



DOCTORAL THESIS

Study of $B \rightarrow \mu^+ \mu^-$ decays with the ATLAS detector

*A thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy*

in the

Experimental Particle Physics Research Group
School of Mathematical and Physical Sciences

Author:

Fabio TRESOLDI

Supervisor:

Prof. Alessandro CERRI

14th January 2020



ACKNOWLEDGEMENTS

The work of presented in this thesis is the final result of an almost four years long effort; I could not have made it without all the poeple around me, both academically and personally speaking.

First, I would like to thank my supervisor, Alex. Thanks for the exceptional supervision, for all the hours spent talking about statistics, for the incredible dinners and for all the cocktails. Sussex would not be the same without the presence of Iacopo, Antonella, Fabrizio and Lily; thanks to all of you for all the support you gave me in the last years.

A special thanks also to Umberto, who helped both while he was at Sussex and also after he moved to Rome.

My time in Brighton would not have been the same without all the other PhD students. I would start with Giuseppe, even if we managed to be almost constantly in different countries, he has always been a great company. A special thank also to mini-Fab, with whom I spent many days in the office and many nights out or on Skype. The other two ATLAS PhD of my years deserve as special mention as well. Thanks to Emma for all the time we spent together in the office, in some random pub and with beers in front of us; thanks also to Sam, I have to admit, I will never get a better bit of butter on my knife. A huge thank also to all the other students: Ioannis, Luigi, Diana, Mario S, Marco, Mario G, Dani, EW-Fab, Meirin, Nicola, Olly, Suf and Zara.

During my PhD I also had the opportunity to spend a lot of time at CERN where I met a lot of extraordinary people. A special thank to Carlo, Barbosa, Nacho, Jose, Ceci and Roger and thanks also to all the others. I would like to mention everyone, but I would definitely end up forgetting someone.

Thanks also to all my friends in Italy: the physicists, the skateboarders and my friends from high school. Even if we have not met often in the last years, your presence is a constant help.

A massive thank to my parents, to my brother Daniele and Mattia, who have always believed in me. Grazie per avermi sempre supportato, non ce l'avrei mai fatta senza di voi.

Finally, thank you Andrea. I honestly do not know how to say how lucky I feel to be with you. Thank you for all the time we spent together and also for the time we spent living away from each other. Gracias por siempre estar ahí.

STATEMENT

I, Fabio TRESOLDI, hereby declare that this thesis has not been and will not be, submitted in whole or in part to another university for the award of any other degree.

*Brighton,
14th January 2020*

Fabio Tresoldi.....
Fabio TRESOLDI

University of Sussex
School of Mathematical and Physical Sciences
Experimental Particle Physics Research Group

DOCTORAL THESIS

Study of $B \rightarrow \mu^+ \mu^-$ decays with the ATLAS detector

by Fabio TRESOLDI

ABSTRACT

This thesis presents the search for the very rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$, based on 26.3 fb^{-1} of $\sqrt{s} = 13 \text{ TeV}$ LHC proton–proton collision data collected in 2015 and 2016 by the ATLAS experiment. $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are considered one of the pillars of the ATLAS experiment searches for New Phenomena with flavour, thanks to an extremely clean experimental signature, a very small and precisely constrained Standard Model (SM) amplitude and the potential sizable interference from New Physics phenomena.

This work greatly improves past ATLAS results, including a new dataset, improved statistical and systematic extraction techniques, up-to-date muon identification tools and background models. The analysis results in the experiment’s most precise measurement to date of the branching fraction $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.2_{-1.0}^{+1.1}\right) \times 10^{-9}$ and upper limit

$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-10}$ at 95% confidence level.

Combining with the ATLAS Run 1 analysis results in $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(2.8_{-0.7}^{+0.8}\right) \times 10^{-9}$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 2.1 \times 10^{-10}$ at 95% confidence level: the most stringent experimental limit available to this date. All measurements are compatible with the SM prediction.

Projections on the future ATLAS sensitivity are also performed.

CONTENTS

Introduction	1
1 Theoretical background	3
1.1 The Standard Model of particle physics	3
1.1.1 Overview	3
1.1.2 Quantum Electrodynamics	5
1.1.3 Electroweak unification	6
1.1.4 The Higgs Mechanism	7
1.1.5 Quantum chromodynamics	9
1.1.6 Quark mixing and the CKM matrix	10
1.2 Limitations of the Standard Model	11
1.3 Dimuon B decays	13
1.3.1 Branching fraction and B mixing	14
1.3.2 Branching fractions computation	16
1.3.3 The SM branching fractions	19
1.4 New physics and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$	19
1.5 Experimental state of the art	21
2 Experimental apparatus	24
2.1 The LHC	24
2.1.1 Acceleration stages	25
2.1.2 Performance of the Large Hadron Collider (LHC)	26
2.2 The ATLAS detector	27
2.2.1 The ATLAS coordinate system	28
2.2.2 The magnet system	29
2.2.3 The Inner Detector	30
2.2.4 The calorimeters	32
2.2.5 The Muon Spectrometer	34
2.2.6 Trigger system	36
2.3 The HL-LHC upgrade	38

3	The ATLAS data analysis and simulation frameworks	41
3.1	Monte Carlo simulation	41
3.1.1	Simulation of physics processes	41
3.1.2	Detector simulation	46
3.2	Object reconstruction	47
3.2.1	Track reconstruction	47
3.2.2	Vertex reconstruction	48
3.2.3	Muons	49
3.2.4	Electrons and photons	50
3.2.5	Jets	51
3.2.6	Taus	52
3.2.7	Missing transverse energy E_T^{miss}	52
4	Studies for trigger improvements	53
4.1	The FTK	53
4.2	Case of study	54
4.3	Samples employed	55
4.3.1	Signal sample	55
4.3.2	Background sample	57
4.4	Trigger selection optimisation	58
4.5	Trigger rate estimation	62
4.6	$B \rightarrow hh'$ signal yield estimation	64
4.7	Results and conclusions	65
5	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis strategy	67
5.1	Expected sensitivity	70
6	Statistical tools	72
6.1	Likelihood function	72
6.2	Hypothesis test and p -value	74
6.2.1	The profiled likelihood ratio	75
6.3	Uncertainties and confidence intervals	76
6.3.1	RCF variance method	77
6.3.2	Likelihood intervals	78
6.3.3	Neyman belt construction	78
6.3.4	Neyman belt implementation	80
6.3.4.1	One-dimensional belt	82
6.3.4.2	Two-dimensional belt	83
6.3.4.3	Systematic uncertainties	85
7	Data and MC samples	88
7.1	Data and trigger selection	88
7.2	MC	90

7.3	Event preselection	91
7.4	Studies on fake muons rate	94
7.5	Primary vertex association	96
7.6	MC samples reweighting	99
7.6.1	Pile-up reweighting	100
7.6.2	Muon offline efficiency weights	101
7.6.3	Muon trigger weights	101
7.6.4	Kinematic corrections	102
7.6.4.1	Quark Level Corrections	103
7.6.4.2	Data Driven Weights	107
8	Background sources and reduction	119
8.1	Main analysis backgrounds	119
8.1.1	Continuum background	120
8.1.2	$B \rightarrow hh'$ background	121
8.1.3	Partially reconstructed background	121
8.2	Background reduction	123
8.2.1	Boosted Decision Tree (BDT) description	124
8.2.2	BDT input variables	125
8.2.3	BDT training and testing	128
9	Data-MC comparison	133
9.1	Reference channel comparison	133
9.2	Control channel comparison	136
9.3	Continuum background comparison	141
10	Reference channel	144
10.1	$B^+ \rightarrow J/\psi K^+$ yield extraction	144
10.1.1	Samples composition	144
10.1.2	Reference channel fit configuration	145
10.1.3	Systematic uncertainties on the $B^+ \rightarrow J/\psi K^+$ fit	147
10.1.4	Result of the reference channel yield extraction	148
10.2	Evaluation of the $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency ratio	149
10.2.1	Efficiency ratio calculation	149
10.2.2	Systematic uncertainties on R_{Ae}	151
10.2.3	Systematic uncertainties on BDT bins	154
11	Signal yield extraction	157
11.1	Non-resonant background models	158
11.1.1	Non-resonant background components	160
11.1.2	Non-resonant backgrounds parameterisation	161
11.1.2.1	Continuum background	162
11.1.2.2	Same-Side and Same-Vertex background	162

11.1.2.3	B_c background	163
11.1.2.4	Semileptonic background	164
11.1.3	Fit to MC and data sidebands with combined model	166
11.1.4	Fit to MC and data sidebands with simultaneous model	169
11.1.5	Summary of the non-resonant background configuration	172
11.2	Signal and peaking background models	173
11.2.1	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ models	173
11.2.2	Peaking background	174
11.2.2.1	Peaking background model	175
11.2.2.2	Normalisation of the $B \rightarrow hh'$ background	177
11.3	Signal yield extraction	180
11.3.1	Fit configuration	182
11.3.2	Fit convergence properties	183
11.3.3	Fit model systematics	187
11.4	Branching fraction extraction approach	190
11.5	Conclusions	192
12	Branching fraction extraction	194
12.1	2015/16 dataset branching fractions extraction	194
12.1.1	Fit on unblinded data	195
12.1.2	Unblinded fit quality	197
12.1.3	Branching fractions uncertainty	197
12.1.4	2015/16 analysis result	202
12.2	Combination with the Run 1 result	202
12.2.1	Comparison with the Run 1 result	202
12.2.2	Analyses combination procedure	203
12.2.3	Branching fraction extraction	204
12.2.4	Comparison with other experiments	207
12.3	Summary of results	208
12.4	LHC experiment $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ combination	209
13	Extrapolations for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis	212
13.1	Projection for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis	212
13.1.1	Extrapolation procedure	212
13.1.1.1	Systematic uncertainties	214
13.1.2	Cross-check on Run 1 result	215
13.1.3	Run 2 yield extrapolation	216
13.1.4	HL-LHC yield extrapolation	217
13.1.5	Results	218
13.1.5.1	Run 2 projections	218
13.1.5.2	HL-LHC projections	219
13.1.6	Comparison with other analyses	220

13.2 ATLAS-CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses comparison	224
13.2.1 Comparison procedure	225
13.2.2 Results	227
13.2.3 Conclusions	228
Conclusions	230
A The ATLAS Run 1 analysis	232
A.1 Analysis overview	232
A.2 Background reduction	233
A.3 Reference channel	235
A.4 Signal yield extraction	237
A.5 Branching fraction extraction	239
B Additional studies on kinematic corrections	242
B.1 Studies on $\hat{p}_{T\text{min}}$ cut	242
B.2 Cross-checks on QLC	246
B.3 Cross-checks on DDW	258
C Summary of personal contributions	267
Glossary	269
Bibliography	273

INTRODUCTION

The Standard Model (SM) of particle physics is currently the most successful theory to explain the fundamental structure and relations of the particles that compose our universe. Despite its successful confirmation with the discovery of the Higgs boson [1, 2], there are still open issues that this model can not address. The most compelling are the lack of dark matter candidates, the inadequacy in describing the matter anti-matter imbalance in the universe, the fine-tuning problem in the Higgs boson mass and neutrino oscillations.

There are several New Physics (NP) models that try explain these and other SM shortcomings, and the search for NP is currently one of the drivers of experimental particle physics. There are two main approaches in the search of Beyond the Standard Model (BSM) phenomena. Direct searches study new possible particles produced in high energy collisions, while indirect searches focus on studying processes that can be affected by the presence of virtual NP particles in loops.

The two rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ are considered golden flavour physics channels in the indirect search for NP, as their branching fractions $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ are very interesting from a theoretical point of view; while these processes are extremely suppressed and well determined in the SM framework, they can be modified by several BSM scenarios, yielding either an enhancement or a further suppression of these decays. The experimental measurement of the two branching fractions can therefore help posing constraints on NP models.

Some of the cutting edge experiments that are currently testing the SM and searching for hints of new phenomena are located around the LHC, at the CERN laboratories (near Geneva, Switzerland). The LHC is, as of today, the most powerful particle collider ever built and the products of the particle collisions are studied by four main experiments: ATLAS, CMS, LHCb and ALICE. The work described in this thesis is focused on the search for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays in the dataset collected by the ATLAS experiment during the 2015 and 2016 data taking period. The results of this physics analysis are also combined with the previous ATLAS analysis, based on the dataset collected during 2011 and 2012, for a further enhancement of the precision of the results. Additionally, a trigger selection aimed at improving the collection of $B \rightarrow hh'$ events is presented, as this is the most irreducible of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis' backgrounds, and the additional events would refine the analysis performance. As part of the outlook and as a contribution to the physics case for the High-Luminosity Large Hadron Collider (HL-LHC) proposal [3], pro-

jections on the sensitivity of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis with the full Run 2 and HL-LHC statistics are also presented.

The work presented in this thesis resulted, to date, in multiple presentations at conferences and workshops, the publication of a JHEP article [4], an ATLAS PUB note [5], several conference proceedings, as well as the CERN HL-LHC *yellow report* [3].

The theoretical background on the SM and the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is described in chapter 1, followed by a description of the experimental tools, in particular an overview on the LHC and a technical description of the ATLAS detector, in chapter 2. The description of the ATLAS analysis and simulation frameworks follows in chapter 3.

The author's contribution with the development of a trigger selection for the $B \rightarrow hh'$ process, which can help the background reduction for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, is described in chapter 4, while the central part of this thesis, the 2015-16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis is introduced in chapter 5. The analysis is then discussed in detail in the following chapters 7 - 12, culminating with the combination of the results with the ATLAS Run 1 analysis.

Finally, the prospects for the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis based on the full Run 2 statistics and on the statistics to be collected at HL-LHC are discussed in chapter 13.

This thesis is complemented with three appendices. An overview of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the Run 1 dataset is provided in appendix A. I was granted exceptionally authorship to this publication thanks to my contribution to the studies performed for the analysis' systematic uncertainties.

The various cross-checks performed during the evaluation of a set of kinematic corrections for the Monte Carlo (MC) samples employed in the 2015-16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis are shown in appendix B.

The work described in this thesis has been developed within the heavy flavour group of the ATLAS Collaboration, in which the author has been collaborating with teams of scientists from different countries and research institutes. The author's individual work has therefore been often part of a bigger collective effort. For this reason, the results obtained in this thesis have been published on behalf of the ATLAS Collaboration, as envisaged by its policies. A detailed list of the author's contribution to the results in this thesis is provided in appendix C.

THEORETICAL BACKGROUND

1

This chapter introduces theoretical background needed to understand the work developed in this thesis. After a summary of the Standard Model (SM) of particle physics and its limitations (respectively sections 1.1 and 1.2), the $B_{(s)}^0$ dimuon decays are introduced and their model-independent branching fraction is evaluated (section 1.3). This quantity is then calculated in the SM framework (section 1.3.3) and effects of BSM theories are considered (section 1.4). Finally, the status of the experimental searches for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes is presented (section 1.5).

1.1 The Standard Model of particle physics

1.1.1 Overview

The Standard Model (SM) of particle physics [6–8] is a renormalisable Quantum Field Theory (QFT) [9] developed to describe the known fundamental components of the universe. It includes all the particles experimentally discovered and three of the four known fundamental interactions: electromagnetic, weak, and strong interactions. Attempts at including gravity in the framework of renormalisable QFT are yet to be successful.

Elementary particles are described as excitations of fundamental quantum fields which live in a four-dimensional Minkowski spacetime [10]. The classification of elementary particles is based on their behaviour under Poincaré [9] transformations; one of the most notable invariant under such transformations is the spin, used to define two main categories: fermions have half-integer spin values in units of $\hbar$, while bosons have integer spin.

Present experimental observations show 12 different fermions, and for each fermion a corresponding anti-fermion, with equal mass but opposite quantum numbers. A further categorisation of fermions is based on which interaction they partake. Six of the known fermions are called quarks, which differ from the other six fermions, the leptons, because they can interact strongly. Quarks and leptons are further organised in three generations, which are characterised by similar quantum numbers but different mass. The first generation is the lightest, while the third generation contains the heavier particles. A summary of the SM fermions and their

Table 1.1: Summary of the fermions of the SM, together with their charges and masses [11]. The quark masses listed are calculated in the $\overline{\text{MS}}$ renormalisation scheme [12], except for the t quark, whose mass comes from direct measurements [11].

SM fermions						
	Quarks			Leptons		
	q	Charge	Mass (MeV)	l	Charge	Mass (MeV)
First Generation	u	+2/3	2.2	e	-1	0.511
	d	-1/3	4.7	ν_e	0	$< 2 \cdot 10^{-6}$
Second Generation	c	+2/3	1275	μ	-1	106
	s	-1/3	95	ν_μ	0	$< 2 \cdot 10^{-6}$
Third Generation	t	+2/3	$173.0 \cdot 10^3$	τ	-1	1777
	b	-1/3	4180	ν_τ	0	$< 2 \cdot 10^{-6}$

properties is provided in table 1.1.

Each lepton generation is composed of a negatively charged particle and a neutrino, which has no charge. These particles only interact via the weak and (if charged) electromagnetic interactions.

Quarks interact via the strong, weak and electromagnetic interactions. Each quark generation is composed of a +2/3 and a -1/3 electrically charged particle. Quarks with a 2/3 positive charge are often referred to as *up-type* quarks, while quarks with a -1/3 charge are referred to as *down-type* quarks. Due to the effect of the strong interaction, quarks can not be observed as isolated particles, but form bound states known as hadrons, which are further divided into mesons, formed of two quarks, and baryons, made of three quarks.

Fermions follow the Dirac Lagrangian

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.1)$$

where ψ is the fermion field, γ^μ are the Dirac matrices [9] and m is the mass of the fermion. The Dirac Lagrangian is invariant under global gauge transformations like

$$\psi \rightarrow \psi' = e^{i\theta} \psi, \quad (1.2)$$

where the equation only shows a $U(1)$ symmetry group global transformation. Imposing also the invariance under local transformations of a given symmetry group requires a modification in the Lagrangian, which results in the introduction of gauge fields (boson fields).

The SM is based on the invariance under the gauge group [9]

$$SU(3)_C \otimes SU(2)_L \otimes U(1)_Y. \quad (1.3)$$

$SU(3)_C$ describes the symmetry group of the strong interaction, mediated by 8 gluons; $SU(2)_L \otimes U(1)_Y$ describes the symmetry group of the electroweak interaction, mediated by the photon, $W^\pm$ and Z bosons.

The mass of the fermions and gauge bosons should appear in the Lagrangian as a term bilinear with respect to the fields and containing the mass value, as shown in the Dirac Lagrangian in formula 1.1. The introduction of such term in the SM Lagrangian for either the fermions or the gauge fields would break the local $SU(2)_L$ invariance [9]. This issue is solved with the spontaneous symmetry breaking mechanism, which requires the introduction of an additional doublet of complex scalar fields, from which an additional boson arises, corresponding to the recently discovered Higgs boson (section 1.1.4). Table 1.2 summarises the property of the bosons in the

Table 1.2: Summary of the bosons of the SM, together with their charges and masses [11].

Name	Mass (GeV)	Electric charge
Gluon($\times 8$) (g)	0	0
Photon (γ)	0	0
$W^\pm$	80	± 1
Z	91	0
Higgs (H^0)	125	0

SM.

The following sub-sections provide a more complete description of the different theories that constitute the SM.

1.1.2 Quantum Electrodynamics

Quantum Electrodynamics (QED), developed by Feynman, Schwinger and Tomonaga [13–15], is the quantum field theory for electromagnetism. The theory is based on the requirement that the Lagrangian 1.1 is invariant under local gauge transformations of the $U(1)$ symmetry group $\psi \rightarrow \psi' = e^{i\theta(x)}\psi$.

In order to attain local invariance, the Lagrangian must be modified in a manner that is equivalent to substituting the partial derivative ∂_μ with

$$\partial_\mu \rightarrow D_\mu = \partial_\mu - iQA_\mu, \quad (1.4)$$

where e is an arbitrary parameter of the theory that will be discussed later. This reformulation of the derivative operator to preserve local gauge invariance is known as “covariant derivative” [9]. The newly introduced A_μ field transforms under gauge transformations according to:

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{Q}\partial_\mu\theta(x). \quad (1.5)$$

Additional modifications must be added to the Lagrangian, as a kinetic term for the A_μ field is needed to describe its spacetime evolution. Such a kinetic term is in the form $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$, with $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$.

The Lagrangian showed in equation 1.1 becomes therefore

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi + Q\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.6)$$

Thanks to the introduction of the covariant derivative, a new term appears in the Lagrangian connecting the fermion field and A_μ , interpreted as the interaction term. No mass term appears in the Lagrangian for the field A_μ ; this is because this term would not be gauge invariant. The A_μ field can be interpreted as the photon field. Applying Noether's theorem [9], the conserved quantity associated to the $U(1)$ symmetry is Q and is interpreted as the electric charge of the fermions. The coupling constant for QED is a running parameter [9] dependent on the momentum exchange of the interaction q^2 . Its value at low q^2 is obtained experimentally [11] and it is expressed in terms of the so-called fine structure constant:

$$\alpha = \frac{e^2}{4\pi\hbar c} \sim \frac{1}{137}, \quad (1.7)$$

where the $\hbar$ and c constants are respectively the reduced Planck's constant and the speed of light in vacuum.

1.1.3 Electroweak unification

The theory of the weak interaction was proposed and unified with QED by Glashow and Weinberg [6, 7]. The so-called electroweak theory is based on the symmetry group $SU(2)_L \otimes U(1)_Y$. The structure $SU(2)_L$ group implies the existence three gauge bosons. The L stands for left-handed, meaning that only fermionic fields with a left handed chirality are involved, where left- and right-handed chiralities are defined as:

$$\psi_{L/R} = \frac{1 \mp \gamma^5}{2} \psi, \quad \text{with} \quad \gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3. \quad (1.8)$$

The associated conserved quantity is the weak isospin. Left- and right-handed fields transform in a different way under $SU(2)_L$. Right-handed particles (and left-handed anti-particles) transform as singlets, while left-handed particles (and right-handed anti-particles) transform as doublets:

$$f_L^i = \begin{pmatrix} \nu_L^i \\ l_L^i \end{pmatrix}, \quad f_R^i = l_R^i, u_R^i, d_R^i, \quad (1.9)$$

where the i index runs over the three generations.

$U(1)_Y$ is the same symmetry group employed in QED, but in this case the associated conserved quantity is called hypercharge, which relates to the weak isospin and the electric charge according to the following formula:

$$Y = 2(Q - T_3), \quad (1.10)$$

where T_3 refers to the weak isospin.

The covariant derivative in this case has the form:

$$D_\mu = \partial_\mu - igT^i W_\mu^i - ig' \frac{Y}{2} B_\mu, \quad (1.11)$$

with W_μ^i and B_μ being the gauge fields and g and g' the coupling constants associated with the $SU(2)_L$ and $U(1)_Y$ symmetries respectively. As for QED, the Lagrangian is further modified by adding the kinetic terms for the new fields, $-\frac{1}{4}B_{\mu\nu}B^{\mu\nu}$ and $-\frac{1}{4}W_{\mu\nu}^i W^{i\mu\nu}$, defined as:

$$B^{\mu\nu} = \partial^\mu B^\nu - \partial^\nu B^\mu, \quad W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i + g\epsilon^{ijk} W_\mu^j W_\nu^k, \quad (1.12)$$

where ϵ^{ijk} is the totally antisymmetric Levi-Civita tensor [9].

Finally, all these ingredients together give the electroweak Lagrangian:

$$\mathcal{L} = \sum_{f=l,q} \bar{f} i \gamma^\mu D_\mu f - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} - \frac{1}{4} W_{\mu\nu}^i W^{i\mu\nu}, \quad (1.13)$$

where the sum runs over the three generations of leptons and quarks.

Two important remarks must be made about equation 1.13. First, the gauge bosons which appear in equation 1.13 are not the actual bosons observed in experiments. W^1 and W^2 are complex fields and in order to obtain the real charged states observed in experiments they must be combined according to

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2). \quad (1.14)$$

The other two fields W^3 and B are real neutral states, but they do not coincide with what is observed by experiments. The photon can not be associated with the B field as the interaction with fermions is different. Similarly, the W^3 field can not be associated with the carrier of the neutral weak interaction. The Z^0 boson and the photon are thus obtained with a linear combination of W^3 and B :

$$\begin{pmatrix} Z_\mu^0 \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix}. \quad (1.15)$$

The θ_W mixing parameter is called Weinberg angle. Its value has been precisely measured in experiments: $\sin^2\theta_W = 0.23155 \pm 0.00005$ [11]. This angle can also be expressed in terms of the coupling constants associated with the $SU(2)_L$ and $U(1)_Y$ symmetries as

$$\cos\theta_W = \frac{g}{\sqrt{g^2 + g'^2}}, \quad \sin\theta_W = \frac{g'}{\sqrt{g^2 + g'^2}}. \quad (1.16)$$

The second remark about the electroweak Lagrangian showed in equation 1.13 regards the mass terms. All the fields involved, both fermionic and bosonic, appear to be massless, because the required mass terms are not invariant under $SU(2)_L \otimes U(1)_Y$. On the other hand, experimental observations strongly show the existence of massive fermions and bosons. An additional term must be included in the electroweak Lagrangian to account for the experimentally observed mass: the mechanism is discussed in the following sub-section.

1.1.4 The Higgs Mechanism

The solution to the mass problem arisen in the previous sub-section was given by R. Brout and F. Englert [16], and P. Higgs [17]. Mass terms that are $SU(2)_L$ invariant can be added to the electroweak Lagrangian via a spontaneous symmetry breaking mechanism, which can be achieved with the introduction of a new scalar complex field

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^1 + i\phi^2 \\ \phi^3 + i\phi^4 \end{pmatrix}, \quad (1.17)$$

which transforms as a doublet in the $SU(2)_L \otimes U(1)_Y$ group.

The Lagrangian of this new field is written as

$$\mathcal{L}_\phi = D_\mu \phi^\dagger D^\mu \phi - V(\phi), \quad (1.18)$$

with $V(\phi)$ being the potential of the field and D_μ the covariant derivative introduced in equation 1.11.

The potential of the field is written as

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (1.19)$$

If μ^2 is negative, the first term of the potential becomes a mass term for the ϕ field. On the other hand, if μ^2 is positive, the potential assumes a sombrero-like shape, illustrated in figure 1.1 and the ϕ field assumes a non-zero Vacuum Expectation Value (VEV) of

$$\phi^\dagger \phi = \frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2}. \quad (1.20)$$

The minimum of the potential is now a circumference with radius $\mu^2/2\lambda$. Due to this sym-

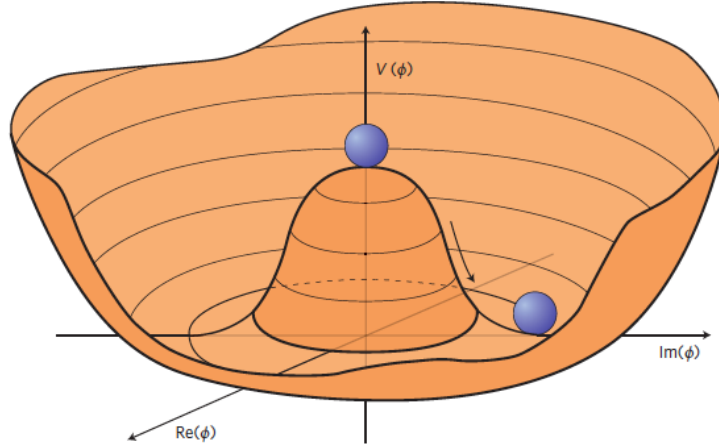


Figure 1.1: Illustration of the Higgs potential in the case that $\mu^2 > 0$, in which case the minimum is at $|\phi|^2 = \mu^2/(2\lambda)$. Choosing any of the points at the bottom of the potential breaks spontaneously the rotational $U(1)$ symmetry. Picture from [18].

metry, there is an infinite amount of points satisfying $\phi^\dagger \phi = v^2/2$; since the Lagrangian only depends on $\phi^\dagger \phi$, any of those points can be chosen as the expectation value for the ϕ field. A massless photon requires:

$$\phi = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}. \quad (1.21)$$

The field can now be perturbatively expanded around this minimum, obtaining

$$\phi = \begin{pmatrix} 0 \\ v/\sqrt{2} + H(x) \end{pmatrix}. \quad (1.22)$$

The field $H(x)$ is referred to as the Higgs field; the excitation of such field is called Higgs boson. Once expression 1.22 is added to the Lagrangian shown in equation 1.18, the mass terms of the gauge bosons naturally appear from the covariant derivatives:

$$m_W = \frac{gv}{2} \quad m_Z = v \frac{\sqrt{g^2 + g'^2}}{2} \quad m_A = 0. \quad (1.23)$$

The mass term for the Higgs boson appears from the second term in the $V(\phi)$ potential 1.19, with a value $m_H = \sqrt{2\lambda}v$.

The masses of the fermions do not arise in such a natural way, but they can be introduced as:

$$\mathcal{L} = \sum_{f=q,l} \lambda_f (\bar{f}_L \phi f_R + \bar{f}_R \phi f_L). \quad (1.24)$$

The λ_f terms, known as Yukawa couplings, contain the coupling constants of the Higgs to the fermions. After substituting equation 1.22 in 1.24, the fermion masses appear in the form

$$m_f = \lambda_f \frac{v}{2}. \quad (1.25)$$

After about 50 years from the formulation of this theory, the Higgs boson discovery was announced in 2012 at the CERN laboratories by the ATLAS [1] and CMS [2] collaborations.

1.1.5 Quantum chromodynamics

The last piece of the SM is a quantum field theory describing the strong interaction. This theory is called Quantum Chromodynamics (QCD) and it is built from a $SU(3)_C$ symmetry, where the C stands for colour [19]. This is defined as an additional quantum number that can assume three different discrete values (conventionally labelled as green, red and blue).

As for the QED and the electroweak theories, QCD is developed requiring the invariance of the Lagrangian under local $SU(3)_C$ gauge transformations. This is achieved by introducing the covariant derivative:

$$D_\mu = \partial_\mu + g_s \frac{\lambda_a}{2} A_\mu^a, \quad (1.26)$$

where g_s is the coupling constant of the strong interaction and λ_a are the Gell-Mann matrices [9], which constitute the eight generators of the $SU(3)$ group. The eight additional A_μ^a fields represent the gluons: by design the mediators of the strong interaction.

Following the approach introduced in section 1.1.2, the kinematic term for the gluons is also added to the Lagrangian in the form $-\frac{1}{4}F_{\mu\nu}^a F^{\mu\nu a}$, with

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g_s f_{abc} A_\mu^b A_\nu^c, \quad (1.27)$$

where f_{abc} are the structure constants of the $SU(3)$ group [9]. The third term in equation 1.27 is the gluon self-interaction, showing that gluons can interact with each other and therefore carry colour charge. In particular gluons carry effectively two colour charges: one colour charge and one “anti-colour” charge.

The resulting Lagrangian for the QCD theory is

$$\mathcal{L}_{\text{QCD}} = \bar{q}_i (i D^\mu \gamma_\mu - m)_{ij} q_j - \frac{1}{4} F_{\mu\nu}^a F^{\mu\nu a}, \quad (1.28)$$

where the i and j indices account for the colour of the quarks.

In analogy with QED, the strong coupling constant is expressed in terms of α_s , with

$$g^2 = 4\pi\alpha_s. \quad (1.29)$$

The α_s coupling is a running constant [9], which depends on the energy scale considered, as for the couplings of the electroweak interaction. On the other hand, the strong coupling constant decreases with increasing momentum exchange, while the other two tend to increase. This effect implies that at low energy scale α_s diverges, making the perturbative theory not applicable. As a consequence of the α_s behaviour, quarks and gluons tend to be free at very small distances (high energies), which is a property known as asymptotic freedom. At the same time, thanks to the very high value of the coupling constant at low energies, colour-charged particles can not be observed isolated. This property is called colour confinement.

1.1.6 Quark mixing and the CKM matrix

The concept of quark mixing was first introduced by N. Cabibbo [20] in the framework developed by M. Gell-Mann [21], which was based on the existence of three quarks: up, down and strange. The main idea is that the mass eigenstates of quarks do not necessarily overlap with interaction eigenstates, therefore the weak charged current, (mediated by the $W^\pm$ bosons) could be expressed as a function of a mixing parameter θ_c , called the Cabibbo angle. However, such theory predicted the existence of strangeness changing neutral currents, which would make the $K \rightarrow \mu^+ \mu^-$ process as probable as $K^+ \rightarrow \mu^+ \nu$, in open contrast with experimental results. The solution to this issue was provided by S. L. Glashow, J. Iliopoulos and L. Maiani [22], who developed the so called Glashow-Iliopoulos-Maiani (GIM) mechanism, predicting the existence of a fourth quark, named charm. Such a quark was then discovered, with the observation of the J/ψ : a $c\bar{c}$ bound state [23, 24]. The GIM mechanism describes quark mixing as follows:

$$\begin{bmatrix} d' \\ s' \end{bmatrix} = \begin{bmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{bmatrix} \begin{bmatrix} d \\ s \end{bmatrix}, \quad (1.30)$$

where the weak interaction eigenstates are on the left and on the right the Cabibbo matrix and the mass eigenstates; the value of the Cabibbo angle is $\sin\theta_c \sim 0.225$.

The remaining two quarks were predicted by M. Kobayashi and T. Maskawa, who noticed that in order to explain the CP violation found in neutral K meson decays [25], at least three quark doublets (six quark fields) are needed [26].

The quark mixing is described with the 3×3 unitary Cabibbo–Kobayashi–Maskawa (CKM) matrix:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix}, \quad (1.31)$$

where CP violation is parametrised with an irreducible complex phase and the probability of a transition from a quark i to a quark j is proportional to $|V_{ij}|^2$.

Experimental observations show that the elements on the diagonal of the CKM matrix (corresponding to transitions within the same quark generation) are $\mathcal{O}(1)$, with decreasing values while getting further away from the diagonal. The hierarchy of the quarks couplings is manifest

in the Wolfenstein parameterisation [27] of the CKM matrix:

$$V_{CKM} = \begin{bmatrix} 1 - \frac{\lambda^2}{2} & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1 - \frac{\lambda^2}{2} & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 & 1 \end{bmatrix} + \mathcal{O}(\lambda^4), \quad (1.32)$$

with $\lambda = V_{us} = \sin\theta_c$.

The latest estimates for the magnitudes of the CKM matrix elements are evaluated with a fit on several measurements [11, 28]:

$$V_{CKM} = \begin{bmatrix} 0.97446 \pm 0.00010 & 0.22452 \pm 0.00044 & 0.00365 \pm 0.00012 \\ 0.22438 \pm 0.00044 & 0.97359^{+0.00010}_{+0.00011} & 0.04214 \pm 0.00076 \\ 0.00896^{+0.00024}_{-0.00023} & 0.04133 \pm 0.00074 & 0.999105 \pm 0.000032 \end{bmatrix}. \quad (1.33)$$

Unlike charged currents, weak neutral currents do not allow any flavour mixing. For this reason Flavour-Changing-Neutral-Current (FCNC) are not present in the SM: only higher order Feynman diagrams can show such effects, which are therefore predicted to be highly suppressed.

1.2 Limitations of the Standard Model

The SM has been extensively tested and validated during the last 50 years [11, 29] but despite its major success, there are still some fundamental questions that have no answer.

The following subsections provide an overview of the most known issues of the SM

The hierarchy problem

One of the problems of the SM is strictly connected to the mass of the Higgs boson [30].

According to the theory, the mass of the Higgs boson can be evaluated as $m_H^2 = m_{H_0}^2 + \delta m_H^2$, where $m_{H_0}^2$ is the so-called bare mass, which is the free mass parameter in the Lagrangian, while δm_H^2 is due to radiative corrections, which depend on the couplings of the Higgs to all the other particles. The presence of fermion loops in the Higgs propagator, such as the one shown in figure 1.2, makes the δm_H^2 term become extremely large. For a fermion f with Yukawa

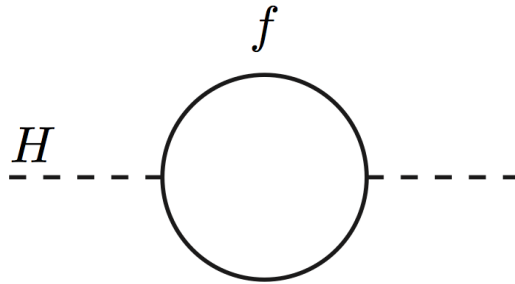


Figure 1.2: One-loop self-energy correction to the Higgs mass due to a fermion f .

coupling λ_f , the size of this correction is

$$\delta m_H^2|_f = -\frac{|\lambda_f|}{8\pi^2} \Lambda_{UV}^2 + \dots \quad (1.34)$$

where Λ_{UV} is the ultraviolet momentum cut-off, chosen to be at the Planck scale ($\sim 10^{18}$ GeV). The correction to the Higgs mass is therefore about 30 order of magnitude higher than the actual Higgs mass of ~ 125 GeV.

The huge difference in scale between the Higgs boson mass and the Planck scale is known as the Higgs hierarchy problem. In the SM framework, this issue is solved by fine-tuning the bare mass parameter so that it cancels the large radiative corrections. This procedure is not forbidden, but it is in open contrast with the argument of naturalness. According to this principle, the existence of these fine cancellations can only be justified with a specific feature of the theory.

Dark matter

Dark matter is a hypothetical form of matter which does not interact via the electromagnetic interaction, but has a substantial gravitational effect. Its presence is implied in several astrophysical observations, which show effects that can not be explained by accepted theories of gravity unless more matter, than can not be seen, is present [31]. Among the main arguments supporting the existence of dark matter one of the most compelling is based on the observation of the rotation curves of galaxies [32], where the velocity of the stars as a function of the distance from the centre of the galaxy is not compatible with the amount and distribution of visible matter present in the galaxy. Furthermore, the decomposition of the anisotropies of the Cosmic Microwave Background (CMB), measured by WMAP [33] and Planck [34], into an angular power spectrum, clearly shows a peak that relates directly to the density of dark matter.

Neutrino mass

Neutrino physics measurements performed by the Super-Kamiokande collaboration [35] and the SNO collaboration [36] gave unambiguous evidence of neutrino lepton flavour oscillations. This process is only possible if neutrinos have mass, in open disagreement with the SM, which implies massless neutrinos and lepton flavour universality.

CP violation

The only source of CP violation in the SM is the complex phase in the CKM matrix, discussed in section 1.1.6. The imbalance in matter and anti-matter in the universe can not be explained only with this CP violation source [37], therefore additional CP violation sources are expected to appear in new physics scenarios.

Particular interest is devoted to the search for CP violation in strong interactions, QCD in fact does not explicitly forbid it. However, no strong CP violation effect has ever been observed.

1.3 Dimuon B decays

The work developed in this thesis is centred on the search for the rare decays of B_s^0 and B_d^0 mesons into two muons. This section is therefore dedicated to the theoretical calculation of the branching fraction of these decays and motivates their experimental search.

The neutral B_s^0 and B_d^0 mesons are composed of a $\bar{b}$ quark and, respectively, a s and d quarks as valence quarks; their properties are summarised in table 1.3. Although extremely rare in the SM, the decays of these mesons to two muons ($B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$) are considered interesting from an experimental point of view, as they can provide insights on the potential presence of BSM physics and its scale. The probability of these decays to happen is of the order of $\sim 10^{-9}$ and $\sim 10^{-10}$ respectively. Such heavy suppression comes from three sources.

Table 1.3: Summary of the properties of the B_s^0 and B_d^0 mesons [11].

Name	Mass (GeV)	Electric charge	Valence quarks
B_s^0	5.367	0	$\bar{b}s$
B_d^0	5.280	0	$\bar{b}d$

The first source of suppression is due to the quark composition of the two mesons. The dimuon decay of B_s^0 and B_d^0 can proceed at the lowest level in the SM through higher order electroweak interactions, since tree-level processes would require $b \rightarrow s$ and $b \rightarrow d$ transitions mediated by Flavour-Changing-Neutral-Currents (FCNCs) (not foreseen in the SM, as discussed in section 1.1.6). Since they are forbidden at tree-level, the decays proceed via more complex Feynman diagrams, such as W -boxes and Z -penguins, as shown in figure 1.3. Higgs-mediated pen-

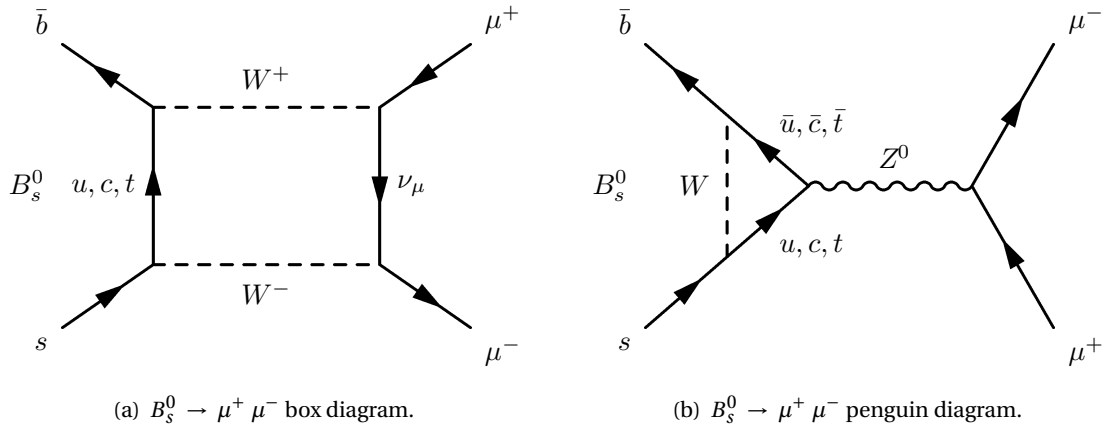


Figure 1.3: Dominant Feynman diagrams for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The $B_d^0 \rightarrow \mu^+ \mu^-$ decay has the same diagrams, but instead of a s quark in the initial state it presents a d quark.

guins are also possible, but their contribution is negligible [38]. Due to the lack of tree-level Feynman diagrams, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes are suppressed with respect to other $B_{(s)}^0$ decays which can occur at tree-level.

The second source of suppression of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is due to the coupling strength

of the weak interaction to different quark flavours. The coupling strengths are described in the CKM matrix, introduced in section 1.1.6. Figure 1.3 shows that the internal quarks lines in the processes have contributions from u , c and t quarks; however, the u and c contributions are basically negligible compared to t contributions. Since a $t-s$ transition couples quarks from the second and third generations, it is less suppressed compared to a $t-d$ transition, which couples quarks from the first and third generation. A relative CKM suppression is therefore the reason why $B_d^0 \rightarrow \mu^+ \mu^-$ decays are more suppressed than $B_s^0 \rightarrow \mu^+ \mu^-$ decays, even though the kinematics of the two decays are basically the same.

The third source of suppression comes from helicity effects. Both B_s^0 and B_d^0 are pseudoscalar mesons ($j^P = 0^-$), therefore, the spins of the two muons must be oppositely aligned in order to conserve the angular momentum in the decay. This leads to the muons having opposite helicities. On the other hand, the weak interaction only couples to particles with a left-handed chirality. In the limit of massless particles, negative helicity states correspond to left-handed chirality states and positive helicity states correspond to right-handed chirality states, making the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays impossible. Since muons are not massless, the decay is not completely forbidden, but heavily suppressed, as the muon mass is much smaller than the B mass, making one of the helicity states highly disfavoured.

The following sub-sections provide the model-independent calculation of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions (sub-sections 1.3.2 and 1.3.3) and the corresponding value in the SM (sub-section 1.3.3).

1.3.1 Branching fraction and B mixing

The branching fraction ($\mathcal{B}$) of a particle decay is defined as the probability of a such decay to occur. Alternatively, it can also be considered as the relative frequency of the decay. The definition of the branching fractions for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ decays presents an additional difficulty with respect to other decays, due to the properties of the $B_{(s)}^0$ mesons.

Neutral B mesons are produced as flavour eigenstates $B_{(s)}^0 - \bar{B}_{(s)}^0$, which are described at $t = 0$ by the states $|B_{(s)}^0\rangle$ and $|\bar{B}_{(s)}^0\rangle$. Due to quark mixing (section 1.1.6), these flavour states do not correspond to the mass eigenstates in which the $B_{(s)}^0$ meson evolve and propagate. The time evolution of the flavour eigenstates is described by the time-dependent Schrödinger equation [9], where only B_s^0 mesons are considered in order to simplify the notation, as [39]:

$$i \frac{d}{dt} \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix} = \left(M - \frac{i\Gamma}{2} \right) \begin{pmatrix} |B_s^0\rangle \\ |\bar{B}_s^0\rangle \end{pmatrix}, \quad (1.35)$$

where M and Γ are hermitian matrices which describe the mass and decay time. The presence of non zero off-diagonal elements in such matrices ensures that for any $t > 0$ the particles are a superposition of the $|B_s^0\rangle$ and $|\bar{B}_s^0\rangle$ states. The eigenstates of the time-dependent Schrödinger equation have different masses and lifetimes with respect to the flavour eigenstates and are known as the heavy $|B_H\rangle$ and light $|B_L\rangle$ mass eigenstates.

The LHC experiments can only measure the time-integrated $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions

without considering the flavour eigenstate of the B mesons, which corresponds to [40]

$$\begin{aligned}\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} &= \frac{1}{2} \int_0^\infty \left(\Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) + \Gamma(\bar{B}_q^0(t) \rightarrow \mu^+ \mu^-) \right) dt \\ &= \frac{1}{2} \int_0^\infty \langle \Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) \rangle dt,\end{aligned}\quad (1.36)$$

where $\langle \Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) \rangle$ is the time-dependent and *un-tagged* (without considering the flavour) decay rate.

On the other hand, the classical theoretical computation of the two branching fractions yields the CP averaged decay rates without considering the time evolution of B mesons [40]:

$$\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{theo}} = \frac{\tau_{B_q}}{2} \langle \Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) \rangle|_{t=0}, \quad (1.37)$$

where τ_{B_q} is the mean lifetime of the B_q meson.

The time-dependent un-tagged decay rate, introduced in equation 1.36, can be written in terms of the decay rate of the two mass eigenstates, which have well-defined lifetimes [39]:

$$\langle \Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) \rangle = R_H^{\mu^+ \mu^-} e^{-\Gamma_H^q t} + R_L^{\mu^+ \mu^-} e^{-\Gamma_L^q t}, \quad (1.38)$$

where the H and L sub-scripts refer to the heavy and light mass eigenstates and $R_H^{\mu^+ \mu^-}$ and $R_L^{\mu^+ \mu^-}$ are the respective decay rates into two muons.

The difference in the decay widths of the light and heavy mass eigenstates is usually reported as

$$y_q = \frac{\Gamma_L^q - \Gamma_H^q}{2\Gamma_q} = \frac{\Delta\Gamma_q}{2\Gamma_q} \quad \text{with} \quad \Gamma_q = \tau_{B_q}^{-1} = \frac{\Gamma_L^q + \Gamma_H^q}{2}. \quad (1.39)$$

The measured magnitude of y_q for B_d^0 mesons (y_d) is smaller than 1%, while for the $B_s^0 - \bar{B}_s^0$ system the measured value is [11, 28]

$$y_s = 0.0645 \pm 0.0045. \quad (1.40)$$

Using the definition of y_q , equation 1.38 can be re-written as [40]:

$$\langle \Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-) \rangle = (R_H^{\mu^+ \mu^-} + R_L^{\mu^+ \mu^-}) e^{-\Gamma_q t} \left[\cosh\left(\frac{y_q t}{\tau_{B_q}}\right) + \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} \sinh\left(\frac{y_q t}{\tau_{B_q}}\right) \right], \quad (1.41)$$

with

$$\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = \frac{R_H^{\mu^+ \mu^-} - R_L^{\mu^+ \mu^-}}{R_H^{\mu^+ \mu^-} + R_L^{\mu^+ \mu^-}}. \quad (1.42)$$

Using this definition, the experimentally measurable branching ratio shown in equation 1.36 can be converted into the CP averaged time-independent branching fraction (equation 1.37) through

$$\mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{theo}} = \left[\frac{1 - y_q^2}{1 + y_q \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}} \right] \mathcal{B}(B_q^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}. \quad (1.43)$$

For $y_q \sim 0$ the two definitions coincide. This is basically realised in the $B_d^0 \rightarrow \mu^+ \mu^-$ process, but the difference is sizeable for $B_s^0 \rightarrow \mu^+ \mu^-$.

According to the SM $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ takes the maximum value of +1, meaning that only the heavy mass

eigenstate contributes to the $B_s^0 \rightarrow \mu^+ \mu^-$ process. This can be understood because the dimuon final state is a CP odd state and the B_s^0 heavy mass eigenstate is CP odd. However, possible new physics effects can modify $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. These effects can be experimentally quantified by measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ effective lifetime and can result in a modification of the branching fraction up to $\mathcal{O}(10\%)$ [40]

1.3.2 Branching fractions computation

This sub-section provides an overview of the computation of the model-independent $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions. A more detailed discussion can be found in Refs. [41–48].

As shown in the previous section, the branching fraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes can be expressed in terms of the instantaneous un-tagged decay rate. This quantity, in turn, is given by Fermi's golden rule [49]:

$$\Gamma(B_q^0(t) \rightarrow \mu^+ \mu^-)|_{t=0} = \frac{1}{16\pi m_{B_{(s)}^0}} \sqrt{1 - 4 \left(\frac{m_\mu}{m_{B_{(s)}^0}} \right)^2} |\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)|^2 \quad (1.44)$$

where m_μ is the mass of the muon, $m_{B_{(s)}^0}$ is the mass of the B meson of interest and $\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is the matrix element which describes the transition processes responsible for the decays. Unlike electroweak interactions, where the small coupling constants allow perturbative calculations at arbitrary energy scales, the strong interaction makes the computation of the matrix element complicated: as introduced in section 1.1.5, the strong coupling constant tends to diverge at low energy scales, ruling out the use of perturbation theory.

The approach followed in this case is to separate the high energy contributions from the low energy contributions. This procedure is referred to as effective approach. For high transferred momentum the perturbative approach can be used, as the strong coupling constant is small, while at low transferred momentum non-perturbative techniques, like Lattice QCD, must be employed.

The separation is based on a factorisation scale μ_S . For energy scales lower than μ_S the effects of the mediators of the interactions are omitted; the interaction is therefore expressed in terms of a set of four-prong vertices, which can be interpreted as initial state particles scattering off static potentials and emerging as final state particles. The presence of the mediators of the interactions is taken into account at the energy scales higher than μ_S , where perturbation theory can be employed. Operator Product Expansion (OPE) [50] is used to factorise the contributions of different interactions between the initial and final state:

$$\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \sum_i^{OPE} \left(\underbrace{A_i}_{\text{Perturbative}} \times \underbrace{\text{Diagram}}_{\text{Non-perturbative "four legs" vertex}} \right). \quad (1.45)$$

Effectively, the OPE is a sum over all the possible local operators between the initial and final state, representing the non-perturbative part, and a set of weights A_i that allow the reproduction of the QFT effects.

In a more formal notation, equation 1.45 can be written as:

$$\begin{aligned} \mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) &= \langle \mu^+ \mu^- | \mathcal{H}_{\text{eff}} | B_{(s)}^0 \rangle + \mathcal{O}\left(\frac{p^2}{M_W^2}\right) \\ &\simeq \frac{G_F}{\sqrt{2}} \sum_i V_{\text{CKM}}^i C_i(\mu_S) \times \langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle, \end{aligned} \quad (1.46)$$

where G_F is the Fermi constant for the weak processes, C_i are the coefficients that take into account the perturbative effects, $\mathcal{O}_i$ are the local operators which consider the non-perturbative effects and V_{CKM}^i are CKM matrix elements. The $\mathcal{O}(p^2/M_W^2)$ term represents higher order operators which are basically negligible given the huge difference between the $W^\pm$ mass and the typical $B_{(s)}^0$ momentum scale p .

The C_i coefficients are called *Wilson coefficients*. They are evaluated by matching the effective theory to the calculations performed using perturbation theory, more details are given in [51]. Regarding the matrix elements $\langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle$, the strong interaction only affects the initial state of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes, as the final state is fully leptonic; it is therefore possible to decouple, from the point of view of the strong interaction, the final state from the initial state. The four-prong matrix element can be therefore factorised as:

$$\langle \mu^+ \mu^- | \mathcal{O}_i | B_{(s)}^0 \rangle = \underbrace{\langle \mu^+ \mu^- | \mathcal{O}_i^{ll} | 0 \rangle}_{\text{Leptonic matrix element}} \times \underbrace{\langle 0 | \mathcal{O}_i^{qq} | B_{(s)}^0 \rangle}_{\text{Hadronic matrix element}}, \quad (1.47)$$

where all the strong interaction effects are embedded in the hadronic term.

The $\mathcal{O}_i^{qq}$ and $\mathcal{O}_i^{ll}$ operators are classified according to their transformations as scalar ($\mathbb{1}$, indicated with S), pseudo-scalar (γ_5 , indicated with P), vector (γ_μ , indicated with V), axial-vector ($\gamma_\mu \gamma_5$, indicated with A) and tensor ($\gamma_\mu \gamma_\nu - \gamma_\nu \gamma_\mu$, indicated with T).

Several possible combinations of operators, expressed as $\mathcal{O} = \mathcal{O}^{ll} \otimes \mathcal{O}^{qq}$, are therefore available. The Lorentz invariance requirement allows one to narrow down all the available combinations to the following four:

$$S^{ll} \otimes P^{qq}, \quad P^{ll} \otimes P^{qq}, \quad V^{ll} \otimes A^{qq}, \quad A^{ll} \otimes A^{qq}. \quad (1.48)$$

These combinations present only two operators for the hadronic component, which account for the strong interaction effects in the initial state. The resulting two hadronic matrix elements can be expressed in terms of a constant term $f_{B_{(s)}^0}$ called the *decay constant*, which encloses all the initial state non-perturbative effects due to the strong interaction; its can be evaluated using Lattice QCD. The two hadronic term are evaluated as follows [52]:

- axial-vector term (A^{qq})

$$\langle 0 | A^{qq} | B_{(s)}^0 \rangle = \pm i f_{B_{(s)}^0} p_\mu^B, \quad (1.49)$$

where p_μ^B is the four momentum of the $B_{(s)}^0$ meson;

- pseudo-scalar term (P^{qq})

$$\langle 0 | P^{qq} | B_{(s)}^0 \rangle = \mp \frac{i m_{B_{(s)}^0}^2}{m_b + m_q} f_{B_{(s)}^0}, \quad (1.50)$$

where m_b and m_q are the masses of the b quark and the of other quark in the B meson (d or s).

Among the four possible combinations of operators shown in formula 1.48, one contains a vector operator for the leptonic term ($V^{ll} \otimes A^{qq}$). It can be expressed as

$$\langle \mu^+ \mu^- | V^{ll} | 0 \rangle \times \langle 0 | A^{qq} | B_{(s)}^0 \rangle = \pm \langle \mu^+ \mu^- | V^{ll} | 0 \rangle i f_{B_{(s)}^0} p_\mu, \quad (1.51)$$

and vanishes, due to the Ward identity [46]. In the SM, the presence of a vector current implies that the process is mediated by a photon; the fact that this term vanishes is the reason why the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ penguin diagram in figure 1.3(b) does not include the photon counterpart.

Finally the three operator combinations which have not been excluded, divided into the left- and right- handed chirality states of the quarks, are:

$$\begin{aligned} S^{ll} \otimes P^{qq} &\Rightarrow \mathcal{O}_S^L = (\bar{\mu}\mu)(\bar{b}P_L q), & \mathcal{O}_S^R &= (\bar{\mu}\mu)(\bar{b}P_R q) \\ P^{ll} \otimes P^{qq} &\Rightarrow \mathcal{O}_P^L = (\bar{\mu}\gamma_5\mu)(\bar{b}P_L q), & \mathcal{O}_P^R &= (\bar{\mu}\gamma_5\mu)(\bar{b}P_R q) \\ A^{ll} \otimes A^{qq} &\Rightarrow \mathcal{O}_{10}^L = (\bar{\mu}\gamma_\mu\gamma_5\mu)(\bar{b}\gamma_\mu P_L q), & \mathcal{O}_{10}^R &= (\bar{\mu}\gamma_\mu\gamma_5\mu)(\bar{b}\gamma_\mu P_R q), \end{aligned} \quad (1.52)$$

Where $q = b, s$. These operators can now be introduced in equation 1.46, obtaining:

$$\mathcal{M}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{G_F}{\sqrt{2}} V_{tq}^\dagger V_{tb} \sum_i^{S,P,10} \left(C_i^L(\mu_S) \mathcal{O}_i^L + C_i^R(\mu_S) \mathcal{O}_i^R \right), \quad (1.53)$$

where $q = b, s$ and only the contributions of the t quark to loops are considered, as the contributions from other up-type quarks are negligible. For this reason, the CKM matrix elements relative to the $b - t$ and $t - q$ vertices are explicitly shown.

Finally, the model-independent experimental branching fraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays can be expressed taking the connection between the experimental and theoretical definitions of the branching fractions from equation 1.43, the theoretical definitions of the branching fractions from equation 1.37, Fermi's golden rule from equation 1.44 and the matrix element calculated using the OPE from equation 1.53. The resulting formula is [53]

$$\begin{aligned} \mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}} &= \frac{\tau_{B_{(s)}^0} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} |C_{10}^{SM} V_{tq}^\dagger V_{tb}|^2 f_{B_{(s)}^0}^2 m_{B_{(s)}^0}^2 m_\mu^2 \\ &\quad \frac{1 + y_q \mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}}{1 - y_q^2} \sqrt{1 - 4 \left(\frac{m_\mu}{m_{B_{(s)}^0}} \right)^2} (|P|^2 + |S|^2), \end{aligned} \quad (1.54)$$

where P and S are defined as

$$\begin{aligned} P &= |P| e^{i\varphi_P} = \frac{C_{10}^R - C_{10}^L}{C_{10}^{SM}} + \frac{m_{B_{(s)}^0}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{C_P^R - C_P^L}{C_{10}^{SM}} \right), \\ S &= |S| e^{i\varphi_S} = \sqrt{1 - 4 \left(\frac{m_\mu}{m_{B_{(s)}^0}} \right)^2} \frac{m_{B_{(s)}^0}^2}{2m_\mu} \left(\frac{m_b}{m_b + m_{(d,s)}} \right) \left(\frac{C_S^R - C_S^L}{C_{10}^{SM}} \right), \end{aligned} \quad (1.55)$$

with C_{10}^R, C_{10}^L and C_{10}^{SM} dimensionless and C_P^R, C_P^L, C_S^R and C_S^L with dimension GeV^{-1} .

1.3.3 The SM branching fractions

In the SM the only non-negligible contributions in the evaluation of the branching fractions of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays come from the axial-vector operator $\mathcal{O}_{10}^R$ from equation 1.52. The scalar ($\mathcal{O}_S$) and pseudo-scalar ($\mathcal{O}_P$) contributions are in fact absent, with the only exception of the Higgs penguin process, which is however negligible due to the small coupling of the Higgs boson with muons. For this reason the $C_P^R, C_P^L, C_S^R, C_P^L, C_{10}^L$ terms in equation 1.55 vanish, leaving $P_{\text{SM}} = 1$ and $S_{\text{SM}} = 0$. Equation 1.54 therefore becomes

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}^{\text{SM}} = \frac{\tau_{B_{(s)}^0} G_F^4 M_W^4 \sin^4 \theta_W}{8\pi^5} \frac{f_{B_{(s)}^0}^2 m_{B_{(s)}^0} m_\mu^2}{1 - y_q} |C_{10}^{\text{SM}} V_{tq}^\dagger V_{tb}|^2 \sqrt{1 - 4 \left(\frac{m_\mu}{m_{B_{(s)}^0}} \right)^2}. \quad (1.56)$$

Where $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-} = +1$, As introduced in section 1.3.1. The SM $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ reference values considered throughout this thesis are [54]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}^{\text{SM}} = (3.65 \pm 0.23) \times 10^{-9} \quad (1.57)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{exp}}^{\text{SM}} = (1.06 \pm 0.09) \times 10^{-10}. \quad (1.58)$$

The largest uncertainties affecting the values in equations 1.57 and 1.58 originate from the CKM coefficients and the decay constants, as shown in table 1.4. A more recent calculation

Table 1.4: Summary of the main uncertainties affecting the calculation of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ in the SM [54].

	CKM [%]	$f_{B_{(s)}^0}$ [%]	Other sources [%]	Total [%]
$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	4.3	4.0	2.3	6.3
$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$	6.9	4.5	2.0	4.8

of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, performed considering improved electromagnetic corrections, yields [55]

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.57 \pm 0.17) \times 10^{-9}, \quad (1.59)$$

showing a slightly lower value of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction together with a smaller uncertainty. In order to be consistent with the approach followed in the studies described, only the values reported in equations 1.57 and 1.58 will be employed throughout this thesis.

1.4 New physics and $B_{(s)}^0 \rightarrow \mu^+ \mu^-$

There are a large number of BSM theories that predict some influence on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes, figure 1.4 shows the effect of some of these. For this reason, the experimental search for these decays is considered a milestone in the indirect search for new physics. This section provides an overview of the most known models that can influence $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays. A detailed discussion on possible BSM models and their effects on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays is given in Refs. [48, 53, 61, 62].

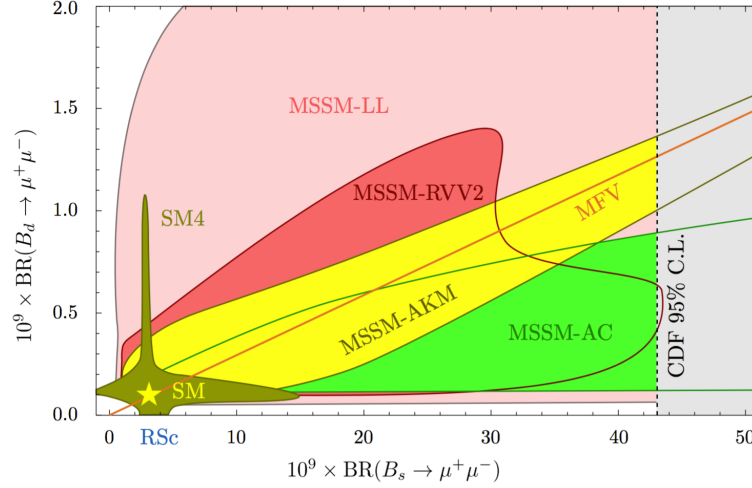


Figure 1.4: Possible values of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ due to different models in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ - $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane. The SM prediction is shown as a star. The BSM models shown are MFV [56], four MSSM [57] and the SM4 model [58]. The grey regions shows the upper limit posed by the CDF experiment on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ [59] before the results from the LHC experiments. Picture from [60].

One of the most popular models is called Minimal Flavour Violation (MFV) [56]. In this model the additional flavour structure and the operators that contribute to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are assumed to be SM-like; new physics effects can arise in the C_{10}^R coefficient. Similar models introduce an additional CP violating phase, which can provoke changes in $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$. For these models the expected effect on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is small. For example, the Littlest Higgs Model with T-Parity (LHT) [63] predicts an enhancement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ of at most 30%.

Other BSM models predict the existence of new particles which could affect $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+ \mu^-}$ because of their contribution in the Feynman diagrams involved in the decay. These new particles can appear in loops, similar to the ones shown in figure 1.3, or can allow FCNC at tree level. The most known theories which imply the introduction of new particles are two Higgs doublet models (2HDMs) [64], supersymmetric models [57] and leptoquark models [65]. The 2HDMs model implies a modification of the SM Higgs sector by introducing two complex scalar fields, both with non-zero VEVs. The effect of spontaneous symmetry breaking implies the production of two charged scalar, two neutral scalar and one pseudoscalar Higgs. All these new particles can enter the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ loops or allow FCNC at tree level. Depending on the property of the particles allowed to interfere in the SM Feynman diagrams the values of the Wilson coefficients in equation 1.54 can be modified, with possible large effects on the branching fractions.

Supersymmetric models (SUSY) introduce an additional symmetry, which gives each SM particle a supersymmetric partner. Since no evidence of SUSY has been found, the symmetry must be broken and the mass of the supersymmetric partners has to be higher than the SM particles and possibly not reachable by the current accelerators. Nevertheless, these new particles might affect $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, appearing in loops or at tree level. Some of the most studied SUSY models are called Minimal Supersymmetric Standard Model (MSSM) which modify the SM Higgs sector in a similar way as 2HDM. The overall effect of MSSM models can modify S , P ,

$S \pm P$ and $\mathcal{A}_{\Delta\Gamma}^{\mu^+\mu^-}$.

Leptoquark models are currently popular as they can explain some anomalies observed in heavy flavour measurements, in particular the ones regarding lepton flavour universality [66–68]. Leptoquarks are additional bosons which carry both leptonic and baryonic quantum number and therefore can imply FCNC at tree-level in $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes. Their exact properties depend on how they couple to SM particles.

BSM effects have not been observed yet in $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes, but, as shown in the next section, the current experimental precision leaves plenty of room for them. The Run 2 and future runs of the LHC will allow us to study these decays with unprecedented precision, given the size of the datasets collected, and to pose increasingly stringent limits on BSM physics.

1.5 Experimental state of the art

The history of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements, up to the beginning of 2018, is shown in figure 1.5. The latest measurements, excluding the ATLAS results discussed later in

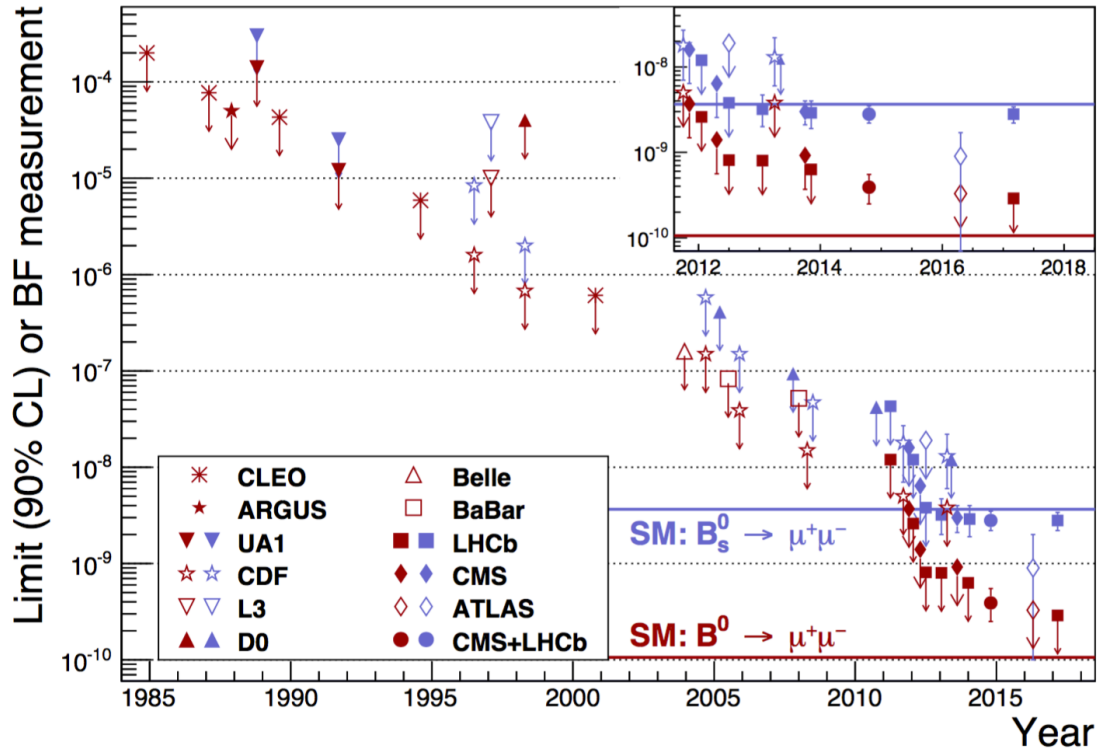


Figure 1.5: History of the limits and measurements of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions across the years [69].

this thesis, come from the ATLAS, CMS and LHCb experiments, which collect pp collisions data at the LHC.

The CMS and LHCb collaborations combined their analysis performed on the data collected during the LHC “Run 1” data taking period (2011–2012) [70], respectively 25 fb^{-1} [71] and 3.0 fb^{-1} [72], leading to the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay. The resulting branching

fraction has significance of 6.2σ and is compatible with the SM prediction [54]:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8_{-0.6}^{+0.7}) \times 10^{-9} \quad (1.60)$$

An evidence for the $B_s^0 \rightarrow \mu^+ \mu^-$ decay was also reported with a statistical significance of 3.2σ :

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (3.9_{-1.4}^{+1.6}) \times 10^{-10}. \quad (1.61)$$

Later also the ATLAS collaboration published the analysis performed on the Run 1 dataset (25 fb^{-1}) [73] (more details are provided in appendix A) leading to:

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (0.9_{-0.8}^{+1.1}) \times 10^{-9} \\ \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) &< 4.2 \times 10^{-10} \quad \text{at 95\% CL.} \end{aligned} \quad (1.62)$$

The LHCb collaboration published also a result combining Run 1 data with the data collected in the first part of the LHC Run2 data taking period, for a total combined integrated luminosity of 4.4 fb^{-1} [74]. The result of the analysis is compatible with the SM prediction and represents the first observation of the $B_s^0 \rightarrow \mu^+ \mu^-$ decay in a single experiment, having a significance of 7.8σ :

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6_{-0.2}^{+0.3}) \times 10^{-9} \quad (1.63)$$

as well as the most stringent upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ available prior to the result discussed in this thesis:

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10} \quad \text{at 95\% CL.} \quad (1.64)$$

The first uncertainty reported for the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction is statistical and the second systematic.

The two-dimensional likelihood contours in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ space obtained from the three analysis described above are shown in figure 1.6, where figure 1.6(a) shows the ATLAS contours in blue and the CMS+LHCb contours in grey and figure 1.6(b) shows

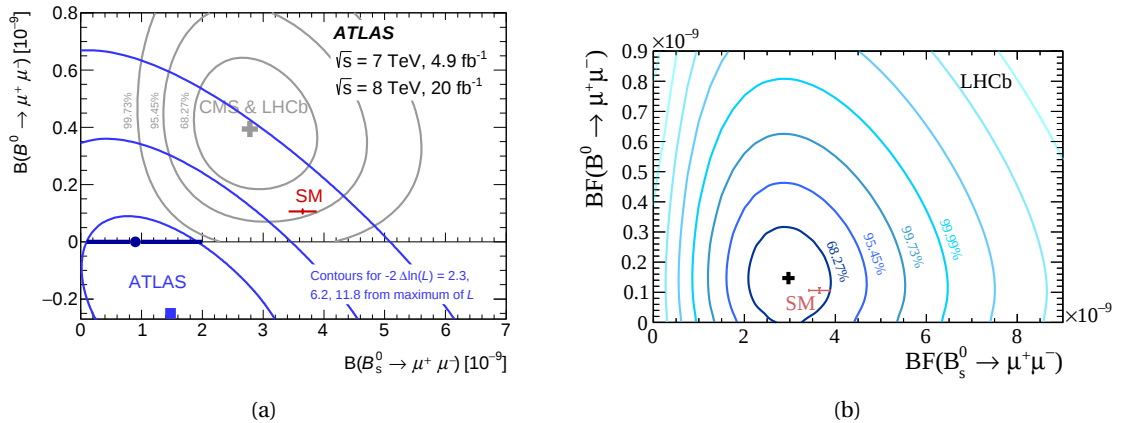


Figure 1.6: Two-dimensional likelihood contours in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ space obtained from the three analysis described in the text. Figure 1.6(a) [73] shows the ATLAS Run 1 contours [73] in blue and the Run 1 combined CMS+LHCb contours [70] in grey. Figure 1.6(b) shows the LHCb Run 2 contours [74]. Both pictures also include the SM theoretical prediction and its uncertainty [54].

the LHCb Run 2 contours.

Comparing the various results in figure 1.5 and in the contour plots shown above, it appears that the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses performed by the three LHC experiments are characterised by different sensitivities, as the various confidence intervals show different behaviours. The experiment with the highest sensitivity, therefore the most precise results, is LHCb, which was designed specifically to perform flavour physics searches at the LHC. The experiment, in fact, presents extremely good tracking and vertexing properties, the possibility of performing particle identification and a lower threshold on the muon transverse momentum acceptance with respect to the other LHC experiments. The two general-purpose experiments, ATLAS and CMS, present therefore a lower sensitivity with respect to LHCb, but, when comparing the two, it seems that the CMS sensitivity is higher than the ATLAS one. This feature will be investigated in this thesis (section 13.2 in chapter 13), showing that the main reason for the higher CMS sensitivity is the different dimuon mass resolution of the two experiments.

Even if all the results presented are consistent with the SM prediction, there are disagreements between the analysis results of the various experiments. The ATLAS $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ Run 1 result is lower compared to the CMS+LHCb combination and the latest LHCb result. Regarding $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, the evidence obtained in the CMS+LHCb combination yields a branching fraction higher compared to the upper limits from both ATLAS and LHCb.

In order to solve these disagreements, new measurements are needed; thanks to the increased size of the datasets collected by the three LHC experiments, improved uncertainties are expected. Additionally, since the analyses presented are highly statistically limited with uncertainties much larger than the theoretical uncertainty on the SM prediction, there is a lot of room for improvements in the precision of the results.

EXPERIMENTAL APPARATUS

2

The work developed in this thesis focuses on a physics analysis performed on pp collision data, delivered by the CERN Large Hadron Collider (LHC) at a centre of mass energy of $\sqrt{s} = 13$ TeV during 2015 and 2016, and collected by the ATLAS experiment. The author also had a leading role in the development of physics projections regarding the HL-LHC upgrade.

In this chapter the LHC apparatus and the ATLAS experiment are described, respectively in sections 2.1 and 2.2. An overview of the HL-LHC upgrade of the LHC accelerator is presented in section 2.3.

Rather than an exhaustive treatment, this chapter will focus on areas and components that are most relevant to the work discussed in this thesis.

2.1 The LHC

The Large Hadron Collider (LHC) [75], built between 1998 and 2008 at the European Organization for Nuclear Research (CERN) laboratories, is, as of today, the world's biggest and most powerful accelerator. It lies in an underground tunnel, at a depth ranging from 50 to 175 metres, beneath the Franco–Swiss border near Geneva. The LHC is composed of a ring about 27 kilometre long that contains two beam pipes, designed to contain two hadron beams which travel in opposite directions around the accelerator. The beams are guided and collimated by strong magnetic fields generated by superconducting magnets that surround the beam pipes. 1232 superconducting dipoles, used to bend the beams, and 392 quadrupole magnets, used to focus the beams, are present. The superconducting magnets are kept at a temperature below 1.7 K, in order to maintain their superconducting properties, which allow them to produce an average magnetic field of 8.3 T. Magnets of higher multipole orders are also present and are used to correct smaller imperfections in the field geometry and in the beams.

The first beam circulated around the LHC in September 2008, but due to an incident that damaged over 50 dipole magnets, the start of the main research programme and the beginning of the first substantial operational run, called Run 1, only began in November 2009.

In early 2013 Run 1 finished and the two following years, called Long Shutdown 1 (LS1), were used to upgrade the collider and the experiments. The second operational run, called Run 2,

started in 2015. After about four years of successful operation, Run 2 finished in 2018. Currently the LHC is not operational, as the accelerator and the experiments are undergoing another upgrade. The end of this second shutdown, named Long Shutdown 2 (LS2), is foreseen for 2021, when Run 3, the third operational run, will start.

The following section provides details on the acceleration stages of the beams (section 2.1.1), while the performance of particle accelerators, and in particular of the LHC, is discussed in section 2.1.2.

2.1.1 Acceleration stages

Before being injected in the LHC and being accelerated to a maximum possible energy of 6.5 TeV each, the beams are accelerated in various stages by smaller machines, which also provide beams to other experiments. Figure 2.1 shows a schematic representation of the CERN's accel-

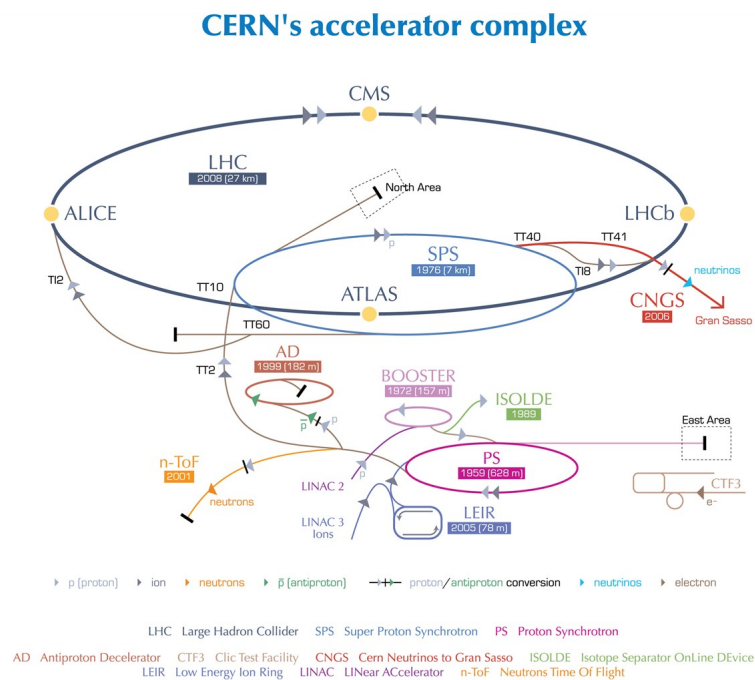


Figure 2.1: CERN Accelerator complex. The LHC is the last ring (dark grey line) in a complex chain of particle accelerators. The smaller machines are used in a chain to help boost the particles to their final energies and provide beams to a whole set of smaller experiments [76].

erator complex.

Protons, obtained by stripping hydrogen atoms of their electrons, are first accelerated to the energy of 50 MeV by the Linear Accelerator 2 (Linac2), then injected into the Proton Synchrotron Booster (PSB), where they reach 1.4 GeV. The next stages are the Proton Synchrotron and the Super Proton Synchrotron (SPS), which boost the protons up to respectively 25 GeV and 450 GeV. The protons are finally injected in bunches with a spacing of 25 ns into the LHC beam pipes, where they are accelerated up to a centre of mass energy of 13 TeV while they travel in

opposite directions. Once the beams have reached the maximum energy, they are made to collide at four points around the LHC ring.

The LHC is designed to accelerate also ion beams, mainly fully-stripped lead ions ($^{208}\text{Pb}^{82+}$), for which the acceleration procedure is slightly different. The first acceleration steps are the Linear Accelerator 3 (Linac3) and the Low Energy Ion Ring (LEIR); the ions are then injected in the Proton Synchrotron after which they follow the same path as the protons.

Located at the four collision points of the LHC, are four main experiments: two multi-purpose detectors A Toroidal LHC ApparatuS (ATLAS) [77] and Compact Muon Solenoid (CMS) [78], Large Hadron Collider beauty (LHCb) [79] which specialises in flavour physics and A Large Ion Collider Experiment (ALICE) [80] which focuses on heavy ions physics. In addition, there are other smaller experiments located near the four caverns. Large Hadron Collider forward (LHCf) [81], located at either side of ATLAS, measures neutral particle production, for use in cosmic rays research. TOTal cross-section, Elastic scattering and diffractive dissociation Measurement (TOTEM) [82], located at either sides of CMS, measures the total elastic and single or double diffractive cross-section of proton-proton collisions. Monopole and Exotics Detector at the LHC (MoEDAL) [83], at the LHCb cavern, searches for magnetic monopoles.

2.1.2 Performance of the LHC

The performance of particle colliders can be quantified in terms of the centre of mass energy of the collisions and the instantaneous luminosity. The centre of mass energy allows one to estimate the energy available for the production of new phenomena, while the instantaneous luminosity provides an estimation of the rate of a physics process a collider is able to produce. The luminosity is defined based on the following formula:

$$\frac{dN_{\text{process}}}{dt} = \mathcal{L} \times \sigma_{\text{process}}. \quad (2.1)$$

Basically, the rate of physics processes depends on the physical properties of the process, enclosed in the cross-section term σ_{process} , and on the features of the accelerator, encapsulated in the instantaneous luminosity $\mathcal{L}$. From equation 2.1 follows that the unit of measurement the instantaneous luminosity is $\text{cm}^{-2}\text{s}^{-1}$, often written as $\text{b}^{-1}\text{s}^{-1}$, using the barn unit of measurement ($1 \text{ b} = 10^{-24}\text{cm}^2$, therefore $10^{34}\text{cm}^{-2}\text{s}^{-1} = 1\text{pb}^{-1}\text{s}^{-1}$).

The luminosity can be calculated using the properties of an accelerator, using the following formula [84]:

$$\mathcal{L} = \frac{N_1 N_2 f_{\text{rev}} n_b}{4\pi\sigma_x\sigma_y}, \quad (2.2)$$

where N_1 and N_2 are the number of particles per colliding bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency and σ_x and σ_y are the horizontal and vertical dimension of the beams, assuming the same Gaussian profile for both beams.

Formula 2.2 assumes ideal head-on collisions of bunches where the particle densities in the three dimensions are uncorrelated; in practice, additional effects need to be taken into account. Such effects include the beams crossing angle, the collision offset and non-Gaussian beam profiles.

The integral over the data taking time of the instantaneous luminosity is the so-called integrated luminosity:

$$\mathcal{L}_{\text{int}} = \int_0^T \mathcal{L}(t') dt', \quad (2.3)$$

which provides an estimation of the total number of processes produced:

$$\mathcal{L}_{\text{int}} \times \sigma_{\text{process}} = N_{\text{process}}. \quad (2.4)$$

This estimation refers only to the total number of physics processes produced and it does not include any effect due to their experimental detection. In order to estimate the number of processes that can be observed in an actual detector, several other effects must be taken into account, such as the acceptance and efficiency of the detector and the dead-time.

The design centre of mass energy and instantaneous luminosity of the LHC are $\sqrt{s} = 14$ TeV and $\mathcal{L} = 1.0 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ [85], while for lead-lead collisions at a centre of mass energy of $\sqrt{s_{NN}} = 5.02$ TeV the design luminosity is $10^{27} \text{cm}^{-2} \text{s}^{-1}$ [86].

Due to technical limitations the design energy has not been reached yet: the highest centre of mass energy in pp collisions reached is $\sqrt{s} = 8$ TeV in Run 1 and $\sqrt{s} = 13$ TeV in Run 2. The design instantaneous luminosity has been reached and overtaken, with the record instantaneous luminosity of $\mathcal{L} = 2.06 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$ reached in 2017 [87].

2.2 The ATLAS detector

ATLAS (A Toroidal LHC ApparatuS) [77] is a general-purpose detector located at the CERN's "Point 1" cavern, the closest access to the LHC from the CERN Meyrin site.

The detector is designed to run at the highest luminosity provided by the LHC. It measures about 44 metres in length and 25 metres in diameter, for a weight of about 7000 tons.

The aim of the ATLAS detector is to exploit the high luminosity pp collisions provided by the LHC to test the SM and search for new physics Beyond the Standard Model; one of the milestones of the ATLAS physics program is the search for the Higgs boson, which was observed in 2012 by both ATLAS and CMS [1, 2].

The ATLAS detector shows the typical design of a general-purpose detector, with a forward-backward symmetry with respect to the interaction point, defined as the region where the two proton beams collide and a structure composed of concentric layers, each designed to accomplish a specific task.

The detector is divided in two regions, a central region known as barrel and a forward region, known as end-cap. There are two endcaps, one on either side of the detector.

The overall ATLAS detector layout is shown in figure 2.2. The innermost layer is the Inner Detector (ID), which allows high-resolution tracking and vertexing on charged particles. It is immersed in a 2 T field generated by a thin superconducting solenoid. The next layers are the electromagnetic and hadronic calorimeters, which allow to perform precision measurements of the energy of photons and electrons, and hadronic jets. The calorimeters are surrounded

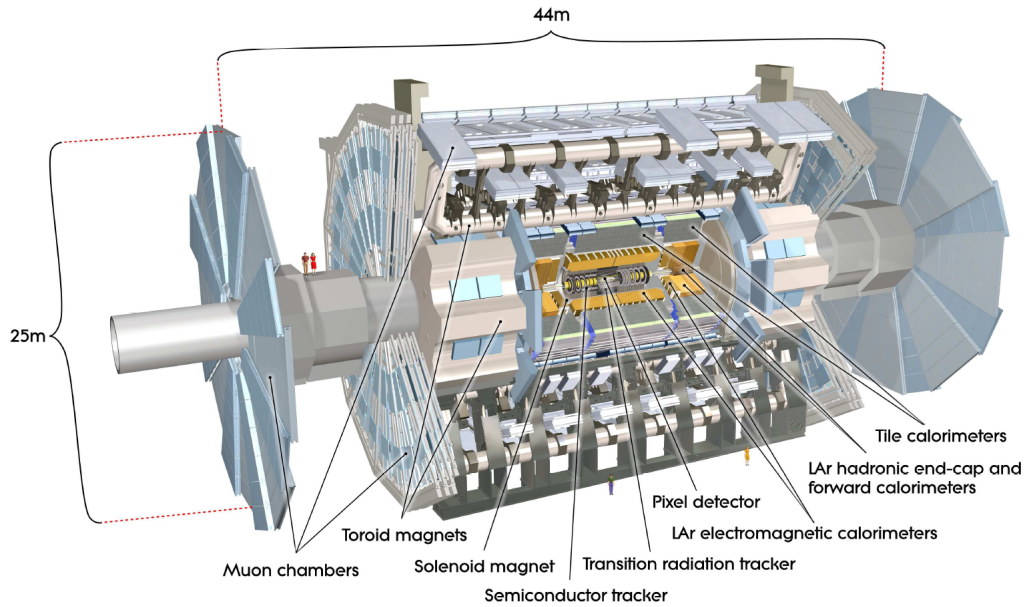


Figure 2.2: Cut-away view of the ATLAS detector. The dimensions of the detector are 25 m in height and 44 m in length. The overall weight of the detector is approximately 7000 tons. Picture taken from [77].

by the Muon Spectrometer (MS), which is the outermost layer, enclosed in a toroidal magnetic field; together with the ID it allows precise measurement of momentum and position of muons. The following sub-section provides an overview of the ATLAS coordinate system, used consistently in all ATLAS publications and in this thesis. The subsequent sub-sections discuss in detail the various ATLAS sub-detectors and systems.

2.2.1 The ATLAS coordinate system

The nominal interaction point is defined as the origin of the ATLAS coordinate system. The beam direction defines the z -axis, with the x - y plane transverse to it. The positive x -axis is defined as pointing from the interaction point to the centre of the LHC ring and the positive y -axis is defined as pointing upwards.

Given the symmetry of the detector and the pp collisions with respect to the interaction point, cylindrical coordinates are employed. The polar angle θ is measured from the beam axis, while the azimuthal angle ϕ is the angle around the beam axis. The polar angle is often reported in terms of a quantity known as pseudorapidity (η), defined as

$$\eta = -\ln\left(\tan(\theta/2)\right), \quad (2.5)$$

with $\eta = 0$ corresponding to the direction perpendicular to the beams and $\eta = \pm\infty$ to the forward regions $\theta = 0, \pi$. This quantity is extensively used in particle physics instead of the polar angle θ , as differences in pseudorapidity are Lorentz invariant under boosts along the longitudinal axis.

2.2.2 The magnet system

The ATLAS magnet system, 22 m in diameter and 26 m in length, is composed of four large superconducting magnets. They generate the magnetic field needed to bend the trajectory of charged particles in order to perform momentum measurements. The four components are a central solenoid, a barrel toroid and two end-cap toroids. Figures 2.3(a) and 2.3(b) show the geometry of the magnet system and its components. The central solenoid is located between

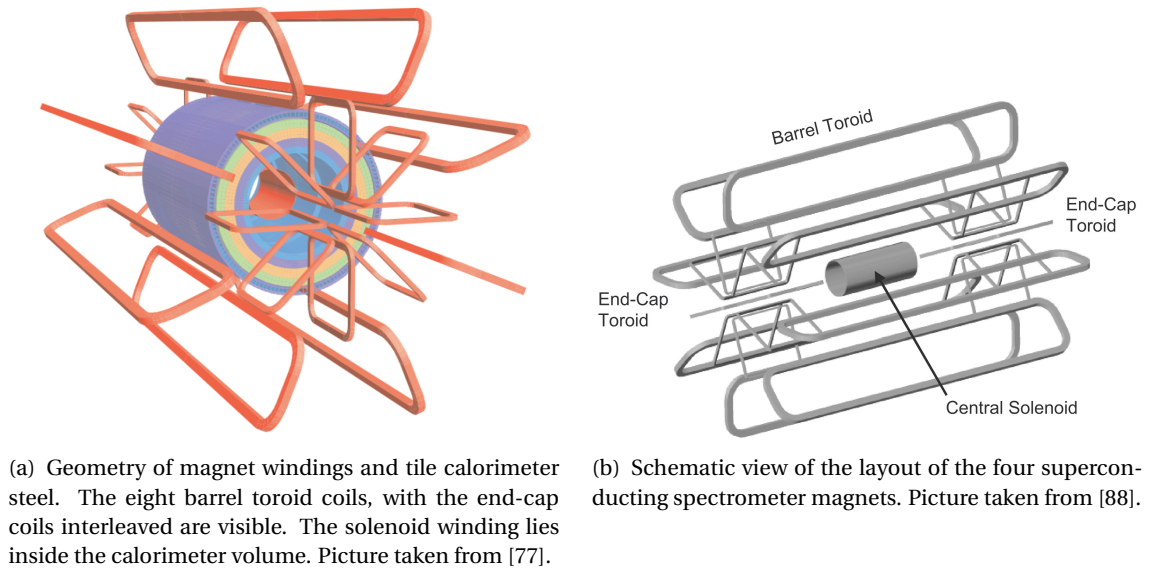


Figure 2.3: The ATLAS magnet system.

the ID and the electromagnetic calorimeter aligned to the beam axis, and has an axial length of 5.8 m, an inner diameter of 2.46 m and an outer diameter of 2.56 m. It provides a 2 T field parallel to the beam axis able to bend the charged particles that go through the ID in the ϕ direction, allowing an accurate momentum measurement up to 100 GeV [88].

The toroid system, used for the measurement of the momentum of muons, is composed of a barrel and two end-cap toroids and surrounds the calorimeters. The barrel toroid is composed of eight coils with a length of 25.3 m and inner and outer diameters of 9.4 m and 20.1 m respectively. It provides a toroidal magnetic field of approximately 0.5 T to the central muon detector. The end-cap toroids, also composed of eight coils each, produce a 1 T toroidal magnetic field for the end-cap regions of the muon spectrometer. They have a length of 5.0 m, an inner diameter of 1.65 m and an outer diameter of 10.7 m. The toroids allow the deflection of charged particles that cross the muon spectrometer in the η direction.

The magnet system composed of both solenoids and toroids is one of the main differences between ATLAS and CMS, which employs a single solenoid magnet [89]. The usage of the toroids allows the generation of a magnetic field over a large volume, for a reduced material budget. This approach helps to minimise the multiple scattering probability for muons crossing the magnets, which is one of the limiting factors of the muon momentum resolution.

2.2.3 The Inner Detector

The ATLAS ID [90] is the innermost sub-detector of ATLAS, designed to detect electrically-charged particles in a high particle flux environment, allowing a precise measurement of their direction, momentum, and charge. It has a length of 6.2 m, a diameter of 2.1 m and is all immersed in a solenoidal magnetic field of 2 T.

The ID is composed of three independent sub-detectors, shown in figure 2.4. The innermost

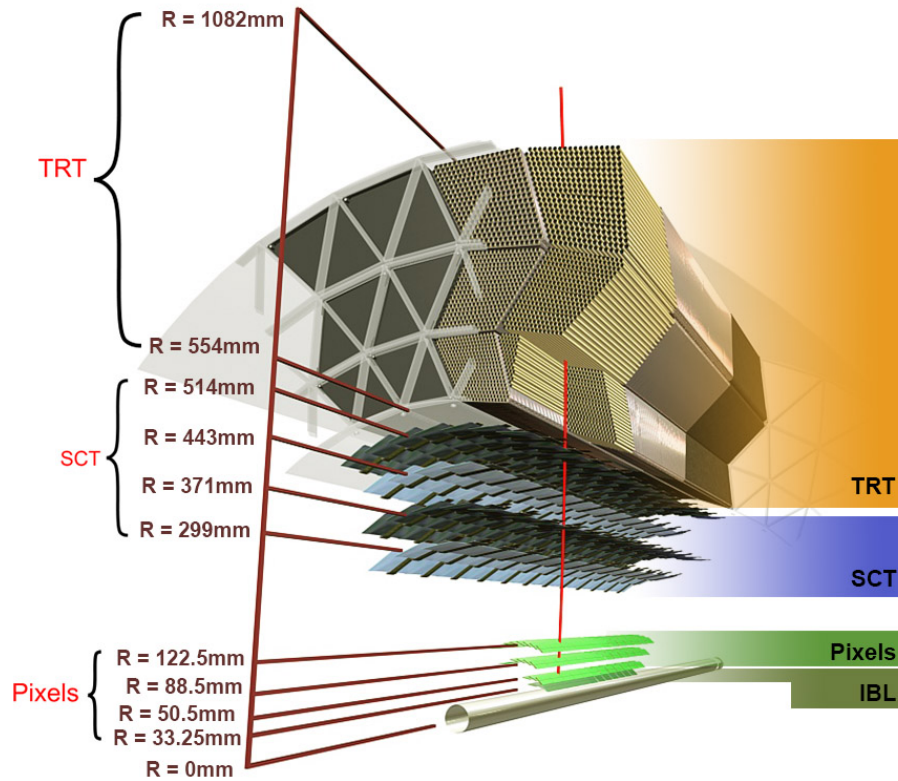


Figure 2.4: Diagram of the ATLAS ID and its sub-detectors. Picture taken from [91].

sub-detector is the Silicon Pixel Tracker (Pixel), composed of silicon pixel layers. Its innermost layer, named Insertable B-Layer (IBL) was added during the LS1. The second sub-detector is the SemiConductor Tracker (SCT), composed of micro-strip silicon detectors. The outer sub-system is the Transition Radiation Tracker (TRT), made of several layers of drift tube (also called straw tube) detectors.

The choice of the usage of three different technologies for the ID is strongly influenced by the need to keep costs within affordable bounds, as silicon detectors provide high-resolution measurements but are highly expensive. For this reason the TRT employs a more economical technology, at the cost of a lower resolution.

The combination of the sub-detectors, which are discussed more in detail in the following paragraphs, allows robust track reconstruction, with accurate impact parameter resolution ($\sim 20\mu\text{m}$) and precise Primary Vertex (PV) and Secondary Vertex (SV) reconstruction for charged particles (tracks) above 500 MeV and within $|\eta| < 2.5$.

IBL

The IBL [92] is the closest layer to the beam pipe, as shown in figure 2.4. It is composed of 6 million channels and each pixel measures $50 \times 250 \mu\text{m}^2$, for a resolution of $8 \times 40 \mu\text{m}^2$, with a pseudorapidity coverage of $|\eta| < 2.5$.

This layer was added during the LS1 of the LHC (2013/2014) at a distance from the beamline of about 3.3 cm. Due to the aging of the former innermost layer of the Pixel detector, named B-layer, the ATLAS detector could suffer efficiency losses in Run 2. Instead of replacing the B-layer, the IBL was designed as an additional pixel layer, improving the vertexing and the impact parameter reconstruction.

Pixel

The Pixel detector contains about 80 million semiconductor silicon pixels, with a nominal size of $50 \times 400 \mu\text{m}^2$ and a resolution of $10 \times 115 \mu\text{m}^2$ [93]. It measures 48.4 cm in diameter and 6.2 m in length, for a pseudo-rapidity coverage of $|\eta| < 2.5$. In the barrel region ($|\eta| < 1.7$) it is composed of three concentric layers placed 50.5 mm, 88.5 mm and 122.5 mm respectively from the beamline. Six disk layers are placed in the end-cap region ($1.7 < |\eta| < 2.5$), three per side, such that a charged particle that crosses the layers will generate at least three hits, where a hit is defined as the energy deposits in a detection cell.

The Pixel detector, including the IBL, is the ID sub-system with the highest granularity, as it needed to perform high-precision momentum and vertex measurements in the region close to the interaction point, where the particle density is extremely large.

SCT

The SCT sub-detector consists of 4088 modules of silicon-strip detectors [94], of which 2112 are arranged in four concentric barrels [95] ($|\eta| < 1.4$), ranging from 299 mm to 514 mm from the beamline, and 1976 are arranged in nine disks per endcap [96] ($1.4 < |\eta| < 2.5$). Each SCT module has two sensor layers and each sensor has 768 strips, so that the precision measurement of the position is performed using four points, one per layer, corresponding to a total of 8 hits in the barrel region. The readout strips are 6.356 cm long with a pitch of $80 \mu\text{m}$ on $285 \mu\text{m}$ thick sensor material, for an intrinsic resolution of $17 \mu\text{m}$ in the $R - \phi$ direction and $580 \mu\text{m}$ in the z direction. Since the SCT is further away from the beam-pipe with respect to the Pixel detector, it is designed to cope with a reduced particle density, allowing a reduced granularity of the detector while maintaining a comparable $R - \phi$ resolution as the Pixel.

TRT

The outermost sub-detector of the ID is the TRT. It is a gaseous detector, composed of 4 mm diameter straw tubes made of a multilayer film reinforced with carbon fibers and containing a $30 \mu\text{m}$ gold plated tungsten wire in the centre. The straws are filled with several gas admixtures,

depending on their position in the detector; the most used gas admixture is 70% Xe, 27% CO₂ and 3% O₂ [97].

As the with other sub-detectors, the TRT is divided into a barrel region [98] ($|\eta| < 1$) and an end-cap region [99] ($1 < |\eta| < 2$). The barrel region comprises 105088 144 cm long straws parallel to the beam axis arranged in 73 layers, while the endcap regions contain each 122880 39 cm long straws oriented radially in 160 layers. The inner radius of the TRT is 554 mm and its maximum radius is 1082 mm.

When a charged particle crosses a straw it ionises the gas. The resulting free electrons drift to the anode wire where they are multiplied in an electric avalanche. Since the electric field in the tube is known, the distance between the particle trajectory and the anode can be calculated using the time that electrons take to drift to the wire. The ions resulting from the ionisation drift towards the walls of the straws, but are not used in the detection of charged particles.

The space between the straws is filled with a material with a different dielectric constants with respect to the straws. Exploiting the Transition Radiation (TR), soft X-ray photon emitted by a charged particle when crossing the boundary between materials with different dielectric constants, the TRT is also capable of performing Particle IDentification (PID). The TR photons are detected in the straw tubes by absorption on xenon atoms and subsequent ionisation. In ATLAS normally only electrons reach the velocity needed for the TR emission, therefore the detection of TR photons indicates the presence of an electron.

The TRT has a spatial resolution of 130 μm , much lower than the previous layers, but it allows an almost continuous tracking, with more than 30 hits per charged particle. The lower resolution is therefore compensated by the high number of hits and the long lever arm (larger than half a meter); this adds a significant improvement in the momentum resolution for charged tracks.

2.2.4 The calorimeters

The ATLAS calorimeter system, which surrounds the solenoid, is composed of two main sub-detectors: the electromagnetic calorimeter [100, 101] and the hadronic calorimeter [102, 103], shown in figure 2.5. The two calorimeters are designed to stop and measure the energy of electrons and photons (electromagnetic calorimeter) and hadrons (hadronic calorimeter). The calorimetric system covers the full ϕ range and $|\eta| < 4.9$. The inner sub-system is the electromagnetic calorimeter, surrounded by the hadronic calorimeter; both detectors are composed of the barrel section and two symmetric endcaps. The following paragraphs provide an overview of the two calorimetric systems.

The electromagnetic calorimeter

The electromagnetic calorimeter is a sampling calorimeter that employs lead-Liquid Argon (LAr) [101] with accordion-shaped kapton electrodes and lead absorber plates over its full coverage.

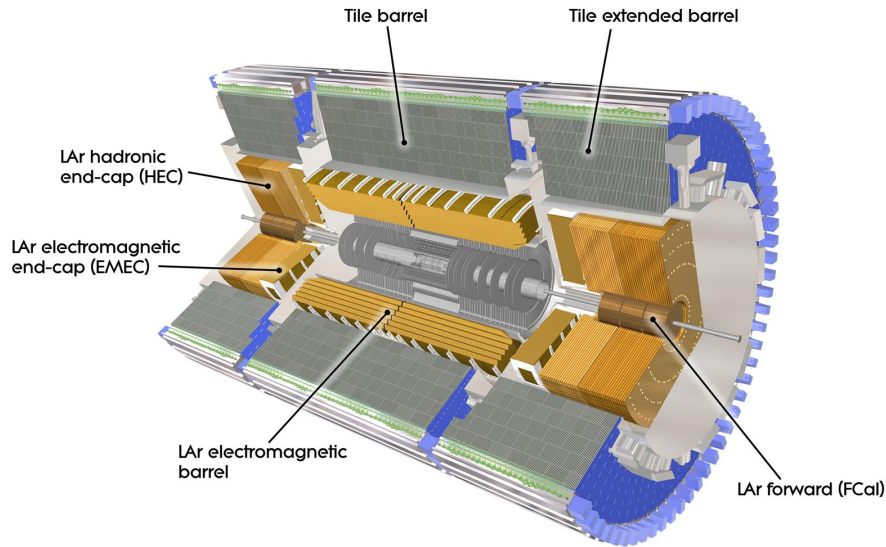


Figure 2.5: Cut-away view of the ATLAS calorimeter system. Picture taken from [77].

The choice of the LAr as the active material was driven by its intrinsic linear response, large signal yield and resistance to radiation. The accordion-shape of the absorber and electrodes, oriented in the radial direction, allows a complete coverage avoiding cracks in the azimuthal direction.

The barrel section is divided in three longitudinal layers, characterised by different depths and $\eta - \phi$ plane segmentations. The first layer has a fine segmentation, allowing it to perform accurate measurements of the coordinates of the particles that enter the calorimeter. It has a thickness of 4.3 radiation lengths (X_0), much smaller than the second layer, which has a thickness of 16 X_0 . This layer stops most of the showers and provides additional information on the shower position. The third layer, 2 X_0 thick, has a coarser granularity; it is used to estimate the leakage of showers beyond the electromagnetic calorimeter.

The total thickness of the electromagnetic calorimeter varies as a function of η : in the barrel region it ranges between 22 and 33 X_0 , while in the endcap it ranges between 24 and 38 X_0 .

The performance of the electromagnetic calorimeter was assessed using electron beams fired against modules which are the same as the ones present in ATLAS. The resulting energy resolution is [104]:

$$\frac{\sigma(E)}{E} = \frac{10\%}{\sqrt{E}} \oplus \frac{0.17\text{GeV}}{E} \oplus 0.7\% \quad (2.6)$$

with E expressed in GeV. The first term in equation 2.6 is due to the stochastic behaviour of the showers, the second term is due to electronic noise and the third term is due to non-uniformities of the calorimeter response. The effect of the last term is independent of the energy of the showers.

The hadronic calorimeter

The hadronic calorimeter surrounds the electromagnetic calorimeter. The barrel region, called the Tile calorimeter [102], is a sampling calorimeter: it uses iron plates as an absorber and plastic scintillating tiles as the active material. Wavelength-shifting fibers are coupled to the tiles and are read out by photomultipliers.

The endcap regions use the same principle as the electromagnetic calorimeter, but a different absorber (copper) and a different geometry. A forward region, the closest possible to the beam, is covered by a LAr forward calorimeter (FCal) [103]. The choice of different technologies for the hadronic calorimeter was dictated by the different particles flux and performance requirements as a function of the pseudorapidity.

The thickness of the hadronic calorimeter is about 7.4 interaction lengths (λ_I) for both the barrel and end-cap regions. Combined with the inner sub-systems of the ATLAS detector, the total thickness is always at least $10 \lambda_I$.

The resolution of the hadronic calorimeter is assessed by firing a pion beam against a prototype of the electromagnetic and hadronic calorimeters. The resulting resolution for the Tile calorimeter is [102]:

$$\frac{\sigma(E)}{E} = \frac{52\%}{\sqrt{E}} \oplus \frac{160\text{GeV}}{E} \oplus 3.0\% \quad (2.7)$$

with E expressed in GeV and the three terms are due to the same effects as for the electromagnetic calorimeter. A similar resolution is obtained also for the forward LAr hadronic calorimeter.

2.2.5 The Muon Spectrometer

The Muon Spectrometer (MS) [105], sketched in figure 2.6, is the outermost sub-detector of ATLAS. It surrounds the calorimeters and is used to measure the momentum of muons with detectors partitioned in fast coarser-precision *trigger* chambers and higher precision detectors of slower readout speed. All chambers provide positional information on the muon trajectories, translated into momentum measurements thanks to the deflection of the three ATLAS toroids. The chambers are arranged such that particles from the interaction point traverse three measuring layers, with a geometry optimised for optimum momentum resolution.

As with the other ATLAS sub-detectors, the MS is divided in a barrel region ($0 < |\eta| < 1$) and two endcap regions ($1 < |\eta| < 2.7$). The barrel region is composed of three cylinders concentric with the beam axis at radii of about 5, 7.5, and 10 m; the endcap regions are composed of four disks each, at distances of 7, 10, 14, and 21–23 m from the interaction point. The MS provides almost complete coverage in the $|\eta| < 2.7$ pseudorapidity range, except for an opening in the central R- ϕ plane at $\eta = 0$, known as “MS crack”, for the passage of cables and services of the rest of the ATLAS detector.

Monitored Drift Tubes (MDT) chambers provide the precision tracking and momentum measurement for both barrel and endcap, except for the innermost layer of the endcap, where Cathode Strip Chambers (CSCs) are used. The MS is also equipped with two types of trigger detectors,

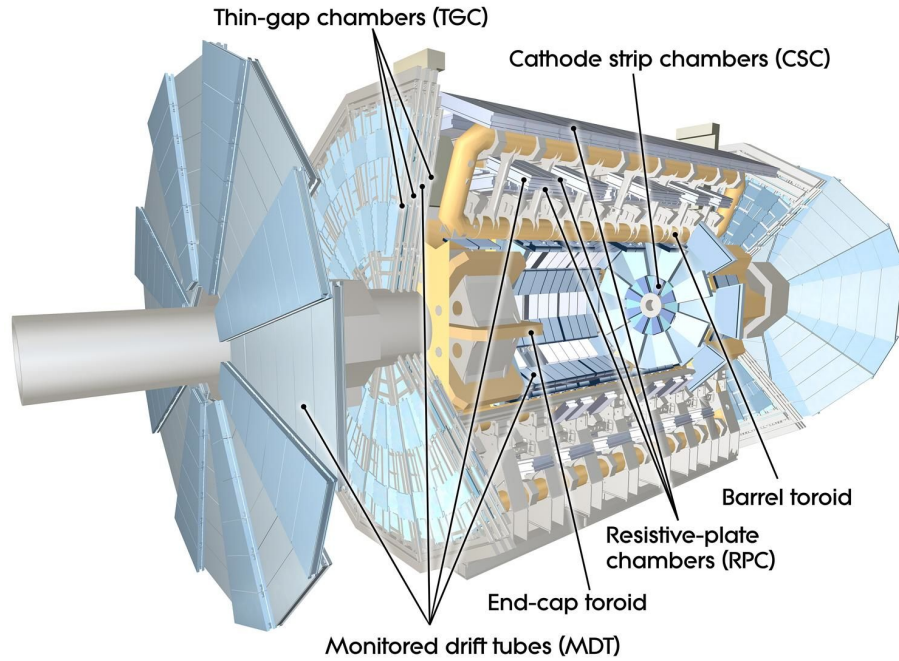


Figure 2.6: Cut-away view of the ATLAS muon system. Picture taken from [77].

the Resistive-Plate Chambers (RPCs) in the barrel region and the Thin-Gap Chambers (TGCs) in the endcap.

MDT chambers are made of aluminium tubes of 30 mm diameter and 400 μm wall thickness, with a 50 μm diameter central W-Re wire, surrounded by a non-flammable Ar-CH₄-N₂ mixture at a pressure of 3 bar. In order to improve the resolution on the particle position beyond the single wire limit (80 μm), the MDT chambers are constructed with 2×4 layers of drift tubes for the inner layer of MDTs and 2×3 layers for the middle and outer MDTs sets. The drift tubes layers are disposed orthogonally with respect to the beam axis, and they only provide a measurement of the η coordinate, while the measurement of the ϕ coordinate is provided by the RPC and TGC chambers.

The CSCs are multiwire proportional chambers with a symmetric cell, with cathode strips positioned above and below the anode wires. In order to enhance the precision of the measurements, a set of wires is orthogonal and the other is parallel to the beam axis, providing a measurement of the transverse coordinate. The gas used in this case is a non-flammable mixture of Ar-CO₂-CF₂.

The RPCs are gaseous detectors, comprised of two parallel resistive plates separated by insulating spacers. The primary ionisation electrons are multiplied into avalanches by a high, uniform electric field, and then detected by aluminium strips separated from the plates by an insulating film. These chambers are mounted in the barrel region, while the endcap is equipped with TGCs, which are multi-wire proportional chambers. RPCs and the TGCs provide a fast signal, $\sim 15 - 20$ ns, making them ideal for triggering purposes.

2.2.6 Trigger system

The LHC is designed deliver a bunch crossing every 25 ns at the interaction points, as discussed in section 2.1. The resulting rate of collisions is much higher than the ATLAS readout and recording capabilities. The purpose of the ATLAS Trigger system [106] is to perform a fast online selection of the physics events¹ and decide whether to record or reject the data from a bunch crossing, effectively reducing the recording rate from more than 30 MHz to about 1kHz.

Due to the high rate of bunch crossings, the ATLAS trigger system is composed of two levels, so that the first level performs a fast but coarse decision, and the second level can perform a more complex decision at lower rates. The two trigger levels are named Level-1 trigger (L1) and High Level Trigger (HLT).

The L1 makes a quick decision based on coarse information from the calorimeters and the MS. If an event is accepted by the L1, it is passed to the HLT. At this level, more information is used to make a decision. If the event passes also the HLT selection, it is permanently stored.

When the output rate of a trigger selection is too high, it can be prescaled by a factor P , which means that only a $1/P$ fraction of the events that pass the selection is accepted and recorded. The collection of the L1 and HLT selections (trigger chains), together with their prescales, defines a *trigger menu*, which is designed based on the physics program taking into account the LHC luminosity.

The ATLAS trigger is part of a larger system, known as Trigger and Data Acquisition (TDAQ) system [107]. It is in charge of the data flow in the experiment, as well as control, configuration and monitoring aspects.

A scheme of the TDAQ system is shown in figure 2.7, highlighting the interplay between the different sub-components. The entire data acquisition process is supervised by the Detector Control System (DCS) [108], which allows the coherent and safe operation of the ATLAS detector hardware and serves as a homogeneous interface to all sub-detectors and to the technical infrastructure of the experiment.

A more detailed description of the two trigger stages is provided in the next sub-sections.

Level-1 trigger

The ATLAS L1 [109, 110] performs the first fast decision using low-granularity information from the calorimeters (L1 calorimeter) and the MS (L1 muon), reducing the rate from more than 30 MHz to a maximum of 100 kHz.

The trigger decision is basically made by the Central Trigger Processor (CTP), which collects information from the L1 muon and L1 calorimeter trigger systems. The entire L1 is based on custom hardware, in order to enhance the speed of the decision time. If an event is accepted by the CTP, it is buffered in the Read-Out System (ROS) and then processed by the HLT.

The L1 system is also in charge of providing the HLT with Regions of Interest (RoIs), defined as

¹ An *event* is defined as the collection of the final state objects associated to a given bunch crossing.

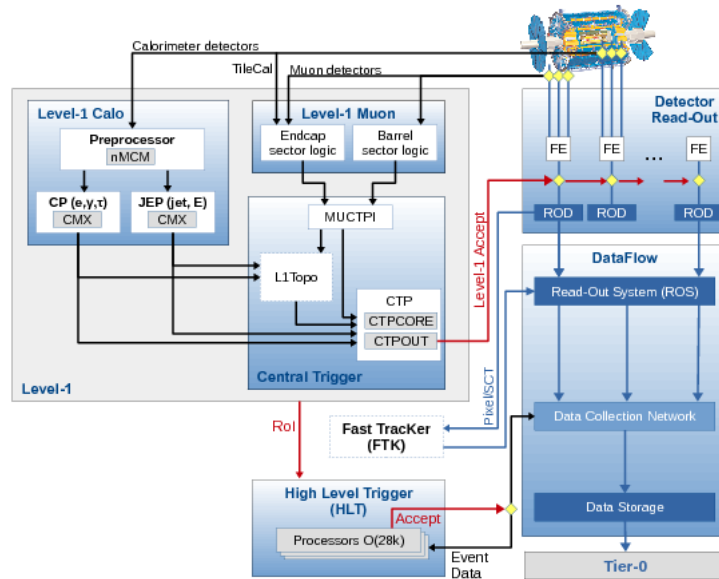


Figure 2.7: The ATLAS TDAQ system in Run 2 with emphasis on the components relevant for triggering. Picture taken from [106].

regions in the $\eta - \phi$ space containing the physics objects of interest. The RoIs are then used as seeds at the HLT stage. This approach allows a reduction of the computational time, as the HLT has to analyse only a small portion of ATLAS, instead of reconstructing the full detector.

During Run 2 the L1 has been upgraded by adding a topological trigger processor (L1Topo) [107], able to combine kinematic information from L1 muon and L1 calorimeter. The addition of L1Topo allows a complex selection based on the topology of the event, in order to improve the L1 selection.

HLT

The HLT [111] performs the second-stage decision of the ATLAS trigger, reducing the rate from 100 kHz to approximately 1 kHz, which is the actual recording rate of the detector.

The system is based on software algorithms, running on a cluster of computers (HLT farm), that reconstruct the physics objects contained in the RoIs received from the L1. The reconstruction uses additional information with respect to L1, such as tracking from the ID and finer granularity calorimetric inputs. If an event passes the HLT selection it is written to disks for the offline analysis.

One of the limiting factors for the HLT is the ID tracks reconstruction, which is a CPU expensive process and takes most of the time available to make a decision. For this reason, a new hardware based Fast Tracker (FTK) [112] is being incorporated in the HLT system for Run 3. It will provide ID track reconstruction to the HLT at the L1 rate, exploiting lookup tables stored in custom associative memory chips. This will allow the HLT to perform more complex selections without having to reconstruct the tracks. Two possible approaches will be available for the tracks reconstruction: regional tracking and full-scan tracking. The former is a faster approach,

which provides only the tracks in the RoIs obtained from the L1, while the latter provides the reconstruction of all the tracks in the ID. More details about the FTK structure and purpose are given in chapter 4.

2.3 The HL-LHC upgrade

The LHC physics program has completed the Run 2 data-taking period in 2018 and is heading towards the third operational run, named Run 3. The beginning of the Run 3 operations is scheduled for 2021, after the end of the Phase-I upgrade of the accelerator. Figure 2.8 shows the timeline for the LHC operations and the planned upgrades. A Phase-II upgrade is sched-

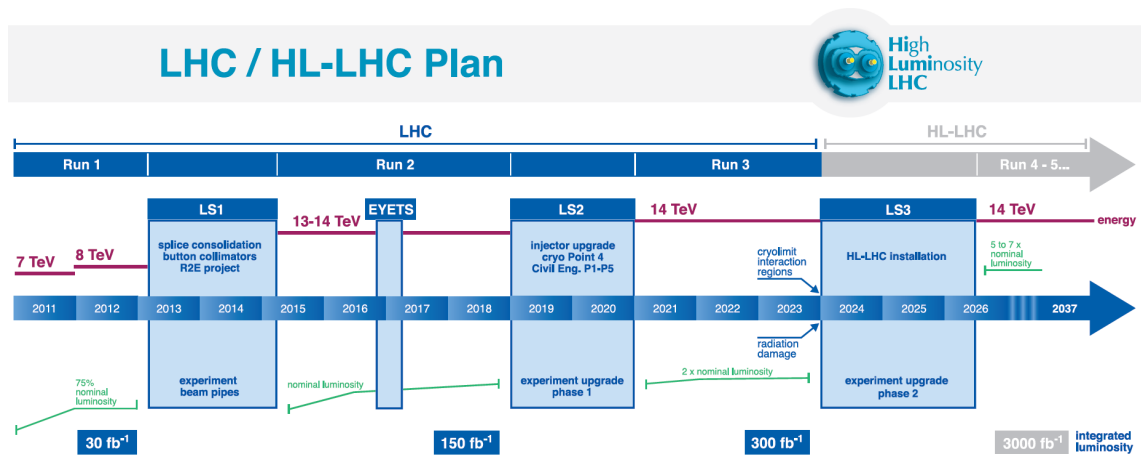


Figure 2.8: Timeline for the LHC accelerator operation and planned upgrades [113].

uled, after the end of Run 3 (2023) to further upgrade the LHC into the High-Luminosity Large Hadron Collider (HL-LHC) [85].

The HL-LHC is designed to deliver proton-proton collisions at the centre-of-mass energy of $\sqrt{s} = 14$ TeV. The design instantaneous luminosity is $5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$, five times the design instantaneous luminosity of the LHC, while the ultimate achievable instantaneous luminosity is $7.5 \times 10^{34} \text{cm}^{-2} \text{s}^{-1}$. In order to be able to reach such instantaneous luminosities, the number of pp collisions per bunch crossing will have to increase from an average of 34 during Run 2 to about 200.

The increase in luminosity will allow the ATLAS experiment to collect an unprecedented amount of data during the entire HL-LHC campaign, reaching about 3000fb^{-1} (3ab^{-1}) of integrated luminosity. The size of the collected dataset will grant the possibility to perform advanced searches for physics BSM and precision measurements of the SM.

In order to reach this ambitious goal, the ATLAS detector must be upgraded: the ATLAS Phase-II upgrade [114, 115] is designed to improve the detector to face the more challenging experimental conditions and to replace aging detector components.

The following sub-sections provide an overview of the upgrades foreseen for the ATLAS detector.

Inner Tracker

The ATLAS ID, described in section 2.2.3, will be completely replaced at the end of Run 3. The new innermost component of the detector, called Inner Tracker (ITk), will be a silicon-only based tracker [116, 117], that will allow improved momentum resolution for reconstructed tracks.

The ITk will be composed of two main parts, shown in figure 2.9, a silicon pixel detector made of 5 barrel layers placed closest to the beamline, and a silicon strip detector with 4 barrel layers that extend out to higher radii. The end-cap region will be covered with a series of rings of both pixels and strips.

The ITk will also improve the $|\eta|$ coverage in the forward region, from a coverage up to $|\eta| = 2.5$ of the current ID to $|\eta| = 4.0$.

The design of the ITk allows a lower material budget than the ID, reducing the uncertainty on tracking due to multiple scattering.

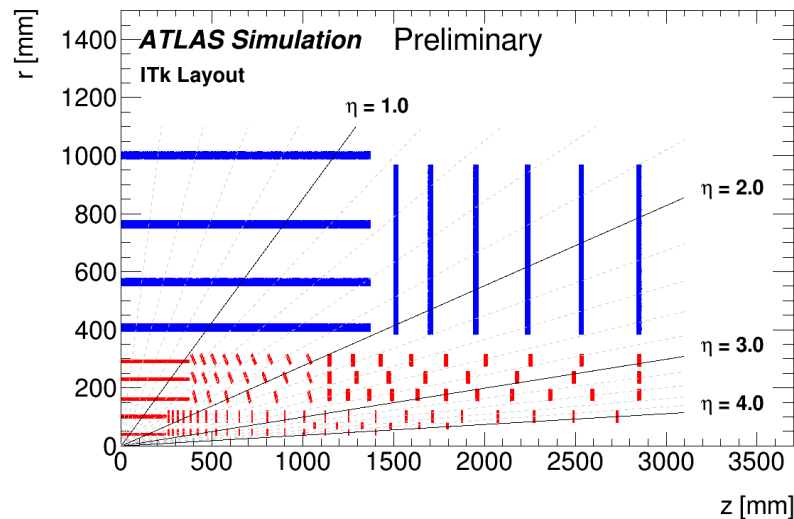


Figure 2.9: Schematic layout of the ITk for the HL-LHC phase of ATLAS as presented in [118]. The active elements of the barrel and end-cap Strip Detector are shown in blue; for the Pixel Detector the sensors are shown in red for both the barrel layers the end-cap rings. Only one quadrant and only active detector elements are shown. The horizontal axis is the axis along the beam line with zero being the interaction point. The vertical axis is the radius measured from the interaction region.

Calorimeters

The ATLAS calorimeters, described in section 2.2.4, will have new frontend and readout electronics, specifically designed to withstand the harder radiation conditions for the duration of the HL-LHC data taking period.

In particular, the electronics architecture of the Liquid Argon Calorimeter will be optimised to allow a full-granularity output at 40 MHz, improving the trigger performance. The upgrades are described in detail in the Liquid Argon Calorimeter Technical Design Report [119] and in the Tile Calorimeter Technical Design Report [120].

Muon Spectrometer

A large fraction of the front-end, read-out and trigger electronics of the ATLAS muon spectrometer, introduced in section 2.2.5, will be replaced, allowing it to cope with higher trigger rates. In order to maintain muon identification and reconstruction performance, additional muon chambers will be also installed.

Following the extension of the acceptance of the inner tracker, the possibility of an extension of the MS up to $|\eta| = 4.0$, called high- η tagger, is currently being considered. The full studies on the upgrade of the ATLAS MS are provided in the Muon Spectrometer Technical Design Report [121].

Trigger and data acquisition

The harder experimental conditions with respect to Run 2 and 3, due to the higher number of interactions per bunch crossing, impose the need of a new TDAQ system. The designed system, shown in the Phase-II TDAQ Technical Design Report [122], presents a two-level architecture. A hardware based Level 0 trigger will operate at an output rate of 1 MHz, followed by a processing farm, named Event Filter (EF), which will reduce the output data rate to 10 kHz, for permanent storage. The new architecture will allow higher data granularity and enhanced flexibility, compared to the current system.

The Phase-II TDAQ upgrade includes also a hardware based tracking pre-processor, the Hardware Track Trigger (HTT), sitting in parallel to the processing farm in the EF. The HTT is designed to provide the EF with the tracks from the ITk when needed, as the EF would not be able to reconstruct them, due to the required computing resources. Tracks can be reconstructed with two approaches: regional tracking, which provides the track reconstruction in regions of interest, and global tracking, which provides the full ITk reconstruction.

High-Granularity Timing Detector

A new sub-detector will be added to the ATLAS detector in order to precisely measure the timings of charged particles in the $2.4 < |\eta| < 4.0$ region. The High-Granularity Timing Detector (HGTD) [123] will be installed in front of the LAr calorimeter in order to reduce the background originated by the high number of jets due to the large number of interactions per bunch crossing.

THE ATLAS DATA ANALYSIS AND SIMULATION FRAMEWORKS

3

A detector like ATLAS provides a huge amount of data, which need to be stored, processed and then analysed. In addition, high energy physics experiments vastly base their analyses also on simulations, which allow one to compare real data with theoretical predictions.

In order to manage all these complicated tasks, the ATLAS software framework Athena [124], based on the Gaudi framework [125] developed by the LHCb collaboration [79], manages all the level of processing of the ATLAS data, from high-level trigger to event simulation, reconstruction and analysis.

The first section of this chapter, section 3.1, provides an overview of the generation of simulations, based on Monte Carlo (MC) techniques. The approach described, used to generate MC simulations of hadron-hadron collisions at high energy, is common to most high energy physics experiments. For this reason, the softwares used are not developed internally in ATLAS, but are developed and maintained by international collaborations.

The second section 3.2 gives a description of the approach followed by the ATLAS experiment for the reconstruction of physics objects in both simulations and real data.

3.1 Monte Carlo simulation

MC generators [126] are heavily used in particle physics to simulate the physics processes, predict their kinematic properties and their behaviour in the detector. Simulations can then be compared to real data, to validate theoretical predictions.

The MC simulation of physics processes at the LHC is a complicated task, usually split into several steps. The two main steps, discussed in the following sub-sections, are the simulation of physics processes and the detector simulation.

3.1.1 Simulation of physics processes

The simulation of the pp collisions requires the understanding of physics processes involving very different energy scales. Starting from the high energy scale of the deep-inelastic scattering between two constituents of the protons (quarks or gluons), down to the low energy scale of the final state partons that evolve into colourless hadrons. The different regimes can not be

described using the same approach; for instance the soft regime can not be described using perturbative QCD, which is used to describe high energy scale processes. This forbids a full analytic description of the process.

In order to overcome this issue, the simulation of physics processes is further divided in several steps, effectively factorising the different energy scales involved. This allows the usage of perturbation theory for the description of the hard regimes and phenomenological models for the soft regimes. Figure 3.1 illustrates the different steps involved in the pp collision simulation. The first part involves perturbation theory, which models the interaction of the partons inside two colliding protons via deep-inelastic scattering. All the partons involved in the pp in-

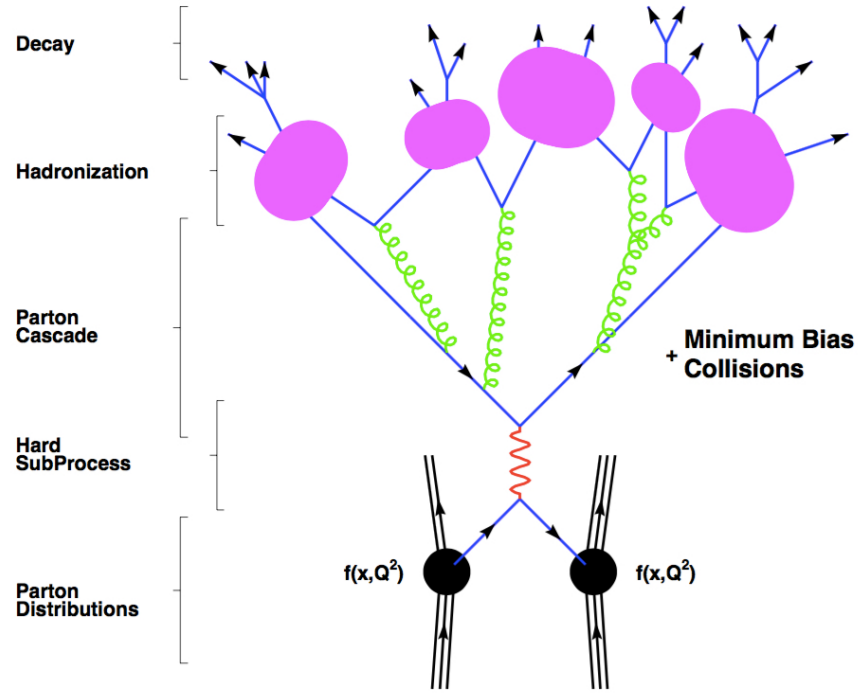


Figure 3.1: Representation of the different steps involved in the simulation of a pp collision at the LHC [127].

teraction are colour charged, therefore they predominantly interact via the exchange of gluons, which, in turn will radiate further gluons or split into quark/anti-quark pairs and so on, leading to the formation of a parton shower. This radiation process continues until the partons involved reach an energy scale of $Q \sim 1$ GeV and a new process, called hadronisation, becomes predominant. At this stage the partons recombine into colourless hadrons, which can be stable particles or can decay into stable final states. These last processes can only be described with phenomenological models.

A more detailed description of the various steps of the MC simulation procedure is provided in the following paragraphs.

Factorisation of the cross-section calculation and PDFs

As already mentioned, the hard scattering in pp collisions occurs between the partons inside the protons, which carry respectively a fraction x_a, x_b of the protons momentum. The total

cross section for a generic $pp \rightarrow X$ process can be written as [128, 129]:

$$\sigma_{pp \rightarrow X} = \sum_{a,b} \int dx_a dx_b f_a(x_a, \mu_F^2) f_b(x_b, \mu_F^2) \hat{\sigma}_{ab \rightarrow X}(p_a x_a, p_b x_b, \mu_F^2, \mu_R^2) \quad (3.1)$$

where the sum is taken over all initial state partons that can produce the given final state.

The $f_i(x_i, \mu_F^2)$ term, which describes the probability of finding a parton of type i carrying a fraction of the proton's momentum x_i , is known as Parton Distribution Function.

Parton Distribution Functions are universal, since they do not depend on the particular process and are usually measured combining information from deep-inelastic scattering experiments and hadron colliders.

The μ_F and μ_R factors are known respectively as factorisation scale and renormalisation scale. The μ_F factor can be described as the scale that separates the long- and short-distance physics, while μ_R is the renormalisation scale of the QCD running coupling. For a more detailed discussion see [128, 129].

The cross-section of the partonic process $\hat{\sigma}_{ab \rightarrow X}(p_a x_a, p_b x_b, \mu_F^2, \mu_R^2)$, where p_a and p_b are the momenta of the two colliding protons, can be explicitly computed at a fixed order in perturbation theory. This step is also known as Matrix Element (ME) calculation.

Matrix element calculation

The computation of the ME involves the calculation of the scattering matrix, or S-matrix, relating the initial and final state particles of the process. It can be performed at different orders in perturbation theory.

A process is said to be calculated at Leading Order (LO) when only tree-level diagrams are involved in the computation, while Next-to-Leading Order (NLO) computations include one loop diagrams or gluon emission.

While tree-level diagrams are generally convergent, NLO computations can lead to divergences. However, the Kinoshita-Lee-Nauenberg (KLN) theorem [130, 131] assures that at fixed orders in perturbation theory the infinities are exactly cancelled, yielding a finite result for the cross-section.

Issues might arise in fixed-order calculations with a large multiplicity of final states. In these cases, the integration over the full phase space will include configurations in which one, or more, final state partons become collinear or soft, leading to singularities. These divergences are not dealt with in the KLN theorem.

In order to remove these divergences, the ME calculation is truncated by the introduction of a cutoff, such that soft objects are excluded and a finite result can be obtained. The remaining part of the phase space, excluded by the introduction of the cutoff, is dealt with in the Parton Shower (PS) generation.

Parton shower

The PS generation step of the physics event simulation consists of the successive emission of soft collinear initial and final state partons, used to account for the processes that are not in-

cluded in the fixed order perturbative calculation. This step is an approximation as it assumes independent parton emissions, without considering virtual corrections. There are three possible QCD processes that contribute to the PS: $q \rightarrow gg$, $g \rightarrow gg$ and $g \rightarrow q\bar{q}$, where g is a gluon and q is a quark. The simulation is based on the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equation [132], which describes the evolution of the Parton Distribution Functions as a function of the transferred momentum. The equation is applied iteratively for each parton involved in the hard interaction, until the energy of the partons reach a regime in which perturbative QCD can not be applied any more.

The matching of the ME and the PS is not straightforward in general. It is performed by defining a procedure that can determine two orthogonal domains in the phase space, in order to avoid possible overlaps between the physics objects generated in the two stages, without causing discontinuities in the observable spectra. This is taken care of by dedicated matching algorithms, the most common implementations are the Catani-Krauss-Kuhn-Webber (CKKW) [133] and the Michelangelo L. Mangano (MLM) [134] schemes.

Hadronisation

While the parton shower evolves, its energy scale Q^2 decreases until it reaches the hadronisation scale $Q_0 \sim 1$ GeV. At this level, the confinement effects of QCD become important and the recombination of the partons into colourless objects (mesons and baryons) must occur. Since this process is dominated by non-perturbative QCD, generators rely heavily on phenomenological models. The most common models implemented in the MC generators are the cluster model [135, 136] and the Lund string model [137, 138].

The Lund model uses a string to represent the confinement potential between a quark-antiquark pair. For such a pair, as the colour charges move apart, the string is stretched, and its potential energy grows. Once the potential energy reaches the order of hadron masses, a colour singlet pair is produced and the string is split into two smaller strings. The process is repeated until all the energy has been converted into quark-antiquark pairs connected by short strings.

The cluster model firstly forces all the gluons to split into quark-antiquark pairs. The pairs are then clustered into forming colourless hadrons, which can decay and split into smaller clusters, depending on the energy available.

Figure 3.2 shows a sketch of the two hadronisation models, the Lund string model on the left and the cluster model on the right.

Final particles decay

After the hadronisation phase a number of stable and unstable hadrons are produced, and the latter must decay into particles that are stable at collider timescales. The simulation of hadron decays involves non-trivial modelling, as one has to consider several effects, such as the possible interference effects which can lead to CP-violation and the various excited states. Historically the standard generators included few matrix elements for hadron decays and at best

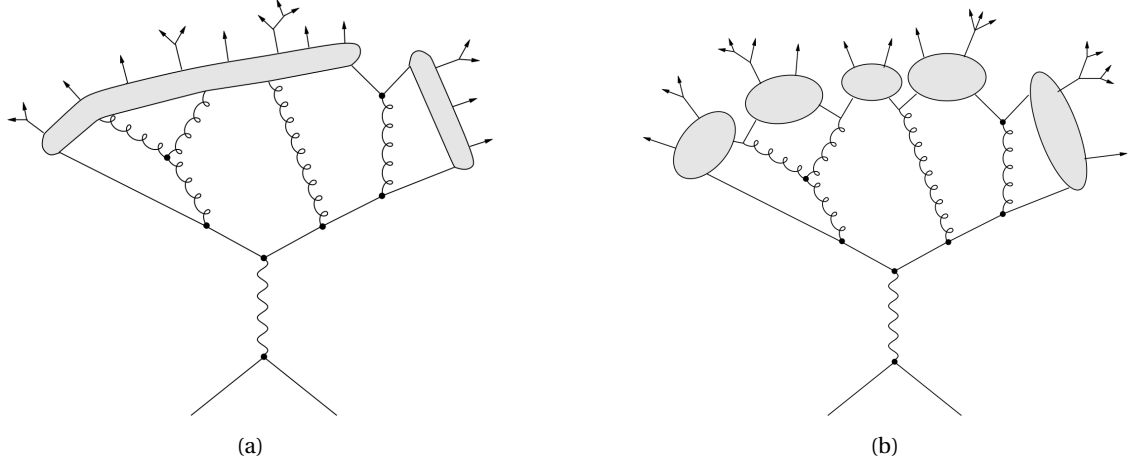


Figure 3.2: Schematic representation of the Lund string model, figure 3.2(a), and of the cluster model, figure 3.2(b). Pictures from [139].

used a naive Breit–Wigner smearing of the masses of the particles, while a more sophisticated simulation can be performed using external packages, such as EvtGen [140] for hadron decays, and TAUOLA [141] for τ decays. The EvtGen package in particular is extensively used in flavour physics because it is the most up to date in terms of experimental measurements and theoretical predictions. It also allows the simulation of $B^0 - \bar{B}^0$ mixing and rare B-meson decays that are of interest for the study of CP-violating phenomena. It also uses a proper spinor algebra, which allows one to take into account polarisation and helicity effects, allowing the calculation of the correct angular distributions of the final states.

The emission of soft QED radiation is also simulated by the standard generators, while for a more precise approach the PHOTOS [142] package provides LO and NLO QED corrections for the decays.

Underlying event

The Underlying Event (UE) [143] refers to the soft interactions that occur among the partons not involved in the hard scattering, also called spectator partons. These interactions are predominantly non-perturbative in nature, therefore the simulations are based on phenomenological models with many parameters, tuned using data [144].

Relatively hard interactions are still present also in the UE. Referred to as Multiple Parton Interactions (MPIs), these are described with a perturbative calculation and tuned with experimental data [145, 146].

Pile-up

When two bunches collide the number of pp interactions may vary and although, in general, only one pp interaction is characterised by a hard scattering, the remaining interactions must be taken into account. They are referred to as in-time Pile-Up (PU), to be distinguished from out-of-time PU, which arises from previous or later bunch crossings merged to the bunch

crossing of interest due to limited time resolution of the readout of the detector. Both in-time and out-of-time PU consist mostly of soft QCD interactions and are modelled in the same manner as the UE.

Monte Carlo generators

Several MC generators have been developed to implement the different steps of the simulation of LHC physics processes.

They can be roughly divided into two main categories, multi-purpose generators, capable of handling the full simulation chain, or ME generators, which are specifically designed to perform a sophisticated simulation of the hard scattering, but need to be interfaced to additional PS generators.

The main MC generators currently used are:

- **Pythia**, multi-purpose MC generator, performs LO calculations for $2 \rightarrow n$ ($n \leq 3$) processes and includes PS. The Lund string model is used for hadronisation and includes the UE simulation. There are currently two versions of this generator, Pythia6 [147], written in Fortran, and Pythia8 [148], written in C++;
- **Herwig** [149], multi-purpose MC generator, can compute LO ME for 2-to-2 processes and includes PS. The cluster model is used for hadronisation and for the UE description. It is usually interfaced with a standalone software, Jimmy [150], for UE and MPI simulation;
- **Sherpa** [151], multi-purpose MC generator, computes LO or NLO calculations for multi-leg processes. It contains its own parton shower algorithm, while the matching between PS and ME is performed with the CKKW method. Hadronisation is simulated with the cluster method and includes the UE simulation;
- **MadGraph** [152], ME generator, computes LO or NLO amplitudes for $2 \rightarrow n$ ($n \leq 6$) processes. Must be interfaced with another generator for PS, hadronisation and UE;
- **Powheg** [153], NLO ME generator. Must be interfaced with another generator for PS, hadronisation and UE.

3.1.2 Detector simulation

The output of a MC generator is a list of all the four-vectors of the particles involved in the generation and the information about their decay vertices, usually provided in the HepMC format [127]. The simulation at this stage is called *truth-level simulation*, used to perform *truth-level* studies on the physics processes.

It is not yet possible to compare the MC simulation with the real data collected by ATLAS or any other high energy physics experiment, as a simulation of the interaction of the particles with the detector and its response is needed.

The detector simulation is performed using Geant4 [154], a C++ based toolkit for the simulation of the passage of particles through matter. It is widely used in high energy, nuclear and accelerator physics, as well as in medical and space science.

The simulation of the ATLAS detector first considers the energy deposit of the particles in the detector sub-systems, taking into account the geometry and the materials. The following step is the conversion of the energy deposit into electric signals and currents (digitisation). The response of the various sub-detectors is then simulated, together with the complete structure of the ATLAS Trigger and Data Acquisition (TDAQ) system. At this stage the output of the simulation and the data are reconstructed using the same software.

The simulation of the full ATLAS detector is an expensive CPU process, therefore a faster detector simulation can be performed using ATLFast-II (or AF2) [155]. It allows a significant reduction of the CPU time necessary to process the events by considering a parametrised description of the behaviour of the detector. The obvious drawback of this fast simulation strategy is a reduced accuracy in the outcome of the simulation.

3.2 Object reconstruction

After the detector simulation, the output of a simulation is the same as the one obtained from real data, but before being able to perform a physics analysis another step is needed to convert the hits in the tracker and in the MS and the energy deposits in the calorimeters into physics objects to be used in an analysis. The main physics objects are tracks, vertices, electrons, photons, muons, jets, taus, and missing transverse energy; the work described in this thesis exploits mainly tracks, vertices and muons, for which a more thorough description is provided, while a brief description is given for the remaining physics objects.

3.2.1 Track reconstruction

When a charged particle travels through the ID, it releases localised ionisation signals (hits and clusters) in the sub-detectors, pixel, SCT and TRT, described in section 2.2.3. These charge deposits are used to trace the particle's trajectory and reconstruct a "track".

Tracks are of particular importance in data and simulation reconstruction, because they are used in the reconstruction of almost every other physics object, such as leptons, jets and Primary Vertices (PVs).

Assuming no multiple scattering and negligible bremsstrahlung radiation, the trajectory of a particle moving in a solenoid magnetic field, as the one present in the ID, can be described as a helix. The ATLAS tracks reconstruction employs 5 helix parameters P , measured at the point of closest approach to the z axis:

$$P = (d_0, z_0, \phi, \Theta, q/p) \quad (3.2)$$

where d_0 and z_0 are the transverse and longitudinal impact parameters, representing respectively the minimum distance between the track and the centre of the detector in the transverse

plane and in the longitudinal direction, ϕ and Θ are the longitudinal and azimuthal angles and q/p is the ratio between the charge and the momentum of the particle.

Several algorithms are available for track reconstruction [156, 157], the most used is the so-called inside-out method. It works starting from the centre of the ID, creating a “seed” by grouping together hits in the pixel and in the SCT; the track candidate is then extended outward to the TRT. The back-tracking algorithm, on the contrary, exploits an outside-in approach, seeding the track with clusters in the TRT and then extrapolating the candidate to the silicon detectors. These approaches result in a list of track candidates.

The following step of the track reconstruction procedure is the search for fake tracks, that do not correspond to an actual charged particle trajectory, and their removal. This cleaning process requires a trade-off between the purity of the track list, defined as the number of fake tracks present with respect to the total, and the track reconstruction efficiency. The candidates are ordered according to a track score, considering the energy of the track, the number of shared clusters with other tracks and the number of hits and holes, where a hole is defined when a hit is expected in a layer of sensors given the fitted trajectory of the track but none is found. The candidates with the lowest score are removed and further requirements are applied in order to reduce the overlap between the tracks.

3.2.2 Vertex reconstruction

Due to the large content of protons in the colliding bunches, several interaction vertices can be reconstructed in each event; their identification is essential for the correct reconstruction of the physics objects and their kinematics.

The PV reconstruction, where a PV is identified as the points where a pp interaction occurred, employs a sub-set of all the reconstructed tracks, selected to minimise the number of fake tracks and the number of tracks coming from secondary interactions.

Once the track's sub-set is defined, the vertices are reconstructed according to the following steps [158].

- A seed for the vertex fitting algorithm is identified. The transverse position of the seed is taken as the centre of the beam spot¹. The mode of the z-coordinates of tracks at their points of closest approach to the centre of the beam spot is used as z-coordinate of the seed.
- The Adaptive Vertex Fitting algorithm [159, 160] estimates the vertex position with an iterative fit of the ID tracks based on the least squares method. The fit starting point is the seed evaluated in the previous step and at the first iteration all tracks are assigned the same weights. As the iterations progress, tracks more compatible with the vertex are assigned higher weights and incompatible tracks are assigned small weights, so that their effect on the fit is minimum. Once the iterative procedure does not show any significant

¹ The beam spot is defined as the spatial region around the interaction point where the profiles of the two beams overlap.

change, the position of the vertex is defined and tracks not compatible are removed and used in the determination of another vertex.

- The procedure is repeated with the remaining tracks in the event, until all tracks are assigned to a PV or no additional vertex can be found in the remaining set of tracks.

The ATLAS standard approach associates the source of the event of interest to the PV with the largest sum of the squared momenta of the associated tracks ($\sum p_T^2$); all the remaining vertices are considered Pile-Up (PU) vertices.

The $\sum p_T^2$ approach is commonly used, but it was proved to be ineffective when low- p_T processes, such as flavour physics events, are considered: in this case a different algorithm is used, based on the kinematics of the event, more details are provided in section 7.5. Once all PVs have been identified, the impact parameters of the tracks associated with them are recalculated and expressed with respect to their position.

Vertices originated in secondary interactions, such as particles decays, are named Secondary Vertices (SVs). They can be reconstructed by looking at the displacement of the tracks with respect to the PV. This allows one to identify processes in which a particle has moved away from the primary interaction before decaying.

3.2.3 Muons

Muons are reconstructed mainly using tracks in the MS and in the ID [161]. Muon tracks in the ID are built like any other charged particles as described in section 3.2.1, while the tracks in the MS are reconstructed in different steps. Hits are searched in the various sub-systems of the MS and a straight line fit is performed in each layer, creating a set of segments. Muon track candidates are then built by fitting together hits from segments in different layers.

The reconstructed muon objects employed in ATLAS are classified into different categories, depending on which sub-detectors are used in the reconstruction:

- **Combined muons**, when tracks with compatible p_T and $\eta - \phi$ coordinates are found in both the ID and the MS a global refit is performed combining the information from the two sub-detectors. The result is a combined muon track which describes the trajectory of the muon through the full detector, taking into account also energy losses in the calorimeters. This procedure provides a good resolution at all p_T ranges, driven by the ID at low p_T and by the MS at high p_T . The acceptance for combined muons is $|\eta| < 2.5$, limited by the ID acceptance.
- **Segment-tagged muons**, ID tracks are classified as muons if a matching hit in the innermost MS layer can be found. This signature is in general associated with low p_T muons, which can not reach the other MS chambers due to the magnetic field, or muons that fell into the MS crack.
- **Calorimeter-tagged muons**, an ID track matched to a energy deposit in the calorimeters consistent with a Minimum Ionising Particle (MIP), with no hits in the MS. This signature

is generally due to muons that fall in the central pseudorapidity region, $|\eta| < 0.1$, where the MS crack is located.

- **Extrapolated muons**, in this case the muon's trajectory is reconstructed using only information from the MS, adding loose requirements on the compatibility with the interaction point. This signature is generally due to muons that fall outside of the ID acceptance, but still inside the MS acceptance, $2.5 < |\eta| < 2.7$.

In case of an overlap between different type of muons, a priority is given to the combined muons, then segment-tagged muons, calorimeter-tagged muons and finally extrapolated muons. Additional requirements are applied in order to separate “real” muons from muons originated by decay-in-flight of pions and mesons or hadrons punch-through, which, in both cases, are called “fake” muons.

A set of Working Points (WPs) is defined, using variables related to the combined track and requirements on the track quality, the charge, the momentum measurements in both the ID and the MS and requirements on the energy deposit in the calorimeters. Four WPs are available, *Loose*, *Medium*, *Tight* and *HighPT* [161]. The first three WPs are characterised by decreasing real muon efficiency and increasing fake muon rejection, the latter WP is designed for muons with $p_T > 100 - 200$ GeV, when the resolution is dominated by the MS and additional requirements are applied. In order to improve the background rejection, other requirements on track-based and calorimeter-based isolation can also be applied.

The performance of the muon reconstruction is studied using a tag-and-probe method in $Z \rightarrow \mu^+ \mu^-$ and $J/\psi \rightarrow \mu^+ \mu^-$ events [162]. Figure 3.3 shows the reconstruction efficiency for low p_T muons, measured with $J/\psi \rightarrow \mu^+ \mu^-$ events.

3.2.4 Electrons and photons

The typical signature of electrons and photons in the ATLAS detector is an electromagnetic shower in the electromagnetic calorimeter. The selection procedure for these objects is described in [163, 164], in particular electron candidates are reconstructed as clusters of energy deposit in the calorimeter matched to a track in the ID, while photons are not associated to an ID track. However, it is possible that a photon produces a $e^+ e^-$ pair before reaching the calorimeter, in this case it is called a converted photon; it is possible to identify also this process, based on the associated ID hits [165]:

- **electron**, characterised by a cluster in the calorimeter with a single associated reconstructed ID track with an associated vertex;
- **unconverted photon**, characterised by a cluster in the calorimeter without an associated ID track;
- **converted photon**, the calorimeter cluster is matched to two opposite-signed tracks (consistent with electrons), associated to a secondary vertex and collinear near it. The detection of conversion track pairs becomes inefficient at large conversion radius, therefore a

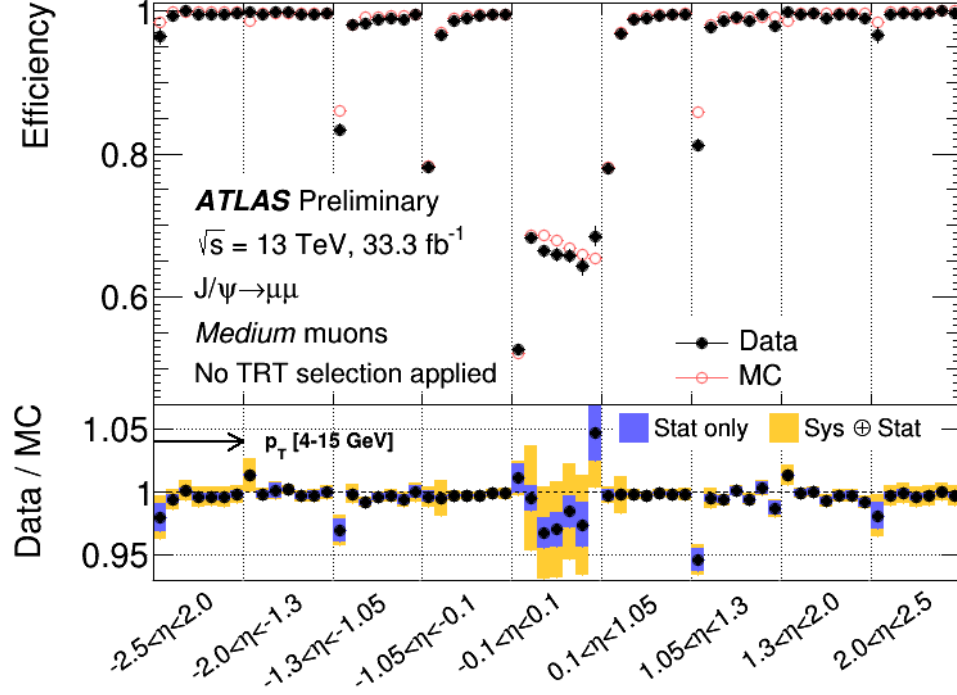


Figure 3.3: Muon reconstruction efficiency in different η regions measured in $J/\psi \rightarrow \mu^+\mu^-$ events for Medium muon selection. Within each η region, the efficiency is measured in seven p_T bins (4-5, 5-6, 6-7, 7-8, 8-10, 10-12, and 12-15 GeV). The resulting values are plotted as distinct measurements in each η bin with p_T increasing from 4 to 15 GeV going from left to right. The error bars on the efficiencies indicate the statistical uncertainty. The panel at the bottom shows the ratio of the measured to predicted efficiencies, with statistical and systematic uncertainties. Figure obtained from [162].

cluster matched to a single track with missing hits in the IBL is also considered a converted photon.

Similarly to muons, different WPs and additional requirements can be defined, with varying signal efficiencies and fake rejections. The calibration and efficiency measurement are performed using $Z \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$ events [163, 164, 166].

A detailed discussion on electrons and photons reconstruction and calibration can be found in Refs. [163–166].

3.2.5 Jets

Due to the confinement properties of QCD, high energy quarks and gluons produced in pp collisions by deep inelastic scattering generate collimated sprays of particles called jets [167]. The typical signature of a jet are showers in the calorimeters with associated tracks in the ID.

Jets are reconstructed by clustering energy deposits in the calorimeters [168]; several jet-finding algorithms have been developed, in particular, the ATLAS approach uses the anti- k_t algorithm [169]. A detailed discussion on jet reconstruction and calibration can be found in [168].

3.2.6 Taus

Tau leptons have a lifetime of $\mathcal{O}(10^{-13})$ [11], which corresponds to a decay length much smaller than the detector resolution. For this reason, no Secondary Vertex (SV) can be identified, consequently they have to be reconstructed via their decay products.

It is hard to separate the product of the leptonic decay channels of the tau from other leptons originated in the event, therefore taus are usually reconstructed using the hadronic decay channels. High momentum taus produce jets when they decay hadronically, that can be distinguished from quark and gluon originated jets because they are narrower and present a lower track multiplicity. The tau identification is enhanced using a BDT multivariate classifier, specifically designed for this purpose.

A detailed description of tau identification and reconstruction can be found in [170, 171].

3.2.7 Missing transverse energy E_T^{miss}

The only standard model particles that can travel through the detector without interacting are neutrinos, and whenever they are produced in a collision they create an imbalance in the total transverse momentum measurement. In addition, many BSM theories predict the existence of particles that do not interact with the detector, producing a similar imbalance.

The momentum imbalance in the transverse plane is measured with a two dimensional vector called missing transverse momentum vector $\mathbf{p}_T^{\text{miss}}$ [172], defined as the negative sum of the transverse component of the four-momenta of all the visible objects in the event. The magnitude of $\mathbf{p}_T^{\text{miss}}$ is defined as missing transverse energy E_T^{miss} .

A detailed discussion on the reconstruction and calibration of these objects can be found in [172].

STUDIES FOR TRIGGER IMPROVEMENTS

4

Most the ATLAS flavour physics analyses are based, as of today, on muon triggers. Due to the high rate of low momentum hadrons, electrons and photons events, the typical low momentum flavour physics events not based on muons can not be properly triggered at L1. The addition of the Fast Tracker (FTK) [112], introduced in section 2.2.6 and discussed in the following section, to the ATLAS HLT will allow the reconstruction of the tracks in the ID granting access all the pp collisions in each event in Run 3. As introduced in section 2.2.6, the FTK is designed to perform regional and global track reconstruction, with an acceptance on the tracks of $p_T^{\text{tracks}} > 1$ GeV and $|\eta^{\text{tracks}}| < 2.5$. The regional reconstruction is seeded by the L1 RoIs and only tracks in these region are reconstructed. The global reconstruction, called *full-scan*, allows the reconstruction of the full ID and is the FTK feature on which these studies are based.

The aim of this section is to explore the use of FTK to enhance the ATLAS B-physics capabilities. The main possibility explored is that of exploiting the FTK to trigger on pile-up collisions, parasitically with respect to the L1 selection.

The FTK features relevant to this study are described first (section 4.1) followed by studies relevant to the main analysis topic of this thesis (section 4.2). After a description of the samples employed (section 4.3), the trigger selection will be optimised (section 4.4) and the corresponding trigger rates will be discussed (section 4.5).

4.1 The FTK

The FTK [112] is a custom electronics system designed to rapidly find and fit tracks in the ATLAS ID for events passing the L1 trigger [173]. The FTK receives data from the full pixel and SCT detectors and outputs a collection of tracks to the HLT.

In order to cope with the high rate, the system is highly parallelised and the detector is divided into 64 $\eta - \phi$ regions, called towers, which are processed in parallel. The track fitting procedure is divided into two steps. The first step (pattern recognition) identifies track candidates comparing coarse-resolution information from the ID to predefined trajectories stored in memory (patterns). The second step (track fitting) determines the charged track trajectory parameters

from the combinations of ID hits identified by matched patterns. This procedure relieves the CPU load that track reconstruction would require in the HLT.

Figure 4.1 shows a complete sketch of the FTK system. Data from the ID are transmitted from

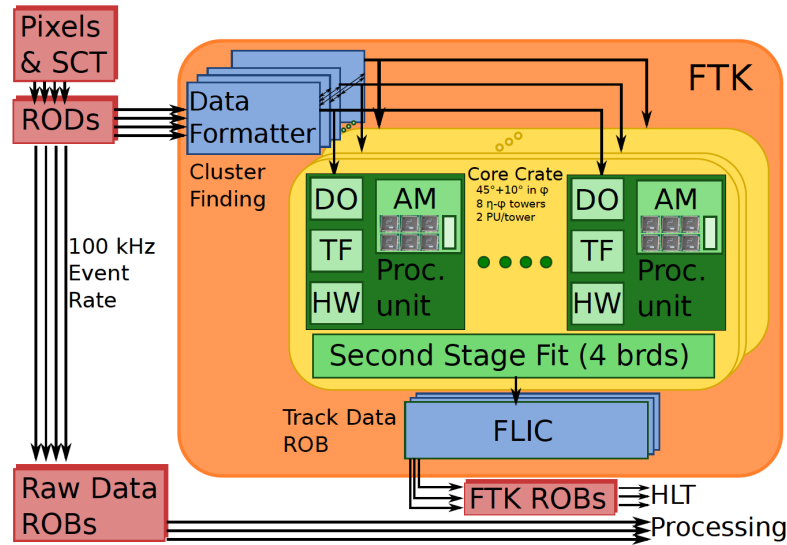


Figure 4.1: Functional sketch of FTK. AM is the Associative Memory, DO is the Data Organizer, FLIC is the FTK-to-Level-2 Interface Crate, HW is the Hit Warrior, ROB is the ATLAS Read Out Buffer, ROD is a silicon detector Read Out Driver, and TF is the Track Fitter. Picture from [112].

the detector ReadOut Drivers (RODs) to the Data Formatters, which identify the clusters corresponding to hits in the ID and distribute them to the Data Organisers (DOs). The DOs feed hits to the pattern recognition ASICS (“Associative Memory chips” or AM) which identify matching patterns (“roads”). The DOs align roads with the corresponding hits and feed them to the Track Fitters (TFs) to perform a first evaluation of the track parameters. The calculation of the track parameters, based on information from 8 silicon layers of the ID, out of 12 available, is performed linearly based on pre-calculated constants using the following formula [112]:

$$p_i = \sum_j c_{ij} x_j + q_j, \quad \forall i = 1 \dots 5. \quad (4.1)$$

The reconstructed tracks are passed to the Hit Warrior (HW), which performs duplicate track removal and discards tracks with a low fit quality.

When a track passes the HW selection, it is sent to the Second Stage Boards (SSBs), which carry out a more accurate evaluation of the track parameters, using all the information from the 12 ID silicon layers. Finally, the tracks are sent to the FTK-to-Level2 Interface Crate (FLIC), which collects and unifies information from the different towers and sends it to the Read-Out Buffer (ROB), ready to be used by the HLT.

4.2 Case of study

In order to understand if the FTK full-scan tracks can be effectively used to trigger on pile-up events, four case studies have been identified:

- $B \rightarrow hh'$, where B can be a B_d^0 or a B_s^0 meson and h and h' can be either a pion or a kaon;
- $B^+ \rightarrow \bar{D}^0(K^+\pi^-)\pi^+$;
- $B_s \rightarrow \phi\phi \rightarrow K^+K^-K^+K^-$;
- $B_d \rightarrow D^{*-}(D^0(K^+\pi^-)\pi^-)\pi^+$,

where the notation refers also to the charge conjugates of these processes. These four processes are characterised by different topologies and final states, allowing coverage of most of the use cases an experiment like ATLAS may want to focus into. The $B \rightarrow hh'$ process presents the simplest topology, with no intermediate decays and only two particles in the final state. This process represent the benchmark for studies on the other channels. The $B^+ \rightarrow \bar{D}^0(K^+\pi^-)\pi^+$ with three particles in the final state and an intermediate decay presents a more complex topology. The $B_s \rightarrow \phi\phi \rightarrow K^+K^-K^+K^-$ process presents an ever more complex topology, but has two equal intermediate decays, which can simplify its detection. Finally, $B_d \rightarrow D^{*-}(D^0(K^+\pi^-)\pi^-)\pi^+$ presents a four particles final state as well, with a series of successive decays. The pion from the $D^{*-} \rightarrow D^0\pi^-$ decay presents an extremely low momentum, which, most of the times, is lower than the FTK acceptance. This process can allow the study of processes in which not all the particles in the final state can be reconstructed.

The study presented in this thesis focuses on the $B \rightarrow hh'$ process. This two-body decay can allow the measurement of CP violation effects and presents the simplest topology among the above modes and is fundamental for improvements in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis. Section 8.1.2 will show how $B \rightarrow hh'$ events with the two hadrons mis-identified as muons are a particularly problematic background, as they are superimposed on the B_d^0 mass peak and present almost identical features to the signal. The main approach currently followed to study this background heavily relies on MC simulations. The possibility to study $B \rightarrow hh'$ decays with a dedicated trigger would allow direct data-driven studies on the fake-muons probability, resulting in a improved estimation of its normalisation.

4.3 Samples employed

The design of a trigger selection for the processes listed above requires two samples for the choice and optimisation of the cuts: a signal sample, to assess the signal efficiency, and a background sample, to evaluate the rejection power and predict trigger rates. Both samples require the emulation of the FTK response in order to evaluate the effect of the proposed selection.

4.3.1 Signal sample

The signal samples are based on MC simulations performed using Pythia8 [148]. Some of the MC samples required are already available in ATLAS, e.g. the $B \rightarrow hh'$ sample employed in the $B_{(s)}^0 \rightarrow \mu^+\mu^-$ analysis (see chapter 7 for more details) but they do not include the FTK emulation and are not tuned on the FTK properties. In fact, such samples are designed on

the properties of the current ATLAS analyses, for instance, the $B \rightarrow hh'$ sample is tuned on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis acceptance, requiring the transverse momentum of the tracks to be higher than 4 GeV. Since the acceptance of the FTK is $p_T^{\text{tracks}} > 1$ GeV and $|\eta^{\text{tracks}}| < 2.5$, the MC samples need to be re-tuned, in order to avoid possible kinematic biases in the selection while maintaining good generation times.

The MC tuning is based on several parameters, for the processes listed in section 4.2, only three are mainly responsible for possible kinematic biases. These parameters are:

- $\hat{p}_T$: lower cut applied to the p_T of the partons produced in the hard scattering in the reference system of the incoming partons;
- $\bar{b}$ quark p_T : lower cut on the p_T of the $\bar{b}$ quark in the pp collision centre of mass energy frame;
- final state particles p_T : lower cut on the p_T of the final products of the decay.

Notably only the $\bar{b}$ quark involved in the $b\bar{b}$ production is subjected to cuts. This is done to avoid the introduction of biases due to the application of cuts on both quarks. In order to optimise the generation, only the B meson is forced to decay into the wanted process, while the $\bar{B}$ meson is not constrained.

The parameters tuning is simplified by maintaining $\hat{p}_T$ and $\bar{b}$ quark p_T to the same value; this choice can, in principle, introduce a kinematic bias. As shown later in this section, only the extremely low p_T region of the B meson spectrum is found to be biased, therefore this assumption does not undermine the validity of these studies.

The tuning of the MC parameters is performed using a set of test samples generated ad-hoc varying the quark cuts, which include $\hat{p}_T$ and $\bar{b}$ quark p_T , and the final state particles p_T cut. The production of such samples is limited to generation, without simulation of the detector response and reconstruction. The kinematic distributions of the particles of interest in the decay are then compared to the same quantities from an unbiased MC sample, generated with extremely loose cuts. For the $B \rightarrow hh'$ MC the unbiased sample is generated using quark cuts = 1 GeV and final state particles p_T cut = 0.1 GeV. The test samples are generated with quark cuts between 1 and 5 GeV with steps of 1 GeV, and final state particles p_T cut of 0.5 and 0.8 GeV.

The B meson and final state particles p_T distributions are compared to the ones from the unbiased sample using the Kolmogorov–Smirnov (KS) test [174]. Since a possible bias is likely to manifest at low p_T , the test is performed on sub-samples of the distributions, considering lower limits on the transverse momentum. The first quantity checked is the p_T of the B meson. Figure 4.2 shows the result of such a test. Each line of the plot represents a different test sample and each column a different lower p_T cut; the height of each bin is the outcome of the KS test. The B meson p_T distribution is shown to be compatible with the unbiased distribution for $p_T^B > 3$ GeV at low quark cut values, and progressively less compatible with increasing values of the generation cuts.

The same check is performed on the transverse momentum of the final state particles. Figure 4.3 shows the result of the various tests performed on the final state particles from the

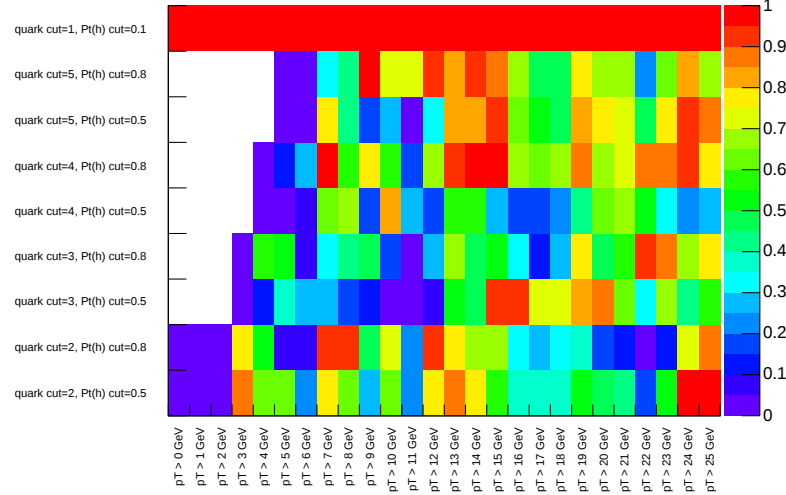


Figure 4.2: Result of the KS test [174] applied to the B meson p_T distribution. Each line of the plot shows a different test sample and each column a different lower p_T cut; the height of each bin is the outcome of the KS test. The white bins indicate an outcome of the KS test = 0.

various $B \rightarrow hh'$ sub-processes. Considering the four plots, the sets of cuts which allow one to avoid any kinematic bias have quark cuts at 2 GeV. Both thresholds considered on the final state particles p_T do not seem to introduce any bias, therefore the higher threshold, 0.8 GeV, is chosen.

4.3.2 Background sample

The ATLAS experiment collects regularly particular datasets, named *enhanced bias samples* [175]. These datasets are recorded at various instantaneous luminosity values and are employed to perform a reliable extrapolation of the output rate of the entire HLT trigger menu (section 2.2.6) and to test the output rate of new trigger selections. Additionally, enhanced bias samples are optimal datasets to be used as background samples for the optimisation of a trigger selection that runs parasitically with respect to the L1 selection. This is because the collection of such samples is based on a variety of L1 triggers of all signatures types, combinations and p_T ranges, designed to contain events which are more likely to be selected by the trigger. Events recorded in enhanced bias samples are only biased by the L1 system, no additional selection is applied at the HLT except for the application of prescales, to control the output rates. This allows to have a sample that emulates the expected background conditions for the $B \rightarrow hh'$ trigger selection being considered.

Few enhanced bias samples from the Run 2 data taking period, reprocessed to include the FTK simulation, were available during the development of the studies described in this chapter. Only one of them will be employed for the optimisation of the trigger selection, while all the available samples will be employed to estimate the rate of events accepted by such selection. The output rate of a trigger can be estimated thanks to the particular trigger menu employed during the data taking of an enhanced bias sample. This trigger menu is, in fact, designed to be

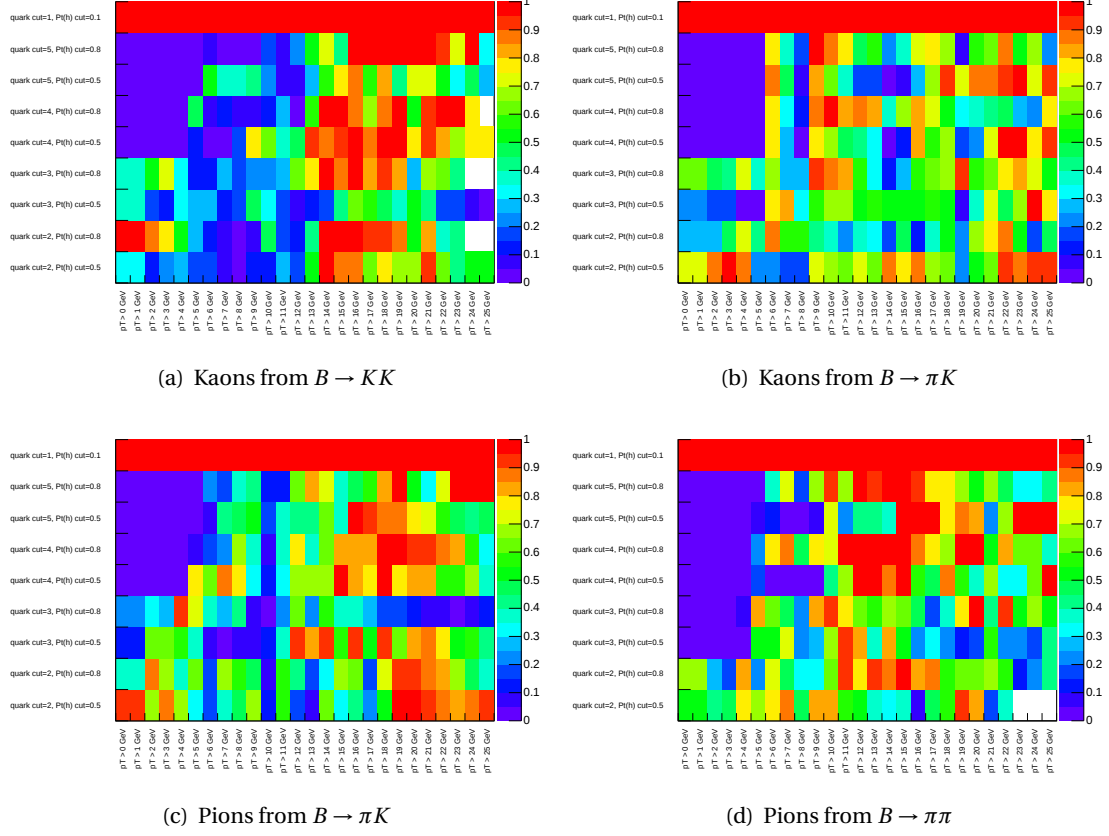


Figure 4.3: Result of the KS test [174] applied to the final state particle of the $B \rightarrow hh'$ process for the different sets of truth-level cuts considered. Each line of the plot shows a different test sample and each column a different lower p_T cut; the height of each bin is the outcome of the KS test. The white bins indicate that the KS test could not be performed, due to the low statistics available in the samples.

“invertible”, such that a single weight (enhanced bias weight) is calculable per event, in order to correct for the prescales applied during the enhanced bias data taking and restore an effective zero bias spectrum [175] (see section 4.5). Once the data taking is concluded, the enhanced bias samples are reprocessed offline including the full HLT trigger menu without any prescale, in order to allow testing of triggers not considered during the data taking.

4.4 Trigger selection optimisation

The main idea guiding the $B \rightarrow hh'$ trigger selection is to employ a fast selection to reduce the huge amount of combinatorial background present in each event.

The design of the trigger selection follows a cut-based approach, which is preferred to a Multivariate Analysis (MVA) based approach, as it is simpler to interpret and is more similar to the preselection of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. The purpose of this trigger selection is, in fact, mostly devoted at fake-rate studies for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis (discussed in section 7.4), therefore the selection variables are derived from the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during Run 1 of the LHC operations [73].

The same mass hypothesis of 140 GeV is used for all tracks; this assumption allows avoidance

of having to run multiple times on the same candidate while varying the mass hypothesis and was proven not to modify the efficiency of the selection. In addition, tracks are approximated as straight lines, described by the five track parameters introduced in section 3.2.1. A vertexing tool could in principle be exploited to precisely identify the position and the properties of the B decay vertex (Secondary Vertex (SV)), however, this would increase the computational burden of the trigger selection. The determination of the position and properties of the SV is therefore performed using a simpler approach: the four-vector of the B candidate is evaluated by summing the four-vectors of the two tracks, while the position of the SV is identified as the middle point of the segment joining the points of closest approach of the two tracks.

The calculation of the position of the event PV is another computationally heavy procedure. Since however this position is extremely useful to select B-physics events from the prompt background, a more expedited method to approximate the PV position is employed. The point of closest approach on the beamline to the B candidate momentum backward projection is assumed as position of the PV. The definition of PV used in the rest of this chapter refers therefore to this modified approach, unless explicitly mentioned.

The order in which the selection cuts are applied can further improve the computational time needed for the trigger selection. The first selection is performed on the single FTK tracks, in order to reduce the possible combinations of tracks employed in the second selection, which is applied to the B candidates.

While the selections of the analyses are usually optimised on a candidate basis, the optimisation for this trigger selection is based on an event basis. This is because the aim of these studies is to minimise the trigger rate. The approach employed for the optimisation is the iterative N-1 plots method. This procedure starts with a loose selection for a set of discriminating variables. One variable at the time is considered and all the cuts on the other discriminating variables are applied to both the signal and background samples; the value of a figure of merit as a function of the variable being studied is then analysed: the value of the variable for which the figure of merit presents the best score is used in the selection for the following iteration. Once the values of the various selections converge, the optimisation is considered finished.

The figure of merit exploited is $S/\sqrt{S+B}$, where S and B refer to the number of signal or background events in which at least one $B \rightarrow hh'$ candidate passes the selection. In particular, the number of signal events S corresponds to the number of signal candidates, as only reconstructed candidates matched to the MC truth-level information are employed.

The selection optimised specifically for the $B \rightarrow hh'$ process is shown in table 4.4, together with an explanation of the quantities involved. The order employed to show the variables reflects the order in which the selection cuts are applied. The plots in the third column show the distribution of the discriminating variables employed; they are obtained by applying the selections not included in the optimisation (requirements on the p_T of the tracks and on the p_T and invariant mass of the B candidates) and without applying any requirements on the remaining variables. The signal efficiency, evaluated as the number of signal events from the MC sample

which pass the selection divided by the number of events containing signal candidates reconstructed by the FTK, is about 3%. The background rejection, evaluated as the number of events from the enhanced bias sample in which at least one candidate passes the selection divided by the total number of events, is 99.997%.

Table 4.1: Trigger selection for the $B \rightarrow hh'$ process. The first column provides the name of the variable and the selection applied. the second column provides a description of the variable and the third column shows a normalised plot of the quantity for the truth-matched MC $B \rightarrow hh'$ candidates and the background candidates from enhanced bias samples. The Primary Vertex (PV) and Secondary Vertex (SV) are evaluated as explained in the text. The order in which the cuts are listed reflects the order in which they are applied in the selection.

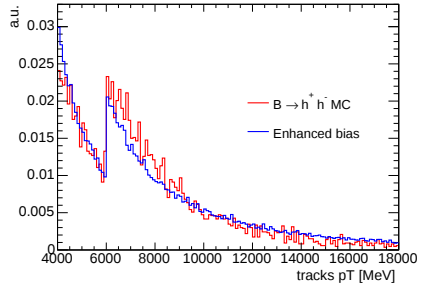
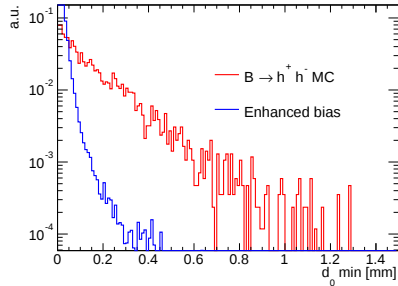
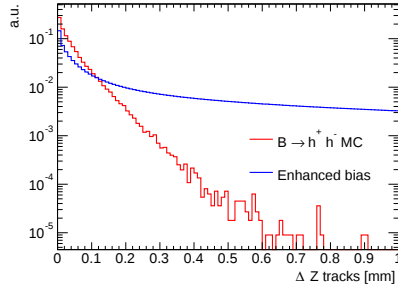
Cut	Description	
tracks $p_T > 6, 4 \text{ GeV}$	Transverse momentum cut on each of the two tracks. The two thresholds refer to the leading and sub-leading tracks respectively. This cut is not included in the optimisation, but is set to the p_T thresholds employed in the 2015-16 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.	
$ d_0 ^{\min} > 0.18$	Smaller absolute value of the impact parameter to the beam-line.	
total charge = 0	Charge of the B candidate, obtained as sum of the charge of the two tracks. No plot is shown for this quantity.	
$\Delta Z < 0.16 \text{ mm}$	Distance between the two tracks projected on the z axis.	

Table 4.1: Trigger selection for the $B \rightarrow hh'$ process. The first column provides the name of the variable and the selection applied. the second column provides a description of the variable and the third column shows a normalised plot of the quantity for the truth-matched MC $B \rightarrow hh'$ candidates and the background candidates from enhanced bias samples. The Primary Vertex (PV) and Secondary Vertex (SV) are evaluated as explained in the text. The order in which the cuts are listed reflects the order in which they are applied in the selection.

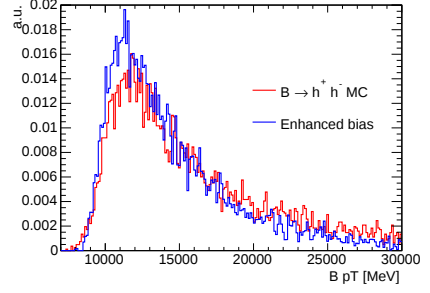
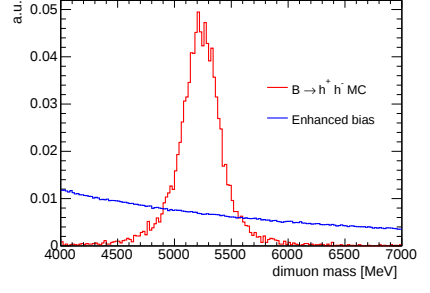
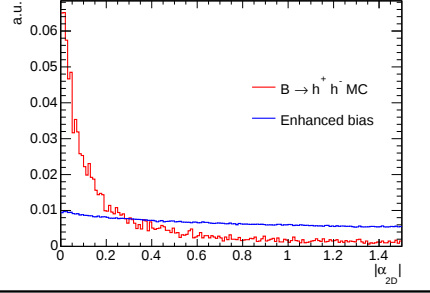
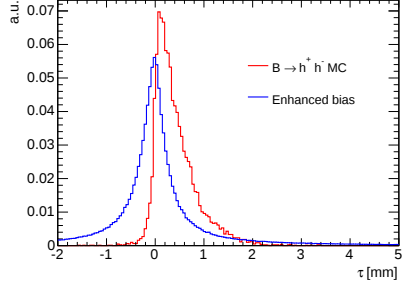
Cut	Description	
$p_T^B > 7 \text{ GeV}$	Transverse momentum of the B candidate. This cut is not included in the optimisation, as it does not improve significantly the selection.	
$4 \text{ GeV} < M^B < 7 \text{ GeV}$	Mass of the B candidate. This cut is not included in the optimisation. The interval is centred on the $B_{(s)}^0$ masses and is large enough to allow the identification of mass sidebands to perform a fit on the M^B distribution.	
$ \alpha_{2D} < 0.09$	Absolute value of the angle in the transverse plane between $\vec{p}_T^B$ and the line connecting the PV and the SV. This quantity is often referred to as “pointing angle”.	

Table 4.1: Trigger selection for the $B \rightarrow hh'$ process. The first column provides the name of the variable and the selection applied, the second column provides a description of the variable and the third column shows a normalised plot of the quantity for the truth-matched MC $B \rightarrow hh'$ candidates and the background candidates from enhanced bias samples. The Primary Vertex (PV) and Secondary Vertex (SV) are evaluated as explained in the text. The order in which the cuts are listed reflects the order in which they are applied in the selection.

Cut	Description	
$\tau > 0.3 \text{ mm}$	Proper decay time of the B candidate. This quantity is evaluated as $L_{xy} \cdot M^B / p_T^B$, where L_{xy} is the projection of the line connecting the PV and the SV along $\vec{p}_T^B$. M^B and p_T^B are the mass and the transverse momentum of the B candidate. Although this quantity should be measured in seconds, the same unit of measurement as L_{xy} is maintained.	

4.5 Trigger rate estimation

The estimation of the trigger rate for the $B \rightarrow hh'$ trigger selection is performed using different enhanced bias samples, characterised by different PU and instantaneous luminosity conditions. These estimations are valid under the assumption that most of the events that trigger the chain are background events.

The HLT output rate of a trigger chain and its error (respectively R and R_{err}) are evaluated according to the following formulas [175]:

$$R = \frac{\sum_{e=1}^N w(e)}{\Delta t}, \quad R_{\text{err}} = \frac{\sqrt{\sum_{e=1}^N w(e)^2}}{\Delta t}. \quad (4.2)$$

Where the sum runs over all $e = 1, 2, \dots, N$ events in the enhanced bias sample and Δt is the time period over which the sample was collected (typically about one hour). The weight $w(e)$ represents the effective number of events passed by the trigger chain. The rate obtained with this formula corresponds to the average instantaneous luminosity over the time period during which the enhanced bias data sample was collected.

The weights $w(e)$ are, in turn, evaluated according to the following formula [175]:

$$w(e) = w_{\text{EB}}(e) \cdot w_{\text{C}}(e) \cdot w_{\mathcal{L}}(e). \quad (4.3)$$

The $w_{\text{EB}}(e)$ term is the enhanced bias weight, introduced in sub-section 4.3.2, used as per-event weight to correct for the prescales employed during the data taking of the enhanced bias sample and restore an effective zero bias spectrum. This weight is always equal to or higher than 1 and is evaluated according to [175]:

$$\frac{1}{w_{\text{EB}}(e)} = 1 - \prod_{j=1}^{\text{EB chains}} \left(1 - \frac{r_{je}}{p_j} \right) \quad (4.4)$$

Where the product runs over the trigger chains used to collect the dataset, r_{je} is the trigger decision before the application of any prescale (0 not passed, 1 passed) and p_j is the total prescale for the chain.

The $w_{\mathcal{L}}(e)$ term is called luminosity extrapolation weight ($w_{\mathcal{L}}(e) > 0$). This quantity is used to extrapolate the estimated rate for a trigger chain or a combination of trigger chains to a different luminosity. There are several ways to evaluate this term, based on the properties of the physics process of interest and of the trigger chain. The trigger rate estimations performed in this thesis assume $w_{\mathcal{L}}(e) = 1$, meaning no luminosity extrapolation is performed. More information on the luminosity extrapolation weights is available in Ref. [175].

The $w_{\text{C}}(e)$ term is the chain/combination weight. This quantity accounts for prescales on the trigger chains involved in the rate estimation and its value is bounded $0 < w_{\text{C}}(e) \leq 1$. The formula used to calculate the chain/combination weights depends on the trigger selection being analysed, and can be evaluated for individual chains, for the union of multiple chains and for the intersection of chains. The $B \rightarrow hh'$ trigger selection is designed to be seeded by all the fired L1 triggers, for this case the formula is [175]:

$$w_{\text{C}}(e) = 1 - \prod_{i=1}^{N^{\text{L1}}} \left(1 - \frac{r_{ie}^{\text{L1}}}{p_i^{\text{L1}}} \right). \quad (4.5)$$

The product runs over all the L1 triggers which seed the HLT selection, r_{ie}^{L1} is the L1 trigger decision before the application of any prescale (0 not passed, 1 passed) and p_i^{L1} is the L1 prescale for the chain. This expression does not take into account possible prescales applied to the HLT selection of interest, which would result in an overall $1/p^{\text{HLT}}$ term.

Ideally, for the $B \rightarrow hh'$ selection, the product in equation 4.5 would run on all the possible L1 triggers. Unfortunately, the presence of a trigger and its prescale are not constant in the trigger menu, but depend on several factors, such as the instantaneous luminosity. For this reason, an estimation of the HLT trigger rate for the $B \rightarrow hh'$ selection seeded by every possible L1 trigger combination is not possible. The estimation performed in this thesis is based on the triggers called *lowest unprescaled triggers*. These are the unprescaled triggers characterised by the lowest thresholds. Almost all the events contained in the enhanced bias sample pass at least one of these triggers, therefore the estimation of the trigger rate takes into account all the events in the sample.

The trigger rate estimation is based on three enhanced bias samples, the only datasets where the FTK emulation was available at the time of these studies. The result of the trigger rate estimation is shown in figure 4.4. The x axis represents the average instantaneous luminosity of

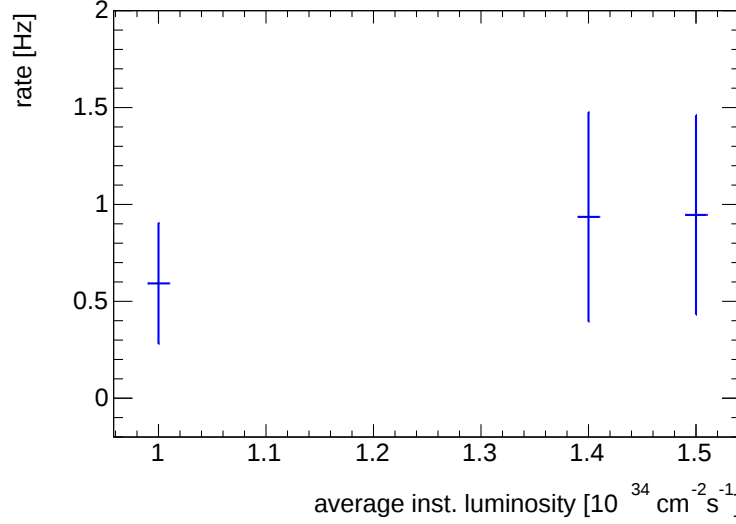


Figure 4.4: Rate estimation for the $B \rightarrow hh'$ trigger selection performed using the procedure explained in the text. The three points show the estimation based on different enhanced bias samples, characterised by different instantaneous luminosities.

the enhanced bias samples, while the y axis is the rate estimation. The estimated HLT output trigger rates range between 0.5 and 1 Hz, with uncertainties of the order of 0.5 Hz. Such rates are optimal, as they are low enough not to occupy too much bandwidth.

An estimation of the dependence of the rate on the instantaneous luminosity is not possible, due to the low number of available enhanced bias samples and the large uncertainty on the estimations.

4.6 $B \rightarrow hh'$ signal yield estimation

The expected signal yield for the $B \rightarrow hh'$ trigger selection is evaluated based on the properties of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during 2015 and 2016 [4], extensively described in the following chapters. Assuming the same efficiency of the trigger selection for all the $B \rightarrow hh'$ processes, only the expected $B_d^0 \rightarrow \pi^+ \pi^-$ yield is calculated. This value is then rescaled according to the total branching fraction of $B \rightarrow hh'$ processes. The other processes are used to validate the estimation.

The values employed in the extraction of the branching fractions for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ (chapter 5) allow the estimation of the number of expected $B_d^0 \rightarrow \pi^+ \pi^-$ decays in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis truth-level acceptance volume $p_T^\mu > 6,4 \text{ GeV}$, $|\eta_\mu| < 2.5$, $p_T^B > 8.0 \text{ GeV}$ and $|\eta_B| < 2.5$, where the cuts are applied to the hadrons instead of the muons. This number, multiplied by the efficiency of the $B \rightarrow hh'$ trigger selection calculated with respect to this acceptance volume gives the number of $B_d^0 \rightarrow \pi^+ \pi^-$ events that would pass the $B \rightarrow hh'$ trigger selection in the 2015/2016 dataset. Considering that the FTK only reconstructs events accepted by the L1 trigger, a factor of $\text{L1}_{\text{rate}}/\text{collision}_{\text{rate}}$ has to also be considered. Equation 4.6 summarises the formula employed for the estimation of expected $B_d^0 \rightarrow \pi^+ \pi^-$ yield. An additional $\mathcal{L}/\mathcal{L}_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}$ factor, where $\mathcal{L}_{B_{(s)}^0 \rightarrow \mu^+ \mu^-} = 36.1 \text{ fb}^{-1}$ is the integrated luminosity of the dataset employed for

the 2015-16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, is also shown and used to rescale the expected yield to a wanted integrated luminosity $\mathcal{L}$.

$$N_{B_d^0 \rightarrow \pi^+ \pi^-}(\mathcal{L}) = \frac{\epsilon_{B_d^0 \rightarrow \pi^+ \pi^-} \cdot \mathcal{B}(B_d^0 \rightarrow \pi^+ \pi^-) \cdot N_{B_s^0 \rightarrow \mu^+ \mu^-}}{\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-} \cdot \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) \cdot f_s / f_u} \cdot \frac{\text{L1}_{\text{rate}}}{\text{collision}_{\text{rate}}} \cdot \frac{\mathcal{L}}{\mathcal{L}_{B_{(s)}^0 \rightarrow \mu^+ \mu^-}}. \quad (4.6)$$

The $\epsilon_{B_d^0 \rightarrow \pi^+ \pi^-}$ term is the efficiency of the trigger selection for $B_d^0 \rightarrow \pi^+ \pi^-$ events with respect to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis acceptance volume, $\mathcal{B}(B_d^0 \rightarrow \pi^+ \pi^-)$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ are the SM branching fractions for the two processes, obtained from Ref. [11] and $\frac{f_s}{f_d} = 0.256 \pm 0.013$ is the b quark hadronisation probability ratio for B_s^0 to B_d^0 mesons, obtained from the latest HFLAV average [28]. The two terms $\epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}$ and $N_{B_s^0 \rightarrow \mu^+ \mu^-}$ are obtained from the 2015-16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [4], which is the core part of this thesis and will be presented later starting from chapter 5, and are respectively the efficiency for the $B_s^0 \rightarrow \mu^+ \mu^-$ channel with respect to the analysis acceptance volume and the expected number of $B_s^0 \rightarrow \mu^+ \mu^-$ events according to the SM prediction.

The resulting $B \rightarrow hh'$ yield, obtained by rescaling $N_{B_d^0 \rightarrow \pi^+ \pi^-}(\mathcal{L})$ according to the total branching fraction for the $B \rightarrow hh'$ process [11] and for the total expected integrated luminosity of Run 3 (300 fb⁻¹) is about 17000 events. The same procedure performed considering the other $B \rightarrow hh'$ processes yields a comparable expected number of events.

4.7 Results and conclusions

The optimisation of a trigger selection that runs parasitically with respect to the L1 selection for the $B \rightarrow hh'$ process has been performed and the expected HLT output rate for this trigger selection and the expected amount of $B \rightarrow hh'$ events to be collected in Run 3 have also been evaluated.

The main result obtained in this chapter is that employing a series of selections, which appear to be affordable from a computational point of view, it is possible to maintain a low HLT output rate and collect a good amount of signal statistics. This represents a proof of principle: pile-up events can be exploited at the HLT level to collect fully hadronic processes, exploiting the FTK full-scan reconstruction.

Regarding the particular trigger selection studied, the estimated $B \rightarrow hh'$ yield for the total expected integrated luminosity of Run 3 is about 17000 events. Unfortunately, this yield does not allow any accurate study. In fact, considering the current sensitivity of CP violation studies on $B \rightarrow hh'$ samples [11], together with the limited mass resolution of the ATLAS experiment and the impossibility of performing particle identification, the resulting studies would not be significant, compared to the current state of the art. Regarding the studies on the muon-fake rate for the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, the average fake-muon rate for hadrons (studied in detail on both simulations and data in sections 7.4 and 11.2.2.2) is found to be $\sim 0.09\%$, implying that only about 32 $B \rightarrow hh'$ events would contain at least one fake muon. Such low fake-muons yield would not allow to properly study the muon mis-identification fraction.

Two main possibilities are being studied to improve the selection in order to be able to employ

the $B \rightarrow hh'$ sample for muon-fake rate studies. Relaxing some of the selections described in section 4.4 would allow the collection of more signal events, at the cost of a higher trigger rate due to background events. The other possibility is to require that at least one of the two tracks associated to a $B \rightarrow hh'$ candidate is matched to a track in the MS. The study of the second track of the process can allow the estimation of the fake rate. The MS matching requirement would strongly reduce both the output rate of the trigger selection and its CPU load, implying the possibility of relaxing most of the selection cuts.

The selection for the remaining processes considered in section 4.2 can be based on the one designed in this chapter, which was proven to highly reduce the effect of the combinatorial background. The $B \rightarrow hh'$ mode, in fact, presents the simplest topology among the processes considered; the additional requirements deriving from a more complex topology will therefore allow a more effective selection.

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ANALYSIS STRATEGY

5

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the data collected by the ATLAS experiment during the first two years of the LHC Run 2 pp collision data-taking period constitutes the core of this thesis. This chapter provides an introduction to the analysis strategy, followed by a description of the main steps of its development. An overview of the content of the following chapters, in which the analysis is discussed in detail, is also provided. Finally, the expected sensitivity is discussed, in order to motivate the publication of results based on partial Run 2 data.

The aim of the analysis is to measure the branching fractions of the two extremely rare decays $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$, which are relevant for indirect searches of BSM physics, as introduced in chapter 1.

The analysis strategy is mostly derived from the previous ATLAS measurement based on the dataset collected during the Run 1 of the LHC [73], described in appendix A: the $B_s^0 \rightarrow \mu^+ \mu^-$ ($B_d^0 \rightarrow \mu^+ \mu^-$) branching fractions are extracted relative to a *reference* decay mode $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$, which is abundant and has a well-measured branching fraction $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$. This procedure allows the cancellation of several systematic uncertainties due to sources affecting both the signal and the reference channels. Additionally, the $B_s^0 \rightarrow J/\psi \phi$ process, with $J/\psi \rightarrow \mu^+ \mu^-$ and $\phi \rightarrow K^+ K^-$, is employed in the analysis as a *control channel* for the signal kinematic variables. The reference and control channels are complementary: while $B^+ \rightarrow J/\psi K^+$ presents a topology more similar to the signal, as it is a 3-prong process, the $B_s^0 \rightarrow J/\psi \phi$ process allows access to a B_s^0 sample. This grants the possibility to check for possible kinematic differences between the signal and the reference channels due to hadronisation effects.

The notation used throughout this thesis refers also to the charge conjugates unless otherwise specified, therefore $B^+ \rightarrow J/\psi K^+$ includes also $B^- \rightarrow J/\psi K^-$. Additionally, in order to avoid mentioning the two signal processes, e.g. $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ notation is used to refer to both.

The branching fraction calculation is performed according to the following formula:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{f_u}{f_{s(d)}} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \times \frac{N_{B_d^0(B_s^0)}}{N_{J/\psi K^+}} \times R_{A\epsilon}, \quad (5.1)$$

where:

- $\frac{f_u}{f_{s(d)}}$ is the ratio of the hadronisation probabilities of a b-quark into B^+ and B_s^0 (B_d^0), this quantity is not measured in the analysis and its value is taken from the latest available HFLAV average [28]: $\frac{f_s}{f_d} = 0.256 \pm 0.013$, which assumes $\frac{f_u}{f_d} = 1$. The value of this quantity may present a dependency on the b quark kinematics, therefore the HFLAV value might need an additional correction. This effect can be studied by analysing the transverse momentum and η dependence of the $B_s^0 \rightarrow J/\psi \phi$ and $B_d^0 \rightarrow K^{*0} J/\psi$ yield ratio. This technique has been employed by the ATLAS experiment to measure $\frac{f_u}{f_s}$ at a centre of mass energy of $\sqrt{s} = 7$ TeV during Run 1, obtaining $\frac{f_s}{f_d} = 0.240 \pm 0.004_{\text{stat}} \pm 0.010_{\text{syst}} \pm 0.017_{\text{theo}}$ [176]. Considering that this value is compatible within the uncertainty with the HFLAV average, it is used in the analysis without any modification;
- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$ is the branching fraction of the $B^+ \rightarrow J/\psi (\rightarrow \mu^+ \mu^-) K^+$ process, its value is taken from world averages of the two branching fractions [11]: $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.010 \pm 0.029) \times 10^{-3}$ and $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033)\%$.
- $N_{J/\psi K^+}$ is the yield of the reference channel, this quantity is extracted from data. Only half of the reference channel sample is employed in the determination of this term, as the other half is used to tune the kinematic distributions of simulated events. This approach avoids unwanted correlations within the analysis;
- $R_{A\epsilon}$ is the ratio of the acceptances (A) and efficiencies (ϵ) for the reference and signal channels, defined as $\frac{A \times \epsilon_{J/\psi K^+}}{A \times \epsilon_{\mu^+ \mu^-}}$; this quantity is extracted from MC simulations corrected with data-driven techniques based on $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ data;
- finally $N_{B_d^0(B_s^0)}$ is the $B_d^0 \rightarrow \mu^+ \mu^-$ ($B_s^0 \rightarrow \mu^+ \mu^-$) signal yield, which is extracted from data.

The whole analysis strategy and tuning is developed maintaining the dimuon mass window that contains the signal blinded (not accessible), to avoid the non-measurable experimenter's bias [177].

The candidates considered in the analysis are required to have fired a low p_T dimuon trigger and pre-selected with a set of loose cuts in order to obtain a uniform reference selection for all the steps of the analysis and manageable sized samples while keeping a very high signal efficiency. These cuts exploit the kinematic of the particles involved in the processes and the topology of the two-body decays $B_{(s)}^0 \rightarrow \mu^+ \mu^-$, in which the B meson travels away from the PV before decaying, allowing the identification of a SV.

On top of this selection, tight requirements on the muon properties are applied, in order to strongly reduce the presence of hadrons mis-identified as muons. In particular the background due to $B \rightarrow hh'$ events, where $h = \pi, K$ and both hadrons are erroneously identified as muons, has the same topology as the signal: a branching fraction four orders of magnitude higher and an invariant mass close to the B_d^0 mass.

At this stage of the analysis it is impossible to distinguish the presence of signal on top of the

background, due to the overwhelming number of background events. The main source of background is the so-called continuum, which is mostly due to muons originating from distinct b quarks in a $b\bar{b}$ pair. A MVA selection, based on a Boosted Decision Tree (BDT) algorithm¹, which makes use of topological and kinematic information of the candidates, is employed to reduce the continuum background contribution.

Finally, a fit to the candidate's invariant mass distribution measures the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ yields. The functional models employed in the fit are studied on MC samples and verified on data when possible. The signal fit is performed solely on data, simultaneously on four datasets obtained by dividing the BDT output in the higher S/B region in contiguous bins with the same signal efficiency. This procedure exploits most of the sample signal events while distinguishing the regions with different signal purities: the lower the BDT range, the lower the signal to background ratio.

The confidence intervals on the signal branching fractions are extracted with the exact Neyman frequentist approach [178]. This approach allows the exact evaluation of confidence intervals and a natural way of switching between upper limits and measurements.

The various steps of the analysis are explained in detail in the following chapters. The content is organised to achieve clarity and minimise forward references.

Before starting the discussion of the analysis, the main statistical tools employed are described (chapter 6), with particular attention given to the Neyman belt construction (sections 6.3.3 and 6.3.4). After describing the collected and simulated datasets employed in the analysis (sections 7.1 and 7.2 in chapter 7), the set of pre-selection cuts, which ensure manageable sized samples and provide a uniform reference for all the analysis steps, is identified (section 7.3).

Once the baseline selection of the analysis is defined, the weighting procedure, designed to correct possible imperfections of the simulated samples, is introduced (section 7.6). After a description of the main backgrounds that affect the analysis (section 8.1 in chapter 8), the final MVA based selection designed to reduce the contribution of the most important background is described (section 8.2).

This concludes the definition and tuning of the analysis selection, therefore a data-MC comparison is performed (chapter 9), to check that simulations properly reproduce real data and to quantify possible residual data-MC discrepancies.

The following steps of the analysis pertain to the various terms present in formula 5.1.

- The evaluation of the $B^+ \rightarrow J/\psi K^+$ yield (section 10.1 in chapter 10).
- The calculation of the R_{Ac} term (section 10.2 in chapter 10).
- The signal yield extraction (chapter 11). Particular attention is given to this important step. First the models used in the fit are accurately defined (sections 11.1 and 11.2 describe the non-resonant and the resonant models respectively), then the fitting procedure is constructed and tested (section 11.3). Finally, the strategy followed to extract the

¹ A Boosted Decision Tree (BDT) is an algorithm based on MVA techniques; a description is provided in chapter 8.

$B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions, based on this fitting procedure, is introduced (section 11.4).

This completes the tuning phase of the analysis, therefore the blinded region in the data is opened (chapter 12). The fitting procedure is applied to the unblinded data and the signal branching fractions are extracted (section 12.1). Additionally, the likelihood of the analysis is combined with the likelihood obtained in the Run 1 analysis to improve the accuracy of the results (section 12.2).

5.1 Expected sensitivity

Before starting the detailed discussion on the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the 2015/16 dataset, its expected sensitivity is evaluated. The significant improvements with respect to the Run 1 analysis that will be shown later in this section motivate the decision of analysing only a part of the Run 2 dataset, while the data taking was still ongoing.

The analysis sensitivity will be illustrated with the expected confidence regions in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane, evaluated with the same procedure employed for the full Run 2 and HL-LHC analysis projections, discussed later in this thesis (section 13.1 in chapter 13).

The expected signal and background statistics are extrapolated relative to the expected number of signal events of the Run 1 analysis [73] according to the SM predictions [54]. Toy-simulations are then generated based on these extrapolations and used to evaluate the confidence regions using the Neyman belt approach (sections 6.3.3 and 6.3.4 in chapter 6).

The expected statistics considers the 2015/16 integrated luminosity, the increase in centre of mass energy of the pp collisions and the different triggers employed with respect to Run 1. The integrated luminosity in the 2015/16 period amounts to 36.2 fb^{-1} , compared to the Run 1 integrated luminosity of 25 fb^{-1} . Additionally, the increase in centre of mass energy of the pp collisions yields an increment in the B mesons production cross-section; the same $\times 1.7$ increment employed in the analysis projections is used to parameterise this effect. Due to the higher instantaneous luminosity that characterises Run 2 with respect to Run 1, the thresholds of the dimuon triggers employed in the analysis had to be increased. This results in a lower signal efficiency. A detailed discussion on the trigger employed in the analysis is provided later in section 7.1.

These ingredients yield a two-fold increment in the available signal statistics with respect to the Run 1 analysis.

The resulting 68.3 %, 95.5 % and 99.7 % confidence regions for the expected statistics, obtained with the Neyman belt approach, are shown in figure 5.1. For comparison, also the confidence regions for the Run 1 expected SM statistics are shown as solid lines. Considering the size of the expected confidence regions, a significant increment in the analysis sensitivity is therefore expected.

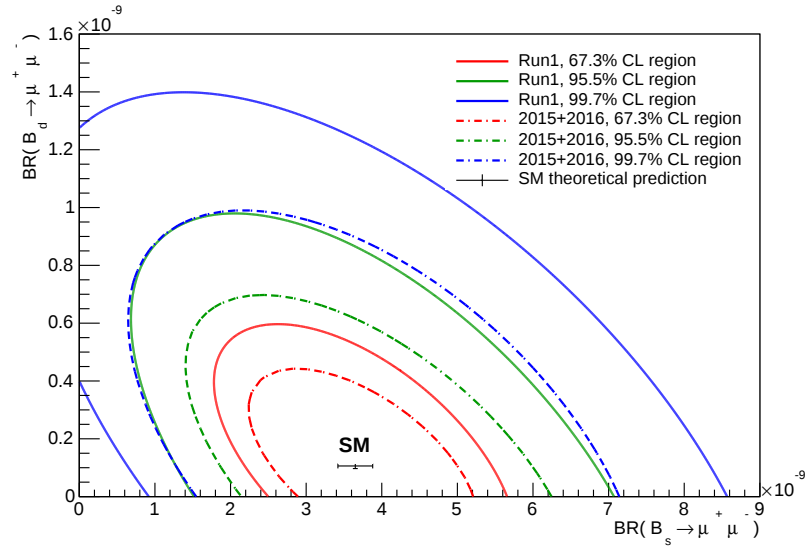


Figure 5.1: Expected 68.3 %, 95.5 % and 99.7 % confidence regions for the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the 2015/16 dataset (dashed lines). The confidence regions for the Run 1 expected statistics according to the SM predictions are shown for comparison (solid lines). The expected 2015/16 statistics is extrapolated based in the Run 1 analysis according to the procedure detailed in section 13.1 in chapter 13. All confidence regions are obtained with the Neyman belt approach [178]. The SM prediction for the signal branching fractions with its uncertainties [54] is also indicated.

STATISTICAL TOOLS

6

The work presented in this thesis relies on several statistical concepts and tools, summarised in this chapter.

The likelihood function and maximum likelihood fits are introduced first (section 6.1), followed by the concepts of hypothesis testing and p -value (section 6.2). The evaluation of the uncertainty on the estimated values with maximum likelihood estimators and the concept of confidence intervals are then introduced (section 6.3). Three approaches for the extraction of the uncertainty (or confidence intervals) are discussed: the Rao-Cramer-Frechet (RCF) variance (sub-section 6.3.1), the likelihood intervals (sub-section 6.3.2) and the Neyman belt construction (sub-section 6.3.3). They will be employed for the extraction of the signal branching fractions uncertainty in sequence, providing a cross-check to the Neyman belt approach, which is chosen to provide the final result of the analysis and its combination with the ATLAS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. Additionally, the evaluation of the result employing the three approaches allows one to cross-check the ATLAS result against other experiment's, depending in the procedure they employ for the uncertainty extraction.

Particular attention is dedicated to the Neyman belt construction (sub-section 6.3.4): while the other approaches are already implemented in the software employed for the estimation of the signal branching fractions (the RooFit toolkit [179] in the ROOT framework [180]), a dedicated 2D construction of the Neyman contours had to be developed for this work.

6.1 Likelihood function

The likelihood concept and the maximum likelihood estimators are extensively employed in high-energy particle physics for the estimation of the parameters of interest in experimental analyses. This section provides an overview on the likelihood and maximum likelihood estimator concepts. A more extended discussion on these topics can be found in Refs. [181–186].

Consider a random variable x , scalar or vector, distributed according to a Probability Distribution Function (PDF) $f(x;\theta)$ [182]; the term θ represents a set of parameters which define the functional form of the PDF. Repeated measurement of x can be performed, leading to a

dataset $x_1, x_2, \dots, x_n$ of size n . By definition, and under the assumption that x is distributed according to $f(x; \theta)$, the probability of finding measurement x_1 in the interval $[x_1, x_1 + dx_1]$ is given by $f(x_1; \theta) dx_1$. Assuming all measurement to be independent, the joint probability that all n measurements fall in $[x_i, x_i + dx_i]$ is given by $\prod_{i=1}^n f(x_i; \theta) dx_i$. If the assumption of the data to be distributed according to $f(x; \theta)$ and the parameter values are correct, the intuitive expectation is a high value of such probability. Vice versa, if there is something not correct in the assumptions one can expect a low value.

Since the parameters do not depend on the dx_i , the likelihood function $\mathcal{L}$ can be therefore introduced as [182]:

$$\mathcal{L}(\theta) = \prod_{i=1}^n f(x_i; \theta). \quad (6.1)$$

Which, For a given dataset $\{x_1, x_2, \dots, x_n\}$, is only a function of the parameter(s) θ .

As discussed in Refs. [181, 182], assuming that the functional form is correct but the value of one, or more parameters is not known, one can estimate their values, using the maximum likelihood estimator:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta_i} = 0 \quad \text{for } i = 1, \dots, n_{\text{parameters}}. \quad (6.2)$$

The estimators of the values that maximise the likelihood are defined as $\hat{\theta}_i$, in order to distinguish them from the real (physical) parameters that are unknown. The act of maximising the likelihood to find the optimal $\hat{\theta}_i$ parameters is called a *maximum likelihood fit*.

Maximum likelihood estimators can be proved to be consistent, meaning that in the limit of large datasets, the estimator $\hat{\theta}_i$ converges to a defined value, and asymptotically unbiased, meaning that a possible bias on the estimator $\hat{\theta}_i$ approaches 0 for increasing size for the dataset. Given their optimal properties and the simplicity of their definitions, maximum likelihood estimators are widely used in particle physics.

Rather than using the likelihood function as defined in equation 6.1, it is usually more convenient to use its logarithm, allowing one to convert the complicated product in $\mathcal{L}$ into a more manageable sum. Since the logarithm is a monotonically increasing function, the maximum of $\mathcal{L}$ is also the maximum of $\log \mathcal{L}$.

There are some examples in which the estimator $\hat{\theta}_i$ can be analytically evaluated, but most of the times numerical techniques are employed.

Based on the properties of the dataset and of the measurement being performed, maximum likelihood fits can be classified as follows [182].

- *Binned* and *unbinned* maximum likelihood fits; the example shown in equation 6.1 assumes an unbinned dataset, meaning that all the data points are explicitly employed in the likelihood, and histograms are not used in the procedure. This fitting procedure is called an unbinned maximum likelihood fit. For very large data samples the unbinned likelihood might become too difficult to compute. Instead of recording each measurement, a histogram can be employed. This reduces the number of data points employed in the fitting procedure to the number of bins of the histogram. The expectation values

of each bin can be estimated by integrating the PDF between the edges of each bin and the joint PDF is now given by a multinomial distribution. The log likelihood distribution becomes therefore:

$$\log \mathcal{L} = \sum_{i=1}^N n_i \log \left(n_{\text{tot}} \int_{x_i^{\text{low}}}^{x_i^{\text{up}}} f(x; \theta) dx \right), \quad (6.3)$$

where additive factors which do not depend on the parameters have been dropped. x_i^{low} and x_i^{up} are the edges of each bin, n_{tot} is the total number of entries in the histogram with N bins and n_i is the content of each bin. The usage of a histograms implies a loss of information on the dataset, therefore the uncertainty on the estimated values (discussed later in sub-section 6.3) is larger compared to the one obtained for unbinned maximum likelihood estimators. In the limit of a small bin size (therefore a large number of bins) binned maximum likelihood estimators become asymptotically the same as the unbinned ones [182].

- *Extended* maximum likelihood fits; in the likelihood shown in function 6.1 the size of the dataset is assumed to be fixed and known. It is often the case that the size of the sample itself is a randomly distributed variable which follows a poisson distribution. In order to take into account this behaviour, the likelihood is modified introducing an additional poissonian PDF as follows:

$$\mathcal{L} = \frac{v^n}{n!} e^{-v} \prod_{i=1}^n f(x_i; \theta), \quad (6.4)$$

where n is the size of the dataset and v is the real and unknown physical parameter of the poisson distribution.

- *Simultaneous* maximum likelihood fits; a likelihood maximum fit can also be performed on two different datasets at the same time. In this case, the two likelihoods relative to the two datasets are multiplied together to obtain a simultaneous likelihood. This quantity is then maximised to estimate the parameters. If the likelihoods employed are not correlated, e. g. do not have any parameter in common, this procedure does not improve the performance of the estimators. On the other hand, if the two likelihoods are correlated, the parameter estimation can be significantly improved. The same approach is often followed to combine experimental results. The likelihoods from different experiments can be combined and maximised, improving the final result.

6.2 Hypothesis test and p -value

One of the frequent goals in a statistical analysis in particle physics is to test whether the predictions of a given model are in agreement with the observed data. A hypothesis H can be defined as a statement for the probability of finding the data x , where x denotes the outcome of a measurement [185]. In case x includes continuous variables, the hypothesis H specifies a PDF for x . The probability of finding a dataset x under the assumption of the hypothesis H can therefore be written as the conditional probability $P(x|H)$ [182].

A statement about the validity of a hypothesis, for instance H_0 , may involve the comparison with some alternative hypothesis $H_1, H_2, H_3, \dots$. A classical example in particle physics is to attribute the H_0 , or null hypothesis, to the “background only” hypothesis, while H_1 , the alternate hypothesis, corresponds to the hypothesis of a signal being present together with the background.

In order to measure the disagreement between the observed data and a given hypothesis, a function of the measured variables is constructed. This function is referred to as test statistic $t(x)$. Each considered hypothesis will therefore imply a given PDF for the test statistic t , that can be written as $g(t|H_0), g(t|H_1)$, etc. In case of a test with two hypotheses, the test statistic can e.g. be designed to assume large values when the data are incompatible with H_0 . The PDF of the test statistic for a given hypothesis can be estimated using asymptotic approximations [184] in case of samples with large statistics, otherwise, the PDF is estimated with toy-MC studies. Toy-MCs, extensively employed in the work described in this thesis, are basically defined as simplified simulations. Instead of employing the full simulation procedure of the physical processes and the detector (section 3.1), toy-MCs rely on the knowledge of the PDF of the observables of the analysis. This allows one to generate quickly and in a CPU affordable way a large amount of different datasets with the same properties as the one composed of real data.

The compatibility of a hypothesis with the data is tested through the introduction of the concept of the p -value. This quantity represents the probability of finding a larger incompatibility between the data and the H hypothesis, under the assumption that H is valid, as illustrated in figure 6.1(a). Considering the hypothesis H_0 , the p -value can therefore be evaluated as [184]:

$$p = \int_{t_{\text{obs}}}^{\infty} g(t|H_0) dt, \quad (6.5)$$

where the integral is taken between the observed value of the test statistic to infinity. It is often convenient to convert the p -value into a significance Z , defined as the number of standard deviations from the mean of a normal distribution for which the integral of the tail of the curve is equal to p . The significance Z is therefore written as

$$Z = \Phi^{-1}(1 - p), \quad (6.6)$$

where Φ^{-1} represents the inverse of the cumulative function of a normal distribution. A graphical representation of the significance is provided in figure 6.1(b). The standard approach in particle physics when e.g. performing a signal+background versus background only hypothesis test, is to declare a *discovery* for a significance $Z \geq 5$ against the background only hypothesis. This corresponds to $p \leq 2.87 \times 10^{-7}$. A significance $3 \leq Z < 5$ is often regarded as *evidence* and corresponds to $2.87 \times 10^{-7} < p \leq 0.0013$.

6.2.1 The profiled likelihood ratio

The choice of an appropriate test statistic is crucial for hypothesis testing. A test statistic widely employed by the LHC experiments is the profiled likelihood ratio, and it is defined as [182, 187]:

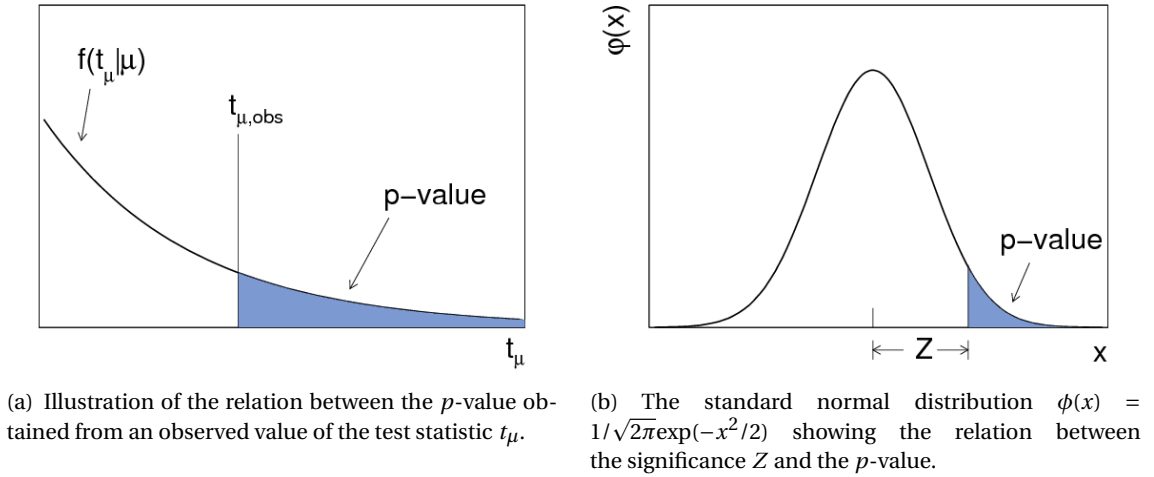


Figure 6.1: Figures from [184].

$$\lambda(s) = \frac{\mathcal{L}(s, \hat{\hat{\theta}})}{\mathcal{L}(\hat{s}, \hat{\hat{\theta}})}, \quad (6.7)$$

where s represents the vector of the parameters of interest, defined as the quantities that the analysis aims to measure, and θ represents the nuisance parameters. These are the remaining parameters of the likelihood and represent quantities not of immediate interest but that must be taken into account in the analysis of the parameters of interest. The $\hat{s}$ and $\hat{\hat{\theta}}$ symbols represent the maximum likelihood estimators of s and θ , while $\hat{\hat{\theta}}$ is the maximum likelihood estimator of θ for a given s ; therefore, the denominator $\mathcal{L}(\hat{s}, \hat{\hat{\theta}})$ corresponds to the absolute maximum of the likelihood, while the numerator $\mathcal{L}(s, \hat{\hat{\theta}})$ corresponds to the maximum of the likelihood for a given value of s .

This quantity ranges between 0 and 1, with higher values corresponding to a better agreement of the data with the s hypothesis.

It can be proved that for samples with large statistics, the function of the profiled likelihood ratio defined as

$$\Lambda(s) = -2 \log \lambda(s) = -2 \log \left(\frac{\mathcal{L}(s, \hat{\hat{\theta}})}{\mathcal{L}(\hat{s}, \hat{\hat{\theta}})} \right) \quad (6.8)$$

behaves as a χ^2 function with the number of degrees of freedom corresponding to the number of parameters of interest [182, 184].

6.3 Uncertainties and confidence intervals

The maximum likelihood approach allows one to obtain an estimation of the parameters of interest, but an evaluation of the uncertainty still has to be discussed. The uncertainty on a measurement is usually reported as its standard deviation σ (or its variance σ^2) and referred to as *standard error*. It measures how widely the estimates would be distributed if the experiment were to be repeated many times with the same number of observations per experiment [182].

The standard error reported for a measurement, $\theta_i \pm \sigma_{\theta_i}$ or $\theta_i^{+\sigma_{\theta_i}^+}_{-\sigma_{\theta_i}^-}$ for asymmetric uncertainties, is conventionally chosen to cover an interval in which 68.3% of the repeated measurements would fall.

The definition of the standard error does not make any statement on the position of the “true” value, the actual physical parameter the experiment is trying to measure. *Confidence intervals*, on the other hand, allow for a quantitative statement about the fraction of times that such an interval would contain the true value of the parameter in a large number of repeated experiments [182]. Confidence intervals are usually extracted to have a coverage of 68.3%, which corresponds to the canonical 1σ interval.

Even if the final result of the analysis discussed in this thesis employs confidence intervals evaluated using the exact frequentist construction known as *Neyman belt* construction [178], two other methods are used to estimate the standard errors and the confidence intervals. This allows one to cross-check the Neyman belt based final result against other procedures and gives a way to validate the ATLAS result against other experiment’s, which may employ a different procedure for the extraction of the uncertainty.

The three approaches employed for the standard errors and confidence intervals evaluation are the are the Rao-Cramer-Frechet (RCF) variance, that estimates the standard error, the likelihood intervals, which allow an approximated evaluation of the confidence intervals, and, as already mentioned, the Neyman belt construction. The three approaches are discussed in the following sub-sections.

6.3.1 RCF variance method

The RCF variance method [182] is the common approach employed to evaluate the covariance matrix when the likelihood is maximised numerically. The approach is derived from the Rao-Cramer-Frechet inequality [188] and assumes a gaussian behaviour of the likelihood as a function of the parameters θ on the vicinity of its maximum. The inverse of the covariance matrix can be obtained as [182]:

$$\left(\widehat{V^{-1}}\right)_{ij} = - \frac{\partial^2 \log \mathcal{L}}{\partial \theta_i \partial \theta_j} \bigg|_{\theta=\hat{\theta}}. \quad (6.9)$$

In case of a single parameter the variance can be therefore derived as:

$$\left(\widehat{\sigma_{\theta}^2}\right) = \left(-1 \bigg/ \frac{\partial^2 \log \mathcal{L}}{\partial \theta^2}\right) \bigg|_{\theta=\hat{\theta}}. \quad (6.10)$$

The MIGRAD and HESSE routines in the MINUIT program [189], extensively employed in the ROOT analysis framework [180], use this approach.

The RCF variance method allows the evaluation of the standard error of a measurement, but does not make any statement on the position of the “true” physical value. Additionally, it depends on the assumption of a gaussian likelihood. For low statistics the likelihood is not expected to show a gaussian behaviour, which implies asymmetric uncertainties on the estimated parameters, therefore the result provided by the RCF variance is not optimal.

6.3.2 Likelihood intervals

Confidence intervals can be estimated with an approximate technique based on the likelihood function. As for the RCF variance, the construction depends on the gaussian assumption for the likelihood as a function of the parameters θ .

In the case of a likelihood function with only one parameter, $\mathcal{L}$ takes the form [182, 185]

$$\mathcal{L}(\theta) = \mathcal{L}_{\max} \exp \left(-\frac{(\hat{\theta} - \theta)^2}{2\sigma_{\theta}^2} \right), \quad (6.11)$$

where $\mathcal{L}_{\max}$ is the maximum value of the likelihood. The following log likelihood ratio can be therefore derived

$$\Lambda(\theta) = -2 \log \frac{\mathcal{L}(\theta)}{\mathcal{L}_{\max}} = \frac{(\hat{\theta} - \theta)^2}{\sigma_{\theta}^2}, \quad (6.12)$$

often written as $-2\Delta \log(\mathcal{L})$, which can lead to a simple prescription to evaluate confidence intervals, by changing the parameter θ by N standard deviations:

$$\Lambda(\hat{\theta} \pm N\sigma_{\theta}) = \pm N^2. \quad (6.13)$$

In particular, the approximated confidence interval with 68.3% coverage can be obtained with $\Lambda(\hat{\theta} \pm \sigma_{\theta}) = \pm 1$.

The procedure just discussed considers a likelihood which depends only on one parameter. In case of one parameter of interest and several nuisance parameters, instead of the likelihood ratio shown in equation 6.12, the profiled likelihood ratio from equation 6.8 is employed.

In the case of a likelihood with several parameters of interest, the profiled likelihood ratio allows the evaluation of approximated confidence regions. The profiled likelihood ratio must be evaluated at a different value to obtain the same approximated coverage as the intervals. For instance, for a likelihood with two parameters of interest the 1, 2 and 3 sigma contours, corresponding to 68.3%, 99.5% and 99.7% coverage, can be obtained with $-2\Delta \log(\mathcal{L}) = 2.3, 6.2$ and 11.8 .

For a gaussian likelihood, the likelihood intervals are symmetric around the likelihood maximum and correspond to the uncertainty evaluated with the RCF variance. In case of a non-gaussian likelihood, the confidence intervals can become asymmetric, as one would expect for a low signal statistics. However, since the hypothesis of gaussianity of the likelihood is lost, the intervals do not necessarily correspond any more to the wanted coverage.

6.3.3 Neyman belt construction

An exact frequentist approach to evaluate confidence intervals was first developed by Neyman [178]. This approach, usually referred to as Neyman belt construction, relies on the knowledge of the PDF of the measured value of a parameter of interest s_m , given the true unknown value of the physical quantity s . This PDF $g(s_m|s)$ can be evaluated with analytical calculations

or, more often, MC based studies.

For each possible value of the true quantity s , one can construct acceptance intervals with a given coverage probability, e. g. 68.3%, defined as $[s_l, s_h](s)$ via the integration [183]:

$$\int_{s_l}^{s_h} g(s_m|s) ds_m = 68.3\%. \quad (6.14)$$

The evaluation of an acceptance interval for each possible value of s leads to the construction of a belt in the $s_m - s$ space. Figure 6.2 shows an illustration of a Neyman belt, where the horizontal lines correspond to the acceptance intervals evaluated for different values of s .

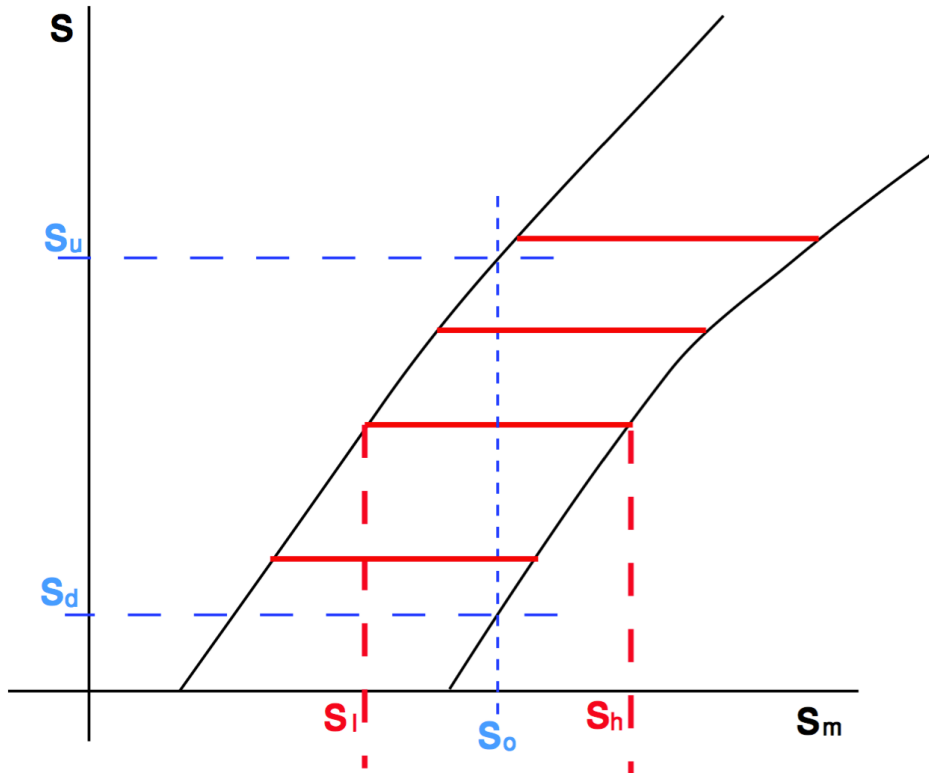


Figure 6.2: Illustration showing the Neyman belt. The horizontal lines are the acceptance intervals in the measured parameter space s_m for a given possible true s , $[s_l, s_h](s)$. Given an observation s_o one can construct the confidence interval $[s_d, s_u]$. Figure from [183].

Given an experimental measure s_o , which lies on the x-axis in figure 6.2, one can extract the 68.3% confidence interval $[s_d, s_u](s_o)$, in the space of the true quantity s , providing an evaluation of the confidence interval.

The same approach can also be employed to obtain confidence regions in the case of more than one parameter of interest. The integration in equation 6.14 is modified and performed over a volume instead of an interval. In this case the Neyman belt does not lie in a 2-dimensional space as the one shown in figure 6.2, but, for the case of two parameters of interest, in a 4-dimensional space, with two dimensions for the true parameters and the other two for the observed parameters.

It must be pointed out that there is no prescription given for the choice of the integration limits

for the acceptance intervals. Given different limits, the size of the intervals might vary. Additionally, one-sided confidence intervals (upper or lower limits) can also be evaluated by changing e.g. the lower bound of the integral in equation 6.14 to minus infinity to obtain upper limits. In order to sort out the integration limits an ordering rule is needed.

A commonly used approach, that allows a unique ordering scheme to evaluate confidence intervals and upper or lower limits without changing the belt construction, is based on the likelihood ratio [181, 190]

$$\lambda(s) = \frac{\mathcal{L}(s, \hat{\theta})}{\mathcal{L}(\hat{s}, \hat{\theta})}. \quad (6.15)$$

This approach, known as Feldman-Cousins ordering, allows seamless switching between an upper limit or a confidence interval for a given coverage based on the measured value s_o .

It is common to come across single-parameter constructions of Neyman belts, for which several software libraries provide pre-packaged solutions. Software solutions for the construction of multi-dimensional belts are not as readily available. As part of the work for this thesis, a software implementation of such a construction in the case of two physical parameters and two observables had to be developed. A description of this implementation is provided in the remainder of this chapter.

6.3.4 Neyman belt implementation

Given the aim of this thesis, the Neyman belt construction will be discussed here with specific reference to the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. In particular, the studies performed to design the Neyman construction procedure are based on the likelihood of the ATLAS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [73], described in appendix A.

The Neyman construction will be employed to extract the belts for the evaluation of the confidence intervals on the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ branching fractions separately (one-dimensional belt) and for the evaluation of the confidence regions in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane (two-dimensional belt). The approach followed for the construction of the belt, which relies on the principles described earlier in this chapter (sub-section 6.3.3), is the same for the one- and two-dimensional cases; for this reason, the main features of the belt construction are explained in this sub-section, while a description of the one- and two-dimensional cases are provided later respectively in sub-sub-sections 6.3.4.1 and 6.3.4.2. The introduction of the systematic uncertainty in the construction procedure is then discussed in sub-sub-section 6.3.4.3.

Toy-MCs are generated according to the analysis likelihood, employing as true value of the parameters of interest the set of values $(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}}, \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{true}})$, or only one of the two for the one-dimensional case. The fitting procedure designed to extract the branching fractions is then applied to evaluate the likelihood maximum and extract the estimated $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ values. A second fit is also performed, fixing the parameters of interest to the true values used in the generation, in order to evaluate the numerator of the profiled like-

likelihood ratio. Once a large set of toy-MCs is ready¹, they are ordered according to the likelihood ratio in formula 6.15.

The toy-MCs estimated branching fractions are taken to build the acceptance interval starting from the ones characterised by a higher value of the likelihood ratio, until the interval presents the wanted coverage. When the confidence band on only one of the branching ratios is sought, the likelihood ratio in formula 6.15 is considered to have only that one parameter of interest, and the resulting Neyman belt construction becomes one-dimensional. When, on the other hand, a constraint on both branching ratios is placed simultaneously from the same dataset, the likelihood ratio presents two parameters of interest and two experimentally observed values of the branching fraction.

In order to ensure exact coverage, two independent sets of toy-MCs are generated for each set of true values. The first set, called sample A, is employed to build the acceptance interval, while the second set, sample B, is employed to measure its coverage. In case under-coverage is detected, toy-MCs from sample A are added to the interval until the coverage measured with sample B reaches the desired value. This procedure could then suffer from biases due to the reliance on sample B for the final choice of the sample A entries that constitute the desired band. For this reason further studies were carried out employing a third simulated sample (sample C) to measure the coverage of the interval obtained with samples A and B. As an example, figure 6.3 shows the unbiased coverage measurement for the construction of the two-dimensional belt with 68.3% (1σ) coverage obtained with the likelihood resulting from the Run 1 analysis. The toy-MCs are generated using the Run 1 likelihood function, varying the expected number of signal events according to the points in a grid in the $(N(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}} - N(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{true}})$ space². For each point on the grid three toy-MC samples are generated (samples A, B and C), each with 10000 events.

The coverage measured using sample A (plots in left column of figure 6.3) is overestimated with respect to the coverage measured with the sample C (plots in right column of figure 6.3), which is unbiased with respect to the belt construction and compatible within its uncertainty with the desired coverage. The coverage measured with sample B (plots in central column of figure 6.3) shows a sharp cut at the wanted coverage, this is due to the construction procedure, as the coverage measured with sample B is not allowed to be lower than the wanted coverage; this choice yields a conservative estimation of the confidence region and its coverage. The compatibility of the measured coverage with sample C with the wanted coverage shows that the procedure employed provides the desired acceptance regions, within the uncertainty discussed later in this section. Additionally, the compatibility of the coverage of sample C with the desired one motivates the usage of two samples for the construction of the acceptance regions, as the usage of sample A alone would bias the coverage measurement.

Due to the limited size of the toy-MC samples employed in the belt construction, statistical

¹ The samples employed are composed of 10000 toy-MCs each, as discussed later in this sub-section.

² The usage of a $(N(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}} - N(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{true}})$ grid instead of a $(\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}} - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{true}})$ grid does not imply any change in the procedure. In fact there is a biunivocal connection between the number of signal events and the branching fraction given the values of the normalisation terms in equation 5.1

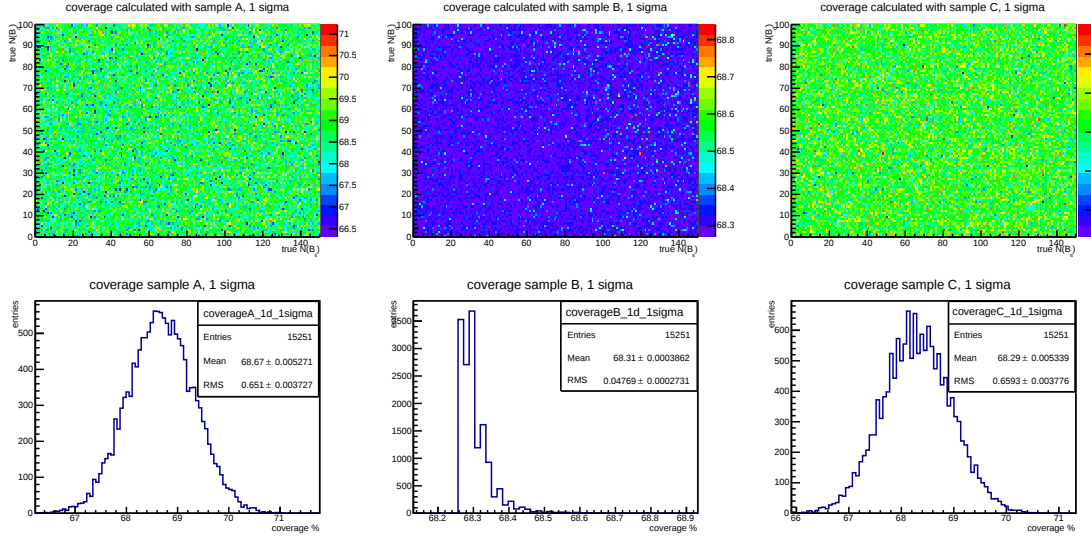


Figure 6.3: Coverage of the 1 sigma (68.27% coverage) acceptance regions built for all the points in a $(N(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}} - N(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{true}})$ grid, all the toy-MC sets are generated using the Run 1 analysis likelihood function. The three columns show the coverage measured with the three samples, respectively, starting from the left, coverage measured with sample A, B and C. The upper plots show the coverage of each point in the grid of possible values used in the generation, while the lower plots show the 1D plot of all measurements of the coverage. The coverage measured with sample C (not used in the region construction) is compatible with the wanted coverage within its statistical uncertainty, while the coverage measured with sample A overestimates the actual value. The coverage measured with sample B shows a sharp cut, this is due to the region construction procedure, as explained in the text.

fluctuation on the measured coverage arise. The uncertainty on the coverage of the 68.3% confidence region due to such fluctuations can be extracted from the Root Mean Square (RMS) of the one-dimensional plots in figure 6.3. The uncertainty on the coverage measured using sample C is comparable with the uncertainty measured with sample A and yields a coverage uncertainty of 0.65 %. The same procedure is performed on the 95.5 % and 99.7 % confidence regions (the canonical 2σ and 3σ), and the resulting uncertainty on the coverage is respectively 0.25 % and 0.07 %.

The uncertainties obtained are considered satisfactory, therefore an increment in the size of the simulated samples in order to reduce statistical fluctuations is not needed. For this reason, all the toy-MCs sets employed for the construction of the Neyman belts shown in this thesis are based on samples with 10000 simulations .

6.3.4.1 One-dimensional belt

The one-dimensional Neyman belt construction follows the procedure previously described, leading to a well defined belt in the plane of measured vs true values of the considered branching fraction. The procedure is tested with the Run 1 likelihood; in the original version of the analysis the confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ was in fact extracted with the Neyman approach.

Figure 6.4 shows the resulting belt in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{true}} - \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{fitted}}$ plane. Both the belts with statistical only uncertainties (blue) and statistical+systematics uncertain-

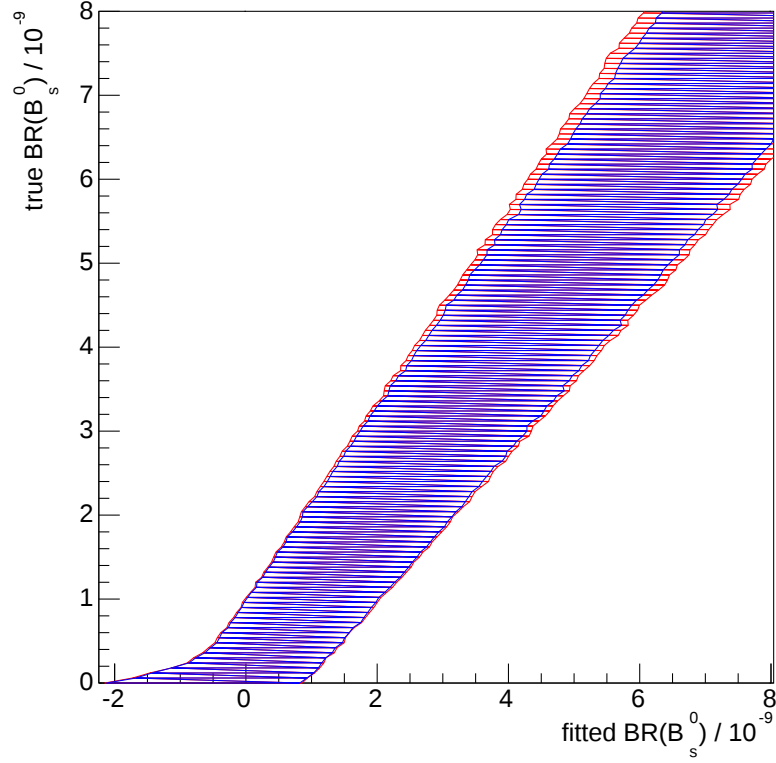


Figure 6.4: One-dimensional Neyman belt built for the extraction of the 68.3% $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ confidence interval using the Run 1 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis likelihood. The blue belt includes statistical uncertainties only, while the red belt includes statistical and systematic contributions, according to the procedure discussed in sub-sub-section 6.3.4.3. The resulting confidence interval for the Run 1 $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ value which maximises the likelihood corresponds to the Run 1 interval.

ties (red) are shown (the inclusions of the systematic uncertainties is discussed later in sub-sub-section 6.3.4.3). The confidence interval obtained for the Run 1 analysis is $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (0.9^{+1.1}_{-0.8}) \times 10^{-9}$ [73], which includes both statistical and systematic uncertainties. Exploiting as experimentally observed value $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0.9 \times 10^{-9}$ and a local approximation of the belt with a first order polynomial in order to reduce the effects of statistical fluctuations, the resulting confidence interval is the same as the one of the Run 1 analysis.

This validates the approach designed for the construction of the Neyman belt, allowing now to extend it, for the evaluation of the two-dimensional confidence regions.

6.3.4.2 Two-dimensional belt

The evaluation of the two-dimensional Neyman belt is a novelty with respect to the Run 1 analysis. In fact, even if the result was in a region in which likelihood contours do not ensure the exact coverage, due to technical reasons the two-dimensional confidence contours were evaluated employing the likelihood contours. Thanks to the tools developed in this thesis, the 2015/16 result is based on a construction which ensures coverage.

In the two-dimensional case the Neyman belt is actually a manifold in a four-dimensional space, where 2 dimensions are dedicated to the fitted values of the branching fractions and

the other two to their true physical values. The acceptance region are therefore 2-dimensional, with respect to the 1-D intervals obtained with the one-dimensional Neyman belt.

This introduces an additional issue in the construction of the acceptance regions. The same likelihood ratio ordering is employed, but, while in the one-dimensional case it naturally provides well defined edges of the acceptance regions, in this case it only provides a set of points in a 2-D space. In order to obtain a well defined contour, a convex hull finding algorithm [191] is employed on the set of points identified.

The two-samples technique to define the acceptance regions and measure their coverage (mentioned in section 6.3.4) is employed also in this case; the observed $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ - $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ values obtained from the toy-MCs fits are ranked according to the likelihood ratio and are taken starting from the ones with the higher ranking score. Every time a point is added the coverage is measured with sample B on the region identified with the convex hull finding algorithm. The procedure stops once the desired coverage is reached.

When all the acceptance regions are defined, the 4-dimensional manifold that constitutes the Neyman belt is ready and the confidence regions can be extracted. Given an experimental value, the two-dimensional confidence regions are evaluated finding the acceptance regions that contain this value. Figure 6.5 shows the confidence region obtained with the construc-

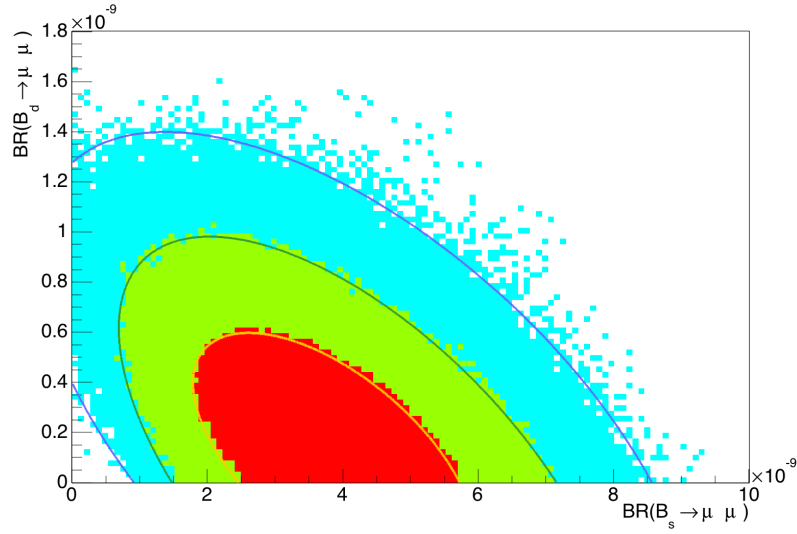


Figure 6.5: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ - $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ two-dimensional confidence regions with coverage 68.3 % (red area), 95.5 % (green area) and 99.7% (blue area), built considering the statistical uncertainty only, based on the Run 1 likelihood and following the procedure described in the text. The expected SM branching fractions [54] are employed as measured value to start the confidence regions construction. Statistical fluctuations are well visible at the edges of the contours, therefore a smoothing technique is applied to derive well defined contours. The three lines superimposed to the filled areas show the resulting confidence regions.

tion procedure based on the Run 1 likelihood, considering only the statistical uncertainty and taking the SM expected branching fractions [54] as experimentally measured values. The filled areas are the confidence regions with 68.3 % (red area), 95.5 % (green area) and 99.7% (blue area) coverage.

Statistical fluctuations are well visible at the edges of the confidence regions, in particular for

the one with 99.7% coverage. As discussed in section 6.3.4, these fluctuations are due to the finite size of the toy-MC samples generated. An additional smoothing procedure is employed, in order to obtain well defined contours. The three lines drawn on top of the filled areas in figure 6.5 correspond to the edges identified with the smoothing algorithm.

6.3.4.3 Systematic uncertainties

The addition of the systematic uncertainties of the analysis to the Neyman construction is not trivial. In fact, systematics are usually implemented in the likelihood as gaussian smearings associated with the nuisance parameters affected by systematic uncertainties. During the likelihood fit, these gaussian smearings increase the uncertainty on the estimated parameters of interest, but do not modify their central values. Since the Neyman approach does not consider the estimated uncertainties, but only the estimated central values of the toy-MCs, an additional step needs to be added in the procedure to include systematic uncertainties.

The generation of the toy-MCs is performed using the likelihood of the analysis, but, before starting the fitting procedure, the central values of the nuisance parameters affected by systematic uncertainties are modified. The gaussian smearings are employed to generate a new central value for the nuisance parameters they are associated to and such values are then used in the fit. In order to verify that this procedure modifies the RMS of the fitted $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ distributions according to the value of the systematic uncertainties, a detailed test was conducted on the Run 1 likelihood. Different sets of toy-MCs are generated using the same truth values of the signal branching fractions (SM prediction) and adding different sources of systematic uncertainties one at the time; the distribution of the various toy-MC fit results are then drawn to evaluate the effect of the systematic uncertainties on the RMS of the distribution.

The systematic uncertainties affecting the Run 1 analysis, discussed in appendix A, can be divided into two main categories, the systematic uncertainties that scale with the estimated signals yield and the systematic uncertainties that are independent from it. Four sets of toy-MCs are generated, according to the following specifications:

- **set 1**, statistical uncertainty only;
- **set 2**, statistical uncertainty and systematic uncertainties independent from the signals yield;
- **set 3**, statistical uncertainty and systematic uncertainties that scale according to the signals yield;
- **set 4**, statistical uncertainty and all systematics considered.

The resulting distributions of the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ estimated values are shown in figure 6.6, together with the measured and expected RMS. The measured RMS is always consistent with the expectations, showing that the procedure for the introduction of the systematic uncertainties affects the distribution of the estimated branching fractions according to the systematics value.

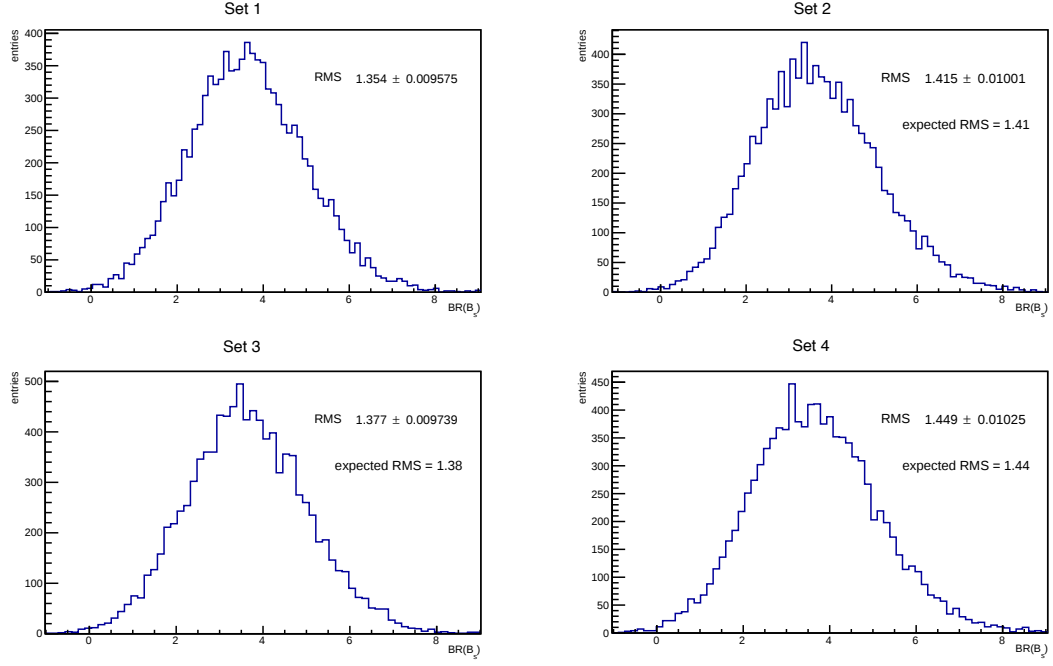


Figure 6.6: Resulting $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ distribution from toy-MC sets generated based the Run 1 likelihood function and using the SM theoretical predictions of the signal branching fractions as true values. The top left plot shows the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ distribution when only the statistical uncertainty is considered (set 1), while the systematic uncertainties do not depend on the signals yield (set 2) are added in the top right plot; in the bottom left plot the systematic uncertainties that scale according to the signals yield are added to the statistical uncertainty (set 3) and in the bottom right plot both systematics are considered (set 4). The measured and expected RMS of the distributions are shown in each plot. The measured RMS is always in agreement with the expectations.

Similar results are obtained for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

The effect of the inclusion of the systematic uncertainties in the Neyman construction is shown in figure 6.7, where the 68.3 %, 95.5 %, and 99.7% confidence regions for the construction based on the Run 1 likelihood and taking the SM expected branching fractions [54] as experimentally measured values are shown. Both the statistical only contours (dashed lines) and the statistical + systematic contours (solid lines) are shown for comparison. Due to the presence of $f_u/f_{s(d)}$ in the evaluation of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and the larger relative statistical uncertainty of $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ compared to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, the effect of the inclusion of the systematic uncertainties is larger on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ with respect to $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

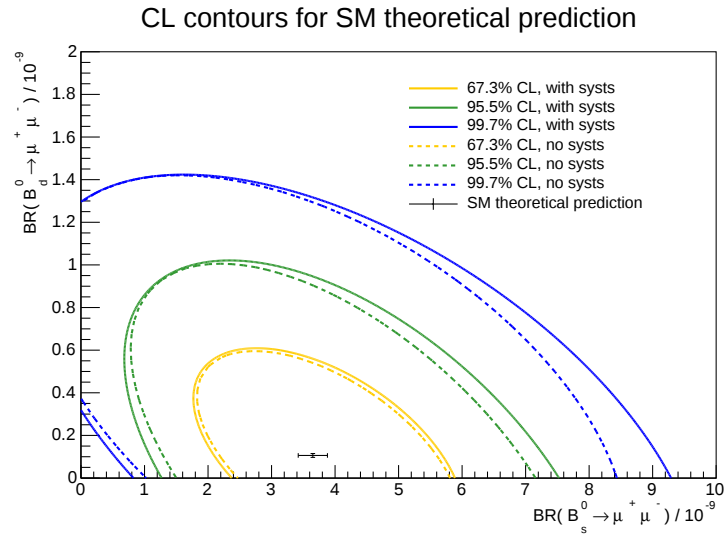


Figure 6.7: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ - $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ two-dimensional confidence regions with coverage 68.3 %, 95.5 % and 99.7%, built considering statistical+systematic uncertainties (solid lines) and statistical only uncertainties (dashed lines). The Belt construction is based on the Run 1 likelihood and follows the procedure described in the text. The SM expected branching fractions are employed as measured values [54].

DATA AND MC SAMPLES

7

A comprehensive list of the data and MC samples used in the analysis is presented in this chapter, together with the preselection applied to both, in order to obtain a uniform reference selection for all the steps of the analysis and manageable sized samples. The weighting procedure applied to the simulated samples in order to obtain a faithful response of the MCs is also described.

The datasets collected by the ATLAS experiment are introduced first, together with the trigger selection employed (section 7.1), followed by the MC samples generated for the analysis (section 7.2). The preselection applied uniformly to all samples is then described (section 7.3); particular attention is devoted to the muons mis-identification probability, which is evaluated on simulations (section 7.4) and validated on data later in the analysis, once the full selection is defined (section 11.2.2.2). The studies carried out to validate the PV-SV association algorithm introduced in section 3.2.2, are then described (section 7.5). Finally, the MC weighting procedure is discussed (section 7.6).

7.1 Data and trigger selection

The analysis uses data from the first part of the ATLAS Run 2 dataset, consisting of $\sqrt{s} = 13$ TeV collision data taken with stable LHC beams in the years 2015 and 2016. The total collected luminosity is 36.2 fb^{-1} with an uncertainty of 2.1%, most of which was collected during 2016. The luminosity measurement is performed using a methodology similar to that detailed in Ref. [192], and based on the LUMinosity Cherenkov Integrating Detector-2 (LUCID-2) detector [193] for the baseline luminosity measurement. Figure 7.1 shows the cumulative luminosity delivered to and recorded by ATLAS during the 2015/2016 data taking period.

The trigger chains (see section 2.2.6 for more details on the ATLAS trigger system) employed in the analysis, referred to as dimuon triggers, are based on the presence of two muons in the MS. Several muon low- p_T thresholds are available, depending on the instantaneous luminosity. All the trigger chains have in common a full track reconstruction of the muon candidates performed by the HLT, where an additional loose selection is imposed on the dimuon invariant mass $m_{\mu\mu}$, accepting candidates in the range 4 to 8.5 GeV.

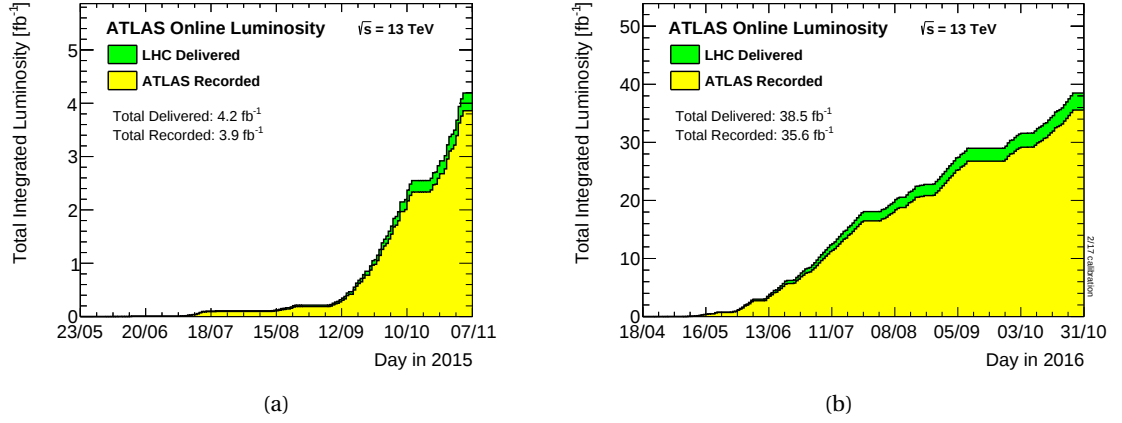


Figure 7.1: Cumulative luminosity versus time delivered to (green) and recorded by ATLAS (yellow) during stable beams for pp collisions at 13 TeV centre-of-mass energy in 2015 7.1(a) and 2016 7.1(b). The “good for physics” luminosity obtained after the application of data quality selection, is not shown and amounts to 36.2 fb^{-1} , with an uncertainty of 2.1%. Plots obtained from [194].

The dimuon triggers employed in the analysis are characterised by muon p_T cuts of 6 GeV and 4 GeV respectively for the leading and sub-leading muon. Two trigger chains with these thresholds are available during the data-taking period considered, both seeded by `L1_MU6_2MU4`¹ at the Level-1 trigger stage:

- `HLT_mu6_mu4_bBmumu`, dimuon trigger with p_T thresholds of 6 and 4 GeV, unprescaled during 2015 and heavily prescaled during 2016, in favour of the next item; this is the baseline trigger of the analysis;
- `HLT_mu6_mu4_bBmumu_Lxy0`, not active in 2015 and prescaled in 2016. This trigger selection applies the same cuts, with the addition of a $L_{xy} > 0$ cut, where L_{xy} is the projection of the distance between the PV and the SV on the transverse momentum of the B candidate evaluated online by the HLT.

The total effective integrated luminosity of these two triggers combined is 26.3 fb^{-1} , where the prescaling of the triggers approximately averages to a reduction by a factor of 1.4.

The addition of another trigger, `HLT_2mu6_bBmumu_Lxy0`, characterised by muon p_T thresholds of 6 GeV, would allow the usage of the full available integrated luminosity, but studies performed with toy-MC simulations show limited improvements on the reach of the analysis, for a major increment in its complexity. For this reason, the trigger chain with muons p_T thresholds of 6 GeV is not used.

Considering the two triggers with muon p_T thresholds at 6 and 4 GeV, a two-fold increment in the available signal statistics with respect to the Run 1 analysis [73] is estimated, exploiting the approach described in chapter 13 for the signal statistics extrapolation, as already mentioned in section 5.1.

¹ L1 trigger which requires a 6 GeV muon p_T threshold to fire once and a 4 GeV muon p_T threshold to fire twice, effectively requiring a muon with $p_T = 6 \text{ GeV}$ and a muon with $p_T = 4 \text{ GeV}$.

The $B^+ \rightarrow J/\psi K^+$ (reference) and $B_s^0 \rightarrow J/\psi \phi$ (control) channels need a different trigger selection, as the dimuon invariant mass of the J/ψ meson would fall outside of the mass window selected. In order to allow the systematic uncertainty reduction in equation 5.1, the trigger selection performed has to be similar to the one used for the signal samples. For this reason, the same L1 trigger is employed, `L1_MU6_2MU4`, and the HLT selection only differs for the dimuon invariant mass $m_{\mu\mu}$ cut, tuned on the J/ψ meson mass ($2.5 < m_{\mu\mu} < 4.3$ GeV). The available triggers with this property are:

- `HLT_mu6_mu4_bJpsimumu`: basic dimuon trigger for muons p_T thresholds of 6 and 4 GeV, only active in 2015;
- `HLT_mu6_mu4_bJpsimumu_Lxy0`: adds an $L_{xy} > 0$ cut on-top of `HLT_mu6_mu4_bJpsimumu`. This trigger was prescaled at the beginning of 2016 and heavily prescaled afterwards;
- `HLT_mu6_mu4_bJpsimumu_Lxy0_delayed`: same as `HLT_mu6_mu4_bJpsimumu_Lxy0` but redirected to a different data stream, named delayed stream.

The total effective collected integrated luminosity for the reference and control channels is 15.1 fb^{-1} , for an effective prescale of about 2.4.

7.2 MC

Several MC samples are required for most of the steps of the analysis: a comprehensive list of the processes simulated is provided in table 7.1. MC samples are categorised “signal” if they are used to emulate the behaviour of the signal processes, “reference” for the MC used for the reference channel simulation, “control” for the control channel and “bkg” for the background samples. In addition, the number of generated events is reported, as well as the MC generator used.

In order to expedite the generation of the MC samples listed in the table, a set of kinematic cuts is employed at the generation level. The usage of such cuts can introduce a bias in the kinematic distributions. For this reason, a re-weighting procedure, discussed in detail in section 7.6.4.1, is applied to the MC samples.

For all the MC samples the detector simulation has been performed using ATLFast-II, described in section 3.1.2, the only exceptions are the samples used for the studies of the peaking background and muon mis-identification probability, for which the full simulation of the detector is exploited.

Most of the dimuon candidates in the data sample, as observed in the Run 1 analysis, originate from the decay of hadrons produced in the hadronisation of $b\bar{b}$ pairs. This includes a large amount of possible processes, which are simulated in the inclusive $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC. The generation of this sample requires the presence of two muons in the final state, with both muons

Table 7.1: Monte Carlo data samples for signal, reference, control and background channels. EvtGen is employed for all the samples except for $B_s^0 \rightarrow J/\psi\phi$, $B \rightarrow hh'$ and $b\bar{b} \rightarrow \mu^+\mu^-X$. This is because the $B_s^0 \rightarrow J/\psi\phi$ sample is shared with other analyses, therefore it is generated with flat angular distributions and then the SM distributions are recovered through a hit-and-miss approach. Regarding the $B \rightarrow hh'$ sample, in order to ????? The $b\bar{b} \rightarrow \mu^+\mu^-X$ is specifically designed to contain only certain decays, therefore the property of such decays are added by hand.

Channel	Type	Events	Generator
$B_s^0 \rightarrow \mu^+ \mu^-$	exclusive signal	1,000,000	PYTHIA+ EvtGen
$B_d^0 \rightarrow \mu^+ \mu^-$	exclusive signal	1,000,000	PYTHIA+ EvtGen
$B^+ \rightarrow J/\psi K^+$ with $J/\psi \rightarrow \mu^+ \mu^-$	exclusive reference	1,997,000	PYTHIA+ EvtGen
$B^- \rightarrow J/\psi K^-$ with $J/\psi \rightarrow \mu^+ \mu^-$	exclusive reference	1,999,500	PYTHIA+ EvtGen
$B^+ \rightarrow J/\psi \pi^+$ with $J/\psi \rightarrow \mu^+ \mu^-$	exclusive bkg	498,000	PYTHIA+ EvtGen
$B^- \rightarrow J/\psi \pi^-$ with $J/\psi \rightarrow \mu^+ \mu^-$	exclusive bkg	500,000	PYTHIA+ EvtGen
$B_s^0 \rightarrow J/\psi\phi$ with $J/\psi \rightarrow \mu^+ \mu^-$, $\phi \rightarrow K^+ K^-$	exclusive control	5,000,000	PYTHIA+ Photos
$B \rightarrow hh'$	exclusive bkg	5,000,000	PYTHIA
$B_s^0 \rightarrow K^- \mu^+ \nu$	exclusive bkg	250,000	PYTHIA+ EvtGen
$B_s^0 \rightarrow \pi^- \mu^+ \nu$	exclusive bkg	500,000	PYTHIA+ EvtGen
$\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}$	exclusive bkg	250,000	PYTHIA+ EvtGen
$b\bar{b} \rightarrow J/\psi X$ with $J/\psi \rightarrow \mu^+ \mu^-$	inclusive bkg	10,000,000	PYTHIA+ EvtGen
$b\bar{b} \rightarrow \mu^+ \mu^- X$	inclusive bkg	650,000,000	PYTHIA+ Photos

originating from the $b\bar{b}$ decay chain. Prompt dimuons and dimuons from $c\bar{c}$ pairs are not included in this simulation, as their contribution would be removed in any case by the MVA selection, discussed in section 8.2.

According to the SM, the main contributor to the $B_s^0 \rightarrow \mu^+ \mu^-$ decay is the $B_{s,H}$ state, whose lifetime is 1.615 ps [11]. On the other hand, some new physics scenarios are expected to modify the decay time distribution of $B_s^0 \rightarrow \mu^+ \mu^-$, which would become different from the one of $B_{s,H}$, with the effect related to the observable $\mathcal{A}_{\Delta\Gamma}^{\mu\mu}$ [48, 53]. Possible changes in the analysis due to lifetime variations can be checked by applying a correction to the acceptance and efficiency ratio present in formula 5.1. In order to minimise the uncertainty due to application of this correction, the $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample is generated using as lifetime of the B_s^0 meson the average of the two mass eigenstates $B_{s,H}$ and $B_{s,L}$, 1.533 ps. A correction is therefore applied to the R_{Ac} term in formula 5.1 for the evaluation of the branching fraction in the SM.

This procedure is not employed for the $B_d^0 \rightarrow \mu^+ \mu^-$ process; due of the small value of $\Delta\Gamma_d$, variations in the branching fraction would not be appreciable. For this reason, the $B_d^0 \rightarrow \mu^+ \mu^-$ MC is generated maintaining the B_d^0 experimentally measured lifetime.

7.3 Event preselection

A first loose selection is applied to the candidates of both MC and data samples allowing manageable dataset sizes and a uniform reference selection for all the analysis steps. Table 7.2

shows the preselection applied to the signal, reference and control channels, on top of the trigger selection described in section 7.1.

The information on the B decay vertex (SV) can be determined either combining the ID and

Table 7.2: Preselection applied to signal, reference and control channels. The first column shows the cut applied, the second column the channel(s) affected and the third column provides a brief description of the quantity involved.

Cut	Type	description
<i>combined</i> muon [161]	all	Muon candidates are required to have compatible tracks in both the ID and the MS.
$p_T(\mu_1) > 6 \text{ GeV}$, $p_T(\mu_2) > 4 \text{ GeV}$	all	Transverse momentum of the muon candidates.
$ \eta(\mu) < 2.5$	all	Pseudorapidity of the muon candidates.
$p_T^B > 8.0 \text{ GeV}$	all	Transverse momentum of the B candidate.
$ \eta^B < 2.5$	all	Pseudorapidity of the B candidates.
$\chi_B^2/NDF < 6$	all	Reduced χ^2 obtained from the vertexing procedure employed to reconstruct the SV.
$4766 \text{ MeV} < m(B) < 5966 \text{ MeV}$	signal and control	Invariant mass of the B candidates.
$4930 \text{ MeV} < m(B) < 5630 \text{ MeV}$	reference	Invariant mass of the B candidates.
<i>loose</i> track quality selection [195]	reference and control	Quality requirement on the tracks.
$p_T(K) > 1.0 \text{ GeV}$	reference and control	Transverse momentum of the kaon candidates.
$ \eta(K) < 2.5$	reference and control	Pseudorapidity of the kaon candidates.
$2915 < m(J/\psi) < 3275 \text{ MeV}$	reference and control	Invariant mass of the J/ψ candidates.
$\chi_{J/\psi}^2/NDF < 6$	reference and control	Reduced χ^2 obtained from the vertexing procedure employed to reconstruct the J/ψ candidate.
$1005 < m(\phi) < 1035 \text{ MeV}$	control	Invariant mass for the ϕ candidate.
$\chi_\phi^2/NDF < 10$	control	Reduced χ^2 obtained from the vertexing procedure employed to reconstruct the ϕ candidate.

MS tracking information (section 2.2.5) or using ID hits alone (section 2.2.3). The first approach was proved to provide a narrower mass distribution of the signal MC samples with respect to the second approach, and therefore it is chosen to evaluate the properties of the B candidates for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ channels.

The reference and control modes final states contain also additional tracks (the kaons in the B decay chain); for these channels the combined ID-MS information was found to provide wider mass distributions than the second approach. For this reason, only the ID hits are employed

for them.

The mass distribution of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates has a blinded region (5166-5526 MeV) chosen to exclude the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ peaks. The reconstructed mass interval ($4766 \text{ MeV} < m(B) < 5966 \text{ MeV}$) is designed to contain this blinded region, allowing studies on the mass sidebands ($4766 \text{ MeV} < m(B) < 5166 \text{ MeV}$ and $5526 \text{ MeV} < m(B) < 5966 \text{ MeV}$). Figure 7.2 shows the expected shapes of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ mass distributions obtained from simulations, together with the width of the blinded region; more detailed studies are provided in section 11.2.1. Data contained in this region was not accessible until the

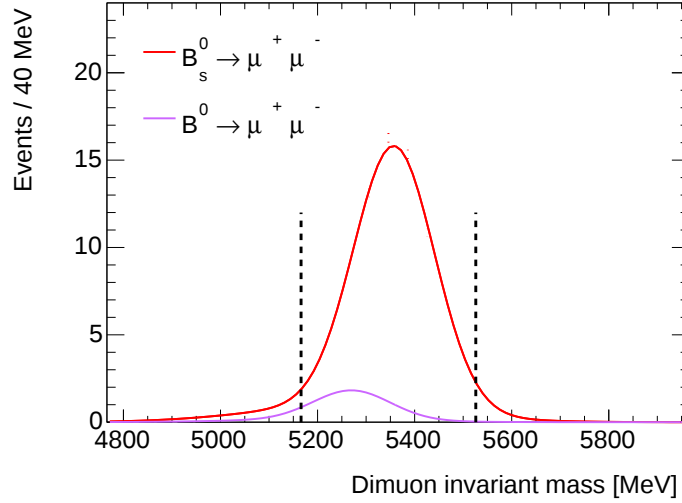


Figure 7.2: Dimuon invariant mass distribution for the B_s^0 and B_d^0 signals shapes obtained from simulations (section 11.2.1), normalised to the SM prediction for the expected yield with an integrated luminosity of 26.3 fb^{-1} . The two dashed lines indicate the blinded region that contains the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ peaks.

analysis was fully tuned.

The *tight* muon [161] working point is also required, in order to enhance the fake-muon rejection. This requirement is relaxed in the studies performed on the $B \rightarrow hh'$ MC, to estimate the muon mis-identification fraction, section 7.4, and on the studies performed on data and MC to estimate the $B \rightarrow hh'$ yield in the signal region, section 11.2.2.2.

Additional cuts are applied to reduce the background yield without significantly affecting the signal. The addition of these cuts reduces the background by more than 20%, maintaining a signal efficiency of about 98%. These are:

- $\Delta R_{\text{flight}} < 1.5$;
- $|\alpha_{2D}| < 1.0$;
- $L_{xy} > 0$, tightened to $L_{xy} > 0.3$ for the $B^+ \rightarrow J/\psi K^+$ Data Driven Weights (DDW) calculation, see section 7.6.4.2.

Where ΔR_{flight} is the three-dimensional opening, defined as $\sqrt{\Delta\phi^2 + \Delta\eta^2}$, between the B candidate reconstructed momentum and the vector between the PV and SV, and $|\alpha_{2D}|$ is the absolute value of its transverse-plane projection. L_{xy} is the projection of the distance between the PV and the SV along the transverse momentum of the B candidate (already mentioned in section 7.1).

7.4 Studies on fake muons rate

Muon fake rates need to be carefully tuned and measured as the backgrounds due to B decays with hadrons mis-identified as muons constitute one of the main background sources of the analysis. This section shows therefore the studies performed on the $B \rightarrow hh'$ MC sample to evaluate the mis-identification fraction (or fake rate) of hadrons into muons. Additional studies on the $B_s^0 \rightarrow \mu^+ \mu^-$ simulated sample are performed, to evaluate the efficiency of real muons. The muon fake rate measured on MC will then be validated on real data in section 11.2.2.2.

The mis-identification fraction is evaluated using the hadrons coming from $B \rightarrow hh'$ decays, as the MC sample generated for this process has large statistics and the hadrons produced present the typical kinematic properties of the particles involved in this analysis. The two hadrons in the decay are considered uncorrelated, as the possible effect of such correlation on the mis-identification fraction has been tested and found to be negligible. The quantity evaluated in this section is therefore the “single-leg” mis-identification fraction.

The calculation is based on the ratio of hadrons mis-identified as muons divided by the total number of hadrons. The preselection described in the previous chapter is applied to both the numerator and the denominator, except for the *tight* muon Working Point (WP) requirement and the kinematic cuts on the reconstructed muons. Since most of the events do not contain any reconstructed muon, the kinematic selection is applied to the tracks in the ID and the p_T of the tracks is required to be higher than 4 GeV. In order to be mis-identified as a muon, a hadron not only has to be reconstructed offline as a muon object, but also needs to be identified as a muon by the online trigger system. For this reason, a single muon trigger, with a p_T threshold of 4 GeV, is required in the numerator, while no trigger requirements are applied to the denominator. A different muon threshold (p_T higher than 6 GeV) has also been tested, applying the threshold to both the offline muon p_T and the trigger threshold, but no significant difference was found.

The muon efficiency is evaluated on the $B_s^0 \rightarrow \mu^+ \mu^-$ MC, using only muons coming from the signal process. The number of muons that pass the muon quality requirements divided by the total number of muons is defined as muon efficiency. Also in this case the preselection described in the previous chapter is applied to both the numerator and the denominator, except for the *tight* muon quality requirement. The dimuon trigger `HLT_mu6_mu4_bBmmumu`, seeded at L1 by `L1_MU6_2MU4` is required to have fired in both the numerator and the denominator. Possible variations in the muon efficiency might be due to the usage of a different trigger. The

other HLT trigger employed in the analysis is therefore tested (HLT_mu6_mu4_bBmumu_Lxy0) and no difference in the muon efficiency is found.

The definitions of the mis-identification fraction and muon efficiency follow the ones used in the Run 1 analysis, so that the performance of the two analyses can be compared.

The Run 1 analysis [73] employed an ad-hoc BDT classifier (“fake-BDT”) to optimise efficiency vs mis-ID probability (section A.2 in appendix A), as the ATLAS default muon identification was found to be not sufficient. The muon requirements have been re-defined for Run 2 [161], improving the muon purity, thanks to the introduction of different Working Points (WPs) which implement some of the most effective variables used in the fake-BDT.

Table 7.3 shows the mis-identification fractions and the muon efficiency for the *loose* and *tight* WPs, for different admixtures of positive and negative π and K . The mis-identification fraction and muon efficiency from the Run 1 analysis are also shown (when available).

Table 7.3: Mis-identification fraction and muon efficiency for the *loose* and *tight* muon Working Points (WPs), measured respectively employing the $B \rightarrow hh'$ and $B_s^0 \rightarrow \mu^+ \mu^-$ exclusive MC samples. For comparison, the same quantities from the Run 1 analysis, when available. Different admixtures of positive and negative π and K are considered. The mis-identification fraction is calculated as the ratio of the number of hadrons after all selections and the quality requirements divided by the total number of hadrons that pass all selections. Muon efficiency is defined as the number of muons that pass all selections and the quality requirements divided by the total number of muons that pass all selections. The uncertainty, when reported, is purely statistical, due to the finite size of the simulated samples.

		Run1 fake-BDT	<i>loose</i> muon WP	<i>tight</i> muon WP
Mis-identification fraction	total	0.00067	$0.002395 \pm 3.0\text{e-}05$	$0.000944 \pm 1.9\text{e-}05$
	total +		$0.00227 \pm 4\text{e-}05$	$0.00091 \pm 3\text{e-}05$
	total -		$0.00252 \pm 4\text{e-}05$	$0.00098 \pm 3\text{e-}05$
	total π	0.0004	$0.00270 \pm 5\text{e-}05$	$0.00106 \pm 3\text{e-}05$
	total K	0.0009	$0.00210 \pm 4\text{e-}05$	$0.00083 \pm 3\text{e-}05$
	π^+	0.00042	$0.00257 \pm 6\text{e-}05$	$0.00103 \pm 4\text{e-}05$
	π^-	0.00044	$0.00282 \pm 7\text{e-}05$	$0.00108 \pm 4\text{e-}05$
	K^+	0.00101	$0.00199 \pm 6\text{e-}05$	$0.00079 \pm 4\text{e-}05$
	K^-	0.00076	$0.00222 \pm 6\text{e-}05$	$0.00088 \pm 4\text{e-}05$
Muon efficiency	μ	0.95	$0.997 \pm 1\text{e-}03$	$0.899 \pm 1\text{e-}03$
	μ^+		$0.997 \pm 1\text{e-}03$	$0.900 \pm 1\text{e-}03$
	μ^-		$0.997 \pm 1\text{e-}03$	$0.898 \pm 1\text{e-}03$

The resulting mis-identification fraction for the *tight* muon WP is a factor of ~ 1.5 higher than the Run 1 fake-BDT, for a comparable muon efficiency.

In order to ensure that the usage of the *tight* muon WP does not compromise the performance of the analysis, a toy-MC based study is performed.

The impact of the usage of *tight* muon WP on the Run 1 analysis is verified first, in order to check the possible degradation of the result. This studies are performed generating toy-MC using the likelihood of the Run 1 analysis in two configurations. The first configuration is the

baseline Run 1 analysis, as described in appendix A; the second configuration uses the same functional forms as the Run 1 likelihood, but the normalisations of the various components are set to the usage of the Run 2 *tight* muon WP. Such expected yields are calculated by applying the mis-identification rates and muon efficiency from table 7.3 to the calculated yields of the Run 1 analysis before the application of the fake-BDT. All the toy-simulations are generated for the SM value of the signal branching fractions [54].

The impact of the *tight* muon WP with respect to the usage of an ad-hoc fake-BDT is then checked scaling the expected signal and background statistics to the one of the 2015/16 analysis, according to the evaluation performed in section 7.1. Two sets of toy-MC are generated also in this case, maintaining the properties of the ones already described, but scaling the statistics by a factor of 2.

The resulting RMS of the signal yield distributions obtained from the four toy-MC studies are shown in table 7.4. No significant effect on analysis sensitivity is visible for B_s^0 , while a $\sim 1\%$ broader RMS is visible on the B_d^0 count.

In view of these result, the *tight* muon WP does not compromise the performance of the analysis, and is therefore employed to reduce the fake muons contamination.

Table 7.4: RMS of distributions of number of fitted B_s^0 and B_d^0 candidates in the four toy-MC studies described in the text. The columns labelled “Run 1 analysis” refer to the statistics available in the previous analysis, modified for the usage of tight muons in case of the second column. Columns labelled “2015/16 analysis” refer to the estimation of the statistics available for this analysis, performed in section 7.1.

	Run 1 analysis		2015/16 analysis	
	fake-BDT	tight muons WP	fake-BDT	tight muons WP
RMS $N_{B_s^0 \rightarrow \mu^+ \mu^-}$	15.20 ± 0.01	15.16 ± 0.01	21.260 ± 0.015	21.260 ± 0.015
RMS $N_{B_d^0 \rightarrow \mu^+ \mu^-}$	13.88 ± 0.01	13.96 ± 0.01	18.850 ± 0.015	19.130 ± 0.015

7.5 Primary vertex association

The standard ATLAS approach to select the PV to be associated with the candidates of interest ($\sum p_T^2$, described in section 3.2.2) was proven to be not optimal for the low momentum scale of the physics objects involved in this search in the previous version of the analysis [73]. The correct association of the B decay vertex to its PV is critical for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. In fact, some of the variables used in the MVA selection for the background reduction, described in section 8.2, heavily depend on the correct PV-SV association. For this reason, the previous analysis relied on an ad-hoc approach to associate PV to B candidates; this approach is named PV_MIN_Z0_BA and is described below.

Given the increased energy and pile-up conditions that characterise the Run 2 of the LHC, the PV-SV association algorithm needs to be tested.

In this section the performance of the Run 1 approach is compared to three alternate methods. The four algorithms considered are:

- PV_MAX_SUM_PT2: predefined in ATLAS, considers the sum of the squared transverse momentum of the tracks associated to each PV, the chosen PV is the one with the highest sum;
- PV_MIN_A0: a backward extrapolation of the B momentum from the decay vertex is considered, the PV is chosen as the one with the shortest 3D distance from the Point Of Closest Approach (POCA) of the B extrapolation to each of the reconstructed PVs;
- PV_MIN_Z0: similar to PV_MIN_A0, but uses the distance along z from the POCA of the B extrapolation to each of the reconstructed PVs;
- PV_MIN_Z0_BA (Run 1 analysis approach): the associated PV is chosen as the one with the shortest separation, along z , from the POCA of the B extrapolation to the beam line.

The performance of these procedures is tested on a sub-sample of the signal MC.

A quantitative estimation of the algorithm's performance is obtained computing, for each PV reconstructed in the events, the χ^2 compatibility of the 3D vertex position relative to the MC truth for the signal. The reconstructed PV with the lowest χ^2 in each event is "truth-matched" i. e. considered to be correctly associated to the B candidate. Figure 7.3 shows the distribution of the χ^2 of all the PVs reconstructed in the events (blue points) and the distribution of the χ^2 of the "truth-matched" PVs (red points). Two clear peaks appear in the plot, the first on the left, which corresponds to the "truth-matched" vertices and the second, much higher, on the right, corresponding to all the remaining vertices. Figure 7.4 shows the distribution of the chosen

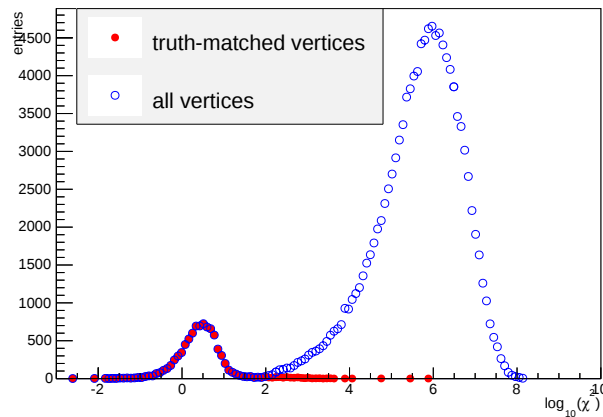


Figure 7.3: χ^2 of all the PVs (blue points) and the distribution of the χ^2 of the "truth-matched" PVs.

PVs using the 4 approaches (green distributions) superimposed to the truth-matched distribution (red). While the PV_MAX_SUM_PT2 distribution shows two clear peaks, corresponding to correct (left peak) and wrong (right peak) associations, the other three algorithms present essentially the single "truth-matched" peak one would expect for the right PV association. Two quantities can be defined to compare the performance of the four algorithms:

- *efficiency*, defined as the ratio of the number of successful PV-SV associations, without considering if it is correct, and the total number of B candidates;

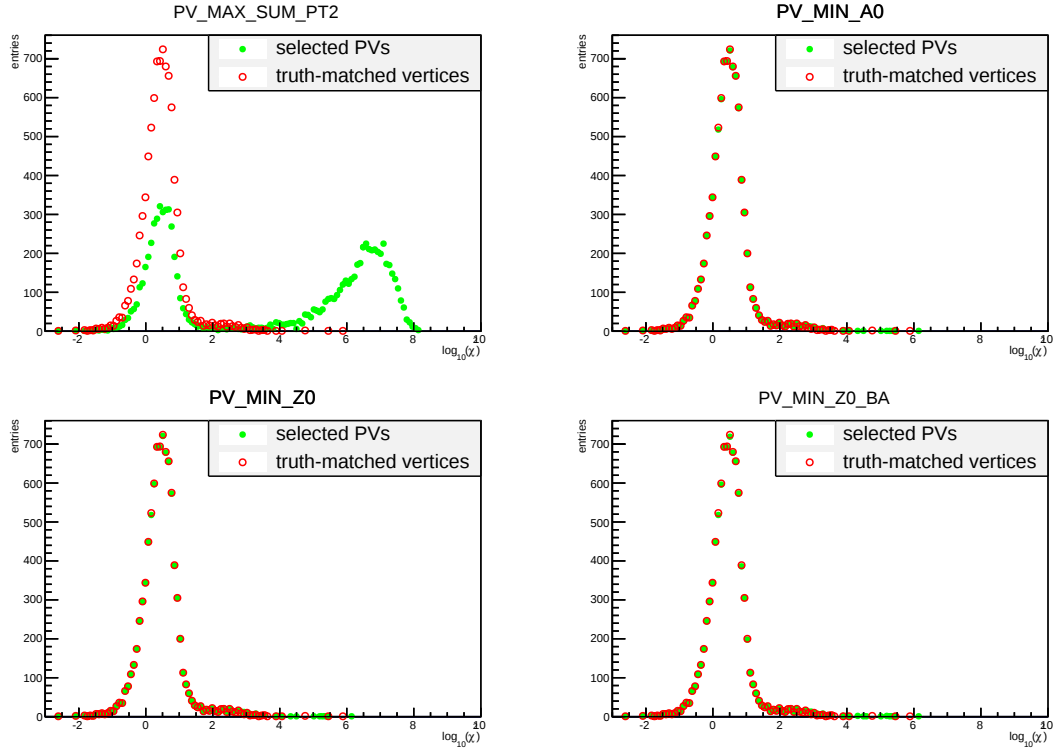


Figure 7.4: χ^2 distribution of the chosen PVs using the 4 approaches (green distributions) superimposed to the truth-matched distribution (red).

- *purity*, defined as the ratio between the number of correct associations and the total number of successful associations.

Given the definition of the four approaches, their efficiency is always 1, as they always provide a PV candidate. On the other hand, the purity of each algorithm, shown in table 7.5, can vary. As

Table 7.5: Purity of the four considered approaches to perform PV-SV association.

Approach	Purity
PV_MAX_SUM_PT2	0.451 ± 0.0053
PV_MIN_A0	0.9938 ± 0.0008
PV_MIN_Z0	0.9937 ± 0.0008
PV_MIN_Z0_BA	0.9931 ± 0.0009

anticipated, the purity for PV_MAX_SUM_PT2 is significantly lower than the other approaches, which have compatible performance with one another.

Due to the increasing pile-up environment in Run 2, the stability of the four approaches as a function of the number of reconstructed primary vertices in each event is also tested. Figure 7.5 shows the purity of each approach as a function of the number of reconstructed primary vertices in the pp collision. While PV_MAX_SUM_PT2's purity depends on the number of reconstructed primary vertices, the other algorithms appear stable. Given the lack of significant

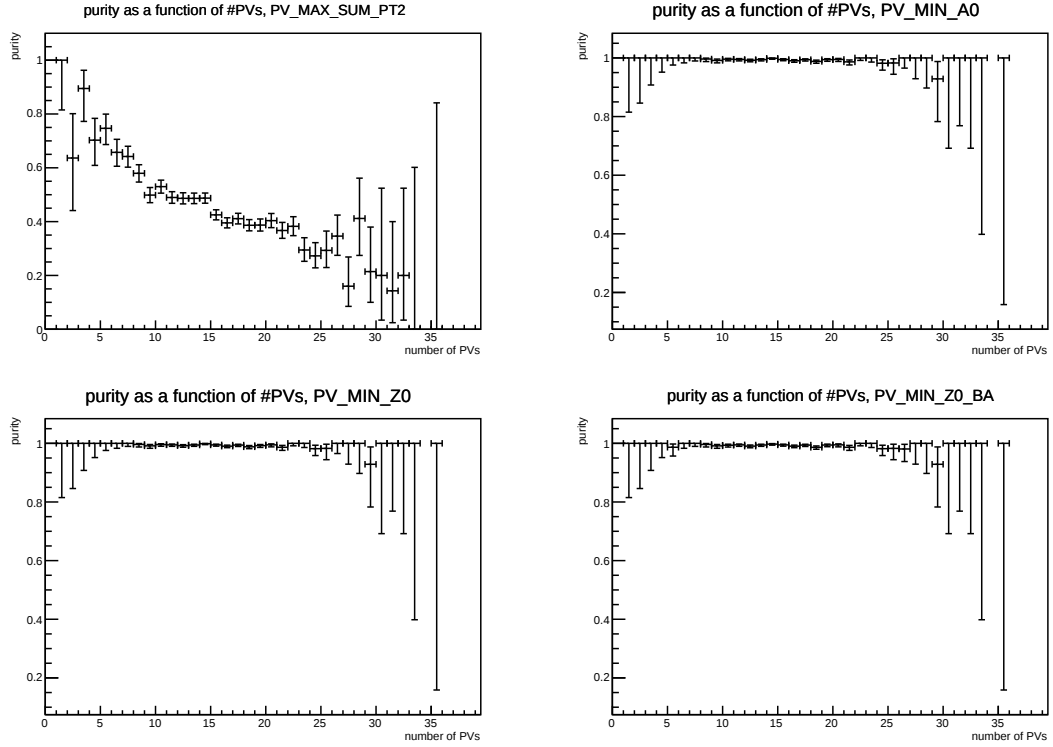


Figure 7.5: Purity as a function of number of reconstructed PVs for the four approaches.

improvements with the two new approaches tested, the same algorithm used in the Run 1 analysis (PV_MIN_Z0_BA) is chosen.

7.6 MC samples reweighting

MC simulations often do not fully reproduce the data kinematics, one common approach applicable in cases where reference samples are accessible is to re-weight them to obtain a realistic response. This procedure allows also to evaluate systematic uncertainties related to the MC imperfections.

This section provides a description of the main causes of MC mis-modelling and the approach followed in the reweighting procedure.

The main causes of data-MC discrepancies are an imperfect detector and trigger response simulation, the usage of a different PU profile in the simulation with respect to data and kinematic discrepancies of the B meson spectrum due to imperfections in the b quark phase space and hadronisation models.

The differences between MC and data are tackled separately, introducing different sets of corrections to be applied on MC simulations. In order to have a realistic response in simulations, all the corrections are evaluated with data-driven approaches.

- **Pile-up reweighting:** this set of weights is used to correct the PU profile of the simu-

lation in order to match the one present in data, these weights are described in sub-section 7.6.1.

- **Muon offline efficiency weights:** these weights correct the detector response to muons in simulations, they are described in sub-section 7.6.2.
- **Muon trigger weights:** used to correct the trigger response to muons in simulations, described in sub-section 7.6.3.
- **Kinematic corrections:** these weights are used to correct kinematic discrepancies in the B candidates spectrum in MC. As introduced in section 7.2, only the exclusive samples are subjected to this procedure. More attention is dedicated to these corrections in the current chapter, as the author had a major role in their calculation. Sub-section 7.6.4 presents an extensive discussions on these weights.

7.6.1 Pile-up reweighting

MC samples are usually produced before data taking has concluded, by this design, only a best guess of the data Pile-Up (PU) conditions can be put into the MC. For this reason, there is the need to reweight the MC PU profile to the actual data taking conditions.

The PU reweighting is based on the average number of pile-up interactions, defined as $\langle\mu\rangle$. This quantity allows one to take into account both the in-time PU (number of interactions in the same bunch crossing, as explained in section 3.1.1) and the out-of-time PU (overlapping signals in the detector from other neighbouring bunch crossings). Figure 7.6 shows the mean number of interactions per bunch-crossing during the 2015-2016 data taking campaign. Since the PU reweighting is a procedure applied to essentially all ATLAS analyses, a common

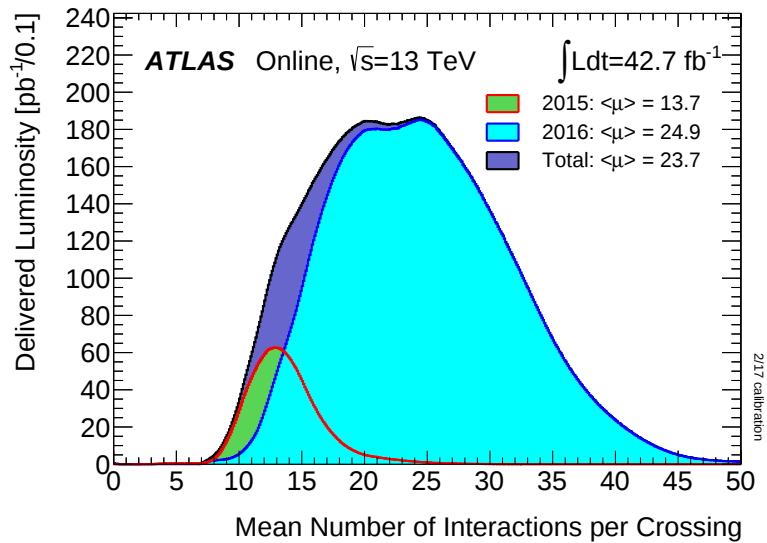


Figure 7.6: Mean number of interactions per bunch-crossing for the 2015 and 2016 pp collision data at 13 TeV centre-of-mass energy. Plot obtained from [194].

tool is provided in the ATLAS software, called *PileupReweighting tool*. This tool basically takes

the ratio of the average number of PU interactions in data and in MC and provides per-event weights to be used on simulations; this approach is designed not to modify the normalisation of the simulations. On top of its main purpose, the *PileupReweighting tool* is also designed to take into account the effect of the prescales applied to the triggers employed in the analyses. In this case the tool is designed to modify the normalisation of the simulation according to the effective collected luminosity for the selected samples; this allows to avoid the association of large weights to the simulated events.

The uncertainty associated with the PU weights, derived from the PU reweighting procedure and from the uncertainty on the collected luminosity, is also provided, to be used as a systematic uncertainty in the ATLAS analyses.

7.6.2 Muon offline efficiency weights

Muon offline efficiency weights are used to correct the detector response to muons in MC. As for the PU weights, these corrections are widely used in all ATLAS analyses which employ muons.

These weights are applied to single muons, therefore each B candidate is associated to two weights. Different sets of weights are available, based on the muon-ID working point used in each analysis.

The weights are calculated as the ratio of the offline reconstruction efficiency in data ϵ^{Data} and in simulations ϵ^{MC} [196]. The efficiencies, in turn, are evaluated using a tag-and-probe technique in both data and MC. Muon samples from $J/\psi \rightarrow \mu^+\mu^-$ and $Z \rightarrow \mu^+\mu^-$ events are employed, in order to study the efficiency for different muon p_T ranges. The samples are collected by requiring one leg of the decay (tag) to be identified as a muon that fires a single muon trigger and the second leg (probe) to be reconstructed by a system independent of the one being studied [196]. An additional selection, tuned on the event topology, is applied to reduce the background contamination.

Since muons are reconstructed using information from the ID, the calorimeters and the MS, three kinds of probes are available. ID tracks and calorimeter-tagged muons (see section 3.2.3 for more details) are used to measure the MS efficiency, while MS tracks are used to evaluate the complementary efficiency of the muon reconstruction in the ID. The final value of the efficiency is obtained combining the different efficiency measurements.

The efficiencies are evaluated in different η and p_T regions, as shown in figure 3.3 in section 3.2.3, allowing to calculate the offline efficiency weights with the same binning scheme.

As for the PU reweighting, these weights are associated with an uncertainty, to be used as a systematic uncertainty in the ATLAS analyses.

7.6.3 Muon trigger weights

Muon trigger weights account for the different response of the muon triggers in real data and simulations. Also in this case these corrections are common to the ATLAS analyses which em-

ploy muon triggers.

The evaluations of the weights is similar to the approach used to calculate the muon offline efficiency weights, taking the ratio of the trigger efficiency measured in data and in MC. Such efficiency measurements are performed for a range of different topologies and muon transverse momentum. Dimuon decays of J/ψ and Z are used to evaluate the trigger efficiency for low and moderate p_T muons respectively. Two topologies targeting semileptonic $t\bar{t}$ candidates and W +jets events are employed for the determination of the efficiency of high- p_T muons. Given the low- p_T regime of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, the most important contribution comes from the weights obtained with the $J/\psi \rightarrow \mu^+ \mu^-$ process.

The efficiencies in data and MC are evaluated with a tag-and-probe method, using single muon triggers to select the tag muon. The probe muon is obtained by reconstructing the physics process of interest, e. g. a J/ψ dimuon decay. In order to obtain the weights for dimuon triggers, the correlation between the two muons at the trigger stage should be taken into account. The correlation term becomes relevant if the opening angle between the two muons is small, therefore they might trigger the same muon chamber. Given the typical opening angle of the dimuon candidates for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, this term is found to be not relevant and therefore is not included.

Two sets of weights are derived. The first set of weights is obtained for 2015 and 2016 separately, divided in intervals of p_T^μ and η^μ . The second set of weights is obtained dividing the datasets in smaller time periods, but maintaining a coarser binning in p_T^μ and η^μ in order to preserve a manageable uncertainty. Since no difference in the effects of the two sets of weights is found, the former is chosen to be used in the analysis.

As for the other corrections described, these weights are associated to an uncertainty, to be used as a systematic uncertainty in the ATLAS analyses.

7.6.4 Kinematic corrections

The remaining set of weights are meant to correct B mesons kinematic discrepancies between data and simulations. Such discrepancies are due to cuts applied to the MC samples at the generation stage or to a non-perfect modelisation of the physics processes in simulations.

Two sets of weights are used to improve the data-MC agreement:

- **Quark Level Corrections (QLC):** account for the $(p_T(B), \eta(B))$ selection bias introduced to expedite the MC generation of b quarks; these weights are described in sub-section 7.6.4.1;
- **Data Driven Weights (DDW):** account for residual data-MC $(p_T(B), \eta(B))$ discrepancies, mostly arising from differences in the b quark phase space and hadronisation models; these weights are described in sub-section 7.6.4.2.

Both sets of weights are evaluated on a $(p_T(B), \eta(B))$ grid and only for the exclusive MC simulations, that are consistently used across the analysis after applying both QLC and DDW. The two sets of weights are meant to be used in combination in order to obtain the most faithful response from the MC, as well as a direct way of accounting for residual data-MC discrepancies

through the propagation of QLC and DDW uncertainties.

While QLC are evaluated using simulations, as explained in section 7.6.4.1, DDW are obtained with data-driven techniques. For this reason, background contributions have to be considered. A background subtraction procedure is employed, based on an additional set of weights called Sideband Weights (SBW), described in detail in section 7.6.4.2.

The full event selection shown in section 7.3 is applied to both data and MC for the SBW and DDW calculation. The simulated samples involved in the DDW calculation are corrected using the PU, offline efficiency, trigger efficiency and QLC weights. This allows one to account for all possible residual data-MC discrepancies. The DDW are evaluated for both the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels. An additional $L_{xy} > 0.3$ mm is employed for the reference channel, in both the SBW and DDW evaluation, in order to remove most of the combinatorial background, which would otherwise overtake the signal, especially in the low p_T region. This cut was shown to highly reduce the background, about 80% reduction verified on $B^+ \rightarrow J/\psi K^+$ data sidebands, while removing about 13% of the signal, verified with a study performed on MC. The $p_T(B)$ and $\eta(B)$ distributions were also found not to be critically modified by this additional cut. The $L_{xy} > 0.3$ mm selection is not employed for the control channel, as the combination of the selections applied to the J/ψ candidate and to the ϕ candidate (both described in section 7.3) allows a high reduction of the background. For this reason, the $B_s^0 \rightarrow J/\psi \phi$ sample has a better signal over background ratio compared to the reference channel sample.

These weights are exclusive to the $B_s^0 \rightarrow \mu^+ \mu^-$ analysis and are not employed in other ATLAS analyses, therefore the full description of their derivation is given in the following sections. QLC are described in sub-section 7.6.4.1, the procedure followed to evaluate the DDW is described in sub-section 7.6.4.2, together with the SBW, used to perform the background subtraction in data.

7.6.4.1 Quark Level Corrections

Signal MC simulations are generated with specific b quark kinematic cuts (quark-level cuts). These cuts are employed to enhance the MC production speed, with the drawback of a bias on the signal kinematics. The Quark Level Corrections (QLC) are introduced to correct this selection bias.

An additional truth-level selection is also applied to the kinematics of the particles present in the final state of the physics processes², in order to maximise the amount of events that pass the kinematic requirements discussed in section 7.3. These truth-level cuts, referred to as Final State (FS) cuts, are looser than the selection applied at reconstruction level and their effect on the MC simulation is overtaken by the offline selections.

Due to the different topologies of the processes considered, three sets of quark-level cuts are employed in the analysis; they are summarised in table 7.6. Charge conjugate processes, such as $B^+ \rightarrow J/\psi K^+$ and $B^\pm \rightarrow J/\psi K^\pm$, are generated separately, applying the same quark level cuts to the b quark responsible for the hadronisation of the decay of interest. Three versions

² Muons $p_T > 3.5$ GeV and muons $|\eta| < 2.6$; in case hadrons are present in the final state, the cuts applied on their kinematic properties are $p_T > 0.8$ GeV and $|\eta| < 2.6$

Table 7.6: Quark-level cuts employed in the generation of the exclusive MC sample. $\hat{p}_T \text{min}$ is the lower cut applied to the p_T of the partons produced in the hard scattering in the reference system of the incoming partons. The last two columns referred to the cuts on p_T and $|\eta|$ of the b quark responsible for the hadronisation that originates the B meson of interest.

Process	$\hat{p}_T \text{min}$	b (anti- b) $ \eta $	b (anti- b) p_T
$B_s^0 \rightarrow \mu^+ \mu^-$, $B_d^0 \rightarrow \mu^+ \mu^-$, $B \rightarrow hh'$	5 GeV	2.6	5 GeV
$B^\pm \rightarrow J/\psi K^\pm$, $B^\pm \rightarrow J/\psi \pi^\pm$, $B_s^0 \rightarrow \pi^- \mu^+ \nu$, $B_s^0 \rightarrow K^- \mu^+ \nu$, $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}$	7 GeV	2.6	7 GeV
$B_s^0 \rightarrow J/\psi \phi$	11 GeV	2.5	9 GeV

of the QLC are therefore calculated, providing a way to correct the different bias introduced by the three sets of quark-level cuts. The processes employed for the evaluation of the QLC are $B_s^0 \rightarrow \mu^+ \mu^-$, $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$.

QLC are evaluated as the ratio of yields in a given $(p_T(B), \eta(B))$ cell, relative to the signal distribution in an unbiased reference phase space volume common to all the different signal simulations. This guarantees that all efficiencies, discussed in chapter 10, are calculated relative to the same common b quark phase space volume, by design containing the kinematic acceptance of the analysis.

Each set of QLC is evaluated using two different MC samples³:

- *unbiased* MC: generated with looser quark-level cuts with respect to the default MCs, described in section 7.2, and no FS cuts;
- *quark biased*: generated with the same quark-level cuts as the default MCs and without FS cuts.

Table 7.7 shows the sets of generation cuts employed for the different MCs.

The QLC calculation is based on two samples because the $\hat{p}_T \text{min}$ cut, which is the lower cut applied to the p_T of the partons produced in the hard scattering in the reference system of the incoming partons, affects the phase space distribution of the events also above the cut itself, due to features of the PYTHIA generator [148]. A more extended discussion is provided in section B.1 of appendix B.

The computation of the QLC is performed using the unbiased and the quark biased samples according to the following formula:

$$W_{\text{QL}} = v_{\text{quarkBiased}}^{\text{FScuts}} \cdot \left(\frac{\sigma_{\text{quarkBiased}}^{\text{Pythia}}}{N_{\text{quarkBiased}}^{\text{tot}}} \right) / \left[v_{\text{unbiased}}^{\text{FScuts}} \cdot \left(\frac{\sigma_{\text{unbiased}}^{\text{Pythia}}}{N_{\text{unbiased}}^{\text{tot}}} \right) \right], \quad (7.1)$$

where v is the number of entries in a $(p_T(B), \eta(B))$ bin from the unbiased or quark biased samples after applying the final state particle cuts. σ^{Pythia} is the Pythia-calculated generation

³ Since QLC are meant to correct generator-level biases, the production of the unbiased and quark biased MCs is limited to generation, without simulation of the detector response and reconstruction.

Table 7.7: Quark-level and Final State particle cuts per MC sample. $\hat{p}_T \text{min}$ is the lower cut applied to the p_T of the partons produced in the hard scattering in the reference system of the incoming partons. Anti- b $|\eta|$ and anti- b p_T are the cuts applied to the anti- b quark produced in the hard scattering. No cuts are applied to the b quark, as it is not responsible for the production of B mesons; when generating the charge conjugate of the processes, the Anti- b and b quarks are swapped, therefore the cuts are applied only to the b quark. Muons $|\eta|$, muons p_T , final h $|\eta|$ and final h p_T are the FS cuts applied respectively to muons and hadrons.

	$\hat{p}_T \text{min}$	anti- b $ \eta $	anti- b p_T	muons $ \eta $	muons p_T	final h $ \eta $	final h p_T
default $B^+ \rightarrow J/\psi K^+$	7 GeV	2.6	7 GeV	2.6	3.5 GeV	2.6	0.9 GeV
unbiased $B^+ \rightarrow J/\psi K^+$	5 GeV	4	2.5 GeV				
quark biased $B^+ \rightarrow J/\psi K^+$	7 GeV	2.6	7 GeV				
default $B_s \rightarrow \mu^+ \mu^-$	5 GeV	2.6	5 GeV	2.6	3.5 GeV		
unbiased $B_s \rightarrow \mu^+ \mu^-$	5 GeV	4	2.5 GeV				
quark biased $B_s \rightarrow \mu^+ \mu^-$	5 GeV	2.6	5 GeV				
default $B_s \rightarrow J/\psi \phi$	11 GeV	2.5	9 GeV	2.6	3.5 GeV	2.6	0.9 GeV
unbiased $B_s \rightarrow J/\psi \phi$	5 GeV	4	2.5 GeV				
quark biased $B_s \rightarrow J/\psi \phi$	11 GeV	2.5	9 GeV				

cross-section; this quantity is evaluated by PYTHIA and encloses the production cross-section evaluated with the Matrix Element (ME), the decay cross-section and the effects of the quark level and FS cuts. Finally, N^{tot} is the number of generated events for the given samples⁴. Effectively the QLC are correcting not only the $p_T(B) - \eta(B)$ shape distribution of the events, but also their normalisation in order to account for the specific generator-level cuts chosen.

The inverse of these weights is used to weight events individually, thus correcting with event-weights the quark level cut biases.

The computation of the QLC is limited to the B fiducial volume of the analysis ($p_T(B) > 8$ GeV and $|\eta(B)| < 2.5$); the binning scheme adopted is not regular, but rather chosen to maintain a comparable statistical uncertainty in every bin. In order to ensure that such a binning scheme allows to properly correct the kinematic bias, a complementary binning has also been tested; it is discussed in section B.2.4 of appendix B. The B meson η spectrum is assumed to be symmetric with respect to 0, allowing a reduction on the uncertainty associated to the corrections; the absence of any $\eta(B)$ asymmetry has been tested by calculating the QLC without folding the distribution, as reported in section B.2.5 of appendix B.

Figure 7.7 shows a 2D ($p_T(B)$, $\eta(B)$) map of the QLC and their uncertainties for the three processes $B^+ \rightarrow J/\psi K^+$, $B_s \rightarrow J/\psi \phi$ and $B_s \rightarrow \mu\mu$. The QLC consistency is checked for the three sets of weights applying QLC calculated using odd-numbered events, from both the quark biased and unbiased samples, to the even-numbered events of the quark biased sample. The resulting $p_T(B)$ and $\eta(B)$ distributions are compared to the unbiased distributions obtained using only even-numbered events. The same test is repeated switching even- with odd-numbered events. The outcome of all these checks is reported in appendices B.2.1 for $B^+ \rightarrow J/\psi K^+$, B.2.2 for $B_s^0 \rightarrow J/\psi \phi$ and B.2.3 for $B_s^0 \rightarrow \mu^+ \mu^-$: all the distributions show good compatibility.

⁴ The $\frac{\sigma^{\text{Pythia}}}{N^{\text{tot}}}$ terms in equation 7.1 effectively account for the relative “integrated luminosity” of the two MC samples considered.

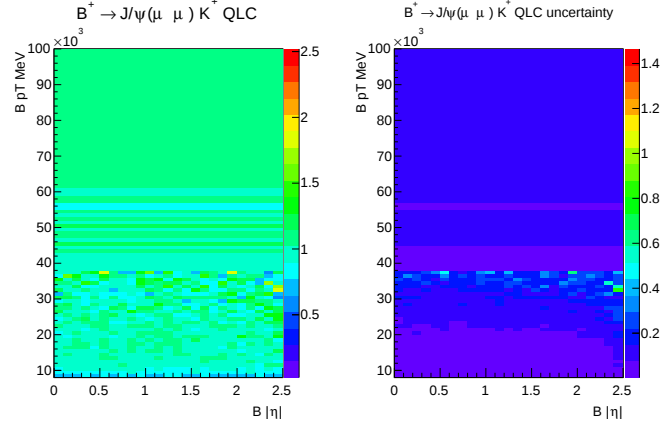
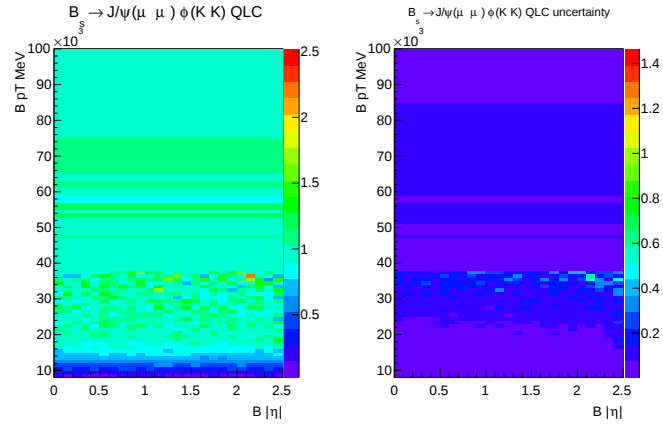
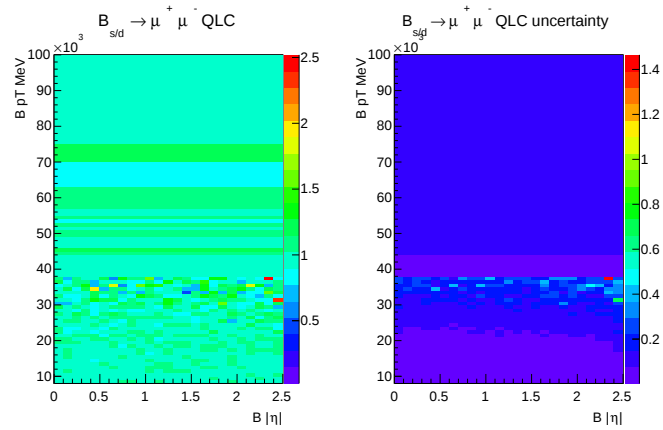
(a) $B^+ \rightarrow J/\psi K^+$ QLC (left) and the corresponding uncertainty (right).(b) $B_s \rightarrow J/\psi \phi$ QLC (left) and the corresponding uncertainty (right).(c) $B_s \rightarrow \mu\mu$ QLC (left) and the corresponding uncertainty (right).

Figure 7.7: The three sets of plots show the QLC (left plot) and the relative uncertainty (right plot), calculated for the three processes, respectively, $B^+ \rightarrow J/\psi K^+$ figure 7.7(a), $B_s \rightarrow J/\psi \phi$ figure 7.7(b) and $B_s \rightarrow \mu\mu$ figure 7.7(c).

7.6.4.2 Data Driven Weights

Data Driven Weights (DDW) aim to account for any residual data-MC discrepancy, after the simulation is tuned at its best applying all the appropriate prescriptions. In particular, all the PU, offline efficiency, trigger efficiency and QLC corrections are applied to the simulated samples used in the calculations of the DDW.

The calculation of the corrections is performed using both real and simulated samples, as mentioned before. In order to perform a meaningful comparison, background contributions need to be removed from the data. A sideband subtraction technique is therefore applied to both data and MC, in order to remove the background contribution for the first and account for possible signal subtractions due to signal leakage in the mass sidebands for the latter. A description of the SBW employed in the background subtraction is provided later in this sub-section.

DDW are evaluated for both the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels. Despite the difference in principle, the B^+ DDW are found to be interchangeable with the $B_s^0 \rightarrow J/\psi \phi$ ones, as was the case for the Run 1 analysis. This is verified by calculating the DDW for the $B_s^0 \rightarrow J/\psi \phi$ process and comparing them with the B^+ DDW. As shown later in this sub-section, the resulting weights are compatible within statistical uncertainty: the DDW corrections are therefore dominated by data-MC differences common to B^+ and B_s^0 , most likely due to b quark kinematics, while differences due to the fragmentation function, which is different for B^+ and B_s^0 , are less important and cannot be resolved with the current statistics. The compatibility of the two sets of DDW confirms also that the additional $L_{xy} > 0.3$ mm cut applied to the $B^+ \rightarrow J/\psi K^+$ samples does not affect the kinematic distributions of the B meson.

The DDW consistency between $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ can be considered also as a consistency check on the QLC: since the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ MCs are generated with different kinematic biases, any mistake in the QLC on either sample would reflect in an inconsistency between the two sets of DDW.

In order to avoid possible correlations between data and MC after the reweighting procedure, only even numbered MC and data events are used in the DDW calculation. The remaining data and MC events are respectively used for the B^+ yield fit and efficiency and acceptance calculation.

Due to the limited statistics available for the $B_s^0 \rightarrow J/\psi \phi$ and $B^+ \rightarrow J/\psi K^+$ datasets, DDW are not computed in a two-dimensional grid of $p_T(B)$ and $\eta(B)$. Instead, they are obtained with an iterative method: $p_T(B)$ and $\eta(B)$ corrections are calculated separately and applied to the MC as if they were independent corrections. The resulting $p_T(B)$ and $\eta(B)$ distributions are then used to evaluate again $p_T(B)$ and $\eta(B)$ corrections, which are applied again to the MC. This is repeated until the corrections become compatible with 1, showing that the procedure converged. This approach was shown to converge in a few iterations in the Run 1 analysis. The convergence will be verified also with the current datasets and shown later.

The following formulas are therefore used in the evaluation of the DDW:

$$W_{p_T}^1 = \frac{\sum_{\eta} D_{\eta(B), p_T(B)}}{\sum_{\eta(B), p_T(B)} D_{\eta(B), p_T(B)}} \frac{\sum_{\eta(B), p_T(B)} MC_{\eta(B), p_T(B)}}{\sum_{\eta} MC_{\eta(B), p_T(B)}} \quad (7.2)$$

$$W_{p_T}^n = \frac{\sum_{\eta} D_{\eta(B), p_T(B)}}{\sum_{\eta(B), p_T(B)} D_{\eta(B), p_T(B)}} \frac{\sum_{\eta(B), p_T(B)} \left(MC_{\eta(B), p_T(B)} \prod_{m=1}^{m=n-1} W_{p_T}^m W_{\eta}^m \right)}{\sum_{\eta} \left(MC_{\eta(B), p_T(B)} \prod_{m=1}^{m=n-1} W_{p_T}^m W_{\eta}^m \right)}, \quad (7.3)$$

where equation 7.2 is the formula for the first iteration, which allows to extract the $W_{p_T}^1$ weights, and equation 7.3 is the formula for the n_{th} iteration, which allows to evaluate the $W_{p_T}^n$ weights. $MC_{\eta(B), p_T(B)}$ refers to a particular $\eta(B), p_T(B)$ bin of the MC sample, weighted as described at the beginning of this section, while $D_{\eta(B), p_T(B)}$ refers to a $\eta(B), p_T(B)$ bin of the data distribution.

The corresponding formulas for the calculation of the η weights can be obtained from equations 7.2 and 7.3 by exchanging the η and p_T indices.

The combined η and p_T weights are obtained multiplying the corresponding weights obtained at each iteration. Assuming a convergence after iteration N , the weights are:

$$\begin{aligned} W_{p_T} &= \prod_{m=1}^{m=N} W_{p_T}^m \\ W_{\eta} &= \prod_{m=1}^{m=N} W_{\eta}^m. \end{aligned} \quad (7.4)$$

The final weights used in the analysis are obtained multiplying the η and p_T weights according to

$$W_{DDW}(p_T, \eta) = W_{p_T} \cdot W_{\eta} \quad (7.5)$$

Due to the elaborate formulas employed, the calculation of the uncertainty associated to the DDW is rather complex, as one has to consider the correlation of the different iterations with the event counts in all the data and MC bins together with the propagation of the sideband subtraction weights' uncertainties. For this reason, an ad-hoc framework was developed, able to perform automatically the error propagation, taking into account all the aspects discussed earlier.

The binning scheme chosen for the DDW follows the one used in the Run 1 analysis, which was proven to be dense enough to properly correct data-MC discrepancies, while maintaining a manageable uncertainty on the weights. Additionally, the DDW calculated for the 2015/16 analysis can be compared with the ones from the Run 1 analysis, as an additional check.

Before proceeding to the actual DDW extraction, the Sideband Weights (SBW) must be introduced and evaluated. The following sub-section provides an extensive description of these weights.

Sideband Weights

The number of signal events in each $(\eta(B), p_T(B))$ bin in data for the DDW calculation is obtained through a sideband subtraction procedure.

The Sideband Weights (SBW) are per event weights defined as 1 if the mass of the B candidate falls into a defined region around the mass peak (signal region). If this is not the case then the

event acquires a negative weight proportional to the background fraction in the region, with the SBW defined as:

$$W_{SBW} = -1 \cdot \frac{\int_{\text{signal region}} \text{PDF}_{bkg}}{\int_{\text{left sideband}} \text{PDF}_{bkg} + \int_{\text{right sideband}} \text{PDF}_{bkg}} \quad (7.6)$$

where PDF_{bkg} is the PDF associated with the background. This PDF is estimated with a binned extended maximum likelihood fit on the mass distribution of the B meson in different $p_T(B)$ and $\eta(B)$ bins.

Unlike the QLC, the DDW are extracted in the full $[-2.5; 2.5]$ $\eta(B)$ interval, as their distribution shows an asymmetry with respect to $\eta(B)$, therefore the SBW are calculated in the same full $\eta(B)$ range. The binning scheme employed is chosen to maintain a comparable statistical uncertainty for the SBW. This binning scheme is coarser than the one employed for the DDW.

The SBW are calculated separately for the $B_s^0 \rightarrow J/\psi\phi$ and $B^+ \rightarrow J/\psi K^+$ processes, and discussed in the following paragraphs; the fits on the data are modified accordingly to the features of the datasets and the definition of the signal and sideband regions is modified according to the position and the width of the mass peak.

Control channel SBW

For the $B_s^0 \rightarrow J/\psi\phi$ process the signal and sideband regions are defined as:

- signal region: [5286, 5456] MeV
- left sideband: [5201, 5286] MeV
- right sideband: [5456, 5541] MeV.

A visual representation of these regions is provided later in figure 7.9.

The fit on the data is performed using the following configuration:

- the signal is described with two Gaussians with equal mean;
- the continuum background is described with a third order Chebychev polynomial.

All shape and amplitude parameters are extracted from the fit. Systematic uncertainties on the fits employed are evaluated by varying the functional model for the background and checking the effect on the SBW. They are found to be small compared to the statistical uncertainty, therefore systematic effects are not included.

The SBW are calculated separately in $p_T(B)$ and $\eta(B)$ bins. Figure 7.9 shows some of the fits performed to extract the $B_s^0 \rightarrow J/\psi\phi$ yield and evaluate the SBW, together with vertical lines which show the signal and sideband regions. The result of the SBW evaluation is shown in figure 7.8.

The $\eta(B)$ SBW show compatibility among each other, confirmed by a linear fit (p -value $\sim 56\%$). For the $B_s^0 \rightarrow J/\psi\phi$ process the sideband subtraction procedure is therefore within the available statistics, independent of the specific η bin chosen, so only the p_T SBW are used in the DDW calculation.

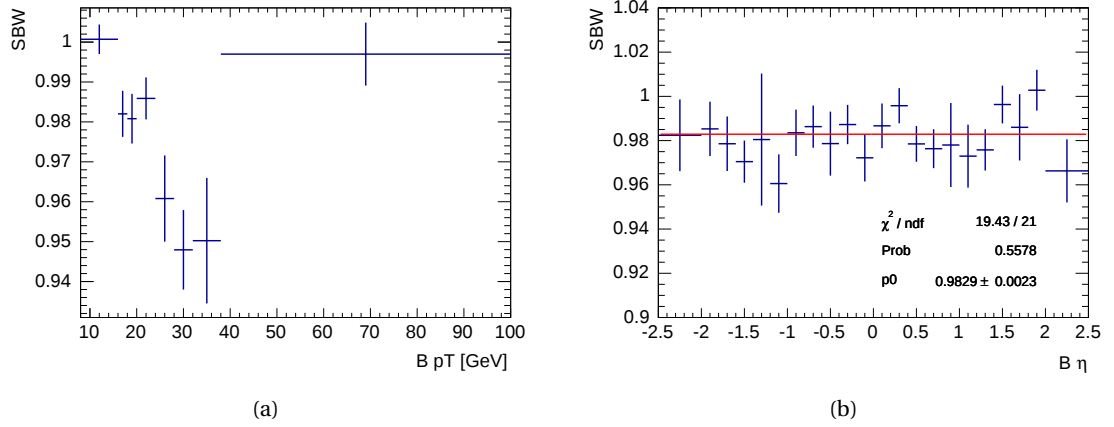


Figure 7.8: SBW calculated for the $B_s^0 \rightarrow J/\psi \phi$ process. The SBW have been calculated separately in $p_T(B)$ 7.8(a) and $\eta(B)$ 7.8(b) bins. The dependence of the SBW on $\eta(B)$ is checked with a linear fit on the SBW, that shows a good compatibility with the histogram (p-value $\sim 56\%$) proving that the SBW for this process do not depend on η .

Reference channel SBW

Signal and sideband regions for the $B^+ \rightarrow J/\psi K^+$ process are defined as:

- signal region: [5182.96, 5382.96] MeV
- left sideband: [5082.96, 5182.96] MeV
- right sideband: [5382.96, 5482.96] MeV.

As for the $B^+ \rightarrow J/\psi K^+$ SBW, a visual representation of these regions is provided later in figure 7.11.

The fit on the data is performed using the following configuration:

- the signal is described with two Gaussians with equal mean;
- the continuum background is described by an exponential;
- the background due to partially reconstructed decays ($B^+ \rightarrow J/\psi X$, with $m(J/\psi h^+) < 5.200$ GeV) is described with a complementary error function. Such a functional form was proven to properly described this background in the Run 1 analysis and it is used as functional form for this background in the systematic uncertainty evaluation of the reference channel yield (section 10.1.3).

All shape and amplitude parameters are extracted from the fit. Possible systematic uncertainties are considered and checked by varying the functional models employed for the background. The variation of the SBW is found to be much smaller than the statistical uncertainty, therefore systematic effects are not included.

The $B^+ \rightarrow J/\psi K^+$ SBW were initially evaluated separately in $p_T(B)$ and $\eta(B)$ bins, as is done for the $B_s^0 \rightarrow J/\psi \phi$ SBW, but a dependence of the SBW on $\eta(B)$ was found. This $\eta(B)$ dependence is

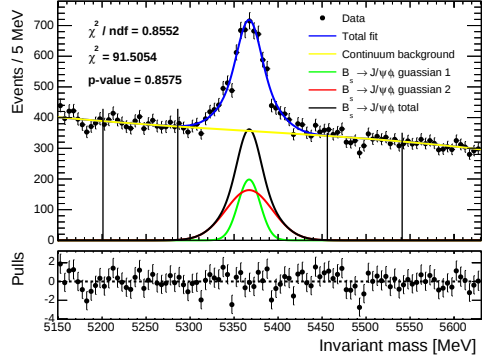
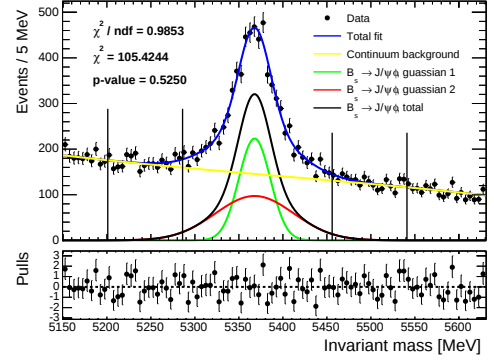
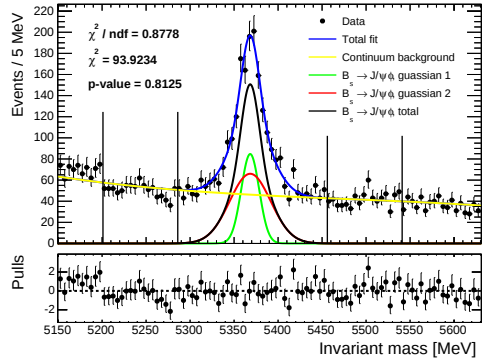
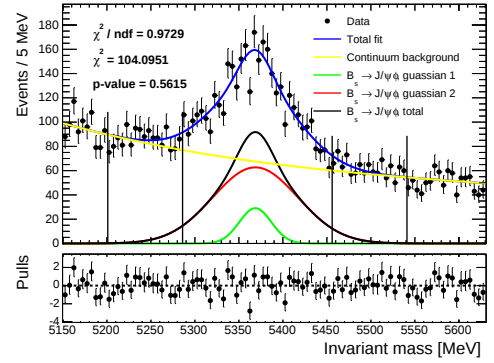
(a) Fit performed on the $8 \text{ GeV} < p_T < 16 \text{ GeV}$ bin.(b) Fit performed on the $38 \text{ GeV} < p_T < 100 \text{ GeV}$ bin.(c) Fit performed on the $0 < \eta < 0.2$ bin.(d) Fit performed on the $2.0 < \eta < 2.5$ bin.

Figure 7.9: Some of the fits performed on the $B_s^0 \rightarrow J/\psi\phi$ sample. Each plot shows the data points (black dots) and the PDFs used to perform the fit: the green and red lines are the two gaussians for the signal PDF, while the black line is the total signal PDF, the yellow line is the PDF for the combinatorial background and the blue line is the total PDF. Each plot shows also the compatibility of the data with the fitted curve and the pulls are shown in the lower frame in each plot. The vertical lines indicate the sideband and signal regions.

shown in figure 7.10, where the SBW are evaluated folding the $\eta(B)$ distribution with respect to zero. Three linear fits are performed on the $\eta(B)$ SBW. The fit performed on the full X axis (red line) range has a low compatibility with the data (p-value $\sim 0.00\%$); two other fits are therefore performed in the two η regions that show a different behaviour. The compatibility with the SBW of these two fits is much higher than the previous (p-value $\sim 14.6\%$ for $\eta < 1.3$, green line, and p-value $\sim 24.9\%$ for $\eta > 1.3$, blue line). The full $\eta(B)$ interval is hence divided in four regions where the SBW are compatible, and the B^+ SBW are extracted in combined $p_T(B) - \eta(B)$ bins: 11 p_T bins and 4 η bins are employed.

Figure 7.11 shows some of the fits performed to extract the B^+ yield and evaluate the SBW. The result of the SBW extraction is shown in figure 7.12.

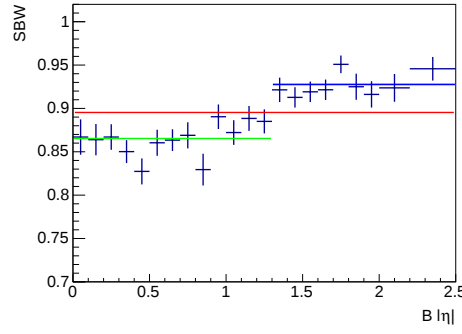
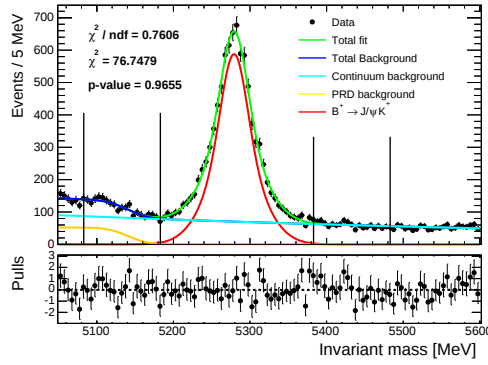
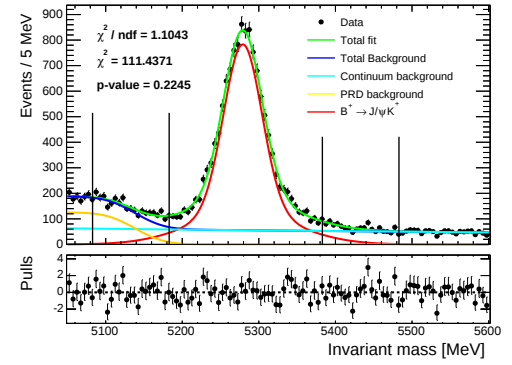


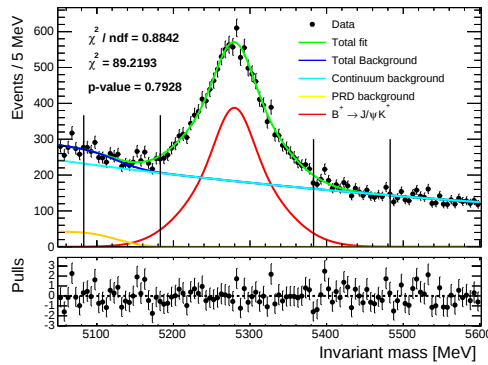
Figure 7.10: $\eta(B)$ SBW calculated for the $B^\pm \rightarrow J/\psi K^\pm$ process folding the $\eta(B)$ distribution with respect to zero. The dependence of the SBW on $\eta(B)$ is checked with a linear fit on the SBW, and it does not show a good compatibility with the histogram (red line, p-value $\sim 0.00\%$). Two other fits are performed, dividing the η range in two, and they show a much higher compatibility, p-value $\sim 13.6\%$ for $\eta < 1.3$, green line, and p-value $\sim 24.9\%$ for $\eta > 1.3$, blue line. The SBW extraction is therefore performed in combined pT- η bin.



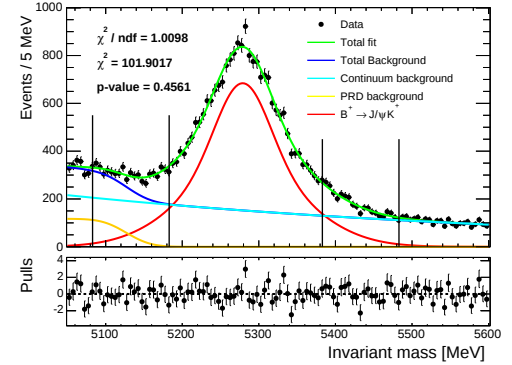
(a) Fit performed on the $0 < \eta < 1.3$, $8 \text{ GeV} < p_T < 13 \text{ GeV}$ bin.



(b) Fit performed on the $0 < \eta < 1.3$, $34 \text{ GeV} < p_T < 100 \text{ GeV}$ bin.



(c) Fit performed on the $1.3 < \eta < 2.5$, $8 \text{ GeV} < p_T < 13 \text{ GeV}$ bin.



(d) Fit performed on the $1.3 < \eta < 2.5$, $34 \text{ GeV} < p_T < 100 \text{ GeV}$ bin.

Figure 7.11: Some of the fits performed on the $B^+ \rightarrow J/\psi K^+$ sample. Each plot shows the data points (black dots) and the PDFs used to perform the fit: the red line is the signal PDF, the light blue line is the PDF for the combinatorial background, the yellow line is the PDF for the partially reconstructed decays, the blue line is the total background PDF and the green line is the total PDF. Each plot shows also the compatibility of the data with the fitted curve and the pulls are shown in the lower frame in each plot. The vertical lines indicate the sideband and signal regions.

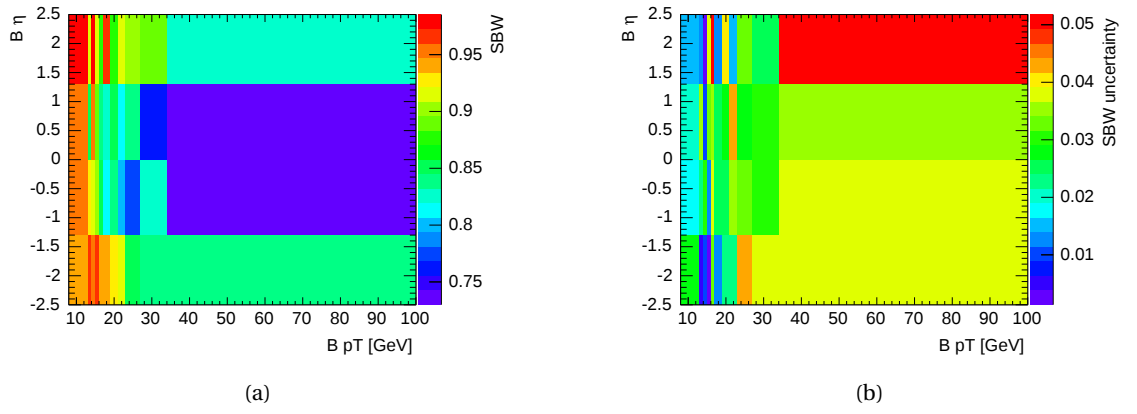


Figure 7.12: SBW calculated for the $B^+ \rightarrow J/\psi K^+$ process 7.12(a) and their uncertainty 7.12(b).

Data Driven Weights extraction

Now that the SBW have been calculated, the DDW can finally be extracted. Figure 7.13 shows

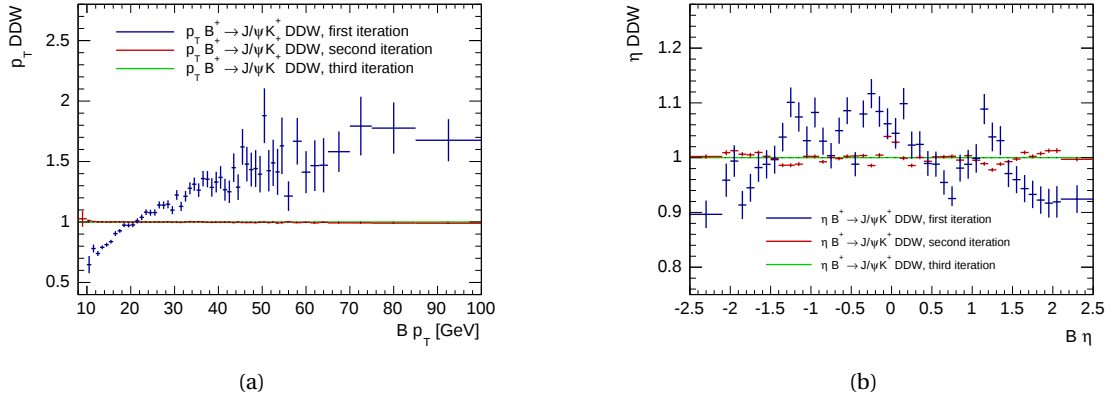


Figure 7.13: First three iterations of the $B^+ \rightarrow J/\psi K^+$ DDW calculation. Figure 7.13(a) shows the p_T DDW, while 7.13(b) shows the η DDW. The first bin of the first iteration in figure 7.13(a) is omitted due to its huge uncertainty, allowing to properly read the remaining weights. The plots clearly show the convergence of the procedure used to calculate the DDW. The first iteration of the weights (blue histogram) provides the core of the correction, as most of the points are away from 1. The second iteration (red histogram) already shows values really close to one, in particular for the p_T weights. Finally, the third iteration (green histogram) is compatible with one. The uncertainties associated to the weights (vertical bars) are evaluated considering the correlation of the different iterations with the event counts in all the data and MC bins together with the propagation of the sideband subtraction weights uncertainties.

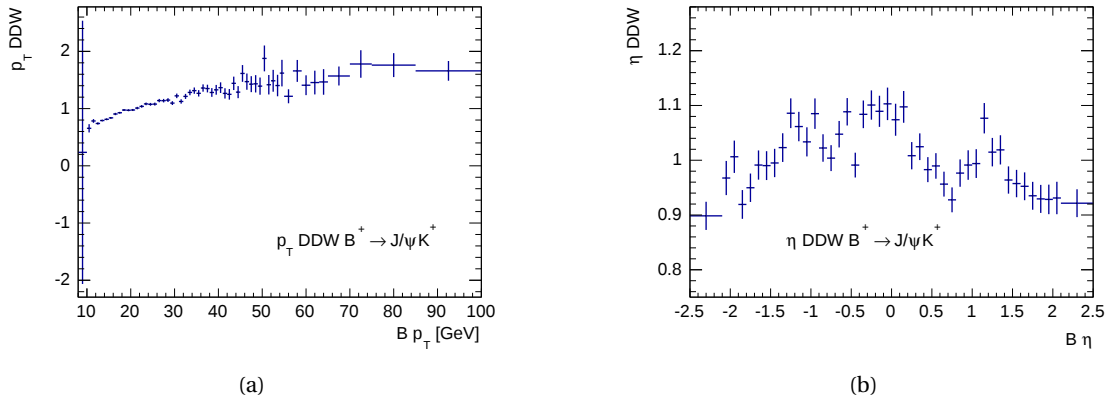


Figure 7.14: Final values of the $B^+ \rightarrow J/\psi K^+$ DDW. The weights are obtained combining the first two iterations shown in figure 7.13, while the third iteration, which is compatible with one, is not employed. In this case the first bin of 7.13(a) is shown.

the first three iterations for the p_T and η weights for the $B^+ \rightarrow J/\psi K^+$ process; the procedure converges at the second iteration, as the weights obtained with the third iteration are compatible with 1. The resulting DDW, obtained combining the first and second iterations are shown in figure 7.14.

The first p_T weight shows a huge uncertainty, due to the low statistics available in the $B^+ \rightarrow J/\psi K^+$ dataset for $p_T(B) < 10$ GeV. This issue did not affect the Run 1 analysis, which was based on a mixture of several triggers, among which were lower momentum ones ($p_T > 4$ GeV for both muons), and the collected statistics was proven to be enough to calculate the first p_T DDW.

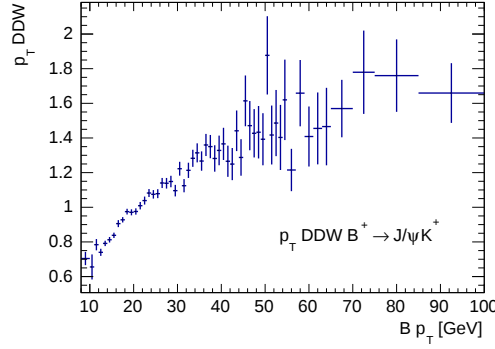


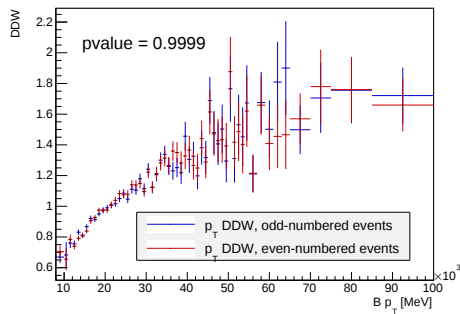
Figure 7.15: Final $p_T B^+ \rightarrow J/\psi K^+$ DDW, the first bin has been replaced with the first bin of the DDW calculated with the $2\mu 4$ trigger in order to reduce the bin's uncertainty.

In order to reduce the impact of this bin's uncertainty, data collected using a different dimuon trigger with p_T thresholds of 4 GeV for both muons (referred to as $2\mu 4$), are used to derive the data-driven weight for the first $p_T(B)$ bin and its uncertainty. This trigger was heavily prescaled in 2015 and disabled in 2016.

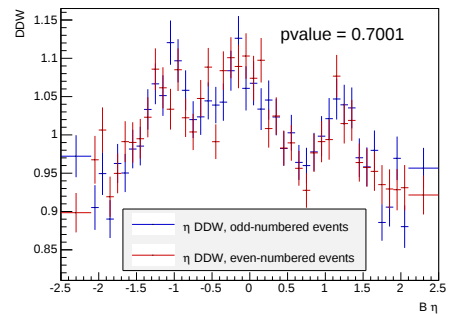
The same procedure for the calculation of the DDW is applied to the $2\mu 4$ sample, in this case changing the trigger and offline muons requirements (p_T thresholds of 4 GeV instead of 4 and 6 GeV). The calculation converges again at the second iteration, and the weights obtained are compatible with the ones shown before.

The first p_T weight calculated with the $2\mu 4$ based dataset is therefore used as the first bin of the DDW used in the analysis. Figure 7.15 shows the resulting set of p_T DDW.

The first consistency check is performed by calculating the DDW using odd-numbered MC and data events and comparing them to the DDW evaluated with even-numbered events. The two sets of weights are superimposed and their compatibility is quantified with a χ^2 test. Figure 7.16 shows the comparison of the two sets of weights, together with the result of the χ^2 test. Both



(a) $B^+ \rightarrow J/\psi K^+$ p_T DDW comparison.



(b) $B^+ \rightarrow J/\psi K^+$ η DDW comparison.

Figure 7.16: Comparison of the $B^+ \rightarrow J/\psi K^+$ DDW calculated using odd-numbered data and MC events with the DDW calculated using even numbered events. The compatibility is quantified with a χ^2 test, which gives a p -value of 0.9999 for the p_T DDW (figure 7.16(a)) and 0.7001 for the η DDW (figure 7.16(b)).

the p_T and η DDW show high compatibility.

The DDW consistency is also checked in the same way as the QLC, by applying DDW calcu-

lated using odd (even) MC and data events to the even (odd) MC events and comparing the result with the even (odd) numbered data events. All these tests, presented in appendix B.3.1, show compatibility between the weighted MC and the data.

DDW for the $B_s^0 \rightarrow J/\psi\phi$ process are calculated using the same approach. The binning scheme employed is the same as the one used for the $B^+ \rightarrow J/\psi K^+$ corrections, in order to allow a comparison between the two sets of weights. Given the smaller size of the datasets used in the calculation, larger statistical uncertainties are expected. Also in this case the procedure converges at the second iteration. The DDW consistency is checked in the same way as the $B^+ \rightarrow J/\psi K^+$

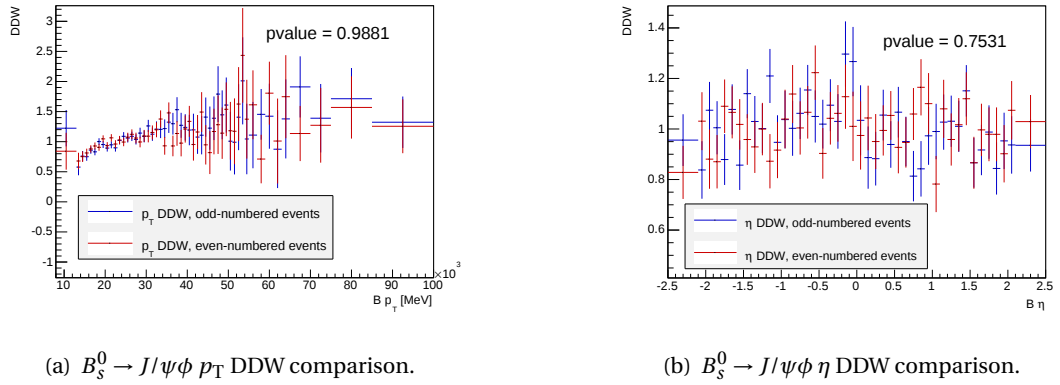


Figure 7.17: Comparison of the $B_s^0 \rightarrow J/\psi\phi$ DDW calculated using odd-numbered data and MC events with the DDW calculated using even numbered events. The compatibility is quantified with a χ^2 test, which gives a p -value of 0.9881 for the p_T DDW (figure 7.17(a)) and 0.7531 for the η DDW (figure 7.17(b)).

DDW, first comparing the DDW calculated using odd-numbered MC and data events with the DDW evaluated with even-numbered events (figure 7.17); the two sets of weights show a good compatibility. The $B_s^0 \rightarrow J/\psi\phi$ DDW are also tested by applying DDW calculated using odd (even) MC and data events to the even (odd) MC events and comparing the result with the even (odd) numbered data events, obtaining a good compatibility between the weighted MC and the data. As before, the results of the checks are shown in appendix B.3.2.

Figure 7.18 compares the $B_s^0 \rightarrow J/\psi\phi$ DDW with the $B^+ \rightarrow J/\psi K^+$ DDW. The $B_s^0 \rightarrow J/\psi\phi$ DDW have a larger uncertainty, as expected; the first p_T bins are merged, in order to avoid large uncertainties or issues in the calculations due to empty bins⁵. Since the two sets of weights show compatibility, as already mentioned, the DDW calculated with the $B^+ \rightarrow J/\psi K^+$ datasets are used to correct also the B_s^0 MCs.

As introduced in section 7.6.4.2, a comparison with the DDW employed in the Run 1 analysis is also performed. It is reported in section B.3.3 of appendix B.

The ad-hoc framework developed to properly calculate the DDW taking into account the un-

⁵ The $B_s^0 \rightarrow J/\psi\phi$ sample presents a harder B meson p_T spectrum compared to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$. This is due to the topology of the decay and the selection cuts applied: in order to have more particles in the final state with the same p_T as the particles in the other processes, the B meson must have a higher p_T .

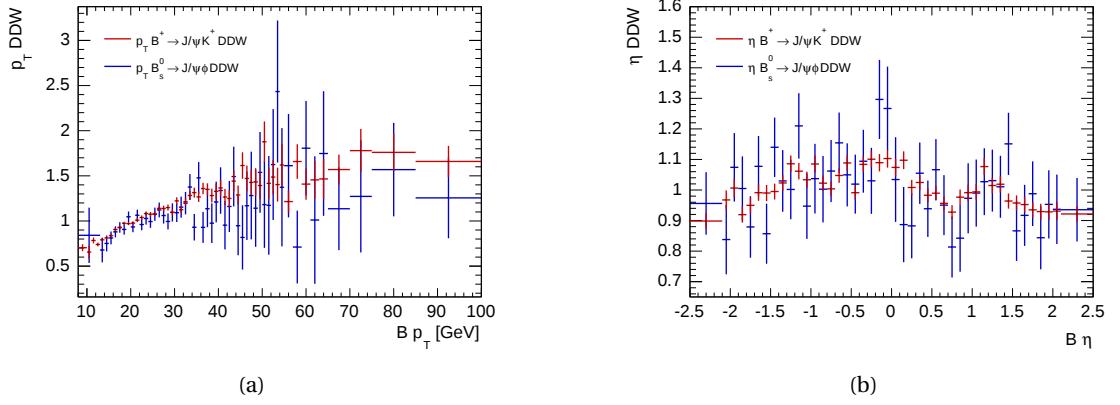


Figure 7.18: Comparison of the DDW calculated with the $B^+ \rightarrow J/\psi K^+$ datasets (red) and with the $B_s^0 \rightarrow J/\psi \phi$ datasets (blue). The plot on the left shows the p_T DDW 7.18(a), while the plot on the right shows the η DDW 7.18(b). The uncertainty of the $B_s^0 \rightarrow J/\psi \phi$ DDW is visibly larger with respect to the $B^+ \rightarrow J/\psi K^+$ weights due to the limited size of the datasets used. The two sets of weights are compatible within their uncertainty.

certainties and the correlations of the various ingredients needed performs also the evaluation of the correlation matrix of the DDW. It is shown in figure 7.19, where the diagonal is set to 0 to make the plot more readable. The main feature noticeable in the correlation matrix is a “block structure”, where groups of DDW show a higher correlation. This is due to the looser binning scheme of the SBW with respect to the DDW, where care is taken in matching the boundaries of the bins of both sets of weights. The block structure corresponds, in fact, to the binning scheme of the SBW. The relative uncertainty between the DDW and the SBW acquires importance in this condition. For the low p_T DDW, the relative uncertainty of the two sets of weights is comparable, therefore higher correlation effects appear (up to 30%). With increasing p_T the uncertainty of the DDW becomes larger than the uncertainty of the SBW, therefore the correlation effect introduced by the latter becomes less relevant.

Due to their size, the correlation terms are taken into account and propagated in the systematic uncertainty evaluation throughout the analysis. A similar study is performed to assess the correlation between the QLC and the DDW. This is found to be negligible with respect to the correlations among the DDW, and therefore can be omitted in the determination of the systematic uncertainties.

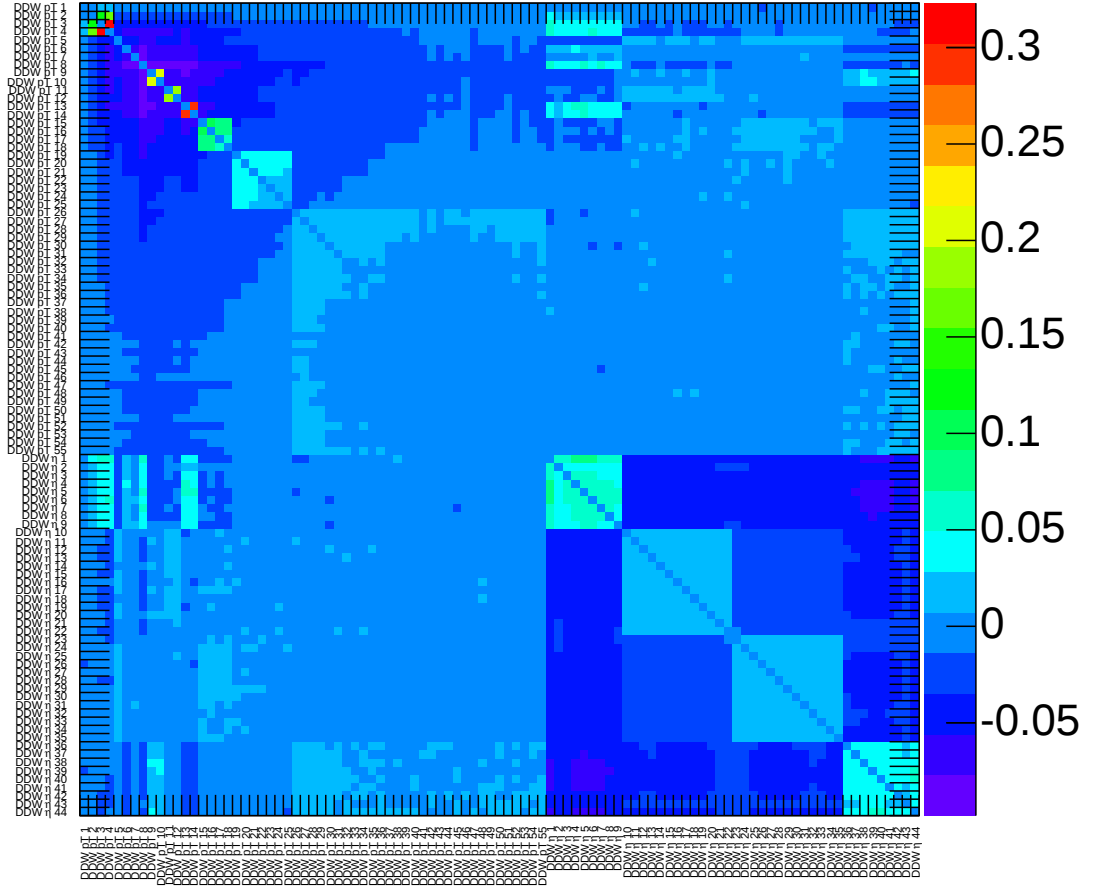


Figure 7.19: Correlation matrix of the DDW calculated for the $B^+ \rightarrow J/\psi K^+$ process. The diagonal has been set to 0 to improve readability. The height of each bin shows the correlation between two DDW. As introduced in the text, most of the weights have small correlation terms, but few of them show an exceptionally large correlations, up to 30%.

BACKGROUND SOURCES AND REDUCTION

8

After the application of the selection described in chapter 7 the dimuon mass region considered in the analysis ($4766 \text{ MeV} < m(B) < 5966 \text{ MeV}$) contains about 3.5×10^6 candidates, of which about a third fall into the signal region. This huge amount of background events is compared to an expected total signal yield of about 200 events¹, considering the sum of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ channels. A further reduction of the background contribution is therefore needed.

Several background categories can be identified in the dimuon mass region, depending on their source and their mass dependence. They are discussed in section 8.1.

In order to enhance the discrimination against the main background source of the analysis, a MVA technique based on a Boosted Decision Tree (BDT) algorithm is employed. Section 8.2 is devoted to a description of the BDT implementation.

8.1 Main analysis backgrounds

This section provides a description of the main categories of background identified in the dimuon mass region considered in the analysis. Real muon pairs are one of the primary background sources in this analysis. Out of these, only non-prompt pairs are relevant, due to lifetime-related selections exploited to isolate the signal effectively, discussed later in section 8.2. This leaves essentially only dimuons coming from B mesons, out of which two main contributions can be distinguished: the first, rapidly falling below the B meson masses, is due to dimuons produced in the decay chain of one single hadronised b quark. The second contribution, commonly referred to as *continuum* background, has the two muons originating from distinct b quarks in a $b\bar{b}$ pair.

Dimuon pairs with at least one fake muon constitute another source of background for the analysis. Two main sources can be identified also in this case. The first, characterised by the

¹ Expected signal yield estimation based on the SM expected yield of the Run 1 analysis (41 B_s^0 events and 5 B_d^0 events). After the preselection a BDT selection with signal efficiency of 54% was applied, therefore the total expected signal yield of the Run 1 analysis after the preselection is about 100 events. Considering the two-fold increase in signal statistics estimated in section 7.1 for the 2015/16 analysis and the fact that the preselection employed is extremely similar to the Run 1 counterpart, an expected total signal yield of about 200 events after the preselection is estimated.

presence of one fake muon, is mainly due to semileptonic b decays. This background presents similar features as the one coming from the decay chain of one single hadronised b quark. The second source of background is due to events with two fake muons. The main contributors are $B \rightarrow hh'$ process, where $h = \pi, K$. As introduced in chapter 5, these decays present the same topology and kinematic properties as the signal processes.

These sources of backgrounds are categorised into three groups, based on the source of the background events and their mass dependence. The following sub-sections provide an overview of these three background categories: continuum (sub-section 8.1.1), $B \rightarrow hh'$ (sub-section 8.1.2) and Partially Reconstructed Decays (PRD), which encloses the dimuons coming from the decay chain of one single hadronised b quark and the semileptonic b decays (sub-section 8.1.3).

8.1.1 Continuum background

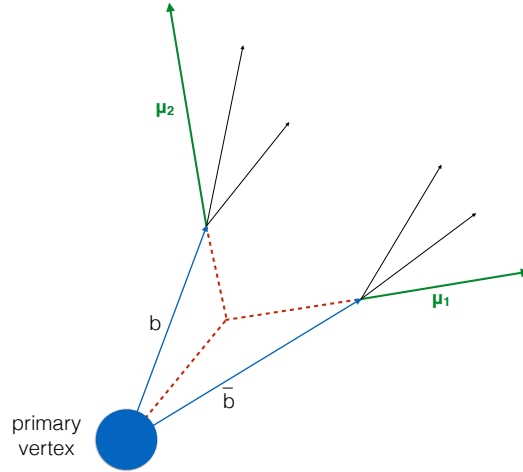


Figure 8.1: Sketch of a continuum background event. The two muons are originated from two decay chains, originated by the b quarks in a $b\bar{b}$ pair. The trajectory of the two muons shows an intersection point, which, at reconstruction level, can be taken as the decay vertex of a B candidate.

The main source of background for the analysis is due to real dimuon events originating from distinct b quarks in a $b\bar{b}$ pair. A common Secondary Vertex (SV) can be identified for the trajectory of the two muons, which are therefore identified as a B candidate.

As already mentioned, this background is called continuum background. A sketch of the typical topology of these dimuon events is shown in figure 8.1. Due to the huge amount of continuum background events, which constitutes more than 99% of the B candidates in the mass sidebands, a dedicated selection is designed and applied to the datasets. As introduced at the beginning of this chapter, it is based on a Boosted Decision Tree (BDT) algorithm and described in section 8.2.

This selection, in particular, removes all the prompt sources of dimuon events. For this reason, the large $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC, introduced in section 7.2, does not contain most of the prompt background sources and events due to $c\bar{c} \rightarrow \mu^+ \mu^- X$ processes.

8.1.2 $B \rightarrow hh'$ background

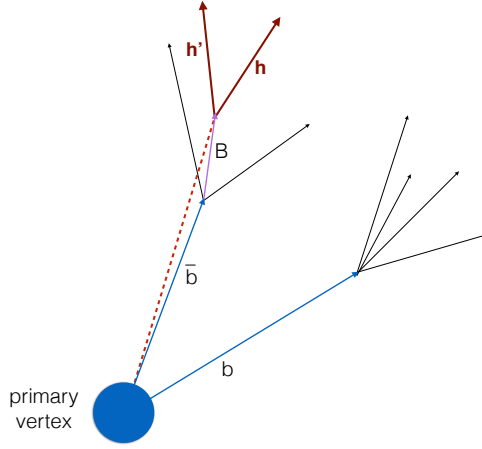


Figure 8.2: Sketch of the peaking background. This decay shows large similarities with the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ processes, making it a problematic background.

Charmless two-body $B \rightarrow hh'$ decays, where $h = \pi, K$ and both hadrons are mis-identified² as muons constitute an important source of background. Since this is the only background which presents a resonant structure, it is called *peaking background*.

This background is particularly problematic, because due to the mass distortion of the $K \rightarrow \mu$ and $\pi \rightarrow \mu$ assignments, the invariant mass distribution of the three main processes, $B_d^0 \rightarrow \pi^\pm K^\mp$, $B_s^0 \rightarrow K^\pm K^\mp$ and $B_d^0 \rightarrow \pi^+ \pi^-$, is basically superimposed on the B_d^0 peak. The invariant mass distribution of remaining processes is not superimposed on the B_d^0 peak, but considering their small branching fraction compared to the three processes mention before, their effect is negligible. Anyway, the presence of these processes is accounted for in these studies.

Section 7.4 describes the studies performed on MC samples to evaluate the probability of a hadron to be mis-identified as a muon. Such studies are validated on data in section 11.2.2, in order to obtain an estimation of the normalisation of the peaking background after the full selection of the analysis is applied. The expected peaking background yield is about 3 events (section 11.2.2.2), to be compared to the estimated $B_d^0 \rightarrow \mu^+ \mu^-$ yield after the BDT selection of about 10 events.

8.1.3 Partially reconstructed background

The remaining background events are gathered into the PRD category. These processes can be described as two muons, of which at least one is a real muon, and additional tracks in the final state coming from a single b quark decay chain. Since only the two muons are considered in the B candidates' reconstruction, the dimuon mass of the candidate will be smaller than the B_s^0

² The hadron mis-identification considers both the case of a punch-through of hadrons to the MS and of hadrons decay-in-flight.

and B_d^0 masses. This background therefore populates predominantly the lower mass sideband. The PRD background can be further classified on the basis of the origin of the two muons:

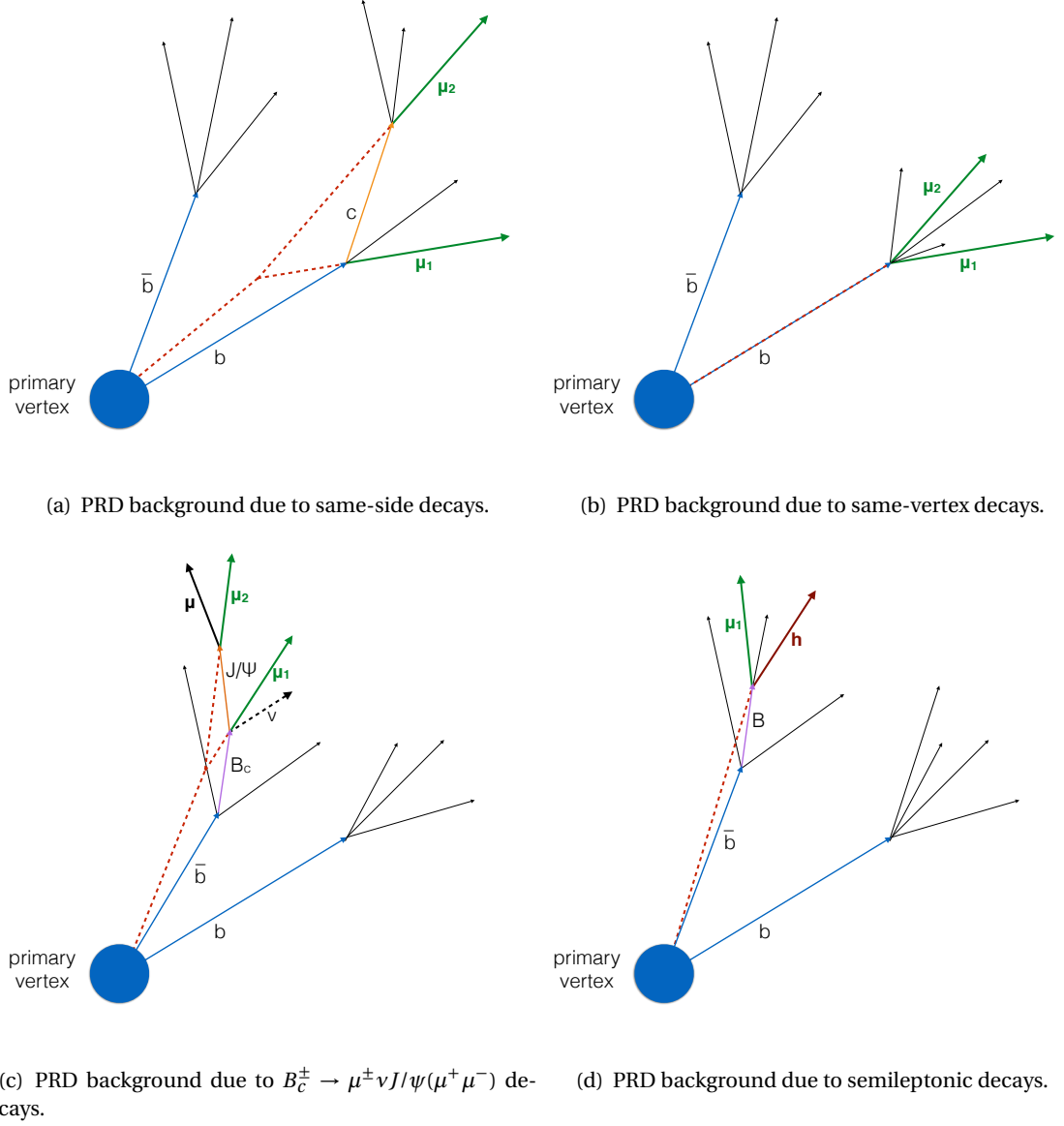


Figure 8.3: Sketches of the PRD background sources. These processes are originated by a single b quark

- **same-side**, (figure 8.3(a)), is composed of muons originated by the same b quark from a cascade like $b \rightarrow c \mu^- \nu \rightarrow s(d) \mu^- \mu^+ \nu \bar{\nu}$. This background is simulated in the large $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC introduced in section 7.2;
- **same-vertex**, (figure 8.3(b)), decays such as $B^\pm \rightarrow K^\pm \mu^+ \mu^-$, where both muons originate from the same decay vertex, belong to this category. Also this background is contained in the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC;
- B_c , (figure 8.3(c)), mainly due to $B_c^\pm \rightarrow \mu^\pm \nu J/\psi(\mu^+ \mu^-)$ decays, where the muon from the B_c decay and the oppositely charged muon from the J/ψ decay are reconstructed into a

B candidate. As the two previous background, also the B_c background is simulated in the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC;

- **semileptonic**, (figure 8.3(d)), is due to few-body semileptonic B decays, as introduced at the beginning of this section, feeding into the final selection through a muon misidentification, in the limit of low energy neutrinos. The processes that contribute the most are $B_d \rightarrow \pi \mu \nu$, $B_s \rightarrow K \mu \nu$ and $\Lambda_b \rightarrow p \mu \nu$. Exclusive MC simulations are generated for these three processes, as shown in section 7.2.

Figure 8.4 shows the invariant mass of the PRD backgrounds from simulation, after the application of the preselection described in section 7.3.

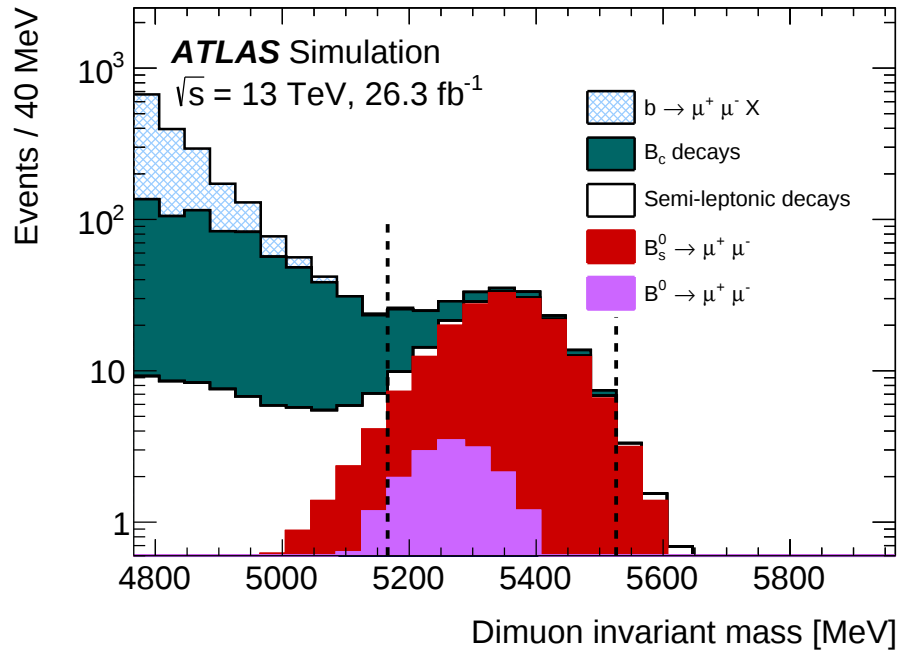


Figure 8.4: Dimuon invariant mass distribution for the partially reconstructed background, from simulation, before the final selection against the continuum background is applied. The different components are shown as stacked histograms, normalised according to world-averaged measured branching fractions. The SM expectations for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals are also shown for comparison. Continuum background is not included here and the Same-Side and Same-Vertex background categories are summed into the $b \rightarrow \mu\mu X$ histogram.

8.2 Background reduction

As introduced at the beginning of this chapter, the amount of background present in the dimuon mass region considered after the application of the preselection cuts is extremely large. An additional background reduction is therefore needed.

The only background source that can be highly reduced without critically affecting the signal is the continuum background, thanks to its kinematic and topologic differences with respect

to the events of interest. Regarding the remaining background sources, PRD backgrounds are mostly made of real dimuons, and get kinematical and topological features very close to the signal when the mass approaches the B mass. The fake-muon component of this background presents similar features and is partially rejected thanks to the muon quality cuts. Also the peaking background is attenuated by the muon quality cuts, but then it is indistinguishable from the signal.

In order to achieve a high continuum background rejection, a MVA technique based on a BDT algorithm is employed, as implemented in the TMVA [197] package of ROOT [180].

After a description of BDT classifiers, (sub-section 8.2.1), the list of the input variables employed is discussed (sub-section 8.2.2). The BDT training and testing phases are then described (sub-section 8.2.3), together with a brief comparison of the BDT performance with respect to its Run 1 counterpart.

8.2.1 BDT description

A Decision Tree is a simple binary decision tool, sketched in figure 8.5. Each binary decision of the tree is called a node and represents cuts applied on individual variables, while the final decision of a tree is referred to as leaf node. The number of nodes before a decision is taken is the depth of the tree. There are two main kinds of decision trees, based on the output variable: *classification trees*, when the outcome is a discrete variable, and *regression trees*, when the outcome is a real number.

Ideally, each node cut would correspond to the optimal signal-background separation in the

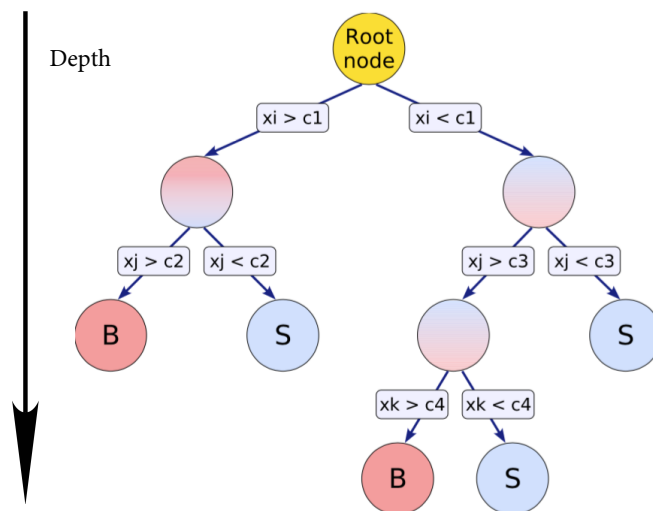


Figure 8.5: Schematic view of a decision tree. Starting from the root node, a sequence of binary splits using the discriminating variables x_i is applied to the data. The same variable may thus be used at several nodes, while others might not be used at all. The leaf nodes at the bottom end of the tree are labelled “S” for signal and “B” for background depending on the majority of events that end up in the respective nodes. Picture obtained from [197].

given variable. This is achieved through a training procedure. The training is based on the

measurement of the impurity of each leaf node, defined as the contamination of signal or background events in such leaf node. Signal purity is defined as $p_{\text{sig}} = \frac{s}{s+b}$, where s and b are the number of signal and background events in the leaf node.³ The background purity is defined as $p_{\text{bkg}} = 1 - p_{\text{sig}}$. There are several functions of the purity used in the training of decision trees. As for other BDT features, different configurations have been tested for the BDT employed in this analysis, and no significant difference was found in the final behaviour. The chosen function for the analysis is called *Gini impurity index*, defined as $\text{Gini} = \sum_{i,j \in \text{sig,bkg}}^{i \neq j} p_i p_j$ when several classes of events are used. In case of a signal-background discrimination, it can be simplified to $\text{Gini} = \frac{2sb}{(s+b)^2}$. This quantity becomes 0 when the leaf node contains only signal or background events, but increases when both are present.

The term *Boosted* in the BDT acronym refers to the usage of several trees, obtaining a *forest*. Basically, the forest is built training one tree at the time; each new tree is trained to improve the signal-background separation of the previous, based on their purity function. This procedure allows one to reduce the sensitivity of the decision tree to fluctuations in the training datasets, and can considerably enhance the performance with respect to the usage of a single tree.

There are several algorithms used for the boosting procedure, one of the most used is called *adaBoost*, short for “adaptive boosting”; the basic idea of this algorithm is that the outcomes of the various decision trees are combined into a weighted sum, allowing a higher discrimination power.

The *adaBoost* boosting procedure is used in combination with the *bagging* approach, short for bootstrap aggregation. The aim of this procedure is to further reduce the sensitivity of the decision tree to fluctuations in the training datasets. The basic idea is to use several collections of random subsets of the original training dataset to train different decision trees.

8.2.2 BDT input variables

The BDT exploited in the analysis is based on *regression trees*, with 15 input variables, summarised in table 8.1, selected in analogy with the Run 1 analysis [73]. A set of complementary and alternative variables has been extensively tested and shown not to increase the separation power of the BDT.

The input variables can be categorised into three main groups:

- ***B meson variables***, related to the reconstructed B candidate, its decays vertex and the collinearity between the candidate’s momentum $\vec{p}^B$ and the flight vector between the production and decay vertices $\vec{\Delta x}$;
- ***muon variables***, related to the two muons that form the B meson candidate, for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ channel, and to the two muons that form the J/ψ candidate for the reference and control channels;

³ When using weighted events s and b are defined as weighted sums.

- **variables related to the rest of the event**, which consider the UE in proximity of the B candidate; a correct PV-SV association of the B candidate is crucial for these variables because, as mentioned in section 7.5, an incorrect association would imply an incorrect estimation of e. g. the isolation.

Table 8.1: Description of the 15 input variables used in a BDT classifier to discriminate between signal and continuum background. When the BDT classifier is applied to $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ candidates, the variables related to the decay products of the B mesons refer only to the muons from the decay of the J/ψ . Horizontal lines separate the classifications into the three groups described in the text. For the last category, additional tracks are required to have $p_T > 500$ MeV. Table obtained from [198].

Variable	Description
p_T^B	Magnitude of the B candidate transverse momentum $\vec{p}_T^B$.
$\chi_{PV,DV}^2$	Compatibility of the separation $\vec{\Delta x}$ between production (i.e. associated PV) and Decay Vertices (DVs) in the transverse projection: $\vec{\Delta x}_T \cdot \Sigma_{\Delta x_T}^{-1} \cdot \vec{\Delta x}_T$, where $\Sigma_{\Delta x_T}$ is the covariance matrix.
ΔR_{flight}	Three-dimensional angular distance between $\vec{p}^B$ and $\vec{\Delta x}$: $\sqrt{\alpha_{2D}^2 + (\Delta\eta)^2}$
$ \alpha_{2D} $	Absolute value of the angle in the transverse plane between $\vec{p}_T^B$ and $\vec{\Delta x}_T$.
L_{xy}	Projection of $\vec{\Delta x}_T$ along the direction of $\vec{p}_T^B$: $(\vec{\Delta x}_T \cdot \vec{p}_T^B) / \vec{p}_T^B $.
IP_B^{3D}	Three-dimensional Impact Parameter (IP) of the B candidate to the associated PV.
$DOCA_{\mu\mu}$	Distance Of Closest Approach (DOCA) of the two tracks forming the B candidate (three-dimensional).
$\Delta\phi_{\mu\mu}$	Azimuthal angle between the momenta of the two tracks forming the B candidate.
$ d_0 ^{\text{max-sig.}}$	Significance of the larger absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
$ d_0 ^{\text{min-sig.}}$	Significance of the smaller absolute value of the impact parameters to the PV of the tracks forming the B candidate, in the transverse plane.
P_L^{min}	The smaller of the projected values of the muon momenta along $\vec{p}_T^B$.
$I_{0.7}$	Isolation variable defined as ratio of $ \vec{p}_T^B $ to the sum of $ \vec{p}_T^B $ and the transverse momenta of all additional tracks contained within a cone of size $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.7$ around the B direction. Only tracks matched to the same PV as the B candidate are included in the sum.
$DOCA_{\text{xtrk}}$	DOCA of the closest additional track to the decay vertex of the B candidate. Only tracks matched to the same PV as the B candidate are considered.
$N_{\text{xtrk}}^{\text{close}}$	Number of additional tracks compatible with the DV of the B candidate with $\ln(\chi_{\text{xtrk,DV}}^2) < 1$. Only tracks matched to the same PV as the B candidate are considered.
$\chi_{\mu,\text{XPV}}^2$	Minimum χ^2 for the compatibility of a muon in the B candidate with any PV reconstructed in the event.

Two of the most discriminating variables employed are ΔR_{flight} and its transverse projection

$|\alpha_{2D}|$. As one can expect from the topology of the signal, signal candidates are characterised by small values of these variables; continuum background events, on the other hand, can show much larger values.

Another important variable is L_{xy} , but in this case, the signal is characterised by large values. For instance, for B mesons $c\tau \sim 470 \mu\text{m}$; given a B meson with an energy of $\sim 50 \text{ GeV}$, the flight length becomes about 5 mm. Only the transverse projection of the flight length is considered, as the accuracy of the ATLAS detector on the $x - y$ plane is much higher compared to the accuracy on the z direction. The L_{xy} variable is highly effective in discriminating the prompt background, as already introduced in section 7.6.4.

The significances of the impact parameters of the two muons, $|d_0|^{\text{min-sig.}}$ and $|d_0|^{\text{max-sig.}}$, are highly connected to L_{xy} . Considering the example provided earlier, the two muons produced by the decay of a B meson with energy $\sim 50 \text{ GeV}$ are characterised by an impact parameter of $d_0 \sim 500 \mu\text{m}$. The usage of the significance allows one to choose only muons whose trajectory is not close to the PV, effectively requiring a B candidate that moves away from it. Additionally, the significances of the impact parameters of the two muons are expected to be correlated for real dimuon decays of B mesons, while it is not the case for continuum events. This allows additional rejection power in the training of the BDT.

Similarly, $\chi_{\text{PV,DV}_{xy}}^2$ requires the Secondary Vertex (SV) not to be compatible with the PV, enforcing the requirement of a B meson that moved away from its generation vertex.

The B meson three-dimensional impact parameter IP_B^{3D} requires the impact parameter of the B meson to be small, requiring the trajectory of the B meson to be compatible with the PV.

The $\text{DOCA}_{\mu\mu}$ variable enforces the vertexing procedure applied to the two muons. In fact, muons coming from the same SV are characterised by a small distance between their tracks. On the other hand, dimuons originated from background sources, in particular the ones from distinct b quarks, are characterised by higher distances.

The four variables $I_{0.7}$, $N_{\text{trk}}^{\text{close}}$, DOCA_{trk} and $\chi_{\mu,\text{xPV}}^2$ are strictly related to the environment surrounding the dimuon candidate. $I_{0.7}$ requires that the B candidate is isolated from other particles. This helps reducing the contribution of the continuum background as well as PRD events. Similarly, $N_{\text{trk}}^{\text{close}}$ acts to ensure that the SV is only associated with two muons.

DOCA_{trk} is complementary to $N_{\text{trk}}^{\text{close}}$; while the latter allows checking of the surrounding of the SV, the former explicitly searches for the most compatible track with the decay vertex.

The three variables just described only consider tracks associated to the same PV as the B meson. This allows the reduction of the pile-up dependence of these variables, while maintaining a high background rejection. From this consideration, it follows that the choice of the correct PV for the B candidate is extremely important, as introduced in section 7.5. The variable $\chi_{\mu,\text{xPV}}^2$ performs an additional check on this. Even if two signal muons do not come explicitly from the PV of the B meson, their compatibility with such a vertex is higher compared to their compatibility with every other PV in the event. This is not the case for the continuum background, in which case one of the two muons can show a high compatibility with other PVs.

The three remaining variables p_T^B , $\Delta\phi_{\mu\mu}$, P_L^{min} carry similar information regarding the boost of the B candidate. In fact, real B mesons are expected to show a harder energy and momentum

spectrum with respect to dimuons from the background.

8.2.3 BDT training and testing

The training and testing of the BDT classifier are performed using the $B_s^0 \rightarrow \mu^+ \mu^-$ exclusive MC and the data sidebands, with the selection described in chapter 7 applied to both.

The usage of data sidebands as background sample in the training and testing phase of the BDT selection is a novelty compared to the previous version of the analysis [73], which employed the large $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC as the background sample. An attempt to train the BDT with the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC as background sample was performed, but the performance of the classifier was not satisfying (see section 9.3 for more details).

In order to avoid biases due to the usage of the same sample to train, test and evaluate the BDT selection, a rotation strategy is implemented. Both the data sidebands and the MC are split into three independent sets of approximately equal statistics and used in turns to train, test and evaluate three separate BDTs. The evaluation phase is only performed on the upper data sideband, as the low mass sideband is contaminated with PRD background. The sample rotation strategy is summarised in table 8.2.3.

Table 8.2: Training, testing, and evaluation of the 3 BDTs. Each data sample contains 1/3 of the sideband data and 1/3 of the signal MC. For the evaluation sample, 1/3 of the signal MC and only the high-mass sideband of the collision data is used.

	Trained	Tested	Evaluated
<i>BDT 1</i>	Sample 1	Sample 2	Sample 3
<i>BDT 2</i>	Sample 2	Sample 3	Sample 1
<i>BDT 3</i>	Sample 3	Sample 1	Sample 2

A dedicated study is performed to obtain the optimal settings for the BDT training. The results are summarised in table 8.2.3, together with a short description. The optimal values for the configuration parameters *MinNodeSize* and *AdaBoostBeta* are found with the help of a grid scan using background rejection at 36% signal efficiency on the Receiver Operating Characteristic (ROC) curve as a figure of merit of classifier performance. *MaxDepth* is studied in a similar manner and the performance of the BDT is found to improve with increasing value of this parameter. However, the discrepancy between the training and testing samples, assessed using the KS test [174], is also found to increase with increasing value of *MaxDepth*, leading to over-training. The *adaBoost* technique is known to work best on trees with a small depth; for this reason, the value *MaxDepth*=4 is chosen, which allows to avoid over-training while maintaining a good performance of the BDTs.

The *NTrees* parameter is found to be highly correlated with the performance of the BDT. Low values imply a low separation between signal and background, while high values can provoke over-training. The chosen value allows the maintaining of a high performance of the BDT,

Table 8.3: Configuration parameters used in the BDT training.

Parameter	Value	Description
<i>NTrees</i>	500	Number of trees employed in the classifier
<i>MinNodeSize</i>	0.2 %	Minimum percentage of training events required in a leaf node
<i>MaxDepth</i>	4	Max depth of the decision tree allowed
<i>BoostType</i>	AdaBoost	Trees boosting strategy
<i>AdaBoostBeta</i>	0.5	Learning rate for AdaBoost algorithm, this is the weight assigned to each subsequent tree
<i>UseBaggedBoost</i>	True	Use Bootstrap aggregating (bagging) approach to improve BDT
<i>BaggedSampleFraction</i>	0.6	Relative size of bagged event sample to original size of the data sample
<i>SeparationType</i>	GiniIndex	Separation criterion for node splitting
<i>nCuts</i>	100	Number of grid points in variable range used in finding optimal cut in node splitting
<i>NormMode</i>	EqualNumEvents	Overall renormalisation of event-by-event weights used in the training

without introducing any over-training effect.

Figure 8.6(a) shows the comparison between the ROC curves of the three BDTs trained and tested on Run 2 data and the BDT employed in the Run 1 analysis applied to both Run 1 and Run 2 data. The three Run 2 classifiers show, as one can expect, a comparable separation power, while the Run 1 BDT applied to Run 2 data shows a decrease on the separation power with respect to its application to Run 1 data. Figure 8.6(b) shows the same Run 1 BDT ROC curves and the ROC curve of the Run 2 BDT obtained merging the three Run 2 classifiers. The Run 2 BDT shows a lower separation power with respect to the Run 1 BDT applied to Run 1 data. This loss in performance of the selection has been investigated and its source was recently identified. The definition of the isolation related variables of the Run 2 analysis is not fully compatible with their Run 1 counterparts in the ATLAS core software. This feature also explains the low performance of the Run 1 BDT applied to Run 2 data. Ongoing studies are trying to assess how to fix this issue, in order to improve the performance of the isolation variables. Figure 8.7 shows the distribution of the BDT output variable for the various MC samples and the data sidebands. The BDT output variable is designed to range between -1 and 1, with higher values corresponding to signal-like candidates. Most of the background events belong to the continuum background category, as shown in figure 8.7(b). The continuum background simulated sample shows a slightly different behaviour compared to the data sidebands. This is due to the presence of prompt dimuon contributions in data, in particular at low BDT values. Such com-

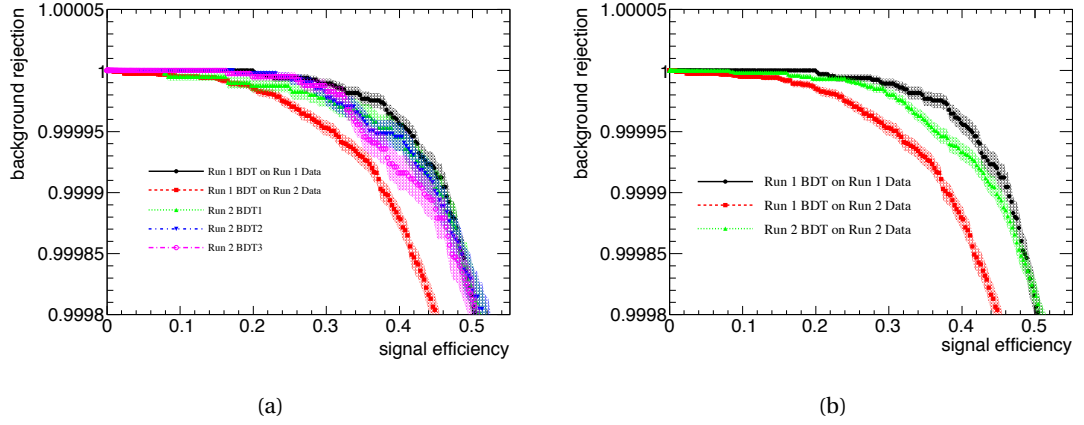


Figure 8.6: Comparison of the ROC curves of the Run 2 BDTs with the ROC curves of the Run 1 BDT applied to both Run 1 and Run 2 data. In figure 8.6(a) the Run 2 BDTs ROC curves are shown, while in figure 8.6(b) the ROC curve of the Run 2 BDT obtained merging the three BDTs is shown. Errors on the background efficiency values are represented by shaded error bands.

ponents are not simulated in the $b\bar{b} \rightarrow \mu^+\mu^-X$ MC, as introduced in section 7.2. The events categorised as PRD background, shown in figure 8.7(a), have a signal-like behaviour, which allows them to acquire a higher BDT output value. Even if these components have a much lower yield compared to the continuum background, their presence becomes relevant in the signal-like BDT region, as discussed later in section 11.1.

The correlation of the BDT input variables are studied using the signal and continuum back-

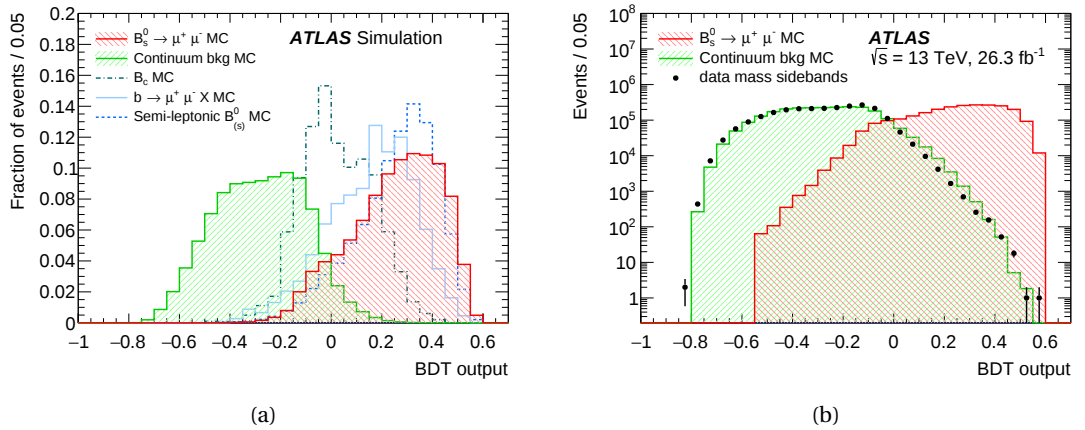


Figure 8.7: BDT output distribution for the signal and background events after the preliminary selection and before applying any reweighting to the BDT input variables described in section 10.2: 8.7(a) simulation distributions for $B_s^0 \rightarrow \mu^+\mu^-$ signal, continuum, partially reconstructed $b \rightarrow \mu^+\mu^-X$ events and B_c decays; 8.7(b) dimuon sideband candidates (which also include prompt contributions, mainly at lower BDT values and not simulated in the continuum MC sample), compared with the continuum MC sample and the simulated signal. All distributions are normalised to unity in 8.7(a) and to data sidebands in 8.7(b).

ground MC samples and the data sidebands. Figure 8.8 shows the correlations matrix for the discriminating variables together with m_B , $|\eta_B|$ and n_{PV} for the signal MC and data sidebands before, 8.8(a) and 8.8(b), and after, 8.8(c) and 8.8(d), applying the BDT selection with a signal

efficiency of 54%; the value of this cut is discussed later in this section.

These four plots provide a large amount of information on the BDT, its training and its validity. The first important piece of information regards the correlation of the various variables with the invariant mass of the dimuons. Any mass-dependence might introduce selection-driven sculpting of the dimuon invariant mass distribution, which can modify the shape of the signal peaks. Since none of the variables shows a relevant correlation with the mass of the dimuon candidates, both before and after the BDT application, the BDT is considered independent from the dimuon invariant mass.

The comparison of the correlation matrices before and after the application of the BDT allows a simple check to verify the proper functioning of the selection. While before the application BDT signal and data sidebands show different features, after the selection the two correlation matrices are extremely similar. This confirms that the BDT is only selecting signal-like candidates.

The expected behaviour of the discriminating variables can be also checked. As introduced in section 8.2.2, the significances of the impact parameters of the two muons, $|d_0|^{\text{min-sig.}}$ and $|d_0|^{\text{max-sig.}}$ are expected to be highly correlated for the signal, but not for the background. The correlation matrices obtained before the application of the BDT show exactly this behaviour.

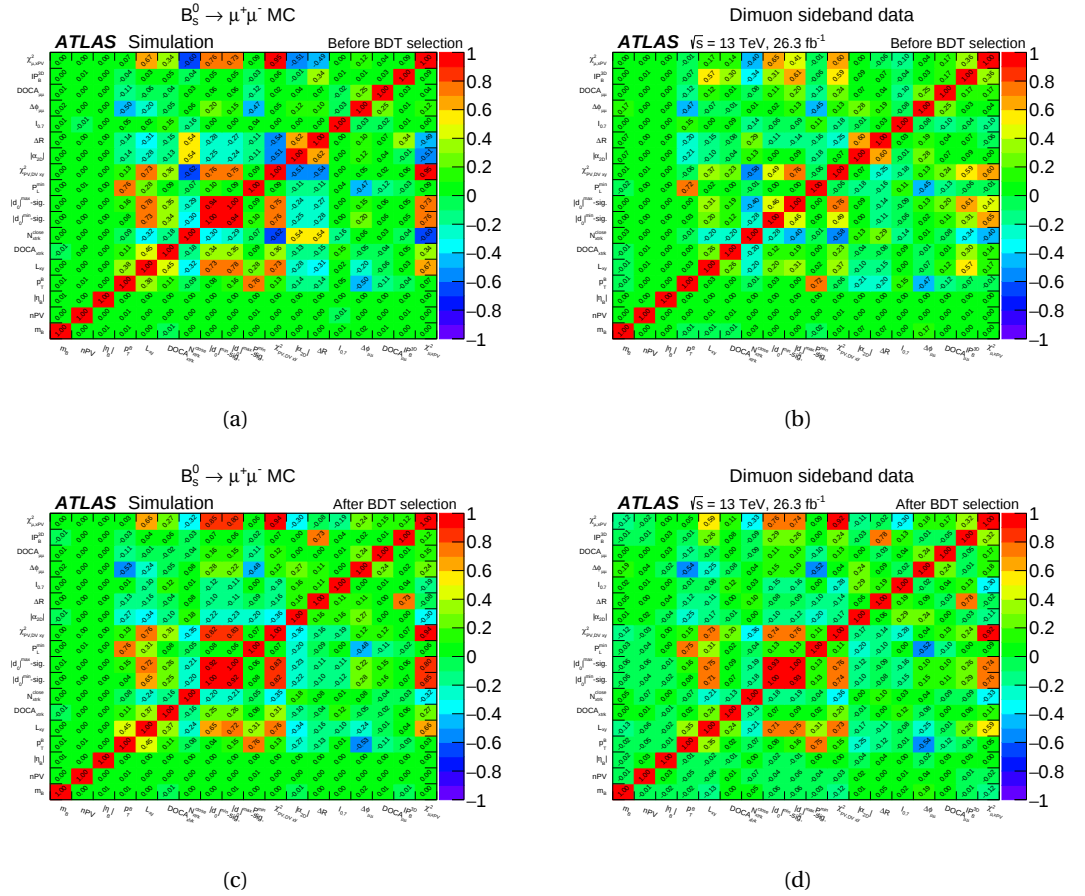


Figure 8.8: Correlation table of the discriminating variables used in the BDT, together with m_B , $|\eta_B|$ and nPV (the number of reconstructed PV in the event), for $B_S^0 \rightarrow \mu^+\mu^-$ MC sample and data sidebands before, 8.8(a) and 8.8(b), and after, 8.8(c) and 8.8(d), applying the BDT selection.

The final BDT selection is not re-optimised with respect to the previous analysis. A cut on the BDT output variable with a signal efficiency of 54% is therefore chosen. This corresponds to BDT values larger than 0.2455 and a background rejection of about 99.97%.

The resulting dataset is divided in three disjoint BDT intervals in order to enhance the signal sensitivity of the analysis. An explicit optimisation of the BDT binning is not repeated: the BDT efficiency binning simply repeats the one chosen for the Run 1 analysis.

The three BDT bins have boundaries 0.2455, 0.3312, 0.4163 and 1, chosen to obtain a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency of 18% each, measured with the $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample.

A novelty with respect the previous analysis is the introduction of an additional bin at lower BDT values. It is chosen to have the same signal efficiency as the other BDT bins (18%): this corresponds to a lower boundary of 0.1439, for an overall signal efficiency of 72% and a background rejection of about 99.7%. Accepted candidates in this BDT bin are dominated by background contributions and are basically insensitive to the signal contribution: as discussed in section 11.3, their role is effectively that of constraining the background models. For this reason, signal and reference channel yields as well as efficiencies are measured with respect to the former selection, implying a signal efficiency of 54%.

Systematic uncertainties related to the BDT selection are taken into account in the R_{Ac} term in formula 5.1. Systematic uncertainties due to the binning scheme chosen for the BDT output distribution are also described in section 10.2, and taken into account in the signal fit procedure.

DATA-MC COMPARISON

9

Simulated signal and background events are used in this analysis to assess signal efficiency, identify suitable background models and other purposes. It is therefore important to gauge the consistency of such simulations with data where possible, and identify potential sources of systematic uncertainties that will need to be taken into account in the final result. This chapter is devoted to the comparison of data and MC samples, in order to quantify relevant discrepancies.

The efficiency of the BDT-based selection will be assessed on signal MC, which cannot be easily compared to data. It will be therefore verified, in sections 9.1 and 9.2 respectively, that the distributions of the BDT input variables introduced in section 8.2.2 are consistent between data and MC when considering the reference channel $B^+ \rightarrow J/\psi K^+$ and the control channel $B_s^0 \rightarrow J/\psi \phi$.

Possible residual discrepancies will be used as sources of systematic uncertainties, as discussed in chapter 10.

Despite its limited use in the analysis, the inclusive background simulation is compared against data sidebands as well, since even though simulated background samples are not used for the training of the BDT, they are still employed in the last stage of the analysis (section 11.1) to identify suitable invariant mass models in the data invariant mass distribution fit. The reliance on the background simulation is limited to the selection of suitable functional models (and not used e. g. to determine background yields or shape parameters) which are then checked against data sidebands, and is complemented with systematic studies challenging these assumptions.

9.1 Reference channel comparison

This section describes the comparison between data and MC for the high-statistics $B^+ \rightarrow J/\psi K^+$ channel, used to estimate the agreement between the simulated samples and real data and to evaluate any discrepancy, used in section 10.2 to estimate the systematic uncertainties on the efficiency and acceptance ratio.

Before comparing it to data, the $B^+ \rightarrow J/\psi K^+$ MC sample is re-weighted according to the full

procedure described in section 7.6 and the signal distribution in data is extracted using the sideband-subtraction procedure, introduced in section 7.6.4.2.

Figures 9.1 and 9.2 show the comparison for the most important variables used in the BDT se-

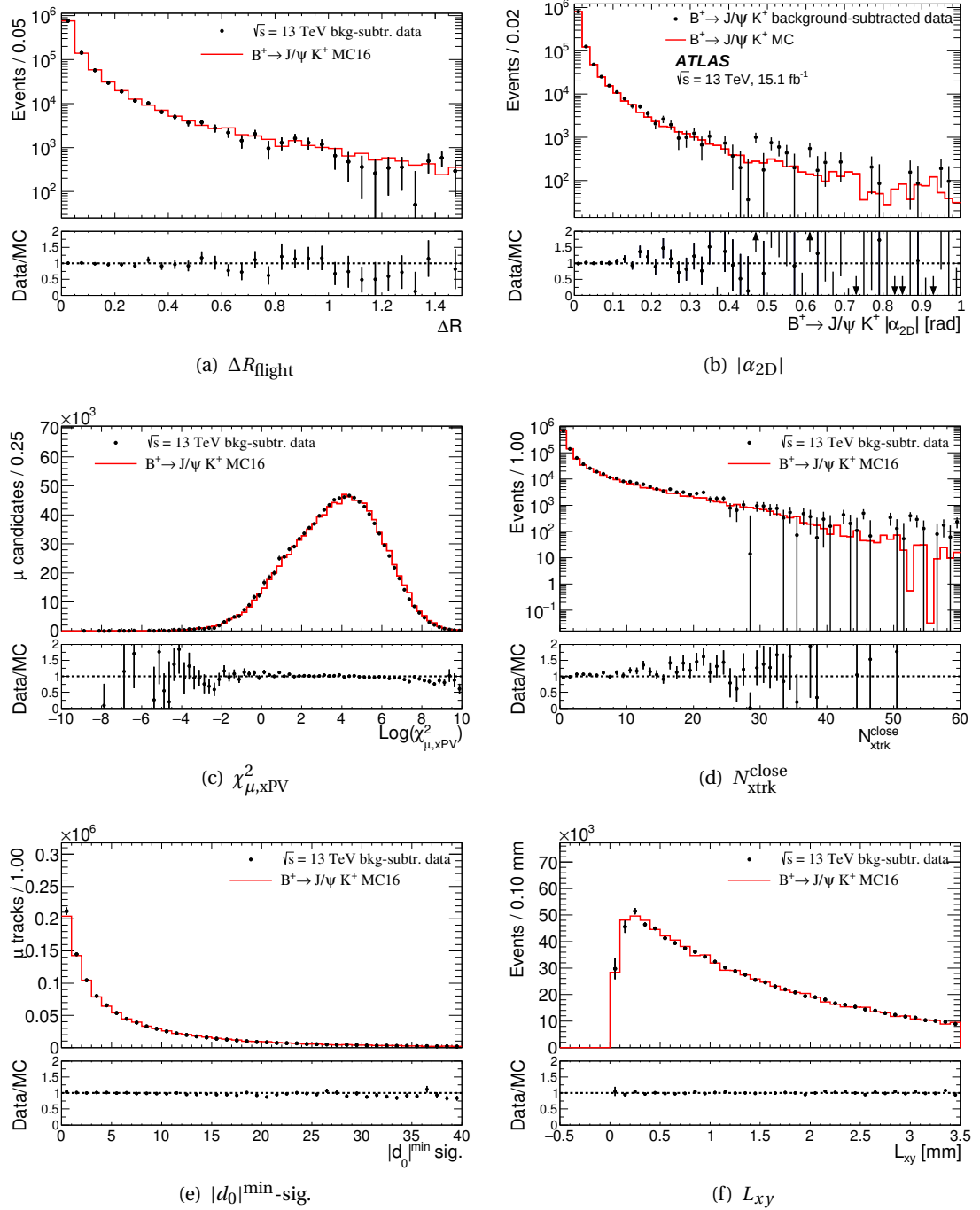


Figure 9.1: Comparison of $B^+ \rightarrow J/\psi K^+$ sideband-subtracted data and MC. The variable compared is shown below each plot. The black filled markers correspond to data and the red histogram corresponds to the $B^+ \rightarrow J/\psi K^+$ MC sample, normalised to the data. The lower plot for each comparison shows the ratio of the data to the MC.

lection (section 8.2). The remaining variables not present in these figures show a comparable agreement. Since the B^+ vertex is 3-prong, certain variables are not directly equivalent to their

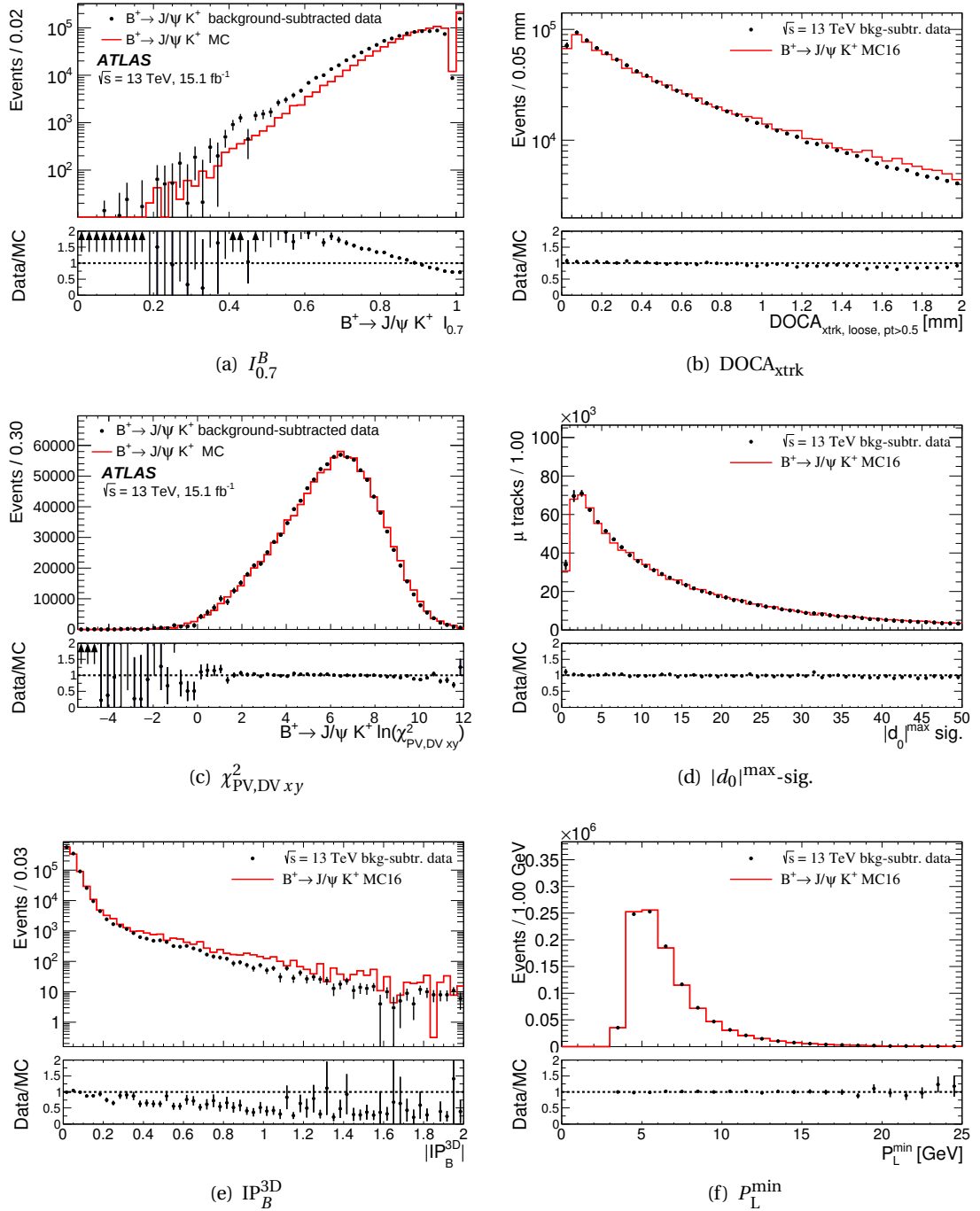


Figure 9.2: Comparison of $B^+ \rightarrow J/\psi K^+$ sideband-subtracted data and MC. The variable compared is shown below each plot. The black filled markers correspond to data and the red histogram corresponds to the $B^+ \rightarrow J/\psi K^+$ MC sample, normalised to the data. The lower plot for each comparison shows the ratio of the data to the MC.

counterparts in the two-prong $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal; for these cases, together with the 3-prong version of the variables, the comparison is performed also mimicking a two-prong counterpart by reconstructing the dimuon vertex in the normalisation and control channel signal candidates.

The result of the data-simulation comparison is not qualitatively different from the one from [73]:

the overall behaviour of the distributions of data and MC is comparable, with possible disagreements especially in the regions with low statistics. A more significant disagreement (also compared to what found in [73]) is present in the isolation-related variables $N_{\text{trk}}^{\text{close}}$ and $I_{0.7}^B$ (figures 9.1(d) and 9.2(a)). This discrepancy was found to be the main source of the disagreement

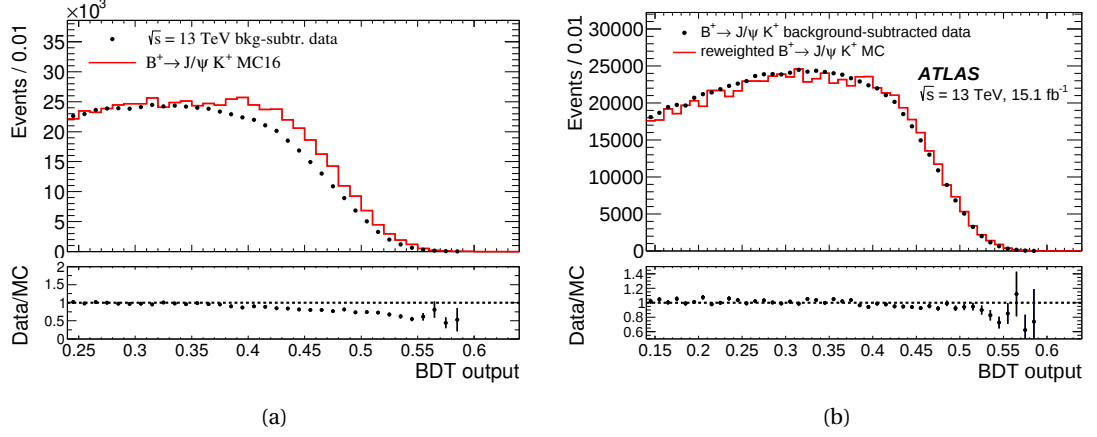


Figure 9.3: Comparison of the BDT output variable for the $B^+ \rightarrow J/\psi K^+$ channel. The black filled markers correspond to the sideband-subtracted data and the red histogram corresponds to the MC sample. The re-weighting procedure discussed in section 7.6 has been applied to the MC distribution in figure 9.3(a), while figure 9.3(b) shows the same distributions, after additionally reweighting the MC according to the $I_{0.7}^B$ distribution observed in the sideband subtracted signal.

in the data-MC comparison of the BDT output variable, shown in figure 9.3. This figure depicts the data-MC comparison of the BDT output variable zoomed in the region of interest for the analysis, before (figure 9.3(a)) and after (figure 9.3(b)) a re-weighting performed on the isolation distribution on sideband-subtracted data, showing the improvement in the comparison after the re-weighting.

Due to the large data-MC disagreement in the isolation-related variables and its extensive effect on the BDT output distribution, the central value of the efficiencies, derived in section 10.2, is calculated applying all the MC corrections as well as the reweighting on the isolation variable applied to figure 9.3(b).

The residual discrepancies in the data-MC comparison are considered in the systematic uncertainties that affect the efficiency ratio (section 10.2.2). Such effects are, in fact, not expected to have a sizeable contribution to the systematic uncertainties affecting the analysis, as was the case for [73].

9.2 Control channel comparison

The same data-MC comparison described in the previous section for $B^+ \rightarrow J/\psi K^+$ channel is performed also on the control channel $B_s^0 \rightarrow J/\psi \phi$.

This second comparison is important for three main reasons:

- it provides a cross-check of the observed discrepancies on $B^+ \rightarrow J/\psi K^+$ and their applic-

ability to B_s^0 decays;

- it allows the checking of the consistencies (or lack thereof) on a sample with a much cleaner signal to background ratio than the reference channel;
- the control channel provides also a check on a final state with a different particle multiplicity, sustaining the applicability of the corrections on the BDT input variables distributions also to two-body B decays.

The approach followed is the same as of the $B^+ \rightarrow J/\psi K^+$ samples: the MC sample is weighted according to the procedure described in section 7.6 and the signal distribution in data is extracted using a sideband-subtraction procedure, already discussed in section 7.6.4.2.

Figures 9.4 and 9.5 show the result of the comparison for the most important variables used in the BDT selection. Analogously to the reference channel, the overall agreement is good, except for a significant discrepancy in isolation-related variables. In this case, the disagreement is less pronounced than the one found for $B^+ \rightarrow J/\psi K^+$. Figure 9.6 shows the data-MC comparison for the BDT output variable of the control channel, zoomed in the region of interest for the analysis. In particular, figure 9.3(a) shows the comparison before the same isolation based re-weighting as the $B^+ \rightarrow J/\psi K^+$ channel is applied, while figure 9.6(b) shows the comparison afterwards. As for the data-MC comparison performed on the single variables, the comparison for the BDT output variable shows a better agreement for $B_s^0 \rightarrow J/\psi \phi$ compared to $B^+ \rightarrow J/\psi K^+$. This effect can be related to the smaller size of the sample, which implies larger statistical uncertainties, but might also point to issues in the sideband subtraction procedure. In fact, the $B_s^0 \rightarrow J/\psi \phi$ has a much lower background contribution compared to $B^+ \rightarrow J/\psi K^+$, allowing a more efficient background subtraction. In any case, the compatibility of the DDW calculated for $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ rule out the possibility of issues in the sideband subtraction procedure.

The overall compatible agreement for the discriminating variables between data and MC is therefore found for the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ processes. This validates the possibility of employing the data-MC discrepancies in the $B^+ \rightarrow J/\psi K^+$ channel for the evaluation of systematic uncertainties related also to signal and control channels, as discussed later in section 10.2.

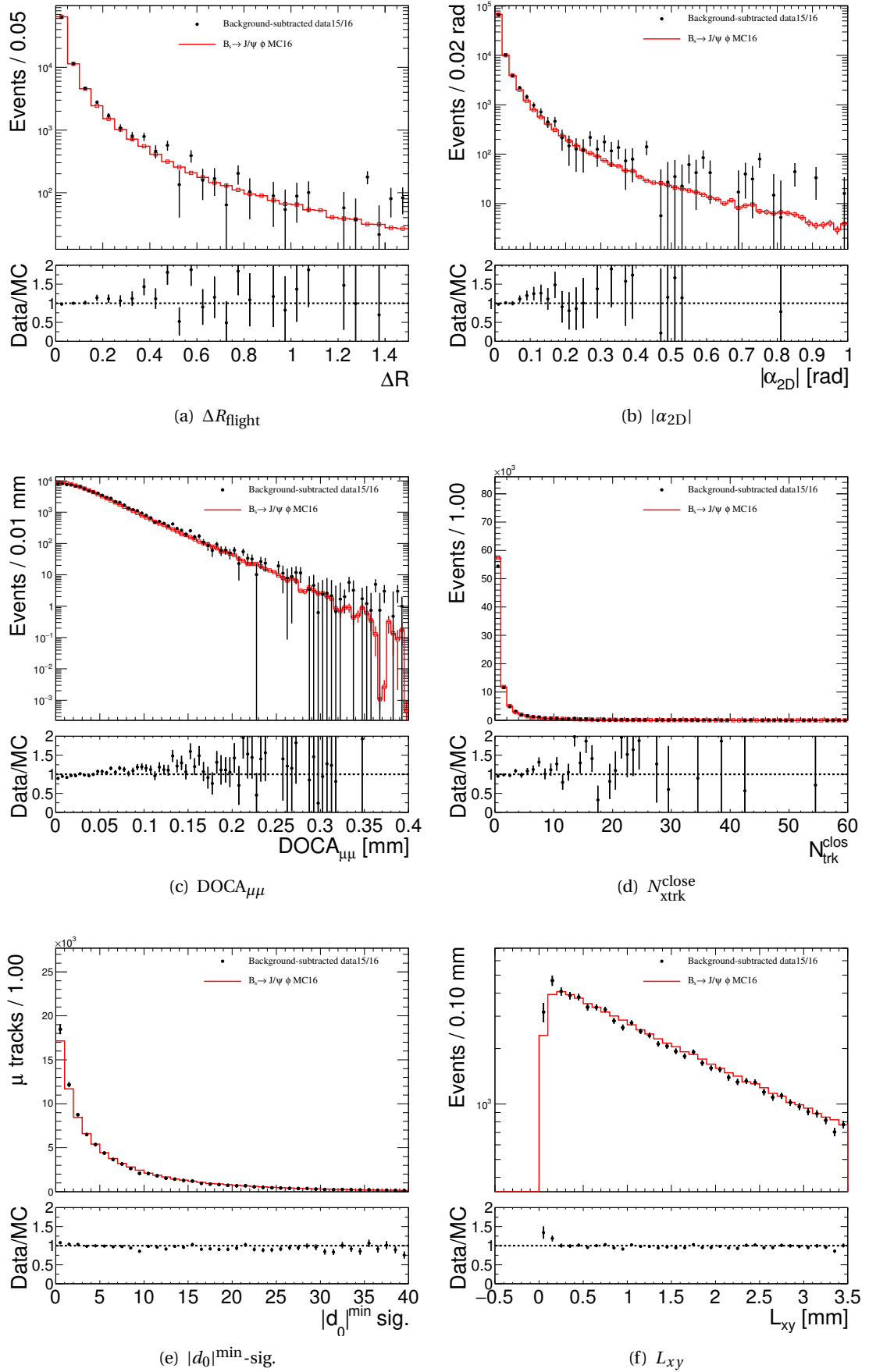


Figure 9.4: Comparison of $B_s^0 \rightarrow J/\psi\phi$ sideband-subtracted data and MC. The variable compared is shown below each plot. The black filled markers correspond to data and the red histogram corresponds to the $B_s^0 \rightarrow J/\psi\phi$ MC sample, normalised to the data. The lower plot for each comparison shows the ratio of the data to the MC.

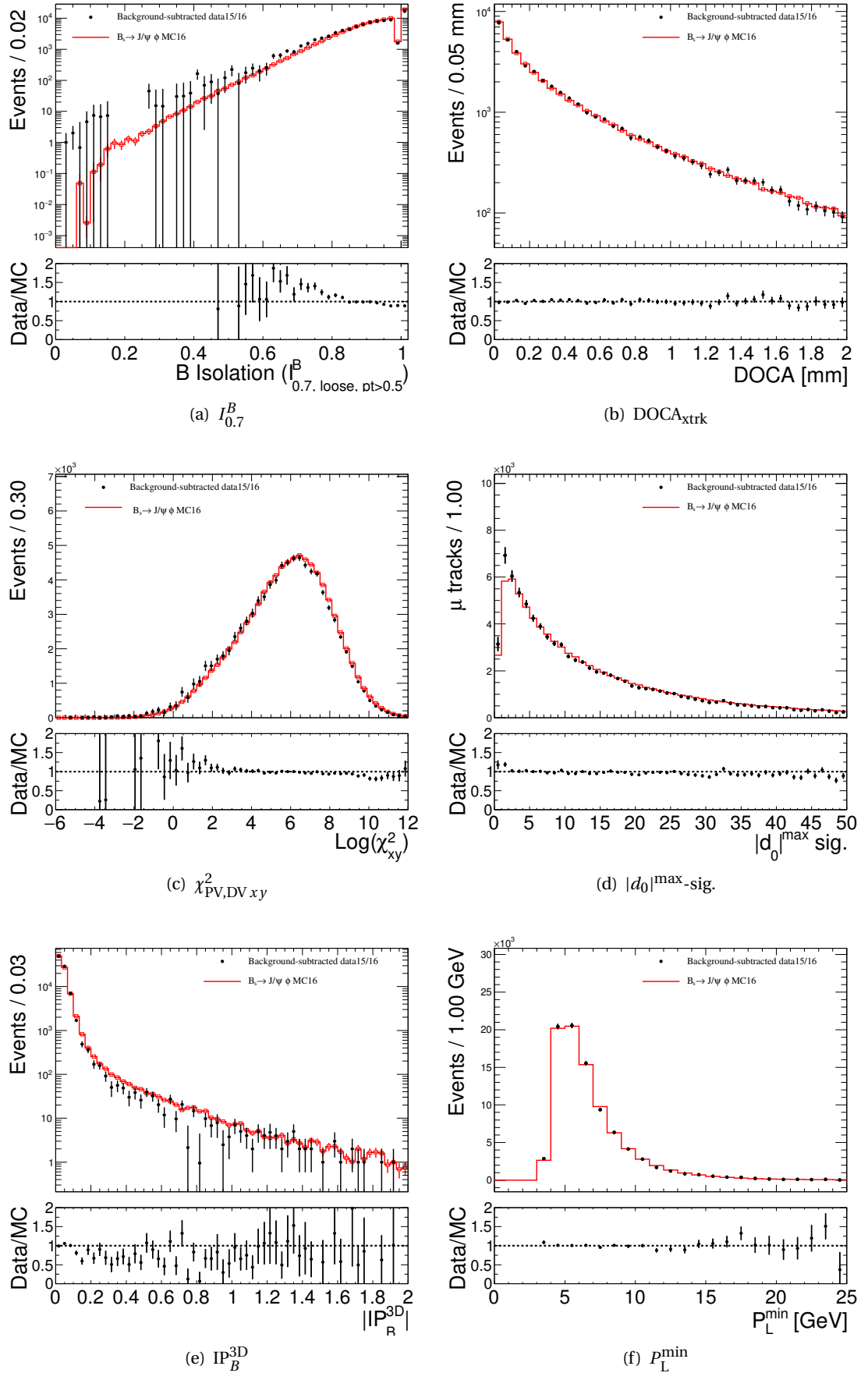


Figure 9.5: Comparison of $B_s^0 \rightarrow J/\psi\phi$ sideband-subtracted data and MC. The variable compared is shown below each plot. The black filled markers correspond to data and the red histogram corresponds to the $B_s^0 \rightarrow J/\psi\phi$ MC sample, normalised to the data. The lower plot for each comparison shows the ratio of the data to the MC.

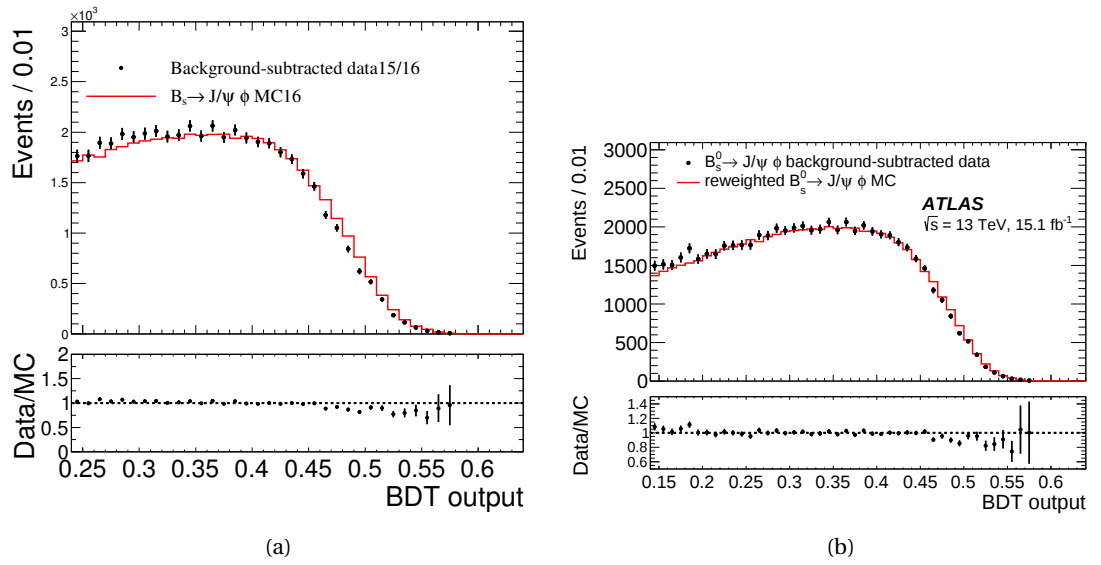


Figure 9.6: Comparison of the BDT output variable for the $B_s^0 \rightarrow J/\psi \phi$ channel. The black filled markers correspond to the sideband-subtracted data and the red histogram corresponds to the MC sample. The reweighting procedure discussed in section 7.6 has been applied to the MC distribution in figure 9.6(a), while figure 9.6(b) shows the same distributions, after additionally reweighting the MC according to the $I_{0.7}^B$ distribution observed in the sideband subtracted signal.

9.3 Continuum background comparison

The data-MC comparison for the validation of the continuum background simulated sample is performed between the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC and the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ data sidebands.

The purpose of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC, at the beginning of the analysis, was two-fold. It was supposed to be used for the background sample for the BDT training and to evaluate the functional forms to be used in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal extraction fit. The BDT was first trained with this sample, while the analysis was being developed, but the resulting BDT showed a low performance. The BDT training is hence based on data; the purpose of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ sample for this analysis is therefore to evaluate the functional forms that will be used in the signal fit on the unblinded data.

The preselection described in chapter 7 is applied to both data and MC, with an additional $L_{xy} > 0.3$ mm cut to suppress the prompt components. This is the same cut described in section 7.6.4: it allows the elimination of the prompt background components, which are not simulated in the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC. Additionally, only the continuum events are retained in the MC sample, via a truth-level selection.

The MC re-weighting procedure, shown in section 7.6, is applied, except for the kinematic weights, which are derived only for the exclusive samples. The kinematic re-weighting is performed instead based on three variables: p_T^B , η_B and $N_{0.7}^{\text{trks}}$. The last variable is the number of tracks contained within a cone of size $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} = 0.7$ around the B candidate direction. It is added to the re-weighting procedure as the same discrepancy present in the previous data-MC comparison is found. The re-weighting is performed simultaneously on all the three variables, employing a gradient boosted re-weighting technique [199]; the re-weighting uses as reference sample the data sidebands.

Figures 9.7 and 9.8 show the distributions of the re-weighted variables, together with the most important discriminating variables used in the BDT selection.

The level of agreement of the variables is found to be worse than the one found for the Run 1 analysis [73]; as already described, an attempt to train the BDT using this simulated sample was made, resulting in a low-performance classifier. For this reason, the BDT is trained on real data, as described in section 8.2

On the other hand, the agreement of the distributions is considered sufficient to employ this sample to identify suitable functional forms to be used in the final signal extraction fit. As already mentioned, the reliance on this MC sample is limited, as the signal fit is performed solely on data (section 11.3). In addition, all the models designed using the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC are checked against data sidebands and systematic uncertainties relative to the functional models chosen are considered.

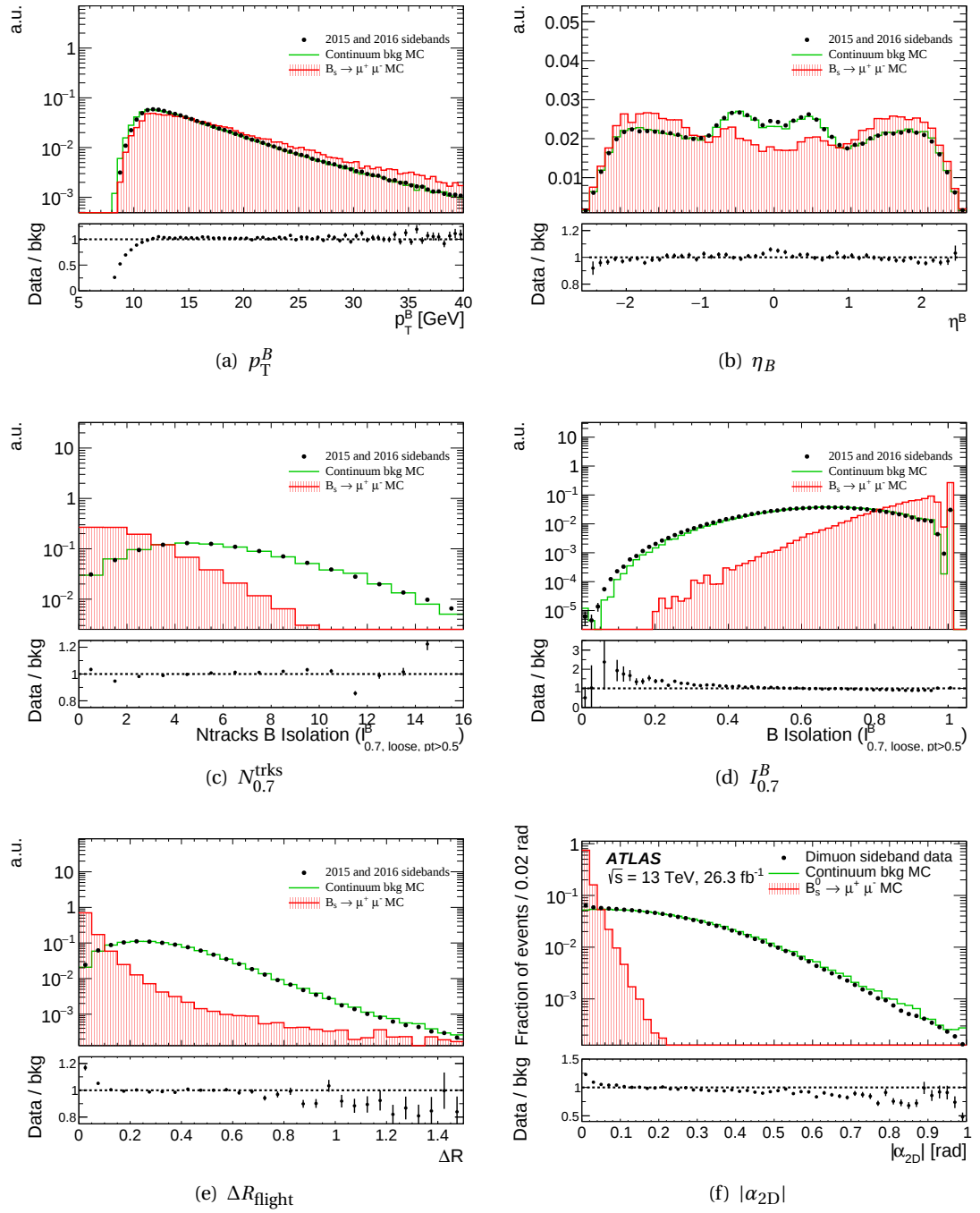


Figure 9.7: Comparison of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ data sidebands and $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC. The variable compared is shown below each plot. The black filled markers correspond to the sideband data and the green histogram corresponds to the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC sample. For reference, the $B_s^0 \rightarrow \mu^+ \mu^-$ signal MC is also displayed in the figures as a filled histogram. All histograms are normalised to unity. The inset below for each comparison shows the ratio of the sideband data to the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC.

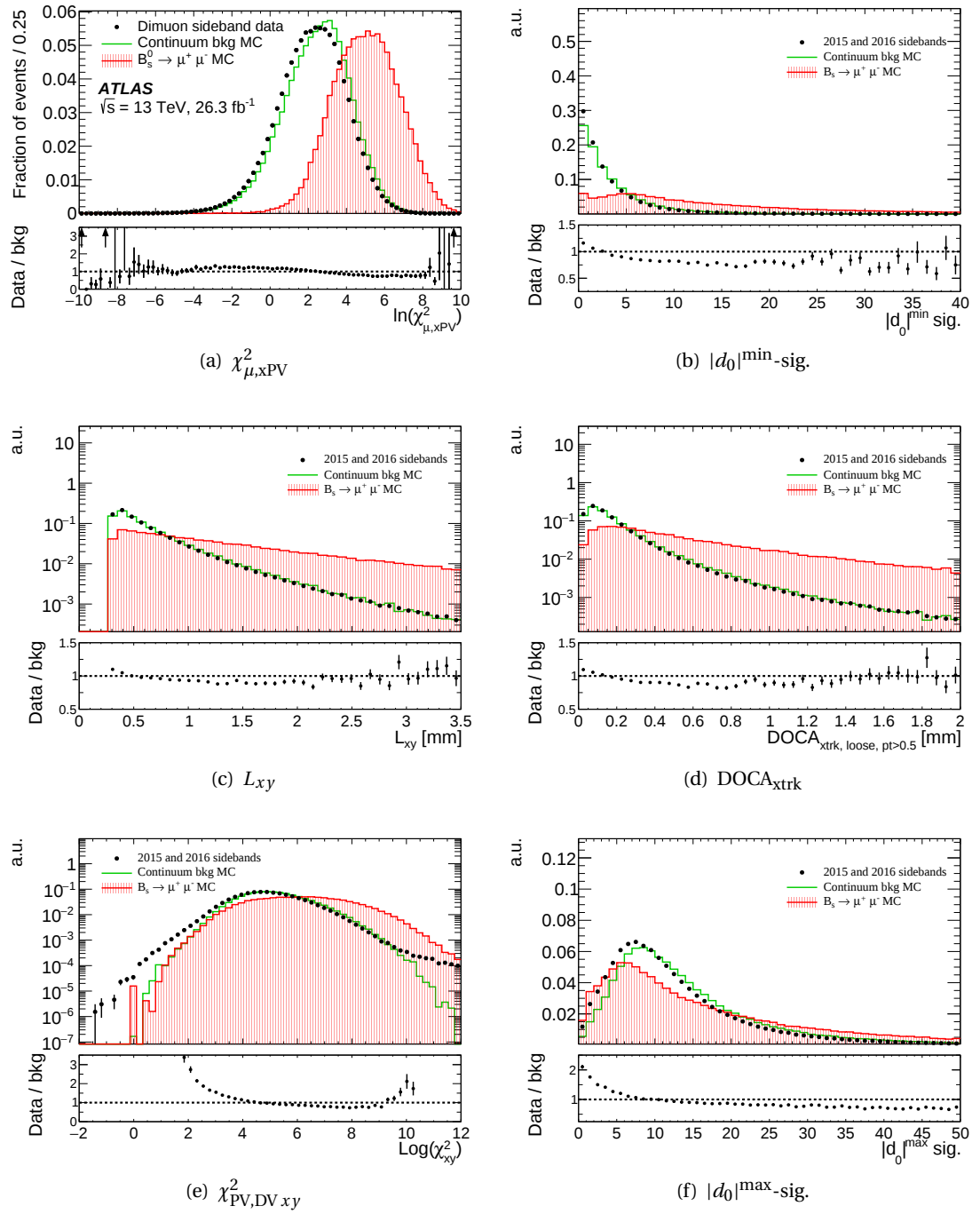


Figure 9.8: Comparison of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ data sidebands and $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC. The variable compared is shown below each plot. The black filled markers correspond to the sideband data and the green histogram corresponds to the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC sample. For reference, the $B_s^0 \rightarrow \mu^+ \mu^-$ signal MC is also displayed in the figures as a filled histogram. All histograms are normalised to unity. The inset below for each comparison shows the ratio of the sideband data to the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC.

REFERENCE CHANNEL

10

The preceding chapters have established the foundations of the analysis: the required real and simulated samples have been introduced in chapter 7, and assessed in terms of potential systematic deviations between the two in section 7.6, with chapter 8 then focusing on the finalisation of the analysis selection.

This chapter focuses on the determination of the terms of equation 5.1 which involve the reference channel: the $B^+ \rightarrow J/\psi K^+$ yield and its relative efficiency compared to the signal.

Section 10.1 will discuss the fit technique employed in the reference channel yield extraction and the result of this procedure, followed by the detailed determination of the relative signal efficiency R_{Ac} (section 10.2).

10.1 $B^+ \rightarrow J/\psi K^+$ yield extraction

This section is dedicated to the extraction of the reference channel yield, which enters in formula 5.1.

The number of $B^+ \rightarrow J/\psi K^+$ candidates is estimated using a fit to the $J/\psi K^+$ invariant mass distribution: the samples employed are described in sub-section 10.1.1, while the fitting procedure is detailed in sub-section 10.1.2. Systematic uncertainties associated with the fitting procedure are reviewed and evaluated in sub-section 10.1.3. The final result of the fit is reported in sub-section 10.1.4.

10.1.1 Samples composition

The $B^+ \rightarrow J/\psi K^+$ fit is extracted from data after the application of a selection as close as possible to the one used on $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ candidates; as discussed in chapter 7, this allows the minimisation of systematic uncertainties in the relative efficiencies of the two signals. Simulated samples are employed to study the functional forms used on data and, in certain cases, to constrain such functional forms in the fit.

The pre-selection described in section 7.3 and the BDT cut at 0.2455 (54% efficiency on $B_s^0 \rightarrow \mu^+ \mu^-$, as described in section 8.2) are applied to both data and MC samples employed in the reference yield extraction. The invariant mass region considered for the fit is defined at preselection level

($4930 \text{ MeV} < m(B) < 5630 \text{ MeV}$ from section 7.3); it is centred on the B^+ mass peak and wide enough to allow the proper modelisation of the background for a good signal-background discrimination, without introducing additional background contributions. Only half of the data and MC signal samples are exploited in the fitting procedure, as the other half is used to tune the kinematic distributions of simulated events; this avoids possible correlations among the various steps of the analysis. The samples are split according to the unique *event number* associated with each recorded pp collision: odd-numbered events are used in the $B^+ \rightarrow J/\psi K^+$ fit and even-numbered events are used in the derivation of the DDW.

After the application of the selection, four main components can be identified in the B^+ candidates' invariant mass distribution in data. The B^+ signal is evident, with a tall peak in the centre of the invariant mass distribution, together with visible contributions from three background categories.

- The **Combinatorial background** spans across the fit mass range, which corresponds to the B^+ invariant mass region defined at preselection level ($4930 \text{ MeV} < m(B) < 5630$ from section 7.3), with a mild dependence on the $J/\psi K^\pm$ invariant mass. Studies performed on the $b\bar{b} \rightarrow J/\psi X$ MC show that this background is mostly due to $J/\psi \rightarrow \mu^+ \mu^-$ decays associated with a random track from the event.
- The $B^+ \rightarrow J/\psi \pi^+$ **background** is shifted away from the B^+ mass peak due to the $\pi \rightarrow K$ mass misassignment, and for this reason appears as a distorted peak to the right of the main signal. The contribution is less noticeable with respect to the other backgrounds, due to the CKM suppression of the $B^+ \rightarrow J/\psi \pi^+$ process. This background is studied with the exclusive $B^+ \rightarrow J/\psi \pi^+$ MC.
- The **Partially Reconstructed Decays (PRD) background** originates from decays in which one or more particles in the final state are not reconstructed, e. g. $B \rightarrow K^*(\pi K) J/\psi(\mu\mu)$. Due to the missing four-momentum, this background populates the left side of the $B^+ \rightarrow J/\psi K^+$ peak. The studies of this background are based on the $b\bar{b} \rightarrow J/\psi X$ MC.

10.1.2 Reference channel fit configuration

The $B^+ \rightarrow J/\psi K^+$ yield extraction is performed with an unbinned extended maximum likelihood fit to the $J/\psi K^+$ mass distribution.

All functional forms used to model signal and background are extracted from studies on MC samples; as part of the systematic studies, these models will be varied to assess the effect of this assumption on the signal yield. Some of the shape parameters of the functional forms are constrained fitting simultaneously data and MC samples, while the yields of both signal and backgrounds are extracted from the data.

The $B^+ \rightarrow J/\psi K^+$ model is studied on the exclusive MC generated for this process; such simulations include also the radiative decays where the B^+ meson emits a γ . The models for the non-radiative and radiative components are studied separately and are both parametrised with the sum of a Johnson S_U [200, 201] and a gaussian function. The total PDF used for the signal

is therefore the sum of the two components. The shape parameters associated with the signal PDF are extracted with a simultaneous fit on data and the exclusive $B^+ \rightarrow J/\psi K^+$ MC.

The three background sources are parametrised as follows:

- the **Combinatorial background** is modelled with an exponential PDF, although the model is chosen on the basis of past analyses and MC studies, the normalisation and shape parameters are extracted as free parameters in the data fit;
- the $B^+ \rightarrow J/\psi \pi^+$ **background** is described with the sum of a Johnson S_U and a gaussian. The shape parameters are extracted from a simultaneous fit on data and the exclusive $B^+ \rightarrow J/\psi \pi^+$ MC;
- the **PRD background** is further divided into sub-categories according to the mass dependence of its various components.
 - *PRD1* produces a “shoulder” at about 5150 MeV; the main contributors to this component are found to be $B^0 \rightarrow J/\psi \rho^0$, $B^+ \rightarrow J/\psi \rho^+$, $B^0 \rightarrow J/\psi K^+ \pi^-$, and $B^+ \rightarrow J/\psi K^+ \pi^0$.
 - *PRD2*, mainly composed of $B^+ \rightarrow \chi_{c1}(J/\psi \gamma) K^+$ decays, due to the different missing four-momentum compared to *PRD1* presents a shoulder in the invariant mass distribution at 5050 MeV.
 - *PRD3* includes the remaining PRD events; this component shows a smoother mass dependence.

The three categories are parametrised with combinations of Fermi-Dirac [202, 203] and exponential functions, whose parameters are extracted from the simultaneous fit on data and MC.

The simultaneous fit on data and MC includes also a gaussian smearing of the data PDF relative to the MC templates, to account for possible data-MC discrepancies in mass scale and resolution.

The starting values of the fit parameters are based on the result of the studies based on MC for the functional shapes, while the relative normalisations are based on the result of the $B^+ \rightarrow J/\psi K^+$ fit of the Run 1 analysis [73].

The result of fit on data and MC performed using the configuration described above is shown in figure 10.1 and yields 334351 ± 975 $B^+ \rightarrow J/\psi K^+$ decays, corresponding to a statistical uncertainty of 0.3%.

The additional smearing parameters introduced in the fit, to account for data-MC discrepancies show a negligible difference in resolution between data and MC (0.1 ± 0.3 MeV) and a mass shift at the level of -2 MeV.

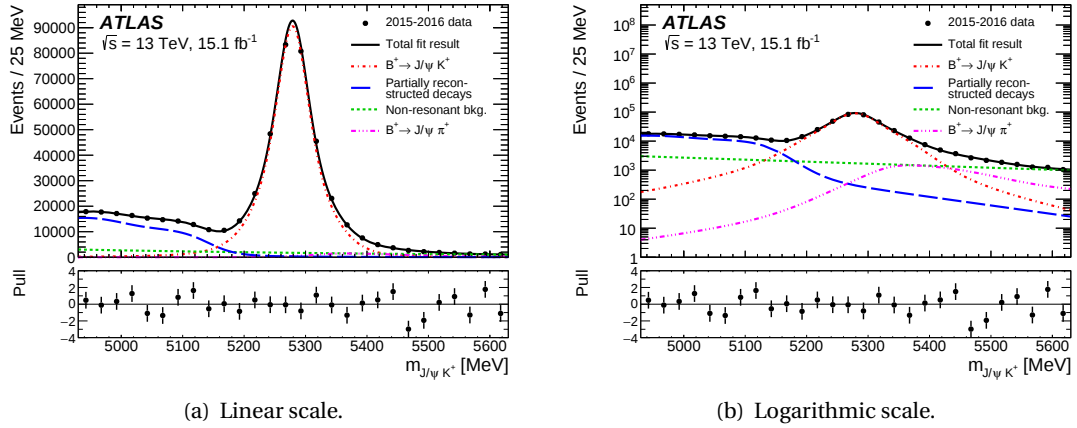


Figure 10.1: Result of the fit to the $B^+ \rightarrow J/\psi K^+$ invariant mass distribution for all B^+ candidates in half of the data events, with linear and logarithmic y axis respectively in 10.1(a) and 10.1(b). The various components of the spectrum are described in the text. The inset at the bottom of the plots shows the bin-by-bin pulls for the fit, where the pull is defined as the difference between the data point and the value obtained from the fit function, divided by the error from the fit.

10.1.3 Systematic uncertainties on the $B^+ \rightarrow J/\psi K^+$ fit

Some of the potential sources of systematic uncertainty are included in the fitting procedure: the effect of the finite size of the simulated samples as well as the differences in mass scale and resolution between data and MC. Additional systematic uncertainties originate from the assumptions on the fit models, which are evaluated by varying the default models described in section 10.1.2.

- **Combinatorial background model.** The choice of the functional form for this background, an exponential PDF, is challenged by changing the model to a linear function.
- **PRD1 and PRD2 models.** Systematic uncertainties associated with the functional form chosen for these backgrounds are evaluated by replacing the Fermi-Dirac PDF employed in the baseline fit with a complementary error function.
- **PRD composition in MC.** The relative normalisation of the PRD components in the data sample is free and not tied to the PRD MC values. However, the relative abundance of partially reconstructed final states within a given PRD mode can affect the corresponding model shape and therefore the signal yield in the combined fit. In order to take this effect into account, the MC sample PRD background composition is modified by varying the relative abundances of the different processes within the uncertainties of the latest Particle Data Group (PDG) [11] values.

Besides shape parameterisations, other sources of systematic uncertainty affect the fit.

- **Kinematic differences between data and MC.** These discrepancies are evaluated and corrected for the exclusive samples by the QLC and DDW, which are applied to the MC

samples in the baseline fit. The systematic uncertainty associated with these weights is evaluated by removing the weighting procedure on the MC.

- $B^\pm \rightarrow J/\psi K^\pm$ **charge asymmetry**. The MC fit for this component is performed on an even generation of $B^+ \rightarrow J/\psi K^+$ and $B^- \rightarrow J/\psi K^-$ events. The fit is varied by including only either $B^+ \rightarrow J/\psi K^+$ or $B^- \rightarrow J/\psi K^-$ signal MC, with the full variation taken as systematic uncertainty on the $J/\psi K$ charge asymmetry.
- **Fit stability**. Finally, the stability of this large sample fit is tested by repeating the fit with different starting values for the parameters, with respect to the baseline starting point described in section 10.1.2.

In each case the variation relative to the baseline fit is recorded, symmetrised with respect to the baseline result and used as an estimate of the systematic uncertainty.

Table 10.1 shows a breakdown of the contribution of the different systematic uncertainties

Table 10.1: Sources of systematic uncertainties associated with the fit performed to extract the normalisation channel yield. The systematics are combined by taking the maximum between the combination in quadrature of the positive and that of negative deviations.

Source of systematics	Value[%]
PRD1 Parameterisation	+1.7
PRD2 Parameterisation	-3
PRD3 Parameterisation	+1
Combinatorial Parameterisation	+1.7
Use B^+ signal sample	+2.2
Use B^- signal sample	+1.4
Weights	+0.7
Starting point	± 1.4
PRD composition	+2.4
Combined	4.8

and the resulting total systematic uncertainty, which amounts to 4.8%; the main contributions come from the functional models of the background components, the PRD composition and the charge asymmetry.

10.1.4 Result of the reference channel yield extraction

The $B^+ \rightarrow J/\psi K^+$ yield obtained is $N_{J/\psi K^+} = 334351 \times (1 \pm 0.3\% \pm 4.8\%)$, where the first uncertainty is statistical and the second is systematic.

As a cross-check of the fit behaviour, the relative $B^\pm \rightarrow J/\psi \pi^\pm / B^+ \rightarrow J/\psi K^+$ yield is assessed. The fit yields a ratio of $(3.71 \pm 0.09)\%$, where the uncertainty reported is statistical only, in agreement with the world average [11] of $(3.84 \pm 0.16)\%$.

10.2 Evaluation of the $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency ratio

The efficiency and acceptance terms for both the reference and signal channels enter in equation 5.1 through their ratio R_{Ac} , evaluated using MC samples corrected with the weighting procedure reported in section 7.6. This term is evaluated for the selection obtained combining the preselection described in section 7.3 and the BDT cut at 54% $B_s^0 \rightarrow \mu^+ \mu^-$ efficiency described in section 8.2.

The technique employed to evaluate the R_{Ac} is discussed first (sub-section 10.2.1). The associated systematic uncertainties are then evaluated (sub-section 10.2.2).

The relative signal fraction for the BDT bins introduced in section 8.2.3 with respect to the selection mentioned is discussed separately, together with the associated uncertainties (sub-section 10.2.3).

10.2.1 Efficiency ratio calculation

The efficiency ratio R_{Ac} is evaluated by separating the contribution of an acceptance term A and an efficiency term ϵ , according to the following formula:

$$R_{Ac} = \frac{A_{B^+ \rightarrow J/\psi K^+} \times \epsilon_{B^+ \rightarrow J/\psi K^+}}{A_{B_s^0 \rightarrow \mu^+ \mu^-} \times \epsilon_{B_s^0 \rightarrow \mu^+ \mu^-}} \quad (10.1)$$

The A term considers the effect of the cuts applied to the final state particles in the preselection (defined in section 7.3) $p_T^\mu > 6,4$ GeV, $|\eta_\mu| < 2.5$, $p_T^K > 1.0$ GeV and $|\eta_K| < 2.5$. These are measured with respect to a kinematic fiducial volume (also defined in section 7.3) $p_T^B > 8.0$ GeV and $|\eta_B| < 2.5$. Acceptances are hence evaluated by taking the ratio of the number of events passing the cuts on the final state particles and the number of events in the fiducial volume in the signal simulations.

The acceptance does not consider any effect due to the detector or the reconstruction procedure, therefore it is evaluated using generator level MC, without simulation of the detector response and reconstruction, applying the fiducial volume and cuts on the final state particles

Table 10.2: A terms for B^+ and B_s^0 channels for the 2015/16 analysis and for the Run 1 analysis. The latter was characterised by two different centre of mass energies, therefore two values are provided (2011 and 2012). The statistical uncertainty due to the finite size of the simulated samples is also reported.

	2011	2012	2015/16
$A_{B^+ \rightarrow J/\psi K^+}$	0.0819 ± 0.0003	0.0865 ± 0.0002	0.0734 ± 0.0001
$A_{B_s^0 \rightarrow \mu^+ \mu^-}$	0.3035 ± 0.0005	0.2902 ± 0.0005	0.2413 ± 0.0005

to the truth-level appropriate quantities. Two simulations are employed, one for the signal and the other for the reference channel, with the same properties of the *quark biased* samples defined in section 7.6.4.1. In order to take into account known MC mis-modellings, the QLC

and DDW corrections, defined in section 7.6.4 are applied to both the numerator and the denominator.

The resulting values for A are reported in table 10.2, together with the values obtained in the Run 1 analysis for comparison. The uncertainties reported account only for the finite size of the simulations employed. The comparison with the Run 1 acceptance terms shows the effect of the higher p_T cut applied to the leading muon, 6 GeV instead of 4 GeV, which lowers the value of both signal acceptances.

The ϵ term accounts for all the detector and reconstruction effects, such as the application of selections cuts to reconstructed quantities, trigger efficiency and reconstruction efficiency. ϵ is evaluated as the ratio between the number of events passing the full analysis selection¹ and the number of events passing the fiducial volume cuts on the final state particles applied to the appropriate truth level quantities (i. e. the numerator of the acceptance defined just before). As for the acceptance, the efficiency is evaluated using MC simulations, with QLC and DDW corrections applied to both numerator and denominator. Trigger and reconstruction efficiency corrections and pile-up reweighting, discussed in sections 7.6.3, 7.6.2 and 7.6.1, are also applied to the numerator, as it considers reconstructed quantities.

As introduced in section 7.2 (page 91), the $B_s^0 \rightarrow \mu^+ \mu^-$ signal efficiency needs to account for uncertainties on the B_s^0 lifetime. The $B_s^0 \rightarrow \mu^+ \mu^-$ signal MC was chosen to be generated with the average B_s^0 meson lifetime (1.615 ps). According to the SM predictions, the mass eigenstate contributing to $B_s^0 \rightarrow \mu^+ \mu^-$ decays is $B_{s,H}$ [53, 54]. The signal MC is thus reweighted to an effective lifetime of 1.533 ps, which increases the efficiency ratio of $B_s^0 \rightarrow \mu^+ \mu^-$ to $B^+ \rightarrow J/\psi K^+$ by 3.3%. Systematic uncertainties arising from this reweighting procedure are considered and found to be negligible compared to the other systematic uncertainties affecting $R_{A\epsilon}$ (discussed later in subsection 10.2.2)

The resulting values of ϵ are $0.0719 \times (1 \pm 0.4\%)$ and $0.1745 \times (1 \pm 0.7\%)$ for the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow \mu^+ \mu^-$ (before the lifetime reweighting) channels respectively. Before calculating the value of $R_{A\epsilon}$ and comparing the results with the same quantities obtained in the Run 1 analysis an additional step is required: as introduced in section 9.1, due to the noticeable data-MC disagreement found for the isolation related variables, the central value of $R_{A\epsilon}$ is corrected according to a re-weighting based on the $I_{0.7}^B$ variable. Table 10.3 summarises the resulting values for ϵ , $A \times \epsilon$ and $R_{A\epsilon}$, before and after the isolation re-weighting.

Where present, the first uncertainty is statistical and the second is systematic. The sources of systematic uncertainties affecting $R_{A\epsilon}$ are discussed in sub-section 10.2.2.

A comparison between the efficiency and $R_{A\epsilon}$ terms between Run 1 and the current analysis is not trivial, as several effects should be considered. The effect of the trigger prescales is not considered in $R_{A\epsilon}$ of the Run 1 analysis but explicitly inserted in the equivalent of formula 5.1, while it is taken into account in the 2015/16 calculation; the overall effect yields a reduction of a factor ~ 1.72 on $R_{A\epsilon}$, compatible with the ratio of the prescale factors of the triggers employed

¹ Which includes the preselection presented in section 7.3 and the cut on the BDT output at 0.2455 (which has 54% efficiency on the signal) presented in section 8.2.

Table 10.3: ϵ , $A \times \epsilon$ and $R_{A\epsilon}$ values for B^+ , B_s^0 and B_d^0 (where present) channels for Run 1 and 2015/16 samples. Where present, the first uncertainty is statistical and second one is systematic. The complete list of systematic uncertainties considered can be found in sub-section 10.2.2. The last row reports the values for the $B_d^0 \rightarrow \mu^+ \mu^-$ channel, evaluated using the appropriate MC sample.

channel	ϵ	$A \times \epsilon$	$R_{A\epsilon}$
B^+ Run 1	0.0928	$0.0080 \pm 0.0001 \pm 0.0013$	$0.180 \pm 0.001 \pm 0.093$
B_s^0 Run 1	0.1522	$0.0441 \pm 0.0002 \pm 0.0045$	
B^+ 2015/16	0.0719 ± 0.0003	$0.0053 \pm 0.0001 \pm 0.0063$	$0.1254 \pm 0.0001 \pm 0.009$
B_s^0 2015/16	0.1745 ± 0.0007	$0.0421 \pm 0.0003 \pm 0.0004$	
B^+ 2015/16 with B^+ isol. rew.	0.0670 ± 0.0003	$0.0049 \pm 0.0001 \pm 0.0005$	$0.1144 \pm 0.0009 \pm 0.0046$
B_s^0 2015/16 with B_s^0 isol. rew.	0.1783 ± 0.0007	$0.0430 \pm 0.0003 \pm 0.0037$	
B^+ 2015/16 with B^+ isol. rew.	0.0670 ± 0.0003	$0.0049 \pm 0.0001 \pm 0.0005$	$0.1176 \pm 0.009 \pm 0.0047$
B_d^0 2015/16 with B_s^0 isol. rew.	0.1734 ± 0.0007	$0.0419 \pm 0.0003 \pm 0.0036$	

in the analysis for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$.

Considering that the acceptances ratio A_{B^+}/A_{B_s} increased by about 3% with respect to the previous analysis, the main contribution to the increase of $R_{A\epsilon}$ is due to the efficiency terms. In particular, the ratio of efficiencies ϵ_{B^+} over ϵ_{B_s} increased by 16% with respect to Run 1. This is compatible with the harder cut applied to the leading muon transverse momentum in the 2015/16 analysis, which favours the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ channels with respect to $B^+ \rightarrow J/\psi K^+$, as the former shows a harder muon p_T spectrum, due to the additional cuts on the kaon.

In conclusion, even if the 2015/16 efficiency ratio calculation shows a different behaviour compared to the Run 1 analysis, the various differences can be addressed. This corroborates the validity of the $R_{A\epsilon}$ term for the 2015/16 analysis.

The last row of table 10.3 shows the ϵ , $A \times \epsilon$ and $R_{A\epsilon}$ values obtained using the $B_d^0 \rightarrow \mu^+ \mu^-$ MC sample, as expected the results are compatible with the $B_s^0 \rightarrow \mu^+ \mu^-$ channel within the uncertainty. The value obtained for the $B_s^0 \rightarrow \mu^+ \mu^-$ channel is therefore employed for both the signal processes².

10.2.2 Systematic uncertainties on $R_{A\epsilon}$

Several sources of systematic uncertainties are considered in the evaluation of $R_{A\epsilon}$; they originate from the reweighting procedure discussed in section 7.6 (muon trigger efficiency, pile-up re-weighting, muon offline efficiency, QLC and DDW), the kaon tracking efficiency and the re-

² The $R_{A\epsilon}$ term enters in the branching fraction extraction fit, discussed in chapter 12. The same efficiency ratio value is used for both B_s^0 and B_d^0 branching fractions extraction, with the additional lifetime corrections applied only to $B_s^0 \rightarrow \mu^+ \mu^-$.

sidual data-MC discrepancies already discussed in chapter 9.

The systematic uncertainties due to the MC reweighting procedure are evaluated singularly for each correction described in section 7.6. A toy-MC based study is run, varying the corrections being studied within their uncertainty and re-computing R_{Ac} after each toy. The RMS of the resulting distribution of the efficiency ratio is taken as systematic uncertainty.

The systematic uncertainties arising from the reweighting procedure are:

- trigger and offline efficiency corrections: 1.0%;
- PU re-weighting: 0.6%;
- kinematic corrections (QLC + DDW): 0.8%.

Notably, the systematic uncertainty due to the kinematic corrections is smaller than its Run 1 counterpart, which was found to be about 1.5%. The reduction in the uncertainty is compatible with the increased size of the $B^+ \rightarrow J/\psi K^+$ dataset used to compute the DDW, which is roughly 4 times larger than the one used in Run 1.

The numerator and denominator processes differ by the presence of one additional hadron in the final state. As a result, the ratio directly depends on the absolute tracking efficiency for this extra particle. The systematic uncertainty related to the kaon tracking efficiency is mostly due to inaccuracies in the modelling of passive material in the ID in simulations. It is evaluated by varying the detector model in simulations. The resulting systematic uncertainty varies between 0.4% and 1.5%, depending on the η range considered [204]. The largest value is conservatively used in the full eta range. This systematic only affects the reference channel, therefore its effect on R_{Ac} is 1.5%.

The last source of systematic uncertainties considered arises from the residual discrepancies between data and MC observed for the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels, after all the corrections discussed in section 7.6 are applied. For each of the BDT input variables (see section 8.2) the data-MC discrepancy observed in the reference channel (section 9.1) is used as a corrective weight. This is applied to the same variable in the MC samples entering the R_{Ac} calculation ($B_s^0 \rightarrow \mu^+ \mu^-$ or $B^+ \rightarrow J/\psi K^+$). The change in R_{Ac} observed when applying this correction for a given BDT input variable is taken as systematic uncertainty for that variable. The only exceptions are the $I_{0.7}$ and $\Delta\phi_{\mu\mu}$ variables:

- the reweighting of the $I_{0.7}$ variable for the $B_s^0 \rightarrow \mu^+ \mu^-$ channel is performed using $B_s^0 \rightarrow J/\psi \phi$ data and MC, as differences in the b quark hadronisation might alter the environment around the B candidate with respect to the reference channel.
- the reweighting of the $\Delta\phi_{\mu\mu}$ variable is not performed for the $B_s^0 \rightarrow \mu^+ \mu^-$ channel, given the difference in this distribution with respect to $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ events,

where the two muons come from a $J/\psi \rightarrow \mu^+ \mu^-$ process. Since this variable shows a large correlation with p_T^B and $P_L^{\min}$, the reweighting on these two is assumed to take into account also the $\Delta\phi_{\mu\mu}$ effects.

The total systematic uncertainty due to discrepancies in the BDT input variables is obtained by summing in quadrature the systematic uncertainty due to each variable. The resulting total systematic uncertainty on $R_{A\epsilon}$ due to this procedure after the isolation reweighting, which contributes an additional 1% effect, is 3.2%. The breakdown of the contribution of the single BDT input variables after the isolation reweighting is provided in table 10.4. The effect of the sys-

Table 10.4: Breakdown of the contribution of the BDT input variables to the systematic uncertainty of $R_{A\epsilon}$ after the isolation reweighting is applied. The systematic uncertainty relative to the $A \times \epsilon$ term for both B_s^0 and B^+ channels is shown, together with the combined effect on $R_{A\epsilon}$. No systematic uncertainty due to the reweighting of the $\Delta\phi_{\mu\mu}$ variable is evaluated for $A \times \epsilon_{\mu^+ \mu^-}$ as explained in the text.

Source	Contribution [%]		
	$A \times \epsilon_{\mu^+ \mu^-}$	$A \times \epsilon_{J/\psi K^+}$	$R_{A\epsilon}$
p_T^B	0.0	0.0	0.0
$\chi_{\text{PV,DV},xy}^2$	0.4	1.2	0.8
ΔR_{flight}	0.5	0.5	0.0
$ \alpha_{2D} $	0.5	0.7	0.2
L_{xy}	0.2	0.2	0.0
IP_B^{3D}	0.5	0.5	0.1
$\text{DOCA}_{\mu\mu}$	0.5	1.1	0.6
$\Delta\phi_{\mu\mu}$	-	0.0	0.0
$ d_0 ^{\text{max-sig.}}$	1.9	2.2	0.2
$ d_0 ^{\text{min-sig.}}$	3.6	1.9	1.7
$P_L^{\min}$	0.0	0.1	0.1
$I_{0.7}$	0.0	0.0	0.0
$\text{DOCA}_{\text{xtrk}}$	5.1	6.9	1.8
$N_{\text{xtrk}}^{\text{close}}$	2.5	2.8	0.4
$\chi_{\mu,\text{xPV}}^2$	3.5	2.1	1.5
Total	7.9	8.5	3.2

tematic uncertainty on the single BDT variables is shown separately for the $A \times \epsilon$ terms of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ channels and for $R_{A\epsilon}$: the combined effect on $R_{A\epsilon}$ is smaller than the effect on the single channel. This clearly shows the effect of the strategy of measuring the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fractions with respect to a reference channel in order to reduce the contribution of the systematic uncertainties.

The total systematic uncertainty for $B_s^0 \rightarrow \mu^+ \mu^-$ amounts to 7.9%, dominated by $\text{DOCA}_{\text{xtrk}}$ (5.1%), $|d_0|^{\text{min-sig.}}$ (3.6%) and $\chi_{\mu,\text{xPV}}^2$ (3.5%); regarding $B^+ \rightarrow J/\psi K^+$, the total systematic uncertainty for $A \times \epsilon$ amounts to 8.5%, dominated by $\text{DOCA}_{\text{xtrk}}$ (6.9%).

Table 10.5 shows a summary of the various contributions to the total systematic uncertainty on $R_{A\epsilon}$, together with the total value, obtained by summing in quadrature the single contributions. As for table 10.4, the systematic uncertainties related to the $A \times \epsilon$ terms are bigger compared to

Table 10.5: Summary of the systematic uncertainties in $R_{A\epsilon}$ after the isolation re-weighting procedure. The contributions to the $A \times \epsilon$ for both $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ are also shown. No systematic uncertainty due to the kaon tracking efficiency is associated with the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ process since it does not contain a kaon.

Source	Contribution [%]		
	$A \times \epsilon_{\mu^+ \mu^-}$	$A \times \epsilon_{J/\psi K^+}$	$R_{A\epsilon}$
Kinematic reweighting (QLC, DDW)	2.1	2.3	0.8
Muon trigger and reconstruction	2.5	2.6	1.0
Isolation reweighting	0.8	0.6	1.0
BDT input variables	7.9	8.5	3.2
Kaon tracking efficiency	-	1.5	1.5
Pile-up reweighting	0.8	1.2	0.6
Total	8.6	9.4	4.0

the systematic uncertainties on $R_{A\epsilon}$, confirming the systematic uncertainty reduction due to the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ measurement relative to a reference channel.

10.2.3 Systematic uncertainties on BDT bins

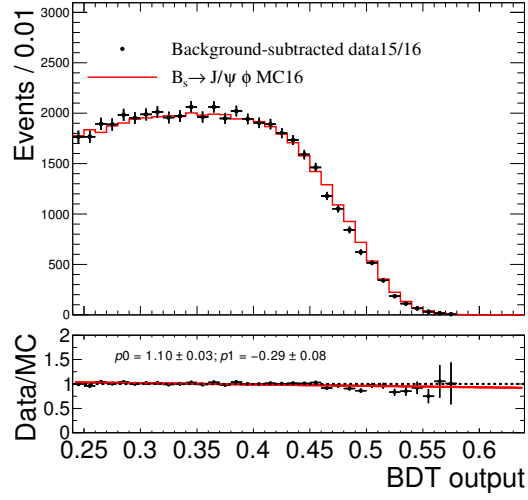
As introduced in chapter 5 and section 8.2, the candidates effectively³ employed in the signal yield extraction procedure are required to have a BDT output value larger than 0.2455, which corresponds to a signal efficiency of 54%. The resulting dataset is further divided in three subsets, referred to as BDT bins, with boundaries 0.2455, 0.3312, 0.4163 and 1, and a $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency of 18% each. Since this binning is based on the $B_s^0 \rightarrow \mu^+ \mu^-$ exclusive MC, additional uncertainties due to the BDT distribution in simulations, which can affect the relative population in the BDT bins, must be considered.

Two approaches are exploited to evaluate these uncertainties:

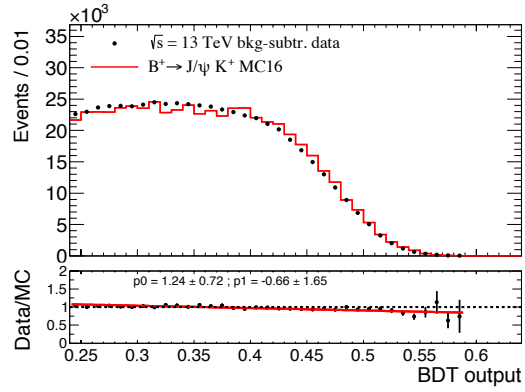
- the same re-weighting procedure of the BDT input variables used to calculate the uncertainty on the $R_{A\epsilon}$ term, described in the previous section, is used to evaluate also the uncertainty on the three BDT bins;
- the BDT output distribution from MC is compared to sideband-subtracted data for the the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels. The linear dependencies observed for the two channels are in turn used to reweight the $B_s^0 \rightarrow \mu^+ \mu^-$ simulated sample. Figure 10.2

³ The first BDT bin, with boundaries $0.1439 < \text{BDT output} < 0.2455$, is highly dominated by the background and does not add any sensitivity to the signal. For this reason, this bin is effectively only used to constrain the background models in the signal yield extraction.

shows the linear fits performed on the data/MC ratios of the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ BDT output distributions.



(a) $B_s^0 \rightarrow J/\psi \phi$ channel.



(b) $B^+ \rightarrow J/\psi K^+$ channel

Figure 10.2: BDT distributions observed on data for $B_s^0 \rightarrow J/\psi \phi$ (figure 10.2(a)) and $B^+ \rightarrow J/\psi K^+$ (figure 10.2(b)) candidates (filled markers), compared to the corresponding distribution from MC (red histograms). The inset at the bottom of the plots shows the data / MC ratio for each bin. The linear fit shown is used to evaluate the dependence of the data/MC ratio to the BDT output.

The first approach shows similar effects for both the $B^+ \rightarrow J/\psi K^+$ and $B_s^0 \rightarrow J/\psi \phi$ channels. Absolute variations of $\pm 1.0\%$, $\pm 2.4\%$, $\pm 4.4\%$ are obtained for the bins with boundaries 0.2455–0.3312, 0.3312–0.4163 and 0.4163–1 respectively. The maximum variations found with the second approach are equal to $+1.7\%$ and -2.3% respectively for the bins with boundaries 0.2455–0.3312 and 0.4163–1 respectively. The middle interval is found to be basically unaffected. In order to avoid correlations with the uncertainty on the R_{Ac} term, the second approach described is used to evaluate the uncertainty on the BDT bins. Figure 10.3 shows the effect of the linear reweighting used in the second approach on the $B_s^0 \rightarrow \mu^+ \mu^-$ BDT output.

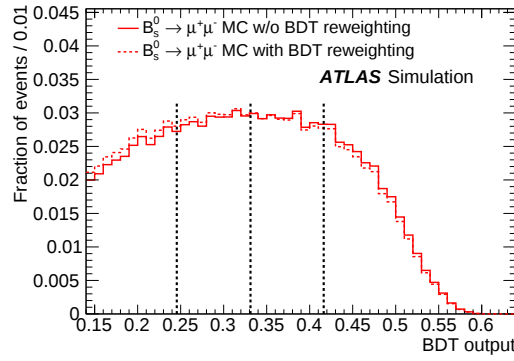


Figure 10.3: BDT output for the $B_s^0 \rightarrow \mu^+ \mu^-$, with the dashed histogram illustrating the effect of the linear reweighting on the BDT output discussed in the text. The vertical dashed lines correspond to the boundaries of the BDT intervals used in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal fit.

SIGNAL YIELD EXTRACTION

11

The extraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal yield is performed using an extended unbinned simultaneous maximum likelihood fit, as introduced in chapter 5. The same fitting procedure is then exploited to evaluate the signal branching fractions, by substituting the number of signal events with $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ and the normalisation terms from formula 5.1. The fit is carried out on the dimuon invariant mass distribution of the four contiguous BDT bins introduced in section 8.2. This is similar to the strategy employed by CMS [70] and LHCb [74]. The fit is performed on the dimuon invariant mass range defined at preselection level ($4766 \text{ MeV} < m(B) < 5966 \text{ MeV}$ from section 7.3). Such mass window is designed to be centred on the signal mass peaks, which fall into the blinded region, and has two sidebands ($4766 \text{ MeV} < m(B) < 5166 \text{ MeV}$ and $5526 \text{ MeV} < m(B) < 5966 \text{ MeV}$) which allow studies on some of the backgrounds that affect the analysis. The size of the sidebands is chosen to be large enough to contain enough candidates to allow accurate studies on the functional shapes of the backgrounds, without introducing additional background sources.

The fit is performed exclusively on data: for this reason all the preparatory studies are based on MC and, if possible, verified on data sidebands. This allows one to maintain the choice of the fitting procedure independent from data, thus preventing biases or double counting effects.

For the sake of readability of this chapter, the four BDT bins employed in the fit are numbered as follows:

- **bin 0**, with $0.1439 < \text{BDT output} \leq 0.2455$;
- **bin 1**, with $0.2455 < \text{BDT output} \leq 0.3312$;
- **bin 2**, with $0.3312 < \text{BDT output} \leq 0.4163$;
- **bin 3**, with $0.4163 < \text{BDT output} \leq 1.0$.

The addition of bin 0 is a novelty with respect to the Run 1 analysis, which employed 3 BDT bins. As explained in section 8.2, although its fit template includes signal components, bin 0 is effectively not contributing to the signal extraction: it was verified on toy-simulations not to have any impact on the signal yield. The reason for its inclusion is mostly to allow a better handling of the background and an overall stabilisation of the fit.

This chapter describes the studies performed to design and validate the fit. The models employed on the non-resonant backgrounds are identified first (section 11.1), followed by the studies performed on the resonant models (section 11.2), that include the signals and the peaking background. The final model designed with these components is then tested and validated (section 11.3). Finally, the procedure to substitute the number of signal events with the signal branching fractions is discussed in section 11.4.

11.1 Non-resonant background models

The models used to describe the non-resonant backgrounds are identified first.

The studies presented in this chapter heavily rely on the exclusive $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC and on the exclusive semileptonic MC, both introduced in section 7.2.

First the non-resonant backgrounds are classified according to their response to the BDT selection and their dimuon mass distribution (sub-section 11.1.1). Each background class is then studied individually on simulations, in order to identify a suitable functional form to be used in the four BDT bins (sub-section 11.1.2); these studies are carried out by splitting the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC in the various background categories, thanks to a MC-truth based selection. Significant background contributions will eventually be grouped in four categories: continuum 11.1.2.1, same-side and same-vertex 11.1.2.2, B_c 11.1.2.3 and semileptonics 11.1.2.4.

The combined fitting model for the non-resonant backgrounds is then obtained combining the different functional forms with individual free scale parameters; the model is chosen and validated first on MC and then on data sidebands before unblinding (sub-section 11.1.3).

The final fit model, designed to perform the simultaneous fit on the four BDT bins, is then introduced and tested (sub-section 11.1.4).

Finally, sub-section 11.1.5 provides a summary of the fit configuration of the non-resonant background components.

While the first part of the studies presented in this section is performed on the full invariant mass range of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC, the tests performed on the combined and simultaneous models (sub-sections 11.1.3 and 11.1.4) imply the application of the fitting procedure to both data and MC. In order to perform a fair comparison between the data and MC fit results, also the MC distribution is blinded for these studies. The absence of additional structures in the invariant mass signal region, that would require the addition of other functional forms, is guaranteed by the studies performed with the “unblinded” MC in sub-section 11.1.2.

In order to be coherent with the final fit, all fits performed in this section, unless explicitly stated, are extended unbinned maximum likelihood fits.

The fit quality of all the fits performed throughout this chapter are validated using a binned χ^2 test based on the uncertainty of the data, according to the following formula:

$$\chi^2 = \sum_{i=0}^{N_{\text{bins}}} \frac{(O_i - E_i)^2}{\sigma_i^2}, \quad (11.1)$$

where N_{bins} is the number of bins in a histograms, O_i is the observed number of events in a bin, σ is the uncertainty on O_i and E_i is the expected number of events in the bin, based on the fit results. Due to the properties of this statistical test, its result when applied to datasets with low statistics might not provide a meaningful estimation of the fit quality, usually over-estimating the compatibility. While a more appropriate approach is exploited for the goodness of fit of the final fit performed on unblinded data (chapter 12), the compatibility of the remaining datasets with the fits is still performed using the χ^2 test. This choice does not have any consequence on the analysis, as all the assumptions made in this chapter are then challenged with systematic uncertainties.

An additional issue in the calculation of the χ^2 test for the evaluation of the goodness of fit is due to the application of the PU weights described in section 7.6.1. As introduced in section 7.1, the datasets employed in the analysis were collected with triggers subjected to prescale. In particular the triggers were deactivated or heavily prescaled in case of high PU, usually at the beginning of the data taking runs, while they were running basically unprescaled at low PU. This effect is noticeable in figure 11.1, which compares the PU profile for the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^+ \rightarrow J/\psi K^+$ MC (respectively figures 11.1(a) and 11.1(b)) before and after the application of the PU weights. In both plots the preselection introduced in section 7.3 is applied to all the distributions and the histograms are normalised with respect to the one representing the PU profile before the reweighting. The PU weights obtained with the *PileupReweighting* tool,

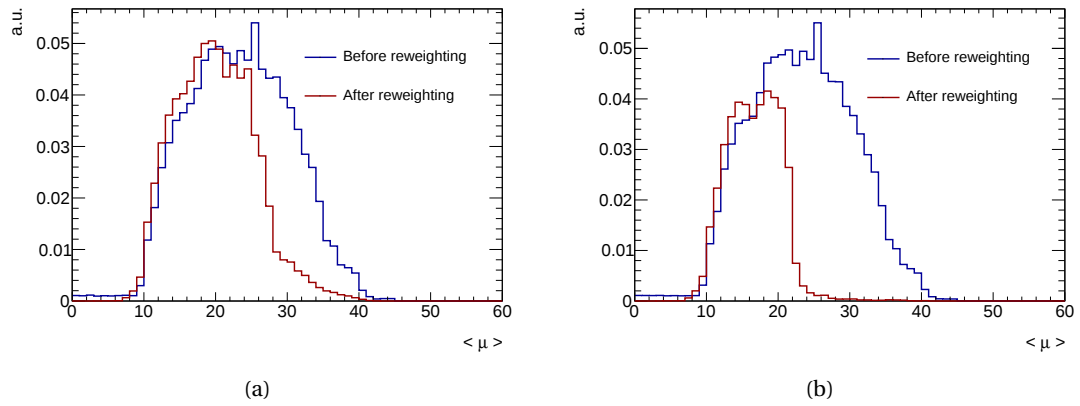


Figure 11.1: Comparison of the PU profile for the $B_s^0 \rightarrow \mu^+ \mu^-$ (figure 11.1(a)) and $B^+ \rightarrow J/\psi K^+$ (figure 11.1(b)) MC before and after the application of the PU weights. In both plots the preselection introduced in section 7.3 is applied to all the distributions and the histograms are normalised with respect to the one representing the PU profile before the reweighting.

in this case, can assume extremely small values when the simulated PU has high values. The effect of such small corrections on a large dataset would be un-noticed, as they do not have any contribution in e.g. the height of a bin or in its uncertainty¹. In case of small datasets, e.g. a bin with only one entry, the presence of these weights can provoke issues in the calculation of the χ^2 , as they count as zero entries with a corresponding uncertainty which is basically zero as well. This is because the evaluation of the uncertainty on a weighted distribution is an asymptotic process, which acquires meaning only if the effective number of weighted events is

¹ The uncertainty on a weighted distribution is evaluated as the sum in quadrature of the weights.

large enough.

This issue does not compromise the maximum likelihood fits performed in this section: since all the fits are unbinned, the contribution of the entries with small weights basically does not contribute to the total likelihood. On the other hand, this issue will appear in the calculation of the χ^2 test. Often the data and the fitted functional form are visually extremely compatible, but the resulting χ^2 is large. For this reason, a bad compatibility due to χ^2 tests applied to distributions which present this feature are not considered harmful. Anyway, in order to validate the fits performed on simulations, the Pearson χ^2 test is also considered. This test is based on the following formula:

$$\chi^2 = \sum_{i=0}^{N_{\text{bins}}} \frac{(O_i - E_i)^2}{E_i}, \quad (11.2)$$

where N_{bins} is the number of bins in a histogram, O_i and E_i are the observed and expected number of events in a bin. This test does not suffer the same issue as the one shown in formula 11.1, as it considers the uncertainty on the expected number of events instead of the uncertainty in the observed number of events.

11.1.1 Non-resonant background components

The large $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC (introduced in section 7.2) is used to identify and model the various sources of non-resonant background. The only exception is represented by the semileptonic backgrounds, which are studied separately using dedicated high-statistics exclusive MC samples introduced in section 7.2.

Events in the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC are classified based on the underlying b-hadron decay mode, using the MC truth information. Figure 11.2 shows the dimuon invariant mass and BDT output distributions for the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC; the analysis preselection (described in section 7.3) is applied, and the simulation is corrected for discrepancies from data according to the weighting procedure described in chapter 7.6. The reweighted $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC passing all the selection cuts is additionally normalised to the number of events in the data sidebands. Two sets of plots are shown, the first row presents the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC without applying the blinding procedure, while the events in the signal region are removed in the plots in the second row.

The various components highlighted in the plot are described in chapter 8, except for “unmatched”, “missing particles” and “b-onia”. “Unmatched” and “missing particles” are originated by dimuon candidates that could not be associated to MC truth particles. A study on the reconstructed muons showed that such events originate from Pile-Up (PU) vertices or from the Underlying Event (UE), for which the MC truth is not available. Given the combinatorial nature and the mass shape similar to the continuum background, these events are merged, for the sake of model building, with the “continuum” background sample.

The “b-onia” component is due to heavier $b\bar{b}$ bound states decaying into two muons and additional particles, e.g. $\Upsilon(1S) \rightarrow \tau^+ (\bar{\nu}_\tau \mu^+ \nu_\mu) \tau^- (\nu_\tau \mu^- \bar{\nu}_\mu)$; the dimuon invariant mass falls into the reconstructed region due to the loss of energy taken by the particles not considered in the dimuon candidate reconstruction. No events from this background source survive the BDT with signal efficiency at 72 %, as shown in figure 11.2(b). This background is therefore not con-

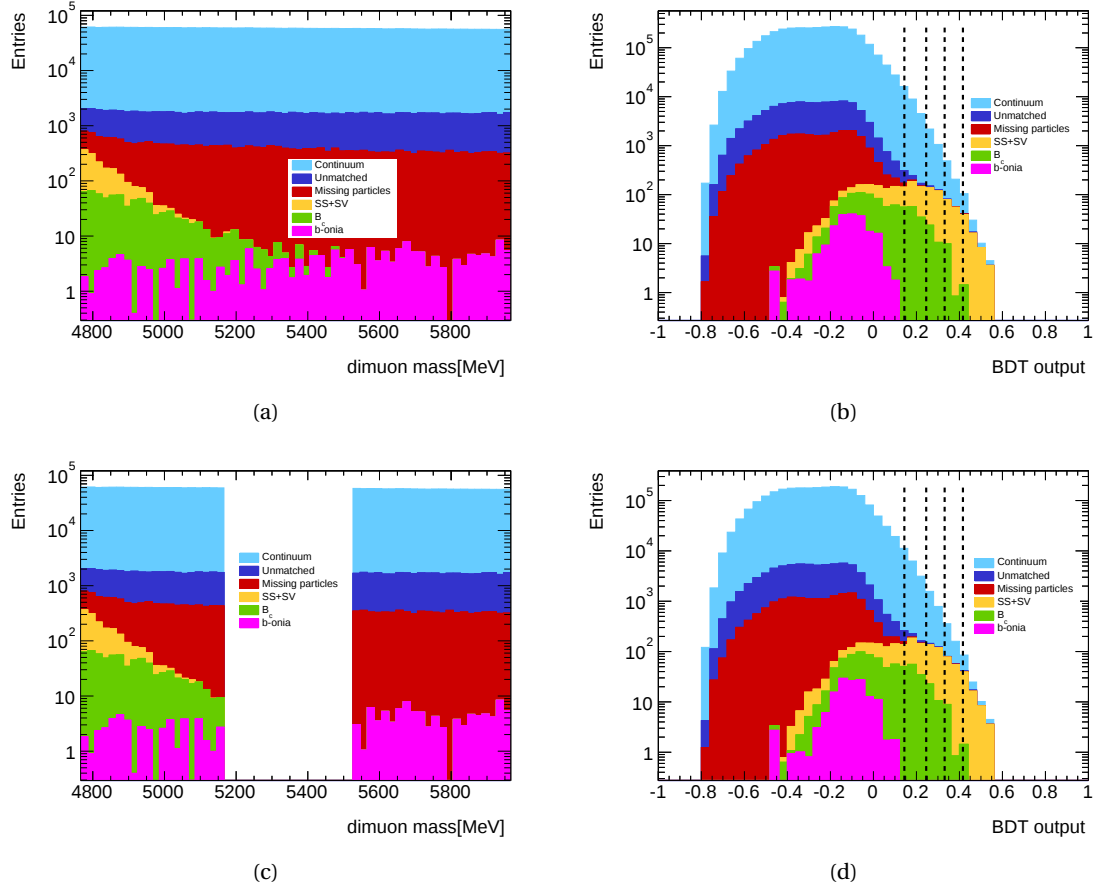


Figure 11.2: Breakdown of the composition of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC into the background components explained in the text. Two sets of plots are shown: the two plots in the first row show the mass distribution (figure 11.2(a)) and the BDT (figure 11.2(b)) output for the same events without applying the blinding procedure to the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC. The same plots are reported in the second row, but removing the events present in the signal region (figure 11.2(c) and 11.2(d)). The four vertical lines in the two plots showing the BDT output distribution mark the lower edge of the BDT bins employed in the analysis.

sidered in the signal yield extraction procedure; its possible small contribution is assumed to be absorbed in the continuum background.

As suggested by figure 11.2, the main contribution to the background is due to continuum events. These events show a smooth dependence on the dimuon invariant mass across the fit region. The other backgrounds show a much steeper dependence on the mass: due to phase space constraints, they populate only the low mass region.

As for the BDT distribution, continuum events clearly dominate the lower BDT ranges used in the analysis, with the other components (same-side and same-vertex, as well as B_c to a smaller extent) becoming more important in the higher BDT bins.

11.1.2 Non-resonant backgrounds parameterisation

This section will identify a suitable functional model for the non-resonant backgrounds. This will rely on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC components as well as the semileptonic b decay samples ($B_s^0 \rightarrow \pi^- \mu^+ \nu$, $B_s^0 \rightarrow K^- \mu^+ \nu$ and $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}$, introduced in section 7.2).

The expected yield of the different backgrounds is obtained by weighting the sidebands of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC to the data sidebands, while for the semileptonic backgrounds a different approach is applied, explained in sub-section 11.1.2.4.

The studies are performed considering the four BDT bins with 18% signal efficiency introduced in section 8.2; similar studies are performed including also BDT bins at lower values, but do not show any relevant feature.

11.1.2.1 Continuum background

The **combinatorial** (opposite-side) background, is described with a Chebychev first order polynomial:

$$f(x) = 1 + \alpha T_1(x), \quad T_1(x) = -1 + 2 \cdot \frac{x - 4766 \text{ MeV}}{5966 \text{ MeV} - 4766 \text{ MeV}}. \quad (11.3)$$

Figure 11.3 shows the fits performed on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC combinatorial sub-sample split in the four BDT bins already discussed: in all fits the Chebychev first order polynomial reproduces satisfactorily the MC shape. The resulting slope of the first order polynomial for all fits as a function of the BDT bin is shown in figure 11.4: the X axis is the average value of the BDT variable in each bin, while the Y axis is the fitted slope of the polynomial. The error bars represent the RMS of the BDT variable distribution in each bin (X axis) and the uncertainty on the fitted parameter extracted by the fit (Y axis). The binned fit applied to figure 11.4 shows that the continuum slope variation as a function of the BDT value is consistent with a linear model. As discussed later in section 11.1.3, this linear behaviour is confirmed on the data sideband fits, therefore the linear function is used as a constraint in the signal extraction fit. The systematic uncertainty arising from this assumption will be presented in section 11.3.3.

11.1.2.2 Same-Side and Same-Vertex background

The mass distribution of the two muons of the combined **Same Side and Same Vertex (SS+SV)** backgrounds peaks far below the signal region. The fit region contains the upper tail of the SS+SV distribution sculpted by kinematic limits and detector resolution effects. This background is represented with an exponential PDF $f(x) = \exp(\alpha x)$. Figure 11.5 shows the fits performed on the four BDT bins of the SS+SV component of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC.

All fits show a similar behaviour, and the exponential well reproduces the MC shape. The resulting slope of the exponential as a function of the BDT value is studied in figure 11.6. In analogy with figure 11.4, the X axis is the average value of the BDT variable in each bin, while the Y axis is the fitted slope of the exponential. Errors bars are the RMS of the BDT distribution in each bin and the uncertainty on the exponential slope obtained in the fit.

The exponential slope seems independent of the BDT bin: this is verified with the constant fit shown in figure 11.6. This feature is confirmed by the fit performed on the data sidebands, section 11.1.3, therefore the slope of the exponential is treated as a common parameter among the bins in the simultaneous fit. A systematic uncertainty, described in chapter 11.3.3, is included to account for this assumption.

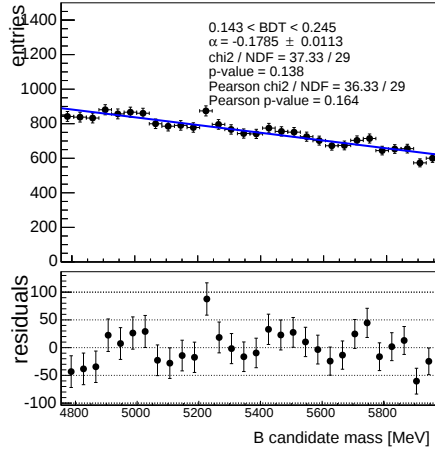
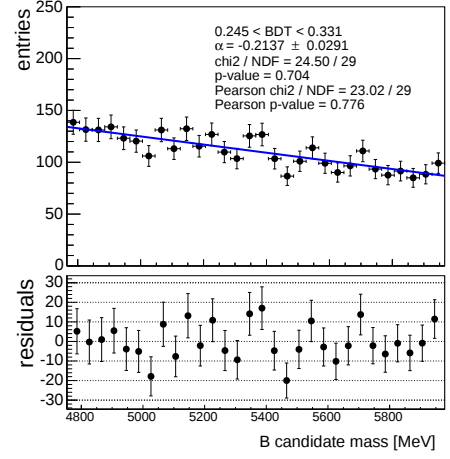
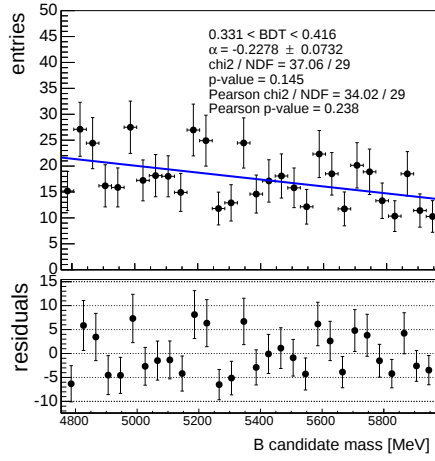
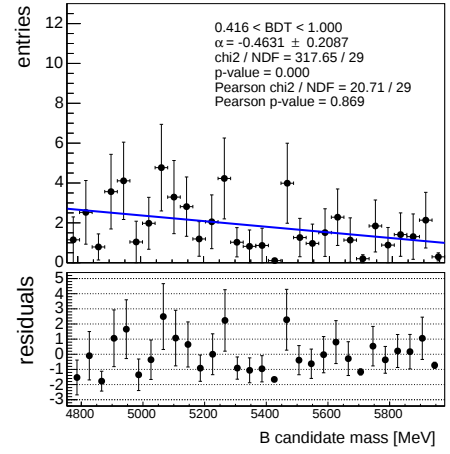
(a) BDT bin 0 ($0.1439 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.3: Fits performed on the combinatorial background component of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC in the four higher BDT bins. In all bins the fitting PDF, a Chebychev first order polynomial, well reproduces the MC. This is confirmed by the χ^2 values for bins 0, 1 and 2. As for bin 3, due to the low value of the PU weights employed on the simulated sample, the χ^2 test shows low compatibility; in any case, the Pearson χ^2 test ensures a good fit quality also for this bin.

11.1.2.3 B_c background

The B_c **background** is characterised by a small number of events; the expected contribution to the highest three BDT bins are respectively 1.5%, 1.8% and 0.2% of the background in the same BDT bins. It can be fitted with an exponential PDF, $f(x) = \exp(\alpha x)$, as shown in figure 11.7; The choice of this functional form is purely empirical. The exponential PDF reproduces the MC distribution reasonably well; there is a possible structure in the mass distribution at about 4950 MeV, that could be caused by a selection-driven sculpting of the dimuon mass distribution. This feature might be taken into account by fitting the distribution with a gaussian+exponential shape. However, given the small amount of expected B_c events and the fact that the usage of this more complicated model would reduce the leakage of B_c events in the signal region compared to the exponential fit, the gaussian+exponential model is conservatively avoided.

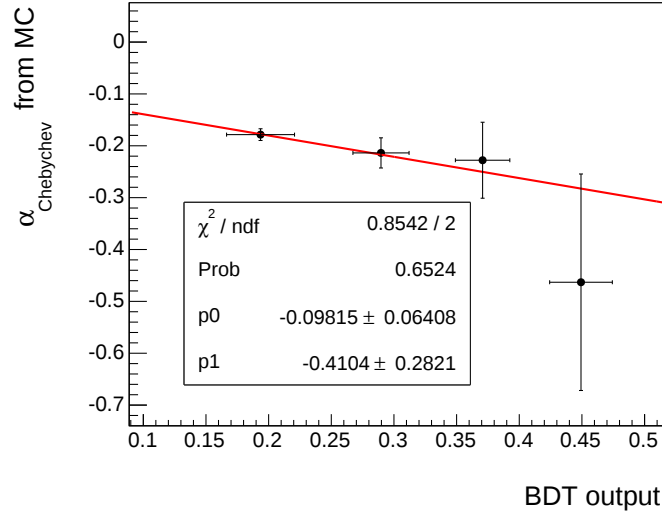


Figure 11.4: Chebychev first order polynomial slopes from the fits performed on the combinatorial component of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC in different BDT bins. The X axis is the average BDT variable for each bin. The horizontal error bars are the RMS of the BDT variable distribution in each bin, while the vertical error bars are the uncertainties on the polynomial slope extracted by the fit. The slope shows a linear behaviour with respect to the BDT variable, confirmed by a binned linear fit.

A similar treatment as the SS+SV slope can be performed in this case, checking the slope as a function of the BDT bins, but given the smaller yield of the B_c background with respect to combinatorial and SS+SV backgrounds, these events are expected to be absorbed by the Chebychev first order polynomial and exponential PDFs. No PDF is added to the fit model for this component of the background: the contribution of the B_c background is taken into account in the evaluation of the systematic uncertainties related to the fit, described in section 11.3.3.

The mass distribution of simulated B_c background events is susceptible to the choice of the form factors exploited in the MC generation. The introduction of a systematic uncertainty might be considered for this effect, but, as discussed earlier, no PDF is added to the fit model for this background, as its expected yield is small and it is expected to be absorbed by the other background components. For this reason, the systematic uncertainty associated to the form factors employed in the simulation is not considered.

11.1.2.4 Semileptonic background

The semileptonic background is characterised by low-multiplicity semileptonic b decays passing the final selections through a $h \rightarrow \mu$ mis-identification, in the limit of low energy neutrinos. In particular $B_d^0 \rightarrow \pi^- \mu^+ \nu$ and $B_s^0 \rightarrow K^- \mu^+ \nu$ are found to contribute significantly, together with $\Lambda_b^0 \rightarrow p \mu^- \bar{\nu}$. The mass distribution for the last process extends the closest to the signal region, but is highly suppressed by the low probability of protons to be mis-identified as muons.

The expected yield of these processes is evaluated counting separately for each BDT bin the number of sideband events passing tight and loose muon quality requirements. Assuming the muon mis-identification probability and efficiency evaluated in section 7.4, the estimated pur-

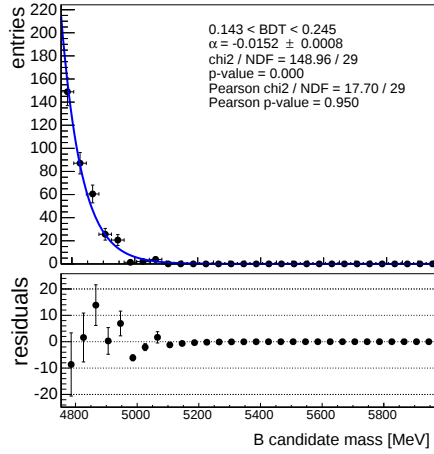
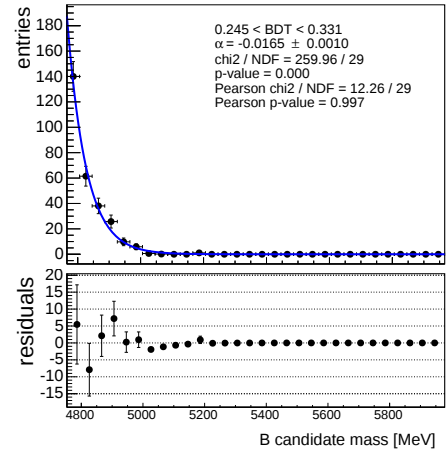
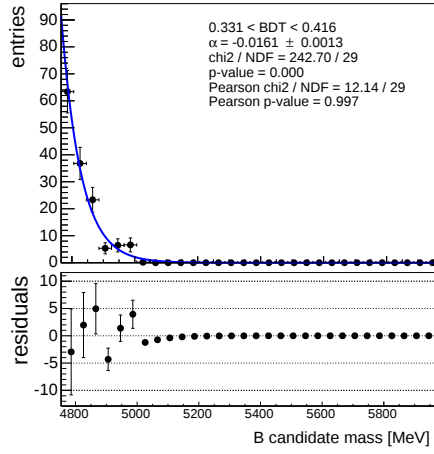
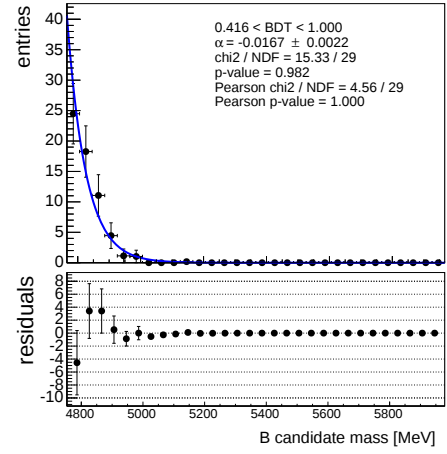
(a) BDT bin 0 ($0.143 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.5: Fits performed on the $SS+SV$ background component of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC in the four BDT bins. Due to the weights applied to the MC, some of the fit qualities (χ^2 tests) do not show a good agreement between the MC and the fits; Anyway, the fit quality estimated with the the Pearson χ^2 test show high compatibility between the data and the fit results.

ity is 97%. The purity is defined as the estimated number of dimuon candidates composed of real muons divided by the total number of candidates after the full analysis selection is applied. Performing the same estimation on the lower sideband, where the semileptonic decays are expected to appear, the number of events with a single-leg fake-muon is 30 ± 3 in the region with BDT output > 0.2455 . Conservatively assuming to saturate these with the semileptonic backgrounds, the contribution from these semileptonic modes is significantly smaller than the $SS+SV$ and the combinatorial background contributions in all bins of the BDT output. The exponential+polynomial background model is therefore assumed to absorb these small contributions in the default fit model. In order to assess the systematic uncertainty due to this assumption, the mass distribution of the combination of these three modes is fitted with a combination of an exponential plus a gaussian tail, without constraints or fixed parameters. This model will be used later in toy-simulations in section 11.3.3.

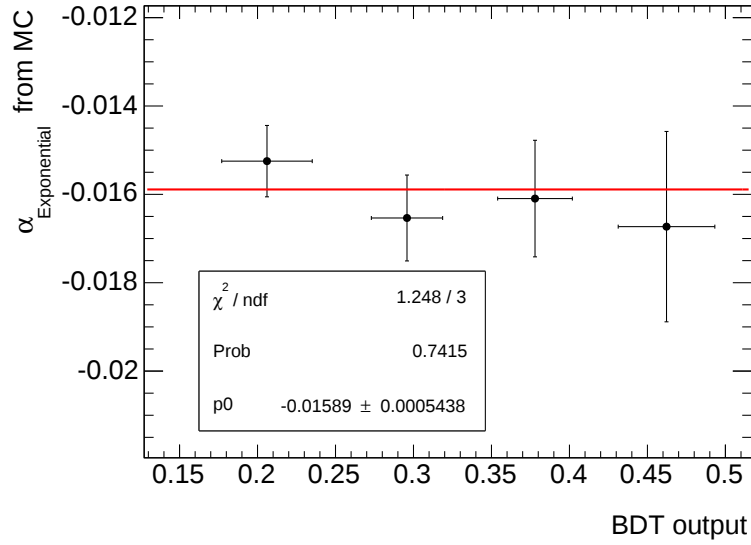


Figure 11.6: Exponential slopes from the fits performed on the SS+SV component of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC in the four BDT bins. The X axis is the average BDT variable for each bin. The horizontal error bars are the RMS of the BDT variable distribution in each bin, while the vertical error bars are the uncertainties on the polynomial slope extracted by the fit. The result of the fit performed on the bin at low BDT values is not shown, because only one event is present in the dataset. The slope shows a linear behaviour with respect to the BDT variable, confirmed by a binned pol0 fit.

Figure 11.8 shows the fits performed on the combination of the semileptonic MC samples merged and normalised according to the expected yields.

As for the B_c background, an additional systematic uncertainty might be considered, due to the uncertainty associated with the form factors employed in the MC generation. Anyway, also in this case, the PDF is not employed in the model, but only considered for systematic uncertainty studies. For this reason, the systematic uncertainty associated with the form factors employed in the MC generation is not considered.

11.1.3 Fit to MC and data sidebands with combined model

The model employed to describe the non-resonant backgrounds, studied in the previous subsection, is a combination of a Chebychev first order polynomial and an exponential PDFs, respectively modelling the smooth background that extends across the mass region and the background at low dimuon masses. The validity of this model is tested in this section.

The model is first used to fit the combinatorial and SS+SV components of the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC separately in each BDT bin, verifying the stability of different fits; the model is then tested on the data sidebands, again fitting each BDT bin separately. The behaviour of the slope parameters of the different PDFs as a function of the BDT bin is compared with what is observed in the MC, corroborating the constraints applied to the simultaneous fit.

Figure 11.9 shows the fits performed on the four BDT bins. The fits well reproduce the shape of the simulated sample, as confirmed by the χ^2 test. The last bin shows a not satisfying fit quality, due to the PU weights employed on the simulated sample. The same model is also used to fit

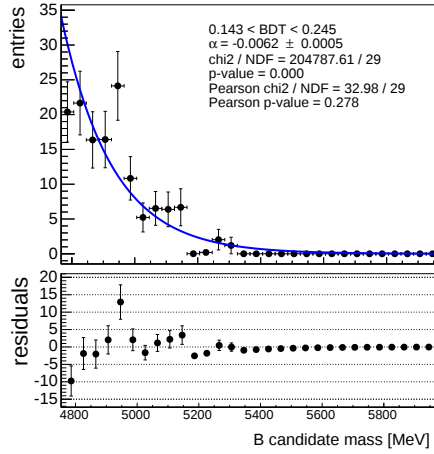
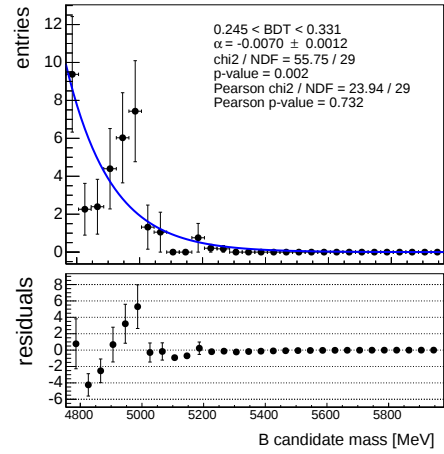
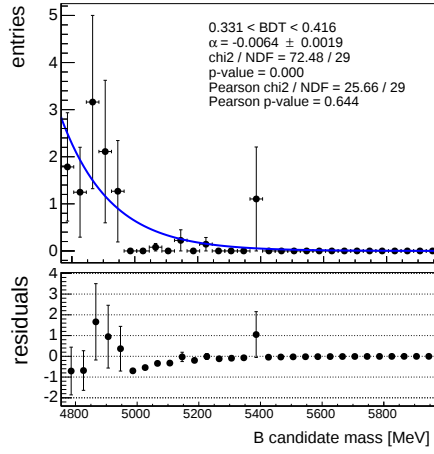
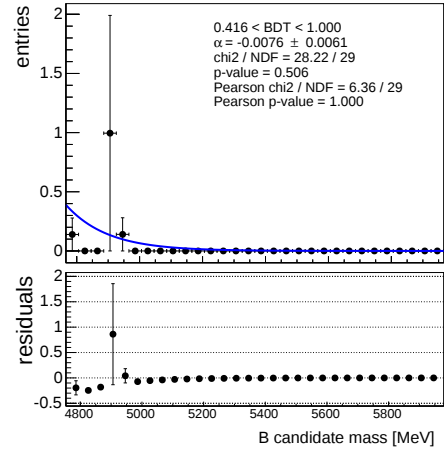
(a) BDT bin 0 ($0.1439 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.7: Fits performed on the B_c background component of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC in the four BDT bins. Due to the weights applied to the MC, some of the fit qualities (χ^2 tests) do not show a good agreement between the MC and the fits, even if the fitted PDF is able to reproduce the shape of the MC. This is confirmed by the Pearson χ^2 test, which shows high compatibility between the data and the fit results.

the data-sidebands, the resulting fits are shown in figure 11.10; again the fit model well reproduces the data, as confirmed by the χ^2 test.

The resulting slope parameters are shown in figure 11.11, where 11.11(a) shows the slopes of the first order Chebychev polynomials and figure 11.11(b) shows the slopes of the exponentials. In both plots the X axis is the average value of the BDT variable in each bin, while the Y axis is the fitted slope of the PDF. The error bars are the RMS of the BDT distribution in each bin and the uncertainty on the slope parameter obtained with the fit. Three sets of points are plotted: the black points are obtained with the fits performed separately on the corresponding MC components (first order Chebychev polynomial for the continuum background, section 11.1.2.1, and exponential for the SS+SV background, section 11.1.2.2), the red points are obtained with the fits performed with the combined model on the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC and

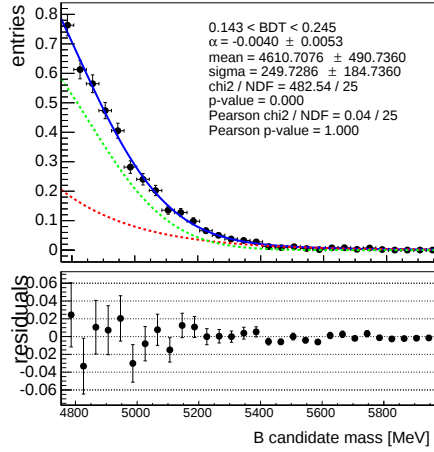
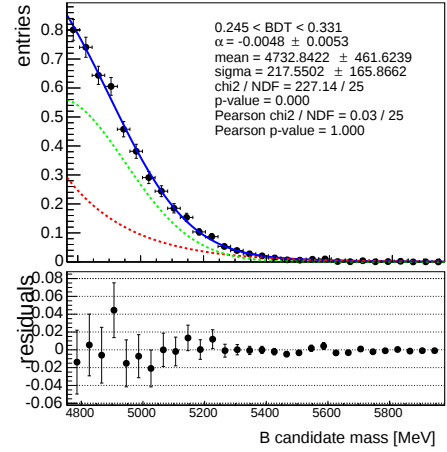
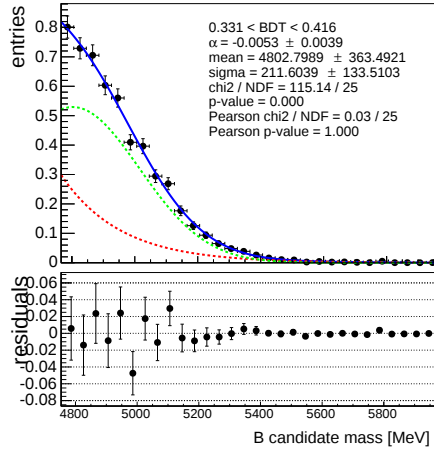
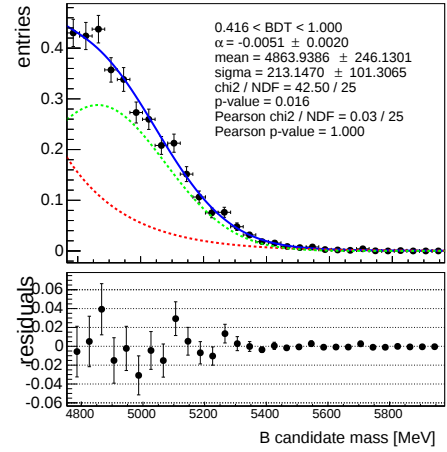
(a) BDT bin 0 ($0.1439 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.8: Fits performed in the four BDT bins to the semileptonic background MC samples merged together and normalised according to what is described in the text. The fit model employed is a combination of an exponential (red dashed line) and a gaussian tail (green dashed line). The solid line shows the combined model. The relative normalisation of the two PDFs is a free parameter in the fit. Due to the weights applied to the MC, some of the fit qualities (χ^2 tests) do not show a good agreement between the MC and the fits, even if the fitted PDF is able to reproduce the shape of the MC. This is confirmed by the Pearson χ^2 test, which shows high compatibility between the data and the fit results.

green points are obtained with the fits on data sidebands.

The slopes of the polynomial, figure 11.11(a), show compatibility between the two sets of fits performed on MC. This ensures that the combined Chebychev+exponential model does not modify the fitting procedure. Additionally, the slope parameters are also compatible with the fits performed on data; the last data bin shows error bars considerably larger than in the MC case, due to the limited statistics available. The polynomial slope as a function of the BDT output is compatible with a line, confirming the choice of this linear constraint in the signal extraction.

The exponential slopes appear again independent on the BDT bin in the fits performed on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC, confirming what was observed in section 11.1.2.2. This constraint is as-

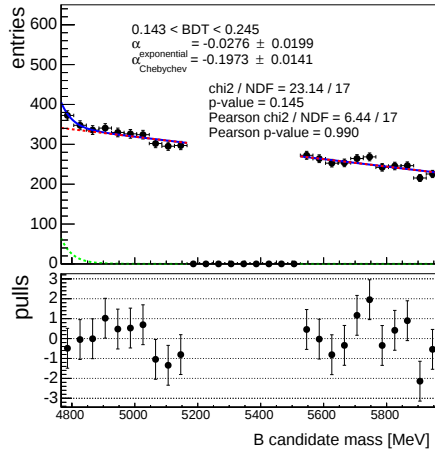
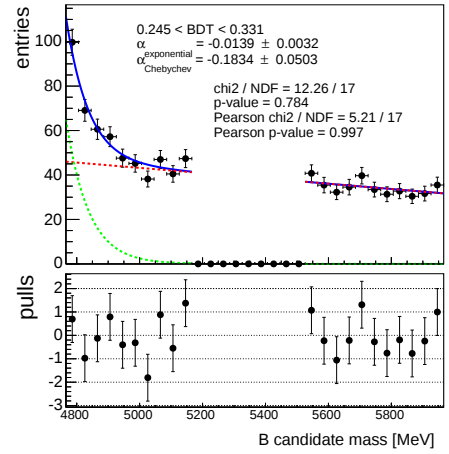
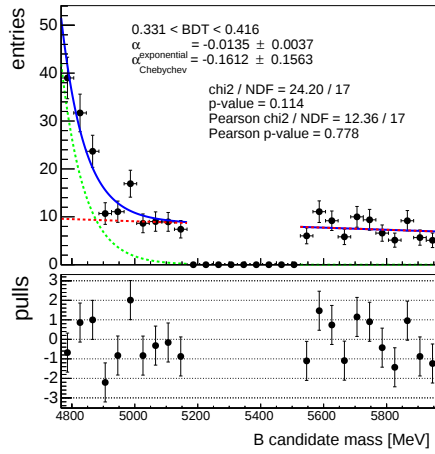
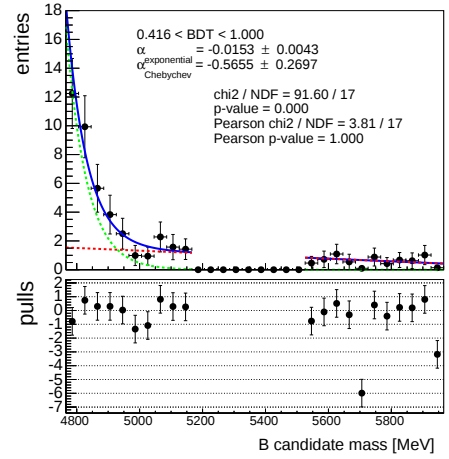
(a) BDT bin 0 ($0.1439 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.9: Fits performed on the combinatorial and SS+SV background components of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC in the four higher BDT bins. The fit model is composed of an exponential PDF and a Chebyshev first order polynomial, and it well reproduces the data. This is confirmed by the χ^2 values for bins 0, 1 and 2. As for bin 3, due to the weights employed on the simulated sample, the χ^2 test shows a low compatibility, even if the fitted PDF is able to reproduce the shape of the MC. This is confirmed by the Pearson χ^2 test, which shows high compatibility between the data and the fit results.

sumed in the signal extraction. Although still constant, the exponential slope seems to differ between data and MC. This is accounted for in the unblinded fit, where the background shape parameters are free to vary, with no input from the MC.

The assumptions regarding the two constraints on the polynomial and exponential PDFs are challenged with a systematic uncertainty, as discussed in section 11.3.3.

11.1.4 Fit to MC and data sidebands with simultaneous model

The final test on the non-resonant background models is the validation of the BDT dependency of the background slopes both on $b\bar{b} \rightarrow \mu^+\mu^- X$ MC and data sidebands when including the constraints on the polynomial and exponential slopes as part of the simultaneous fit of the

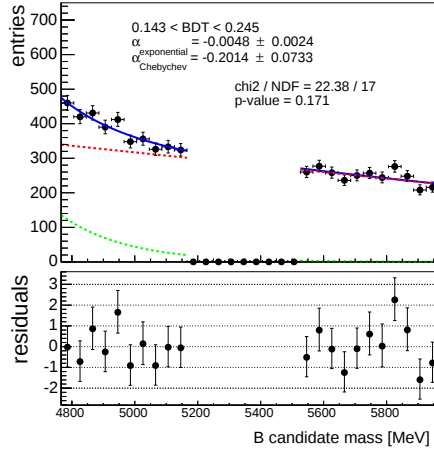
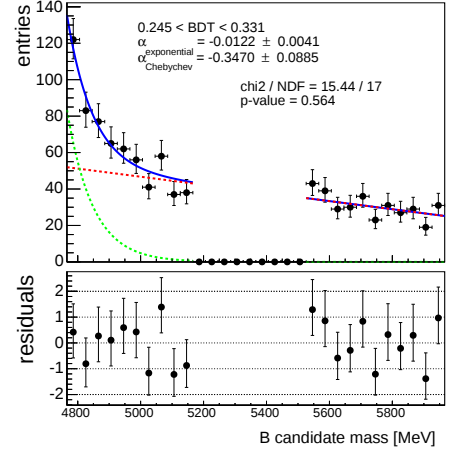
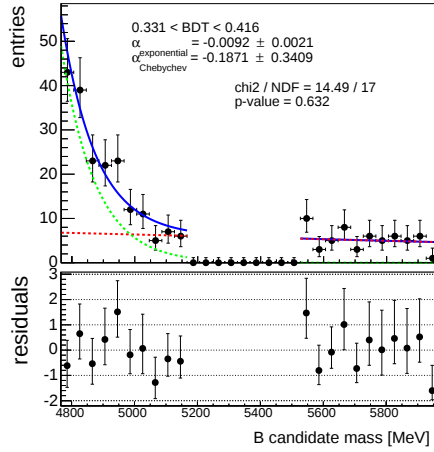
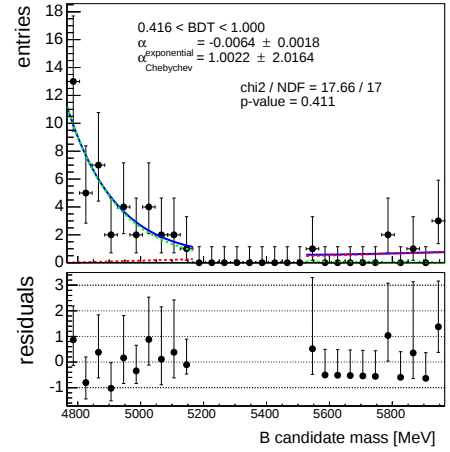
(a) BDT bin 0 ($0.1439 < \text{BDT value} < 0.2455$)(b) BDT bin 1 ($0.2455 < \text{BDT value} < 0.3312$)(c) BDT bin 2 ($0.3312 < \text{BDT value} < 0.4163$)(d) BDT bin 3 ($0.4163 < \text{BDT value} < 1$)

Figure 11.10: Fits performed on the data sidebands in the four higher BDT bins. The fit model is composed of an exponential PDF and a Chebychev first order polynomial. The model well reproduces the data, as shown by the fit quality (χ^2 test).

background shapes to the analysis four BDT bins. This will allow the comparison of the results on data to the ones on MC, as well as the simultaneous fit against the fits to the four individual bins. The resulting parameters are then compared with the results of the fits performed in the previous sections.

The slopes of the exponential PDFs involved in the simultaneous fit are constrained to be the same in all BDT bins, while the polynomial slopes are constrained to depend linearly on the average BDT value of the data points in each bin. No constraints are applied to the normalisations of the PDFs.

Both the fits on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC sidebands and on the data sidebands fully converge and show a good fit quality. The individual mass plots of the fits are not shown, as they do not differ visually from the ones shown in the previous section.

Figure 11.12 shows the slope of the Chebychev polynomial 11.12(a) and the slope of the exponential 11.12(b) as a function of the average BDT output in each BDT bin. In each plot the

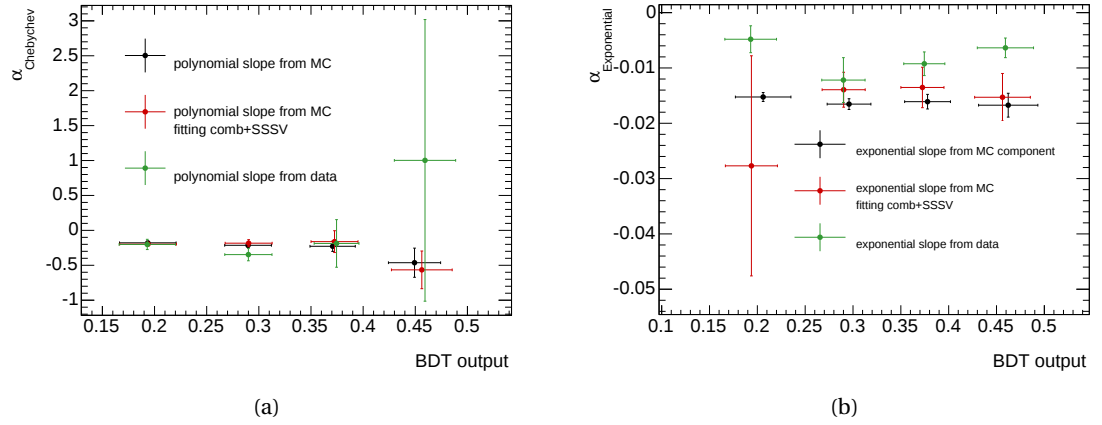


Figure 11.11: The plots show the slopes of the first order Chebychev polynomial 11.11(a) and exponential 11.11(b) as a function of the average BDT value of each BDT bin. The red points are the values of the slopes obtained fitting the combinatorial and SS+SV components of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC with the combined model, the black points are the slopes obtained with the fits on the single MC components and the green points are obtained with the fit on the data sidebands.

results from the fits on the single bins on the MC sidebands (black points) and on the data sidebands (red points) are shown, superimposed on the results of the linear fits performed on data (green) and MC (blue).

The lines showing the simultaneous fit results for the polynomial slopes and the exponential

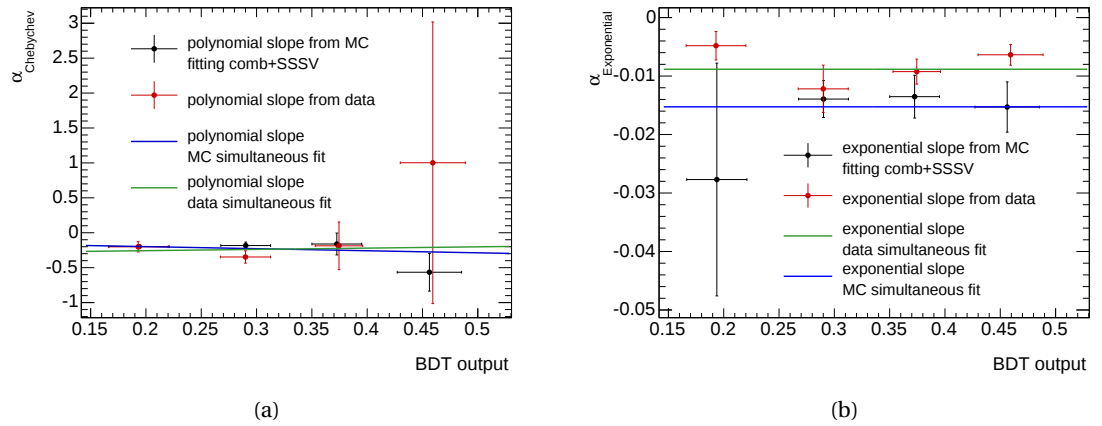


Figure 11.12: Slopes of the first order Chebychev polynomial 11.12(a) and exponential 11.12(b) as a function of the average BDT value of each BDT bin. The black points are obtained fitting the combinatorial and SS+SV components of the $b\bar{b} \rightarrow \mu^+\mu^- X$ MC with the combined model, the red points are obtained with the fits on the data sidebands and the two lines are obtained with the simultaneous fit on the MC (blue) and data sidebands (green).

slopes show compatibility with the points obtained with the single fits on the same datasets. The level of agreement between the results of the simultaneous fits on data and MC is shown in figure 11.13. The parameters of the lines that constrain the polynomial slopes are compared in figure 11.13(a): the two sets of parameters are consistent within their uncertainty. The slopes of the exponential PDFs obtained with the simultaneous fits are shown in figure 11.13(b): the two parameters are not compatible. This is interpreted as residual inconsistencies between MC and data. Such disagreement does not constitute an issue for the analysis, as the fit on unblinded

data does not depend on any of these results. In fact, the parameters used as constraints are left free to be determined by the fit. These effects are in any case considered in the evaluation of the systematic uncertainties arising from the fitting procedure (section 11.3.3).

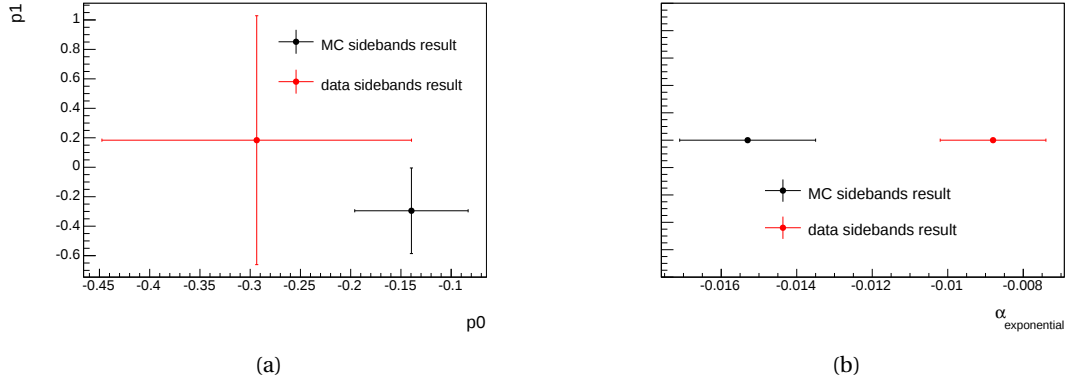


Figure 11.13: Data - MC comparison of the parameters that constrain the combinatorial and exponential slopes in the simultaneous fit. Parameters of the line that constrains the polynomial slope ($p0 + p1 \cdot < \text{BDT} >$) for data (red) and MC (black) shown in the $p0 - p1$ plane with their uncertainty in figure 11.13(a). The common exponential slope to all the BDT bins is shown in figure 11.13(b).

11.1.5 Summary of the non-resonant background configuration

The baseline configuration of the non-resonant models in signal yield is, in conclusion, made of the following two components:

1. **linear background PDF**; first order Chebychev polynomial with the normalisation extracted independently in each BDT bin. The slopes of the polynomials depend linearly on the bin averaged BDT value. Systematic uncertainties on the final fit result are derived for this assumption;
2. **exponential background PDF**: exponential dependence on the mass. The normalisation is extracted independently in each BDT bin, while the slope is assumed to be uniform. This assumption is challenged as part of the systematic studies.

The following parameters are free in the model:

- the number of linear background events in each BDT bin (4 free parameters);
- the number of exponential background events in each BDT bin (4 free parameters);
- the two parameters describing the linear constraints for the polynomial slopes;
- the common slope of the low-mass background.

11.2 Signal and peaking background models

This section is dedicated to the models employed in the signal yield fit to describe the $B_s^0 \rightarrow \mu^+ \mu^-$ and the peaking background shapes.

As for the non-resonant backgrounds, the functional forms are studied on MC, but in this case it is not possible to check the behaviour of the PDFs on data, because they all end up in the blinded region. Systematics uncertainties are added to account for assumptions on the models.

The first functional models studied are the signal models for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ (sub-section 11.2.1). The model employed for the peaking background is then discussed (sub-section 11.2.2), with a dedicated study performed to estimate its normalisation (sub-section 11.2.2.2).

11.2.1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ models

The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ mass shapes are described with a superposition of two gaussians with independent means, in order to take into account radiative tails. The parameters of these distributions are extracted from MC and the mass PDF is found to be independent from the output of the BDT used to reduce the combinatorial background. The signal MC samples have much higher statistics than the $\mathcal{O}(100)$ events expected on data, and would be sensitive to nuances in the signal shape parameterisation which are totally irrelevant for the fit on data. For this reason, a bootstrap technique is employed in the determination of the parameters of the functional forms. This approach allows the evaluation of the functional forms of the signals with a simple model, maintaining its independence from possible statistical fluctuations that might appear when fitting a single sample with a small size.

Four random subsets of $\mathcal{O}(1000)$ events are extracted from the MC, representing the four BDT bins of the analysis. These events are then fitted with a simultaneous extended unbinned maximum likelihood fit to extract the parameters of the double gaussian PDF, with the constraint of maintaining the same shapes across the four bins. This procedure is repeated 1000 times and the averages of the fit results are taken as parameters of the signal PDFs.

As introduced in section 7.3, the information on the B vertex can be evaluated combining the ID and MS tracking information or using the ID hits alone. The invariant mass distribution of the $B_s^0 \rightarrow \mu^+ \mu^-$ MC sample is found to be narrower when using the first approach, which is therefore employed in the analysis. The resulting measurement on MC of the B mesons ID and MS tracking information based invariant mass shows a bias of about 9 MeV with respect to the expected PDG values [11]. This offset is not observed when using the ID-based tracking and is thus most likely due to calibration effects on the combined muon tracking.

The bias on the combined-muon-based B invariant mass is checked on the J/ψ mass distribution from $B^\pm \rightarrow J/\psi K^\pm$ data and MC. In both cases the J/ψ mass is displaced with respect to the PDG mass of about 5 MeV. The J/ψ mass difference between MC and data is about 0.5 MeV, smaller than the uncertainties on the difference of the two means.

The same mass offset is observed in both data and MC Υ samples [205] when evaluating the

dimuon invariant mass with the combined-muon-based approach.

These checks assure that the same mass bias is present consistently in both data and MC. The means of the gaussians for the signal models are therefore fixed to the values extracted from the MC. Possible shifts will be taken into account as part of the mass scale systematic uncertainty (section 11.3.3).

All the BDT bins employed in the analysis are designed to each contain the same number of signal events (section 8.2.3). Given the insensitivity of bin 0 to the signal contribution, its signal content is fixed, for consistency, to the average content of the other three bins. Systematic uncertainties on the relative fractions of signal in the BDT bins are derived from the studies discussed in section 10.2.3 and are implemented as gaussian constraints. Since the systematic uncertainties associated to the BDT with a 54% efficiency on the $B_s^0 \rightarrow \mu^+ \mu^-$ signal are already taken into account in the R_{Ac} term, the sum of the relative efficiencies of bins 1, 2 and 3 is constrained to 1. This is effectively achieved by implementing the gaussian constraints to the yield of bins 2 and 3 relative to bin 1.

The resulting shape parameters for the functional models of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals are summarised in table 11.1; the uncertainty on the parameters is the RMS of the various fit results obtained with the bootstrap technique. The widths of the gaussians employed are larger than what found in the Run 1 analysis ($\sigma_1 = 67.7$ MeV and $\sigma_2 = 142.9$ MeV for $B_s^0 \rightarrow \mu^+ \mu^-$ and $\sigma_1 = 64.9$ MeV and $\sigma_2 = 135.5$ MeV for $B_d^0 \rightarrow \mu^+ \mu^-$). This is consistent with the higher amount of material in the ID with respect to Run 1, that increases the multiple scattering of particles at low p_T , effectively resulting in a loss of resolution.

Figure 11.14 shows the B_d^0 and B_s^0 PDFs superimposed on the invariant mass distribution for

Table 11.1: Shape parameters for the PDFs of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals. The values are obtained with the bootstrap technique described in the text. The uncertainties are obtained taking the RMS of the various fit results.

	$B_s^0 \rightarrow \mu^+ \mu^-$	$B_{(s)}^0 \rightarrow \mu^+ \mu^-$
μ_1	5357.7 ± 1.5 MeV	5270.6 ± 1.5 MeV
σ_1	83.0 ± 1.9 MeV	79.2 ± 2.2 MeV
μ_2	5257.0 ± 18.3 MeV	5197.1 ± 14.5 MeV
σ_2	193.8 ± 10.5 MeV	172.9 ± 9.6 MeV
fraction of gaussian 1	0.88 ± 0.02	0.83 ± 0.03

the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ MC samples. Both templates are normalised to the number of events expected according to the SM: $N_{B_s^0} = 91$ and $N_{B_d^0} = 10$ (obtained inverting formula 5.1).

11.2.2 Peaking background

The peaking background is composed of $B \rightarrow hh'$ decays, mainly $B_s \rightarrow K^+ K^-$, $B_d \rightarrow K^\pm \pi^\mp$ and $B_d \rightarrow \pi^+ \pi^-$, with both hadrons mis-identified as muons. Due to the mass distortion related to the $K/\pi \rightarrow \mu$ assignment the mass distribution of these events is substantially superimposed

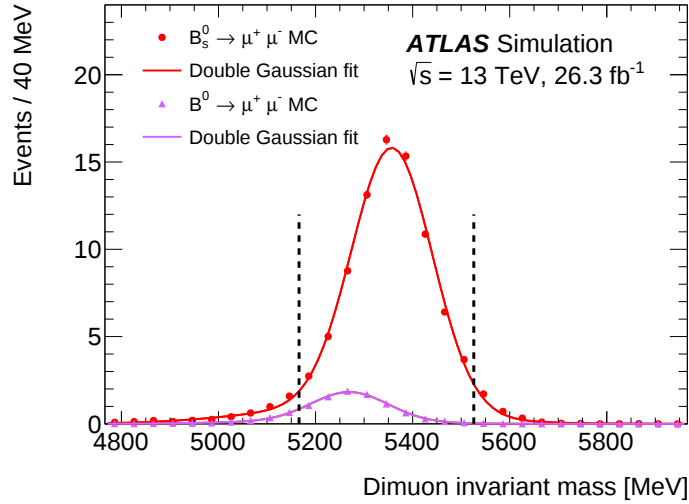


Figure 11.14: Dimuon invariant mass distribution for the B_s^0 and B_d^0 signals from simulation, normalised to the SM prediction for the expected yield with an integrated luminosity of 26.3 fb^{-1} . The double-Gaussian $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ PDFs obtained with the bootstrap technique discussed in the text are superimposed.

with the B_d^0 signal.

The functional model for $B \rightarrow hh'$ is evaluated on simulations (sub-section 11.2.2.1) and particular attention is dedicated to its normalisation (sub-section 11.2.2.2). In fact, the yield of this background needs to be constrained in the unblinded fit, in order to avoid the leakage of events from the B_d^0 peak to this background (or the opposite).

11.2.2.1 Peaking background model

As for the signal functional forms, the PDF for the peaking background is evaluated on simulations. The $B \rightarrow hh'$ MC sample, however, does not present sufficient statistics to directly study the invariant mass distribution of the double-fake events by applying the same muon quality requirements used on data. This means that the invariant mass calculated combining ID and MS tracking information for the $B \rightarrow hh'$ candidates can not be directly accessed. The invariant mass distribution of the $B \rightarrow hh'$ background is therefore studied using the ID-based calculation of the vertex mass, without requiring the presence of fake-muons in the final state.

Before evaluating the shape parameters for the $B \rightarrow hh'$ PDF, several effects have to be taken into account. The offset found in the combined ID and MS (combined-muon) based mass distribution of the signals (section 11.2.1) is not present in the ID-only based mass calculation. As discussed later in this section, this effect is taken care of by applying the B_d^0 mass offset to the $B \rightarrow hh'$ mass distribution. Additionally, the requirement of two muons in the final state can modify the invariant mass distribution. This effect can be due to the trigger selection (section 7.1), which requires the presence of two muons in the MS, and to the *tight* muon-ID requirements applied in the analysis preselection (section 7.3).

The first effect is studied on the B_s^0 and B_d^0 mass distributions in simulation as a function of the B mass itself. The combined-muon-based – ID-based mass shift obtained from the

$B_d^0 \rightarrow \mu^+ \mu^-$ MC is applied to the ID-based mass distribution of the $B_s^0 \rightarrow \mu^+ \mu^-$ simulation. The $B_s^0 \rightarrow \mu^+ \mu^-$ MC combined-muon-based distribution and ID-based mass distribution before the application of the B_d^0 mass shift are shown in figure 11.15(a), while figure 11.15(b) shows the same $B_s^0 \rightarrow \mu^+ \mu^-$ combined-muon-based mass distribution, superimposed on the ID-based mass distribution shifted according to what found in the B_d^0 MC. The resulting shape for the shifted ID-based mass is consistent with the combined-muon-based $B_s^0 \rightarrow \mu^+ \mu^-$ mass distribution. The same test is repeated inverting the roles of B_s^0 and B_d^0 , and the same result is obtained.

The mass offset that affects the fake dimuon invariant mass distribution might differ from the

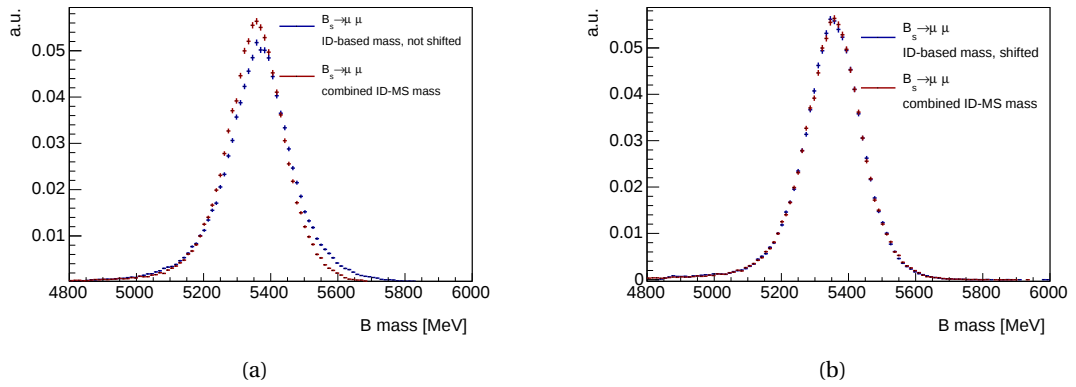


Figure 11.15: Application of the $B_d^0 \rightarrow \mu^+ \mu^-$ combined-muon-based mass offset measured in simulations on the ID-based mass distribution of the $B_s^0 \rightarrow \mu^+ \mu^-$ MC. Figure 11.15(a) shows the combined-muon-based and ID-based mass distributions for before the application of the B_d^0 mass shift, while figure 11.15(b) show the same combined-muon-based mass distribution together with the shifted B_s^0 ID-based mass distribution.

one that affects the invariant mass distribution of real dimuons, due to the possible differences between the kinematic properties of the reconstructed fake muon in the MS and the track in the ID. This is checked on the $B \rightarrow hh'$ simulation by comparing the reconstructed kinematic properties of the fake muons to the kinematic properties of the relative tracks in the ID. These are found to be compatible, therefore additional effects on the mass offset, if present, are extremely small.

The potential distortion of the invariant mass distribution of the $B \rightarrow hh'$ candidates due to the requirements on the presence of two muons is checked by separating the possible contributions due to trigger and muon identification requirements. The first effect is checked on the $B_s^0 \rightarrow \mu^+ \mu^-$ mass distribution in simulations, which is compared before and after the application of the trigger requirements discussed in section 7.1. The result of this check is shown in figure 11.16(a): the two distributions are compatible.

The effect due to the muon-ID requirements is checked on the invariant mass distribution of the $B \rightarrow hh'$ MC. As discussed earlier, the size of the simulation does not allow a comparison of the mass before and after the application of the muon-ID requirements on both hadrons. For this reason, only one hadron in the $B \rightarrow hh'$ decay is required to produce a fake-muon and the *tight* muon requirement is applied. The remaining hadrons are weighted according to the

fake-muon rates calculated on MC in section 7.4 and shown in table 7.3. Both distributions are shifted according to the B_d^0 mass offset discussed earlier in this section. The resulting mass distributions, shown in figure 11.16(b), are compatible.

In conclusion, the $B \rightarrow hh'$ invariant mass distribution is not significantly affected by the

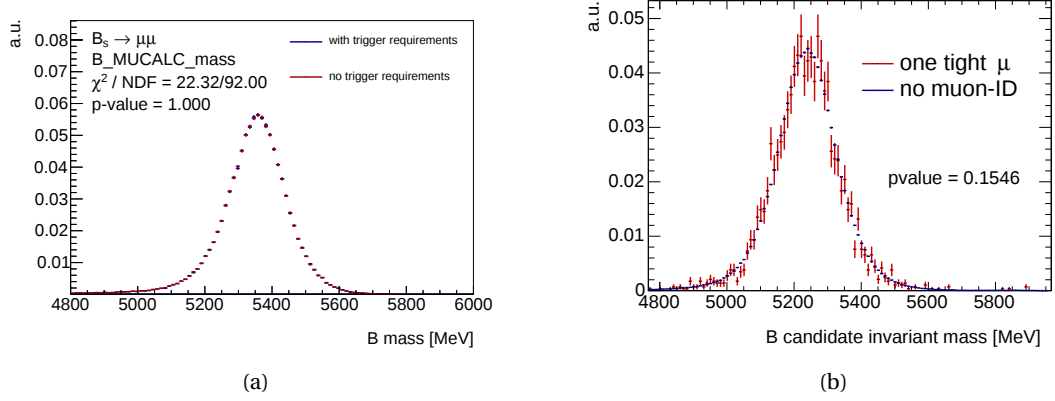


Figure 11.16: Checks on the potential distortion of the invariant mass distribution due to the requirements on the presence of two muons in the final state. The dimuon trigger requirements are checked in figure 11.16(a), where their effect on the $B_s^0 \rightarrow \mu^+ \mu^-$ mass distribution are shown. The distribution before and after the requirement of the triggers discussed in section 7.1 are compared. The potential mass distortion due to the muon-ID requirements discussed in section 7.3 is shown in figure 11.16(b), where the invariant mass distribution of the $B \rightarrow hh'$ sample without any muon-ID requirement is compared to the same distribution with one of two hadrons required to produce a fake muon. The mass shift obtained from the B_d^0 simulation is applied to both.

muon quality requirements. The $B \rightarrow hh'$ mass distribution, corrected as described above, is therefore fit with the same model developed for the signal PDFs. Also in this case the parameters of the functional models are found to be independent from the output of the BDT.

Figure 11.17 shows the resulting distribution of the $B \rightarrow hh'$ MC events after the complete analysis selection is applied, normalised according to the procedure described in 11.2.2.2.

The fit parameters on the $B \rightarrow hh'$ MC template obtained as described above are:

- μ_1 : 5239.5 ± 1.3 MeV
- σ_1 : 85.90 ± 0.2 MeV
- μ_2 : 5216 ± 10 MeV
- σ_2 : 188.7 ± 3.2 MeV
- fraction of gaussian 1: 0.85 ± 0.02

11.2.2.2 Normalisation of the $B \rightarrow hh'$ background

Due to the irreducible nature of the $B \rightarrow hh'$ background and its similarity in shape with the $B_d^0 \rightarrow \mu^+ \mu^-$ signal, the unblinded fit will need the yield for this background component to be constrained.

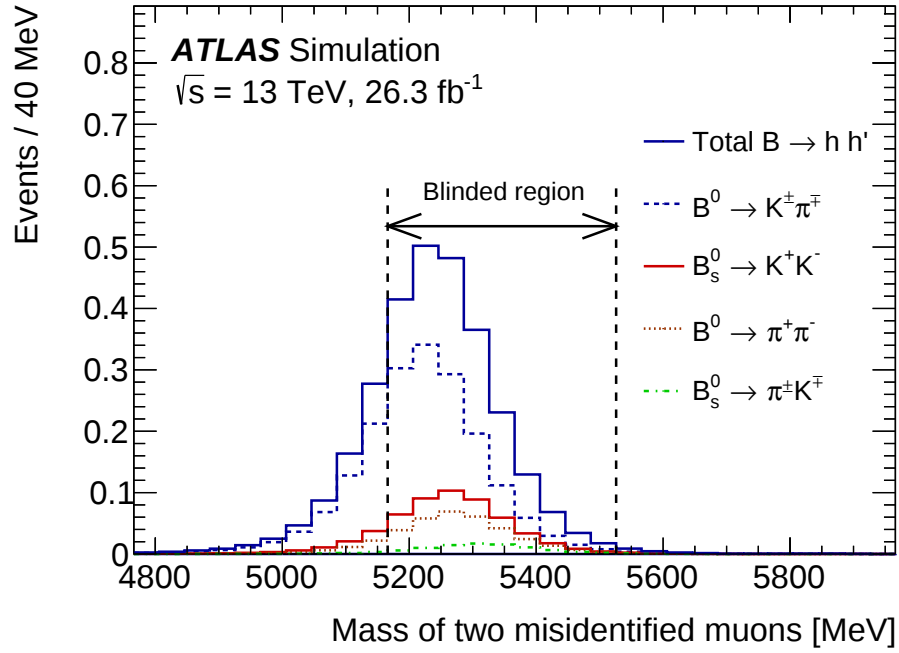


Figure 11.17: Invariant mass distribution of the $B \rightarrow hh'$ peaking background components after the complete signal selection is applied. The $B_s^0 \rightarrow \pi^+\pi^-$ and $B^0 \rightarrow K^+K^-$ contributions are negligible on this scale. The vertical dashed lines indicate the blinded analysis region. Distributions are normalised to the expected yield for the integrated luminosity of 26.3 fb^{-1} .

Two techniques are employed to estimate this yield: the first is based on the muon mis-ID rates discussed in section 7.4. The second technique is data-driven and exploits event counts in the invariant mass signal region obtained for muons of lower purity than was used for the definition of the blinded signal region.

The first estimation of the $B \rightarrow hh'$ yield is performed following the approach used to estimate the SM expected $N_{B_s^0}$ and $N_{B_d^0}$ yields. In analogy to formula 5.1, it can be found that:

$$N_{B \rightarrow hh'} = \frac{f_{s(d)}}{f_u} \times \frac{N_{J/\psi K^+}}{\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)} \times \frac{\mathcal{B}(B \rightarrow hh')}{R_{Ac}}. \quad (11.4)$$

This is combined with the PDG branching fractions [11], obtaining an estimate of each individual $B \rightarrow hh'$ contribution before the application of the muon identification requirements, shown in the second column of table 11.2. Assuming the mis-identification rates of table 7.3 in section 7.4, separated per particle type and charge, the yields listed in the last column of table 11.2 are obtained for each decay mode. The total predicted yield is 2.7 ± 1.3 events, equally distributed in BDT bins 1, 2 and 3.

The second technique for the estimation of the $B \rightarrow hh'$ yield exploits real data events in the invariant mass signal region that do not pass the *tight* muon-ID requirements discussed in section 7.4. The expected yield for these events is calculated using the same approach used for the values in table 11.2. The blinding procedure is not applied on these events, therefore a fit is performed to evaluate their yield and the result is compared with the prediction.

Table 11.2: Breakdown of the expected $B \rightarrow hh'$ yield for the different processes. The numbers in the second column, obtained with equation 11.4 show the estimated yield before the application of the muon identification requirements. Assuming the mis-identification rates of table 7.3, the expected yield of the last column are obtained. The last row shows the total expected $B \rightarrow hh'$ background yield for the analysis.

Process	$B \rightarrow hh'$ expected yield	
	Before muon-ID requirements	After muon-ID requirements
$B_d^0 \rightarrow \pi^+ \pi^-$	$4.0 \cdot 10^5$	0.38
$B_d^0 \rightarrow \pi^\pm K^\mp$	$1.9 \cdot 10^6$	1.68
$B_d^0 \rightarrow K^+ K^-$	$7.6 \cdot 10^3$	0.01
$B_s^0 \rightarrow \pi^+ \pi^-$	$1.3 \cdot 10^4$	0.01
$B_s^0 \rightarrow \pi^\pm K^\mp$	$1.1 \cdot 10^5$	0.09
$B_s^0 \rightarrow K^+ K^-$	$5.4 \cdot 10^5$	0.51
total		2.7

Based on the effect of the muon-ID requirements, muons are defined *not-tight* (NT), if they satisfy the *loose* muon-ID requirement but do not satisfy the *tight* one, and *tight* (T), if they satisfy the *tight* muon-ID. Four possible combinations are obtained taking the two muon categories and applying all possible permutations to the particles in the final state²:

- **NT-NT** (Not Tight - Not Tight), where both muons must satisfy the *loose* criteria but do not satisfy the *tight* criteria;
- **NT-T** (Not Tight - Tight), where the leading muon satisfies the *loose* criteria and does not satisfy the *tight* criteria, and the sub-leading muon satisfies the *tight* criteria;
- **T-NT** (Tight - Not Tight), same as the previous, but swapping the leading and sub-leading muons;
- **T-T** (Tight - Tight), both *tight* muons, this is the actual blinded signal region.

The mis-identification fractions of table 7.3 allow the estimation of the expected number of events for the four categories, using the same approach employed to calculate the values in table 11.2. Table 11.3 shows the expected $B \rightarrow hh'$ yields for the four combinations, as well as the expected $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields.

An unbinned extended maximum likelihood fit is then performed on the control region characterised by NT-NT events in real data. The PDFs developed for the background and signal components in the previous sections are exploited in the same fitting procedure discussed in section 11.3. The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ contributions are fixed to their SM prediction and the background shapes are fixed to what was determined in the sideband fits in section 11.1. The normalisations of all the backgrounds are left free in the fit. Figure 11.18

² Where μ_1 - μ_2 are the leading and sub-leading muons

Table 11.3: Estimated yields of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ and $B \rightarrow hh'$ events for the possible combinations of the following requirements: *tight* muon and *loose* muon inverting the *tight* muon requirement. The two requirements are shown in the table as “*tight*” and “*loose* and not *tight*”. All the expected yields are obtained inverting formula 5.1, using the appropriate branching fraction from the PDG [11], and exploiting the mis-identification rates and muon efficiencies estimated on simulations in section 7.4.

μ_1 WP \ μ_2 WP	$B_s^0 \rightarrow \mu^+ \mu^-$		$B_d^0 \rightarrow \mu^+ \mu^-$		$B \rightarrow hh'$	
	<i>tight</i>	<i>loose</i> and not <i>tight</i>	<i>tight</i>	<i>loose</i> and not <i>tight</i>	<i>tight</i>	<i>loose</i> and not <i>tight</i>
<i>tight</i>	90.7	10.2	10	1.1	2.7	4.2
<i>loose</i> and not <i>tight</i>	10.2	1.15	1.1	0.11	4.2	6.4

shows the invariant mass of distributions of the four BDT bins, together with the projection of the likelihood.

This procedure yields a total number of $B \rightarrow hh'$ events $N_{hh}^{NT-NT} = 6.8 \pm 3.7$, in good agreement with the 6.4 expected events from table 11.3. This number is translated into a prediction of the $B \rightarrow hh'$ yield in the *tight-tight* signal region, by taking the ratio between the expected yields in simulations:

$$N_{hh}^{T-T} = N_{hh}^{NT-NT} \times \frac{MC_{hh}^{T-T}}{MC_{hh}^{NT-NT}} = (2.9 \pm 2.0) \text{ events.} \quad (11.5)$$

As a loose cross-check, a fit is also performed in the data in the control region obtained summing the T-NT and NT-T datasets. The same fit procedure as the one used in the NT-NT region is employed, with the inclusion of a constraint on the $B_d^0 \rightarrow \mu^+ \mu^-$ SM expectation of 2.2 ± 0.4 events. The result of this check yields $N_{hh}^{N-NT, NT-T} = 5.8 \pm 7$ events, compatible with the expected number of events of 8.4.

Additionally, the ratio $\frac{N_{hh}^{NT-NT}}{N_{hh}^{T-NT}} = 1.2 \pm 1.6$ is used to measure directly the single-leg hadron species-averaged fake rate for hadrons from $B \rightarrow hh'$ decays. Figure 11.19 shows the per-species value obtained from table 7.3. This ratio is remarkably constant across hadron species, and consistent with the data-based estimation.

In conclusion, the expected $B \rightarrow hh'$ yield in the three higher BDT bins, evaluated with simulations (2.7 ± 1.3 events) and by fitting the “NT-NT” control region (2.9 ± 2.0 events) are in agreement. The more conservative value of 2.9 ± 2.0 events, obtained in equation 11.5, is used to constrain the total yield of the peaking background in the unblinded fit.³

11.3 Signal yield extraction

The combination of the non-resonant functional forms from section 11.1 and the signal and peaking background functional forms from section 11.2 allows the obtaining of a full model

³ The value 2.9 ± 2.0 is used to constrain the total $B \rightarrow hh'$ yield in bins 1, 2 and 3. Since this background presents the same BDT response as the signal, the relative yield in the three bins are set to be the same. Regarding bin 0, its $B \rightarrow hh'$ content is set to the average content of the other three.

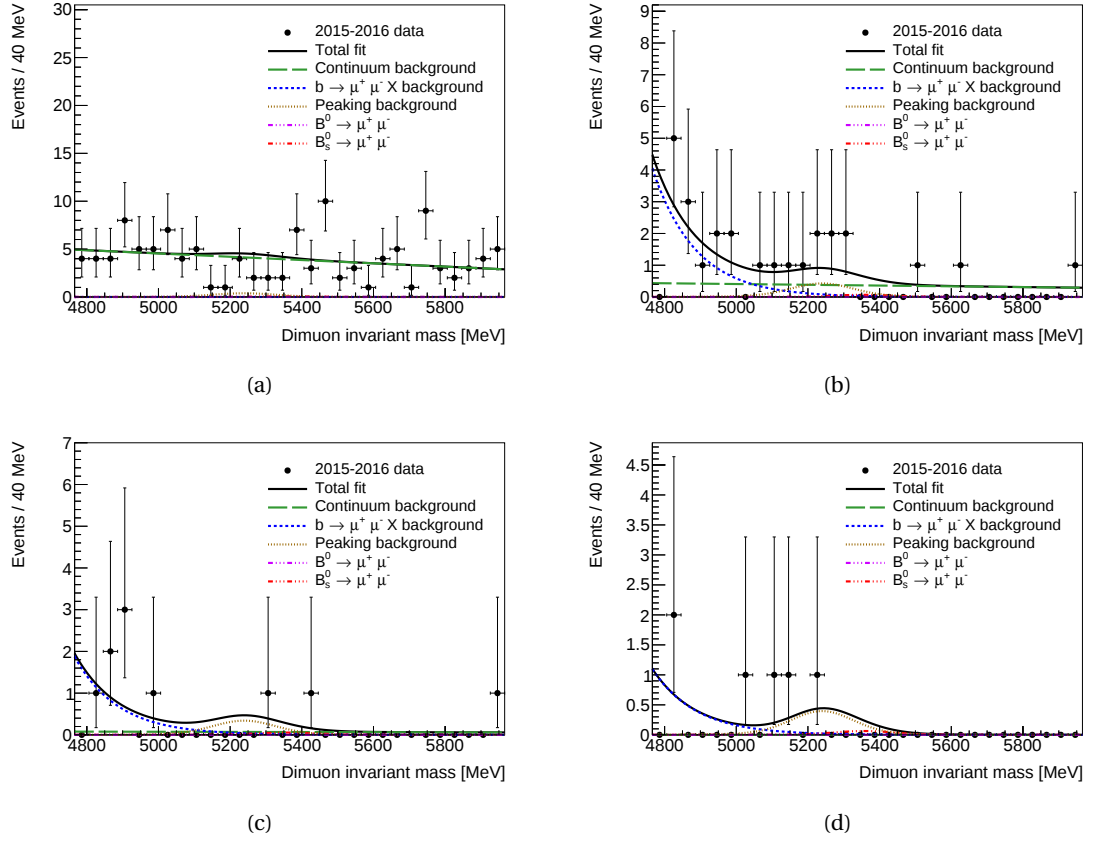


Figure 11.18: Di-muon invariant mass distributions of the control region characterised by NT-NT events in real data, in the four intervals of BDT output. Superimposed is the result of the maximum-likelihood fit. The total fit is shown as a continuous line, with the dashed lines corresponding to the observed signal component, the $b \rightarrow \mu^+ \mu^- X$ background, and the continuum background. The $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ signal components are represented with dot-dashed red and violet lines respectively. Considering the low expected yield for the $B_d^0 \rightarrow \mu^+ \mu^-$ process, the corresponding component is basically not visible. The curve representing the peaking $B \rightarrow hh'$ background is represented with a fine-dashed line.

to be used in the unblinded fit on data. This section is dedicated to the studies performed on the full fit model. First, the description of the model is provided (sub-section 11.3.1), then the convergence properties are tested in sub-section 11.3.2 by means of a toy-MC study. Exploiting a similar toy-MC study, the systematic uncertainties arising from the fitting procedure are evaluated in sub-section 11.3.3; the introduction of the systematic uncertainties related to the other terms in formula 5.1 is discussed later in section 11.4.

The studies performed in this section rely heavily on the usage of toy-MCs. The initial values used for the generation of the mock datasets for the non-resonant backgrounds are the ones found in the simultaneous fit on data sidebands in section 11.1, while for the signal and peaking background the values used are obtained in section 11.2. Regarding the signal normalisation, the studies are performed on a grid of possible values, usually ranging between 0 and about twice the SM expectation.

The number of events expected given $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ is calculated by inverting formula 5.1, obtaining $N_{B_s^0} = 91$ and $N_{B_d^0} = 10$ for the nominal SM branching fractions [54].

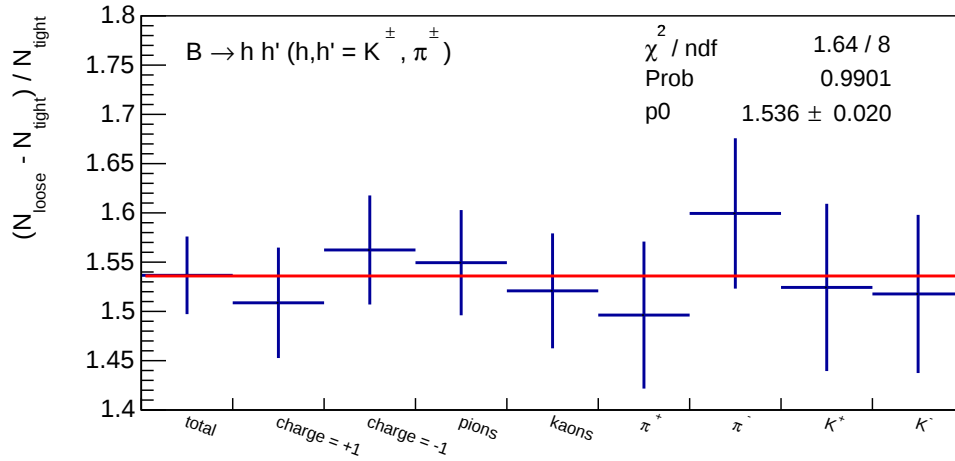


Figure 11.19: Comparison of the relative muon-ID fraction for different particle species as determined from the $B \rightarrow hh'$ MC.

11.3.1 Fit configuration

This section provides a summary of the PDFs used in the signal yield extraction.

1. **Signal:** the mass dependence for $B_s^0 \rightarrow \mu^+ \mu^-$ is described with the sum of two gaussians. The means and the widths of the gaussians and their relative fraction are taken from MC with their values fixed in the fit. The same configuration, but with different parameters is used for the $B_d^0 \rightarrow \mu^+ \mu^-$ signal.

The three higher BDT bins are designed to contain each 1/3 of the total number of events, with the uncertainty discussed in section 10.2.3. Such systematic uncertainties are implemented as gaussian constraints applied to the relative fractions of signal in the BDT bins.

The lower BDT bin (bin 0), also designed to be as efficient for the signal as the three bins with higher S/B, was verified not to affect the signal sensitivity of the analysis, but allows a better background estimation. Its signal yield is fixed, for consistency, to the average content of the other three BDT bins and it is not considered in the branching fractions extraction.

2. **Linear background:** the mass dependence is a first order Chebychev polynomial. The normalisation is independent in each BDT bin. The polynomials' slope is constrained to depend linearly on the average content of each BDT bin. Systematic uncertainties on the final fit result are derived for this assumption.
3. **Exponential background:** exponential dependence on the mass. The normalisation is extracted independently in each BDT bin, while the shape is assumed to be independent of the BDT bin value. This assumption is challenged as part of the systematic studies.

4. **Peaking background** ($B \rightarrow hh'$): the mass dependence is described with the same double gaussian model used for the signals, but independent parameters determined from simulation. A gaussian constraint is applied to the normalisation, as discussed in section 11.2.2.2.

The following free parameters are to be determined by the fit:

- the total number of B_s^0 events, and the total number of B_d^0 events in bins 1, 2 and 3;
- the number of events associated with the linear background in each BDT bin (four parameters);
- the number of events associated with the exponential background in each BDT bin (four parameters);
- the parameters describing the linear constraints for the polynomial slopes (two parameters);
- the parameter describing the shape of the low-mass background;

The following parameters subject to gaussian constraints are also determined by the fit:

- the fractions of signal events contained in each bin of the BDT bin;
- the total number of peaking-background events (see constraint derived in section 11.2.2.2);

No boundaries are applied to the normalisations of the various components⁴. This means that the yields extracted by the fit can fluctuate and assume non-physical values. Fit results with a negative number of events, especially for the $B_d^0 \rightarrow \mu^+ \mu^-$ component which is affected by large uncertainties, are therefore possible. This possibility does not constitute a problem for the validity of the analysis. In fact, the estimated values and uncertainties obtained with maximum likelihood estimators do not relate to the true physical parameter the experiment is trying to measure [182]. The maximum likelihood fit result is, however, employed in the extraction of confidence intervals, which allow for a quantitative statement about the fraction of times that such an interval would contain the true value of the parameter in a large number of repeated experiments [182]. The approach employed in the analysis, the Neyman construction [178], is an exact frequentist method for the confidence intervals and regions extraction.

11.3.2 Fit convergence properties

The convergence properties of the fit are tested by means of a toy-MC based study. Two main aspects are investigated, the fraction of toy-MC in which the fitting procedure converges and the possible presence of a bias in the estimated signal yields.

⁴ Except for the $B \rightarrow hh'$ yield which is subjected to a gaussian constraint.

Toy-simulations are generated according to the values of the slopes and normalisations obtained in the previous chapters and then fitted with the baseline configuration of the simultaneous fit. Since both the fit convergence and the signal yield bias can show a dependence on the estimated number of events, 10000 toy-simulations are run in each point of a 20×20 grid in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ space, with values ranging between 0 and about twice the SM expectations.

Figure 11.20 shows the invariant mass distributions of the four BDT bins for one of the toy-MCs generated at the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ SM prediction. Each plot reports also the projection of the fitted model with its various components.

The first property of the fit investigated is the fraction of toys for which the fitting procedure

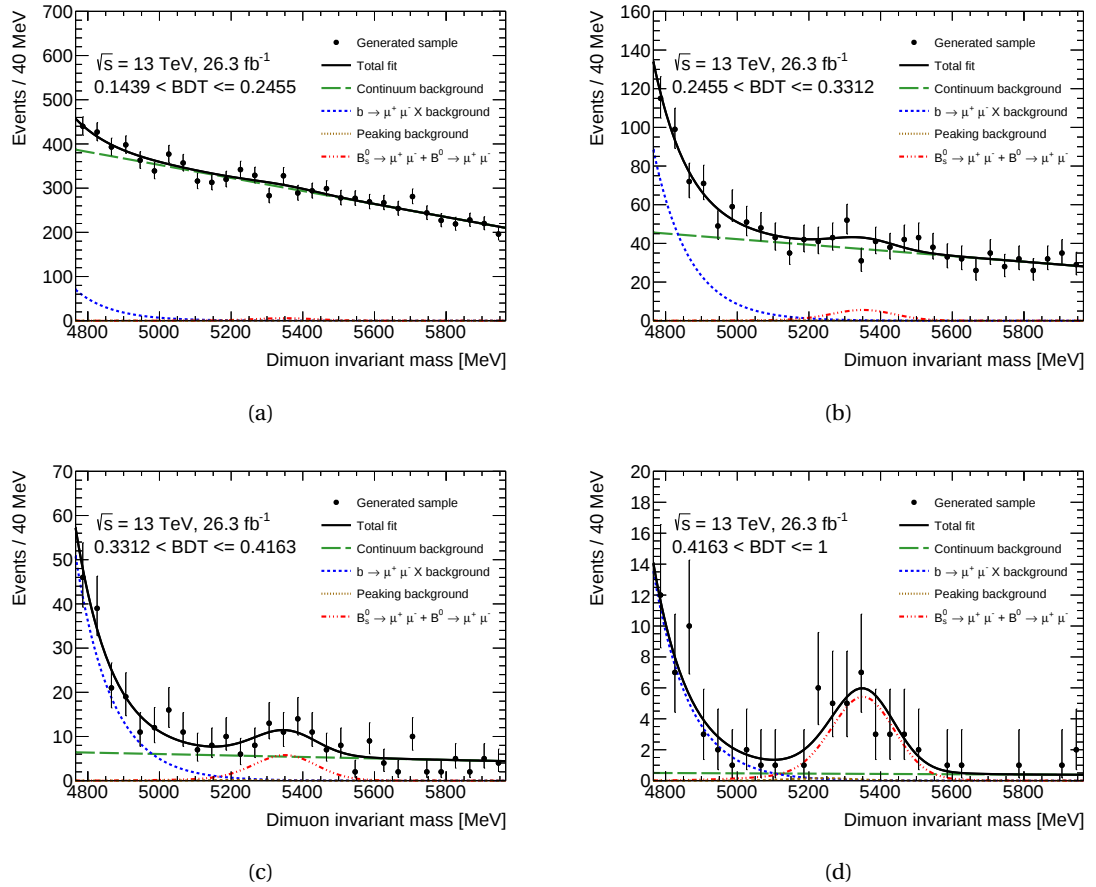


Figure 11.20: Di-muon invariant mass distributions in the four intervals of BDT output of one of the toy-MCs generated for the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ SM prediction. Superimposed is the result of the maximum-likelihood fit. The total fit is shown as a continuous black line, with the dashed lines corresponding to the observed signal component, the low-mass background, and the continuum background. The signal components are grouped in one single curve, including both the $B_s^0 \rightarrow \mu^+ \mu^-$ and the $B_d^0 \rightarrow \mu^+ \mu^-$ component. The curve representing the peaking $B \rightarrow hh'$ background lies very close to the horizontal axis in all BDT bins.

converges.

The fraction of toy-MCs that converge, providing a maximum of the likelihood function and returning a covariance matrix⁵ for the fit parameters is shown in figure 11.21, as a function of the

⁵ The likelihood maximisation is performed numerically by the MINUIT [189] program. While MINUIT provides the maximum of the likelihood, its routine HESSE provides the covariance matrix of the fit parameters. The estim-

branching fraction of the signal processes used in the generation. The overall fit convergence is good, with more than 99.5% of the toy-MCs fully converging. In case of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$, the fraction of converging toys decreases; a dependence on the B_d^0 branching fractions is visible in this case, with the fraction of converging toy-MCs getting down to 96% for $(N_{B_s^0}, N_{B_d^0}) = (0, 0)$. This behaviour shows a clear dependence of the fit on the number of events fitted. Due to the degeneracy of the B_s^0 and B_d^0 peaks, which are almost superimposed (figure 11.14), once the available signal statistics drops, the fit has too many degrees of freedom to manage, resulting in an instability of its convergence.

Another important quantity regarding the fit convergence is the fraction of toy-MCs for which

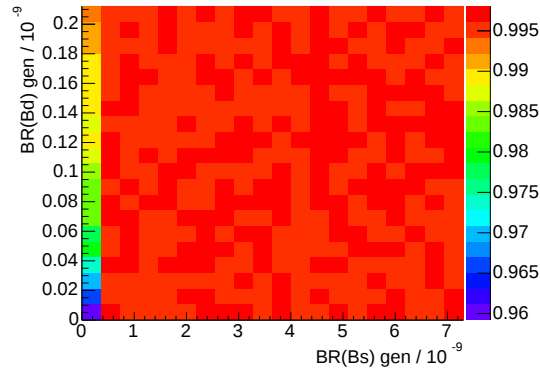


Figure 11.21: Fraction of toy-MC in which the maximum of the likelihood is identified and the covariance matrix of the fit parameters is returned by the fit, as a function of the signals branching fractions used in the generation of the toy-MCs.

the likelihood maximum is found, without requiring a covariance matrix. This quantity is particularly interesting, as the analysis confidence interval extraction is performed using the Neyman belt approach [178]. The construction of the belt, in fact, does not require the extraction of the uncertainty on toy-MCs or on the fitted value of the analysis: the entire construction is based on the central values obtained in the fits. In this case the average fraction of toy-MCs that converge is higher than 99.99%. An effect similar to what is shown in figure 11.21 is also found, with the fraction of converging toys dropping to about 99% at $(N_{B_s^0}, N_{B_d^0}) = (0, 0)$.

These two tests on the fit convergence show a possible issue at extremely low signal statistics. In this case, the best approach would be to fit both signals with a single PDF, avoiding the degeneracy of the two signal models. Considering the recent experimental results in the search for $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays, the probability of such low signal statistics is extremely low. Using toy-MCs generated for the expected SM signal yields, the probability of an underfluctuation for which the fitted number of $B_s^0 \rightarrow \mu^+ \mu^-$ events is zero, is estimated to be at the level of 5σ . Additionally, in this case only a small fraction of fits would show convergence issues. For this reason, an analysis result with a fitted $B_s^0 \rightarrow \mu^+ \mu^-$ yield of 0 events would point more towards an issue in the analysis, than to a possible indication of BSM physics.

In conclusion, the convergence of the fit is considered satisfactory, with the large majority of the toy-MC in the region of interest of the analysis converging.

ation of the covariance matrix is performed using a gaussian approximation of the likelihood near its maximum.

The presence of a bias in the fitting procedure is checked using the same set of toy-MC employed to study the fraction of converging fits. The average value of the pull with respect to the generated value in each point of the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ grid is calculated as $(N_{\text{gen}} - N_{\text{fit}})/\sigma_{N_{\text{fit}}}$. The uncertainty employed in this calculation is evaluated with a gaussian approximation around the likelihood function maximum. For this reason, only the toy-simulations for which the likelihood maximum is found and the covariance matrix of the fitted parameter is returned are used. Figure 11.22 shows the resulting distributions for the pulls. At

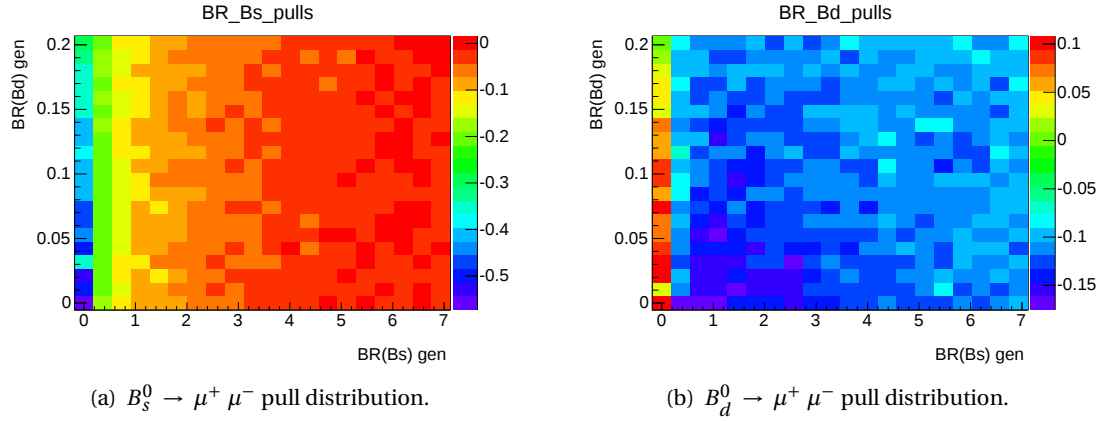


Figure 11.22: Toy-MC assessment of the bias on the position of the likelihood maximum as a function of the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ branching fractions. The bias is estimated as the pull on the fitted number of events, with respect to the generated number of events: $(N_{\text{gen}} - N_{\text{fit}})/\sigma_{N_{\text{fit}}}$. The uncertainty on the fitted number of events is calculated using a gaussian approximation around the likelihood function maximum.

the SM value, the bias on the $B_s^0 \rightarrow \mu^+ \mu^-$ ($B_d^0 \rightarrow \mu^+ \mu^-$) branching fraction is at the level of 5% (10%) of the uncertainty on the branching fraction itself, with the bias on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ becoming higher than 10% below $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 1 \times 10^{-9}$.

In the regions of low generated branching fractions the bias of the fit becomes more pronounced, and as large as 50% of the uncertainty for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$. Regarding $B_d^0 \rightarrow \mu^+ \mu^-$, the bias behaves similarly as $B_s^0 \rightarrow \mu^+ \mu^-$, expect that for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$ it becomes positive. These effects can be caused by the lack of asymptoticity of the likelihood fit itself or by the instability of the fit caused by the degeneracy of the B_s^0 and B_d^0 peaks. In particular, the large effects that appear for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$ are most likely due to the degeneracy of the signals, as $B_s^0 \rightarrow \mu^+ \mu^-$ is the main peak present in the fit and when it disappears the fit becomes unstable.

As for the fraction of converging toy-MCs, the behaviour of the fit is satisfactory, with undesirable effects at low signal yields. The same argument discussed earlier in this section can be used also in this case: the possibility of a huge underfluctuation of the signal events is a remote possibility, at the level of a 4-5 σ fluctuation. This eventuality would therefore mostly point towards an issue in the analysis.

In conclusion, the convergence properties of the fit are tested, in particular the fraction of toy-

MC that converge and the presence of a possible bias in the maximum likelihood estimation. Both tests show satisfactory properties for the expected B_s^0 and B_d^0 statistics. In case the signal statistics approaches $(N_{B_s^0}, N_{B_d^0}) = (0, 0)$ the fitting procedure shows instability, with large biases and a larger fraction of non-converging toy-MCs, due to the degeneracy of the B_s^0 and B_d^0 mass peaks. The possibility of such low signal events yields is tiny (at the level of a 4-5 σ fluctuation), and would be most likely explained with an issue in the analysis procedure. Therefore, the fitting procedure is considered validated.

11.3.3 Fit model systematics

The systematic uncertainties associated with the fitting procedure are extracted with a similar toy-MC based study.

Basically, the model used in the generation of the toy-simulations is modified according to systematic variations and the baseline fit model (described in section 11.3.1) is used to fit the resulting datasets. The corresponding variations in the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields, estimated as described in this section, are analysed and taken as systematic uncertainties associated with the fitting procedure and introduced in the fitting procedure as gaussian constraints on the number of fitted signal events.

Three main classes of systematic uncertainties are considered: systematics arising from the resonant models, systematics arising from the choice of the non-resonant models and constraints, and systematics arising from the non-resonant models neglected in the baseline fit.

The systematic uncertainties associated with the resonant models are evaluated by applying mass scale and resolution variations, derived from the known features of the detector and from previous analyses. The systematic uncertainties on the choice of the non-resonant models are derived by using different functional models in the generation of the toy-simulations. The new models are derived following the same approach described in section 11.1.2. The systematic uncertainties associated with the constraints used in the baseline fit are evaluated by generating the toy-MCs varying the quantities subjected to a constraint according to the values found in data and MC. Finally, systematics arising from the neglected models in the baseline fit are evaluated by adding such models in the generation of the toys. The shape and normalisations of the functional models employed are obtained in section 11.1.2.

The following list summarises the various systematic variations considered.

- **Mass scale ± 5 MeV:** a momentum scale uncertainty of ± 0.05 % on the di-muon mass has been estimated at the mass of the Υ meson [205] (± 2.7 MeV at the B_s^0 mass). When comparing the J/ψ mass peak position in MC and data (section 10.1.2), a mass scale variation of about 2 MeV is found. The sensitivity of the signal fit to the mass scale uncertainty is conservatively tested by shifting the means of the peaking models by 5 MeV.
- **Mass resolution $\pm 5\%$:** an average systematic uncertainty of 5 % on the dimuon mass resolution has been estimated employing the $J/\psi \rightarrow \mu^+ \mu^-$ and $Z \rightarrow \mu^+ \mu^-$ [161]; the mass

resolution uncertainty is thus probed varying the width of the resonant PDFs increased or reduced by 5%.

- **Combinatorial bkg PDF model and low mass bkg PDF model:** the assumptions on the functional models for the exponential and polynomial components are (separately) tested using respectively a gaussian-tail and exponential model.
- **Exponential slope (data) and exponential slope (MC):** the assumption of the constant dependence of the exponential background slope on the BDT value is challenged by generating independent slopes for each BDT bin. A gaussian PDF is used to generate the different slopes in each bin, based on the results of the fits on data and MC in section 11.1.3. The mean and width of the gaussian are set to the fit results for the exponential slope and its uncertainty; the values used in the generation are shown in figure 11.9 for the MC and 11.10 for the data.
- **Polynomial slope (data), polynomial slope (MC) and flat polynomial slope:** the assumption of the linear dependence of the linear background slope on the average BDT value in each BDT bin is challenged by generating independent slopes for each BDT bin, according to the values measured in data and on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ MC in section 11.1.3. The same procedure employed for the exponential slope variations is used. In addition the slope is also generated assuming a flat dependence of the background slope with respect to the BDT value.
- **Inclusion of semileptonic background:** the semileptonic background is neglected in the baseline fit. Its inclusion in the toy generation is therefore used to assess the systematic uncertainty on the presence of this contribution using the model and scales defined in section 11.1.2.4.
- **Inclusion of B_c background:** B_c decays are neglected in the baseline fit. Their inclusion in the toy generation is tested with the model described in section 11.1.2.3.

In order to test the possible dependence of these systematic uncertainties on the B_s^0 and B_d^0 yields, a 5×5 points grid of values of $B_d^0 \rightarrow \mu^+ \mu^-$ and $B_s^0 \rightarrow \mu^+ \mu^-$ branching ratios, ranging between 0 and about twice the SM expectation, is explored. 10k toy-MC samples are generated for each variation listed above in each grid point. Additionally, toy-MCs are also generated using the baseline fit in both the generation and fit (baseline toys), for a total of 140k toy-MC samples generated for each grid point.

The generated toy-MCs show convergence fractions similar to the ones discussed in section 11.3.2. Only the converging fits are employed in the systematic uncertainties determination.

The variation between the average signal yields obtained with the baseline toys and the modified toys is taken as $\widehat{N}_{B(s)}^{baseline} - \widehat{N}_{B(s)}^{modified}$. Positive and negative variations are then summed in quadrature separately; the larger between the combination of negative and positive variations is taken as systematic uncertainty for the $(N_{B_s^0}, N_{B_d^0})$ grid point considered.

Since the invariant mass peaks of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signals overlap, the yields extracted by the

fit are correlated; this introduces a correlation also in the systematic uncertainties discussed in this section. For this reason, the correlation between the systematic uncertainties associated with the fitting procedure is also evaluated.

The resulting systematic uncertainties on the B_s^0 and B_d^0 yields and their correlation are shown

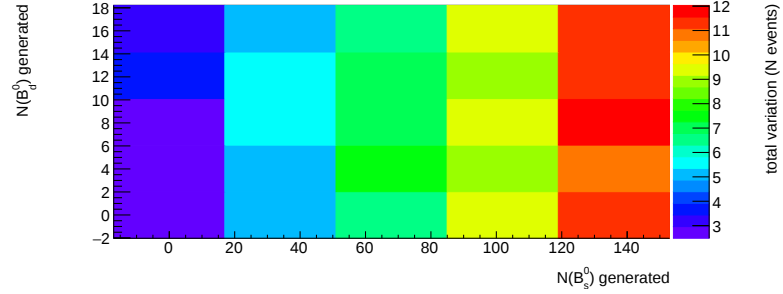


Figure 11.23: Systematic uncertainty associated with the $B_s^0 \rightarrow \mu^+ \mu^-$ fitted yield as a function of the generated B_s^0 and B_d^0 yields. The systematic variations and their combinations are obtained as explained in the text.

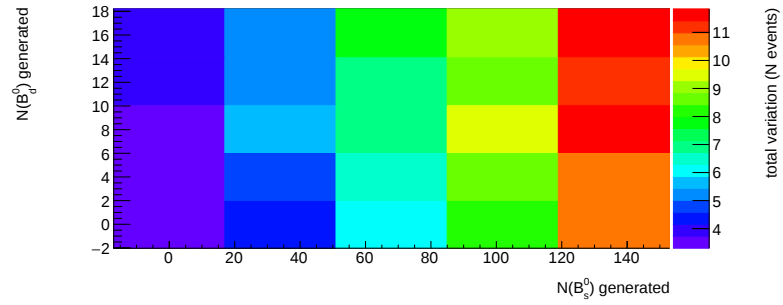


Figure 11.24: Systematic uncertainty associated with the $B_s^0 \rightarrow \mu^+ \mu^-$ fitted yield as a function of the generated B_s^0 and B_d^0 yields. The systematic variations and their combinations are obtained as explained in the text.

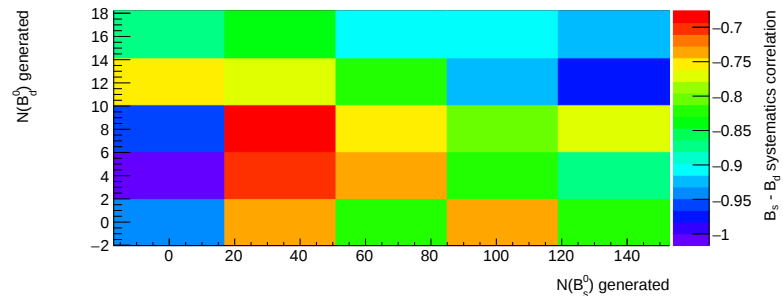


Figure 11.25: Correlation of the systematic uncertainties associated with the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ fitted yields as a function of the generated B_s^0 and B_d^0 yields.

respectively in figures 11.23, 11.24 and 11.25.

The relative ranking of the various systematic sources implemented is shown in table 11.4 for the SM expectation on the signal yields. Systematic variations associated with the B_d^0 yield are larger compared to the B_s^0 ones. This is due to the fact that the B_d^0 yield is much smaller

Table 11.4: Fit modelling systematic uncertainties expressed as fraction of the expected signal yield, evaluated for the SM expectation on the signal yields.

Systematic source	Impact on B_s yield	Impact on B_d yield
Mass scale -5 MeV	5.0%	64.0%
Mass scale +5 MeV	4.9%	64.4%
Mass resolution -5%	3.5%	4.6%
Mass resolution +5%	2.9%	2.9%
Exponential slope (data)	1.2%	0.5%
Exponential slope (MC)	0.2%	7.6%
Flat polynomial slope	0.1%	1.3%
Polynomial slope (data)	0.1%	0.9%
Polynomial slope (MC)	0.9%	1.6%
Combinatorial bkg PDF model	0.5%	2.2%
Low mass bkg PDF model	5.8%	53.6%
Inclusion of B_c background	0.2%	1.3%
Inclusion of semil. background	0.4%	12.1%

compared to the B_s^0 counterpart and it is closer to the lower mass sideband, implying a higher contamination of background events from PRD events.

The dependence of the systematic variations as a function of $(N_{B_s^0}, N_{B_d^0})$ is fit with a 2D linear function. A similar parameterisation is used to account for the correlation coefficient between the systematic uncertainties associated with the $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ event counts, resulting in the determination of the full 2D gaussian systematic uncertainty as a function of $(N_{B_s^0}, N_{B_d^0})$. The dependence of the combined systematics on the number of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$, as well as their correlation, are fit with a linear model $aN_{B_s^0} + bN_{B_d^0} + c$. The resulting coefficients are summarised in table 11.5, together with the actual dependence of the gaussian parameters implemented in the signal fit.

The resulting 2D gaussian systematic uncertainty as a function of $(N_{B_s^0}, N_{B_d^0})$ is implemented in the fit model described in section 11.3.1 as a gaussian smearing parameter on the number of fitted B_s^0 and B_d^0 events.

11.4 Branching fraction extraction approach

The signal branching fraction extraction is based on the same fitting procedure discussed in the previous sections. The number of signal events is replaced with the branching fractions together with the normalisation terms from formula 5.1.

The approach followed to replace the signal yields in the fitting procedure is discussed in this section, together with a summary of the terms needed in formula 5.1. An estimation of the expected 2D likelihood contours for the SM $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ hypothesis [54] is also presented.

The number of signal events $N_{B_s^0}$ and $N_{B_d^0}$, after the application of the systematic uncertainties

Table 11.5: The first three lines show the parameters obtained with a linear fit of the combined systematic uncertainties as a function of $N_{B_s^0}$ and $N_{B_d^0}$. The last three lines show the dependence of the parameters used to implement the 2D gaussian constraint in the fitting procedure.

	$N_{B_s^0}$ Coefficient	$N_{B_d^0}$ Coefficient	Constant
$\sigma_{N_{B_s^0}}$	0.049 ± 0.001	0.008 ± 0.011	2.99 ± 0.16
$\sigma_{N_{B_d^0}}$	0.046 ± 0.001	0.046 ± 0.011	2.92 ± 0.16
ρ	-0.0001 ± 0.0003		-0.827 ± 0.034
$\sigma_{N_{B_s^0}} = 3 + 0.05N_{B_s^0}$			
$\sigma_{N_{B_d^0}} = 2.9 + 0.05N_{B_s^0} + 0.05N_{B_d^0}$			
$\rho = -0.83$			

arising from the fitting procedure (section 11.3.3), are replaced with the corresponding branching fractions divided by the normalisation terms obtained rearranging formula 5.1. Effectively, the signal yields are replaced as follows:

$$N_{B_d^0(B_s^0)} \rightarrow \left(\frac{\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) \times N_{J/\psi K^+}}{\frac{f_u}{f_{s(d)}} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \times R_{Ac}} \right). \quad (11.6)$$

Additional gaussian smearing factors are added to the likelihood function to take into account the uncertainties of the terms in formula 11.6. The only terms not associated with gaussian constraints are the signal branching fractions, which are left free to be determined by the fitting procedure. This approach allows the evaluation of the two signal branching fractions taking into account the correlation between the two signal yields arising from the fitting procedure as well as a clean way to propagate all the uncertainties.

The terms needed in formula 11.6 for the evaluation of the signal branching fractions have been evaluated in chapter 10 or taken from the most recent experimental measurements [11, 28].

The normalisation terms included in equation 11.6 are the following:

- $\frac{f_s}{f_d} = 0.256 \pm 0.013$, which assumes $\frac{f_u}{f_d} = 1$, taken from the latest HFLAV average [28];
- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.010 \pm 0.029) \times 10^{-3}$ and $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033)\%$, obtained from the world average of the two branching fractions [11];
- $N_{J/\psi K^\pm} = 334351 \times (1 \pm 0.3\% \pm 4.8\%)$, obtained in section 10.1;
- $R_{Ac} = 0.1144 \pm 0.8\% \pm 4.0\%$, used for both B_s^0 and B_d^0 branching fraction extraction. Additionally, the lifetime correction discussed in section 10.2.1 applied to R_{Ac} term for the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ evaluation.

The modified fitting procedure for the extraction of the signal branching fractions is employed to evaluate the 2D profiled likelihood contours in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane

for the toy-MC shown in figure 11.20. Only likelihood contours are extracted in order to limit the CPU consumption, as the Neyman approach [178] requires a large number of toy-simulations. The toy-MC is generated for the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ SM prediction, and the contours are drawn for values of $-2\Delta\ln(\mathcal{L})$ equal to 2.3, 6.2 and 11.8, as discussed in section 6.3.2. Figure 11.26 shows the resulting contours: the blue lines are obtained considering both the statistical and systematic uncertainties, while the red lines are obtained considering only the statistical uncertainty. As expected, the statistical uncertainty largely dominates, in fact the two sets of contours are extremely similar.

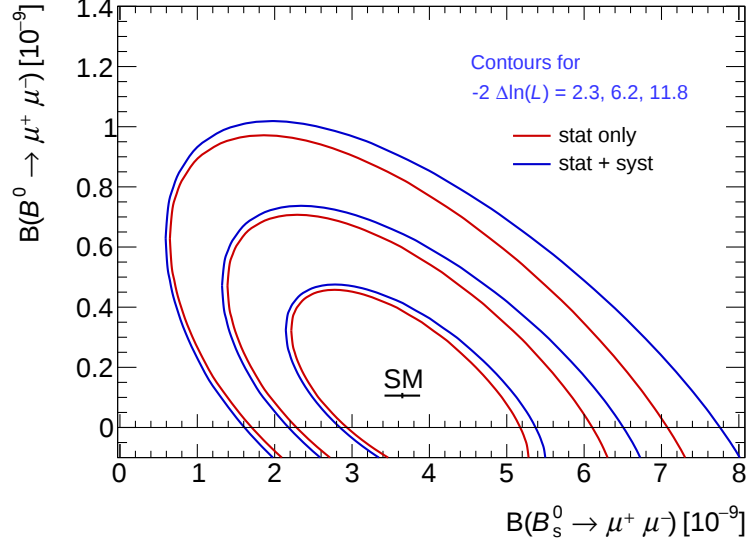


Figure 11.26: Likelihood contours obtained with the fit on simultaneous fit to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ on the toy-MC shown in figure 11.20, generated for the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ SM prediction. The contours are drawn for values of $-2\Delta\ln(\mathcal{L})$ equal to 2.3, 6.2 and 11.8. Two sets of contours are shown: one including statistical and systematic uncertainties (blue lines) and the other including only statistical uncertainties (red lines). Also the SM prediction with uncertainties is indicated.

11.5 Conclusions

The fit for the extraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal yields has been designed and tested.

The choice of the functional models and their constraints are based on simulations and completely independent from data. The final fit, on the other hand, will be performed solely on data, without any input from simulations.

The fitting procedure has been tested with a large number of toy-MC simulations, in particular studying the convergence of the fit and the possible bias in the signal yields. Systematic uncertainties arising from the assumptions and choices made during the development of the fitting procedure have also been evaluated and considered in the fit.

The procedure for the extraction of the signal branching fractions has also been designed: it is based on the fit for the extraction of the signal yields, where the number of signal events is substituted with the signal branching fractions and the normalisation terms from equation 5.1. The implementation of the fit on data for the extraction of the signal yields was the last ingredi-

ent needed for the definition and tuning of the analysis. Since all the procedures are now set, the blinded mass signal region can be opened. The following chapter therefore discusses the fit performed on real data and the final result of the analysis.

BRANCHING FRACTION EXTRACTION

12

The preceding chapters have provided all the ingredients needed to evaluate the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions using formula 5.1: the hadronisation probabilities of a b-quark into B^+ and $B_s^0(B_d^0)$ and the reference channel branching fraction, which are taken from the latest available experimental measurements [11, 28], the reference channel yield (section 10.1), the $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency ratio (section 10.2) and a procedure for the extraction of the branching fractions employing the fit on data to extract the signal yields (chapter 11). The final steps of the analysis are discussed in this chapter, with the extraction of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions using the 2015/16 dataset discussed in section 12.1. This result is then combined with the result of the Run 1 analysis [73] in section 12.2, in order to improve the statistical precision of the result.

The results of the analysis performed on the 2015/16 dataset and its combination with the ATLAS Run 1 analysis are finally summarised in section 12.3.

Additionally, the ongoing effort to combine the latest $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ results from the LHC experiments is presented in section 12.4.

12.1 2015/16 dataset branching fractions extraction

The first step towards the final result of the analysis is the unblinding procedure, in order to access the mass signal region, not accessible during the analysis definition and tuning phases. Once the signal region is accessible, the number of signal events can be extracted (sub-section 12.1.1), applying the fitting procedure described in chapter 11, followed by the evaluation of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. Given the low statistics regime of the analysis, a reliable strategy for the extraction of the uncertainty on the signal branching fractions is needed. The three approaches for the evaluation of the uncertainty described in chapter 6 are applied and discussed in sub-section 12.1.3, converging to the Neyman belt construction method, which is used to provide the final result of the analysis. This result is finally summarised in sub-section 12.1.4.

12.1.1 Fit on unblinded data

As introduced at the beginning of this chapter, the analysis is fully tuned and the signal region can be unblinded. Table 12.1 summarises the event counts in the data mass sidebands and signal region for the four BDT bins after the unblinding procedure.

The fitting procedure is then applied to the unblinded data. From the sideband-only fit on

Table 12.1: Event counts in the four BDT bins in data after the unblinding procedure.

	bin 0	bin 1	bin 2	bin 3
left sideband (4766 MeV < m(B) < 5166 MeV)	3800	639	191	42
signal region (5166 MeV < m(B) < 5526 MeV)	2683	354	63	35
left sideband (5526 MeV < m(B) < 5966 MeV)	2730	337	58	7

data (section 11.1.4) 2708, 331, 56 and 8.2 events are interpolated in the blinded signal region for BDT bins 0, 1, 2 and 3 respectively. The number of background events observed in the data blinded region after the application of the fitting procedure is respectively 2685 ± 37 , 330 ± 14 , 51 ± 6 and 7.9 ± 2.6 for BDT bins 0, 1, 2 and 3, in good agreement with the interpolations. As already mentioned, the number of expected signal events according to the SM prediction [54] is $N_{B_s^0} = 91$ and $N_{B_d^0} = 10$. The number of expected signal events and the number of background events observed in the signal regions in the three BDT bins effectively employed in the signal yield extraction are used to compare the performance of the Run 1 and the 2015/16 analyses. The value of the figure of merit $S/\sqrt{S+B}$, where S and B are the expected SM yield of the signal and the interpolated background yield in the signal region respectively, is evaluated for both the analyses and shown in table 12.2. Even if the BDT selection of the 2015/16 analysis shows a lower performance with respect to the one employed in Run 1, the value of the figure of merit

Table 12.2: Comparison of the performance of the 2015/16 and Run 1 analyses. The performance is quantified with the figure of merit $S/\sqrt{S+B}$, where S and B are the yields of the SM expected signal and interpolated background in the signal region of the three higher BDT bins respectively. The number of expected signal events includes both $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ contributions.

	Run 1			Run 2		
	N_{bkg}	Expected N_{sig}	$\frac{S}{\sqrt{S+B}}$	N_{bkg}	Expected N_{sig}	$\frac{S}{\sqrt{S+B}}$
bin 1	509	15.3	0.7	330	33.7	1.8
bin 2	32	15.3	2.2	51	33.7	3.7
bin 3	4.8	15.3	3.4	7.9	33.7	5.2

in the three BDT bins is higher. This is most likely due to the additional precision on the measurement of e. g. the decay length resolution, granted by the introduction of the IBL in the ID.

Without applying any bounds on the values of the fitted parameters, the number of signal events obtained by the fit are $N_{B_s^0} = 80 \pm 22$ and $N_{B_d^0} = -12 \pm 20$, compatible with the SM expectation, with a correlation of $\rho = 0.82\%$. The uncertainties are obtained using a gaussian approximation in the neighbourhood of the likelihood maximum and include the contributions of the statistical uncertainty and the systematic uncertainties associated with the fitting procedure; statistical uncertainties largely dominate.

The non-physical result obtained for the $B_d^0 \rightarrow \mu^+ \mu^-$ yield does not constitute an issue for the validity of the analysis. Given its underfluctuation, the Neyman construction for the extraction of the confidence intervals on the branching fractions will provide an upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

Figure 12.1 shows the invariant mass of distributions of the four BDT bins, together with the projection of the likelihood.

The fitting procedure for the extraction of the branching fractions, discussed in section 11.4, is

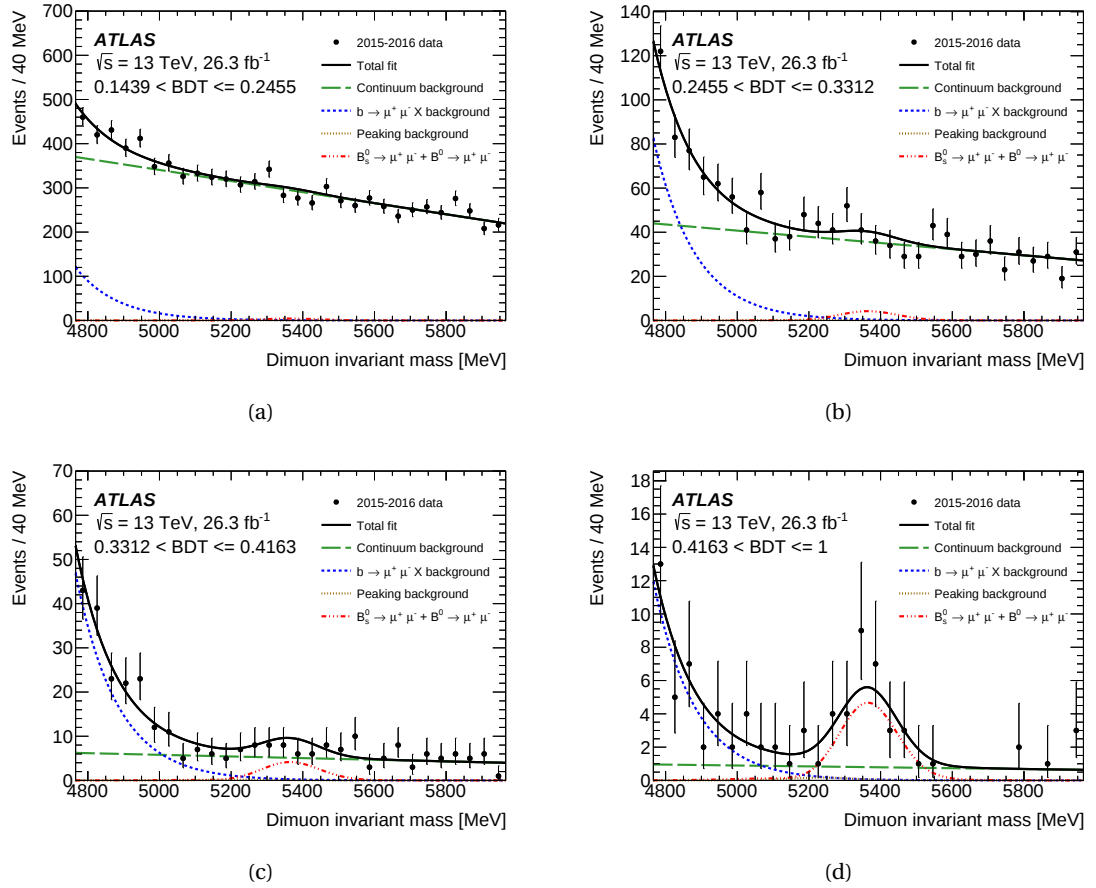


Figure 12.1: Di-muon invariant mass distributions in the unblinded data, in the four intervals of BDT output. Superimposed is the result of the maximum-likelihood fit. The total fit is shown as a continuous line, with the dashed lines corresponding to the observed signal component, the $b \rightarrow \mu\mu X$ background, and the continuum background. The signal components are grouped in one single curve, including both the $B_s^0 \rightarrow \mu^+ \mu^-$ and the (negative) $B_d^0 \rightarrow \mu^+ \mu^-$ component. The curve representing the peaking $B \rightarrow hh'$ background lies very close to the horizontal axis in all BDT bins.

then applied to evaluate $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, obtaining

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 3.21 \times 10^{-9} \quad (12.1)$$

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = -1.3 \times 10^{-10}. \quad (12.2)$$

Before evaluating the uncertainty on the measurements, the quality of the fit performed on data needs to be checked. The following sub-section provides a discussion on the evaluation of the goodness-of-fit.

12.1.2 Unblinded fit quality

The estimation of the quality of the fitting procedure performed on the unblinded data represents an interesting problem. Due to the simultaneous fit and the small statistics available in the last BDT bin, the usage of a standard binned χ^2 test would not provide a meaningful result. Binned χ^2 tests are, in fact, known to provide a meaningful estimation of the fit quality only if all the bins in the histogram present sufficient statistics to assume symmetric errors on the bin content.

The fit quality is therefore estimated with a modified Kolmogorov–Smirnov (KS) test [174]. The KS test is applied to each BDT bin between the normalised unbinned data distribution and the PDF resulting from the fit on data; the maximum KS distance found among the four bins is taken as resulting KS distance. The same procedure is applied to a set of toy-MC generated according to the result of the fit to data PDF parameters and fitted with the baseline fit.

Figure 12.2 shows the distribution of the maximum KS distance obtained from the toy-MC study and the result of the test applied to the fit to the real data (red arrow). The fit quality is extracted as the integral of the distribution on the right-hand side of the arrow, normalised by the integral of the entire distribution. The result is p -value is 0.84, which shows good agreement between the data and the fit.

12.1.3 Branching fractions uncertainty

As introduced at the beginning of this section, given the low-statistics regime of the analysis, a reliable approach for the evaluation of the uncertainty on the branching fractions is needed. The three methods for the extraction of the uncertainty (or confidence interval) described in chapter 6 are employed. The three approaches are applied successively, starting from the one with the most requirements on the properties of the likelihood (RCF variance [182]) and converging to the most accurate and reliable, the Neyman belt construction [178].

This procedure is employed for two main reasons. The first two approximate approaches can provide a cross-check for the Neyman belt construction. In fact, knowing the expected behaviour of the different procedures allows a gauge of the consistency of the final result of the analysis, e.g. the Neyman approach is expected to provide a larger uncertainty compared to the others, which underestimate the coverage. Additionally, providing the uncertainty with all three approaches gives a way to validate the ATLAS result against the result from other experiments, based on the uncertainty extraction procedure they employ.

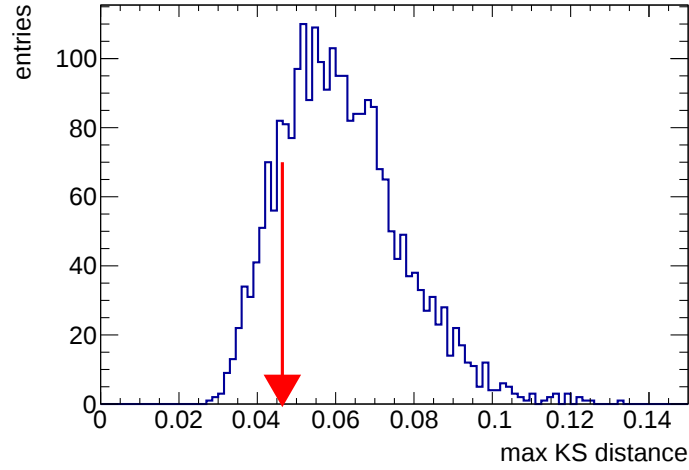


Figure 12.2: Distribution of the KS distance calculated as described in the text for a set of toy-MC generated using the result of the fit on data as generation PDF parameters. The red arrow shows the KS distance value obtained for the data fit. The fit quality is extracted as the integral to the right of the arrow of the normalised histogram. The resulting p -value is 0.84.

The first approach employed for the extraction of the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ uncertainty is the RCF variance, introduced in section 6.3.1. Assuming a gaussian behaviour of the likelihood in the neighbourhood of its maximum, the uncertainty on the signal branching fractions is evaluated applying formula 6.10. Formulas 12.1 and 12.2 therefore become:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.21 \pm 0.85 \pm 0.37) \times 10^{-9} = (3.21 \pm 0.93) \times 10^{-9} \quad (12.3)$$

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (-1.3 \pm 2.0 \pm 0.6) \times 10^{-10} = (-1.3 \pm 2.1) \times 10^{-10}. \quad (12.4)$$

Statistical and systematic uncertainties (shown separately at first, and then combined) are obtained by repeating the likelihood fit after removing the contribution of the gaussian smearing terms in the likelihood function. Table 12.3 shows the breakdown of the systematic uncertainties on the two branching fractions. In both cases the dominant systematic uncertainty is due to the fitting procedure. The last row shows the statistical uncertainty, which is the main source

Table 12.3: Breakdown of the systematic uncertainties in $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. The measurements are dominated by statistical uncertainty, followed by the systematic uncertainty from the fit. The latter is dominated by contributions from the mass scale uncertainty and the parameterisation of the low mass background. The statistical uncertainties reported here are obtained from the maximisation of the fit likelihood and are meant only as a reference for the relative scale uncertainties.

Source	B_s^0 [%]	B_d^0 [%]
f_s/f_d	5.1	-
B^+ yield	4.8	4.8
R_{Ac}	4.1	4.1
$\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$	2.9	2.9
Fit systematic uncertainties	8.7	65
Stat. uncertainty (from likelihood est.)	27	150

of uncertainty in the analysis.

As introduced in section 6.3.1, the resulting uncertainties are symmetric around the central value. Given the low statistical regime of the analysis, asymmetric uncertainties are expected. Additionally, the RCF variance does not provide any information about the “true” physical value of the branching fraction.

The second approach described in chapter 6 is therefore employed: approximated confidence

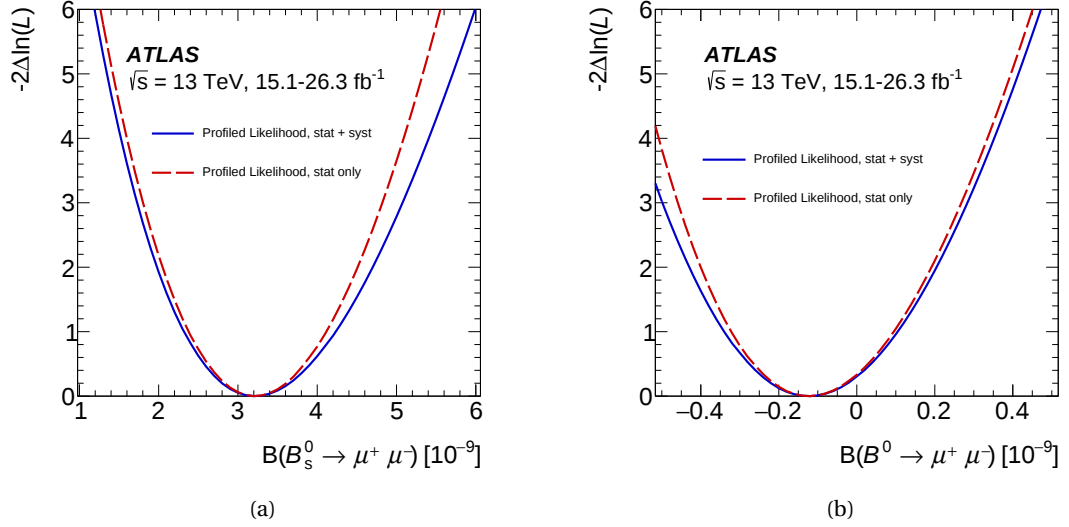


Figure 12.3: Profiled likelihood ratio scan of $-2\ln\left(\frac{\mathcal{L}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{L}(\max)}\right)$ as a function of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, for $B_s^0 \rightarrow \mu^+ \mu^-$ (left), and $B_d^0 \rightarrow \mu^+ \mu^-$ (right). The continuous curve includes statistical and systematic uncertainties, while the dashed one represents the effect of statistical uncertainties alone. The corresponding $-2\ln\left(\frac{\mathcal{L}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{L}(\max)}\right) = 1$ intersections are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.21_{-0.83-0.31}^{+0.90+0.48}\right) \times 10^{-9}$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = \left(-1.3_{-1.9-0.8}^{+2.2+0.7}\right) \times 10^{-10}$.

intervals built using the profiled likelihood ratio. Exploiting the profiled likelihood ratio scans shown in figure 12.3, the approximated confidence intervals on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ can be evaluated taking $-2\Delta\log(\mathcal{L}/\mathcal{L}(\max)) = 1$:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.21_{-0.83-0.31}^{+0.90+0.48}\right) \times 10^{-9} = \left(3.21_{-0.89}^{+1.02}\right) \times 10^{-9} \quad (12.5)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = \left(-1.3_{-1.9-0.8}^{+2.2+0.7}\right) \times 10^{-10} = \left(-1.3_{-2.1}^{+2.3}\right) \times 10^{-10}. \quad (12.6)$$

When two sets of errors are reported, they refer to the statistical and systematic uncertainties, obtained repeating the profiled likelihood ratio scan after removing the contribution of the gaussian smearing terms in the likelihood function. In both scans the likelihood function is profiled with respect to all other parameters, including the other branching fraction.

The 2D likelihood contours are also evaluated by profiling the likelihood function with respect to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ and taking $-2\Delta\log(\mathcal{L}/\mathcal{L}(\max)) = 2.3, 6.2$ and 11.8 . Such contours are shown in figure 12.4, where the red contours include statistical uncertainty only and the blue contours statistical+systematic uncertainty.

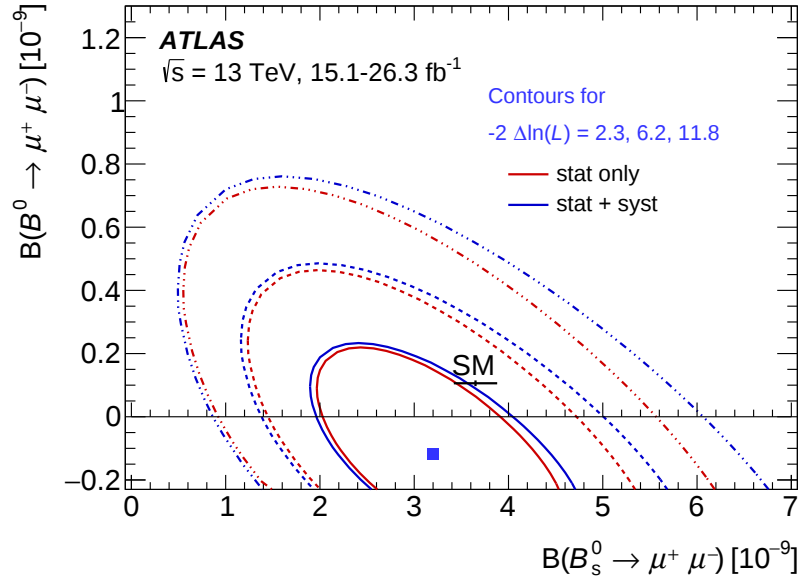
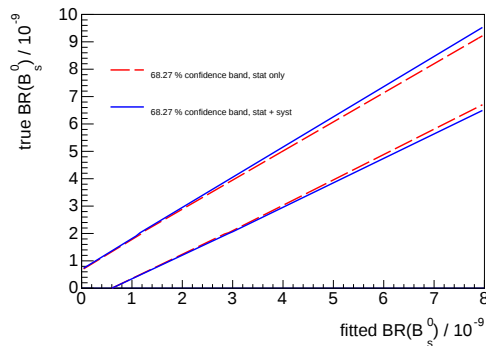


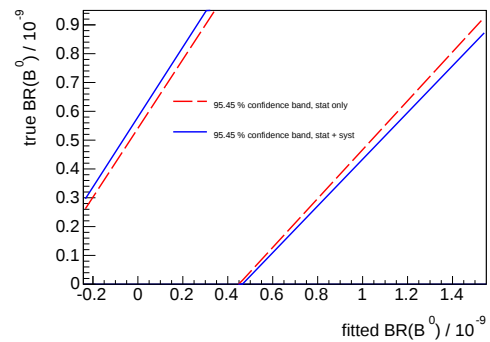
Figure 12.4: Two-dimensional likelihood contours for the simultaneous fit to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, for values of $-2\Delta \ln(\mathcal{L}/\mathcal{L}(\text{max}))$ equal to 2.3, 6.2 and 11.8. The SM prediction with uncertainties is indicated.

Compared to the RCF variance, the profiled likelihood scans produce asymmetric uncertainties. As introduced in section 6.3.2, the intervals obtained can be interpreted as confidence intervals with a given coverage only if the likelihood can be assumed to have a gaussian behaviour. Considering the scans in figure 12.3 and the fact that this approach provided asymmetric uncertainties, the likelihood can not be considered gaussian, hence the $-2\Delta \log(\mathcal{L}) = 1$ interval can not be assumed to be a 68.3 % confidence level interval.

The third and final approach described in chapter 6, the exact frequentist Neyman belt construction [178], is then employed to extract exact confidence intervals. Toy-MCs are generated to construct the 1D and 2D Neyman belts for the extraction of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$. The resulting 1D belts are shown in figures 12.5(a) and 12.5(b) for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$. Both belts



(a) 68.3% confidence bands for $B_s^0 \rightarrow \mu^+ \mu^-$.



(b) 95.5% confidence bands for $B_d^0 \rightarrow \mu^+ \mu^-$.

Figure 12.5: 68.3 % Confidence band for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (figure 12.5(a)) and 95.5 % confidence band for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ (figure 12.5(b)), obtained with toy-MCs, following the Neyman belt construction approach [178]. The continuous belt includes statistical and systematic uncertainties, while the dashed one represents the effect of statistical uncertainties alone.

are built considering statistical uncertainty only (red dashed lines) and statistical+systematic uncertainties (blue solid lines). The $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ belt has a coverage of 68.3 % and is employed to extract the confidence interval

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.21_{-0.91-0.30}^{+0.96+0.49}\right) \times 10^{-9} = \left(3.2_{-1.0}^{+1.1}\right) \times 10^{-9}. \quad (12.7)$$

When multiple errors are reported, they refer to the statistical and systematic uncertainties, obtained by building the Neyman belt after removing the contribution of the gaussian smearing terms in the likelihood function.

Given the underfluctuation of the $B_d^0 \rightarrow \mu^+ \mu^-$ yield, its belt is built with a coverage of 95.5 %, in order to extract the upper limit

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-10}. \quad (12.8)$$

The Neyman construction is also exploited to extract the expected result of the analysis assuming the SM predicted branching fractions [54], obtaining $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.6_{-1.0}^{+1.1}\right) \times 10^{-9}$ and an upper limit at 95% CL of 7.1×10^{-10} for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

The results shown in formulas 12.7 and 12.8 are compatible with the SM predictions.

Finally, the 2D belt is also constructed: the resulting 68.3 %, 95.5 % and 99.7 % confidence regions, corresponding to the 1, 2 and 3 σ contours, are shown in figure 12.6.

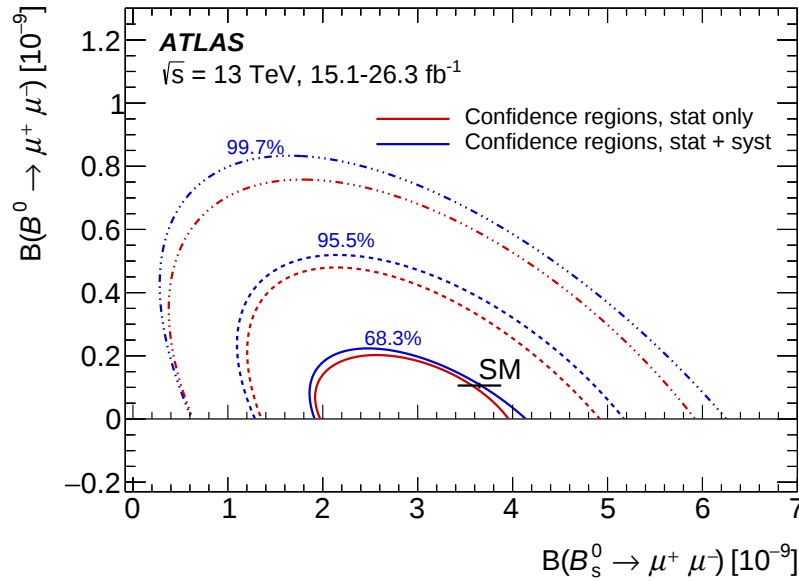


Figure 12.6: Neyman contours in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane for 68.3%, 95.5% and 99.7% coverage. The inner (red) contours are statistical uncertainty only, while the outer ones (blue) include statistical and systematic uncertainties. The construction of these contours makes use of both the dimuon (26.3 fb^{-1}) and the reference channel (15.1 fb^{-1}) datasets. The SM prediction for the signal branching fractions with its uncertainties [54] is also indicated.

Compared to the results obtained with the previous approaches, the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ confidence intervals extracted with the Neyman approach are different, enforcing the initial statement of this chapter, that an exact approach for the evaluation of the uncertainty on the branching fractions is needed. The Neyman approach allows also to extract in a natural way the upper

limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, without requiring the usage of additional statistical tools.

The resulting confidence interval (or upper limit) obtained with the Neyman construction constitutes therefore the final result of the analysis, summarised in the following sub-section.

12.1.4 2015/16 analysis result

In conclusion, the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during 2015 and 2016 by the ATLAS experiment yields a 68.3 % confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.2_{-1.0}^{+1.1}\right) \times 10^{-9}. \quad (12.9)$$

obtained exploiting the Neyman belt approach.

A similar construction is exploited for the evaluation of the upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ at 95.5% coverage of

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-10}. \quad (12.10)$$

Both results are compatible with the SM predicted branching fractions [54].

12.2 Combination with the Run 1 result

The result obtained in the previous section employs a large part of the data ever collected by the ATLAS experiment until the end of 2016, but, as already mentioned, an analysis was also performed on the dataset collected during the Run 1 of the LHC data taking period. The two analyses are combined to improve the results already obtained.

In order to limit the CPU consumption to obtain the results in this section, the uncertainties derived, unless explicitly reported, always include both statistical and systematic contributions. This section is dedicated to the combination of the two analyses. After a comparison between the two results (sub-section 12.2.1), the procedure employed to perform the combination is described (sub-section 12.2.2). The result of the combination is finally obtained in sub-section 12.2.3 and compared to the most recent experimental results (sub-section 12.2.4).

12.2.1 Comparison with the Run 1 result

The two ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses performed on the full Run 1 dataset [73] and on the 2015/16 dataset, described in the previous chapters, yield the following results:

$$\begin{aligned} \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{\text{Run 1}} &= (0.9_{-0.8}^{+1.1}) \times 10^{-9} & \mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)_{2015/16} &= \left(3.2_{-1.0}^{+1.1}\right) \times 10^{-9} \\ \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)_{\text{Run 1}} &< 4.2 \times 10^{-10} & \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)_{2015/16} &< 4.3 \times 10^{-10}. \end{aligned} \quad (12.11)$$

Both analyses show a final result which is lower than the expected SM branching fractions [54]. In particular, both analyses show an underfluctuation on the measured $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, while regarding $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, the 2015/16 analysis shows a result which is compatible with the SM and the Run 1 analysis reports a lower branching fraction.

Such underfluctuations can be motivated in different ways. A possible interplay between the two signals in the fitting procedure can explain fluctuations, but in such a case one of the two would overfluctuate and the other underfluctuate. Since both channels show underfluctuations, the possibility of such interplay must be discarded.

In case the $B \rightarrow hh'$ contribution is overestimated, the $B_d^0 \rightarrow \mu^+ \mu^-$ branching fraction could be underestimated, as signal events would leak into the peaking background PDF. In both analyses the $B \rightarrow hh'$ expected yield is cross-checked on data, enforcing the validity of the gaussian smearing term associated with the $B \rightarrow hh'$ PDF. Even in case of an issue in the $B \rightarrow hh'$ evaluation, the expected $B \rightarrow hh'$ contribution is extremely small (1 event for Run 1 and 2.9 for 2015/16) not enough to motivate the B_d^0 underfluctuation.

Effects due to new physics can modify the branching fractions, as discussed in section 1.4, but this possibility would add tension between the ATLAS result and the latest results from other experiments, which do not show any significant discrepancy with the SM expectation.

Finally, statistical fluctuations in the signal and/or the background components can motivate the results obtained in both analyses. This is considered the most probable possibility.

The compatibility of the two analyses is gauged with a χ^2 test, which yields a compatibility of 1.16 standard deviations, giving a fair agreement between the results. The combination of the results is expected to provide a central value smaller than the one found in the 2015/16 analysis alone, due to the lower values of the branching fractions found in Run 1. Anyway, thanks to the higher statistics available, the central value is expected to be closer to the 2015/16 result.

12.2.2 Analyses combination procedure

The combination of the Run 1 and 2015/16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses is based on the likelihood combination. This procedure, discussed later in this sub-section, allows the combination two likelihoods into one, while considering possible correlated effects.

The basic concept of a likelihood combination is simple, the two likelihood are multiplied with each other maintaining the two signal branching fractions as common parameters. The resulting likelihood is then maximised with respect to the signal branching fractions, as was done for the likelihoods of the two analyses. This is the same idea employed in the simultaneous likelihood fits, discussed in section 6.1 and largely used in both the Run 1 and 2015/16 analyses.

Correlations between the two analyses due to the usage of the same parameters are considered. The values employed in formula 5.1 and in its Run 1 counterpart which are not evaluated in the analyses ($\frac{f_u}{f_{s(d)}}$ and $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$) are set to be fully correlated and to have the most recent measured values (employed in the 2015/16 analysis):

- $\frac{f_u}{f_{s(d)}} = 0.256 \pm 0.013$, which assumes $\frac{f_u}{f_d} = 1$ [28];
- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.010 \pm 0.029) \times 10^{-3}$ and $\mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) = (5.961 \pm 0.033)\%$ [11].

The full correlation of these parameters is implemented employing the same smearing term for the systematics uncertainty association in both likelihoods.

The two likelihoods are therefore combined as follows

$$\begin{aligned}
 & \mathcal{L}_1(s_1, \theta_1^{\text{common}}, \theta_1^{\text{only } \mathcal{L}_1}); \mathcal{L}_2(s_1, \theta_2^{\text{common}}, \theta_2^{\text{only } \mathcal{L}_2}) \\
 & \quad \downarrow \\
 & \mathcal{L}_1(s, \theta^{\text{common}}, \theta_1^{\text{only } \mathcal{L}_1}) \times \mathcal{L}_2(s, \theta^{\text{common}}, \theta_2^{\text{only } \mathcal{L}_2}) \\
 & = \mathcal{L}_{\text{tot}}(s, \theta^{\text{common}}, \theta_1^{\text{only } \mathcal{L}_1}, \theta_2^{\text{only } \mathcal{L}_2}),
 \end{aligned} \tag{12.12}$$

where $\mathcal{L}_i$ represents the likelihood of one of the two experiments ($i = 1, 2$), with s_i being the parameters of interest, θ_i^{common} the nuisance parameters in common between the two likelihoods and $\theta_i^{\text{only } \mathcal{L}_i}$ the nuisance parameters not in common. The resulting likelihood only has one set of parameters of interest and common nuisance parameters, which are shared between the two likelihoods.

12.2.3 Branching fraction extraction

The approach for the combination of the likelihoods discussed in the previous sub-section is implemented and the combined likelihood is then maximised. The results obtained from this procedure are discussed in this sub-section; first the signal yields and branching fractions which maximise the likelihood are discussed, followed by the extraction of their uncertainties and confidence intervals, as was performed for the 2015/16 analysis in section 12.1.3.

The maximum of the likelihood is found for the following signal branching fraction values:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 2.8 \times 10^{-9} \tag{12.13}$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = -1.9 \times 10^{-10}. \tag{12.14}$$

The corresponding number of signal events obtained for each channel and period are summarised in table 12.4, where the uncertainties reported, evaluated assuming a gaussian behaviour

Table 12.4: Number of signal events fitted with the the combination of the 2015+2016 Run 2 likelihood with the one for the full Run 1 result. The uncertainties reported, which consider the statistical uncertainty and the systematic uncertainty arising from the fitting procedures employed in the two analyses, are evaluated assuming a gaussian behaviour of the likelihood in the vicinity of its maximum.

Period	Channel	Fitted yield
Run 1	$B_s^0 \rightarrow \mu^+ \mu^-$	30.7 ± 8.1
	$B_d^0 \rightarrow \mu^+ \mu^-$	-8.7 ± 7.2
2015/16	$B_s^0 \rightarrow \mu^+ \mu^-$	72.0 ± 18.1
	$B_d^0 \rightarrow \mu^+ \mu^-$	-17.8 ± 15.0

of the likelihood in the neighbourhood of its maximum, consider the statistical uncertainty and the systematic uncertainty arising from the fitting procedures employed in the two analyses. As expected the central values of the measured branching fractions are lower compared to the ones found in the 2015/16 analysis.

The same approach for the extraction of the uncertainties and confidence intervals exploited for the 2015/16 analysis is used in this case. The uncertainties are extracted employing the three techniques described in section 6.3, starting from the one with the most requirements on the properties of the likelihood and ending with the exact Neyman belt approach.

The uncertainty extracted with the RCF variance (section 6.3.1) are the following:

$$\begin{aligned}\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) &= (2.8 \pm 0.7) \times 10^{-9} \\ \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) &= (-1.9 \pm 1.6) \times 10^{-10}.\end{aligned}$$

Where the uncertainties include statistical and systematic contributions. Similar considerations as the ones made for the 2015/16 result can be made in this case. This approach does not provide any information regarding the “true” values of the branching fractions and the uncertainties are symmetric with respect to the central value. This is not expected, in fact, given the low statistics regime of the analysis, asymmetric uncertainties should appear.

The likelihood intervals (section 6.3.2) are employed to estimate the approximated confidence intervals and contours. Figure 12.7 shows the likelihood contours obtained profiling the likelihood with respect to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ and drawing the contours for $-2\Delta\log(\mathcal{L}/\mathcal{L}(\max)) = 2.3, 6.2$ and 11.8 . The combined likelihood contours are shown with black lines and shaded areas, together with the 2015/16 likelihood contours (blue lines) and, in

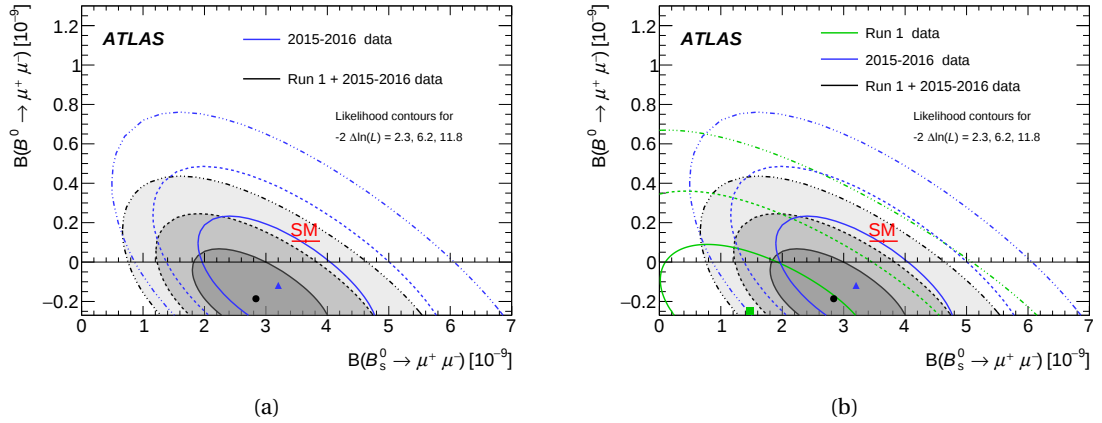


Figure 12.7: Comparison of the Run 1 and 2015/16 combined likelihood contours (black lines and shaded areas) with the likelihood contours obtained using the 2015/16 likelihood (figure 12.7(a)) and also with the likelihood contours obtained using the Run 1 likelihood (green lines in figure 12.7(b)). All contours include the contribution of statistical and systematic uncertainties. The three sets of contours for each result are obtained by profiling the likelihood with respect to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ and taking $-2\Delta\log(\mathcal{L}/\mathcal{L}(\max)) = 2.3, 6.2$ and 11.8 . The SM expected branching fractions [54] are also indicated.

figure 12.7(b), also the Run 1 likelihood contours (green lines). The compatibility of the 2015/16 and Run 1 analyses is cross-checked by taking the value of $-2(\log \mathcal{L}_{\text{Run 1}} + \log \mathcal{L}_{2015/16})$ at the maximum value of the combined likelihood, with $\mathcal{L}_{\text{Run 1}}$ and $\mathcal{L}_{2015/16}$ set to have maxima at zero. This quantity can be interpreted as a χ^2 with two degrees of freedom [184], leading to a compatibility of 1.2 standard deviations, in good agreement with what was found in subsection 12.2.1.

The likelihood ratio scans profiled with respect to only one of the two signal branching fractions, shown in figure 12.8, are then employed to obtain the approximated confidence intervals

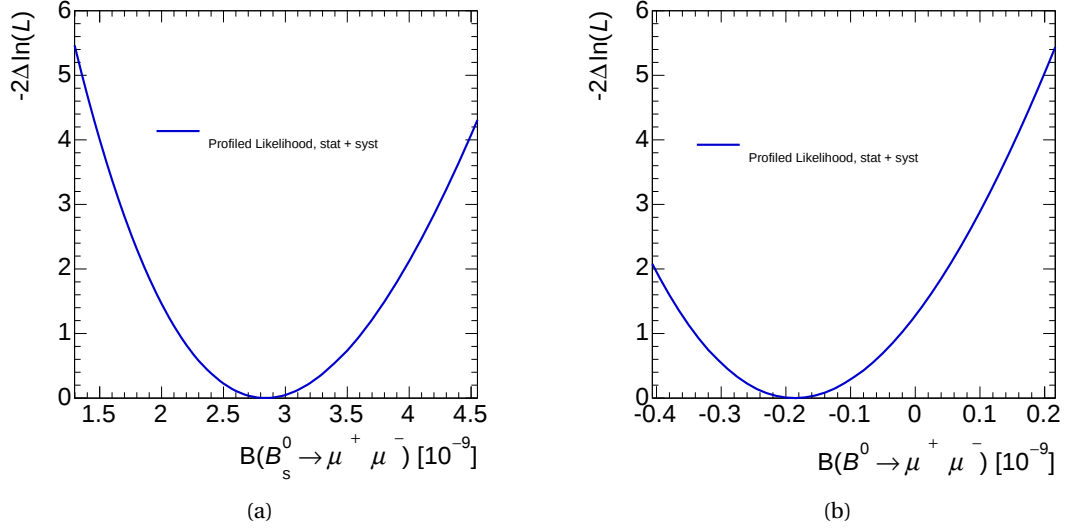


Figure 12.8: Profiled likelihood ratio scan of $-2\ln\left(\frac{\mathcal{L}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{L}(\max)}\right)$ as a function of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, for $B_s^0 \rightarrow \mu^+ \mu^-$ (left), and $B_d^0 \rightarrow \mu^+ \mu^-$ (right), obtained using the Run 1 + 2015/16 combined likelihood. The curves include statistical and systematic uncertainties. The corresponding $-2\ln\left(\frac{\mathcal{L}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)}{\mathcal{L}(\max)}\right) = 1$ intersections are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (-1.9^{+1.7}_{-1.5}) \times 10^{-10}$.

taking $-2\Delta\log(\mathcal{L}) = 1$:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9} \quad (12.15)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (-1.9^{+1.7}_{-1.5}) \times 10^{-10}, \quad (12.16)$$

where the uncertainties reported include statistical and systematic contributions. As for the 2015/16 result, the profiled likelihood ratio scans do not show a gaussian behaviour, as visible in figure 12.8 and by the asymmetric uncertainties on the branching fractions. This means that the intervals obtained for $-2\Delta\log(\mathcal{L}) = 1$ do not ensure a coverage of 68.3 %.

The Neyman belt approach is therefore employed to extract the exact confidence intervals. Only the one-dimensional Neyman belts for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ are obtained, following the approach discussed in sections 6.3.3 and 6.3.4; in order to limit the CPU consumption, the two-dimensional belt is not constructed. Figure 12.9 shows the one-dimensional belts for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ (figure 12.9(a)), built with a coverage of 68.3 %, and the one-dimensional belt for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ (figure 12.9(b)), built with a coverage of 95.5 %; in both cases the belts include both statistical and systematic uncertainties. The 68.3 % confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is therefore

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8^{+0.8}_{-0.7}) \times 10^{-9}, \quad (12.17)$$

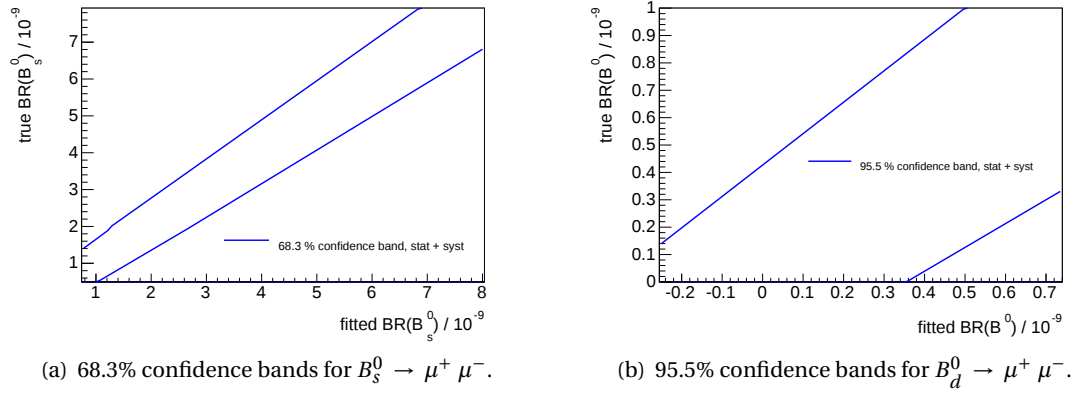


Figure 12.9: 68.3 % confidence band for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$, figure 12.9(a), and 95.5 % confidence band for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, figure 12.9(b); both are obtained with toy-MCs, following the Neyman belt construction approach discussed in sections 6.3.3 and 6.3.4. Both include the contribution of statistical and systematics uncertainties.

and the upper limit at 95.5% CL on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ is

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 2.1 \times 10^{-10}. \quad (12.18)$$

For comparison, using the predicted SM branching fractions [54], the analysis is expected to yield on average a measurement of $(3.6^{+0.9}_{-0.8}) \times 10^{-9}$ for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and an upper limit of 5.6×10^{-10} for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$. The two results shown in formulas 12.17 and 12.18 are compatible with the SM expectations.

The significance of the result obtained for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is estimated evaluating the profiled likelihood ratio shown in figure 12.8(a) in the hypothesis of no signal ($\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$) [184]. This test yields a profiled likelihood ratio value of 21.5, which, for one degree of freedom corresponds to a significance of 4.6 standard deviations. This measurement is therefore the first ATLAS evidence of the $B_s^0 \rightarrow \mu^+ \mu^-$ process.

The compatibility of the result obtained with the SM expected branching fractions [54] is measured in a similar way, evaluating the likelihood ratio profiled with respect to $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ at the SM expected value [184]. The resulting significance in this case is 2.4 standard deviations, considering two degrees of freedom.

12.2.4 Comparison with other experiments

The result obtained in the previous sub-section can be compared with the most recent experimental results, which are obtained by the other LHC experiments.

As introduced in section 1.5, the most recent results are the combined CMS and LHCb analyses based on the full Run 1 dataset [70] and the LHCb analysis performed on the Run 1 dataset together with to first part of the dataset collected in Run 2 [74]. The CMS and LHCb combination yields the following results:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9} \quad (12.19)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}. \quad (12.20)$$

While the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ result shows a remarkable compatibility with the ATLAS result discussed in the previous section, the estimated $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ shows tensions. Such disagreement is not present when comparing the ATLAS result with the most recent LHCb result, as shown in figure 12.10, where the profiled likelihood contours drawn in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane for the ATLAS 2015/16 and Run 1 combined analyses and for the LHCb

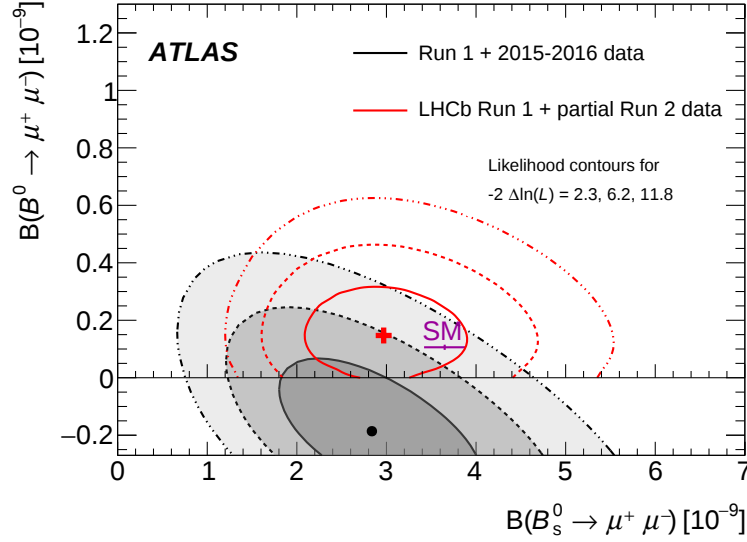


Figure 12.10: Comparison of the Run 1 and 2015/16 combined likelihood contours (black lines and shaded areas) with the likelihood contours obtained by the LHCb experiment (red lines) [74]. The SM expected branching fractions [54] are also indicated.

analysis [74] are compared. The LHCb analysis yields

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9} \quad (12.21)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 3.4 \times 10^{-10} \quad \text{at 95\% CL.} \quad (12.22)$$

Also in this case $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ result shows good compatibility with the ATLAS result. The upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ estimated by LHCb is compatible, even if less stringent¹, with the ATLAS one shown in formula 12.18. The ATLAS $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ upper limit obtained with the combination of the 2015/16 and Run 1 analyses is therefore, as of today, the most stringent upper limit ever found on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

12.3 Summary of results

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during 2015 and 2016 with the ATLAS experiment is concluded. The results of the analysis and its combination with the one performed on the Run 1 dataset are summarised in this section.

The analysis performed on the 2015/16 dataset yields a 68.3 % confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$

¹ While the LHCb sensitivity is higher than ATLAS', the underfluctuation on the $B_d^0 \rightarrow \mu^+ \mu^-$ yield allows the ATLAS upper limit to be more stringent.

of

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(3.2_{-1.0}^{+1.1}\right) \times 10^{-9}, \quad (12.23)$$

and an upper limit at 95.5 % CL on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ of

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 4.3 \times 10^{-10}. \quad (12.24)$$

Both results are obtained with the exact frequentist Neyman belt approach [178] and are compatible with the SM expectations [54].

The combination of the 2015/16 analysis with the ATLAS analysis performed on the Run 1 dataset leads to the first ATLAS evidence of the $B_s^0 \rightarrow \mu^+ \mu^-$ process, with a 68.3 % confidence interval obtained with the same Neyman construction:

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = \left(2.8_{-0.7}^{+0.8}\right) \times 10^{-9}. \quad (12.25)$$

The Neyman approach is also employed to derive the upper limit at 95.5 % CL on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, leading to the most stringent upper limit ever found, thanks to its underfluctuation:

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 2.1 \times 10^{-10}. \quad (12.26)$$

Both results are compatible with the SM expected branching fractions [54].

12.4 LHC experiment $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ combination

In order to increase the reach of the current published analysis performed by the LHC experiments, an effort to combine their results is ongoing. The results shown in this section are still work-in-progress therefore will be subject to changes before being published.

The combination is based on the following analyses:

- ATLAS analysis [4]: performed on the first two years of the Run 2 data-taking period and combined with the likelihood from the ATLAS Run 1 analysis; this is the analysis described in chapters 5 to 12.2;
- CMS analysis [71]; performed on the full Run 1 dataset, this analysis has been briefly introduced in section 1.5;
- LHCb analysis [74]: performed on the dataset collected during the Run 1 and the first two years of the Run 2 data-taking periods; this analysis has been briefly introduced in section 1.5 as well.

The chosen approach for the combination is a sum of the profiled likelihoods. Two-dimensional histograms with the profiled likelihood ratios of the three experiments are shared among the experiments and summed. The minimum of the resulting histogram is then shifted to zero, obtaining the profiled likelihood ratio of the combination.

This approach allows an easy and quick way to combine the results from different experiments, but does not take into account the effect of common systematic uncertainties, such as the reference channel branching fraction and the hadronisation probability. While the correlation effect on the reference channel branching fraction can be neglected, thanks to its small relative uncertainty, the correlation effect for the hadronisation probability has to be taken into account. For this reason, studies for its inclusion are currently ongoing.

Figure 12.11 shows the result of the combination, without considering the common systematics. Several likelihood contours are shown, representing the $-2\Delta\log(\mathcal{L}) = 2.3, 6.2, 11.8$ contours of the ATLAS experiment in blue, the CMS experiment in red, the LHCb experiment in green and the combination of the three in black with grey shades. The central value of the combination is shown with a black round marker, while the central values of the analyses of the three experiments are omitted, in order to reduce the amount of information in an already overcrowded plot. The SM theoretical prediction [54] is also shown.

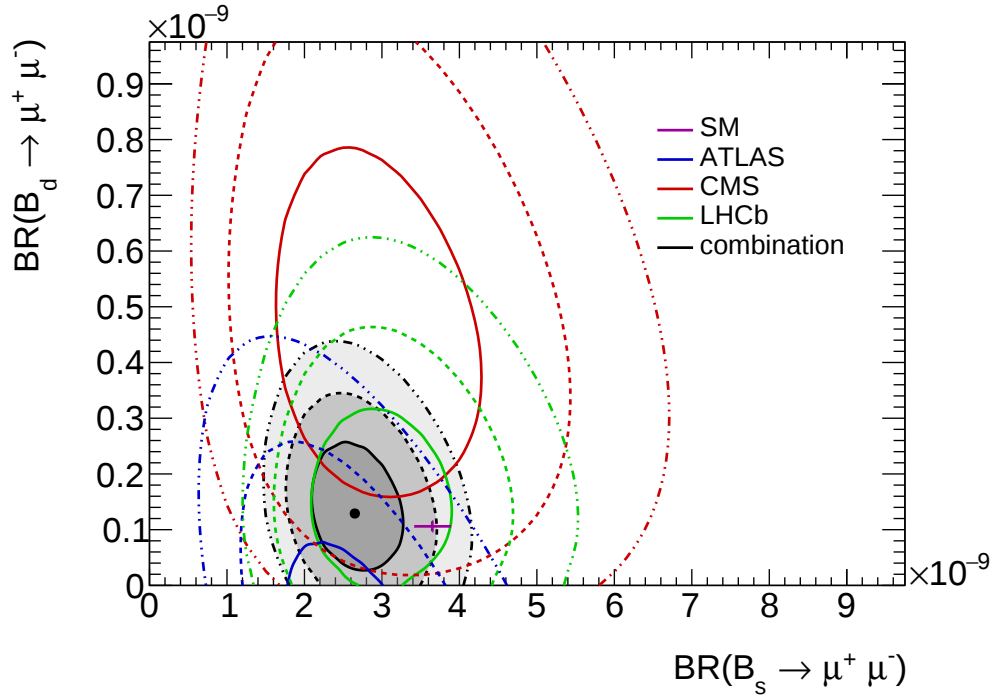
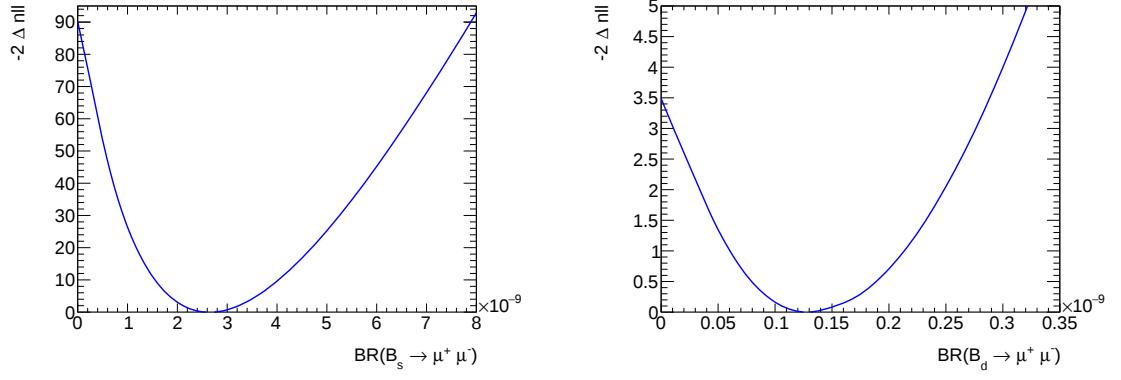


Figure 12.11: Two-dimensional likelihood contours for the ATLAS (blue), CMS (red) and LHCb (green) $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses and for their combination (black with shades of grey). The SM theoretical prediction [54] for the two branching fractions is also shown together with its uncertainty. The contours are obtained taking $-2\Delta\log(\mathcal{L}) = 2.3, 6.2, 11.8$ on the single experiments likelihoods, while the combination is obtained summing the profiled likelihood ratios of the three experiments. Common systematic uncertainties are not considered.

The approach followed for the combination of the likelihoods allows also to obtain the one-dimensional profiled likelihood ratio scans of the combination. Figure 12.12 shows the resulting plots for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$. The result of the combination is obtained taking the minimum of the two likelihood scans as central values and $-2\Delta\log(\mathcal{L}) = 1$ as uncer-



(a) One-dimensional profiled likelihood scan for $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$. (b) One-dimensional profiled likelihood scan for $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

Figure 12.12: One-dimensional profiled likelihood scans for the combination of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses of the ATLAS, CMS and LHCb experiments.

tainty. This yields in

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.65 \pm 0.4) \times 10^{-9} \quad (12.27)$$

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = \left(1.3_{-0.7}^{+0.8}\right) \times 10^{-10}. \quad (12.28)$$

The significance of the results is quantified with the same likelihood ratio approach employed in section 12.2.3. The significance on the $B_s^0 \rightarrow \mu^+ \mu^-$ measurement is 9σ , while the significance on $B_d^0 \rightarrow \mu^+ \mu^-$ is 1.85σ . The two results are compatible with the SM prediction within 2.3σ .

EXTRAPOLATIONS FOR THE $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ANALYSIS

13

This chapter presents simulation-based studies on the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis sensitivity. Two sets of studies are performed. The first study is carried out to estimate the sensitivity of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis employing the statistics collected during the Run 2 of the LHC data taking period (section 13.1); the second study compares the performance of the ATLAS and CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses performed on the Run 1 dataset, in order to understand the main drivers in the different performance of the two experiments (section 13.2).

13.1 Projection for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

The simulation-based projections of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis sensitivity using the statistics collected in Run 2 and to be collected at HL-LHC with the ATLAS detector are discussed in this section; the results are published as an ATLAS PUB note [5] and from an integral part of [206] and [3].

These studies are based on the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the Run 1 dataset [73] as they were carried out before the most recent analysis was published.

The projection strategy is described in sub-section 13.1.1, with the systematic uncertainties discussed in sub-sub-section 13.1.1.1. After validating the technique against the Run 1 ATLAS result (sub-section 13.1.2), the extrapolation proceeds with projections to the full Run 2 statistics (sub-section 13.1.3) and to the HL-LHC projected sensitivity (sub-section 13.1.4). The results are then discussed (sub-section 13.1.5) and compared against the extrapolations of other LHC experiments (sub-section 13.1.6).

13.1.1 Extrapolation procedure

The analysis sensitivity is going to be quantified through the expected $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - \mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ contour plots. This will require the choice of a reference yield for the extrapolations, which is chosen to be based on the expected SM branching ratios [54]. Rather than performing absolute extrapolations, the analysis statistical power gain is expressed relative to the ATLAS Run 1 result. This procedure leverages on a well established result based on

actual data, requiring to account for the possible differences in analysis performance and the relative gain in number of signal and background events.

This gain is then implemented in a PDF that recreates the Run 1 analysis likelihood, with further modifications discussed later in this sub-section. The resulting PDF is then used to generate toy-MC samples of the observable candidates invariant mass distribution. The resulting distributions are then fitted employing the same likelihood, via a simultaneous maximum-likelihood fit in invariant mass.

The evaluation of the branching ratios contours will be performed with two approaches. Due to the expected statistics regime of the Run 2 analysis the contours need to be derived using the Neyman belt construction technique [178] described in sections 6.3.3 and 6.3.4. Such contours will be compared with the likelihood based contours (discussed in section 6.3.2), in order to probe the asymptotic gaussian behaviour of the likelihood. This comparison represents a pivotal point in these studies, justifying the usage of the likelihood contours asymptotic approach for the HL-LHC projections.

Since likelihood contours are based on the the value which maximises the likelihood, confidence intervals are built by generating a large number of toy-MC samples and constructing the contours using one toy-MC, for which the maximum of the likelihood is the closest to the reference yield for the extrapolations. The remaining toy-MC samples are employed to gauge the variance of the contours.

This extrapolation of the signal and background statistics implicitly assumes the same selection and reconstruction efficiencies as the Run 1 analysis. The statistics extrapolation will be performed considering the increase in centre of mass energy and integrated luminosity, accounting for a different trigger selection. The main source of background of the Run 1 analysis was due to $b\bar{b}$ events, which are also responsible for the production of the B mesons of interest for the analysis. Thanks to the selection variables employed, e. g. vertexing quality and pointing angle requirements, the analysis is robust against pile-up, therefore heavy flavour decays from a single pp interaction are the dominant source of background after the final selection. Signal and background yields are therefore expected to scale in the same way, which allows to assume the same S/B as the Run 1 analysis for these studies.

The effects of the most relevant detector upgrades with respect to Run 1 are also considered. The insertion of the IBL in the ID, discussed in section 2.2.3, affects the signal mass resolution; this effect is taken into account in the Run 2 projections. The Phase-II upgrade, discussed in section 2.3, foresees major modifications to the detector. In particular, the ITk upgrade entails improvements in the vertex and mass determination, partly enhanced by the increased muon trigger momentum thresholds, necessary to cope with the higher luminosity and pile-up conditions. While the decay length resolution improvements are neglected in favour of a conservative S/B assumption, the signal mass resolution variations are taken into account in the projections. Figure 13.1, from [117], compares the $B_s^0 \rightarrow \mu^+ \mu^-$ mass spectrum obtained from Run 2 simulations and from simulations of the HL-LHC detector.

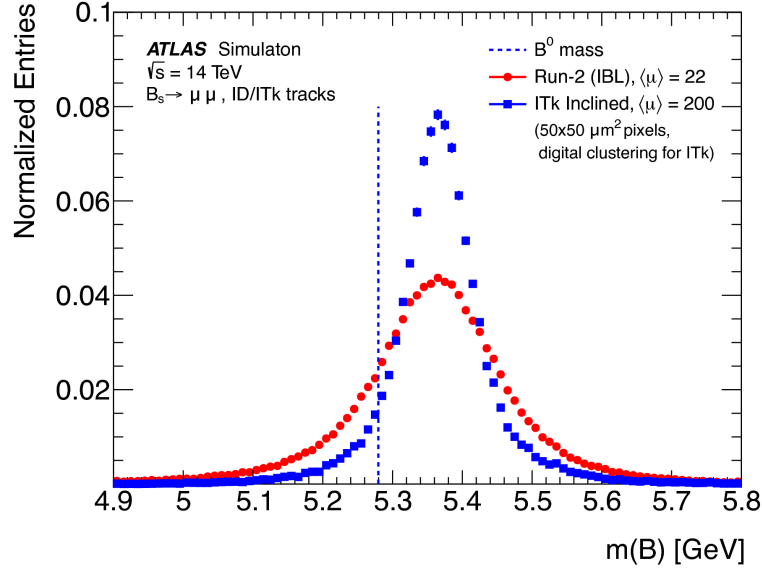


Figure 13.1: HL-LHC and Run 2 reconstructed $B_s^0 \rightarrow \mu^+ \mu^-$ mass spectra for muons with $|\eta| < 2.5$. For reference, the B_d^0 mass value [11] is indicated with a dotted line to emphasise the limited B_d^0 - B_s^0 resolving abilities of the ATLAS detector. The figure shows the mass resolution using only the ID (red)/ITk (blue) track parameter measurement, evaluated at the fitted B-vertex.

The expected gain in statistics is evaluated assessing separately three factors:

$$\begin{aligned}
 (\text{Projected statistics gain}) = & \frac{(\text{projected } B \text{ cross-section})}{(\text{Run 1 } B \text{ cross-section})} \\
 & \times \frac{(\text{projected Luminosity})}{(\text{Run 1 Luminosity})} \times \frac{(\text{projected trigger efficiency})}{(\text{Run 1 trigger efficiency})}. \quad (13.1)
 \end{aligned}$$

The B production cross-section scale factors used will assume, conservatively, that Run 1 data have been all collected at a centre of mass energy of 8 TeV. Sub-sections 13.1.3 and 13.1.4 are dedicated to the estimation of the different terms of equation 13.1.

13.1.1.1 Systematic uncertainties

Two main classes of systematic uncertainties, characterised by different properties, are parametrised in the Run 1 analysis and considered in the projections.

- *External systematics*: these systematic uncertainties are due to uncertainties on external inputs, namely $\frac{f_\mu}{f_{s(d)}}$ and $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$. It is plausible to expect these the uncertainties on these quantities to be reduced with other measurements, however, the present study conservatively assumes their values to be those used in the Run 1 analysis.
- *Internal systematics*: these systematic uncertainties depend on internal analysis effects, e. g. uncertainties on invariant mass fit shapes and efficiencies. In the Run 1 analysis they were parametrised as a function of the signal yields where dependencies were found to be significant. The same parameterisation is maintained for these extrapolations.

The same systematic uncertainties that affect the Run 1 analysis are therefore maintained.

Most of the sources of systematic uncertainties are not parametrised as a function of the signal yields, these quantities are:

- $f_s/f_d = 0.240 \pm 0.020$ [176];
- $\mathcal{B}(B^+ \rightarrow J/\psi K^+) = (1.027 \pm 0.031) \times 10^{-3}$ and $J/\psi \rightarrow \mu^+ \mu^- = (5.961 \pm 0.033)\%$ [207];
- the normalisation term $D_{\text{norm}} = (2.77 \pm 0.16) \times 10^6$, which encloses the reference channel yield, the ratio of the acceptances and efficiencies for the reference and signal channels, and the uncertainty associated with the total integrated luminosity.

Only the systematic uncertainties arising from the uncertainties on the fitting procedure present a dependence on the signals yield. Following the parameterisation obtained in Ref. [73], the systematic uncertainty for the fitted number of B_s^0 events is parametrised as $\sigma_{\text{syst}}(N_s) = \sqrt{4 + (0.06N_s)^2}$, while the uncertainty on B_d^0 yield is parametrised as $\sigma_{\text{syst}}(N_d) = 3$. The correlation between the B_s^0 and B_d^0 systematics is found to be $\rho_{\text{syst}} = -0.7$.

All the sources of systematic uncertainties are taken into account in the likelihood as gaussian smearing parameters; the systematic uncertainties associated with the fitting procedure, in particular, are included as two smearing parameters for the number of fitted B_s^0 and B_d^0 events constrained by a two-dimensional Gaussian distribution parametrised by the values of $\sigma_{\text{syst}}(N_s)$, $\sigma_{\text{syst}}(N_d)$ and ρ_{syst} .

13.1.2 Cross-check on Run 1 result

Since these studies are extensively based on the Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, the extrapolation technique is first cross-checked by reproducing the published ATLAS Run 1 profiled likelihood contour¹.

Toy-simulations are generated and fitted, based on the procedure described in sub-section 13.1.1, with all the factors in formula 13.1 set to one. Only the toy-simulations with a central value close to the one obtained in the Run 1 analysis are retained and employed to draw the $-2\Delta\log(\mathcal{L}/\mathcal{L}(\text{max})) = 2.3, 6.2$ and 11.8 profiled likelihood contours. Considering the amount of signal statistics, variations at the level of $\sim 15\%$ on such contours are expected; the contours obtained with toy-simulations are found to be compatible with the Run 1 result within such expectations. Figure 13.2 shows the Run 1 analysis result (left) and the result of one of the toy-MC generated (right).

Given the compatibility of the Run 1 likelihood contours with the contours obtained with the procedure described in sub-section 13.1.1, the toy-MC generation and fitting procedure are considered validated.

¹ The final result of the analysis was provided employing profiled likelihood contours, therefore the same procedure is used to test the extrapolation technique.

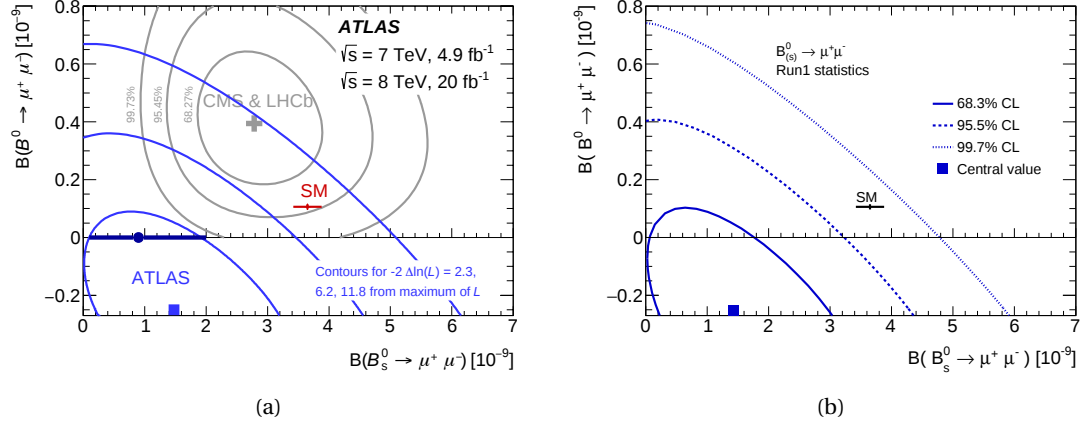


Figure 13.2: Comparison of the contour plot of the published Run 1 analysis 13.2(a) and the contours obtained with a toy simulation 13.2(b) which present a central value compatible with the Run 1 result.

13.1.3 Run 2 yield extrapolation

The three ingredients of equation 13.1 for the Run 2 extrapolation are computed as follows.

1. B production cross section with respect to Run 1:

Run 2 is characterised by a higher centre of mass energy with respect to Run 1. The increase in energy, 8 to 13 TeV, corresponds to an increase of approximately $\times 1.7$ in B production cross-section, according to studies performed using FONLL [208]. This factor is assumed for the first term of formula 13.1.

2. Expected collected luminosity for the selected triggers in Run 2:

an integrated luminosity of 130 fb $^{-1}$ is assumed for the full Run 2 data taking period. These studies were performed at the beginning of 2018, when Run 2 data were still being taken. Similar trigger conditions as the ones present in 2017 are projected to 2018. The ratio of integrated luminosities between the Run 2 assumption and Run 1 (130fb $^{-1}$ /25fb $^{-1}$ = 5.2) is therefore used in the second term of formula 13.1. These assumptions are, a posteriori, found conservative: the total integrated luminosity for the full Run 2 is 139 fb $^{-1}$ [194] and, when allowed by the pile-up conditions, on top of the triggers projected from 2017, events from additional low dimuon p_T triggers have been collected. These additional effects are not considered in these studies.

3. Efficiency of the difference in dimuon triggers available in Run 2 with respect to Run 1:

this is evaluated exploiting $B_s^0 \rightarrow \mu^+ \mu^-$ MC simulations for the Run 2 data-taking conditions. The computation of the relative Run 2 / Run 1 trigger efficiency is based on fully reconstructed simulated signal events, after the application of the same preselection as the Run 1 analysis. All trigger efficiencies are normalised with respect to the trigger with the lowest muon p_T thresholds available in Run 1; such trigger has p_T thresholds of 4 GeV for both muons, similar to the ones described in section 7.1.

The dataset employed in the Run 1 analysis was based on an admixture of different dimuon

triggers, as described in appendix A, in order to maximise the signal statistics. The combination² of several dimuon triggers is therefore considered also for the full Run 2 analysis. The third term of formula 13.1 is evaluated taking the ratio of the combined trigger efficiencies of Run 2 and Run 1, weighted according to their prescales. This term amounts to ~ 0.81 .

The evaluation of formula 13.1 based on these ingredients yields a 7-fold increase with respect to the Run 1 statistics.

13.1.4 HL-LHC yield extrapolation

The HL-LHC $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ signal yields is extrapolated as follows:

1. **B production cross section with respect to Run 1:**

any further increase in the centre of mass energy relative to Run 2 is neglected, conservatively assuming the same $\times 1.7$ factor as for the Run 2 extrapolation.

2. **Expected collected luminosity for the selected triggers at HL-LHC:**

an integrated luminosity of $\sim 3 \text{ ab}^{-1}$ (3000 fb^{-1}) is assumed to be collected during the whole HL-LHC data taking period (section 2.3);

3. **Efficiency of the di-muon triggers available at HL-LHC with respect to the Run 1 triggers:**

the efficiency of the HL-LHC triggers is evaluated exploiting the same simulations used in sub-section 13.1.3. The pile-up conditions in these simulations differ from what is expected for HL-LHC, but this is secondary when assessing truth-matched signal efficiencies. Depending on the delivered instantaneous luminosity, three possible dimuon triggers are foreseen [122]. Similar topological selections as those employed in Run 2 are assumed, with di-muon transverse momentum thresholds ($p_T^{\mu_1}, p_T^{\mu_2}$): (6 GeV, 6 GeV), (6 GeV, 10 GeV) and (10 GeV, 10 GeV). A combination of these three triggers is the most likely option for the full HL-LHC $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ ATLAS analysis, but an estimation of such a combination would require further assumptions on the instantaneous luminosity profile delivered by HL-LHC. Instead, three scenarios are considered, each characterised by the usage of only one of these triggers. The resulting trigger efficiencies relative to the Run 1 trigger combination are 0.37, 0.28 and 0.07, respectively for the three sets of muon p_T thresholds.

² The Run 2 data taking period can be divided into two main parts, separated by the introduction of the topological trigger processor L1Topo (introduced in section 2.2.6). Three dimuon triggers are considered for the first main part (2015/16): the two triggers employed in the ATLAS 2015/16 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, described in section 7.1, and a dimuon trigger with higher muon p_T thresholds ($p_T > 6 \text{ GeV}$ on both muons), which was basically unprescaled. These three dimuon triggers were deactivated or heavily prescaled in 2017, in favour of their counterparts with additional $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ between the muons and dimuon invariant mass requirements at the L1 stage of the ATLAS trigger. Such dimuon triggers present a lower signal efficiency, due to the additional trigger selection, but also a much higher background rejection, which allows to maintain them active at higher PU conditions.

Depending on the di-muon trigger scenario chosen, three Working Points (WPs) are evaluated:

- Conservative: $\times 15$ Run 1 statistics (trigger thresholds (10 GeV, 10 GeV));
- Intermediate: $\times 60$ Run 1 statistics (trigger thresholds (6 GeV, 10 GeV));
- High-yield: $\times 75$ Run 1 statistics (trigger thresholds (6 GeV, 6 GeV)).

13.1.5 Results

The preceding sub-sections have provided a approach for the evaluation of the expected statistics (sub-section 13.1.1) and discussed all the ingredients needed to implement such a procedure (sub-sections 13.1.3 and 13.1.4 respectively for Run 2 and HL-LHC). The results of the projections are therefore discussed in this section.

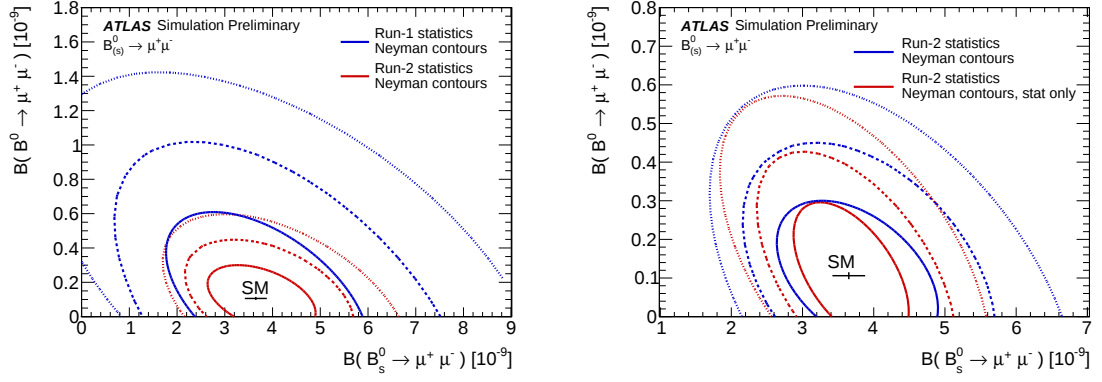
The Run 2 projections are evaluated first and compared to the confidence contours evaluated for the Run 1 statistics (sub-section 13.1.5.1). As introduced in sub-section 13.1.1, the Run 2 confidence regions are also compared to the ones obtained with likelihood contours. This comparison shows that the likelihood presents a gaussian behaviour close to its maximum, which becomes progressively less adequate with increasing $\Delta \log \mathcal{L}$, but is deemed consistent within sample fluctuations for the lowest $\Delta \log \mathcal{L}$ value.

This justifies the usage of the likelihood contours approach for the HL-LHC projections (sub-section 13.1.5.2). In this extrapolation three sets of contours are obtained, one for each WP defined in sub-section 13.1.4. On top of the two-dimensional contours, also the one-dimensional projected uncertainties on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ are evaluated for every extrapolation performed.

13.1.5.1 Run 2 projections

The confidence contours of the combined measurement of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ for the Run 2 projections are first extracted using the Neyman belt construction. Figure 13.3 shows the result of the projections for the Run 2 expected statistics. The 68.3%, 95.5% and 99.7% confidence level Neyman-based contours are compared with the contours for the Run 1 statistics in figure 13.3(a), both sets of contours are centred on the SM prediction of $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ and include systematic uncertainties. The effects of the systematic uncertainties on the projected Run 2 analysis are shown in figure 13.3(b), where the same Run 2 68.3%, 95.5% and 99.7% confidence level Neyman-based contours are compared with their counterpart built without considering the systematic uncertainties.

As anticipated in sub-section 13.1.1, the Neyman-based confidence contours can be compared with the likelihood contours in order to test the asymptotic behaviour of the likelihood. This comparison is shown in figure 13.4, where the 68.3%, 95.5% and 99.7% confidence level Neyman contours are drawn together with the likelihood contours obtained taking $-2\Delta \log \mathcal{L} = 2.3, 6.2, 11.8$, which correspond to confidence levels of 68.3%, 95.5% and 99.7% in gaussian approximation. Both sets of contours include statistical and systematic uncertainties. The expected variance on the likelihood contours is estimated to be at the level of $\sim 5\%$, therefore the



(a) Comparison of the stat.+syst. confidence regions for the Run 1 statistics (blue, outermost) and the extrapolated Run 2 statistics (red, innermost).

(b) Comparison of the confidence contours for the Run 2 case. Red contours are statistical only; blue contours include systematic uncertainties from the ATLAS Run1 analysis [73] extrapolated to Run 2 statistics.

Figure 13.3: Comparison of 68.3% (solid), 95.5% (dashed) and 99.7% (dotted) stat.+syst. confidence level contours for the Run 2 expected statistics with the same confidence regions evaluated for the Run 1 statistics (figure 13.3(a)) and for the Run 2 statistics without including systematic uncertainties (figure 13.3(b)). All confidence regions are obtained with the 2D Neyman belt construction based on toy-MC experiments and the Run 1 analysis likelihood. The Run 2 toy-MCs reproduce the expected signal mass resolution and have been scaled with respect to their Run 1 counterpart according to the triggers available in Run 2, the different integrated luminosity and the different B production cross section. The black point shows the SM theoretical prediction and its uncertainty [54].

consistency for the 68.3% confidence level contours is good, but it degrades with increasing $\Delta \log \mathcal{L}$ (with inconsistencies at the level of 15-20%). This effect is probably due to the progressively less adequate gaussian approximation of the likelihood maximum.

As introduced in sub-section 13.1.1, this comparison represents a crucial point for these studies. The 68% contour, most relevant in the understanding of the extrapolated ATLAS experimental capabilities is already well approximated by $-2\Delta \log \mathcal{L} = 2.3$ at the level of Run 2 statistics. As for the 95.5% and 99.7% contours, their approximation with likelihood contours is expected to improve in the extrapolations at higher integrated luminosities. In light of this, the HL-LHC projections are performed using constant $-2\Delta \log \mathcal{L}$ contours.

13.1.5.2 HL-LHC projections

For each of the three HL-LHC working points discussed in sub-section 13.1.4 the 68.3%, 95.5% and 99.7% confidence level contours are obtained taking $\Delta \log \mathcal{L} = 2.3, 6.2, 11.8$. Both statistical only and statistical + systematic uncertainties are considered. Figure 13.5 shows the resulting sets of contours. In particular figures 13.5(a), 13.5(b) and 13.5(c) show respectively the *Conservative*, *Intermediate* and *High-yield* WPs, corresponding to the dimuon trigger selections with muon p_T thresholds of (10 GeV, 10 GeV), (6 GeV, 10 GeV) and (6 GeV, 6 GeV).

Table 13.1 shows a comparison of the profiled likelihood ratio uncertainties separately for the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ measurements at the various data taking points discussed in the previous paragraphs. Such uncertainties are not obtained from the two-dimensional

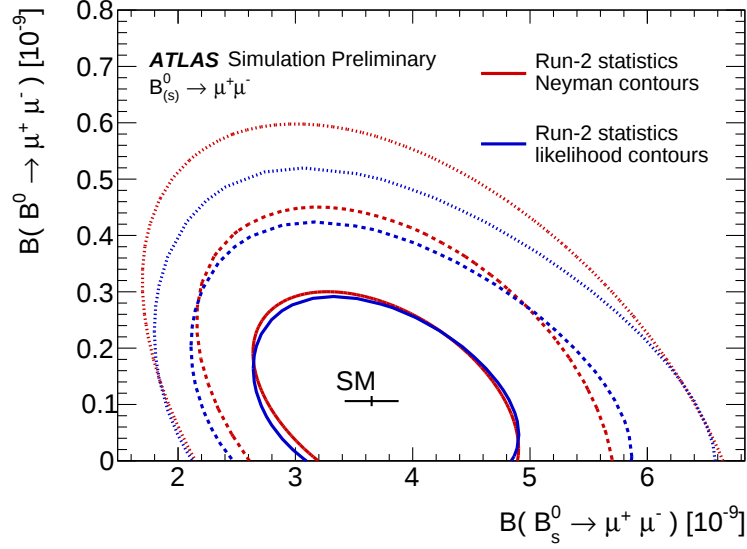


Figure 13.4: Comparison of the 68.3% (solid), 95.5% (dashed) and 99.7% (dotted) stat.+syst. confidence regions for the extrapolated Run 2 statistics. Red contours are obtained exploiting the 2D Neyman belt construction based on toy-MC experiments, while blue contours are drawn at constant $\Delta \log \mathcal{L}$ in the gaussian maximum approximation. The Run 2 toy-MCs reproduce the expected signal mass resolution and have been scaled with respect to their Run 1 counterpart according to the triggers available in Run 2, the different integrated luminosity and the different B production cross section. The black point shows the SM theoretical prediction and its uncertainty [54].

likelihood contours shown before, but rather employing the one dimensional likelihood intervals (section 6.3.2).

Table 13.1: Projected uncertainty on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ as reported by the fitting procedure applied to the toy simulations. The results are centred on the SM theoretical prediction [54]. For each extrapolation performed, statistical and statistical + systematic uncertainties are reported in units of 10^{-10} . The table reports a sufficient number of significant digits to highlight the difference between statistical+systematics and systematics-only uncertainties.

	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$		$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$	
	stat [10^{-10}]	stat + syst [10^{-10}]	stat [10^{-10}]	stat + syst [10^{-10}]
Run 2	7.0	8.3	1.42	1.43
HL-LHC: <i>Conservative</i>	3.2	5.5	0.53	0.54
HL-LHC: <i>Intermediate</i>	1.9	4.7	0.30	0.31
HL-LHC: <i>High-yield</i>	1.8	4.6	0.27	0.28

13.1.6 Comparison with other analyses

The results shown in the previous sub-section can be compared to the latest $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ experimental results, including the one discussed in this thesis, and with the projections performed by the other LHC experiments.

Without considering the ATLAS experiment, as the comparison with its most recent analysis (described in this thesis) will be performed later in this sub-section, the most recent exper-

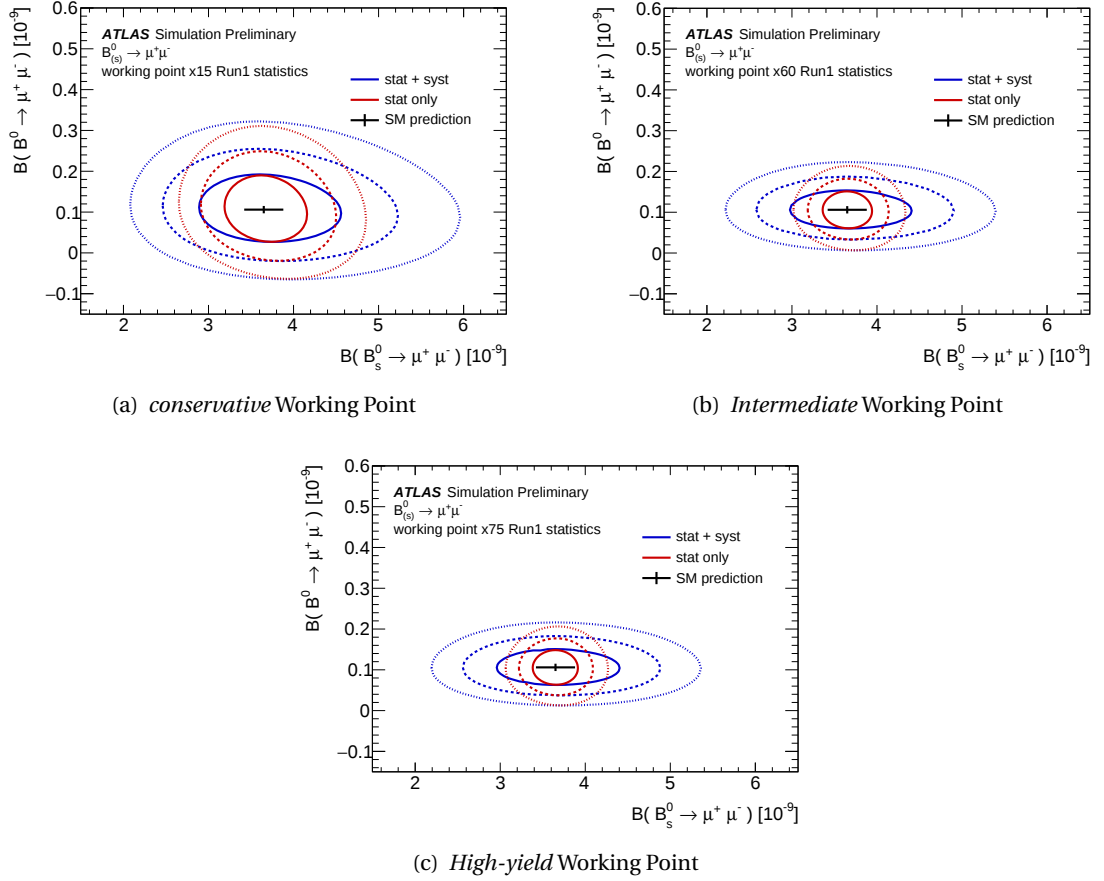


Figure 13.5: The three plots show the comparison of 68.3% (solid), 95.5% (dashed) and 99.7% (dotted) confidence level profiled likelihood ratio contours for the three Working Points (WPs) discussed in subsection 13.1.4. Red contours do not include the systematic uncertainties, which are then included in the blue ellipsoids. The black points shows the SM theoretical prediction and its uncertainty [54]. Figure 13.5(a) corresponds to the *conservative* WP, figure 13.5(b) to the *Intermediate* WP and figure 13.5(c) to the *High-yield* WP.

imental results come from the CMS and LHCb experiments. Such results are the combined Run 1 measurement of CMS and LHCb [70], which yields $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8^{+0.7}_{-0.6}) \times 10^{-9}$ ($\sim 23\%$ relative uncertainty) and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) = (3.9^{+1.6}_{-1.4}) \times 10^{-10}$ ($\sim 38\%$ relative uncertainty), and the latest LHCb result based on 4.4 fb^{-1} of integrated luminosity [209], which yields $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (3.0 \pm 0.6^{+0.3}_{-0.2}) \times 10^{-9}$ ($\sim 22\%$ relative uncertainty). The projections performed in this chapter lead to an estimated relative uncertainty of 23 % and 13 % on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and 135 % and 29 % on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, considering respectively the Run 2 and HL-LHC intermediate WP extrapolations.

The expected sensitivity of the ATLAS Run 2 analysis appears to be the same as the two analyses mentioned above. Similarly, the confidence intervals obtained with the 2015/16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis discussed in this thesis appear to be comparable to the Run 2 ATLAS projection.

Considering the confidence interval on $B_s^0 \rightarrow \mu^+ \mu^-$, there is in fact an apparent inconsistency between the projections and the result of the analysis presented in the previous chapters. The result of the analysis, combined with the one performed on the Run 1 dataset, yields a confid-

ence interval of $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8_{-0.7}^{+0.8}) \times 10^{-9}$, comparable (if not smaller) with the one projected considering the full Run 2 statistics. The increment in statistics due to the combination with the Run 1 analysis is not enough to motivate such improvement. In fact, the Run 2 dataset is estimated to present a 7-fold increase in statistics compared to Run 1, while the combined 2015/16 + Run 1 dataset can be estimated to have only about three times the Run 1 statistics (two-fold increase for the 2015/16 dataset, section 5.1, and the actual Run 1 dataset). This apparent inconsistency is explained with the conservative assumptions made in the procedure followed for the projections (sub-section 13.1.1). Only two main effects are considered in the projections, the increase in available statistics and the changes in dimuon mass resolution, while additional effects are neglected. Such effects include the improvements in the decay length resolution, due to the addition of the IBL in the ATLAS ID, which allow a better discrimination of the background. Additionally, given the higher trigger thresholds employed in Run 2 and that will be employed at HL-LHC with respect to Run 1 imply, several discriminating variables are expected to improve their discriminating power. Such effects can already be noticed comparing the performance of the Run 1 and 2015/16 analyses, as done in table 12.2 in section 12.1.1. The performance of the analyses is quantified with the value of the figure of merit $S/\sqrt{S+B}$ in the three higher BDT bins, where S and B are the expected signal yield and the interpolated background yield in the signal region respectively. The performance of the 2015/16 analysis is significantly higher than the Run 1's, while, according to the assumptions made for the projections (same S/B ratio), the value of the figure of merit should have been only a factor $\times \sqrt{2}$ higher. In conclusion, the apparent inconsistency between the projections and the result of the combined 2015/16 and Run 1 analyses is an effect of the conservative assumptions made in sub-section 13.1.1.

HL-LHC projections have been performed also by CMS [210] and LHCb [211], extrapolating the performance of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis.

The CMS projections are based on an extrapolation of the Run 1 analysis performance to the HL-LHC scenario. The trigger efficiency is expected to be comparable to the one in Run 2 [212], and the signal selection efficiency is not expected to be reduced by the higher pile-up conditions. Phase-II upgrades of the CMS detector are also foreseen [212]: the CMS HL-LHC inner tracker will allow a 40-50% improvement relative to Run 2 on the di-muon mass resolution, resulting in an improved separation of the B_s^0 and B_d^0 signals.

The LHCb experiment is already highly optimised for the detection of $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays; the foreseen LHCb Phase-II upgrade will insure improvements to the tracking system and to the muon detector shielding, allowing it to maintain a high muon reconstruction purity and efficiency.

The treatment of the systematic uncertainties is slightly different in the projections of the three experiments. The ATLAS projections assume conservatively that the external systematics will be at the same level as the ones in the Run 1 analysis [73]. The CMS and LHCb projections assume a reduction of such systematics, based on assumptions about additional Belle II inputs

and improvements in branching fraction measurements.

Table 13.2, obtained from [3], summarises the projected statistical and systematic uncertainties

Table 13.2: Projected ATLAS, CMS and LHCb uncertainty on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$. The HL-LHC scenario corresponds to an integrated luminosity of 300 fb⁻¹ for LHCb and 3 ab⁻¹ for ATLAS and CMS. For each extrapolation the total (statistical+systematic) uncertainties are reported. The SM prediction for the two branching fractions [54] is also shown for comparison. Table obtained from [3]

Experiment	Scenario	$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$	$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$
		stat + syst %	stat + syst %
LHCb	23 fb ⁻¹ , Run 3	8.2	33
LHCb	300 fb ⁻¹ , HL-LHC	4.4	9.4
CMS	300 fb ⁻¹ , Run 3	12	46
CMS	3 ab ⁻¹ , HL-LHC	7	16
ATLAS	130 fb ⁻¹ , Run 2	22.7	135
ATLAS	3 ab ⁻¹ , Conservative HL-LHC	15.1	51
ATLAS	3 ab ⁻¹ , Intermediate HL-LHC	12.9	29
ATLAS	3 ab ⁻¹ , High-Yield HL-LHC	12.6	26
SM		6.4	8.5

for the three experiments, together with the theoretical uncertainty on the SM predictions [54] for comparison. Similarly figure 13.6 [3] shows the expected 1 sigma contours for the HL-LHC ATLAS intermediate WP, CMS and LHCb, together with the branching fraction predictions from a particular class of BSM models [213].

The expected performance of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis seems to be much lower compared to the others. This effect is due to three main reasons. The trigger thresholds that ATLAS is planning to employ at HL-LHC are higher than the ones planned by the other two experiments. This largely affects the signal statistics that the experiment is going to collect. Additionally the ATLAS dimuon mass resolution is lower compared to CMS (as shown later in section 13.2) and LHCb; this affects negatively the B_s^0 - B_d^0 separation and the background discriminating power of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. Finally, as already pointed out in the comparison with the 2015/16 ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, the assumptions made in the extrapolation procedure (sub-section 13.1.1) are extremely conservative. This already affects the ATLAS Run 2 projections, which show comparable expected confidence intervals to the ones found for the combined Run 1 + 2015/16 result. Possible further improvements can be foreseen at HL-LHC, thanks to improvements in the discriminating variables due to the ITk upgrade.

Table 13.2 and figure 13.6 show that the projected HL-LHC experimental sensitivity is close to the current theoretical SM uncertainty. The small uncertainty on the measurements will allow one not only to discriminate between SM and BSM models, but also to discriminate within the parameter space of those BSM models, in case of inconsistency with the SM.

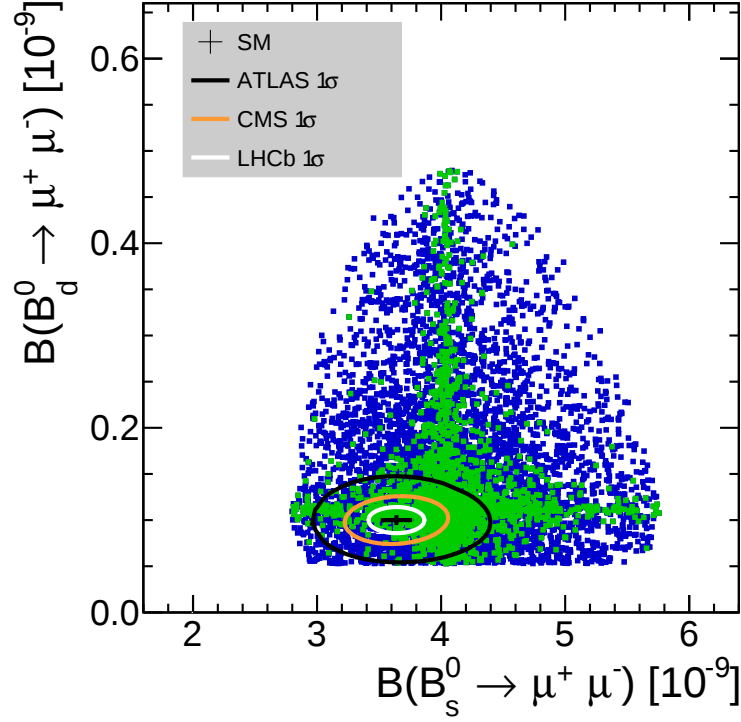


Figure 13.6: $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ branching ratios as computed using new sources of Flavour-Changing-Neutral-Current (FCNC), as discussed in [213]. The green points are the subset consistent with other measurements. The black cross point is the SM prediction, while the coloured contours show the expected 1-sigma HL-LHC sensitivities of ATLAS (Intermediate WP), CMS, and LHCb.

13.2 ATLAS-CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses comparison

The overview of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ experimental state of the art (section 1.5) and the comparison of the projections for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses performed by the LHC experiments (section 13.1.6) show that the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses carried out by the LHC experiments present very different performances.

The LHCb experiment is highly optimised for the detection and analysis of heavy flavour physics events, therefore it is expected to have a higher performance compared to e. g. ATLAS and CMS. However, despite ATLAS and CMS being similar general-purpose detectors, the CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performance seems to be much higher compared to the ATLAS one.

This section is dedicated to the studies performed to understand the main differences between the ATLAS and CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses sensitivities, aiming at potential improvements to the ATLAS analysis strategy. The main idea guiding these studies is to emulate the result of the CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [71] exploiting the same tools used in the Run 2 and HL-LHC projections (section 13.1). The ATLAS toy-MCs are modified according to the characteristics of the CMS analysis and the resulting performance is gauged with the two-dimensional confidence regions in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane, obtained with the Neyman belt construction. The confidence regions highlight the prominent role of two key aspects of the CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis (dimuon mass resolution and background discrimination), as they

are found to be reasonably similar to the ones published in CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis. The procedure followed to perform the analysis performance comparison (sub-section 13.2.1) is followed by a discussion of the results obtained (sub-section 13.2.2). The conclusions of these studies are then discussed in sub-section 13.2.3.

13.2.1 Comparison procedure

The procedure followed to compare the ATLAS and CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses performance is based on the tools exploited for the Run 2 and HL-LHC projections of the ATLAS analysis (section 13.1). Two aspects believed to be the possible source of the difference in performance are implemented as modifications to the ATLAS toy-simulations, in order to reproduce the CMS analysis result. These two aspects are the different dimuon mass resolution and background discrimination of the CMS analysis with respect to the ATLAS counterpart. The resulting performance of the analyses is gauged with the two-dimensional confidence regions in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane. The Neyman belt construction technique (sections 6.3.3 and 6.3.4) is employed for the evaluation of the confidence regions. Other effects that could contribute to the difference in performance between the two experiments, such as the different design of the fit for the extraction of the signal yield, are assumed to be of minor importance. The CMS fit in fact employs a simultaneous fit on 12 datasets and uses the per-event mass resolution as a conditional variable, but these features are not considered the main source of difference between the two analyses. The assumption will be proved correct later in sub-section 13.2.2.

The CMS dimuon mass resolution for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ peaks is obtained from Ref. [71], where the signal peaks are modelled with Crystal Ball functions [214]. The two CMS Crystal Ball functions are approximated with the same model employed by ATLAS (double gaussians with the same mean), with a suitable modification of the model RMS. Table 13.3 summarises the para-

Table 13.3: Comparison of the parameters employed to describe the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ invariant mass peaks in the ATLAS and CMS-like toy-MCs. The four parameters fully describe the double gaussian with the same mean models employed. The parameters shown for ATLAS are the values actually employed in the Run 1 analysis [73]. The CMS analysis exploits Crystal Ball functions to described the signal peaks [214]. In order to be able to compare the dimuon mass resolution of the two experiments, the Crystal Ball functions are approximated with the same models employed for ATLAS.

	ATLAS PDFs		CMS PDFs	
	$B_s^0 \rightarrow \mu^+ \mu^-$	$B_d^0 \rightarrow \mu^+ \mu^-$	$B_s^0 \rightarrow \mu^+ \mu^-$	$B_d^0 \rightarrow \mu^+ \mu^-$
mean [MeV]	5366	5279	5366	5279
σ_1 [MeV]	67.7	64.9	42	42
σ_2 [MeV]	142.9	135.5	75	75
fraction	0.86	0.82	0.71	0.71

eters employed in the various PDFs. The resulting width of the models employed to describe the CMS signal peaks are much smaller compared to the ATLAS ones. In order to consider this

effect, the widths of the gaussian PDFs employed in the toy-MCs is modified according to the values listed in table 13.3.

The CMS signal and background yields are taken from Ref. [215] and used in the toy-MC. The total number of signal events observed by CMS is assumed to be respectively 30 $B_s^0 \rightarrow \mu^+ \mu^-$ and 18 $B_d^0 \rightarrow \mu^+ \mu^-$, compared to an expected yield of the ATLAS Run 1 analysis of 41 $B_s^0 \rightarrow \mu^+ \mu^-$ and 5 $B_d^0 \rightarrow \mu^+ \mu^-$ events. The $\mathcal{D}_{\text{norm}}$ term of equation A.1, which encloses the $B^+ \rightarrow J/\psi K^+$ yield, the ratio of the efficiency and acceptance for signals and reference channel and the effect of the trigger prescale factors and integrated luminosities of the signal and normalisation channels, is modified, so that the observed number of signal events of the CMS experiment matches the observed branching fractions. The total number of background events generated in the toy-MCs is modified as well, without modifying either the relative normalisation of the background PDFs or the structure of the ATLAS Run 1 signal yield extraction fit (described in section A.4 of appendix A) implemented in the toy-simulations.

In order to factor possible dependencies on the observed branching fraction, the comparison between the resulting ATLAS and CMS-like confidence regions in the 2D $\mathcal{B}$ plane is performed assuming in both cases the branching ratio values observed by the CMS experiment.

As shown later in section 13.2.2, the resulting confidence regions obtained with the CMS-like toys resemble the actual CMS Run 1 result; such confidence region is much smaller compared to the one obtained with the ATLAS toys. Two additional confidence regions are therefore constructed, in order to verify if one of the two modified features of the CMS-like toys is mostly responsible for the different performance of the ATLAS and CMS analyses. Each region is obtained modifying *either* the dimuon mass resolution *or* the background discrimination. This will finally show that the main feature responsible for the higher CMS performance is its better dimuon mass resolution, which allows a better $B_s^0 - B_d^0$ separation.

All confidence regions discussed in this section are obtained employing the Neyman construction approach (sections 6.3.3 and 6.3.4), but since these studies were conducted while the Neyman construction procedure was under development, the resulting contours differ from the ones shown in sections 12.1.3 (2015/16 analysis result) and 13.1 ($B_{(s)}^0 \rightarrow \mu^+ \mu^-$ Run 2 and HL-LHC projections) because of two reasons:

- the smoothing procedures for the acceptance and confidence regions defined in section 6.3.4.2 are not included; for this reason the resulting confidence regions present boundaries that are not-well defined. As shown later in sub-section 13.2.2, this does not constitute an issue for the 68.3 % coverage regions, which present a defined contour, while the 95.5 % coverage region shows large fluctuations; the 99.7 % region is not shown as it presents even bigger fluctuations;
- the procedure for the inclusion of the systematic uncertainties was under development, therefore the systematic uncertainties are not considered in these studies. Since both the ATLAS and CMS Run 1 analyses were largely statistically limited, such an effect is not critically relevant.

13.2.2 Results

The procedure defined in sub-section 13.2.1 is employed to construct the confidence regions based on the ATLAS toys and on the CMS-like toys. Both employ the CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis result [71] as the experimentally observed value. Figure 13.7 shows the resulting confidence regions, superimposed to the confidence regions obtained in the CMS Run 1 analysis.

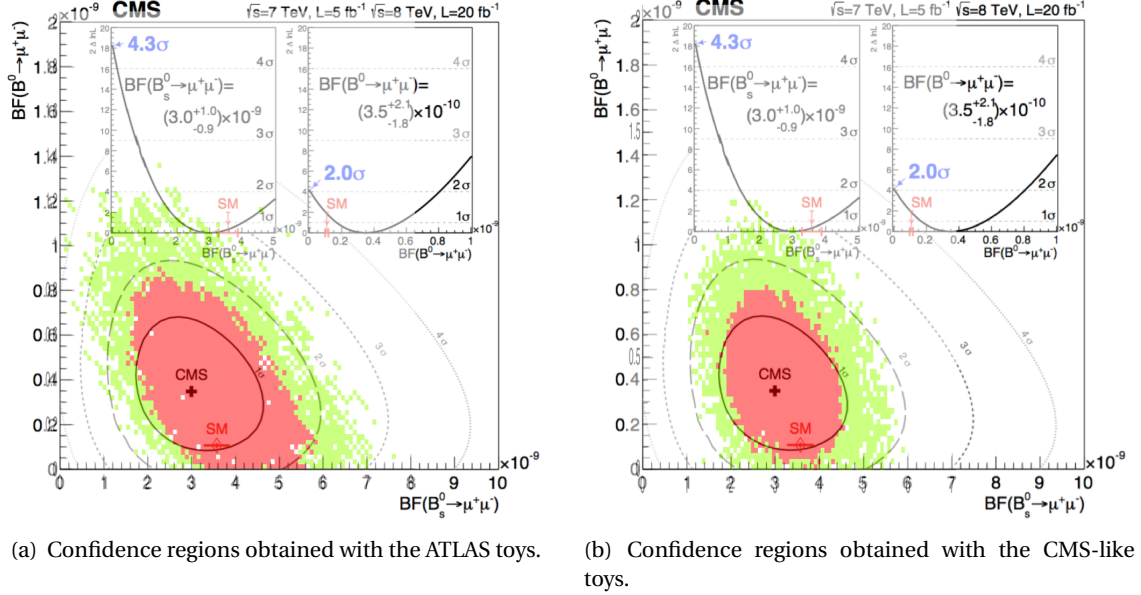


Figure 13.7: (a): 68.3 % and 95.5 % coverage confidence regions based on the toy-MCs generated with the ATLAS Run 1 analysis likelihood (ATLAS toys). (b): 68.3 % and 95.5 % coverage confidence regions based on the toy-MCs generated with the same likelihood modifying the dimuon mass resolution and the background discrimination, according to the CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [71] (CMS-like toys). Both sets of confidence regions are superimposed to the two-dimensional likelihood contours of the CMS analysis and employ its result as experimentally observed value.

The confidence regions obtained with the ATLAS toys (figure 13.7(a)) show a different behaviour compared to the result of the CMS analysis. Basically, in the direction of the bisecting line of the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ – $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ plane the confidence regions show a similar width, while in the perpendicular direction the ATLAS region is much larger. This behaviour does not appear for the confidence regions obtained with the CMS-like toys (figure 13.7(b)), which show a similar shape as the contours resulting from the CMS analysis. As one can expect the confidence regions do not show an exact overlap with the CMS results, as the procedure employed to mimic the CMS analysis is only approximated.

The two plots in figure 13.8 clearly show that the two features considered in the CMS-like toys are responsible for most of the better CMS performance with respect to ATLAS. In order to understand if one of the two is dominant over the other, two additional sets of confidence regions are built. One is based on the ATLAS toys, with only the dimuon mass resolution modified according to the one found in the CMS analysis, while the second is built considering only the different background discrimination. The resulting contours, superimposed to the confidence regions obtained in the CMS Run 1 analysis, are shown in figure 13.8. The behaviour of the confidence regions resembles the one shown in figure 13.7: the regions based on the toy-MCs

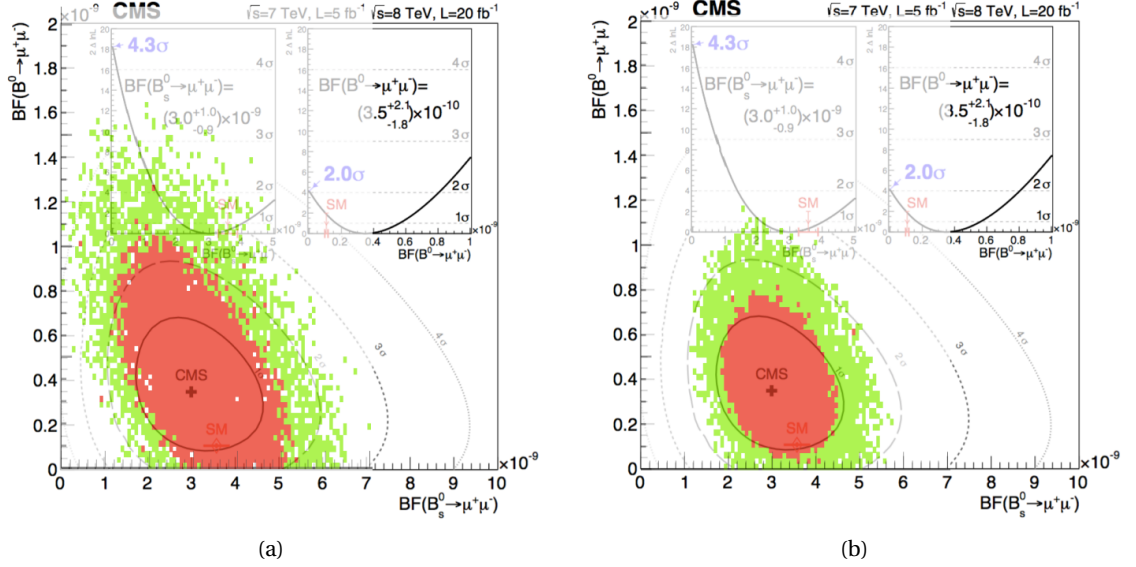


Figure 13.8: 68.3 % and 95.5 % coverage confidence regions based on the toy-MCs generated with the ATLAS Run 1 analysis likelihood modifying the background discrimination (figure (a)) or the dimuon mass resolution (figure (b)) according to the CMS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [71]. Both sets of confidence regions are superimposed to the two-dimensional likelihood contours of the CMS analysis and employ its result as experimentally observed value.

modified according to the CMS background discrimination present the same shape as the one obtained with the ATLAS toys. On the other hand, the regions based on the toy-MCs that only take into account the different dimuon mass resolution show a behaviour extremely similar to the CMS result and to the confidence regions shown in figure 13.7(b).

The main contributor to the difference in performance between the ATLAS and CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses is therefore the higher dimuon mass resolution of the CMS experiment.

13.2.3 Conclusions

A comparison of the performance of the ATLAS and CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analyses based on the dataset collected during the LHC Run 1 data taking period has been performed. Since the CMS analysis shows a higher performance, the reasons for such difference in the two analysis are investigated.

The possible sources of discrepancy in the performance are narrowed down to two possibilities, the higher dimuon mass resolution and the different background discrimination of the CMS analysis. Employing the same technique used for the projections of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performance at Run 2 and HL-LHC (section 13.1), the main source of discrepancy is found to be the dimuon mass resolution.

Thanks to its high magnetic field the CMS experiment has a dimuon mass resolution for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ peaks which is better compared to the one of the ATLAS detector, as shown in table 13.3. The corresponding ATLAS analyses are wider, resulting in a large correlation of the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ branching fraction and an effective increased background contamination. This effect is visible in all the ATLAS confidence regions shown in this thesis, e.g. figure A.4

in appendix A, figures 13.3 and 13.7(a) in this chapter and the results of the analysis on the 2015/16 dataset in chapter 12, which show elliptic shapes. The branching ratio correlation is less present in the CMS contours, shown in all the plots of this section.

Since the dimuon mass resolution is strictly related to the properties of the detector, improvements in the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis can not overcome this effect, while possible upgrades of the ATLAS detector can improve the dimuon mass resolution. In fact, as discussed in section 13.1, the ITk upgrade will allow an improvement in the dimuon mass resolution. Anyway, a more refined analysis techniques could be used to advance the analysis performance. A possibility being considered is the modification of the signal extraction fit of the analysis, e. g. differentiating events according to their dimuon mass resolution or employing per-candidate mass resolution as a conditional variable in the fit. This would result in a fit where the bins with low background and low uncertainty on the dimuon mass contribute to reducing the correlation between the two signals.

Due to time constraints, studies on how to improve the fitting procedure of the 2015/16 analysis were not pursued, although studies are ongoing exploiting the full Run 2 ATLAS dataset.

CONCLUSIONS

The core of this thesis is the first ATLAS Run 2 search for the rare decays of B_s^0 and B_d^0 mesons into oppositely charged muons in pp collisions at $\sqrt{s} = 13$ TeV delivered by the LHC. The dataset was collected during 2015 and 2016 and corresponds to a total integrated luminosity of 36.2 fb^{-1} , for an effective integrated luminosity of 26.3 fb^{-1} .

The branching fractions of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ decays are evaluated relative to the reference channel $B^+ \rightarrow J/\psi K^+$, allowing a reduction of the systematic uncertainties. The efficiency of the selection on signal and reference channels are evaluated on simulations corrected with data-driven techniques, while the reference channel yield is evaluated with a simultaneous fit on data and MC. Several background sources can affect the signal topology. In particular, events characterised by uncorrelated muons from the $b\bar{b}$ decay chain constitute the largest background. A MVA selection, based on a BDT algorithm, is optimised to separate $B_{(s)}^0$ candidates from this background from the signal. Events with $B \rightarrow hh'$ decays where both hadrons are mis-identified as muons constitute an irreducible background, as they present the same topology as the $B_d^0 \rightarrow \mu^+ \mu^-$ signal. Standard ATLAS muon-id requirements are employed to reduce the mis-identification probability of hadrons to muons of a factor 0.39, while still maintaining a good efficiency on muons.

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields are extracted simultaneously with an unbinned extended maximum likelihood fit on four independent data categories, designed using the BDT response to have the same signal efficiency.

The Neyman belt exact frequentist approach is employed to extract the final results, obtaining a 68.3% confidence interval on the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction of $(3.2_{-1.0}^{+1.1}) \times 10^{-9}$ and a 95% CL upper limit of 4.3×10^{-10} on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$.

The likelihood of this analysis is combined with the likelihood of the previous ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the full Run 1 dataset. The Neyman belt approach provides $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (2.8_{-0.7}^{+0.8}) \times 10^{-9}$, which is the first ATLAS evidence of the $B_s^0 \rightarrow \mu^+ \mu^-$ process with a significance of 4.7σ , and $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 2.1 \times 10^{-10}$ at 95.5% CL, which is, as of today, the most stringent upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$. All the results presented are compatible with the SM prediction and with the current available experimental results.

Figure 13.9 shows the history of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fraction measurements, already shown in chapter 1, but modified to contain also the result of the analysis presented in this

thesis.

Since no major deviations from the SM prediction are found, NP effects on $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$, if they exist, are small. The ATLAS analysis, as well as the analyses performed by other experiments, is statistically limited, therefore the new data provided by the LHC collisions will allow an improvement in the experimental precision on both decays. Therefore, a study of the sensitivity of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis based on the statistics collected in Run 2 and to be collected at HL-LHC is also presented. Projections on the analysis sensitivity are performed and compared with the ones from the other LHC experiments.

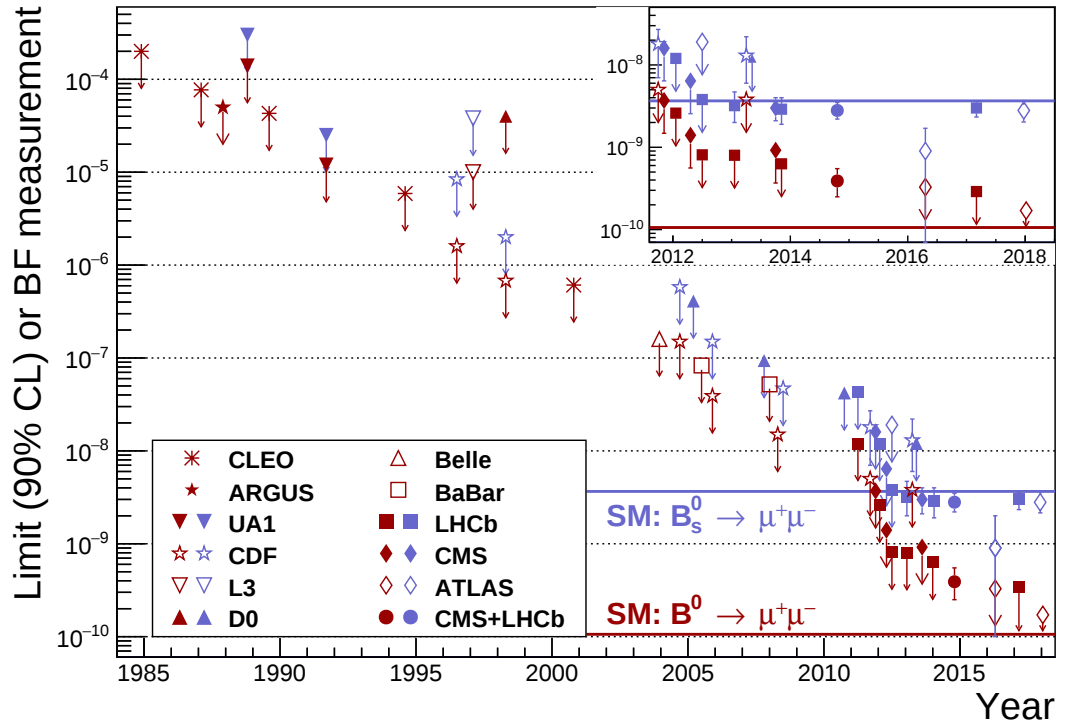


Figure 13.9: History of the limits and measurements of the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions across the years [69] modified adding the result of the ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during 2015 and 2016 and combined with the result of the Run 1 analysis.

THE ATLAS RUN 1 ANALYSIS



The ATLAS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the full Run 1 dataset is particularly relevant for the work developed in this thesis, in fact, a large part of the 2015/16 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis strategy is derived from it. In addition, the author had a role in the evaluation on the systematic uncertainties arising from the fitting procedure.

A.1 Analysis overview

The ATLAS Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis [73] employs a dataset made of $\sqrt{s} = 7$ TeV pp and $\sqrt{s} = 8$ TeV pp collision data taken with stable LHC beams in 2011 and 2012 by the ATLAS experiment, for a total integrated luminosity of about 25 fb^{-1} .

The dimuon invariant mass region between 5166 and 5526 MeV, where the signal peaks fall, is kept blinded until the procedures for event selection and the details of signal yield extraction are completely defined.

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ branching fractions are extracted relative to a reference decay mode $B^+ \rightarrow J/\psi(\rightarrow \mu^+ \mu^-)K^+$. The branching fraction calculation is performed according to the following formula:

$$\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-) = \frac{f_u}{f_{s(d)}} \times \mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-) \times N_{B_d^0(B_s^0)} \times \frac{1}{\mathcal{D}_{\text{norm}}}, \quad (\text{A.1})$$

where $f_u/f_{s(d)}$ is the hadronisation probability of a b -quark into B^+ and $B_s^0(B_d^0)$, $\mathcal{B}(B^+ \rightarrow J/\psi K^+) \times \mathcal{B}(J/\psi \rightarrow \mu^+ \mu^-)$ is the branching fraction of the reference channel $\mathcal{B}(B^+ \rightarrow J/\psi(\mu^+ \mu^-)K^+)$ and $N_{B_d^0(B_s^0)}$ is the number of signal events. The first two quantities are not measured in the analysis, but are obtained from recent measurements. The normalisation term $\mathcal{D}_{\text{norm}}$ is defined as:

$$\mathcal{D}_{\text{norm}} = \sum_k N_{J/\psi K^+}^k \alpha_k \left(\frac{\epsilon_{\mu^+ \mu^-}}{\epsilon_{J/\psi K^+}} \right)_k, \quad (\text{A.2})$$

where the k index runs over four mutually exclusive data samples, defined with different trigger selections and data-taking periods. The $N_{J/\psi K^+}^k$ term refers to the reference channel yield, $(\epsilon_{\mu^+ \mu^-}/\epsilon_{J/\psi K^+})_k$ is the ratio of the efficiency and acceptance for signals and reference channel and α_k takes into account the different trigger prescale factors and integrated luminosities in the signal and normalisation channels.

The various triggers employed require two muons and a full track reconstruction of the candidates, together with a loose cut on the dimuon invariant mass $4 < m_{\mu\mu} < 8.5$ GeV, while they differ on the transverse momentum and pseudorapidity cuts on the muons. The three trigger categories for the 2012 dataset are characterised by different prescale factors, therefore their combination is used to maximise the available statistics. They are:

- T_1 : “higher threshold” trigger, with dimuon p_T threshold of 6 and 4 GeV;
- T_2 : “barrel” trigger, with $p_T > 4$ GeV for both muons and at least one of them with $|\eta_\mu| < 1.05$. In addition, the T_1 category can not be satisfied;
- T_3 : “basic” trigger, with $p_T > 4$ GeV for both muons. The categories T_1 and T_2 can not be satisfied.

A fourth category is defined for events from the 2011 dataset. The trigger employed is characterised by muon p_T thresholds of 4 GeV. This category does not present any prescale. The triggers employed to collect the reference channel dataset have the same requirements, but a dimuon invariant mass cut of $2.5 < m_{\mu\mu} < 4.3$ GeV, tuned on the J/ψ mass.

Several simulated samples are employed in the analysis. These MCs present the same properties as the ones listed in table 7.1, but a lower amount of events; a similar reweighting procedure as the one presented in section 7.6 is applied to all the MC samples.

A.2 Background reduction

After the offline reconstruction a preliminary selection is applied to the candidates. This selection uses the same cuts as the ones shown in table 7.2, except for the muon p_T thresholds, which are both set to 4 GeV.

The main backgrounds that affect the analysis can be divided into three categories.

- *Continuum background*, which is the dominant background, is due to pairs of uncorrelated muons. This background is highly reduced with the application of a MVA based selection; specifically, a BDT, named *c-BDT* is employed. It is based on 15 variables, the same as the ones listed in table 8.1, and it is trained and tested on the $b\bar{b} \rightarrow \mu^+ \mu^- X$ and $B_s^0 \rightarrow \mu^+ \mu^-$ MCs.
- *Partially reconstructed background*, composed of several topologies. Same-side background, from decay cascades like $b \rightarrow c\mu^-\nu \rightarrow s(d)\mu^-\mu^+\nu\bar{\nu}$, same-vertex background, from B decays containing a muon pair and additional particles, B_c background, mostly due to $B_c \rightarrow J/\psi(\mu\mu)\mu\nu$ events and semileptonic background, due to semileptonic b-hadron decays where the final-state hadron is mis-identified as muon. Figure A.1(a) shows the dimuon mass distribution of these backgrounds obtained from simulations.

- *Peaking Background*, due to $B \rightarrow hh'$ decays with both hadrons mis-identified as muons. It populates the signal region below the B_d^0 peak. Figure A.1(b) shows the invariant mass distribution for this background obtained from simulation.

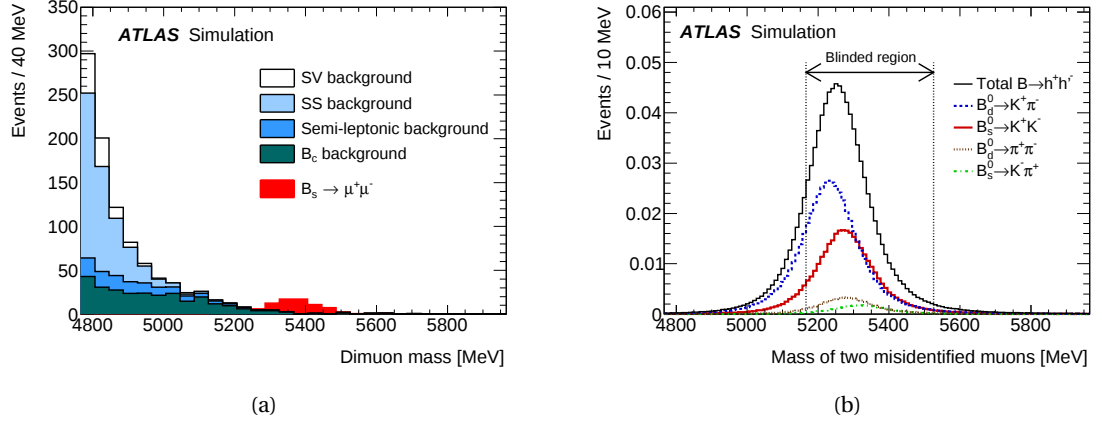


Figure A.1: A.1(a) Dimuon mass distribution for the partially reconstructed background, from simulation, before the final selection against continuum is applied but after all other requirements. The different components are shown as stacked histograms, normalised according to world-averaged measured branching fractions. The SM expectation for the $B_s^0 \rightarrow \mu^+ \mu^-$ signal is also shown for comparison (non-stacked). Continuum background is not included here. A.1(b) Invariant mass distribution of the peaking background components $B \rightarrow hh'$, after the complete signal selection is applied. In both plots the distributions are normalised to the expected yield for the integrated luminosity of 25fb^{-1} . Figures taken from [73].

In order to reduce the background contribution from hadrons mis-identified as muons, an additional MVA selection is employed. Such selection uses a BDT, named *fake-BDT*, based on eight input variables, listed in table A.1. The usage of the fake-BDT allows a reduction of the

Table A.1: Description of the eight variables used in the discrimination between signal muons and those from hadron decays in flight and punch-through. Table taken from [73].

- | | |
|----|--|
| 1. | Absolute value of the track rapidity measured in the ID. |
| 2. | Ratio q/p (charge over momentum) measured in the MS. |
| 3. | Scattering curvature significance: maximum variation of the track curvature between adjacent layers of the ID. |
| 4. | χ^2 of the track reconstruction in the MS. |
| 5. | Number of hits used to reconstruct the track in the MS. |
| 6. | Ratio of the values of q/p measured in the ID and in the MS, corrected for the average energy loss in the calorimeter. |
| 7. | χ^2 of the match between the tracks reconstructed in the ID and MS. |
| 8. | Energy deposited in the calorimeters along the muon trajectory obtained by combining ID and MS tracks. |

muon mis-identification probability by a factor 0.37, leading to a number of expected peaking background events in the signal region of 1.0 ± 0.4 .

A.3 Reference channel

The reference channel yield and the efficiency ratio terms in equation A.1 are evaluated separately in each trigger category.

A.3.1 Reference channel yield

The reference channel yield is extracted with an unbinned extended maximum likelihood fit on the $J/\psi K^+$ mass distribution, similar to the one described in section 10.1. All the yields are extracted from the fit to data, while the shape parameters are determined from a simultaneous fit to data and MC samples. The fit models and the systematic uncertainties considered have already been described in sections 10.1.2 and 10.1.3.

Figure A.2 shows the fit performed on the T_1 category, while table A.2 summarises the result of the fits on the four categories.

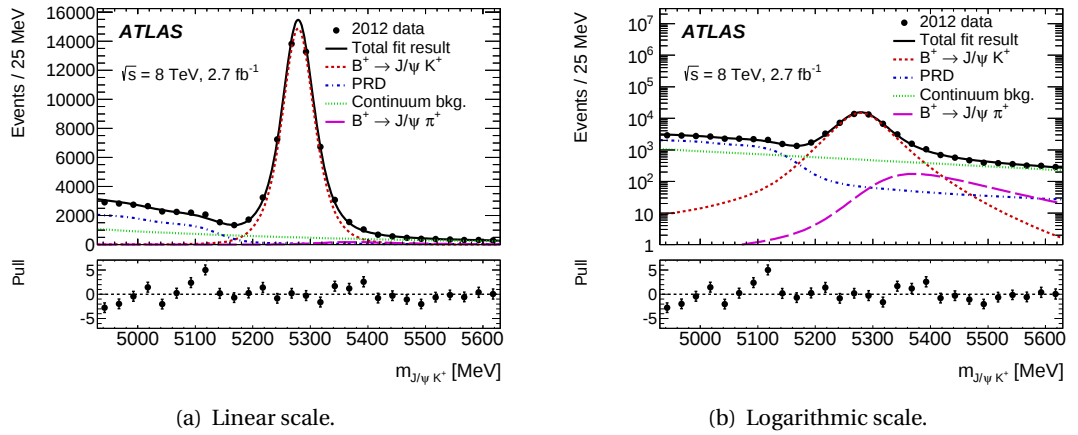


Figure A.2: $J/\psi K^+$ invariant mass distribution for all B^+ candidates in the T_1 trigger category in 2012 data in linear A.2(a) and logarithmic A.2(b) scale. The result of the fit is overlaid. The insets at the bottom of the plots show the bin-by-bin pulls for the fits, where the pull is defined as the difference between the data point and the value obtained from the fit function, divided by the error from the fit. Figure obtained from [73].

Table A.2: Results of the fits to the events reconstructed as $B^+ \rightarrow J/\psi K^+$ in each trigger and data category. Uncertainties are statistical and systematic, respectively. Table obtained from [73].

Category	$N_{J/\psi K^+}$	$N_{J/\psi \pi^+}$
T_1	$46860 \pm 290 \pm 280$	$1420 \pm 230 \pm 440$
T_2	$5200 \pm 84 \pm 100$	$180 \pm 51 \pm 89$
T_3	$2512 \pm 91 \pm 42$	$85 \pm 77 \pm 30$
2011	$95900 \pm 420 \pm 1100$	$3000 \pm 340 \pm 1140$

$B^+ \rightarrow J/\psi\pi^+ / B^+ \rightarrow J/\psi K^+$ ratio measurement

In addition to the $B^+ \rightarrow J/\psi K^+$ yield, the fits on the $J/\psi K^+$ mass distribution are also exploited to extract the branching fraction ratio between $B^+ \rightarrow J/\psi\pi^+$ and $B^+ \rightarrow J/\psi K^+$. The measurement is performed separately in the four trigger categories and combined into an error-weighted mean $\rho_{\pi/K}$.

Systematic uncertainties are considered in the evaluation of $\rho_{\pi/K}$. Several systematic sources cancel in the measurement of this ratio, while residual systematic uncertainties arise from the K^-/K^+ , π^-/π^+ and K^+/π^+ relative efficiencies.

The final result on the $\rho_{\pi/K}$ term is:

$$\rho_{\pi/K} = \frac{\mathcal{B}(B^+ \rightarrow J/\psi\pi^+)}{\mathcal{B}(B^+ \rightarrow J/\psi K^+)} = 0.035 \pm 0.003 \pm 0.012, \quad (\text{A.3})$$

where the first error is statistical and the second is systematic. The result is in good agreement with the current world average [11] of $(3.84 \pm 0.16) \%$.

A.3.2 Evaluation of the $B^+ \rightarrow J/\psi K^+$ to $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ efficiency ratio

The evaluation of the efficiency and acceptance ratios that appear in equation A.1 follows the same steps of the one described in section 10.2.

Table A.3 shows the values of the efficiency and acceptance ratios for the four trigger categories, together with their statistical and systematic uncertainties. The same efficiency ratio can be used for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$, as they are proven to agree within the MC statistical uncertainty of $\pm 0.5\%$.

A correction for the $B_s^0 \rightarrow \mu^+ \mu^-$ efficiency ratio is needed due to the width difference $\Delta\Gamma_s$ between the two B_s^0 mass eigenstates. This is the same correction described in section 10.2.1 and its effect is found to modify the $B_s^0 \rightarrow \mu^+ \mu^-$ efficiency by $+4\%$.

Table A.3: Values of the efficiency ratios $\epsilon_{J/\psi K^+}/\epsilon_{\mu^+ \mu^-}$ for the 2012 trigger categories and the 2011 sample, and their relative contributions to D_{norm} (Eq. A.1). The first uncertainty is statistical and the second systematic. The systematic component includes the uncertainties from the MC reweighting and from data-MC discrepancies. The correction due to the B_s^0 effective lifetime value is not applied to the numbers shown. Table obtained from [73].

Category	$\epsilon_{J/\psi K^+}/\epsilon_{\mu^+ \mu^-}$	Relative contribution to D_{norm}
T_1	$0.180 \pm 0.001 \pm 0.009$	68.3%
T_2	$0.226 \pm 0.004 \pm 0.014$	6.0%
T_3	$0.189 \pm 0.005 \pm 0.022$	3.5%
2011	$0.156 \pm 0.002 \pm 0.009$	22.2%

A.4 Signal yield extraction

The signal yield is extracted with a simultaneous extended unbinned maximum likelihood fit on the mass distribution of three datasets defined taking three intervals in the c-BDT output: 0.240-0.346, 0.346-0.446 and 0.446-1. Each interval is designed to have a signal efficiency of 18% and in each bin the four trigger categories are merged.

The following models and constraints are used in the fit, where the shape parameters and normalisations are, unless explicitly said, extracted from data.

- $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ PDFs: a superposition of two gaussians is used to describe the mass distribution of the signals. The parameters are extracted from MC simulation, and they are taken to be uncorrelated with the c-BDT output.
- $B \rightarrow hh'$ background PDF: the mass distribution of the peaking background is similar to the B_d^0 signal, as shown in figure A.1(b). The same double gaussian model is therefore used, with fixed shape parameters obtained from simulations and a normalisation of 1.0 ± 0.4 events, as introduced in section A.2. As for the signal, the $B \rightarrow hh'$ contribution is equally distributed among the three c-BDT intervals.
- Continuum background PDF: the dependence of this background on the dimuon mass is described with a Chebychev first order polynomial. The slope parameter is independent in the three bins but subjected to a loose gaussian constraint within $\pm 40\%$ between the first and second interval, and $\pm 80\%$ between the first and the third.
- PRD background PDF: the dependence of this background on the dimuon mass is described with an exponential. The same slope parameter is used in the three c-BDT bins.

A.4.1 Systematic uncertainties in the fit

Two sets of systematic uncertainties are included in the fit: uncertainties due to the signal efficiency in the three c-BDT intervals and uncertainties due to the fitting procedure.

The former is obtained as explained in section 10.2.3, values of ± 0.026 , ± 0.010 , ± 0.023 are found respectively in the first, second and third c-BDT bin. Gaussian terms are included in the fitting procedure in order to describe these uncertainties. Care is taken in constraining the sum of the efficiencies in the three intervals to 1, since the uncertainty on the c-BDT selection is already included in the efficiency ratio term.

The systematic uncertainties due to the fitting procedure are evaluated with toy-MC experiments. Variations in the description of the components of the fit are considered in the generation of the toy-MC, and the generated sample is fitted with the nominal fit configuration. The corresponding variations in the number of fitted signal events with respect to the toy-MC study performed with the nominal fit configuration used in both generation and fit are taken as systematic uncertainties.

The normalisation and shape parameters of the different components in the toy-MC experiments are set to what extracted from the mass sidebands of the signal region in data or from simulations.

The following variations in the baseline model are considered:

- **Mass scale ± 5 MeV and Mass resolution $\pm 5\%$:** the sensitivity of the signal fit to the mass scale and mass resolution uncertainty uncertainties is tested by separately shifting the signal mean by 5 MeV and varying the width of the signal PDF increased or reduced by 5%.
- **Combinatorial bkg PDF model and low mass bkg PDF model:** the assumption on the functional models for the exponential and polynomial are (separately) probed using respectively a gaussian-tail and an exponential.
- **Inclusion of semil. background and Inclusion of B_c background:** the semileptonic and B_c backgrounds are neglected in the baseline fit. Systematic uncertainties associated with this choice are evaluated by including (separately) the semileptonics or the B_c backgrounds in the generation of the toy-MC. Their PDF and the shape parameters and normalisations are evaluated from simulations.
- **Polynomial slope:** the introduction of the gaussian constraints on the polynomial slope is challenged by varying the mass dependence among the different c-BDT intervals.
- **Exponential slope:** the usage of the same exponential slope in all c-BDT bins is challenged by varying the mass dependence among the different c-BDT intervals.

The different variations are evaluated in four points in the $N(B_s^0) - N(B_d^0)$ space, three close to the unphysical region and the SM expectation corresponding to $N(B_s^0) = 41, N(B_d^0) = 5$. Samples of 10000 toy experiments per variation are generated in each point, together with the baseline configuration toys.

For each point, the quadratic sums of the deviations computed separately for the positive and the negative contributions are evaluated for both B_s^0 and B_d^0 ; the larger value between the positive and negative variations is taken as systematic uncertainty in the fit. The correlation between the $N(B_s^0)$ and $N(B_d^0)$ variations is also evaluated.

For B_s^0 the total systematic uncertainty is found to increase with the assumed size of the signal, while for B this dependence is not observed. The systematic uncertainty on B_s^0 is parametrised as $\sigma_{\text{syst}}(N_s) = \sqrt{4 + (0.06N_s)^2}$, while the systematic uncertainty on B_d^0 is parametrised as $\sigma_{\text{syst}}(N_d) = 3$. The correlation among the B_s^0 and B_d^0 variations is found to be approximately constant $\rho_{\text{syst}} = -0.7$.

These systematic uncertainties are included in the fit as two smearing parameters for $N(B_s^0)$ and $N(B_d^0)$, constrained by a two-dimensional Gaussian distribution parametrised by the values of $\sigma_{\text{syst}}(N_s)$, $\sigma_{\text{syst}}(N_d)$ and ρ_{syst} .

A.4.2 Fit on unblinded data

The fit on unblinded data, shown in figure A.3 yields to $N(B_s^0) = 16 \pm 12$ and $N(B_d^0) = -11 \pm 9$, where the uncertainties include the systematics on the fit discussed in section A.4.1, but statistical uncertainties largely dominate. As mention in section A.4.1, the expected yields for the SM branching fractions are $N(B_s^0) = 41$, $N(B_d^0) = 5$.

The primary result of this analysis is obtained by applying the natural boundary of non-negative yields, for which the fit returns the values $N(B_s^0) = 11$ and $N(B_d^0) = 0$.

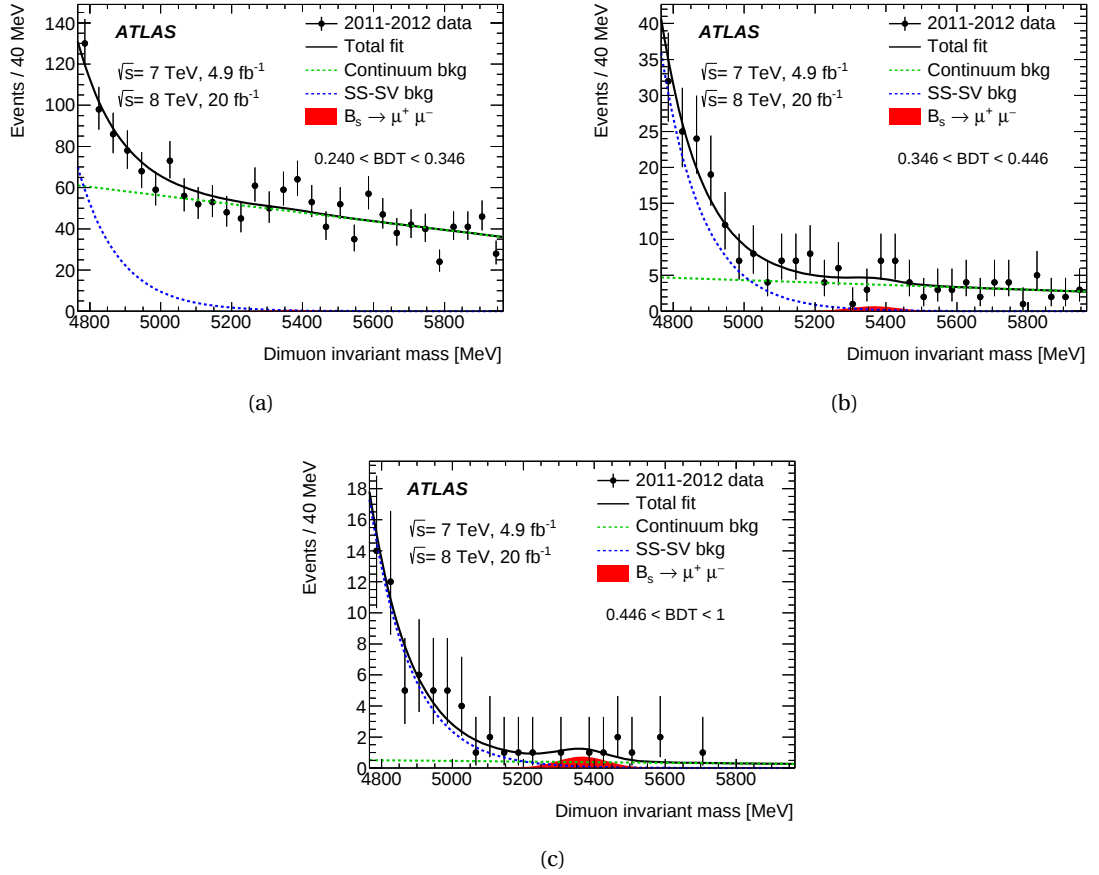


Figure A.3: Dimuon invariant mass distributions in the unblinded data, in the three intervals of continuum-BDT output. Superimposed is the result of the maximum-likelihood fit, obtained imposing the boundary of non-negative signal contributions. The total fit is shown as a black continuous line, the filled area corresponds to the observed signal component, the blue dashed line to the SS+SV background, and the green dashed line to the continuum background. Figures taken from [73].

A.5 Branching fraction extraction

The branching fractions of $B_s^0 \rightarrow \mu^+ \mu^-$ and $B_d^0 \rightarrow \mu^+ \mu^-$ are extracted from data using the fit described in section A.4 and replacing $N(B_s^0)$ and $N(B_d^0)$ with the corresponding branching fractions divided by the normalisation terms shown in equation A.1. The uncertainties on the normalisation terms are included as gaussian smearings.

The $\mathcal{B}(B^+ \rightarrow J/\psi(\mu^+ \mu^-)K^+)$ term is obtained from [207] as the product of $\mathcal{B}(B^+ \rightarrow J/\psi K^+) =$

$(1.027 \pm 0.031) \times 10^{-3}$ and $J/\psi \rightarrow \mu^+ \mu^- = (5.961 \pm 0.033)\%$. The hadronisation probability is one for B_d^0 and for B_s^0 is obtained from the ATLAS measurement $f_s/f_d = 0.240 \pm 0.020$ [176], which uses $f_u/f_d = 1$ [216].

The D_{norm} term, obtained by combining the quantities shown in equation A.2 is $D_{\text{norm}} = (2.88 \pm 0.17) \times 10^6$ for B_s^0 and $D_{\text{norm}} = (2.77 \pm 0.16) \times 10^6$ for B_d^0 , where the former includes the correction due to the width difference between the B_s^0 mass eigenstates.

The values of the branching fractions that maximise the likelihood function within the constraint of non-negative values are $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0.9 \times 10^{-9}$ and $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = 0$.

The 68.3% confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ obtained with the Neyman construction approach [178] is

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) = (0.9_{-0.8}^{+1.1}) \times 10^{-9}, \quad (\text{A.4})$$

where the uncertainty includes both statistical and systematic contributions.

Figure A.4, already shown in section 1.5 and included here for completeness, shows the 2-dimensional profiled likelihood contours in the $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - BR(B_s^0 \rightarrow \mu^+ \mu^-)$ plane drawn for $-2\Delta\ln(L/L_{\text{max}}) = 2.3, 6.2$ and 11.8 , allowing negative values of the branching fractions. The 68.3% confidence interval on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ is also shown, together with the corresponding contours of the combination of the results of the CMS and LHCb experiments [70] and the SM prediction [54].

The CL_s method [217] implemented with toy-MC is employed to set an upper limit on both

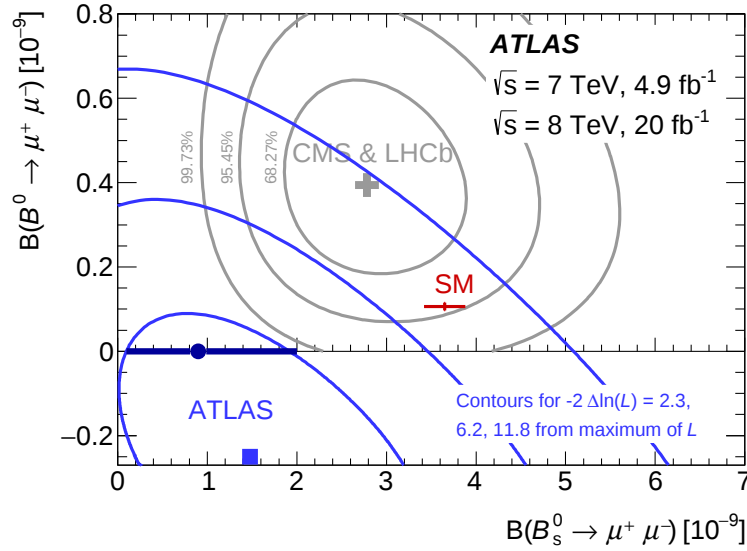


Figure A.4: Contours in the plane $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) - BR(B_s^0 \rightarrow \mu^+ \mu^-)$ for intervals of $-2\Delta\ln(L)$ equal to 2.3, 6.2 and 11.8 relative to the absolute maximum of the likelihood, without imposing the constraint of non-negative branching fractions. Also shown are the corresponding contours of the combined result of the CMS and LHCb experiments, the SM prediction, and the maximum of the likelihood within the boundary of non-negative branching fractions, with the error bars covering the 68.3% confidence range for $B_s^0 \rightarrow \mu^+ \mu^-$. Picture from [73].

branching fractions. The 95% CL upper limit on $\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-)$ obtained under the back-

ground only hypothesis, with $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$ left free to be determined by the fit is

$$\mathcal{B}(B_s^0 \rightarrow \mu^+ \mu^-) < 3.0 \times 10^{-9} \quad (95\% \text{ CL}), \quad (\text{A.5})$$

for an expected upper limit of $1.8_{-0.4}^{+0.7} \times 10^{-9}$.

The 95% CL upper limit on $\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-)$, obtained generating the toy-MC according to the observed amplitudes of backgrounds and B_s^0 signal is

$$\mathcal{B}(B_d^0 \rightarrow \mu^+ \mu^-) < 4.2 \times 10^{-10} \quad (95\% \text{ CL}), \quad (\text{A.6})$$

for an expected upper limit of $5.7_{-1.5}^{+2.1} \times 10^{-9}$.

ADDITIONAL STUDIES ON KINEMATIC CORRECTIONS

B

This appendix reports several additional studies performed on the kinematic reweighting designed for the inclusive simulations (QLC and DDW, extensively described in section 7.6.4).

Section B.1 shows a study on the $\hat{p}_T$ min cut applied at generator level to all the simulated samples.

Section B.2 reports the plots of all the cross-checks performed to validate the QLC: the corrections are evaluated using only half of the samples, while the remaining half samples are weighted and used for the comparison.

Section B.2.4 shows additional studies performed on the binning scheme chosen for the QLC. Section B.3 reports the plots of all the cross-checks performed on the DDW. As for the checks on the QLC, the DDW are evaluated using only half of the MC and data samples, while the remaining halves are weighted and used for comparison.

B.1 Studies on $\hat{p}_T$ min cut

Considering the generator-level cuts applied to the exclusive simulations at generator level, QLC might be evaluated using only one sample generated with unbiased quark-level cuts and no final state particle cuts. The formula used in this case would be:

$$W_{QL} = \frac{\nu^{\text{unbiased+QLsel+FScuts}}(p_T(B), \eta(B))}{\nu^{\text{unbiased+FScuts}}(p_T(B), \eta(B))}, \quad (\text{B.1})$$

where ν is the number of entries in a $(p_T(B), \eta(B))$ bin and the quark-level selection (QLsel) and final state particle cuts (FScuts), implicit in the default MC sample, are explicitly applied to the unbiased sample.

This approach was tested by evaluating the QLC for the $B^+ \rightarrow J/\psi K^+$ process using the unbiased sample introduced in section 7.6.4.1. As for the QLC evaluated in section 7.6.4.1, the weights have been cross-checked by calculating the QLC using odd events from the unbiased MC and applying them to the quark biased sample. The resulting $(p_T(B), \eta(B))$ distributions are compared with the distributions of the even events from the unbiased sample. The outcome of this cross-check is reported in figure B.1. The B meson η distribution shows good agreement, while the $p_T(B)$ distribution shows an inconsistency at low p_T . The same effect is visible swapping odd and even unbiased events.

The source of this discrepancy was found to be the $\hat{p}_T$ min cut introduced at generator level. Figure B.2 shows the $\hat{p}_T$ distribution of the unbiased sample and the $\hat{p}_T$ distribution from a new sample, called semi biased, generated with the same cuts as the unbiased sample but with tighter $\hat{p}_T$ (7 GeV instead of 5 GeV). The distribution of the unbiased sample is cut at 7 GeV in order to compare the two distributions, that are clearly not compatible.

This is due to a regularisation in the Pythia generation for low $\hat{p}_T$: basically the parton-parton cross section becomes too high as $\hat{p}_T$ decreases, causing a violation of unitarity. The Pythia generator smoothly takes care of this divergence applying a regularisation for decreasing $\hat{p}_T$ values [148], effectively modifying the $\hat{p}_T$ distribution.

Due to this feature QLC calculated using a single unbiased sample are not usable, as they would introduce an additional bias due to this $\hat{p}_T$ discrepancy. The correct approach for the QLC evaluation is therefore based on the usage of two samples, as described in 7.6.4.1.

An additional check is needed, before being able to evaluate the QLC. The set of quark-level cuts employed in the generation of the unbiased sample has to be tested to verify that it is not affected by the $\hat{p}_T$ bias. A new sample with the same quark-level cuts as the unbiased but with looser $\hat{p}_T$ (3 GeV instead of 5) is generated and its $\hat{p}_T$ distribution is compared with the one from the unbiased sample. Figure B.3 shows this comparison. The distribution of the new sample is cut at 5 GeV in order to compare the two distributions. Quark-level cuts, final state particle cuts and B fiducial volume cuts ($p_T(B) > 8$ GeV and $|\eta(B)| < 2.5$) are applied, as the main interest is to avoid a bias in the parameter space used in the analysis. The two distributions are compatible, therefore the cuts chosen for the unbiased sample do not introduce any additional bias.

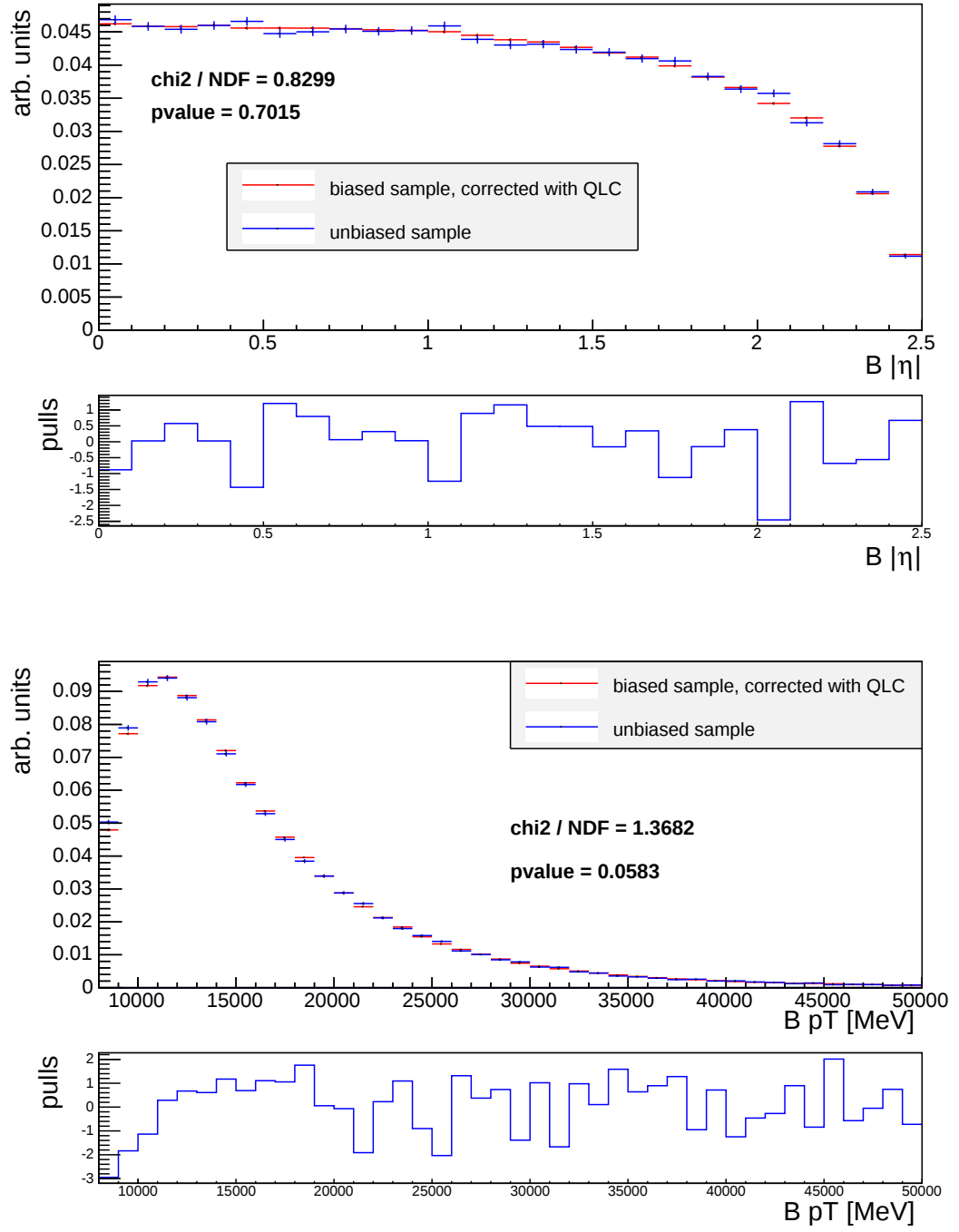


Figure B.1: Comparison of the $\eta(B)$ (top plot) and $p_T(B)$ (bottom plot) QLC corrected quark biased distribution and the unbiased distribution. The QLC have been calculated using the approach based on one sample described in the text. In order to avoid correlations between the distributions, QLC have been calculated using odd numbered events from the unbiased sample and the QLC corrected quark biased distributions are compared with even numbered unbiased events.

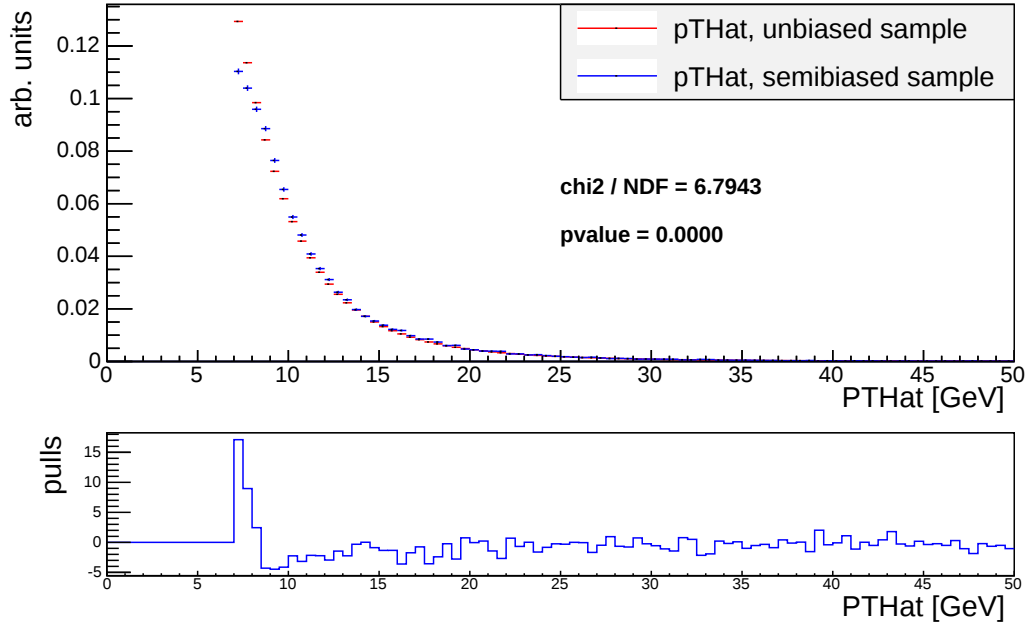


Figure B.2: Comparison of the $\hat{p}_T$ distribution from the unbiased sample and the $\hat{p}_T$ distribution from a new sample, named semi biased. This new sample has the same quark-level cuts as the unbiased but tighter $\hat{p}_T$ (7 GeV instead of 5). In order to compare the two distribution, the $\hat{p}_T$ spectrum of the unbiased sample is cut at 7 GeV.

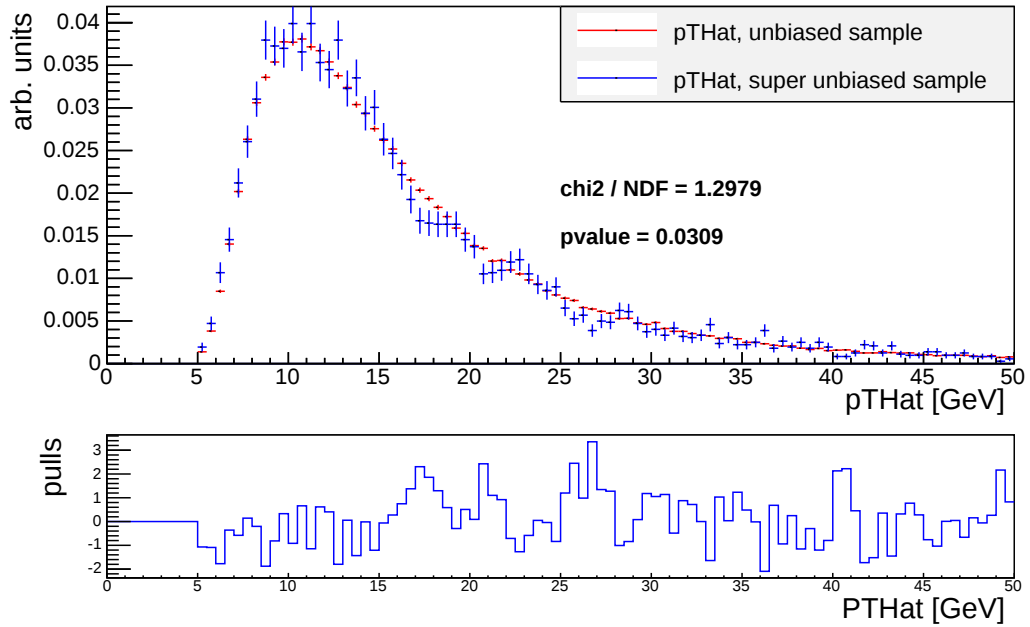


Figure B.3: Comparison of the $\hat{p}_T$ distribution from the unbiased sample and the $\hat{p}_T$ distribution from a new sample characterised by the same quark-level cuts as the unbiased but with looser $\hat{p}_T$ (3 GeV instead of 5). The $\hat{p}_T$ distribution from this sample is cut at 5 GeV in order to compare the two distributions. Quark-level cuts, final state cuts and B fiducial volume cuts ($p_T(B) > 8$ GeV and $|\eta(B)| < 2.5$) are applied to both distributions.

B.2 Cross-checks on QLC

B.2.1 $B^+ \rightarrow J/\psi K^+$ QLC consistency

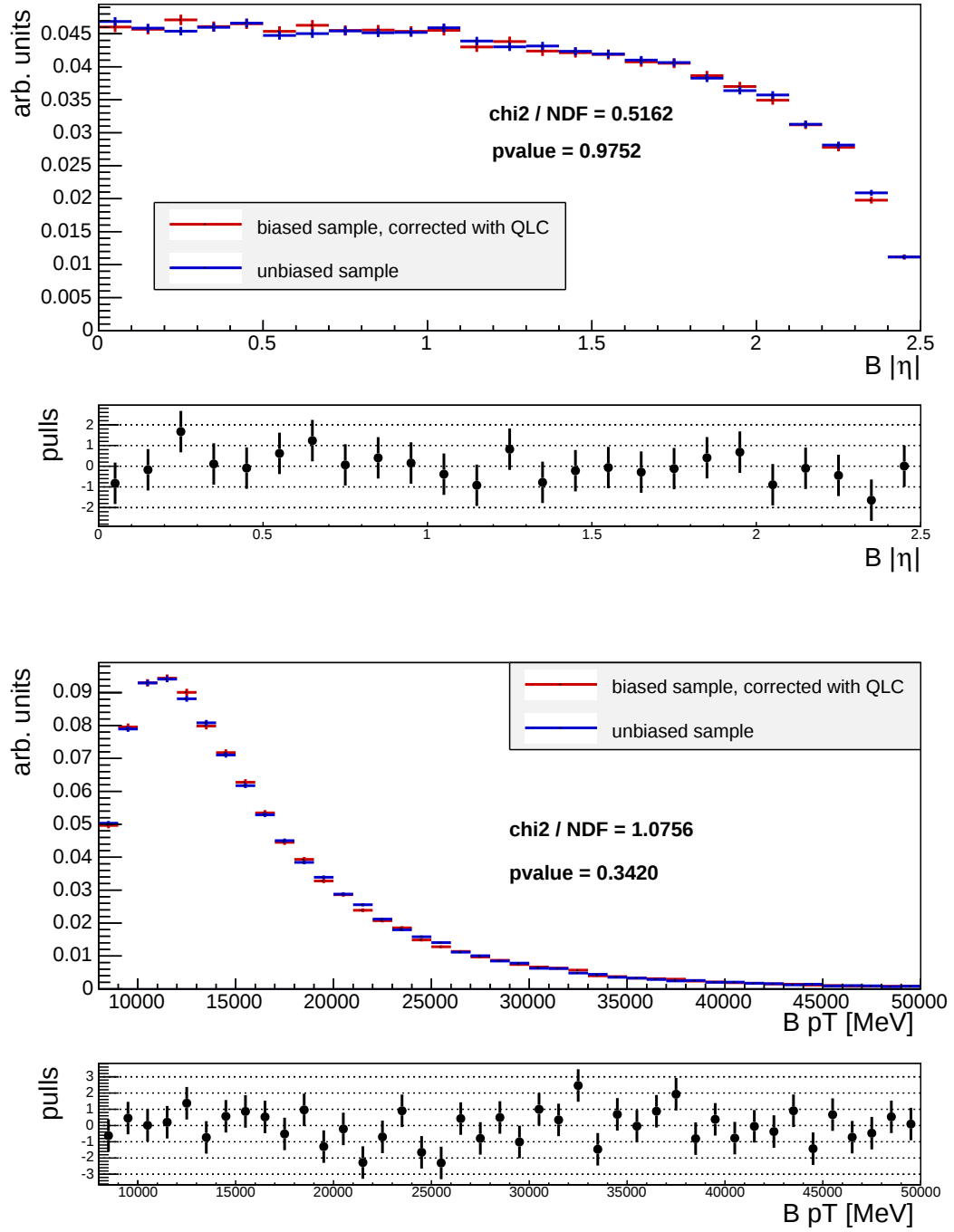


Figure B.4: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B^+ \rightarrow J/\psi K^+$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using odd-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

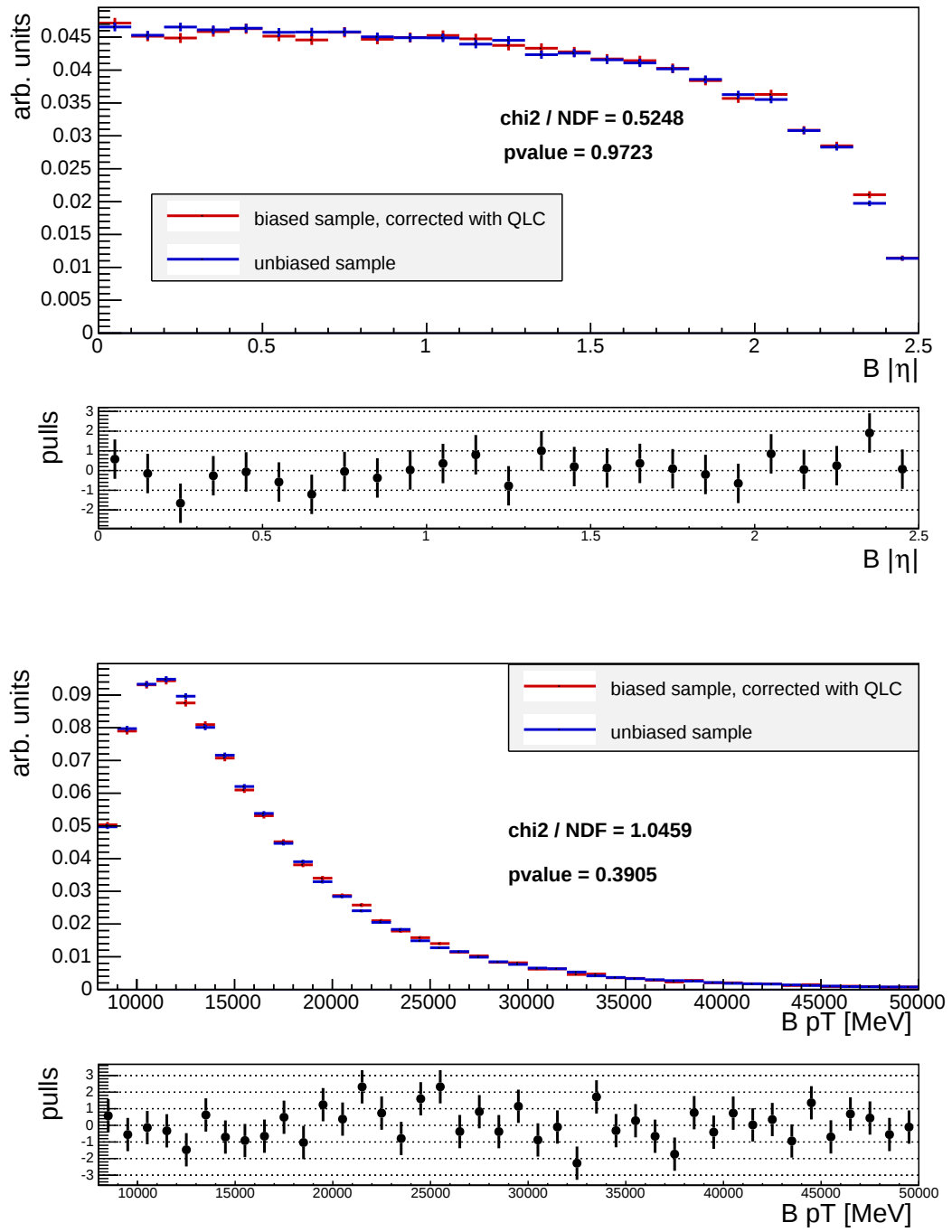


Figure B.5: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B^+ \rightarrow J/\psi K^+$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using even-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

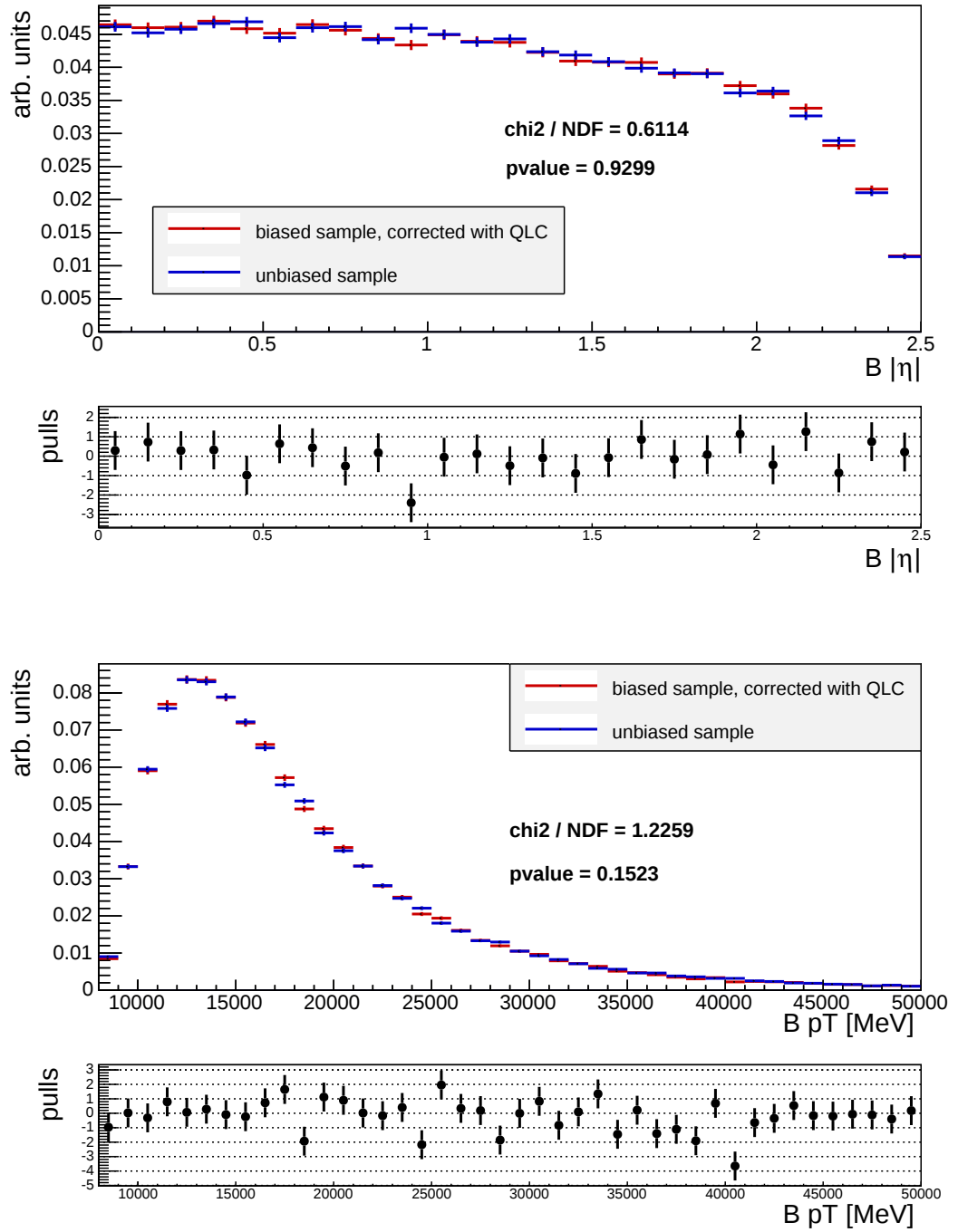
B.2.2 $B_s^0 \rightarrow J/\psi\phi$ QLC consistency

Figure B.6: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B_s^0 \rightarrow J/\psi\phi$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using odd-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

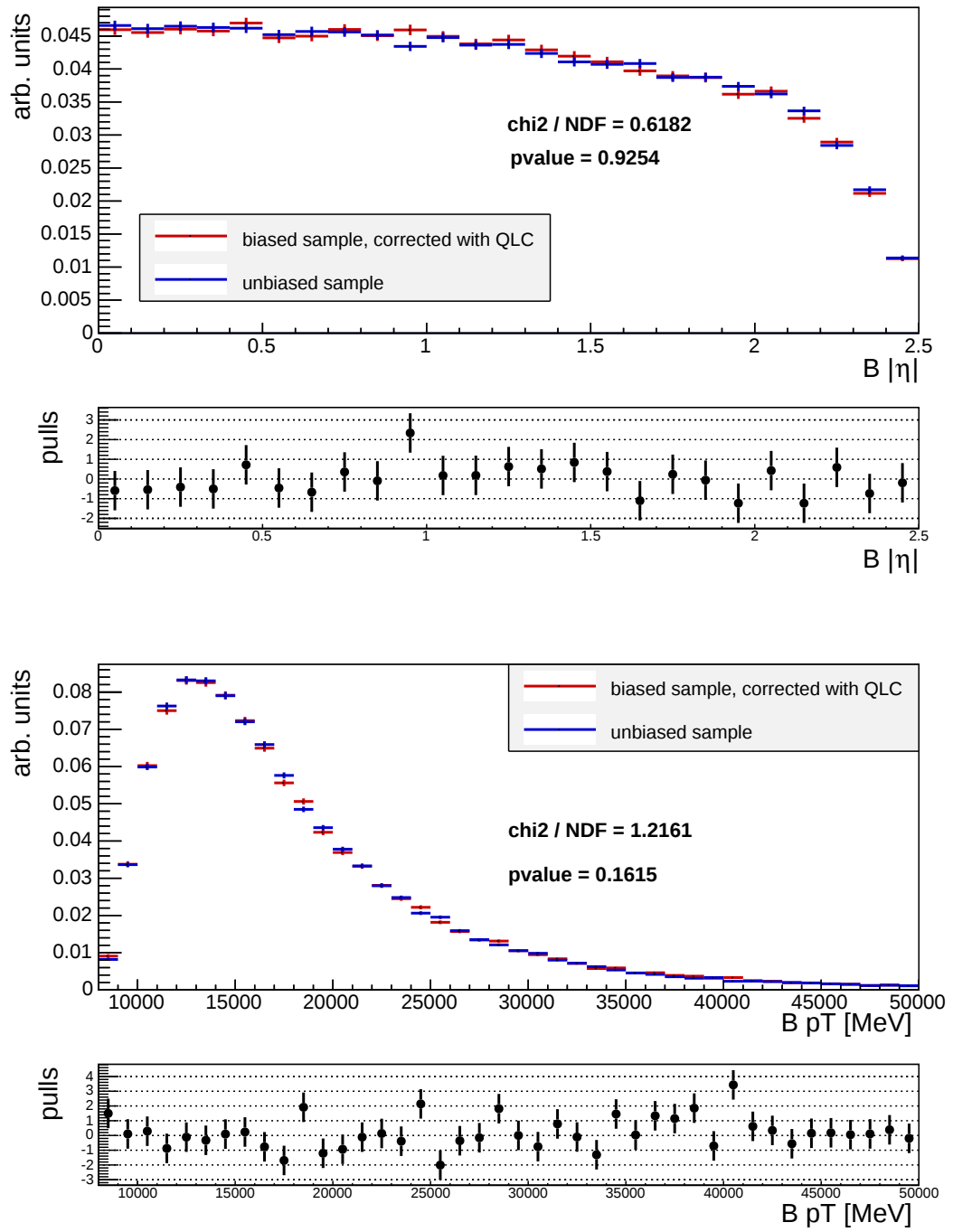


Figure B.7: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B_s^0 \rightarrow J/\psi\phi$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using even-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

B.2.3 $B_s^0 \rightarrow \mu^+ \mu^-$ QLC consistency

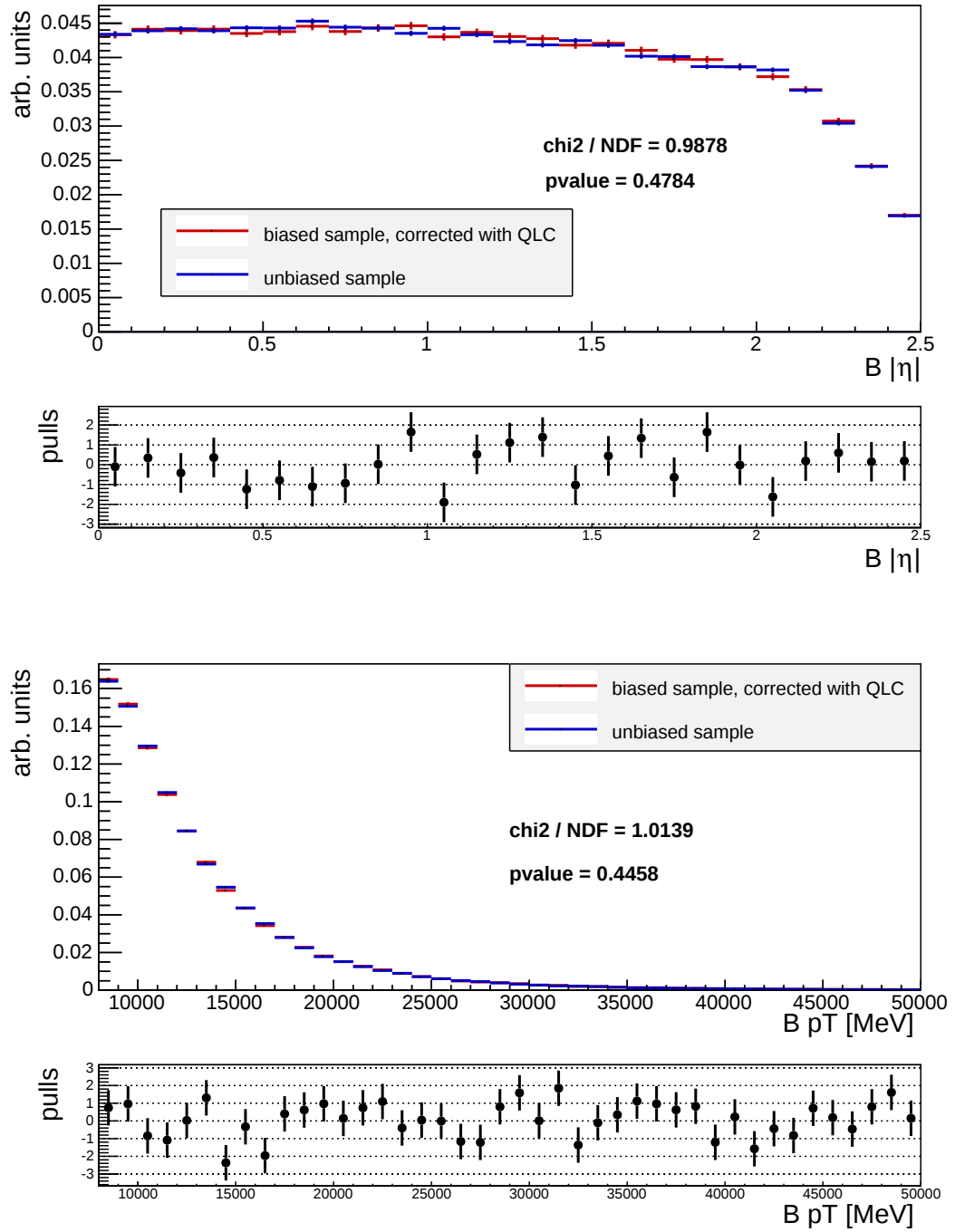


Figure B.8: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B_s^0 \rightarrow \mu^+ \mu^-$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using odd-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

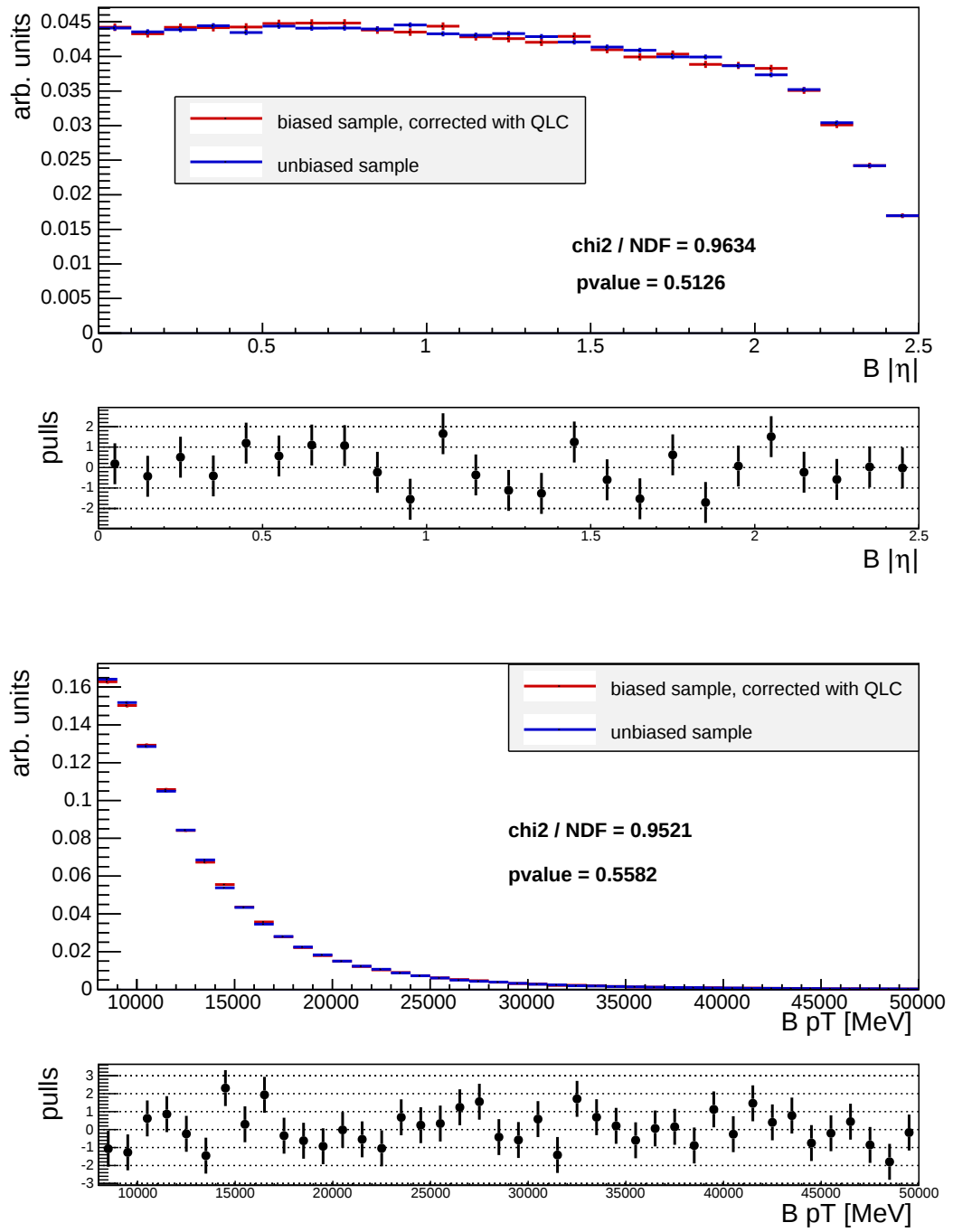


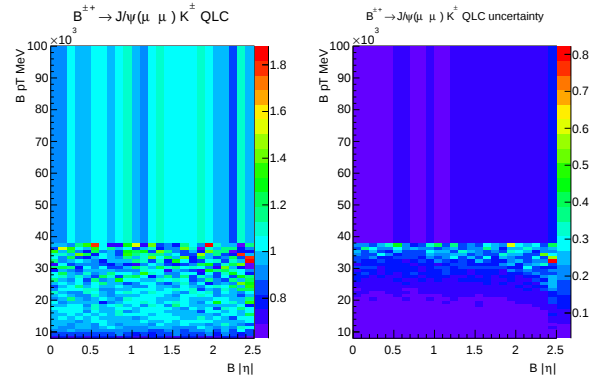
Figure B.9: Comparison of the $\eta(B)$ (*top plot*) and $p_T(B)$ (*bottom plot*) QLC corrected $B_s^0 \rightarrow \mu^+ \mu^-$ quark biased distribution with the respective unbiased distributions. In order to avoid correlations between the distributions, the QLC are calculated using even-numbered events from the unbiased and quark biased samples, and the remaining events in the two samples are weighted and used for the comparison.

B.2.4 Check on QLC binning scheme

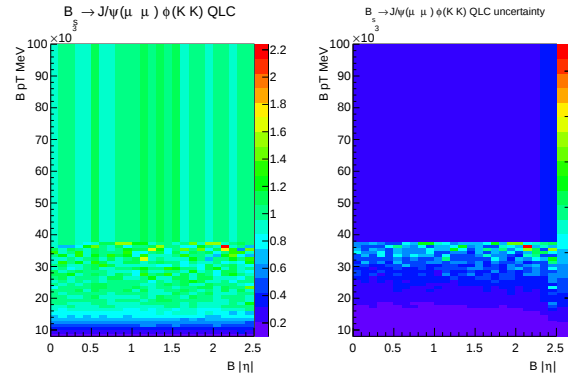
The $(p_T(B), \eta(B))$ binning scheme chosen for the QLC calculation assures a comparable statistical uncertainty in every bin. Basically the binning scheme is composed of two regions, one with a regular binning, in the $p_T < 38$ GeV region, with bins of size $(0.1 \times 1$ GeV), and bins that cover the entire $|\eta(B)|$ spectrum and increasing size in p_T in the $p_T > 38$ GeV region.

In order to check that a unique $|\eta(B)|$ bin in the region above $p_T(B) = 38$ GeV is sufficient to provide an adequate correction to the simulations, the QLC are evaluated using a different binning scheme. While the $p_T < 38$ GeV region maintains the same binning scheme as before, one bin in the $p_T(B)$ direction for $p_T > 38$ GeV and bins with a width of 0.1 in the $|\eta(B)|$ direction are used. The resulting QLC are shown in figure B.10.

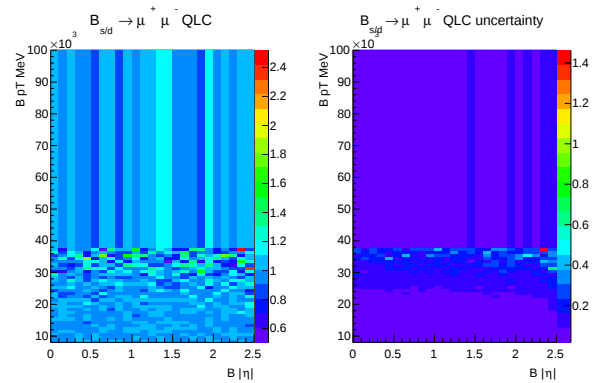
In order to verify the absence of a $|\eta(B)|$ dependence of the QLC for $p_T > 38$ GeV, the QLC distribution for said region is shown as a 1 dimensional histogram and a linear fit is performed on them. The results are shown in figure B.11, the histograms are compatible with a flat line, therefore the QLC distributions do not show a $|\eta(B)|$ dependence.



(a) $B^{++} \rightarrow J/\psi(\mu^+ \mu^-) K^{++}$ QLC (left) and the corresponding uncertainty (right) calculated with the modified binning scheme in the $p_T(B) > 38$ GeV region.

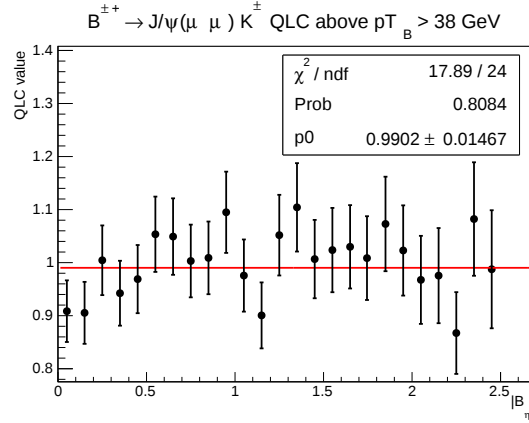


(b) $B_s \rightarrow J/\psi(\mu^+ \mu^-) \phi(K^+ K^-)$ QLC (left) and the corresponding uncertainty (right) calculated with the modified binning scheme in the $p_T(B) > 38$ GeV region.

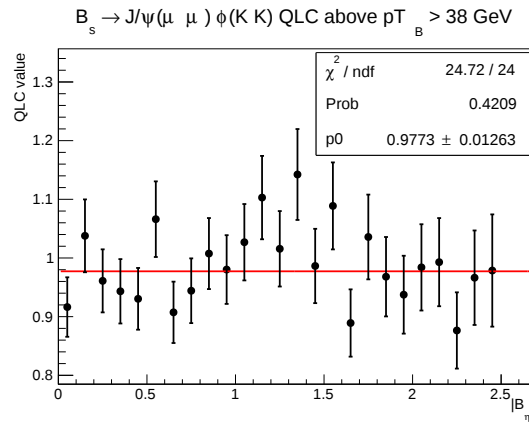


(c) $B_s \rightarrow \mu^+ \mu^-$ QLC (left) and the corresponding uncertainty (right) calculated with the modified binning scheme in the $p_T(B) > 38$ GeV region.

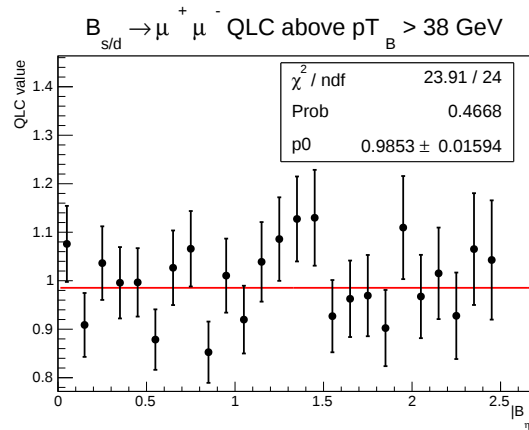
Figure B.10: QLC (left plot) and the relative uncertainty (right plot), calculated for three processes, respectively, $B^{++} \rightarrow J/\psi K^{++}$ figure B.10(a), $B_s \rightarrow J/\psi \phi$ figure B.10(b) and $B_s \rightarrow \mu\mu$ figure B.10(c). The binning scheme in the $p_T(B) > 38$ GeV region has been modified according to the text.



(a)



(b)



(c)

Figure B.11: QLC for the $p_T(B) > 38$ GeV region calculated using the modified binning scheme described in the text for three processes, respectively $B^{\pm} \rightarrow J/\psi K^{\pm}$, figure B.11(a), $B_s^0 \rightarrow J/\psi \phi$, figure B.11(b) and $B_{s/d}^0 \rightarrow \mu^+ \mu^-$, figure B.11(c). A linear fit is performed on each distribution to verify the compatibility with a flat line, showing the absence of a $|\eta|(B)$ dependence.

B.2.5 Check on QLC $\eta(B)$ symmetry

The QLC employed in the analysis assume the B meson η spectrum to be symmetric with respect to 0, therefore the QLC are evaluated for $|\eta(B)| < 2.5$; this allows to increase the statistics in each $(p_T(B), \eta(B))$ bin, reducing the uncertainty associated to the QLC.

The assumption on the QLC symmetry with respect to B meson η is tested by calculating the QLC in the full $\eta(B)$ range. The binning scheme chosen is the same as the one used for the correction employed in the analysis, but symmetrised with respect to $\eta(B) = 0$. Figure B.12 shows a 2D $(p_T(B), \eta(B))$ map of the QLC evaluated in the full $\eta(B)$ range and their uncertainties for the three processes $B^+ \rightarrow J/\psi K^+$, $B_s \rightarrow J/\psi \phi$ and $B_s \rightarrow \mu\mu$. The compatibility of the $\eta(B) > 0$ QLC with the $\eta(B) < 0$ QLC is quantified with a χ^2 test. Figure B.13 shows the pulls evaluated as $(\text{QLC}^{\eta>0} - \text{QLC}^{\eta<0})^2 / ((\sigma_{\text{QLC}}^{\eta>0})^2 + (\sigma_{\text{QLC}}^{\eta<0})^2)$ for the QLC at $(p_T(B), |\eta(B)|)$ and $(p_T(B), -|\eta(B)|)$; the pulls are calculated for all the three processes considered in the QLC extraction. The χ^2 is evaluated as the sum of all the pulls, while the number of degrees of freedom is equal to the number of bins. The resulting p -values for the compatibility are 0.998 for $B^+ \rightarrow J/\psi K^+$, 0.997 for $B_s^0 \rightarrow J/\psi \phi$ and 0.891 for $B_s^0 \rightarrow \mu^+ \mu^-$, confirming the absence of a $\eta(B)$ asymmetry for the QLC.

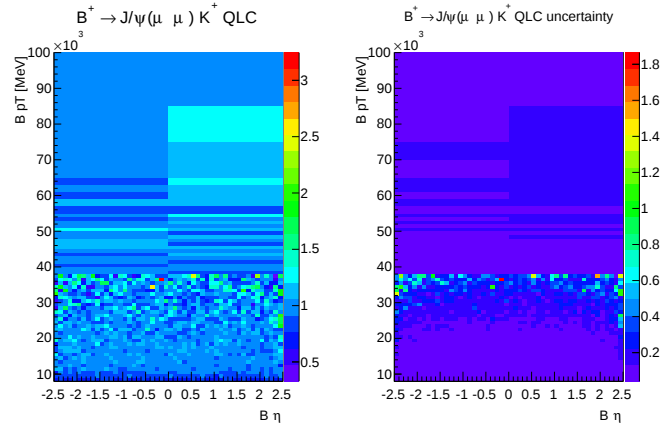
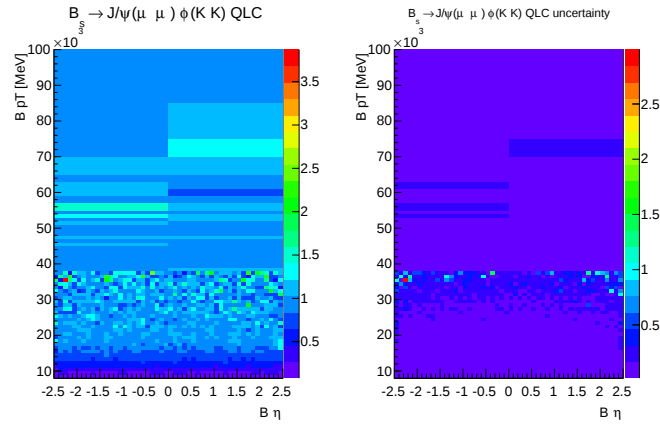
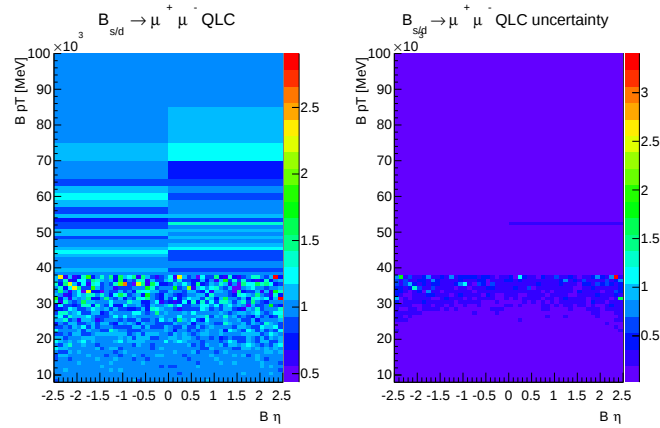
(a) $B^+ \rightarrow J/\psi K^+$ QLC (left) and the corresponding uncertainty (right).(b) $B_s \rightarrow J/\psi \phi$ QLC (left) and the corresponding uncertainty (right).(c) $B_s \rightarrow \mu\mu$ QLC (left) and the corresponding uncertainty (right).

Figure B.12: The three sets of plots show the QLC evaluated in the full $\eta(B)$ range (left plot) and the relative uncertainty (right plot), calculated for the three processes, respectively, $B^+ \rightarrow J/\psi K^+$ figure B.12(a), $B_s \rightarrow J/\psi \phi$ figure B.12(b) and $B_s \rightarrow \mu\mu$ figure B.12(c).

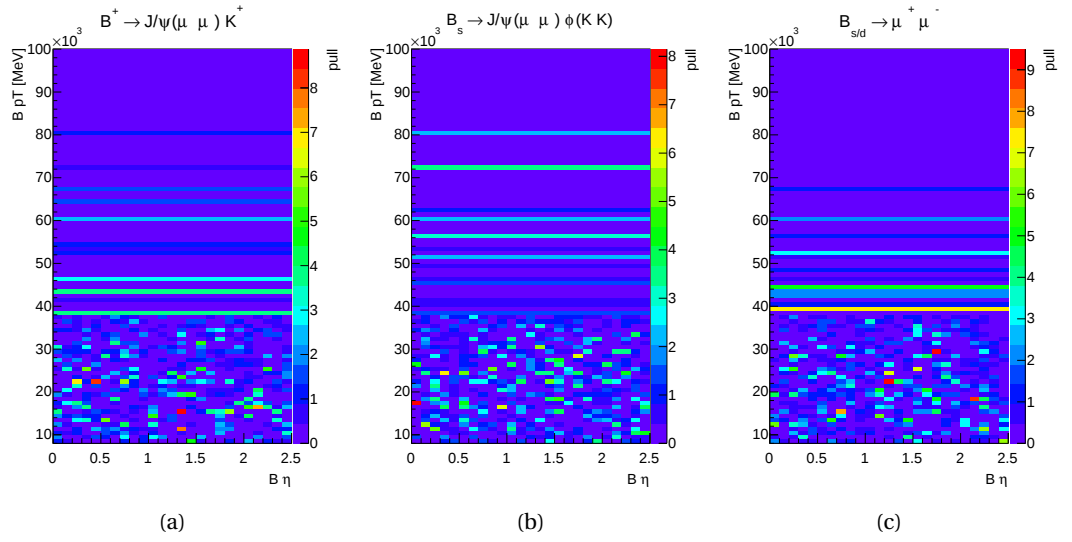


Figure B.13: Pulls evaluated to test the compatibility of the $\eta(B) > 0$ QLC with the $\eta(B) < 0$ QLC. The pulls are calculated as $(\text{QLC}^{\eta>0} - \text{QLC}^{\eta<0})^2 / (\sigma_{\text{QLC}}^{\eta>0} + \sigma_{\text{QLC}}^{\eta<0})$ for the QLC at $(p_T(B), \eta(B))$ and $(p_T(B), -\eta(B))$. The three plots are referred to the three processes involved in the QLC evaluation, respectively, $B^+ \rightarrow J/\psi K^+$ in figure B.13(a), $B_s \rightarrow J/\psi \phi$ in figure B.13(b) and $B_s \rightarrow \mu\mu$ in figure B.13(c).

B.3 Cross-checks on DDW

B.3.1 $B^+ \rightarrow J/\psi K^+$ DDW consistency

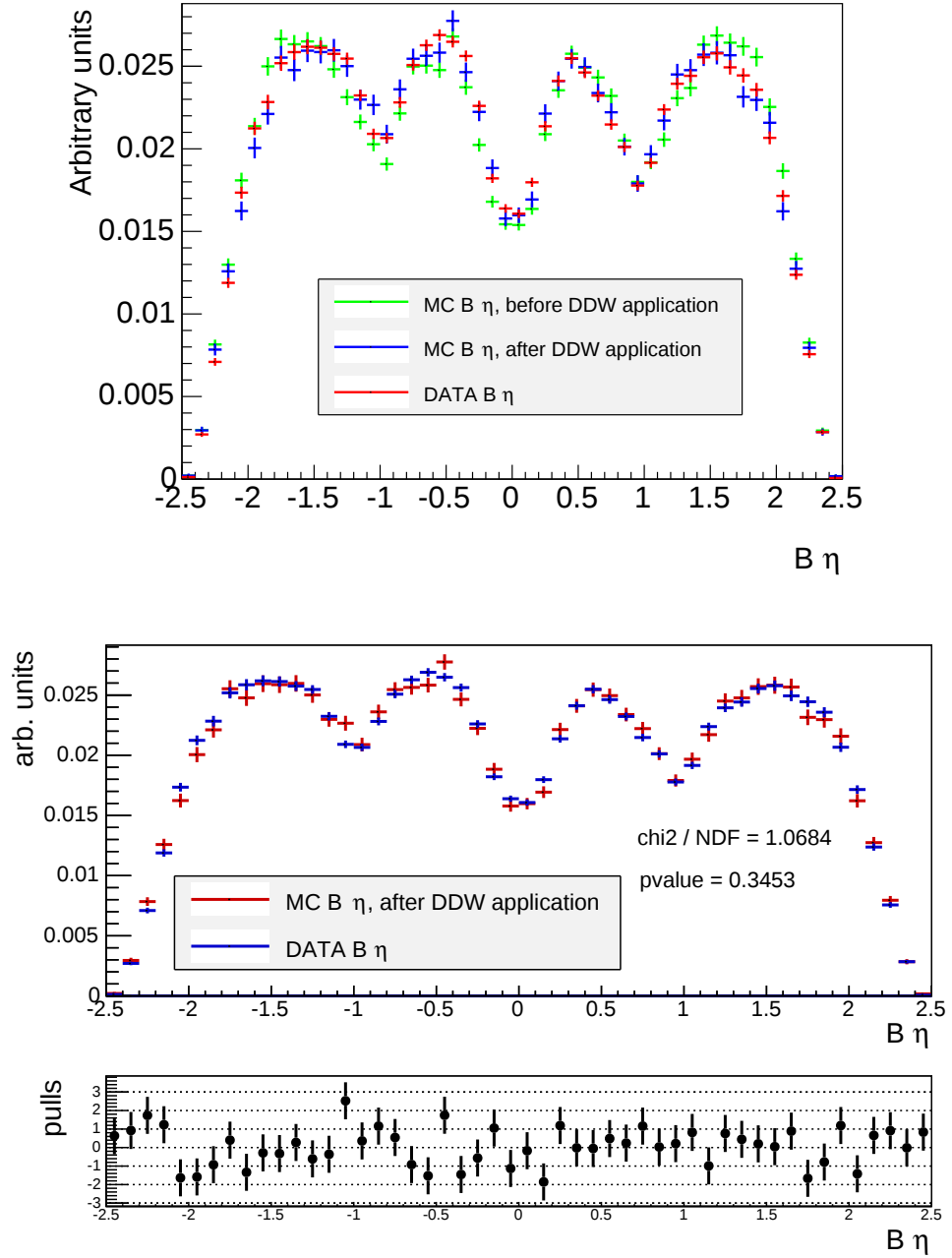


Figure B.14: *Top plot:* comparison of the $\eta(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot:* quantitative comparison between the $\eta(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using odd-numbered events from both data and $B^+ \rightarrow J/\psi K^+$ MC, and the remaining events in the two samples are weighted and used for the comparison.

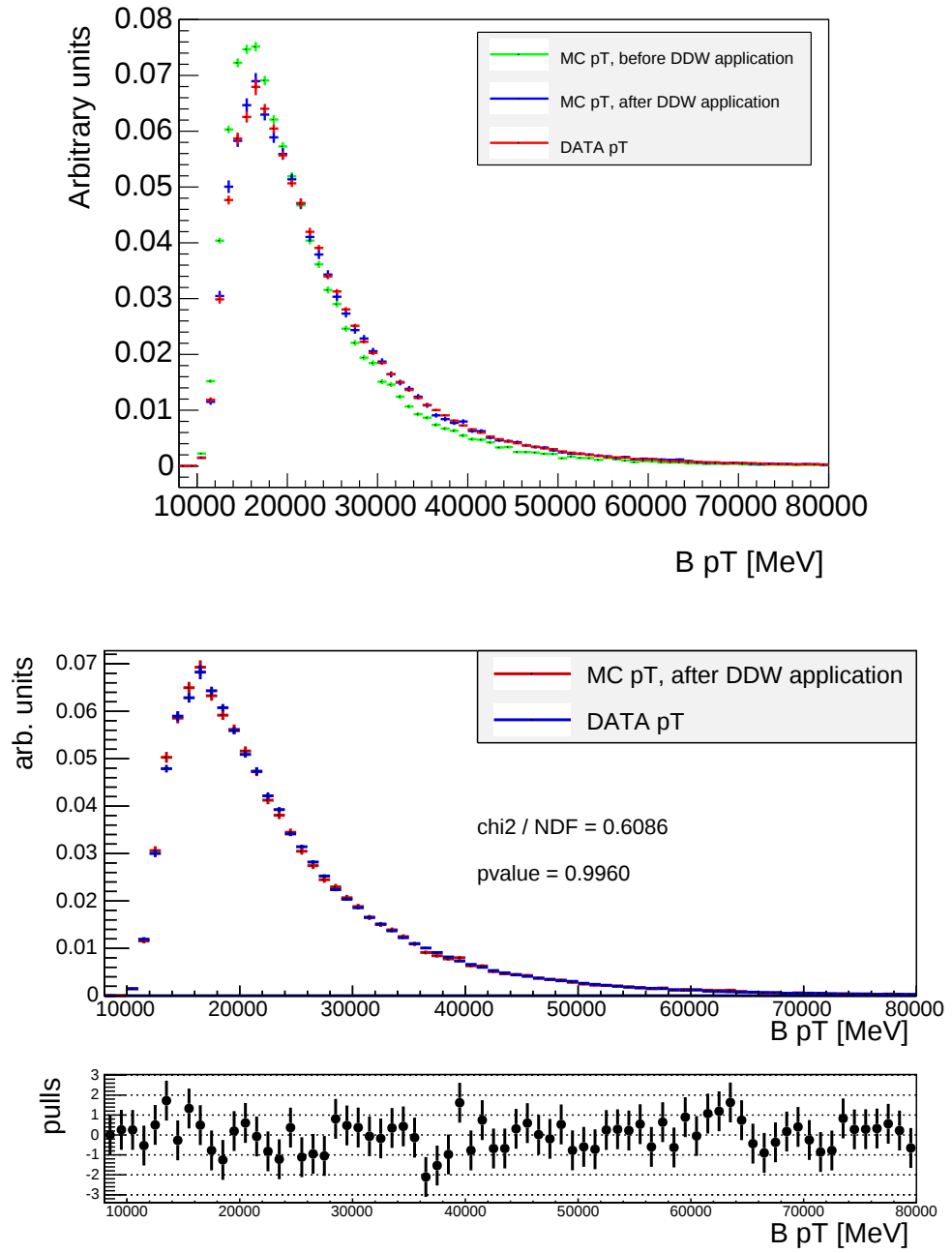


Figure B.15: *Top plot*: comparison of the $p_T(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot*: quantitative comparison between the $p_T(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using odd-numbered events from both data and $B^+ \rightarrow J/\psi K^+$ MC, and the remaining events in the two samples are weighted and used for the comparison.

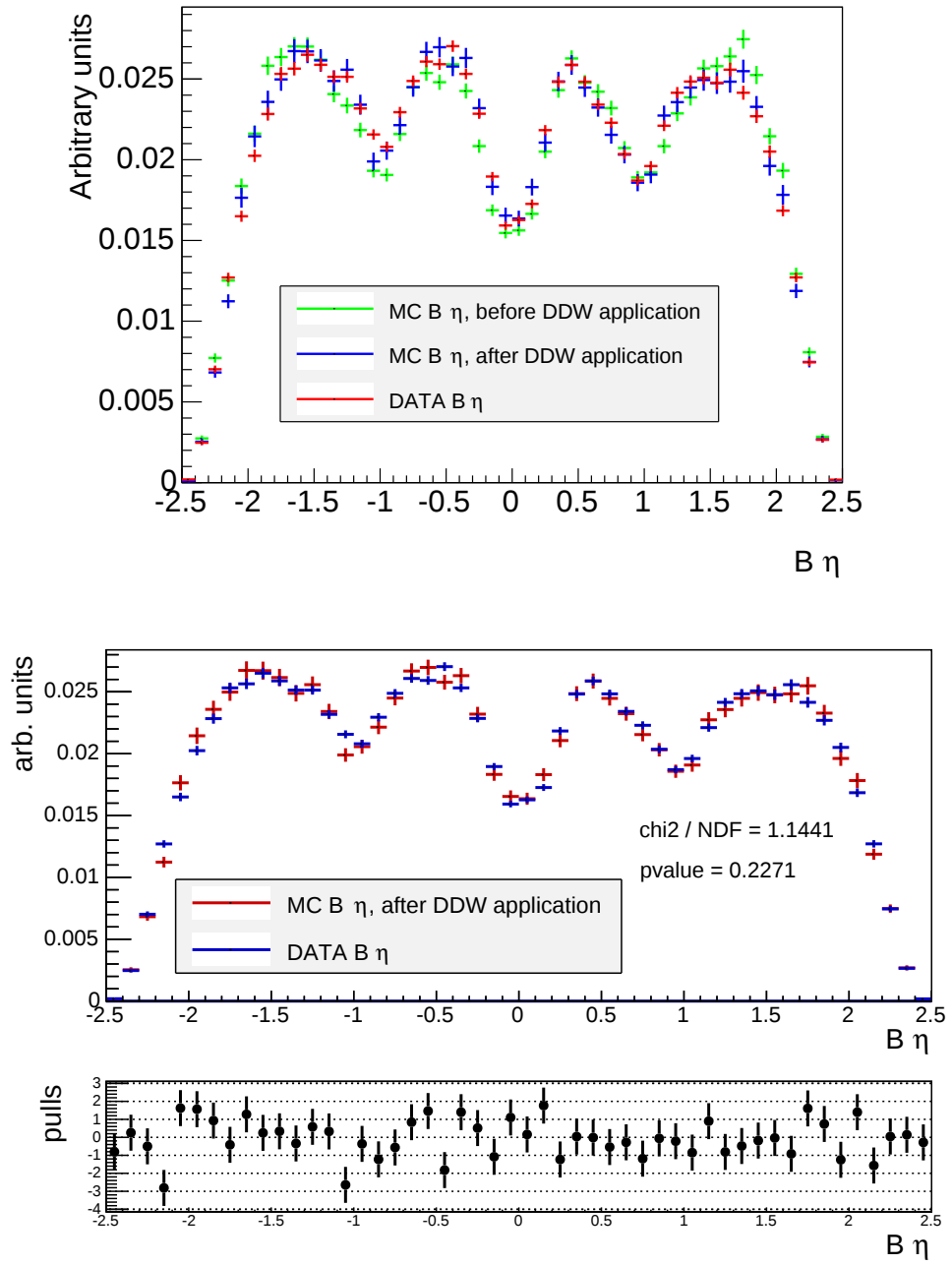


Figure B.16: *Top plot:* comparison of the $\eta(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot:* quantitative comparison between the $\eta(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using even-numbered events from both data and $B^+ \rightarrow J/\psi K^+$ MC, and the remaining events in the two samples are weighted and used for the comparison.

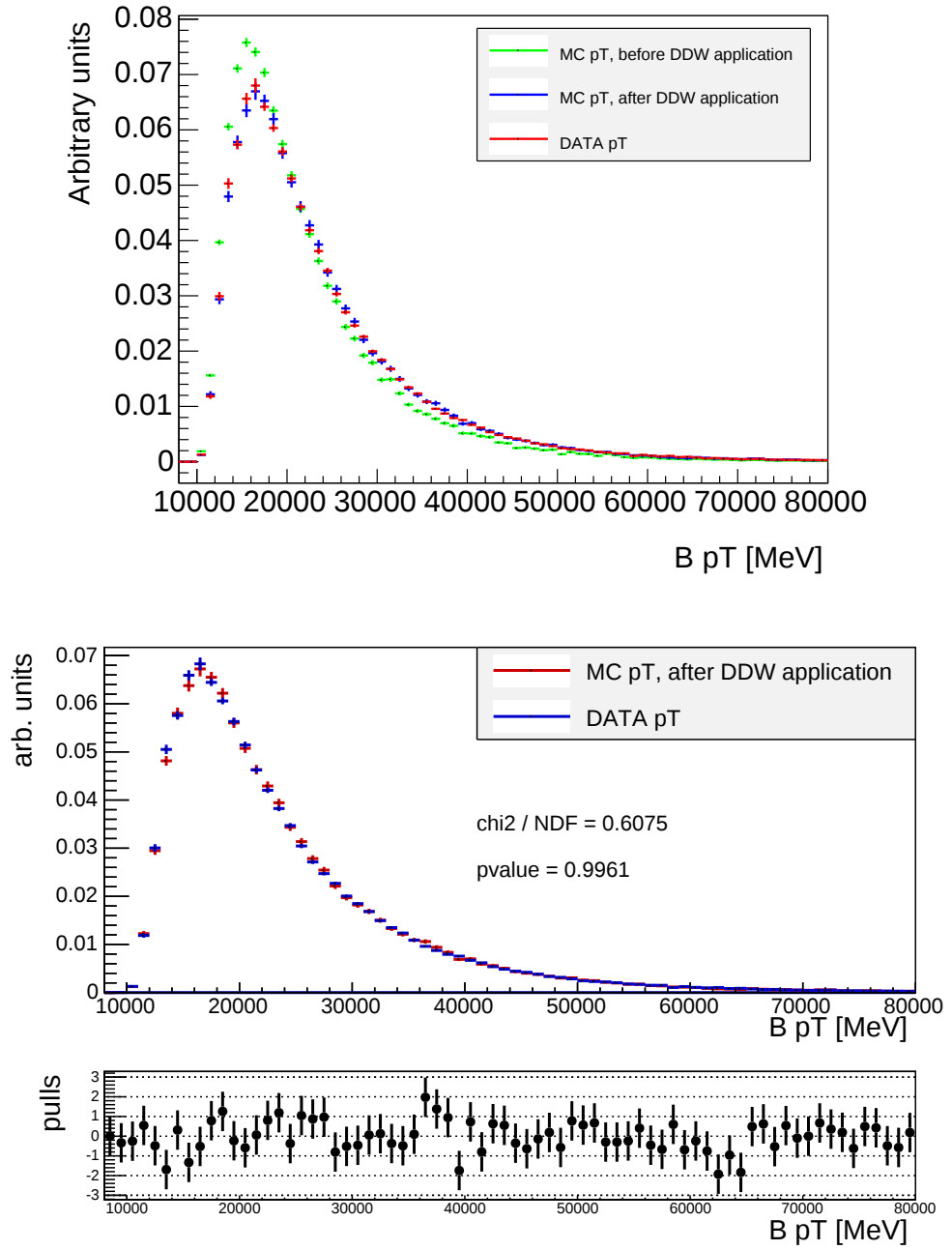


Figure B.17: *Top plot*: comparison of the $p_T(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot*: quantitative comparison between the $p_T(B)$ DDW corrected $B^+ \rightarrow J/\psi K^+$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using even-numbered events from both data and $B^+ \rightarrow J/\psi K^+$ MC, and the remaining events in the two samples are weighted and used for the comparison.

B.3.2 $B_s^0 \rightarrow J/\psi\phi$ DDW consistency

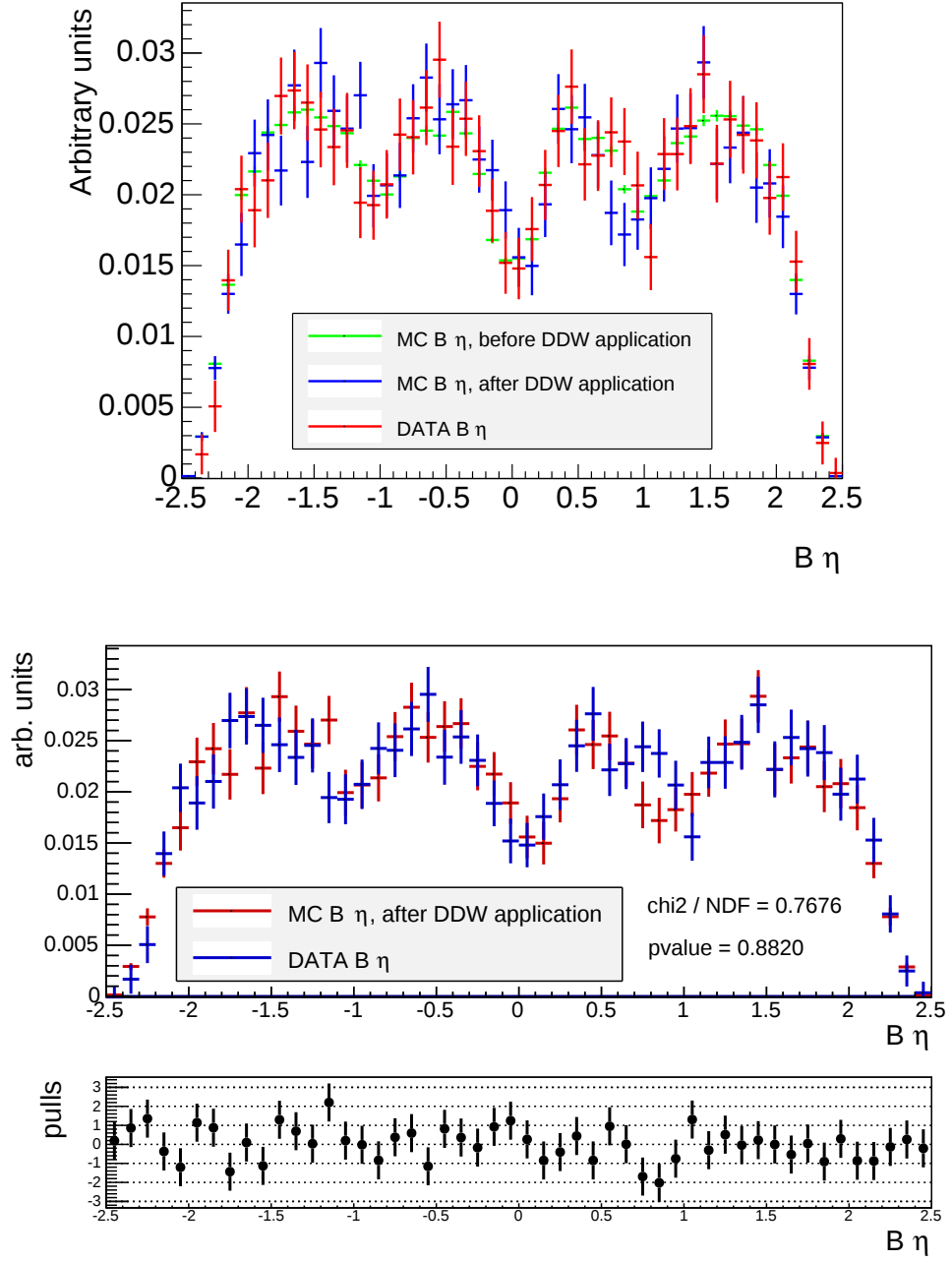


Figure B.18: *Top plot:* comparison of the $\eta(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot:* quantitative comparison between the $\eta(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls. In order to avoid correlations between the distributions, the DDW are calculated using odd-numbered events from both data and $B_s^0 \rightarrow J/\psi\phi$ MC, and the remaining events in the two samples are weighted and used for the comparison.

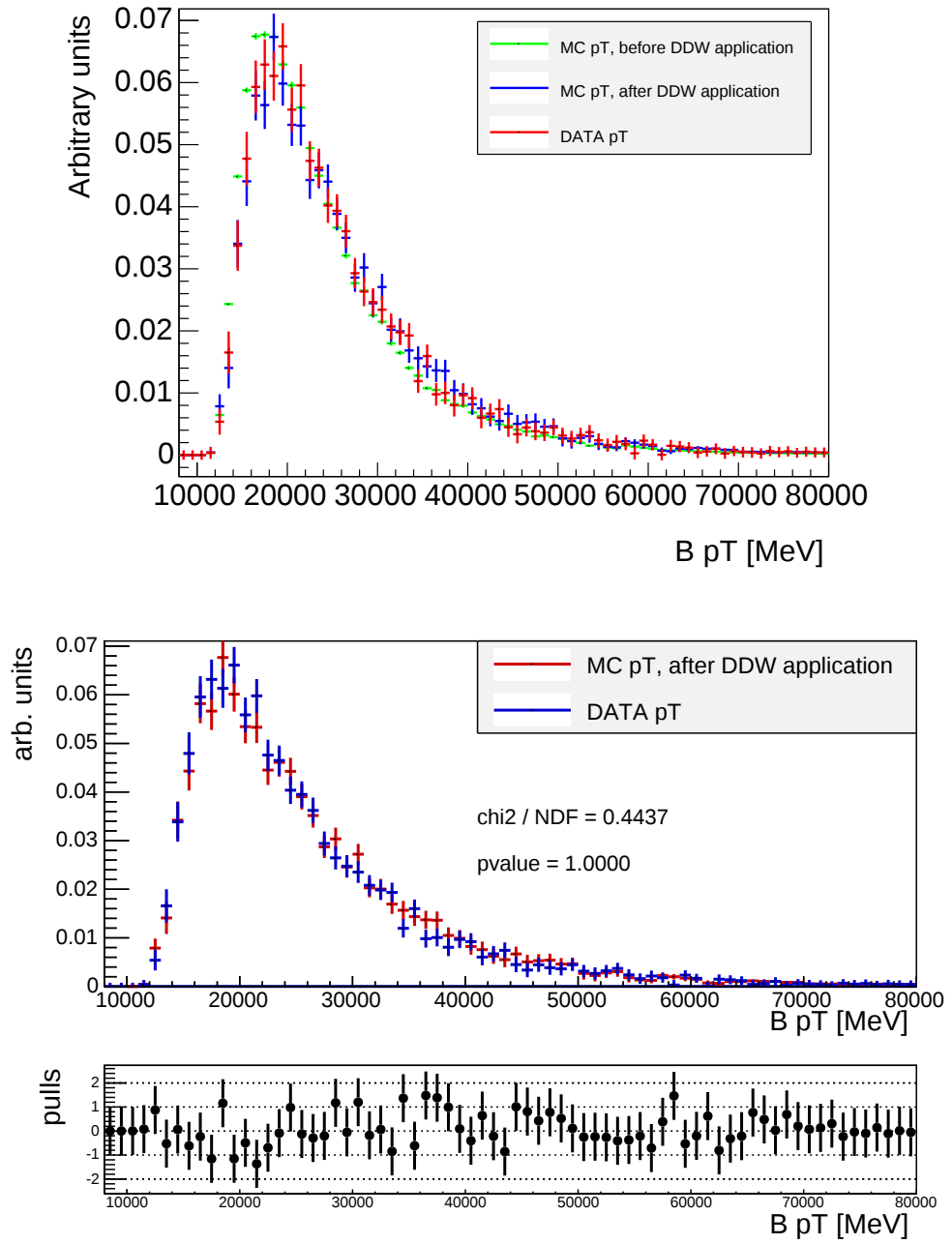


Figure B.19: *Top plot*: comparison of the $p_T(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot*: quantitative comparison between the $p_T(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using odd-numbered events from both data and $B_s^0 \rightarrow J/\psi\phi$ MC, and the remaining events in the two samples are weighted and used for the comparison.

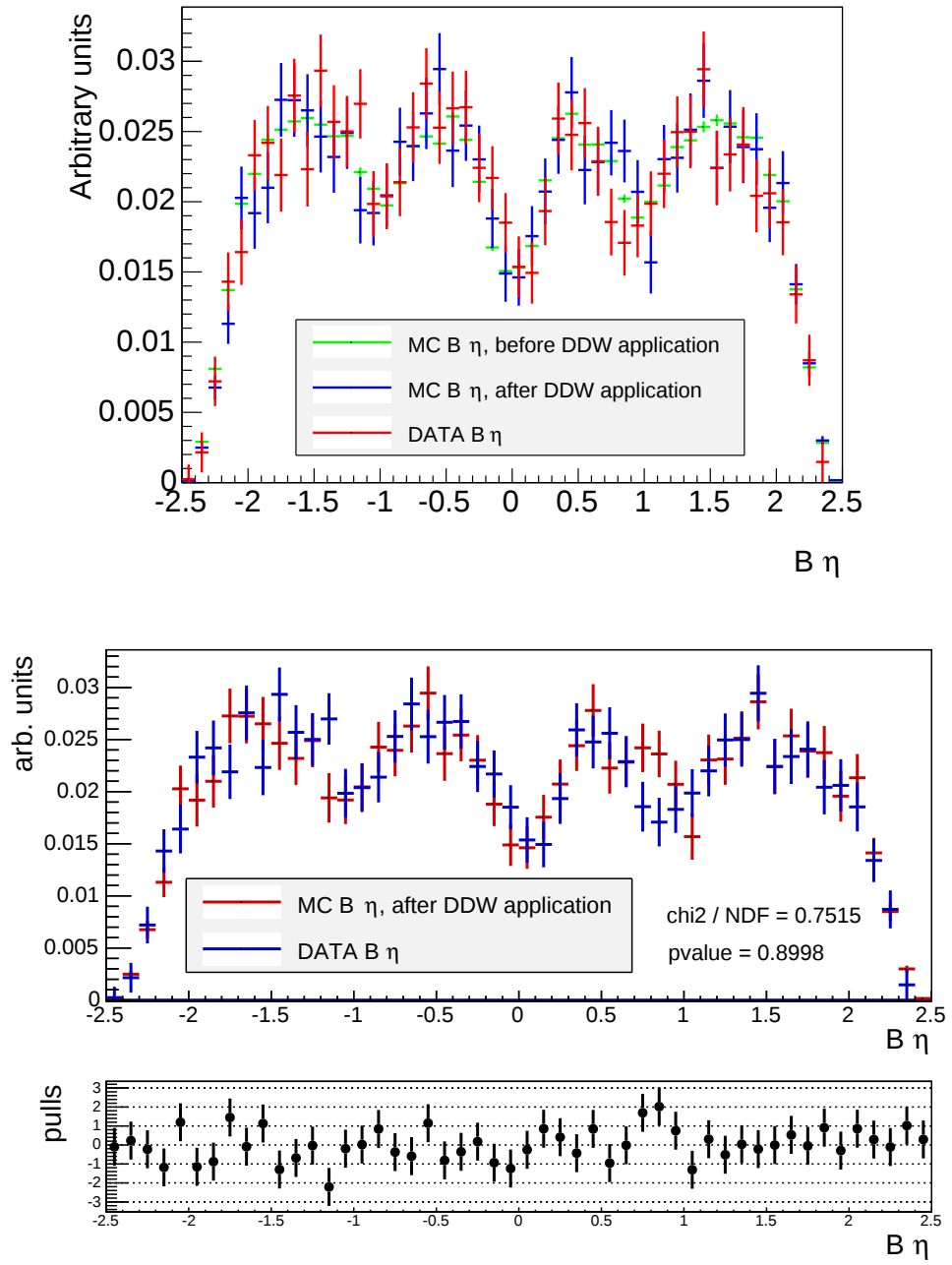


Figure B.20: *Top plot:* comparison of the $\eta(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot:* quantitative comparison between the $\eta(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using even-numbered events from both data and $B_s^0 \rightarrow J/\psi\phi$ MC, and the remaining events in the two samples are weighted and used for the comparison.

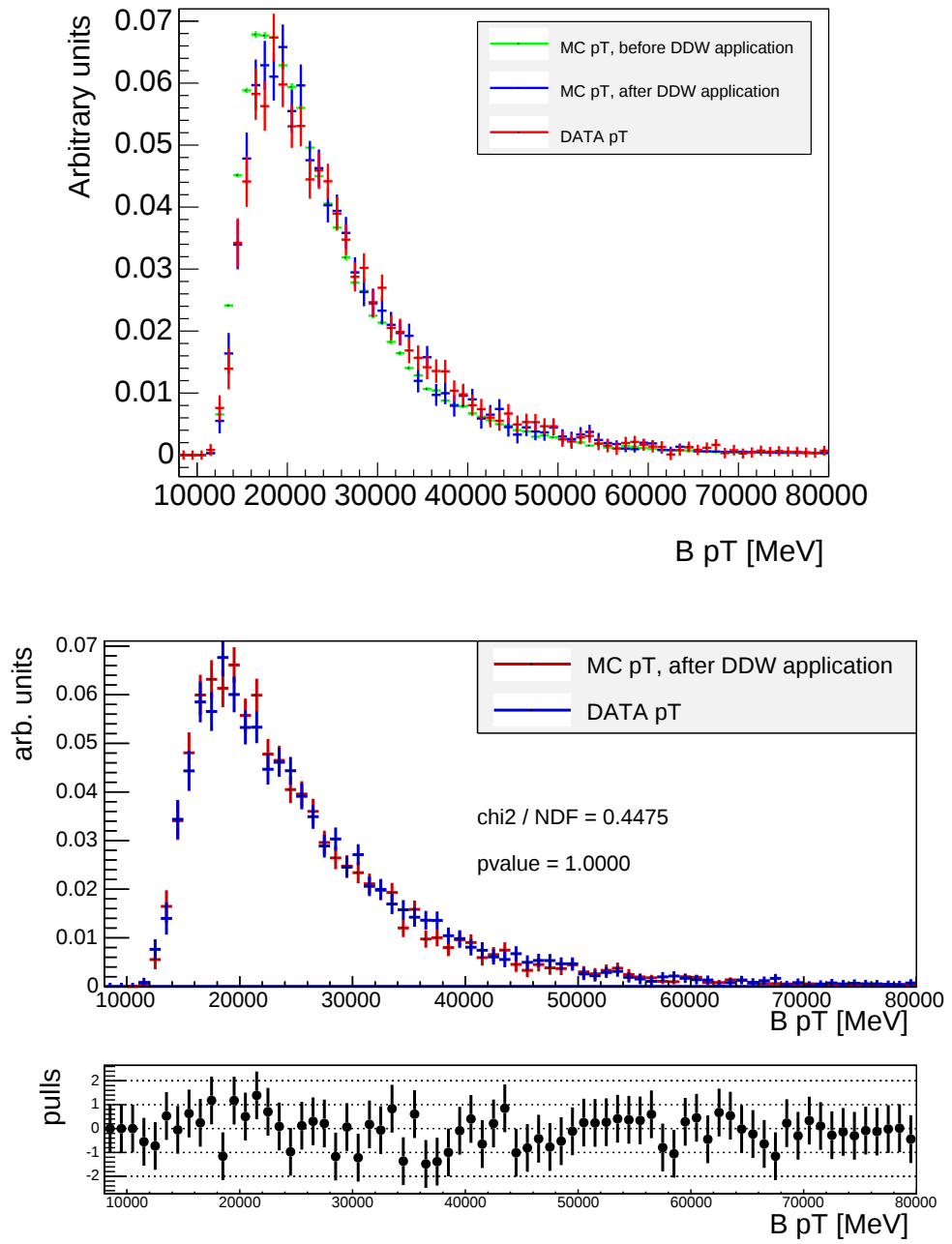


Figure B.21: *Top plot*: comparison of the $p_T(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data and original MC distributions. The muon offline efficiency weights, the muon trigger weights and the QLC are applied to both the original and the corrected MCs. *Bottom plot*: quantitative comparison between the $p_T(B)$ DDW corrected $B_s^0 \rightarrow J/\psi\phi$ MC distribution with the respective data distribution, already shown also in the *top plot*. The bottom panel shows the bin-by-bin pulls.

In order to avoid correlations between the distributions, the DDW are calculated using even-numbered events from both data and $B_s^0 \rightarrow J/\psi\phi$ MC, and the remaining events in the two samples are weighted and used for the comparison.

B.3.3 DDW comparison with Run 1 analysis

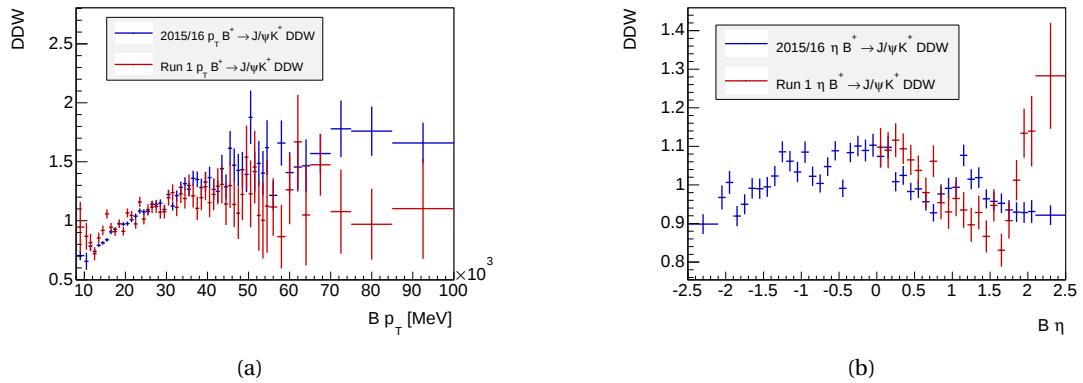


Figure B.22:

As a further cross-check, the $B^+ \rightarrow J/\psi K^+$ DDW are compared to the DDW calculated for the Run 1 analysis, figure B.22. A full compatibility is not expected, as the two analysis are based on data collected at different energy and a different release of the ATLAS software was used to generate and / or reconstruct the events.

The p_T DDW show a good compatibility. The DDW calculated using the 2015/16 data and MC show a more regular behaviour with smaller uncertainties with respect to the Run 1 DDW, thanks to the increased size of the samples used. The η DDW are more complicated to compare: the Run 1 weights were calculated folding the η distribution with respect to 0, as no asymmetry was found in the DDW. This is not the case for the new DDW, which show a clear η asymmetry. The behaviour of the weights as a function of η is also different, this is probably due to improvements in the muon reconstruction, which was highly increased in Run 2 with respect Run 1, especially in the end-cap regions [161].

SUMMARY OF PERSONAL CONTRIBUTIONS



This appendix presents a detailed list of my contributions to the results presented in the thesis.

Studies for trigger improvements

I carried out the studies on the $B \rightarrow hh'$ trigger selection using the FTK full-scan for my qualification task. I had a leading role in the design and optimisation of the trigger selection, as well as in the estimation of the trigger rates (chapter 4).

2015/16 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

The $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis performed on the dataset collected during the first two years of Run 2 has been the core activity of my PhD. I was one of the main analysers, the main PhD student leading the analysis and one of the editors and authors of the internal documentation and publication.

In particular, I had a leading role in the following studies:

- selection of the triggers used in the analysis (section 7.1);
- evaluation of the muon mis-identification fraction (section 7.4);
- validation of the PV-SV association algorithm (section 7.5);
- evaluation and validation of the kinematic corrections for simulations (section 7.6.4 and appendix B);
- design, implementation and validation of the fit used to extract the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ yields (chapter 11);
- statistical extraction of the final results, using different statistical approaches (section 12.1);
- combination of the analysis likelihood with the Run 1 likelihood and statistical extraction of the results (section 12.2).

LHC experiments $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ combination

I have been an active contributor of the ongoing effort to combine the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis results from the ATLAS, CMS and LHCb experiments (section 12.4).

Extrapolations for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

I have been the main analyser and the main editor of the prospects for the $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis at Run 2 and HL-LHC (section 13.1).

I have also been the main analyser in the ATLAS-CMS $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ performance comparison (section 13.2).

Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis

During my first year I contributed to the Run 1 $B_{(s)}^0 \rightarrow \mu^+ \mu^-$ analysis, by co-working on the evaluation of the systematic uncertainties on the fitting procedure with another PhD student (sub-section A.4.1).

GLOSSARY

2HDM	two Higgs doublet model	20
ALICE	A Large Ion Collider Experiment	26
ATLAS	A Toroidal LHC ApparatuS	26
BDT	Boosted Decision Tree	vii
BSM	Beyond the Standard Model	1
CERN	European Organization for Nuclear Research	24
CKKW	Catani-Krauss-Kuhn-Webber	44
CKM	Cabibbo-Kobayashi-Maskawa	10
CMB	Cosmic Microwave Background	12
CMS	Compact Muon Solenoid	26
CSC	Cathode Strip Chamber	34
CTP	Central Trigger Processor	36
DCS	Detector Control System	36
DDW	Data Driven Weights	93
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi	44
DOCA	Distance Of Closest Approach	126
DO	Data Organiser	54
DV	Decay Vertex	126
EF	Event Filter	40
FCNC	Flavour-Changing-Neutral-Current	11
FLIC	FTK-to-Level2 Interface Crate	54
FS	Final State	103
FTK	Fast Tracker	37
GIM	Glashow-Iliopoulos-Maiani	10

HGTD	High-Granularity Timing Detector	40
HL-LHC	High-Luminosity Large Hadron Collider	1
HLT	High Level Trigger	36
HTT	Hardware Track Trigger	40
HW	Hit Warrior	54
IBL	Insertable B-Layer	30
ID	Inner Detector	27
IP	Impact Parameter	126
ITk	Inner Tracker	39
KLN	Kinoshita-Lee-Nauenberg	43
KS	Kolmogorov-Smirnov	56
L1	Level-1 trigger	36
LAr	Liquid Argon	32
LEIR	Low Energy Ion Ring	26
LHCb	Large Hadron Collider beauty	26
LHCf	Large Hadron Collider forward	26
LHC	Large Hadron Collider	v
LHT	Littlest Higgs Model with T-Parity	20
LO	Leading Order	43
LS1	Long Shutdown 1	24
LS2	Long Shutdown 2	25
LUCID-2	Luminosity Cherenkov Integrating Detector-2	88
Linac2	Linear Accelerator 2	25
Linac3	Linear Accelerator 3	26
MC	Monte Carlo	2
MDT	Monitored Drift Tubes	34
ME	Matrix Element	43
MFV	Minimal Flavour Violation	20
MIP	Minimum Ionising Particle	49
MLM	Michelangelo L. Mangano	44
MPI	Multiple Parton Interaction	45
MSSM	Minimal Supersymmetric Standard Model	20
MS	Muon Spectrometer	28

MVA	Multivariate Analysis	58
MoEDAL	Monopole and Exotics Detector at the LHC	26
NLO	Next-to-Leading Order	43
NP	New Physics	1
OPE	Operator Product Expansion	16
PDF	Probability Distribution Function	72
PDG	Particle Data Group	147
PID	Particle IDentification	32
POCA	Point Of Closest Approach	97
PRD	Partially Reconstructed Decays	120
PSB	Proton Synchrotron Booster	25
PS	Parton Shower	43
PU	Pile-Up	45
PV	Primary Vertex	30
QCD	Quantum Chromodynamics	9
QED	Quantum Electrodynamics	5
QFT	Quantum Field Theory	3
QLC	Quark Level Corrections	102
RCF	Rao-Cramer-Frechet	72
RMS	Root Mean Square	82
ROB	Read-Out Buffer	54
ROC	Receiver Operating Characteristic	128
ROD	ReadOut Driver	54
ROS	Read-Out System	36
RPC	Resistive-Plate Chamber	35
RoI	Region of Interest	36
SBW	Sideband Weights	103
SCT	SemiConductor Tracker	30
SM	Standard Model	1
SPS	Super Proton Synchrotron	25
SS+SV	Same Side and Same Vertex	162
SSB	Second Stage Board	54
SV	Secondary Vertex	30

TDAQ	Trigger and Data Acquisition	36
TF	Track Fitter	54
TGC	Thin-Gap Chamber	35
TOTEM	TOTal cross-section, Elastic scattering and diffractive dissociation Measurement ..	26
TRT	Transition Radiation Tracker.....	30
TR	Transition Radiation	32
UE	Underlying Event	45
VEV	Vacuum Expectation Value.....	8
WP	Working Point	50

BIBLIOGRAPHY

- [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv: 1207.7214 [hep-ex].
- [2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: 1207.7235 [hep-ex].
- [3] A. Cerri et al., *Opportunities in Flavour Physics at the HL-LHC and HE-LHC*, (2018), arXiv: 1812.07638 [hep-ph].
- [4] ATLAS Collaboration, *Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector*, JHEP **04** (2019) 098, arXiv: 1812.03017 [hep-ex].
- [5] ‘Prospects for the $\mathcal{B}(B_{(s)}^0 \rightarrow \mu^+ \mu^-)$ measurements with the ATLAS detector in the Run 2 and HL-LHC data campaigns’, tech. rep. ATL-PHYS-PUB-2018-005, CERN, 2018, URL: <https://cds.cern.ch/record/2317211>.
- [6] Sheldon L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579, ISSN: 0029-5582, URL: <http://www.sciencedirect.com/science/article/pii/0029558261904692>.
- [7] Steven Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [8] Abdus Salam, *Gauge unification of fundamental forces*, Rev. Mod. Phys. **52** (3 1980) 525, URL: <https://link.aps.org/doi/10.1103/RevModPhys.52.525>.
- [9] Michael E. Peskin and Daniel V. Schroeder, *An Introduction to quantum field theory*, Addison-Wesley, 1995, URL: <http://www.slac.stanford.edu/~mpeskin/QFT.html>.
- [10] Peter Louis Galison, *Minkowski’s Space-Time: From Visual Thinking to the Absolute World*, Historical Studies in the Physical Sciences **10** (1979) 85, ISSN: 00732672, URL: <http://www.jstor.org/stable/27757388>.
- [11] M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. D **98** (3 2018) 030001, URL: <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.

- [12] G. 't Hooft, *Dimensional regularization and the renormalization group*, Nuclear Physics B **61** (1973) 455, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/0550321373903763>.
- [13] R. P. Feynman, *Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction*, Phys. Rev. **80** (3 1950) 440, URL: <https://link.aps.org/doi/10.1103/PhysRev.80.440>.
- [14] Julian Schwinger, *Quantum Electrodynamics. I. A Covariant Formulation*, Phys. Rev. **74** (10 1948) 1439, URL: <https://link.aps.org/doi/10.1103/PhysRev.74.1439>.
- [15] S. Tomonaga, *On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields**, Progress of Theoretical Physics **1** (1946) 27, ISSN: 0033-068X, eprint: <http://oup.prod.sis.lan/ptp/article-pdf/1/2/27/24027031/1-2-27.pdf>, URL: <https://doi.org/10.1143/PTP.1.27>.
- [16] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- [17] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [18] John Ellis, *Higgs Physics*, (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report, URL: <http://cds.cern.ch/record/1638469>.
- [19] H. Fritzsch, M. Gell-Mann and H. Leutwyler, *Advantages of the color octet gluon picture*, Physics Letters B **47** (1973) 365, ISSN: 0370-2693, URL: <http://www.sciencedirect.com/science/article/pii/0370269373906254>.
- [20] Nicola Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (12 1963) 531, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.10.531>.
- [21] Murray Gell-Mann, *The Eightfold Way: A Theory of strong interaction symmetry*, (1961).
- [22] S. L. Glashow, J. Iliopoulos and L. Maiani, *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. D **2** (7 1970) 1285, URL: <https://link.aps.org/doi/10.1103/PhysRevD.2.1285>.
- [23] J. J. Aubert et al., *Experimental Observation of a Heavy Particle J*, Phys. Rev. Lett. **33** (23 1974) 1404, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1404>.
- [24] J. -E. Augustin et al., *Discovery of a Narrow Resonance in e^+e^- Annihilation*, Phys. Rev. Lett. **33** (23 1974) 1406, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.33.1406>.
- [25] J. H. Christenson et al., *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13** (4 1964) 138, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.138>.

- [26] Makoto Kobayashi and Toshihide Maskawa, *CP-Violation in the Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics **49** (1973) 652, ISSN: 0033-068X, eprint: <http://oup.prod.sis.lan/ptp/article-pdf/49/2/652/5257692/49-2-652.pdf>, URL: <https://doi.org/10.1143/PTP.49.652>.
- [27] Lincoln Wolfenstein, *Parametrization of the Kobayashi-Maskawa Matrix*, Phys. Rev. Lett. **51** (21 1983) 1945, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.51.1945>.
- [28] Y. Amhis et al., *Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2016*, Eur. Phys. J. **C77** (2017) 895, arXiv: 1612.07233 [hep-ex].
- [29] Robert N. Cahn and Gerson Goldhaber, *The Experimental Foundations of Particle Physics*, 2nd ed., Cambridge University Press, 2009.
- [30] Steven Weinberg, *Implications of dynamical symmetry breaking*, Phys. Rev. D **13** (4 1976) 974, URL: <https://link.aps.org/doi/10.1103/PhysRevD.13.974>.
- [31] Gianfranco Bertone, Dan Hooper and Joseph Silk, *Particle dark matter: Evidence, candidates and constraints*, Phys. Rept. **405** (2005) 279, arXiv: hep-ph/0404175 [hep-ph].
- [32] V. C. Rubin and W. K. Ford Jr., *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, Astrophysical Journal **159** (1970) 379.
- [33] D. Larson et al., *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-derived Parameters*, The Astrophysical Journal Supplement Series **192** (2011) 16.
- [34] Planck Collaboration et al., *Planck 2013 results. XVI. Cosmological parameters*, A&A **571** (2014) A16, URL: <https://doi.org/10.1051/0004-6361/201321591>.
- [35] Y. Fukuda et al., *Evidence for Oscillation of Atmospheric Neutrinos*, Phys. Rev. Lett. **81** (8 1998) 1562, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.81.1562>.
- [36] Q. R. Ahmad et al., *Measurement of the Rate of $\nu_e + d \rightarrow p + p + e^-$ Interactions Produced by 8B Solar Neutrinos at the Sudbury Neutrino Observatory*, Phys. Rev. Lett. **87** (7 2001) 071301, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.87.071301>.
- [37] M. B. Gavela et al., *Standard model CP violation and baryon asymmetry*, Mod. Phys. Lett. **A9** (1994) 795, arXiv: hep-ph/9312215 [hep-ph].
- [38] A. Arbey et al., *Supersymmetry confronts $B_s \rightarrow \mu^+ \mu^-$: Present and future status*, Phys. Rev. **D87** (2013) 035026, arXiv: 1212.4887 [hep-ph].
- [39] Isard Dunietz, Robert Fleischer and Ulrich Nierste, *In pursuit of new physics with B_s decays*, Phys. Rev. **D63** (2001) 114015, arXiv: hep-ph/0012219 [hep-ph].
- [40] Kristof De Bruyn et al., *Branching Ratio Measurements of B_s Decays*, Phys. Rev. **D86** (2012) 014027, arXiv: 1204.1735 [hep-ph].

- [41] Gerhard Buchalla, Andrzej J. Buras and Markus E. Lautenbacher, *Weak decays beyond leading logarithms*, Rev. Mod. Phys. **68** (4 1996) 1125, URL: <https://link.aps.org/doi/10.1103/RevModPhys.68.1125>.
- [42] Andrzej J. Buras and Robert Fleischer, *Quark mixing, CP violation and rare decays after the top quark discovery*, Adv. Ser. Direct. High Energy Phys. **15** (1998) 65, [,65(1997)], arXiv: hep-ph/9704376 [hep-ph].
- [43] Andrzej J. Buras, *Operator product expansion, renormalization group and weak decays*, Lect. Notes Phys. **558** (2000) 65, [,65(1999)], arXiv: hep-ph/9901409 [hep-ph].
- [44] Gerhard Buchalla, 'Heavy quark theory', *Heavy flavor physics: Theory and experimental results in heavy quark physics and CP violation. Proceedings, 55th Scottish Universities Summer School in Physics, SUSSP 2001, St. Andrews, UK, August 7-23, 2001*, 2002 57, arXiv: hep-ph/0202092 [hep-ph], URL: <http://weblib.cern.ch/abstract?CERN-TH-2002-018>.
- [45] Robert Fleischer, 'Flavour Physics and CP Violation: Expecting the LHC', *High-energy physics. Proceedings, 4th Latin American CERN-CLAF School, Vina del Mar, Chile, February 18-March 3, 2007*, 2008 105, arXiv: 0802.2882 [hep-ph], URL: <http://weblib.cern.ch/abstract?CERN-PH-TH-2008-034>.
- [46] Athanasios Dedes, Janusz Rosiek and Philip Tanedo, *Complete One-Loop MSSM Predictions for $B \rightarrow \text{lepton lepton}$ at the Tevatron and LHC*, Phys. Rev. **D79** (2009) 055006, arXiv: 0812.4320 [hep-ph].
- [47] T. Blake, T. Gershon and G. Hiller, *Rare b hadron decays at the LHC*, Ann. Rev. Nucl. Part. Sci. **65** (2015) 113, arXiv: 1501.03309 [hep-ex].
- [48] Andrzej J. Buras et al., *Probing new physics with the $B_s \rightarrow \mu^+ \mu^-$ time-dependent rate*, JHEP **07** (2013) 77, arXiv: 1303.3820 [hep-ph].
- [49] Mark Thomson, *Modern Particle Physics*, Cambridge University Press, 2013.
- [50] Edward Witten, *Short distance analysis of weak interactions*, Nuclear Physics B **122** (1977) 109, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/055032137790428X>.
- [51] Kenneth G. Wilson, *Non-Lagrangian Models of Current Algebra*, Phys. Rev. **179** (5 1969) 1499, URL: <https://link.aps.org/doi/10.1103/PhysRev.179.1499>.
- [52] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, Eur. Phys. J. **C77** (2017) 112, arXiv: 1607.00299 [hep-lat].
- [53] Kristof De Bruyn et al., *Probing New Physics via the $B_s^0 \rightarrow \mu^+ \mu^-$ Effective Lifetime*, Phys. Rev. Lett. **109** (2012) 041801, arXiv: 1204.1737 [hep-ph].
- [54] Christoph Bobeth et al., *$B_{s,d} \rightarrow l^+ l^-$ in the Standard Model with Reduced Theoretical Uncertainty*, Phys.Rev.Lett. **112** (2014) 101801, arXiv: 1311.0903 [hep-ph].

- [55] Martin Beneke, Christoph Bobeth and Robert Szafron, *Enhanced electromagnetic correction to the rare B-meson decay $B_{s,d} \rightarrow \mu^+ \mu^-$* , Phys. Rev. Lett. **120** (2018) 011801, arXiv: 1708.09152 [hep-ph].
- [56] G. D'Ambrosio et al., *Minimal flavor violation: An Effective field theory approach*, Nucl. Phys. **B645** (2002) 155, arXiv: hep-ph/0207036 [hep-ph].
- [57] Stephen P. Martin, *A Supersymmetry primer*, (1997) 1, [Adv. Ser. Direct. High Energy Phys.18,1(1998)], arXiv: hep-ph/9709356 [hep-ph].
- [58] Wei-Shu Hou, *Source of CP Violation for the Baryon Asymmetry of the Universe*, Chin. J. Phys. **47** (2009) 134, arXiv: 0803.1234 [hep-ph].
- [59] T. Aaltonen et al., *Search for $B_s^0 \rightarrow \mu^+ \mu^-$ and $B^0 \rightarrow \mu^+ \mu^-$ decays with the full CDF Run II data set*, Phys. Rev. **D87** (2013) 072003, [Erratum: Phys. Rev.D97,no.9,099901(2018)], arXiv: 1301.7048 [hep-ex].
- [60] David M. Straub, *New Physics Searches in Flavour Physics*, Nuovo Cim. **C035N1** (2012) 249, arXiv: 1107.0266 [hep-ph].
- [61] Andrzej J. Buras et al., *The Anatomy of Neutral Scalars with FCNCs in the Flavour Precision Era*, JHEP **06** (2013) 111, arXiv: 1303.3723 [hep-ph].
- [62] Wolfgang Altmannshofer and David M. Straub, *New physics in $b \rightarrow s$ transitions after LHC run 1*, Eur. Phys. J. **C75** (2015) 382, arXiv: 1411.3161 [hep-ph].
- [63] Monika Blanke et al., *FCNC Processes in the Littlest Higgs Model with T-Parity: a 2009 Look*, Acta Phys. Polon. **B41** (2010) 657, arXiv: 0906.5454 [hep-ph].
- [64] G. C. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **516** (2012) 1, arXiv: 1106.0034 [hep-ph].
- [65] W. Buchmuller, R. Ruckl and D. Wyler, *Leptoquarks in Lepton - Quark Collisions*, Phys. Lett. **B191** (1987) 442, [Erratum: Phys. Lett.B448,320(1999)].
- [66] Simone Bifani et al., *Review of Lepton Universality tests in B decays*, J. Phys. **G46** (2019) 023001, arXiv: 1809.06229 [hep-ex].
- [67] T. Faber et al., *A unified leptoquark model confronted with lepton non-universality in B-meson decays*, Phys. Lett. **B787** (2018) 159, arXiv: 1808.05511 [hep-ph].
- [68] Lorenzo Calibbi, Andreas Crivellin and Tianjun Li, *Model of vector leptoquarks in view of the B-physics anomalies*, Phys. Rev. **D98** (2018) 115002, arXiv: 1709.00692 [hep-ph].
- [69] Flavio Archilli, *Observation of B_s^0 to $\mu^+ \mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data. Observation of B_s^0 to $\mu^+ \mu^-$ and first measurement of its effective lifetime with LHCb Run-2 data*, (2017), URL: <https://cds.cern.ch/record/2253469>.
- [70] CMS Collaboration, *Observation of the rare $B_s^0 \rightarrow \mu^+ \mu^-$ decay from the combined analysis of CMS and LHCb data*, Nature **522** (2015) 68, arXiv: 1411.4413 [hep-ex].
- [71] CMS Collaboration, *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ Branching Fraction and Search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS Experiment*, Phys. Rev. Lett. **111** (2013) 101804, arXiv: 1307.5025 [hep-ex].

- [72] *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ Branching Fraction and Search for $B^0 \rightarrow \mu^+ \mu^-$ Decays at the LHCb Experiment*, Phys. Rev. Lett. **111** (10 2013) 101805, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.111.101805>.
- [73] ATLAS Collaboration, *Study of the rare decays of B_s^0 and B^0 into muon pairs from data collected during the LHC Run 1 with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 513, arXiv: 1604.04263 [hep-ex].
- [74] Roel Aaij et al., *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and effective lifetime and search for $B^0 \rightarrow \mu^+ \mu^-$ decays*, Phys. Rev. Lett. **118** (2017) 191801, arXiv: 1703.05747 [hep-ex].
- [75] Oliver Sim Bruning et al., *LHC Design Report*, CERN Yellow Reports: Monographs, CERN, 2004, URL: <https://cds.cern.ch/record/782076>.
- [76] Christiane Lefèvre, 'The CERN accelerator complex. Complexe des accélérateurs du CERN', 2008, URL: <http://cds.cern.ch/record/1260465>.
- [77] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.
- [78] CMS Collaboration, *The CMS experiment at the CERN LHC*, JINST **3** (2008) S08004.
- [79] The LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008) S08005, URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08005>.
- [80] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of Instrumentation **3** (2008) S08002, URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08002>.
- [81] O Adriani et al., *LHCf experiment: Technical Design Report*, Technical Design Report LHCf, CERN, 2006, URL: <http://cds.cern.ch/record/926196>.
- [82] The TOTEM Collaboration, *The TOTEM Experiment at the CERN Large Hadron Collider*, Journal of Instrumentation **3** (2008) S08007, URL: <http://stacks.iop.org/1748-0221/3/i=08/a=S08007>.
- [83] Pinfold, James, *The MoEDAL experiment at the LHC*, EPJ Web Conf. **145** (2017) 12002, URL: <https://doi.org/10.1051/epjconf/201614512002>.
- [84] Werner Herr and B Muratori, *Concept of luminosity*, (2006), URL: <https://cds.cern.ch/record/941318>.
- [85] Apollinari G. et al., *High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1*, CERN Yellow Reports: Monographs, CERN, 2017, URL: <https://cds.cern.ch/record/2284929>.
- [86] John Jowett et al., *The 2015 Heavy-Ion Run of the LHC*, (2016) TUPMW027. 4 p, URL: <https://cds.cern.ch/record/2207379>.
- [87] CERN, 'Record luminosity: well done LHC', URL: <https://home.cern/news/news/accelerators/record-luminosity-well-done-lhc>.

- [88] A. Yamamoto et al., *The ATLAS central solenoid*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **584** (2008) 53, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900207020414>.
- [89] *The CMS magnet project: Technical Design Report*, Technical Design Report CMS, CERN, 1997, URL: <http://cds.cern.ch/record/331056>.
- [90] The ATLAS Collaboration, *The ATLAS Inner Detector commissioning and calibration*, The European Physical Journal C **70** (2010) 787, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-010-1366-7>.
- [91] The ATLAS Collaboration, 'ATLAS Experiment - Photos', URL: <http://atlasexperiment.org/photos/inner-detector-combined.html>.
- [92] The ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, (), URL: <https://cds.cern.ch/record/1291633>.
- [93] G Aad et al., *ATLAS pixel detector electronics and sensors*, Journal of Instrumentation **3** (2008) P07007, URL: <http://stacks.iop.org/1748-0221/3/i=07/a=P07007>.
- [94] A. Ahmad et al., *The silicon microstrip sensors of the ATLAS semiconductor tracker*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **578** (2007) 98, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900207007644>.
- [95] A. Abdesselam et al., *The barrel modules of the ATLAS semiconductor tracker*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **568** (2006) 642, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S016890020601388X>.
- [96] A. Abdesselam et al., *The ATLAS semiconductor tracker end-cap module*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **575** (2007) 353, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900207003270>.
- [97] Ahmet Bingöl, *The ATLAS TRT and its Performance at LHC*, Journal of Physics: Conference Series **347** (2012) 012025, URL: <http://iopscience.iop.org/article/10.1088/1742-6596/347/1/012025/meta>.
- [98] The ATLAS TRT collaboration et al., *The ATLAS TRT Barrel Detector*, Journal of Instrumentation **3** (2008) P02014, URL: <http://stacks.iop.org/1748-0221/3/i=02/a=P02014>.
- [99] The ATLAS TRT collaboration et al., *The ATLAS TRT end-cap detectors*, Journal of Instrumentation **3** (2008) P10003, URL: <http://stacks.iop.org/1748-0221/3/i=10/a=P10003>.
- [100] The ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, (), URL: <https://cds.cern.ch/record/331061>.

- [101] Henric Wilkens and the ATLAS LArg Collaboration, *The ATLAS Liquid Argon calorimeter: An overview*, Journal of Physics: Conference Series **160** (2009) 012043, URL: <http://stacks.iop.org/1742-6596/160/i=1/a=012043>.
- [102] *ATLAS tile calorimeter: Technical Design Report*, Technical Design Report ATLAS, CERN, 1996, URL: <https://cds.cern.ch/record/331062>.
- [103] A Artamonov et al., *The ATLAS Forward Calorimeter*, Journal of Instrumentation **3** (2008) P02010, URL: <http://stacks.iop.org/1748-0221/3/i=02/a=P02010>.
- [104] The ATLAS Collaboration, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, The European Physical Journal C **70** (2010) 723, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-010-1354-y>.
- [105] The ATLAS Collaboration, *ATLAS muon spectrometer: Technical design report*, (), URL: <https://cds.cern.ch/record/331068>.
- [106] Atlas Collaboration, *Performance of the ATLAS trigger system in 2015*, The European Physical Journal C **77** (2017) 317, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-017-4852-3>.
- [107] ‘Technical Design Report for the Phase-I Upgrade of the ATLAS TDAQ System’, tech. rep. CERN-LHCC-2013-018. ATLAS-TDR-023, Final version presented to December 2013 LHCC., 2013, URL: <https://cds.cern.ch/record/1602235>.
- [108] A Barriuso Poy et al., *The detector control system of the ATLAS experiment*, Journal of Instrumentation **3** (2008) P05006, URL: <http://stacks.iop.org/1748-0221/3/i=05/a=P05006>.
- [109] *ATLAS level-1 trigger: Technical Design Report*, Technical Design Report ATLAS, CERN, 1998, URL: <https://cds.cern.ch/record/381429>.
- [110] Will Buttinger, *The ATLAS Level-1 Trigger system*, J. Phys. Conf. Ser. **396** (2012) 012010.
- [111] M. zur Nedden, *The LHC Run 2 ATLAS trigger system: design, performance and plans*, Journal of Instrumentation **12** (2017) C03024, URL: <http://stacks.iop.org/1748-0221/12/i=03/a=C03024>.
- [112] M Shochet et al., ‘Fast TracKer (FTK) Technical Design Report’, tech. rep. CERN-LHCC-2013-007. ATLAS-TDR-021, ATLAS Fast Tracker Technical Design Report, 2013, URL: <https://cds.cern.ch/record/1552953>.
- [113] The HL-LHC project, ‘HL-LHC High Luminosity Large Hadron Collider’, URL: <http://hilumilhc.web.cern.ch/about/hl-lhc-project>.
- [114] Collaboration ATLAS, ‘Letter of Intent for the Phase-II Upgrade of the ATLAS Experiment’, tech. rep. CERN-LHCC-2012-022. LHCC-I-023, Draft version for comments: CERN, 2012, URL: <https://cds.cern.ch/record/1502664>.
- [115] ‘ATLAS Phase-II Upgrade Scoping Document’, tech. rep. CERN-LHCC-2015-020. LHCC-G-166, CERN, 2015, URL: <https://cds.cern.ch/record/2055248>.

- [116] ATLAS Collaboration, ‘Technical Design Report for the ATLAS Inner Tracker Strip Detector’, tech. rep. CERN-LHCC-2017-005. ATLAS-TDR-025, CERN, 2017, URL: <https://cds.cern.ch/record/2257755>.
- [117] ATLAS Collaboration, ‘Technical Design Report for the ATLAS Inner Tracker Pixel Detector’, tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030, CERN, 2017, URL: <https://cds.cern.ch/record/2285585>.
- [118] ‘Expected Tracking Performance of the ATLAS Inner Tracker at the HL-LHC’, tech. rep. ATL-PHYS-PUB-2019-014, CERN, 2019, URL: <https://cds.cern.ch/record/2669540>.
- [119] ATLAS Collaboration, ‘Technical Design Report for the Phase-II Upgrade of the ATLAS LAr Calorimeter’, tech. rep. CERN-LHCC-2017-018. ATLAS-TDR-027, CERN, 2017, URL: <https://cds.cern.ch/record/2285582>.
- [120] ATLAS Collaboration, ‘Technical Design Report for the Phase-II Upgrade of the ATLAS Tile Calorimeter’, tech. rep. CERN-LHCC-2017-019. ATLAS-TDR-028, CERN, 2017, URL: <https://cds.cern.ch/record/2285583>.
- [121] ATLAS Collaboration, ‘Technical Design Report for the Phase-II Upgrade of the ATLAS Muon Spectrometer’, tech. rep. CERN-LHCC-2017-017. ATLAS-TDR-026, CERN, 2017, URL: <https://cds.cern.ch/record/2285580>.
- [122] ATLAS Collaboration, ‘Technical Design Report for the Phase-II Upgrade of the ATLAS TDAQ System’, tech. rep. CERN-LHCC-2017-020. ATLAS-TDR-029, CERN, 2017, URL: <https://cds.cern.ch/record/2285584>.
- [123] ATLAS Collaboration, ‘Technical Proposal: A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade’, tech. rep. CERN-LHCC-2018-023. LHCC-P-012, CERN, 2018, URL: <http://cds.cern.ch/record/2623663>.
- [124] *ATLAS Computing: technical design report*, Technical Design Report ATLAS, CERN, 2005, URL: <https://cds.cern.ch/record/837738>.
- [125] G. Barrand et al., *GAUDI—A software architecture and framework for building HEP data processing applications*, Computer Physics Communications **140** (2001) 45, CHEP2000, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/S0010465501002545>.
- [126] Andy Buckley et al., *General-purpose event generators for LHC physics*, Physics Reports **504** (2011) 145, ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/S0370157311000846>.
- [127] Matt Dobbs and Jørgen Beck Hansen, *The HepMC C++ Monte Carlo event record for High Energy Physics*, Computer Physics Communications **134** (2001) 41, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/S0010465500001892>.

- [128] John C. Collins, Davison E. Soper and George F. Sterman, *Factorization of Hard Processes in QCD*, Adv. Ser. Direct. High Energy Phys. **5** (1989) 1, arXiv: hep-ph/0409313 [hep-ph].
- [129] J M Campbell, J W Huston and W J Stirling, *Hard interactions of quarks and gluons: a primer for LHC physics*, Reports on Progress in Physics **70** (2007) 89, URL: <http://stacks.iop.org/0034-4885/70/i=1/a=R02>.
- [130] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, Phys. Rev. **133** (6B 1964) B1549, URL: <https://link.aps.org/doi/10.1103/PhysRev.133.B1549>.
- [131] Toichiro Kinoshita, *Mass Singularities of Feynman Amplitudes*, Journal of Mathematical Physics **3** (1962) 650, eprint: <https://doi.org/10.1063/1.1724268>, URL: <https://doi.org/10.1063/1.1724268>.
- [132] G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nuclear Physics B **126** (1977) 298, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/0550321377903844>.
- [133] Stefano Catani et al., *QCD Matrix Elements + Parton Showers*, Journal of High Energy Physics **2001** (2001) 063, URL: <http://stacks.iop.org/1126-6708/2001/i=11/a=063>.
- [134] Michelangelo L. Mangano et al., *Matching matrix elements and shower evolution for top-pair production in hadronic collisions*, Journal of High Energy Physics **2007** (2007) 013, URL: <http://stacks.iop.org/1126-6708/2007/i=01/a=013>.
- [135] B.R. Webber, *A QCD model for jet fragmentation including soft gluon interference*, Nuclear Physics B **238** (1984) 492, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/055032138490333X>.
- [136] G. Marchesini and B.R. Webber, *Monte Carlo simulation of general hard processes with coherent QCD radiation*, Nuclear Physics B **310** (1988) 461, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/0550321388900892>.
- [137] B. Andersson et al., *Parton fragmentation and string dynamics*, Physics Reports **97** (1983) 31, ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/0370157383900807>.
- [138] Torbjörn Sjöstrand, *Jet fragmentation of multiparton configurations in a string framework*, Nuclear Physics B **248** (1984) 469, ISSN: 0550-3213, URL: <http://www.sciencedirect.com/science/article/pii/0550321384906072>.
- [139] Bora Isildak, 'Measurement of the differential dijet production cross section in proton-proton collisions at $\sqrt{s} = 7$ tev', PhD thesis: Bogazici U., 2011, arXiv: 1308.6064 [hep-ex].

- [140] David J. Lange, *The EvtGen particle decay simulation package*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **462** (2001) 152, BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900201000894>.
- [141] S. Jadach et al., *The τ decay library TAUOLA, version 2.4*, Computer Physics Communications **76** (1993) 361, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/001046559390061G>.
- [142] E. Barberio and Z. Was, *PHOTOS - a universal Monte Carlo for QED radiative corrections: version 2.0*, Computer Physics Communications **79** (1994) 291, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/0010465594900744>.
- [143] Rick D. Field, *The Underlying event in hard scattering processes*, eConf **C010630** (2001) P501, arXiv: hep-ph/0201192 [hep-ph].
- [144] The ATLAS Collaboration, *Measurement of the underlying event in jet events from 7 TeV proton-proton collisions with the ATLAS detector*, The European Physical Journal C **74** (2014) 2965, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-014-2965-5>.
- [145] Peter Z. Skands, *Tuning Monte Carlo generators: The Perugia tunes*, Phys. Rev. D **82** (7 2010) 074018, URL: <https://link.aps.org/doi/10.1103/PhysRevD.82.074018>.
- [146] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, URL: <https://cds.cern.ch/record/1966419>.
- [147] Torbjorn Sjostrand, Stephen Mrenna and Peter Skands, *PYTHIA 6.4 physics and manual*, Journal of High Energy Physics **2006** (2006) 026, URL: <http://stacks.iop.org/1126-6708/2006/i=05/a=026>.
- [148] Torbjörn Sjöstrand, Stephen Mrenna and Peter Skands, *A brief introduction to PYTHIA 8.1*, Computer Physics Communications **178** (2008) 852, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/S0010465508000441>.
- [149] G. Marchesini et al., *HERWIG 5.1 - a Monte Carlo event generator for simulating hadron emission reactions with interfering gluons*, Computer Physics Communications **67** (1992) 465, ISSN: 0010-4655, URL: <http://www.sciencedirect.com/science/article/pii/0010465592900554>.
- [150] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Zeitschrift für Physik C: Particles and Fields **72** (1996) 637, ISSN: 1431-5858, URL: <https://doi.org/10.1007/s002880050286>.
- [151] T. Gleisberg et al., *Event generation with SHERPA 1.1*, Journal of High Energy Physics **2009** (2009) 007, URL: <http://stacks.iop.org/1126-6708/2009/i=02/a=007>.
- [152] Johan Alwall et al., *MadGraph 5: going beyond*, Journal of High Energy Physics **2011** (2011) 128, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128).

- [153] Simone Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, Journal of High Energy Physics **2010** (2010) 43, ISSN: 1029-8479, URL: [https://doi.org/10.1007/JHEP06\(2010\)043](https://doi.org/10.1007/JHEP06(2010)043).
- [154] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S0168900203013688>.
- [155] Coll ATLAS et al., ‘The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim’, tech. rep. ATL-PHYS-PUB-2010-013, CERN, 2010, URL: <https://cds.cern.ch/record/1300517>.
- [156] M. Aaboud et al., *Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2*, The European Physical Journal C **77** (2017) 673, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-017-5225-7>.
- [157] T Cornelissen et al., *The new ATLAS track reconstruction (NEWT)*, Journal of Physics: Conference Series **119** (2008) 032014, URL: <http://stacks.iop.org/1742-6596/119/i=3/a=032014>.
- [158] S. Boutle et al., *Primary vertex reconstruction at the ATLAS experiment*, J. Phys. Conf. Ser. **898** (2017) 042056.
- [159] ATLAS Collaboration, *Vertex Reconstruction Performance of the ATLAS Detector at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-026, 2015, URL: <https://cds.cern.ch/record/2037717>.
- [160] Wolfgang Waltenberger, Rudolf Frühwirth and Pascal Vanlaer, *Adaptive vertex fitting*, Journal of Physics G: Nuclear and Particle Physics **34** (2007) N343, URL: <http://stacks.iop.org/0954-3899/34/i=12/a=N01>.
- [161] Atlas Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, The European Physical Journal C **76** (2016) 292, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-016-4120-y>.
- [162] The ATLAS Collaboration, ‘ATLAS Muon Combined Performance with the full 2016 dataset’, URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/MUON-2017-001/index.html>.
- [163] The ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data*, The European Physical Journal C **74** (2014) 3071, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-014-3071-4>.
- [164] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2016-015, 2016, URL: <https://cds.cern.ch/record/2203514>.
- [165] ATLAS Collaboration, *Photon identification in 2015 ATLAS data*, ATL-PHYS-PUB-2016-014, 2016, URL: <https://cds.cern.ch/record/2203125>.

- [166] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, URL: <https://cds.cern.ch/record/2157687>.
- [167] Gavin P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-010-1314-6>.
- [168] A. Hrynevich, *ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2*, Journal of Instrumentation **12** (2017) C06038, URL: <http://stacks.iop.org/1748-0221/12/i=06/a=C06038>.
- [169] Matteo Cacciari, Gavin P. Salam and Gregory Soyez, *The anti- k_t jet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, URL: <http://stacks.iop.org/1126-6708/2008/i=04/a=063>.
- [170] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*, ATL-PHYS-PUB-2015-045, 2015, URL: <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-045>.
- [171] C.F. Galea, *Tau Lepton Reconstruction in ATLAS*, Nuclear and Particle Physics Proceedings **287-288** (2017) 111, The 14th International Workshop on Tau Lepton Physics, ISSN: 2405-6014, URL: <http://www.sciencedirect.com/science/article/pii/S2405601417301165>.
- [172] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- [173] A. Annovi et al., *Design of a hardware track finder (Fast TracKer) for the ATLAS trigger*, JINST **9** (2014) C01045.
- [174] A.N. Kolmogorov, *Sulla determinazione empirica di una legge di distribuzione*, G. Ist. Ital. Attuari **4** (1933) 83.
- [175] ATLAS Collaboration, *Trigger monitoring and rate predictions using Enhanced Bias data from the ATLAS Detector at the LHC*, ATL-DAQ-PUB-2016-002, 2016, URL: <https://cds.cern.ch/record/2223498>.
- [176] Georges Aad et al., *Determination of the ratio of b -quark fragmentation fractions f_s/f_d in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Phys. Rev. Lett. **115** (2015) 262001, arXiv: 1507.08925 [hep-ex].
- [177] Aaron Roodman, *Blind analysis in particle physics*, eConf **C030908** (2003) TUIT001, [166(2003)], arXiv: physics/0312102 [physics.data-an].
- [178] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*, Phil. Trans. Roy. Soc. Lond. **A236** (1937) 333.
- [179] Wouter Verkerke and David P. Kirkby, *The RooFit toolkit for data modeling*, eConf **C0303241** (2003) MOLT007, [186(2003)], arXiv: physics/0306116 [physics].

- [180] Rene Brun and Fons Rademakers, *ROOT — An object oriented data analysis framework*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **389** (1997) 81, New Computing Techniques in Physics Research V, ISSN: 0168-9002, URL: <http://www.sciencedirect.com/science/article/pii/S016890029700048X>.
- [181] M. G. Kendall, A. Stuart and J. K. Ord, eds., *Kendall's Advanced Theory of Statistics*, Oxford University Press, Inc., 1987, ISBN: 0-195-20561-8.
- [182] G. Cowan, *Statistical Data Analysis*, Oxford science publications, Clarendon Press, 1998, ISBN: 9780198501558, URL: <https://books.google.co.uk/books?id=ff8ZyW0n1JAC>.
- [183] Eilam Gross, 'LHC statistics for pedestrians', *Statistical issues for LHC physics. Proceedings, Workshop, PHYSTAT-LHC, Geneva, Switzerland, June 27-29, 2007*, 2007 205, URL: <http://cds.cern.ch/record/1099994/files/p205.pdf>.
- [184] Glen Cowan et al., *Asymptotic formulae for likelihood-based tests of new physics*, The European Physical Journal C **71** (2011) 1554, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-011-1554-0>.
- [185] Glen Cowan, 'Statistics for Searches at the LHC', *Proceedings, 69th Scottish Universities Summer School in Physics : LHC Phenomenology (SUSSP69): St.Andrews, Scotland, August 19-September 1, 2012*, 2013 321, arXiv: 1307.2487 [hep-ex].
- [186] Kyle Cranmer, 'Practical Statistics for the LHC', *Proceedings, 2011 European School of High-Energy Physics (ESHEP 2011): Cheile Gradistei, Romania, September 7-20, 2011*, [247(2015)], 2015 267, arXiv: 1503.07622 [physics.data-an].
- [187] Jerzy Neyman and E.S. Pearson, *On the Problem of the Most Efficient Tests of Statistical Hypotheses*, Philosophical Transactions of the Royal Society, A **231** (1933) 289.
- [188] Siegmund Brandt, *Statistical and computational methods in data analysis; 2nd ed.* North-Holland, 1976, URL: <https://cds.cern.ch/record/98616>.
- [189] F. James, *MINUIT Function Minimization and Error Analysis: Reference Manual Version 94.1*, (1994).
- [190] Gary J. Feldman and Robert D. Cousins, *A Unified approach to the classical statistical analysis of small signals*, Phys. Rev. **D57** (1998) 3873, arXiv: physics/9711021 [physics.data-an].
- [191] A.M. Andrew, *Another efficient algorithm for convex hulls in two dimensions*, Information Processing Letters **9** (1979) 216, ISSN: 0020-0190, URL: <http://www.sciencedirect.com/science/article/pii/0020019079900723>.
- [192] M. Aaboud et al., *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, The European Physical Journal C **76** (2016) 653, ISSN: 1434-6052, URL: <https://doi.org/10.1140/epjc/s10052-016-4466-1>.

- [193] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, Journal of Instrumentation **13** (2018) P07017, URL: <http://stacks.iop.org/1748-0221/13/i=07/a=P07017>.
- [194] The ATLAS Collaboration, ‘Luminosity public results Run 2’, URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_AN5.
- [195] ATLAS Collaboration, *Early Inner Detector Tracking Performance in the 2015 Data at $\sqrt{s} = 13$ TeV*, ATL-PHYS-PUB-2015-051, 2015, URL: <https://cds.cern.ch/record/2110140>.
- [196] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv: 1603.05598 [hep-ex].
- [197] Andreas Hoecker et al., *TMVA: Toolkit for Multivariate Data Analysis*, PoS ACAT (2007) 040, arXiv: physics/0703039.
- [198] ‘Study of the rare decays of B0s and B0 into muon pairs from data collected during 2015 and 2016 with the ATLAS detector’, tech. rep. ATLAS-CONF-2018-046, CERN, 2018, URL: <https://cds.cern.ch/record/2639673>.
- [199] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, J. Phys. Conf. Ser. **762** (2016) 012036, arXiv: 1608.05806 [physics.data-an].
- [200] N. L. Johnson, *Systems of Frequency Curves Generated by Methods of Translation*, Biometrika **36** (1949) 149, ISSN: 00063444, URL: <http://www.jstor.org/stable/2332539>.
- [201] M. C. Jones and Arthur Pewsey, *Sinh-arcsinh distributions*, Biometrika **96** (2009) 761, eprint: oup/backfile/content_public/journal/biomet/96/4/10.1093/biomet/asp053/2/asp053.pdf, URL: <http://dx.doi.org/10.1093/biomet/asp053>.
- [202] A. Zannoni, *On the Quantization of the Monoatomic Ideal Gas*, eprint arXiv:cond-mat/9912229 (1999), eprint: cond-mat/9912229.
- [203] *On the theory of quantum mechanics*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **112** (1926) 661, ISSN: 0950-1207, eprint: <http://rspa.royalsocietypublishing.org/content/112/762/661.full.pdf>, URL: <http://rspa.royalsocietypublishing.org/content/112/762/661>.
- [204] Internal ATLAS communication based on <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TrackingCPrecsEarly2018>.
- [205] ‘Validation of the muon momentum corrections for the ATLAS simulation using the $\Upsilon \rightarrow \mu\mu$ channel based on 36.5 fb^{-1} of pp collision data collected in 2015 and 2016’, tech. rep. ATL-PHYS-PUB-2019-018, CERN, 2019, URL: <http://cds.cern.ch/record/2674152>.

- [206] ATLAS and CMS Collaborations, ‘Report on the Physics at the HL-LHC and Perspectives for the HE-LHC’, *HL/HE-LHC Physics Workshop: final jamboree Geneva, CERN, March 1, 2019*, 2019, arXiv: 1902.10229 [hep-ex].
- [207] K. A. Olive et al., *Review of Particle Physics*, Chin. Phys. **C38** (2014) 090001.
- [208] Matteo Cacciari, Mario Greco and Paolo Nason, *The $P(T)$ spectrum in heavy flavor hadroproduction*, JHEP **05** (1998) 007, arXiv: hep-ph/9803400 [hep-ph].
- [209] R. Aaij et al., *Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ Branching Fraction and Effective Lifetime and Search for $B^0 \rightarrow \mu^+ \mu^-$ Decays*, Phys. Rev. Lett. **118** (19 2017) 191801, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.118.191801>.
- [210] ‘Measurement of rare $B \rightarrow \mu^+ \mu^-$ decays with the Phase-2 upgraded CMS detector at the HL-LHC’, tech. rep. CMS-PAS-FTR-18-013, CERN, 2018, URL: <http://cds.cern.ch/record/2650545>.
- [211] Roel Aaij et al., *Physics case for an LHCb Upgrade II - Opportunities in flavour physics, and beyond, in the HL-LHC era*, (2018), arXiv: 1808.08865.
- [212] ‘B Physics analyses for the Phase-II Upgrade Technical Proposal’, tech. rep. CMS-PAS-FTR-14-015, CERN, 2015, URL: <https://cds.cern.ch/record/2036007>.
- [213] Bhaskar Dutta and Yukihiro Mimura, *Enhancement of $\text{Br}(B_d \rightarrow \mu^+ \mu^-)/\text{Br}(B_s \rightarrow \mu^+ \mu^-)$ in supersymmetric unified models*, Phys. Rev. D **91** (9 2015) 095011, URL: <https://link.aps.org/doi/10.1103/PhysRevD.91.095011>.
- [214] M. Oreglia, ‘A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi'$ ’, see Appendix D details on the Crystal Ball function., PhD thesis: SLAC, 1980, URL: <http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-236.html>.
- [215] Luca Martini, ‘Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for the $B^0 \rightarrow \mu^+ \mu^-$ decay with the CMS experiment’, 2014, URL: <https://cds.cern.ch/record/1641195>.
- [216] Y. Amhis et al., *Averages of b -hadron, c -hadron, and τ -lepton properties as of summer 2014*, (2014), arXiv: 1412.7515 [hep-ex].
- [217] Alexander L. Read, *Presentation of search results: The $CL(s)$ technique*, J. Phys. **G28** (2002) 2693, [11(2002)].

This thesis was typeset using the $\text{\LaTeX}$ typesetting system created by Leslie Lamport.
The body text size is set to 11 pt with *Utopia Regular* with *Fourier* font, part of $\text{\TeX}$ Live.
The bibliography was typeset using the ATLAS-paper style.