Probing Quark Hadronization with B mesons at the LHC

Simão Moreira Costa

Thesis to obtain the Master of Science Degree in

Physics Engineering

Supervisor(s): Prof. Nuno Teotónio Viegas Guerreiro Leonardo

Examination Committee

Chairperson: Prof. Dr. Mário João Martins Pimenta
Supervisor: Prof. Dr. Nuno Teotónio Viegas Guerreiro Leonardo
Member of the Committee: Prof. Dr. Alessio Boletti
Member of the Committee: Prof. Dr. Rute Costa Batalha Pedro
Member of the Committee: Prof. Dr. Sandro Fonseca de Souza

October 2023

“Agir, eis a inteligência verdadeira. Serei o que quiser. Mas tenho de querer o que for. O êxito está em ter êxito, e não em ter condições de êxito. Condições de palácio tem qualquer terra larga, mas onde estará o palácio se não o fizerem ali?” — Fernando Pessoa, Livro do Desassossego

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

I express my heartfelt gratitude to LIP (*Laboratório de Instrumentação e Física Experimental de Partículas*) and the LIP-CMS group for providing not only all the necessary tools for the realization of this project, as well as a source of advice and guidance from all the wonderful people that constitute it. Likewise, I express my gratitude to our colleagues of the Heavy Ion physics analysis group (HIN PAG) of the CMS collaboration, for their constant feedback and the discussion they provided during our regular Spectra and Heavy Flavour working group meetings.

I also acknowledge the support of the “Collaboration in the operation and physics data analysis at the CMS experiment at CERN” (CERN/FIS-PAR/0005/2021) grant, funded through Fundação para a Ciência e Tecnologia (FCT).

This work wouldn’t have been possible without the care and supervision of Professor Dr. Nuno Leonardo. Both the scope and quality of this project greatly improved thanks to his eye for detail and expertise in this field, and for that I am very grateful.

Finally, but certainly not less important, I would like to thank all of my friends and family, of which I am fortunate enough to have too many to list. For constant care and love, for all the companionship and the support, for giving me a home, I thank you from the bottom of my heart.

A special thanks must of course go to my Joana, for her constant show of love, patience and how to make prettier graphs. Thank you for always believing in the best of me.

Resumo

Na natureza, os quarks não podem existir por si mesmos. Quando são produzidos, como, por exemplo, em colisões de alta energia, os quarks juntam-se a outros quarks para formar hadrões. Este processo, conhecido como hadronização, é guiado pela força forte (QCD) e ainda não é completamente compreendido. Medições dos mesões B em colisores oferecem sondas únicas do processo de hadronização, pelo qual os quarks individuais formam hadrões de cor neutra. Neste trabalho, os sinais dos mesões B^+ e B_s^0 são estudados e identificados nas colisões do LHC, utilizando métodos avançados de aprendizagem automática. Os mesões B são reconstruídos através dos canais de decaimento $B^+ \to J/\psi\, K^+ \to \mu^+\mu^- K^+$ e $B_s^0 \to J/\psi\, \phi \to \mu^+\mu^- K^+ K^-$, em dados de colisões protão-protão a uma energia de $\sqrt{s} = 5.02$ TeV coletados em 2017 pelo Compact Muon Solenoid (CMS) no LHC.

Esta tese abrange todo o trabalho desenvolvido que engloba uma análise completa. Isto inclui a seleção de eventos, realizada aqui utilizando o XGBoost para a criação de um modelo de árvore de decisão, ajuste de parâmetros das distribuições de massa invariante obtidas e determinação da eficiência do detetor e da seleção, aplicando um método inovador que minimiza a dependência da cinemática de simulações Monte Carlo e remove a contaminação de ruído na região de sinal. Também apresentamos comparações de desempenho entre os modelos BDT e redes neuronais (NN), com estudos que visam investigar a região de baixo p_T, sendo esta mais desafiadora.

Com estes ingredientes, medimos a secção eficaz de produção diferencial dos mesões B em função das variáveis cinemáticas de sinal (momento transverso e rapidez) e variáveis ambientais (multiplicidade de partículas carregadas), obtendo resultados consistentes com as previsões teóricas FONLL e resultados anteriores. Calculamos as razões da secção eficaz entre os dois mesões B, investigando as dependências do rácio entre frações de fragmentação, f_s/f_u, nessas variáveis. Os resultados fornecem indícios de um aumento da produção de B_s^0 em relação ao B^+ para multiplicidade de partículas carregadas elevada, como seria esperado ao considerar a coalescência de quarks como um mecanismo complementar de hadronização.

Palavras-chave: CMS, QCD, Sabor pesado, Fragmentação, Aprendizagem automática, eficiência de detetor

Abstract

In nature, quarks cannot exist by themselves. When produced, like for instance in high energy collisions, quarks join together with other quarks to form hadrons. This process, referred to as hadronization, is driven by the strong force (QCD) and is yet not fully understood. Measurements of B mesons at colliders offer unique probes of the hadronization process by which single quarks form color-neutral hadrons. In this work, B^+ and B_s^0 meson signals are studied and identified in the LHC collisions, employing state-of-the-art machine learning methods. The B mesons are reconstructed through the decay channels $B^+ \to J/\psi\ K^+ \to \mu^+\mu^- K^+$ and $B_s^0 \to J/\psi\ \phi \to \mu^+\mu^- K^+ K^-$, in proton-proton (pp) collision data at $\sqrt{s} = 5.02$ TeV collected in 2017 by the Compact Muon Solenoid experiment (CMS) at the LHC.

This thesis goes over all the developed work that encompasses a full data analysis. This includes the event selection, achieved here by utilizing the XGBoost package for the creation of a boosted decision tree model, parameter fitting of the obtained invariant mass distributions, and detector and selection efficiency determination, in which we apply a novel method that minimizes the reliance on Monte Carlo simulation kinematics and removes background contamination in the signal region. We also present performance comparisons between BDT and neural network (NN) models, along with studies that aim at probing the more challenging low p_T region.

With these ingredients, we measure the B mesons' differential production cross section as function of signal kinematic variables (transverse momentum and rapidity) and environment variables (charged-particle multiplicity), obtaining results consistent with FONLL theoretical predictions and previous results. We calculate the cross section ratios between the two B mesons, investigating dependencies of the fragmentation fraction ratio, f_s/f_u, on these variables. The results provide hints of B_s^0 production enhancement relative to B^+ at high charged-particle multiplicities, as would be expected when considering quark coalescence as a complementary hadronization mechanism.

Keywords: CMS, QCD, Heavy Flavour, Fragmentation, Machine Learning, detector efficiency

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Standard model

Particle physics is the basis of our understanding of the laws of nature. It studies the fundamental constituents of the universe (particles) and the interactions between them (forces). The theory framework that describes the particles and the forces is the Standard Model (SM), which provides a unified picture where the interactions are described by the exchange of particles.

The fundamental constituents of the universe are elementary particles, which, as the name suggests, cannot be further divided. These fundamental building blocks are divided into fermions (quarks and leptons), which generally are matter or anti-matter particles, as well as the bosons, which are generally the force particles that mediate the interactions between the other particles. In Fig. 1.1 we can see the particles belonging to the SM.

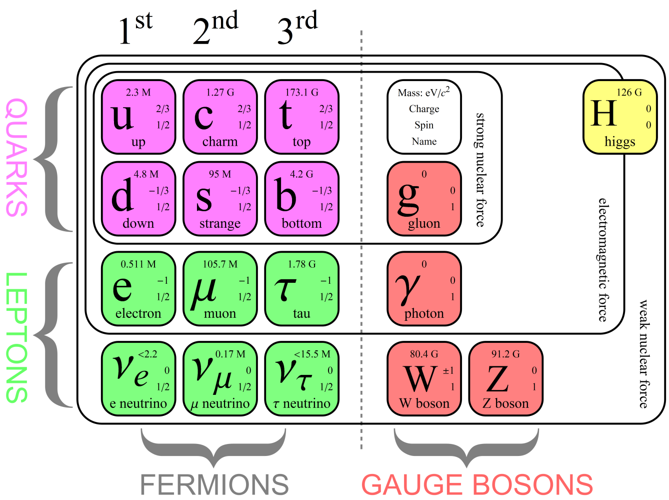

Figure 1.1: Elementary particles in the Standard Model.

In the same vein as the elementary particles, the fundamental forces are the interactions that cannot be reduced to more basic interactions. There are four known forces: electromagnetic, strong, weak and gravitational interactions (the effect of the latter is considered to be negligible at the energy scales here explored and is not further mentioned in this work).

The SM is the most successful theory to date, describing the behaviour of all known fundamental particles with unparalleled precision. It is a non-abelian quantum field theory (QFT) that combines quantum chromodynamics (QCD), based on the $\mathcal{SU}(3)$ gauge group, with the electroweak theory, described by the $\mathcal{SU}(2) \times \mathcal{U}(1)$ gauge group. The $\mathcal{SU}(3)$ component contains a gauge coupling constant g_s and eight gauge bosons known as gluons, $\{G^i\}$, which exclusively act on the colour indices of the quarks fields. QCD is non-chiral, in contrast to electroweak theory, in which the $\mathcal{SU}(2)$ component, with coupling constant g and three gauge bosons, $\{W^i\}$, only acts on left-handed (L) chiral fields, while the $\mathcal{U}(1)$ factor, with coupling constant g' and gauge boson B, acts on both L and R-chiral fields but with differing hypercharges, Y. Because $\mathcal{SU}(2)$ and $\mathcal{SU}(3)$ commute, electroweak interactions do not change colour and QCD interactions do not change flavour.

Within the SM each force is described by a different QFT, each of which has associated a spin-1 force-carrying particle, known has gauge boson. The exchange of this gauge boson is what mediates the corresponding force. Electromagnetism is described by Quantum Electrodynamics (QED), where the interactions between charged particles are mediated by the exchange of virtual photons[1]. This force creates electric and magnetic fields, responsible for the attraction between orbital electrons and atomic nuclei which hold atoms together, as well as chemical bonding and electromagnetic waves, including visible light, and forms the basis for electrical technology. The weak force is mediated by the charged W^+ and W^- bosons and by the neutral Z^0 boson, and acts on the order of the atomic radius, mediating radioactive decay. Finally, the strong force is described by QCD, being mediated by the gluons, and is responsible for quarks binding together to form hadrons, such as protons and neutrons. As a residual effect, it is further responsible for the nuclear force that binds the latter particles to form atomic nuclei.

The properties of the twelve fundamental fermions are divided by the type of interactions they experience, as seen again in Fig. 1.2. All the fundamental fermions experience the weak force and undergo weak interaction. From these, only the neutrinos are electrically neutral, being the only ones that don't participate in electromagnetic interactions. Given that only the quarks and gluons carry colour charges, which are the QCD equivalent to electric charge, only they feel the strong force.

The SM was jump-started by Yang Chen-Ning and Robert Mills, who in 1954 extended the idea of gauge theory for abelian groups to non-abelian groups to explain strong interactions [1]. Future advancements were done by Chien-Shiung Wu, demonstrating in 1957 that parity was not conserved in the weak interaction [2], and by Sheldon Glashow who combined the electromagnetic and the weak interactions [3]. Finally, in 1967 the Higgs mechanism, responsible for giving mass to the elementary particles in the SM, was incorporated into Glashow's electroweak interaction by Steven Weinberg and Abdus Salam [4], giving us the Standard Model we have today.

In 1973, we finally obtained experimental confirmation for the electroweak theory, being the neutral weak currents caused by Z^0 boson exchanges discovered in the Gargamelle experiment at the Proton Synchrotron [5], granting Glashow, Salam and Weinberg the Nobel Prize in Physics in 1979. Later, in 1983, the $W^{\pm}$ and Z^0 bosons were observed by the UA1 and UA2 Collaborations [6, 7], and the ratio of

[1] A virtual particle is a particle that only appears in the intermediary state of particle interactions and is therefore never observed.

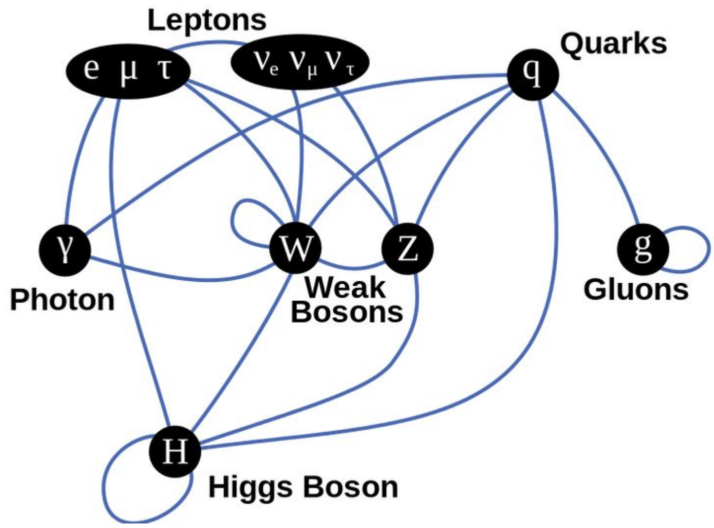

Figure 1.2: Elementary particles and interactions in the Standard Model.

their masses was found to be as the Standard Model predicted [8].

The Large Electron-Positron collider, LEP, was brought online in 1989 to precisely measure the electroweak theory predictions, quickly producing results that suggested that there are only three generations of matter, as shown in Fig. 1.1. The existence of six quark flavours was also established in the meantime, being the top quark the last to be found in 1995 at the Tevatron. The only missing piece of the SM at this point, and a very important one, was the Higgs boson. Only after the construction of the Large Hadron Collider (LHC), did the Compact Muon Solenoid (CMS) and A Toroidal LHC ApparatuS (ATLAS) collaborations announce the observation of a Higgs-like particle that was revealed to be the SM scalar [9].

1.2 Quantum chromodynamics

Quantum Chromodynamics corresponds to a $\mathcal{SU}(3)$ local gauge symmetry of the universe, embedded in the $\mathcal{SU}(3) \times \mathcal{SU}(2) \times \mathcal{U}(1)$ SM gauge group. Because the generators of SU(3) are represented by 3×3 matrices, the quark/anti-quark fields, ψ, must now have three extra degrees of freedom represented by a three-component vector which is termed as "colour", being the three different states labelled as red ($r/\bar{r}$), blue ($b/\bar{b}$) and green ($g/\bar{g}$), whilst the gluons have eight independent colour states, being each represented by a 3×3 unitary matrix. The $\mathcal{SU}(3)$ local phase transformation can then be seen as simply "rotating" the states in the colour space about an axis whose direction is different at every point in space-time. The $\mathcal{SU}(3)$ colour symmetry is exact and QCD is invariant under unitary transformations in colour space.

This theory is described by the Lagrangian density:

$$\mathbf{L}_{QCD} = -\frac{1}{4} F^{\alpha}_{\mu\nu} F^{\mu\nu}_{\alpha} + \bar{\psi}(i\not{D} - M)\psi, \tag{1.1}$$

where ψ and $\bar{\psi}$ are the Dirac spinor and its adjoint, and

$$F^{\alpha}_{\mu\nu} = \partial_\mu A^{\alpha}_{\nu} - \partial_\nu A^{\alpha}_{\mu} + g_s f^{\alpha bc} A^{b}_{\mu} A^{c}_{\nu} \tag{1.2}$$

are the field-strength tensors of the eight gluon fields, $\{A^{\mu}_{\alpha}\}$, $D_\mu = \partial_\mu - ig_s T^{\alpha} A^{\alpha}_{\mu}$ is the covariant derivative, with $\not{D} = \gamma^\mu D_\mu$, where γ^μ are the gamma matrices connecting the spinor representation to the vector representation of the Lorentz group, f_{abc} are the structure constants of the adjoint representation of $\mathcal{SU}(3)$, g_s is the coupling constant of QCD, and M is the quark mass, while $\{T^\alpha\}$ are the generators of the $\mathcal{SU}(3)$ Lie Algebra.

1.2.1 Hadronization

Although there is a lot of experimental evidence for the existence of quarks, they were never seen directly in isolation. This is understood as a result of a long distance, confining property of the strong force. This is explained by the hypothesis of colour confinement, which states that coloured objects are always confined to colour singlet states and that objects with non-zero colour charge can't propagate as free particles. It is to note that there is currently still no analytic proof for this hypothesis on any non-abelian gauge theory.

As a consequence of colour confinement, quarks are not allowed to propagate freely and form jets [2] of colourless particles. The process by which these jets are produced is called hadronization. Figure 1.3 gives a qualitative description of this process. Firstly the quark and antiquark produced in an interaction separate at high velocities, with the colour field restricted to a narrow flux tube with energy density of approximately 1 GeV/fm, and as the quarks separate further the energy stored in this field becomes sufficient to produce new $q\bar{q}$ pairs, and breaking the colour field into smaller "strings" becomes energetically favourable. This process continues until all the quarks and antiquarks have sufficiently low energy to combine to form colourless hadrons.

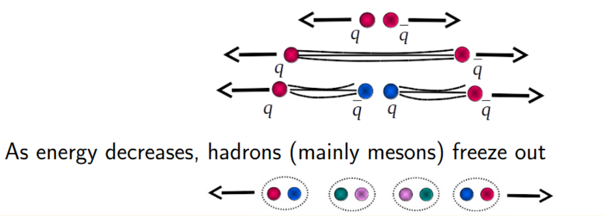

Figure 1.3: Simple depiction of the hadronization process.

The probability of a $\bar{b}$ quark to combine with another quark q into a B_q meson is given by the fragmentation fraction, f_q. These fractions are in turn determined through the hadronization process, being thus senstitive observables. One mechanism for this hadronization process is fragmentation, where showers of partons produced by outgoing quarks form into hadrons. This is described by fragmentation functions, $D^h_i(x, \mu^2)(i = q, \bar{q}, g)$, which are universal functions representing the probability density that an outgoing

[2] A jet is a narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon, that can be detected in a particle experiment.

parton produces a hadron h. Here, x is the fraction of the parton's momentum transferred to the hadron, and μ is a resolution scale known as factorization scale.

A complementary hadronization process is quark coalescence [10]. This process can occur when a quark produced in the collision combines with a nearby quark to form a colour singlet hadron. Coalescence calculations generally require multiple quark wavefunctions to overlap in position and velocity, so the fraction of hadrons produced by this mechanism is expected to increase with the number of quarks produced in the collision.

1.3 B mesons production

B mesons are composed of a bottom antiquark, $\bar{b}$, and another quark [3]. Since the bottom and charm quarks are the heaviest quarks that can comprise observable particles, earning the name of "heavy flavour" quarks, we have that B mesons are some of the heaviest hadrons.

In this work we study two B meson states, B^+ and B_s^0, reconstructed through the decay channels

$$B^+ \to J/\psi\, K^+ \to \mu^+\mu^- K^+\,, \tag{1.3}$$

$$B_s^0 \to J/\psi\, \phi \to \mu^+\mu^- K^+ K^-\,, \tag{1.4}$$

corresponding to the leading-order Feynman diagrams depicted in Fig. 1.4. The J/ψ and ϕ candidates are reconstructed as $J/\psi \to \mu^+\mu^-$ and $\phi \to K^+K^-$ being the total branching fractions given by $\mathcal{B}(1.3) = 6.08 \times 10^{-5}$ and $\mathcal{B}(1.4) = 3.04 \times 10^{-5}$.

Throughout this thesis unless otherwise specified, charged conjugate states are included.

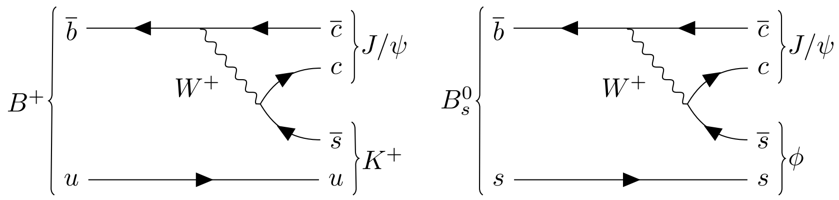

Figure 1.4: The leading-order Feynman diagram of the studied B meson decays, $B^+ \to J/\psi\, K^+$ (left) and $B_s^0 \to J/\psi\, \phi$ (right).

The b-quark hadrons are relatively long-lived, with lifetimes of order 1.5×10^{-12} s. The associated decay modes are suppressed due to the decay $b \to t$ being kinetically forbidden ($m_t \gg m_b$) and due to the $b \to c$ and $b \to u$ decays having a very low probability of happening ($|V_{cb}|^2$ and $|V_{ub}|^2 \ll 1$). When produced in high-energy collisions, this relatively long lifetime, combined with the Lorentz time-dilation factor, means that B hadrons travel on average a few millimetres before decaying. These decays often produce more than one charged particle and because of the relatively large mass of the b-quark, the decay products can be produced at a relatively large angle to the original b-quark direction.

Therefore, there will be a measurable displacement between the point of collision (the primary vertex)

[3]With the exception of the b quark, since the $\bar{b}b$ state is referred to as bottomonia; the top quark is excluded from possible pairings due to its extremely short lifetime.

and the B meson decay point (secondary vertex), allowing us to identify the B hadrons by identifying this secondary vertex from the primary vertex.

The differential production cross section is obtained as

$$\frac{d\sigma}{d\phi} = \frac{1}{\mathcal{L}\mathcal{B}}\frac{1}{\epsilon}\frac{dN_S}{d\phi}\,, \tag{1.5}$$

with respect to a certain variable, ϕ. Here, ϵ denotes the efficiency times acceptance of the detector, $\mathcal{L}$ the integrated luminosity of the data set, $\mathcal{B}$ the branching fraction of the decay chain and N_S the number of signal events. The efficiency ϵ is normally determined from Monte Carlo simulations of the decay process and of the detector, whereas the yield N_S is extracted by fitting the data.

The hadronization fraction ratio

$$\frac{f_s}{f_u}(\phi) = \frac{d\sigma^{B_s^0}}{d\phi}/\frac{d\sigma^{B^+}}{d\phi}\,, \tag{1.6}$$

constitutes an important physics quantity that has to be measured from data. It forms a central ingredient for the study of rare decays at the LHC [11, 12]. Here, the investigation of its dependencies may shed light on the nature of the underlying quark hadronization mechanisms, being in particular sensitive to the presence of complementary effects such as quark coalescence.

1.4 Data selection

1.4.1 Machine Learning

The term “Machine Learning” was coined by Arthur Samuel, an early American leader in the field of computer gaming and artificial intelligence, in 1959 while at IBM [13]. He defined machine learning as “the field of study that gives computers the ability to learn without being explicitly programmed”.

Tom Mitchell [14] provides a simple definition for learning: “A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E”.

Machine learning is an artificial intelligence subfield that involves the creation of algorithms and statistical models that allow computers to improve their performance in tasks through experience. These algorithms and models are intended to learn from data and make predictions or decisions in the absence of explicit instructions.

This field is usually divided into three main types: supervised learning, where our goal is to learn a mapping from a set of inputs x to outputs y given a training set that consists of some set of inputs and their desired outputs (which in principle can be anything, like a letter, a name or a vector); unsupervised learning, where the training set is only consisted of some set of inputs, and we let our model find patterns in the data by itself; and reinforcement learning whereby interacting with its surroundings, an agent learns to make successive decisions. Based on its actions, the agent receives feedback in the form of incentives or punishments. Through trial and error, the agent’s goal is to discover optimal tactics that maximize cumulative rewards over time.

In order for such algorithms to function properly, they must typically be refined many times until they accumulate a comprehensive list of instructions. Algorithms that have been sufficiently trained eventually become “machine learning models”, which are essentially algorithms that have been trained to perform specific tasks such as image sorting, housing price prediction, or chess moves. In some cases, algorithms are layered on top of each other to form complex networks that can perform increasingly complex, nuanced tasks such as text generation and chatbot powering via a technique known as “deep learning”.

We are interested in using ML algorithms to distinguish our signal from the background in our data set, and so we will focus on supervised learning, namely on classification problems, mapping from a set of inputs x to outputs y, where $y \in 1, 2, ..., C$ being C the number of classes.

1.4.2 Neural Networks

Artificial neural networks (ANN) are a special kind of machine learning structure that is inspired by the biological neural networks that constitute the brains of living things. This structure is based on a collection of connected units which model neurons in a biological brain, each connection works the same as synapses, transmitting a signal to neighbouring neurons, which in turn processes it and outputs another signal to other neurons connected to it.

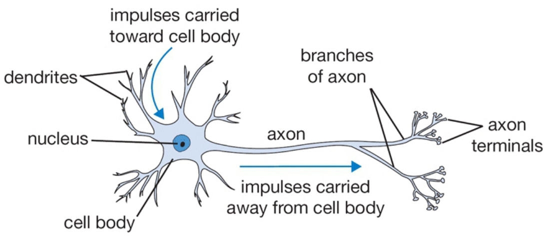

Figure 1.5: A biological neuron is composed of three main body parts: main body (soma), dendrites and axon. The neuron receives input signals from the dendrites, and after processing them outputs signals through the axon. The structure of neural networks is analogous, being the equivalents to the soma, dendrites and axon the hidden layers, input layer and output layer, respectively.

In practice how this works is that each of these artificial neurons functions is a different model, which are usually grouped into layers, being the first layer the so aptly called input layer, and the final one the output layer, being all the intermediary ones called hidden layers. When applying these we are simply changing the representation of our data, and we can go to a representation of higher or lower order than our original one depending on how we define these layers, dubbing as hidden units the dimension of a hidden layer. We now apply our training set to one of these layers to output some prediction values, feeding the output of this layer into the next one and so on and so forth until we get to the final output.

Let us now look at this in a more mathematical sense. Be x our input set and y the output set of our model. We define our model as $y = f(x)$ for some function f that depends on our input and on a set of parameters that we call weights. The goal of our model is to learn these weights in a way that it can predict correctly the output of any data set we give it, and to do this we define a Loss function, $L(y, \hat{y})$, that depends on the correct output, y, and the output predicted by our model, $\hat{y}$. This function will depend on our model and will represent how "close" the output of our model is with reality during training, and therefore our task now becomes simply choosing the weights of f as to minimize this Loss function which can be done by various different methods like for example gradient descent. In the case where we have multiple layers in our network, then we need to perform this minimization with regard to weights of multiple models, which is typically done using backpropagation. After this training step we can finally use our model to predict in what class our input will be, in our case if it will be a signal or background event. The model will output a value which we call score to each of the possible classes and using a softmax transformation [4] we can turn these scores into probabilities, $P(y = c|x)$, and so our predicted value will be the one with the highest probability.

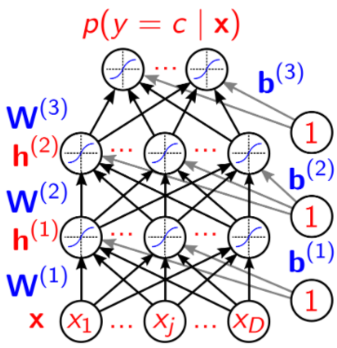

Figure 1.6: Example of a neural network for the simple case of a linear model, $y = Wx + b$. Our inputs are x being W^i and b^i the weights of the respective layer h^i, having the output as a probability of x being of class c $p(y = c|x)$ for every possible class.

This method massively increases the expressiveness of our algorithm because now we stop being confined to the limitations of the model we choose. As an example, take that our model is a simple linear transformation from the inputs to the outputs, $y = Wx + b$, where W and b are the weights of the model, and we wish to use it to model some non-linear data. This is obviously impossible if we just directly apply the model; however, if we use an ANN then we can insert something called an "activation function" at the end of each layer, which transforms the outputs of the given layer before going into the next one. If we simply choose this activation function to be a non-linear function, then each layer will compute a non-linear representation of their inputs and propagate it forward, yielding by the end a more complex output that will better describe our data. As a matter of fact, it was proven in 1989 by Cybenko that an NN with one hidden layer and a linear output can approximate arbitrarily well any continuous function if utilizing a non-linear activation function, given enough hidden units [15]. When we apply these neural networks with many layers, we say that they are Deep Neural Networks (DNN).

1.4.3 Decision Trees

Another relevant ML structure is the decision tree (Fig. 1.7). A decision tree, in its most basic form, is a type of flowchart that depicts a clear path to a decision. It is a type of algorithm that uses conditional 'control' statements to classify data. A decision tree begins with a single point and branches out in two

[4] It is to note that applying this transformation isn't necessary in our model, but it is a common practice.

or more directions. Each branch has a different set of possible outcomes, involving a variety of decisions and chance events until a final result is reached.

The single point with which the decision tree starts is called the root node. The internal nodes, also referred to as decision nodes, are fed by the root node's outgoing branches. These can include decision and chance nodes. Both node types conduct evaluations based on the available features to create homogeneous subsets, which are represented by leaf nodes or terminal nodes. All the outcomes within the dataset are represented by the leaf nodes.

By conducting a greedy search to identify the optimal split points within a tree, decision tree learning employs a divide and conquer strategy. This splitting process is then repeated top-down and recursively until all or the majority of records have been classified under specific class labels. The decision tree's complexity determines whether all data points are classified as homogeneous sets. Smaller trees can more easily achieve pure leaf nodes, or data points in a single class. However, as a tree grows in size, maintaining this purity becomes increasingly difficult, and this usually results in too little data falling within a given subtree. Occam's Razor defines parsimony as "entities should not be multiplied beyond necessity". To put it another way, decision trees should only add complexity when absolutely necessary, because the simplest explanation is often the best. Pruning is commonly used to reduce complexity and prevent overfitting; this is a process that removes branches that split on features of low importance. The model's fit can then be evaluated using the cross-validation procedure. Another method for decision trees to maintain accuracy is to form an ensemble using a random forest algorithm; this classifier predicts more accurate results, particularly when the individual trees are uncorrelated with each other.

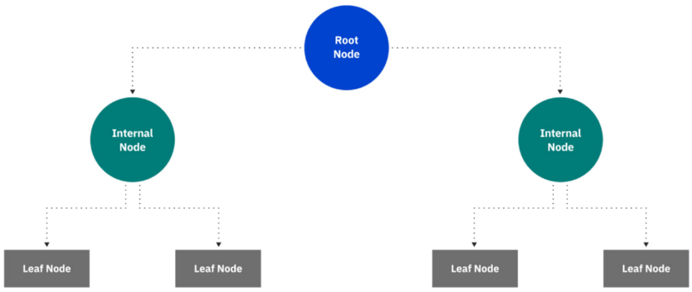

Figure 1.7: Example of the structure of a decision tree.

We can make an analogy between Neural Networks and decision trees [16] by noting that the process of a decision tree is similar to that of a neural network, where instead of simply propagating the signal to the next nodes, the model chooses the nodes to which the signal is propagated to based on the last node's output.

A particularly important type of decision trees are boosted decision trees, which consist of sequentially trained decision trees: that is, we make it so that each decision tree learns from the mistakes of the previous one. Individually, each of these decision trees are weak prediction models that perform only slightly better than random chance, but if we sequentially add up these weak learners, we can filter out the correct observations of a learner at every step, providing us with a much stronger model as a whole.

1.5 State of the art

B mesons are premier choices for probing the hadronization process. In contrast to light quarks, their large mass suppresses their production via non-perturbative processes. Since there is no b content in the valence quark distribution of the incoming beam particles, then the production of $b\bar{b}$ pairs at hadron colliders is dominated by hard parton-parton interactions in the initial stages of the collisions, and is well described by perturbative QCD calculations.

Quantities like the b-hadron production fractions, lifetimes, and neutral B-meson oscillations were studied in the 1990s at LEP and SLC, using e^+e^- colliders at $\sqrt{s} = m_Z$, at DORIS II and CESR, using e^+e^- colliders at $\sqrt{s} = m_{\Upsilon(4S)}$, as well as at the Tevatron, using a proton-proton (pp) collider at $\sqrt{s} = 1.8$ TeV. After these, precise measurements of the B^0 and B^+ mesons were performed at the asymmetric B factories, KEKB and PEPII, using e^+e^- colliders at $\sqrt{s} = m_{\Upsilon(4S)}$, as well as measurements related to the heavier b hadrons, performed at an upgraded Tevatron now with $\sqrt{s} = 1.96$ TeV [17].

Today the most precise measurements are coming from the LHC experiments, specifically the LHCb, CMS and ATLAS, utilizing pp collisions with a $\sqrt{s}$ value of 5 TeV, 7 TeV, 8 TeV and 13 TeV.

$\bar{b}$ quarks can pair with an s quark to form B_s^0 mesons, or pair with a light d quark to form B^0 mesons. The fractions at which these mesons are produced, f_s and f_d respectively, are determined through the hadronization process. Recent measurements of B hadron production in e^+e^- collisions at the $\Upsilon(5S)$ [18–20] and Z^0 [21–24] resonances give consistent values for the ratio f_s/f_d, which can be interpreted as evidence for the universality of b quark fragmentation as assumed by QCD factorization theorems [25]. Other measurements done at the LHC contradict this hypothesis, however. It was shown that the f_s/f_d ratio has a dependence on the collision center-of-mass energy and on the B meson transverse momentum, p_T. It was also noticed that the fraction of b quarks that hadronize into baryons also varies with p_T [26, 27]. Finally, other recent measurements have shown that charm quark hadronization differs between e^+e^- and pp collisions [28]. The reasons for this are not yet fully established, but it might be explained by hadronization mechanisms beyond fragmentation.

The CMS experiment has fully reconstructed B mesons for the first time in nuclear collisions. Recent CMS measurements have shown hints that B_s^0 mesons are enhanced relative to B^0 mesons in PbPb collisions [29]. In pp collisions, CMS has also demonstrated that the efficiency-corrected yield ratio, $\mathcal{R}_s$, has no dependence on rapidity and has a clear p_T dependence at low p_T [30], displayed in Figs. 1.9 and 1.10, respectively. Also, recently, LHCb measurements have shown evidence of such enhanced production of B_s^0 relative to B^0 mesons with high charged particle multiplicity [31], shown in Fig. 1.8. For collisions with relatively low charged-particle multiplicity, and for B mesons with $p_T > 6$ GeV/c, the rate of production is consistent with the results mentioned above in e^+e^- collisions. This makes these results consistent with the expectations based on the quark coalescence as an additional hadronization process, and could indicate that the b quark interactions with the local hadronic environment influence the hadronization process.

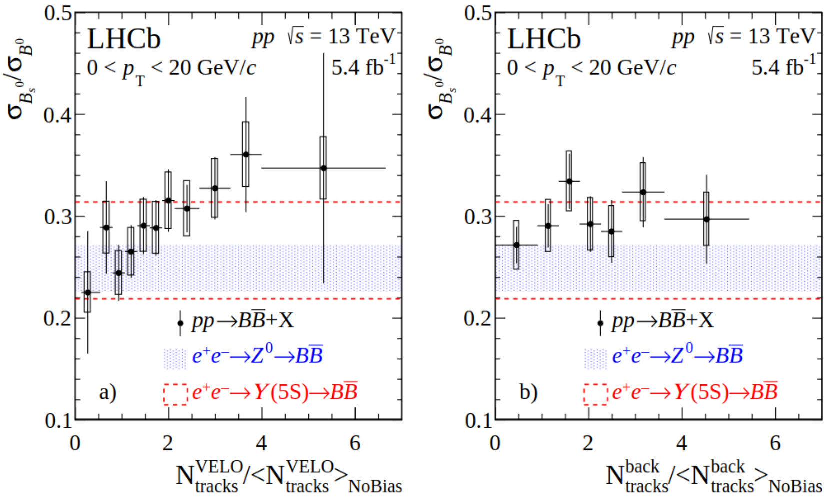

Figure 1.8: Ratio of cross sections $\sigma_{B_s^0}/\sigma_{B^0}$ versus the normalized multiplicity measured by LHCb, for (left) all VELO tracks, and (right) backward VELO tracks. The VELO tracks are tracks coming from the Vertex Locator (VELO), a silicon microstrip detector that surrounds the proton-proton interaction region in the LHCb experiment. The vertical error bars (boxes) represent point-to-point uncorrelated (fully correlated) uncertainties. The horizontal bands show the values measured in e^+e collisions [31].

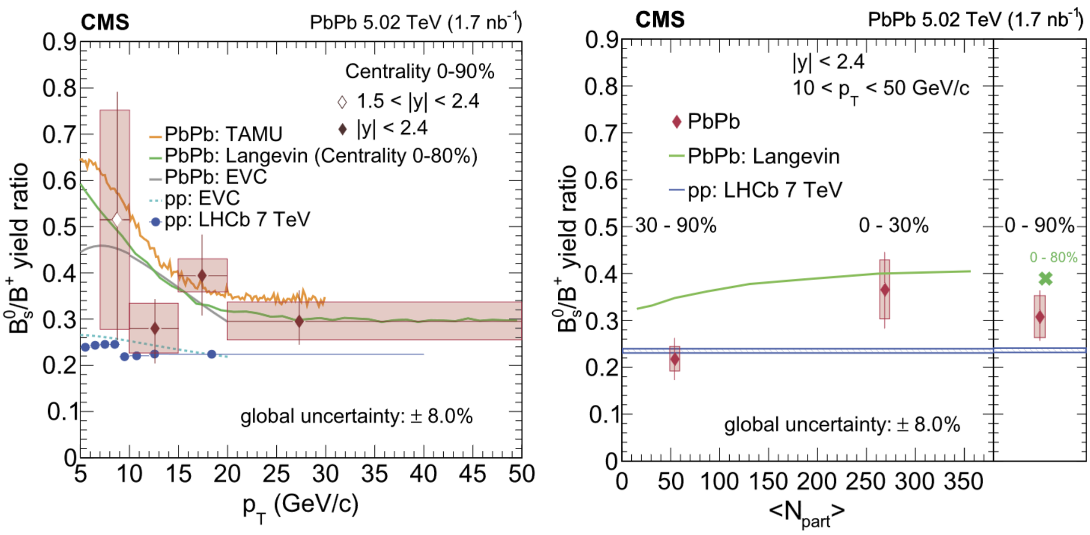

Figure 1.9: The ratio of B_s^0 and B^+ meson production yields, as a function of p_T (left) and of collision centrality (right), measured by CMS in pp collisions. The vertical bars (boxes) correspond to statistical (systematic) uncertainties. Predictions from a Langevin hydrodynamic model [32] and from the TAMU transport model [33] and the EVC quark combination model are overlaid [29].

1.6 Analysis overview

This analysis is based on the $\sqrt{s} = 5.02$ TeV pp dataset collected by CMS in 2017 and performed as function of three variables: transverse momentum, p_T, and rapidity, y, of the B mesons, and the event multiplicity, N_{trk} (defined in Section 3.5).

We explore the LHC data using B mesons, aiming at investigating the interplay of the two competing hadronization mechanisms, fragmentation and coalescence, by measuring the relative production of different B meson species, namely B^+ and B_s^0, through the decay channels shown in Eqs. 1.3 and 1.4

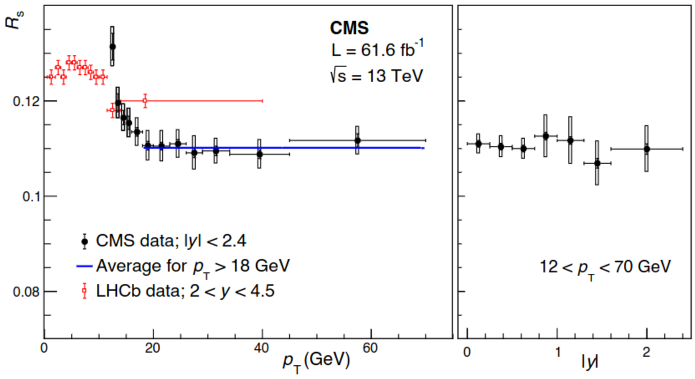

Figure 1.10: Efficiency-corrected yield ratio $\mathcal{R}_s$, as a function of p_T (left) and $|y|$ (right), measured by CMS and LHCb in pp collisions. The blue line represents the average for $p_T > 18$ GeV, obtained from the CMS data [30].

respectively, and its dependence on the relevant kinematic and environment variables.

In Chapter 2 we can find a brief description of the CMS detector, alongside a description of the kinematic variables used. Chapter 3 presents the CMS data and MC samples used and how they are processed, being the selection used explained in Chapter 4 alongside studies for the low p_T region and a comparison of two different selection procedures, one using Boosted Decision Trees (BDT) and another using Neural Networks (NN). Chapters 5, 6 and 7 then discuss the yields, efficiencies, and systematic uncertainties respectively, being the final results and their discussion present in Chapter 8. Additionally, Chapter 6 also presents a novel method for efficiency calculation alongside an analytic and empirical proof. Finally, we present in the appendixes additional plots and tables that could not fit in the main text.

Chapter 2

The CMS detector

At a particle accelerator, the colliding beams produce individual interactions, which we refer to as events. The particle detectors employed at colliders use a wide variety of systems to detect and measure the properties of the particles produced in these high energy collisions, with the aim of reconstructing the primary particles produced in the interaction.

The Large Hadron Collider (LHC) is the world's largest and highest-energy particle collider. It is a 26.7 km circumference ring, 50 to 175 m underground, where two counter-rotating proton or lead beams are accelerated up to 99.99999% the speed of light, making them collide and forming new particles. In this work, we make use of proton beams, in which the protons are squeezed together into bunches that cross each other every 25 nanoseconds. Each of these bunches contain more than 10^{11} protons, occurring about 30 collisions per crossing, with a 40 MHz bunch collision frequency. Due to this, the LHC is able to produce around 600 million collisions per second [34].

The accelerated particles collide at four crossing points along the collider ring. There are nine experiments [35] at the LHC that are distinct and defined by the detectors employed to record these collisions. In addition to CMS, the largest LHC detectors include A Toroidal LHC Apparatus (ATLAS), also a general-purpose detector, the Large Hadron Collider Beauty (LHCb), designed to investigate the charge-parity violation arising in beauty flavoured hadron interactions, and A Large Ion Collider Experiment (ALICE), a detector dedicated to heavy ion collisions for QCD research. The most recent CERN experiment, the Scattering and Neutrino Detector at the LHC (SND@LHC), has been designed and deployed to study neutrinos for the first time in the LHC environment.

This work is developed with the Compact Muon Solenoid (CMS) detector. It is a general-purpose detector at the LHC, capable of triggering on and identifying electrons, muons, photons and both charged and neutral hadrons. A superconducting solenoid with an internal diameter of 6 meters provides a magnetic field of 3.8 T. Within the solenoid volume are the silicon pixel and strip tracker, a crystal electromagnetic calorimeter, and a brass and scintillator hadron calorimeter. Muons are identified with gas-ionization detectors, called muon stations, embedded in the steel flux-return yoke outside the solenoid. Between these, the ones of most importance for this work are the silicon tracker and the muon stations. Events of interest are selected using a two-tiered trigger system, explored in Sec. 2.3. In Fig. 2.1 we show an illustration of a slice of the CMS detector.

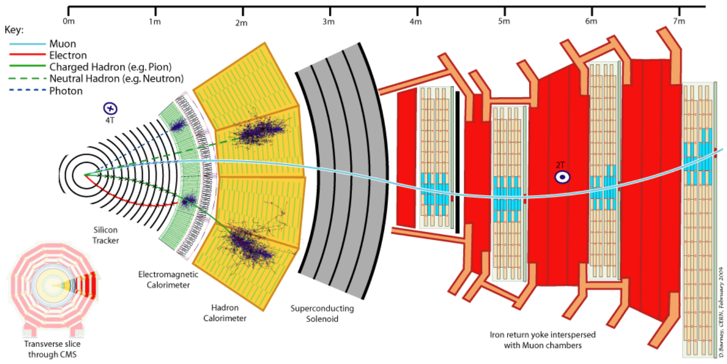

Figure 2.1: Simplified illustration representing a transverse slice of the CMS detector.

2.1 CMS coordinate system and parameters

The CMS experiment employs a Cartesian coordinate system with the detector's centre as its origin. The x-axis is set to point radially inwards toward LHC's centre, the y-axis is set to point vertically upwards, and the z-axis is set to point counter-clockwise along the beam direction. Particles that follow a trajectory along the z-axis also tend to be those that did not participate in a collision, while the hard-scatters tend to produce particles with measurable transverse momentum[1], $p_T = \sqrt{p_x^2 + p_y^2} = \sqrt{|\vec{p}|^2 - p_z^2}$, being therefore this quantity preferentially used over the total tree-momentum, $|\vec{p}|$.

The azimuthal angle $\phi \in [-\pi, \pi]$ is measured from the x-axis in the xy plane, being the radial coordinate in this plane denoted by $R = \sqrt{x^2 + y^2}$. It follows that the polar angle, $\theta \in [0, \pi]$, is measured in relation to the z axis. Finally, we also have another variable of importance, the rapidity, defined as

$$y = \frac{1}{2} \ln \frac{E - p_z}{E + p_z} \tag{2.1}$$

Although this quantity is not Lorentz invariant, its differences are under boosts along z. To measure this quantity, we would need to know the mass of the particle a priori and measure the z component of the momentum, which cannot be done with precision. Alternatively, we utilize pseudorapidity, defined as

$$\eta = \frac{1}{2} \ln \frac{|\vec{p}| - p_z}{|\vec{p}| + p_z} = -\ln tan \frac{\theta}{2} \tag{2.2}$$

being this one equal to rapidity in the relativistic limit $|\vec{p}| \gg m$, where m is the mass of the particle. In Fig. 2.2 we show a transverse section of the CMS detector with the relevant coordinates displayed.

[1] p_T is also invariant with respect to Lorentz boosts along z

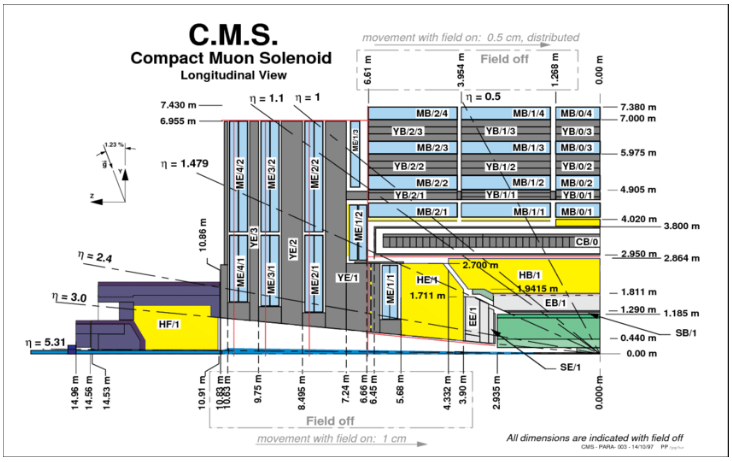

Figure 2.2: Longitudinal cross section view of the CMS detector.

2.2 Silicon tracker and muon stations

To be able to precisely and efficiently measure the trajectories of charged particles, CMS possesses an inner detector that consists of a silicon tracker. This system also allows measuring the primary and secondary vertexes[2] and impact parameters. It is located at the heart of the CMS detector, to better reconstruct the momentum of the particles, and is divided into silicon pixels and silicon strips, being the pixels the closest to the beam and reaching a total of 135 million read-out channels. As particles pass through the silicon tracker, they lose energy to the medium via ionization, forming electron-hole pairs that generate a tiny electric current that can be amplified and detected. As a result, the path of passing charged particles can be reconstructed. Furthermore, because the silicon tracker is surrounded by the magnetic field created by the solenoid, any charged particle will have a bent trajectory based on its positive or negative charge. The particle transverse momentum can thus be easily calculated using the Lorentz force equation, $p_T = qrB$, where q is the particle's electric charge, r is the radius of curvature, and B is the magnetic field generated by the solenoid.

The silicon tracker is able to measure the path of a track with a spatial resolution of 10-12 μm, and can reconstruct high p_T isolated electrons and muons with an efficiency of about 99%, with a fake track[3] rate lower than 1% [36].

Muons are identified in muon stations, forming the detector's outer layers, only accessible to muons and neutrinos, with the latter escaping detection. A muon's path is determined by fitting a curve to the hits recorded by the four muon stations, as well as the hits in the silicon tracker. The muon system is used for muon identification, momentum measurement and triggering, being the steel return yoke plates that interleave the muon stations used as hadron absorbers.

A total of 1400 muon chambers are used to track the path that muons take. Three technologies of gas detectors are employed: drift tubes, cathode strip chambers, and resistive plate chambers. When a muon enters one of these detectors, it knocks out electrons along its path, which then excite other

[2]Being the primary vertex the point where two protons collide, and the secondary vertex the point where a produced long-lived particle decays.
[3]A fake track represents the detector measurement of a track that never existed.

electrons creating an avalanche that gives rise to an electric signal, allowing the muon's trajectory to be reconstructed.

2.3 Trigger

The CMS detector employs a two-tiered trigger system, designed to reduce the amount of collected data by a factor of 10^6, due to the large storage space and bandwidth[4] that would be required to store all the collisions that occur at the interaction point.

The first level (L1), composed of custom hardware processors, uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4 μs. The Calorimeter Trigger provides information regarding muon isolation[5] and compatibility with minimally ionizing particles[6]. The muon trigger system on the other hand identifies and reconstructs tracks in the muon chambers, and estimates their p_T. This information is then used by the Global Muon Trigger, that gathers regional information from the three muon sub-systems. The L1 trigger decision is finally issued by the Global Trigger system, based on the specified L1 requirements that form the trigger menu.

Events that pass the hardware-level trigger are then processed by the ensuing software-level trigger, denoted by High-Level Trigger (HLT), that is executed in a farm of commercial processors running a faster version of the full event reconstruction software of the experiment. It was designed to operate at a maximum of 50 kHz L1 output rate, making use of the full-event information, from all detector subsystems including the silicon tracker, reducing the event rate to around 1 kHz, before data storage.

[4] Maximum rate of data transfer.

[5] A muon is isolated if its energy deposit is below a certain threshold in the calorimeter region from which it emerged.

[6] A minimally ionizing particle, or MIP for short, is a particle whose average energy loss rate approaches the minimum in the Bethe-Bloch curve.

Chapter 3

Data and MC samples

3.1 Data samples

The analysis is carried out on data acquired with the CMS experiment during the 2017 LHC pp run at $\sqrt{s} = 5.02$ TeV. This data is characterized by low pile-up and reduced trigger thresholds, which allow to explore the low p_T region. The trigger algorithm, with reference HLT_HIL1DoubleMu0, requires the presence of at least two muons in the event, with no explicit p_T threshold. The events selected online integrate a primary dataset, that is stored offline, and then undergoes thorough event reconstruction. The total integrated luminosity of the data set is $302.3\,pb^{-1}$.

Following event reconstruction, dedicated selection is applied, to ensure high purity physics objects, namely muons and tracks, that form the basis of the ensuing B meson reconstruction. Events must contain at least one reconstructed primary vertex (PV). The primary vertex is created by two or more associated tracks that must be less than 15 cm along the beam axis and less than 0.15 cm in the transverse plane from the nominal interaction zone. To exclude beam-gas and beam-halo events, which are effectively collisions with beam remnants, pp events are subjected to a beam-background filter.

3.2 MC samples

Various MC samples are used during the analysis to achieve various goals, such as evaluating the presence of physics backgrounds, characterizing fitting models for obtaining signal yields from data, estimating selection efficiency and detector acceptance, and training the Machine Learning algorithms developed for signal-background classification. The MC samples used include signal simulations of the B meson decay channels, specified in Eqs.1.3 and 1.4. Inclusive simulations of background processes that may mimic the signals are also included. For this, B hadron decays to J/ψ mesons are considered.

The MC samples are generated with `PYTHIA` 8 [37] and propagated through the CMS detector with the `GEANT4` package [38]. Only signal events with at least one B meson candidate, with $p_T > 3.0$ GeV/c and $|y| < 2.4$ are saved. `EVTGEN` 1.3.0 [39] is used to model the decay of these B mesons through the considered channels. `PHOTOS` 2.0 [40] is used to simulate final-state photon radiation. In addition to the signal MC samples, an inclusive non-prompt J/ψ MC sample is generated, which contains simulations of various B mesons decaying into J/ψ mesons. These are used to investigate and characterize the peaking backgrounds visible in final state mass spectra.

3.3 Fiducial region

Low momentum muons may not reach the muon sub-detectors in the CMS experiment outermost layers. As a result, they do not leave a track there and cannot be triggered. For this reason, data events in the low p_T and low $|y|$ regions are scarce.

Very few signal candidate events have been reconstructed for B_s^0 meson candidates in the kinematic region limited by $p_T < 10$ GeV/c and $|y| < 1.5$. As a result, the data events contained therein are most likely noise and should be avoided from interfering with the yield extraction step of the analysis. For B^+ meson candidates in the same kinematic region slightly more signal candidate events are reconstructed, particularly towards this region limits. A single fiducial region to carry out the analysis of both B meson data samples is then identified and given by

$$1.5 < |y| < 2.4 : \ p_T > 5 \text{ GeV/c}\,, \qquad 0.0 < |y| < 1.5 : \ p_T > 10 \text{ GeV/c}\,. \tag{3.1}$$

3.4 B candidate reconstruction

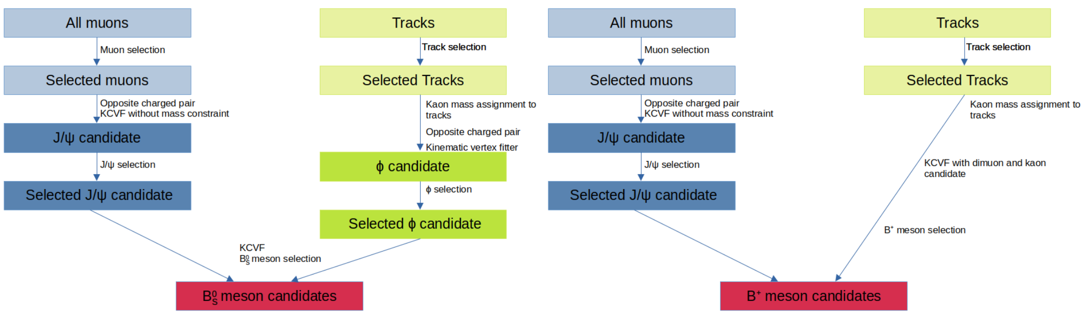

Figure 3.1: Schematic diagrams for the B_s^0 (left) and B^+ (right) mesons reconstruction workflow.

The B mesons' signal reconstruction workflow is depicted schematically in Fig. 3.1. Muon and kaon tracks must meet a number of quality selection criteria in order to have a good signal-to-noise ratio. The J/ψ candidates are reconstructed by vertexing oppositely charged pairs of muon tracks. The ϕ meson is similarly reconstructed by vertexing oppositely charged pairs of tracks. Combining the J/ψ meson candidates with the $K^\pm$ or ϕ meson candidates yields the relevant B meson candidate. A kinematic fit to the $J/\psi K^\pm$ and $J/\psi K^- K^+$ systems is performed, constraining the mass of the dimuon pair to closely resemble the nominal J/ψ world average mass.

The topology of a B hadron decay is illustrated in Fig. 3.2.

The muon quality cuts begin by requiring each candidate to be both a global and a tracker muon. In the first situation, this means matching muon trajectories detected on the muon chambers with tracks left

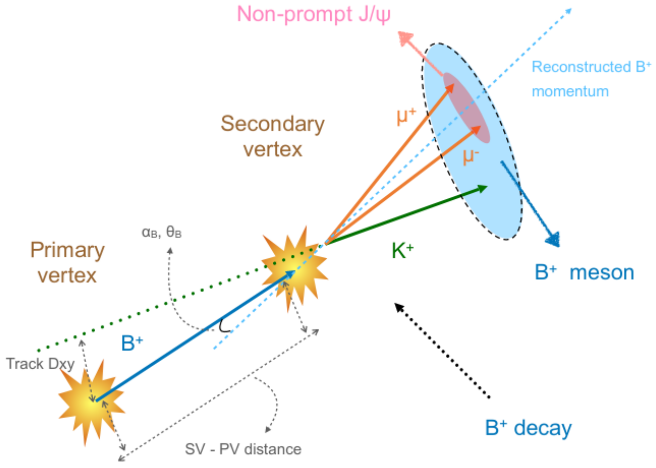

Figure 3.2: Schematic of the B^+ meson decay.

in the silicon detector, while in the second case, it means the reversed method. The muon candidates must also meet the following requirements:

- the transverse distance between track and the primary vertex normalized by its uncertainty (transverse impact parameter), $D_{xy} < 0.3\ cm$,
- the longitudinal distance between track and the primary vertex normalized by its uncertainty (longitudinal impact parameter), $D_z < 20\ cm$,
- tracker tracks must match with at least one muon segment in any station .

Due to detector acceptance, the region of phase space within which the final state particles are detectable, requirements to the muon phase space are also needed. Particles with high pseudorapidity cannot reach the muon stations and may evade detection by the CMS detector. Furthermore, in the case of muons, a minimum p_T is required in order to traverse a sufficient amount of the silicon tracker and reach the muon stations. To provide a good identification efficiency ($> 10\%$), muons must therefore meet the following acceptance criteria:

$$
\begin{aligned}
&p_T^\mu > 3.5 \text{ GeV/c} && \text{for } |\eta^\mu| < 1.2\,,\\
&p_T^\mu > (5.47 - 1.89 \times |\eta^\mu|) \text{ GeV/c} && \text{for } 1.2 \leq |\eta^\mu| < 2.1\,,\\
&p_T^\mu > 1.5 \text{ GeV/c} && \text{for } 2.1 \leq |\eta^\mu| < 2.4\,.
\end{aligned}
$$

In addition to these single-muon quality selection criteria, for a pair of muon candidates, we apply the following additional selection:

- the two muons must have opposite charge ,
- dimuon invariant mass be within 0.15 GeV of the J/ψ meson PDG mass ,
- probability of the two muon tracks to have originated in the same decay vertex be bigger than 1%.

A set of criteria must also be met by the kaon candidate tracks. Their pseudorapidity must be less than 2.4 and their transverse momentum, $p_{T,trk}$, must be greater than 1 GeV/c, with a relative uncertainty, $\frac{\Delta p_{T,trk}}{p_{T,trk}}$, of less than 10%. The number of valid hits recorded on the silicon tracker must be greater than ten, and the χ^2/ndf value must be less than 18%. Pairs of different tracks for the B_s^0 meson must also have an invariant mass within 15 MeV/c^2 of the ϕ meson PDG mass and have opposite charge.

Following the above selection, the resulting data set will be enriched in the specified B meson signals. In addition to genuine B meson signals candidates, the data also contains background events. Indeed, the selected candidates may involve some tracks that are present in the event, but originate from other processes. These events form what is referred to as the combinatorial background.

3.5 Charged-particle multiplicity

In addition to the B meson kinematics, we wish to study the effect of the environment on the hadronization process. The event-by-event charged particle multiplicity, N_{trk}, is the number of charged particles produced in the primary collision. We define it as the number of stable charged particles with $|y| < 2.4$ and $p_T > 0.4$ GeV/c, using tracks coming from the PV and satisfying high-purity criteria [41], also known as primary tracks. Some extra criteria are enforced to ensure the primary tracks and improve the track quality. These include the impact parameter significance to be less than 3, measured with respect to the PV, both along the beam axis $\frac{d_z}{\sigma(d_z)}$ and transverse to the beam direction, $\frac{d_T}{\sigma(d_T)}$. Additionally, the relative p_T uncertainty, $\frac{\sigma(p_T)}{p_T}$, must be less than 10% and finally to ensure high tracking efficiency and to reduce the rate of misconstrued tracks we once again must have that $|y| < 2.4$ and $p_T > 0.4$ GeV/c.

Not all charged tracks produced in a collision are reconstructed, and the efficiency for detection and reconstruction needs to be considered. This can be done by weighting each reconstructed track by the inverse of the efficiency factor,

$$\epsilon_{trk}(p_T, y) = \frac{E}{(1-F)}, \tag{3.2}$$

as a function of the track's pseudoradidity and transverse momentum. This efficiency factor corrects the number of tracks by the track reconstruction efficiency, $E(p_T, y)$, given by the number of reconstructed tracks associated to the simulated tracks over the number of simulated tracks. Also included is the correction due to the fake rate, $F(p_T, y)$, given by the number of reconstructed tracks not associated with simulated particles over the number of reconstructed tracks.

Chapter 4

Data selection and Machine Learning

The pre-selection described in Sec. 3.4 results in a B-enriched data sample, which contains however still a considerable level of background contamination. As a result, the sample still lacks any bold mass resonance that can be fitted to correctly extract the number of observed decays. We shall therefore employ machine learning (ML) techniques to improve our signal-background separation.

For training our ML models, the reconstructed candidates matched to the generated signal in MC samples are used as the signal training sample. The background training sample is on the other hand obtained directly from data, namely it corresponds to the mass sideband region, given by ($0.15\ GeV/c^2 < |M_{B^+} - M_{B^+}^{PDG}| < 0.25\ GeV/c^2$ and $0.25\ GeV/c^2 < |M_{B_s^0} - M_{B_s^0}^{PDG}| < 0.30\ GeV/c^2$).

Following the optimization and training of the ML algorithms, these are employed to improve the signal-to-noise ratio of our data sample. This is achieved by optimizing the cut value on the final score of the ML model. Such optimal choice is taken as the one that maximizes the figure of merit, significance $= \frac{S}{\sqrt{S+B}}$, being S and B the number of signal and background events, in the signal region defined by $|M_B - M_B^{PDG}| < 0.08\ GeV/c^2$.

We have a particular interest in the low p_T region since we expect the coalescence mechanism to be more predominant here; however, this data is also more difficult to access and analyse, as can be inferred from Fig. 4.1. In general, there is an exponentially larger number of particles randomly produced

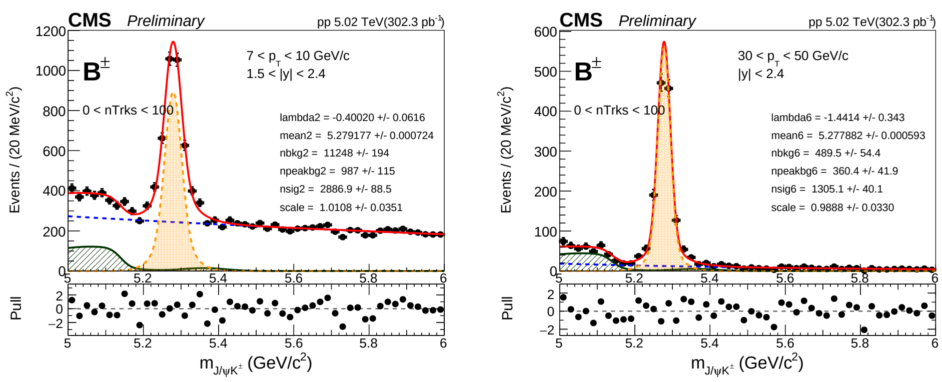

Figure 4.1: Fit to the B^+ meson candidate's invariant mass for the ranges $7 < p_T < 10$ GeV/c (left) and $30 < p_T < 50$ GeV/c (right). We can see that the lower p_T plot has substantially more background.

at lower p_T in a collision. These form a large combinatorial background that buries the signals of interest. Analyses are for this reason generally performed above a given p_T threshold.

Previous selections applied to our data were insufficient to perform proper signal-background separation for $p_T < 5$ GeV/c. For this reason, we here employ more powerful machine learning tools, to both improve the selection of previous studies and to extend the sensitivity of our data.

In this work, we aim at probing dependencies of the production cross-sections versus p_T, y and N_{trk}, which will be achieved by binning the data in these variables. Instead of devising distinct algorithms for different bins, which could result in undesirable asymmetries, the algorithms will be devised and trained in the full samples. The optimization of the working point will then be carried out for maximizing the signal significance in each bin.

4.1 Feature engineering

From our data samples, we must now construct and choose the variables that would help us differentiate between signal and background amongst our B meson candidates. During pre-selection, the muon and track candidates in a given combination are required to approximately originate from the same vertex, and so it follows that the vertex fitting performance, or more aptly named, the fitting χ^2 value, is a naturally good selection feature. Additionally, the decay length of the B meson is measurable, making the distance between the collision PV and the decay SV a feature of interest. The opening angle between daughter muons and tracks can be very small comparing to combinatorial background candidates, also contributing separation power. Since we want to consider the possible relations between decay daughters, we also utilize the track p_T, η and the distance of closest approach between the tracks and the PV.

The topology of B meson decays, depicted in Fig. 3.2, illustrates several of the characteristic features. The full list of features employed in the training process for the B^+ meson is the following:

- dls: Normalized decay length (SV-PV distance), the distance between the primary and the B decay (secondary) vertex (on a 3D plane), normalized by its uncertainty ,
- Balpha: The angle between B^+ meson displacement and B meson momentum in 3D ,
- Btrk1Pt: Track p_T , the transverse momentum of the track ,
- Bchi2cl: The χ^2 probability of the secondary decay vertex fitting ,
- Trk1DCAxy: The transverse impact parameter, D_{xy} ,
- Trk1DCAz: The longitudinal impact parameter, D_z ,
- dls2D: Normalized decay length (SV-PV distance), the distance between the primary and the B decay (secondary) vertex (on a 2D plane), normalized by its uncertainty ,
- Btrk1Eta : Track η, the pseudo-rapidity of the track ,

- Bpt: B candidate p_T, the transverse momentum of the B candidate .

For the B_s^0 meson, in addition to these, we use 4 more variables for the second daughter track of ϕ:

- Btrk2Pt: Track p_T, the transverse momentum of the track ,
- Trk2DCAxy: The transverse impact parameter, D_{xy} ,
- Trk2DCAz: The longitudinal impact parameter, D_z ,
- MassDis: The absolute difference between the invariant mass of the track pair and the ϕ PDG mass .

In Figs. 4.2 and 4.3 we can find, for each variable, the corresponding B^+ and B_s^0 normalized distributions for signal and background, illustrating their individual discriminating power.

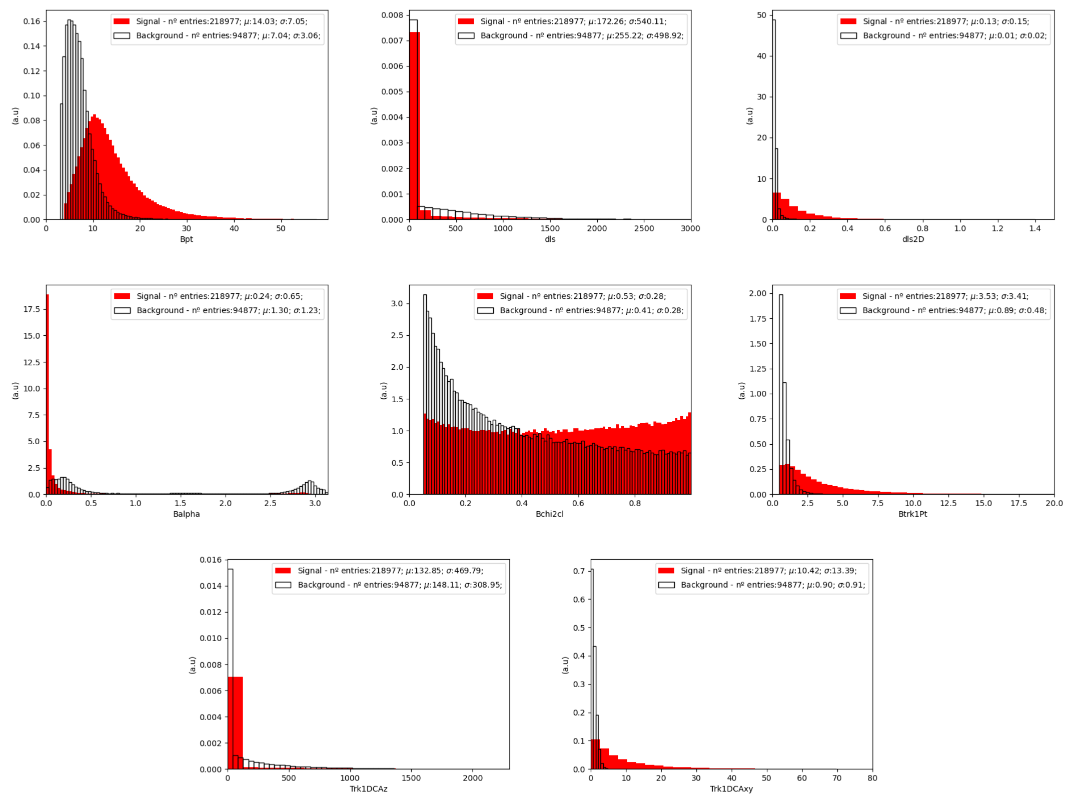

Figure 4.2: Variable distributions for the B^+ meson.

Figs 4.4 and 4.5 display the correlations between the individual variables, for both mesons, evaluated separately in the signal and background samples and in the merged sample. The ML algorithms usefully explore such correlations. The B meson mass is not part of the set of features employed in the training process, but has been included in the correlation study. This is done to inspect and confirm the lack of correlation with the selection features. This is something we wish to happen, as the B candidate invariant mass is the discriminating variable that will be used in our final fits to the data to extract the parameters of interest. Its distribution should accordingly remain undistorted following the selection procedure.

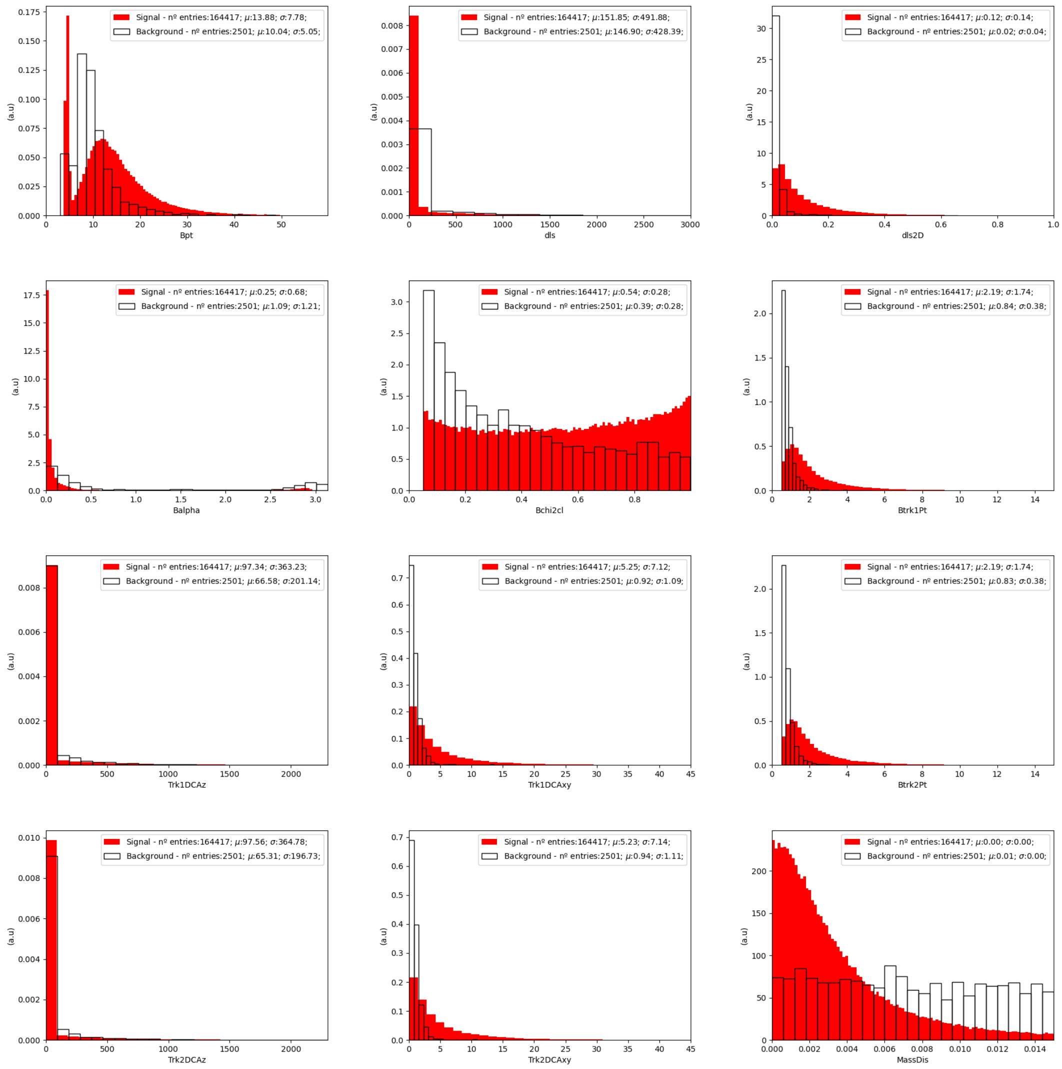

Figure 4.3: Variable distributions for the B_s^0 meson.

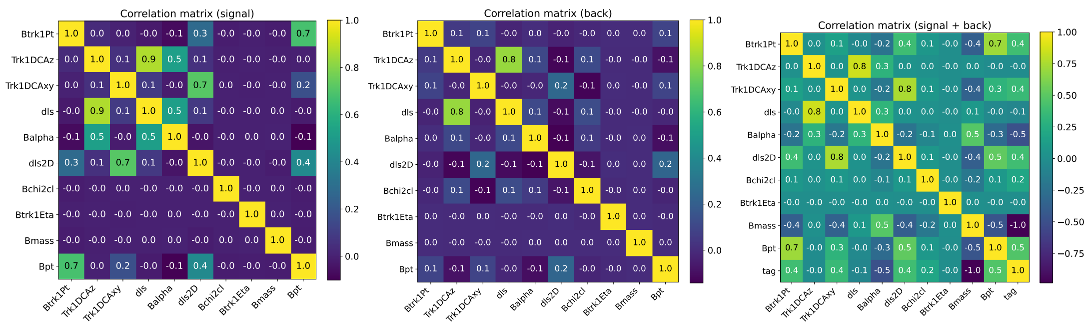

Figure 4.4: Correlation matrices, for the B^+ meson, between our variables on signal (left), background (middle) and on signal+background (right) for $3 < p_T < 60$ GeV/c. This last variation has an extra variable called "tag" which represents the true label of the event (either signal or background).

4.2 ML model optimization

We explore decision trees (BDT) and neural network (NN) approaches. For the training of BDT models, we use the XgBoost package [42], an optimized distributed gradient boosting library that im-

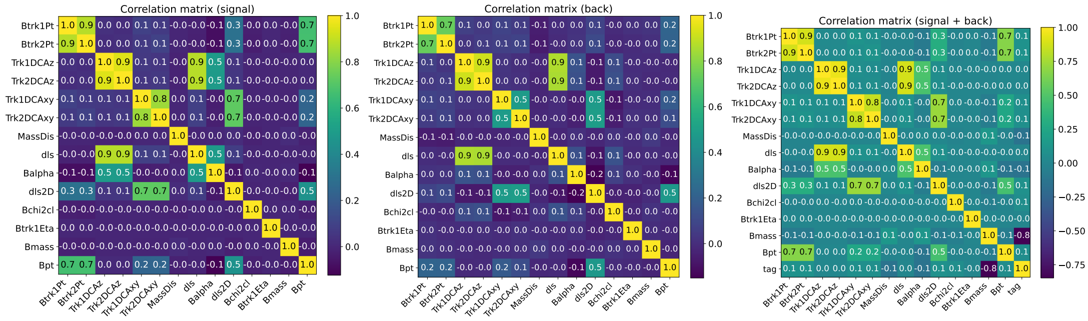

Figure 4.5: Correlation matrices, for the B_s^0 meson, between our variables on signal (left), background (middle) and on signal+background (right) for $3 < p_T < 50$ GeV/c. This last variation has an extra variable called “tag” which represents the true label of the event (either signal or background).

plements machine learning algorithms under the gradient boosting framework. For the training of neural networks, we use the pytorch package [43], a machine learning framework based on the Torch library.

For hyperparameter optimization, we use Optuna [44], an automatic optimization software framework. Given search ranges for different parameters of our models, Optuna searches for the configuration that maximizes or minimizes our set of objectives. In our case, the objective is to maximize the area under the Receiving Operating Characteristic (ROC) curve obtained with our models.

For the neural network, we let the number of *layers* vary between 1 and 5, the size of each *hidden layer* can fluctuate between 128 and 256 units, the *learning rate* varies from 1×10^{-6} to 1×10^{-2}, the *dropout rate* at each *layer* can also change between 0 and 0.5 and finally the *weight decay* of the applied *L2 regularization* can shift between 1×10^{-4} to 0.05.

The best performing NN model was obtained after 103 trials with pruning, and is given in Table 4.1.

Table 4.1: NN model summary for the B^+ meson.

	Layer nº			
Parameters	1	2	3	4
Hidden size	159	164	182	205
Dropout	0.138	0.1	0.3	0.12
Learning rate	0.0013			
Weight decay	0.00012			
Nº of Epochs	20			
Batch size	256			
Activation function	Relu			
Optimizer	Adam			

For the BDT, we let the *booster* (the method utilized for performing gradient boosting) either be *dart* or *gbtree*, with both utilizing tree based models, the *max depth* of each tree is allowed to change between 1 and 20. For the rest of the parameters of our model, we have *eta*, the step size shrinkage used in updates to prevent overfitting, *gamma*, the minimum loss reduction required to make a further partition on a lead node of the tree, *lambda*, the L2 regularization term on weights and *alpha*, the L1 regularization term on weights all set to vary between 1×10^{-8} and 1. The *grow policy*, that is how the model controls the way new nodes are added to the tree, can either be *depthwise*, splitting the nodes closest to the root, or *lossguide*, splitting at nodes with the highest loss change. There are additionally 4 extra parameters when the chosen *booster* is *dart*. The *dropout rate* and *skip drop*, the probability of skipping the dropout procedure during a boosting iteration, are also allowed to change from 1×10^{-8} to 1. Finally, we have the sampling algorithm, which can either be uniform or weighted, making the dropped trees selected uniformly or in proportion to weight, respectively, and the normalization algorithm, which is allowed to be the *tree* or *forest* algorithm, making so that new trees have the same weight as each of the dropped ones or have the same weight as their sum. A more complete description of all the BDT parameters can be found in the XGBoost documentation [45].

Table 4.2: BDT model summary for the B^+ and B_s^0 mesons.

Parameters	B^+				B_s^0
Booster	Dart				gbtree
Booster parameters	Skip drop	Dropout rate	Sample type	Normalize type	Nan
	0.0028	2.08	weighted	tree	
Lambda	0.0001				0.001
Alpha	0.13				1.05×10^{-8}
Max depth	9				5
Eta	0.32				0.5
Gamma	0.0011				5.4×10^{-6}
Grow policy	Depthwise				Depthwise
Nº of epochs	20				20

The best performing BDT models for the B^+ and B_s^0 mesons were obtained after 100 trials with pruning, and are given in Table 4.2.

On Fig. 4.6 is presented the trial progression for our B^+ BDT model. Each line represents one iteration of Optuna, showing the options chosen at that specific time. Places where the density of lines is higher represent the more favoured values during optimization. The full results pertaining to the optimization performed by Optuna are presented on Appendix D.

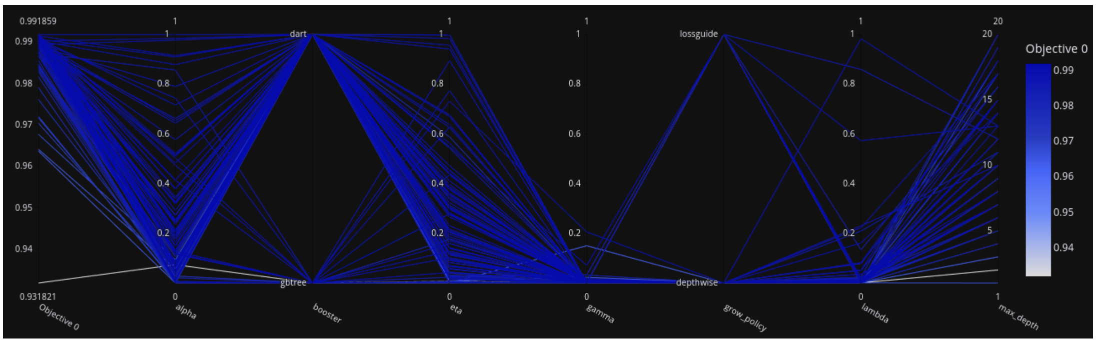

Figure 4.6: Trial progression for Optuna during our B^+ BDT model optimization.

4.3 Training and performance

We start by training both a NN and a BDT on the B^+ meson, due to the sample larger statistics, with the aim of comparing performances between these two and select the one more apt to perform our selection. Therefore, the results shown on this section are pertaining to the B^+ meson unless otherwise stated. The results pertaining to all models can be found on Appendix E.1.

Figs. 4.7 and 4.8 show the accuracy, defined as the number of events our model correctly labels over the total number of events, per epoch of our models, comparing this for the training set and a validation set[1]. We are able to verify that there is no overfitting occurring for our NN since the behaviour of both curves is the same, however there is slight overtraining occurring for our BDT model, although it is of no significant amount, as can be seen from the AUC plot where this one still keeps increasing for the last few iterations unlike the accuracy.

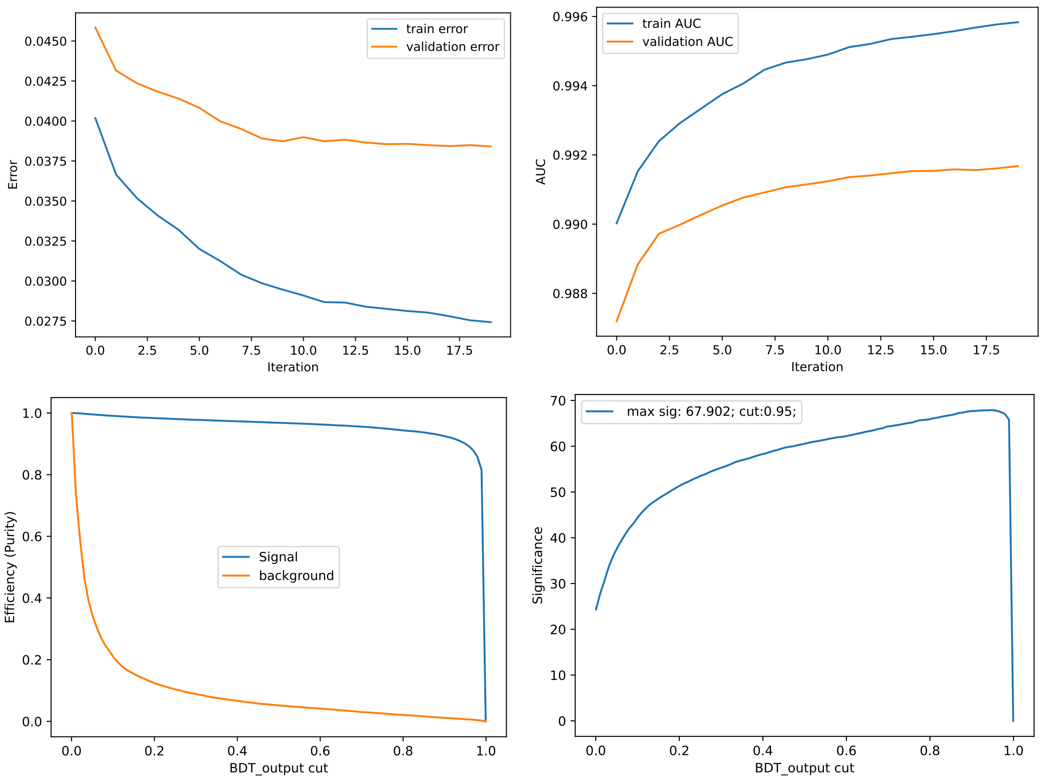

Figure 4.7: Top row: Accuracy (left) and area under the ROC curve (right) on every iteration of the BDT model during training. Bottom row: Purity of signal and background events (left) and Significance (right) vs cut value in our BDT score.

It can also be observed that, during training, our BDT model is able to give a higher significance after cutting than the NN model, having the BDT a maximum significance of 68 and the NN a maximum of 47. This is consistent with the area under the ROC curve, shown in Fig. 4.9, where the highest value is also obtained by the BDT, being this one an area of 0.992, versus the 0.989 given by our NN.

Finally, Fig. 4.10 shows the feature importance calculated through the capabilities of *Xgboost* for our BDT model, and for our NN model we calculate it through shuffling the respective variable, making a

[1]We create our validation and training sets from randomly splitting our samples. The training set corresponds to 64% of our full dataset whilst the validation set corresponds to 16%. The remaining 20% of our total number of events is labelled as the test set.

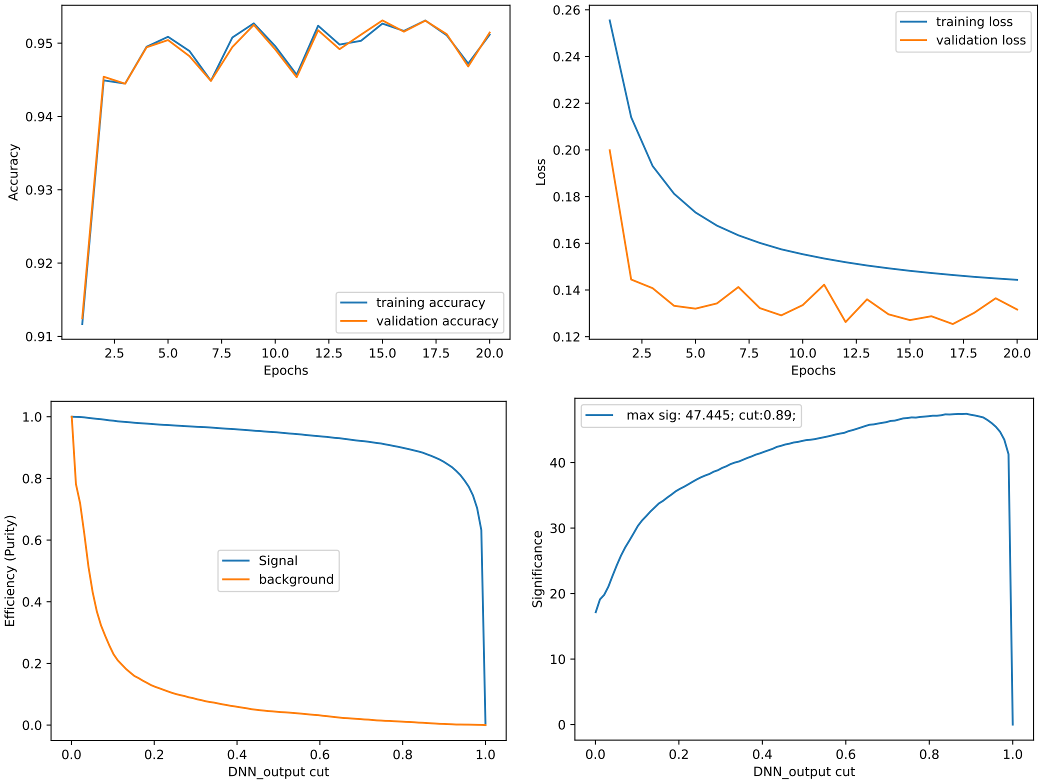

Figure 4.8: Top row: Accuracy (left) and loss function value (right) on every iteration of the NN model during training. Bottom row: Purity of signal and background events (left) and Significance (right) vs cut value in our NN score.

new dataset with mismatched entries, passing it through our trained model multiple times, calculating the average difference of the predictions with the ones obtained with the original dataset and taking this as a measure of the importance of the feature. It is to note that through all the variations the models have gone through, the feature importance never suffered any substantial changes.

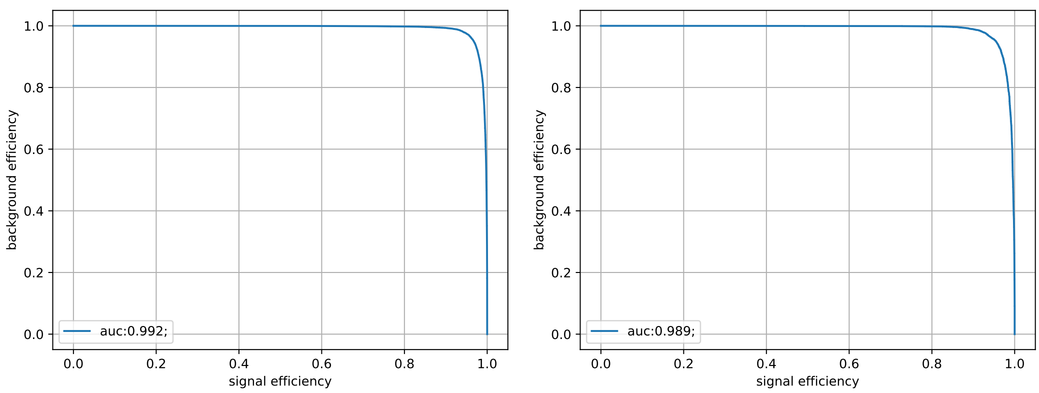

Figure 4.9: ROC curve for our BDT model (left) and for our NN model (right).

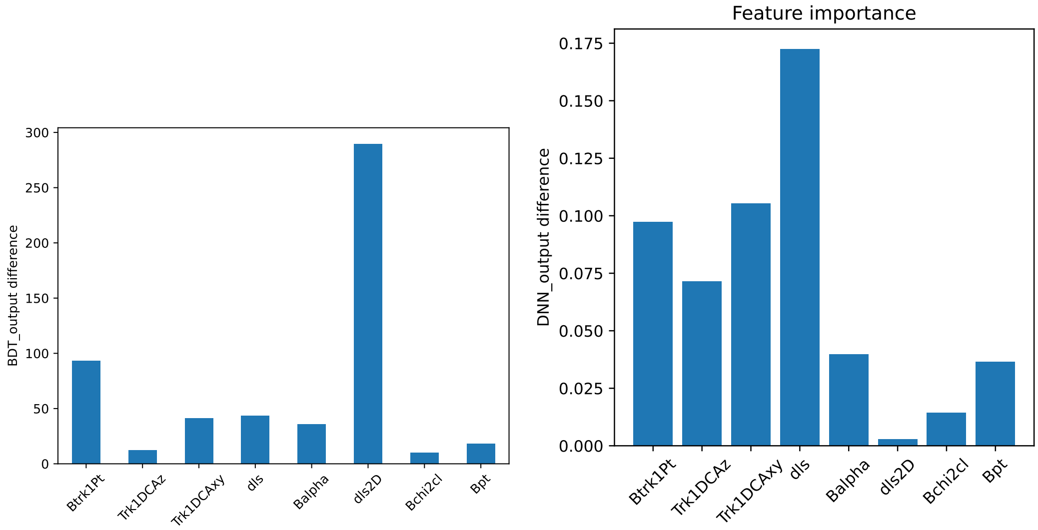

Figure 4.10: Feature importance for our BDT model (left) and for our NN model (right).

4.4 Working point

After having carried out the training in the full dataset, we apply our models to all events. We find the cuts that optimize the figure of merit in each p_T bin, given the dependence of the background level on p_T.

In our figure of merit, $\frac{S}{\sqrt{S+B}}$, S and B are the normalized number of signal and background in the signal region. This normalization is due to the fact that we do not wish to depend on the number of events in our samples, MC or data, for the calculation of the significance. Therefore, we took the ratio of the number of signal events in a dataset with a reference selection already applied and the number of signal events in our MC sample, whilst for background we took a ratio of the number of background events within the signal region and the background events within the sidebands. For both cases, we obtained the number of events from the datasets by performing a fit to the invariant mass plot. For use as reference, we chose the selection described in Ref. [46].

On Figs. 4.11, 4.13 and 4.15 we show the resulting ROC curves for the p_T bins 5-7, 10-15 and 30-50 GeV/c respectively. As we can see, the BDT model shows an overall better performance for every bin, maintaining the trend we saw at training. The results pertaining to all bins can be found on Appendix E.2.

On Figs. 4.12, 4.14 and 4.16 we show the purity and significance versus cut values for the bins 5-7, 10-15 and 30-50 GeV/c respectively. Here we notice the same as with the ROC curves, the BDT model shows a better performance, and additionally we can more clearly see through the significance that this selection has a more noticeable impact for low p_T, where there is a higher amount of background compared to the other bins, not existing a very high increase in significance per cut for a bin like 30-50 GeV/c for example, where the amount of noise was small to begin with.

Having done this procedure for all p_T bins, we are able to obtain the cut that yields the highest significance, the working point. These values are presented in Table 4.3. From these results, it is clear to see that the BDT has a higher performance than the NN model (this is made even clearer in

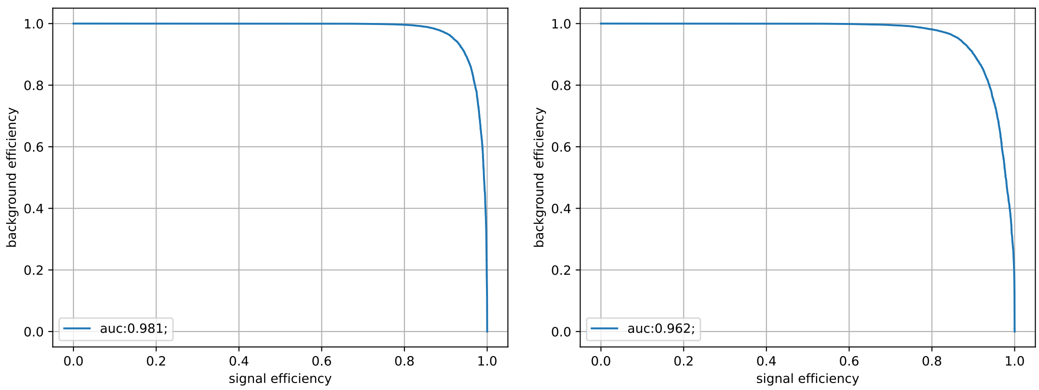

Figure 4.11: ROC curve for our BDT model (left) and for our NN model (right), for $5 < p_T < 7$ GeV/c.

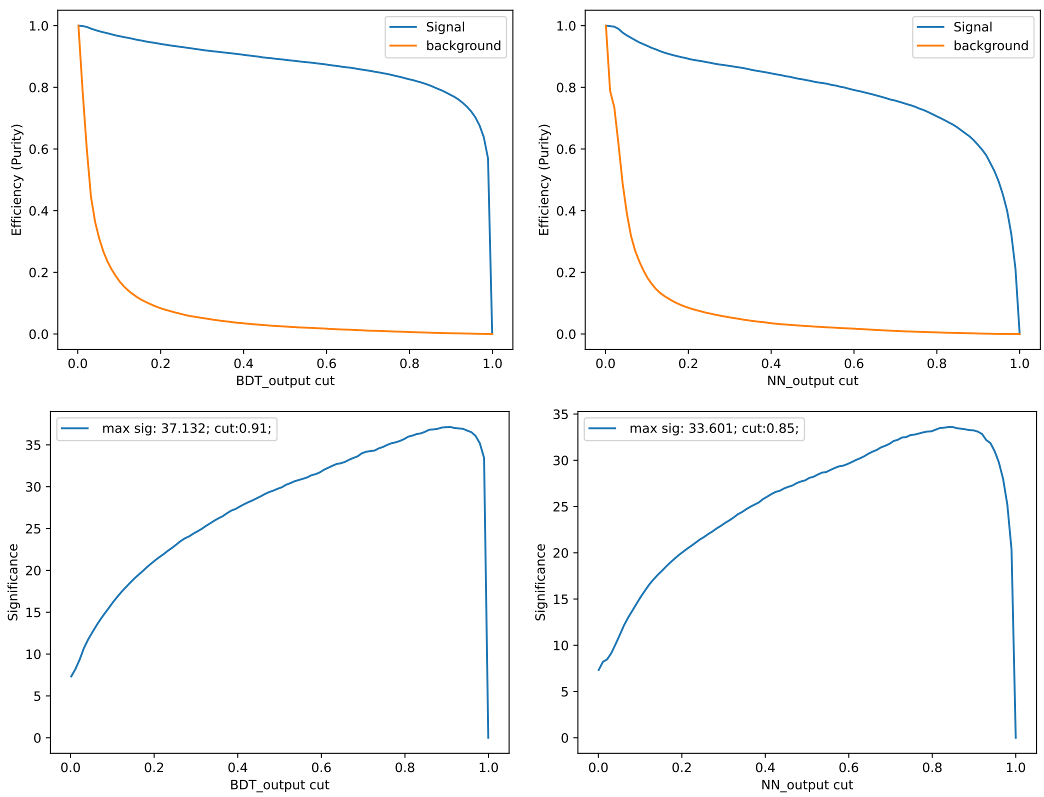

Figure 4.12: Top row: Signal and background purity vs cut value for our BDT model (left) and for our NN model (right). Bottom row: Significance vs cut value for our BDT model (left) and for our NN model (right). All plots are for $5 < p_T < 7$ GeV/c.

Sec. 5.4.1), therefore, the BDT model is chosen to perform the analysis selection. The full training and selection results for the B_s^0 meson are shown in Appendix E, alongside the full B^+ results, with the resulting working point presented in Table 4.4.

Table 4.3: NN and BDT working points, for the B^+ meson.

B^+ p_T [GeV/c]	5–7	7–10	10–15	15–20	20–30	30–50	50–60
BDT working point	0.91	0.95	0.95	0.9	0.77	0.95	0.97
NN working point	0.85	0.88	0.89	0.88	0.67	0.89	0.61

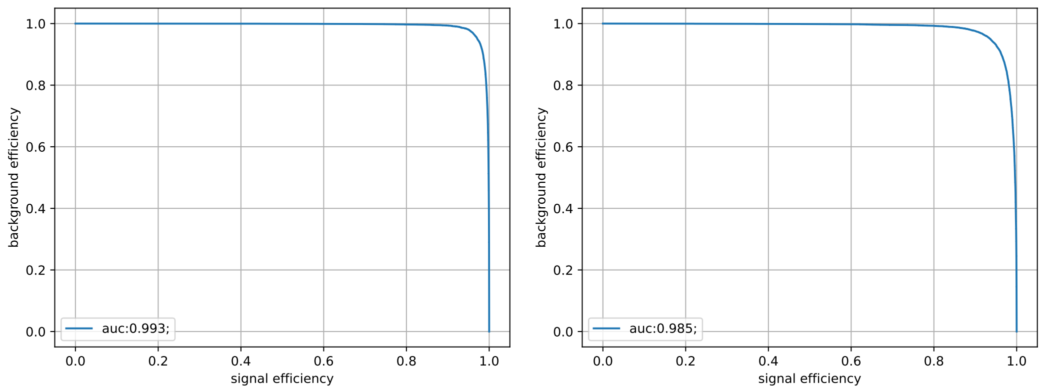

Figure 4.13: ROC curve for our BDT model (left) and for our NN model (right), for $10 < p_T < 15$ GeV/c.

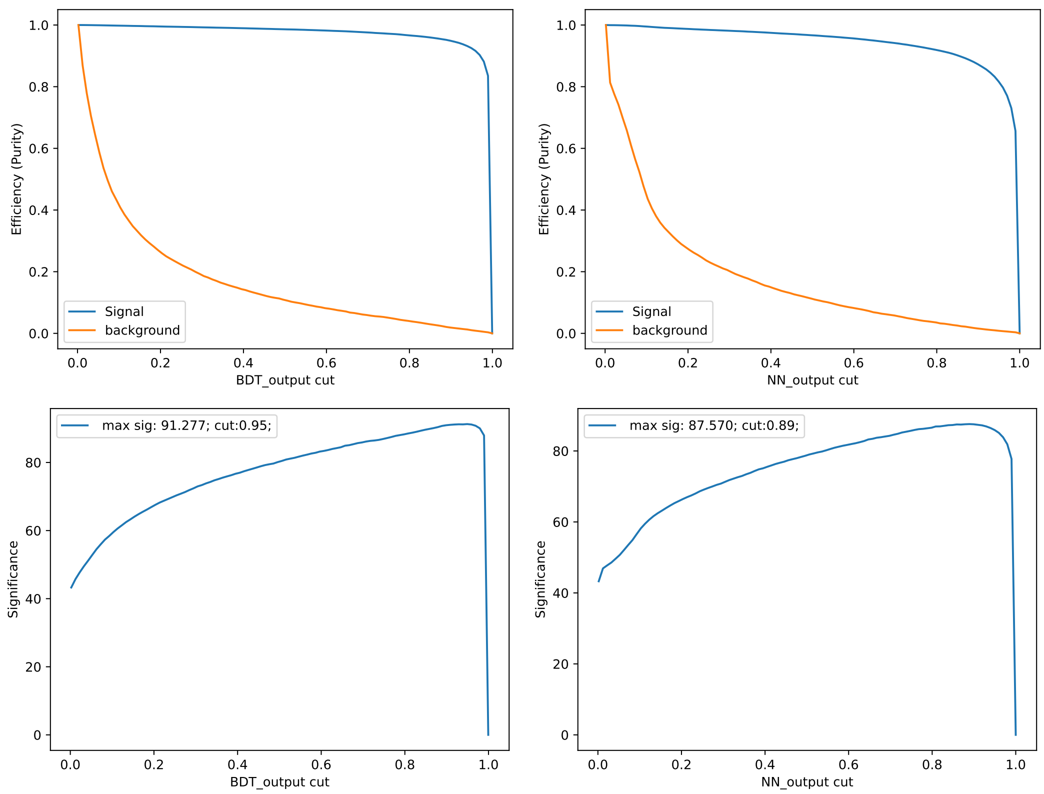

Figure 4.14: Top row: Signal and background purity vs cut value for our BDT model (left) and for our NN model (right). Bottom row: Significance vs cut value for our BDT model (left) and for our NN model (right). All plots are for $10 < p_T < 15$ GeV/c.

Table 4.4: BDT working points, for the B_s^0 meson.

B_s^0 p_T [GeV/c]	7–10	10–15	15–20	20–50
BDT working point	0.99	0.99	0.99	0.991

With these values calculated, we associated to each event in our data a probability of being signal, given by the output of our models, performing cuts on this new variable according to the working point obtained for each bin.

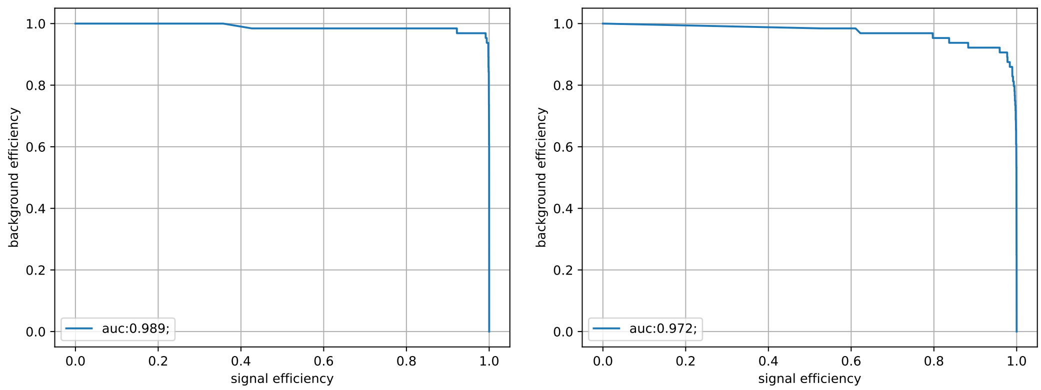

Figure 4.15: ROC curve for our BDT model (left) and for our NN model (right), for $30 < p_T < 50$ GeV/c.

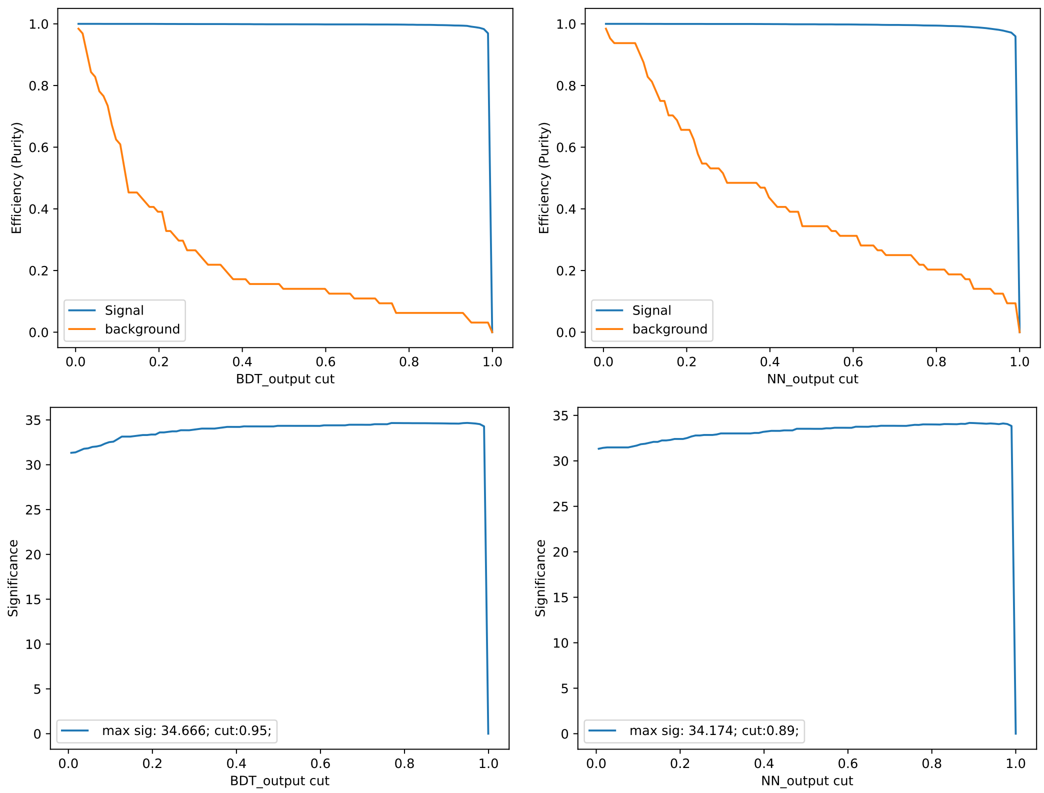

Figure 4.16: Top row: Signal and background purity vs cut value for our BDT model (left) and for our NN model (right). Bottom row: Significance vs cut value for our BDT model (left) and for our NN model (right). All plots are for $30 < p_T < 50$ GeV/c.

4.5 Low p_T studies

It is desirable to expand our analysis to the low p_T region. We apply the same method as before for obtaining a working point for the $[3, 5]$ GeV/c bin (and for the $[5, 7]$ GeV/c bin for the B_s^0 meson), with a slight difference, since the chosen reference selection has no data in this region it's impossible to use it as a reference to normalize our significance. Therefore, we performed an arbitrary cut on our unselected dataset of only selecting events with a BDT or NN score higher than 0.5 to serve as the reference. The results are shown on Figs. 4.17 and 4.18. In Sec. 5.4.1 we present a comparison between the BDT, NN and reference selection, validating our BDT model as the nominal choice, as well as conclude the studies of the low p_T region.

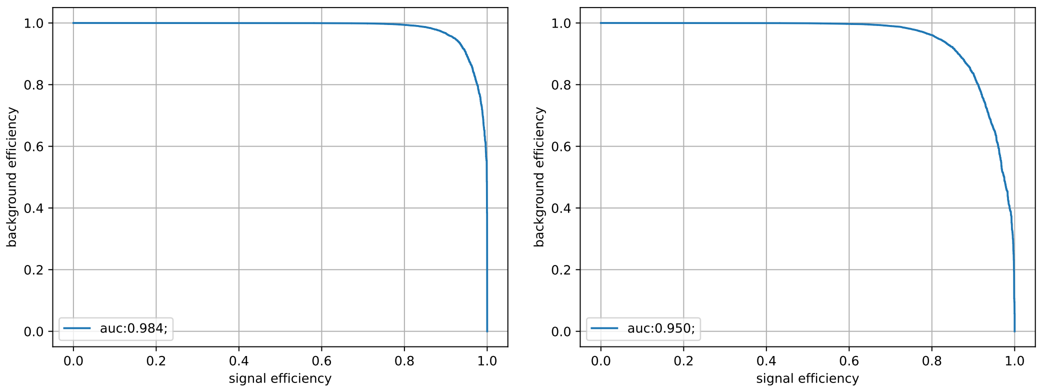

Figure 4.17: ROC curve for our BDT model (left) and for our NN model (right), for $3 < p_T < 5$ GeV/c.

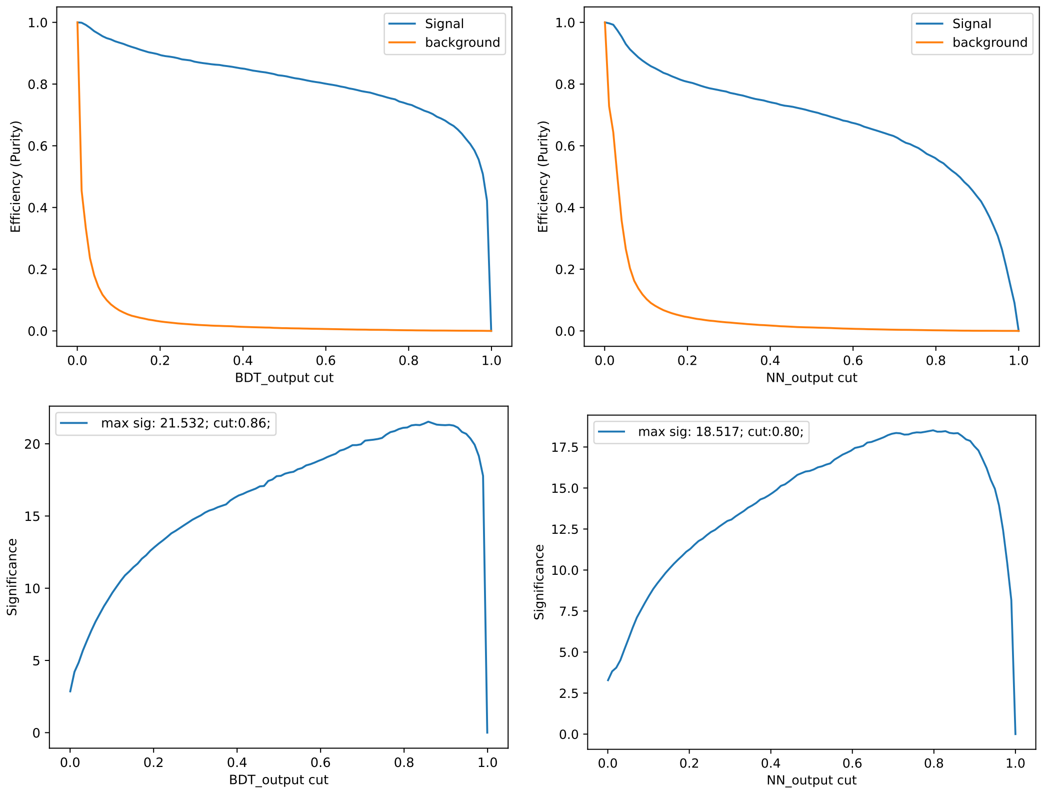

Figure 4.18: Top row: Signal and background purity vs cut value for our BDT model (left) and for our NN model (right). Bottom row: Significance vs cut value for our BDT model (left) and for our NN model (right). All plots are for $3 < p_T < 5$ GeV/c.

Table 4.5: NN and BDT working points, for the B^+ meson.

B^+ p_T [GeV/c]	3–5
BDT working point	0.86
NN working point	0.8

Chapter 5

Differential signal yields

5.1 Extended unbinned maximum likelihood fitting

The differential signal yields, Y_i, appearing in the B meson production cross-section expression, given by Eq. 1.5, are obtained via extended unbinned maximum likelihood (EUML) fits to the invariant mass distribution of the final states appearing in the decay chains represented by the Feynman diagrams shown in Fig.1.4. This is a well-established method for estimating the parameters that maximize the likelihood function, $\mathcal{L}$, which gives the probability of observing a data sample, characterized by a set of n statistically independent quantities, $\vec{m} = (m_1, ..., m_n)$, knowing the model $l(\vec{Y}, m_i, \vec{\theta})$, with arguments $\vec{\theta} = (\theta_1, ..., \theta_p)$, composed of a sum of normalized probability density functions (PDFs), $Y_\alpha f_\alpha(m_i, \vec{\lambda}_\alpha)$, whose arguments $\vec{\lambda}_\alpha$ reside in a subspace of $\mathbb{R}^p$, with p representing the number of our parameters, that is, the number of entries of $\vec{\theta}$. We are interested in obtaining the signal yields and in order to get a reliable estimate of their uncertainty the likelihood function must be extended by adding a multiplicative term corresponding to the Poison distribution, $\frac{Y^n e^Y}{n!}$, where $Y = \sum_\alpha Y_\alpha$.

Each event that passed the trigger and the selections described in Sec. 3.4 and 4.4 is reconstructed as a B_s^0 or B^+ meson candidate with an associated invariant mass, m_i. Because this is a statistically independent quantity that characterizes the data sample, the quantity $\vec{x}$ is identified with the invariant mass of each B meson candidate event in the data in this analysis. Therefore, we write the likelihood function as

$$\mathcal{L}(Y_1, ..., Y_\alpha, \vec{m}; \vec{\theta}) = \frac{Y^n e^Y}{n!} \prod_{i=1}^{n} \left\{ \sum_\alpha Y_\alpha f_\alpha(m_i, \vec{\lambda}_\alpha) \right\}. \tag{5.1}$$

The exponential in Eq. 5.1 will be a very large number for very large data samples. As a result, it is more practical to consider the minimization of the negative logarithm of the likelihood, $-ln\mathcal{L}$. To accomplish this, the ROOFIT [47] package is used, and the minimization is performed by MINUIT [48].

5.2 The fitting models

Each B meson event's mass distribution will have different structures that must be addressed individually when building their respective models. The signal peak and the combinatorial background are two such features shared by both B meson mass spectra. The former is composed by a mass resonance, while the latter is a continuous structure formed by the incorrect combination of particle tracks. The B_s^0 mass distribution involves only these two structures. On the other hand, there are potential background feed-down sources in the B^+ mass distribution from other B meson decays that can form peaking structures in the region of interest and thus need to be properly treated in order to not bias the yield extraction procedure. Candidates in the inclusive B^+ meson MC sample that are matched to a genuine signal are discarded in order to study these components. The resulting mass spectrum confirms the existence of two peaking structures, one very robust below 5.20 GeV/c^2 and one very feeble around 5.34 GeV/c^2. These two peaking structures have different origins. The first occurs when a track from any B meson 4-body decays through similar channels, such as $B_s^0 \to J/\psi\ K^+K^-$. As a result, we refer to it as the partially reconstructed background. The second is formed by misidentifying the pions from $B^+ \to J/\psi\ \pi^+$ as kaons. Because the associated flavour transition is suppressed by the V_{CKM} matrix, we call it the Cabibbo suppressed background.

For the decay channel $B_s^0 \to J/\psi\phi \to K^+K^-\mu^+\mu^-$ the model used to fit the invariant mass distribution of the final state system, $m_{\mu^+\mu^-K^+K^-}$, is composed by an exponential function, $\mathcal{P}_{exp}(m_i;\lambda_{cb}) = e^{-m_i\lambda_{cb}}$, and a sum of two Gaussians PDFs, $\mathcal{P}_{2G}(m_i;\vec{\lambda}_{2G})$, with $\vec{\lambda}_{2G} = (\sigma_1, \sigma_2, \mu, \alpha)$ and

$$\mathcal{P}_{2G}(m_i;\vec{\lambda}_{2G}) = \frac{\alpha}{\sqrt{2\pi}\sigma_1}\text{exp}\left(\frac{-(m_i-\mu)^2}{2\sigma_1^2}\right) + \frac{(1-\alpha)}{\sqrt{2\pi}\sigma_2}\text{exp}\left(\frac{-(m_i-\mu)^2}{2\sigma_2^2}\right) \tag{5.2}$$

where σ_1 and σ_2 are the widths of the respective Gaussian functions, α is a parameter relating them and μ is their mean, physically standing for the B meson mass and being for that reason the same in the two Gaussian functions. The $\mathcal{P}_{exp}(m_i;\lambda_{cb})$ PDF accounts for the combinatorial background while $\mathcal{P}_{2G}(m_i;\vec{\lambda}_{2G})$ accounts for the mass resonance. Hence, this model parameters are $\vec{\theta}_{B_s^0} = (\sigma_1, \sigma_2, \mu, \alpha, \lambda_{cb})$ and its full expression is given by

$$l(m_i, \vec{\theta}_{B_s^0}) = Y_s\left[\frac{\alpha}{\sqrt{2\pi}\sigma_1}\text{exp}\left(\frac{-(m_i-\mu)^2}{2\sigma_1^2}\right) + \frac{(1-\alpha)}{\sqrt{2\pi}\sigma_2}\text{exp}\left(\frac{-(m_i-\mu)^2}{2\sigma_2^2}\right)\right] + Y_{cb}\,\text{exp}(-\lambda_{cb}m_i) \tag{5.3}$$

where Y_s and Y_{cb} are the signal and background yields.

For the decay channel $B^+ \to J/\psi K^+ \to \mu^+\mu^-K^+$, the mass resonance and the combinatorial background are also described by a sum of two Gaussians PDF and by an exponential PDF. The partial reconstructed background is modelled through a complementary error function PDF, $\mathcal{P}_{Ec}(m_i, C_{sc}, C_{sh})$,

$$\mathcal{P}_{Ec}(m_i, C_{sc}, C_{sh}) = 1 - \frac{2}{\sqrt{\pi}}\int_0^{\left(\frac{m_i - C_{sh}}{C_{sc}}\right)} e^{t^2}\, dt\,, \tag{5.4}$$

where C_{sh} and C_{sc} are, respectively, shifting and scaling parameters that allow a more comprehensive construction of this PDF. The Cabibbo suppressed background is modeled by a PDF, $\mathcal{P}_{2\mathcal{A}_G}(m_i, \vec{\lambda}_{2\mathcal{A}_G})$, with $\vec{\lambda}_{2\mathcal{A}_G} = (\mu_{aG}, \beta, \sigma_{1L}, \sigma_{2L}, \sigma_{1R}, \sigma_{2R})$, that consists in the sum of two asymmetric Gaussians,

$$\mathcal{A}_G(m_i; \mu_{aG}, \sigma_L, \sigma_R) = \frac{2}{(\sigma_L + \sigma_R)\sqrt{2\pi}} \begin{cases} \exp\Big(-\frac{(m_i-\mu_{aG})^2}{2\sigma_L^2}\Big) & \text{if } m_i < \mu_{aG} \\ \exp\Big(-\frac{(m_i-\mu_{aG})^2}{2\sigma_R^2}\Big) & \text{otherwise} \end{cases} \tag{5.5}$$

related by the coefficient β according to

$$\mathcal{P}_{2\mathcal{A}_G}(m_i, \vec{\lambda}_{2\mathcal{A}_G}) = \beta\mathcal{A}_G(m_i, \mu_{aG}, \sigma_{1L}, \sigma_{1R}) + (1-\beta)\mathcal{A}_G(m_i, \mu_{aG}, \sigma_{2L}, \sigma_{2R})\,. \tag{5.6}$$

Since this background structure arises from a specific B meson decay its yield, $Y_{J/\psi\pi^+}$, should be proportional to the signal yield instead of a free parameter. Thus, the full model to fit the B^+ meson data has parameters $\vec{\theta}_{B^+} = (Y_s, Y_{cb}, Y_{J/\psi X}, \sigma_1, \sigma_2, \sigma_{1L}, \sigma_{1R}, \sigma_{2L}, \sigma_{2R}, \mu, \mu_{aG}, \alpha, \beta, \lambda_{cb})$, and is given by

$$\begin{aligned} l(m_i, \vec{\theta}_{B^+}) = {} & Y_s\left[\frac{\alpha}{\sqrt{2\pi}\sigma_1}\exp\left(\frac{-(m_i-\mu)^2}{2\sigma_1^2}\right) + \frac{(1-\alpha)}{\sqrt{2\pi}\sigma_2}\exp\left(\frac{-(m_i-\mu)^2}{2\sigma_2^2}\right)\right] + Y_{cb}\exp\big(-\lambda_{cb}m_i\big) \\ & + Y_{J/\psi X}\,\mathcal{P}_{Ec}\Big(\frac{m_i - C_{shift}}{C_{scale}}\Big) + Y_{J/\psi\pi^+}\,\mathcal{P}_{2\mathcal{A}_G}(m_i, \mu_{aG}, \sigma_{1L}, \sigma_{1R}, \sigma_{2L}, \sigma_{2R})\,, \end{aligned} \tag{5.7}$$

where $Y_{J/\psi\pi^+} = R_{PDG} \times Y_s$ with $R_{PDG} = \frac{\mathcal{B}(B^+\to J/\psi\pi^+)}{\mathcal{B}(B^+\to J/\psi K^+)} = 0.038 \pm 0.001$.

5.3 PDF shape constrains from MC studies

The sole purpose of fitting the data is to obtain the number of signal events, Y_s, and in order to do so as accurately as possible, the shape of the two fitting models introduced in Sec. 5.2 shall be constrained via MC studies. This reduces the number of floating parameters, resulting in a more reliable adjustment between individual PDFs and data structures.

To characterize the shape of the signal PDF of both B meson models, Eq. 5.2, the respective parameters, $\vec{\lambda}_{2G}$, shall be learned and fixed by fitting the signal MC sample. The top row of Fig. 5.1 shows the two inclusive bin results, and the variable binned results can be found on Appendix A. For the B^+ model, both the partially reconstructed and Cabibbo suppressed backgrounds PDFs must be constrained. The shape of the Cabibbo suppressed background PDF was found to be consistent across the analysis bins after a binned and inclusive MC sample study, and thus its parameters are fixed once through an inclusive analysis.

In contrast, the shape of the partially reconstructed background was discovered to vary slightly across the p_T and $|y|$ bins, and thus the respective PDF, like the signal PDF, is particularly constrained to each bin. Due to the last bin of the B^+ meson, 50-60 GeV/c^2, containing very few events, the shape characterization becomes unreliable. To circumvent this problem, we consider the parameters fixed in the previous bin.

The inclusive bin results for these backgrounds are shown in the bottom row of Fig. 5.1.

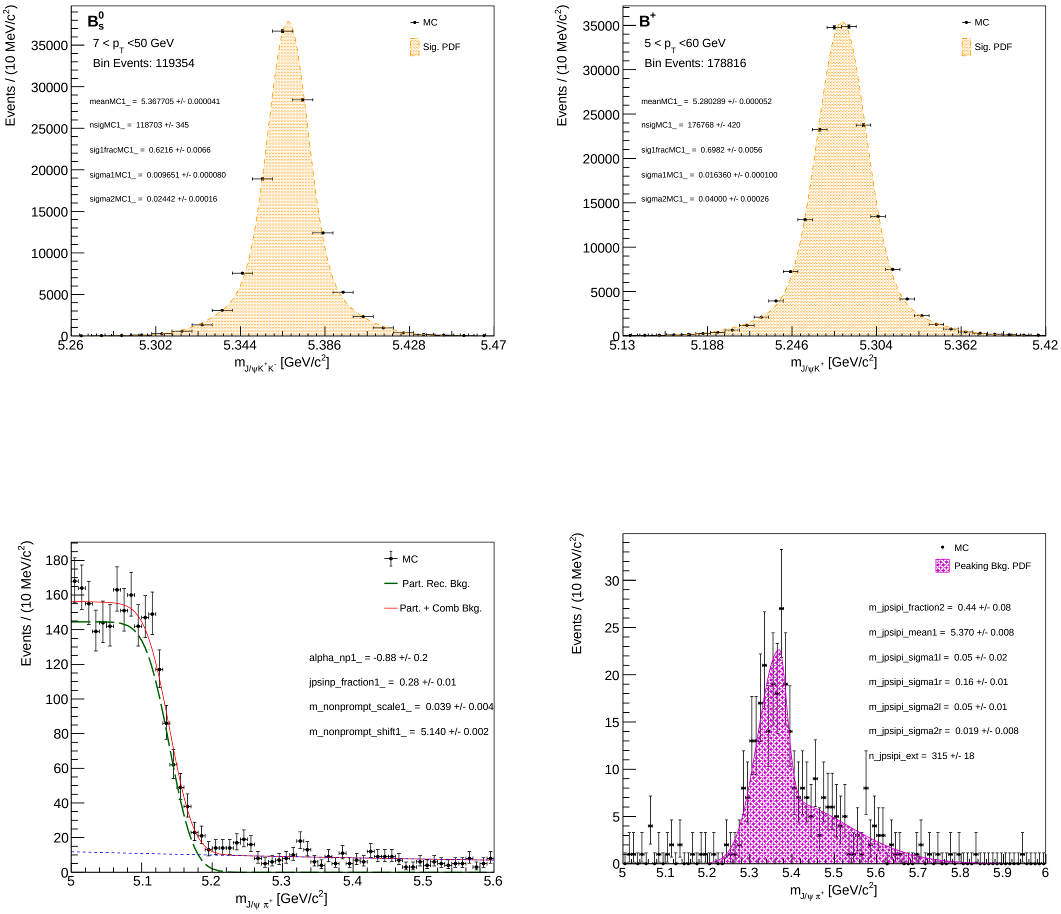

Figure 5.1: On the top row are the fits to the signal MC inclusive bin for the B_s^0 meson (left) and the B^+ meson (right). On the bottom row are the fits to MC inclusive bin for the partial reconstructed background, $B \to J/\psi\, X$ (left) and for the Cabibbo suppressed background $B^+ \to J/\psi\, \pi^+$ (right).

5.4 Invariant mass fits

We let the widths of the signal PDFs to float through the introduction of a common scaling factor, C, so that $\sigma_{1,2} \to C\sigma_{1,2}$ and thus

$$(5.2) \to \mathcal{P}_{2G}(m_i; \vec{\lambda}_{2G}) = \frac{\alpha}{\sqrt{2\pi}C\sigma_1}\text{exp}\left(\frac{-(m_i-\mu)^2}{2C\sigma_1^2}\right) + \frac{(1-\alpha)}{\sqrt{2\pi}C\sigma_2}\text{exp}\left(\frac{-(m_i-\mu)^2}{2C\sigma_2^2}\right). \tag{5.8}$$

This allows for a more accurate description of the data while preserving the overall shape of the fitting model and accounting for mismodelling of detector resolution effects in MC affecting the mass resonance width. In addition, the mean of the PDF signal is released and a model variation where this is instead considered fixed is taken to study the systematic uncertainties. The mass interval $[5, 6]$ GeV/c^2 is fitted to both B meson data samples. The inclusive bin results are shown in Fig. 5.2. The pulls, P_i, are defined as $P_i = \frac{Y_i^{data} - Y_i^{fit}}{\sigma_i^{data}}$, where Y_i^{data} is the total number of data events in the ith-bin, σ_i^{data} is the respective uncertainty, and Y_i^{fit} is the total number of fit events.

A χ^2 test was run, and the results are shown in the top left corner of each plot. For normally distributed data points around the fit result, the χ^2 value should be equal to the number of degrees of freedom (ndf), or χ^2/ndf~ 1. With the exception of the B^+ low data bin, [50, 60] GeV/c^2, the scaling parameter, C, was found to be consistent with unity. Because the respective MC bin also has a low number of events, the addition of this parameter effectively prevented a poor signal shape characterization from affecting yield extraction. The values obtained for the unfixed mean were very close to the respective B meson mass, indicating that the PDF components were properly shifted during the fit.

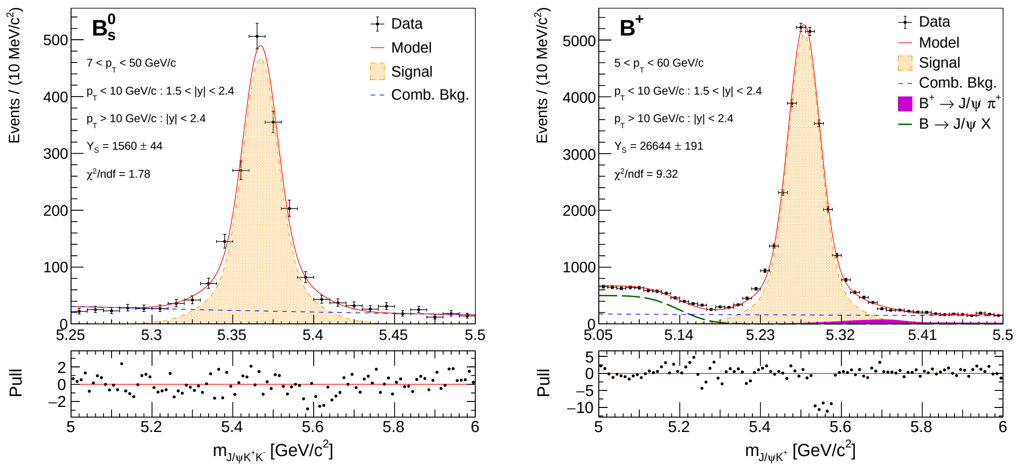

Figure 5.2: Fit to B_s^0 meson (left) and B^+ meson (right) invariant mass distribution.

In the Tables 5.1 to 5.6 the obtained B_s^0 and B^+ meson signal yields normalized by the respective bin widths, $\frac{Y_s}{\Delta p_T}$, $\frac{Y_s}{\Delta|y|}$, and $\frac{Y_s}{\Delta N_{trk}}$ are summarized. Appendix C.1 shows a plot of this values with their respective statistical and systematic uncertainty.

Table 5.1: Obtained Y_s normalized by the p_T bin width for the B_s^0 meson.

p_T [GeV/c]	7–10	10–15	15–20	20–50
$\frac{Y_s}{\Delta p_T}$ [c/GeV]	46$\pm$ 5	122$\pm$ 5	82$\pm$ 4	13$\pm$ 1

Table 5.2: Obtained Y_s normalized by the p_T bin width for the B^+ meson.

p_T [GeV/c]	5–7	7–10	10–15	15–20	20–30	30–50	50–60
$\frac{Y_s}{\Delta p_T}$[c/GeV]	804$\pm$ 27	979$\pm$ 23	2071$\pm$ 24	1217$\pm$ 18	421$\pm$ 7	64 $\pm$ 2	10 $\pm$ 1

Table 5.3: Obtained Y_s normalized by the $|y|$ bin width for the B_s^0 meson.

$\|y\|$	0.0–0.5	0.5-1.0	1.0–1.5	1.5–2.0	2.0–2.4
$\frac{Y_s}{\Delta\|y\|}$	643$\pm$37	685$\pm$39	645$\pm$39	872$\pm$48	284$\pm$33

Table 5.4: Obtained Y_s normalized by the $|y|$ bin width for the B^+ meson.

$\|y\|$	0.0–0.5	0.5-1.0	1.0–1.5	1.5–2.0	2.0–2.4
$\frac{Y_s}{\Delta\|y\|}$	10458 $\pm$ 151	10310$\pm$154	10694$\pm$166	18325$\pm$240	6046$\pm$178

Table 5.5: Obtained Y_s normalized by the N_{trk} bin width for the B_s^0 meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
$\frac{Y_s}{\Delta N_{trk}}$	17$\pm$1	47$\pm$2	46$\pm$2	22$\pm$2	11$\pm$1	6$\pm$1	1$\pm$0.2

Table 5.6: Obtained Y_s normalized by the N_{trk} bin width for the B^+ meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
$\frac{Y_s}{\Delta N_{trk}}$	350$\pm$5	884$\pm$11	703$\pm$10	399$\pm$8	183$\pm$5	70$\pm$3	10$\pm$1

5.4.1 Selection comparison and the low p_T region

In Sec. 4.4 we described how we normalized the number of signal and background events according to a previously done selection [46]. Of course, it is now of interest to draw a comparison between these two to see if the one we obtained is indeed an improvement.

In Fig. 5.3 we show the invariant mass fit plots of the B^+ data with our selections, compared to the invariant mass fit plots of our reference selection for the bins 5-7, 10-15 and 30-50 GeV/c. The full comparison can be seen on Appendix F.

Our optimized selections show a considerable reduction in the amount of background, especially at low p_T where it is more abundant, and as expected we see diminishing returns, when compared to the reference selection, for higher p_T, where the number of background events is lower. Comparing the invariant mass plots obtained with the BDT and NN models, we can also clearly see that there seems to be slightly less background events for our NN selection, but this comes at the cost of a disproportionately larger amount of signal events being cut, ending up giving us a smaller significance than the BDT model selection.

Tables 5.7 to 5.12 show the significance values for on each bin for each selection, being in the B_s^0 only shown the reference and nominal selection since no NN model was trained for this meson. The values observed here are consistent with the plots showed in Fig. 5.3, validating the BDT as our nominal model and showing that both the BDT and NN offer a considerable improvement over the reference selection.

Table 5.7: Significance vs p_T on each model, for the B^+ meson.

p_T [GeV/c]	5–7	7–10	10–15	15–20	20–30	30–50	50–60
Ref.	14.3	24.5	64.2	60.1	55.6	29.6	7.3
NN	24.8	39.5	81.8	67.0	54.6	30.8	8.3
Nominal	25.4	40.6	82.5	64.8	55.6	31.4	8.8

Finally, for the low p_T region we do not possess reference results to compare to since fitting in this region wasn't previously done, however this very fact proves the effectivity of our selection, and opens up

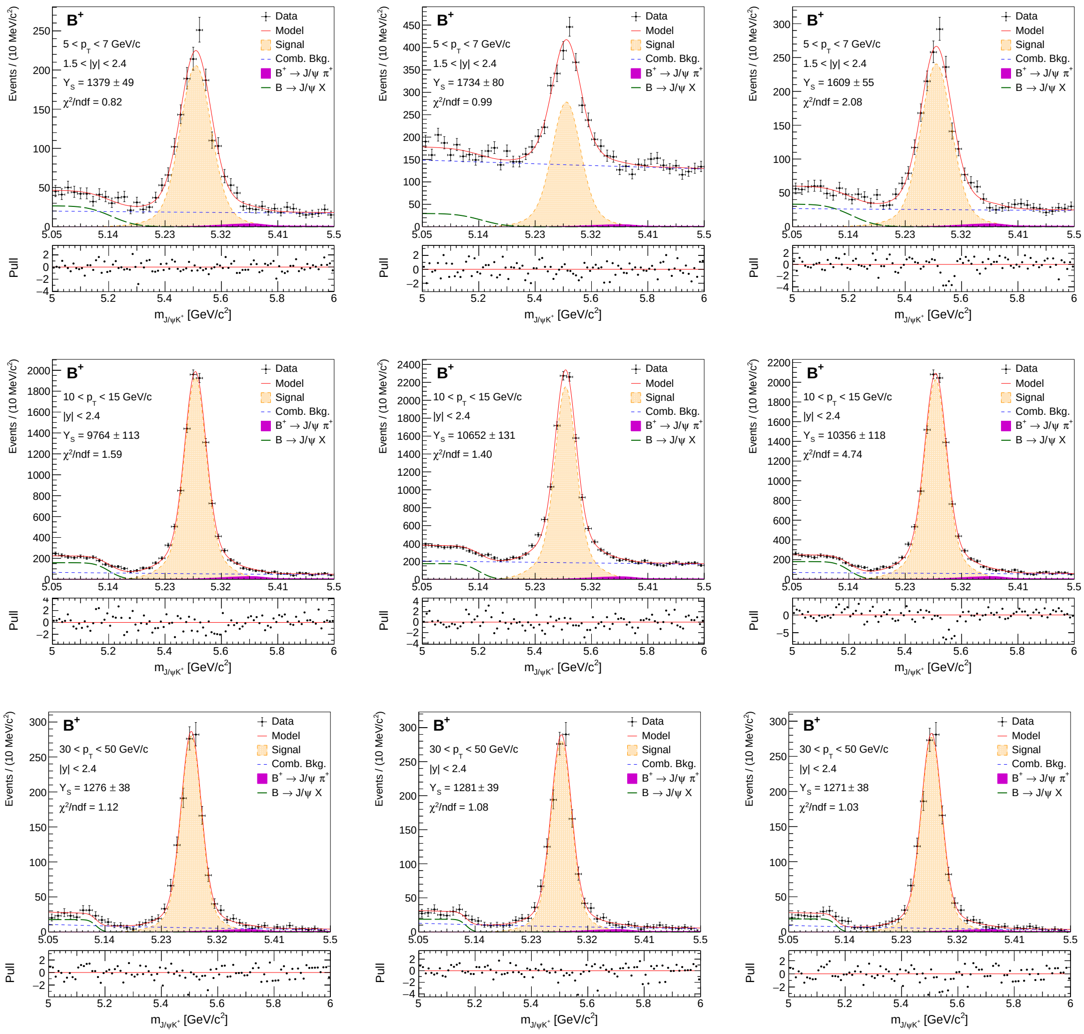

Figure 5.3: Fit to B^+ meson invariant mass distribution, on the bins 5-7 (first row), 10-15 (second row), and 30-50 (third row) GeV/c. On the right column we show the mass fits for our nominal selection, whilst on the first and second columns we show the fits for our optimized selection with the NN model and reference selections, respectively.

Table 5.8: Significance vs $|y|$ on each model, for the B^+ meson.

$\|y\|$	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
Ref.	54.6	52.5	50.4	48.6	22.6
NN	61.5	60.5	60.9	71.9	35.8
Nominal	61.5	60.9	60.8	71.9	35.9

Table 5.9: Significance vs N_{trk} on each model, for the B^+ meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
Ref.	63.2	61.4	47.9	32.4	20.5	12.2	6.3
NN	72.8	78.4	66.2	46.2	29.8	18.3	11.1
Nominal	73.8	78.0	65.7	46.5	29.7	17.3	10.8

Table 5.10: Significance vs p_T on each model, for the B_s^0 meson.

p_T [GeV/c]	7–10	10–15	15–20	20–50
Ref.	4.9	13.3	13.9	9.3
Nominal	6.6	15.7	13.7	14.1

Table 5.11: Significance vs $|y|$ on each model, for the B_s^0 meson.

$\|y\|$	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
Ref.	9.6	10.1	9.7	10.4	5.1
Nominal	12.0	12.7	11.8	11.8	6.0

Table 5.12: Significance vs N_{trk} on each model, for the B_s^0 meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
Ref.	12.6	12.9	11.5	6.5	4.1	3.5	1.7
Nominal	13.3	14.6	13.8	8.2	5.5	4.1	2.1

the possibility of conducting an analysis in this region. Through the fits, we obtain a significance value of 15.2 for our BDT model and 15.8 for our NN model, on the B^+ meson for the $[3, 5]$ GeV/c bin. In Fig. 5.4 we present the mass fit plots of this region. Due to a lack of MC samples for $p_T < 4$ GeV/c we were unable to perform a selection on an even lower p_T region on our B^+ meson, whilst for our B_s^0 meson an attempt to reach the $[3, 5, 7]$ GeV/c bins was made. Given the ML framework and tools developed in this thesis, we will be able to probe into this region with upcoming MC and data samples.

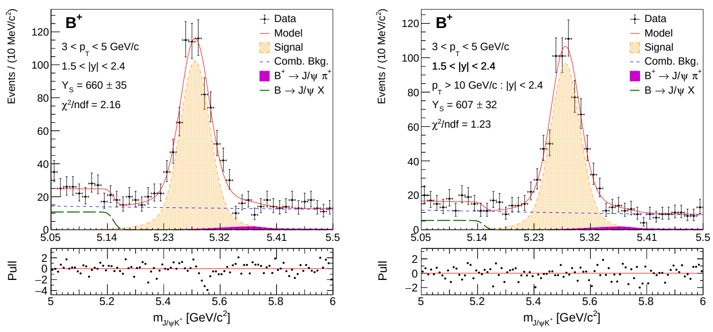

Figure 5.4: Fit to B^+ meson invariant mass distribution, on the 3-5 GeV/c bin. We show the fits for our optimized selection with the BDT (left) and NN (right) models.

Chapter 6

Efficiency

6.1 Detector acceptance and selection efficiency

The signal yields obtained in Sec. 5.4 must be corrected for selection efficiency, ϵ, and detector acceptance, α, in order to obtain the differential production cross sections of the B mesons, Eq. 1.5. The MC samples introduced in Sec. 3.2 are used to estimate both factors. The overall analysis is restricted to the fiducial region. Accordingly, unlike the selection criteria, the fiducial criteria specified in Eq. 3.1 is not corrected for in the efficiency determination.

The acceptance, α, corresponds to the ratio between the number of generated events passing the *muon* and *kaon* acceptance thresholds described in Sec. 3.4, N_{pass}, and the total amount of generated events, N_{GEN}, according to

$$\alpha = \frac{N_{pass}}{N_{GEN}} . \tag{6.1}$$

In turn, the selection efficiency, ϵ, is the fraction of accepted events that pass all the analysis selection criteria, N_{ANpass}, according to

$$\epsilon = \frac{N_{ANpass}}{N_{pass}} . \tag{6.2}$$

The total efficiency factor is then obtained by multiplying the detector acceptance and the selection efficiency factors. We then use its inverse,

$$\frac{1}{\alpha \times \epsilon} = \frac{N_{GEN}}{N_{ANpass}}, \tag{6.3}$$

to correct the previously obtained raw signal yields, following Eq. 1.5.

The total efficiency factor, $\alpha \times \epsilon$ may have a dependence on the event kinematics. When carrying out the ratio in Eq. (6.3) we are implicitly relying not only on the accurate description of the relevant detector effects but also on the simulated kinematic distributions.

For this reason, we produce a fine-grained efficiency map as function of relevant kinematic variables; more precisely, a 2D fine-grained map of our inverse total efficiency vs p_T and rapidity, y, $\mathcal{E}(p_T, y)$. Each entry in this map is obtained from simulation, following Eq. 6.3, and then assigned to candidate events in the data set, using the corresponding (p_T, y) values.

However, this average should be performed using only B meson signal candidates while the data samples contain both signal and background events. To address this issue, we must employ some method to perform background subtraction. This is usually done by restricting the events entering the average to those with an invariant mass close to the B meson nominal mass. However, while this approach does mitigate the problem, it does not fully address it. We therefore suggest a novel method in this work, that involves applying background subtraction utilizing the ${}_s\mathcal{P}lot$ method [1]. More specifically, based on the nominal likelihood fit to the B candidate mass distribution, *s-weights* associated to the signal component are derived. These s-weights can in turn be used to re-weight any event variable, thus projecting out the signal distribution for said variable. In our case, the variable of interest is the inverse efficiency, and we obtain in this way the inverse efficiency distribution for signal only.

The inverse total efficiency, $\frac{1}{\alpha\times\epsilon}$, is finally obtained as the weighted average

$$\langle\frac{1}{\alpha\times\epsilon(p_T, y)}\rangle = \frac{1}{\sum_{i=1}^{N_{Fid}} w_i}\sum_{i=1}^{N_{Fid}} w_i \frac{1}{\alpha\times\epsilon(p_T, y)_i}, \tag{6.4}$$

where the sum runs over the data candidates in our fiducial region, N_{Fid}, and where each event is weighted by its associated s-weight, w.

The results of the intermediate step where the $\alpha\times\epsilon(p_T, y)$ 2D maps are formed are shown in Sec. 6.4 after discussing the improvement made to the $\epsilon(p_T, y)$ 2D map through the application of the *tag-and-probe* method.

In Fig. 6.1, the results for $\langle\frac{1}{\alpha\times\epsilon}\rangle$ of each B meson are shown as functions of the analysis' p_T, y and N_{trk} bins and in Tables 6.1 to 6.6 each bin value can be consulted.

Table 6.1: Average inverse total efficiency factor vs p_T for the B_s^0 meson.

p_T [GeV/c]	7–10	10–15	15–20	20–50
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	57.5 ± 0.5	25.9 ± 0.08	9.24 ± 0.03	5.55 ± 0.04

Table 6.2: Average inverse total efficiency factor vs p_T for the B^+ meson.

p_T [GeV/c]	5–7	7–10	10–15	15–20	20–30	30–50	50–60
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	53.0 ± 0.1	24.0 ± 0.04	11.9 ± 0.006	5.65 ± 0.005	3.72 ± 0.006	3.02 ± 0.03	2.86 ± 0.1

Table 6.3: Average inverse total efficiency factor vs y for the B_s^0 meson.

y	0–0.5	0.5–1	1–1.5	1.5–2	2–2.4
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	15.2 ± 0.09	15.8 ± 0.08	14.9 ± 0.05	19.0 ± 0.09	45.8 ± 0.7

[1] We refer to Sec. 6.5 for a more detailed look at this method.

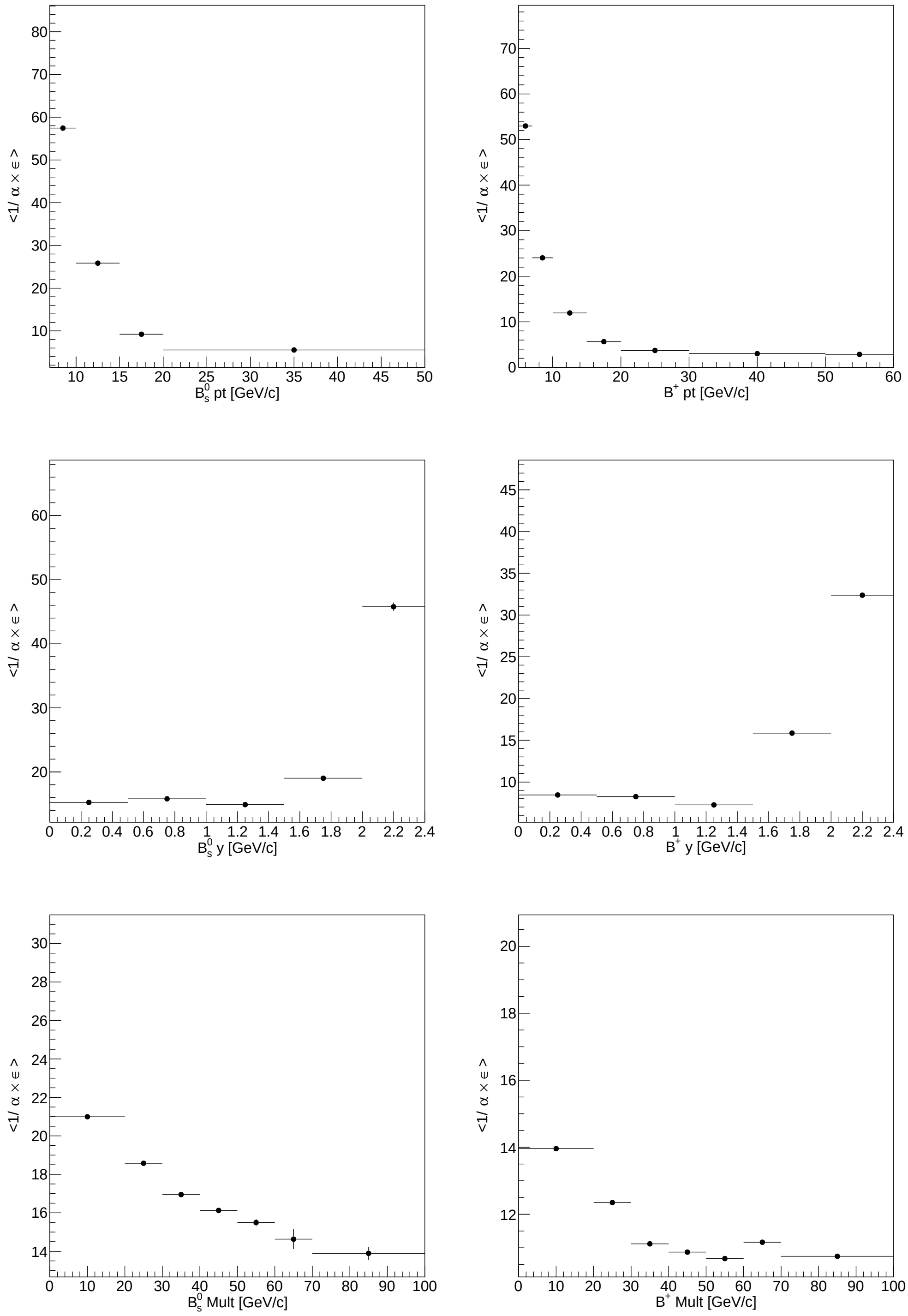

Figure 6.1: The B_s^0(left) and B^+ (right) average of the inverse total efficiency factor as function of the differential p_T (top row), y (middle row) and multiplicity (bottom row) bins used in this analysis.

Table 6.4: Average inverse total efficiency factor vs y for the B^+ meson.

y	0–0.5	0.5–1	1–1.5	1.5–2	2–2.4
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	8.44 ± 0.007	8.24 ± 0.007	7.25 ± 0.005	15.9 ± 0.01	32.4 ± 0.09

Table 6.5: Average inverse total efficiency factor vs N_{trk} for the B_s^0 meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	21.0 ± 0.1	18.6 ± 0.1	16.9 ± 0.1	16.1 ± 0.1	15.5 ± 0.2	14.6 ± 0.5	13.9 ± 0.3

Table 6.6: Average inverse total efficiency factor vs N_{trk} for the B^+ meson.

N_{trk}	0–20	20–30	30–40	40–50	50–60	60–70	70–100
$\langle\frac{1}{\alpha\times\epsilon}\rangle$	14.0 ± 0.02	12.3 ± 0.02	11.1 ± 0.02	10.9 ± 0.02	10.7 ± 0.03	11.2 ± 0.07	10.7 ± 0.08

6.2 Analytic proof of the method

Let an arbitrary dataset have an efficiency value of $\epsilon_0 = \frac{N^{reco}}{N^{Gen}}$, where N^{reco} and N^{Gen} are the number of reconstructed and generated events in our sample, respectively. Consider an efficiency map as a function of any two variables (e.g. a fine-grained map vs p_T and y). Each entry in this map is obtained from MC simulation directly as $\frac{N^{reco}}{N^{Gen}}$. From this 2D map, it is then possible to determine an efficiency value, for a given sample of events, by averaging over our events.

Going back to our previous sample with an efficiency of ϵ_0. We now consider that there exists an M number of events, and that our map is divided among K bins, each associated with a different efficiency value, ϵ_j. Of course, for any of these K bins we will also have corresponding numbers of reconstructed and generated events, N_j^{reco} and N_j^{Gen}, such that $\sum_{i=1}^{K} N_i^{reco} = N^{reco}$ and $\sum_{i=1}^{K} N_i^{Gen} = N^{Gen}$, being the bin efficiency given by $\epsilon_j = \frac{N_j^{reco}}{N_j^{Gen}}$.

The average inverse efficiency is given by

$$< \frac{1}{\epsilon} >= \frac{1}{M}\sum_{j=1}^{K} N_j \frac{1}{\epsilon_j} = \frac{1}{M}\sum_{j=1}^{K} N_j \frac{N_j^{Gen}}{N_j^{reco}}, \tag{6.5}$$

where N_j the number of events that belong to the j^{th} bin. Obviously, we are averaging over the reconstructed events on our sample, and so we get that $N_j = N_j^{reco}$. Henceforth,

$$\frac{1}{M}\sum_{j=1}^{K} N_j \frac{N_j^{Gen}}{N_j^{reco}} = \frac{1}{M}\sum_{j=1}^{K} N_j^{reco} \frac{N_j^{Gen}}{N_j^{reco}} = \frac{1}{\sum_{j=1}^{K} N_j^{reco}}\sum_{j=1}^{K} N_j^{Gen} = \frac{N^{Gen}}{N^{reco}} = \frac{1}{\epsilon_0}. \tag{6.6}$$

We have therefore obtained closure, demonstrating the validity of the $< \frac{1}{\epsilon} >$ calculation method. Note that this is already the term that enters the cross section expression

$$\sigma \sim \frac{N}{L} \times < \frac{1}{\epsilon} > . \tag{6.7}$$

Case study: a failing alternative, $\frac{1}{\langle\epsilon\rangle}$

For completeness, let's investigate an alternative procedure where instead of averaging the inverse efficiency, one would average directly the efficiency values $\frac{1}{<\alpha\times\epsilon>}$. We can see that such a procedure would fail to close,

$$< \epsilon >= \frac{1}{M}\sum_{i=1}^{M}\epsilon_i = \frac{1}{M}\sum_{j=1}^{K}N_j\epsilon_j = \frac{1}{M}\sum_{j=1}^{K}N_j^{reco}\frac{N_j^{reco}}{N_j^{Gen}} = \frac{1}{N^{reco}}\sum_{j=1}^{K}\frac{(N_j^{reco})^2}{N_j^{Gen}}\,, \tag{6.8}$$

which is of course different from ϵ_0, since

$$\begin{aligned}\epsilon_0 = \frac{1}{N^{reco}}\sum_{j=1}^{K}\frac{(N_j^{reco})^2}{N_j^{Gen}} \Leftrightarrow \frac{N^{reco}}{N^{Gen}} &= \frac{1}{N^{reco}}\sum_{j=1}^{K}\frac{(N_j^{reco})^2}{N_j^{Gen}}\\ \Leftrightarrow \frac{(N^{reco})^2}{N^{Gen}} &= \sum_{j=1}^{K}\frac{(N_j^{reco})^2}{N_j^{Gen}}\\ \Leftrightarrow \frac{(\sum_{i=1}^{K}N_i^{reco})^2}{\sum_{i=1}^{K}N_i^{Gen}} &= \sum_{j=1}^{K}\frac{(N_j^{reco})^2}{N_j^{Gen}}.\end{aligned} \tag{6.9}$$

This lack of closure shows that the alternative computation method would fail.

6.3 Closure test

Having analytically demonstrated the working principle of the method, we now check its closure with actual MC samples. The efficiency 2D maps are formed first, and then alternative computations are carried out. Since we are working with MC samples, the s-weights method is not employed here, being all the MC events signal events this method would have no effect here.

Three efficiency computations are obtained:

- **Baseline: 1D**: baseline calculation that takes directly N^{reco} over N^{Gen}; corresponding to ϵ_0 ;
- **Method 2D** $< \frac{1}{\alpha\times\epsilon} >$: the method that is adopted ;
- **Naive 2D** $< \alpha \times \epsilon >$: the alternative method that is not adopted .

The results, obtained without using TnP correction, explained in Sec. 6.4, are shown in Fig. 6.2. We can see that the $< \frac{1}{\alpha\times\epsilon} >$ method converges to 1D results, demonstrating closure. Whilst the alternative $< \alpha \times \epsilon >$ method diverges, showing its non-applicability.

We conclude that the proposed efficiency computation method is suitable in our analyses, with dependence on MC kinematics thus avoided.

Table 6.7: MC averaged Efficiency obtained vs p_T [GeV/c] for each of the 3 different methods, for the B^+ meson.

Eff. values	$5< p_T <7$	$7< p_T <10$	$10< p_T <15$	$15< p_T <20$	$20< p_T <30$	$30< p_T <50$	$50< p_T <60$
1D	0.0185± 0.0002	0.0406± 0.0003	0.0826± 0.0003	0.177± 0.001	0.269±0.002	0.351±0.005	0.394±0.02
2D	0.0209± 0.00001	0.0505± 0.00002	0.099± 0.00001	0.187± 0.00005	0.281±0.0002	0.373±0.001	0.426±0.006
2D Inv.	0.0185± 0.00001	0.0407± 0.00002	0.0826± 0.00001	0.177± 0.00006	0.269±0.0002	0.352±0.001	0.397±0.008

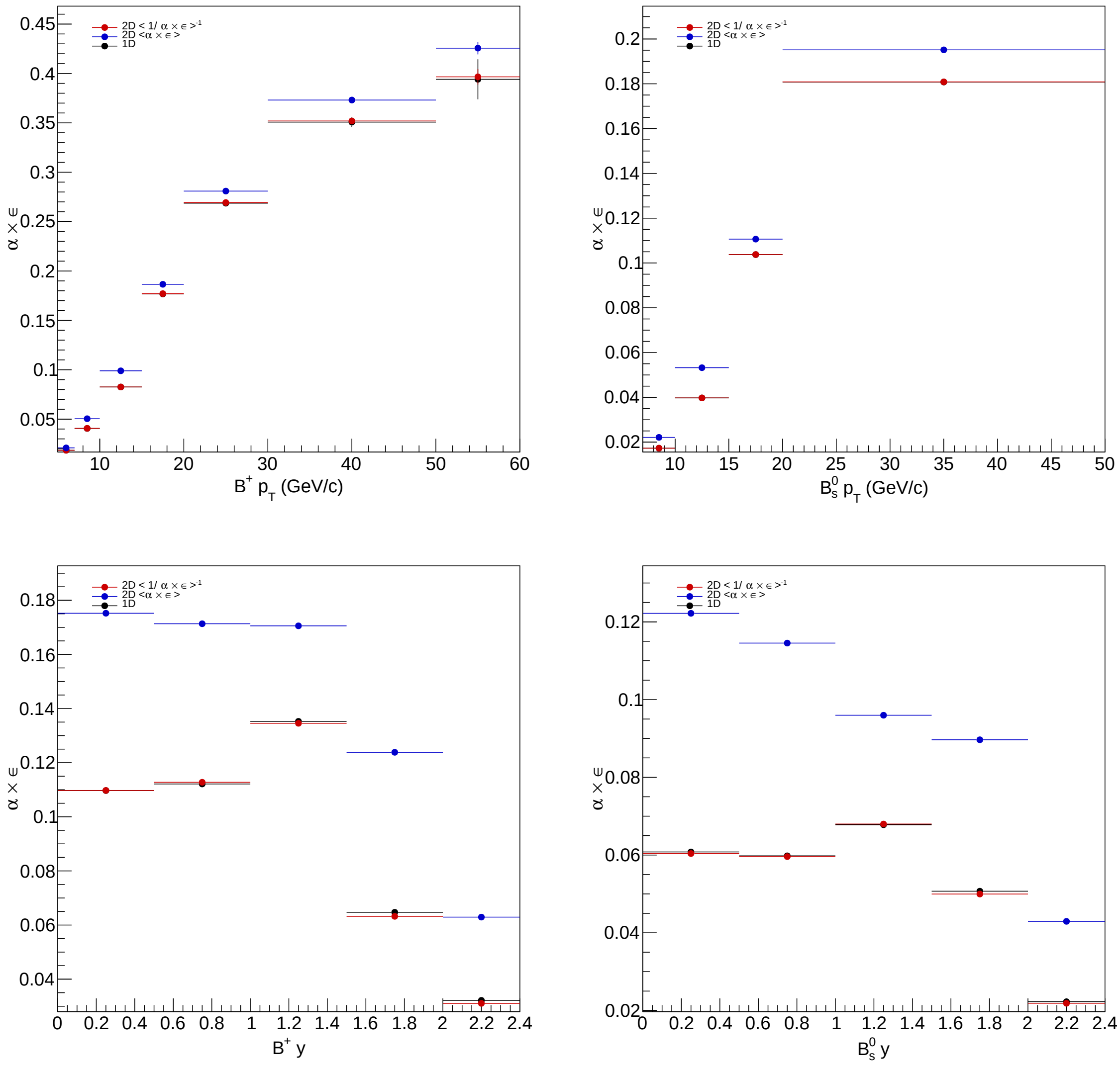

Figure 6.2: Efficiency computations for B^{+}(left) and B_s^0 (right) vs p_T (top) and vs y (bottom). The overlapping black and red points denote the method's closure.

Table 6.8: MC averaged Efficiency vs p_T [GeV/c] obtained for each of the 3 different methods, on the B_s^0 meson.

Eff. values	$7< p_T <10$	$10< p_T <15$	$15< p_T <20$	$20< p_T <50$
1D	0.0173± 0.0002	0.0397± 0.0002	0.104± 0.0006	0.181± 0.001
2D	0.0221± 0.00001	0.0532± 0.00005	0.111± 0.00004	0.195± 0.0002
2D Inv.	0.0173± 0.00002	0.0398± 0.00001	0.104± 0.00004	0.181± 0.0002

Table 6.9: 1D vs 2D efficiency vs p_T [GeV/c] comparisons, for the B^{+} meson.

Relative difference 1D vs:	$5< p_T <7$	$7< p_T <10$	$10< p_T <15$	$15< p_T <20$	$20< p_T <30$	$30< p_T <50$	$50< p_T <60$
2D	13.0%	24.3%	19.9%	5.52%	4.56%	6.37%	7.99%
2D Inv.	0.116%	0.302%	0.06%	0.162%	0.237%	0.336%	0.632%

Table 6.10: 1D vs 2D efficiency vs p_T [GeV/c] comparisons, for the B_s^0 meson.

Relative difference 1D vs:	$7< p_T <10$	$10< p_T <15$	$15< p_T <20$	$20< p_T <50$
2D	28.0%	34.0%	6.66%	7.95%
2D Inv.	0.0891%	0.0277%	0.0277%	0.0373%

Table 6.11: MC averaged Efficiency vs y obtained for each of the 3 different methods, for the B^+ meson.

Eff. values	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
1D	0.11± 0.0009	0.112± 0.0009	0.135± 0.001	0.0647± 0.0004	0.0322±0.0004
2D	0.175± 0.0001	0.171± 0.0001	0.171± 0.0001	0.124± 0.00007	0.0629±0.0002
2D Inv.	0.11± 0.00003	0.113± 0.00003	0.135± 0.00003	0.0632± 0.00002	0.0311±0.00003

Table 6.12: MC averaged Efficiency vs y obtained for each of the 3 different methods, for the B_s^0 meson.

Eff. values	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
1D	0.0608± 0.0006	0.0598± 0.0006	0.0678± 0.0006	0.0507± 0.0004	0.0223±0.0003
2D	0.122± 0.0001	0.115± 0.0001	0.096± 0.0001	0.0897± 0.00007	0.0429±0.0002
2D Inv.	0.0604± 0.00004	0.0596± 0.00003	0.068± 0.00002	0.05± 0.00002	0.0219±0.00003

Table 6.13: 1D vs 2D efficiency vs y comparisons, for the B^+ meson.

Relative difference 1D vs:	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
2D	59.7%	52.8%	26.1%	91.4%	95.6%
2D Inv.	0.0459%	0.562%	0.527%	2.31%	3.44%

Table 6.14: 1D vs 2D efficiency vs y comparisons, for the B_s^0 meson.

Relative difference 1D vs:	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
2D	101%	91.5%	41.6%	76.9%	92.9%
2D Inv.	0.673%	0.331%	0.273%	1.44%	1.81%

6.4 Tag-and-probe correction

In order to reduce the analysis dependence on simulations, it is desirable to obtain the selection efficiency 2D map including data-driven corrections. Hence, the *tag-and-probe* (TnP) method [49] is used to improve the selection efficiency 2D map using weights obtained from data comparison against the MC.

The TnP method reconstructs a well-known resonance from its decay into two daughter particles. This analysis takes into account a J/ψ resonance and its di-muon decay. One of the muons is labelled "tag", while the other is labelled "probe". To ensure that the tag muons are indeed muons, strict selection criteria must be met. In contrast, because probe muons are used to study the efficiency associated with a specific selection criteria, they must meet a very loose selection criteria in order to avoid bias. Probe muons are paired with tag muons so that their invariant mass is similar to that of the J/ψ meson.

There are two invariant mass distributions formed based on whether the probes meet the selection criteria for which the selection efficiency is being estimated, one for successful probes and one for failed probes. The distributions are then fitted within fine transverse momentum and rapidity bins. The resulting signal yields of the mass peaks, $Y_{ok}(p_T^\mu, y^\mu)$ and $Y_{ko}(p_T^\mu, y)^\mu$, are used to calculate the 2D efficiency map of the selection criteria under study, $\epsilon_{TnP}^{sel}(p_T^\mu, y^\mu) = \frac{Y_{ok}(p_T, y)}{Y_{ok}(p_T^\mu, y^\mu)+Y_{ko}(p_T^\mu, y^\mu)}$. Then, the ratio between $\epsilon_{TnP}^{sel}(p_T^\mu, y^\mu)$ and the resulting 2D map following the application of the TnP method to the associated MC sample, $\epsilon_{TnP}^{mc}(p_T^\mu, y^\mu)$, is used to form a 2D muon scaling factor map, $SF_\mu(p_T^\mu, y^\mu) = \frac{\epsilon_{TnP}^{sel}(p_T^\mu, y^\mu)}{\epsilon_{TnP}^{mc}(p_T^\mu, y^\mu)}$. Since each event features two muons, the product between their scaling factors, $SF_{evt} = SF_{\mu_1} \times SF_{\mu_2}$, gives the final scaling factor for that event.

Considering each event scaled by the associated SF_{evt} value results in a re-weighted MC, from it the TnP weighted 2D selection efficiency map, $\epsilon_{TnP}(p_T, y)$, is obtained and preferred over the 2D map simply obtained from Eq. 6.2. The p_T bin widths used when generating the map are $1/8$ GeV/c, $1/4$ GeV/c, $1/2$ GeV/c, and 1 GeV/c in the respective momentum regions of 5 to 10 GeV/c, 10 to 40 GeV/c, 40 to 50 GeV/c and 50 to 60 GeV/c. The considered rapidity binning is $[0\,;\,0.5\,;\,1\,;\,1.5\,;\,2\,;\,2.4]$. Finally, according to the discussion in Sec. 6.1, multiplying by $\alpha(p_T, y)$ and taking its inverse the TnP weighted $\frac{1}{\alpha(p_T,y)\times\epsilon(p_T,y)}$ 2D map for the B^+ and B_s^0 mesons are obtained and shown in Fig. 6.3.

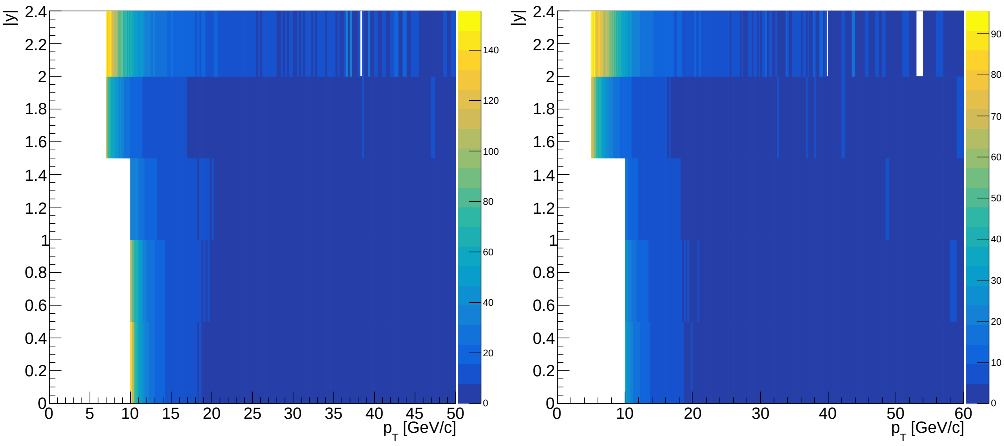

Figure 6.3: TnP weighted 2D maps of $\frac{1}{\alpha\times\epsilon}$ as a function of p_T and $|y|$ for B_s^0 (left) and B^+ (right).

6.5 MC validation

MC samples are used in this section of the analysis to estimate the selection efficiency and detector acceptance. However, there is no reason to believe that the MC samples faithfully reproduce the data distributions. In fact, many factors, such as incorrect physics process modelling in the event generation step or unaccounted detector effects in the reconstruction step, can cause MC simulations to deviate from nature and have a significant impact on the final results. Therefore, a comparison of data and MC is required. Because the MC samples only contain signal events, such a comparison can only be performed after separating the signal events from the background events in the data. This is achieved using $_s\mathcal{P}lot$ [50], a technique that has the key advantage of using the full data sample for the comparison.

The $_s\mathcal{P}lot$ is a likelihood-based technique for extracting the distributions of relevant variables, such as p_T, y, and so on, for signal or background events in a data sample. These variables must be unrelated to the variable entering the likelihood expression, the final state system's invariant mass. It is used to assign two weights to each event: a probability of being signal, $\mathcal{W}_s(m_i)$, and a probability of being background, $\mathcal{W}_b(m_i)$, with $\mathcal{W}_s(m_i) + \mathcal{W}_b(m_i) = 1$ and $\mathcal{W}_n(m_i) = \frac{\sum_j V_{nj} f_j(m_i)}{\sum_k Y_k f_k(m_i)}$,, where V_{nj} is the covariance matrix and the sum runs over the j-th and k-th model components. It is possible to plot each desired variable distribution for the respective data component by applying the weights to the respective event.

The `RooStats::SPlot` class implements the $_s\mathcal{P}lot$ technique. The distributions are normalized across the range of each variable because it is their shape that should be considered. Each plot includes a bin-by-bin ratio calculated with the `TRatioPlot` class in the bottom panel. Figure 6.4 shows a comparison of the data signal p_T, y, and multiplicity distributions against the corresponding MC distributions.

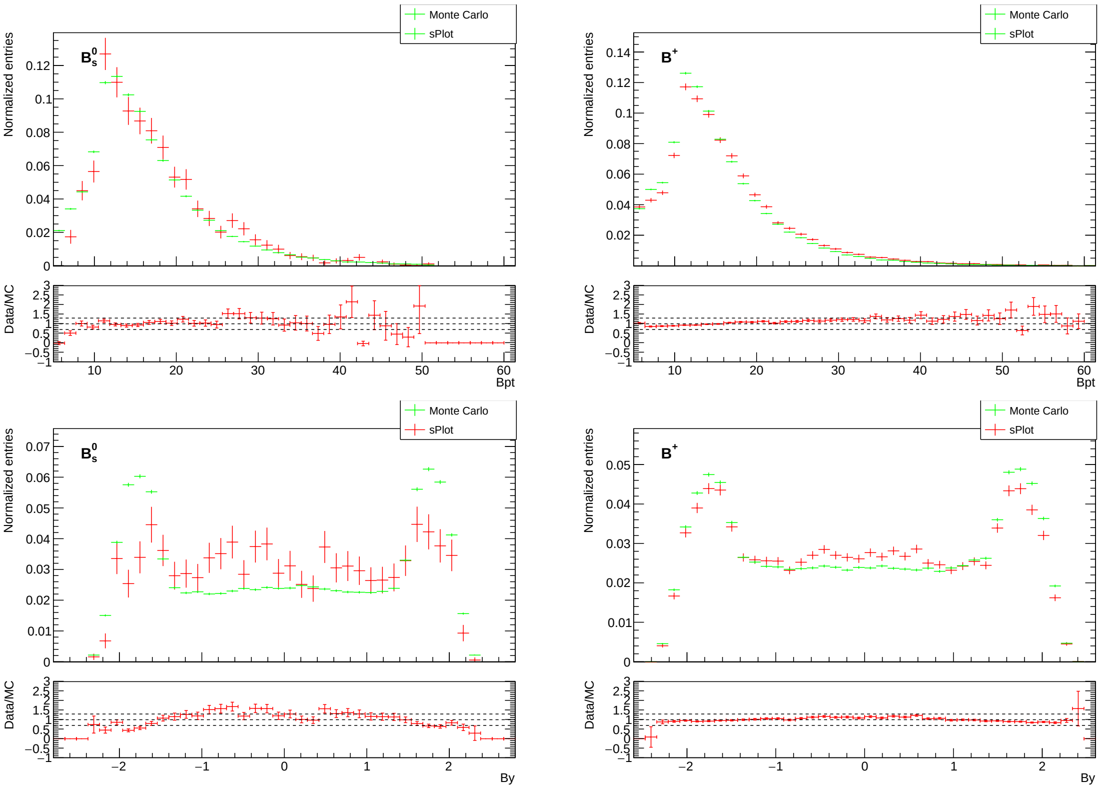

Figure 6.4: Comparison of B_s^0 (left) and B^+ (right) p_T (top row), rapidity (bottom row) distributions in data and MC using the $_s\mathcal{P}lot$ technique.

6.6 1D method comparison

It is, of course, of interest to compare the efficiency values we are obtaining to those that we would get from relying solely on the MC samples. If both Data and MC samples are in agreement, then we expect to obtain close results between these two methods. The comparisons are shown on Fig. 6.5 with the relative differences shown on Tables 6.15 to 6.18.

Like it was to be expected, since in Sec. 6.5 we saw the MC and signal distributions were in reasonably good agreement, we see that for p_T the 2D method results are in accordance with the 1D results, being the highest discrepancy for the central rapidity bins as predicted since it is in this region that we see a higher discrepancy between the data and MC distributions.

Table 6.15: 1D vs 2D efficiency vs p_T [GeV/c] comparisons, for the B^+ meson.

Relative difference 1D vs:	5< p_T <7	7< p_T <10	10< p_T <15	15< p_T <20	20< p_T <30	30< p_T <50	50< p_T <60
2D	2.34%	0.67%	0.445%	0.26%	1.12%	3.61%	8.88%

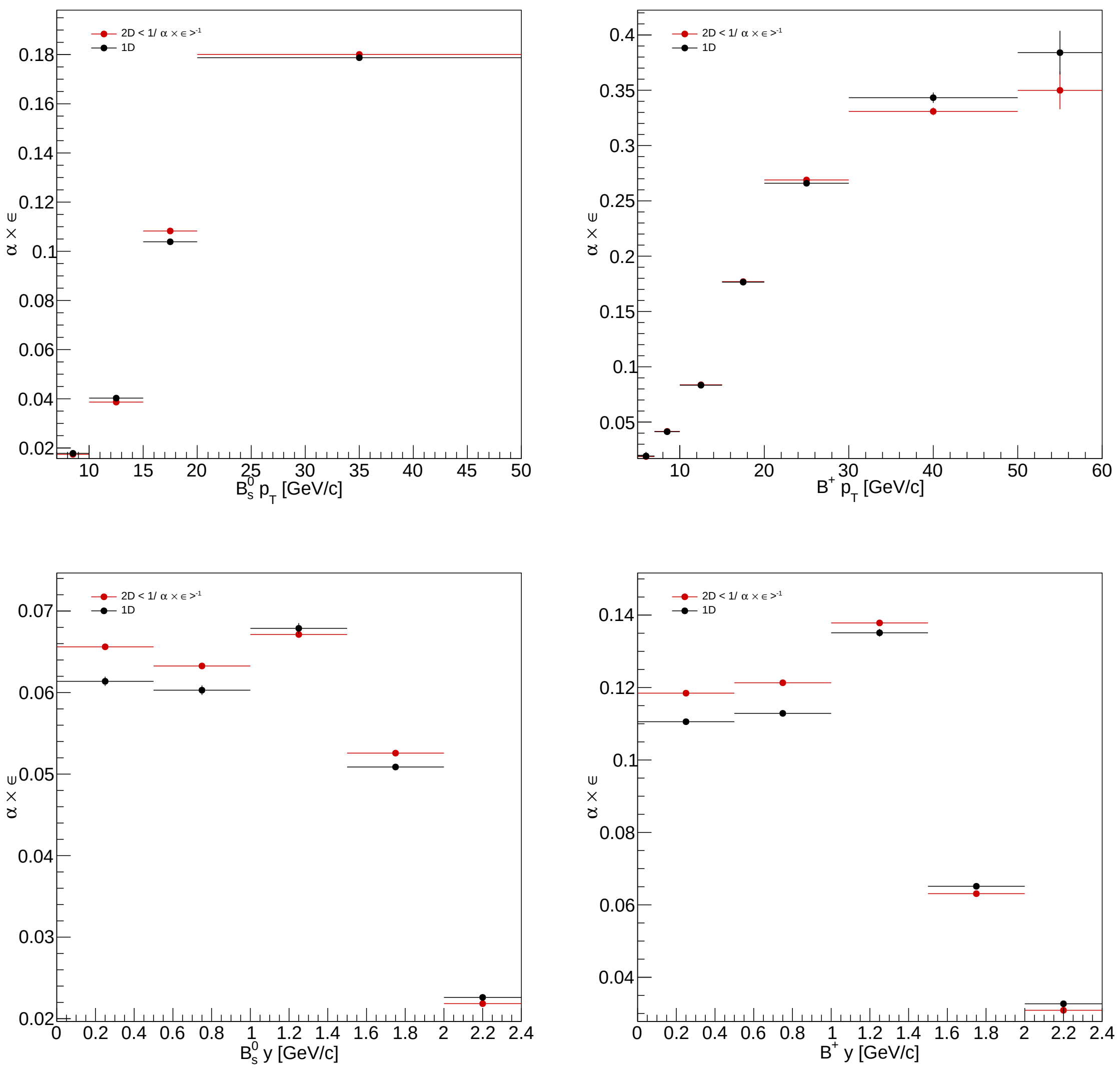

Figure 6.5: Comparison of B_s^0 and B^+ efficiency vs p_T (top row), rapidity (middle row) and multiplicity (bottom row) between the 1D and 2D methods.

Table 6.16: 1D vs 2D efficiency vs p_T [GeV/c] comparisons, for the B_s^0 meson.

Relative difference 1D vs:	7< p_T <10	10< p_T <15	15< p_T <20	20< p_T <50
2D	2.3%	4.0%	4.23%	0.746%

Table 6.17: 1D vs 2D efficiency vs $|y|$ comparisons, for the B^+ meson.

Relative difference 1D vs:	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
2D	7.13%	7.48%	2.02%	3.14%	5.48%

Table 6.18: 1D vs 2D efficiency vs $|y|$ comparisons, for the B_s^0 meson.

Relative difference 1D vs:	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
2D	6.89%	4.93%	1.11%	3.35%	3.39%

Chapter 7

Systematic uncertainties

Various systematic uncertainties arise from the differential signal yields extraction and the total efficiency computation steps, as well as two global (to all bins equally) uncertainty sources, that affect our B mesons' differential production cross sections, given by Eq. 1.5. All of these uncertainty sources are discussed in Sections 7.1, 7.2 and 7.3.

The total systematic uncertainties are computed as the sum in quadrature of the different contributions, summarized in Tables 7.1 to 7.6 for the B^+ and B_s^0 mesons.

Table 7.1: Summary of relative systematic uncertainties, in %, of p_T bins from the B^+ cross section calculations.

	p_T [GeV/c]						
Sources	5-7	7-10	10-15	15-20	20-30	30-50	50-60
Tracking efficiency	2.4	2.4	2.4	2.4	2.4	2.4	2.4
Data vs MC	0.14	0.027	0.14	0.050	0.28	0.46	0.0023
PDF variation	2.2	0.72	8.8	1.6	0.84	2.1	7.3
Track selection	0.29	0.76	0.36	0.40	0.37	0.15	1.4
p_T shape	0.008	0.004	0.01	0.0096	0.004	0.006	0.02
TnP variations	0.47	0.45	0.37	0.36	0.43	0.62	0.66
sum	3.31	2.7	9.1	2.9	2.6	3.3	7.8
Luminosity, $\mathcal{L}$				1.9			
Branching fractions				2.9			
Sum				3.5			

7.1 Yield systematics

The uncertainty resulting from signal yield extraction is assessed by taking into account variations of the fitting model components one at a time. Using the Occam's razor principle, their complexity should not be excessively increased, and so three variations are considered for the signal peak, where the model PDF is given by (i) a sum of a Gaussian function and a crystal ball function, (ii) a sum of three Gaussian functions, and (iii) fixing the mean to its MC learned value. Another three variations

Table 7.2: Summary of relative systematic uncertainties, in %, of p_T bins from the B_s^0 cross section calculations.

	p_T [GeV/c]			
Sources	7-10	10-15	15-20	20-50
Tracking efficiency	4.8	4.8	4.8	4.8
Data vs MC	0.42	0.41	0.015	0.31
PDF variation	4.1	2.5	4.3	4.2
Track selection	0.44	0.85	2.2	0.36
p_T shape	0.04	0.02	0.004	0.008
TnP variations	0.46	0.38	0.35	0.45
sum	6.4	5.5	6.8	6.4
Luminosity, $\mathcal{L}$	1.9			
Branching fractions	7.5			
Sum	7.7			

Table 7.3: Summary of relative systematic uncertainties, in %, of $|y|$ bins from the B^+ cross section calculations.

	$\|y\|$				
Sources	0.0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.4
Tracking efficiency	2.4	2.4	2.4	2.4	2.4
Data vs MC	0.32	0.28	0.14	0.033	0.031
PDF variation	0.41	0.99	0.45	10	0.97
Track selection	0.63	1.3	0.16	0.77	0.43
p_T shape	0.0003	0.0005	0.001	0.03	0.04
TnP variations	0.30	0.31	0.42	0.47	0.46
sum	2.6	2.9	2.5	10.4	2.7
Luminosity, $\mathcal{L}$	1.9				
Branching fractions	2.9				
Sum	3.5				

for the background model are considered: (iv) a first order polynomial function, (v) a second order polynomial function, and (vi) varying the mass fitting range from $[5, 6]$ GeV/c^2 to $[5.19, 6]$ GeV/c^2 and excluding the partially reconstructed background model if it is the B^+ meson. In addition to these, the B^+ meson will see one more variation involving the PDF modelling the Cabibbo suppressed model where $R_{PDG} \rightarrow R_{AN} = 0.0517 \pm 0.0004$. Systematic uncertainty is calculated by adding the maximum uncertainty of the signal and background variations in quadrature. The systematic uncertainties resulting from each considered signal and background model variation are shown in Fig. 7.1. In Appendix C.1 we present the tabled results, alongside the extracted differential signal yields normalized by the respective bin width for each B meson with their systematic and statistical uncertainties.

To quantify the systematic uncertainty associated with the B^+ meson Cabibbo suppressed background model, R_{PDG} is changed to the uncertainty weighted average value obtained in the differential analysis performed on the $B^+ \rightarrow J/\psi\pi^+$ MC sample. Although the used MC has a good distribution shape, the branching fraction of the $B^+ \rightarrow J/\psi\pi^+$ in `EVTGEN` is incorrect, resulting in a $\sim 30\%$ differ-

Table 7.4: Summary of relative systematic uncertainties, in %, of $|y|$ bins from the B_s^0 cross section calculations.

	$\|y\|$				
Sources	0.0-0.5	0.5-1.0	1.0-1.5	1.5-2.0	2.0-2.4
Tracking efficiency	4.8	4.8	4.8	4.8	4.8
Data vs MC	1.03	0.93	0.28	0.26	0.75
PDF variation	1.8	1.5	2.6	4.2	9.6
Track selection	5.9	2.1	1.7	0.42	2.21
p_T shape	0.0008	0.0009	0.0006	0.01	0.14
TnP variations	0.30	0.32	0.42	0.46	0.46
sum	7.9	5.5	5.7	6.44	11.0
Luminosity, $\mathcal{L}$			1.9		
Branching fractions			7.5		
Sum			7.7		

Table 7.5: Summary of relative systematic uncertainties, in %, of N_{trk} bins from the B^+ cross section calculations.

	N_{trk}						
Sources	0-20	20-30	30-40	40-50	50-60	60-70	70-100
Tracking efficiency	2.4	2.4	2.4	2.4	2.4	2.4	2.4
Data vs MC	0.098	0.095	0.12	0.071	0.010	0.066	0.079
PDF variation	1.2	1.4	1.9	5.5	2.5	0.72	4.7
Track selection	0.17	0.81	0.81	0.31	0.077	1.1	1.3
p_T shape	0.007	0.002	0.01	0.001	0.002	0.0004	0.04
TnP variations	0.42	0.42	0.42	0.42	0.42	0.41	0.41
sum	2.7	2.9	3.2	6.1	3.5	2.8	5.5
Luminosity, $\mathcal{L}$				1.9			
Branching fractions				2.9			
Sum				3.5			

ence between $R_{AN} = \frac{Y_{J/\psi\pi^+}}{Y_{J/\psi K^+}}$ and the value of R_{PDG} obtained from the PDG. Nonetheless, because it represents a very small portion of the data, its impact on the final results is negligible, and no further uncertainty studies involving the variation of the chosen PDF are required. From Fig. 7.1 it follows that using the R_{AN} value results in a very small relative difference between the nominal and newly extracted signal yields. This variation did not contribute to any B^+ bin's total systematic uncertainty.

Inspecting the systematic uncertainties in the B_s^0 and B^+ yields resulting from the (vi) variation reveals that the fit and the model are not dependent on the shape of the background to the left of the mass resonance to successfully extract the signal yields. Furthermore, for the B^+ meson, this means that no additional systematic uncertainty studies involving the discarded background PDF, a complementary error function, are required. In fact, few bins saw this variation contribute to the overall systematic uncertainty. It was, however, small, as high as 3.5%, and statistical uncertainty covers it.

Increasing the flexibility of the respective B meson fitting model by using a triple Gaussian as the PDF of the signal model, variation (ii), resulted in a good overall convergence of the fit. According to

Table 7.6: Summary of relative systematic uncertainties, in %, of N_{trk} bins from the B_s^0 cross section calculations.

	N_{trk}						
Sources	0-20	20-30	30-40	40-50	50-60	60-70	70-100
Tracking efficiency	4.8	4.8	4.8	4.8	4.8	4.8	4.8
Data vs MC	0.63	0.090	00.36	0.39	0.37	0.026	1.1
PDF variation	2.9	3.2	3.2	3.6	5.0	2.2	9.5
Track selection	0.37	2.2	2.2	3.0	4.2	1.4	8.3
p_T shape	0.02	0.02	0.02	0.02	0.0009	0.03	0.03
TnP variations	0.40	0.40	0.41	0.41	0.39	0.40	0.42
sum	5.7	6.2	5.8	6.7	8.1	5.5	13.6
Luminosity, $\mathcal{L}$				1.9			
Branching fractions				7.5			
Sum				7.7			

the overview in Fig. 7.1, this signal variation did not produce large systematic uncertainties, even in the bins with the most deviated values, and thus contributed to the final value of the bin systematic uncertainty. The conclusion drawn from this variation is that the remarkably simple nominal model considered describing the signal peak, a sum of two Gaussian PDFs, accurately describes the data. This conclusion is supported by the signal variation involving the sum of a crystal ball and a Gaussian PDF, which is found to contribute to the final systematic uncertainty only in 2 bins, yielding at worst close to the smallest systematic uncertainties in any other bin. A crystal ball PDF is made up of a Gaussian PDF and a power-law low-end tail PDF and is commonly used to model processes in which the final state particles lose energy through radiation. However, the very low systematic uncertainties discovered suggest that this scenario did not apply. In this case, the crystal ball PDF is reduced to a single Gaussian PDF, and summing another Gaussian function produces the nominal signal PDF.

7.2 Efficiency systematics

7.2.1 Data vs MC comparison

The Data and the MC are not entirely in agreement. To estimate this effect in the efficiency calculation, we use ${}_s\mathcal{P}lot$ techniques to create a MC-correction weight from the Data/MC ratio. The study yields such Data/MC comparisons for various kinematic variables, as well as the BDT output score. We use the normalized-to-unity Data/MC ratio as a weight to quantify the MC/Data discrepancy in the efficiency calculation procedure. Essentially, the Data/MC ratio is calculated as a function of the BDT score. This ratio in then employed to re-weight the MC sample, and the efficiency is re-computed using this re-weighted MC sample. The BDT-weighted efficiency's percent deviation from the nominal unweighted efficiency is determined as the associated systematic effect. Figure 7.2 depicts the BDT-weighted and unweighted $\frac{1}{\alpha\times\epsilon}$ for B^+ and B_s^0. The systematic uncertainties are presented in Appendix C.2.1.

As can be seen, this is a small source of systematic uncertainty, as would be expected if we note

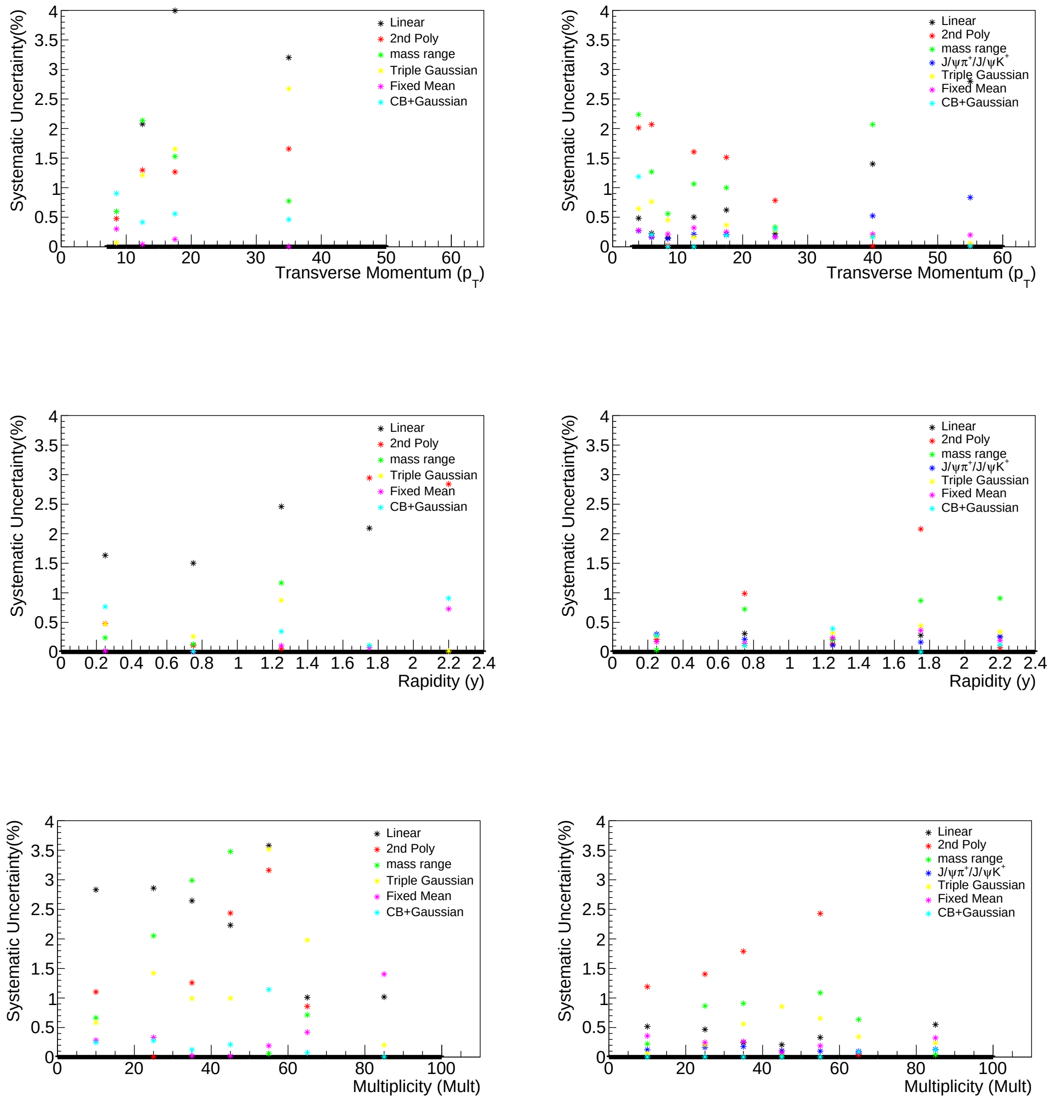

Figure 7.1: Summary of the p_T (top row), $|y|$ (middle row) and multiplicity (bottom row) signal and background systematic uncertainties arising from each model variation for the B_s^0 (left) and B^+ (right) mesons.

that there is a good agreement between our MC and signal distributions, as can be seen in Fig. 6.4. In accordance to this, the central rapidity bins show a higher uncertainty since this region also presents the higher discrepancy between signal and MC.

7.2.2 Muon efficiency: tag-and-probe

Because we use tag-and-probe scale factor weighted efficiency as the nominal, a common approach is to vary the TnP scale factor up and down based on its statistical and systematic uncertainties. Our systematic uncertainties are defined as the difference between the nominal and up/down varied values. In Fig. 7.3, the nominal efficiency and the comparison with the up/down TnP scale factor for B^+ and B_s^0

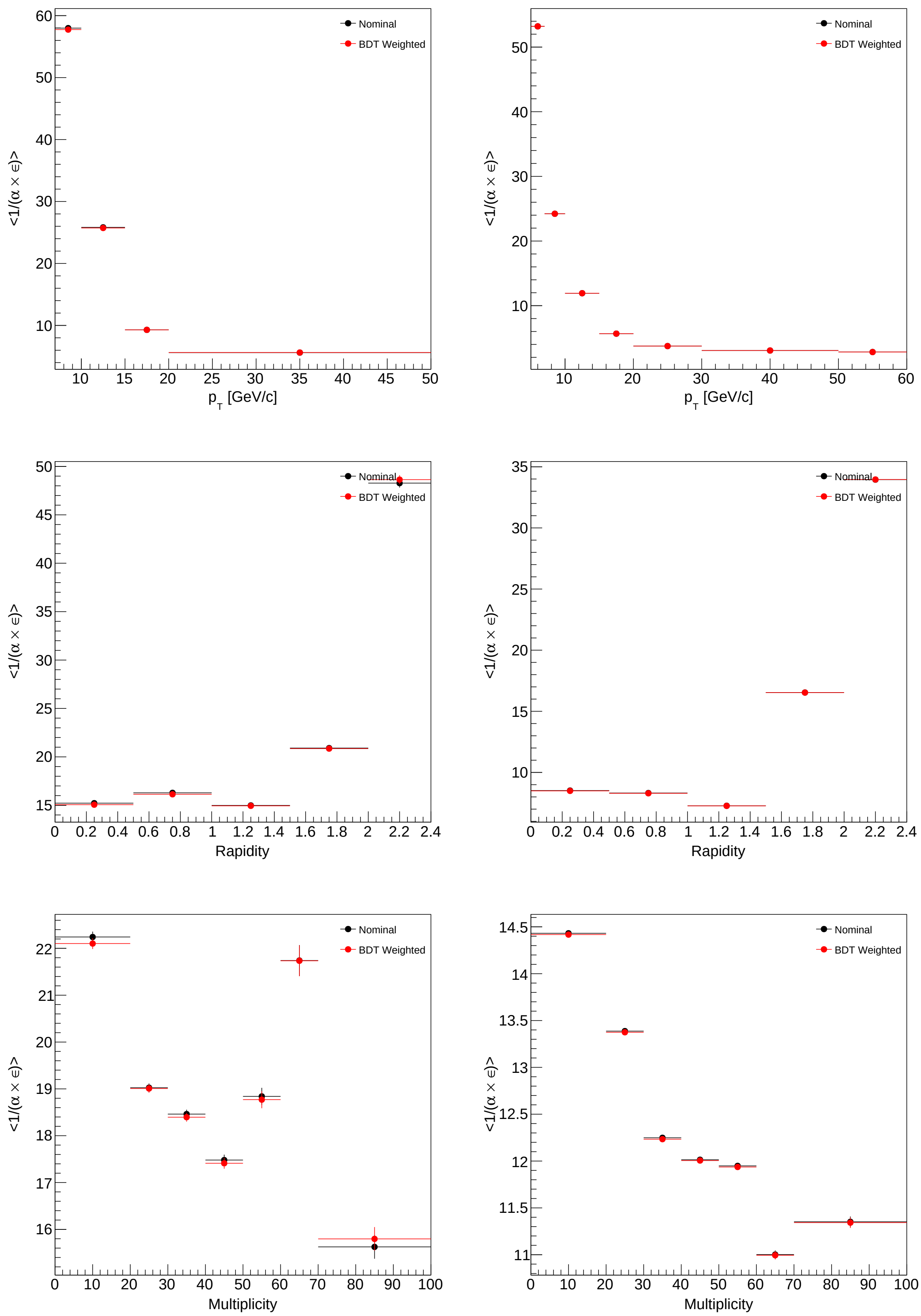

Figure 7.2: Comparison of B_s^0 (first column) and B^+ (second column) efficiencies vs p_T (first row), vs rapidity (second row) and vs multiplicity (third row) with BDT weights.

as a function of B-meson p_T are shown. The systematic uncertainties associated to the TnP correction are depicted in Appendix C.2.2.

7.2.3 p_T shape

The assumption of the B meson p_T shape in our MC samples has an effect in our final results. To quantify its effect on the efficiency calculation, we can compare our MC p_T distribution with another MC sample that presents a differing p_T shape. We do this by weighting our samples by a p_T weight, obtained from the FONLL/MC ratios, similarly to what we did for the Data/MC discrepancy. The ratio of their normalized distributions is then calculated. Then, based on the kinematics of the B-meson candidates, we fit the FONLL/MC ratio and use the fit function as a weight in the efficiency calculation.

Figure 7.4 shows the p_T-weighted and unweighted $\frac{1}{\alpha\times\epsilon}$ factor for B^+ and B_s^0. The p_T weight systematic uncertainties are shown in Appendix C.2.3.

We highlight that this is a very small source of uncertainty, which reinforces the robustness of our efficiency calculation method. Since we are utilizing fine-grained 2D maps and averaging over the data kinematics, we are already minimizing the dependency on MC kinematics, including its p_T shape, making so that changing this distribution yields negligible differences.

7.2.4 Track selection

The systematic uncertainty from the track selection criteria is computed by comparing the nominal cross-sections with those obtained using a tighter selection, whose difference is $\frac{\Delta_{p_T}}{p_T} < 0.05$ and $\frac{\chi^2/ndf}{N_{layers}} < 0.15$, where Δ_{p_T} is our events' p_T uncertainty and N_{layers} is the sum of the number of pixel and strip layer hits. The resulting deviation is quoted as the systematic uncertainty.

7.3 Global systematics

There exist two sources of uncertainties that have a global effect across the analysis bins. The uncertainties associated to the branching fraction of the decay $B^+ \to J/\psi K^+$, with $J/\psi \to \mu^+\mu^-$, is calculated by adding in the quadrature the uncertainties of each sub-channel. The resulting uncertainty for the full decay chain is 2.9% for B^+ and 7.5% for B_s^0.

The difference in the track reconstruction efficiency in data and simulation is estimated by comparing 3-prong and 5-prong D^+ decays, $D^* \to K\pi\pi(\pi\pi)$, resulting in a 2.4% uncertainty for each hadronic track.

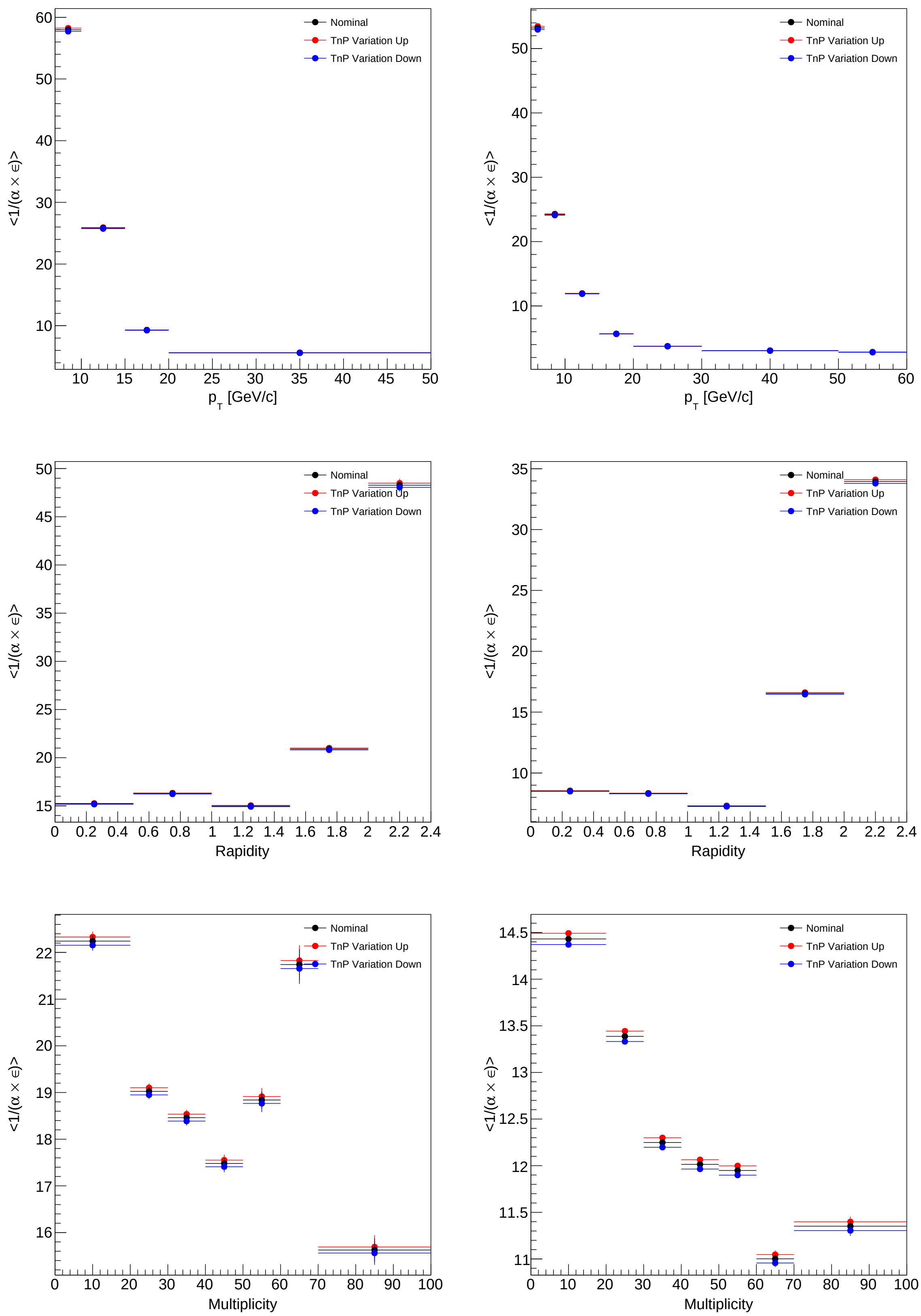

Figure 7.3: Comparison of B_s^0 (first column) and B^+ (second column) efficiencies vs p_T (first row), vs rapidity (second row) and vs multiplicity (third row) with tag and probe variation up and down.

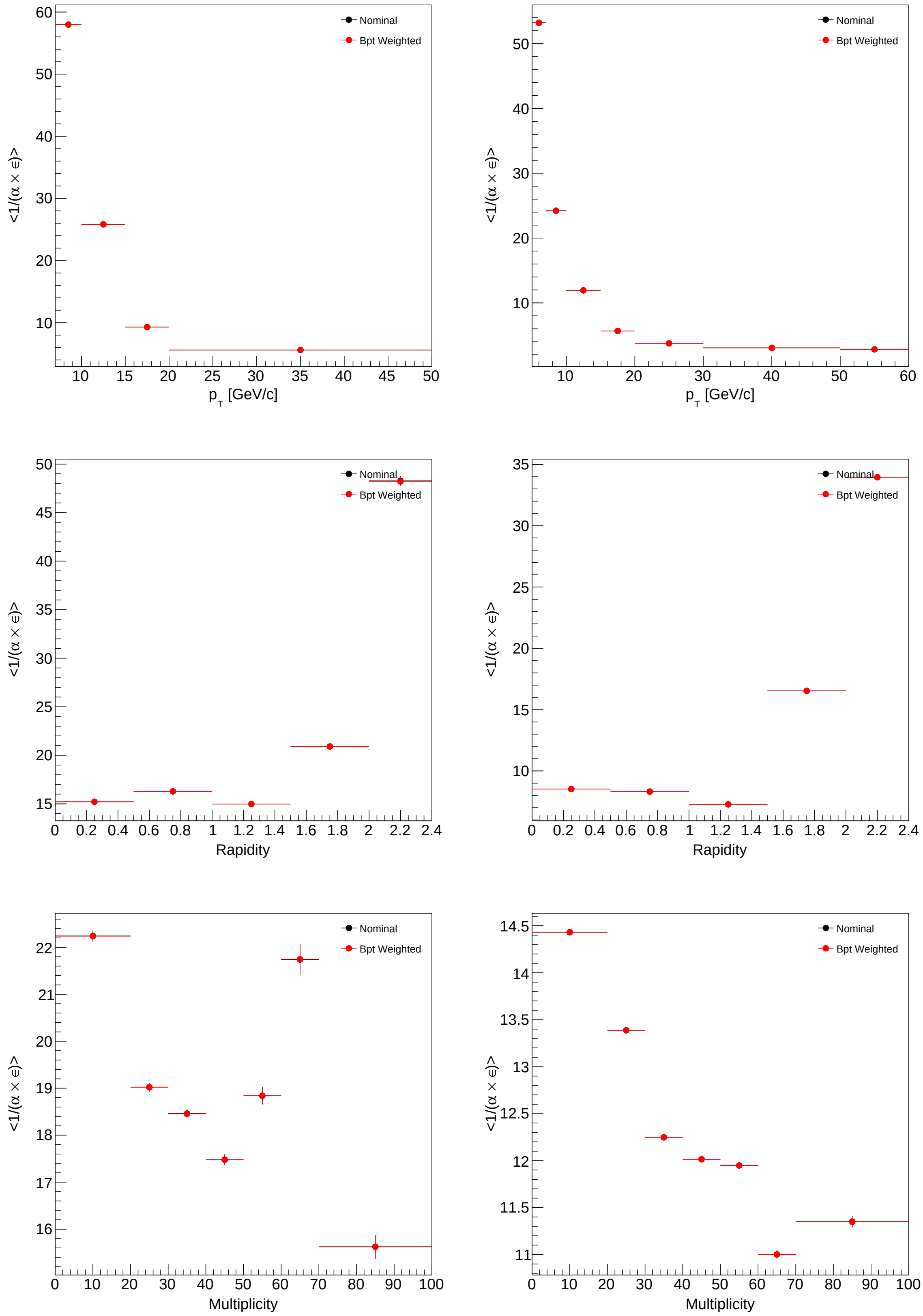

Figure 7.4: Comparison of B_s^0 (first column) and B^+ (second column) efficiencies vs p_T (first row), vs rapidity (second row) and vs multiplicity (third row) with p_T weights.

Chapter 8

Results

8.1 Cross-sections

Having obtained the p_T, $|y|$ and N_{trk} differential signal yields, in Sec. 5.4, and the respective efficiency factors, $\frac{1}{\epsilon\times\alpha}$, in Sec. 6.1, along with $\mathcal{L}$ from Sec. 3.1 and $\mathcal{B}$ from Sec. 1.3, we can now compute the B^+ and B^0_s meson production cross section, following Eq. 1.5. The measurements are provided for the p_T bins $[5, 7, 10, 15, 20, 30, 50, 60]$ GeV/c for the B^+ meson and $[7, 10, 15, 20, 50]$ GeV/c for the B^0_s meson as well as the $|y|$ bins $[0.0, 0.5, 1.0, 1.5, 2.0, 2.4]$ and the N_{trk} bins $[0, 20, 30, 40, 50, 60, 70, 100]$.

The measured differential cross sections are presented in Fig. 8.1. We present the FONLL reference production spectra vs p_T for our B mesons, to serve as a comparison between our measured pp spectra and the FONLL theoretical predictions. We can see that the one sigma theoretical uncertainty of each bin covers the corresponding experimental results and their associated uncertainties, with the exception of the B^+ $[50, 60]$ GeV/c bin and the B^0_s $[7, 10]$ and $[20, 50]$ bins, which border the outskirts of this theoretical uncertainty. Nevertheless, we see a very good agreement between our results and the theory predictions, as can be further seen by the ration between the measured production cross sections and the FONLL reference production cross sections, presented in the lower panel of the p_T plots in Fig. 8.1, which oscillate around unity.

For the rapidity differential cross sections it is more difficult to evaluate the possible dependencies at play since here we employ two different fiducial regions, $|y| < 1.5 : p_T > 10$ GeV/c and $|y| > 1.5 : 5 < p_T < 60$ GeV/c for the B^+ meson or $|y| > 1.5 : 7 < p_T < 50$ GeV/c for the B^0_s meson. This renders direct comparisons for the two regions unfeasible.

8.2 Fragmentation fractions

The cross-sections presented in Sec. 8.1 have an unmatching p_T binning between the B^+ and B^0_s mesons. For computing the ratio between the two mesons, we re-compute the B^+ cross-sections using the same p_T binning as adopted for the B^0_s meson. The results are shown in Figs. 8.2, 8.3 and 8.4, with the relevant cross-sections presented in Appendix B.

p_T [GeV/c]	$\frac{d\sigma}{dp_T}$ [pb c/GeV]	Relative Stat.	Unc.(%) Syst.
5–7	1170581	3.40	4.82
7–10	646581	2.36	4.39
10–15	679555	1.14	9.75
15–20	188829	1.47	4.55
20–30	43019	1.73	4.37
30–50	5275	3.17	4.79
50–60	810	11.9	8.58

Table 8.1: Summary of the measured B^+ differential cross section vs p_T.

p_T [GeV/c]	$\frac{d\sigma}{dp_T}$ [pb c/GeV]	Relative Stat.	Unc.(%) Syst.
7–10	138776	10.0	9.99
10–15	167190	4.48	9.45
15–20	40054	5.43	10.3
20–50	3972	5.40	10.0

Table 8.2: Summary of the measured B_s^0 differential cross section vs p_T.

$\|y\|$	$\frac{d\sigma}{d\|y\|}$ [pb]	Stat. Unc.(%)	Syst. Unc.(%)
0.0–0.5	2425451	1.44	4.33
0.5–1.0	2334683	1.50	4.57
1.0–1.5	2131203	1.55	4.29
1.5–2.0	7980027	1.31	11.0
2.0–2.4	5377197	2.97	4.40

Table 8.3: Summary of the measured B^+ differential cross section vs $|y|$.

$\|y\|$	$\frac{d\sigma}{d\|y\|}$ [pb]	Stat. Unc.(%)	Syst. Unc.(%)
0.0–0.5	519576	5.85	11.0
0.5–1.0	574375	5.70	9.48
1.0–1.5	509576	6.10	9.61
1.5–2.0	879535	5.55	10.0
2.0–2.4	688615	11.5	13.4

Table 8.4: Summary of the measured B_s^0 differential cross section vs $|y|$.

N_{trk}	$\frac{d\sigma}{dN_{trk}}$ [pb]	Stat. Unc.(%)	Syst. Unc.(%)
0–20	134167	1.92	4.45
20–30	299766	1.83	4.57
30–40	214578	1.96	4.73
40–50	119190	2.38	7.00
50–60	53719	3.19	4.95
60–70	21408	5.12	4.46
70–100	3030	7.50	6.5

Table 8.5: Summary of the measured B^+ differential cross section vs N_{trk}.

N_{trk}	$\frac{d\sigma}{dN_{trk}}$ [pb]	Stat. Unc.(%)	Syst. Unc.(%)
0–20	19253	8.29	9.56
20–30	46729	7.77	9.87
30–40	41252	7.84	9.62
40–50	18794	9.72	10.2
50–60	9232	12.7	11.2
60–70	4443	16.6	9.46
70–100	559	26.1	15.6

Table 8.6: Summary of the measured B_s^0 differential cross section vs N_{trk}.

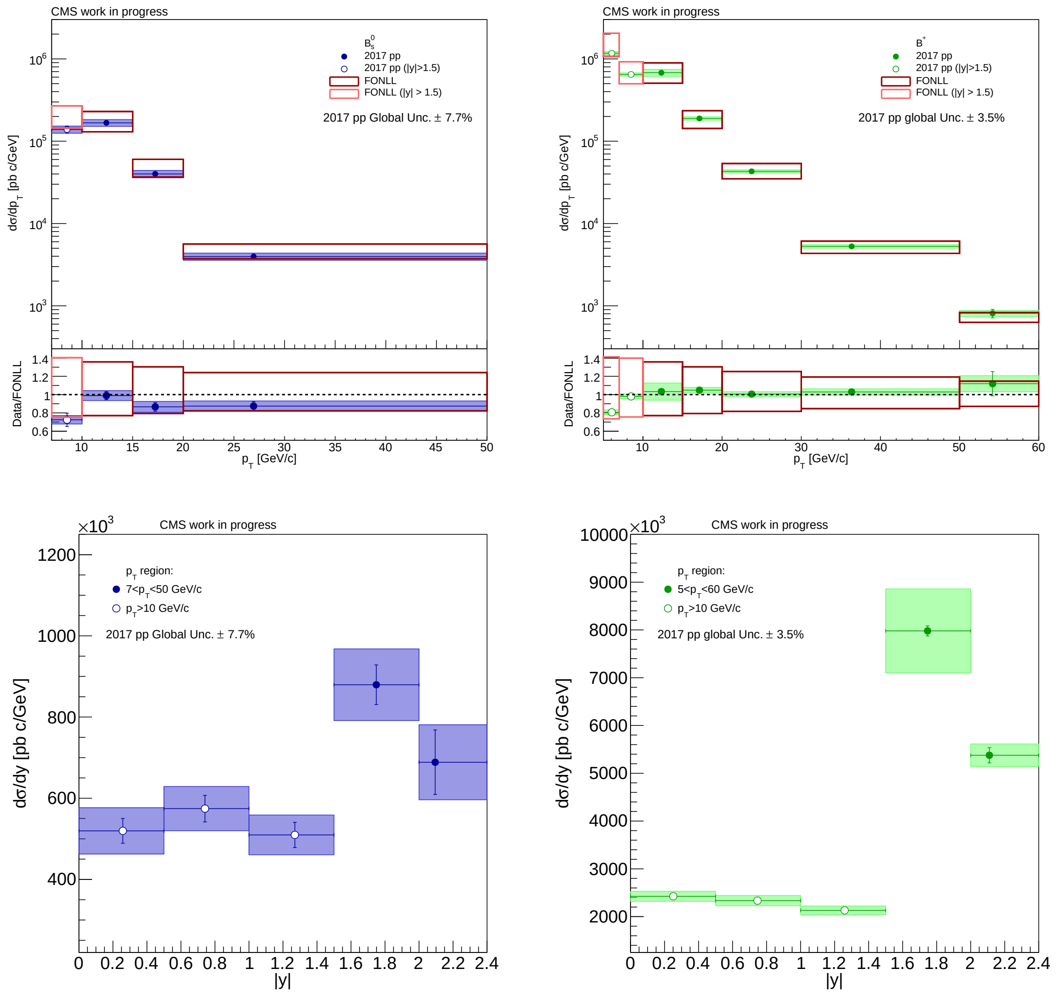

Figure 8.1: Differential cross section vs p_T (top row) and $|y|$ (bottom row) for the B_s^0 (left) and B^+ (right) mesons.

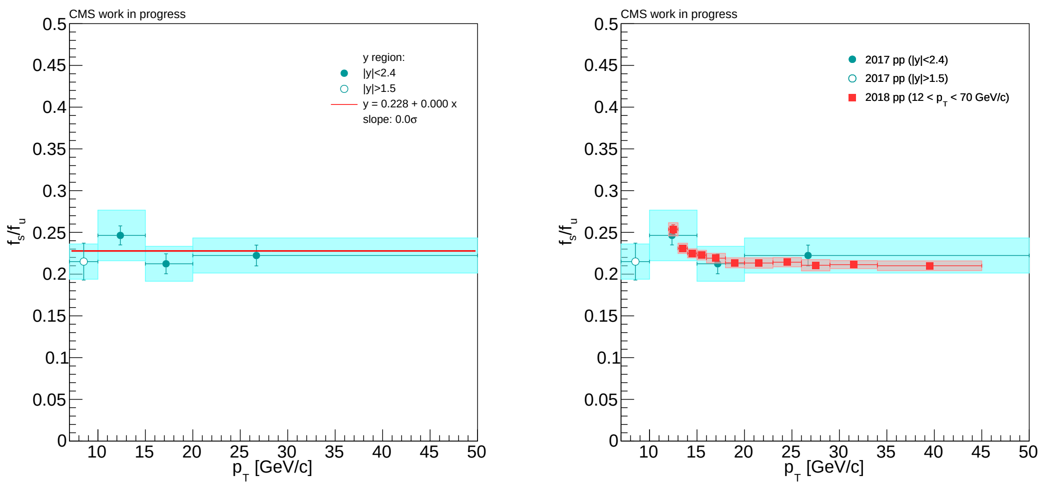

Figure 8.2: Ratio of the measured B_s^0 and B^+ differential cross sections vs p_T. Comparison with a fit to the obtained results using a linear model (left) and with the 2018 pp CMS results at $\sqrt{s} = 13$ TeV (right).

The measured fragmentation fraction ratio displays hints of a small decrease with the meson's p_T, within the available precision in both coordinate and abscissae axes. Such a dependency has been first reported by LHCb in the low-p_T region, and corroborated by recent CMS results on the measurement of the hadron production fraction ratios f_s/f_u on B meson kinematic variables in pp collisions at $\sqrt{s} = 13$ TeV [30]. Here, a p_T range of 12 to 70 GeV/c was used with $|y| < 2.4$, having measured a clear p_T dependence at $p_T < 15$ GeV/c being the higher p_T region devoid of any dependence.

Our results are consistent with the absence of dependence in the higher p_T region. The $[10, 15]$ GeV/c bin displays a higher value, being thus compatible with the increasing trend towards lower p_T observed by LHCb and CMS.

However, in our case, the smaller size of B_s^0 sample limits the p_T range and binning we can probe. Also, the presence of a different fiducial region in our first bin further restricts the p_T region available, hindering a more clear conclusion. It is of note that although the 13 TeV dataset boasts higher statistics compared to the 5.02 TeV dataset used in this analysis, which, due to trigger limitations, is not able to reach the lower p_T where the dependence is more pronounced. Given the tools developed in this thesis, we will be able to further probe into this region with upcoming data.

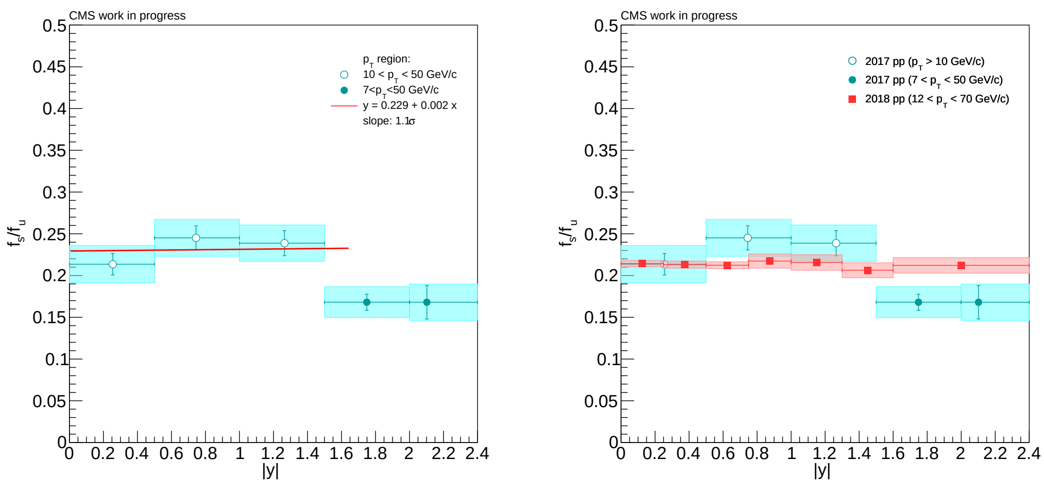

Figure 8.3: Ratio of the measured B_s^0 and B^+ differential cross sections vs $|y|$. Comparison with a fit to the obtained results using a linear model in the shown range (left) and with the 2018 pp CMS results at $\sqrt{s} = 13$ TeV (right).

No noticeable dependence is observed as function of rapidity. This is also consistent with previous results, as can be on the right side plot of Fig. 8.3, which shows a very good agreement between our results and those obtained with the 13 TeV dataset for similar fiducial regions. It is noted that in Fig. 8.3, the measured points are split into two different fiducial regions, being the high rapidity bins a bad reference for comparison, given the distinct fiducial criteria used in the two measurements.

Finally, the measured hadronization fraction ratio results as function of charged particle multiplicity are presented in Fig. 8.4. Here, N_{trk} stands for the raw number of charged tracks observed in data, that pass the selection presented in Sec. 3.5 but not corrected for tracking efficiency effects. We observe a slight increase of f_s/f_u with multiplicity. This is consistent with the f_s/f_d results that were also recently

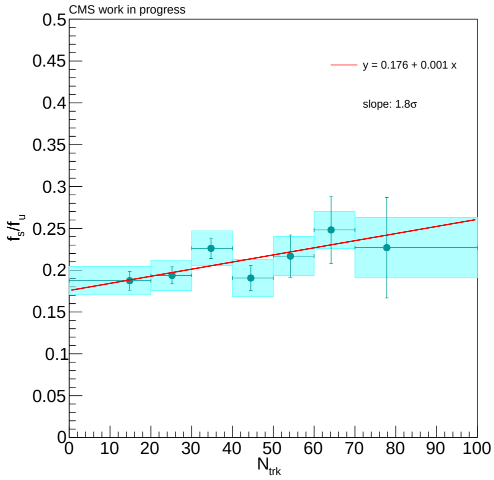

Figure 8.4: Ratio of the measured B_s^0 and B^+ differential cross sections vs N_{trk}. The result of a linear fit to the obtained results is overlaid.

reported by LHCb, relative to the ratio of cross sections of the B_s^0 and B^0 mesons, in pp collisions at $\sqrt{s} = 13$ TeV [31].

Accordingly, in collisions with high charged-particle multiplicity, we see evidence for B_s^0 production enhancement relative to B^+, indicating that strangeness enhancement is present in B hadron production. These findings are consistent with expected contributions from quark coalescence to the hadronization process, in addition to fragmentation.

Chapter 9

Summary

In this thesis, we carried out studies of B meson production, and investigated the interplay of two competing hadronization mechanisms, fragmentation and coalescence. The B^+ and B^0_s mesons were studied in $\sqrt{s} = 5.02$ TeV proton-proton collision data collected by CMS, through the decay channels $B^+ \to J/\psi\, K^+ \to \mu^+\mu^- K^+$ and $B^0_s \to J/\psi\, \phi \to \mu^+\mu^- K^+ K^-$. The mesons' production cross sections were measured as a function of relevant kinematic (transverse momentum, p_T, and rapidity, y) and environment (charged-particle multiplicity, N_{trk}) variables. Finally, the ratio of the two mesons' production was determined, along with its dependence on the same variables.

Machine learning methods were explored to obtain signal-enriched samples, that formed the basis for the ensuing data analysis. The ML models were trained using MC samples for signal and a subset of the data from mass-sideband as background. We utilized boosted decision trees with the help of the XGBoost package, achieving a signal-background separation in our data that far surpassed any selection employed by previous studies on this data. Our selection not only vastly increases the signal significance, but it also allows us to reach into lower p_T regions, previously inaccessible due to their high amount of background that made it impossible to perform mass fits. We were successful in fitting the $[3, 5]$ GeV/c bin of the B^+ meson and see a clear mass peak in this region, although due to limitations in our MC samples we were unable to fully perform an analysis on this bin. Likewise, due to lack of suitable simulated samples and lower B^0_s yields in data, we were unable to perform mass fits at $p_T < 3$ GeV/c for the B^+ meson and at $p_T < 7$ GeV/c for the B^0_s meson. However, a machine learning framework was created that lets one easily apply these methods to any dataset, making it a trivial task to repeat the model optimization and selection for new and improved datasets, with enough information to perform a full analysis in this region. This would be of great importance since it is in this region that we expect a more prominent effect of coalescence as a hadronization mechanism, and the further extension of the analysis to this region would increase its physics sensitivity.

A novel method for efficiency determination was also devised, by utilizing the ${}_s\mathcal{P}lot$ method to weigh our events in conjunction with a fine-grained 2D inverse-efficiency map as a function of p_T and rapidity, y, minimizing our reliance in the correct kinematic description of our data by the MC samples and removing background contamination. We have also provided an analytical proof of the method accompanied by a closure test performed over MC samples to verify that the method shows closure and prove its validity.

With these tools at hand, we carried out the measurement of the B^+ and B_s^0 meson production cross sections as functions of p_T, y and N_{trk}. The p_T dependency is found to be consistent with theoretical predictions. Our results lie within the one sigma FONLL prediction, uncertainty with the exception of 3 bins that border the outskirts of this uncertainty, being the overall agreement still very good overall.

From the obtained differential cross section results, we attained a measurement of the hadronization fraction ratio, f_s/f_u, and its dependencies on the above-mentioned variables. We observed hints of a slight decrease with p_T, consistent with previous CMS and LHCb results. No indication of dependence on $|y|$ is found in our results, excluding the difference between the two different fiducial regions. This is also consistent with the behaviour reported in previous LHCb and CMS studies. Agreement is obtained with previous CMS results in comparable fiducial regions.

Finally, we have our measurement of f_s/f_u versus N_{trk}. These are the first such results by CMS. We observe a slight increase with multiplicity. This shows agreement with recent results obtained by LHCb. The results hint that in events characterized by large charged-particle multiplicity we have B_s^0 production enhancement relative to B^+, which is consistent with the emergence of quark coalescence as an additional hadronization mechanism, rather than fragmentation alone.

Bibliography

[1] Yang C. N. and Mills R. L. Conservation of isotopic spin and isotopic gauge invariance. *Phys. Rev.*, 96:191–195, Oct 1954. doi: 10.1103/PhysRev.96.191. URL https://link.aps.org/doi/10.1103/PhysRev.96.191.

[2] Adrian Cho. Postage stamp to honor female physicist who many say should have won the nobel prize. URL https://www.science.org/content/article/postage-stamp-honor-female-physicist-who-many-say-should-have-won-nobel-prize.

[3] Glashow, Sheldon L. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579–588, Feb. 1961. doi: 10.1016/0029-5582(61)90469-2.

[4] Weinberg, Steven. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. doi: 10.1103/PhysRevLett.19.1264. URL https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[5] Hasert et al. Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment. *Nuclear Physics B*, 73(1):1–22, Apr. 1974. doi: 10.1016/0550-3213(74)90038-8.

[6] UA1 collaboration, G. Arnison et al. Experimental observation of isolated large transverse energy electrons with associated missing energy at s=540 gev. *Physics Letters B*, 122(1):103–116, 1983. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(83)91177-2. URL https://www.sciencedirect.com/science/article/pii/0370269383911772.

[7] UA2 collaboration, M. Banner et al. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider. *Physics Letters B*, 122(5):476–485, 1983. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(83)91605-2. URL https://www.sciencedirect.com/science/article/pii/0370269383916052.

[8] Gaillard, Mary K. and Grannis, Paul D. and Sciulli, Frank J. The standard model of particle physics. *Reviews of Modern Physics Supplement*, 71(2):S96–S111, Mar. 1999. doi: 10.1103/RevModPhys.71.S96.

[9] CMS collaboration, S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, 2012. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2012.08.021. URL https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[10] Taesoo Song and Gabriele Coci. Prerequisites for heavy quark coalescence in heavy-ion collisions. *Nuclear Physics A*, 1028:122539, 2022. ISSN 0375-9474. doi: https://doi.org/10.1016/j.nuclphysa.2022.122539. URL https://www.sciencedirect.com/science/article/pii/S0375947422001592.

[11] Bruno Afonso Fontana Santos Alves. Measurement of b-quark fragmentation fraction ratios at the CMS experiment: a key ingredient for the $B_s^0 \to \mu^+\mu$ rare decay analysis. URL https://cds.cern.ch/record/2649927/files/CERN-THESIS-2018-274.pdf.

[12] CMS collaboration, Tumasyan, Armen and others. Measurement of the $\mathrm{B_s^0}{\to}\mu^+\mu^-$ decay properties and search for the $\mathrm{B^0}{\to}\mu^+\mu^-$ decay in proton-proton collisions at $\sqrt{s}$ = 13 TeV. *Phys. Lett. B*, 842: 137955, 2023. doi: 10.1016/j.physletb.2023.137955.

[13] Samuel, A. L. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959. doi: 10.1147/rd.33.0210.

[14] Tom Mitchell. Machine learning. URL https://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html.

[15] Kurt Hornik and Maxwell Stinchcombe and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

[16] Caglar Aytekin. Neural networks are decision trees. URL https://deepai.org/publication/neural-networks-are-decision-trees.

[17] Y. Amhis et al. Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018. *The European Physical Journal C*, 81(3), mar 2021. doi: 10.1140/epjc/s10052-020-8156-7. URL https://doi.org/10.1140%2Fepjc%2Fs10052-020-8156-7.

[18] CLEO collaboration, M. Artuso et al. First evidence and measurement of $B_s^* \ \bar{B}_s^*$ production at the $\Upsilon(5s)$. *Physical Review Letters*, 95(26), dec 2005. doi: 10.1103/physrevlett.95.261801. URL https://doi.org/10.1103%2Fphysrevlett.95.261801.

[19] BELLE collaboration, A. Drutskoy et al. Measurement of inclusive D_s, D^0, and J/ψ rates and determination of the $B_s^{(*)}\bar{b}_s^{(*)}$ production fraction in $b\bar{b}$ events at the $\Upsilon(5S)$ resonance. *Physical Review Letters*, 98(5), jan 2007. doi: 10.1103/physrevlett.98.052001. URL https://doi.org/10.1103%2Fphysrevlett.98.052001.

[20] CLEO Collaboration, G. S. Huang et al. Measurement of $\mathcal{B}(\Upsilon(5s) \to \mathrm{B}_s^{(*)}\bar{B}_s^{(*)})$ using ϕ mesons. *Physical Review D*, 75(1), jan 2007. doi: 10.1103/physrevd.75.012002. URL https://doi.org/10.1103%2Fphysrevd.75.012002.

[21] L3 collaboration, M. Acciarri et al. Measurements of the production cross section and forward–backward asymmetry at centre-of-mass energies above the z pole at LEP. *Physics Letters*

B, 485(1-3):71–84, jul 2000. doi: 10.1016/s0370-2693(00)00676-6. URL https://doi.org/10.1016%2Fs0370-2693%2800%2900676-6.

[22] DELPHI collaboration, J. Abdallah et al. A measurement of the branching fractions of the b-quark into charged and neutral b-hadrons. *Physics Letters B*, 576(1-2):29–42, dec 2003. doi: 10.1016/j.physletb.2003.09.070. URL https://doi.org/10.1016%2Fj.physletb.2003.09.070.

[23] OPAL collaboration, P.D. Acton et al. Evidence for the existence of the strange b-flavoured meson B_s^0 in Z^0 decays. *Physics Letters B*, 295(3):357–370, 1992. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(92)91578-W. URL https://www.sciencedirect.com/science/article/pii/037026939291578W.

[24] D Buskulic et al. Measurement of the B_s0 lifetime and production rate with $D_s^- l^+$ combinations in Z decays. *Physics Letters B*, 361(1):221–233, 1995. ISSN 0370-2693. doi: https://doi.org/10.1016/0370-2693(95)01173-N. URL https://www.sciencedirect.com/science/article/pii/037026939501173N.

[25] John C. Collins and Davison E. Soper and George Sterman. Factorization of hard processes in qcd, 2004.

[26] LHCb collaboration, R. Aaji et al. Measurement of b hadron production fractions in 7 tev pp collisions. *Physical Review D*, 85(3), feb 2012. doi: 10.1103/physrevd.85.032008. URL https://doi.org/10.1103%2Fphysrevd.85.032008.

[27] LHCb collaboration, R. Aaij et al. Measurement of b hadron fractions in 13 tev pp collisions. *Physical Review D*, 100(3), aug 2019. doi: 10.1103/physrevd.100.031102. URL https://doi.org/10.1103%2Fphysrevd.100.031102.

[28] ALICE collaboration, S. Acharya et al. Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC. *Physical Review D*, 105(1), jan 2022. doi: 10.1103/physrevd.105.l011103. URL https://doi.org/10.1103%2Fphysrevd.105.l011103.

[29] CMS Collaboration. Observation of B_s^0 mesons and measurement of the B_s^0/B^+ yield ratio in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV. *Phys. Lett. B*, 829:137062, 2022. doi: 10.1016/j.physletb.2022.137062.

[30] CMS collaboration, Tumasyan, A. et al. Measurement of the dependence of the hadron production fraction ratios f_s/f_u and f_d/f_u on b meson kinematic variables in proton-proton collisions at $\sqrt{s} = 13\ \mathrm{TeV}$. *Phys. Rev. Lett.*, 131:121901, Sep 2023. doi: 10.1103/PhysRevLett.131.121901. URL https://link.aps.org/doi/10.1103/PhysRevLett.131.121901.

[31] LHCb Collaboration. Evidence for modification of b quark hadronization in high-multiplicity pp collisions at $\sqrt{s} = 13$ tev, 2022. URL https://arxiv.org/abs/2204.13042.

[32] Shanshan Cao and Kai-Jia Sun and Shu-Qing Li and Shuai Y.F. Liu and Wen-Jing Xing and Guang-You Qin and Che Ming Ko. Charmed hadron chemistry in relativistic heavy-ion collisions. *Physics*

Letters B, 807:135561, 2020. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2020.135561. URL https://www.sciencedirect.com/science/article/pii/S0370269320303658.

[33] Min He and Rainer J. Fries and Ralf Rapp. Heavy flavor at the large hadron collider in a strong coupling approach. *Physics Letters B*, 735:445–450, 2014. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2014.05.050. URL https://www.sciencedirect.com/science/article/pii/S0370269314003591.

[34] CERN. LHC machine outreach. URL https://lhc-machine-outreach.web.cern.ch/.

[35] CERN. CERN. URL https://home.cern/science/experiments.

[36] The CMS Collaboration, S. Chatrchyan et al. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, aug 2008. doi: 10.1088/1748-0221/3/08/S08004. URL https://dx.doi.org/10.1088/1748-0221/3/08/S08004.

[37] Torbjörn Sjöstrand et al. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, jun 2015. doi: 10.1016/j.cpc.2015.01.024. URL https://doi.org/10.1016%2Fj.cpc.2015.01.024.

[38] J. Allison et al. Recent developments in Geant4. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 835:186–225, 2016. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2016.06.125. URL https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[39] 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(1):152–155, 2001. ISSN 0168-9002. doi: https://doi.org/10.1016/S0168-9002(01)00089-4. URL https://www.sciencedirect.com/science/article/pii/S0168900201000894. BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines.

[40] Elisabetta Barberio and Bob van Eijk and Zbigniew Was. Photos — a universal monte carlo for QED radiative corrections in decays. *Computer Physics Communications*, 66(1):115–128, 1991. ISSN 0010-4655. doi: https://doi.org/10.1016/0010-4655(91)90012-A. URL https://www.sciencedirect.com/science/article/pii/001046559190012A.

[41] The CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009–P10009, oct 2014. doi: 10.1088/1748-0221/9/10/p10009. URL https://doi.org/10.1088%2F1748-0221%2F9%2F10%2Fp10009.

[42] XgBoost. Xgboost. URL https://xgboost.readthedocs.io/en/stable/.

[43] PyTorch. Pytorch. URL https://pytorch.org/.

[44] Optuna. Optuna. URL https://optuna.org/.

[45] XGBoost. Xgboost parameters. URL https://xgboost.readthedocs.io/en/latest/parameter.html.

[46] The CMS collaboration. Constraining beauty quark energy loss and hadronization with B^+ and B^0_s nuclear modification factors in PbPb collisions at 5.02 TeV with CMS. Technical report, CERN, Geneva, 2023. URL https://cds.cern.ch/record/2871505.

[47] CERN. Roofit, . URL https://root.cern/manual/roofit/.

[48] CERN. Minuit: Function minimization and error analysis reference manual, . URL https://cds.cern.ch/record/2296388.

[49] CMS collaboration, Sirunyan et al. Relative modification of prompt $\psi(2S)$ and J/ψ yields from pp to PbPb collisions at $\sqrt{s_{NN}} = 5.02\ \mathrm{TeV}$. *Phys. Rev. Lett.*, 118:162301, Apr 2017. doi: 10.1103/PhysRevLett.118.162301. URL https://link.aps.org/doi/10.1103/PhysRevLett.118.162301.

[50] M. Pivk and F.R. Le Diberder. Plots: A statistical tool to unfold data distributions. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 555(1):356–369, 2005. ISSN 0168-9002. doi: https://doi.org/10.1016/j.nima.2005.08.106. URL https://www.sciencedirect.com/science/article/pii/S0168900205018024.

Appendix A

Binned invariant mass fits results

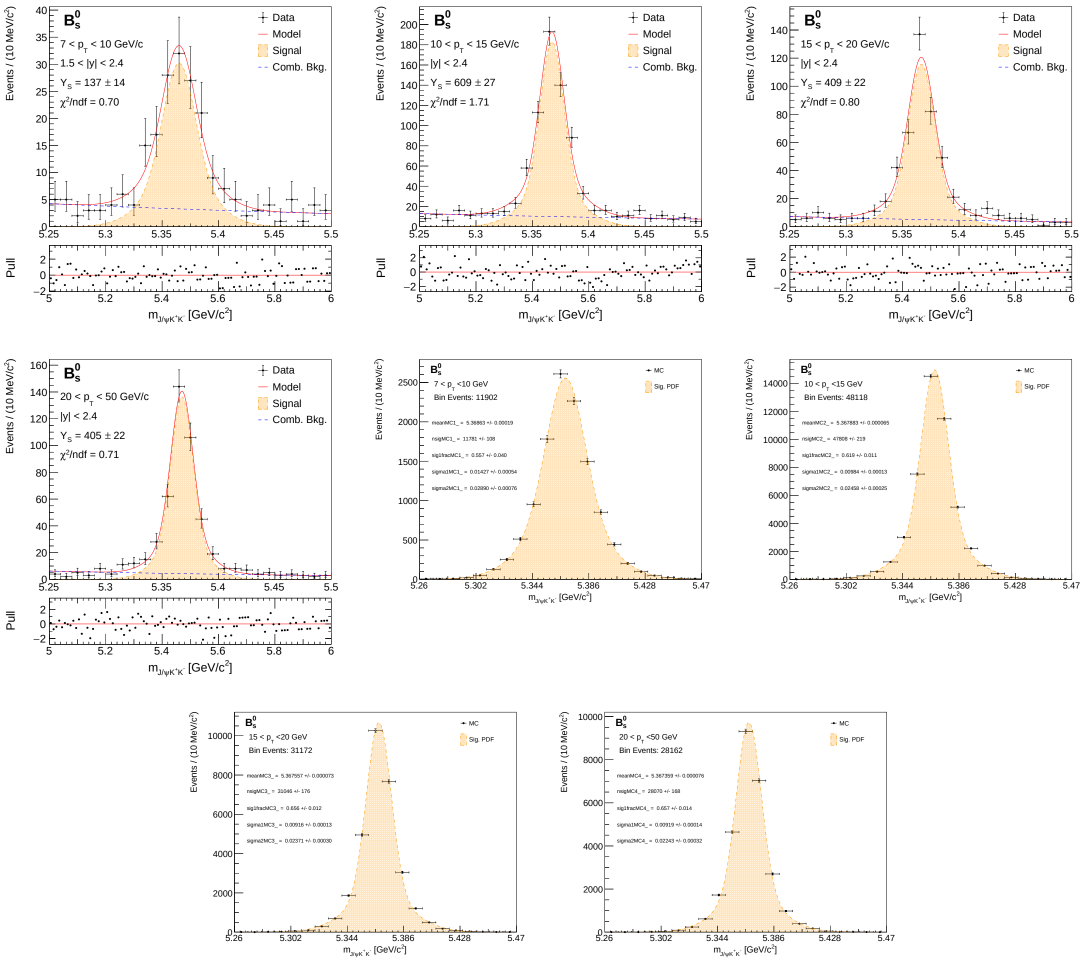

Figure A.1: Invariant mass fits vs p_T to the data and MC for the B_s^0 meson.

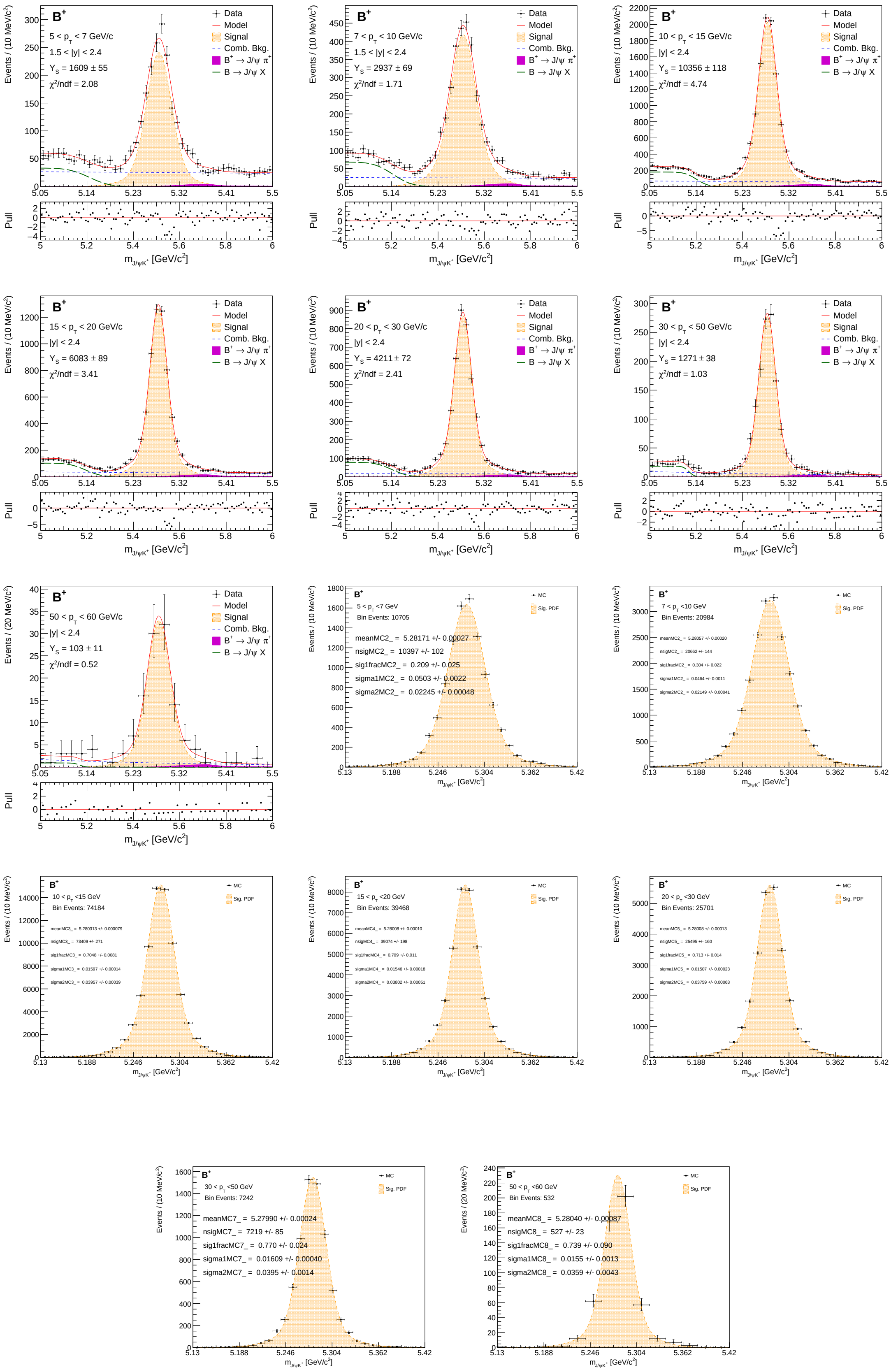

Figure A.2: Invariant mass fits vs p_T to the data and MC for the B^+ meson.

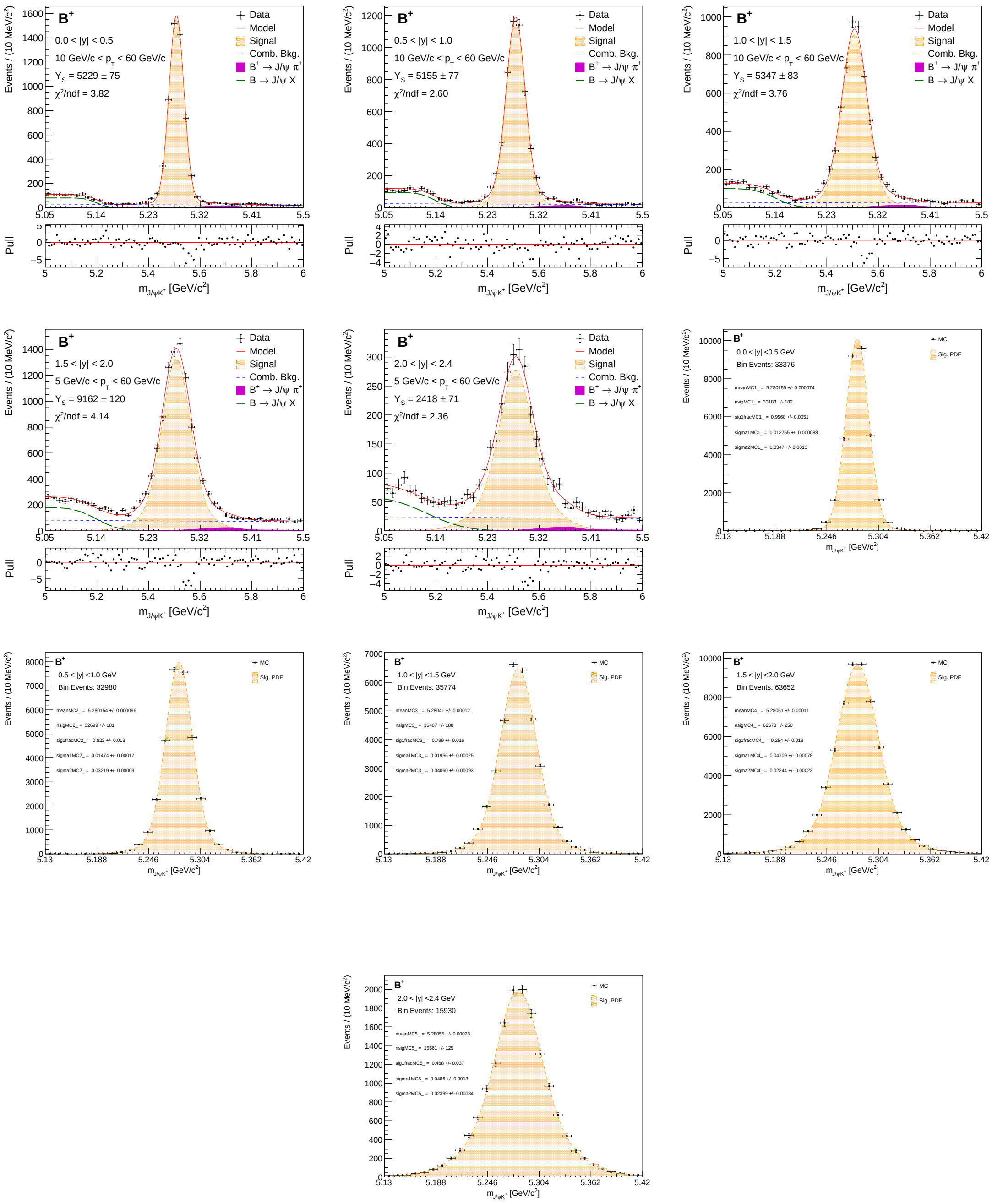

Figure A.3: Invariant mass fits vs y to the data and MC for the B^+ meson.

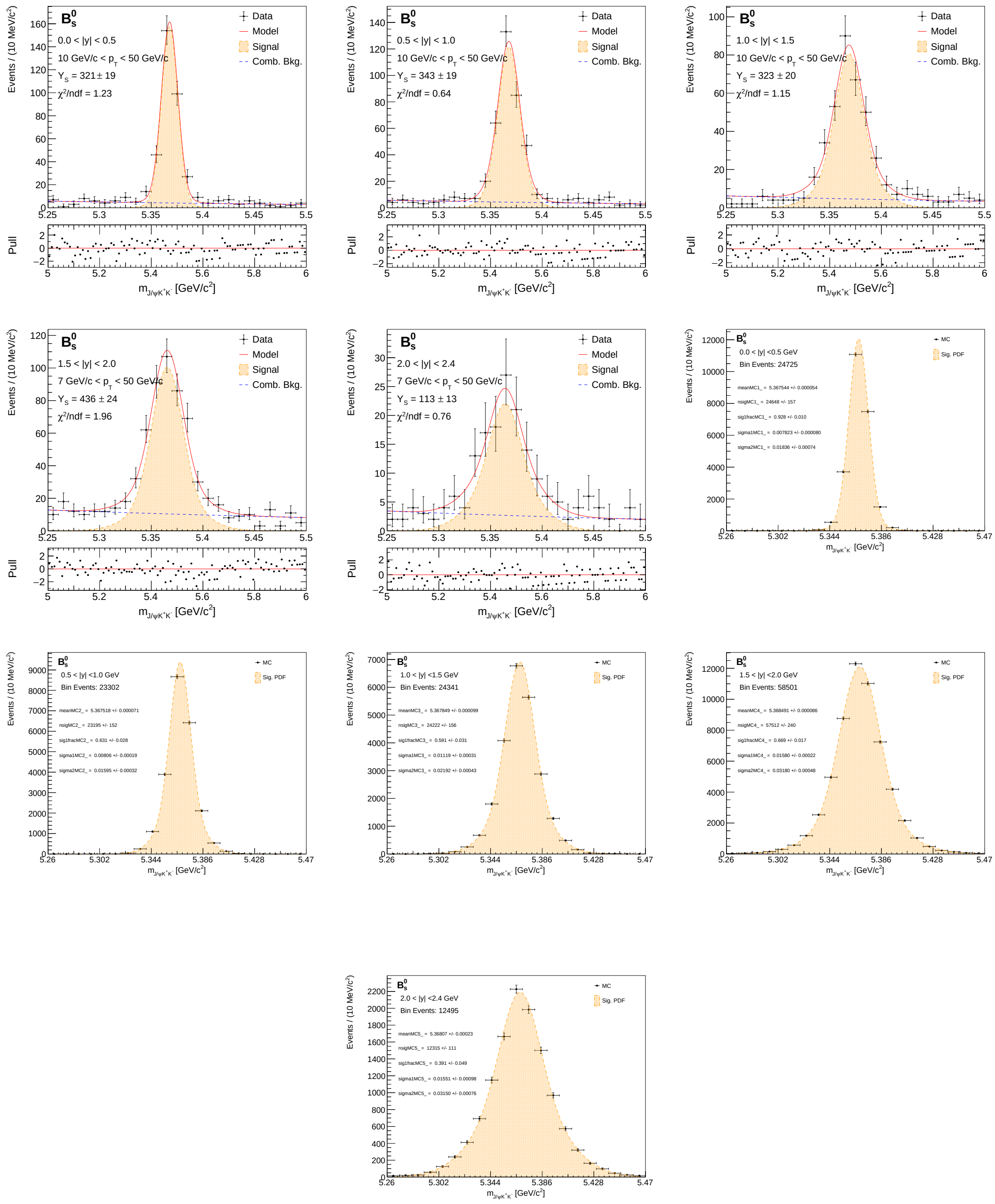

Figure A.4: Invariant mass fits vs y to the data and MC for the B_s^0 meson.

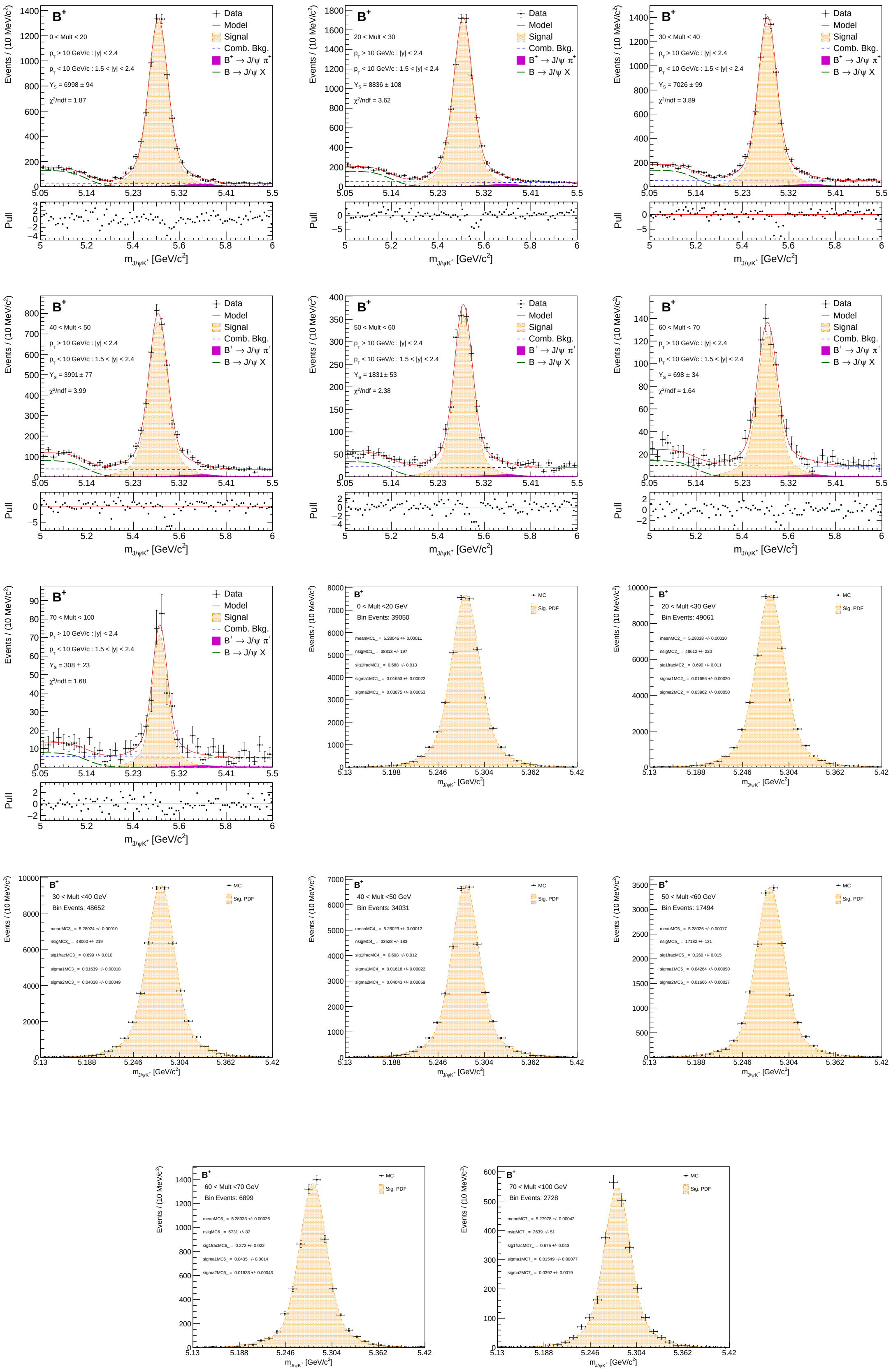

Figure A.5: Invariant mass fits vs N_{trk} to the data and MC for the B^+ meson.

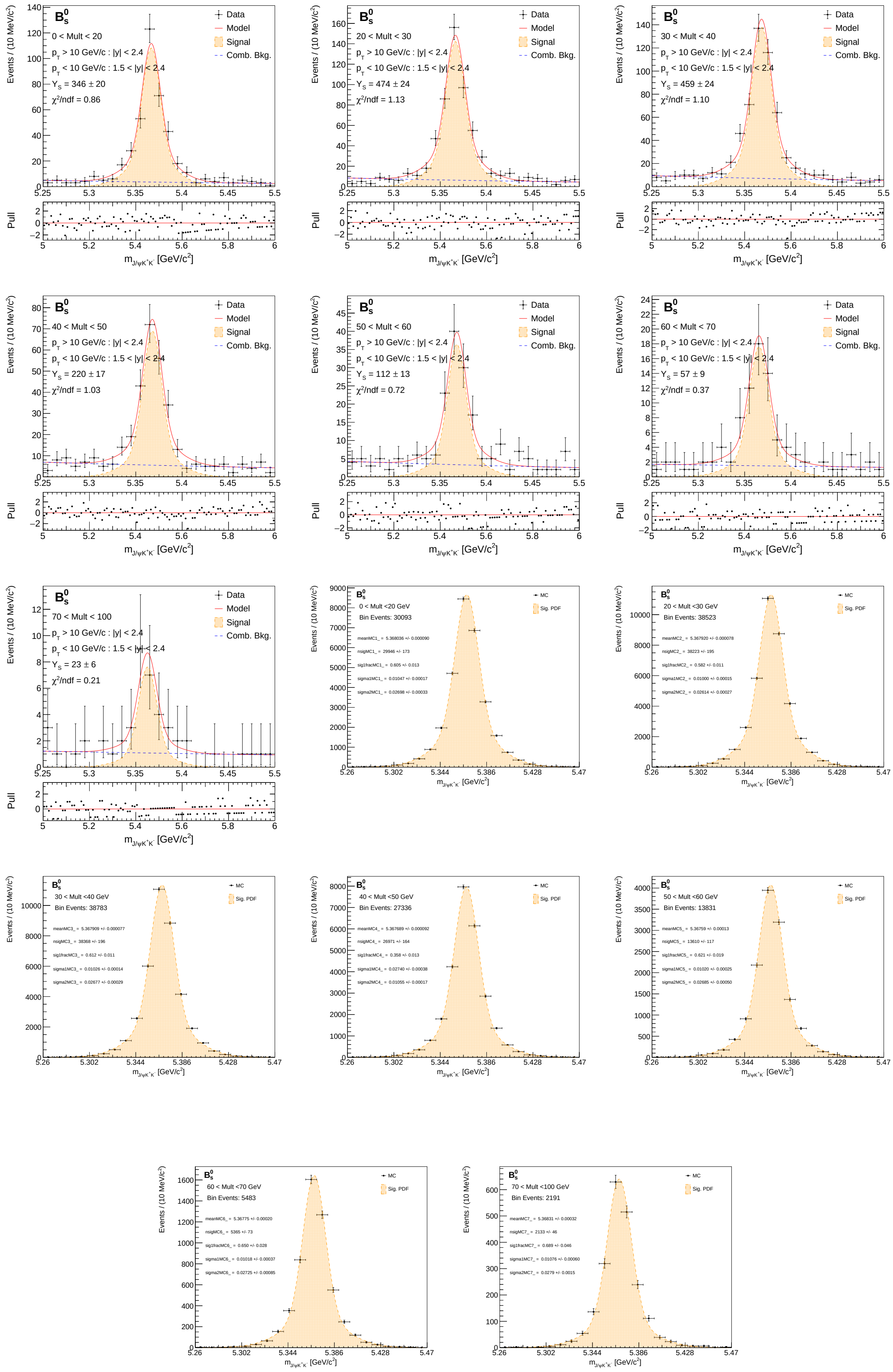

Figure A.6: Invariant mass fits vs N_{trk} to the data and MC for the B_s^0 meson.

Appendix B

Cross-section plots

B.1 Cross sections for the fragmentation ratios calculation

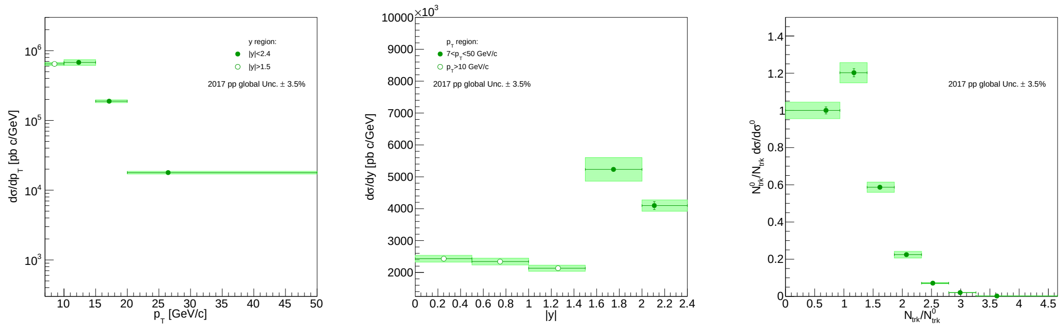

Figure B.1: Differential cross section vs p_T (top row), y (middle row) and multiplicity (bottom row), for the B^+ meson.

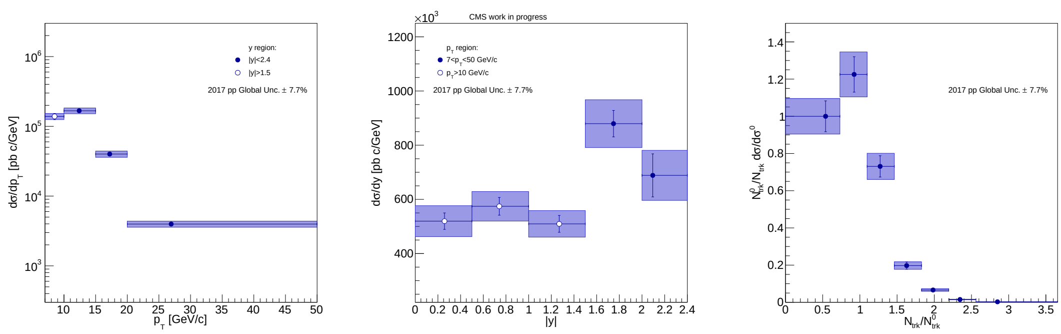

Figure B.2: Differential cross section vs p_T (top row), y (middle row) and multiplicity (bottom row), for the B_s^0 meson.

B.2 Cross section 1D vs 2D comparison

In the main work, we discussed and implemented a 2D fine-grained map for our efficiency calculation, but one might wonder by how much does this actually change our final results. In this section follows the comparisons plots between the cross section obtained with each method alongside the tabled relative difference between them.

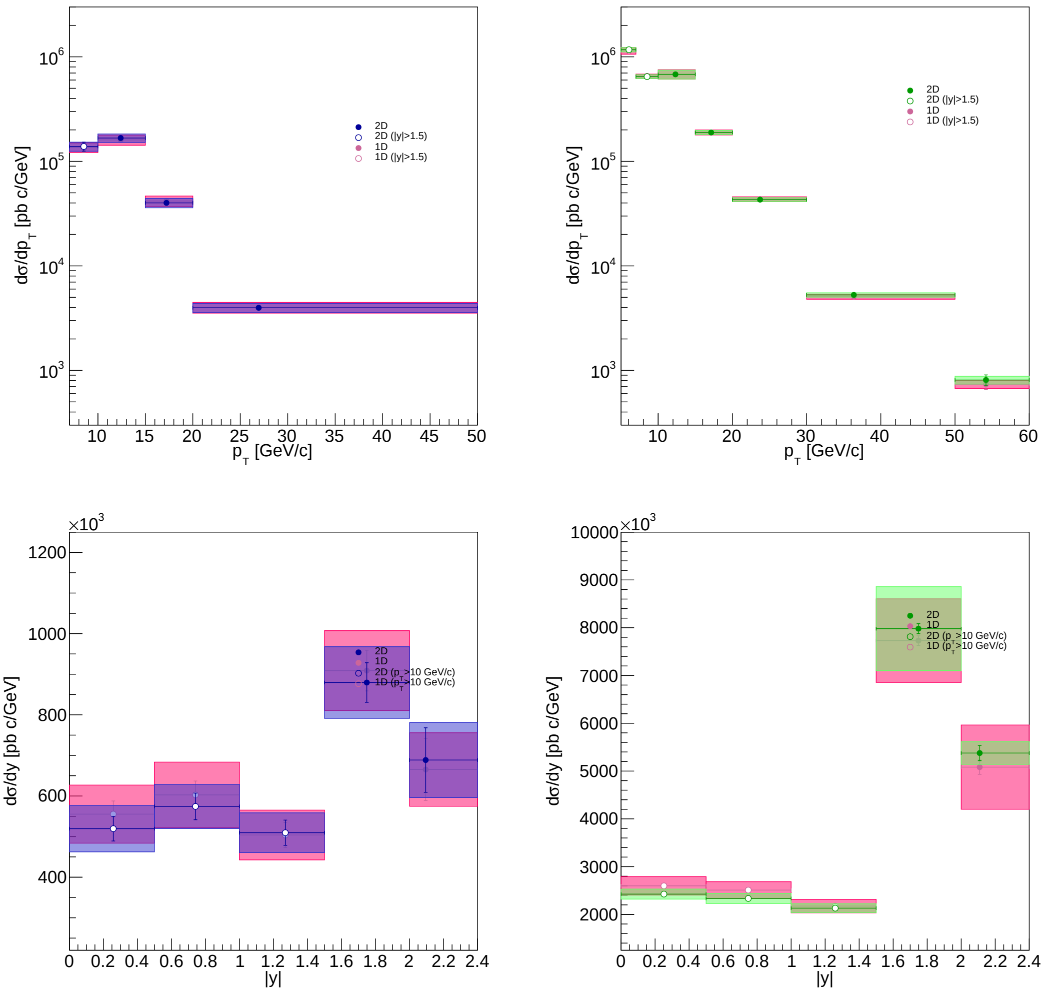

Figure B.3: Comparison between 1D and 2D differential cross section vs p_T (top row) and y (bottom row), for the B_s^0 (left) and B^+ (right) mesons. Cross-section vs N_{trk} was excluded from this comparison since we lacked the GEN level info necessary to compute its 1D efficiency.

Table B.1: 1D vs 2D cross section as a function of p_T [GeV/c], for the B^+ meson.

	5< p_T <7	7< p_T <10	10< p_T <15	15< p_T <20	20< p_T <30	30< p_T <50	50< p_T <60
Relative difference 1D vs 2D	2.34%	0.67%	0.445%	0.26%	1.12%	3.61%	8.88%

Table B.2: 1D vs 2D cross section as a function of p_T [GeV/c], for the B_s^0 meson.

	7< p_T <10	10< p_T <15	15< p_T <20	20< p_T <50
Relative difference 1D vs 2D	2.3%	4.0%	4.23%	0.746%

Table B.3: 1D vs 2D cross section as a function of y, for the B^+ meson.

	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
Relative difference 1D vs 2D	7.13%	7.48%	2.02%	3.14%	5.48%

Table B.4: 1D vs 2D cross section as a function of y, for the B_s^0 meson.

	0.0< y <0.5	0.5< y <1.0	1.0< y <1.5	1.5< y <2.0	2.0< y <2.4
Relative difference 1D vs 2D	6.89%	4.93%	1.11%	3.35%	3.39%

As can be seen, the discrepancy between these two methods is non-negligible ($< 9\%$) and therefore we see that the reliance on the MC kinematics has an effect on our final results. This further validates our use of the fine-grained 2D map.

Appendix C

Systematic uncertainties

C.1 Yield systematics

Table C.1: Summary of relative systematic uncertainties, in %, of p_T bins from the B_s^0 fitting model variations.

B_s^0 p_T [GeV/c]	7–10	10–15	15–20	20–50
Background Unc.	4.01	2.14	3.99	3.20
Signal Unc.	0.90	1.21	1.65	2.67
PDF variation Unc.	4.11	2.46	4.32	4.17

Table C.2: Summary of relative systematic uncertainties, in %, of p_T bins from the B^+ fitting model variations.

B^+ p_T [GeV/c]	5–7	7–10	10–15	15–20	20–30	30–50	50–60
Background Unc.	2.07	0.559	1.61	1.51	0.783	2.07	7.28
Signal Unc.	0.763	0.451	0.320	0.365	0.298	0.213	0.199
PDF variation Unc.	2.205	0.718	1.64	1.56	0.838	2.08	7.29

Table C.3: Summary of relative systematic uncertainties, in %, of y bins from the B_s^0 fitting model variations.

B_s^0 y	0–0.5	0.5–1	1–1.5	1.5–2	2–2.4
Background Unc.	1.63	1.50	2.46	4.24	9.55
Signal Unc.	0.766	0.262	0.873	0.111	0.909
PDF variation Unc.	1.80	1.52	2.61	4.24	9.59

Table C.4: Summary of relative systematic uncertainties, in %, of y bins from the B^+ fitting model variations.

B^+ y	0–0.5	0.5–1	1–1.5	1.5–2	2–2.4
Background Unc.	0.302	0.988	0.214	2.08	0.908
Signal Unc.	0.277	0.144	0.394	0.439	0.342
PDF variation Unc.	0.410	0.999	0.448	2.12	0.970

Table C.5: Summary of relative systematic uncertainties, in %, of N_{trk} bins from the B_s^0 fitting model variations.

B_s^0 mult	0–20	20–30	30–40	40–50	50–60	60–70	70–100
Background Unc.	2.83	2.86	2.99	3.48	3.58	1.01	9.37
Signal Unc.	0.584	1.42	0.992	0.996	3.52	1.98	1.40
PDF variation Unc.	2.89	3.19	3.15	3.62	5.02	2.22	9.47

Table C.6: Summary of relative systematic uncertainties, in %, of N_{trk} bins from the B^+ fitting model variations.

B^+ mult	0–20	20–30	30–40	40–50	50–60	60–70	70–100
Background Unc.	1.19	1.40	1.79	0.208	2.43	0.636	4.70
Signal Unc.	0.359	0.247	0.560	0.855	0.656	0.344	0.325
PDF variation Unc.	1.24	1.43	1.87	0.880	2.52	0.723	4.71

In Fig. C.1 the extracted p_T ,y and N_{trk} signal yields normalized by the respective bin width are plotted for each B meson along their systematic and statistical uncertainties.

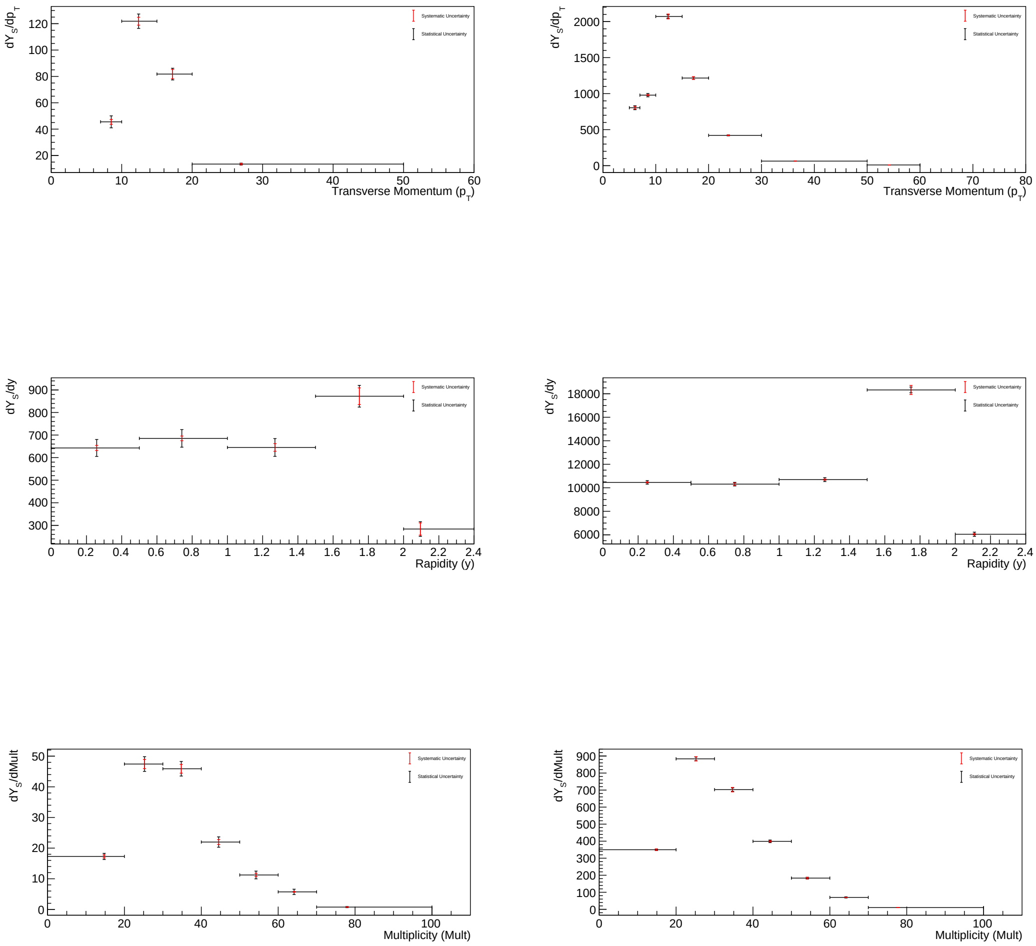

Figure C.1: Summary of the p_T (top row), y (middle row) and multiplicity (bottom row) normalized signal yields and respective uncertainties, for the B_s^0 (left) and B^+ (right) meson.

C.2 Efficiency systematics

C.2.1 Data vs MC comparison

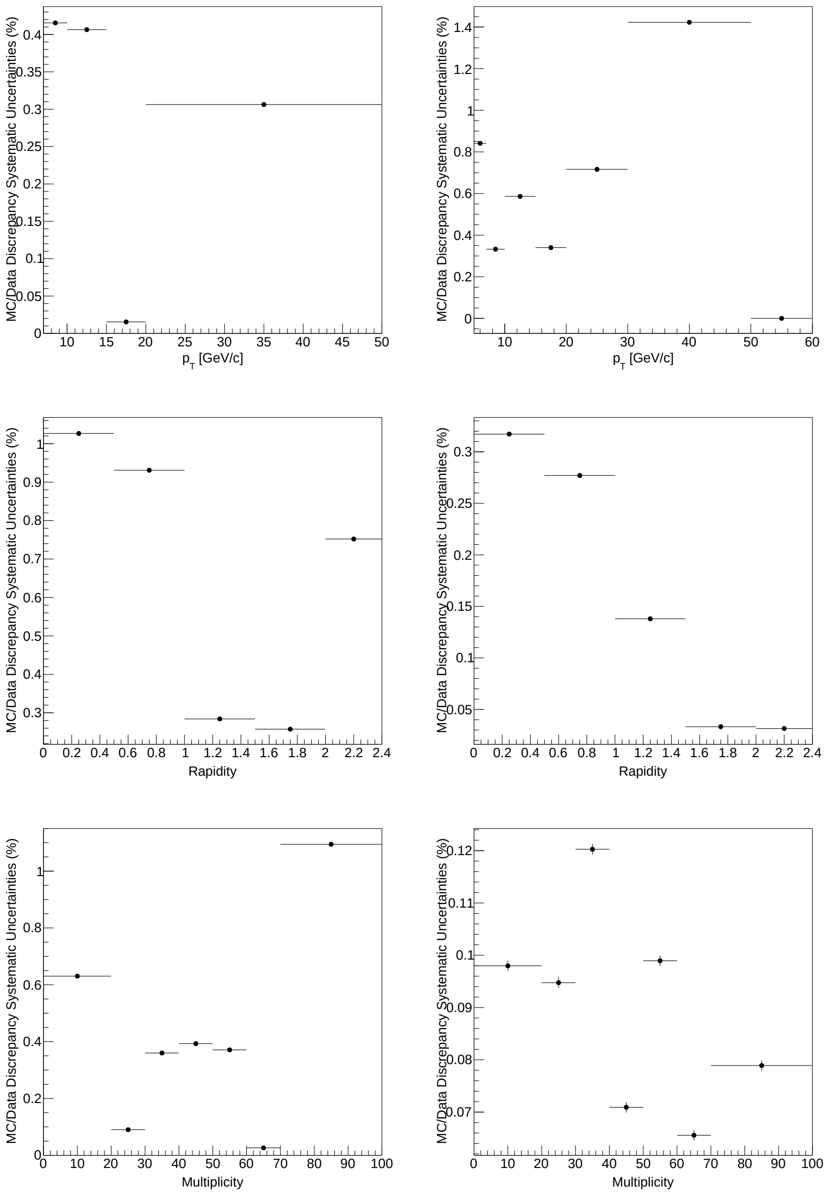

Figure C.2: Systematic uncertainties in total efficiency vs p_T (first row) and vs rapidity (second row) for BDT weights variation, on the B_s^0 (first column) and B^+ (second column) mesons.

C.2.2 Muon efficiency: tag-and-probe

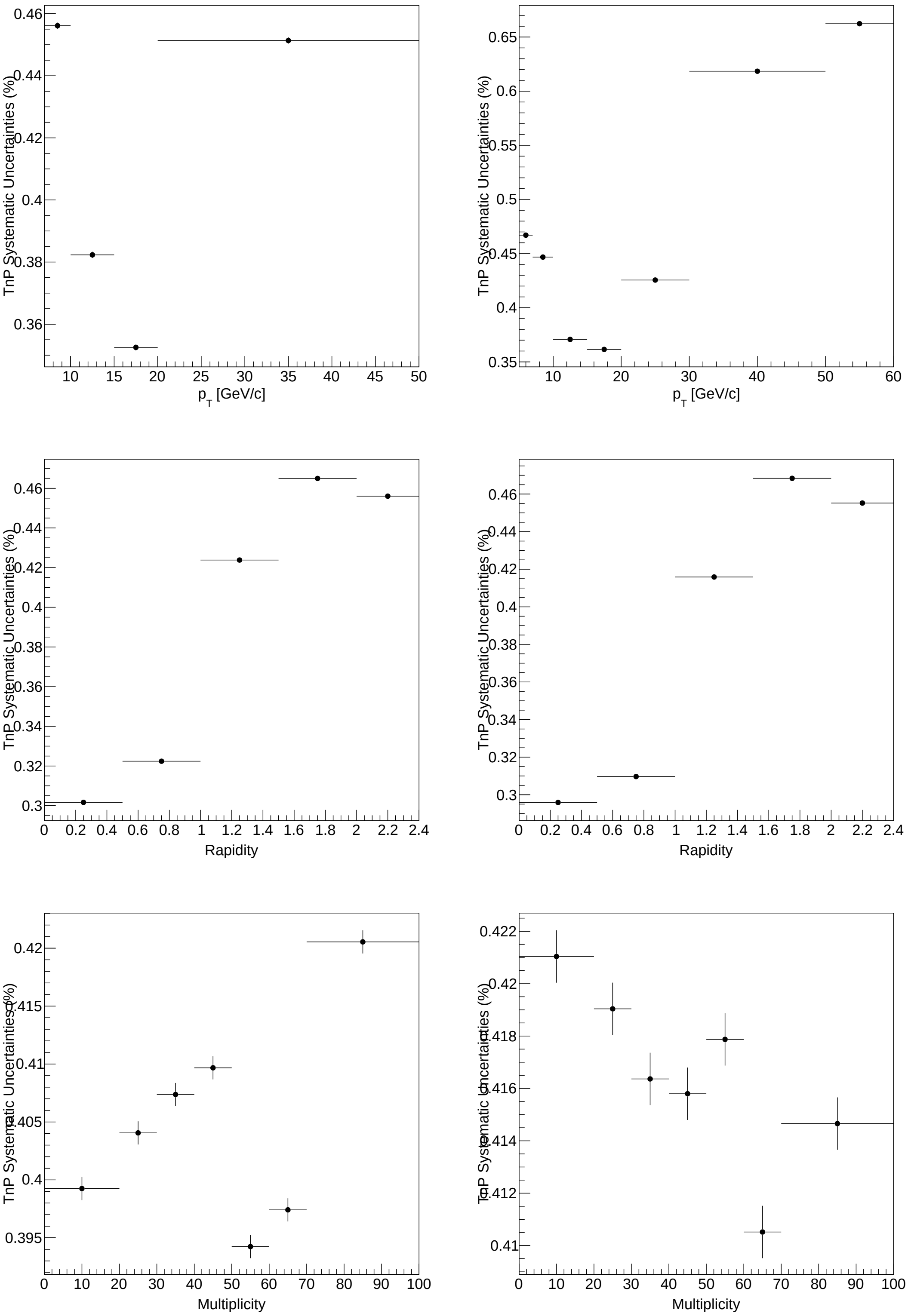

Figure C.3: Systematic uncertainties in total efficiency vs p_T (first row), vs rapidity (second row) and vs multiplicity (third row) for the tag and probe variation up and down, on the B_s^0 (first column) and B^+ (second column) mesons.

C.2.3 p_T shape

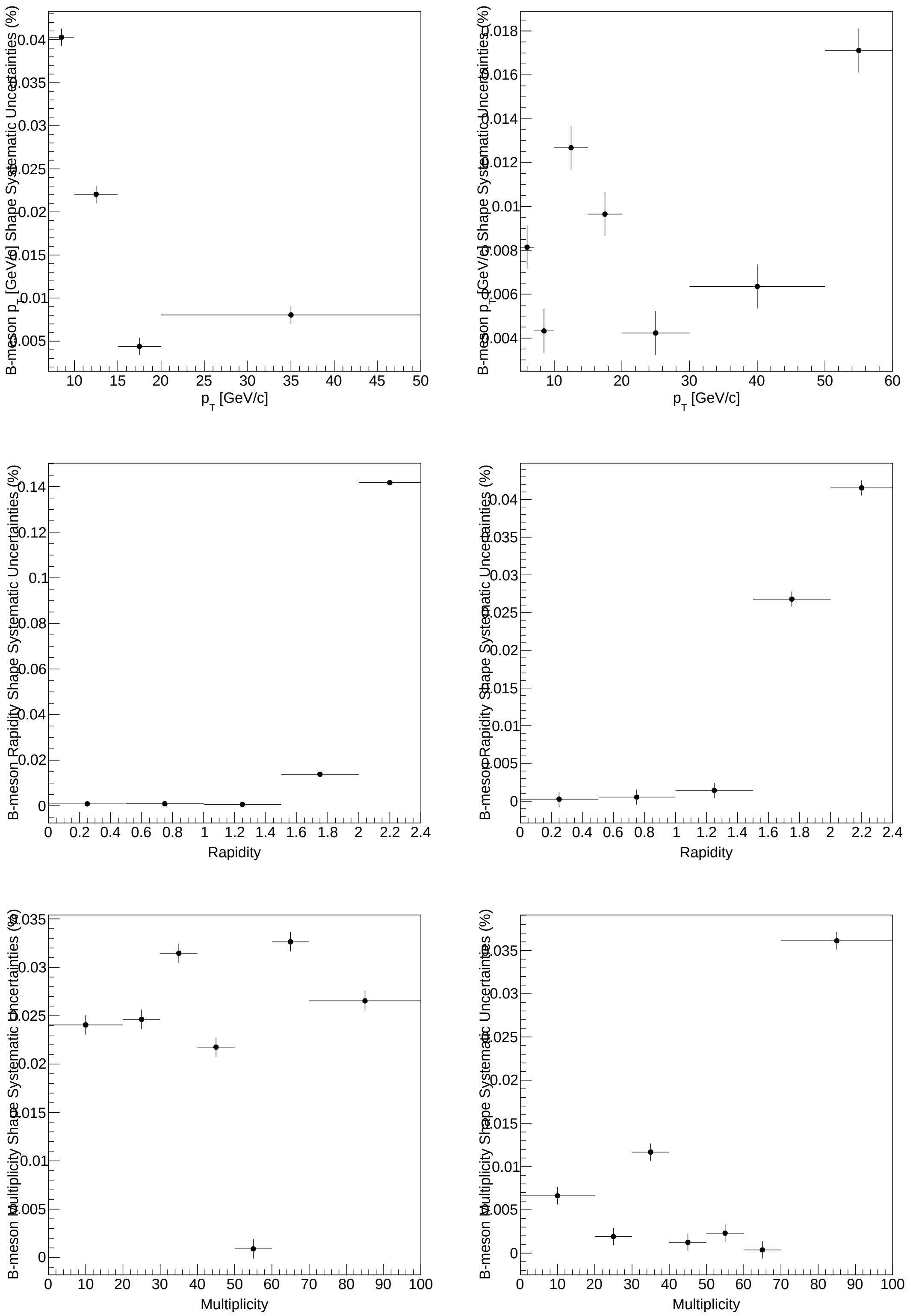

Figure C.4: Systematic uncertainties in total efficiency vs p_T (first row) and vs rapidity (second row) for p_T weights variation, on the B_s^0 (first column) and B^+ (second column) mesons.

Appendix D

Optuna optimization results

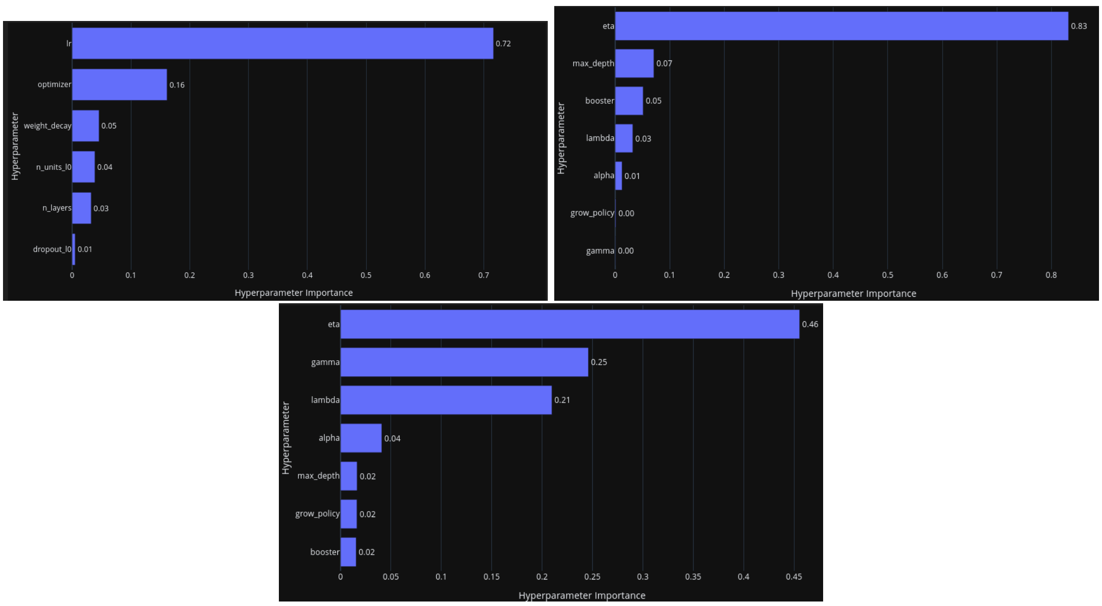

Figure D.1: Hyperparameter importance for Optuna during our B^+ NN (left) and BDT (middle) models and our B_s^0 BDT (right) model optimization.

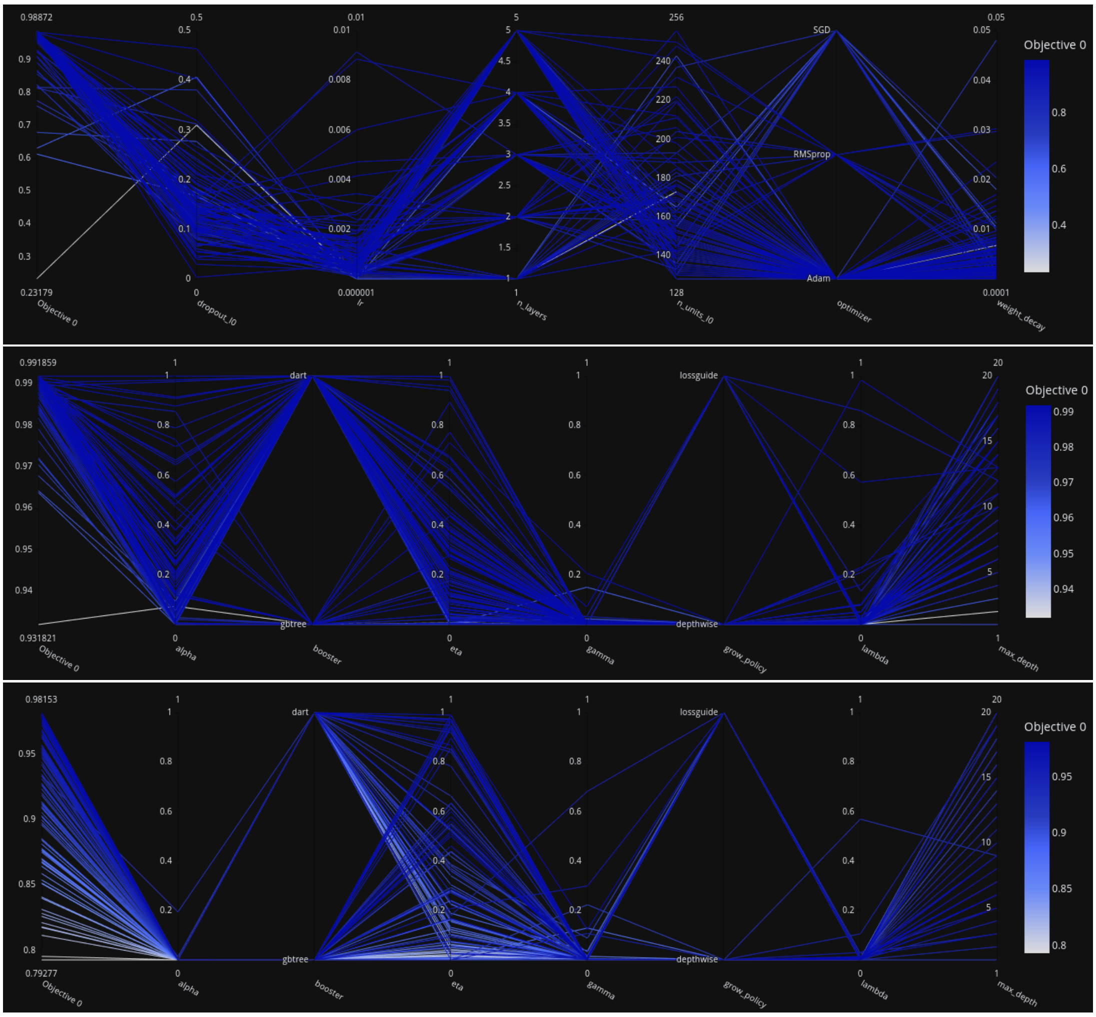

Figure D.2: Trial progression for Optuna during our B^+ NN (top) and BDT (middle) models and our B_s^0 BDT (bottom) model optimization. Each line represents one iteration of Optuna, showing the options chosen at that specific time. Places where the density of lines is higher represent the more favoured values during optimization.

Appendix E

Full results of models training and performance

E.1 training

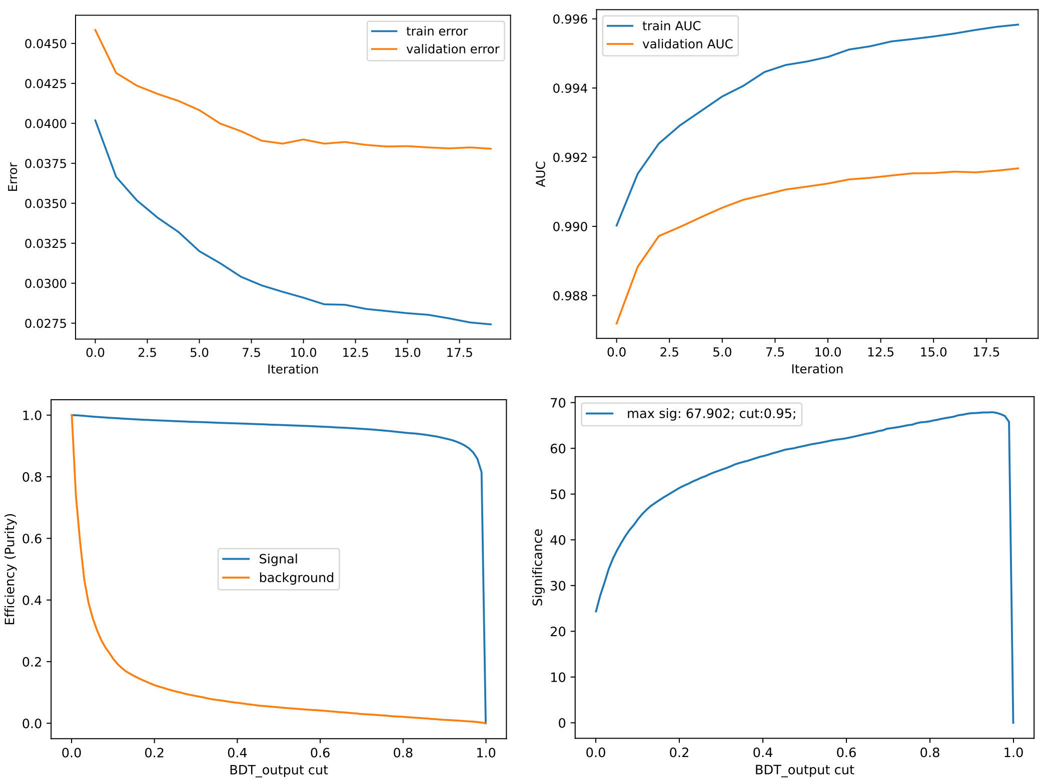

Figure E.1: Top row: Accuracy (left) and area under the ROC curve (right) on every iteration of the B^+ BDT model during training. Bottom row: Purity of signal and background events (left) and Significance (right) vs cut value in our BDT score.

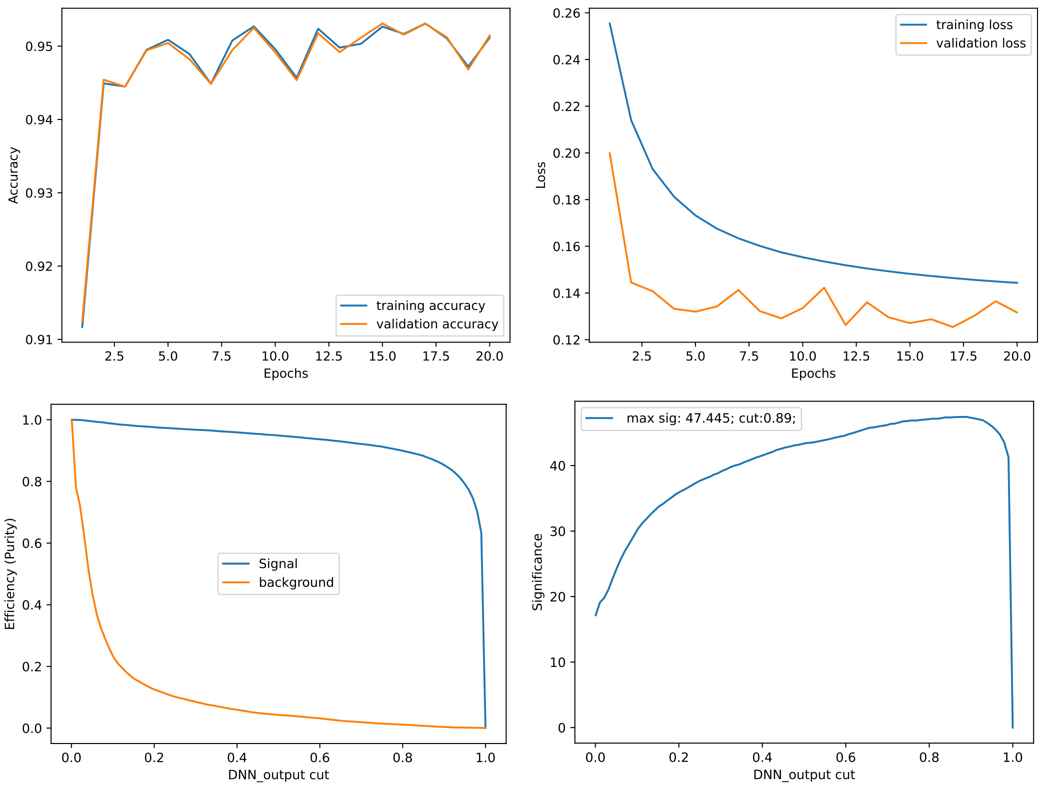

Figure E.2: Top row: Accuracy (left) and loss function value (right) on every iteration of the B^+ NN model during training. Bottom row: Purity of signal and background events (left) and Significance (right) vs cut value in our NN score.

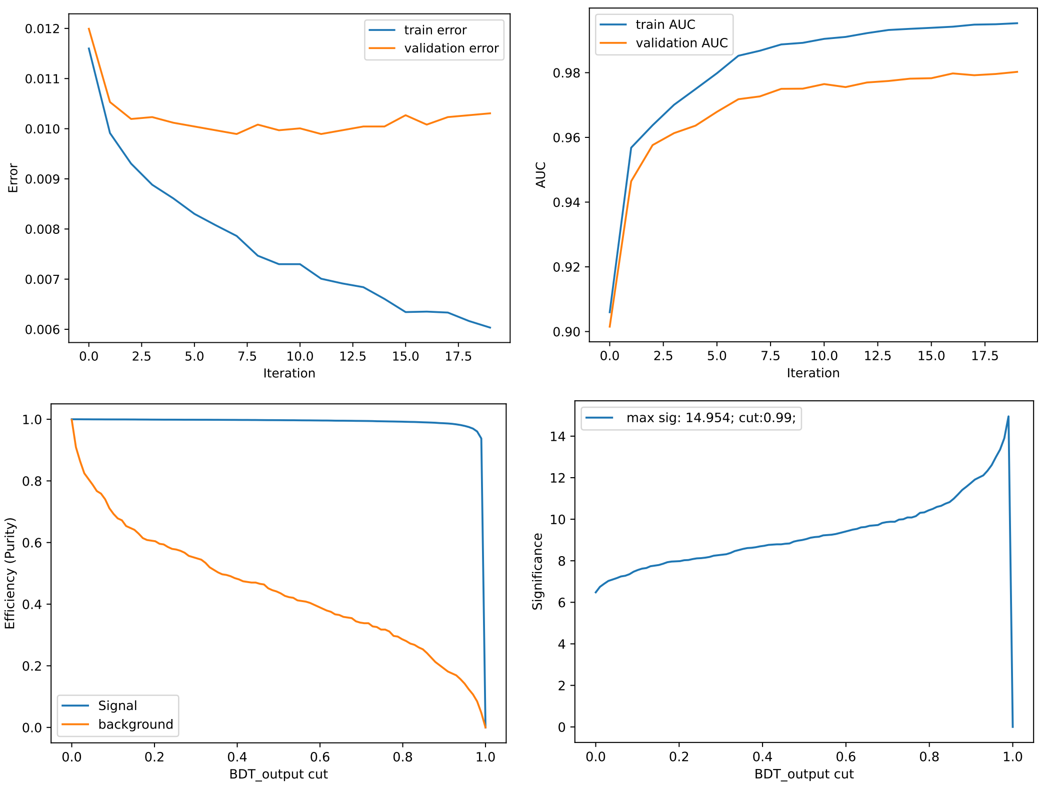

Figure E.3: Top row: Accuracy (left) and area under the ROC curve (right) on every iteration of the B_s^0 BDT model during training. Bottom row: Purity of signal and background events (left) and Significance (right) vs cut value in our BDT score.

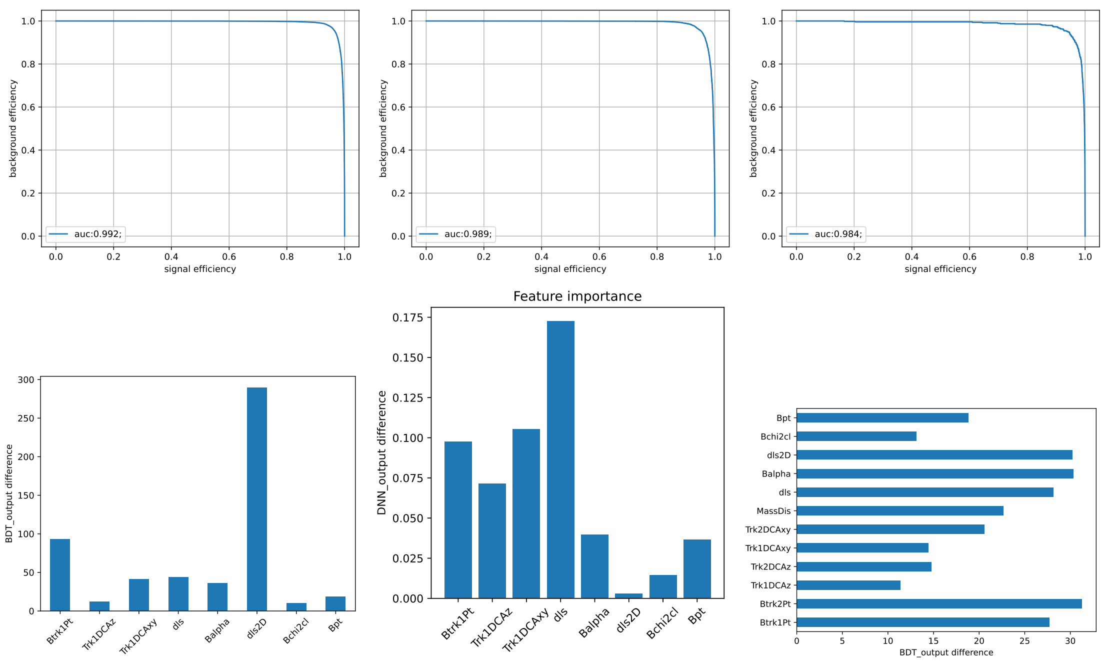

Figure E.4: Top row: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Feature importance for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right).

E.2 Working point

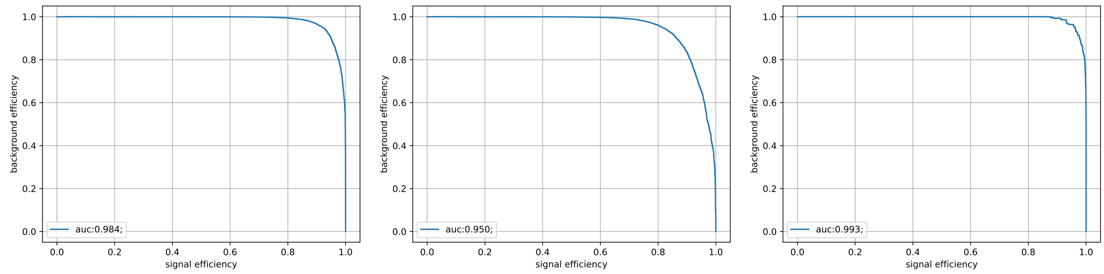

Figure E.5: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $3 < p_T < 5$ GeV/c.

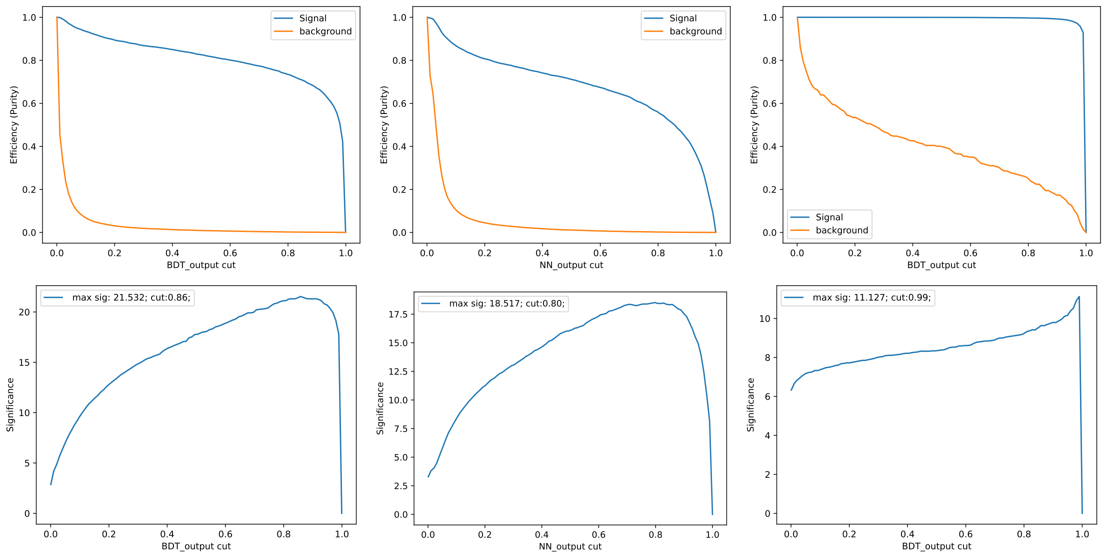

Figure E.6: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). All plots are for $3 < p_T < 5$ GeV/c.

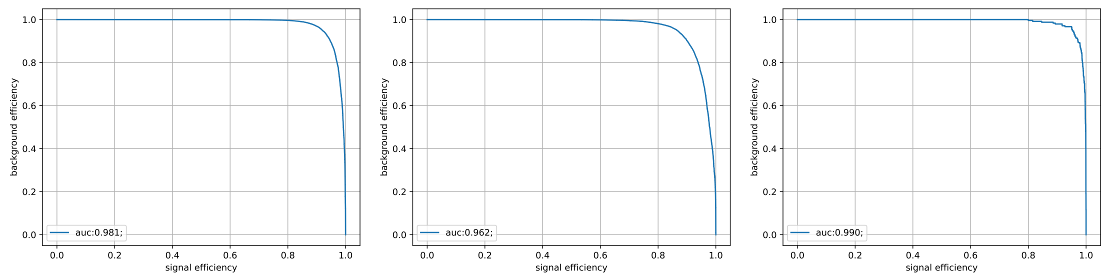

Figure E.7: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $5 < p_T < 7$ GeV/c.

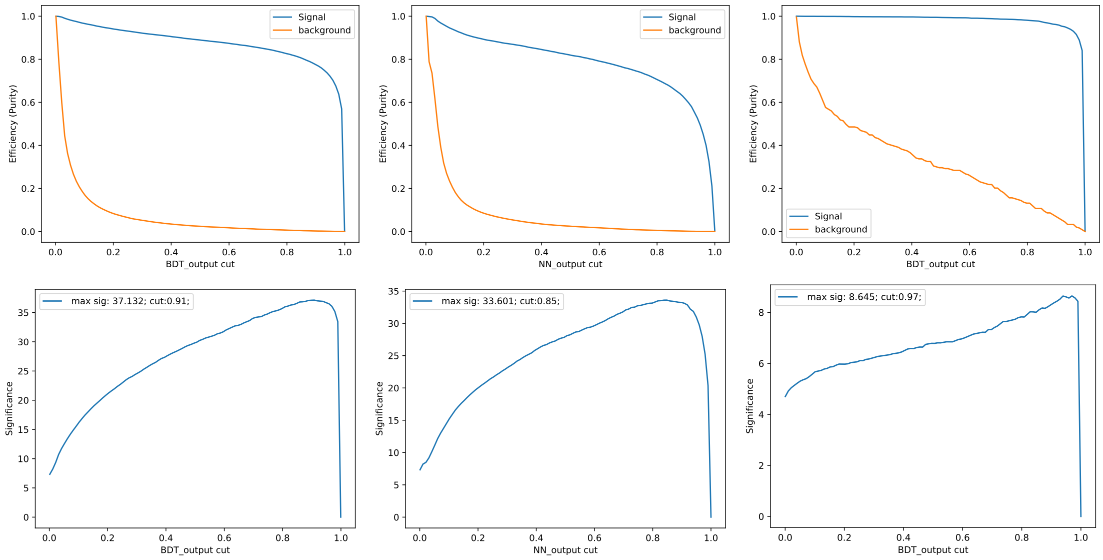

Figure E.8: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). All plots are for $5 < p_T < 7$ GeV/c.

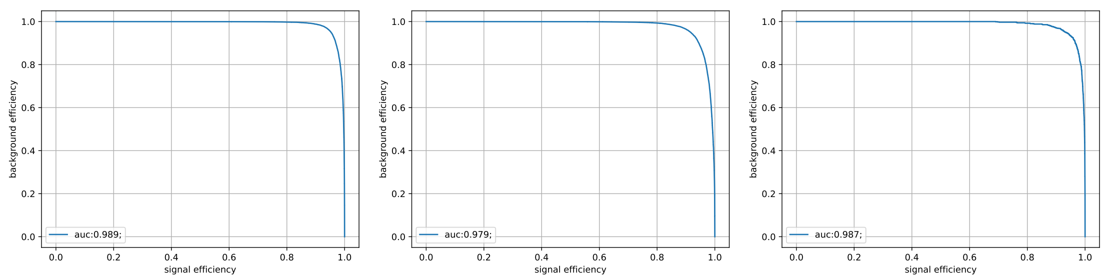

Figure E.9: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $7 < p_T < 10$ GeV/c.

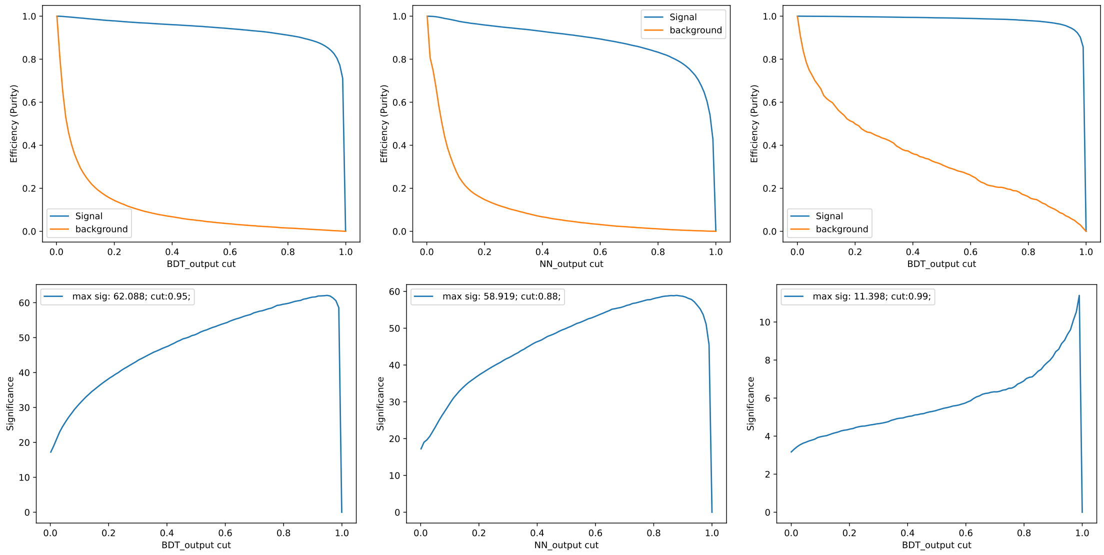

Figure E.10: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). All plots are for $7 < p_T < 10$ GeV/c.

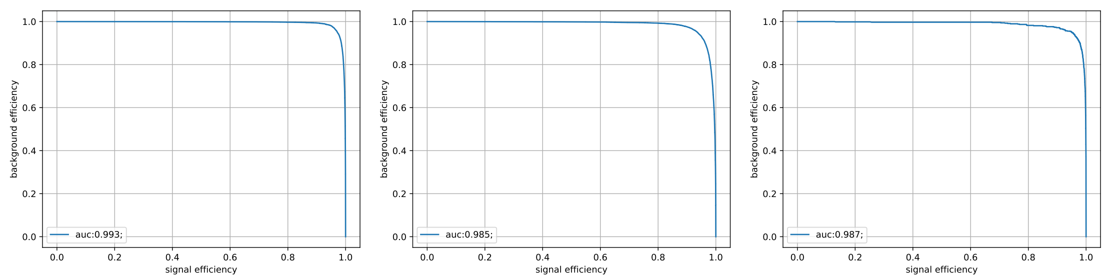

Figure E.11: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $10 < p_T < 15$ GeV/c.

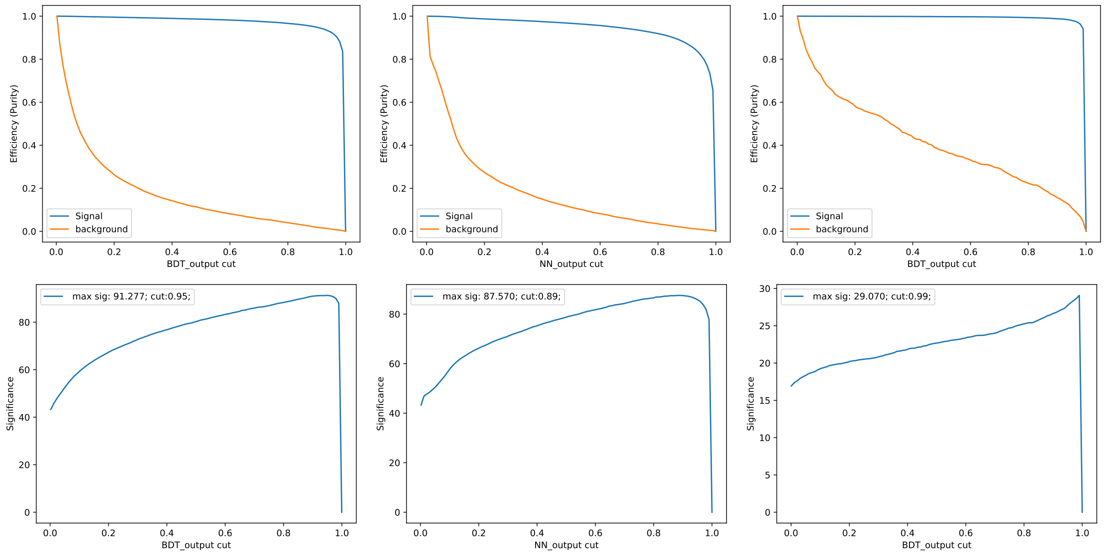

Figure E.12: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). All plots are for $10 < p_T < 15$ GeV/c.

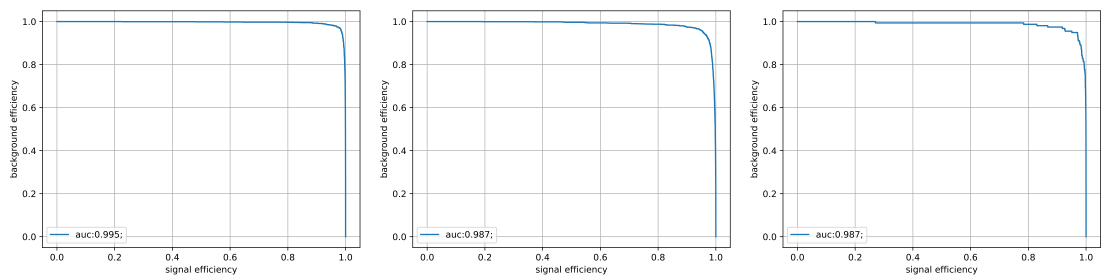

Figure E.13: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $15 < p_T < 20$ GeV/c.

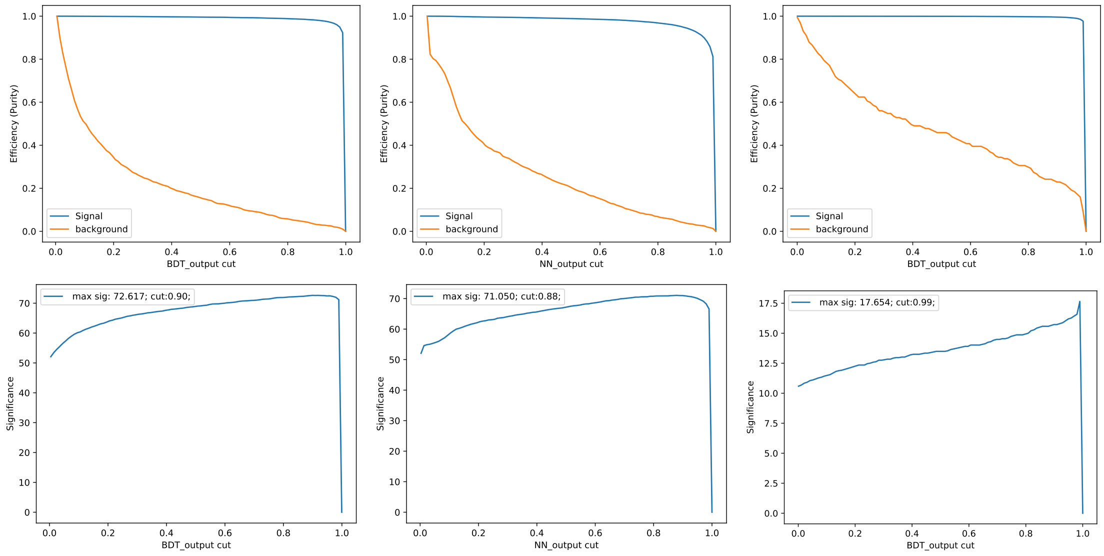

Figure E.14: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). All plots are for $15 < p_T < 20$ GeV/c.

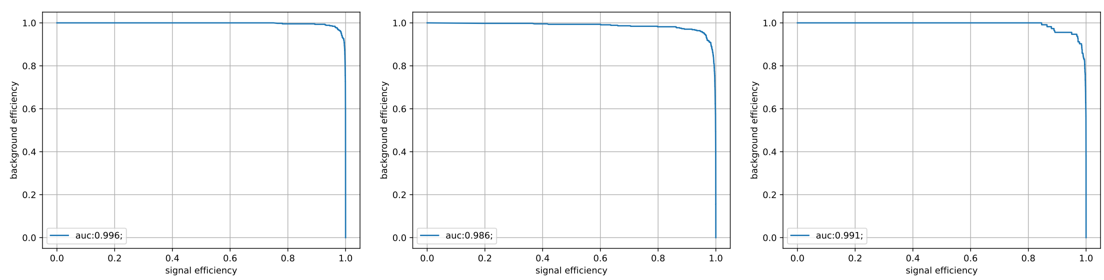

Figure E.15: ROC curve for our B^+ BDT model (left) and our NN model (middle), for $20 < p_T < 30$ GeV/c and for our B_s^0 BDT model (right), for $20 < p_T < 50$ GeV/c.

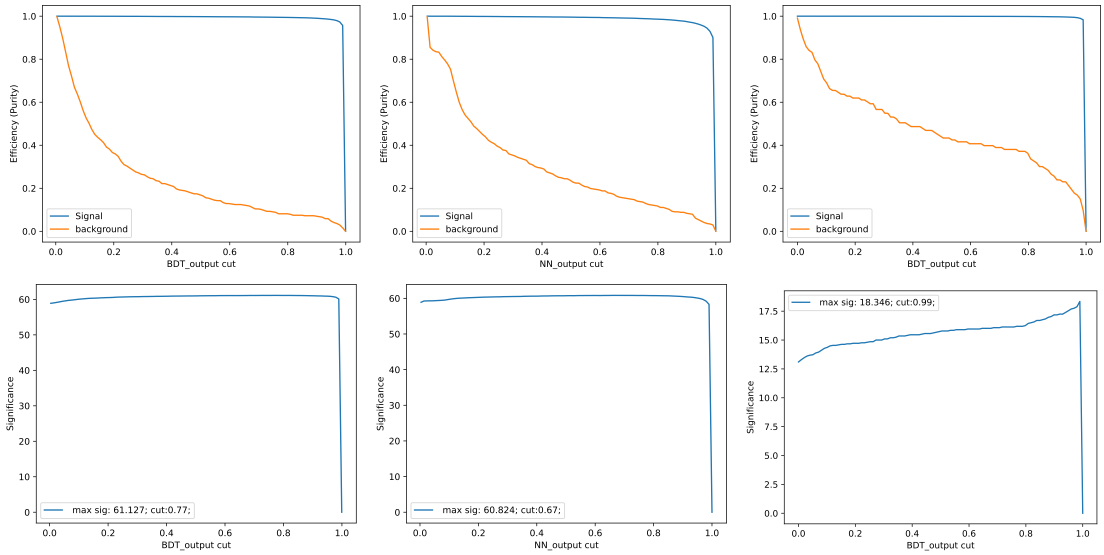

Figure E.16: Top row: Signal and background purity vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right). The B^+ meson plots are for $20 < p_T < 30$ GeV/c, whilst the B_s^0 plots are for $20 < p_T < 50$.

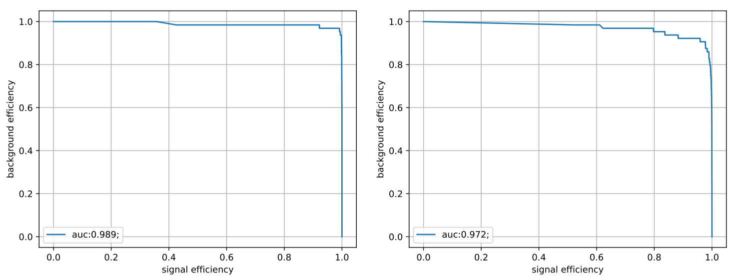

Figure E.17: ROC curve for our B^+ BDT model (left), our NN model (middle) and for our B_s^0 BDT model (right), for $30 < p_T < 50$ GeV/c.

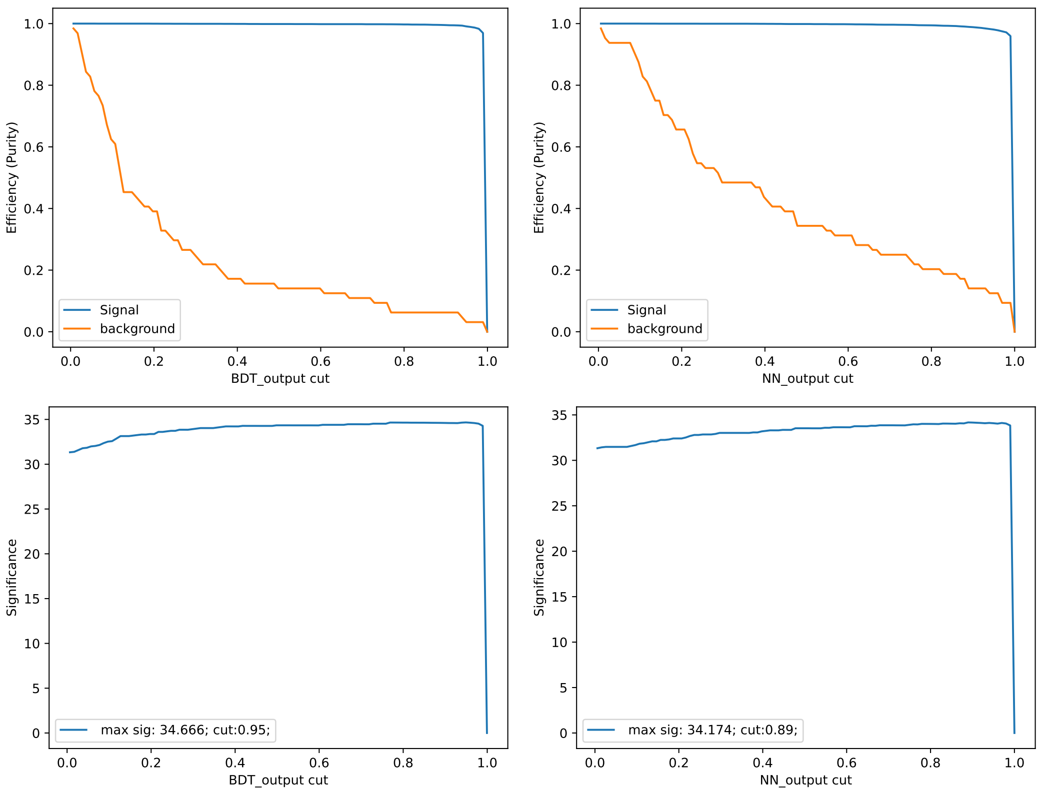

Figure E.18: Top row: Signal and background purity vs cut value for our B^+ BDT model (left) and our NN model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left) and our NN model (right). All plots are for $30 < p_T < 50$ GeV/c.

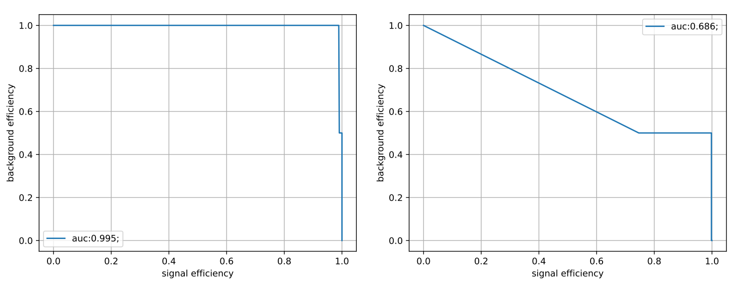

Figure E.19: ROC curve for our B^+ BDT model (left) and our NN model (right), for $50 < p_T < 60$ GeV/c.

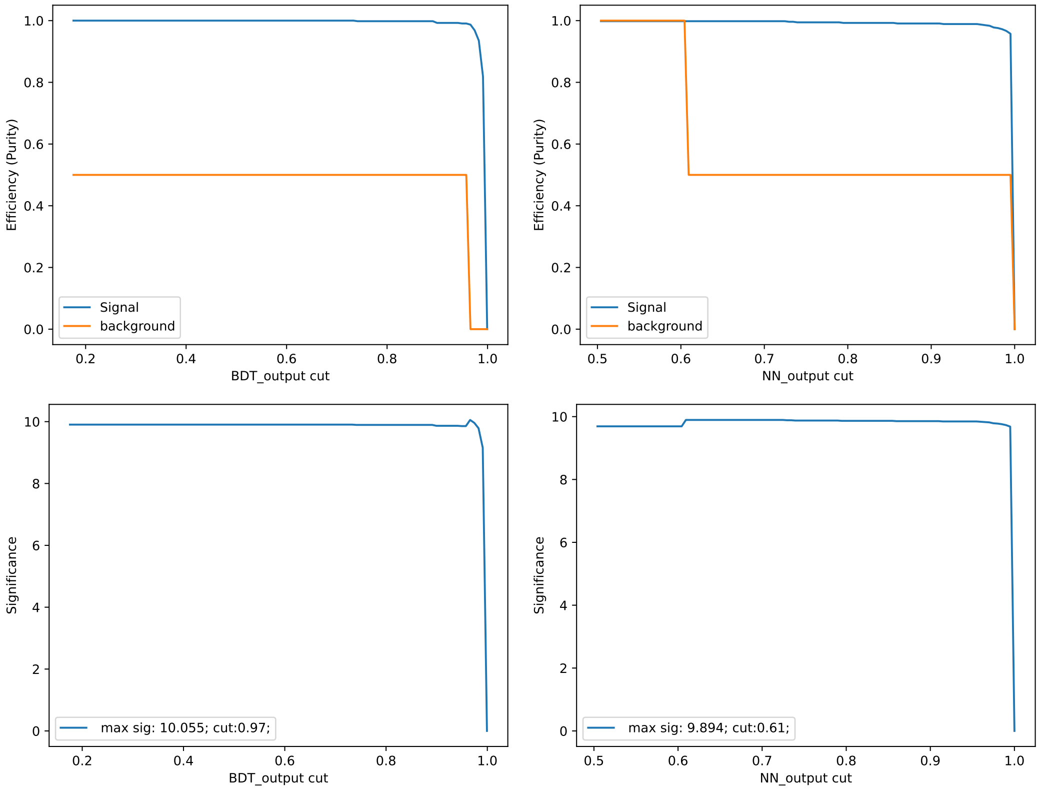

Figure E.20: Top row: Signal and background purity vs cut value for our B^+ BDT model (left) and our NN model (right). Bottom row: Significance vs cut value for our B^+ BDT model (left) and our NN model (right). All plots are for $50 < p_T < 60$ GeV/c.

Appendix F

Comparison with reference selection

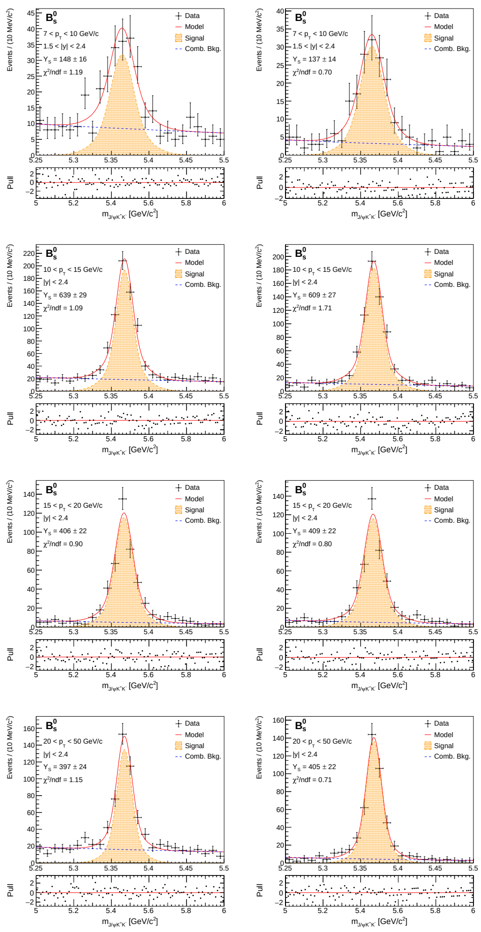

Figure F.1: Fit to B_s^0 meson invariant mass distribution, on the p_T bins. On the right column we show the mass fits for our nominal selection, whilst on the left column we show the fit with the reference selection.

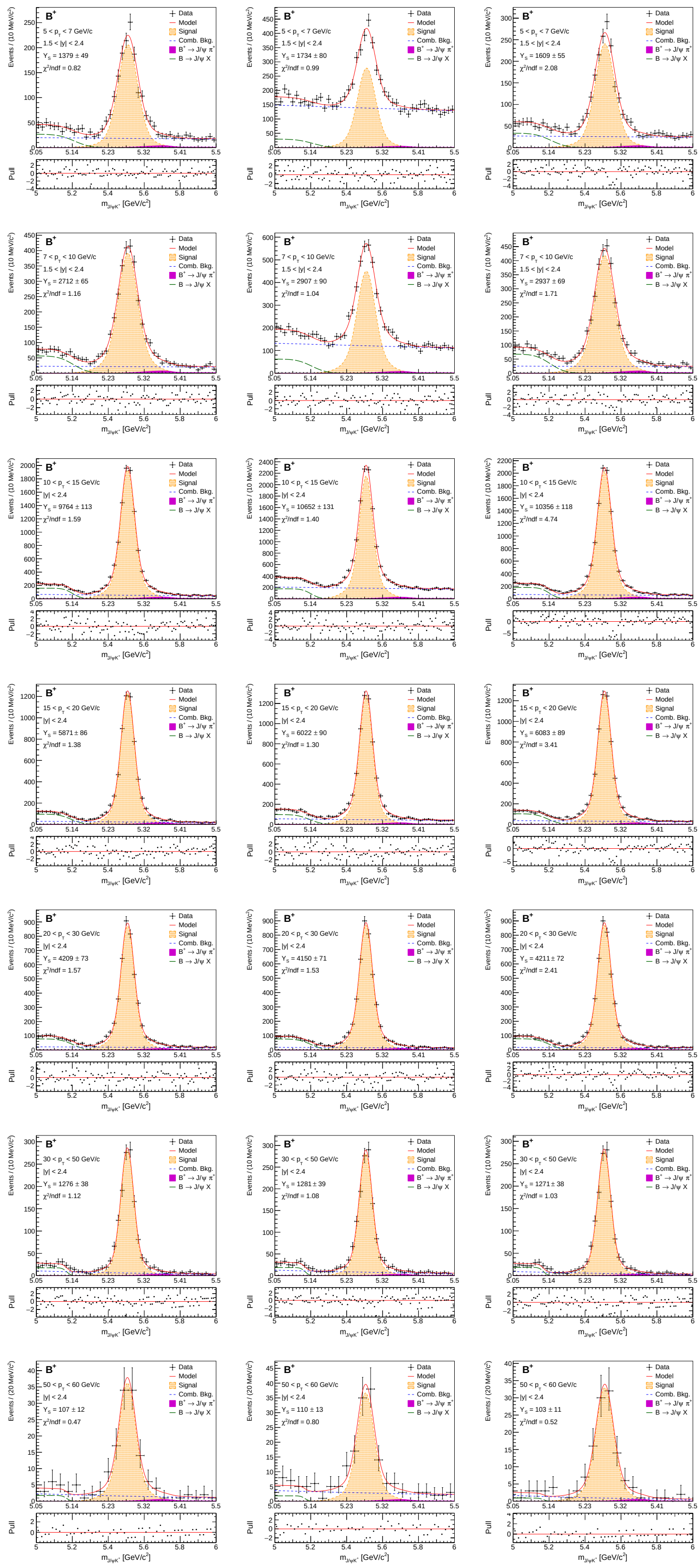

Figure F.2: Fit to B^+ meson invariant mass distribution, on the p_T bins. On the right column we show the mass fits for our nominal selection, whilst on the first and second columns we show the fits for the reference and NN model selections, respectively.

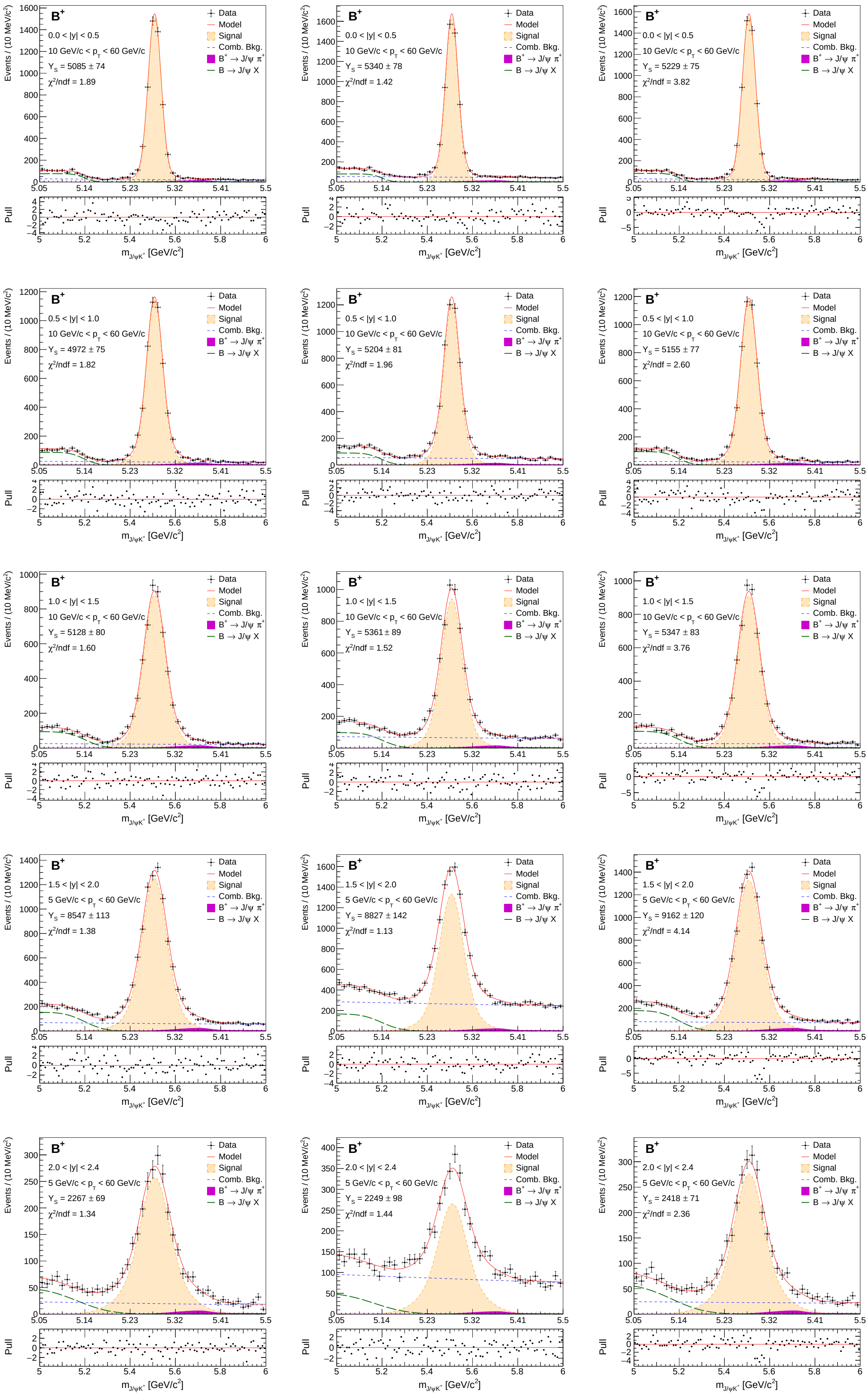

Figure F.3: Fit to B^+ meson invariant mass distribution, on the y bins. On the right column we show the mass fits for our nominal selection, whilst on the first and second columns we show the fits for the reference and NN model selections, respectively.

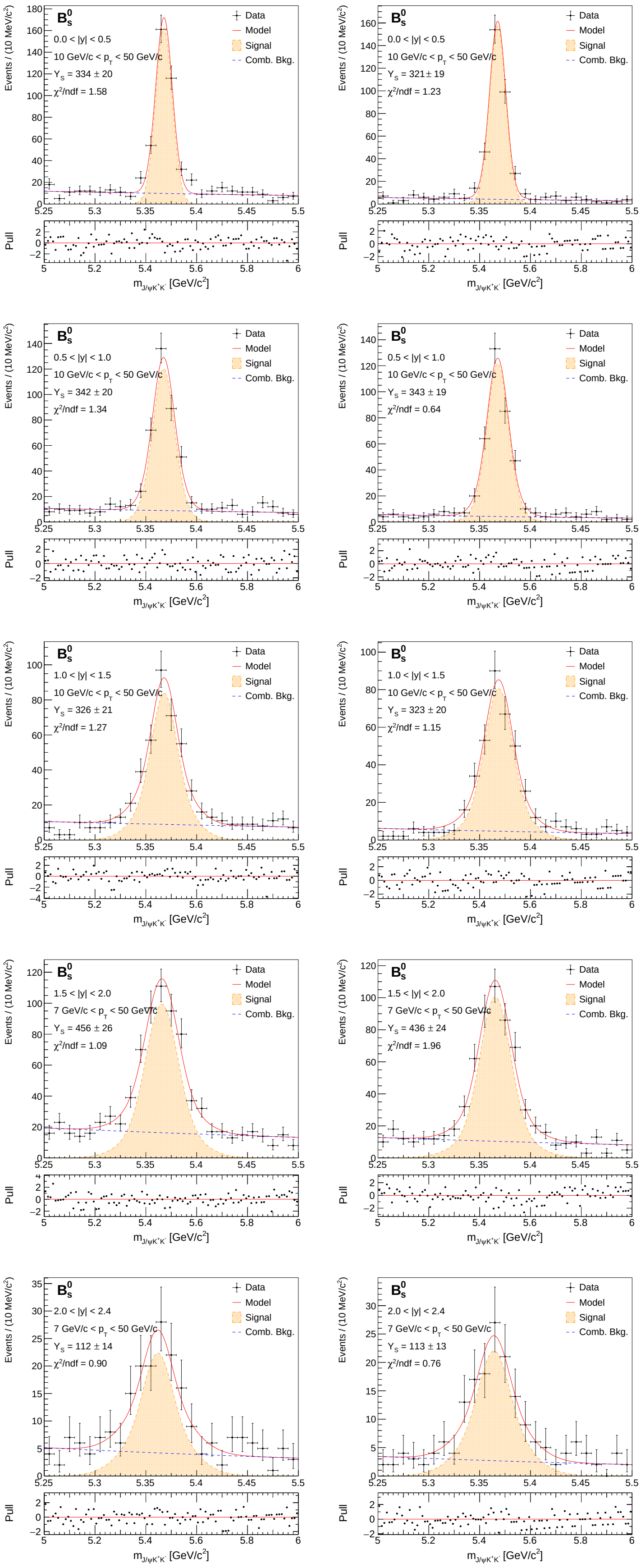

Figure F.4: Fit to B_s^0 meson invariant mass distribution, on the y bins. On the right column we show the mass fits for our nominal selection, whilst on the left column we show the fit with the reference selection.

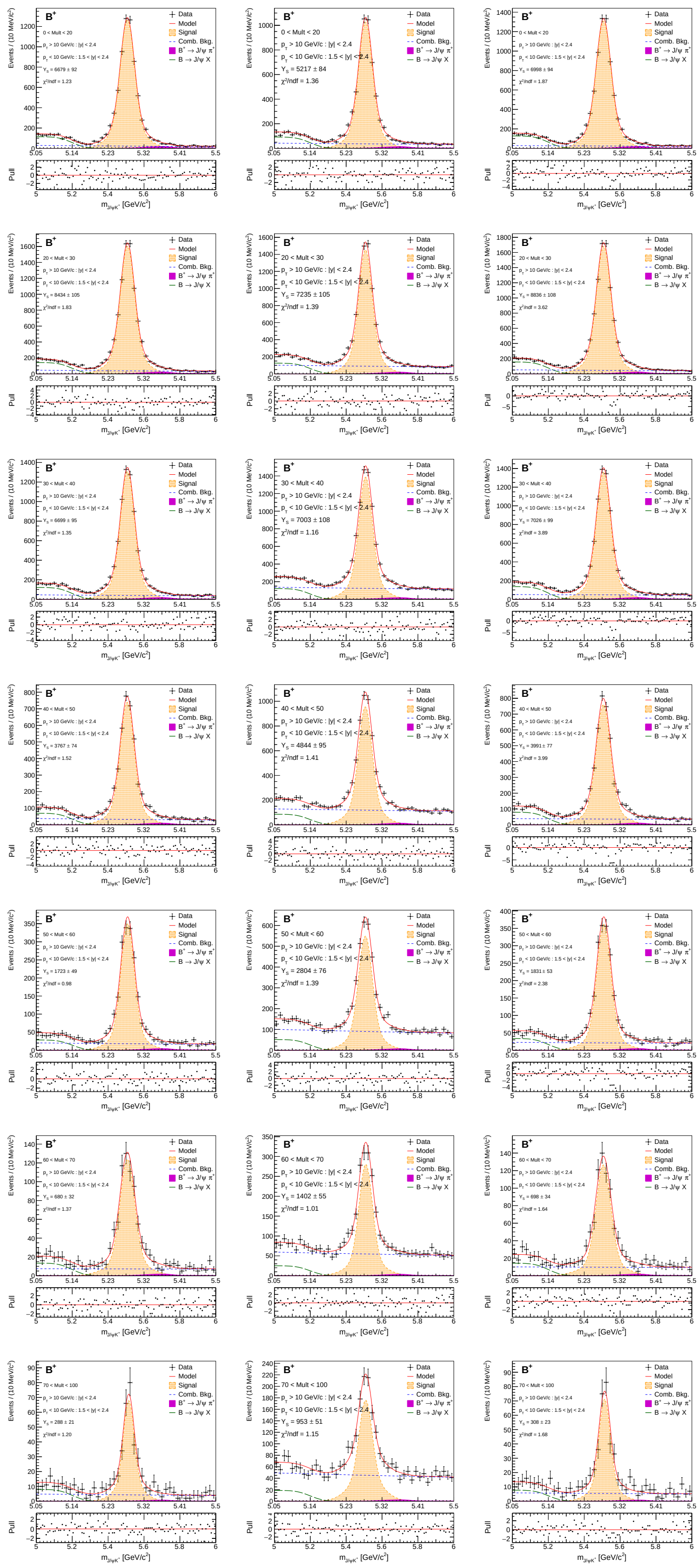

Figure F.5: Fit to B^+ meson invariant mass distribution, on the N_{trk} bins. On the right column we show the mass fits for our nominal selection, whilst on the first and second columns we show the fits with the reference and NN model selections, respectively.

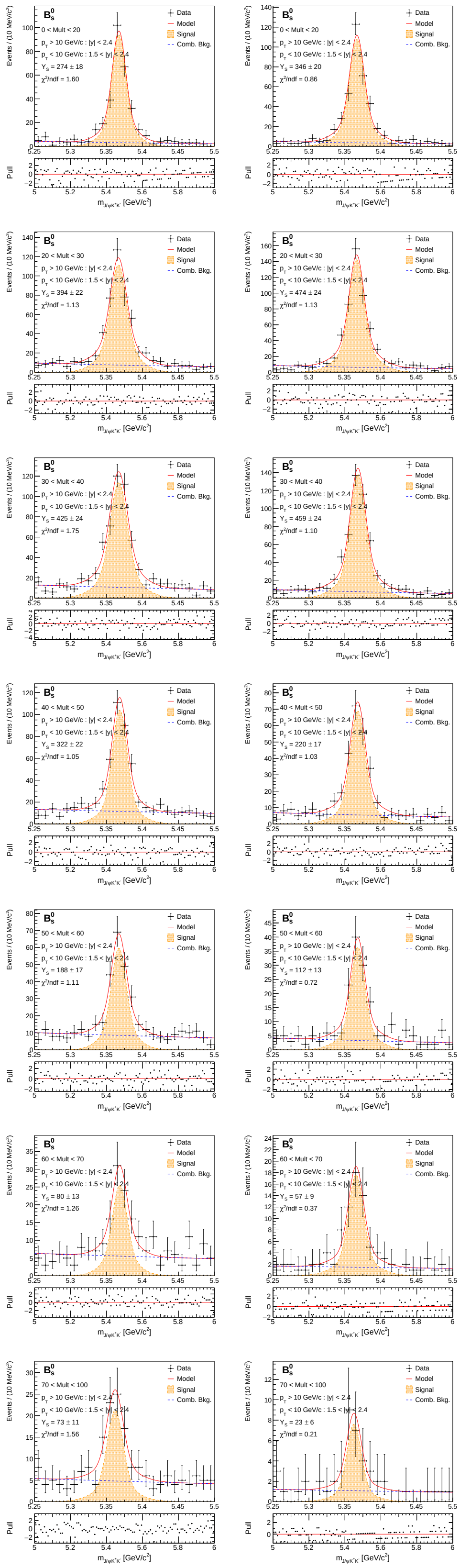

Figure F.6: Fit to B_s^0 meson invariant mass distribution, on the N_{trk} bins. On the right column we show the mass fits for our nominal selection, whilst on the left column we show the fit with the reference selection.