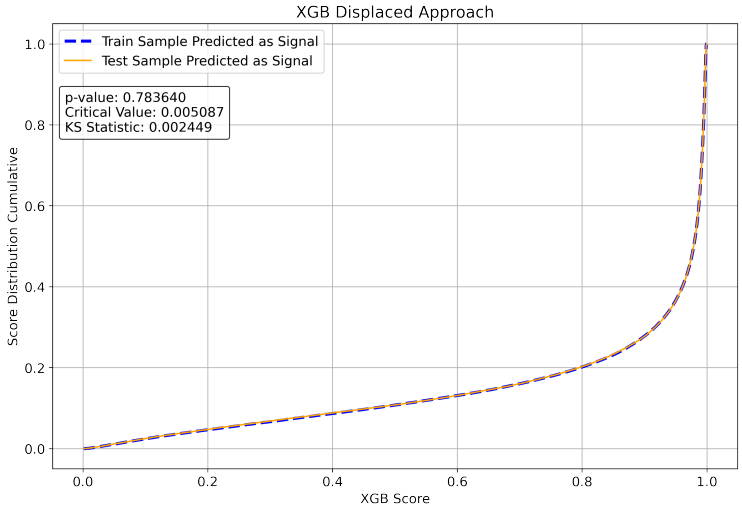


Figure 3.9: Results of the KS test on the signal sample for the XGB displaced approach. CDFs of the score density distributions (see cyan/blue histograms in Fig. 3.8): the blue curve is for the training sample, while the orange for the test sample. The maximum distance between the two CDFs is the KS-distance, used to evaluated the p-value for the compatibility of the two distributions. The two CDFs are almost superimposed and the KS test is passed, meaning that distributions are compatible with each other.

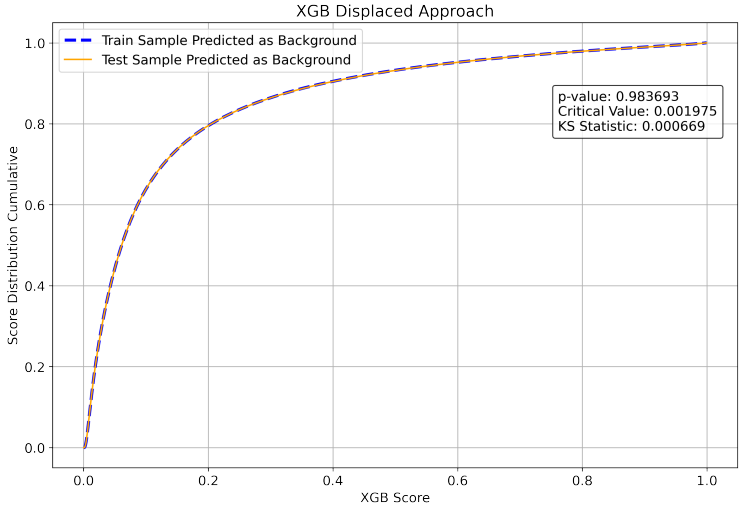


Figure 3.10: Results of the KS test on the background sample for the XGB displaced approach. CDFs of the score density distributions (see red/orange histograms in Fig. 3.8): the blue curve is for the train sample, while the orange for the test sample. The maximum distance between the two CDFs is the KS-distance, used to evaluated the p-value for the compatibility of the two distributions. The two CDFs are almost superimposed and the KS test is passed, meaning that distributions are compatible with each other.

3.11) is placed below the ROC curve (orange curve) - clearly observable in log-scale - thus meaning that XGB outperforms the cut-based approach as expected.

Once the training is done, different working points have been studied and

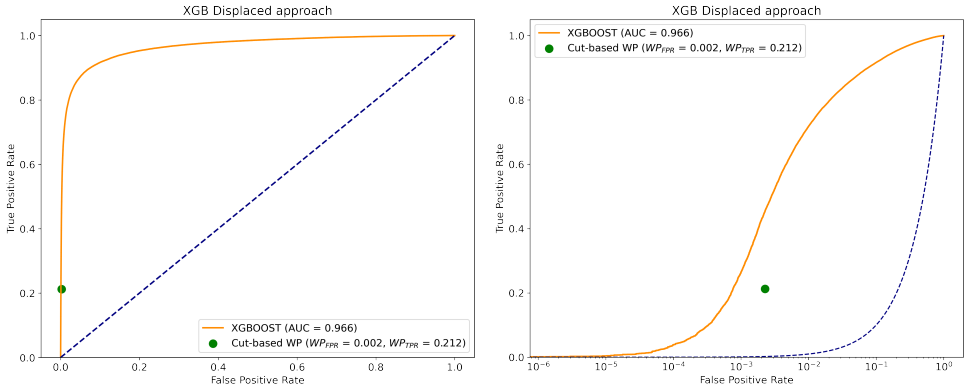


Figure 3.11: The orange line is the ROC curve (linear x-axis on left, logarithmic x-axis on right) for the XGB displaced model. The green point indicates the performances of the cut-based displaced selection. The blue dashed line represents the random output as a reference.

the corresponding signal is extracted by means of an Unbinned Maximum Likelihood fit to the mass spectrum. One working point is chosen so that it leads to a signal yield similar to the one obtained from the cut-based reference, thus providing a better purity (Same Signal yield working point, SS WP); the other working point is instead chosen to have the same signal purity of the cut-based approach, resulting in a larger signal yield (Same Purity working point, SP WP).
The selected WPs are shown on the ROC curve in Figure 3.12.

The invariant mass distribution for the B_c candidates selected with SP WP is shown in Figure 3.13: the fit models the signal with a Johnson p.d.f., the combinatorial background with a "falling" exponential, and an additional contribution from the partially reconstructed background is described with a sigmoid function. The signal mass and resolution obtained are $\mu = 6.2763 \pm 0.0013$ GeV and $\sigma = 0.0339 \pm 0.0028$ GeV, respectively. The mass result is compatible with the PDG value $M_{PDG}(B_c) = 6.27447 \pm 0.00032$ GeV and consistent with previous CMS results considering the experimental mass resolution [28, 68]. The extracted signal yield is $N_{B_c} = 11445 \pm 789$ events, which is about double of the signal yield from the reference selection, while the purity $p = 0.663$ is the same for the two selections.

For the SS WP, the invariant mass distribution for the B_c candidates selected with SS WP is shown in Figure 3.14. Unlike the previous cases, the signal model here uses the sum of two Gaussians with a common mean rather than a single Gaussian. The obtained mass value and resolution for the two Gaussians are $\mu = 6.27070 \pm 0.00057$ GeV, $\sigma_1 = 0.0183 \pm 0.0035$ GeV, and $\sigma_2 = 0.0347 \pm 0.0036$ GeV, respectively. The weighted resolution (where the weights are derived from the yields of the two Gaussians) is $\sigma = 0.0304 \pm 0.0035$ GeV. The estimated mass result is consistent with the PDG value and previ-

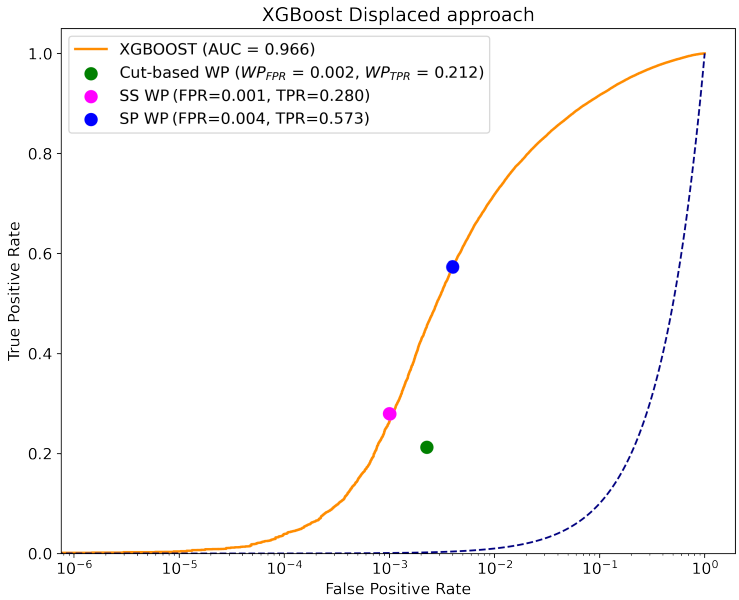


Figure 3.12: ROC curve for the XGB displaced model. The green point indicates the performances of the cut-based reference selection performance. Two WPs with similar signal yield (SS WP) and similar purity (SP WP) have been found for the XGB model and shown in the curve for comparison.

ous CMS results [28, 68]. The extracted signal yield is matching the reference yield, while the obtained purity is $p = 0.8988 \pm 0.0030$, which is significantly higher than the "reference" purity of 0.663.

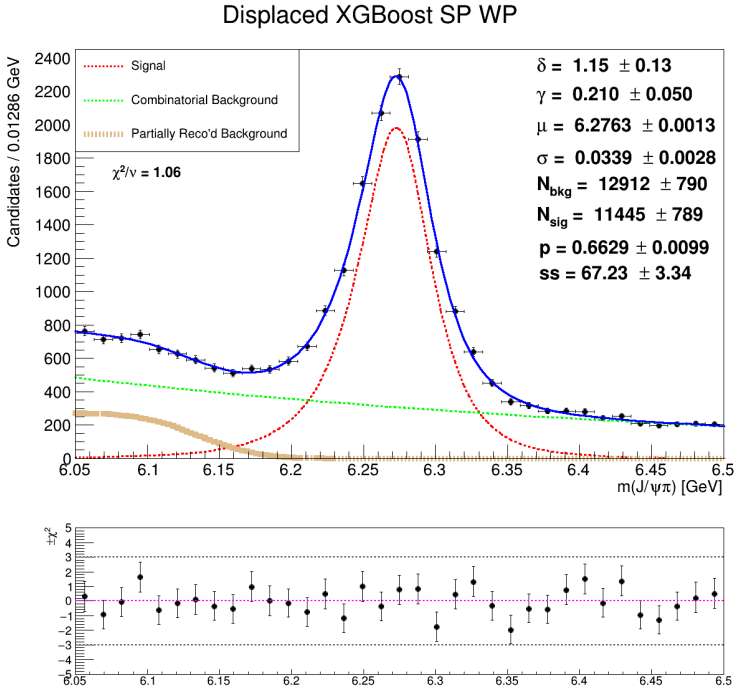


Figure 3.13: Distribution of the invariant mass of B_c candidates selected by the displaced XGB model with the SP WP. The result of the UML fit (blue curve) is superimposed. The fit model components are a Johnson p.d.f. for the signal (red dashed curve), an exponential for the combinatorial background (green dashed curve), and a sigmoid for the partially reconstructed background (brown dashed curve). Bottom plot: Bin-by-bin pull distribution, providing residuals for each bin of the chosen binning.

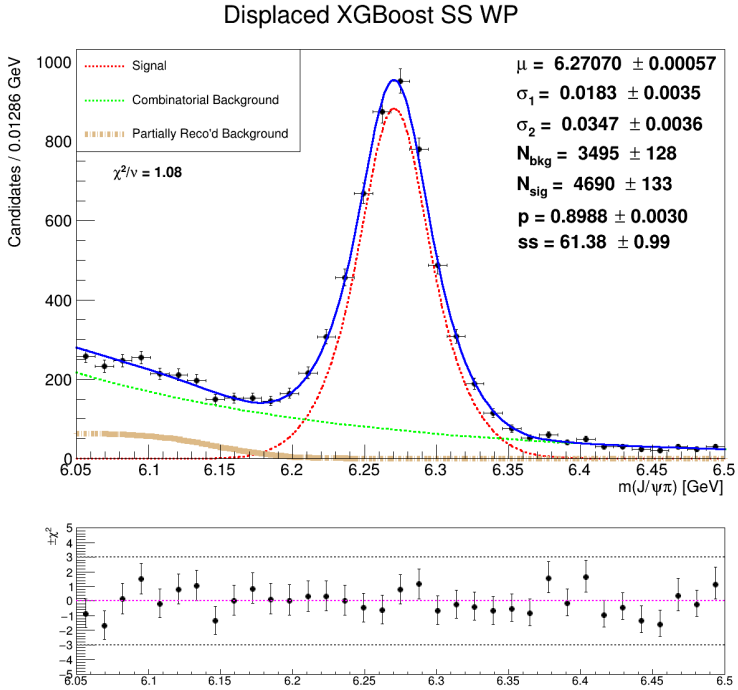


Figure 3.14: Distribution of the invariant mass of B_c candidates selected by the displaced XGB model with the SS WP. The result of the UML fit (blue curve) is superimposed. The fit model components are: a sum of two Gaussians with common mean for the signal (red dashed curve), an exponential for the combinatorial background (green dashed curve), and a sigmoid for the partially reconstructed background (brown dashed curve). Bottom plot: Bin-by-bin pull distribution, providing residuals for each bin of the chosen binning.

3.4.2 Performance of XGBoost inclusive model

In this section the results of an XGBoost model trained on the inclusive set of variables (as listed in Table 3.5) is presented. The workflow for the data preparation is discussed in Sec. 3.4.

The *log loss* curve for the inclusive approach is shown in Figure 3.15: the curves for the training and testing sets are not significantly separated, indicating that there is no overtraining. Moreover the trend of the *log loss* function is considered safe as it steadily decreases. No important gain in the *log loss* function was observed for the same model when more than 500 trees are considered.

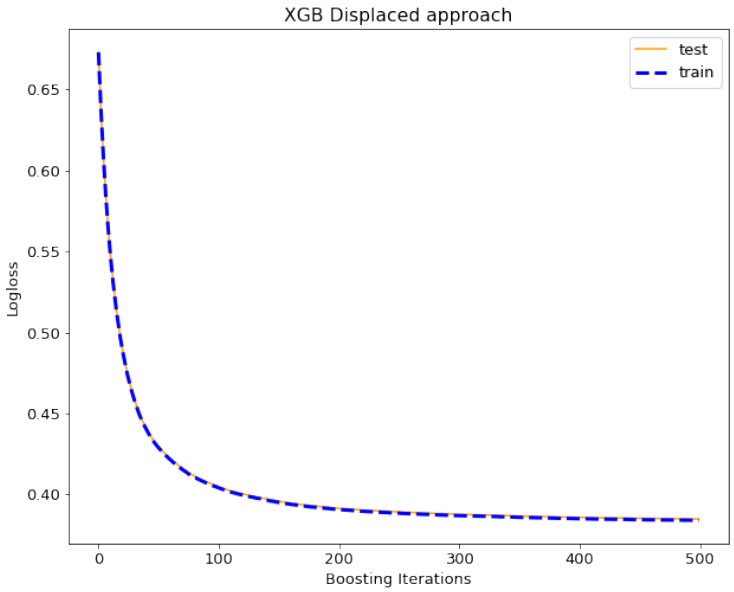


Figure 3.15: *Log loss* curve for the inclusive XGB model. The *log loss* function is evaluated on the training sample (blue curve) and the test sample (orange curve). The two superimposed curves are very close to each other (practically superimposed), thus indicating no overtraining.

The XGB output (*score*) on the signal and background samples is shown in Figure 3.16. Separate (normalized) distributions are shown for the training and test sample: the signal (background) distribution is peaking at 1 (0) as expected. The signal (background) training and test distributions are well compatible with each other, as confirmed by the KS test on both the signal and background samples, respectively shown in Figures 3.17 and Figure 3.18. In both cases a large p-value for the compatibility of the distributions is observed ($p \simeq 0.76$ on the signal sample, $p \simeq 0.42$ on the background sample): since both the p-values are larger than 5%, the model is considered properly trained.

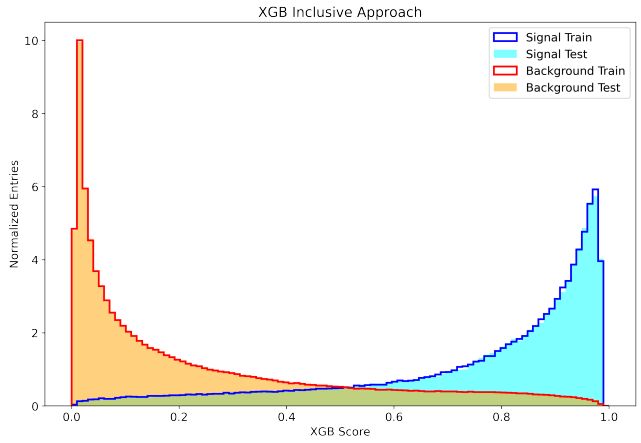


Figure 3.16: Score distribution curves for the XGB inclusive model output on the train and test samples, separately for signal and background candidates. A clear separation between the background and signal distributions indicates the effectiveness of the classification model. The compatibility between signal (background) training and test histograms - evaluated with the KS test - indicates that the model is properly trained (no overfitting).

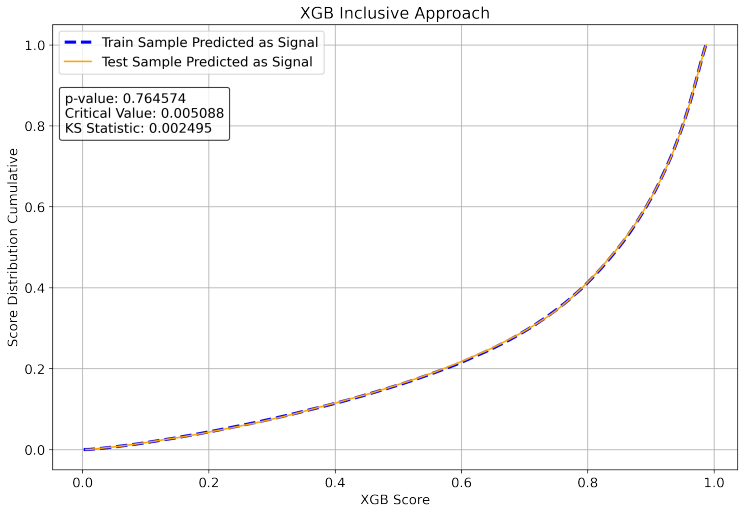


Figure 3.17: Results of the KS test on the signal sample for the XGB inclusive approach are summarized in the legend. The test was the CDFs of the score density distributions (see cyan/blue histograms in Fig. 3.16): the blue curve is for the train sample, while the orange for the test sample. The maximum distance between the two CDFs, the KS-distance, is used to evaluated the p-value for the compatibility of the two distributions. The two CDFs are almost superimposed and the high p-value motivates that the KS test is passed, meaning that distributions are compatible with each other.

The clear separation between the signal and background distributions in Figure 3.16 illustrates the classification power of the XGB inclusive model. This can be quantified by inspecting the resulting ROC curve, shown in Figure 3.19, and evaluating the AUC FoM.

The ROC AUC for the XGB inclusive model is found to be 0.910 (uncertainty

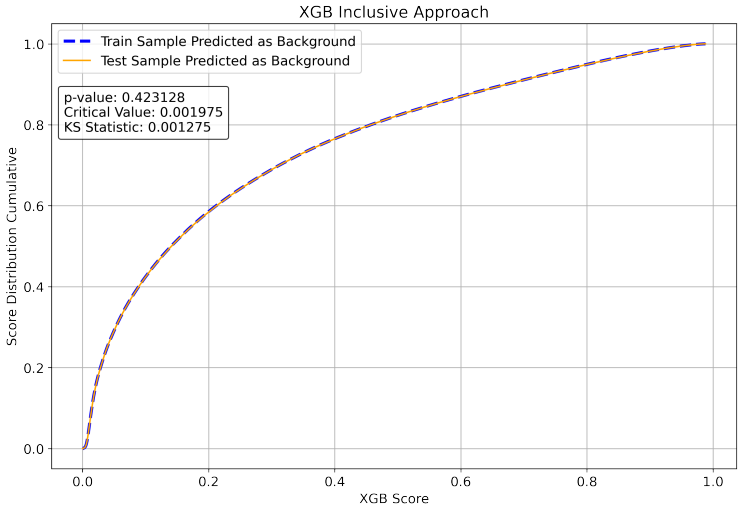


Figure 3.18: Results of the KS test on the background sample for the XGB inclusive approach are summarized in the legend. The test was the CDFs of the score density distributions (see red/orange histograms in Fig. 3.16): the blue curve is for the train sample, while the orange for the test sample. The maximum distance between the two CDFs, the KS-distance, is used to evaluated the p-value for the compatibility of the two distributions. The two CDFs are almost superimposed and the high p-value motivates that the KS test is passed, meaning that distributions are compatible with each other.

less then permill), and it is compared with the results from cut-based inclusive reference (discussed in Sec. 3.3.2). The reference working point (red dot in Figure 3.19) is placed below the ROC curve (orange curve) - clearly observable in log-scale - thus meaning that XGB outperforms the cut-based approach as expteced also in this case. However, the smaller AUC with respect to the XGB displaced model hints that the ML algorithm can learn much about the B_c mesons candidate from information on its displacement from the PV.

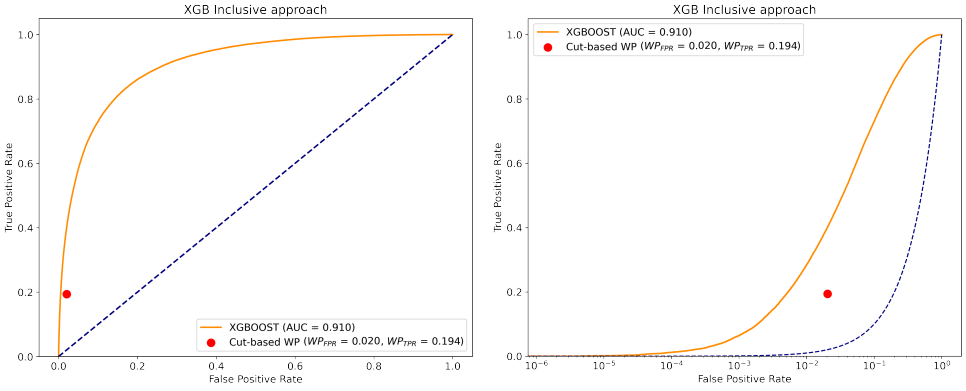


Figure 3.19: The orange line is the ROC curve (linear x-axis on left, logarithmic x-axis on right) for the XGB inclusive model. The red point indicates the performances of the cut-based inclusive selection. The blue dashed line represents the random output as a reference.

Once the training is done, different working points have been studied and the corresponding signal is extracted by means of an Unbinned Maximum Likelihood fit to the B_c invariant mass. In analogy to the displaced study, a same signal (SS) and a same purity (SP) WPs have been identified. The selected WPs are shown on the ROC curve in Figure 3.20.

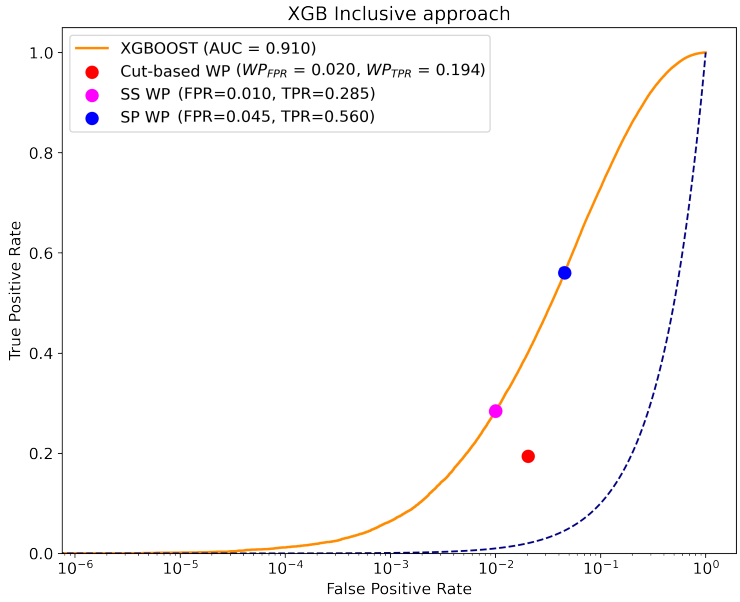


Figure 3.20: ROC curve (with logarithmic x-axis) for the XGB inclusive model. The red point indicates the performances of the cut-based reference selection performance. Two WPs with similar signal yield (SS WP) and similar purity (SP WP) have been found for the XGB model and shown in the curve for comparison.

The fit model in this case is the sum of a Gaussian and a "falling" exponential for both working points: the larger combinatorial background contribution overwhelms other background sources (such as partially reconstructed candidates), therefore the fit model has not been improved as this goes beyond the scope of the simple comparison of the XGB tool with a cut-based selection.

The invariant mass distribution of B_c candidates selected by the XGB inclusive model with the SP WP are shown in Figure 3.21. The obtained mass value and resolution are $\mu = 6.27103 \pm 0.0007$ GeV and $\sigma = 0.02657 \pm 0.0008$ GeV, respectively. The mass estimate is well compatible with the PDG value $M_{PDG}(B_c) = 6.27447 \pm 0.00032$ GeV, WPs are as the width is consistent with the experimental mass resolution estimated by means of the simulated signal. The extracted signal yield is $N_{B_c} = 10465 \pm 307$ events, which about doubles the reference yield obtained in the cut-based approach.

For the SS WP instead, the obtained mass value and resolution are $\mu = 6.27105 \pm 0.0007$ GeV, $\sigma = 0.02618 \pm 0.0007$ GeV respectively (results in Figure 3.22). The mass result is again consistent with the PDG value. A signal of

approximately 5 thousands candidates is extracted with a purity of $p = 0.3647$, which is almost double the purity achieved with the cut-based selection, $p = 0.1962$.

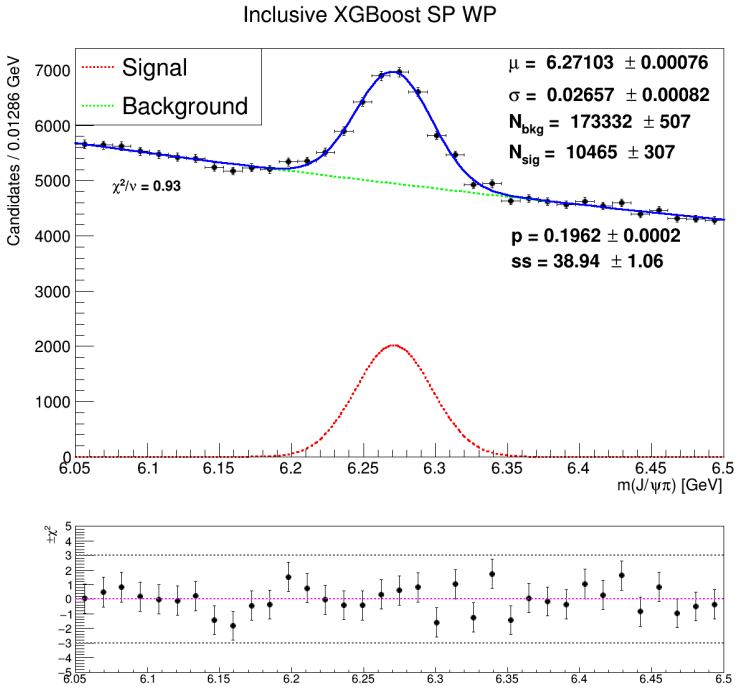


Figure 3.21: Distribution of the invariant mass of B_c candidates selected by the inclusive XGB model with the SP WP. The result of the UML fit (blue curve) is superimposed. The fit components are a Gaussian for the signal (red dashed curve), a falling exponential for the combinatorial background (green dashed curve), and a sigmoid for the partially reconstructed background (brown dashed curve). Bottom plot: Bin-by-bin pull distribution.

The results presented and discussed in this chapter are summarized in the following Table 3.7. As expected, the usage of XGBoost generally well enhances the signal extraction performance with respect to the cut-based "reference" selection in both cases: For a similar purity the signal significance is much better, whereas for a similar SS the purity is much better. This holds for both topologies.

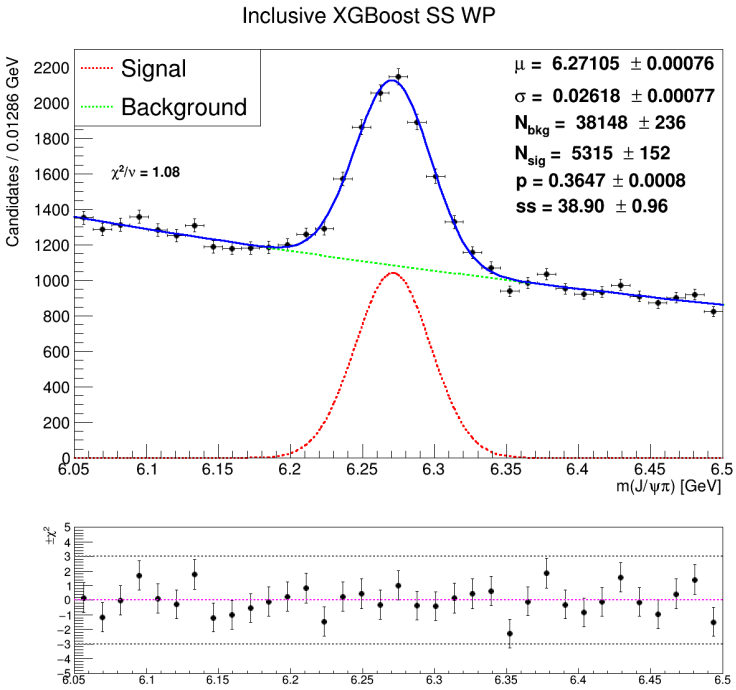


Figure 3.22: Distribution of the invariant mass of B_c candidates selected by the inclusive XGB model with the SP WP. The result of the UML fit (blue curve) is superimposed. The fit components are a Gaussian for the signal (red dashed curve), a falling exponential for the combinatorial background (green dashed curve), and a sigmoid for the partially reconstructed background (brown dashed curve). Bottom plot: Bin-by-bin pull distribution.

	Mass (GeV)	Yield	SS	p
Cut-based displaced	6.2712 ± 0.0006	4691 ± 99	52.08 ± 0.79	0.6655
XGB displaced SS	6.2707 ± 0.0006	4690 ± 133	61.38 ± 0.99	0.8988
XGB displaced SP	6.2763 ± 0.0013	11445 ± 789	67.23 ± 3.34	0.6629
Cut-based inclusive	6.2700 ± 0.0011	5313 ± 217	27.74 ± 1.05	0.1962
XGB inclusive SS	6.2711 ± 0.0008	5315 ± 152	38.90 ± 0.96	0.3647
XGB inclusive SP	6.2710 ± 0.0008	10465 ± 307	38.94 ± 1.06	0.1962

Table 3.7: Summary of mass, signal yield, signal significance, and purity for different approaches.

Chapter 4

Optimisation of the XGBoost selection

The usage of Machine Learning algorithms such as the Boosted Decision Trees (in this case XGBoost) allows to easily design a selection for a physics analysis, but this comes with the price of loosing the direct control on the selection: an effort must always be dedicated to the understanding of the handles available to correctly set up the ML tool. As already discussed in Chapter 3 and Appendix A, the physics observables provided as input to XGB have been carefully chosen in order to extract most of the physics information. On the other hand the studies previously discussed in Chapter 3 are based on a case where the settings of the XGB model itself, the so-called *hyperparameters*, where arbitrarily chosen by looking for a working configuration with a trial-and-error approach (see Table 3.6).

An automatized optimization of the XGB hyperparameter can be achieved thanks to the Optuna library: Optuna [75] is a modern and flexible hyperparameter optimization framework that uses advanced algorithms to efficiently search for the best hyperparameters [76] [77]. Unlike traditional methods, Optuna offers a smart and adaptive approach to hyperparameter tuning, making it an ideal choice for complex machine learning models.

Optuna operates by defining an objective function, which can be either maximized or minimized depending on the user's requirements. In this work, the metrics aimed to maximize are referred to as Optuna Figures of Merit (Optuna FoM). Each iteration of the optimization is called "trial" and each trial corresponds to a different set of hyperparameters. Optuna evaluates the performance of each set by computing the objective function on the training sample and then uses the result to guide the search in future trials with a partially gradient-based algorithm. This iterative approach allows Optuna to learn from previous trials and adaptively focus on the most promising region(s)

of the hyperparameter space, making the search process both efficient and effective. The Optuna optimization is stopped when a pre-set number of trials is reached and it returns the hyperparameter set that achieves the best result according to the specified objective function.

In this study, the XGBoost algorithm was used to extract a sample of $B_c^+ \to J/\psi\pi^+$ decays. In order to improve the performance of XGBoost, a proper adjustment of the hyperparameters is of crucial importance. For this purpose the available data samples used are the same described in Section 3.2: the minimal selection called preprocess (see Table 3.3) is applied in order to homogenize the real data and MC samples, that is removing differences due to the filtering present in the MC generation process.

The investigation on how to choose an optimal working point is also discussed in the current chapter, for both the *displaced* and *inclusive* approaches (Secc. 4.4 and 4.5) similarly to the previous Chapter.

4.1 Optuna's Timing Performance

In the first part of this study (discussed in Chapter 3), hyperparameter optimization was not implemented for the XGBoost algorithm. A common method for hyperparameter optimization is Grid Search, available in the scikit-learn library. This approach systematically evaluates all the possible combinations of hyperparameter values within a predefined set of values or intervals. However, Grid Search's exhaustive nature can be computationally expensive, particularly with large hyperparameter spaces, as it does not intelligently prioritize potentially better-performing combinations. In this study, we compared the performance of scikit-learn's Grid Search with two alternative methods available in the Optuna framework: Optuna's Grid Search (replicating the traditional approach) and Optuna's Tree-structured Parzen Estimator (TPE) [78]. TPE is a Bayesian optimization algorithm designed to efficiently explore the hyperparameter space by learning from previous evaluations and focusing on promising regions. The hyperparameters at each iteration (trial) are updated partly driven by the Optuna FoM gradient and partly randomly, in order to avoid local minima.

To assess the effectiveness of these methods, an experimental setup was designed using the entire MC signal and 80% of the data sidebands after the preprocess selection to train an XGBoost inclusive model (considering a further 70/30 split for training and test purposes). The considered Optuna FoM is the AUC on the test sample, while no particular choice of the Working Point is considered, since it is not relevant for the timing comparison.

For the Grid Search methods, specific values for each of the max_depth, learning_rate, and gamma hyperparameters were defined, while for TPE the search space was defined as a range of values for each hyperparameter as reported in Table 4.1. In all cases, the num_boost_round (called n_estimators in the scikit-learn library) was kept fixed at 500. The results of this comparative analysis, focusing on the time taken to maximize the AUC metric across 24, 42, 63, and 240 trials, are presented in Figure 4.1.

Number of Trials	learning_rate	gamma	max_depth
24	0.07, 0.08, 0.09, 0.1, 0.2, 0.3	0, 1	2, 3
42	0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3	0, 1	2, 3, 4
63	0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3	0, 1, 2	2, 3, 4
240	0.05, 0.06, 0.07, 0.08, 0.09, 0.15, 0.25, 0.35, 0.4, 0.45	0, 1, 2, 3, 4, 5	2, 3, 4, 5

Table 4.1: Hyperparameter values tested in the comparative analysis of scikit-learn's Grid Search, Optuna's Grid Search, and Optuna's TPE sampler. For Grid Search methods, specific values were evaluated, while for TPE, an analogous (continuous) range of values was defined for each hyperparameter. The number of trials indicates the total number of hyperparameter combinations or evaluations performed by each optimization method.

In this analysis, all three optimization methods achieved nearly identical AUC scores (considering an uncertainty of order of 0.01%, better discussed later in Sec. 4.2) for each trial count, highlighting that the primary difference lies in their computational efficiency. To quantify this, the *speed-up* is defined as the ratio of the time taken by a baseline method (in this case, scikit-learn's Grid Search) to the time taken by an alternative method.

As depicted in Figure 4.1, Optuna's TPE sampler consistently outperforms both Grid Search methods in terms of speed-up. For 24 trials, TPE is 1.12 times faster than scikit-learn's Grid Search, while for 42 trials, this speed-up increases to 1.38x. The efficiency gain becomes even more pronounced with 63 trials, where TPE achieves a 1.56x speed-up. For the most extensive search space of 240 trials, TPE demonstrates a remarkable 1.58x speed-up compared to scikit-learn's Grid Search.

These results underscore the superior efficiency of Optuna's TPE sampler in navigating the hyperparameter space, enabling faster model development and iteration without sacrificing performance. This efficiency gain is particularly valuable in scenarios with larger datasets or more complex models, where the

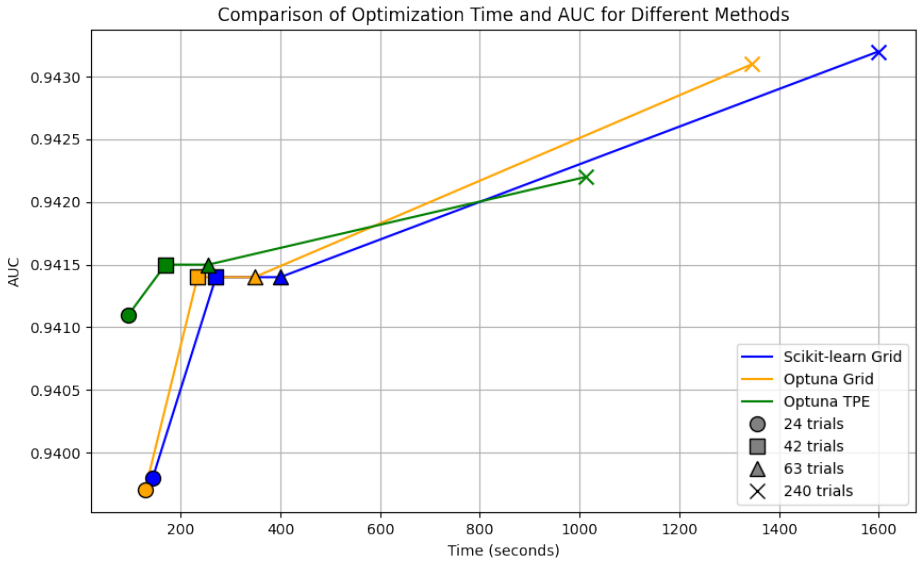


Figure 4.1: Scatter plot of the computational time to execute the task and resulting AUC scores for `scikit-learn`'s Grid Search (blue), `Optuna`'s Grid Search (orange), and `Optuna`'s TPE sampler (green) across different numbers of trials. Different markers represent the number of trials: circles for 24 trials, squares for 42 trials, triangles for 63 trials, and crosses for 240 trials. `Optuna`'s TPE sampler consistently outperforms both Grid Search methods in terms of computational time, achieving comparable or even slightly better AUC scores in all scenarios. Notably, for 24 trials, `Optuna`'s TPE yields the highest AUC, while for the remaining trial counts, both Grid Search methods achieve similar AUC values (considering an uncertainty of order of 0.01%), with `Optuna`'s Grid Search being slightly faster than scikit-learn's implementation.

computational cost of hyperparameter optimization using traditional methods like Grid Search can become prohibitive. Considering these results, `Optuna`'s TPE sampler was chosen as the preferred hyperparameter optimization method for the study mentioned in this chapter because of its superior efficiency and comparable performance to other methods.

4.2 AUC estimate and its uncertainty

As shown in the preliminary study performed with `Optuna` in Section 4.1, the algorithm evaluates the performance of each trial based on the objective function. Since AUC is a key component of this objective function, accurately estimating its uncertainty is crucial for controlling the optimization process and ensuring the reliability of the algorithm's decisions.

To quantify this uncertainty, a simple `XGBoost` set-up was used separately for both the `displaced` and `inclusive` approaches, using samples that had

undergone the `preprocess` selection. Five independent `XGBoost` trainings were conducted for each approach. The ROC curves obtained from these trainings, as shown in Figures 4.2 and 4.3, demonstrate that the AUC values were consistently close to one another in both approaches.

The Root Mean Square Error (RMSE) of the AUC values was calculated to determine the appropriate level of precision, thereby providing an estimate of the typical uncertainty in the AUC with the current sample size (see Table 4.2).

4.2.1 AUC uncertainty in the displaced set-up

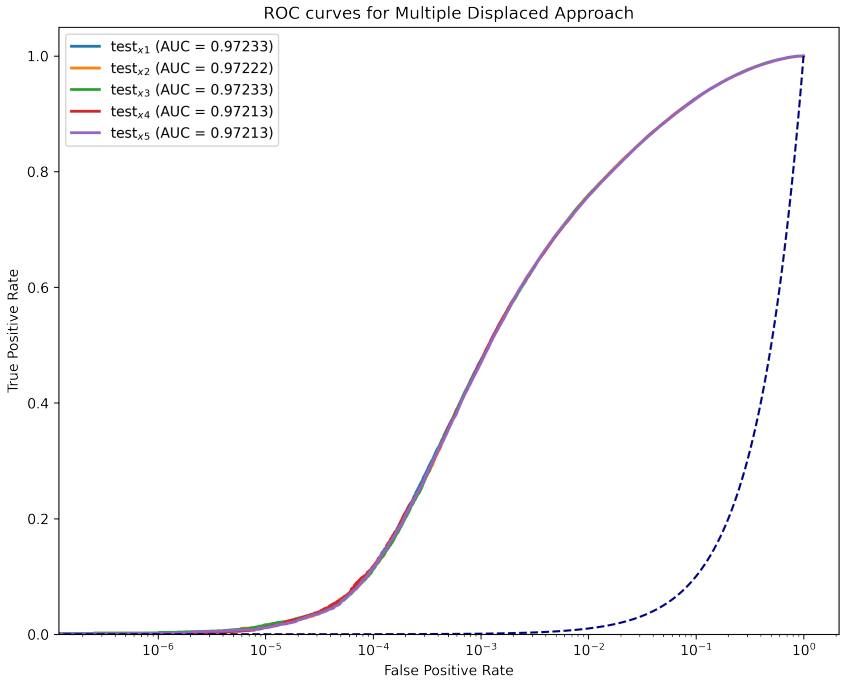


Figure 4.2: ROC curves for five repeated `XGBoost` trainings in the `displaced` approach, showing very close ROC curves and consequently AUC values (indicated in the legend).

The ROC curves in Figure 4.2 illustrate the performance of five repeated `XGBoost` trainings for the `displaced` approach. Each curve represents a different training run, and the AUC values are noted in the legend.

Table 4.2 displays the AUC values, their deviations from the average, and the calculated RMSE. The AUC values range from 0.97213 to 0.97233, with an average AUC of 0.97223. This consistency in AUC values suggests stable model performance across different runs. The RMSE of 0.00009 indicates that meaningful deviations in the AUC values start to appear at the fourth/fifth decimal place. Consequently, the AUC values are reported with precision up to at least the fifth decimal place to capture these small variations.

Names	AUC	Dev. from avg.
test_{x1}	0.97232	0.00009
test_{x2}	0.97222	0.00001
test_{x3}	0.97233	0.00010
test_{x4}	0.97213	0.00010
test_{x5}	0.97213	0.00010
Average	0.97223	-
RMSE	0.00009	-

Table 4.2: AUC values and their deviations from the average for the `displaced` approach after five repeated `XGBoost` trainings. The RMSE value indicates that the fourth/fifth decimal digit is the last significant digit.

4.2.2 AUC uncertanity in the inclusive set-up

In the `inclusive` approach, as shown in Figure 4.3, the AUC values are also very close to each other. The same methodology was applied to assess the uncertainty and determine the appropriate level of precision; five independent `XGBoost` trainings were conducted, and the results are summarized in Table 4.3.

The ROC curves in Figure 4.3 illustrate the performance of five repeated `XGBoost` trainings for the inclusive approach.

Names	AUC	Dev. from avg,
test_{x1}	0.93459	0.00018
test_{x2}	0.93464	0.00013
test_{x3}	0.93452	0.00025
test_{x4}	0.93490	0.00013
test_{x5}	0.93522	0.00045
Average	0.93477	-
RMSE	0.00026	-

Table 4.3: AUC values and their deviations from the average for the inclusive approach after five repeated `XGBoost` trainings. The RMSE value indicates that the fourth decimal digit is the last significant digit.

Table 4.3 shows the AUC values and their deviations from the average for five repeated `XGBoost` trainings in the inclusive approach. The AUC values range from 0.93459 to 0.93522, with an average of 0.93477. The RMSE value is calculated to be 0.00026. This indicates that the variability in AUC values is getting at the fourth decimal cypher only. Therefore, it is reasonable to report AUC values up to five decimal cypher, as this precision captures the significant variations in performance. The slight variations at the fourth decimal cypher

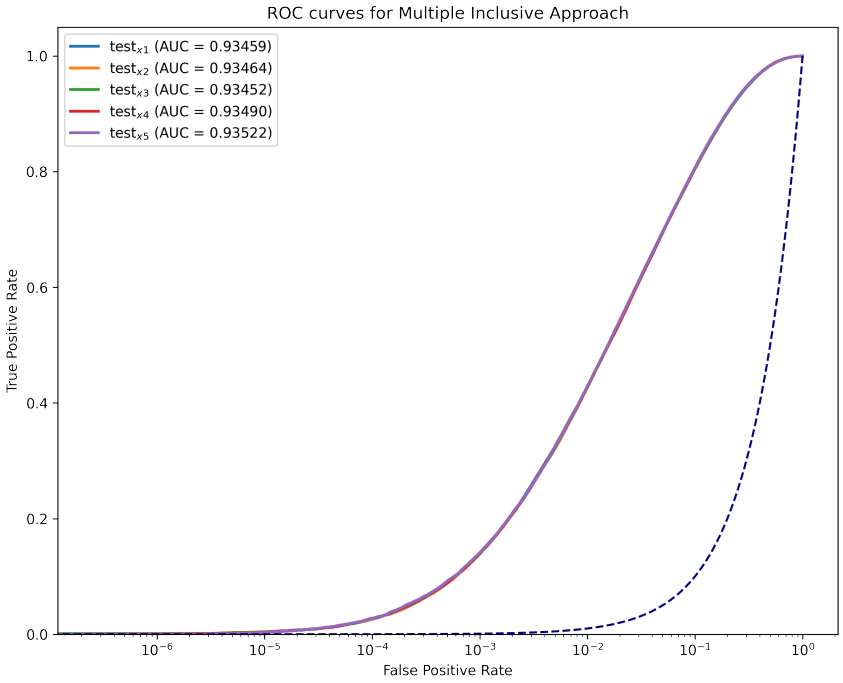


Figure 4.3: ROC curves for five repeated `XGBoost` trainings in the inclusive approach, showing very close AUC values (motivated in the legend).

reflect the model's fluctuations, which are considered to be overall negligible.

4.3 XGBoost optimization and signal extraction

4.3.1 The Optuna Figure of Merit

In hyperparameter optimization, various evaluation metrics can be used to assess the performance of machine learning models. Commonly used metrics include F-score (a measure combining precision, the proportion of true positive instances among those predicted as positive, and recall, the proportion of true positive instances among those actually positive), accuracy (the proportion of true results among the total cases), and AUC. These standard metrics provide a solid foundation for model evaluation.

Additionally, custom Figures of Merit (FoMs) can be defined to better suit specific application needs. In this study, the primary objective was to maximize the AUC; however, solely focusing on maximizing AUC can lead to overfitting. Therefore, an additional contribution was incorporated alongside AUC. After evaluating various methods, the p-value from the KS test p_{KS} (described in Sec. 3.1.2) was found to be the most suitable. The AUC is the primary factor `Optuna` attempts to maximize. To avoid an abrupt cut-off effect around the p_{KS} threshold (chosen to be 5%), the p-value is weighted using an error function centered on the threshold value. The specific FoM, referred to as the Optuna FoM, is calculated as follows:

$$\text{Optuna FoM} = w_{KS} \times \text{AUC} \tag{4.1}$$

where, the weight term w_{KS} is derived from the p-value obtained from the KS test and obtained using the error function:

$$w_{KS} = \frac{\text{erf}\left(\frac{p_{KS}-0.05}{0.025}\right) + 1}{2} \tag{4.2}$$

When the KS p-value is below the threshold of 0.05, w_{KS} is significantly lower (almost zero), penalizing the Optuna FoM and disfavouring the model. As the p-value is above the threshold of 0.05, w_{KS} quickly approaches 1, reducing its influence on the Optuna FoM: since the Optuna FoM becomes practically equivalent to the AUC. As a matter of fact the penalty term w_{KS} acts as a safeguard against overtraining. This dual use of AUC and p-value in the optimization equation (Eq. 4.1) helps to maintain a balance, thus ensuring the model to be both high-performing and robust, without suffering overtraining.

4.3.2 The Signal Extraction Figure of Merit

`Optuna` strives to find the best hyperparameters that maximize the Optuna FoM from Eq. 4.1, strongly favouring models that pass the KS test.

Once the best XGBoost model is identified, a BDT score threshold (or working point) must be selected to apply the model to the sample. Identifying the best threshold is critical for evaluating the model's performance.

For this purpose, a predefined list of working points is considered, mostly focusing on a strong background reduction. Within each trial, Optuna applies the model to the sample according to each working point in the list. A validation sample (20% of the real data) has been devoted to the study of the working point optimization.

For investigation purposes, all the models each coming from a trial are tested and applied on the validation sample. The signal and background in the signal region are estimated by means of an Unbinned Maximum Likelihood fit on the $J/\psi\pi$ invariant mass distribution (better discussed later in Sections 4.4 and 4.5, for the *displaced* and *inclusive* approaches respectively).

The choice of the working point is based on a further figure of merit, which in this work will be referred to as *Signal Extraction* (SE) FoM. The following four SE FoMs are investigated within this work:

- the Signal Significance, SS (Eq. 3.6),
- the purity, p (Eq. 3.7),
- the product of purity and Signal Significance, PxSS,
- the Punzi FoM [79] where the latter is defined as

$$\text{Punzi SE FoM} = \frac{S}{\frac{463}{13} + 4\sqrt{B} + 5\sqrt{25 + 8\sqrt{B} + 4B}} \tag{4.3}$$

where S and B are defined in a 3σ signal region around the B_c invariant mass peak value.

As it will be clearer later in this Chapter, the purity SE FoM tends to pick very high thresholds to obtain a strong background rejection. On the other hand, the Significance SE FoM tends to focus on low threshold values in order to retain more signal candidates. In the present study this choice is not very effective in rejecting the background, which would survive abundantly. A possible trade-off is given by the PxS FoM, which produces adequate but sub-optimal results. Consequently, the Punzi SE FoM resulted to be the most suitable metric.

In this thesis all the four SE FoMs have undergone a thorough study in order to assess a reliable SE FoM choice. The results can be also compared to the studies in Chapter 3, where the working point choice was guided by the comparison with the cut-based approach in order to enhance either the purity or the signal yield (see Figures 3.5 and 3.6). Different WP intervals

were defined for the inclusive and displaced approaches, and these intervals were scanned.

Following the optimization of purity and signal yield values, the workflow for the hyperparameter tuning process and the subsequent validation steps is discussed in the next subsection. This workflow encompasses data preparation, model training, threshold selection, and the iterative optimization process using Optuna.

4.3.3 Optimization workflow

The entire signal MonteCarlo dataset is utilized for the XGBoost training. To avoid any bias, instead, the real data are randomly shuffled and then split into an 80% training set and a 20% validation set. The validation sample is set aside for later use in the threshold selection process. A control region, consisting solely of background events, is created by identifying the $J/\psi\pi$ invariant mass candidates to in 80% training data, being guided by a study on the available MonteCarlo sample of $B_c \to J/\psi\pi$ candidates. The signal and background samples are then combined, shuffled again, and split into 70% for training and 30% for testing, forming the final datasets for XGBoost model training. Optuna initializes this process with a random set of hyperparameters. This process up to the completion of XGBoost training is referred to as **STEP 1** and is illustrated in Figure 4.4.

The process of selecting the optimal BDT working point (threshold), referred to as **STEP 2** and illustrated in Figure 4.5, involves the application of the trained XGBoost model to the validation sample (set aside in the previous step). A predefined list of 20 thresholds is iteratively applied, ranging from 0.970 to 0.979 with a precision of one per mille (0.001), and from 0.980 to 0.999 with a precision of two per mille (0.002). For each threshold, the Punzi SE FoM is calculated. The threshold that yields the highest Punzi SE FoM is designated as the "Best WP", marking the conclusion of the threshold selection loop. To calculate the Punzi SE FoM for each threshold, an UML fit of the $J/\psi\pi$ invariant mass is performed, yielding the values of S and B, which are used to compute the Punzi SE FoM. Consequently, for each trial, 20 fits are conducted and SE FoM values are computed. Although this process is computationally intensive and time-consuming, it is advantageous compared to alternative methods as discussed earlier (see Figure 4.1). Understanding the behavior of SE FoMs in the optimization process is critical despite the computational load.

Upon completion of **STEP 2**, the Optuna FoM is calculated based on the best Punzi SE FoM achieved during the threshold selection. This Optuna FoM guides the subsequent hyperparameter optimization process. XGBoost

then retrains the model with the new hyperparameters provided by Optuna, effectively returning to **STEP 1** and initiating a new iteration of the threshold selection loop **STEP 2**. This process continues iteratively (**LOOP|STEP 3**, with each trial generating a single Optuna FoM based on the best Punzi SE FoM achieved during the threshold selection. Figure 4.6 presents a flowchart illustrating **LOOP** and the the steps involved in this iterative optimization process.

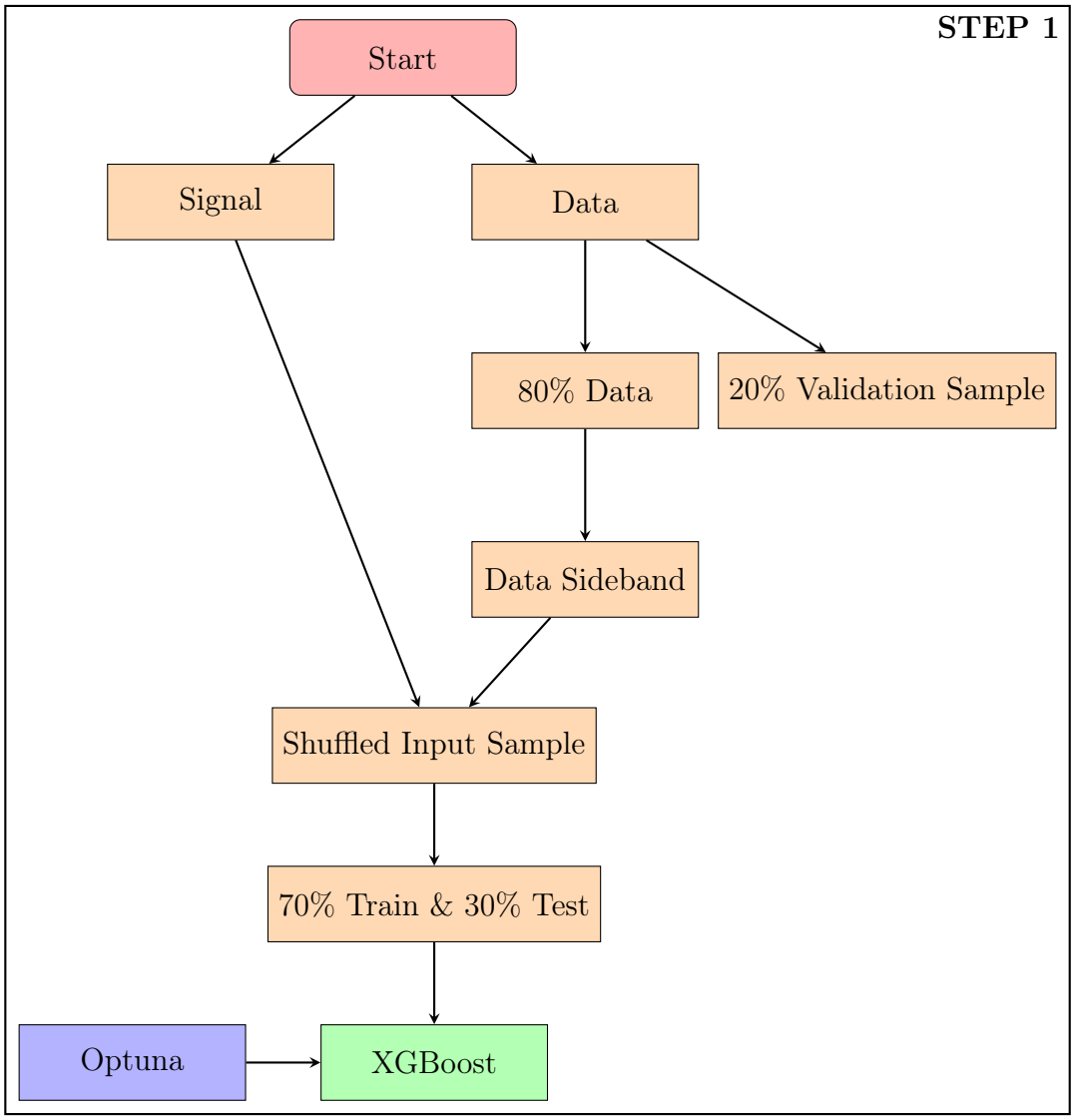


Figure 4.4: Flowchart illustrating the data preparation process for XGBoost training, which is then optimized by Optuna.

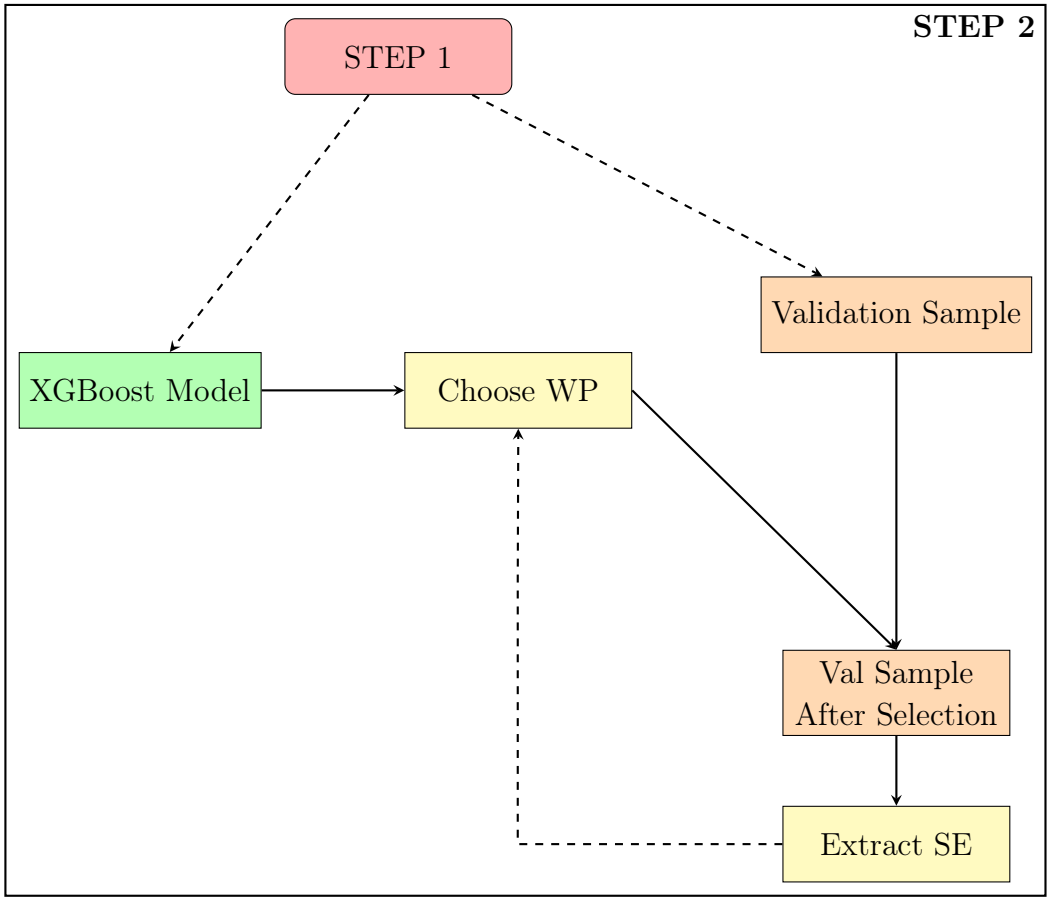


Figure 4.5: Flowchart illustrating the iterative process of applying different WPs (thresholds) to the `XGBoost` model and calculating the Punzi SE FoM for each WP on the validation sample.

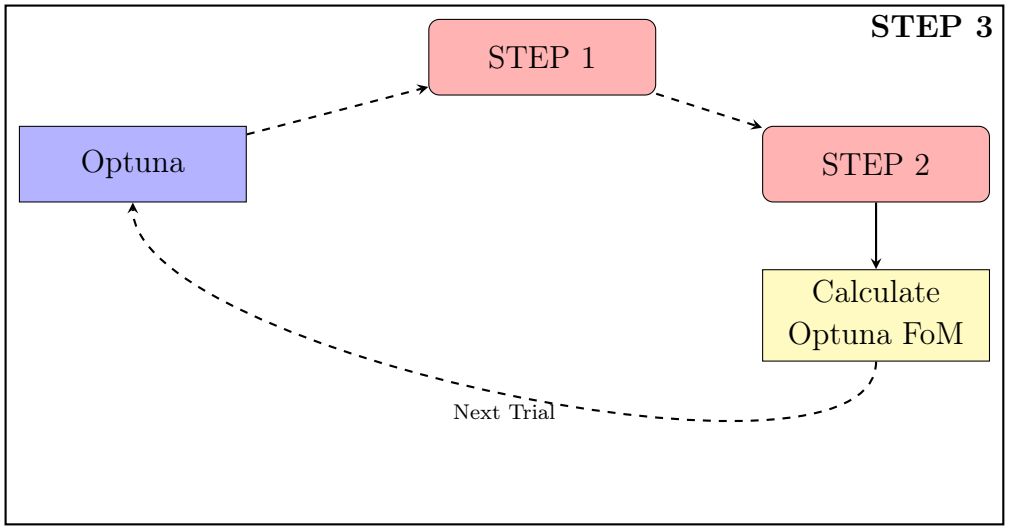


Figure 4.6: Flowchart illustrating the iterative process of calculating the Optuna FoM and the subsequent determination of the next set of hyperparameters by Optuna (**LOOP**.

4.4 Hyperparameter Optimisation for Displaced Approach

4.4.1 BDT optimization with Optuna

In order to deal with the Displaced Approach, the variables listed in Table 3.5 are employed for the purpose of hyperparameter optimization. The following three hyperparameters were optimized using Optuna, as shown in Table 4.4 which reports the inverstigated range of their values.

Hyperparameter	Range
`learning_rate`	[0.01, 0.5]
`gamma`	[0, 20]
`max_depth`	[2, 6]

Table 4.4: Hyperparameters and their ranges chosen for the `Optuna` displaced approach.

Each `XGBoost` BDT is composed of 500 decisional nodes (`num_boost_round`), and `Optuna` iteratively creates a new BDT (with progressively learning hyperparameters) 250 times (`n_trials`). Due to the preprocess selection and the subsequent 80%-20% data split, the `scale_pos_weight` parameter was adjusted to 20.977, differing from the value of 6.636 reported in Table 3.6. These parameters were chosen to ensure a thorough exploration of the hyperparameter space and a robust model training. Based on preliminary results, 500 decisional nodes were considered sufficient for an effective training: for example in the results shown in Figure 3.7 the log loss curve stabilized at 500 boosting rounds, indicating no further significant improvement. The number of trials was limited to 250; as can be appreciated in Figure 4.1, this number is adequate for achieving optimal hyperparameter configurations without excessive computational expense.

`Optuna` provides several visualization tools to assess its performance. The *optimization history* plot was used to observe the overall performance of the model (Figure 4.7). It illustrates the performance of `Optuna`'s TPE sampler in navigating the hyperparameter space: the x-axis represents the trial number (which is progressive), while the y-axis indicates the Optuna FoM (Eq. 4.1). The blue dots depict the Optuna FoM for each trial, and the red line traces the best Optuna FoM achieved up to that point. The plot reveals a rapid convergence of the optimization process, with the Optuna FoM stabilizing after a few initial trials. This early stabilization indicates that `Optuna`'s TPE sampler efficiently identifies promising regions within the hyperparameter space, leading to the discovery of an optimal or near-optimal set of hyperparameters early in the optimization process.

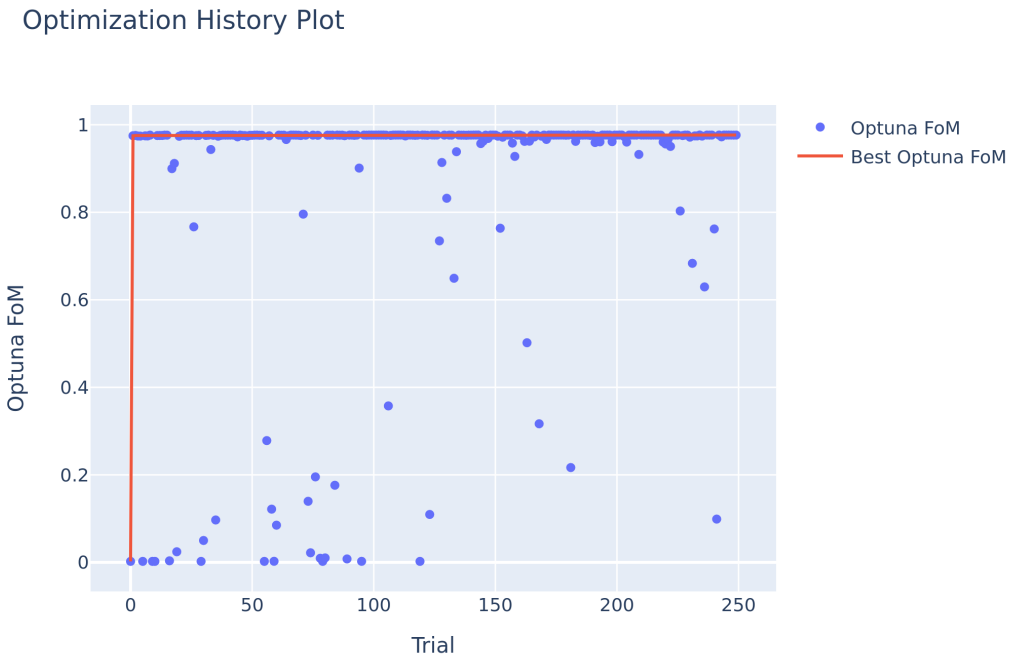


Figure 4.7: Optimization history plot for the Displaced Approach using `Optuna`'s TPE sampler. The blue dots represent the Optuna FoM achieved for each trial, while the red line indicates the best FoM obtained up to that point. The rapid convergence of the Optuna FoM after the initial trials demonstrates the effectiveness of the TPE algorithm in efficiently identifying optimal hyperparameter values.

Moreover, due to the need for more detailed results regarding the p-value of the KS test (p_{KS}) in the Optuna FoM, an additional plot was prepared. Figure 4.8 shows a scatter plot of the AUC values obtained from the BDT ROC curve(x-axis) and the p-value for the KS test (y-axis), namely the two factors of the chosen Optuna FoM (see Eq. 4.1). Each marker on the plot corresponds to different trials, with the `max_depth` values indicated by the shape of the markers. The colors of these markers represent the learning rate, based on the right color scale. The dashed red line at $p_{KS} = 0.05$ serves as the turn-on threshold; all points above this line have passed the KS test, indicating successful trials, while those below have failed due to overtraining. The trial with the highest Optuna FoM is highlighted within a red circle, recommended by `Optuna` as the optimal trial.

To further analyze the performance and characteristics of the optimization process, additional visualizations provided by `Optuna` were inspected. These visualizations help to understand the relationships between hyperparameters and their impact. The Parallel Coordinate Plot in Figure 4.9 illustrates the relationship between different hyperparameters and the Optuna FoM. Each line in

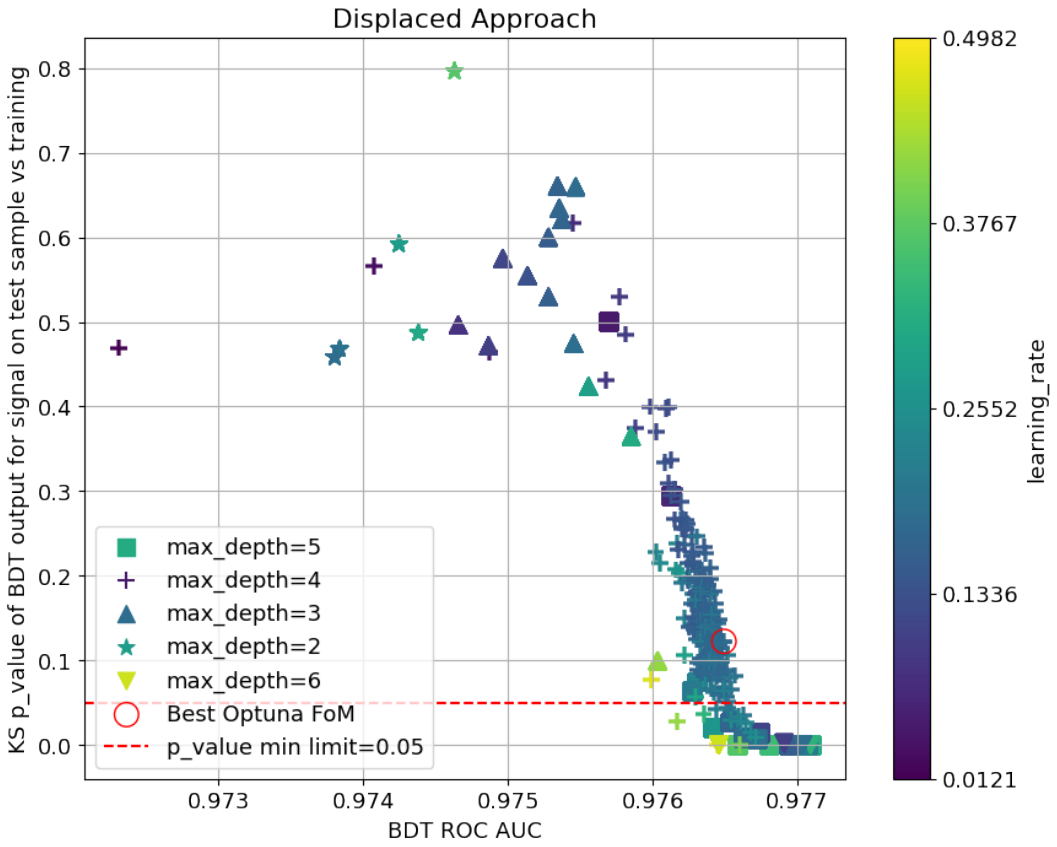


Figure 4.8: Model performance visualization for the Displaced Approach. The x-axis represents the AUC values obtained from the BDT, and the y-axis represents the p-value of the KS test on the signal train/test samples (p_{KS}). Each marker corresponds to different trials with varying `max_depth` values and learning rates, the latter indicated by the color scale. The red dashed line at $p_{KS} = 0.05$ marks the KS test threshold. The red-circled point indicating the best Optuna FoM.

the plot represents a trial, with the color intensity indicating the Optuna FoM value. This customized visualization provides insights into the interactions and dependencies between hyperparameters, enabling the identification of optimal regions within the hyperparameter space. From the plot, it can be observed that lower values of `learning_rate` (in the range between 0.15 to 0.2) tend to correspond to higher Optuna FoM values. Additionally, `max_depth` values clustered around 4 show a clear trend towards higher Optuna FoM, suggesting that deeper trees than the one tentatively considered initially (`max_depth` 2 or 3) can contribute to improve the model performance. A further increase in the `max_depth` instead reasonably leads to overtraining, therefore penalised in the Optuna FoM by the p_{KS} factor (see Figure 4.8. These trends help to identify the most promising combinations of hyperparameters for optimizing the model, emphasizing regions that contribute to higher Optuna FoM val-

ues and, consequently, better overall performance. The evident green "band" present in Figure 4.9 corresponds to a region of hyperparameter space where the Optuna's TPE algorithm has focused since it was considered promising.

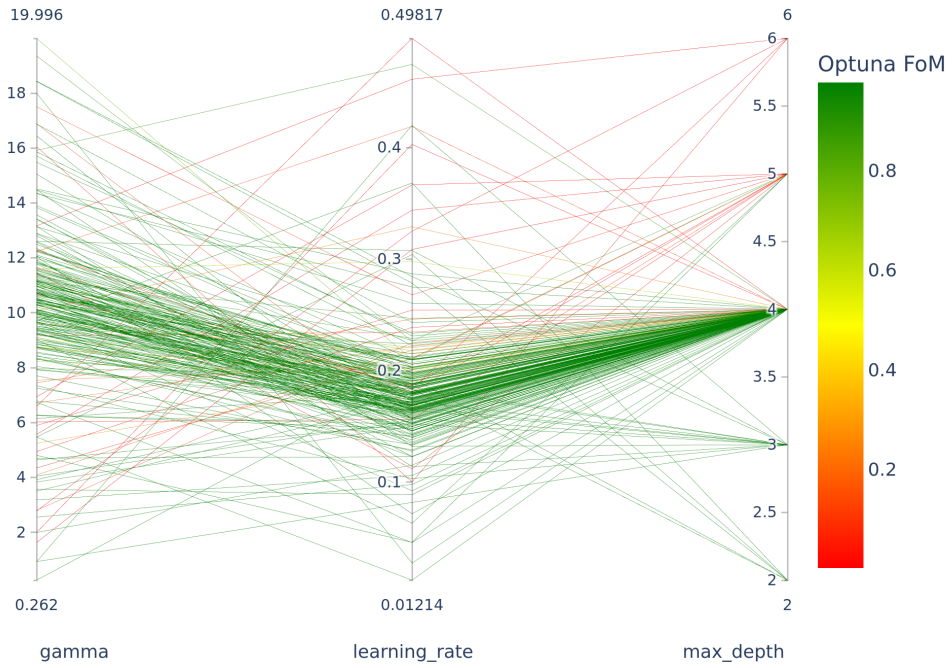


Figure 4.9: Parallel Coordinate Plot for the Displaced Approach showing the relationships between different hyperparameters and the Optuna FoM. Each line represents a trial, and the color intensity indicates the Optuna FoM value.

The Slice Plot (Figure 4.10) is a different visualization that displays the Optuna FoM against each hyperparameter individually, thus helping in understanding how changes in individual hyperparameters affect the Optuna FoM. The color intensity information allows to understand how the algorithm navigates through the hyperparameter space. It is possible to qualitatively note that:

- optimal value for `max_depth` (`max_depth` = 4) is reached earlier with respect to the other parameters, as there are very few tests with different values;

- `learning_rate` and `gamma` parameters, instead, converge less quickly to the final values: the former shows, in this plot a turn-off curve for values above 0.2, while the latter exhibits a less clear pattern, even though the convergence towards the range 8 - 12 is evident.



Figure 4.10: Slice Plot for the Displaced Approach showing the Optuna FoM against individual hyperparameters. Each dot represents a trial, with the color intensity indicating the trial number.

Figure 4.11 displays the Rank Plot, providing a comprehensive overview of the distribution of hyperparameters (including possible correlations between them) and their corresponding Optuna FoM values. In this plot, each dot represents a trial, colored according to its rank in terms of the Optuna FoM. The color gradient ranges from blue (lowest Optuna FoM) to red (highest Optuna FoM). This plot helps to identify the distribution of high-performing trials and the hyperparameter settings that contribute to these results. For instance, the plot shows that `gamma` values in the range between 5 and 15 tend to yield higher Optuna FoM when the `learning_rate` is already optimized (between 0.1 and 0.2). Similarly, `learning_rate` values around 0.1 to 0.2 are optimal only when associated to `max_depth` $= 4$. This visualization aids in understanding which hyperparameter ranges are more likely to yield optimal results.

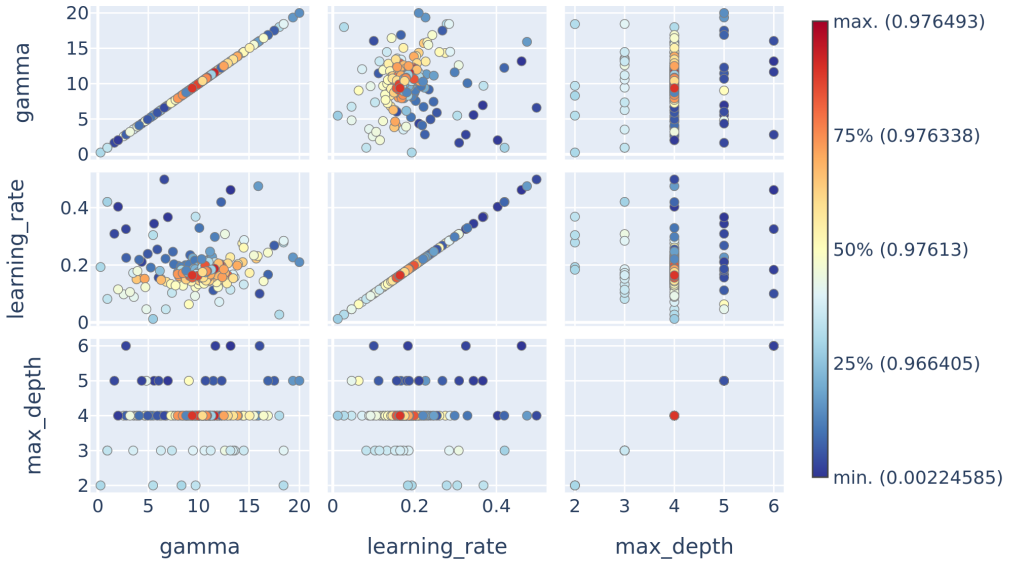


Figure 4.11: Rank Plot for the Displaced Approach showing the distribution of hyperparameters and their corresponding Optuna FoM values. Each dot represents a trial, colored according to its rank, with the gradient ranging from blue (lowest Optuna FoM) to red (highest Optuna FoM). The z-axis scale is expressed in quantiles and the corresponding thresholds are displayed in the legend.

4.4.2 Signal extraction optimization

As the study in Figure 4.7 shows, most of the top results are very close to each other, with differences typically appearing at the fourth decimal place, namely consistent with typical uncertainty on the AUC on the used data sample, previously shown in Table 4.2. Moreover Figure 4.7 shows that the progress from the TPE algorithm is not linear, since there is a random contribution to the TPE learning gradient which prevents it from focusing on local minima, and therefore the best model, in general, is not the one from the last trial.

Indeed it is possible to pick the model with the highest Optuna FoM but, since several models are not significantly different from the best according to the Optuna FoM (called `Best Optuna FoM` later in text), the selection is instead made between the top 25 models (sorted by the Optuna FoM) and then use an additional criterium to pick it. The first 25 models are chosen since they are characterized by an Optuna FoM compatible with the `Best Optuna FoM` within the uncertainty.

The additional criterium involved in the selection is the Signal Extraction

Figure of Merit (SE FoM). Within this work, the latter plays two roles: firstly for each model a WP is selected by optimizing a given SE FoM, secondly the top 25 models (equivalent according to the Optuna FoM) are sorted by their SE FoM corresponding to their optimized WP. The chosen SE FoM is the Punzi Figure of Merit (Eq. 4.3), but other FoMs have been evaluated for comparative purposes: purity p, Signal Significance SS, and their product PxSS.

The calculation of these SE FoMs requires accurate values for the signal yield (S) and the background yield (B), which are derived from an UML fit on the invariant mass distribution of the $J/\psi\pi$ from a validation sample, composed of 20% of the real data, which were preliminarily separated from the training sample. This distribution is fitted with a simple fit function where the signal component is modeled by a Gaussian function and the background is modeled using a 2^{nd}-degree Chebyshev polynomial. The Gaussian model was chosen for its stability and robustness within the automated optimization loop, unlike the Johnson p.d.f., which in some cases tended to accommodate part of the background in its tails. An example of a fit result obtained from this process, for the displaced approach, can be seen in Figure 4.12: in this case the chosen WP threshold (0.996) is selected in order to optimize the Punzi FoM.

For this fit, the Gaussian σ is fixed to 31 MeV, based on the mass resolution obtained from a fit to the signal MC sample (see Fig. 3.2), in order to ensure consistency across different trials and thresholds, thus allowing for a reliable extraction of S and B values. Once the fit is performed for each threshold, the signal yield (S) and the background yield (B): are obtained;specifically, for the computation of the purity, S and B are estimated in a 2σ window around the B_c mass peak, while for the SS and the Punzi FoM within a 3σ.

In this study, it was found that the trial with the highest Optuna FoM also had the highest Punzi SE FoM, even though this can not be considered a general property. This is illustrated in the two visualizations provided in Figure 4.13. The first visualization (Figure 4.13a) shows a general overview that includes the results from all the trials that passed the KS test ($p_{KS} > 0.05$), while the second one (Figure 4.13b) provides a zoomed-in view of the top-performing 25 trials to allow for a clearer distinction among the trials. By examining the top 25 trials, it is clear that the best model was selected based on both the Optuna FoM and Punzi SE FoM. Namely highlighting a scenario where the same trial optimized both metrics. The hyperparameters for this optimal trial are reported in Table 4.5.

The performance measurement plots obtained from `Optuna` illustrates that these optimized hyperparameters consistently produced optimized results. The WP value that resulted from the highest Punzi SE FoM was corresponding to a threshold equal to 0.996. This approach ensured that the selected model

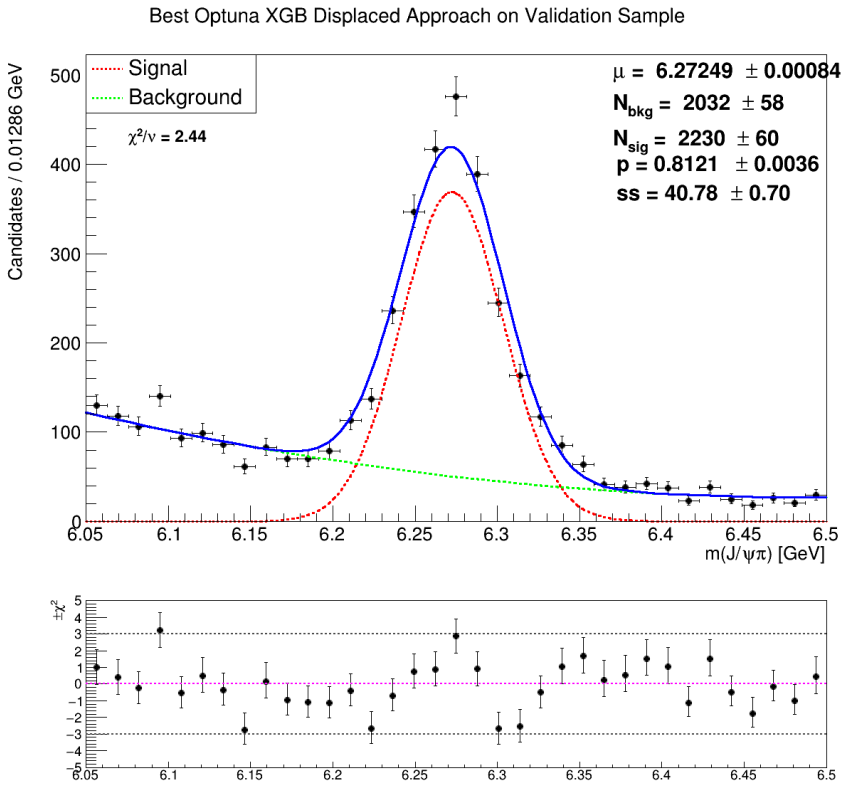


Figure 4.12: Invariant mass distribution for the validation sample using the best Optuna model for the displaced approach. The fit includes a Gaussian signal model and a 2nd-degree Chebyshev polynomial for the background.

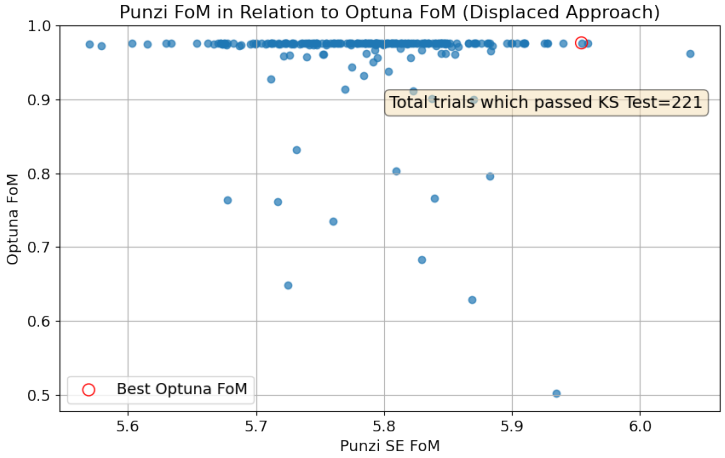
was not only optimal according to the primary metric (Optuna FoM) but also robust and reliable according to the secondary metric (Punzi SE FoM).

Model	`max_depth`	`gamma`	`learning_rate`
Unoptimized XGBoost Displaced Approach	2	0	0.05
`Optuna` Range	[2, 6]	[0, 20]	[0.01, 0.5]
Best Optuna FoM	4	10.53	0.1743

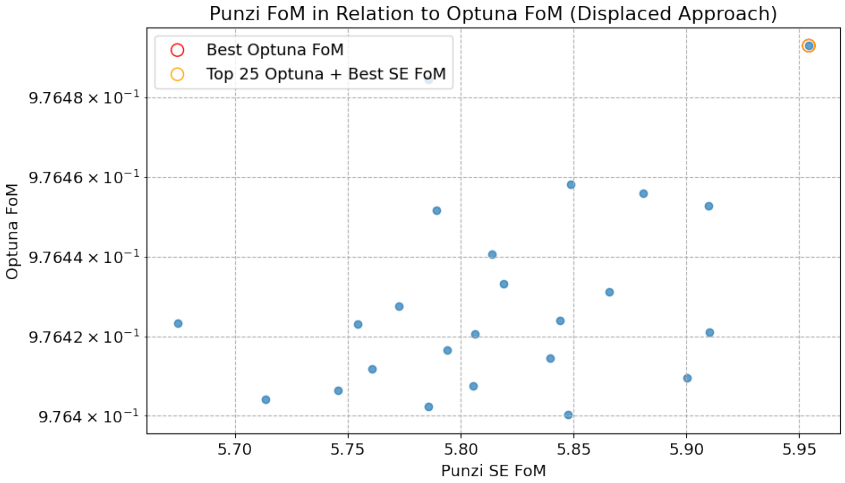
Table 4.5: Comparison of hyperparameters for Unoptimized initial `XGBoost` Displaced Approach, `Optuna` Range allowed in the optimization, and Best Optuna FoM.

4.4.3 Comparison between different SE FoMs

The Figure 4.14 illustrates the comparison of four different SE FoMs for various threshold on the best model, obtained from considering both the Optuna FoM and the Punzi SE FoM. The SE FoMs evaluated are: Signal Sig-



(a)



(b)

Figure 4.13: Comparison of Punzi SE FoM and Optuna FoM. (a) General overview of Punzi SE FoM in relation to Optuna FoM, and (b) Zoomed-in view of the top-performing trials using a log scale, highlighting how different trials rank when sorted by Punzi SE FoM after achieving similar Optuna FoM values. The red circle indicates the trial with the best Optuna FoM and Punzi SE FoM.

nificance (SS), Purity (p), PxSS, and Punzi. Each SE FoM is plotted against the threshold values (representing possible WPs) to visualize their performance and behavior. In this study, the focus is primarily on the Punzi SE FoM, as it provides a balanced approach for maximizing signal extraction while minimizing background interference. The other SE FoMs were included for comparison purposes to understand their behavior.

The scanned threshold score range stops at 0.998: tighter thresholds would result in a very small sample where the fit could not be performed success-

fully. In the developed algorithm the fit is not attempted on samples with less than 100 candidates after the selection, in order to have only reliable signal extraction calculations from the fit results.

The *Displaced* Approach effectively rejects the background and while keepig a good signal yield, which is evident for high WP threshold values (close to 1). Typically, when maximizing the purity, extremely high WPs are generally obtained. The overall distribution of WPs where other SE FoMs reach their highest values is presented in Figure 4.15: this plots shows the results obtained from all the models passing the KS test, in order to increase the available statistics. By comparing these SE FoMs, the study demonstrates that on one side the purity criterium results in a too tight selection which rejects also a large amount of signal, while the WPs obtained from the *SS* criterium show a broad distribution at lower values, corresponding to selected samples with a still large background contamination. The product PxSS and the Punzi FoM distributions instead are relatively narrow and therefore considered stable against minor changes in the underlying BDT model. Since the Punzi distribution is slightly narrower than the PxSS distribution, it is considered as the best from the physics point of view with in the Displaced Approach.

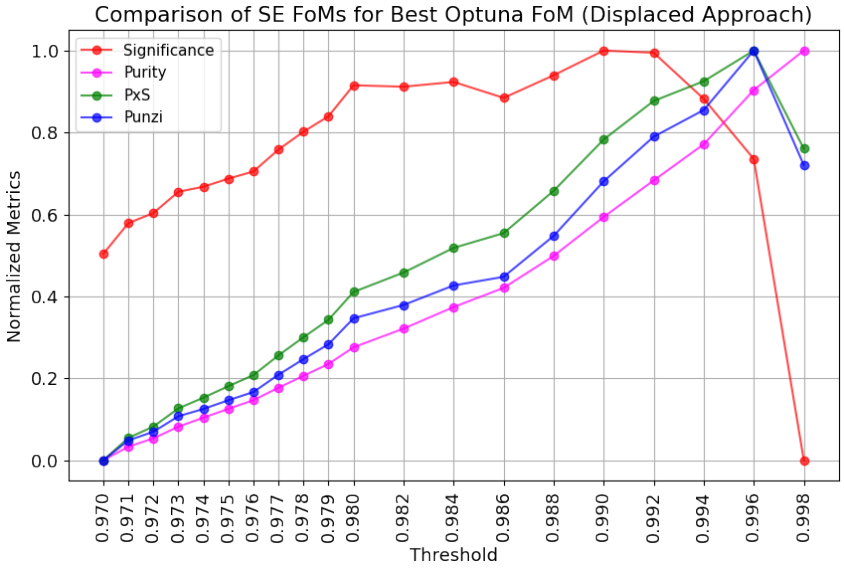


Figure 4.14: Comparison of SE FoMs: Significance, Purity, PxS, and Punzi for the Displaced Approach. The plot shows their performance against working points (thresholds), highlighting the Punzi SE FoM's effectiveness and stability in the optimization process. The general distribution of WPs where other SE FoMs reach their highest values is presented in Figure 4.15.

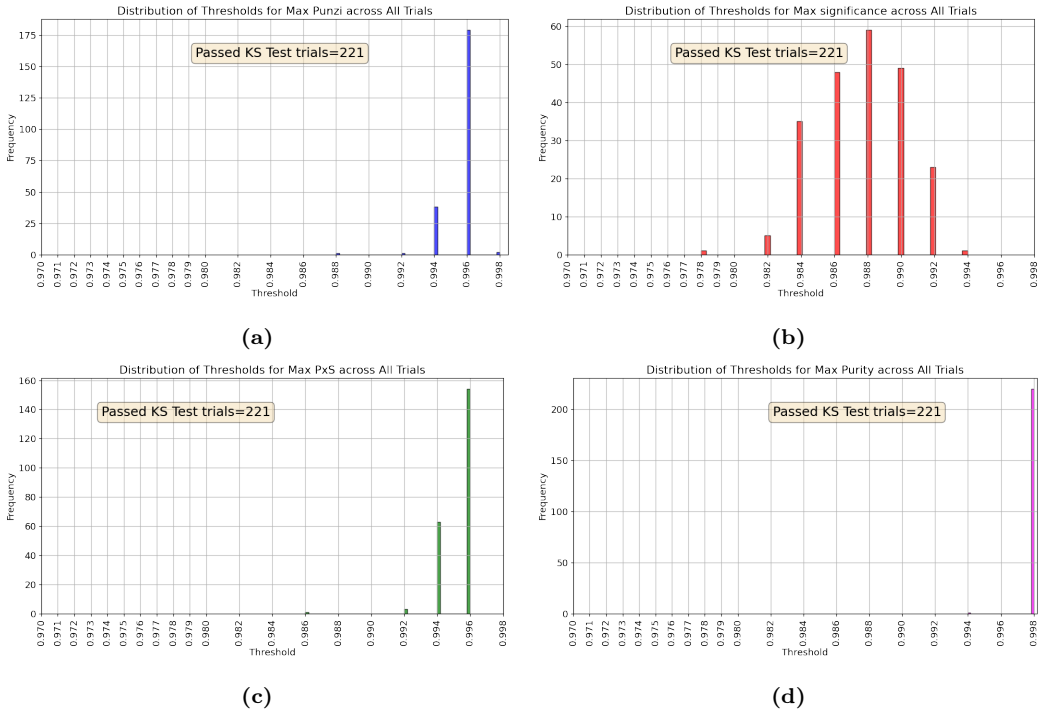


Figure 4.15: Comparison of different SE FoMs: (a) Punzi SE FoM, (b) Significance SE FoM, (c) PxS SE FoM, and (d) Purity SE FoM. For each trial, the maximum value obtained for each SE FoM is shown. The distribution of these maximum values across all trials highlights the WP at which each SE FoM is most frequently maximized. While the primary focus is on Punzi SE FoM, this comparative analysis provides insight into the behavior of other SE FoMs over the trial's population.

4.4.4 Results from the optimized model

The best model, as indicated by the highest Optuna FoM and the corresponding highest Punzi SE FoM, demonstrates robust performance on the validation sample. The `XGBoost` score distribution (Figure 4.16) illustrates clear separation between signal and background, and the best the WP (at 0.996) is shown.

The ROC curve (Figure 4.17) further validates the model's efficacy, showing an AUC of 0.9765. The best WP is highlighted on the curve; it corresponds to a low contamination (≈ 0.001) and a good efficiency (≈ 0.440).

The invariant mass distribution for the B_c candidates selected with the Optuna-tuned model on the full data is displayed in Figure 4.18. The signal is modeled with a Johnson p.d.f., while the background includes combinatorial components and a partially reconstructed background modeled with a sigmoid function. The fit results provide a signal mass of $\mu = 6.2770 \pm 0.0012$ GeV and a resolution of $\sigma = 0.0314 \pm 0.0023$ GeV, and the signal yield is 13697 ± 771 events, with a purity of $p = 0.853 \pm 0.010$.

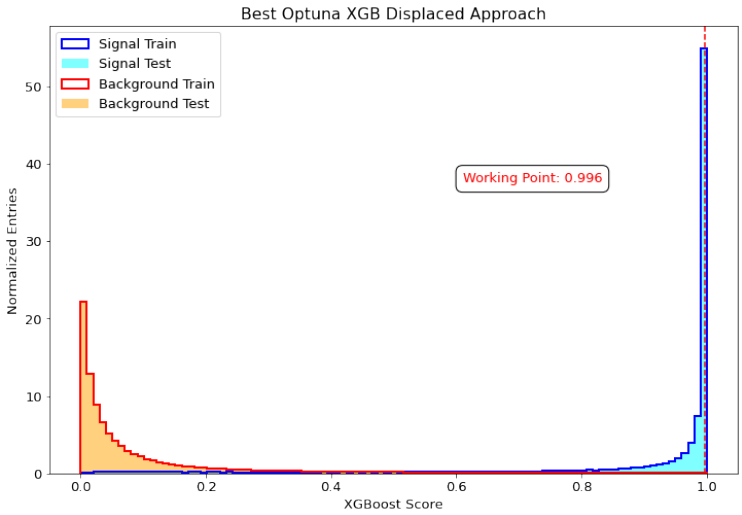


Figure 4.16: Score distribution for the best model obtained with `Optuna` for the Displaced Approach, using the validation sample. The WP is indicated as 0.996. The histogram shows the `XGBoost` scores for signal and background in both training and test samples.

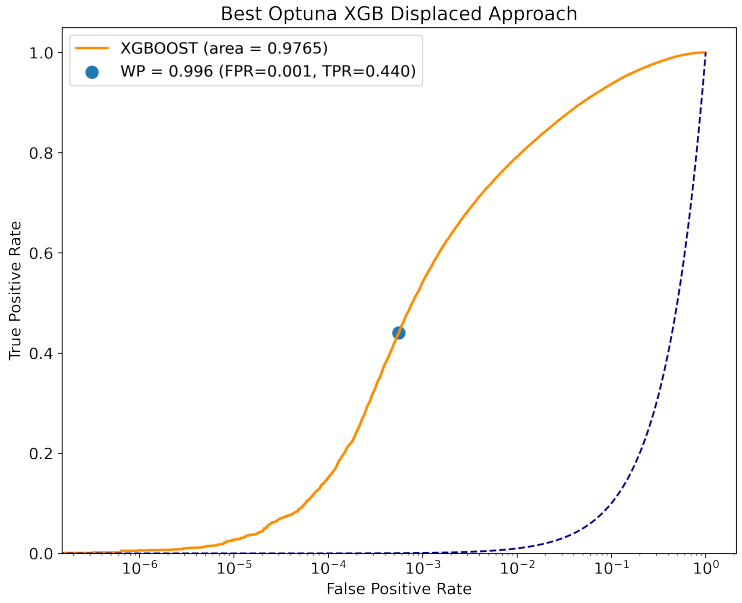


Figure 4.17: ROC curve for the best model obtained with `Optuna` for the displaced approach, evaluated on the validation sample. The AUC value is 0.9765, and the WP of 0.996 is highlighted, showing the corresponding FPR and TPR.

As a result, it turns put that applying the optimized model to the full dataset shows significant improvements as one could have expected. The signal yield is approximately 1.4 times the unoptimized reference selection, and the purity has increased from 0.6655 to 0.8527. As a cross-check the mass result from the optimized model is consistent with the PDG value $M_{PDG}(B_c) = 6.27447 \pm 0.00032$ GeV and previous CMS results.

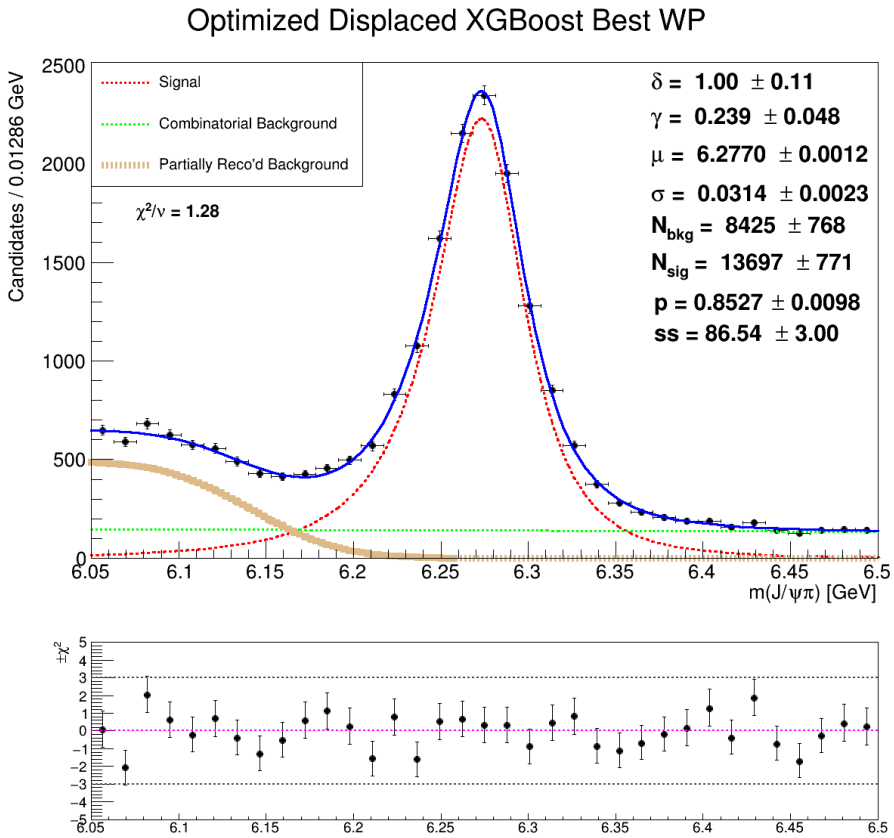


Figure 4.18: Distribution of the invariant mass of B_c candidates selected by the optimized displaced XGB model with the best WP applied to the full dataset. The result of the UML fit (blue curve) is superimposed. The fit components are Johnson p.d.f. for the signal (red dashed curve), a falling exponential for the combinatorial background (green dashed curve), and a sigmoid for the partially reconstructed background (brown dashed curve). Bottom plot: Bin-by-bin pull distribution, providing the pseudo-χ^2 for each bin.

	Unoptimized XGB Displaced Approach	**Optimized XGB Displaced Approach Best WP**
Signal Yield	9764 ± 340	13697 ± 771
Background Yield	11528 ± 342	8425 ± 768
Purity	0.6399 ± 0.0049	0.8527 ± 0.0098
Significance	61.84 ± 1.58	86.54 ± 3.00
χ^2/ν	1.45	1.28

Table 4.6: Comparison of different selections for the displaced approach: signal and background yields, purity (p), significance (ss), and χ^2/ν values.

The comparison between the unoptimized and optimized Displaced `XGBoost` models on the full dataset reveals significant improvements obtained with the

optimized model. For both models, a Johnson p.d.f. was used for the signal, an exponential function for the combinatorial background, and a sigmoid function for the partially reconstructed background.

The optimized model, tuned using Optuna, results in a higher signal yield of 13697 ± 771 compared to the unoptimized model (9764 ± 340). This increase in signal yield is accompanied by a decrease in background yield, from 11528 ± 342 in the unoptimized model to 8425 ± 768 in the optimized model.

As a consequence, the purity (p) and significance (SS) metrics show marked improvements. The optimized model achieves a purity of 0.8527 ± 0.0098, significantly higher than the unoptimized model's purity of 0.6399 ± 0.0049. Similarly, the significance improves from 61.84 ± 1.58 in the unoptimized model to 86.54 ± 3.00 in the optimized model. The χ^2/ν value also shows a better fit in the optimized model (1.28) compared to the unoptimized model (1.45), indicating a more accurate modeling of the data. This comparison, summarized in Table 4.6, quantifies the performance gain motivated by of the optimized model.

4.5 Hyperparameter Optimisation for Inclusive Approach

The same logic of the study discussed in the previous section is adopted also studying the optimization in the case of the Inclusive Approach.

4.5.1 BDT optimization with Optuna

In order to implement an Inclusive Approach, the variables listed in Table 3.5 were employed for the purpose of hyperparameter optimization. The hyperparameters that were optimized using Optuna, are reported in Table 4.7 together with their explored ranges.

Hyperparameter	Range
`learning_rate`	[0.01, 0.5]
`gamma`	[0, 10]
`max_depth`	[2, 4]

Table 4.7: Hyperparameters and their ranges of variability defined for the `Optuna` Inclusive Approach. The numbers within the specified ranges are the allowed values with the optimization process.

For `XGBoost`, `num_boost_round` was set to 500, and `n_trials` for `Optuna` was set to 250, exactly as for the `Displaced` Approach.

The *optimization history* plot was used to observe the overall performance of the model (Figure 4.19). The optimization history plot illustrates the resulting Optuna FoM of each model versus its trial number: the blue dots depict the Optuna FoM for each trial, and the red line traces the best Optuna FoM achieved up to that point. The plot reveals a rapid convergence of the optimization process, with the Optuna FoM stabilizing after a few initial trials. This early stabilization indicates that `Optuna`'s TPE sampler efficiently identifies promising regions within the hyperparameter space, leading to the discovery of a near-optimal set of hyperparameters early in the optimization process.

To further analyze the performance and characteristics of the optimization process, a visualization of the trials was generated. Figure 4.20 illustrates the model performance based on the p_{KS} and AUC of the BDT output. Each marker corresponds to different trials with varying `max_depth` values, and the colors represent the `learning_rate` based on the color scale on the right. The red dashed line at $p_{KS} = 0.05$ reflects the threshold in the weight w_{KS} in the Optuna FoM (Eq. 4.1), indicating that points above this line have passed the KS test. The trial highlighted within the red circle represents the best trial selected based on the procedure described in Section 4.4.2, namely considering

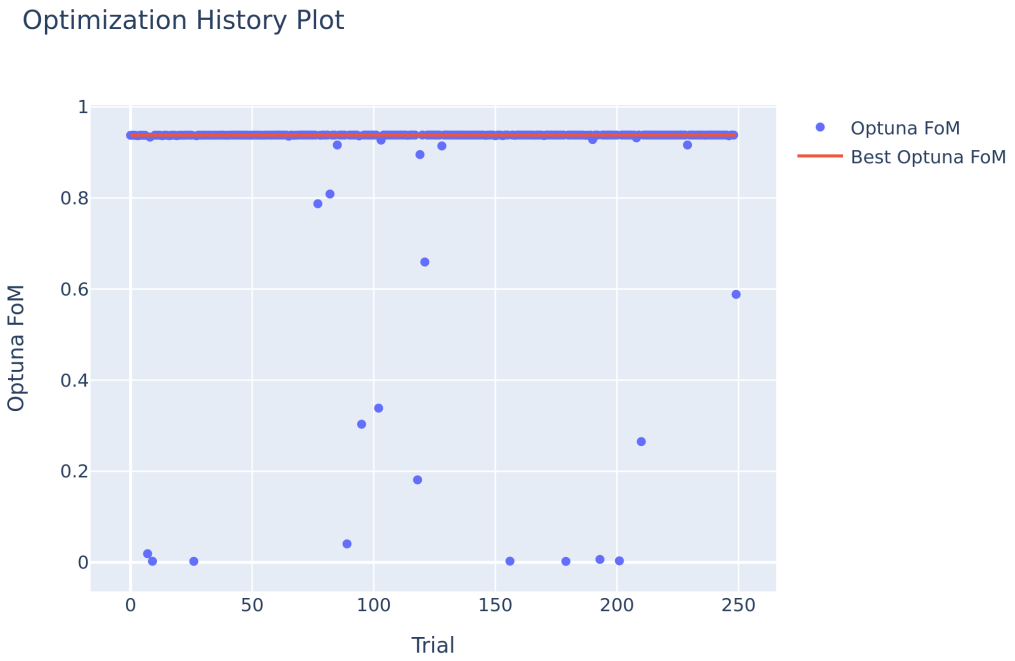


Figure 4.19: Optimization history plot for the Inclusive Approach using `Optuna`'s TPE sampler. The blue dots represent the Optuna FoM achieved for each trial, while the red line indicates the best SE FoM obtained up to that point.

the model with the best Punzi SE FoM amongst top-100 according to the Optuna FoM (details in the next Section 4.5.2).

The Parallel Coordinate Plot in Figure 4.21 shows that good model performances are achieved for `max_depth` = 4, `learning_rate` between 0.1 and 0.2 and `gamma` between 0 and 2. On one hand decent results are possible also outside the hyperparameters' region on which the TPE algorithm tend to focus, and on the other hand poor models are possible also with a maximum tree depth of 4. Further information is obtained from other plots produced by means of the `Optuna` library.

The Slice Plot in Figure 4.22 shows that sub-optimal model were achieved in an early stage (light-coloured points) with relatively high learning rates (above 0.2), but later the algorithm focuses on a different hyperparameters' region where learning rates are below 0.2.

Further information on the correlations between variables is obtained from the Rank Plot in Figure 4.23. From the `learning_rate` versus `max_depth` plot is evident that the optimal model is obtained for a maximum depth of 4 only with learning rates around 0.15. Similarly low `gamma` values lead to high Optuna FoM values, only when the learning rate and the maximum depth are

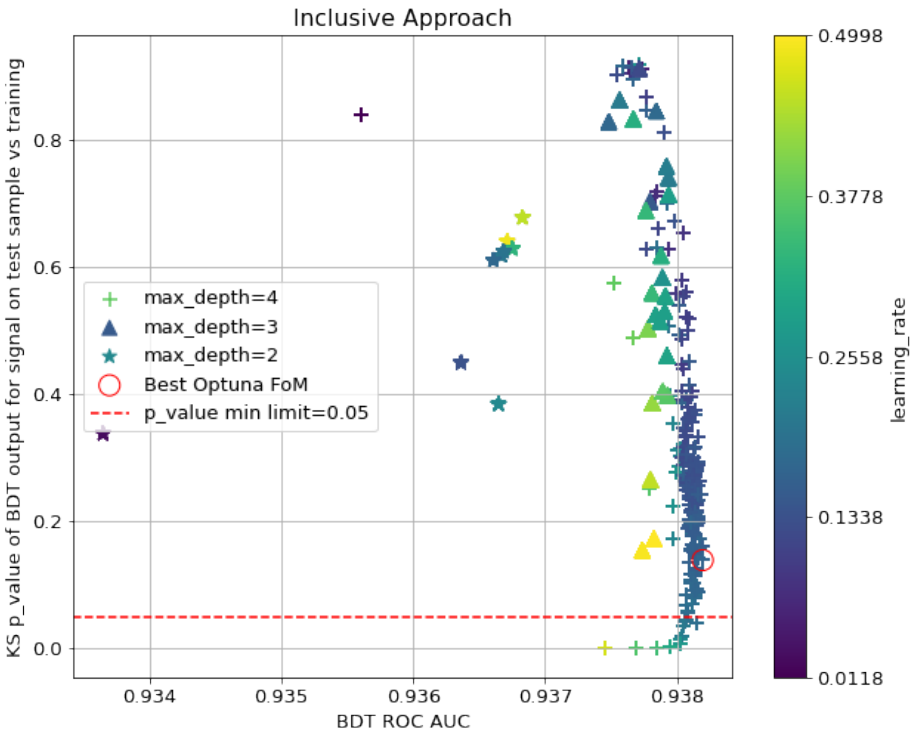


Figure 4.20: Model performance visualization for the Inclusive Approach. The x-axis represents the AUC values obtained from the BDT, and the y-axis represents the p-value of the BDT output. Each marker corresponds to different trials with varying `max_depth` values, and the variable color represent the `learning_rate`. The red dashed line at $p_{KS} = 0.05$ marks the KS test threshold, whereas the red-circled point indicates the best trial according to the `Optuna` FoM.

already at their optimal values, reasonably expected for a lower impact in the optimization procedure.

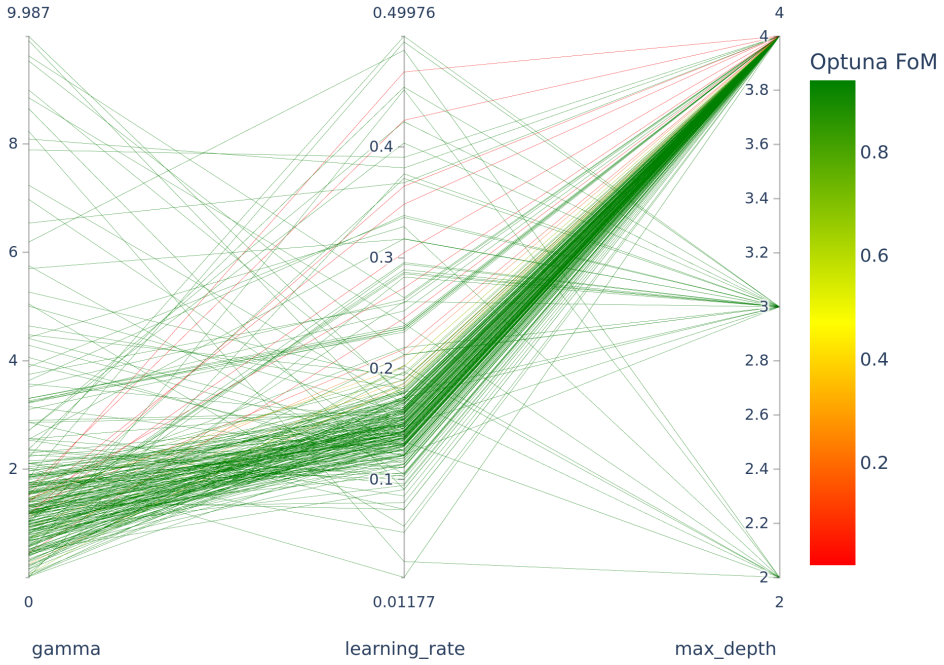


Figure 4.21: Parallel Coordinate Plot for the Inclusive Approach showing the relationships between different hyperparameters and the Optuna FoM. Each line represents a trial, and the color intensity indicates the Optuna FoM value.



Figure 4.22: Slice Plot for the Inclusive Approach showing the Optuna FoM against individual hyperparameters. Each dot represents a trial, with the color intensity indicating the trial number.

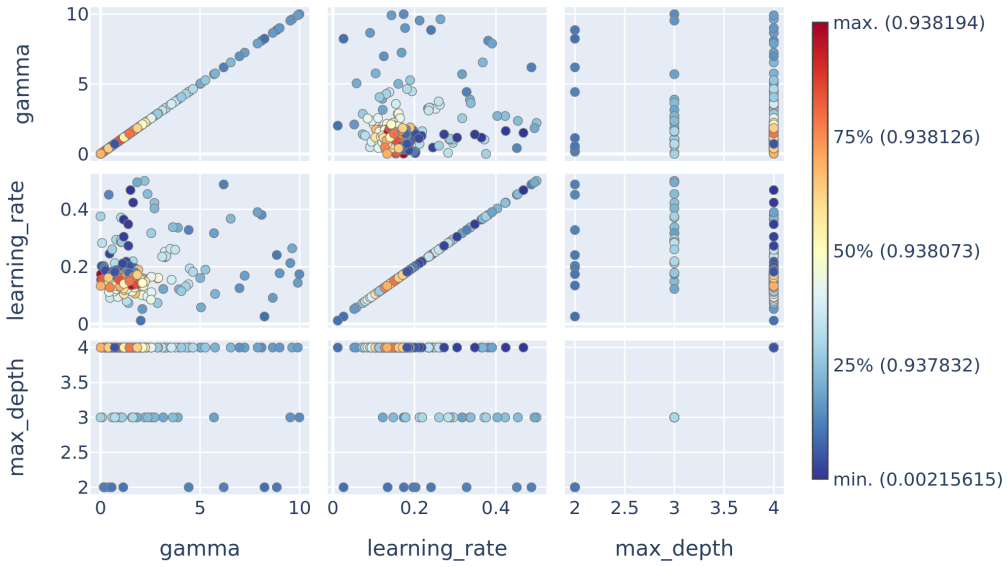


Figure 4.23: Rank Plot for the Inclusive Approach showing the distribution of hyperparameters and their corresponding Optuna FoM values. Each dot represents a trial, colored according to its rank, with the gradient ranging from blue (lowest Optuna FoM) to red (highest Optuna FoM).

4.5.2 Signal extraction optimization

The plots in the previous section show that, analogously to the study for the Displaced Approach, there are several models close in the hyperparameter space with results compatible with each other considering the uncertainty of the order of 0.0003 (see Table 4.3, three times larger than for the displaced study-case). The number of trials characterized by Optuna FoMs within 1σ from the best Optuna FoM value is 95, and therefore the top-100 trials according to the Optuna FoM are considered for the next step. The Punzi SE FoM is then evaluated on the validation sample and it is used to furtherly sort the considered top-100 trials.

The signal extraction procedure consists of UML fits of the invariant mass distribution of the B_c candidates on the validation sample (which comprises 20% of the total data) after the selection with a good model (each of the top-100 Optuna FoM) with a Working Point optimized to obtain the maximum Punzi FoM possible (in this case WP = 0.976) . Figure 4.24 shows the resulting fit for the best Optuna model. The signal model is represented by a Gaussian function with resolution fixed at 31 MeV (obtained from MC study, Fig. 3.2), while the background is modeled using a second-order Chebyshev polynomial. The fit result indicates a significant signal presence with an interesting purity ($p \approx 0.34$).

The best model, as indicated by the highest Optuna FoM combined with the corresponding highest Punzi SE FoM, is then checked by inspecting some relevant distributions such as its score distribution and its ROC curve. The score distribution (Figure 4.25) illustrates clear separation between signal and background: the applied score threshold (vertical dashed red line at 0.984) is relatively close to 1, thus leading to an interesting background rejection. Considering that in this approach the training does not use displacement variables. The ROC curve (Figure 4.26) further validates the model's efficacy, showing an AUC of 0.9381. The WP of 0.976 is highlighted on the curve; it responds to an efficiency of 0.27 with a contamination of 0.3%. All metrics confirm that the model maintains a good discrimination power, even in a situation with large backgrounds characterizing most of the B-Physics analyse with an inclusive topology.

In this inclusive case, the final chosen model, the one that maximises the Punzi FoM, is not the same as the best one obtained by ranking only with the Optuna FoM. Figures 4.27a and 4.27b illustrate this circumtance. The first provides a overview of the Punzi SE FoM versus the Optuna FoM for all the trials passing the KS test. The scatter plot shows that while many trials have similar Optuna FoM values, their Punzi SE FoM values can vary significantly. The trial with the highest Optuna FoM (highlighted in red) does not have the

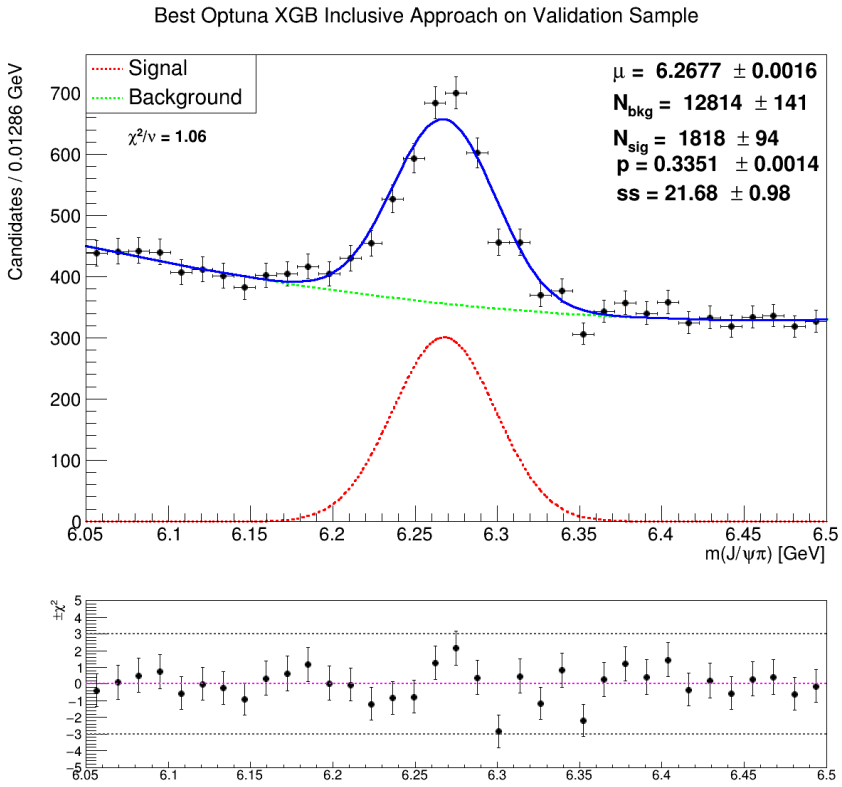


Figure 4.24: Invariant mass distribution of the $J/\psi\pi$ candidates for the validation sample using the best `Optuna` model for the inclusive approach. The fit includes a Gaussian signal model and a second-order Chebyshev polynomial for the background.

highest Punzi SE FoM as shown by a closer examination possible in Figure 4.27b that provides a zoomed-in view of the top-performing trials. Here, it is evident that another trial, marked in orange, exhibits the highest Punzi SE FoM while still having a very high Optuna FoM. This trial can be considered as the optimal model after considering both metrics. The selected trial with the best Punzi SE FoM has an Optuna FoM value of approximately 0.9381 and a Punzi SE FoM of about 2.0004, compared to the highest Optuna FoM trial which has a slightly higher Optuna FoM but a lower Punzi SE FoM like 1.9639. The hyperparameter values for the best model (orange-circled) are reported in the Table 4.8, which also includes the unoptimized `XGBoost` Inclusive Approach for comparison.

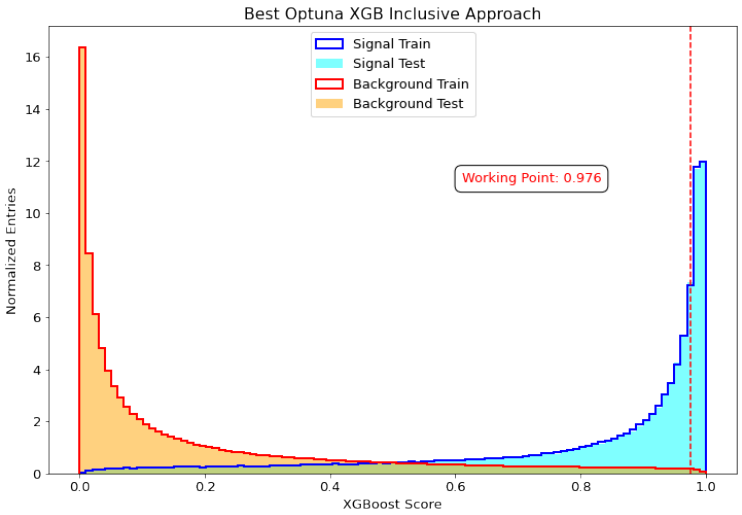


Figure 4.25: `XGBoost` score distribution for the best model obtained with `Optuna` for the Inclusive Approach, using the validation sample. The WP threshold gets the value of 0.976. The histogram shows the `XGBoost` scores for signal and background in both training and test samples.

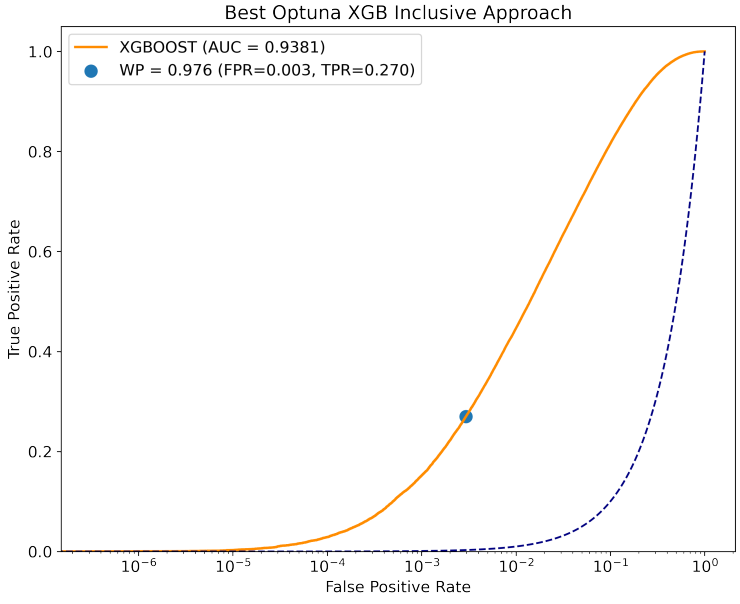
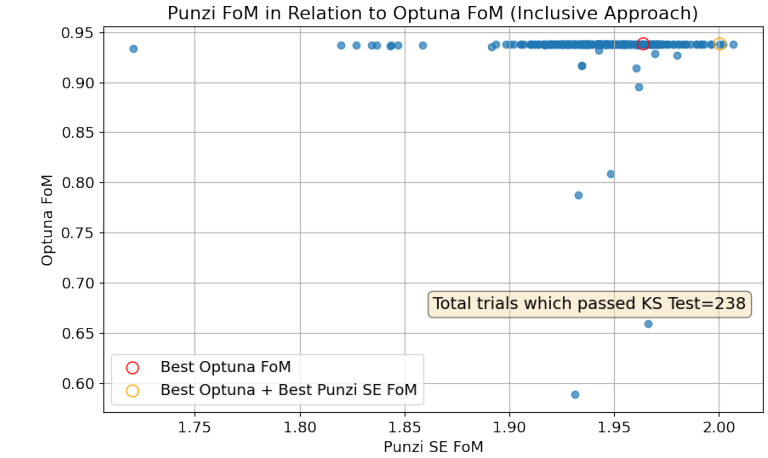
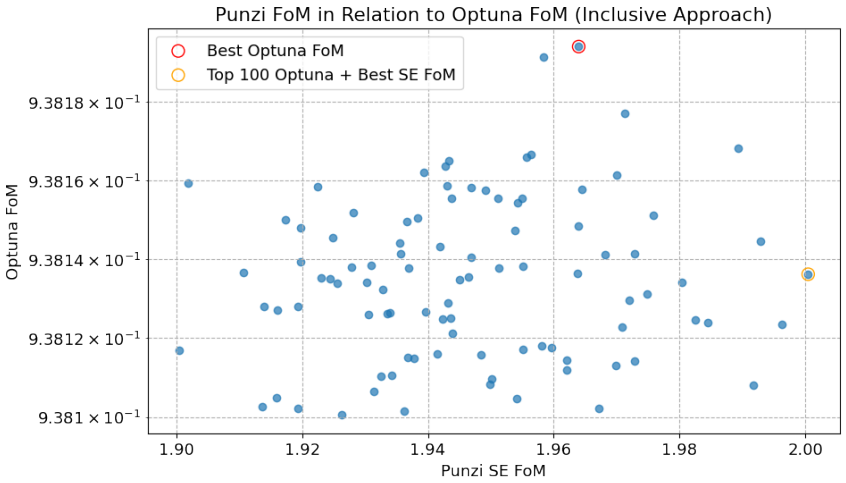


Figure 4.26: ROC curve for the best model obtained with `Optuna` for the Inclusive Approach, evaluated on the validation sample. The AUC value is 0.9381, and the WP of 0.976 is highlighted, showing the corresponding FPR and TPR.



(a)



(b)

Figure 4.27: Comparison of Punzi SE FoM and Optuna FoM in the inclusive approach. (a) General overview showing the relationship between Punzi SE FoM and Optuna FoM for each trial. (b) Zoomed-in view of the top 100 performing trials using a log scale, highlighting how different trials rank when sorted by Punzi SE FoM after achieving similar Optuna FoM values. The red circle indicates the trial with the highest Optuna FoM, and the yellow circle marks the trial recommended by considering the best Punzi SE FoM among the top 100 trials.

Model	max_depth	gamma	learning_rate
Unoptimized XGBoost Inclusive Approach	2	0	0.04
Optuna Range	[2, 4]	[0, 20]	[0.01, 0.5]
Best Optuna FoM	4	0.019862	0.173406

Table 4.8: Comparison of hyperparameters for Unoptimized initial XGBoost Inclusive Approach, Optuna Range allowed in the optimization, and Best Optuna FoM.

4.5.3 Comparison of different SE FoMs

In order to establish the choice of the Punzi FoM among the possible FoMs, the behaviour of other SE FoMs is tested for comparison purposes. The considered SE FoMs are: Signal Significance SS, purity p, their product PxSS, beyond the Punzi FoM.

Figure 4.28 illustrates the normalized trends of the four different SE FoMs while scanning over several score threshold values. The scanned threshold score range stops at 0.996 because tighter thresholds would result in a very small sample where the fit could eventually not be performed successfully. In the developed algorithm the fit is not attempted on samples with less than 100 candidates after the selection, in order to have only reliable signal extraction calculations from the fit results.

While the Purity SE FoM increases steadily, similarly as for the Displaced Approach (Figure 4.14), in this case the Signal Significance exhibits a bump around 0.976 before rapidly decreasing. The PxSS FoM balances the two behaviours and present its maximum at an intermediate working point with respect to p and SS. Here the behaviour of the Punzi SE FoM is similar to that of Signal Significance, while for the displaced approach it was closer to the PxSS one. This shows that the Punzi FoM can represent a compromise between the Signal Significance and the PxSS FoMs by adapting automatically to the sample and the level of background contamination.

The distribution of maximum SE FoM values for the trials' population is presented in Figure 4.29. Each subfigure corresponds to a different SE FoM: Punzi, PxSS, Purity, and Significance. These plots illustrate the frequency of the maximum SE FoM values over all trials, providing insights into the optimization for each metric.

The Punzi SE FoM (Figure 4.29a) and the Significance SE FoM (Figure 4.29d) show similarly a relatively wide distribution: this suggests that the choice of working point is sensitive to the model, even though the considered models are all close to each other in the parameter space.

The Purity SE FoM (Figure 4.29c) presents instead a steep peak, reflecting a

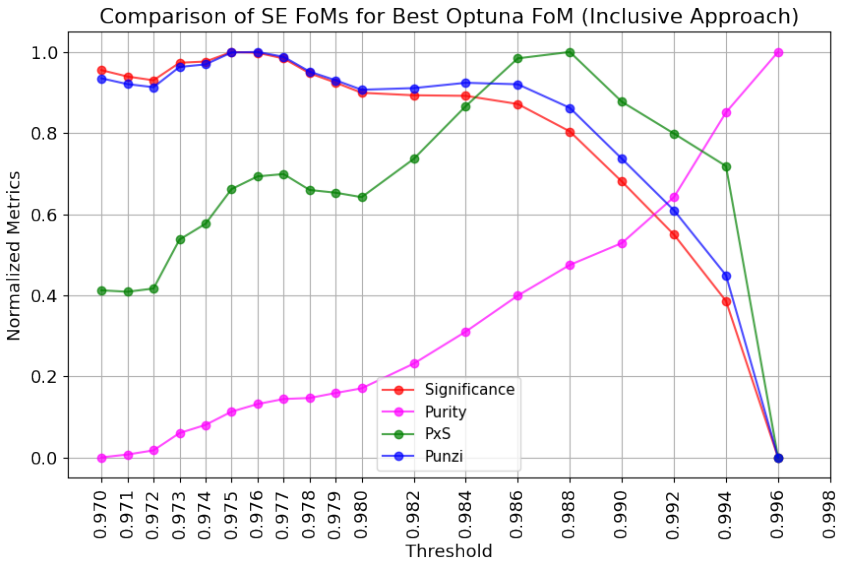


Figure 4.28: Comparison of different SE FoMs for the best model for the inclusive approach: Significance, Purity, PxS, and Punzi for the Inclusive Approach. The plot shows the SE FoMs values against varying working points (score thresholds) for the best model according to the previously described dual-ranking.

strong preference for tight working points that reject almost all the background, and also much of the signal as expected.

The PxSS SE FoM (Figure 4.29b) displays a more concentrated distribution with an intermediate peak, suggesting a more stable optimization process for the PxSS metric compared to the Punzi SE FoM. Nonetheless the working point selected by the PxSS FoM is still a tight working point that rejects effectively much background at the cost of a reduced signal yield. Therefore the choice between the Punzi FoM and the PxSS FoM in an Inclusive Approach must be guided by the kind of analysis that is going to be performed, that is according to the preference for a larger sample or a purer sample.

The results obtained from the Punzi and PxSS FoMs optimization will be both discussed respectively in Sections 4.5.4 and 4.5.5.

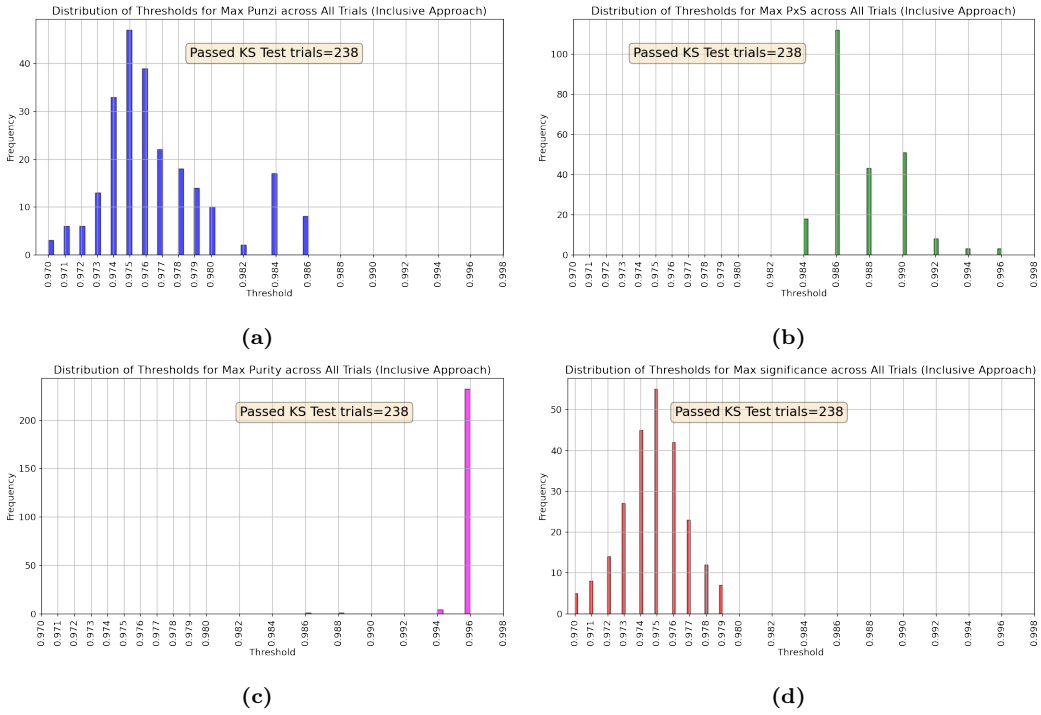


Figure 4.29: Comparison of different SE FoMs: (a) Punzi SE FoM, (b) PxSS SE FoM, (c) Purity SE FoM, and (d) Significance SE FoM. Each plot shows the distribution of the maximum value obtained for each SE FoM over all trials, showing the variability and stability of the optimization process for each metric.

4.5.4 Results with the Punzi FoM optimization

The best model coming from the choice of the Punzi FoM for the signal extraction is then applied to the entire dataset, in order to estimate the signal yield and the background contamination. The resulting invariant mass distribution of the B_c candidates is shown in Figure 4.30. For the full dataset, the background model was modelled with an exponential to better accommodate the larger background volume. The fit results show a signal mass of $\mu = 6.27141 \pm 0.00072$ GeV and a resolution of $\sigma = 0.02665 \pm 0.00076$ GeV. The signal yield is $N_{sig} = 7235 \pm 197$ events, with a purity of $p = 0.3086 \pm 0.0006$. The fit result for the full dataset indicates a further improvement in statistical significance as expected.

Table 4.9 compares the results with and without optimization, thus quantifying improvements. The purity has increased by approximately 1.37 times, from 0.2257 to 0.3086, and the significance has slightly increased from 39.94 ± 1.05 to 41.36 ± 1.00. The B_c mass estimate from the optimized model is consistent with the PDG value ($M_{\mathrm{PDG}}(B_c) = 6.27447 \pm 0.00032$ GeV) and previous CMS results.

The Table 4.9 shows that the results from the unoptimized XGB training and

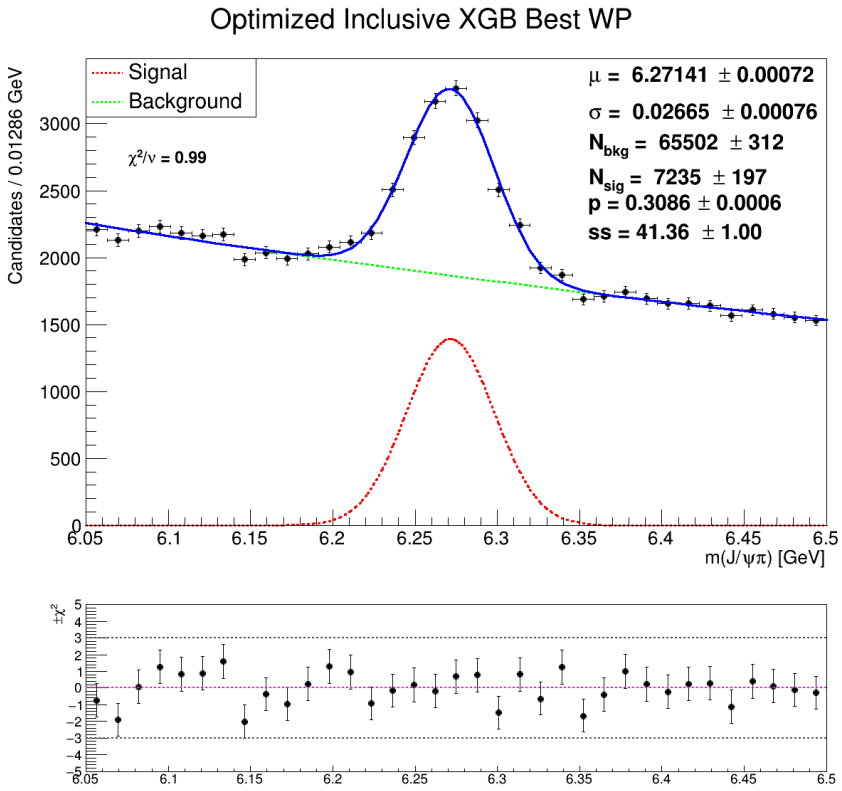


Figure 4.30: Distribution of the invariant mass of B_c candidates selected by the optimized inclusive XGB model with the best WP for the Punzi FoM applied to the full dataset. The result of the UML fit (blue curve) is superimposed. The fit components are a single Gaussian for the signal (red dashed curve) and an exponential model for the background (green dashed curve). Bottom plot: Bin-by-bin pull distribution, providing a pseudo-χ^2 for each bin.

	Unoptimized XGB Inclusive	**Optimized XGB Best WP**
Signal Yield	9476 ± 271	7235 ± 197
Background Yield	129877 ± 440	65502 ± 312
Purity	0.2257 ± 0.0003	0.3086 ± 0.0006
Significance	39.94 ± 1.05	41.36 ± 1.00
χ^2/ν	1.10	0.99

Table 4.9: Comparison of different selections for the displaced approach: signal and background yields, purity (p), significance (ss), and χ^2/ν values.

the Optuna-based hyperparameter optimization in the Inclusive Approach are not too far from each other. This proximity may be attributed to the initial model, which was already enough good based on insights from previous studies. However, as manual optimization reaches its limitations, a need for a more systematic and universally applicable method arises. To address this, the

SE FoM method was employed, followed by meticulous hyperparameter tuning using Optuna. This rigorous approach resulted in an improvement in performance. This outcome highlights the efficacy of the Optuna-based optimization in refining the model to achieve more performant and reliable results.

4.5.5 Results with the PxSS FoM optimization

In this section, the results obtained using the PxSS FoM are presented, a metric particularly valuable in scenarios where balancing signal purity and background rejection is critical. Unlike the Punzi FoM, PxSS emphasizes the maximization of the product of these two FoMs, resulting in a model that shows robust performance in complex datasets.

The model selection process for the PxSS FoM follows a similar methodology as described in Section 4.5.4, but with the PxSS FoM as the ranking metric. The top-performing models based on the Optuna FoM were further evaluated using the PxSS FoM to identify the optimal model. The scatter plot in Figure 4.31 highlights the best trials according to both Optuna FoM and PxSS FoM, with the selected model (highlighted in orange) showing the highest PxSS FoM.

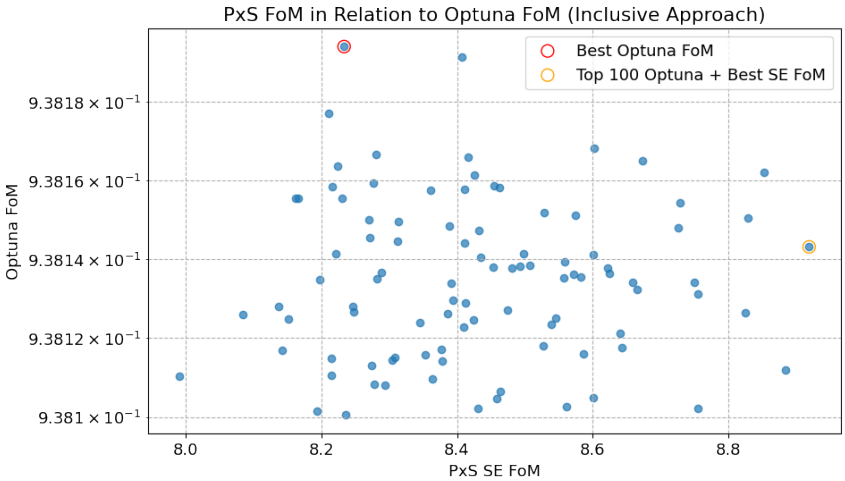


Figure 4.31: Comparison of PxSS SE FoM and Optuna FoM in the inclusive approach. The scatter plot shows the relationship between PxSS SE FoM and Optuna FoM for all the trials passing the KS test. The orange-circled point marks the trial selected as the optimal model according to the PxSS SE FoM within the top 100 trials.

After selecting the model with the highest PxSS FoM, its performance was checked through the score distribution and ROC curve. The score distribution (Figure 4.32) the same as in Figure 4.25, now shows the new chosen working point (WP = 0.986) marked by the vertical dashed red line. The ROC curve (Figure 4.33) is of course the same as in Figure 4.26, with an AUC of 0.9381.

The new working point (WP = 0.986) is highlighted on the curve; it corresponds to efficiency of 0.175 with a contamination of 0.1%. The PXSS FoM underscores the model's capability to balance signal purity and background rejection effectively, making it well-suited for applications that require a good signal extraction in high-background topologies.

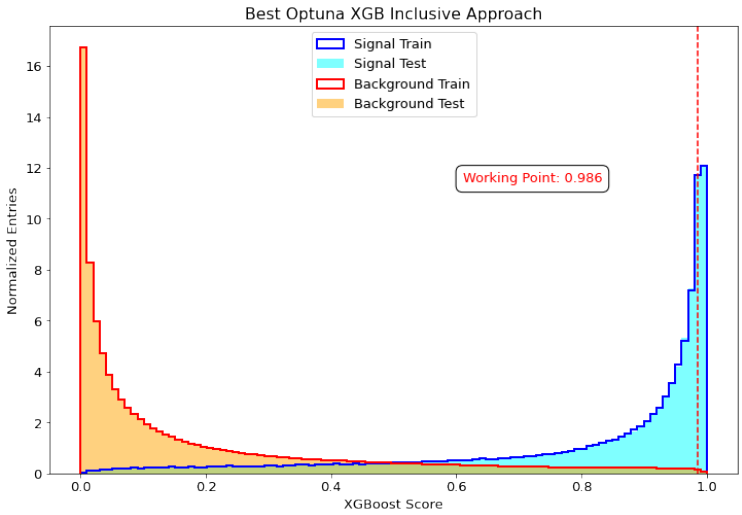


Figure 4.32: Score distribution for the best model obtained with `Optuna` for the inclusive approach using PxSS FoM. The WP threshold gets the value of 0.986. The histogram shows the `XGBoost` scores for signal and background in both training and test samples.

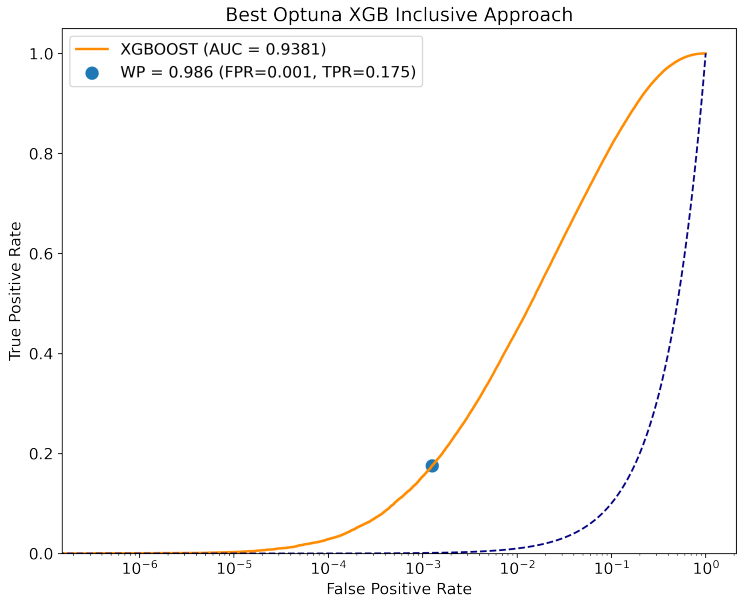


Figure 4.33: ROC curve for the best model obtained with `Optuna` for the inclusive approach using best PxSS FoM, evaluated on the validation sample. The AUC value is 0.9381, and the WP of 0.986 is highlighted.

The best model identified using the PxSS FoM was further validated by applying it to the validation sample, as shown in Figure 4.34. The invariant mass distribution of the $J/\psi\pi$ candidates was fitted using a Gaussian function for the signal and a second-order Chebyshev polynomial for the background. The fit results indicate a signal yield of $N_{sig} = 1164 \pm 65$ events, with a background yield of $N_{bkg} = 5342 \pm 92$ events. The signal purity was calculated to be $p = 0.4406 \pm 0.0026$ in the 2σ window, and the Signal Significance was found to be $ss = 20.24 \pm 0.94$.

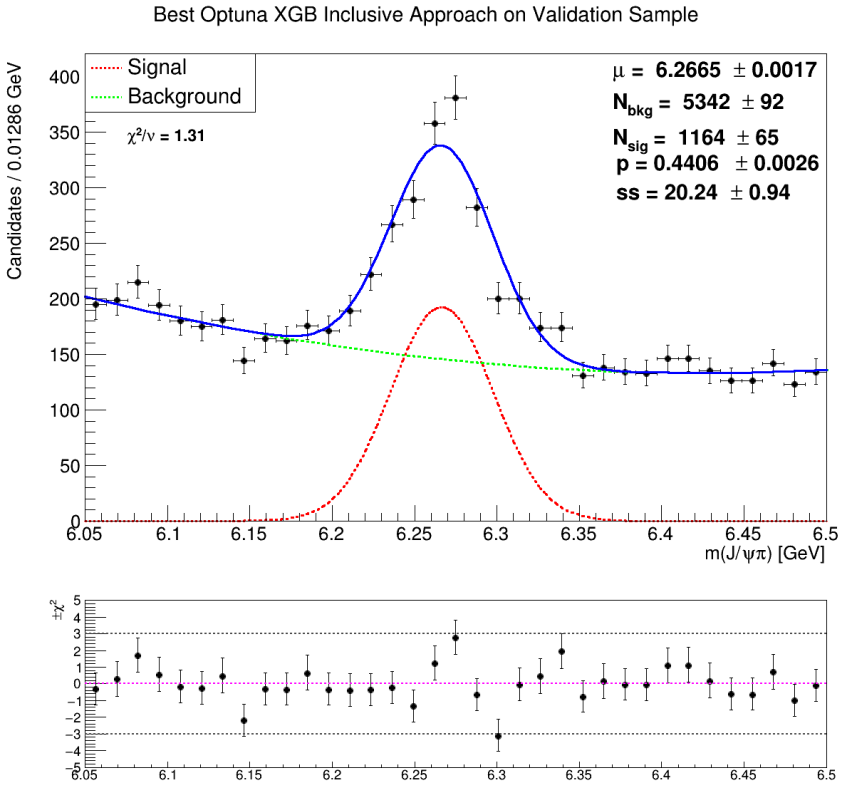


Figure 4.34: Invariant mass distribution of the $J/\psi\pi$ candidates for the validation sample using the best `Optuna` model for the Inclusive Approach, optimized with PxS FoM. The fit includes a Gaussian signal model and a second-order Chebyshev polynomial for the background.

The comparison of SE FoMs for the PxSS-optimized model, as shown in Table 4.10, provides further insight into the performance and reliability of the chosen hyperparameters. The distribution of the different SE FoMs reveals several important aspects of the model's behavior across varying thresholds.

Model	max_depth	gamma	learning_rate
Unoptimized XGBoost Inclusive Approach	2	0	0.04
Optuna Range	[2, 4]	[0, 20]	[0.01, 0.5]
Best Optuna FoM (Punzi)	4	0.019862	0.173406
Best Optuna FoM (PxSS)	4	0.025033	0.154289

Table 4.10: Comparison of hyperparameters for Unoptimized initial XGBoost Inclusive Approach, Optuna Range allowed in the optimization, and Best Optuna FoM for Punzi and PxSS.

Similar to the model selected based on the Punzi SE FoM (refer to Figure 4.28), the optimized model also demonstrates a consistent trend where SS and Punzi SE FoMs again behave very closely and reach their maxima at the same threshold. The PxSS SE FoM metric reaches its peak at a single defined threshold value (0.986), which indicates a more definitive and reliable selection criterion for the working point choice. Interesting to note that for the best PxSS WP trial, SS, Punzi and PxSS show approximately the same threshold value.

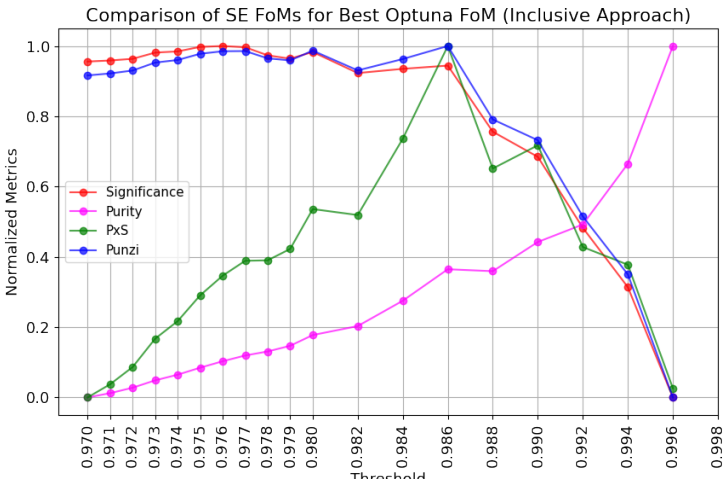


Figure 4.35: Comparison of different SE FoMs for the best model for the Inclusive Approach: Signal Significance, Purity, PxSS, and Punzi. The plot shows the SE FoMs values against varying working points (score thresholds) for the best model according to the PxSS optimization.

The best model derived from the PxSS FoM optimization was subsequently applied to the entire dataset to extract the signal yield and assess background contamination. The resulting invariant mass distribution of the B_c candidates is shown in Figure 4.36. In this case, the background model was again adjusted to an exponential fit to accommodate in a simpler model the characteristics of

the larger dataset.

The fit results show a signal mass of $\mu = 6.27006 \pm 0.00075$ GeV and a resolution of $\sigma = 0.02644 \pm 0.00077$ GeV. The signal yield is $N_{sig} = 4788 \pm 135$ events, with a purity of $p = 0.4128 \pm 0.0011$.

The application of the PxSS-optimized model to the full dataset demonstrates significant improvements, particularly in terms of signal purity that has increased to 0.4128. This increase in purity can be attributed to the higher threshold value selected during the optimization process, which prioritizes a relatively better background rejection. The mass estimate from this optimized model is consistent with the expected value.

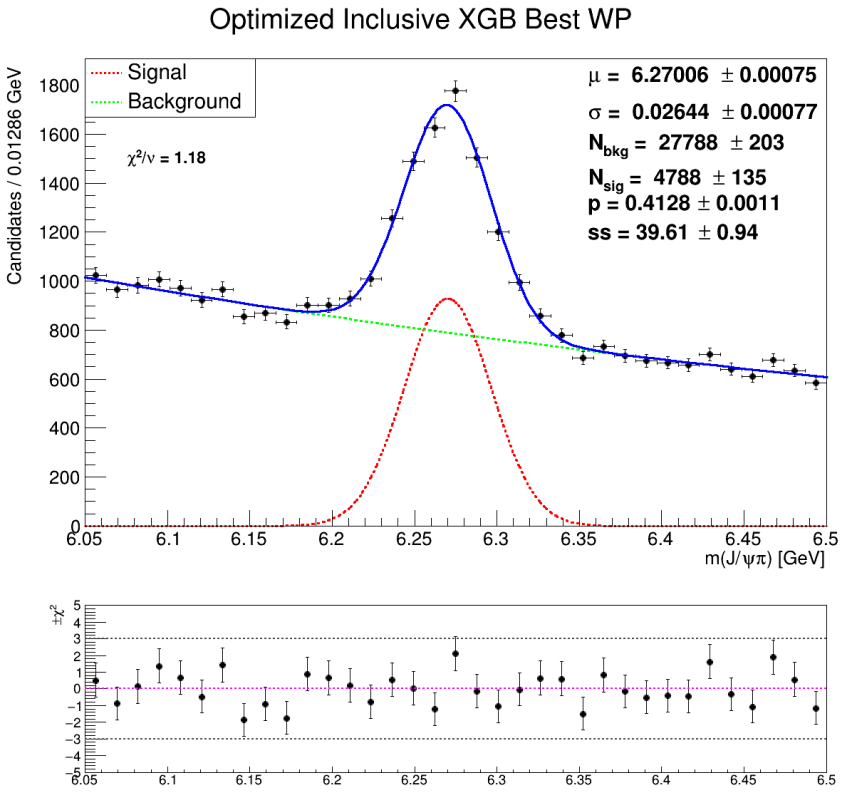


Figure 4.36: Distribution of the invariant mass of B_c candidates selected by the PxS-optimized inclusive XGB model with the best WP for the PxS FoM applied to the full dataset. The result of the UML fit (blue curve) is superimposed. The fit components include a single Gaussian for the signal (red dashed curve) and an exponential model for the background (green dashed curve). Bottom plot: Bin-by-bin pull distribution, providing a pseudo-χ^2 for each bin.

The table 4.11 summarizes the results obtained from the invariant mass fits for the unoptimized XGB inclusive approach and the optimized models using two different FoMs: Punzi and PxSS. The signal yields for the unoptimized, Punzi-optimized, and PxSS-optimized models are 9476 ± 271, 7235 ± 197, and

4788 ± 135 events, respectively. Despite the reduction in signal yield as we move towards higher thresholds (as indicated by the lower signal yield in the PxS-optimized model), the significance remains relatively stable, with values of 39.94 ± 1.05 for the unoptimized model, 41.36 ± 1.00 for the Punzi-optimized model, and 39.61 ± 0.94 for the PxSS-optimized model. This trend in signal significance suggests that while fewer signal events are retained at higher thresholds, the events that are kept are more likely to be true signal rather than background as expected. Of course, this is further corroborated by the increase in purity, which rises from 0.2257 ± 0.0003 in the unoptimized model to 0.3086 ± 0.0006 in the Punzi-optimized model, and to 0.4128 ± 0.0011 in the PxSS-optimized model. As already mentioned, the higher purity in the PxSS-optimized model is a direct result of the higher working point threshold applied, which effectively filters out more background while sacrificing same signal yield.

	Unoptimized XGB Inclusive	**Optimized Inclusive XGB Punzi**	**Optimized Inclusive XGB PxSS**
Signal Yield	9476 ± 271	7235 ± 197	4788 ± 135
Background Yield	129877 ± 440	65502 ± 312	27788 ± 203
Purity	0.2257 ± 0.0003	0.3086 ± 0.0006	0.4128 ± 0.0011
Significance	39.94 ± 1.05	41.36 ± 1.00	39.61 ± 0.94
χ^2/ν	1.10	0.99	1.18

Table 4.11: Comparison of different selections for the inclusive approach: signal and background yields, purity (p), significance (ss), and χ^2/ν values. The first two columns are, of course, the same given in Table 4.9.

Chapter 5

Preliminary B_C lifetime measurement

The measurement of the lifetime of the B_c^+ meson is a subject of significant interest in hadron physics. As already discussed earlier, this meson presents a unique opportunity to study the dynamics and decay processes inherent to systems containing two different heavy quarks. The lifetime of the B_c^+ meson is determined by a combination of QCD and weak interactions, making it a critical parameter for testing both theoretical models and experimental techniques with high precision. In this analysis, the proper decay length, denoted as ct, is the observable used to extract the lifetime of the B_c^+ meson; it is calculated using the following equation:

$$ct = \frac{l_{dec}^{\perp} \cdot m}{p_T}$$

where $l_{dec}^{\perp}$ represents the transverse flight distance between the Primary Vertex (PV) and the Secondary Vertex (SV), p_T is the transverse momentum of the B_c^+ meson, and m is its nominal (PDG) mass.

The determination of l_{dec} implies accurately identifying the correct PV, which from the B_c^+ meson originated. This is achieved by selecting the PV that minimizes the distance to the decay vertex (SV) along the flight direction of the meson. The "minimum distance" criterion (described in Appendix B) is particularly crucial to ensure that the selected PV is closest to the true production point of the meson, thereby minimizing the uncertainty of the proper decay length ct. This method has been found to provide a more reliable and unbiased estimate of the B_c^+ meson's lifetime. Further details on the calculation of ct and the methodology used for selecting the PV are given in Appendix B.

In this study, the flight time measurements for the `displaced` and `inclusive` selections (Table 3.3) will be presented and compared, along with the results for the `XGBoost`-based approaches.

As demonstrated in Tables 4.6 and 4.9, the optimized `XGBoost displaced` and inclusive approach models have successfully extracted the signal. This chapter focuses on two main aspects of the preliminary lifetime extraction:

- the distortion of the decay length distribution due to the selection,
- the possible effect introduced by the choice of the PV (amongst all the reconstructed ones) by means of the "minimum distance" method.

The two effects are quantified respectively in terms of *selection efficiency* and *reconstruction efficiency*. Effects coming from the trigger selection are not considered at this stage.

5.1 Selection Efficiency

The study of the selection efficiency has been carried out on the signal MC sample. In order to decouple the *selection efficiency* from the *reconstruction efficiency*, the study utilizes the flight length computation at gen-level (later called as *gen-ct*), which uses the true PV information available at generator-level.

Generator-level (gen-level) information refers to the simulated events without accounting for detector effects yet. At this level, the gen-level flight distance (gen-ct) distribution reflects the pure kinematic properties of the B_c meson decay, unaffected by any smearing or inefficiencies introduced by the detector simulation. This makes the gen-level distribution a good reference point, as it represents the ideal theoretical behavior of the system, allowing for a direct comparison with simulated events at reconstruction level to assess the impact of detector effects.

On real data instead the available ct values are obtained by using the reconstructed vertices: this is called later in the text *reco-ct* (details in Appendix B).

In the selection efficiency the reference denominator consists of all the gen-matched candidates selected before any kinematic selections or additional `preprocessing` criteria. The numerator, instead, consists of the candidates surviving the considered selection. It must be recalled that `preprocessing` is applied before any `XGBoost`-based model. The selection efficiency expression is given by:

$$\epsilon_{sel} = \frac{N_{\text{sel}}}{N_{\text{gen}}} \tag{5.1}$$

where N_{sel} is the number of events passing the optimized `XGBoost` approach, and N_{gen} is the number of gen-matched events. This efficiency will be inspected as a function of the proper decay length ct.

Figure 5.1 illustrates the generated ct distribution for B_c meson candidates that pass the gen-matching filter. This plot serves as the baseline for comparison with further steps, representing the generated transverse flight length distribution.

The fit model is defined by a pure `RooExponential` function, which represents the exponential decay. The fit to the generated ct distribution yields a decay constant $\lambda_{gen} = -66.74 \pm 0.05$ cm^{-1}, corresponding to a $c\tau_{gen} = 149.84 \pm 0.11$ μm and a lifetime of $\tau_{gen} = 0.4998 \pm 0.0004$ ps. This lifetime is in good agreement with the generation lifetime $\tau_{B_c}^{gen} = 0.507$ ps used for the generated MC sample. The bottom plot shows the bin-by-bin pull distribution, indicating the goodness of fit (with $\chi^2/\text{ndf} = 0.98$) suggesting a well-fitting model.

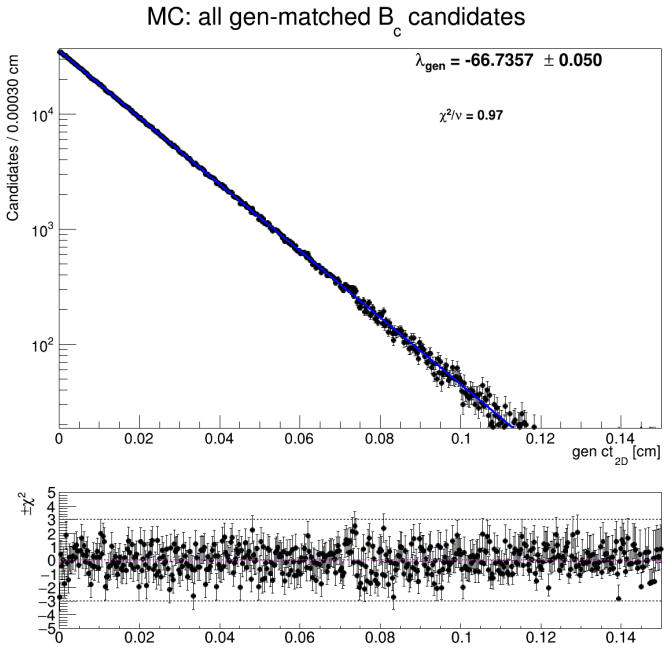


Figure 5.1: Top plot: Distribution of the generated ct for B_c meson candidates passing the gen-matching filter. The distribution is modeled with a `RooExponential` function representing a pure exponential decay. The extracted decay constant is $\lambda_{\text{gen}} = -66.74 \pm 0.05$ cm^{-1}, corresponding to a lifetime of $\tau_{gen} = 0.4998 \pm 0.0004$ ps. Bottom plot: Bin-by-bin pull distribution indicating the goodness of fit.

5.1.1 Inclusive cut-based approach

Figure 5.2 illustrates the generated proper flight distance distribution for B_c candidates after applying the `inclusive` cut-based approach on the Monte Carlo sample. The fit model employed is the `RooDecay` function, which is a decay model that includes the smearing due to detector effects. The distribution is modeled using a convolution of an exponential decay with a Gaussian resolution function to account for detector resolution effects. The fit allows to extract a lifetime $\tau_{B_c} = 0.514 \pm 0.002$ ps.

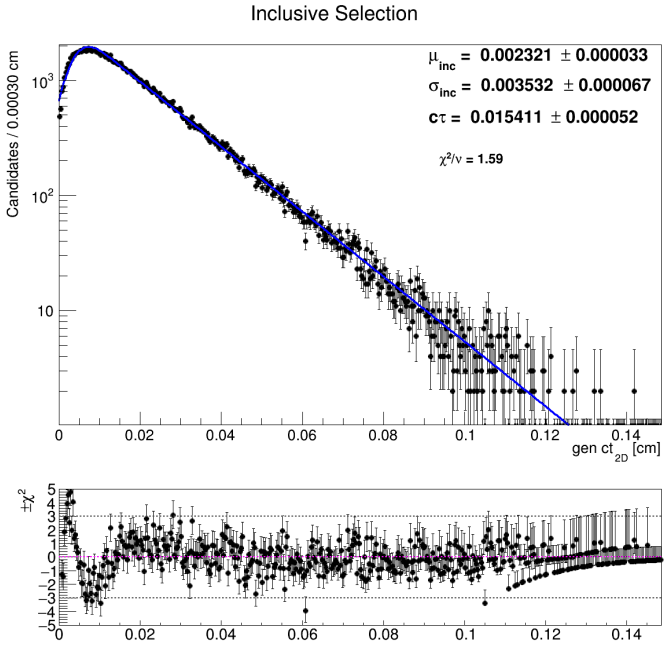


Figure 5.2: Generated proper decay length distribution of B_c mesons after the `inclusive` selection on the MC sample. The distribution is fitted with a `RooDecay` function, yielding to a lifetime of $\tau_{B_c} = 0.514 \pm 0.002$ ps.

In Figure 5.3, the top plot illustrates the proper decay length distribution of B_c meson candidates before any selection and after the `inclusive` approach. The blue histogram represents the gen-matched selection, while the red histogram indicates the `inclusive` one. The gen-matched distribution shows, as expected, the exponential shape, whereas the `inclusive` approach exhibits a significant reduction in candidates, especially at the very lower decay lengths values, indicating the selection's impact. The selection efficiency as a function of gen-ct is obtained as the ratio of the two distributions (see Eq. 5.1) and is shown in the bottom plot.

The selection efficiency curve is expected to exhibit a turn-on shape, thus

it is fitted with a generalized error function:

$$\epsilon(ct) = x \cdot \mathrm{erf}(y \cdot (ct - z)) + w \tag{5.2}$$

where x represents the amplitude of the efficiency, y controls the width of the transition, z corresponds to the center of the transition, and w accounts for the baseline offset. The plateau start is defined as the point where the efficiency reaches a given fraction of the maximum efficiency value, calculated as:

$$ct_{\mathrm{plateau}} = f \cdot \max(\epsilon(ct)) \tag{5.3}$$

where f is the fraction of the maximum efficiency: in this study, $f = 0.95$ for the cut-based selections and $f = 0.90$ for the `XGBoost` approach, unless otherwise stated.

Thus, the efficiency shape is obtained using this error function model to capture the transition from low to high efficiency. The start of the plateau, where the efficiency stabilizes, was determined by finding the point where the efficiency reaches 95% of its maximum value, as described in Eq. 5.3. Using this method, the start of the plateau was found to be at $ct = 0.0147$ cm. The vertical dashed red line in the plot marks this value, indicating where the efficiency can be considered to stable. The efficiency is relatively low at short decay lengths and increases with longer decay lengths, reaching a plateau beyond this point.

After determining the start of the plateau, the gen-ct distribution after the `inclusive` approach was redrawn starting from the chosen point to fit the distribution ct a new fit to the efficiency plateau region (Figure 5.4). The fit model can be still defined by the `RooExponential` function, which represents a simple exponential decay. The new value for the corresponding lifetime was found to be $\tau_{B_c} = 0.503 \pm 0.002$ ps.

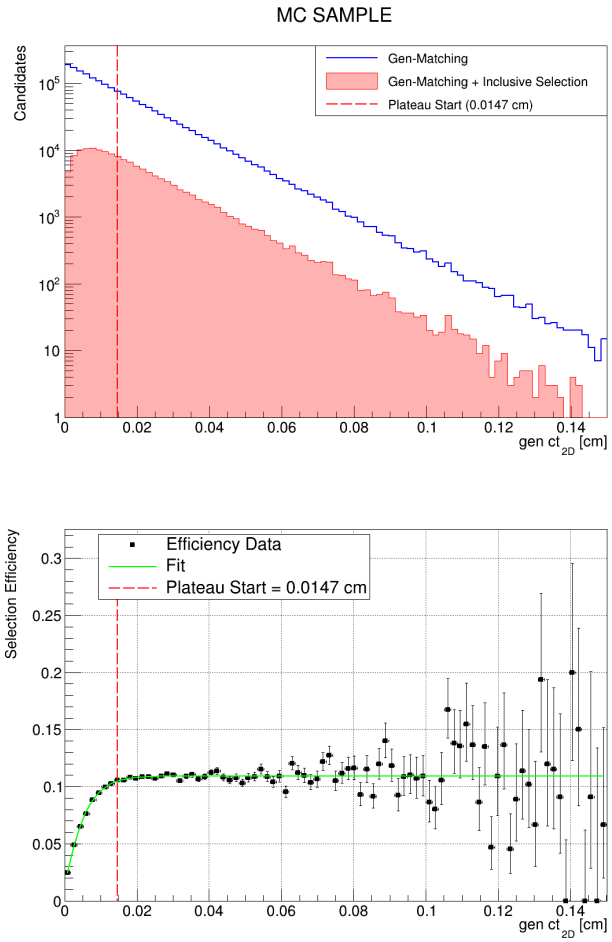


Figure 5.3: Comparison of the proper transverse decay length distribution of B_c meson candidates for gen-matching (blue) and `inclusive` approach (red). The top plot shows the candidate distribution, while the bottom plot depicts the selection efficiency (see Eq. 5.1) as a function of the proper decay length ct.

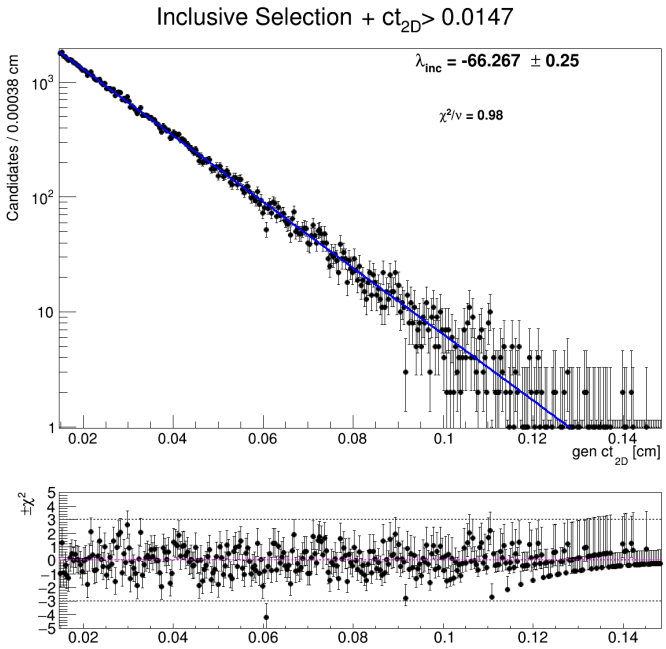


Figure 5.4: Proper decay length distribution of B_c meson candidates after applying the `inclusive` cut-based selection, starting from the selection efficiency plateau point. The simple exponential fit result is superimposed.

5.1.2 Inclusive `XGBoost`-based approach

The proper decay length distribution for B_c meson candidates after applying the optimized `XGBoost inclusive` approach is shown in Figure 5.5. The distribution is fitted using a `RooDecay` function, which represents an exponential decay convoluted with a Gaussian resolution function. The lifetime of $\tau_{B_c} = 0.580 \pm 0.002$ ps. These values are higher compared to the expected B_c meson lifetime due to the `XGBoost inclusive` approach, which can introduce additional biases in the decay length measurement.

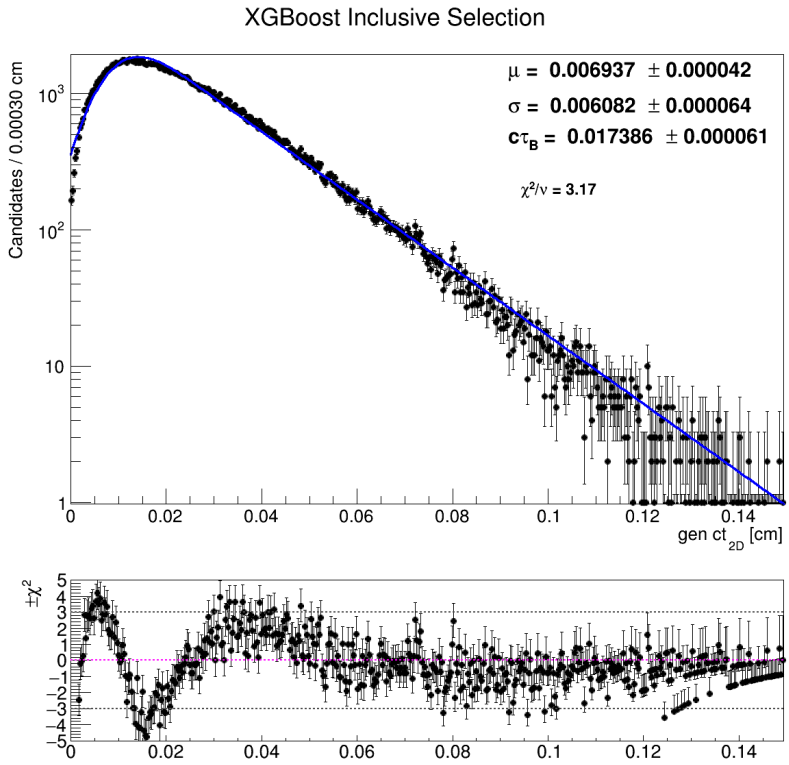


Figure 5.5: Proper decay length distribution of B_c meson candidates for `XGBoost inclusive` approach on the MC sample. The distribution is fitted with a `RooDecay` function; the fit return a lifetime of $\tau_{B_c} = 0.580 \pm 0.002$ ps.

In Figure 5.6, the top plot compares the proper decay length distributions for B_c meson candidates selected by gen-matching (blue) and the `XGBoost inclusive` approach (red). While the gen-matching selection displays the expected exponential decay, the `XGBoost inclusive` approach results in a noticeable reduction in the number of candidates at lower decay lengths, reflecting the effect of the selection criteria.

The bottom plot in Figure 5.6 presents the selection efficiency of the `XGBoost inclusive` approach relative to gen-matching as a function of the proper decay

length ct. Fitting the efficiency curve using the error function model (see Eq. 5.2), the start of the plateau is determined by the point where the efficiency reaches 90% of its maximum value (see Eq. 5.3). With this assumption, the plateau begins at $ct = 0.0440$ cm.

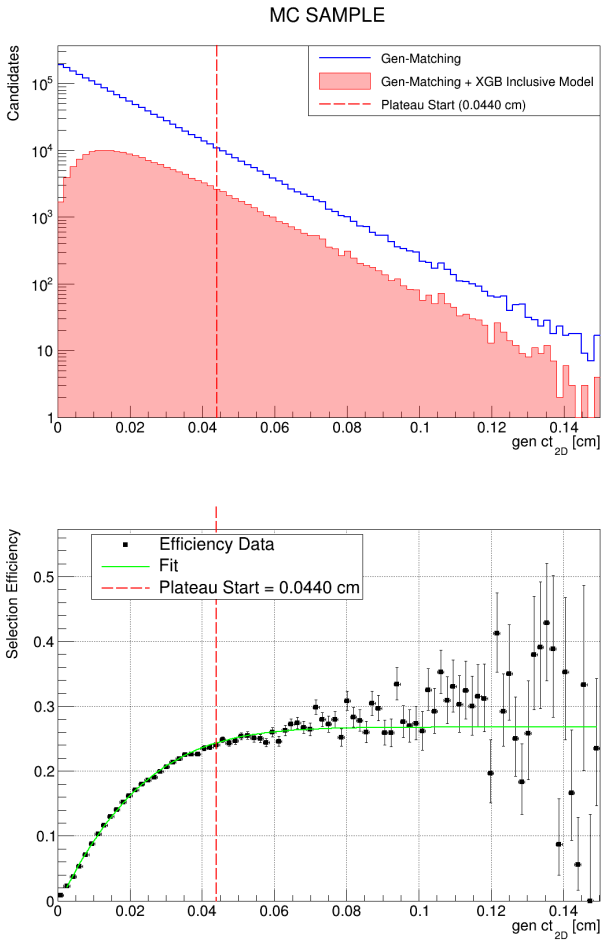


Figure 5.6: Comparison of the proper decay length distribution of B_c meson candidates for gen-matching (blue) and `XGBoost inclusive` approach (red). The top plot shows the candidate distribution, while the bottom plot depicts the selection efficiency as a function of proper decay length ct(see Eq. 5.1).

Once the plateau "start" was identified, the gen-ct distribution for the `XGBoost inclusive` approach was recalculated from this point onward and fitted (Figure 5.7) using the `RooExponential` function, which models the decay as a simple exponential. The newly estimated lifetime is $\tau_{B_c} = 0.526 \pm 0.003$ ps.

When comparing the `inclusive` cut-based and `XGBoost inclusive` approaches, the most striking difference is the efficiency and its impact on the ct stabilization point. The `inclusive` approach achieves an efficiency of approximately 0.10, while the `XGBoost inclusive` approach increases this efficiency

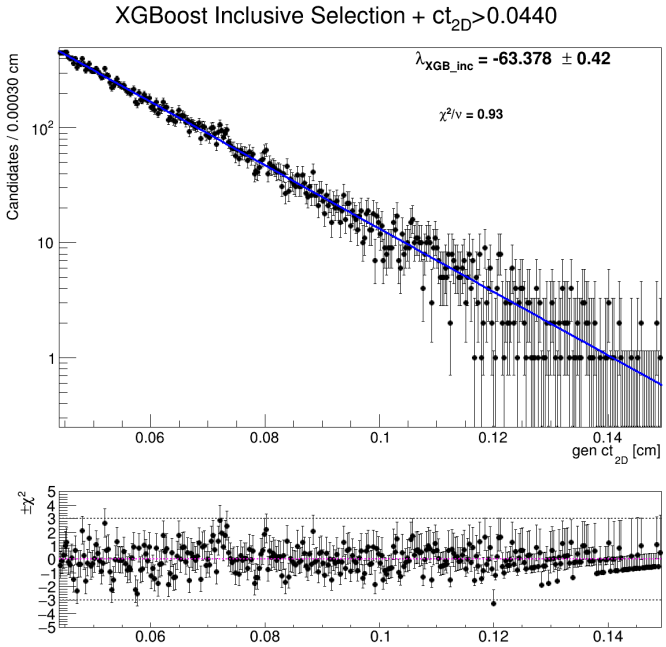


Figure 5.7: Proper decay length distribution of B_c meson candidates for `XGBoost inclusive` approach starting from the plateau point. The fit provides the decay length and corresponding lifetime.

to around 0.30, representing a relevant improvement in candidate retention. However, this increase in efficiency comes with a trade-off: the stabilization point for the `XGBoost inclusive` approach occurs later, at $ct = 0.0440$ cm, compared to the `inclusive` cut-based approach, which stabilizes much earlier at $ct = 0.0147$ cm.

This delay in stabilization results in a bias towards candidates with longer decay lengths in the `XGBoost inclusive` approach: while the `inclusive` approach yields a lifetime of $\tau_{B_c} = 0.503 \pm 0.002$ ps, which is in close agreement with the reference generation value of $\tau_{B_c}^{gen} = 0.507$ ps, the `XGBoost inclusive` approach produces a larger lifetime of $\tau_{B_c} = 0.526 \pm 0.003$ ps.

This deviation from the reference lifetime value in the `XGBoost inclusive` approach indicates that while higher efficiency improves candidate selection, it also shifts the decay length to slightly higher values, introducing a potential bias. In summary, although the `XGBoost inclusive` approach provides a significant gain in efficiency, the earlier stabilization and closer alignment with the reference lifetime in the `inclusive` cut-based approach demonstrate its reliability in maintaining unbiased the B_c meson proper decay length distribution.

5.1.3 Displaced cut-based approach

The proper flight distance distribution for B_c candidates, selected using the `displaced` cut-based approach criteria outlined in Table 3.3, is shown in Figure 5.8. Similar to the `inclusive` approaches, the `RooDecay` function is applied here, thus incorporating the convolution of the decay model with a Gaussian resolution function. A lifetime of $\tau_{B_c} = 0.536 \pm 0.002$ ps is obtained from the fit.

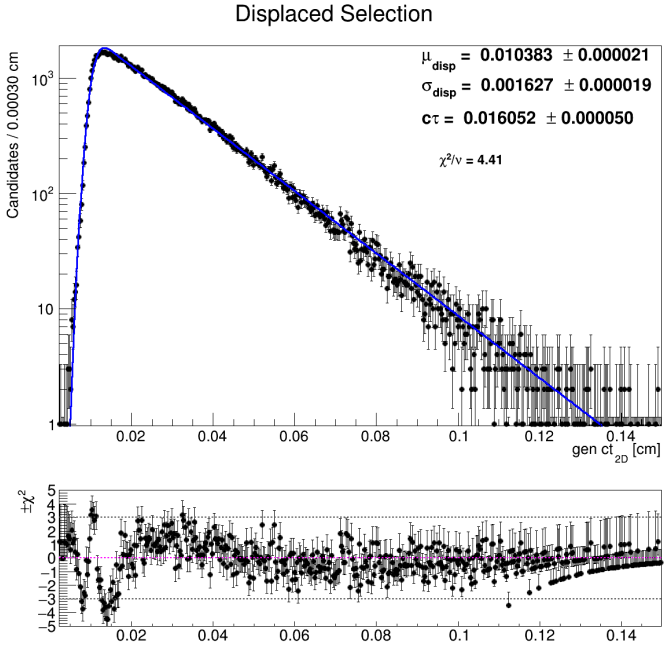


Figure 5.8: Proper flight distance distribution of B_c mesons after the `displaced` cut-based approach on the MC sample. The distribution is fitted with a `RooDecay` function, and a lifetime of $\tau_{B_c} = 0.536 \pm 0.002$ ps is obtained.

Figure 5.9 compares the proper decay length distributions of B_c candidates for both gen-matching and displaced selection models. The gen-matching distribution follows an exponential decay shape, while the displaced cut-based approach selection significantly reduces the number of candidates lengths. The efficiency versus gen-ct is obtained as the ratio of the two distributions (bin-by-bin). In the bottom plot of Figure 5.9, the efficiency of the displaced selection is shown relatively to the gen-matching selection. The plateau starts where the efficiency reaches 95% of its maximum value (Eq. 5.3), which was found at $ct = 0.0223$ cm. At short decay lengths, the efficiency is lower, gradually increasing and stabilizing as it reaches this plateau.

After determining the plateau start, the gen-ct distribution was redrawn starting from this point, providing a clearer representation of the efficiency in

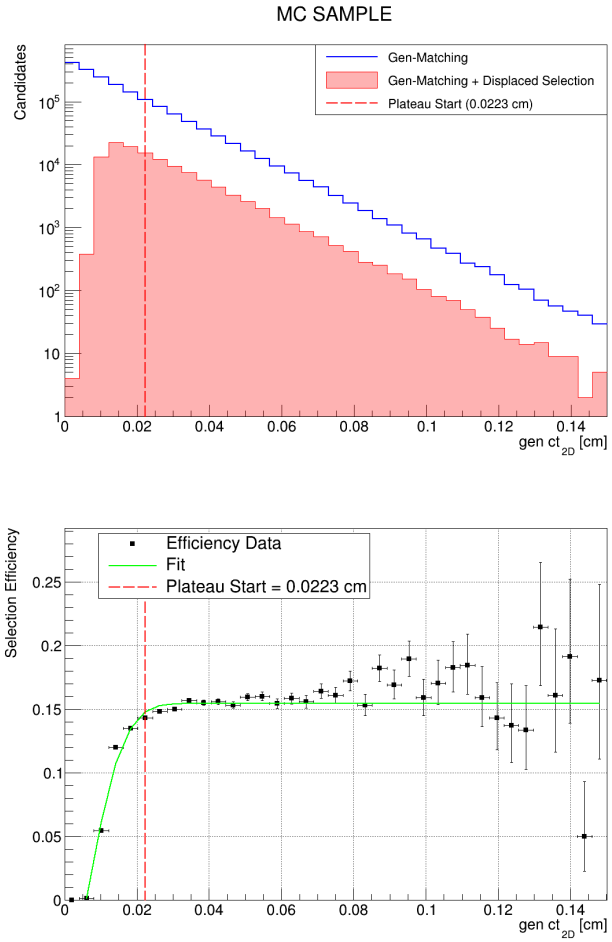


Figure 5.9: Comparison of the proper decay length distribution of B_c meson candidates for gen-matching (blue) and `displaced` approach (red). The top plot shows the candidate distribution, while the bottom plot depicts the selection efficiency as a function of proper decay length ct (see Eq. 5.1).

the stable region (Figure 5.10). The re-estimated lifetime, using the `RooExponential` function, is $\tau_{B_c} = 0.516 \pm 0.002$ ps.

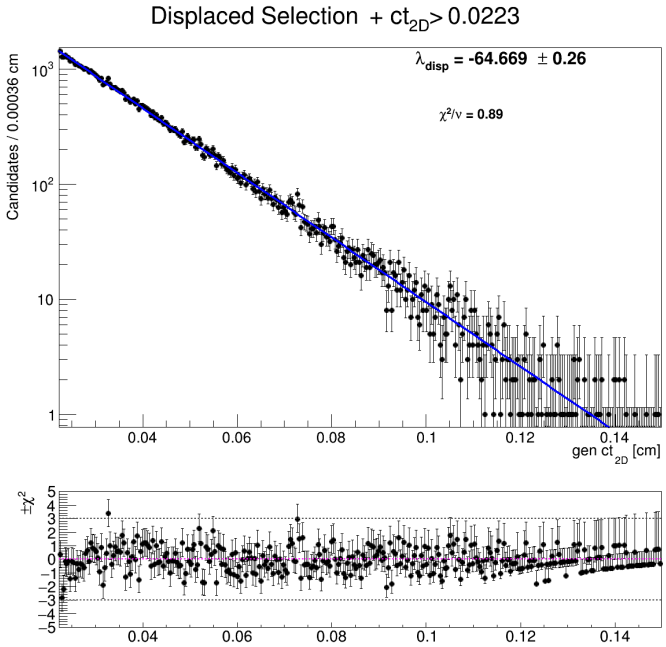


Figure 5.10: Proper decay length distribution of B_c meson candidates for `displaced` cut-basedapproach starting from the plateau point. The fit provides the decay length and corresponding lifetime.

5.1.4 Displaced `XGBoost`-based approach

The results of the `XGBoost displaced` approach are presented in Figure 5.11. The distribution is fitted using the `RooDecay` function, which models the exponential decay convoluted with a Gaussian resolution to incorporate detector effects. The obtained lifetime is $\tau_{B_c} = 0.582 \pm 0.001$ ps. These values are higher than expected, most likely due to the displaced selection method introducing additional biases in the decay length measurement. Notably, the fit does not work perfectly for small ct values.

In Figure 5.12, the top plot presents the proper decay length distribution for B_c meson candidates selected by both gen-matching and the `XGBoost displaced` approach. The gen-matching selection (blue) follows the expected exponential decay, while the `XGBoost displaced` approach (red) shows a noticeable reduction in candidates at shorter decay lengths, highlighting the selection's effect. The efficiency, shown in the bottom plot, was fitted using an error function model (Eq. 5.2), though the fit struggles, particularly in the tail region, indicating potential room for improvement. The plateau start, determined where the efficiency approximately stabilizes, was found at $ct = 0.0450$ cm, and is marked by the vertical dashed red line.

After identifying the start of the plateau, the gen-ct distribution for the `XGBoost displaced` approach was redrawn from this point in the plateau re-

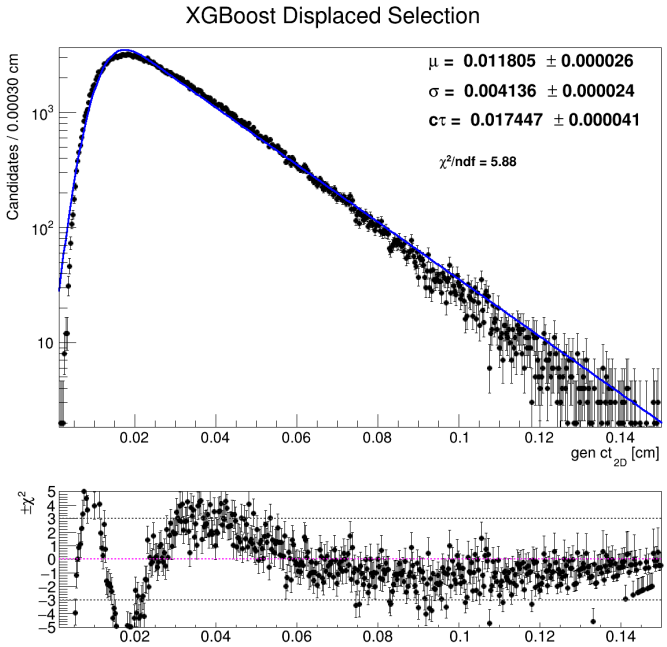


Figure 5.11: Proper decay length distribution of B_c meson candidates for `XGBoost displaced` approach on the MC sample. The distribution is fitted with a `RooDecay` function, yielding a lifetime of $\tau_{B_c} = 0.582 \pm 0.001$ ps.

gion (Figure 5.13). The refitted lifetime value is $\tau_{B_c} = 0.526 \pm 0.002$ ps, using as fit model is the `RooExponential` function.

The comparison between the cut-based `displaced` and `XGBoost displaced` approaches reveals distinct efficiency behaviors. The efficiency of the `displaced` selection stabilizes at around $ct = 0.0223$ cm, while the `XGBoost displaced` approaches the plateau later at $ct = 0.0450$ cm. This delay in reaching stabilization suggests that the `XGBoost displaced` model retains more candidates at short decay lengths but may introduce a larger systematic shift in the lifetime measurement.

In terms of efficiency, the `XGBoost displaced` approach shows a significant improvement, with a plateau efficiency reaching nearly 70%, compared to approximately 15% for the standard `displaced` selection. This increase in efficiency comes at the cost of a bias in the measured decay length. Specifically, the `XGBoost displaced` model yields a lifetime of $\tau_{B_c} = 0.526 \pm 0.002$ ps, while the `displaced` cut-based selection yields a $\tau_{B_c} = 0.515 \pm 0.002$ ps. By comparing these results with the reference B_c generation lifetime of $\tau_{gen}(B_c) = 0.507$ ps, it becomes clear that both approaches overestimate the lifetime. However, the bias is more pronounced in the `XGBoost displaced` selection, likely due to its higher efficiency and later stabilization point.

Thus, while the `XGBoost displaced` approach provides a more efficient

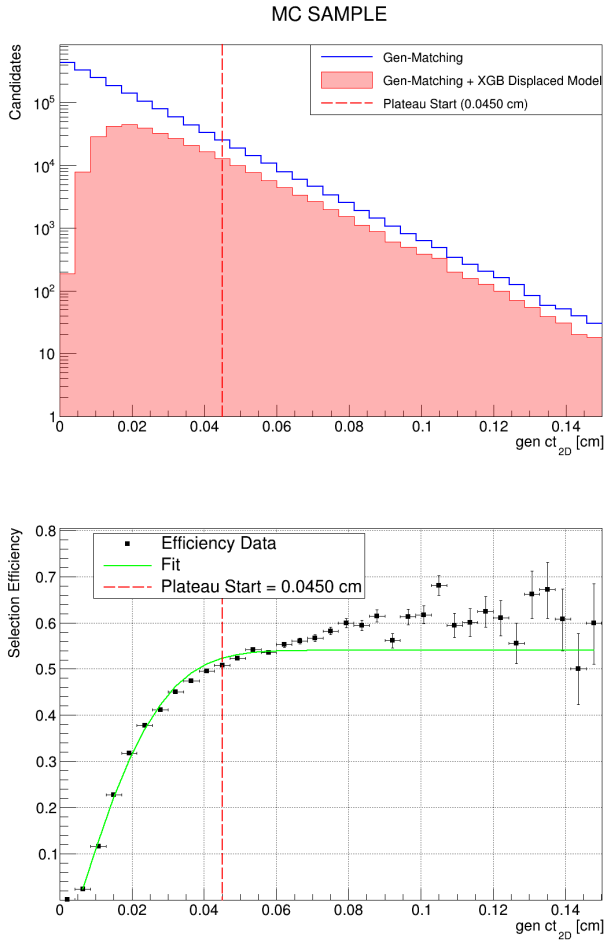


Figure 5.12: Comparison of the proper decay length distribution of B_c meson candidates for gen-matching (blue) and `XGBoost displaced` approach (red). The top plot shows the candidate distribution, while the bottom plot shows the selection efficiency as a function of proper decay length ct (see Eq. 5.1).

candidate selection, it also introduces a larger deviation from the expected lifetime, highlighting the trade-off between efficiency and distortion.

5.1.5 Summary table

It is generally observed that `XGBoost`-based approaches show larger efficiencies when compared to standard cut-based approaches. On the other hand, a non-neglibible distortion of the signal distribution is introduced especially at small ct values and the fit results from a `RooDecay`model are generally poor and lead to a shift of the extracted lifetime to higher values. This effect is mitigated by applying a cut-off at the turn-on of the selection efficiency. A summary of lifetimes for all the approaches discussed in this section is presented in Table 5.1.

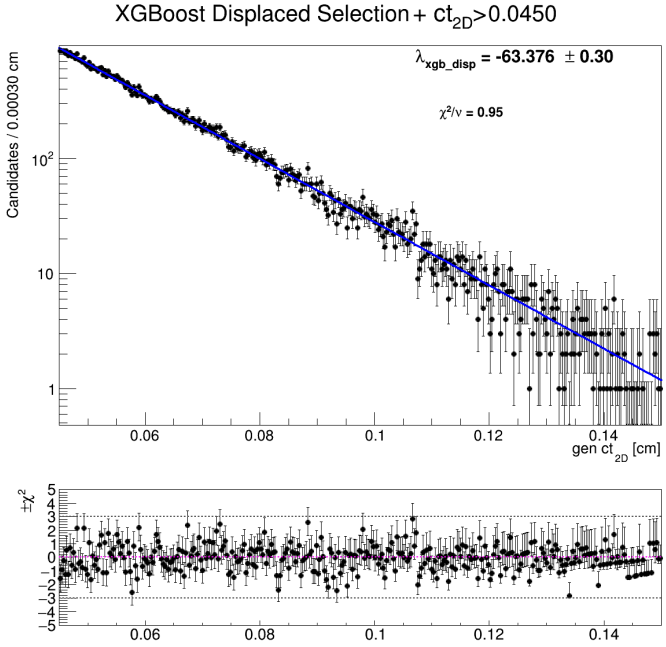


Figure 5.13: Proper decay length distribution of B_c meson candidates for `XGBoost displaced` approach starting from the plateau point. The fit provides the lifetime estimate $\tau_{B_c} = 0.526 \pm 0.002$ ps.

Selection Approach	Lifetime (ps)
No Selection (Reference)	0.4998 ± 0.0004
Inclusive Cut-Based Sel	0.514 ± 0.002
Inclusive Sel+($ct > 0.0147$ cm)	0.503 ± 0.002
XGB Inclusive	0.580 ± 0.002
XGB Inclusive Sel+($ct > 0.0440$ cm)	0.526 ± 0.003
Displaced Cut-Based Sel	0.536 ± 0.002
Displaced Sel+($ct > 0.0223$ cm)	0.515 ± 0.002
XGB Displaced	0.582 ± 0.001
XGB Displaced Sel+($ct > 0.0450$ cm)	0.526 ± 0.002

Table 5.1: Comparison of the lifetime values for B_c candidates obtained by fitting gen-ct for various selection approaches. For reference, the lifetime used in generation is $\tau_{gen}(B_c) = 0.507$ ps.

5.2 Reconstruction efficiency

Reconstructed level (reco-level) refers to the stage in Monte Carlo simulations where the generated particles are passed through the detector simulation, and they are reconstructed based on the detectors' response. This process accounts for the unavoidable experimental dilution (effects of detectors' resolution, smearing, inefficiencies ...), thus offering an actual representation of how the measurements would appear in a realistic experimental setting. In the following sections, the results for the reconstructed proper decay length *reco-ct* will be presented and analyzed.

In order to evaluate alone the effect of the calculation of the flight length using reconstructed information, the signal MC sample is used as it can provide both the reco-ct and the gen-ct. Each selection analyzed throughout the present thesis work is considered and the reco-ct distribution is compared against the gen-ct distribution later in this section. A reconstruction efficiency (ϵ_{reco}) can be defined as a ratio of the reco-ct distribution (N_{reco}) and gen-ct distribution (N_{gen}) in ct bins:

$$\epsilon_{reco} = \frac{N_{\text{reco}}}{N_{\text{gen}}} \tag{5.4}$$

The reconstructed ct distribution for B_c meson candidates passing the gen-matching filter is shown in Figure 5.14. The fit to the reconstructed ct distribution yields a lifetime of $\tau = 0.5041 \pm 0.0004$ ps, with the fit model being a `RooDecay` function. While a slight fluctuation is observed near $ct = 0$, the overall fit quality is excellent as shown by the bin-by-bin pulls.

Figure 5.15 shows the ratio for the reconstruction efficiency as defined in Equation 5.4. All gen-matched candidates are used at this stage, before any further selection: the numerator is the distribution of the flight length with respect to the reconstructed primary vertex (*reco-ct*), while the denominator is the distribution of the flight length with respect to the gen-level primary vertex (*gen-ct*). The ratio is well compatible with 1 (considering the uncertainties) and without appreciable shift effects, as should be expected in case of no biases.

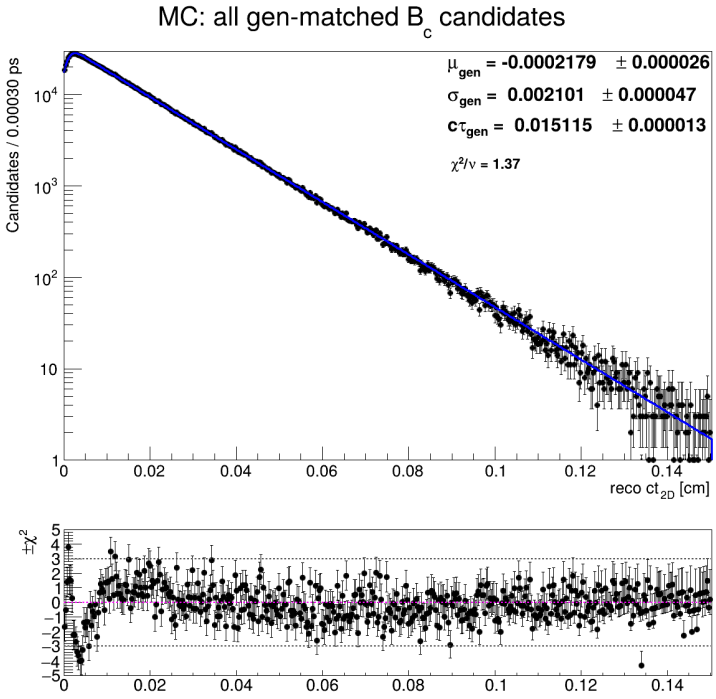


Figure 5.14: Distribution of the reconstructed ct for B_c meson candidates passing the gen-matching filter. The distribution is fitted with a `RooDecay` function, yielding a lifetime of $\tau = 0.5041 \pm 0.0004$ ps. The bottom plot shows the bin-by-bin pull distribution, indicating the goodness-of-fit level.

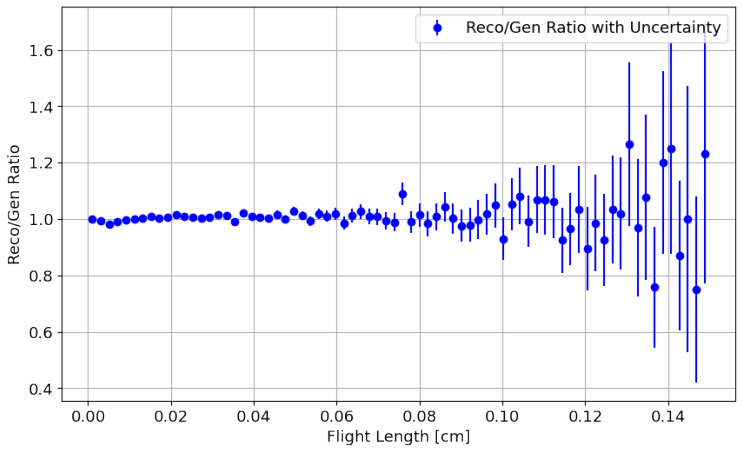


Figure 5.15: Ratio of reconstructed to generated ct distributions, once gen-matching has been applied. The reconstruction efficiency is stable and well compatible with 1 within fluctuations.

5.2.1 Inclusive cut-based approach

The proper decay length distribution for B_c candidates, selected using the `inclusive` approach on the MC sample, is presented in Figure 5.16. The distribution is fitted using the `RooDecay` function, which models an exponential decay convoluted with a Gaussian resolution to account for resolution effects, yielding a lifetime estimate of $\tau_{B_c} = 0.519 \pm 0.002$ ps.

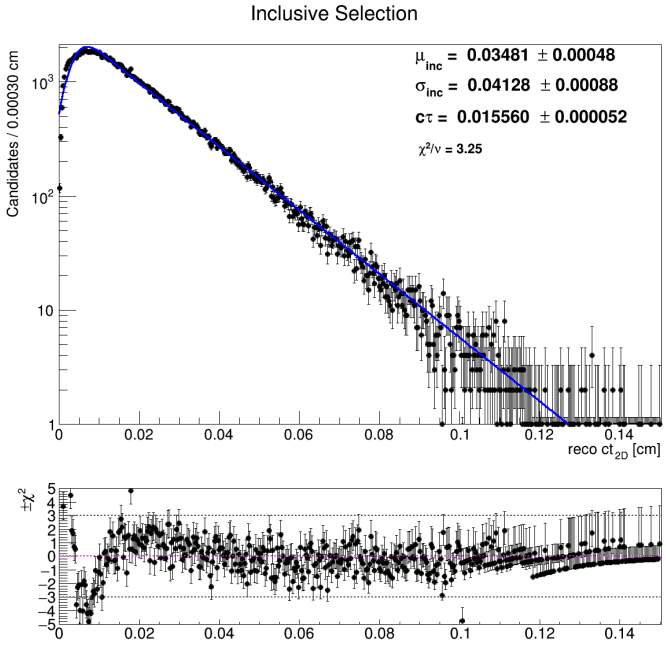


Figure 5.16: Proper flight distance distribution of B_c candidates after the `inclusive` approach on the MC sample, based on the selection criteria from Table 3.3. The distribution is fitted with a `RooDecay` function, yielding a lifetime of $\tau_{B_c} = 0.519 \pm 0.002$ ps.

Using the plateau value determined in the previous section on selection efficiency (Figure 5.3), the reco-ct distribution for the `inclusive` cut-based approach was redrawn starting from that point (a cut-off at $ct = 0.0147$ cm was applied). The fit to the resulting reco-ct distribution provides a lifetime of $\tau_{B_c} = 0.504 \pm 0.002$ ps, using a pure `RooExponential` function.

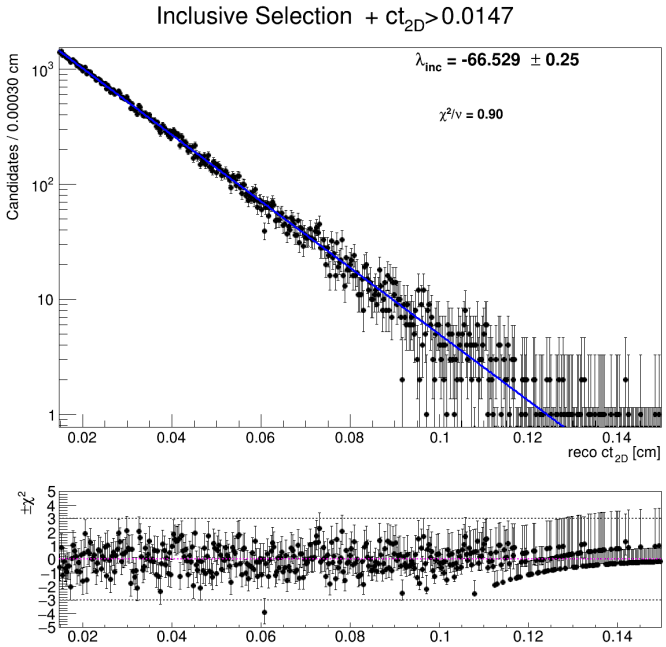


Figure 5.17: Proper decay length distribution of B_c meson candidates for `inclusive` approach in the plateau selection efficiency region. The fit provides the lifetime estimate $\tau_{B_c} = 0.504 \pm 0.002$ ps.

5.2.2 Inclusive XGB-based approach

The B_c meson candidates' proper decay length distribution, obtained using the `XGBoost inclusive` approach on the Monte Carlo sample, is displayed in Figure 5.18. The fit model assumes the pure `RooDecay` function and provides a lifetime o$\tau_{B_c} = 0.585 \pm 0.002$ ps.

The plateau start value from the previous section (Figure 5.6) was used to redraw the gen-ct distribution for the `XGBoost inclusive` approach, focusing on the plateau region only. After applying the `RooExponential` fit, an updated lifetime of $\tau_{B_c} = 0.524 \pm 0.003$ ps.

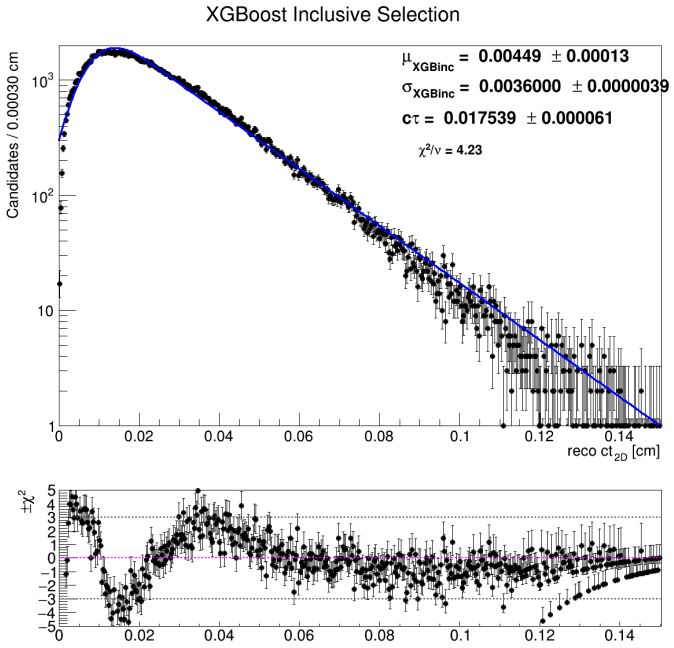


Figure 5.18: Proper decay length distribution of B_c meson candidates for `XGBoost inclusive` approach on the MC sample. The distribution is fitted with a `RooDecay` function, estimating a lifetime of $\tau_{B_c} = 0.585 \pm 0.002$ ps.

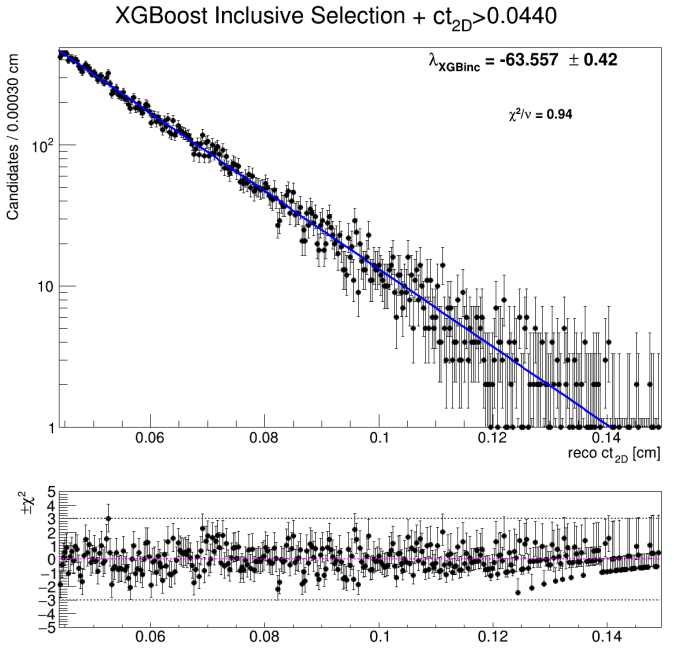


Figure 5.19: Proper decay length distribution of B_c meson candidates for `XGBoost inclusive` approach starting from the plateau point previously identified. The fit provides an estimate of the lifetime given by $\tau_{B_c} = 0.524 \pm 0.003$ ps.

5.2.3 Displaced cut-based approach

The ct distribution for B_c mesons selected using the `displaced` cut-based approach on the Monte Carlo sample is presented in Figure 5.20. This selection follows the criteria described in Table 3.3, where a specific cut of reco-ct > 0.01 cm has been applied for the `displaced` approach. A `RooDecay` function is used as fit model and the estimated lifetime is $\tau_{B_c} = 0.513 \pm 0.002$ ps.

After considering the start of the plateau borrowed by the previous study, the reco-ct distribution was redrawn from that point to provide a more accurate view of the efficiency in the plateau region (Figure 5.21). The updated fit yields a lifetime of $\tau_{B_c} = 0.518 \pm 0.002$ ps, using the `RooExponential` function to model the simple exponential decay.

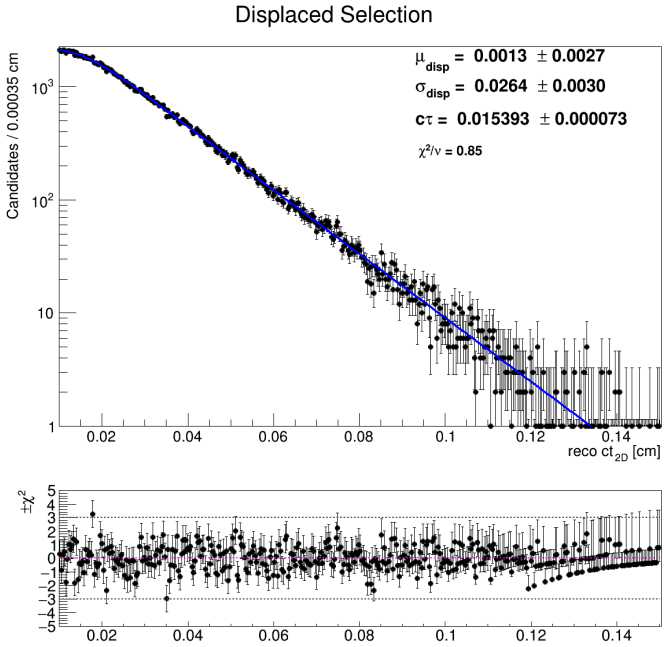


Figure 5.20: Proper flight distance distribution of B_c meson candidates after the `displaced` cut-based approach on the MC sample (based on the selection criteria from Table 3.3). The distribution doesn't start from zero because of the pre-existing cut (at HLT level). The distribution is fitted with a `RooDecay` function, yielding a lifetime estimate $\tau_{B_c} = 0.513 \pm 0.002$ ps.

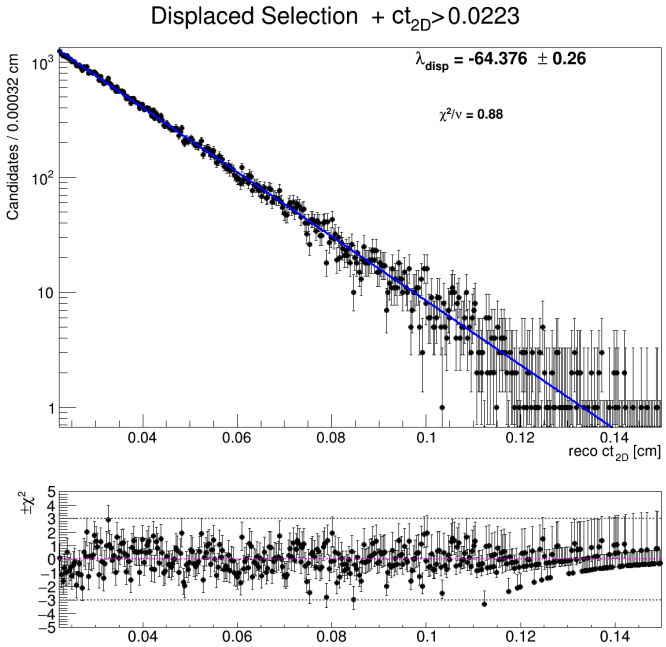


Figure 5.21: Proper decay length distribution of B_c meson candidates for `displaced` cut-based approach in the plateau region. The fit provides an estimate of the lifetime given by $\tau_{B_c} = 0.518 \pm 0.002$.

5.2.4 Displaced XGB-based approach

The proper decay length distribution of B_c meson candidates after applying the `XGBoost displaced` approach is shown in Figure 5.22. Using a `RooDecay`model, the estimated lifetime is $\tau_{B_c} = 0.585 \pm 0.001$ ps; the quality of this fit is poor because the distribution appears to be distorted at low ct values so that the `RooDecay`model turns out to be not adequate.

After borrowing from the previous study the plateau start point, the gen-ct distribution was redrawn from this point, giving a more accurate representation of the plateau region (Figure 5.23). The updated fit yields a lifetime $\tau_{B_c} = 0.523 \pm 0.002$ ps, using the `RooExponential` model.

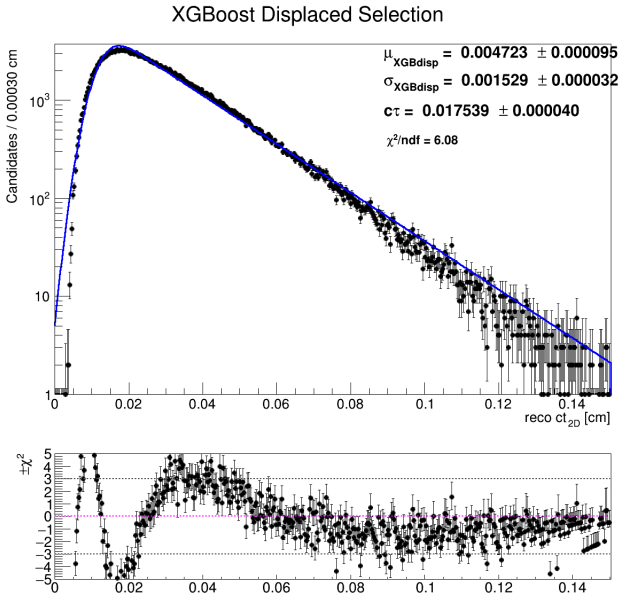


Figure 5.22: Proper decay length distribution of B_c meson candidates for `XGBoost displaced` approach on the MC sample. The distribution is fitted with a `RooDecay` function yielding a lifetime of $\tau_{B_c} = 0.585 \pm 0.001$ ps.

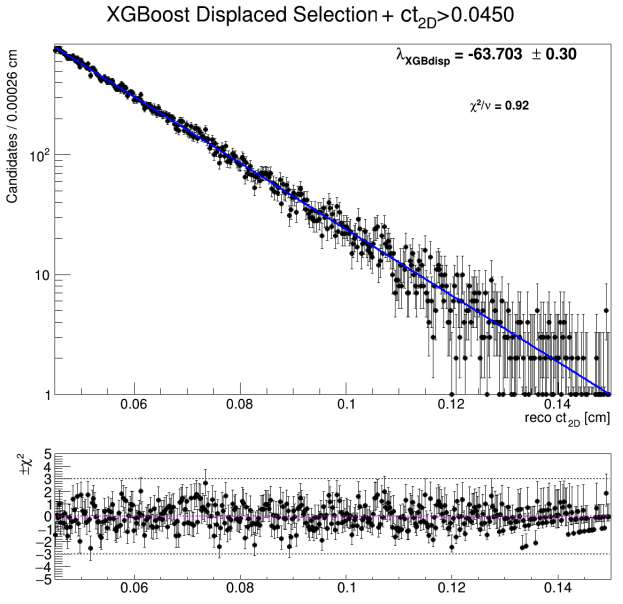


Figure 5.23: Proper decay length distribution of B_c meson candidates for `XGBoost displaced` approach starting from the plateau point. The fit provides the lifetime estimate $\tau_{B_c} = 0.523 \pm 0.002$ ps.

5.2.5 Summary table

The Table 5.2 summarizes the fit results obtained for the different considered selection approaches after reconstruction. The first row is a reference for the application of the several approaches and also measures how much the estimated lifetime after reconstruction (and before any selection) differs from the estimated lifetime using at generation level (given in the first raw of Table 5.1).

Selection Approach	**Lifetime (ps)**
No Selection (Reference)	0.5041 ± 0.0004
Inclusive Cut-Based Sel	0.519 ± 0.002
Inclusive Sel+($ct > 0.0147$ cm)	0.504 ± 0.002
XGB Inclusive	0.585 ± 0.002
XGB Inclusive Sel+($ct > 0.0440$ cm)	0.524 ± 0.003
Displaced Cut-Based Sel	0.513 ± 0.002
Displaced Sel+($ct > 0.0223$ cm)	0.518 ± 0.002
XGB Displaced	0.585 ± 0.001
XGB Displaced Sel+($ct > 0.0450$ cm)	0.523 ± 0.002

Table 5.2: Comparison of the lifetime values for B_c candidates estimated by fitting the reco-ct distributions obtained for the considered different selection approaches.

5.3 Correction of the measured $c\tau$ values

The studies carried out on a high-statistics signal MC sample in the previous two sections can allow to determine an overall correction factor as well as its factorizable contributions associated to selection and reconstruction. The overall correction factor is meant to be be applied to the measured lifetimes reported in the Table 5.2 and check the compatibility with generation lifetime value. After applying the correction to the fit-based estimated lifetime, the corrected lifetime value is expected to be enough close to the lifetime value used at generator level if the biases introduced by the reconstruction and selection chains can be recovered by this overall correction method. It is worthy to stress here that the correction is not obtained by weighting *ct* distributions but consists in a less sophisticated approach, namely in applying an overall correction factor to the results from the fits to the (unweighted) distributions.

This overall correction factor, labelled as k_{corr}, can be imagined to be factorized in three contributions k_{gen}, k_{reco} and k_{sel} according to this formula

$$k_{corr} = k_{gen} \cdot k_{reco} \cdot k^*_{sel} \tag{5.5}$$

where each factor is defined as follows:

- k_{gen} : it accounts for the difference between the lifetime value used in generation ($\tau^{\rm MC}_{gen} = 0.507$ ps) and the value effectively extracted from the simulation by fitting the gen-ct distribution before any kinematical selection (apart from the *gen-matching*, that ensures the purity of the extracted signal sample):

  $$k_{gen} = \frac{\tau^{\text{no sel}}_{gen}}{\tau^{\rm MC}_{gen}} \tag{5.6}$$

  In other words it represents a correction for the bias introduced by the methodology used for the lifetime estimation. In this specific case: $k_{gen} = 0.9858 \pm 0.0004$, being $\tau^{\text{no sel}}_{gen} = 0.4998 \pm 0.0004$ ps taken from the first raw of Table 5.1.

- k_{reco} : it corrects for the reconstruction effects being calculated as the ratio of the τ value extracted by fitting the reco-ct distribution before any kinematical selection and the value effectively extracted from the simulation by fitting the gen-ct distribution before any kinematical selection:

  $$k_{reco} = \frac{\tau^{\text{no sel}}_{reco}}{\tau^{\text{no sel}}_{gen}} \tag{5.7}$$

  In this specific case: $k_{reco} = 1.0086 \pm 0.0011$, being $\tau^{\text{no sel}}_{reco} = 0.5041 \pm 0.0004$ ps taken from the first row of Table 5.2. Note that $k_{gen} \cdot k_{reco}$ is very close to unity (within uncertainties).

- k^*_{sel} : it corrects, for each adopted selection approach (that's why the $*$ label), for the distortion introduced to the *ct* distributions by the application of the selection criteria (either cut-based or XGBoost-based):

$$k^*_{sel} = \frac{\tau^{\text{sel}*}_{reco}}{\tau^{\text{no sel}}_{reco}} \tag{5.8}$$

  where the values of $\tau^{\text{sel}*}_{reco} \equiv \tau^*_{meas}$ are taken from Table 5.2 (all the rows except the first).

Therefore the overall correction factors can be coherently expressed as

$$k^*_{corr} = k_{gen} \cdot k_{reco} \cdot k^*_{sel} = \frac{\tau^{\text{no sel}}_{gen}}{\tau^{\text{MC}}_{gen}} \cdot \frac{\tau^{\text{no sel}}_{reco}}{\tau^{\text{no sel}}_{gen}} \cdot \frac{\tau^{\text{sel}*}_{reco}}{\tau^{\text{no sel}}_{reco}} = \frac{\tau^{\text{sel}*}_{reco}}{\tau^{\text{MC}}_{gen}} \tag{5.9}$$

Ideally $\tau^{\text{sel}*}_{reco} \equiv \tau^*_{meas}$ should be equal to τ^{MC}_{gen} so that $k_{corr} = 1$ (i.e. no correction is needed at all). Practically τ^*_{meas} should be corrected through the k^*_{corr} factors in order to get the obtained τ^*_{corr} all equal to τ^{MC}_{gen}. To formalize this:

$$\tau^*_{corr} = \frac{\tau^*_{meas}}{k^*_{corr}} \equiv \frac{\tau^{\text{sel}*}_{reco}}{k^*_{corr}} \tag{5.10}$$

Thus the corrected lifetime (τ_{corr}) - for each type of selection - is obtained by multiplying the measured τ_{meas} by the inverse of the overall correction factor k_{corr} which is the product of 3 correction factors k_{gen}, k_{reco}, and k_{sel} (Eq. 5.9), that account for the corresponding biases introduced by effects of generation, reconstruction, and selection. Table 5.3 reports, for each selection approach, the k^*_{sel} and k^*_{corr} values; the latter corrections factors can be then applied to correct the lifetime values measured in data (see next section).

Selection Approach	k^*_{sel}	k^*_{corr}
Inc. cut-based Sel.	1.0285 ± 0.0036	1.0236 ± 0.0034
Inc. c-b Sel. on plateau	0.9987 ± 0.0392	0.9932 ± 0.0038
XGB Inc. Sel.	1.1603 ± 0.0042	1.1539 ± 0.0040
XGB Inc. Sel. on plateau	1.0410 ± 0.0069	1.0351 ± 0.0068
Disp. cut-based Sel.	1.0713 ± 0.0034	1.0127 ± 0.0048
Disp. c-b Sel. on plateau	1.0320 ± 0.0042	1.0219 ± 0.0041
XGB Disp. Sel.	1.1644 ± 0.0029	1.1539 ± 0.0026
XGB Disp. Sel. on plateau	1.0386 ± 0.0050	1.0328 ± 0.0049

Table 5.3: Summary of k^*_{sel} and k^*_{corr} values. The word "on plateau" indicates the presence of an additional cut (*ct* values greater than a threshold value, specific to each selection) so that the selection efficiency versus *ct* gets stable (see Sec. 5.1). The correction $k_{gen} \cdot k_{reco}$, common to all cases, is not displayed but used to multiply k^*_{sel} values to get k^*_{corr} values.

5.3.1 Additional test on corrections' factorization

In addition to the selection correction computed at the reconstruction level (k^*_{sel}), a correction factor related to the selection effects can be also defined using generated quantities instead of reconstructed ones. The focus is set on the B_c production vertex, which is known precisely in Monte Carlo sample (gen-PV), while in real data needs to be determined amongst the many reconstructed primary vertices (reco-PV). Therefore, this generator-level selection correction factor, denoted as k^{gen*}_{sel}, can be defined, similarly to k^*_{sel} but by using generator-level quantities, as follows:

$$k^{gen*}_{sel} = \frac{\tau^{\text{sel}*}_{gen}}{\tau^{\text{no sel}}_{gen}} \tag{5.11}$$

where $\tau^{\text{sel}*}_{gen}$ and $\tau^{\text{no sel}}_{gen}$ refer to the lifetimes extracted from the generator-level signal with and without the application of selection criteria, respectively. Table 5.4 compares the latter selection correction factors computed at generator level, $k^{\text{gen}*}_{sel}$, with those obtained at reconstruction level, as reported in the first column of Table 5.3 for k^*_{sel}. The two sets of correction factors are in good agreement with each other, indicating a consistent behavior of the selection effect across both levels (generation and reconstruction).

The resulting correction factor k'^*_{corr} can be applied to the measured τ^*_{meas} values using the equation:

$$\tau'^*_{\text{corr}} = \frac{1}{k_{\text{gen}} \cdot k_{\text{reco}} \cdot k^{gen*}_{sel}} \cdot \tau^*_{meas} = \frac{\tau^*_{meas}}{k^{\text{gen}*}_{corr}} \tag{5.12}$$

Selection Approach	k^*_{sel}	k^{gen*}_{sel}
Inc. cut-based Sel.	1.0285 ± 0.0036	1.0285 ± 0.0036
Inc. c-b Sel. on plateau	0.9987 ± 0.0392	1.0070 ± 0.0038
XGB Inc. Sel.	1.1603 ± 0.0042	1.1603 ± 0.0042
XGB Inc. Sel. on plateau	1.0410 ± 0.0069	1.0529 ± 0.0069
Disp. cut-based Sel.	1.0713 ± 0.0034	1.0713 ± 0.0034
Disp. c-b Sel. on plateau	1.0320 ± 0.0042	1.0319 ± 0.0043
XGB Disp. Sel.	1.1644 ± 0.0029	1.1643 ± 0.0028
XGB Disp. Sel. on plateau	1.0386 ± 0.0050	1.0530 ± 0.0049

Table 5.4: Comparison between the selection correction factors derived at generator level (k^{gen*}_{sel}) and at reconstruction level (k^{reco*}_{sel}).

Selection Approach	τ^*_{meas} **(ps)**	$k^{\mathbf{gen*}}_{corr}$	$\tau'^{\mathbf{*}}_{\mathbf{corr}}$ **(ps)**
Inc. cut-based Sel.	0.519 ± 0.002	1.0227 ± 0.0069	0.5075 ± 0.0025
Inc. c-b Sel. on plateau	0.504 ± 0.002	1.0014 ± 0.0068	0.5028 ± 0.0028
XGB Inc. Sel.	0.585 ± 0.002	1.1538 ± 0.0078	0.5070 ± 0.0026
XGB Inc. Sel. on plateau	0.524 ± 0.003	1.0471 ± 0.0071	0.5012 ± 0.0047
Disp. cut-based Sel.	0.513 ± 0.002	1.0654 ± 0.0072	0.4820 ± 0.0028
Disp. c-b Sel. on plateau	0.518 ± 0.002	1.0262 ± 0.0069	0.5049 ± 0.0030
XGB Disp. Sel.	0.585 ± 0.002	1.1579 ± 0.0078	0.5053 ± 0.0018
XGB Disp. Sel. on plateau	0.523 ± 0.002	1.0472 ± 0.0071	0.5000 ± 0.0034

Table 5.5: The overall correction factor $k^{\mathrm{gen}*}_{corr}$ (middle column), computed using $k^{\mathrm{gen}*}_{sel}$ reported in Table 5.4, is used to correct the measured lifetime τ^*_{meas} (left column) thus obtaining - for each selection - the τ'^*_{corr} values given in right column.

This correction factor for generator-level selection is found to be compatible with the values obtained from reconstruction-level corrections, suggesting that the selection effect is independent of the choice of primary vertex. The corrected lifetime values (τ'^*_{corr}) are presented in Table 5.5.

The correction factor for generator-level selection, along with the precomputed k_{gen} and k_{reco}, has been combined to yield the overall correction factor $k^{\mathrm{gen}*}_{corr}$ reported in Table 5.5. This table presents the measured τ^*_{meas} values and the corrected τ'^*_{corr} values after using the $k^{\mathrm{gen}*}_{corr}$ correction for each selection. The resulting τ'^*_{corr} values are closely aligned or are almost compatible with the generated reference value ($\tau^{\mathrm{MC}}_{gen} = 0.507$ ps), thus indicating that the correction factors can effectively account for systematic biases (i.e. shifts) introduced - in particular - during the application of the selection method.

These results demonstrate that the selection correction computed at generator level remains compatible with the exact reconstruction-level correction (k^*_{corr}) (computed assuming factorization of reconstruction and selection effects), thus suggesting that the selection effects can be considered, with good approximation, independent from the way the primary vertex is reconstructed/chosen, and consequently that the selection and crucial reconstruction effects effectively factorize.

5.4 Lifetime Measurement for Optimized `XGBoost inclusive` Approach

In this section, a very preliminary lifetime measurement of the B_c^+ meson is presented, utilizing an optimized `XGBoost inclusive` approach applied to the full dataset of 68.1 fb^{-1} of pp collisions collected by CMS during 2022 and 2023. An important goal of this thesis is to explore the performance of the optimized `XGBoost inclusive` approach in providing the proper decay length distribution of the B_c^+ meson from which to attemot the lifetime estimation. In this section most of the study focuses on considering both the signal and background contributions in real data and disentangling them.

The choice to focus on the `XGBoost` inclusive approach, rather than pursuing the `XGBoost` displaced approach, stems from several considerations. First, it was observed that the `XGBoost` displaced approach creates significant distortion in the proper decay length (ct) distribution, particularly in the region where ct is close to zero. Additionally, the `XGBoost displaced` approach involves more restrictive selection criteria that further reduce the candidate pool, introducing a bias that can skew the final measurements, especially for the subsample of the "short-lived" B_c^+ candidates. As a result, the `inclusive` approach was chosen for its broader selection efficiency and consequent ability to retain more candidates, thus providing the conditions for a more reliable measurement of the B_c^+ meson lifetime.

As discussed in the Chapter 4, the `XGBoost inclusive` approach has been trained using a variety of discriminant features, and its selection performance has already been checked through invariant mass fits. In this section, the analysis is extended to include the measurement of the B_c^+ meson proper decay length in both the signal region and the background-dominated sidebands (control region).

The final result of this study is derived from the 1D fit of the ct distribution, which should provide a robust measurement of the B_c^+ meson lifetime. The application of optimized `XGBoost inclusive` approach is expected to allow a precise lifetime measurements by relying the model's enhanced background rejection capabilities. The sideband and signal region fits will be presented and discussed in detail in the nextcoming part of the current section.

5.4.1 Flight length distribution in control region

The invariant mass plot obtained by the selected `XGBoost inclusive` model with optimized working point is shown in Figure 5.24 together with the two distinct regions highlighted : the *signal window* and *sidebands*. In this analysis, the signal window is defined as the region within 2σ around the signal peak

(dashed fuchsia), while the sidebands encompass the broader region extending out of the 3σ window (dashed light blue). These regions are used to separate the signal from the background.

The fit components are modeled using a Gaussian function for the signal based. The background is modeled using an exponential function, which provides a robust approximation for an overwhelming combinatorial background. The results of the fit provide a signal yield of $N_{\text{sig}} = 7235 \pm 197$ (and background yield $N_{\text{bkg}} = 65502 \pm 312$); the B_c mass fit estimate ($\mu = 6271.41 \pm 0.72$ MeV) is consistent with the expected (PDG) B_c meson mass (6274.47 ± 0.22 MeV) whereas the width ($\sigma = 26.65 \pm 0.76$ MeV) is on the low side of the mass resolution ($\sigma_{res} = 30.9 \pm 0.2$ MeV) taken from the simulated signal (but estimated by fitting the mass spectrum without applying any selection, see Figure 3.2).

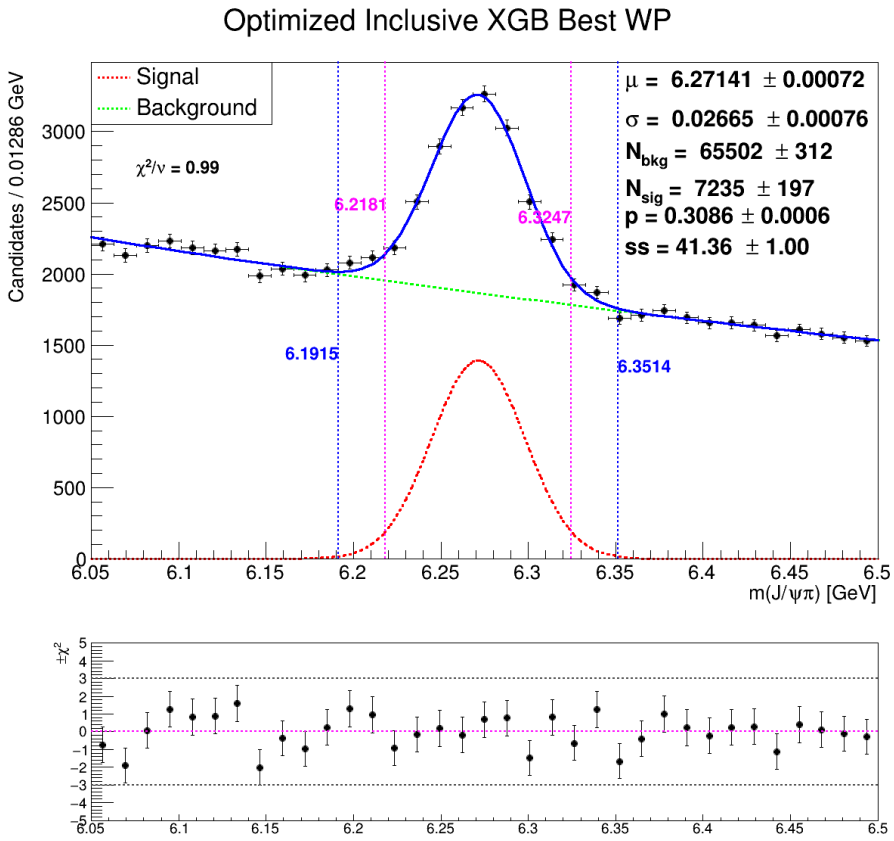


Figure 5.24: Invariant mass distribution for the optimized `XGBoost inclusive` approach. The signal 2σ-window (within dashed fuchsia lines) and sidebands (outside the dashed light blue lines displaying the 3σ-window) are indicated. The fit includes a Gaussian signal model (dotted red) and an exponential function for the background (dotted green).

The defined sidebands are used to further analyze the background distribution in a broader region assuming the sidebands' candidates to have the same characteristics of the background in the signal region. Specifically we tend to assume that we can deduce the proper flight length distribution of a background candidate in the signal region from that of sidebands' candidates, as discussed in detail below. In order to understand the contribution of the background to the proper flight length distribution, the sidebands are further divided into left and right regions (LSB and RSB), corresponding to the areas below (Left) and above (Right) with respect to the 3σ window.

Figure 5.25 shows the proper decay length distributions for the left (lepf plot) and right (right plot) sidebands separately. A fit is performed on each sidebands region to model the prompt and non-prompt background contributions: the common fit model uses two `RooDecay` functions to deal separately with the contributions from prompt and non-prompt background components. The `RooDecay` function is suitable to represent either a pure combinatorial background (prompt) and a non-prompt combinatorial contribution: simply the (exponential) tails will look very different to account for their different feature. It is worthy to explicitly note that the non-prompt contribution is the most dangerous one since it mimics a genuine signal contribution. These fits are crucial for determining the background model when later fitting the signal region, where both prompt and non-prompt backgrounds will contribute alongside the signal.

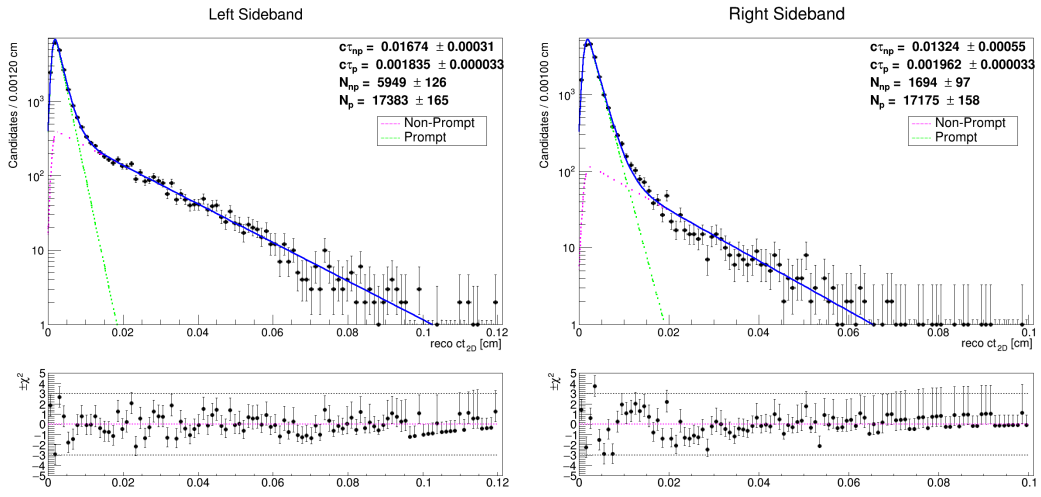


Figure 5.25: Proper decay length distribution for the left (left plot) and right (right plot) sidebands. The fits include a prompt background contribution (green) and a non-prompt background one (magenta).

The overall background behaviour within the signal region is expected to be intermediate between the shapes and compositions observed separately in the left and right sidebands. Comparing the two fits in Figure 5.25 it doesn't go unnoticed that in the left(right) sideband the non-prompt component amounts to $\sim 25\%(\sim 9\%)$ of the total background. This may hint a hidden non-prompt component representing the physical background associate to partially reconstructed B_c meson decays.

Additionally an overall fit is then performed on the entire sidebands' region, namely including both the left and right sidebands. Figure 5.26 shows the proper decay length distribution for the combined left and right sidebands. Each one of the prompt and non-prompt backgrounds are modeled with a `RooDecay`, similarly to the fits performed on the individual sidebands. The results from this fit will be used to model the background in the signal window when performing the fit over the candidates of the signal region. The information extracted from this fit and used late concerns the non-prompt and prompt "lifetime" parameters of the background and the relative proportion between these two components (the non-prompt component is the $\sim 18\%$ of the total).

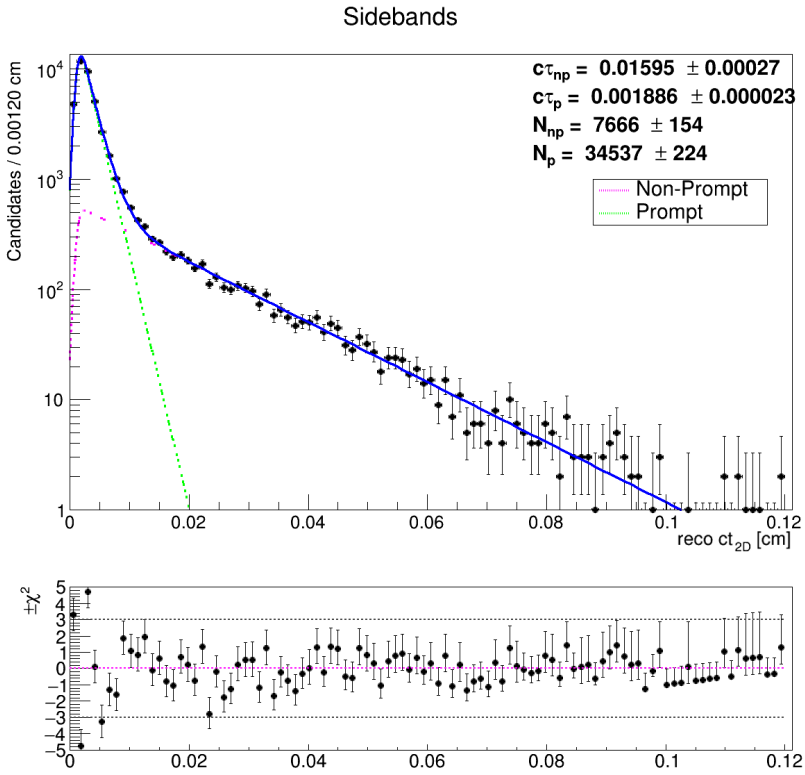


Figure 5.26: Proper decay length distribution for the combined sidebands region. The fit includes contributions from prompt background (green) and non-prompt background (magenta), each of the two modeled by means of a `RooDecay` fucntion.

5.4.2 Flight length distribution in signal region

After modeling the background using the combined sidebands fit, the focus shifts to the signal window, which is, as already said, defined within the 2σ region around the B_c meson invariant mass mode. The goal is to extract the lifetime information, while simultaneously accounting for signal yield and the background contributions.

In the signal window, both the prompt and non-prompt backgrounds are modeled based on the sidebands fit results, ensuring that the same background ct shape is considered when fitting the ct distribution within the signal window. Additionally, the signal yield is fixed to the number of signal events determined from the invariant mass fit within the 2σ signal region (see Fig. 5.24). This approach leverages the information obtained from both the sidebands fit and the invariant mass analysis to provide a realistic description of the background and accurately constrain the signal yield within the signal region. The signal itself is modeled using a `RooDecay` function to represent the exponential decay characteristic of B_c mesons, convoluted with a Gaussian resolution function to account for resolution effects. In this fit, the prompt and non-prompt background fractions are fixed according to the sidebands fit, while the signal lifetime is left as a free parameter to be determined from the data. The ratio between signal yield and background yield is derived from the fit to the invariant mass distribution and applied within the signal window.

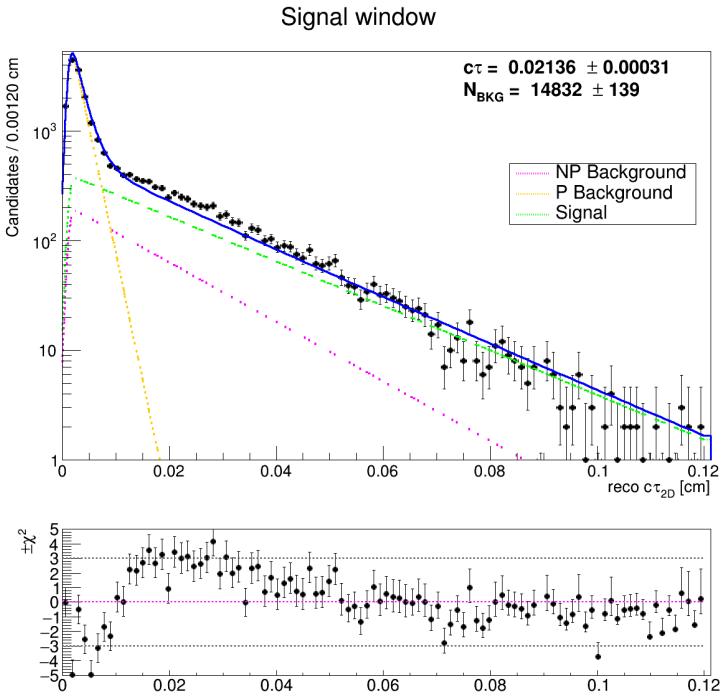


Figure 5.27: Proper decay length distribution for the signal window. The fit includes the B_c signal (green) and the prompt (yellow) and non-prompt (magenta) background contributions based on the sidebands fit.

Figure 5.27 presents the total fit in the signal region (solid blue curve). The signal component (dashed green) is clearly visible, alongside the prompt (dashed yellow) and non-prompt (dashed magenta) background contributions. The fit provides an estimated lifetime $\tau_{B_c} = 0.713 \pm 0.010$ ps and suggests to try to perform the fit in the efficiency plateau region, as discussed below.

5.4.3 Preliminary extraction of the B_c lifetime

The previous fit strategy is then repeated considering a further selection, namely $ct > 0.440$ cm, which ensures a relatively high and stable signal efficiency together with a strong rejection of the prompt background. After applying the $ct > 0.440$ cm cut, the resulting invariant mass distribution in Figure 5.28 shows a significant improvement in signal purity compared to Figure 5.24.

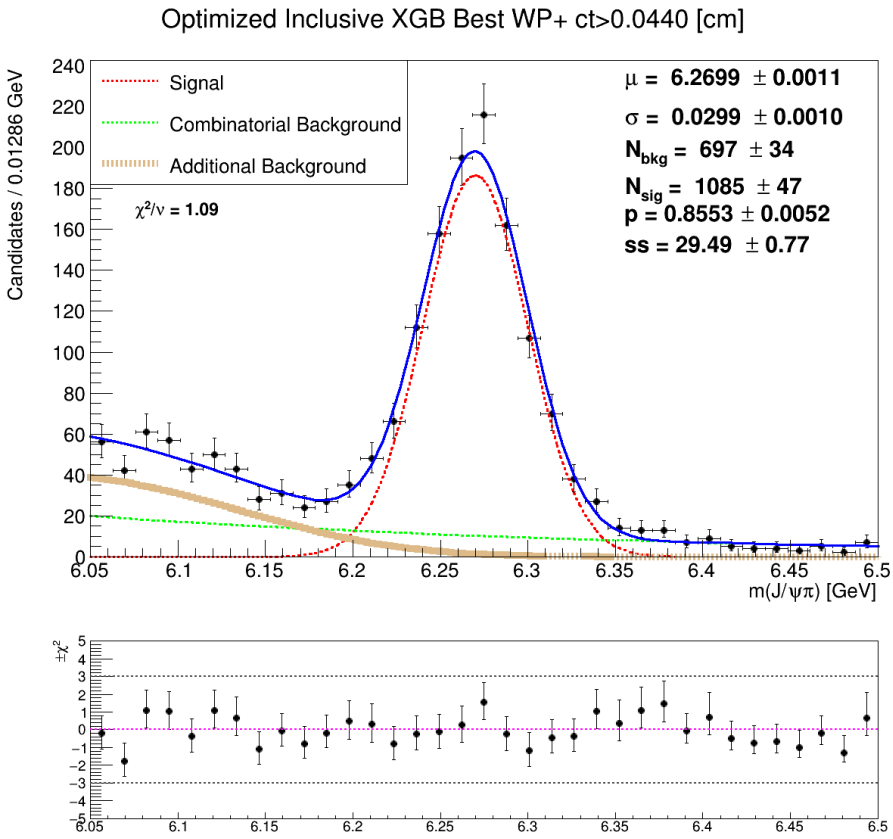


Figure 5.28: Invariant mass distribution for the optimized `XGBoost inclusive` approach with the additional requirement $ct > 0.440$ cm. The B_c signal is modeled using a Gaussian function (red dotted), while the background is modeled as a combination of an exponential function for the combinatorial background (green dotted) and an error function for the additional background (thick brown line).

In the fit of the mass spectrum of Figure 5.28 the B_c^+ signal is modeled using a Gaussian function, while two background contributions are now considered: a combinatorial background modeled by an exponential function and an additional background that accounts for partially reconstructed decays (e.g. $B_c \to J/\psi\pi + X$, where X is a non-reconstructed state) that is modeled by an error function.

The proper decay length distribution in Figure 5.29 shows the result of an Extended Unbinned Maximum Likelihood fit after applying the $ct > 0.0440$ cm cut for the candidates in the signal window. This cut completely kills the prompt background. The remaining Non-Prompt (NP) background and the signal are separately modeled using `RooDecay` functions. The signal yield is fixed to the number of signal events determined from the invariant mass fit within the 2σ signal region (see Fig. 5.28). The fit provides a lifetime of $\tau_B = 0.5375 \pm 0.0197$ ps.

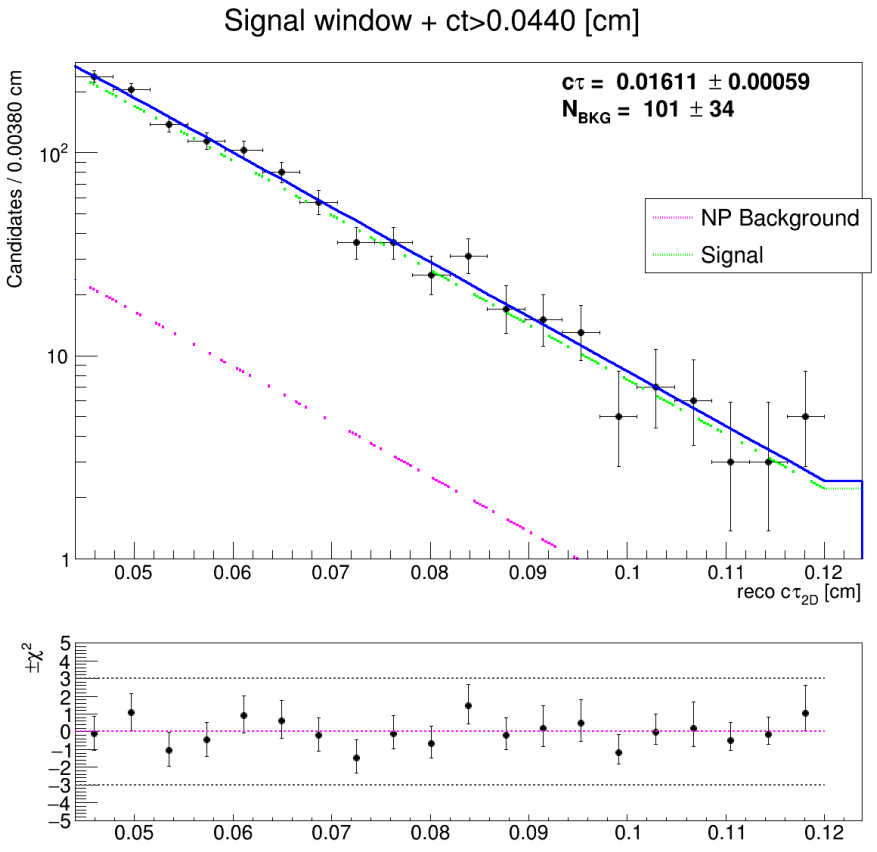


Figure 5.29: Proper decay length distribution for the signal window with $ct > 0.0440$ cm. Both the signal (green) and the non-prompt (NP) background (magenta) are modeled using a `RooDecay` function to capture their exponential behavior. The prompt background contribution is no longer present because of the cut. The fit provides the lifetime estimate $\tau_{B_c} = 0.5375 \pm 0.0197$ ps.

The final fit result is then corrected by the overall correction factor k^*_{corr} (specifically the one reported in the fourth row of Table 5.3) according to Eq. 5.10; the correction procedure yields a lifetime $\tau_{corr} = 0.5203 \pm 0.0193$ ps. This result is in fair agreement with the value reported by the PDG [1] (and dominated by the LHCb result), i.e. $\tau^{PDG}_{B_c} = 0.510 \pm 0.009$ ps, and it is also well compatible with the value earlier measured by CMS (with $\sqrt{s} = 8$ TeV data) [69], $\tau^{\text{CMS}}_{B_c} = [0.5414 \pm 0.0260(\text{stat}) \pm 0.014(\text{syst})]$ ps.

The presented result is well far from being conclusive since a thorough analysis will require a better understanding of many aspects of the B_c triggering and reconstruction effects, together with the full assessment of several sources of systematic uncertainties affecting the measurement. Indeed, the presented work aims to collect many tools to validate the usage of Machine Learning in High Energy Physics and provide strategies for a better understanding of its aspects, by studying the extraction of the B_c signal as a realistic use-case of interest for the CMS B-Physics strategy for Run-3 and beyond.

Conclusions

The physics of B_c mesons started to be explored at the Tevatron but it is at the LHC that can get a relevant boost. Due to the relatively very low production cross section of B_c mesons, most of the rich spectrum of the $\bar{b}c$ system is still unexplored. Below the $B^{(*)}D^{(*)}$ thresholds, the excited states are expected to start decay chains characterized by dipion and/or photon-emission transitions till reaching the ground state (the B_c meson) which decays weakly. In CMS, because of trigger constraints, the relatively cleaner, more copious and suitable decay to be reconstructed is the Cabibbo-favoured $B_c^{\pm} \to J/\psi\,\pi^{\pm}$. The extraction of a sample of $B_c^{\pm} \to J/\psi\,\pi^{\pm}$ candidates, as large and clean as possible, is crucial both to explore the excited states of the $\bar{b}c$ spectrum and to search for the double beauty hadrons (the conventional expected baryons and possibly the exotic double-heavy tetraquarks). For the former investigation, a selection using B_c reconstructed candidates well displaced from the production vertex is more suitable, whereas, for the latter search, an `inclusive` (i.e. without any displacement requirement) selection of B_c reconstructed candidates is preferred to maximize the signal and minimize any bias on the proper decay time distribution. For this reasons, the study of the extraction of the B_c signal is carried out in both configurations, namely with two different topologies (`inclusive` and `displaced`), throughout the whole thesis, thus allowing, at each step, a constant comparison and highlighting of the differences. This study is now possible at CMS with Run-3 data thanks to the development of a new trigger dedicated to the `inclusive` selection of $B_c^{\pm} \to J/\psi\,\pi^{\pm}$ sample; previous results were instead based on `displaced JpsiTrk` trigger 2.5.

The extraction of the $B_c^{\pm} \to J/\psi\,\pi^{\pm}$ signal is first carried out with a standard cut-based selection guided by experience and without any particular optimization; this is used as reference for a trained `XGBoost`-based (BDT-like) selection. In the `inclusive` topology any displacement variable is excluded in the training, in which signal-like input is taken by a signal Monte Carlo sample and background-like input is instead data-driven by exploiting the a dedicated control region. A relevant gain performance in the signal selection is introduced by this Machine Learning (ML) technique and it is quantified in terms of comparison of the cut-based Working Point with the `XGBoost` ROC

curve (Figures 3.12 and 3.20) and furthermore in terms of standard FoMs like signal significance and purity (Table 3.7). The whole software machinery is developed within a python-based Jupyter Notebook and, together with the obtained results, it can be used in the future, in an advanced data analysis teaching course, as a complete and standalone example of the gain introduced by a ML approach in the extraction of a physical signal (nowadays `XGBoost` is becoming a standard tool in the HEP context).

The optimization of the `XGBoost` hyperparameters is then carried out using a cutting-edge python-based open software tool, the `Optuna` library, which is not commonly used in the HEP context yet, but it is already well developed and mature and looks very promising for HEP tasks. From the computational time point of view, `Optuna` results to be more powerful with respect to a traditional Grid Search, especially when using its TPE (Tree-structured Parzen Estimator) sampler algorithm (Figure 4.1).

This optimization is performed in two main steps. The first step is carried out by implementing a user-defined FoM built to maximize the `XGBoost` AUC but considering an extra protection, namely a factor based on the Kolmogorov-Smirnov test, to avoid overtraining. On top of this, a further optimized choice of the final best Working Point can be obtained in terms of additional FoMs such as the Punzi FoM or the product of purity times the signal significance (PxSS). In this second step, the process of selecting an optimal Working Point involves the application of the trained `XGBoost` model to the validation sample. Upon completion of the second step the optimized models are further ranked based ıt the Punzi FoM. The performance gain of this optimization approach has been reported and compared with the unoptimized `XGBoost` results previously obtained (Tables 4.6 and 4.9 respectively in the `displaced` and the `inclusive` scenario). In both `displaced` and `inclusive` scenarios the Punzi FoM and the product PxSS are better performing with respect to other standard FoMs like signal significance or purity alone. The Punzi FoM can be considered the best in the `displaced` approach whereas Punzi FoM and PxSS are both well performing in the `inclusive` approach and the choice between them must be guided by the kind of analysis that is going to be performed, namely according to the preference for a larger sample or a purer sample.

This reported kind of study with `Optuna` has very likely been performed for the first time and tracks the path to a new way to perform the hyperparameter optimization of a BDT-based selection. In general the fully optimized `XGBoost` selection is very well performing in extracting, from the CMS data, the signal of the $B_c^{\pm} \to J/\psi\,\pi^{\pm}$ decay, thus providing an optimized capability to obtain samples of B_c candidates useful for the CMS B_c physics program with Run-3 data and beyond.

Finally all the considered selections are critically examined in the framework of the extraction of B_c meson lifetime: the efficiencies of the different selections are investigated and crucial aspects of the reconstruction of the flight length of the meson are investigated in detail. These efficiencies are key elements for the lifetime measurement and it has been studied the role that they play in a very preliminary lifetime measurement.

Appendix A

Selection of the physics variables for XGBoost

An essential tool for evaluating the performance of a model is the *feature ranking* or *feature importance* table. This table shows the effectiveness of each feature (namely a physics observable) when training the model. By ranking the contributions of each feature, it provides valuable insight into their impact on predictions. This information helps to identify and eliminate irrelevant features, ultimately improving the accuracy and efficiency of the model. In fact the removal of unnecessary features helps to avoid the overfitting of the model on the training sample.

In data analysis and feature selection, the Correlation Matrix (CM) plays a crucial role. This matrix provides a comprehensive view of the relationships between different variables in a dataset. It allows for the identification of patterns, dependencies, and potential redundancies by calculating pairwise correlations between all the variables. The matrix values range from -1 to $+1$, where $+1$ indicates a complete (positive) correlation, and -1 indicates a complete negative correlation (or anti-correlation). When a couple of variables are strongly correlated, the least important variable (according to the feature ranking) is removed.

The CM is a symmetric matrix, where each row/column represents a variable: the variables are sorted with importance decreasing from top to bottom and from left to right.

The process of selecting the most significant variables for the XGBoost model involves iteratively removing the least important and highly correlated variables. The procedure begins with a long list of input variables which are used commonly for the reconstruction of similar decays in HEP. In each iteration, the model is trained, and the CM is analyzed for both the signal and background test datasets. The variable with the lowest importance is removed if it

satisfies the following criteria:

- It is highly correlated for both signal and background datasets.
- Its removal does not lead to a significant decrease in the AUC value.
- It does not adversely affect the overall performance of the model, passing all performance tests.

The process continues until the least important variable is neither highly correlated in both the signal and background CMs nor significantly affects the AUC score. The Pearson CM was created using the `seaborn` library [80] to analyze linear relationships between variables. By systematically removing the least important and highly correlated variables, the model's complexity has been reduced without compromising its predictive power. This approach ensured that the remaining variables provides the most significant contribution to the model's performance, thus enhancing the robustness and interpretability of the final `XGBoost` model.

Through repeated training sessions (5 times per each iteration), it was observed that the uncertainty on the AUC measurement was approximately 1 per mil. This uncertainty was determined by calculating the deviations of the AUC values from the mean across multiple iterations of independent training with the same settings.

This appendix is divided into two sections, respectively dedicated to the `displaced` and `inclusive` approaches introduced in Chapter 3. Each section provides the full list of initial input variables, a summary table of the iterative process, and the corresponding plots for each iteration.

A.1 Variables' selection for Displaced Approach

The first training for the displaced approach used all the variables and resulted in an AUC of 0.973. The variables used in this training are as follows:

- `B_eta`
- `B_Prob`
- `B_pt`
- `bCosAlpha_bestDist`
- `bFlightLen_bestDist_2DSig`
- `cosAngle_J_pi`

- d0SigPi
- deltaRmumu
- dR_J_pi
- J_eta
- J_Prob
- J_pt
- J_pt1
- J_pt2
- jCosAlpha_bestDist
- pi_eta
- pi_pt

Details on the variables are available in the Glossary.

The overall process of variable elimination is summarized in Table A.1. In each iteration, the least important and highly correlated variables are removed in sequence. The final list of variables after all iterations of elimination is as follows:

- B_pt
- B_Prob
- bCosAlpha_bestDist
- bFlightLen_bestDist_2DSig
- d0SigPi
- deltaRmumu
- dR_J_pi
- jCosAlpha_bestDist
- pi_eta
- pi_pt

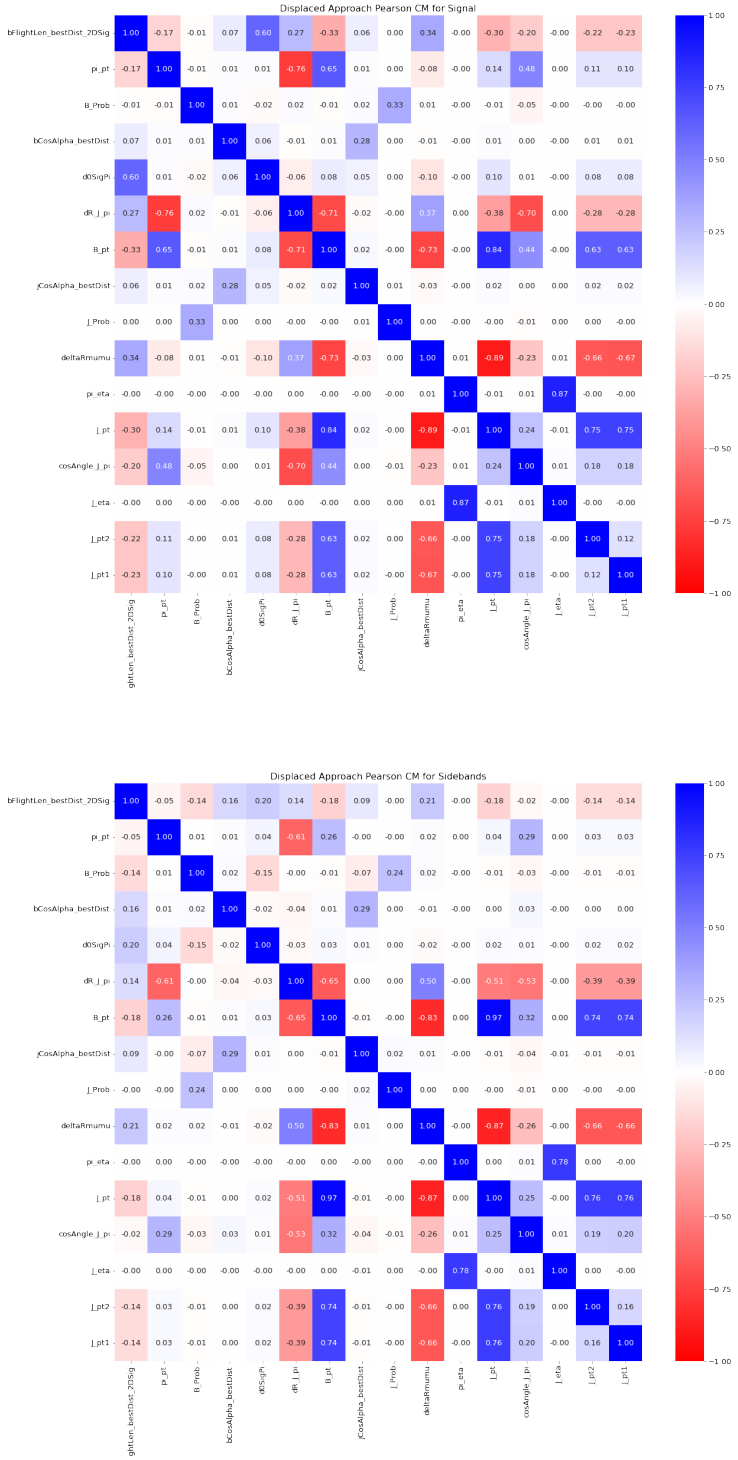


Figure A.1: CMs for the Displaced Approach: (Top) Signal dataset, (Bottom) Background dataset. These matrices illustrate the correlations between variables, where highly correlated and less important variables, such as `J_pt1`, were considered for removal.

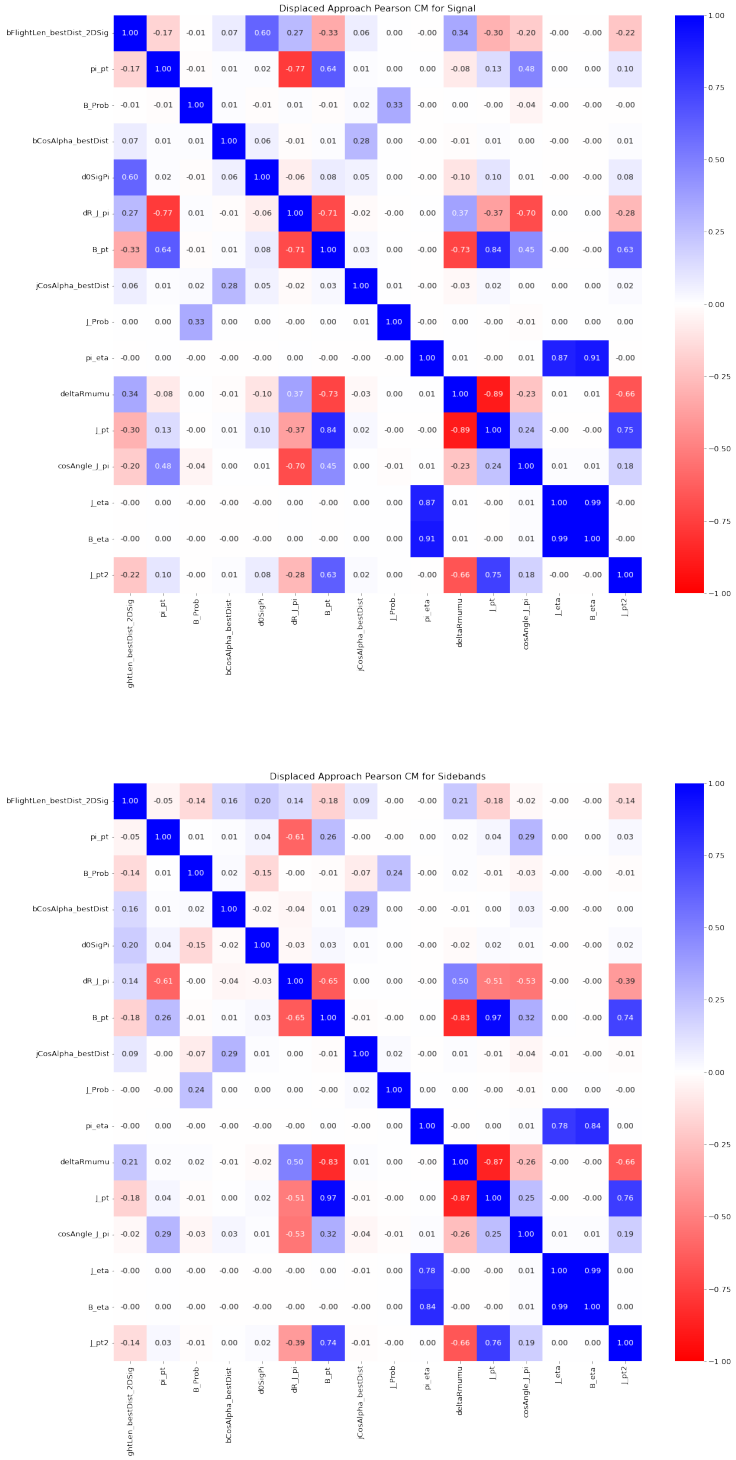


Figure A.2: CMs for the Displaced Approach after removing `J_pt1`: (Top) Signal dataset, (Bottom) Background dataset. The results of this training showed an AUC of 0.972, with `J_pt2` identified for removal in the next iteration.

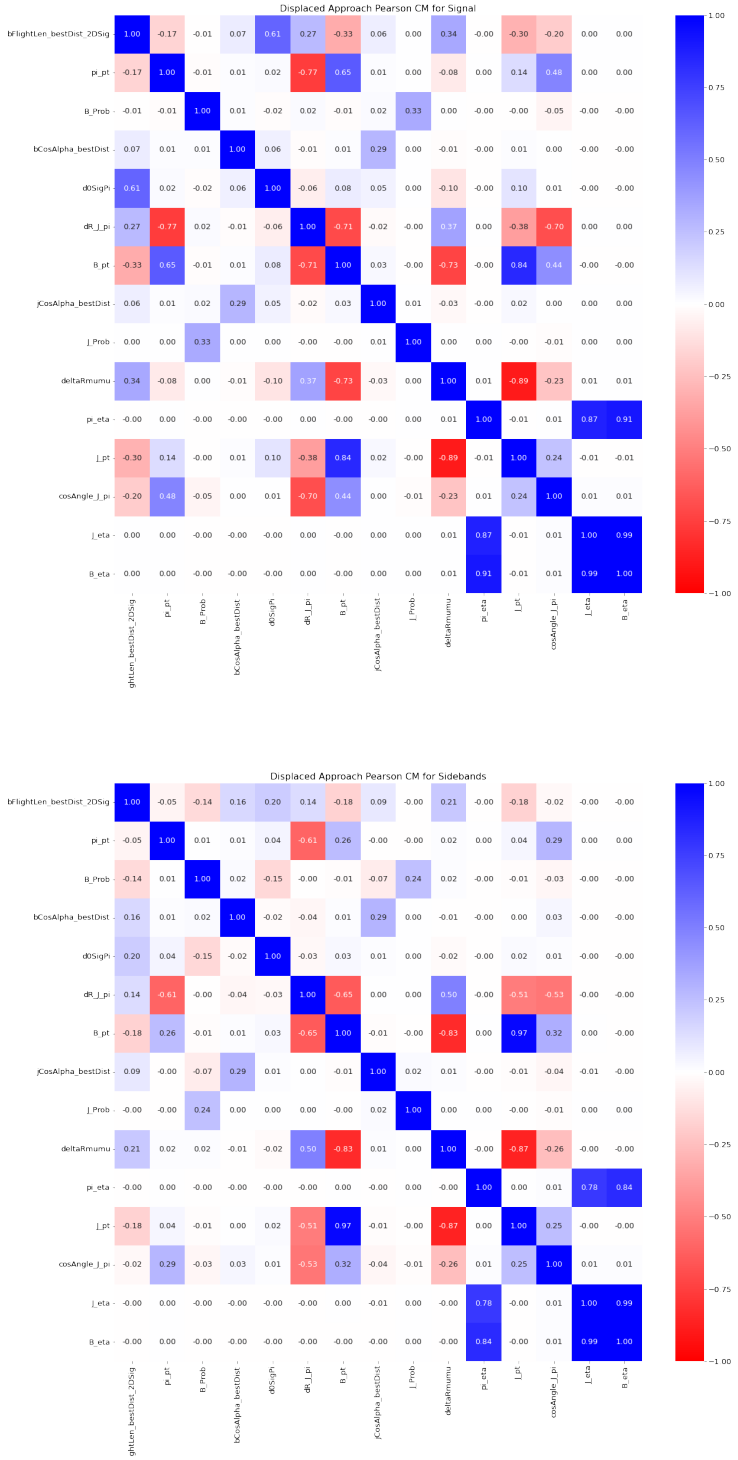


Figure A.3: CMs for the Displaced Approach after removing `J_pt2`: (Top) Signal dataset, (Bottom) Background dataset. The results show that `B_eta` should be the next variable to be removed due to its high positive correlation with `J_eta` and significantly lower importance.

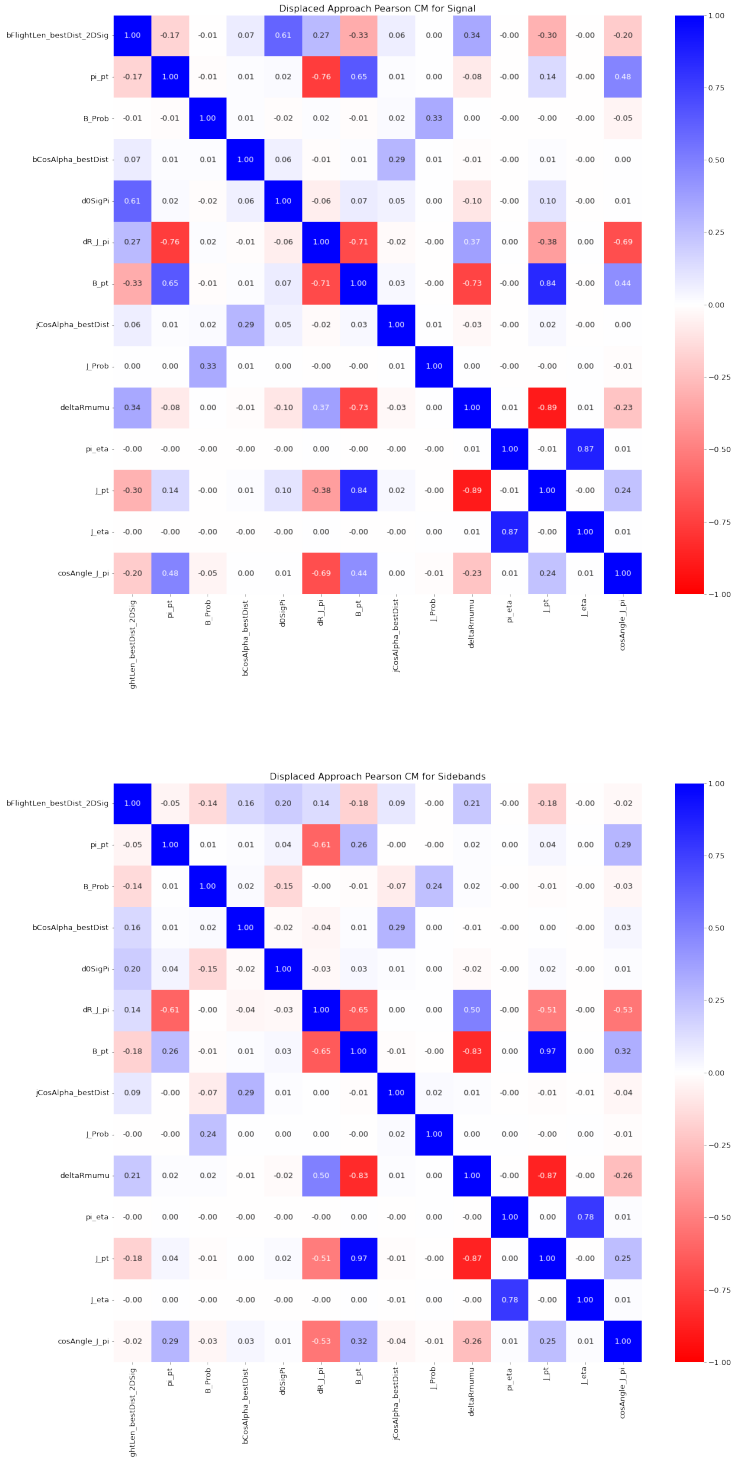


Figure A.4: CMs for the Displaced Approach after removing `B_eta`: (Top) Signal dataset, (Bottom) Background dataset. These matrices show the correlations between variables, where `B_eta` was removed due to its low importance and high correlation with several other variables.

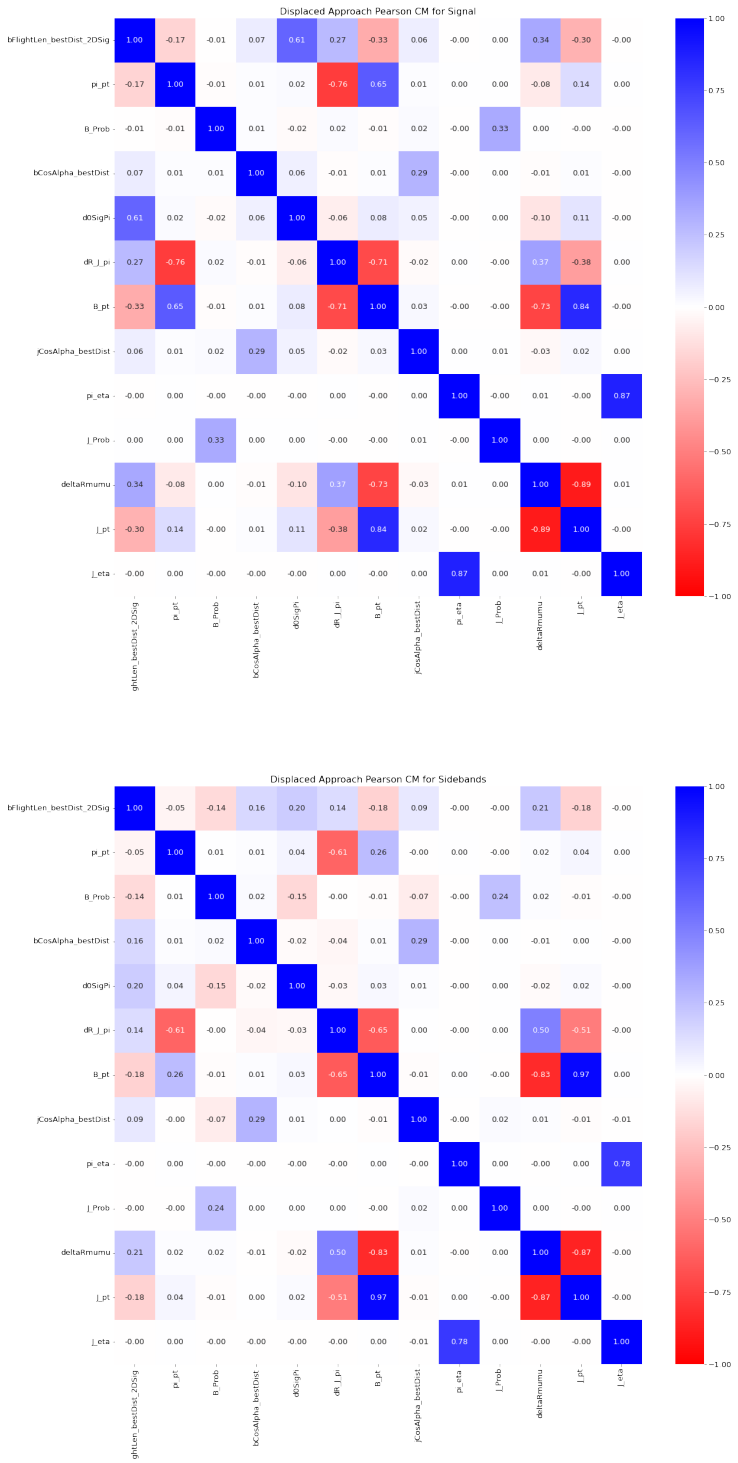


Figure A.5: CMs for the Displaced Approach after removing `cosAngle_J_pi`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_eta`, which shows high positive correlation with `pi_eta`, was identified for removal in the next iteration.

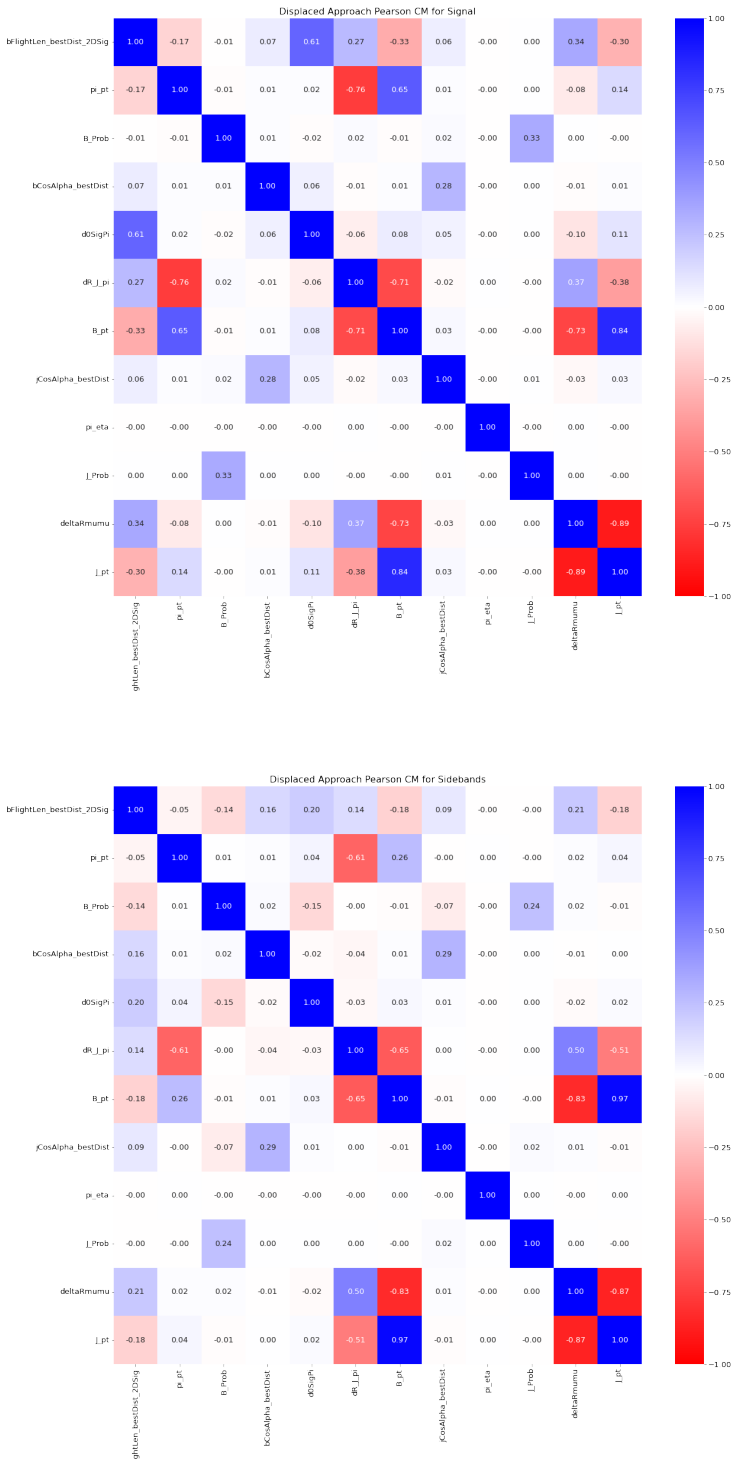


Figure A.6: CMs for the Displaced Approach after removing `J_eta`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_pt`, which is strongly correlated with several other variables, was identified for removal in the next iteration.

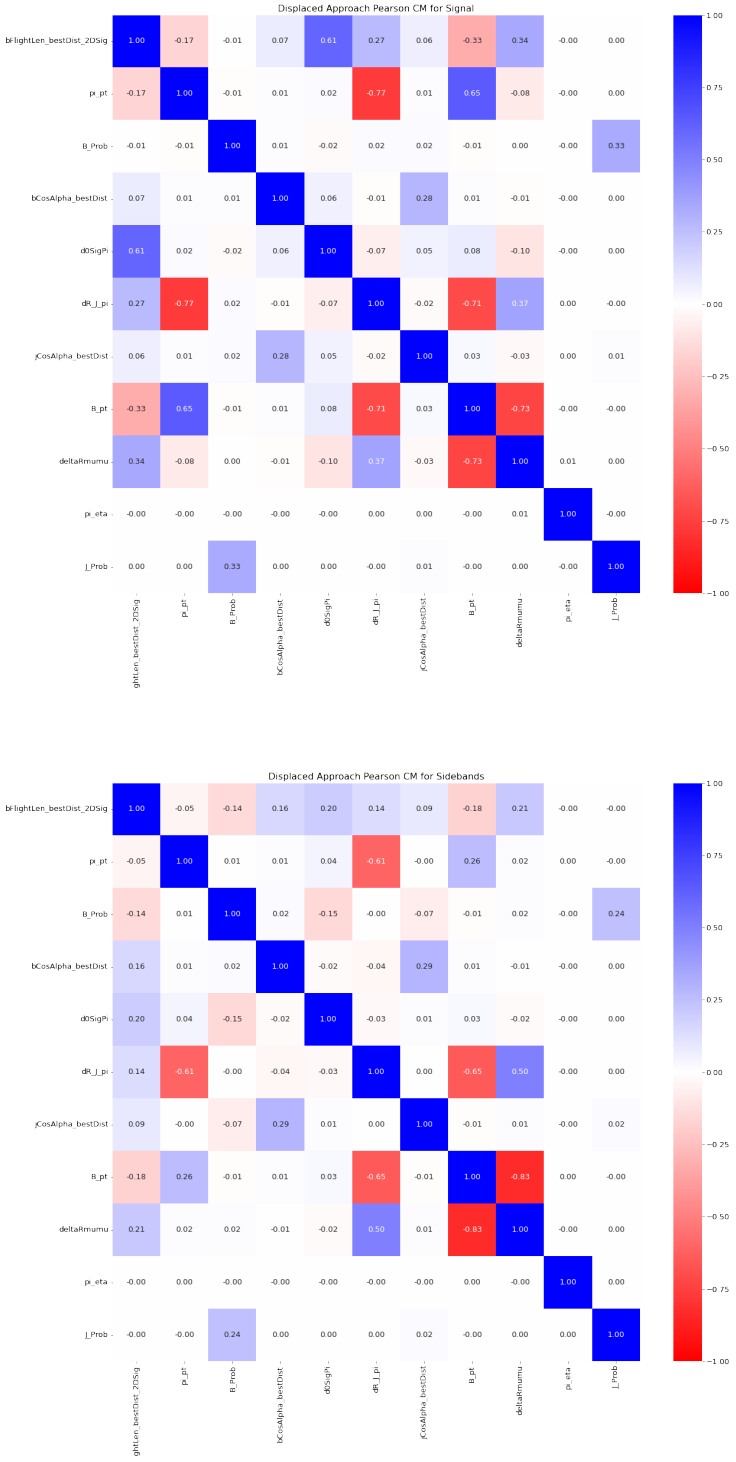


Figure A.7: CMs for the Displaced Approach after removing `J_pt`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_Prob` was identified for removal in the next iteration.

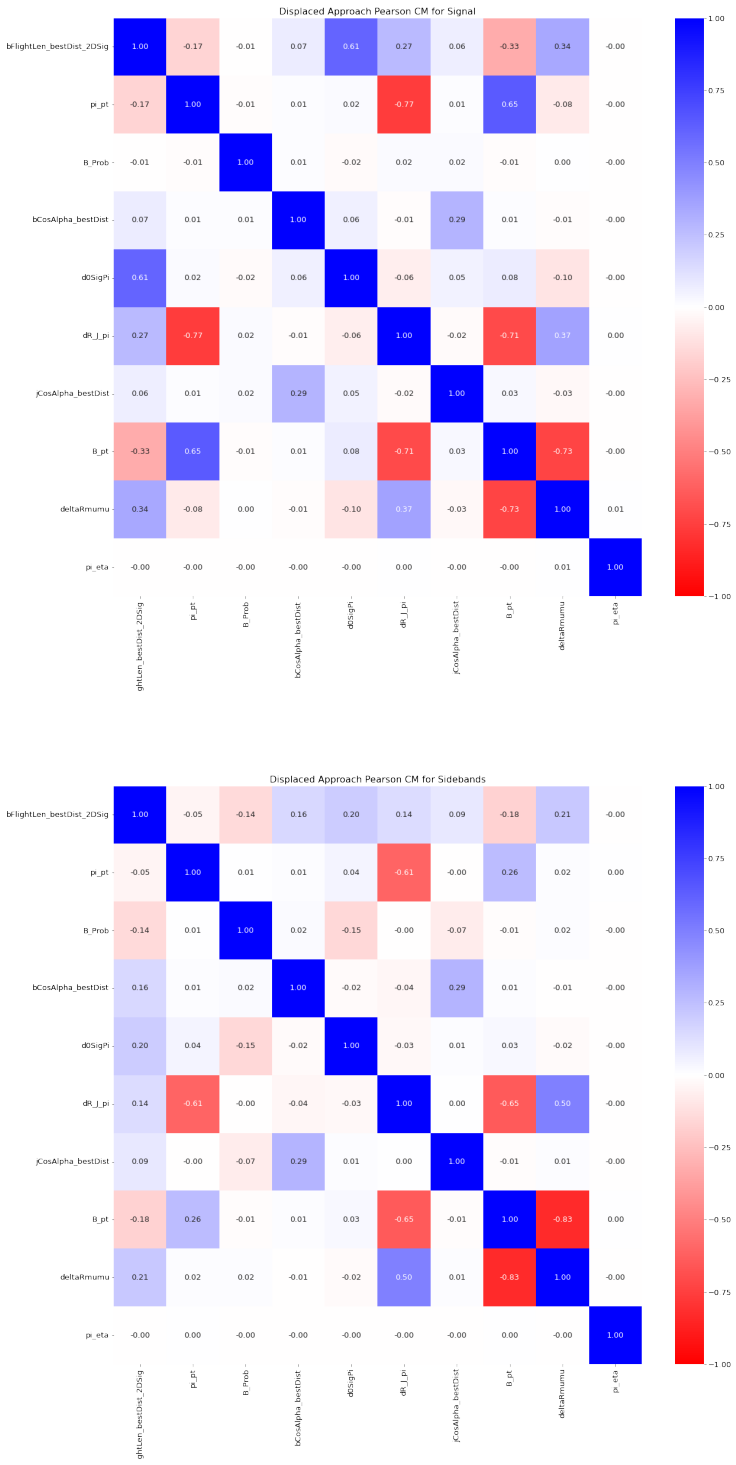


Figure A.8: CMs for the Displaced Approach after removing `J_Prob`: (Top) Signal dataset, (Bottom) Background dataset. The process of variable dropping was halted as no further highly correlated variables were identified.

Number of variables removed	Pearson CM	AUC	Name of removed variable
0	Figure A.1	0.973	None
1	Figure A.2	0.972	J_pt1
2	Figure A.3	0.972	J_pt2
3	Figure A.4	0.973	B_eta
4	Figure A.5	0.973	cosAngle_J_pi
5	Figure A.6	0.972	J_eta
6	Figure A.7	0.972	J_pt
7	Figure A.8	0.972	J_Prob

Table A.1: Overview of variable removal iterations for the Displaced Approach. The table shows the number of variables removed, Pearson CM figure references, AUC values, and the specific variable removed in each iteration. The physical meaning of the mentioned variables is available in the Glossary.

A.2 Variables' selection for Inclusive Approach

The initial training for the inclusive approach, using all the variables but those carrying information from the displacement of the B_c candidate from the PV, resulted in an AUC of 0.935. The variables used in this training are the followings (with reference to the Glossary):

- B_eta
- B_pt
- B_Prob
- bCosAlpha_bestDist
- cosAngle_J_pi
- deltaRmumu
- dR_J_pi
- J_eta
- J_pt
- J_pt1
- J_pt2
- J_Prob

Number of variables removed	Pearson CM	AUC	Name of removed variable
0	Figure A.9	0.935	None
1	Figure A.10	0.935	cosAngle_J_pi
2	Figure A.11	0.936	deltaRmumu
3	Figure A.12	0.935	B_eta
4	Figure A.13	0.935	J_pt2
5	Figure A.14	0.935	J_pt1
6	Figure A.15	0.936	J_eta
7	Figure A.16	0.936	dR_J_pi
8	Figure A.17	0.935	J_pt
9	Figure A.18	0.935	J_Prob

Table A.2: Overview of variable removal iterations for the Inclusive Approach. The table shows the number of variables removed, Pearson CM figure references, AUC values, and the specific variable removed in each iteration.

- jCosAlpha_bestDist
- pi_eta
- pi_pt

The overall process of variable elimination is summarized in Table A.2. In each iteration, the least important and highly correlated variables are removed in sequence. The final list of variables after all the iterations of elimination is as follows:

- B_pt
- B_Prob
- bCosAlpha_bestDist
- jCosAlpha_bestDist
- pi_eta
- pi_pt

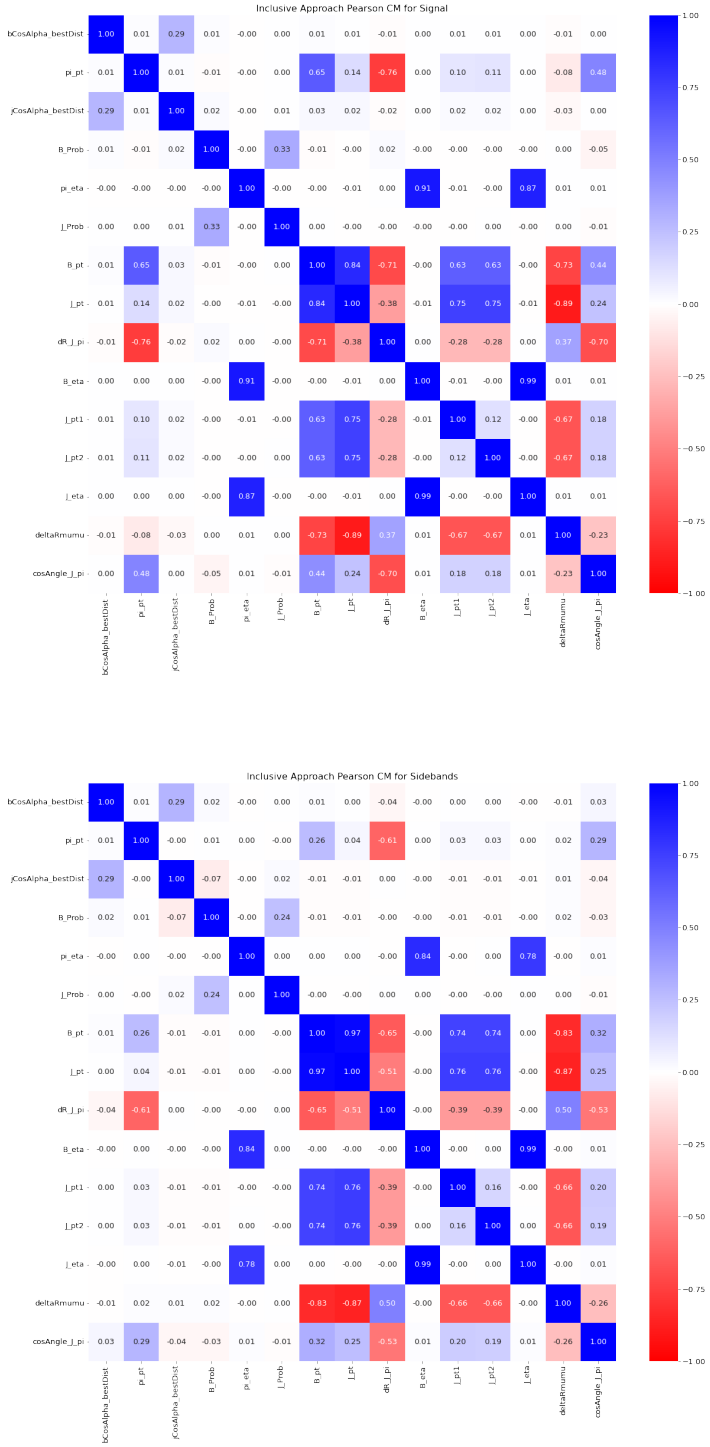


Figure A.9: CMs for the Inclusive Approach: (Top) Signal dataset, (Bottom) Background dataset. These matrices illustrate the correlations between variables, where highly correlated and less important variables, such as`cosAngle_J_pi`, were considered for removal.

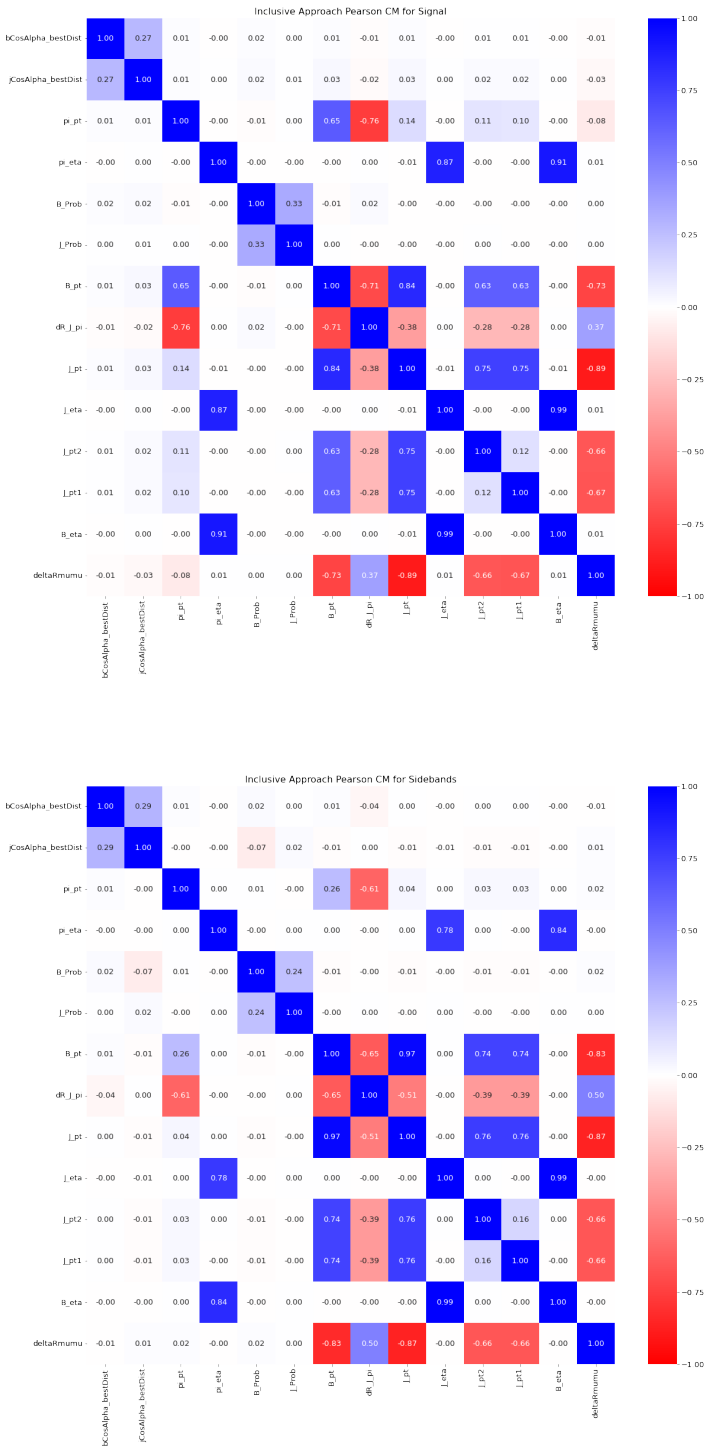


Figure A.10: CMs for the Inclusive Approach after removing `cosAngle_J_pi`: (Top) Signal dataset, (Bottom) Background dataset. The variable `cosAngle_J_pi` was identified for removal in the next iteration.

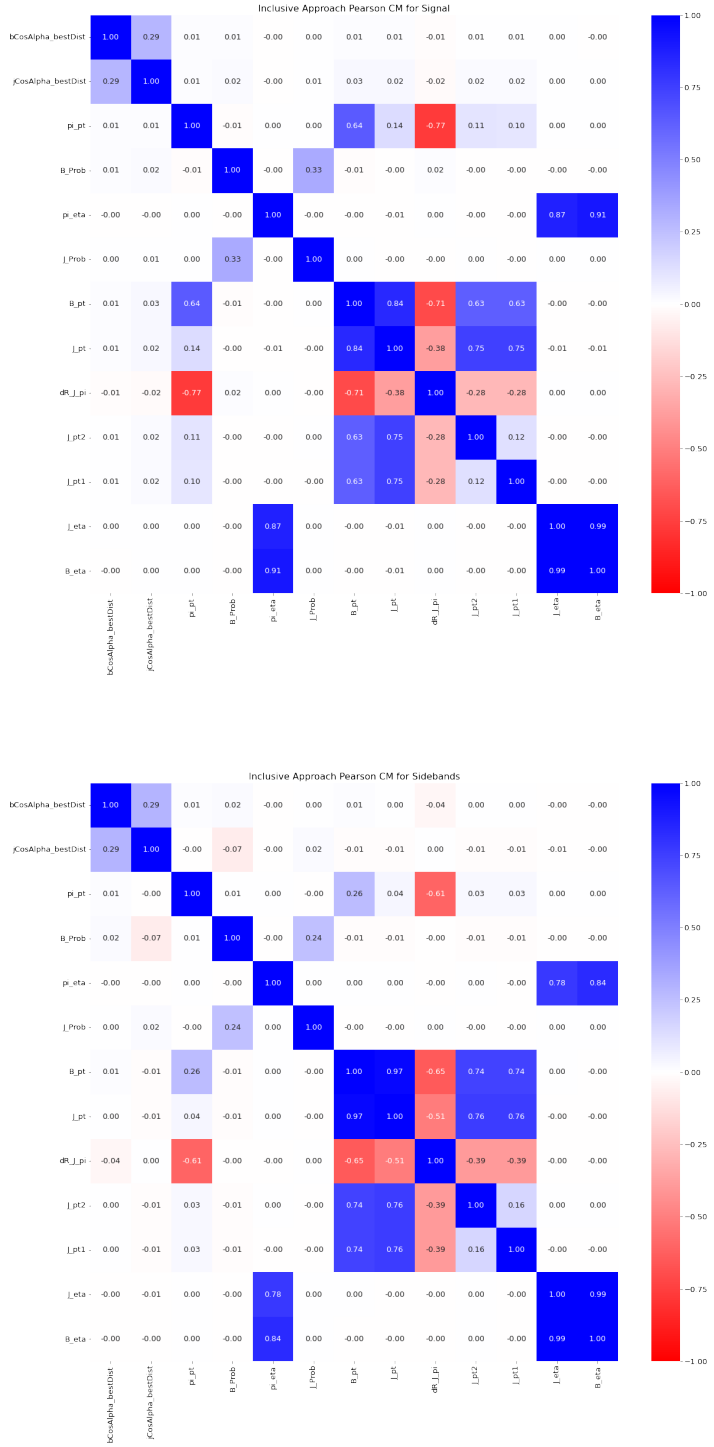


Figure A.11: CMs for the Inclusive Approach after removing `deltaRmumu`: (Top) Signal dataset, (Bottom) Background dataset. The variable `B_eta` was identified for removal in the next iteration.

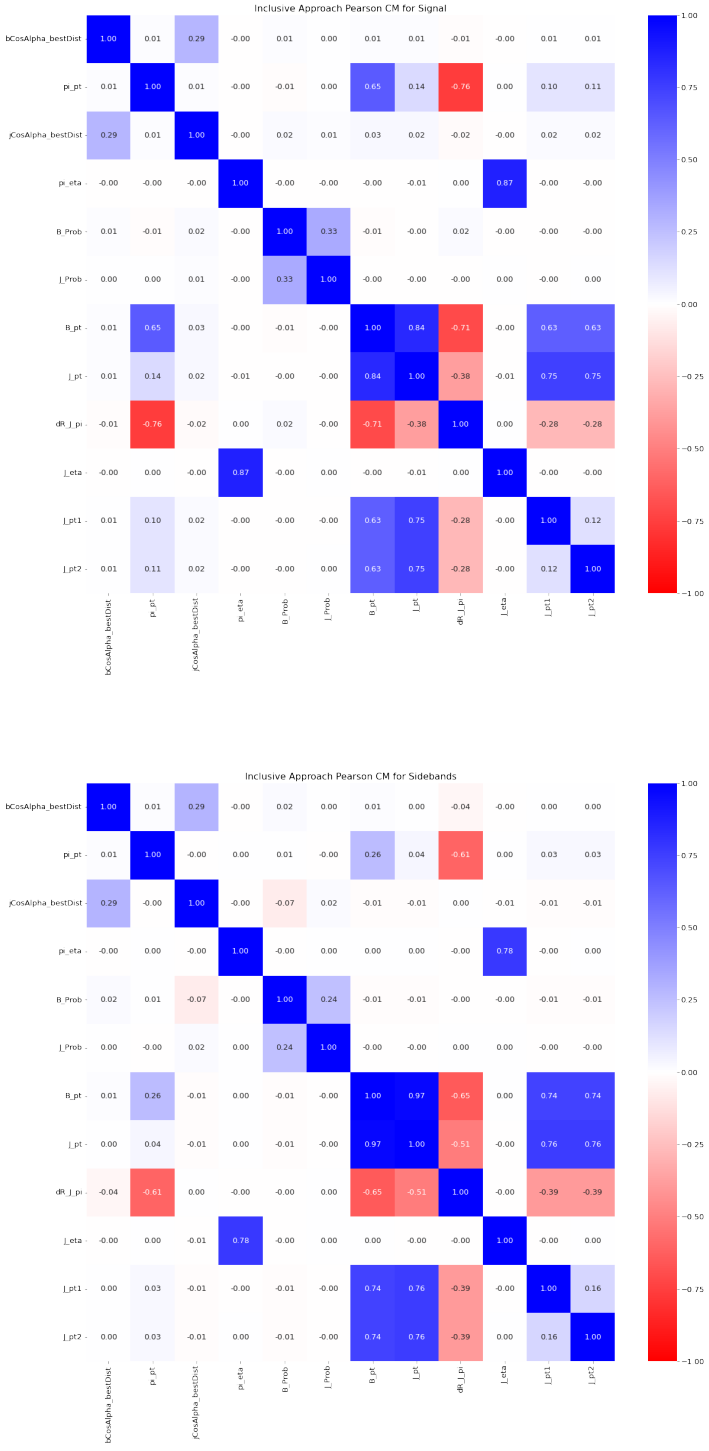


Figure A.12: CMs for the Inclusive Approach after removing `B_eta`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_pt2` was identified for removal in the next iteration.

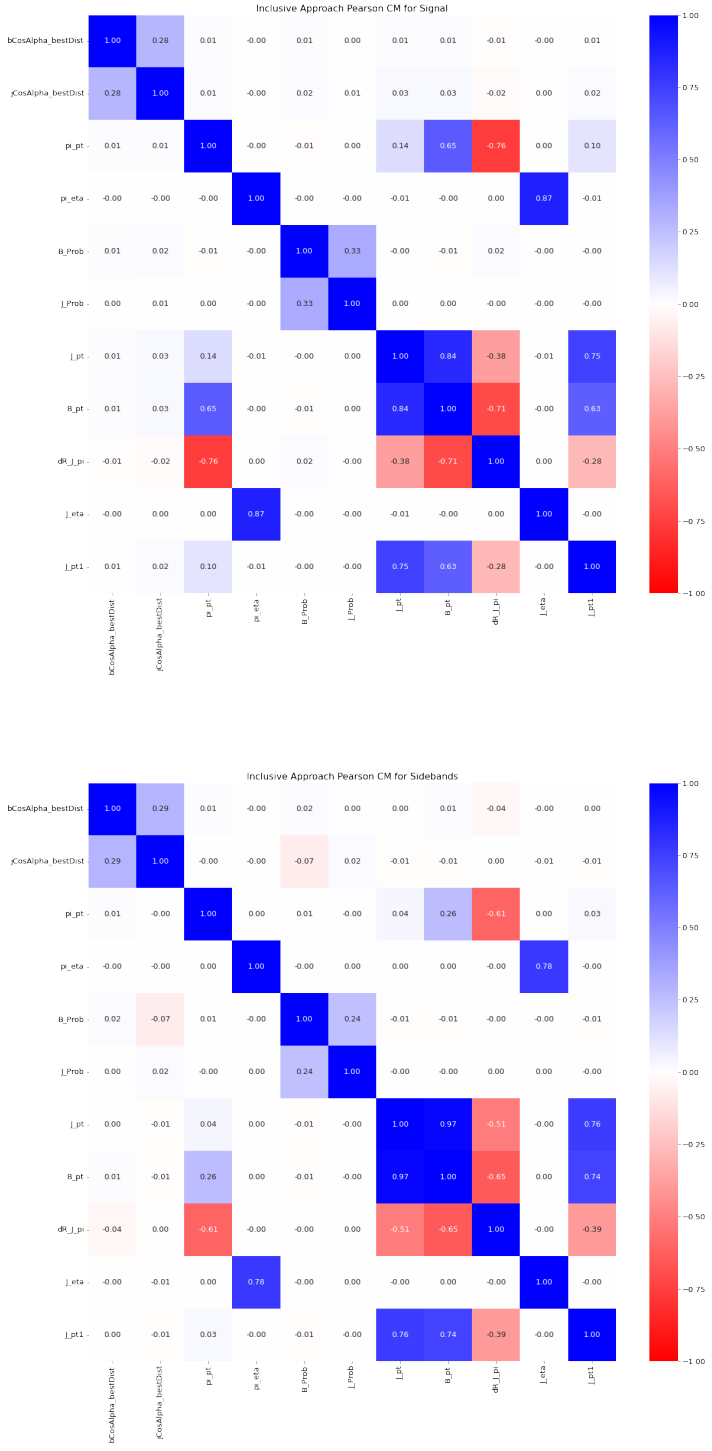


Figure A.13: CMs for the Inclusive Approach after removing `J_pt2`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_pt1` was identified for removal in the next iteration.

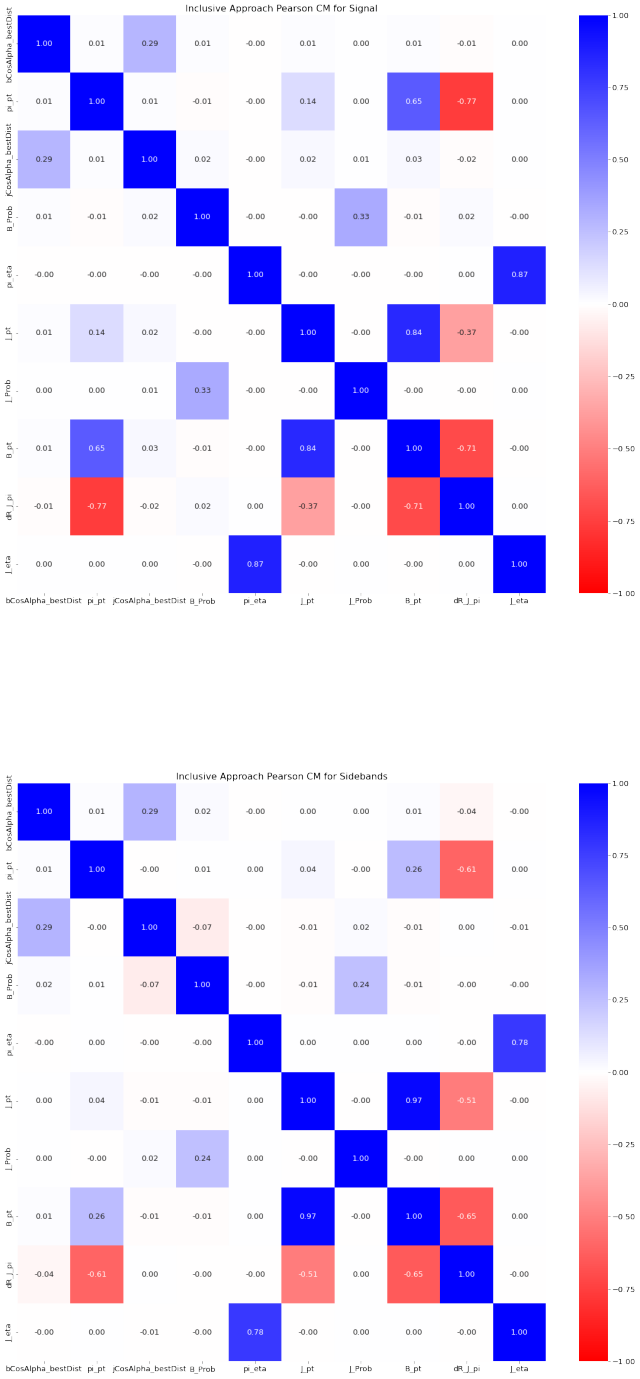


Figure A.14: CMs for the Inclusive Approach after removing `J_pt1`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_eta` was identified for removal in the next iteration.

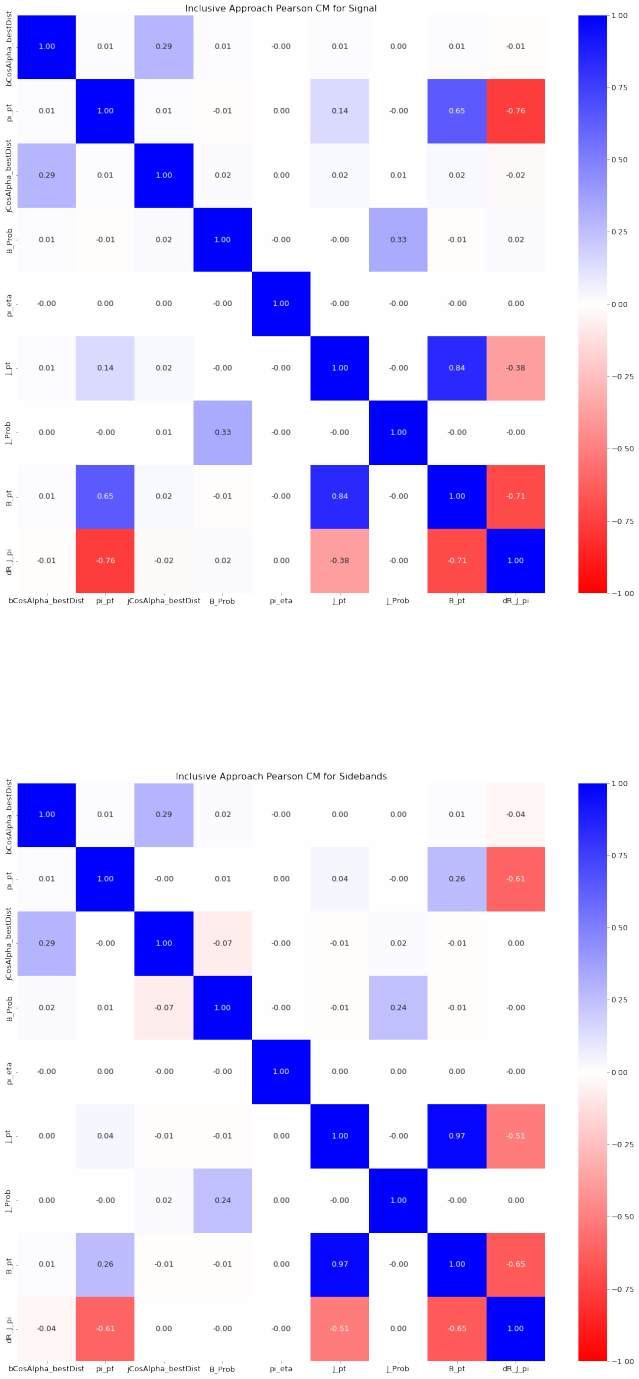


Figure A.15: CMs for the Inclusive Approach after removing `J_eta`: (Top) Signal dataset, (Bottom) Background dataset. The variable `dR_J_pi` was identified for removal in the next iteration.

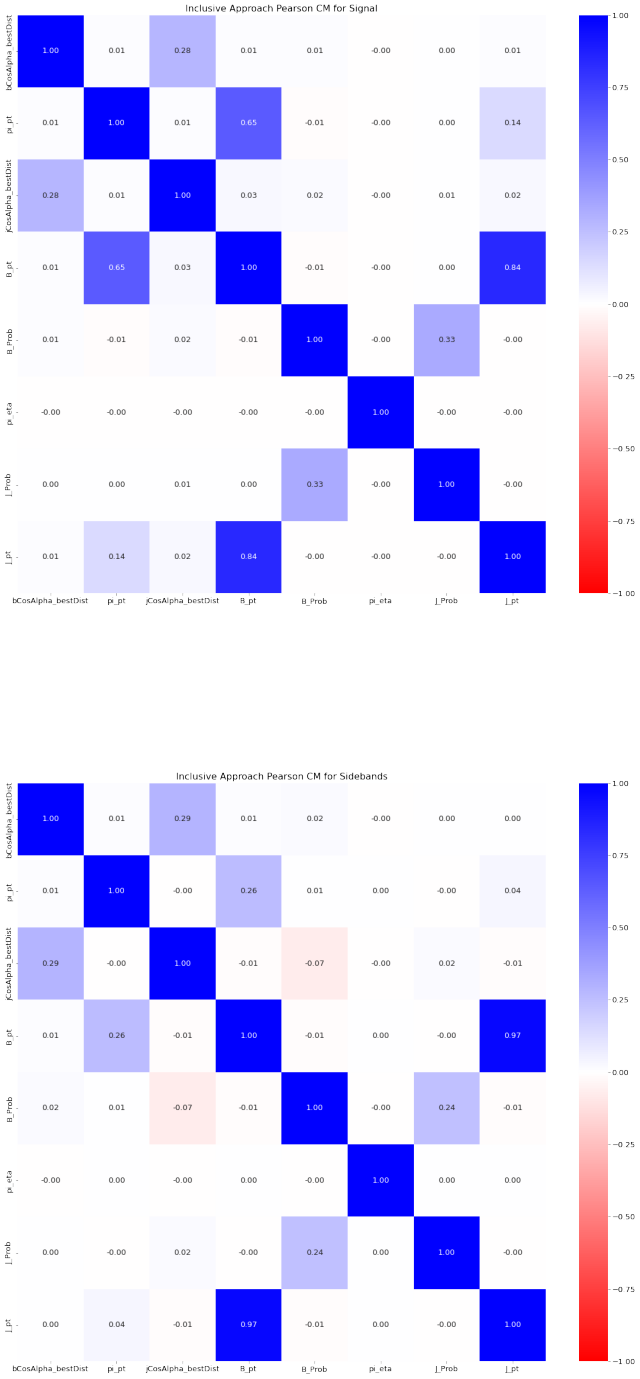


Figure A.16: CMs for the Inclusive Approach after removing `dR_J_pi`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_pt` was identified for removal in the next iteration.

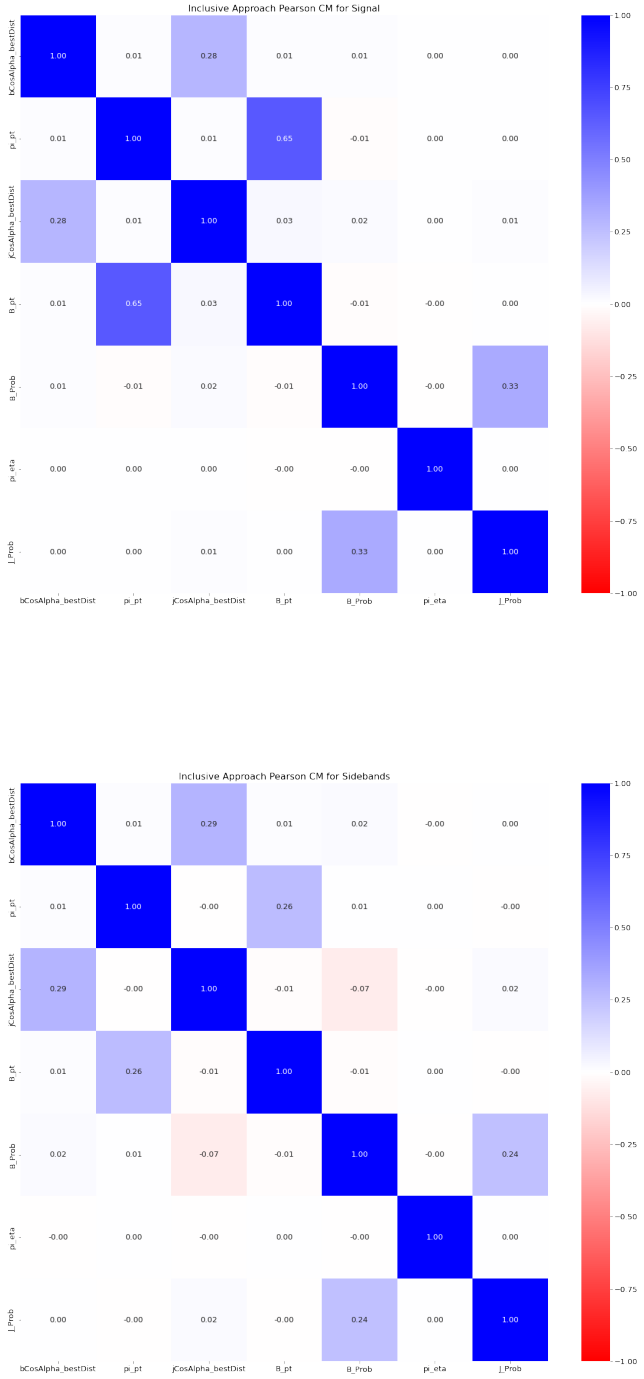


Figure A.17: CMs for the Inclusive Approach after removing `J_pt`: (Top) Signal dataset, (Bottom) Background dataset. The variable `J_Prob` was identified for removal in the next iteration.



Figure A.18: CMs for the Inclusive Approach after removing `J_Prob`: (Top) Signal dataset, (Bottom) Background dataset. The variable `B_Prob` was identified as the least important, showing only low negative correlation in the sidebands and no significant correlation in the signal. Therefore, the variable removal process was terminated at this step.

Appendix B

Strategy for the determination of the B_C primary vertex

In order to study the decay $B_c^+ \to J/\psi(\to \mu^+\mu^-)\pi^+$ a Monte Carlo sample has been privately produced by using `PYTHIA8` to generate high-energy proton-proton collision events and `EvtGen` to handle the heavy flavour decay chain under study. About 800 thousand events of B_c mesons - in an equal amount of positively and negatively charged B_c mesons - decays have been produced considering proton-proton collisions at 13.6 TeV and the conditions characterizing the 2022 data-taking (calibrations, pileup profile, etc.) and are available at the end of the production chain, after the application of an efficiency filter to take into account standard minimal acceptance and kinematic requirements.

As already mentioned in Section 3.2, the Monte Carlo sample was necessary to determine the detector resolution on the invariant mass of the B_c signal and therefore define a signal region and control region. Furthermore the Monte Carlo sample was used to prepare the preliminary measurement of the B_c proper decay time presented in Chapter 3.2.1.

The two key items are to correctly define the proper decay length and to determine the Primary Vertex (PV), namely the vertex that originated the B_c meson. The study on Monte Carlo is mandatory to disentangle the comprehension of each ingredient.

In proton-proton collisions many interactions take place: each interaction point originates several charged particles, whose tracks in the CMS silicon detector are used to reconstruct the many primary vertices. The list of PVs can be sorted by assigning to each PV the sum of the squared-transverse-momentum of all the tracks contributing to the vertex fit.

$$p_T^2(PV_i) = \sum_{tracks} p_T^2 \tag{B.1}$$

This quantity evaluates the *hardness* of the interaction that produced the i-th

Primary Vertex.

B.1 Definition of the proper decay length

The physical observable of the proper decay length ct is derived by considering the lifetime t in the rest frame of the particle:

$$t = \frac{t^{lab}}{\gamma} \tag{B.2}$$

where t^{lab} is the lifetime measured in the laboratory frame of reference and γ the Lorentz factor of the particle in the laboratory frame. This time interval is measured as:

$$t^{lab} = \frac{l_{dec}}{\beta c} \tag{B.3}$$

where l_{dec} is the flight distance of the particle before decaying, and β is the particle velocity in units of the speed of light c. The flight distance is calculated by taking into account both the geometrical distance between the PV and the SV and the direction of the momentum p of the particle:

$$l_{dec} = \frac{(\vec{SV} - \vec{PV}) \cdot \vec{p}}{|\vec{p}|} \tag{B.4}$$

When the particle flies for a significant distance, the SV and PV are significantly detached and it is reasonable to assume that the flight distance and the momentum are aligned; on the contrary for a particle which shortly decays after its production, the measurement of the SV is strongly affected by its uncertainty and it might happen that the SV is "behind" the PV with respect to momentum direction. This is considered by the adopted definition by assigning a negative ct (coming from the scalar product in the formula).

From special relativity, the momentum p of a particle with mass m and velocity βc is calculated as:

$$p = m\beta\gamma c \tag{B.5}$$

Eq. B.2 can be rewritten using Eqq. B.3 and B.5:

$$ct = c\,\frac{l_{dec}}{p} \cdot m \tag{B.6}$$

Since Eq. B.3 still holds if the distance and velocity are considered only in the transverse plane, the proper decay time can also be calculated as:

$$ct = c\,\frac{l^{\perp}_{dec}}{p_T} \cdot m \tag{B.7}$$

where $l_{dec}^{\perp}$ is the transverse component of the flight distance of the particle, while p_T is its transverse momentum. Uncertainties on the proper decay time are computed by propagating the uncertainties on the PV and SV positions, and on the particle's momentum, while the uncertainty on the mass of the particle is neglected. In previous CMS Heavy Flavour Physics analyses in some cases the mass value from the PDG is used as mass m, while in others the invariant mass of the candidate is used. In the present study the latter definition is used, where the invariant mass of the B_c candidate is obtained from a kinematically constrained vertex fit, where the di-muon is constrained to the invariant mass of the J/ψ.

In this appendix two interconnected aspects of the definition of ct are discussed:

- the usage of the 2D or the 3D ct definitions ct_{2D} and ct_{3D} (respectively in Eqq. B.7 and B.6);
- the identification of the PV (in Sec. B.2).

B.1.1 Comparison between 2D and 3D lifetime definitions

The B_c Monte Carlo sample is used in order to verify if the 2D and 3D definitions of the proper decay length ct (respectively defined in Eqq. B.7 and B.6) are equivalent or not. On the MC sample the generator-level information is available, thus knowing exactly the geometrical PV and SV of each B_c candidate.

A simple selection is applied to the MC sample (`preprocess` selection, as described in Table 3.2). Only reconstructed B_c candidates whose children are geometrically matched to the corresponding particles at generator level are kept (angular separation $\Delta R < 0.01$): this procedure is often called *gen-matching*. Each reconstructed candidate is saved in the MC ROOT file together with its information at generator level.

For the selected signal candidates the both the 2D and 3D ct are calculated by using the generator level information ($gen{-}ct_{2D}$ and $gen{-}ct_{3D}$): the plots in Figure B.1 shows there is no significant difference between the two calculations. The average difference between the two quantities for each candidate is of the order of 1 part per million. The shape of both the distributions is consistent with a falling exponential, as expected.

As expected, the two definitions are in principle equivalent, yet the situation changes when the reconstructed information is used: as a first approach to showcase the issue the hardest PV is always chosen as PV. The comparison between the reconstructed ct_{2D} and ct_{3D} is shown in Figure B.2. The

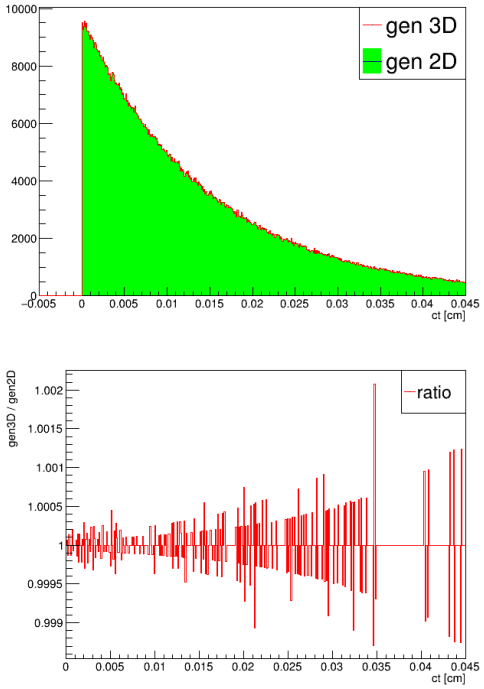


Figure B.1: Top plot: Distribution of gen-ct_{2D} and gen-ct_{3D} for the B_c candidates selected by the `preprocess` selection on the MC signal sample. Bottom plot: Ratio of the two distributions in the plot above. No significant difference between the two formulae.

reconstructed ct_{2D} distribution appears to be very similar to the generator-level distribution for ct with at least a mild displacement ($ct > 30\mu m$). The ct measurement is affected by the resolution effect, which can be modeled with a Gaussian smearing. As a result, the smearing effect is more prominent for small ct values and it happens that the distribution exhibits a negative tail. On the contrary the reconstructed ct_{3D} distribution shows a non-negligible distortion of its shape with respect to the expected falling exponential: this can be due by the wrong choice of the Primary Vertex and - in general - by a worse detector resolution on the z-axis.

The reconstructed ct_{2D} appears to be more robust with respect to the PV choice and it is ultimately favoured against the 3D definition.

An Unbinned Maximum Likelihood fit is performed on the ct_{2D} distribution at generator level: the signal model is a RooDecay function, namely a convolution of an exponential with a Gaussian resolution. The parameter of the exponential is the lifetime constant of the B_c decay, $c\tau$. The fit on the full range (Figure B.3, left) shows a poor quality on the rising part of the distribution at negative values, but still the extracted lifetime constant is compatible with the generation value of 150 μm. The fit is consequently tested on the restricted range $ct_{2D} > -10\mu m$, in order to remove the poorly populated rising

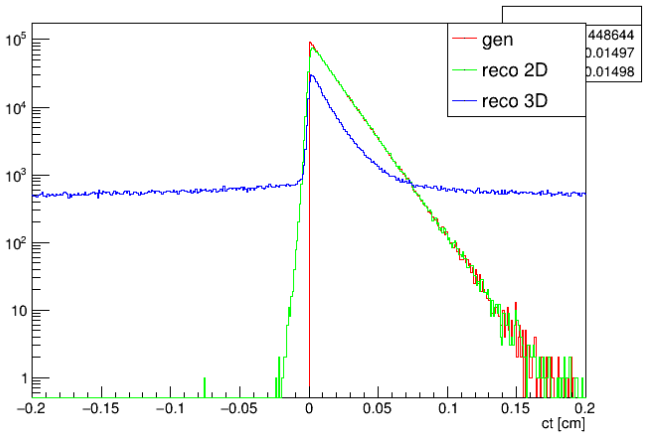


Figure B.2: Distribution for the reconstructed ct (2D in green, 3D in blue), compared with the generator-level ct. The reconstructed ct_{2D} shows a Gaussian smearing with respect to the generator-level distribution, which is easily understandable as an effect of the detector resolution. On the contrary the ct_{3D} is distorted and the method is not considered reliable.

part of the distribution, where additional resolution effects might be needed to improve the description. The result of the second fit is improved in terms of fit quality, while the extracted lifetime constant is compatible with the generation value similarly to the first one.

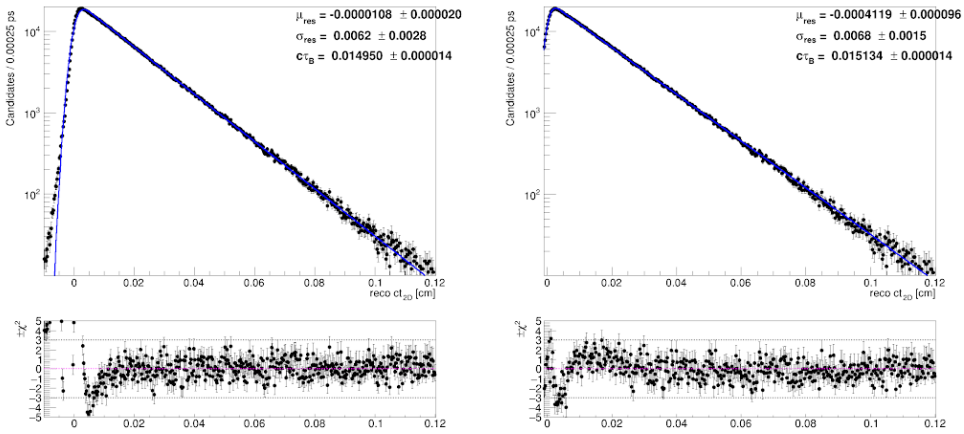


Figure B.3: UML fit with a RooDecay model of the generator-level ct_{2D} for the B_c candidates of the MC sample surving the `preprocess` on the full range (left) and on a slightly restricted range, gen-$ct_{2D} > -10\mu m$ (right). The fit is improved when the rising part of the fit is partially removed, nevertheless both the fit results show a lifetime constant relatively close to the generation value (150 μm).

B.2 Determination of the Primary Vertex choice

The previous discussion allowed to validate the definition of the lifetime ct, yet this requires in any case information on the position of the candidate's production vertex (PV): at generator-level the information is directly available, while this is not the case on real data. In real data all the proton-proton interaction points (primary vertices) are reconstructed and the analyzer needs a criterion to choose which PV among them is the one from which the B_c candidate originated.

Many possibilities have been considered and compared:

- to choose always the hardest vertex according to B.1 (`highest-pt-sum`);
- to choose the vertex that maximises the cosine of the pointing angle (angle between the flight length and the momentum of the particle) computed in the transverse plane (`pointing-angle-2d`);
- to choose the vertex that maximises the cosine of the pointing angle computed using all the three spatial components (`pointing-angle-3d`);
- to choose the vertex that minimises the distance between the particle's flight direction and the vertex itself (`minimum-distance`).

Figure B.4 shows the full ct_{2D} distribution (right) for each of the possible PV choices, together with the ratios with respect to the `highest-pt-sum` (left). All the distributions are compatible with each and, moreover, compatible with the distribution using the generator-level PV for values above $ct_{2D} > 50\mu m$ (one third of the B_c lifetime constant). It can be concluded that when the decay vertex (SV) is significantly displaced, the determination of the PV is not particularly relevant.

An effect is instead is clearly present near the turn-on: a focus on a small range around zero is shown in Figure B.5 for each of the possible PV choices (top plot) together with the ratios with respect to the `highest-pt-sum` distribution. The distributions using reconstructed PVs show a tail at negative ct_{2D} values as expected because of resolution effects; in addition the distributions based on the pointing angle, either 2D or 3D, show a non-trivial shape near 0 which favours vertices resulting in small positive ct_{2D} values rather than small negative ct_{2D} values. The `highest-pt-sum` and the `minimum-distance` distributions instead show a simple turn-on curve: this is expected for the first choice since it is naturally unbiased, while for the latter is a promising hint that the `minimum-distance` criterion can be a reliable, unbiased, method for the PV estimation.

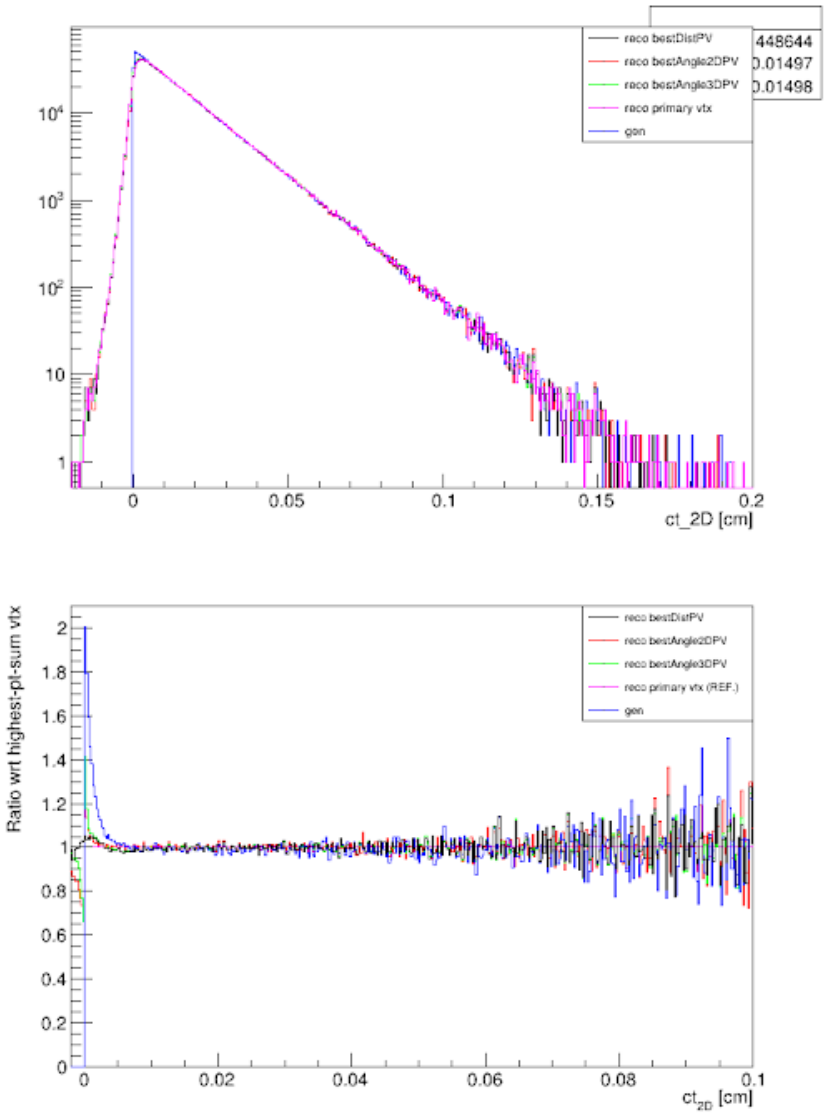


Figure B.4: The top plot shows the full-spectrum distribution of the ct_{2D} distributions using different PV choices, including the generator-level PV (blue curve). The ratios relative to the `highest-pt-sum` distribution are shown on the bottom plot.

Since the generator-level PV is available, it is possible to determine the 3D spatial distance between the true PV and the reconstructed one coming from each method, shown in Figure B.6. The right plot shows the bulk of the distributions: all the distributions rapidly decrease with increasing ct_{2D}, indicating that in any case the average uncertainty is of the order of few tens of micrometers. The left plot shows instead the full distribution: all the distributions show a certain amount of cases where the reconstructed PV is far from the true one. The `minimum-distance` is the only method that exhibits a falling behaviour also at very large values, while the others show an almost uniform

(or slowly decreasing) behaviour that indicates that the PV choice procedure has failed, assigning instead a random PV.

Of course different methods can still result is picking the same PV: the plot in Figure B.7 displays how many times each method picks the reconstructed PV closest to the true PV and shows that the `minimum-distance` method appears to be the more reliable.

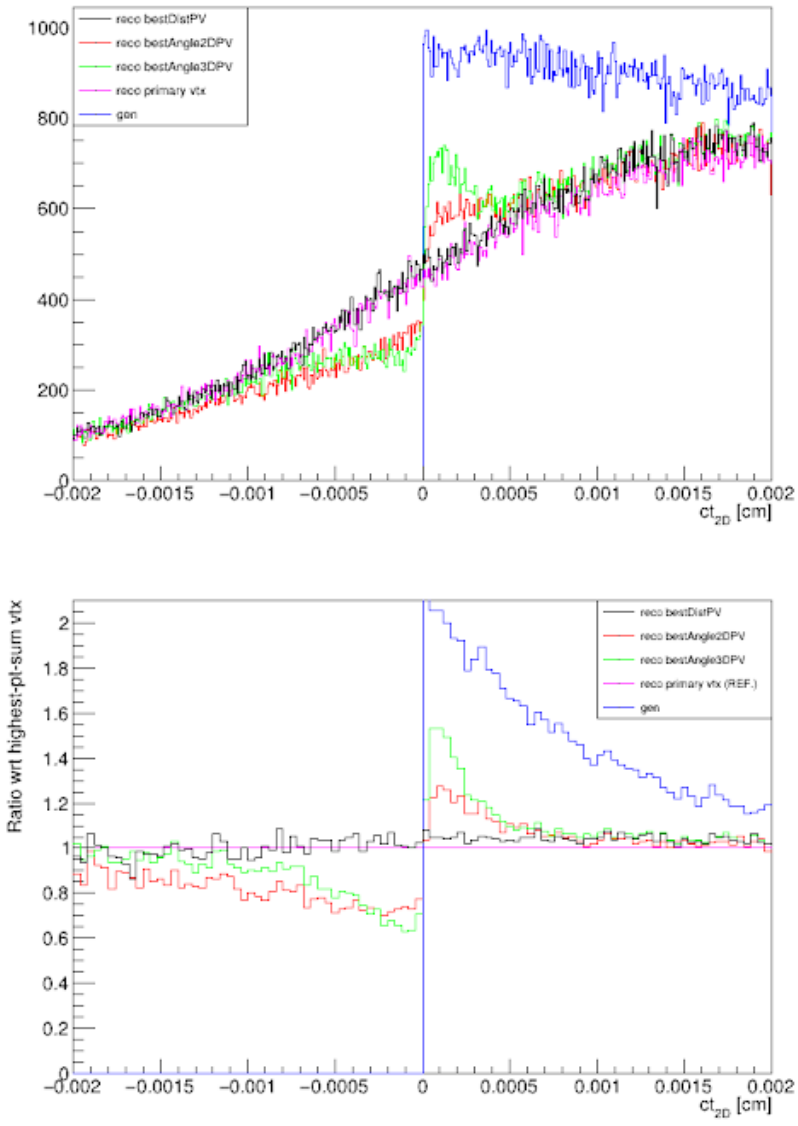


Figure B.5: The top plot shows the distribution of the ct_{2D} distributions using different PV choices, including the generator-level PV (blue curve) zoomed in the range ± 20 μm. The ratios relative to the `highest-pt-sum` distribution are shown on the bottom plot.

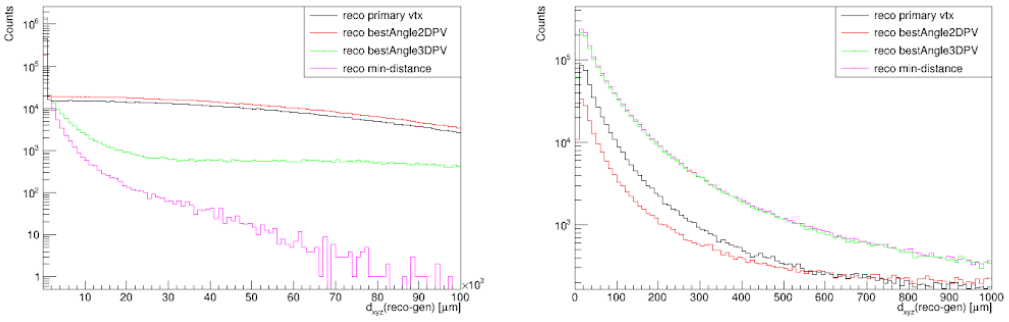


Figure B.6: The left plot shows the distribution of the distance between the true PV and the reconstructed PV from different methods. The right plot shows instead a focus on the first bin, where most of the distribution is focused.

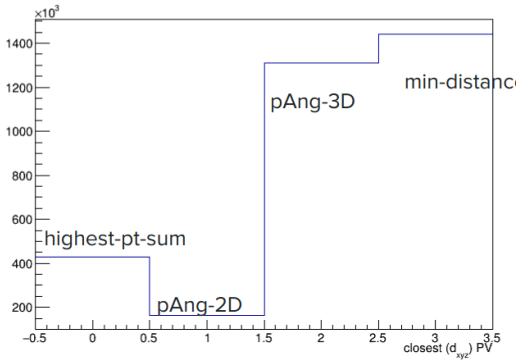


Figure B.7: The histogram shows for each method the number of times that the corresponding algorithm picks the reconstructed PV closest to the true PV.

Appendix C

Cabibbo suppressed $B_C \to J/\psi K$ signal in CMS Run-3 data

In the Section 1.6 of Chapter 1 (Figure 1.8) it was discussed how $B_c \to J/\psi\, K$ is a Cabibbo suppressed decay of type-I, to be compared to $B_c \to J/\psi\, \pi$ which is Cabibbo favoured instead. The two decays are completely indistinguishable for CMS due to the lack of an hadronic identification capability.

When observing the two radially excited states in the $B_c^+\pi^+\pi^-$ invariant mass spectrum, as shown in Fig. 1.9, the Cabibbo-suppressed contribution was considered both in this fit and in the fit model of the $B_c \to J/\psi\, \pi$ invariant mass as shown in Fig. 1.10. In that published CMS analysis the shape of the *reflection* of the $B_c \to J/\psi\, K$ signal in the $J/\psi\, \pi$ invariant mass final state was derived from simulation studies and this component enters in the fit with a weight of $5.0 \pm 0.5\%$; this weight is the result of the ratio of the branching fractions 0.079 ± 0.008 borrowed from LHCb [81] (in good approximation $(8 \pm 1)\%$) multiplied by the ratio of reconstructed efficiencies, 1/0.94, scaled to the mass window of $[6.2, 6.355]$ GeV in the B_c^+ invariant mass distribution. An associated systematic uncertainty has been considered by varying the central value of the weight by $\pm 10\%$.

In the framework of this thesis it was explored, with Run-3 data, the possible contribution coming from the Cabibbo-suppressed $B_c \to J/\psi K$ decay. With this aim, a dedicated $B_c \to J/\psi K$ signal Monte Carlo sample was produced and then the $B_c \to J/\psi K$ candidates, once matched with generated candidates to reject some small possible background associated to random combinations, underwent the same selection applied to extract the $B_c \to J/\psi\pi$ signal. The invariant mass of the final state $J/\psi\pi$ is then interpolated (Binned ML - BML - fit) with a fit model to derive the shape of the Cabibbo-suppressed signal in the $J/\psi\pi$ invariant mass, which represents what is commonly called *reflection* in the HEP jargon. It was checked that the best shape model for

this reflection is provided by a Johnson p.d.f. model (which turns out to fit better than a double Gaussian with common mean). This derived shape is handled as an additional signal in the BML fit model used to interpolate - in the data - the invariant mass of the reconstructed $B_c \to J/\psi\pi$ candidates. The normalization of this contribution (i.e. its relative yield with respect to the dominant Cabibbo-favoured signal) can be fixed to a certain value in the fit (and the related goodness-of-fit is evaluated) or left free to float in the fit within a reasonable interval. This procedure has been implemented in two different ways, as illustrated in the following two sections, with the difference consisting in the different choice of both the candidates' selection and the fit model to extract the shape of the reflection. Specifically, in Section C.1 the `XGBoost` inclusive approach model is considered, whereas in Section C.2 the `displaced` selection used in the CMS paper [82] is chosen. The main goal here is to verify - within Run-3 data - whether the relative yields of the Cabibbo-Suppressed signal obtained in the two approaches are compatible between each other.

C.1 Estimation of the $B_c \to J/\psi K$ contribution with the `XGBoost` inclusive approach

The `XGBoost` inclusive approach model discussed in Chapters 4 is applied to the $B_c \to J/\psi K$ signal MC sample and to the 2022-23 datasets. To derive the reflection shape, a Johnson p.d.f. model is used as signal model in the BML fit of the $J/\psi\pi$ invariant mass obtained by applying the mentioned selection to the $B_c \to J/\psi K$ simulated sample, as shown in Fig. C.1. Then, when fitting the invariant mass distribution of the $J/\psi\pi$ final state in the Run-3 data, shown in Fig. C.2, the additional Cabibbo-Suppressed component is modelled with the Johnson function with its parameters settled to those estimated from the simulation, and leaving only the normalization (i.e. the relative yield fraction with respect to the total signal, namely Cabibbo-Favoured plus Cabibbo-Suppressed) as free parameter. This parameter turns out to be relatively small but still somehow different from zero: $(3.8 \pm 2.9)\%$. Other fits have also been tried injecting a fixed amount of Cabibbo-Suppressed contribution and evaluating the quality of the resulting fit. The results of this procedure are summarized in the Table C.1. The CS-fraction result provided by the free fit is consistent with the trends of the total yield estimate and the normalized chi-squared of the BML fit when reducing the injected amount of the CS relative contribution from 10% to 4%.

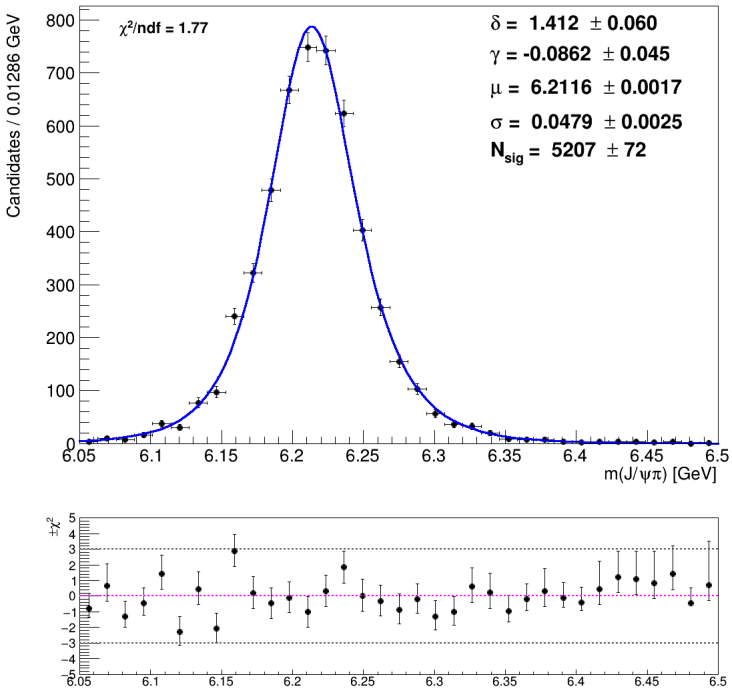


Figure C.1: BML fit to the reconstructed $J/\psi\pi$ invariant mass obtained by the selected simulated $B_c \to J/\psi K$ signal after matching with the generated candidates and applying the `XGBoost` inclusive approach. The resulting shape is not relevantly asymmetric (γ compatible with 0) but shows a clear left tail (δ is definitely different from 0).

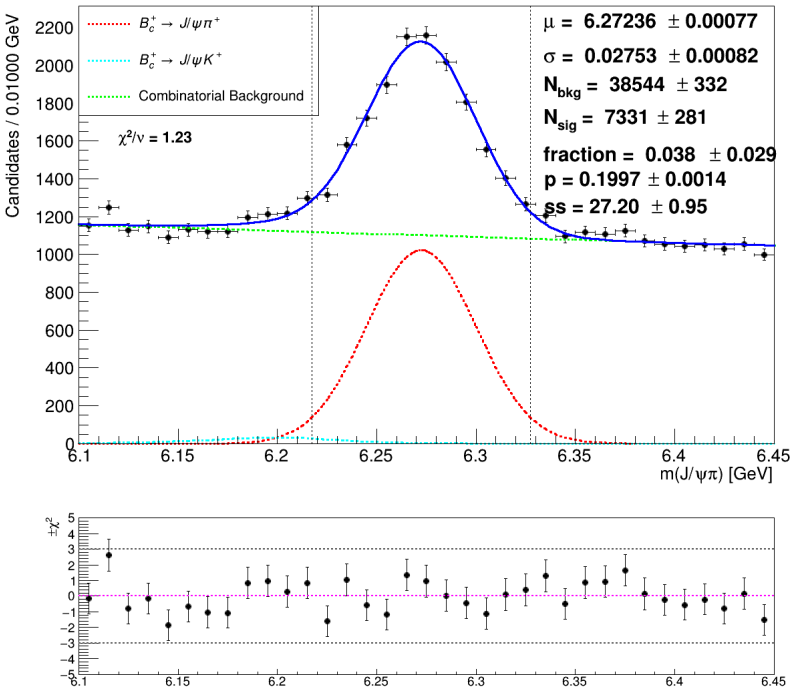


Figure C.2: The invariant mass of the final state $J/\psi\pi$ in the Run-3 data undergoes a BML fit with a model including a Cabibbo-Suppressed contribution (cyan) with the shape borrowed from the fit shown in Fig. C.1 and its normalization left reasonably free to float. The dotted red component represents the Cabibbo-Favoured signal, whereas the green dotted component represents the exponential (combinatorial) background. The relative fraction with respect to the total signal (Cabibbo-Favoured plus Cabibbo-Suppressed) yield is estimated to be $(3.8 \pm 2.9)\%$.

Injected CS fraction	Total signal yield	$\tilde{\chi}^2_{fit}$
10%	7758 ± 208	1.37
8%	7619 ± 206	1.27
6%	7483 ± 203	1.21
4%	7348 ± 201	1.19
Fit estimate of free fraction	Total signal yield	$\tilde{\chi}^2_{fit}$
$3.8\% \pm 2.9\%$	7331 ± 281	1.23

Table C.1: Four fits with an injected Cabibbo-Suppressed (CS) fraction of the total signal are compared with the fit where the CS-fraction is left free, namely with the possibility to be in the widest possible interval, i.e. $[0.1, 99]\%$, starting from the initial value of 8%. The BML fit with the free CS-fraction has one more degree of freedom with respect to the other fits but its effect on $\tilde{\chi}^2_{fit}$ is marginal. These results are obtained when using the `XGBoost`-based inclusive selection model discussed earlier.

An alternative modelization of the reflection shape, using two Gaussians with common mean instead of a Johnson p.d.f. model has also been tried and a similar fit to the data provides an estimate of the CS-fraction compatible with zero when left free to float in the BML fit.

C.2 Estimation of the $B_c \to J/\psi K$ contribution

In the following study a cut-based (`displaced`) selection borrowed from the CMS paper in [82] is applied to the $B_c \to J/\psi K$ signal MC sample and to the 2022-23 datasets. The followed procedure is exactly the same described in the previous Section. Briefly, to derive the CS-reflection shape, a Johnson p.d.f. model is - again - successfully used as CS-signal model in the BML fit of the $J/\psi\pi$ invariant mass obtained by applying the mentioned selection to the $B_c \to J/\psi K$ simulated sample, as shown in Fig. C.3. Then, when fitting the invariant mass distribution of the $J/\psi\pi$ final state in the Run-3 data, shown in Fig. C.4, the additional Cabibbo-Suppressed component is modelled with the Johnson p.d.f. model with its parameters settled to those estimated from the simulation, and leaving only the normalization (i.e. the relative yield fraction with respect to the total signal, namely Cabibbo-Favoured plus Cabibbo-Suppressed) as free parameter. This parameter turns out to be relatively small but still somehow different from zero: $(4.8 \pm 3.2)\%$. It is also consistent with the estimated value obtained in the previous Section, i.e. $(3.8 \pm 2.9)\%$. These two approaches suggest the presence of a marginal Cabibbo-Suppressed signal in the analyzed CMS datasets, on the low-side of the CS-signal fixed contribution considered in the CMS paper of [82] , but still compatible with the latter.

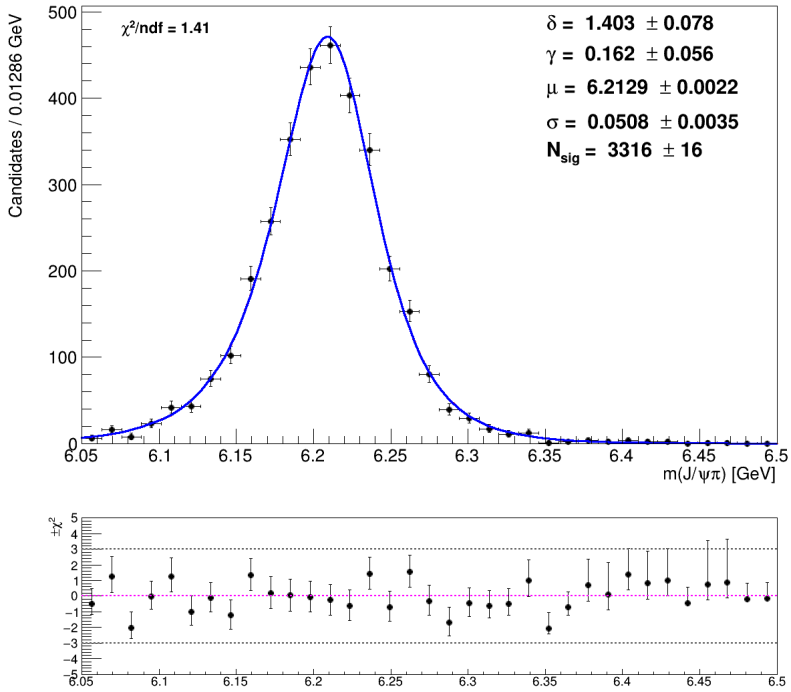


Figure C.3: BML fit to the reconstructed $J/\psi\pi$ invariant mass obtained by the selected simulated $B_c \to J/\psi K$ signal after matching with the generated candidates and applying a cut-based selection. The resulting shape is slightly asymmetric (γ compatible with 0 only at 3σ) but shows a clear left tail (δ is definitely different from 0).

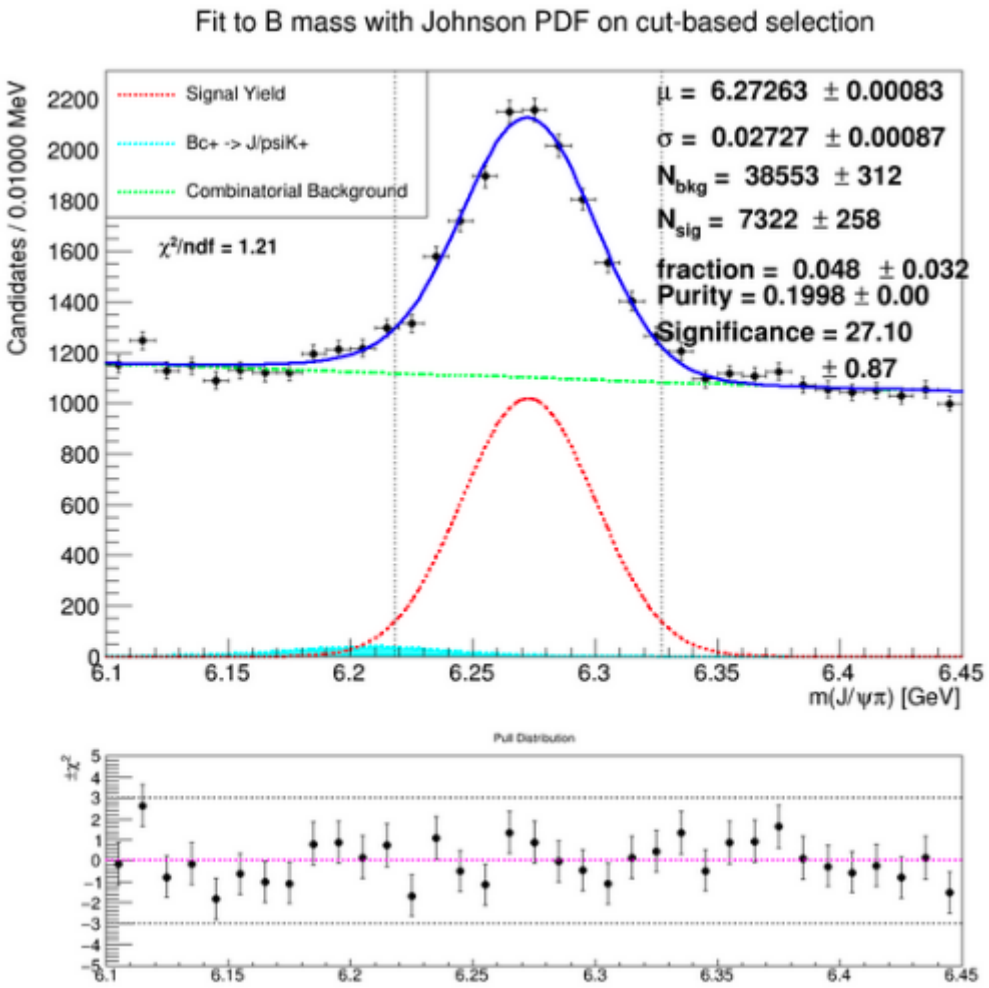


Figure C.4: The invariant mass of the final state $J/\psi\pi$ in the Run-3 data undergoes a BML fit with a model including a Cabibbo-Suppressed contribution (cyan) with the shape borrowed from the fit shown in Fig. C.3 and its normalization left reasonably free to float. The dotted red component represents the Cabibbo-Favoured signal, whereas the green dotted component represents the exponential (combinatorial) background. The relative fraction with respect to the total signal (Cabibbo-Favoured plus Cabibbo-suppressed) yield is estimated to be $(4.8 \pm 3.2)\%$.

Injected CS fraction	Total signal yield	$\tilde{\chi}^2_{fit}$
10%	7589 ± 202	1.27
8%	7484 ± 201	1.21
6%	7382 ± 200	1.18
4%	7279 ± 199	1.17
Fit estimate of free fraction	Total signal yield	$\tilde{\chi}^2_{fit}$
$4.8\% \pm 3.2\%$	7326 ± 259	1.21

Table C.2: Four fits with an injected Cabibbo-Suppressed (CS) fraction of the total signal are compared with the fit where the CS-fraction is left free, namely with the possibility to be in the widest possible interval, i.e. $[0.1, 99]\%$, starting from the initial value of 8%. The BML fit with the free CS-fraction has one more degree of freedom with respect to the other fits but its effect on $\tilde{\chi}^2_{fit}$ is marginal. These results are obtained when applying a cut-based selection almost identical to the one used in the [82] .

Other fits have also been tried by injecting a fixed amount of Cabibbo-Suppressed contribution and evaluating the quality of the resulting fit. The results of this procedure are summarized in the Table C.2. The CS-fraction result provided by the free fit is consistent with the trends of the total yield estimate and the normalized chi-squared of the BML fit when reducing the injected amount of the CS relative contribution from 10% to 4%.

Also in this case an alternative modelization of the reflection shape, using two Gaussians with common mean instead of a Johnson function, has been tried; the corresponding BML fit to the data provides an estimate of the CS-fraction compatible with zero when left free to float in the fit. In general the Johnson function seems to be better representing the Cabibbo-Suppressed reflection shape.

Appendix D

Glossary of Variables

In this appendix, a detailed descriptions of the key variables used in this study is provided. Each observable is listed with its physical representation, the corresponding variable name in the code, and a brief explanation of its meaning. This information is meant to help understanding the role and importance of each variable in the analysis. The variables are presented in alphabetical order for easy reference.

- $\cos(\alpha^{3D}_{B_c})$, `"bCosAlpha_bestDist"`: The selection on the candidate's pointing angle, computed for each candidate with respect to its chosen PV.
- $L_{xy}/\sigma_{L_{xy}}$, `"bFlightLenSig_bestDist"`: the ratio of the displacement in the transverse plane from the beamspot over its error.
- $|\eta(\mu\mu trk)|$, `"B_eta"`: Absolute value of the pseudorapidity of the combined system of the dimuon ($\mu\mu$) and the additional track (trk).
- $P_{vtx}(\mu\mu trk)$, `"B_Prob"`: Vertex probability of the combined system of the dimuon ($\mu\mu$) and the additional track (trk).
- $p_T(\mu\mu trk)$, `"B_pt"`: Transverse momentum of the combined system of the dimuon ($\mu\mu$) and the additional track (trk).
- $\cos(\alpha_{\mu\mu trk})$, `"cosAngle_J_pi"`: Cosine of dimuon-plus-track pointing angle.
- $\Delta R(J/\psi, track)$, `"dR_J_pi"`: The distance between the J/ψ and the track.
- $\frac{d_0(trk)}{\sigma_{d_0}}$, `"d0SigPi"`: Significance of the $d_0(trk)$, which is the transverse displacement of the track from the beam line.
- $\Delta R(\mu^+\mu^-)$, `"deltaRmumu"`: The distance between μ^+ and μ^- that form the J/ψ meson.

- $|\eta(\mu^+\mu^-)|$, `"J_eta"`: Represents the absolute pseudorapidity of the J/ψ meson, composed of the two muons.
- $P_{vtx}(\mu^+\mu^-)$, `"J_Prob"`: The vertex probability of the J/ψ, representing the fit quality of the dimuon vertex.
- $p_T(\mu^+\mu^-)$, `"J_pt"`: Represents the transverse momentum of the J/ψ meson, composed of the two muons.
- $p_T(\mu^\pm)$, `"J_pt1"`: : Represents the transverse momentum of one of the muons constituting the J/ψ meson.
- $p_T(\mu^\pm)$, `"J_pt2"`: Represents the transverse momentum of the other muon constituting the J/ψ meson.
- $\cos(\alpha^{2D}_{\mu^+\mu^-})$, "jCosAlpha_bestDist": The cosine of the dimuon pointing angle. The angle in the transverse plane between the dimuon momentum and the flight distance.
- $|\eta(trk)|$, `"pi_eta"`: The absolute value of the pseudorapidity of the track.
- $p_T(trk)$, `"pi_pt"`: Transverse momentum of the additional track.

After the `preprocessing` step, each variable is visualized by comparing the signal and sideband distributions. These visualizations provide a clear comparison between the signal and background characteristics for each variable, highlighting the effectiveness of the `preprocessing` step in distinguishing between signal and background events. The following figures illustrate the distributions of each variable for both signal and sideband samples.

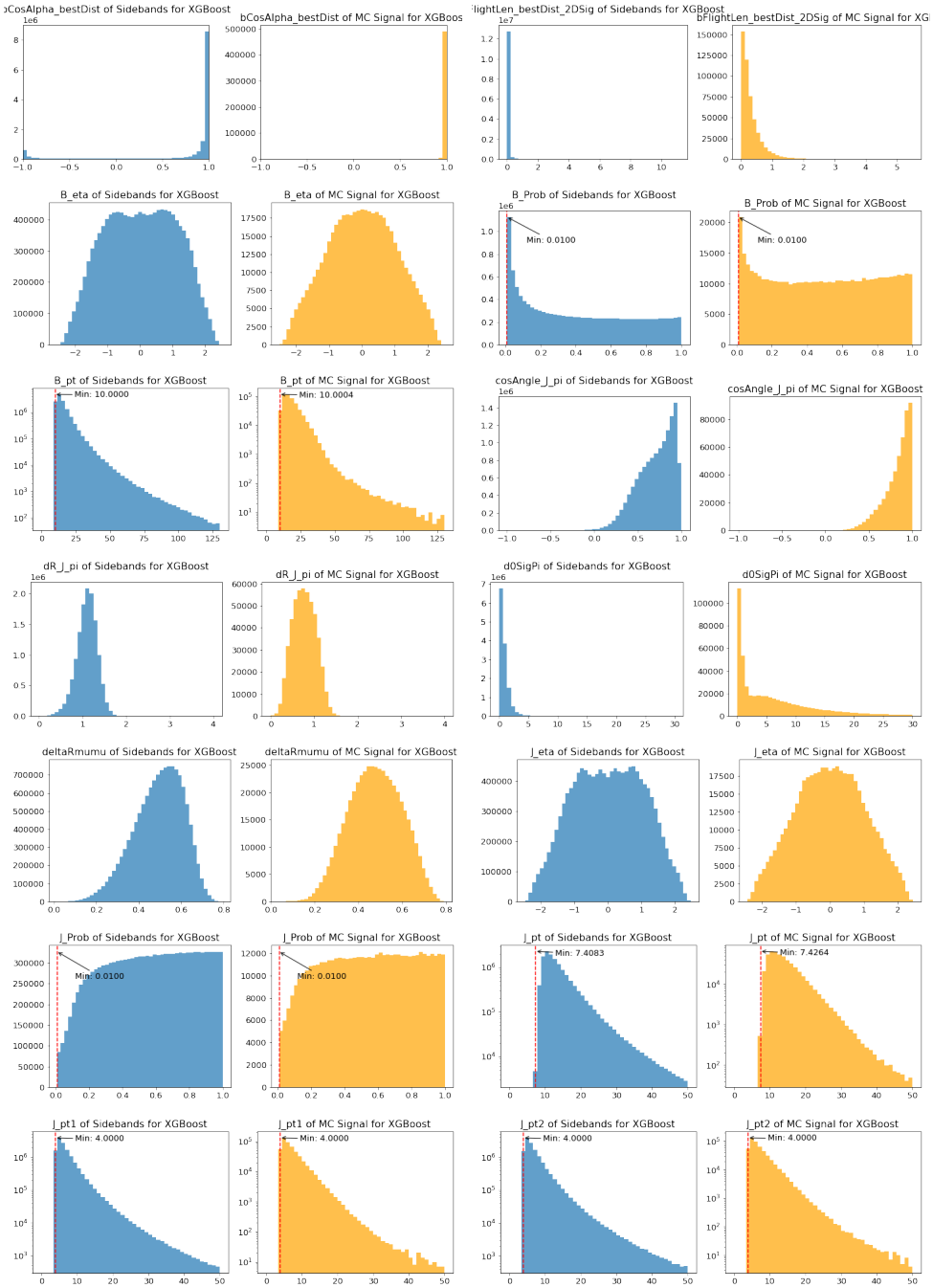


Figure D.1: Variable distributions for the B_c meson. Blue represents sidebands distributions, while orange represents signal distributions.

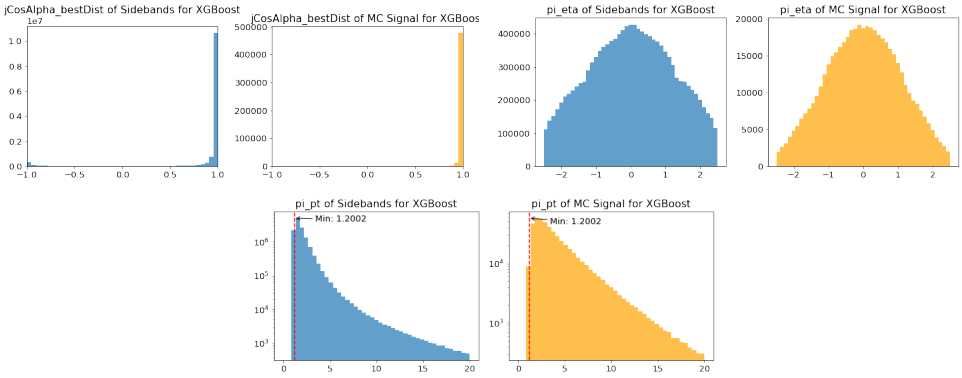


Figure D.2: Variable distributions for the B_c meson. Blue represents sidebands distributions, while orange represents signal distributions.

Bibliography

[1] R. L. Workman et al. "Review of Particle Physics". In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[2] Roel Aaij et al. "Observation of the doubly charmed baryon Ξ_{cc}^{++}". In: *Phys. Rev. Lett.* 119.11 (2017), p. 112001. DOI: 10.1103/PhysRevLett.119.112001. arXiv: 1707.01621 [hep-ex].

[3] Benjamin Fuks et al. "Signatures of toponium formation in LHC run 2 data". In: *Phys. Rev. D* 104.3 (2021), p. 034023. DOI: 10.1103/PhysRevD.104.034023. arXiv: 2102.11281 [hep-ph].

[4] F. Englert and R. Brout. "Broken symmetry and the mass of gauge vector mesons". In: *Phys. Rev. Lett.* 13 (1964), p. 321. DOI: 10.1103/PhysRevLett.13.321.

[5] P. W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), p. 132. DOI: 10.1016/0031-9163(64)91136-9.

[6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global conservation laws and massless particles". In: *Phys. Rev. Lett.* 13 (1964), p. 585. DOI: 10.1103/PhysRevLett.13.585.

[7] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[8] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys.Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[9] Murray Gell-Mann. "A Schematic Model of Baryons and Mesons". In: *Phys. Lett.* 8 (1964), pp. 214–215. DOI: 10.1016/S0031-9163(64)92001-3.

[10] G Zweig. "An SU_3 model for strong interaction symmetry and its breaking; Version 2". In: CERN-TH-412 (Feb. 1964). Version 1 is CERN preprint 8182/TH.401, Jan. 17, 1964, 80 p. URL: http://cds.cern.ch/record/570209.

[11] Chen-Ning Yang and Robert L. Mills. "Conservation of Isotopic Spin and Isotopic Gauge Invariance". In: *Phys. Rev.* 96 (1954). [,150(1954)], pp. 191–195. DOI: 10.1103/PhysRev.96.191.

[12] R. L. Workman et al. *Review of Particle Physics*. 2022. URL: https://pdg.lbl.gov/2022/reviews/rpp2022-rev-qcd.pdf.

[13] Vardan Khachatryan et al. "Measurement of the inclusive 3-jet production differential cross section in proton–proton collisions at 7 TeV and determination of the strong coupling constant in the TeV range". In: *Eur. Phys. J. C* 75.5 (2015), p. 186. DOI: 10.1140/epjc/s10052-015-3376-y. arXiv: 1412.1633 [hep-ex].

[14] Robert L. Jaffe. "Multi-Quark Hadrons. 1. The Phenomenology of (2 Quark 2 anti-Quark) Mesons". In: *Phys. Rev.* D15 (1977), p. 267. DOI: 10.1103/PhysRevD.15.267.

[15] Robert L. Jaffe. "Multi-Quark Hadrons. 2. Methods". In: *Phys. Rev.* D15 (1977), p. 281. DOI: 10.1103/PhysRevD.15.281.

[16] F. Abe et al. "Observation of the B_c meson in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV". In: *Phys. Rev. Lett.* 81 (1998), pp. 2432–2437. DOI: 10.1103/PhysRevLett.81.2432. arXiv: hep-ex/9805034.

[17] T. Aaltonen et al. "Observation of the Decay B^+ -(c) $\to J/\psi\pi^\pm$ and Measurement of the B^+ -(c) Mass". In: *Phys. Rev. Lett.* 100 (2008), p. 182002. DOI: 10.1103/PhysRevLett.100.182002. arXiv: 0712.1506 [hep-ex].

[18] Chao-Hsi Chang, Yu-Qi Chen, and Robert J. Oakes. "Comparative study of the hadronic production of B(c) mesons". In: *Phys. Rev. D* 54 (1996), pp. 4344–4348. DOI: 10.1103/PhysRevD.54.4344. arXiv: hep-ph/9602411.

[19] N. Brambilla et al. "Heavy quarkonium physics". In: (Dec. 2004). DOI: 10.5170/CERN-2005-005. arXiv: hep-ph/0412158.

[20] Chao-Hsi Chang and Xing-Gang Wu. "Uncertainties in estimating hadronic production of the meson B_c and comparisons between TEVATRON and LHC". In: *Eur. Phys. J. C* 38 (2004), pp. 267–276. DOI: 10.1140/epjc/s2004-02015-0. arXiv: hep-ph/0309121.

[21] Estia J. Eichten and Chris Quigg. "Mesons with Beauty and Charm: New Horizons in Spectroscopy". In: *Phys. Rev. D* 99.5 (2019), p. 054025. DOI: 10.1103/PhysRevD.99.054025. arXiv: 1902.09735 [hep-ph].

[22] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175.

[23] Chao-Hsi Chang et al. "BCVEGPY: An Event generator for hadronic production of the B_c meson". In: *Comput. Phys. Commun.* 159 (2004), pp. 192–224. DOI: 10.1016/j.cpc.2004.02.005. arXiv: hep-ph/0309120.

[24] Chao-Hsi Chang, Jian-Xiong Wang, and Xing-Gang Wu. "BCVEGPY2.0: A Upgrade version of the generator BCVEGPY with an addendum about hadroproduction of the P-wave B(c) states". In: *Comput. Phys. Commun.* 174 (2006), pp. 241–251. DOI: 10.1016/j.cpc.2005.09.008. arXiv: hep-ph/0504017.

[25] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth. A* 462 (2001). Ed. by S. Erhan, P. Schlein, and Y. Rozen, pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[26] S. Agostinelli et al. "G4–a simulation toolkit". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250 –303. ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)01368-8. URL: http://www.sciencedirect.com/science/article/B6TJM-48TJFY8-5/2/23ea98096ce11c1be446850c04cfa498.

[27] Stephen Godfrey. "Spectroscopy of B_c mesons in the relativized quark model". In: *Phys. Rev. D* 70 (5 Sept. 2004), p. 054017. DOI: 10.1103/PhysRevD.70.054017. URL: https://link.aps.org/doi/10.1103/PhysRevD.70.054017.

[28] Albert M Sirunyan et al. "Observation of Two Excited B_c^+ States and Measurement of the $B_c^+(2S)$ Mass in pp Collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 122.13 (2019), p. 132001. DOI: 10.1103/PhysRevLett.122.132001. arXiv: 1902.00571 [hep-ex].

[29] Ikaros I. Y. Bigi and Nikolai Uraltsev. "Pauli interference in the 't Hooft model: Heavy quark expansion and quark hadron duality". In: *Phys. Lett. B* 457 (1999), pp. 163–169. DOI: 10.1016/S0370-2693(99)00540-7. arXiv: hep-ph/9903258.

[30] V. V. Kiselev. "Exclusive decays and lifetime of B_c meson in QCD sum rules". In: (Nov. 2002). arXiv: hep-ph/0211021.

[31] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[32] Roel Aaij et al. "Measurement of the B_c^+ meson lifetime using $B_c^+ \to J/\psi\mu^+\nu_\mu X$ decays". In: *Eur. Phys. J. C* 74.5 (2014), p. 2839. DOI: 10.1140/epjc/s10052-014-2839-x. arXiv: 1401.6932 [hep-ex].

[33] Roel Aaij et al. "Measurement of the lifetime of the B_c^+ meson using the $B_c^+ \to J/\psi\pi^+$ decay mode". In: *Phys. Lett. B* 742 (2015), pp. 29–37. DOI: 10.1016/j.physletb.2015.01.010. arXiv: 1411.6899 [hep-ex].

[34] A. V. Berezhnoy, I. N. Belov, and A. K. Likhoded. "Production of D-wave states of $\bar{b}c$ quarkonium at the LHC". In: *Phys. Rev. D* 103.11 (2021), p. 114001. DOI: 10.1103/PhysRevD.103.114001. arXiv: 2101.09784 [hep-ph].

[35] A. M. Sirunyan et al. "Observation of the $\chi_{b1}(3P)$ and $\chi_{b2}(3P)$ and measurement of their masses". In: *Phys. Rev. Lett.* 121 (2018), p. 092002. DOI: 10.1103/PhysRevLett.121.092002. arXiv: 1805.11192 [hep-ex].

[36] Pablo G. Ortega et al. "Spectroscopy of B_c mesons and the possibility of finding exotic B_c-like structures". In: *The European Physical Journal C* 80.3 (Mar. 2020). DOI: 10.1140/epjc/s10052-020-7764-6. URL: https://doi.org/10.1140\%2Fepjc\%2Fs10052-020-7764-6.

[37] LHCb Collaboration. "Observation of the doubly charmed baryon decay $\Xi_{cc}^{++} \to \Xi_c'^+\pi^+$". In: *Journal of High Energy Physics* 2022.5 (May 2022). DOI: 10.1007/jhep05(2022)038. URL: https://doi.org/10.1007\%2Fjhep05\%282022\%29038.

[38] Ahmed Ali, Qin Qin, and Wei Wang. "Discovery potential of stable and near-threshold doubly heavy tetraquarks at the LHC". In: *Phys. Lett. B* 785 (2018), pp. 605–609. DOI: 10.1016/j.physletb.2018.09.018. arXiv: 1806.09288 [hep-ph].

[39] T. Gershon and A. Poluektov. "Displaced B_c^- mesons as an inclusive signature of weakly decaying double beauty hadrons". In: *JHEP* 01 (2019), p. 019. DOI: 10.1007/JHEP01(2019)019. arXiv: 1810.06657 [hep-ph].

[40] Alexis Pompili. *Search for double beauty hadrons*. Presentation at the CMS B-Physics Jamboree for Run-III, 15.12.2020.

[41] Alexander K. Ridgway and Mark B. Wise. "An Estimate of the Inclusive Branching Ratio to $\bar{B}_c$ in Ξ_{bbq} Decay". In: *Phys. Lett. B* 793 (2019), pp. 181–184. DOI: 10.1016/j.physletb.2019.04.020. arXiv: 1902.04582 [hep-ph].

[42] S. Chatrchyan et al. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[43] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001. URL: http://stacks.iop.org/1748-0221/3/i=08/a=S08001.

[44] G. Antchev et al. "First measurement of elastic, inelastic and total cross-section at $\sqrt{s} = 13$ TeV by TOTEM and overview of cross-section data at LHC energies". In: *Eur. Phys. J.* C79.2 (2019), p. 103. DOI: 10.1140/epjc/s10052-019-6567-0. arXiv: 1712.06153 [hep-ex].

[45] Gábor I. Veres. "Soft QCD at the CMS And ATLAS Experiments". In: *Proceedings, 51st Rencontres de Moriond on QCD and High Energy Interactions: La Thuile, Italy, March 19-26, 2016*. ARISF. ARISF, 2016, pp. 143–146.

[46] Merijn H. F. Van De Klundert. "Soft QCD at CMS and ATLAS". In: *Proceedings, 52nd Rencontres de Moriond on QCD and High Energy Interactions: La Thuile, Italy, March 25-April 1, 2017*. CERN. Geneva: CERN, 2017, pp. 231–234.

[47] CMS Collab. *CMS Luminosity - Public Results*. 2023. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[48] CMS Collab. *CMS OMS Website*. 2023. URL: https://cmsoms.cern.ch/cms/.

[49] Tai Sakuma and Thomas McCauley. "Detector and Event Visualization with SketchUp at the CMS Experiment". In: *J. Phys. Conf. Ser.* 513 (2014), p. 022032. DOI: 10.1088/1742-6596/513/2/022032. arXiv: 1311.4942 [physics.ins-det].

[50] *SketchUpCMS*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SketchUpCMS.

[51] Matthias Schott and Monica Dunford. "Review of single vector boson production in pp collisions at $\sqrt{s} = 7$ TeV". In: *Eur. Phys. J.* C74 (2014), p. 2916. DOI: 10.1140/epjc/s10052-014-2916-1. arXiv: 1405.1160 [hep-ex].

[52] David Aaron Matzner Dominguez et al. "CMS Technical Design Report for the Pixel Detector Upgrade". In: (2012). DOI: 10.2172/1151650.

[53] Vardan Khachatryan et al. "Measurement of the CP-violating weak phase ϕ_s and the decay width difference $\Delta\Gamma_s$ using the $B_s^0 \to J/\psi\phi(1020)$ decay channel in pp collisions at $\sqrt{s} = 8$ TeV". In: *Phys. Lett.* B757 (2016), pp. 97–120. DOI: 10.1016/j.physletb.2016.03.046. arXiv: 1507.07527 [hep-ex].

[54] Unofficial preliminary result. Courtesy of Giacomo Fedi. *AN-2019-035*.

[55] Armen Tumasyan et al. "The Phase-2 Upgrade of the CMS Tracker". In: (June 2017). Ed. by K. Klein. DOI: 10.17181/CERN.QZ28.FLHW.

[56] CMS Collaboration. "Development of the CMS detector for the CERN LHC Run 3". In: *JINST* 19.05 (2024), P05064. DOI: 10.1088/1748-0221/19/05/P05064. arXiv: 2309.05466. URL: https://cds.cern.ch/record/2870088.

[57] CMS Collaboration et al. "Commissioning of CMS HLT". In: *Journal of Instrumentation* 5 (Mar. 2010), T03005. DOI: 10.1088/1748-0221/5/03/T03005.

[58] T. Speer et al. "Track reconstruction in the CMS tracker". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 559.1 (2006). Proceedings of the X International Workshop on Advanced Computing and Analysis Techniques in Physics Research ACAT 05X International Workshop on Advanced Computing and Analysis Techniques, pp. 143 –147. ISSN: 0168-9002. DOI: http://dx.doi.org/10.1016/j.nima.2005.11.207. URL: http://www.sciencedirect.com/science/article/pii/S0168900205022576.

[59] Steven Lantz et al. "Speeding up particle track reconstruction using a parallel Kalman filter algorithm". In: *JINST* 15.09 (2020), P09030. DOI: 10.1088/1748-0221/15/09/P09030. arXiv: 2006.00071 [physics.ins-det].

[60] *Performance of Run 3 track reconstruction with the mkFit algorithm.* https://cds.cern.ch/record/2814000 and https://twiki.cern.ch/twiki/bin/view/CMSPublic/TRKmkFitRun3. 2022.

[61] *Performance of Run-3 HLT Track Reconstruction.* https://cds.cern.ch/record/2814111 and https://twiki.cern.ch/twiki/bin/view/CMSPublic/Run3HLTTracking. 2022.

[62] K. Rose. "Deterministic annealing for Clustering, COmpression, CLassification, Regression and related Optimisation Problems". In: *Proc.IEEE* 86 (1998).

[63] W. Waltenberger et al. "Adaptive vertex fitting". In: *J. Phys. G* 34 (12) (2007). URL: http://stacks.iop.org/0954-3899/34/i=12/a=NO1.

[64] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[65] Albert M Sirunyan et al. "Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 16 (2021), P07001. DOI: 10.1088/1748-0221/16/07/P07001. arXiv: 2102.04790 [hep-ex].

[66] A. M. Sirunyan et al. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[67] *Tracking performance using Tag and Probe with $Z \to \mu^+\mu^-$ in 2022 and 2023*. https://cds.cern.ch/record/2904367. 2024.

[68] CMS Collaboration. "Measurement of $B_c(2S)^+$ and $B_c^*(2S)^+$ cross section ratios in proton-proton collisions at $\sqrt{s} = 13$TeV". In: *Phys. Rev. D* 102 (9 Nov. 2020), p. 092007. DOI: 10.1103/PhysRevD.102.092007. URL: https://link.aps.org/doi/10.1103/PhysRevD.102.092007.

[69] Albert M Sirunyan et al. "Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J. C* 78.6 (2018). [Erratum: Eur.Phys.J.C 78, 561 (2018)], p. 457. DOI: 10.1140/epjc/s10052-018-5929-3. arXiv: 1710.08949 [hep-ex].

[70] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System". In: *CoRR* abs/1603.02754 (2016). arXiv: 1603.02754. URL: http://arxiv.org/abs/1603.02754.

[71] A. N. Kolmogorov. "Sulla Determinazione Empirica di una Legge di Distribuzione". In: *Giornale dell'Istituto Italiano degli Attuari* 4 (1933), pp. 83–91.

[72] Hartigan, Patrick. *Critical Values for the Two-sample Kolmogorov-Smirnov test (2-sided)*. 2013. URL: https://sparky.rice.edu/astr360/kstest.pdf.

[73] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[74] ROOT. *Johnson's SU-distribution — Wikipedia, The Free Encyclopedia*. https://root.cern/doc/master/classRooJohnson.html.

[75] Takuya Akiba et al. "Optuna: A Next-generation Hyperparameter Optimization Framework". In: (July 2019). arXiv: 1907.10902 [cs.LG].

[76] Srinath Bulusu et al. "Equivariance and generalization in neural networks". In: *EPJ Web Conf.* 258 (2022), p. 09001. DOI: 10.1051/epjconf/202225809001. arXiv: 2112.12493 [hep-lat].

[77] Cheng-Wei Chiang et al. "Deep Learning to Improve the Sensitivity of Di-Higgs Searches in the $4b$ Channel". In: (Jan. 2024). arXiv: 2401.14198 [hep-ph].

[78] James Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in neural information processing systems* 24 (2011).

[79] *Pseudo-significance for cut optimization (Punzi criterion)*. URL: https://twiki.cern.ch/twiki/bin/viewauth/CMS/PunziFom.

[80] *seaborn: statistical data visualization*. URL: https://seaborn.pydata.org/index.html.

[81] Roel Aaij et al. "Measurement of the ratio of branching fractions $\mathcal{B}(B_c^+ \to J/\psi K^+)/\mathcal{B}(B_c^+ \to J/\psi \pi^+)$". In: *JHEP* 09 (2016), p. 153. DOI: 10.1007/JHEP09(2016)153. arXiv: 1607.06823 [hep-ex].

[82] Albert M Sirunyan et al. "Measurement of $\mathrm{B_c}(2\mathrm{S})^+$ and $\mathrm{B_c^*}(2\mathrm{S})^+$ cross section ratios in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. D* 102.9 (2020), p. 092007. DOI: 10.1103/PhysRevD.102.092007. arXiv: 2008.08629 [hep-ex].