Centro Brasileiro de Pesquisas Físicas
Programa de Doutorado

Search for CP violation on $D_s^+ \to \pi^-\pi^+K^+$ decay with $LHCb$ run 2 data

Leonardo Santoro de Oliveira

Rio de Janeiro/RJ

2026

Search for CP violation on $D_s^+ \to \pi^-\pi^+K^+$ decay with $LHCb$ run 2 data

Leonardo Santoro de Oliveira

Doutorado submetida a coordenação de formação científica do Centro Brasileiro de Pesquisas Físicas (CBPF) como parte dos requisitos para a obtenção do título de Doutor em Física.

Orientador: Alberto Correa dos Reis

Coorientador: Felipe Luan Souza de Almeida

Rio de Janeiro/RJ

2026

"SEARCH FOR CP VIOLATION ON $D^+_S \to \pi^-\pi^+K^+$ DECAY WITH LHCb RUN 2 DATA"

LEONARDO SANTORO DE OLIVEIRA

Tese de Doutorado em Física apresentada no Centro Brasileiro de Pesquisas Físicas do Ministério da Ciência Tecnologia e Inovação. Fazendo parte da banca examinadora os seguintes professores:

Alberto Correa dos Reis - Orientador/CBPF

Diego Torres Machado

Diego Torres Machado – UERJ

Documento assinado digitalmente
gov.br
THIAGO RAFAEL FERNANDEZ PEREZ TOMEI
Data: 30/03/2026 12:49:51-0300
Verifique em https://validar.iti.gov.br

Thiago Rafael Fernandez Perez Tomei – UNESP

Sandro Fonseca de Souza - UERJ

Documento assinado digitalmente
gov.br
MARCELO GAMEIRO MUNHOZ
Data: 06/04/2026 11:04:25-0300
Verifique em https://validar.iti.gov.br

Marcelo Gameiro Munhoz – USP

Documento assinado digitalmente
gov.br
FELIPE LUAN SOUZA DE ALMEIDA
Data: 06/04/2026 15:07:35-0300
Verifique em https://validar.iti.gov.br

Felipe Luan Souza de Almeida - Coorientador/ Universidade de Syracuse

Rio de Janeiro, 26 de março de 2026.

Acknowledgements

Quando penso em começar os agradecimentos deste trabalho, por algum motivo a frase que vem à mente é “Se vi mais longe foi por estar apoiado nos ombros de gigantes”, proferida por Isaac Newton em uma de suas correspondências com Robert Hooke. Apesar do tom irônico dessa frase, para mim ela tem outro sentido. Se eu cheguei até aqui, foi por ter o apoio de muita gente. Nenhuma seção de agradecimentos estaria completa sem agradecer devidamente aos meus pais, Idimar Dias de Oliveira e Catia Cristina Santoro, que fizeram tudo o que acharam certo para apoiar toda a minha trajetória até aqui.

Também gostaria de agradecer aos meus orientadores, Alberto Correia dos Reis e Felipe Luan Souza de Almeida, por não só me orientarem nesses quatro anos, mas também por me aturarem durante todo esse processo. Também preciso agradecer às professoras Erica Polycarpo e Sandra Amato por todo o suporte e por todas as contribuições neste trabalho. Gostaria ainda de deixar um agradecimento especial à professora Carla Göbel, não só pelo suporte e por todas as inúmeras contribuições neste trabalho, mas também por ter me apresentado à área no mestrado e por ter aceitado a dura missão de me orientar anteriormente.

Também preciso deixar um agradecimento especial aos meus amigos da graduação, que ainda hoje me apoiam, ouvem minhas lamúrias e minhas piadas sem graça. Sem o apoio de vocês durante esses mais de dez anos, acredito que eu nunca nem teria saído do ciclo básico. Agradeço, então, à Bruna Gonçalves do Rego por sempre ser minha melhor amiga e estar comigo em todo esse processo; ao Vinicius Lessa Farinha, por ser meu eterno parceiro de discussões e lamúrias; ao Enzzo Pessanha Carvalho, que sempre me colocava para cima com uma piada ruim ou um desafio que me ajudava a evoluir; e ao Gabriel dos Santos Bezerra, cuja empolgação com a Física sempre me inspirou a não desistir. Eu nunca teria chegado aqui sem essa galera, por todas as horas e dias estudando juntos, aprendendo juntos, rindo e chorando juntos.

Também não poderia me esquecer dos amigos que fui fazendo no caminho, amigos que, de certo modo, acompanharam essa trajetória de perto e que até hoje são próximos de mim e colaboraram ativamente nesse processo. Gostaria de agradecer ao Arthur da Cunha Paes Gil Ventura por todas as conversas fora da casinha e pelas coisas aleatórias que aprendi de outras áreas; ao André Luiz Machado Pessanha, por sempre me inspirar e motivar a ir mais longe; ao Luis Gustavo Soares de Oliveira, que foi meu primeiro amigo na graduação e com quem até hoje continuo a compartilhar apoio nos acontecimentos da vida. Também agradeço ao Ricardo Teixeira, que sempre ouve e compra minhas ideias de maluco há mais de dez anos, e ao Pedro Ivo Barcelos Moreira, por todas as discussões fora

da casinha desde a nossa adolescência.

Também não poderia deixar de agradecer aos “charmosinhos”, que direta ou indiretamente colaboraram com todo o processo de formação. Gostaria de agradecer ao Juan Baptista de Souza Leite, que foi peça-chave na ponte que me permitiu ir ao CERN. Agradeço ao Lucas Nicholas Falcão Ferreira pelas horas de conversa e por suas sempre brilhantes sugestões, que me ajudaram a continuar caminhando; ao nosso “traidor” Lucas Cordeiro Romão, que muito me ajudou a seguir no mestrado; ao Ivanildo Rui Barbosa, por todos os papos e trocas que tivemos durante o doutorado; ao Gustavo Alejandro Loachamin Ordoñes, por sempre ser um ótimo ouvinte das minhas lamúrias; e ao Alexandre Cardoso Campos, por todos os papos e trocas que tivemos.

Nenhum agradecimento estaria completo sem agradecer às inúmeras pessoas que me apoiaram durante minhas estadias no CERN. Gostaria de agradecer ao Lucas Meyer Garcia por impedir que eu ficasse louco no CERN e por todos os hambúrgueres que comemos; ao Jean-François Merchand, por tudo o que me ensinou sobre o experimento e por todo o apoio que sempre me deu; ao Florian (Flo) Reiss e à Mariana Fontana por me convidarem a ir ao CERN mais uma vez para terminar meu trabalho, pelas incríveis discussões sobre o projeto e por toda a paciência comigo. Também gostaria de agradecer ao Rosen Matev por todo o suporte e por tudo o que me ensinou sobre programação, e a todos os amigos que fiz enquanto trabalhava na calibração do calorímetro eletromagnético: Giulia Tuci, Giulia Frau, Niels Breer e Miguel Ruiz Diaz.

Claro que não poderia deixar de agradecer à minha esposa, Kathleen Matheus Freitas. Sem seu apoio, incentivo e amor, eu nunca teria terminado a graduação. Agradeço por essa força que sempre me faz andar e não desistir. Mesmo tentando, eu não conseguiria colocar em palavras os sentimentos que tenho quando penso nela, muito menos expressar o quanto sou grato por ela ter entrado na minha vida.

Apesar de soar um pouco arrogante, também preciso agradecer a mim mesmo. Nunca é fácil, para alguém com baixa autoestima, reconhecer uma vitória. Este trabalho não teria sido possível se eu não tivesse superado inúmeras adversidades, se não tivesse dado a volta por cima nas vezes que tentaram me derrubar e por ter sido teimoso o suficiente para nunca ceder aos pensamentos intrusivos de desistir.

Por fim, preciso agradecer à pessoa que estaria mais orgulhosa com esta tese, mas que infelizmente não me viu sequer terminar a graduação. A cada passo da minha vida acadêmica, agradeço à minha avó, Maria das Graças Santoro, que sempre fez de tudo para que eu chegasse aonde estou e por todas as lições que aprendi com ela. Sei que, se ela estivesse aqui, com certeza seria a pessoa mais orgulhosa em me chamar de físico, mestre e doutor. Agradeço também à CAPES e o CNPq por todo o apoio financeiro neste trabalho.

Resumo

Entender e explicar a Violação de CP historicamente permitiu inúmeros trabalhos promoverem novas descobertas no campo de altas energias, incluindo até a previsão de novos quarks. A Violação de CP é um fenômeno que está diretamente ligado ao fato de se observarem diferentes taxas de decaimento de partículas e antipartículas. Atualmente, o conhecimento que se tem das fontes desse fenômeno ainda não permite explicar completamente a bariogênese, sendo necessário que existam mais fontes de violação de CP para conseguir explicar a assimetria de matéria-antimatéria observada no Universo. Buscas por essas novas fontes de violação de CP hoje estão ligadas à "nova física" além do Modelo Padrão de partículas.

O setor do quark *charm*, diferente do setor do quark *beauty*, ainda hoje é, de certo modo, inexplorado. Apenas uma medida nesse setor havia sido observada até o momento da publicação deste trabalho, no ano de 2019, pela colaboração LHCb. As assimetrias esperadas nesse setor são da ordem de 10^{-3} ou menores em decaimentos ditos suprimidos por Cabibbo, tornando estudos desse tipo bastante desafiadores do ponto de vista experimental. Porém, medidas maiores que essa podem indicar nova física além do Modelo Padrão.

Desse modo, este trabalho se dedica à busca de Violação de CP no setor do *charm* por meio do decaimento $D_s^+ \to \pi^-\pi^+K^+$, que é um decaimento suprimido por Cabibbo. Para essa busca, este trabalho utilizou dados coletados durante o *Run* 2 do LHCb, produzidos através de colisões próton-próton com energia de centro de massa de 13 TeV, obtendo uma amostra a ser analisada com cerca de 32 M de candidatos a eventos de sinal. A análise foi realizada utilizando uma técnica independente de modelo para o espaço de fase, chamada mirandizing, que computa a diferença de significância estatística entre as populações de D_s^+ e D_s^- em bins do espaço de fase.

Além dessa análise, este trabalho também apresenta a contribuição feita para a tomada de dados do *Run* 3 do LHCb, realizada por meio da finalização do *upgrade* e da otimização do processo de calibração do calorímetro eletromagnético.Crucial para identificação de leptons produzidos o LHCb.

Palavras-chave: física de charme; Decaimentos hadronicos; Violção de CP;

Abstract

Understanding and explaining CP violation has historically enabled numerous studies to drive new discoveries in high-energy physics, including the prediction of new quarks. CP violation is a phenomenon directly associated with the observation of different decay rates for particles and their corresponding antiparticles. At present, the known sources of this phenomenon are still insufficient to fully explain baryogenesis, implying that additional sources of CP violation must exist in order to account for the matter–antimatter asymmetry observed in the Universe. Searches for such new sources of CP violation are therefore closely connected to physics beyond the Standard Model.

The charm-quark sector, unlike the beauty-quark sector, remains relatively unexplored to this day. At the time of publication of this work, only a single measurement of CP violation in this sector had been reported, by the LHCb collaboration in 2019. The asymmetries expected in charm decays are of the order of 10^{-3} or smaller for Cabibbo-suppressed decays, making such studies particularly challenging from an experimental point of view. However, measurements significantly larger than this expectation could indicate the presence of new physics beyond the Standard Model.

In this context, this work is devoted to the search for CP violation in the charm sector through the study of the Cabibbo-suppressed decay $D_s^+ \to \pi^-\pi^+K^+$. This search is performed using data collected during Run 2 of the LHCb experiment, corresponding to proton–proton collisions at a center-of-mass energy of 13 TeV. The final dataset used in the analysis contains approximately 32 million signal candidate events. The analysis is carried out using a model-independent phase-space technique known as mirandizing, which evaluates the statistical significance of the difference between the D_s^+ and D_s^- populations in bins of the Dalitz plot.

In addition to the physics analysis, this work also presents the author's contributions to Run 3 data taking at LHCb, achieved through the completion of the upgrade and the optimization of the electromagnetic calorimeter calibration procedure. This calibration is a crucial component for the identification of leptons produced in the LHCb experiment.

Key words: Charm physics. Hadronic decays. charge-parity violation.

Sumário

Lista de ilustrações

Lista de tabelas

Não se pode aprender nada de uma lição que não seja acompanhada por dor, já que não se pode conseguir nada sem um sacrifício. Mas quando você aguenta essa dor e a supera, as pessoas conseguem um coração forte que não perde para nada.

Hiromu Arakawa, Fullmetal Alchemist

1 Introduction

The Standard Model is the most successful theory to date for describing the smallest known quantum structures and was developed collaboratively over many decades, essentially since the advent of quantum mechanics. It stands as one of the most successful physical models ever formulated. However, despite its many successes, the Standard Model still fails to answer several fundamental questions, such as the quantum nature of gravity, the origin of neutrino masses, and the matter–antimatter asymmetry observed in the Universe.

The generation of a matter–antimatter asymmetry is directly related to processes of CP violation, which constitute one of the necessary conditions for baryogenesis according to the Sakharov conditions. The sources of CP violation described within the Standard Model are not sufficient to account for the observed asymmetry, implying that additional sources must exist.

A particularly promising arena for the study of CP violation is the charm-quark sector, where the expected effects are of the order of 10^{-3} according to theoretical predictions. Any observation of CP violation significantly larger than this expected order of magnitude could be an indication of physics beyond the Standard Model. Experimental observation of CP violation in this sector are still scarce, with the only measurement reported in 2019 in neutral charm mesons by the LHCb collaboration [10].

In this context, this work focuses on the analysis of the decay of the D_s meson into the $\pi^-\pi^+K^+$ final state. This is the D_s decay with the largest branching fraction among its Cabibbo-suppressed three-body final states. Moreover, this decay is Cabibbo-suppressed, meaning that CP violation is already expected at some level within the Standard Model. An observation of CP violation in this channel would constitute the first measurement of this phenomenon in a charged charmed meson.

This analysis uses data collected during Run 2 of the LHCb experiment, which took place from 2016 to 2018 and corresponds to the same data-taking period used for the first measurement of CP violation in the charm-quark sector. LHCb is an experiment specifically designed to study CP violation and rare decays, and during Run 2 it collected data from proton–proton collisions at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 5.6 fb^{-1}.

This collision energy enabled the production of data samples of unprecedented size. As an example, for the decay channel studied in this work, the final selected sample contains approximately 32 million signal candidates for the D_s meson. This represents a sample vastly larger than that used by the FOCUS collaboration in 2005 [11] and is

statistically comparable to the dataset used in the 2019 publication reporting CP violation in the charm sector.

The strategy adopted to search for CP violation in this channel is based on a model-independent technique that maps the significance of asymmetries across the decay phase-space, represented by the Dalitz plot, this technique is called Mirandizing. In this approach we divide the Dalitz plot into bins and, using the signal yields in each bin, computes an observable that quantifies the statistical significance of the difference between the D_s^+ and D_s^- populations in that region. The observable defines a χ^2 test statistic under the null hypothesis of CP symmetry. The corresponding p-value is computed as the probability, assuming CP symmetry, of obtaining a test statistic at least as extreme as the one observed in data. Very small p-values can be translated into a Gaussian-equivalent significance; a significance of five standard deviations ($\sigma = 5$), corresponding to a p-value of order 10^{-7}, is conventionally required to claim observation of CP violation.

This work therefore focuses on describing how the extraction of the D_s^+ and D_s^- decay yields, as well as their asymmetries, is performed using simultaneous fits to the $\pi^-\pi^+K^+$ invariant mass spectrum. In addition, results from studies of asymmetries in the background region are presented.

The presence of spurious asymmetries, arising from instrumental or production effects, may bias the measurement. Therefore, in order to allow a reliable interpretation of the results in terms of CP conservation within the framework of the method described above, it is necessary to ensure that such effects are negligible for this analysis. In general, control channels featuring the same final-state particles are employed for this purpose. In the absence of a suitable control channel for this analysis, this work also presents the strategies adopted to study the presence of spurious asymmetries, either through simulations or via cross-checks using two Cabibbo-favored decay channels.

Following LHCb policy, analyses are conducted in a blinded, and this analysis is still under review by the corresponding working group. As a result, authorization to unblind and analyze the signal region has not yet been granted.

Accordingly, Chapter 2 provides a theoretical introduction to CP violation, how it arises within the Standard Model, the motivation for its study in the charm-quark sector, and the formalism of three-body decays.

Chapter 3 describes the LHCb experiment, its subdetectors, and its trigger system. This chapter also includes a section dedicated to the LHCb Upgrade I, implemented for Run 3 data taking, which enabled the experiment to operate under significantly higher luminosity conditions.

The data selection is described in Chapter 4, detailing the full selection procedure starting from the trigger level, including the study of charm-related backgrounds in this

channel and the multivariate analysis used for event selection. This process ultimately yields a sample of approximately 32 million signal candidates.

The analysis method, the strategy adopted for simultaneous fits, and the validation of both the observable and the fitting procedure are presented in Chapter 5. This chapter explains the Mirandizing technique and how it is applied in the context of this analysis. And presents results from studies of the background region, aimed at evaluating the presence of asymmetries in this region and assessing whether such effects could contaminate the signal-region analysis under the proposed strategy.

Chapter 6 describes the full simulation-based studies performed using RapidSim to investigate possible spurious production and detection asymmetries. This chapter also presents additional cross-check results obtained from two Cabibbo-favored decay channels.

Finally, in addition to the physics analysis described above, Chapter 7 documents the contributions made to Run 3 data taking through the calibration of the electromagnetic calorimeter. This chapter details the calibration procedure itself and presents the results obtained for the 2024 and 2025 data-taking periods.

2 Theoretical Framework

2.1 An overview of the Standard Model

The most successful model to describe the smallest constituents of nature is the standard model (SM) of particles. This framework was designed unifying new discoveries in the beginning of the 1940s [12, 13, 14] and the theoretical models to describe them. The SM is a non-Abelian theory based on the $SU(3)_C \times SU(2)_L \times U(1)_Y$ group with spontaneous symmetry breaking. The $SU(3)_C$ here is related to the strong interactions in the form of the Quantum Chromodynamics (QCD), while the $SU(2)_L \times U(1)_Y$ are the electroweak sector before the spontaneous symmetry breaking.

In the SM, the smallest "bricks" of nature can be recognized in three sectors: fermionic, bosonic, and the Higgs doublet. In the fermionic sector, we find half-spin particles in the form of six flavored quarks and three charged lepton flavors (and their respective neutrinos). In the bosonic sector, we find integer-spin particles responsible for describing the fundamental forces of nature: electromagnetic (photon), strong (gluons), and weak ($W^{\pm}$ and Z^0) forces. Finally, in addition, there is the Higgs doublet, related to the Higgs mechanism, which is responsible for providing the masses of all massive particles. A pictorial representation of all these particles is shown in Fig. 1.

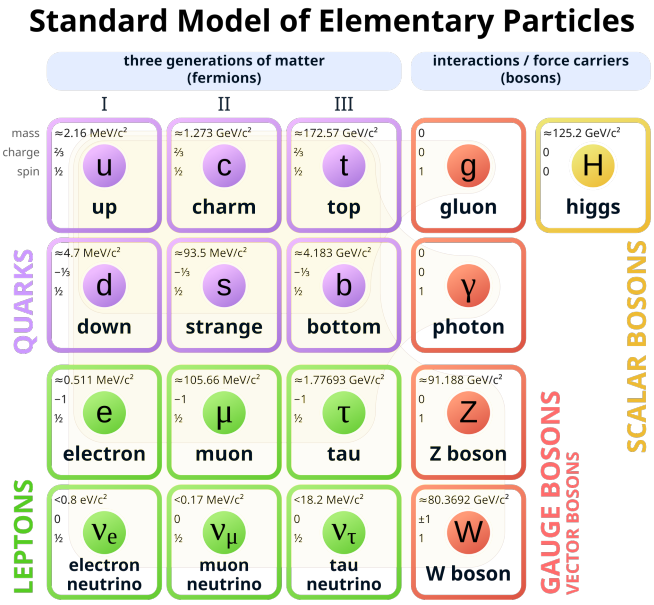

Figura 1 – Standard model of particles.[1]

In general, unlike leptons, quarks cannot be found freely in nature. Due to

confinement, quarks can only be found in the form of mesons and baryons, which can be understood as bound states of a quark–antiquark pair[1] and a bound state of three quarks, respectively. This happens due to the fact that only color-singlet states can exist[2].

Despite its impressive success, there are still many mysteries that the SM cannot explain, for example, the matter–antimatter asymmetries of the universe, the origins of neutrino masses, and the origin of dark matter. This does not imply that the Standard Model is incorrect; rather, it indicates that it is an incomplete theory. A significant effort is devoted to exploring "new physics" beyond the SM.

2.2 CP violation

2.2.1 Discrete transformations

The central concept that permeates all this work is CP violation (CPV). An important step before describing this is to define what a symmetry is and its role in quantum theory, and, as a consequence, to define the C and P operators.

Symmetries in nature are associated with the invariance properties of physical systems. This happens when all the physics remains unchanged under a transformation. These transformations are operators that can act on eigenstates, and a symmetry can be defined if the eigenstate is unchanged under these operations.

With this in mind, we can introduce the charge conjugation (C), parity (P), and time-reversal (T) transformations, which together form one of the most fundamental symmetries in nature: the CPT symmetry.

- **Charge conjugation (C)**: the operation that invert additive quantum numbers, such as charge, lepton number, and strangeness.
- **Parity (P)**: the operation that changes a dextrorotatory coordinate system[3] into a levorotatory[4] coordinate system, it acts by inverting the coordinate system with respect to the origin.
- **Time reversal (T)**: the operation that takes $t \to -t$. In practice, this can be understood as a reversal of the motion.

[1] An antiparticle is a concept originally developed by Dirac and arises for all particles described in the Standard Model. These particles are almost exactly the same as their counterparts, except for the inversion of all internal quantum numbers (charge, parity, etc.).

[2] Recent studies conducted by LHCb [15] have found tetraquark states (bound states of two mesons) and pentaquark states (bound states of a meson and a baryon).

[3] system orientation in which the vector product follows the usual right-hand rule convention.

[4] system in which the vector product changes sign.

In the 1940s, Feynman [16], Tomonaga [17], and Schwinger [18] developed Quantum Eletrodynamics, a theory that conserves C, P, and T. It was natural to assume that weak interactions could also have these same symmetries. However, the ^{60}Co beta-decay experiments conducted by Chien-Shiung Wu in the 1950s [19] provided the first evidence that parity is not conserved in weak interactions. At the same time, Yang, Lee, and Oehme [20] questioned whether parity could truly be regarded as a good symmetry in weak interactions. However, CP violation has implications that extend beyond the matter–antimatter asymmetry. The CPT theorem states that CPT is an exact fundamental symmetry that must be respected by any Lorentz-invariant local quantum field theory with a Hermitian Hamiltonian. Consequently, if CPT symmetry holds, CP violation implies that T must also be violated, which in turn implies that irreversible processes must exist at the fundamental level.

Nonetheless, in the 1960s, CP symmetry violation was measured for the first time in the neutral-kaon experiments by Cronin and Fitch [21]. What was known at the time was that the neutral kaon was observed in two different states, K_S^0 and K_L^0, which were distinguished by their decay widths. The K_S^0 is a CP-even state, while the K_L^0 is a CP-odd state. If CP were a symmetry, one would expect to see K_L^0 decaying into 3π and K_S^0 into 2π. What Cronin and Fitch measured was a small fraction of K_L^0 decaying into 2π. Thus, CP violation was measured for the first time.

2.2.2 CKM matrix

Once it is understood that nature tells us that CP is not a conserved symmetry, it is necessary for the theory to account for and express this phenomenon. Part of this effect is due to the existence of multiple quark generations, as can be seen in Fig. 1, and how they couple to each other. All this work is related to what we know as flavor physics. In other words, flavor physics is the area of particle physics that studies different types of quarks and leptons, focusing on how they transform into one another through weak interactions. The CP violation can be seen arise in the Yukawa sector of the SM Lagrangian, this sector contains the couplings of the Higgs to the fermions. Gauge symmetry implies massless gauge bosons. Therefore, an additional ingredient was required: the Higgs boson and the mechanism of spontaneous symmetry breaking, which give rise to the masses of $W^{\pm}$ and Z^0 bosons. This part of the Lagrangian can be seen in Eq. 2.1.

$$\mathcal{L}_{SM}^{yukawa} = \bar{Q}_L Y^d \phi D_R + \bar{Q}_L Y^u \bar{\phi} U_R + \bar{L}_L Y^e \phi E_R + h.c. \tag{2.1}$$

where $\bar{Q}_L$ represents the left-handed quark doublets, U_R and D_R represent the right-handed quark singlets, ϕ is the Higgs doublet, and $Y^{d,u}$ are the 3×3 complex Yukawa matrices. The first two terms are related to how the down quarks and up quarks interact with the Higgs field, while the last one expresses how the charged leptons interact with the Higgs.

The Yukawa matrices can be diagonalized through a bi-unitary transformation as:

$$(V_L^d)^\dagger Y^D V_L^d = Y_{diag}^d \ ; \quad (V_L^u)^\dagger Y^u V_L^u = Y_{diag}^u \ ; \quad (V_L^e)^\dagger Y^e V_L^e = Y_{diag}^e \tag{2.2}$$

where V_L^d, V_R^d, V_L^u, and V_R^u are 3×3 unitary matrices in the mass eigenstate basis. We are only interested in the quark sector, so we will ignore the leptonic term. The first three terms can be obtained by performing the field transformations that make the Lagrangian invariant under $U(3)$:

$$Q_L \to V_L^u Q_L \quad U_R \to V_R^u U_R \quad D_R \to V_R^d D_R \tag{2.3}$$

Applying these rotations, we obtain the Lagrangian for the quarks as :

$$\mathcal{L}_{SM}^{yukawa} = -\phi \bar{Q}_L (V_L^u)^\dagger V_L^d Y_{diag}^d D_R - \phi \bar{Q}_L Y_{diag}^u U_R + h.c. \tag{2.4}$$

This matrix is the so-called Cabibbo–Kobayashi–Maskawa (CKM) matrix [22]. It was proposed by Kobayashi and Maskawa in the 1973s to explain the origin of CPV. This matrix was based on the Cabibbo mixing matrix, which mixed the down-quark states already known at the time, the d and s quarks. In their work, Kobayashi and Maskawa proposed the existence of a third mass doublet so that the Cabibbo matrix [23], which was 2×2 and real, would become a 3×3 complex matrix with a complex phase responsible for CP violation. Essentially, they not only described the mechanism responsible for CP violation but also predicted the existence of the t and b quarks. Historically, around the same time, the failure to observe flavor-changing neutral currents led Glashow, Iliopoulos, and Maiani to propose the GIM mechanism [24], which predicted the existence of the c quark, completing the second quark doublet. The c quark, however, was only observed in the 1974.

The Cabibbo matrix is described as:

$$\begin{pmatrix} d' \\ s' \end{pmatrix} = \begin{pmatrix} cos(\theta_c) & sin(\theta_c) \\ -sin(\theta_c) & cos(\theta_c) \end{pmatrix} \begin{pmatrix} d \\ s \end{pmatrix}. \tag{2.5}$$

Kobayashi and Maskawa proposed to extend this matrix to become 3×3 to include a third family of quarks with a higher mass hierarchy.

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.6}$$

Since the CKM matrix connects flavor wave functions with mass wave functions, it can be defined as a complex matrix. The unitary nature of the matrix imposes restrictions on its structure, so it is composed of three real parameters and one complex phase, with the latter being responsible for CP violation. These properties give:

$$\sum_{i=1}^{3} |V_{ij}|^2 = 1, \tag{2.7}$$

$$\sum_{i=1}^{3} V_{ji} V_{ki}^* = \sum_{i=1}^{3} V_{ii} V_{ik}^* = 0, j \neq k. \tag{2.8}$$

The elements of this matrix are not predicted by the SM, being provided by experiment. We can write the CKM matrix by the 3 mixing angles and the complex phase as:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} C_{12}C_{13} & C_{12}C_{13} & C_{13}e^{-i\delta} \\ -S_{12}C_{23} - C_{12}S_{23}S_{13}e^{i\delta} & C_{12}C_{23} - S_{12}S_{23}S_{13}e^{i\delta} & S_{23}C_{13} \\ S_{12}S_{23} - C_{12}C_{23}S_{13}e^{i\delta} & -C_{12}S_{23} - S_{12}C_{23}S_{13}e^{i\delta} & C_{23}C_{13} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}. \tag{2.9}$$

where s_{ij} and c_{ij} denote $\sin\theta_{ij}$ and $\cos\theta_{ij}$, respectively. The phase δ carries the CP violation in hadronic decays. The three mixing angles are obtained experimentally and have the values $\theta_{12} = 12.9°$, $\theta_{23} = 2.4°$, and $\theta_{13} = 0.22°$ [25]. These values are very small, which makes the sines very close to zero and the cosines very close to one.

Another way to look for the CKM matrix is expand its terms around $sin\theta_{12} = \lambda$, and three other real parameters (η, ρ and A). One can define :

$$\lambda = s_{12} = \frac{|V_{us}|}{\sqrt{|V_{ud}|^2 + |V_{us}|^2}}, \quad S_{23} = A\lambda^2 = \lambda|\frac{V_{cb}}{V_{us}}|,$$

$$s_{13}e^{i\delta} = V_{ub}^* = A\lambda^3(\rho + i\eta) = \frac{a\lambda^3(\bar{\rho} + i\bar{\eta})\sqrt{1 - A^2\lambda^4}}{\sqrt{1 - \lambda^2[1 - A^2\lambda^4(\rho + \bar{} i\bar{\eta})]}},$$

This expansion is the Wolfenstain parametrization [26], and has the form :

$$V_{ckm} \approx \begin{pmatrix} 1 - \frac{1}{2}\lambda^2 - \frac{1}{8}\lambda^4 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda + \frac{1}{2}A\lambda^5[1 - 2(\rho + i\eta)] & 1 - \frac{\lambda^2}{2} - \frac{1}{8}\lambda^4(1 + 4A^2) & A\lambda^2 \\ A\lambda^3(1 - \rho - i\eta) & -A\lambda^2 + \frac{1}{2}A\lambda^4[1 - 2(\rho + i\eta)] & 1 - \frac{1}{2}A^2\lambda^4 \end{pmatrix} + O(\lambda^6). \tag{2.10}$$

From this expansion, we can see that the complex terms arise at order $\mathcal{O}(\lambda^3)$ in V_{td} and V_{ub}. These terms are related to CP violation in the b-quark sector, but the

corresponding effects in the charm sector only appear at order $\mathcal{O}(\lambda^5)$, which shows that the effect of CP violation in charm is two orders smaller than in beauty.

Depending on the matrix elements in the decay amplitudes, we can classify the decays as:

- Cabibbo-favored: occur within the same quark family, that is, only diagonal terms of the CKM matrix are involved.
- Cabibbo-suppressed: occur in decays that include factors V_{us}, V_{cb}, V_{cd}, and V_{ts}.
- Double Cabibbo-suppressed: occur in decays that include factors V_{ub} and V_{td}.

2.2.3 Weak and Strong phases

Current knowledge about CP violation allows us to understand that the phenomenon can manifest in two different ways: directly and indirectly. Indirect CP violation occurs in processes in which we observe flavor mixing. These effects can arise in two distinct ways: from differences in the oscillation frequencies between particles and antiparticles (which have never been observed), or from the interference between decay amplitudes with and without mixing. Both mechanisms can only be observed in neutral meson systems. In direct CP violation, we observe a difference between the decay widths of particles and antiparticles in their final states, $\Gamma(P \to f) - \Gamma(\bar{P} \to \bar{f}) \neq 0$. In this work, we are interested in studying direct CP violation processes.

To observe differences in the decay widths of particles and antiparticles, it is necessary that two phases exist: one that changes sign under CP and another that remains invariant under this operation. The first, as already discussed, is the weak phase that arises from the CKM matrix in the Lagrangian density. The second is associated with strong processes and is called the strong phase. Our knowledge of the strong phase is limited, but it is believed to arise from second-order processes that occur at short distances or from final-state interactions.

In order to illustrate how the difference between the strong and weak phases is necessary to observe CP violation, we consider a decay process of a state P into a final state f that occurs through two different amplitudes:

$$|A(P \to f)| = |A_1|e^{i\phi_1}e^{i\delta_1} + |A_2|e^{i\phi_2}e^{i\delta_2} \tag{2.11}$$

Here, A_1 and A_2 have different weak phases, ϕ_1 and ϕ_2, and different strong phases, δ_1 and δ_2. We can view the full decay amplitude as a process in which the same final state can be reached through two different quantum paths. For example, a decay may reach the

final state directly, or it may proceed through an intermediate process, such a resonance. The conjugate process can be written as:

$$|A(\bar{P} \to \bar{f})| = |A_1|e^{-i\phi_1}e^{i\delta_1} + |A_2|e^{-i\phi_2}e^{i\delta_2}. \tag{2.12}$$

Direct CPV implies that:

$$|A(P \to f)|^2 - |A(\bar{P} \to \bar{f})|^2 \neq 0, \tag{2.13}$$

So, using Eqs. 2.11 and 2.12, we get:

$$|A(P \to f)|^2 - |A(\bar{P} \to \bar{f})|^2 = 2|A_1||A_2|sin(\phi_1 - \phi_2)sin(\delta_1 - \delta_2). \tag{2.14}$$

Finally, we can define the CP asymmetry as:

$$A_{cp} = \frac{\Gamma(P \to f) - \Gamma(\bar{P} \to \bar{f})}{\Gamma(P \to f) + \Gamma(\bar{P} \to \bar{f})} = \frac{|A(P \to f)|^2 - |A(\bar{P} \to \bar{f})|^2}{|A(P \to f)|^2 + |A(\bar{P} \to \bar{f})|^2} \tag{2.15}$$

Therefore, by substituting the amplitudes and simplifying the expression, we obtain:

$$A_{cp} = \frac{2|A_1A_2|sin(\delta_2 - \delta_1)sin(\phi_1 - \phi_2)}{|A_1|^2 + |A_2|^2 + 2|A_1A_2|cos(\delta_2 - \delta_1)cos(\phi_1 - \phi_2)}. \tag{2.16}$$

Equation 2.16 shows, through a simple example, that CP violation can only be observed if a phase difference between the weak and strong phases exists in the decay processes of particles and antiparticles. Specifically, in three-body decays, these differences can arise, for example, from resonances (intermediate decay states).

2.3 CP violation on Charm

In the c-quark sector, as can be seen from the CKM matrix, we have the presence of a weak phase in processes of the order $\mathcal{O}(5)$ on $(V_{cb}V_{ub}^*)$ in the Wolfenstein parametrization. With this, it is interesting to discuss how CPV manifest in this sector. Charm sector provides the only opportunity to study CP violation in the up-type quark sector, which provides complementary clues for the b and s quark sector. Various authors propose that CPV in the c quark sector can be of the order of 10^{-3} or less [27, 28, 29, 30]. Results significantly larger than this expectation could be associated with effects from physics beyond the Standard Model.

In the charm sector, direct CP violation can only occur in Cabibbo-suppressed decays, where the final states can be reached through amplitudes carrying different weak phases.

There is only one observation of CP violation in the c-quark sector, measured in neutral mesons in 2019 by the LHCb collaboration using Run 2 data. The measurement was obtained from the difference between the A_{CP} values of $D^0 \to \pi^-\pi^+$ and $D^0 \to K^-K^+$, yielding $\Delta A_{CP} = -1.64 \pm 0.28 \times 10^{-3}$. More recent studies of these channels indicate that the $\pi\pi$ channel is more relevant for the result and suggest that CP violation in $\pi\pi$ may be larger than in KK [31]. A plausible interpretation is that final-state interactions play an important role in these decays. In particular, strong rescattering processes can transform a KK final state into a $\pi\pi$ state. As a consequence, CP-violating effects may be effectively transferred from the KK channel to the $\pi\pi$ channel, enhancing the asymmetry observed in $\pi\pi$ decays while suppressing it in KK decays.

Now, considering charged charm mesons, in our case the D_s decaying to $\pi\pi K$. This is a Cabibbo-suppressed decay; in other words, it is a channel in which we expect a weak phase from V_{cd} and, with a strong phase arising from final-state interactions, we expect to observe CP violation in this channel within the SM. This channel also has characteristics that make it a very promising mode for CP-violation studies. It has the highest branching fraction among three-body Cabibbo-supressed D_s meson decays, $(6.2 \pm 0.19) \times 10^{-3}$ [32]. This makes this channel, among channels suppressed by cabibbo of the Ds meson in three bodies, the one with the most statistically significant data for analysis. This might allow for sufficient sensitivity to observe CP violations in this channel.

This channel can reach its final state through different topologies, as shown in Fig. 2. Here, we expect the tree diagram to be the dominant topology, which can interfere with itself through resonances, such as the f_0 (which can be generated through rescattering effects.) and K^{*0} in the $\pi\pi$ system, providing a weak phase through V_{cd}. We can also expect interference arising from the penguin topology, which contains a weak phase associated with V_{ub}.

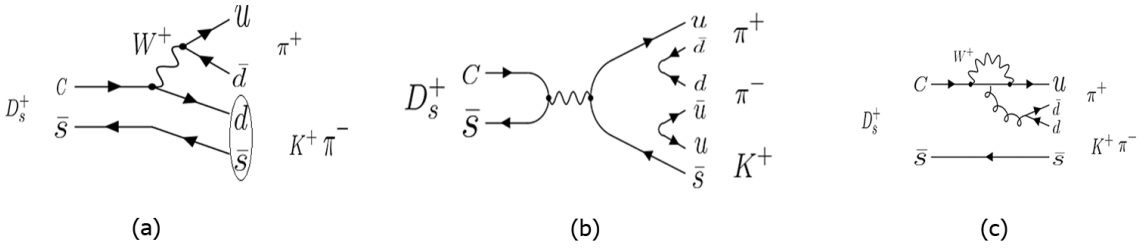

Figura 2 – Decay topologies: (a) Tree, (b) Annihilation e (c) Penguin.

Another interesting feature of this channel is that it is a final state with two opposite-charge pions. This means we may see in this channel a similar effect to what is seen in D^0: the $\pi\pi$ can produce greater CP asymmetries than the conjugate final state with two opposite-charge kaons like $D_s \to KKK$. The search for CP violation in the $D_s \to KKK$ decay has already been published [33]. Using a data sample 30 times smaller than that employed in this work, with about 1M signal events, no evidence of CP violation was observed in this channel. [33].

2.4 Thee-body decays formalism

Three-body decays carry the inherent difficulties of the analysis of many-body dynamics, where the difficulty is proportional to the number of bodies involved in the process. To study systems like these, it is necessary to adopt more sophisticated tools, and in the case of three-body decays, an approach can be conducted by analyzing the phase space.

In two-body decays, as a consequence of the energy-momentum conservation, the phase space of this type of process reduces to a single point, with only one possible kinematic configuration. In three-body decays, on the other hand, the momentum space has nine degrees of freedom in principle. However, energy and momentum conservation impose four constraints, and for three-body decays of spinless particles into three spinless particles, isotropy introduces three additional constraints. This makes it possible to describe the momentum space of three-body decays as a two-dimensional distribution [34].

A very useful description of the phase space for the studies conducted in this work is the *Dalitz plot*. In this representation, the phase space is projected in terms of a pair of the squared invariant masses of any particle pair, s_{12}, s_{13}, and s_{23}. These can be defined as:

$$s_{ij} = (P - p_k)^2 = m_{ij}^2 \tag{2.17}$$

with i, j, and k defined in $(1, 2, 3)$, and all indices being different. P total four-momentum of the initial system. These variables can be related through the decay products as:

$$s_{12} + s_{23} + s_{13} = m_1^2 + m_2^2 + m_3^2 \tag{2.18}$$

In this way, the Dalitz plot defines a region of phase space that contains all the information about the possible kinematic configurations for three-body decays, setting the limits of the phase space when projected onto:

$$\begin{aligned}(m_1+m_2)^2 \leqslant s_{12} \leqslant (M-m_3)^2,\\ (m_2+m_3)^2 \leqslant s_{23} \leqslant (M-m_1)^2,\\ (m_1+m_3)^2 \leqslant s_{13} \leqslant (M-m_2)^2.\end{aligned} \tag{2.19}$$

with M been the mass of the parent (decaying) particle.

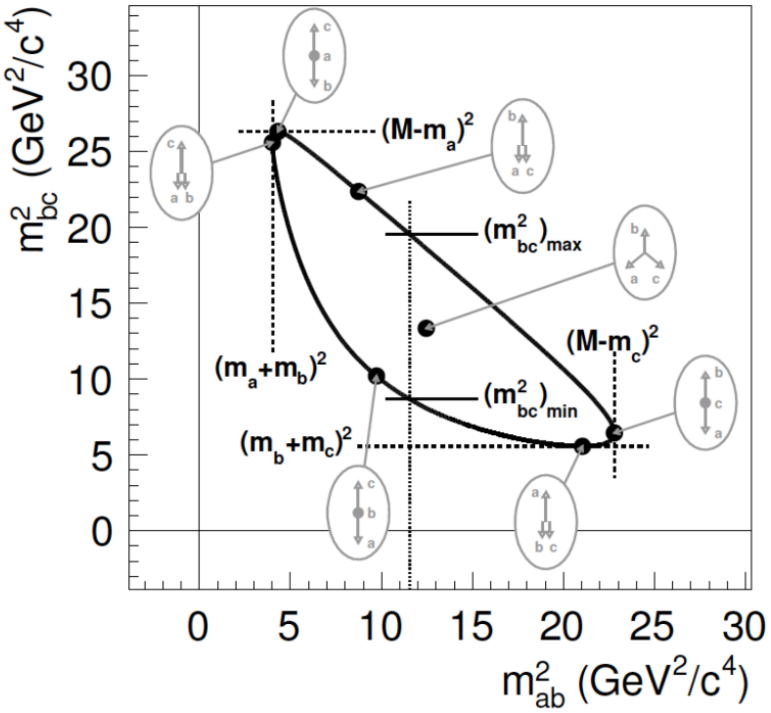

Figura 3 – Dalitz plot For $D \to p_a p_b p_c$.

Each point in the Dalitz plot represents a different momentum configuration, as shown in Fig. 3. By choosing the invariants s_{12} and s_{13}, it is possible to describe the Dalitz plot as:

$$\begin{aligned}s_{13}^{\pm} = m_1^2 + m_j^3 - \frac{1}{2s_{12}}[(s_{12}-m^2-m_3^2)(s_{12}-m_2^2-m_1^2)\\ \mp \lambda^{\frac{1}{2}}(m^2,s_{12},m_3^2)\lambda^{\frac{1}{2}}(s_{12},m_2^2,m_1^2)]\end{aligned} \tag{2.20}$$

where $s_{13}^{\pm}$ are functions that delimit the Dalitz plot, and $\lambda(x,y,z) = (x-y-z)^2 - 4yz$.

The decay rate transition in a three-body decay $D \to h_1 h_2 h_3$ can be written as:

$$d\Gamma = \frac{1}{(2\pi)^3}\frac{1}{32M^3}|\mathcal{A}|^2 ds_{12} ds_{13} \tag{2.21}$$

where $|\mathcal{A}|^2$ is the magnitude of the decay amplitude.

There are many models that can be used to describe the Dalitz plot structure, and perhaps one of the simplest approaches is the isobar model. In this model, we assume that the total decay amplitude is given by the sum of a non-resonant amplitude and all resonant contributions to the decay:

$$\mathcal{A}(s_{ij}, s_{ik}) = a_{no-res}e^{i\delta_{no-res}} + \sum_n a_n e^{i\delta_n}\mathcal{A}_n \tag{2.22}$$

where $a_{no-res}e^{i\delta_{no-res}}$ is a non-resonant amplitude, the contribution of each resonance is represented by a $\mathcal{A}_n$ weighted by a complex number of magnitude a_n and phase δ_n.

This model works with a two-body resonance plus one meson approximation. We assume that two of the three particles composing the final state can be treated as a system independent of the third particle, the resonant state, while the third acts as a spectator in the process. This model has limitations, being more suitable for narrow resonances in situations where there are no overlapping resonances with the same spin.

Assuming that the decay can be written as $P \to R, d_k$, where R is a resonant state that decays into $d_i d_j$, we can write the individual resonant amplitude as:

$$\mathcal{A}_n^R = N \times T(p_j, q_k) \times F_P(p_j, r_{BW}^P) \times F_R(p_k, r_{BW}^R) \times \mathcal{R}(m_{ij}, m_0, L) \tag{2.23}$$

Here, N denotes a normalization factor; F_P and F_R are the Blatt–Weisskopf form factors, used to model centrifugal barrier effects associated with orbital angular momentum, and r^PBW and r^RBW are the meson radii of the decaying particle and the resonance, respectively. The term $T(p_j, q_k)$ carries the angular probability distribution, which reflects the conservation of angular momentum, and can be computed using the Zemach tensor formalism for spinless final-state particles, or the helicity formalism in other cases. Finally, $\mathcal{R}(m_{ij}, m_0, L)$ is the propagator of the dynamical function describing the resonance mass term, where m_0 is the resonance mass and L is the orbital angular momentum between the resonance decay products. A commonly adopted function to describe this term is the relativistic Breit–Wigner function:

$$\mathcal{R} = \frac{1}{(m_0^2 - m_{ij}^2) - im_0\Gamma(m_{ij}, m_0, L)} \tag{2.24}$$

where Γ is the decay width,

$$\Gamma = \Gamma_0 \left(\frac{q}{q_0}\right)^{2L+1} \times \left(\frac{m_0}{m}\right) \times F_R(p_j, r_{BW}^R) \tag{2.25}$$

with q being the momentum of the two decay products in the resonance center-of-mass frame at mass m, and q_0 the momentum at the resonance mass peak m_0.

An interesting conclusion that can be inferred from this model, and which can be observed in the Dalitz plot, is that the angular factor produces different shapes for different

spin values, allowing the spin of the resonances to be identified by the number of nodes present in the distribution. Scalar resonances have spin 0, resulting in an accumulation of events in the form of a band around the resonance mass in the invariant mass of its decay products. Vector resonances (spin 1) produce a similar pattern but with the presence of a minimum in the distribution, while tensor resonances (spin 2) exhibit two minima throughout the distribution. Figure 4 provides a clear example of how these resonances appear in the Dalitz plot. Phase differences between resonances can also become evident, as illustrated in Fig. 4, both in scalar cases with no phase difference and when a phase difference of π is introduced between the resonances.

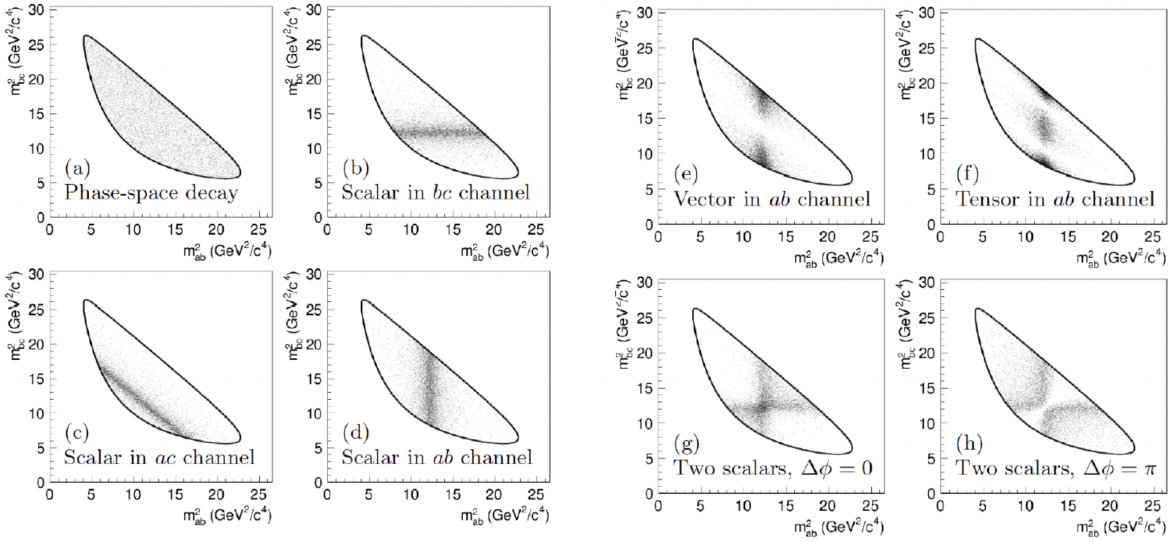

Figura 4 – Forms and examples of manifestation of resonances in a generic decay $M \to ABC$: (a) decay without dynamics (non-resonant); (b), (c) and (d): production of a spin 0 (scalar) resonance in channels bc, ac and ab, respectively; (e) production of a spin 1 resonance in channel ab; (f) production of a spin 2 resonance in channel ab; (g) interference between two scalar resonances without phase difference; (h) interference between two scalar resonances with a phase difference of $\Delta\phi = \pi$.

An interesting aspect of this approach is that it allows the study of local asymmetries that can arise in the Dalitz plot. These asymmetries may occur due to interference between resonances present in this phase space. If these asymmetries change sign across the Dalitz plot, they can be larger than the integrated asymmetry. Effects of this type have already been widely documented in studies of CP violation in B mesons, where local asymmetries larger than 75% have been observed [35, 36, 37].

The Dalitz plot for the $D_s^+ \to \pi^-\pi^+K^+$ for the signal region is shown in Fig. 5. In this figure, it is possible to observe resonant structures in the phase space. Some of these

can be attributed to the $\rho^0(770)$ and $f_0(980)$ resonances in $s_{12}(\pi\pi)$, and to the $K^*(892)$, $K^*(1410)$, and $K_0^*(1430)$ resonances in $s13(\pi K)$ [38].

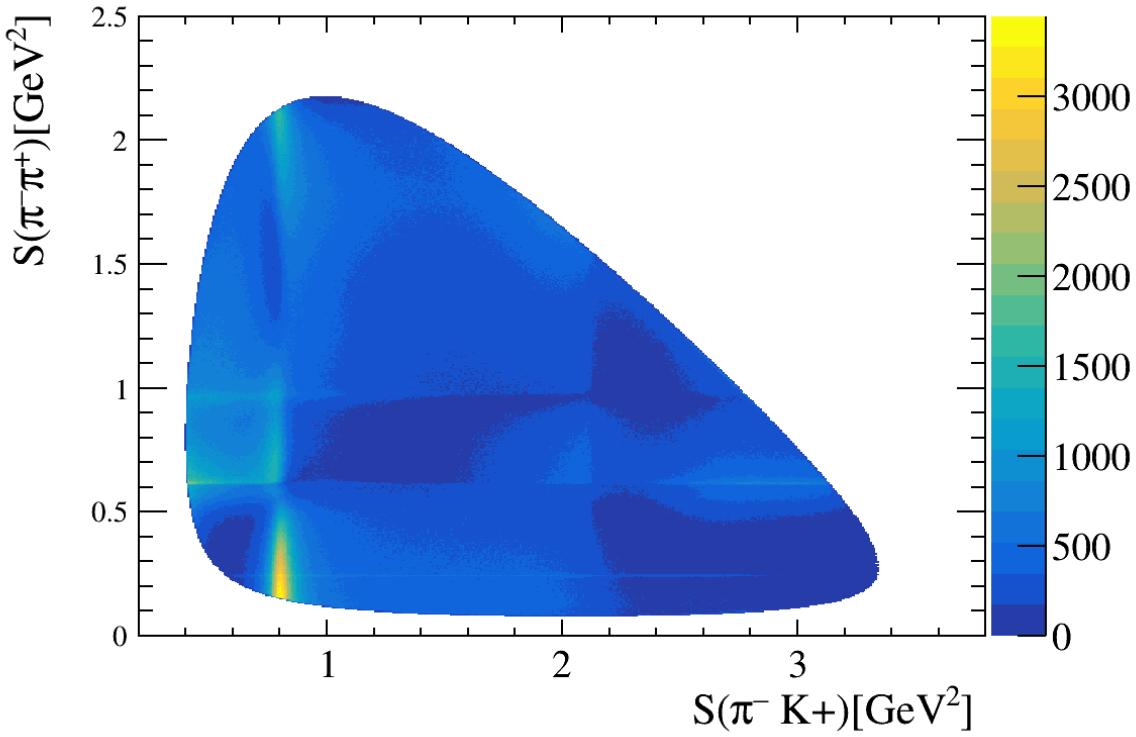

Figura 5 – Dalitz plot for $D_s^+ \to \pi^-\pi^+K^+$ after all the selection process.

Other resonances may be present in this Dalitz plot, suppressed relative to the magnitudes of those already described. Only a full amplitude analysis of the resonant structure can confirm which resonances are present in this decay.

3 The LHCb Experiment

3.1 LHC - the Large Hadron Collider

The LHC (Large Hadron Collider) is the biggest circular particle accelerator currently in operation in the world, with a length of 27 km long and is located close to Geneva on the border between France and Switzerland. Its development, and its detectors were designed to provide fundamental particle physics studies and their interactions. Today, it aims to provide validation studies of the SM and search for physics beyond it.

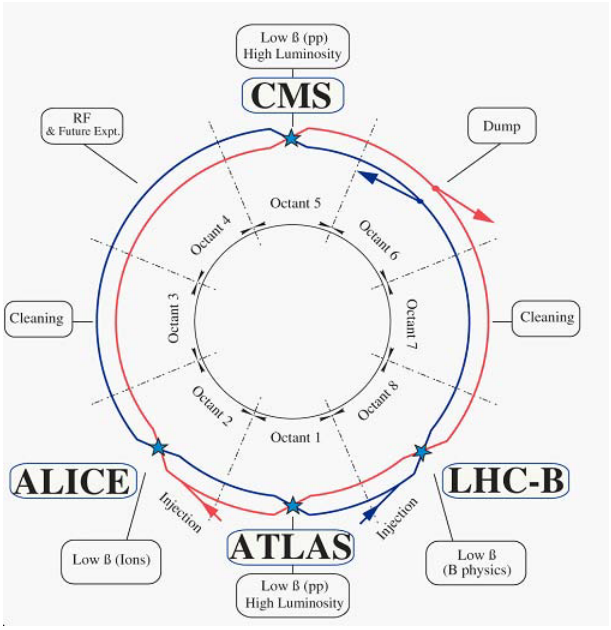

Figura 6 – LHC scheme

The accelerator consists of two beams of high-energy protons traveling at speeds close to the speed of light, with each beam traveling in opposite directions in separate tubes under ultra-vacuum conditions, reaching energies of up to 7 TeV, on the figure 6 this scheme is shown. resulting in a center-of-mass energy of 14 TeV. These beams are collimated by an impressive set of superconducting electromagnets that work in temperatures that reach 1.9 K. To maintain the operational conditions the whole system is connected to a robust liquid helium refrigerator system that operates to meet the demands of all 1232 dipoles magnets that are 15 meters long that guide the beam along the tube and the 292 quadrupole magnets that are 5-7 meters long that are used to focus the beams across the beam pipes and on the 4 interact points where we have the experiments installed, these experiments are:

- **LHCb**: Experiment originally designed to study hadron decays on the b quark sector, including searches for CP violation and studies of rare decays[3].
- **ATLAS**: Is the biggest CERN experiment, and is designed to general purposes. Its main objective initially was the search for the Higgs boson[39]. Today the experiment is still working to provide precision SM measurements and to search for new physics[40].
- **CMS**: Like ATLAS, CMS is a general-purpose experiment[41], and is one of those responsible for the discovery the Higgs boson[42]. It is still working to search for new physics and to provide precision SM measurements.
- **ALICE**: experiment designed to study Pb-Pb and Pb-p collisions with the objective of studying QCD in extreme conditions[43].

For the LHC to reach this impressive milestone, the accelerator takes advantage of all the available infrastructure in the CERN accelerator complex, using accelerators built in the past to pre-accelerate the proton beams. Initially, these protons are obtained from hydrogen gas nuclei using an electromagnetic field to remove them from the electrons. They are accelerated using a radio frequency quadrupole before injection into LINAC2 where they reach the energy of 50 MeV. Then, they proceed to the Booster[1], where they are accelerated to reach the energy of 1.4 GeV. They then follow on to the PS (Proton Synchrotron) to reach the energy of 25 GeV. After that, they are then sent to the SPS (Super Proton Synchrotron) where they reach an energy of 450 GeV before being sent to the LHC to reach the expected 7 TeVs of energy. All these systems are shown in Figure 7. This figure, and more details of the LHC operation, can be found in [44, 45].

About this impressive structure, the LHC aims to maximize the center-of-mass energy and the collision rate. The collision rate is measured in terms of luminosity, defined as:

$$L = \frac{N^2 f}{4\pi\sigma_{eff}} \tag{3.1}$$

where N is the number of protons in each bunch, f is the bunch crossing rate, and σ_{eff} is the effective RMS radius of the beams. During the Run 2[2], the period in which these data were collected, a luminosity of $L = 10^{34} cm^{-2} s^{-1}$ was obtained. For comparison, Fermilab's TEVATRON [46], which until 2011 was the most powerful accelerator in the world, achieved an energy close to 1 TeV and a luminosity of $L = 10^{29} cm^{-2} s^{-1}$. The LHCb already operates with a lower luminosity, around $L = 10^{32} cm^{-2} s^{-1}$, since the

[1] Or PSB (Proton Synchrotron Booster)
[2] The data collection period from 2015 to 2018.

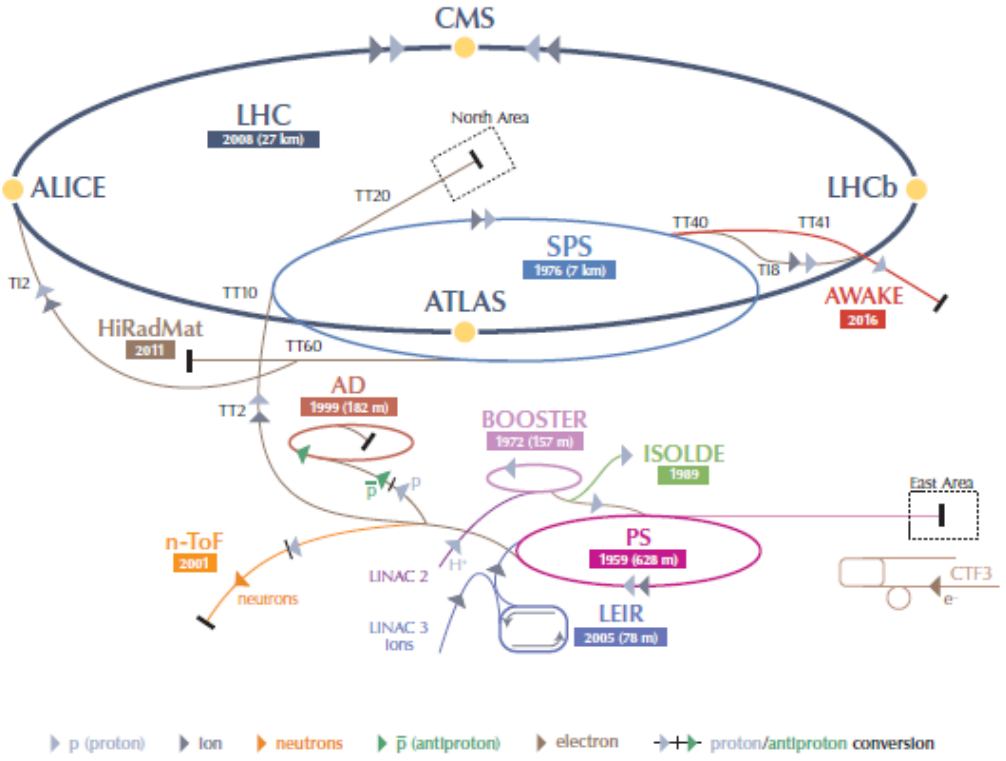

Figura 7 – The LHC accelerator complex.

experiment works with a precise reconstruction of the primary and secondary vertices. This is important to prevent wear and tear on the apparatus and its premature aging.

3.2 LHCb experiment

The LHCb is an experiment initially designed for CPV studies involving the b quark and rare decays. The experiment today has a broad physics program encompassing rare decays, charm physics, and exotic states. To conduct these studies, the LHCb is a single-arm spectrometer with an angular opening of 10-300 mrad in the vertical axis and 10-250 mrad in the horizontal axis. This configuration is adopted because it is expected that the main production of the $b\bar{b}$ pairs is primarily produced at angles close to the beam, as we can see from the simulated pseudorapidity distribution of $b\bar{b}$ produced in LHC collisions at 14 TeV (Figure 8).

$$\eta = -log\left(tan\frac{\theta}{2}\right) \tag{3.2}$$

where θ is the angular aperture of the particles produced in relation to the beam. The LHCb operates with values of pseudorapidity between $2 < \eta < 5$.

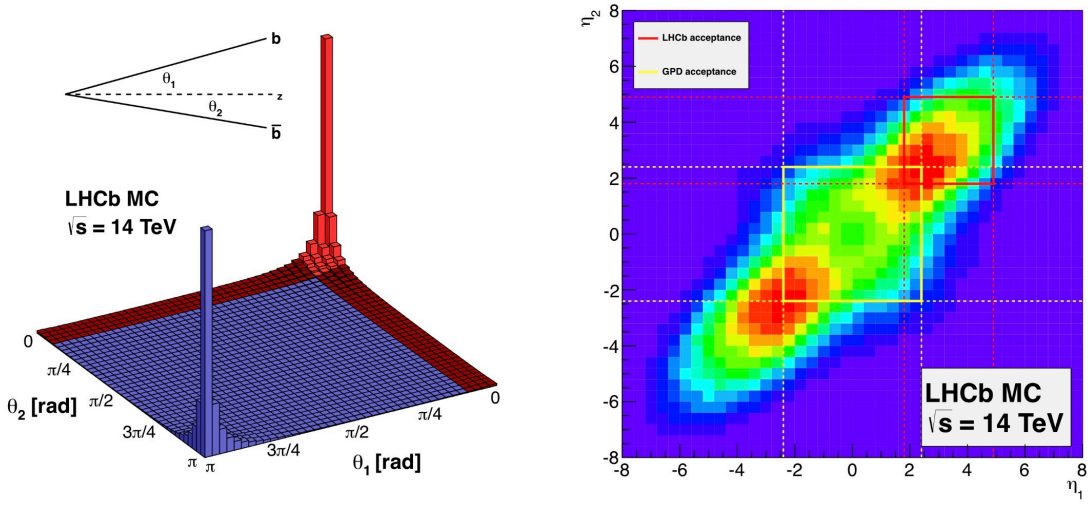

Figura 8 – The angular distribution of $b\bar{b}$ production of the angle to the beam axis (left) and as a function of the pseudorapidity (right) at $\sqrt{s} = 14$TeV centre-of-mass energy.[2]

The LHCb is primarily designed for the study of to heavy flavor hadrons, whose decay processes occur through weak interactions. What is largely measured are the ground states of hadrons containing b and c quarks (for example, $B^{\pm}$, B_s, D^0, $D^{\pm}$ and $D_s^{\pm}$) which tend to have long lifetimes. Therefore, heavy flavor hadrons produced at the LHC will be highly boosted in the laboratory frame and will travel several millimeters before decaying. Basically, the signatures of these processes are related to a set of charged tracks meeting at a vertex displaced from the proton-proton collision point.

The precise determination of the decay vertex and of the trajectories of the particles produced in the collisions is crucial for the accurate reconstruction and identification of the processes of interest. The LHCb is an experiment designed to cover all these points, in order to produce a large amount of data to be analyzed.

On the figure 9 we have the composition of the LHCb, as well his subsystem.

3.2.1 track system

One of biggests highlights of the LHCb is its tracking system. During the Run 2 were composed by the Vertex Locator (VELO), the Tracking stations and the magnet. Everything working together to identify with precision the position of the decay vertex, the trajectory of each decay product and its momentum.

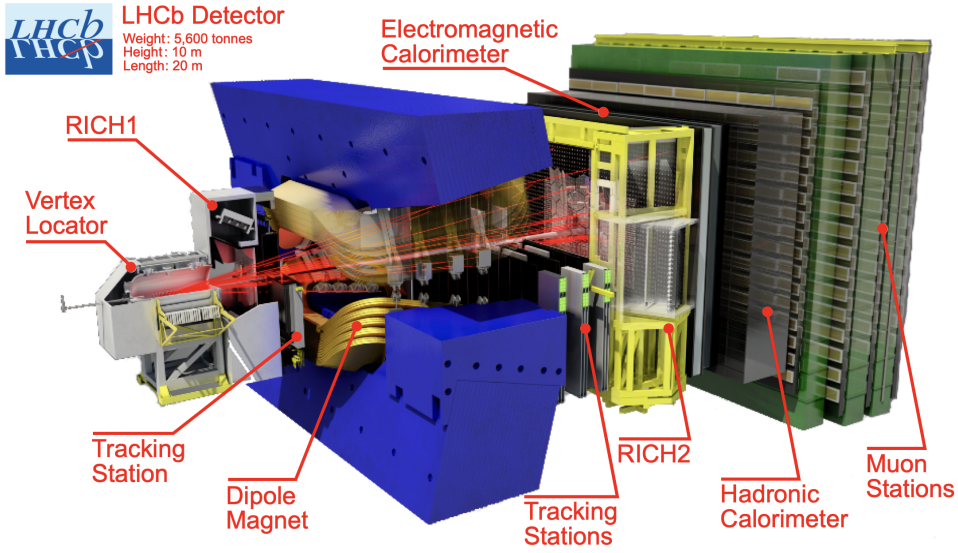

Figura 9 – LHCb 3D layout.

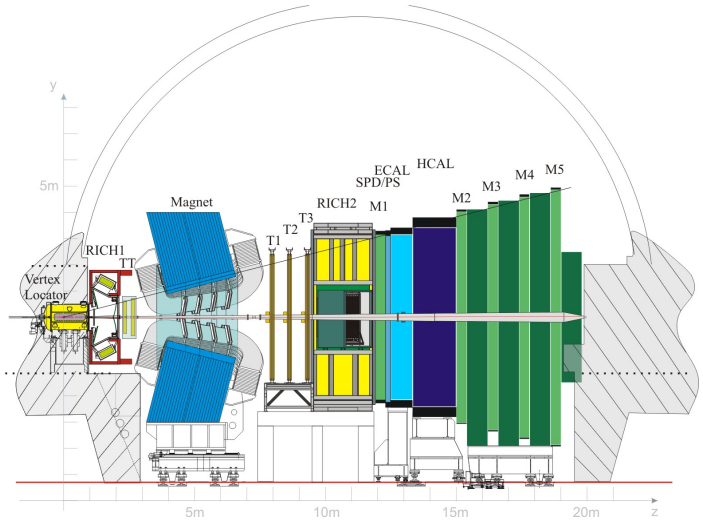

Figura 10 – The LHCb detector with side sectional view.

3.2.1.1 Velo

The VELO [47] is composed of 42 modules with a semi-circular geometry. These modules are arranged to form a cylindrical structure, such that the beam passes along the direction of the modules, thereby defining the z-axis of the experiment. The modules are made of silicon, which is distributed along their radial and azimuthal coverage. This design allows the VELO to achieve a resolution of about 40 μm and 10 μm for the primary vertex, corresponding to the proton–proton interaction point, and about 150 μm and 300 μm for the secondary vertex, corresponding to the decay point of particles containing c

and b quarks.

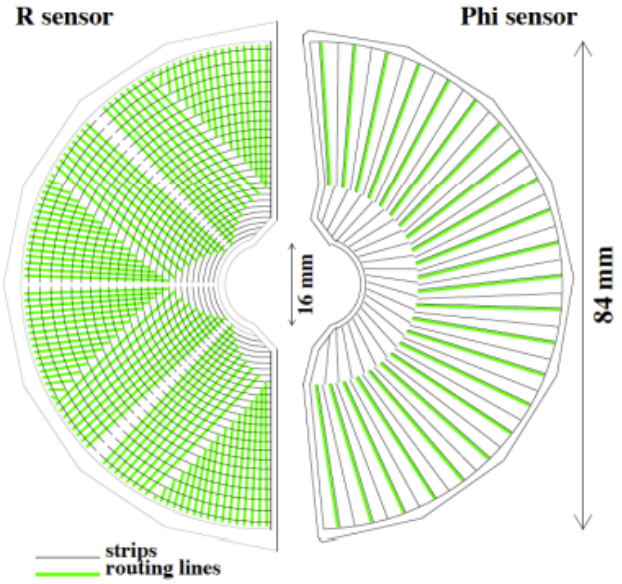

Figura 11 – Schematic view of VELO r and ϕ strips

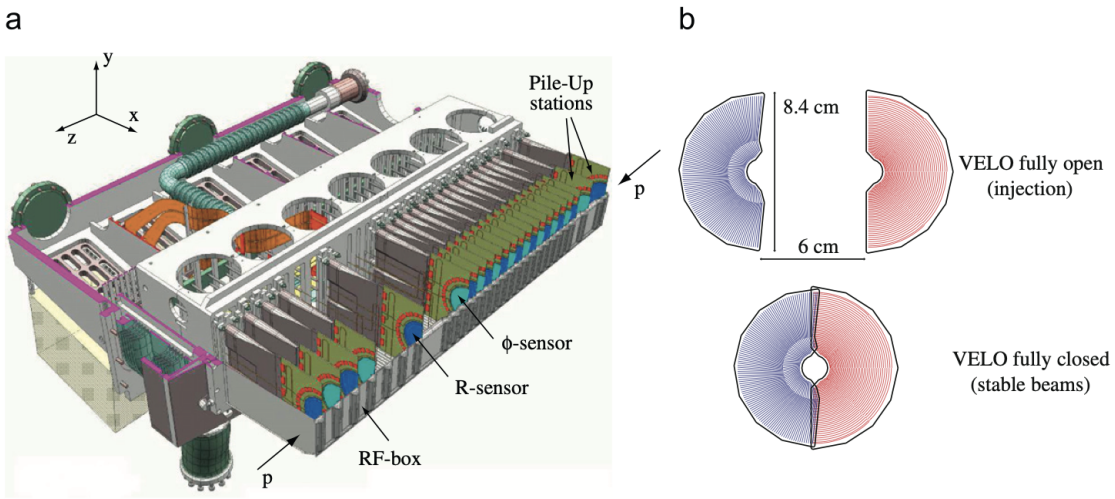

Figura 12 – Schematics of one Velo Half (a) and Position of the sensor with respect to the beam during the injection and physics running conditions (B) .

As a subdetector of LHCb, the VELO must be placed as close as possible to the collision point in order to minimize uncertainties. The sensitive area of the detector starts at a distance of 8 mm from the beam axis. To preserve the VELO's operational lifetime, it has two operating configurations. The so-called VELO open configuration is used during beam injection and acceleration; in this mode, the detector halves are retracted to avoid

exposure to the beam. In the VELO closed configuration, the detector is positioned at its minimum distance of approach when the beams reach stable-beam conditions, as illustrated in figure 12.

The sub-detector provides precise measurements of the track impact parameter, defined as the distance of closest approach to the primary vertex. Long-lived particles tend to have large impact parameters, which allows the use of minimum cuts on the impact parameter itself and on the corresponding χ^2_{IP}[3] to suppress background from short-lived particles.

3.2.1.2 Track Stations

The tracking stations are subdetectors that operate in conjunction with the VELO to reconstruct the trajectories of charged particles. During Run 2, they were composed of the Tracker Turicensis (TT) [48] (depicted in the figure 13) station, located upstream of the magnet, and the T1–T3 stations, positioned downstream of the magnet. The T1–T3 stations are divided into two regions: the Inner Tracker [49], located close to the beam pipe, and the Outer Tracker [50], which covers the region farther away. This division exists because the detectors in each region employ different technologies in their construction: the Inner Tracker, as well as the TT, is built from silicon detectors, while the Outer Tracker detectors are based on straw-tube technology.

The TT provides information for low-momentum particles, with momenta of the order of 2 GeV, which are bent out of the spectrometer acceptance, as well as for particles that do not leave a signal in the VELO. It is composed of four layers, with the two inner layers rotated by -5° and +5° with respect to the outer layers. Each of these layers provides a spatial resolution of about 200 μm. This allows the TT to improve the momentum resolution of tracks reconstructed using the VELO and the downstream tracking stations.

The Inner Tracker is composed of detector boxes arranged in a cross-shaped geometry around the beam axis, covering an effective area of about 4 m^2 and achieving a single-hit resolution of approximately 50 μm. The Outer Tracker covers the largest area of the T stations and consists of straw tubes filled with a gas mixture of 70% Ar, 28.5% CO_2, and 1.5% O_2. The full tracking system achieves an excellent momentum resolution, ranging from about 0.5% at low momentum to about 1% at 200 GeV.

3.2.1.3 Magnet

Finally, the LHCb magnet [51] is used to bend the trajectories of charged particles, allowing their momenta to be determined in combination with the tracking detectors. The magnet consists of a saddle-shaped coils with a window-frame yoke with sloping poles and generates a magnetic field of 4 Tm strength along 10 m in the y direction.

[3] more details about the variables adoted on LHCb can be founded on the chapter 4

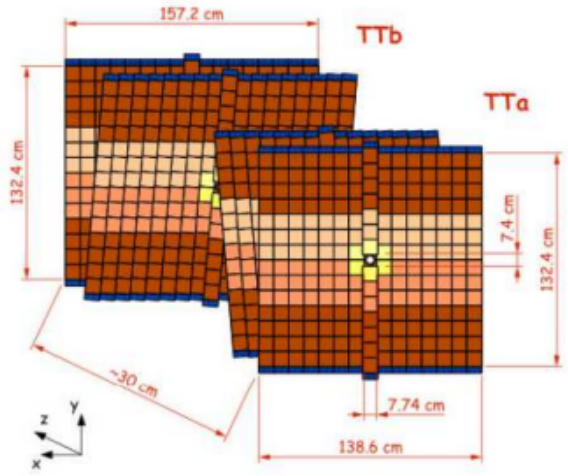

Figura 13 – Illustrative image of Tracker Turicensis.

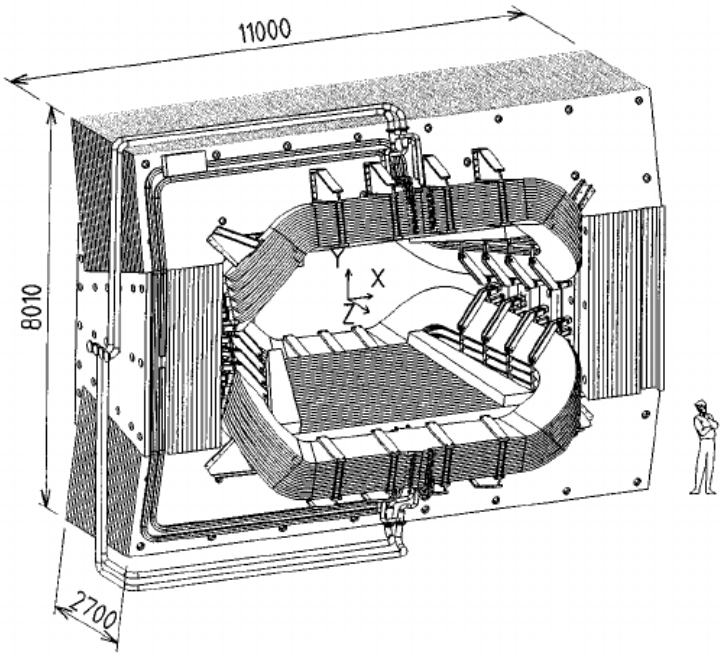

Figura 14 – Illustrative image of LHCb magnet.

Occasionally, the deflection of charged particles can lead to systematic effects. To mitigate this, the polarity of the LHCb magnet is periodically reversed in order to minimize charge-asymmetry effects when samples from the two configurations are combined. These operating conditions are referred to as MagUp and MagDown, according to the direction of the magnetic field.

3.2.2 Particle Identification System

3.2.2.1 Rich

The RICH detector [52] is the main subsystem of LHCb used for particle identification. It is arranged in two sections: one located downstream of the VELO and the other downstream of the outer tracking system, referred to as RICH1 (Figure 15) and RICH2 (Figure 16), respectively. The function of the RICH detectors is to distinguish among kaons, pions, and protons in the final state. When a charged particle enters the detector with a velocity greater than the speed of light in the medium, Cherenkov radiation is emitted in the form of a cone. From this cone, the particle velocity can be measured. Knowing both the velocity and the momentum of the particle, its mass can be estimated, allowing particle identification.

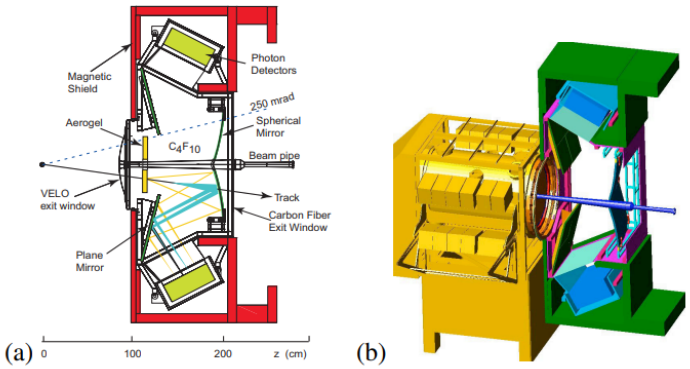

Figura 15 – Illustrative image of the RICH1 detector. (a) Side view of the detector; (b) a 3D model [3].

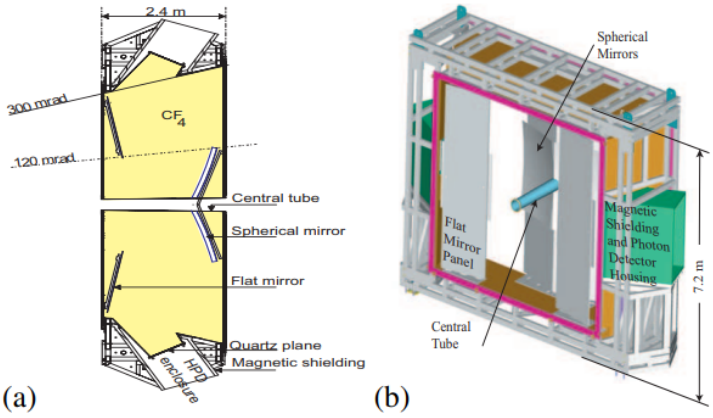

Figura 16 – Illustrative image of the RICH2 detector. (a) Side view of the detector; (b) a 3D model [3].

RICH1 is designed to detect low-momentum particles, covering a momentum range of approximately 1–60 GeV, and is composed of an aerogel radiator and a C_4F_{10} gas chamber, as illustrated in Figure 15. In contrast, RICH2 is optimized for high-momentum particles in the range of approximately 15–100 GeV and consists solely of a CF_4 gas chamber (Figure 16).

3.2.2.2 Calorimeters

In LHCb, the calorimeters[53] are used to identify photons, electrons, and hadrons, as well as to measure their positions and energies. In addition, during Run 2 this part of the LHCb detector provided transverse energy values for photon, electron, and hadron candidates to the first-level trigger, L0, which makes a decision approximately 4 μs after the interaction. During Run 2, the calorimeter system was composed of four components: the Scintillating Pad Detector (SPD), the Pre-Shower detector (PS), the Electromagnetic Calorimeter (ECAL), and the Hadronic Calorimeter (HCAL), as illustrated in Figure 17.

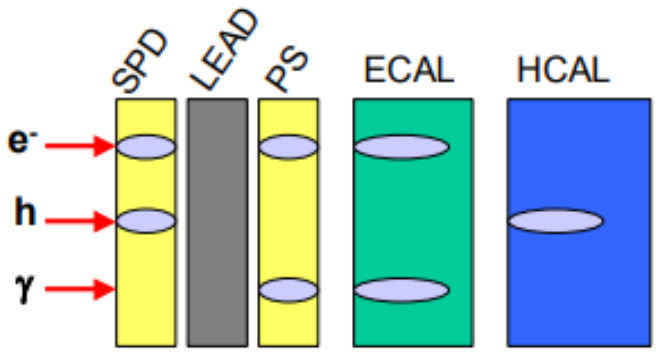

Figura 17 – Illustrative image of the Calorimeters system of the LHCb. showing which particles leave a signal in each detector.[4]

During data taking, when an event reaches the SPD, which consists of a 12 mm thick lead wall, it provides measurements of the event multiplicity. This information can be used to reject or accept events with low or high multiplicity. The PS and the ECAL measure the energy of photons and electrons, while the HCAL measures the energy of hadrons. The ECAL and the HCAL are constructed from alternating layers of absorber material, made of lead or iron, and scintillator plates.

These detectors are designed to absorb particles passing through them and to measure the deposited energy in the process. During the energy loss, particle showers are produced, consisting of secondary particles which, when traversing a scintillating medium, generate light that is transmitted to photomultiplier tubes (PMTs), converting the optical signal into an electrical one. An example of the structure of this detector can be seen in Figure 18.

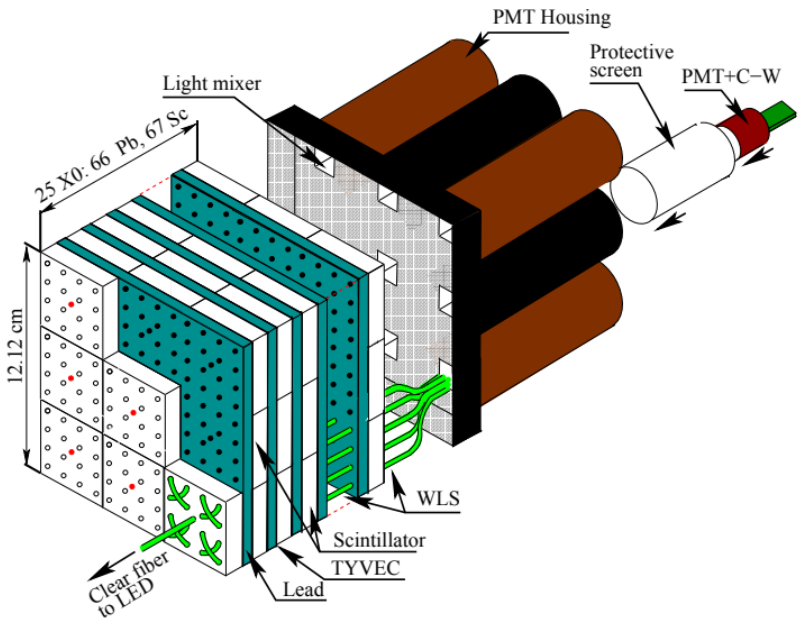

Figura 18 – ECAL cell scheme.[5]

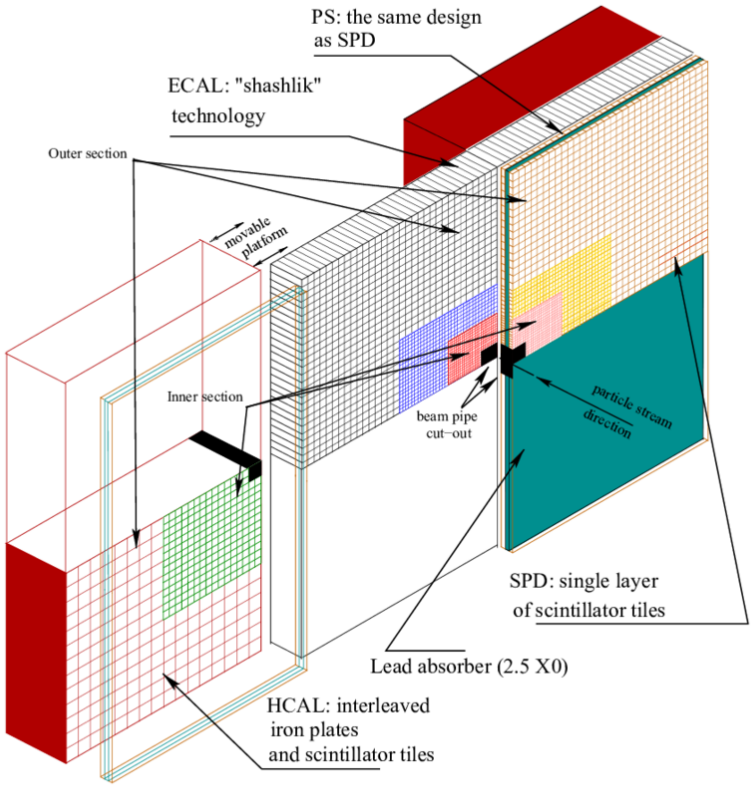

Figura 19 – Layout of the calorimeter system.[5]

The calorimeters are segmented into cells arranged around the beam pipe. The ECAL contains 6016 cells distributed over three regions, distinguished by the cell size, while the HCAL consists of 1488 cells distributed over two cell-size regions, as depicted in the figure 19. The corresponding energy resolutions are: $\sigma_E/E = 10\%/\sqrt{E} \bigoplus 1\%$ for the ECAL and $\sigma_E/E = 69\%/\sqrt{E} \bigoplus 9\%$ in HCAL, with E in GeV.

3.2.2.3 Muon systems

The final sub-detector of LHCb is the muon system[54], which consists of five stations located downstream of the calorimeter system, with the exception of the first station, which is positioned upstream. A high accuracy in muon identification is required, since muons appear as final states in many decays of interest to LHCb. In addition, muons are used for flavor tagging of neutral B and D mesons in measurements of CP asymmetries and oscillations.

The five stations, labeled M1–M5, are based on similar technologies and are primarily composed of Multi-Wire Proportional Chambers (MWPCs). The exception is station M1, which is equipped with a triple-GEM (Gas Electron Multiplier) detector, adopted to prevent detector aging due to the higher radiation exposure in that region.

The MWPCs consist of four layers with a common 5 mm gas gap, containing thin gold-plated tungsten wires with a diameter of about 30 µm and a spacing of 2 mm. They are filled with a gas mixture of 40% Ar, 55% CO_2 and 5%CF_4. The triple-GEM detectors are composed of three gas electron multiplier foils sandwiched between anode and cathode planes, filled with the same gases but in different proportions, 45% Ar, 15% CO_2 and 40%CF_4. The muon system is depicted on the figure 20.

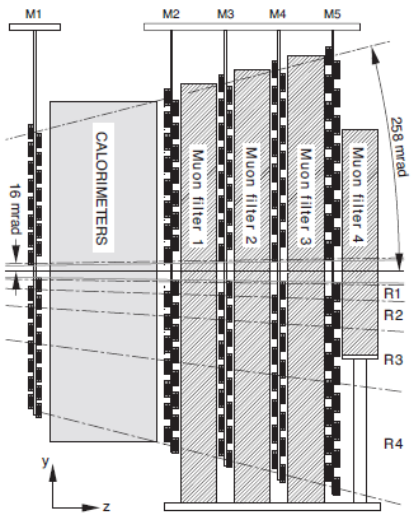

Figura 20 – Schematic of the LHCb muon system[3].

3.2.3 LHCb trigger system

During Run 2, the LHC bunch-crossing rate was 40 MHz, which far exceeds the readout and storage capabilities of the LHCb detector. In high-energy physics analyses, it is common to employ trigger systems, which consist of algorithms that classify events (or parts of events) as interesting or not for subsequent analysis. Their purpose is to reduce the data rate to a manageable level at the front-end electronics of the subdetectors, which during Run 2 operated at a rate of 1 MHz.

During Run 2 data taking at LHCb, the trigger system was divided into two levels. The first level, L0, operated at the hardware level using information from the calorimeter and the muon system, requiring either high transverse momentum muons or high transverse energy deposits from photons, hadrons, or electrons in the calorimeters. The High Level Triggers (HLT), on the other hand, operated at the software level and were themselves divided into two stages. In the first stage, HLT1, a partial reconstruction of the decay is performed for events that passed L0, reconstructing quantities such as transverse momentum and impact parameters. These are used to apply further selection criteria before forwarding events to HLT2, where a full event reconstruction is performed using dedicated trigger lines designed to select candidates of interest for each decay channel.

During operation, data are collected in smaller units known as runs, which are grouped into larger units called fills. Some runs within each fill are taken using specific trigger lines dedicated to the real-time alignment and calibration of the LHCb detector during data taking. These data, together with all events processed by the trigger system, are sent to the event filter farms, which are computing clusters where the trigger software runs. There, new alignment and calibration metrics are computed and automatically applied if the newly determined parameters differ significantly from the current values.

One of the main differences between Run 1 and Run 2 of the LHC was the collision energy. While Run 1 operated at centre-of-mass energies of 7 and 8 TeV, Run 2 ran at 13 TeV, which significantly increased the production cross-sections and, consequently, the event rates. This motivated LHCb to introduce the Turbo stream during Run 2. In this approach, high-quality signal candidates are written directly to permanent storage, removing the need to save the full raw detector data for the rest of the event. This allows events to be stored with a much reduced size and eliminates the need for a full offline reconstruction.

Events that pass the trigger lines and are stored can be classified into different trigger categories depending on whether the trigger decision was caused by the signal candidate itself or by other activity in the event with respect to the dedicated trigger line. Events are therefore classified as Triggered on Signal (TOS), Triggered Independent of Signal (TIS), or both.

3.3 LHCb Upgrade I

Following the success of LHCb during Runs 1 and 2, the experiment demonstrated excellent capabilities in several additional areas, such as electroweak physics, heavy-ion physics, and fixed-target physics. To fully exploit this potential in the higher-luminosity conditions of Run 3, a comprehensive program of hardware and software upgrades was implemented at LHCb during the period between Run 2 and Run 3, enabling the experiment to evolve towards a general-purpose detector.

As the first major change introduced during Upgrade I, LHCb abandoned its hardware trigger level, L0, and transitioned to a fully software-based trigger system. Experience from Runs 1 and 2 showed that L0 imposed overly restrictive limitations on the experiment, resulting in a significant loss of efficiency for processes with hadronic final states. In addition, L0 was responsible for introducing charge asymmetries, which became a challenge in CP-violation studies due to the difficulty of accurately reproducing these effects in simulation. With the removal of the L0 level, the entire detector front-end electronics had to be redesigned to cope with the increased event rate. This led LHCb to adopt the TELL40 (citação) readout system, which was developed to operate at the full 40 MHz bunch-crossing rate of the LHC.

In Upgrade I, a new version of the VELO was installed [55], capable of operating at a readout rate of 40 MHz and at luminosities of up to $2 \times 10^{33} cm^{-2} s^{-1}$. Among the most relevant hardware upgrades to the VELO, the adoption of hybrid silicon pixel sensors can be highlighted. These are expected to improve the resolution of both primary and secondary vertex reconstruction, allowing for a better rejection of combinatorial background. In addition, the upgraded VELO now operates even closer to the beam, with the active pixel sensors positioned at only 5.5 mm from the beam axis, resulting in a substantial improvement in the impact parameter resolution of reconstructed particles.

The tracking stations used during Run 2 were completely replaced. The Tracker Turicensis (TT) was substituted by the Upstream Tracker (UT) [56], and the downstream tracking stations T1–T3 were fully replaced by the Scintillating Fiber Tracker (SciFi) [57]. Both detectors operate with continuous readout at 40 MHz. The UT is a detector composed of four planes of silicon microstrips located upstream of the magnet, already immersed in the fringe field of the magnet. Among its main functions, it is designed to reduce the number of ghost tracks and to provide momentum estimates before the SciFi detector. The SciFi, in turn, completely replaces both the Inner Tracker and the Outer Tracker. It is a homogeneous detector composed of plastic scintillating fibers with SiPM-based readout technology and is responsible for the full measurement of the trajectories and momenta of charged particles.

In the particle identification system, the RICH detectors were upgraded to operate

at the same readout rate as the rest of the detector. The system now employs multi-anode photomultiplier tubes. In particular, RICH1 underwent a major optical redesign in order to reduce the large hit occupancy expected in the central region of the detector.

Finally, for the calorimeter and muon systems, apart from upgrading the front-end electronics to be compatible with the 40 MHz readout rate, only a few additional changes were implemented. In the calorimeter system, the pre-shower detector and the SPD were removed in order to reduce detector ageing under the increased luminosity conditions. In the muon system, station M1 was removed, as it was the main station providing information to the hardware trigger L0.

4 Selection of $D_s \to \pi\pi K$

4.1 Variables

During the proton–proton collisions that occur within the VELO, a large number of particles can be produced at the primary vertex. Among these, the present analysis is interested in the subset of events in which a D_s^+ meson is produced and subsequently decays at a secondary vertex into a $\pi^-\pi^+K^+$ final state. In LHCb analyses, the goal is to construct data samples containing a set of variables related to physically meaningful quantities, such as kinematic and geometric properties of the particles forming the final state. These variables can be used both to reconstruct physical observables associated with the decaying meson and to characterize the probability that a given trajectory is correctly reconstructed for a given particle. Figure 21 illustrates some of the variables used in this analysis.

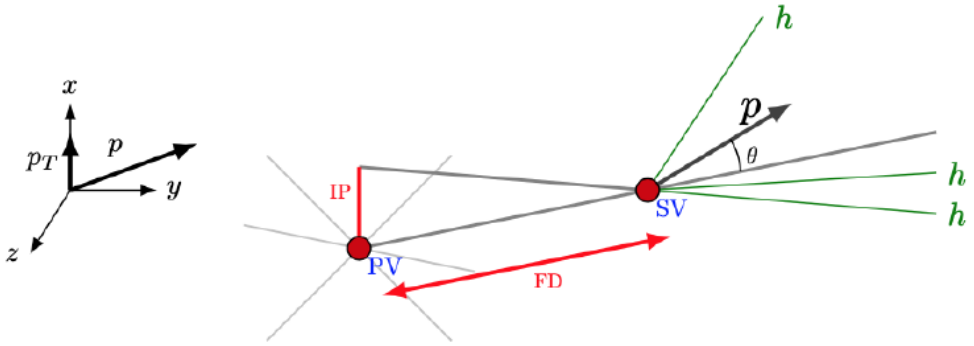

Figura 21 – Illustration of the topological structure of a 3-body decay.

The variables adopted in this analysis are:

- **Momenta** $\vec{p}_D, \vec{p}_i$: Reconstructed linear momenta of the mother particle ($\vec{p}_D$) and of the daughter particles ($\vec{p}_i$) in LHCb. The z-axis is defined along the direction from the VELO to the rest of the detector; the xy plane is perpendicular to it, with the y-axis horizontal and the x-axis vertical.
- **Transverse momentum** (p_T): Projection of the particle momentum onto the xy plane. It is defined both for the D_s^+ candidate (p_T) and for its decay products (p_T^i).

- **Invariant mass ($\pi^-\pi^+K^+$)**: Mass of the D_s^+ candidate reconstructed from the daughter momenta $\vec{p}_i$, assigning the π^-, π^+, and K^+ masses to tracks 1, 2, and 3, respectively. Using the masses of the three daughter particles and their momenta, the mass of the mother particle can be reconstructed.

- **Flight distance (FD)**: Distance between the primary vertex (PV) and the secondary vertex (SV).

- **Impact parameter (IP)**: Minimum distance between the particle trajectory and the primary vertex.

- **DIRA**: Defined as the cosine of the angle between the reconstructed momentum of the mother particle and the direction connecting the PV and the SV.

- **Vertex χ^2**: Variable that provides information on the quality of the secondary-vertex reconstruction.

- **Impact-parameter χ^2 ($\chi^2_{\rm IP}$)**: Difference in the χ^2 of the primary-vertex fit with and without the particle under consideration. It can be computed for the mother particle ($\chi^2_{{\rm IP}D}$) or for the daughter particles ($\chi^2_{{\rm IP}i}$).

- **logIP**: Defined as
$$\log{\rm IP} = \log\left(\frac{\prod_i^3 \chi^2_{{\rm IP}_i}}{\chi^2_{{\rm IP}_D}}\right). \tag{4.1}$$

- **LogFDchi2**: Logarithm of $\chi^2_{\rm FD}$, where $\chi^2_{\rm FD}$ is the flight distance divided by its squared uncertainty.

- **Pointing**: Defined as
$${\rm Pointing} = \frac{p\sin\theta}{p\sin\theta + \sum_i^3 p_{Ti}}, \tag{4.2}$$
where θ is the angle between the momentum of the D_s^+ candidate and the flight-distance vector.

- **Track χ^2/ndf**: Track-fit χ^2 per number of degrees of freedom.

- **PIDK**: Likelihood-based variable constructed from RICH information, defined as the difference between the kaon and pion hypotheses.

- **ProbNNK**: Particle-identification variable obtained from a multivariate analysis, providing a single probability value for a reconstructed particle to be a kaon.

4.2 Data selection

The $D_s^+ \to \pi^-\pi^+K^+$ candidates are first selected from a sample of events accepted by the `photon, hadron, muon` and `dimuon` L0 TIS trigger decision lines. An exclusive HLT2 Turbo Line *Hlt2CharmHadDspToKpPimPip* selects the dataset for $D_s^+ \to \pi^-\pi^+K^+$ decay. The selection criteria used in HLT2 can be seen in Table 1. The specific HLT2 trigger line requires the D_s^+ candidates to be TOS (Triggered On Signal) on any two HLT1 Track Lines, namely *Hlt1TrackMVA* and *Hlt1TwoTrackMVA*.

The HLT1 lines use an MVA algorithm based on a neural network to select the candidates. Table 2 shows the requirements for one and two tracks before the training, as well as the variables used in the HLT1 MVA stage.For more information, please refer to [58]

Final particle cuts	
Track χ^2/ ndf	< 3.0
p_{Ti} [MeV]	> 250
χ^2_{IP}	> 4.0
PIDK (kaons)	> 5
PIDK (pions)	< 5
Combination cuts	
Mass [MeV]	1879 - 2059
$\sum p_{Ti}$ [MeV]	> 3000
p_{Ti} (at least one track) [MeV]	> 1000
P_{Ti} (at least two tracks)[MeV]	> 400
χ^2_{IPi} (at least one tracks)	> 50
χ^2_{IPi} (at least two tracks)	> 10
Decaying particle cuts	
χ^2 vertex/ndf	< 6
Mass [MeV]	1889 – 2049
lifetime (ps)	> 0.2
acos(DIRA) [mrad]	< 14.1
TisTosSpec HLT1.*Track.*Decision %TOS	

Tabela 1 – HLT2 selection criteria.

Requirements used in Hlt1TrackMVA	
Track pre-selections	
Track χ^2/ndf	< 3.0
p_T[MeV]	> 500
IP χ^2 wrt PV	> 4
Variables for NN	
p_T, χ^2 of Impact Parameter	
Requirements used in Hlt1TwoTrackMVA	
Track preselections	
Track χ^2/ndf	< 2.5
p_T[MeV]	> 500
IP χ^2 wrt PV	> 4
TwoTrack preselections	
Track χ^2/ndf	> 2
p_T[MeV]	< 10
mcor [GeV]	> 1
η	2 – 5
Variables for NN	
$\sum p_T$,vertex χ^2/DOF,FD χ,N(tracks with IP $\chi^2 < 16$)	

Tabela 2 – HLT1 selection criteria.

These data then undergo a tupling process, in which the files are converted into a format compatible with the ROOT framework [59], the environment in which this analysis is performed. At this stage, a set of loose selection cuts is introduced in order to partially reduce the combinatorial background present in the samples. This step is performed centrally for all analyses of this type, and the corresponding code is documented in [60]. The requirements adopted for the $D_s^+ \to \pi^-\pi^+K^+$ channel are listed in Table 3, and the number of events obtained after the preselection is shown in Table 4, organized by year and polarity.

NSPDHits	< 1000
$\eta_{daughter}$	$1.5 < \eta_i < 5$
$P_{daughter}$	<100 GeV
D_s χ^2_{IP}	< 12
$1905 <$Mass$(\pi^-\pi^+K^+) < 2035$ MeV	

Tabela 3 – Central production requirements for $D_s^+ \to \pi^-\pi^+K^+$ decays.

4.3 Monte Carlo samples

Full Monte Carlo (MC) simulation samples are used to determine parameters of signal PDFs for mass fits. A multivariate analysis is performed in order to reduce the

Year	Polarity	N. of candidates after pre-selection
2016	Up	107191592
	Down	107231186
2017	Up	98246955
	Down	101869527
2018	Up	110950532
	Down	109410192

Tabela 4 – Number of candidates after the central production requirements.

combinatorial background. The simulation is used as the signal proxy in the MVA training steps.

To generate large $D_s^+ \to \pi^-\pi^+K^+$MC samples, DecFiles with generator-level cuts are applied to all tracks and the parent. A trigger filter (code available at here) is also applied. The generator cuts, defined to match the thresholds observed in data after full selection, are summarized in Table 5. DecayTreeTuples with all those events are produced using the momentum scale calibration tool. More details about the procedure can be found in [61]. These MC samples are generated according to phase space (no dynamics). The corresponding number of simulated events is presented in Table 6. The only selection requirements that are not explicitly applied are those related to particle identification. To account for the effects of these selections, PID weights are applied using PID calibration tables [1] corresponding to the respective requirements.

Variable	Requirement
p_i [GeV]	> 2.0
p_{T_i} [GeV]	> 0.25
$p(D_s^+)$ [GeV]	> 14.0
$p_T(D_s^+)$ [GeV]	> 2.1

Tabela 5 – Requirements applied at generator level. Momentum cuts are applied to all decay products.

Year	Polarity	N. of candidates
2016	Up	1703151
	Down	1717377
2017	Up	1884780
	Down	1889980
2018	Up	1504161
	Down	1512843

Tabela 6 – Number of candidates of $D_s^+ \to \pi^-\pi^+K^+$MC samples.

[1] The PID tables are parameterizations of the response of the particle identification system, used to statistically reproduce its identification performance.

4.4 Offline selection

The $\pi\pi K$ invariant-mass spectrum after the preselection is shown in Figure 22 . Despite the optimizations of the trigger system and the preselection requirements, the sample still exhibits a high level of background, as well as a structure on the right-hand side of the spectrum, which is attributed to a specific background contribution. For the purposes of this analysis, it is important to reduce this background such that its impact on the final result is minimized.

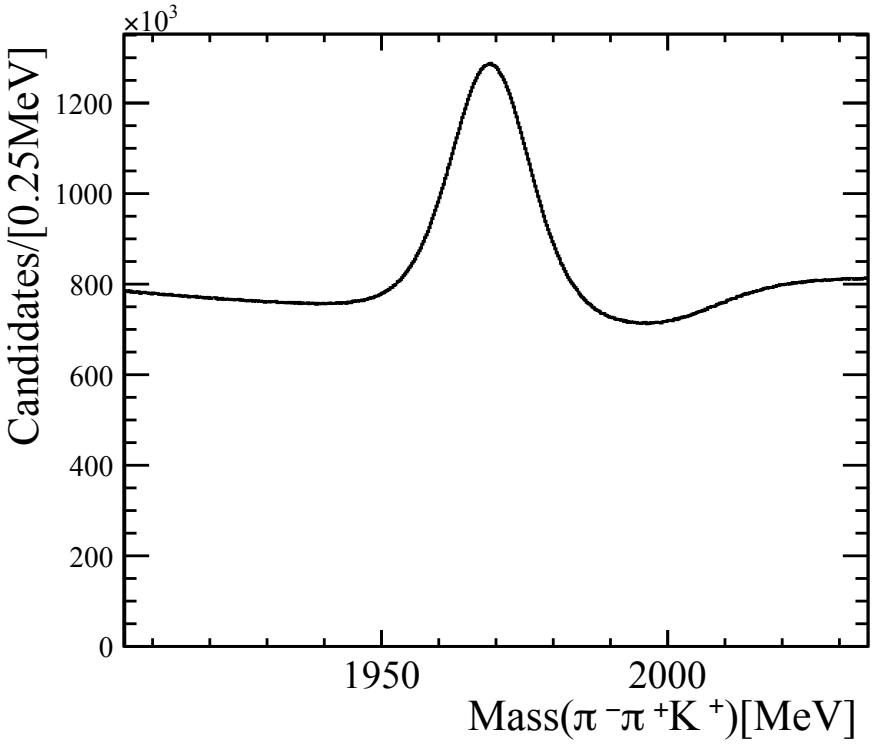

Figura 22 – Invariant mass distribution for de $D_s^+ \to \pi^-\pi^+K^+$ of candidates which satisfy all the selection criteria up to central Working Group production.

To this end, two complementary approaches are adopted in this analysis to suppress the background: studies of specific background sources and a multivariate analysis. The first approach aims to identify possible sources of spurious specific backgrounds and to define selection requirements that not only reduce their contamination but also further optimize the preselection cuts. The second approach is employed as a general strategy to maximize background rejection while minimizing the loss of signal candidates.

4.4.1 Specific background studies

The source of the structure observed in the upper mass sideband, depicted in Fig. 22, can be traced to the decay $\bar{D}^0 \to \pi^-K^+$, as shown in the invariant mass (π^-K^+) distribution in Fig. 23, with the $\bar{D}^0$peak located at 1.86 GeV. This contamination arises in

our data when a random π^+, produced in an unrelated process, is erroneously associated with the D^0 decay products, forming a three-track vertex that passes the $D_s^+ \to \pi^-\pi^+K^+$ selection requirements. To mitigate this contamination, we apply a requirement that the invariant mass of the π^-K^+ system be less than 1.84 GeV.The invariant mass distribution without this contamination is shown on the figure 24.

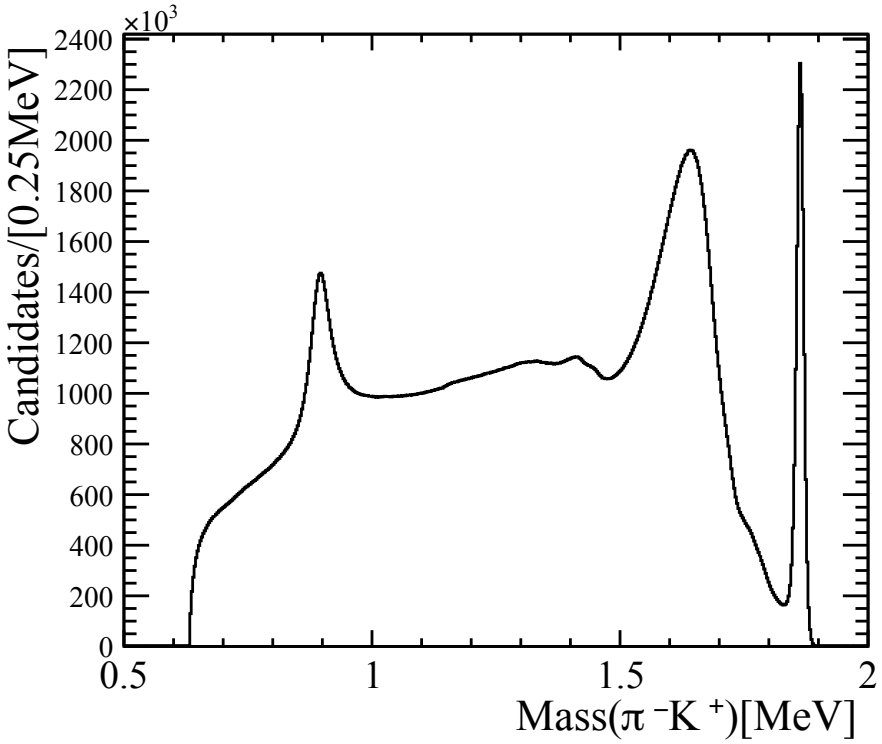

Figura 23 – Invariant mass spectrum π^-K^+ of candidates which satisfy all the selection criteria up to central Working Group production. The clean peak at the end of the spectrum corresponds to the $\bar{D}^0 \to \pi^-K^+$ decay.

Fig. 23 also shows a broad structure between 1.5 GeV and 1.7 GeV which cannot be associated to a known $K\pi$ resonance.

Since the decay $\bar{D}^0 \to \pi^-K^+$ is observed in the invariant-mass spectrum, it is reasonable to assume that the decay $\bar{D}^0 \to \pi^0\pi^-K^+$ could also appear in the spectrum. In this case, the $K\pi$ from the D^0 decay could be wrongly associated with a random π^+ from the event, forming the same final state of interest, $\pi^-\pi^+K^+$. Because the π^0 is not reconstructed, the signal from this decay would appear at a reconstructed mass than the nominal $\bar{D}^0$ mass and with a relatively larger uncertainty. This contribution could overlap with the regions of the $K^*(1410)$, and $K_0^*(1430)$ resonances in the spectrum.

This hypothesis is confirmed by a study using fast simulations, using RapidSim [62] to generate $\bar{D}^0 \to \pi^0\pi^-K^+$ decays within LHCb geometric acceptance. We generated 10 million events using EvtGen DDalitz model based on the decay model from the E691

experiment [63], which includes contributions from the ρ, K^{*0}, K^{*-}, $K_0(1430)$, $\bar{K}_0(1430)$, $\rho(1700)$ and $K^{*-}(1680)$ resonances.

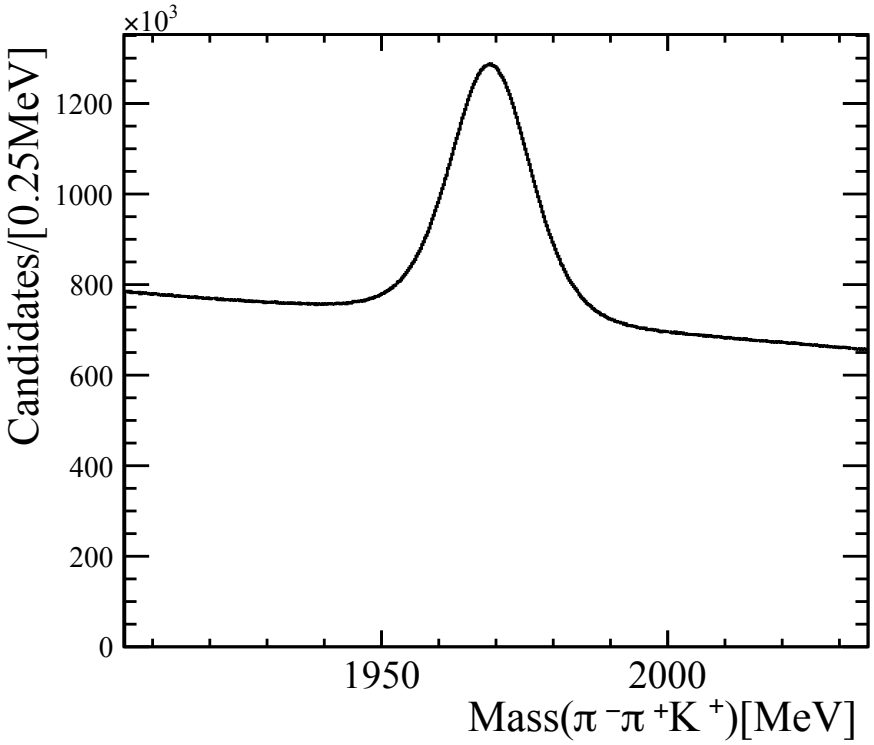

Figura 24 – Invariant mass distribution for de $D_s^+ \to \pi^-\pi^+K^+$ of candidates after appling the cut to remove $\bar{D}^0 \to \pi^-K^+$contamination

In the simulated samples, the neutral pion from the $\bar{D}^0 \to \pi^0\pi^-K^+$ decay is replaced by a positive-charged pion, forming a fake $\pi^-\pi^+K^+$ combination. To reproduce this effect, we reconstruct this fakes combinations using the charged products of the RapidSim (π^- and K^+) with a π^+ in the place of π^0. The π^+ are selected from data, using particles from signal candidates that satisfy the selection cuts up to offline selection and lie in the background region, $1905 < m(\pi^-\pi^+K^+) < 1930$ MeV or $2010 < m(\pi^-\pi^+K^+) < 2035$ MeV. Moreover, a $\pi^+\chi^2_{IP} < 30$ cut is applied to enrich the sample with π^+ candidates with higher probability to come from the interaction point. The idea is see if we can reproduce the same pattern found on the Dalitz plot of background region in $D_s^+ \to \pi^-\pi^+K^+$ decay (left plot on Figure 25).

The right plot of Fig. 25 shows how the misreconstructed $\bar{D}^0 \to \pi^0\pi^-K^+$ decays appear in the $D_s^+ \to \pi^-\pi^+K^+$ phase space, using the RapidSim simulation. This can be compared with the pattern observed in data, for candidates from the background region (left plot). This similarity allows us to associate this structure with the decay $\bar{D}^0 \to \pi^0\pi^-K^+$.

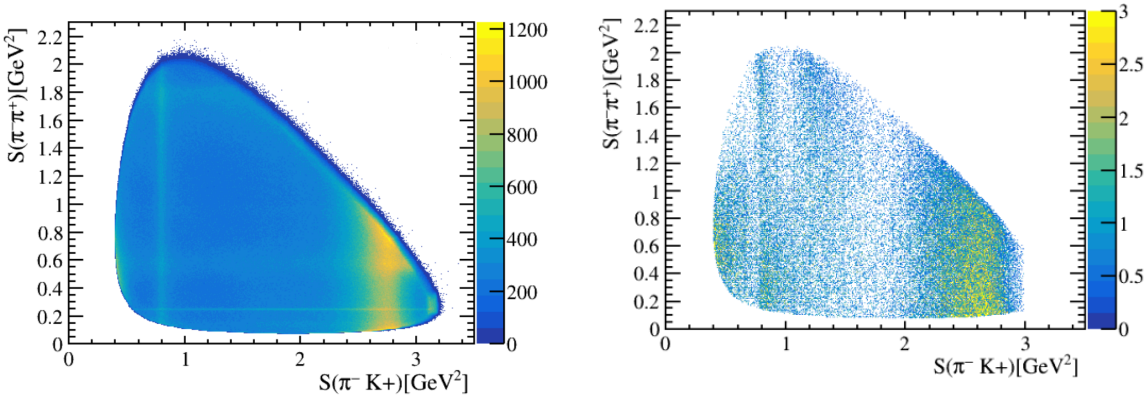

Figura 25 – On left, the Dalitz plot for the background region(left and right sidebands) from $D_s^+ \to \pi^-\pi^+K^+$ data. On right, the RapidSim $\bar{D}^0 \to \pi^0\pi^-K^+$ sample reconstructed as $\pi^-\pi^+K^+$.

From this study, this contamination is distributed evenly across the $\pi^-\pi^+K^+$ invariant mass interval, as shown in Fig. 26. Unlike the $\bar{D}^0 \to \pi^-K^+$ decay, we cannot remove it with a mass cut without affecting a large region of the phase space. Instead, as better discussed in the next section, we exploit variables that are expected to help reducing this contribution, in particular IP χ^2 and the p_T of the π^+ candidate, using cuts and including them in a multivariate analysis selection.

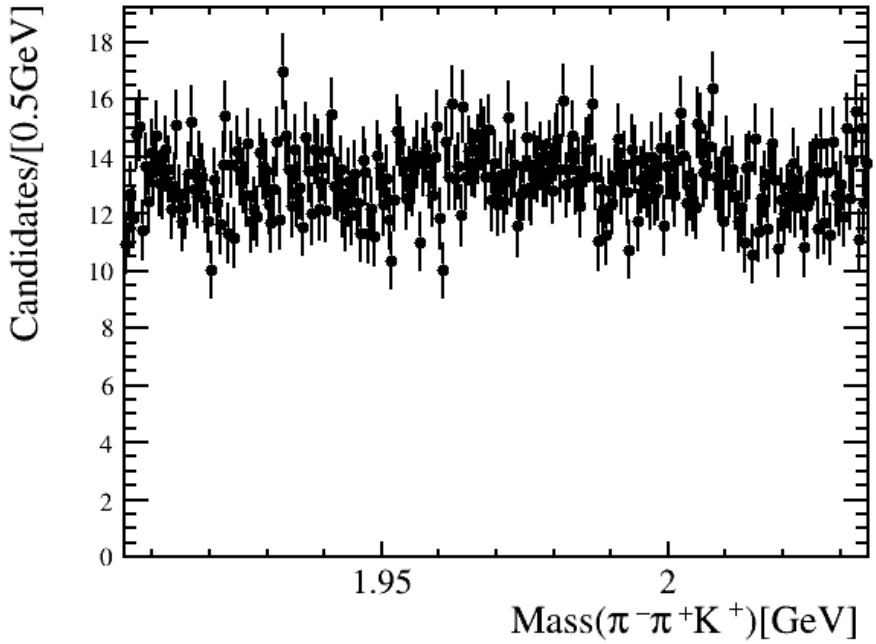

Figura 26 – Invariant mass distribution $\pi^-\pi^+K^+$ for $\bar{D}^0 \to \pi^0\pi^-K^+$ RapidSim reconstructed as $\pi^-\pi^+K^+$.

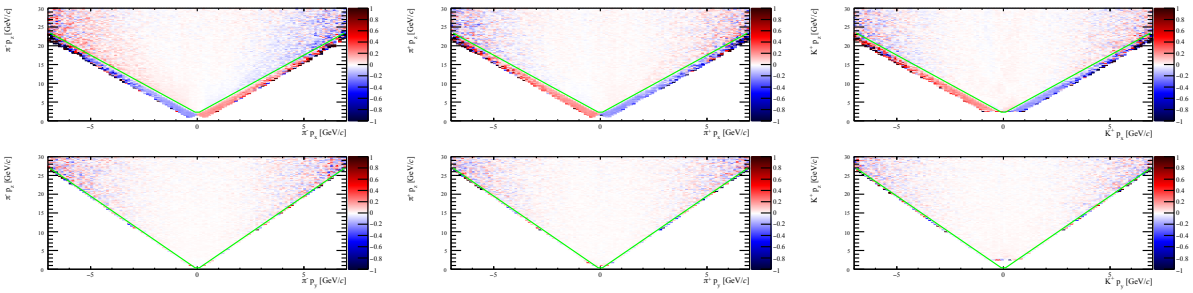

Figura 27 – Charge asymmetry across the $p_z \times p_x$ (top) and $p_z \times p_y$ (bottom) planes, for the π^-(left), π^+(middle) and K^+(right) in data. Positive charge asymmetries are represented in red, negative asymmetries in blue, and the fiducial cuts used in the analysis are shown in green.

To address this contamination, we studied the impact of selection cuts on several variables and identified the IP χ^2 and the p_T of the π^+ candidate as particularly effective in reducing this background. In addition, these variables were included among the inputs to the multivariate analysis (MVA), allowing the method to further eliminate this contamination in a more efficient way. The selection requirements applied to these variables are listed in Table 7.

4.4.2 Additional cuts

Although the HLT2 already incorporates particle-identification requirements (PID), this analysis applies slightly tighter PID selections in order to ensure the correct identification of the particles composing the final state. For the purposes of this analysis, PID requirements with the largest impact on background reduction were selected, while minimizing the loss of signal events. An arbitrary criterion was adopted whereby the PID cut is chosen to maximize background suppression while retaining a signal efficiency of 97%. The resulting PID requirements are listed in Table 7.

As explained in the previous chapter, the polarity of the LHCb magnet is periodically reversed during data taking, producing data samples classified as *MagUp* and *MagDown* according to the detector polarity. This strategy is necessary due to the detector geometry, which leads to acceptance regions in which particles of a given charge are more likely to be deflected out of the acceptance than particles of the opposite charge. By combining the *MagUp* and *MagDown* samples, part of these asymmetries tends to cancel; however, this cancellation is not perfect, resulting in residual asymmetries in momentum space. These asymmetries can be observed in Figures 27.

To reduce the impact of these effects in this analysis, the regions exhibiting significant asymmetries are removed by applying fiducial cuts in momentum space. The selection requirements used to suppress these asymmetries are listed in Table 7.

Variable	Requirement
ProbNNK (for the kaon)	> 0.4
ProbNNPi (for both pions)	> 0.4
$P_T(\pi^+)$ [MeV]	> 400
$\pi^+\chi^2_{IP}$	> 10
Fiducial cuts	
$p_z > 3.9 \times \lvert p_y \rvert$ GeV	
$p_z > 3.1 \times \lvert p_x \rvert + 2$ GeV	

Tabela 7 – Additional cuts applied to all three final state particles before the MVA.

4.4.3 MVA Selection

. The final stage of data selection involves processing the data through a multivariate analysis selection aimed at reducing both the combinatorial background and contamination from the decay $\bar{D}^0 \to \pi^0\pi^-K^+$ in our dataset, thereby increasing the sample purity. For this process, the Toolkit for Multivariate Analysis (TMVA) package [64] is used.

TMVA is an framework that performs an event-by-event classification, assigning to each candidate a probability of representing signal or background. For its operation, representative samples of signal and background must be provided, along with a set of variables to be used in the classification process. The procedure is divided into two stages. In the training stage, the algorithm learns a decision function from the provided list of variables using a reduced subsample of the data. In the application stage, the function learned during training is applied to the remaining sample, producing an output variable associated with the classification process, which can be used to define selection cuts and to determine the final sample.

For the purposes of this analysis, Monte Carlo samples were adopted as a proxy for the signal, while candidates selected from the sidebands of the $\pi^-\pi^+K^+$ invariant mass distribution were used as a proxy for the background. More specifically, events within the mass windows $1905 < m(\pi^-\pi^+K^+) < 1935$ MeVor $2005 < m(\pi^-\pi^+K^+) < 2035$ MeV. were used. The variables employed in the classification were D_s^+IP, $D_s^+\chi^2_{IP}$, $LogIP$, D_s^+FD, $LogFDchi2$, $Pointing$, Vertex χ^2, $DIRA$, D_s^+P ,D_s^+PT, π^+Pt and $\pi^+\chi^2_{IP}$.

The classifier used in this work was the BDT (Boosted Decision Tree), which is a supervised learning method based on the combination of several simple decision trees[2]. Each internal node of a tree represents a splitting criterion on a given variable, while the terminal nodes (leafs) assign a classification or a probability value to the event.

To ensure the MC variables provide an accurate representation of the signal present

[2] A decision tree consists of a sequence of binary cuts applied to an input variable in order to separate events into regions that are purer with respect to the quantity of interest, in this case signal and background.

in the $D_s^+ \to \pi^-\pi^+K^+$ decay, the MC samples were subjected to a reweighting process to align their distributions with those observed in the data, using sPlot [65] for background subtraction. The Gradient Boost Reweighter algorithm [66] was employed to obtain these equalization weights. The variables used for this process are the same set adopted in the MVA training process.

The reweighting was performed independently for each year and polarity. The hyperparameters used in this process are detailed in Table 8. The comparisons between the MC and sPlot data variables before and after the reweighting are shown in Fig. 28 and 29, respectively. Although the original variables' distributions exhibited similar behavior, a clear improvement is observed after the reweighting.

Hyper-parameter	Value
n_estimator	200
learning_rate	0.1
MaxDepth	3
n_folds	8

Tabela 8 – Hyper-Parameters for the reweithing process of the MC samples.

To fully exploit the available data sample in the MVA, a k-folding approach was adopted. In this method, the full dataset is divided into smaller subsamples. During the training stage, each subsample is used independently to produce a distinct decision function, while in the application stage each subsample receives the classifier outputs obtained from the trainings performed on the other subsamples. The main advantage of this approach is that it allows the use of the entire dataset, in contrast to the traditional approach in which the events used for training are discarded.

In this analysis, the dataset was divided into six subsamples. These were constructed by randomly splitting the MagUp and MagDown events of each data-taking year (2016, 2017, and 2018) into two new samples, labeled A and B. As a result, each subsample receives the application of the results from the other five trainings. The final output variable used for the event selection, denoted as valBDTmean, is defined as the arithmetic mean of the individual classifier outputs obtained from each training. As shown in Fig. 30, this averaging procedure effectively captures the performance of all five trainings.

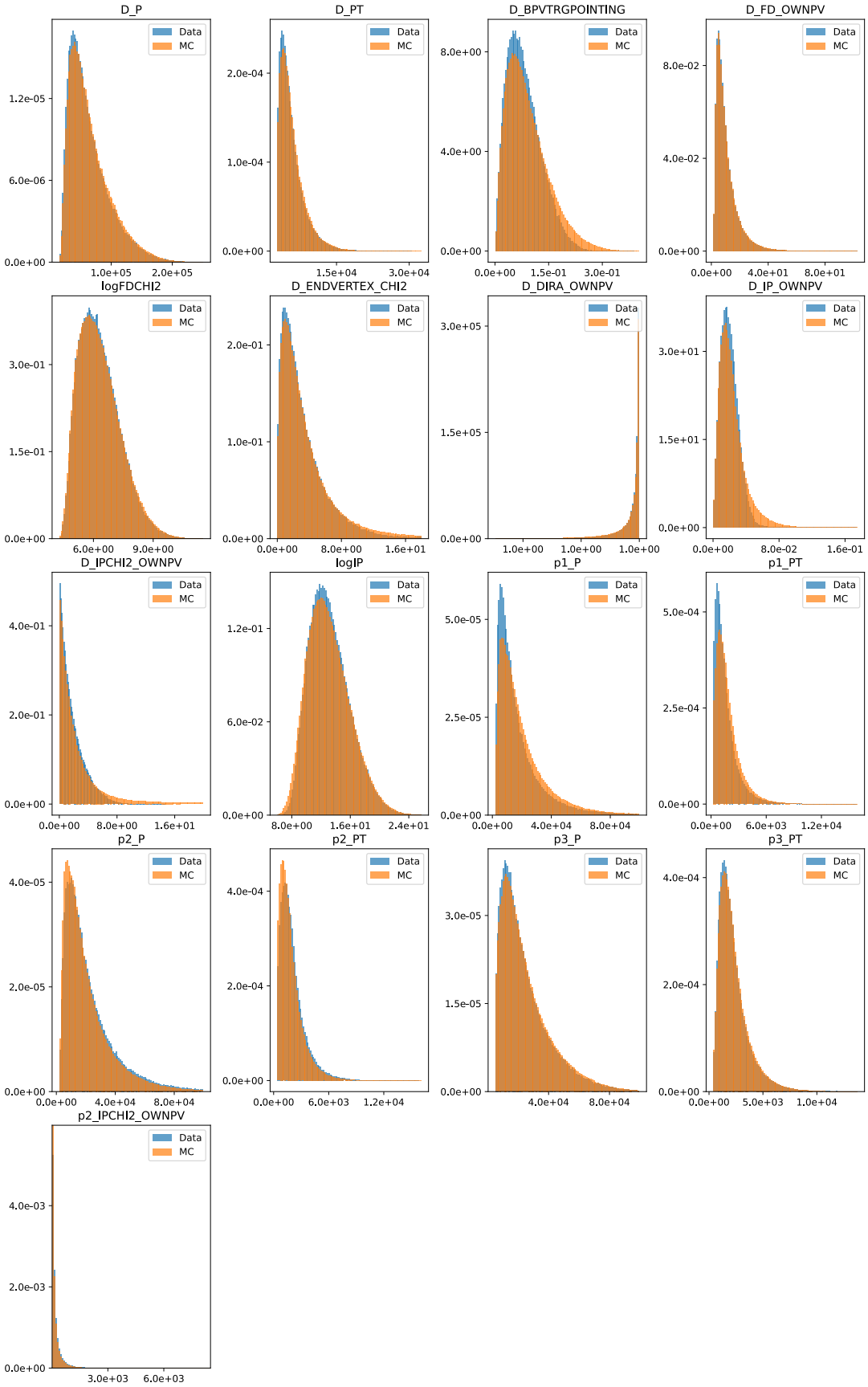

Figura 28 – Comparison between the signal (blue) and MC distributions (orange) before the reweighting process for $D_s^+ \to \pi^-\pi^+K^+$: kinematic variables of the final state particles.

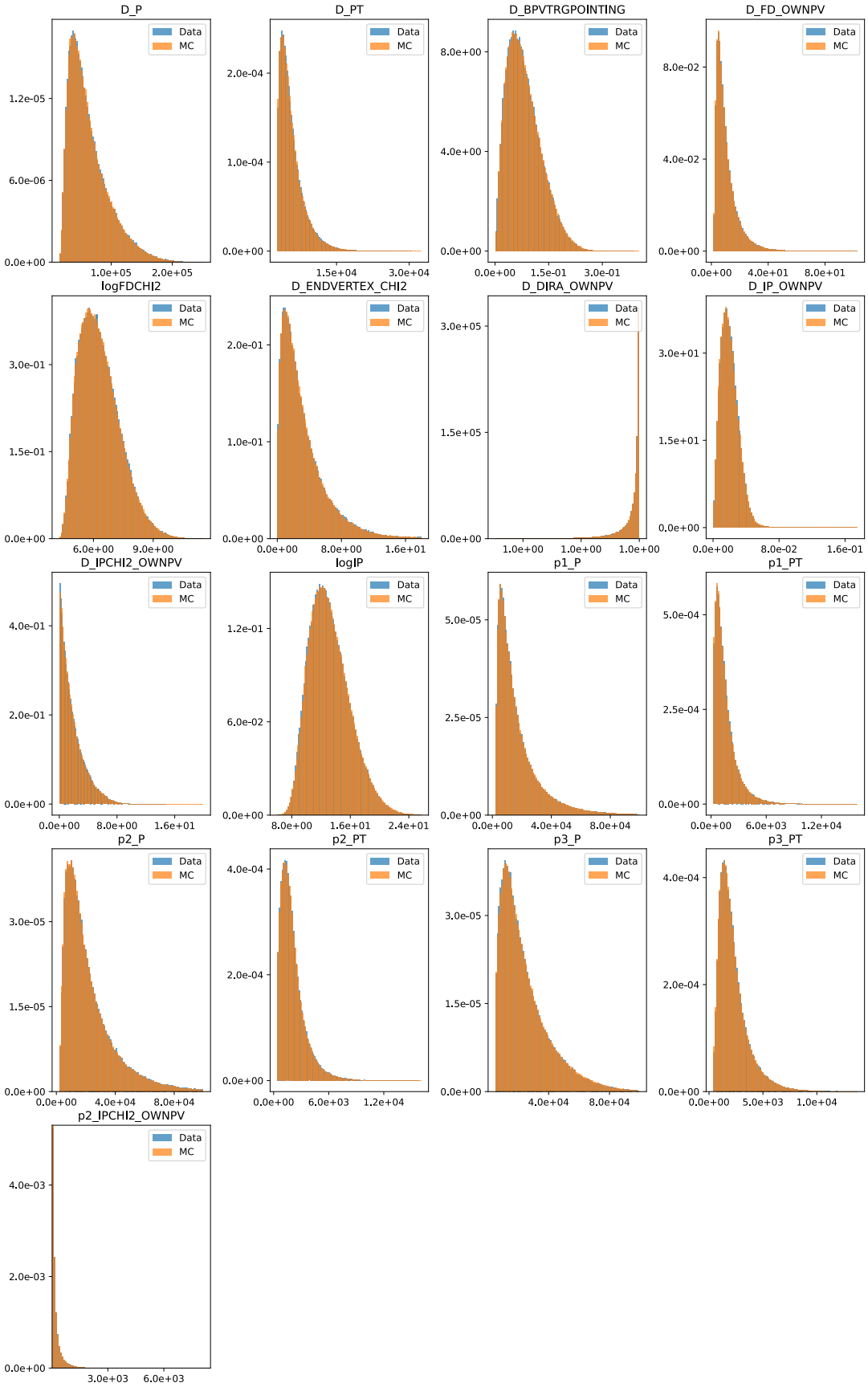

Figura 29 – Comparison between the signal (blue) and MC distributions (orange) after the reweighting process for $D_s^+ \to \pi^-\pi^+K^+$: kinematic variables of the final state particles.

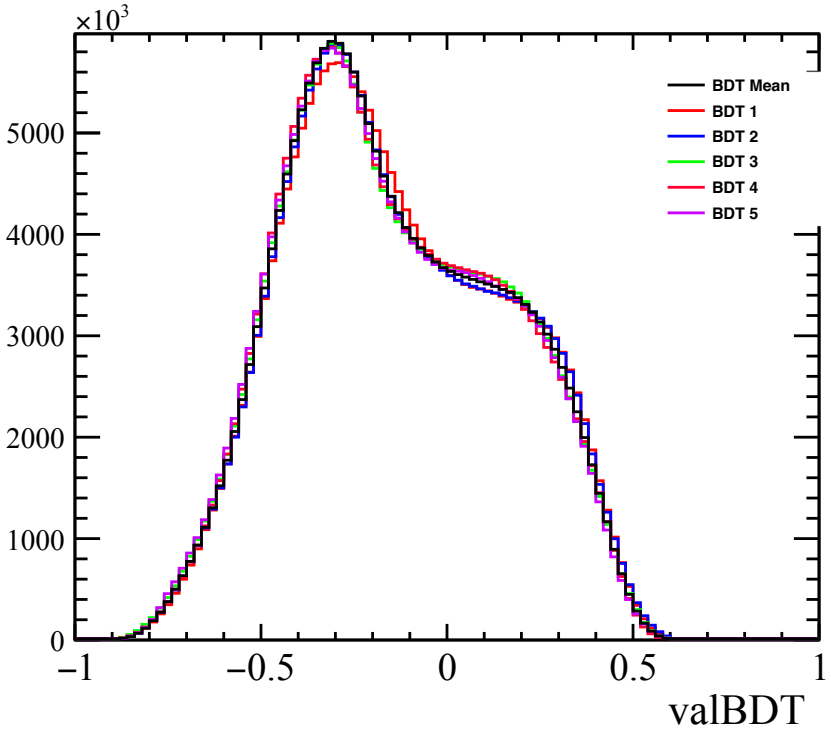

Figura 30 – Comparison between the BDT output of each MVA applied on 2016 A sample training and the mean of the five distributions.

Hyper-parameter	Value
NTrees	1000
MinNodeSize	3%
MaxDepth	4
IgnoreNegWeightsInTraining	True
AdaBoostBeta	0.5
BaggedSampleFraction	0.5
SeparationType	GiniIndex
nCuts	20

Tabela 9 – Hyper-Parameters for the BDT classifiers.

The hyperparameters adopted in the MVA training are listed in Table 9. The BDT response obtained from this procedure, shown in Fig. 31, exhibits a good separation between signal and background, while the ROC curve presented in Fig. 32 indicates that the MVA is highly efficient in distinguishing signal from background. Figure 33 shows the signal and background distributions for the variables used in the training, while Figs. 34 and 35 present the correlation matrices among the input variables for signal and background, respectively.

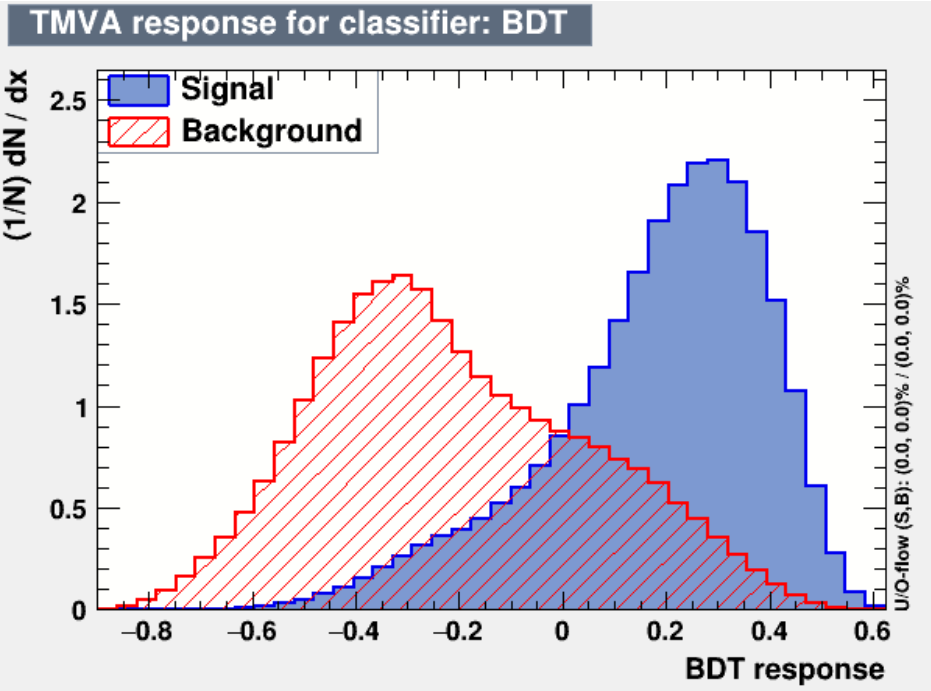

Figura 31 – BDT response obtained for 2016 A sample.

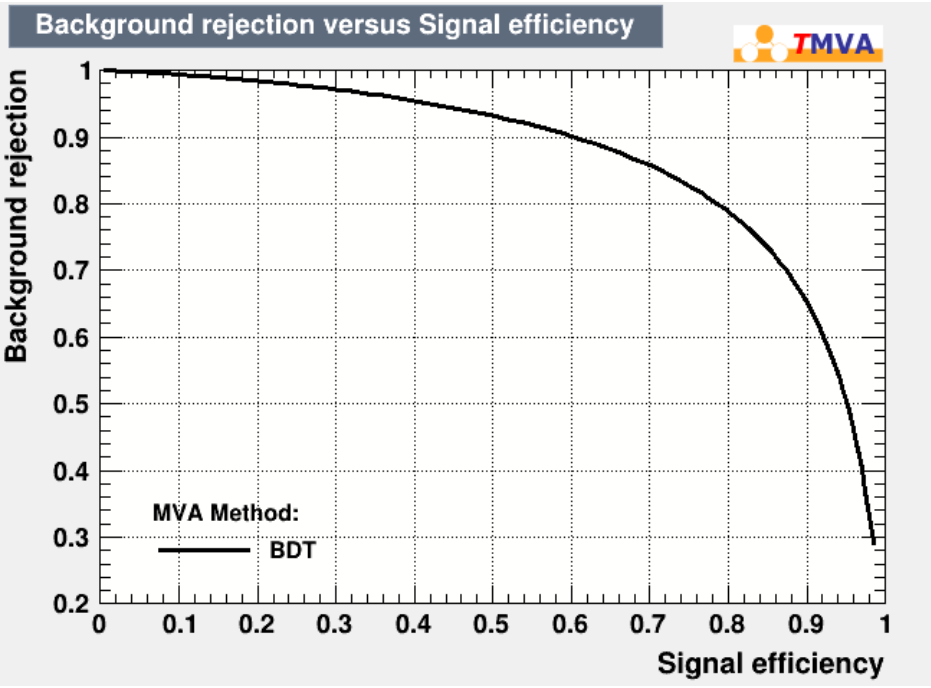

Figura 32 – ROC curve for 2016 A sample.

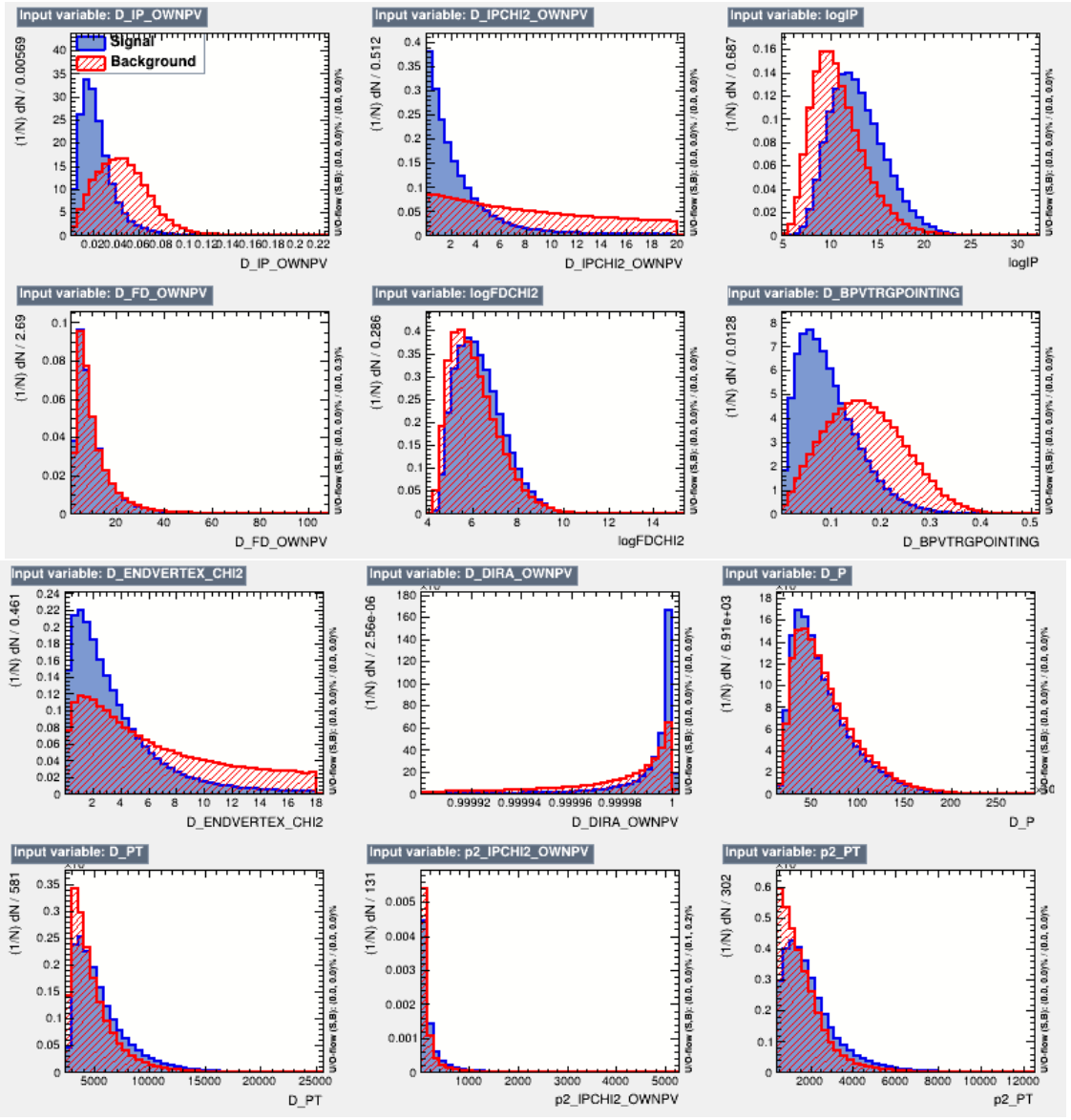

Figura 33 – Comparison between signal and background input variables used for MVA training of 2016 A.

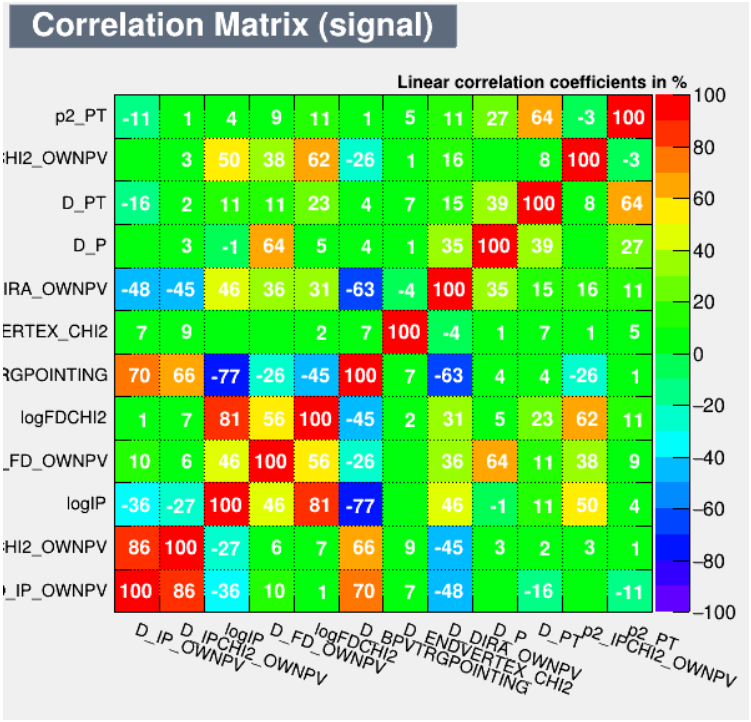

Figura 34 – Correlation matrix of the variables used to train the BDT for signal in 2016 A sample.

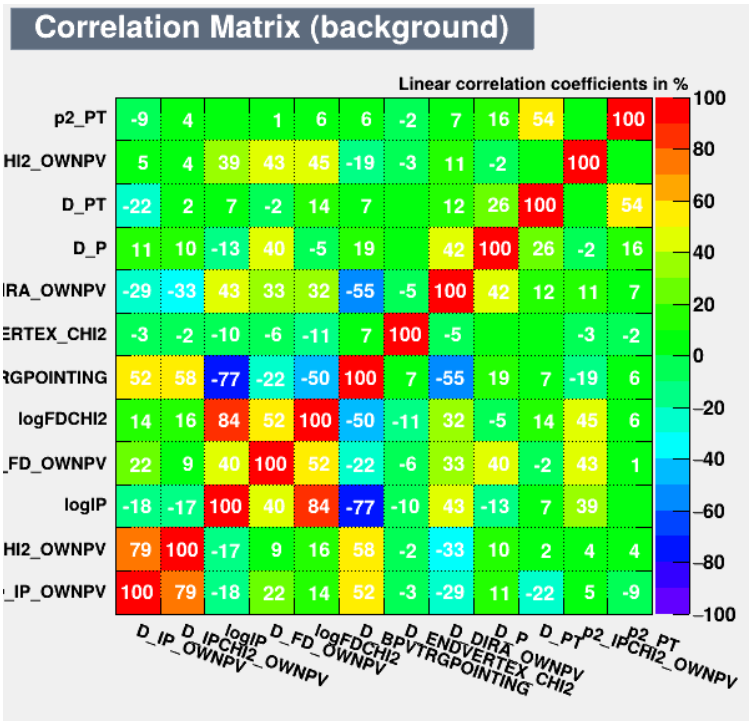

Figura 35 – Correlation matrix of the variables used to train the BDT for background in 2016 A sample.

To define the optimal selection point, the statistical significance was adopted as the figure of merit, as defined in Equation:

$$Significance = \frac{S}{\sqrt{S+B}} \tag{4.3}$$

Here, S and B represent the estimated signal and background yields in a $\pm 3\sigma$ window around the fitted D_s^+ mass, according to the integrals of the PDF fitted to the 3-body invariant mass distribution. The optimal selection point is determined at the maximum value of the statistical significance, which in this case is achieved with $valBDTmean >$ -0.056. The dependence of the statistical significance with the BDT cut is illustrated in Fig. 36.

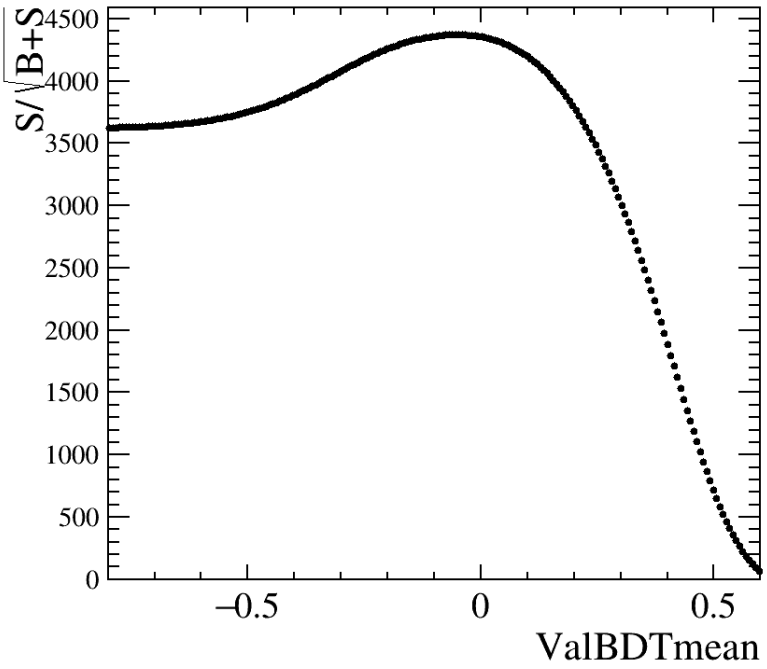

Figura 36 – Significance of the $D_s^+ \to \pi^-\pi^+K^+$ signal yield as a function of the BDT cut.

4.5 Yields of the sample.

The signal yields are determined through invariant mass fits applied to the samples after the BDT selection. To ensure an accurate estimation of the signal yields from data, maximum likelihood fits are first performed on MC samples, separated by year and polarity, to extract the signal shape. For this purpose, and for all fits used in this analysis, the RooFit framework [67] was employed.

The PDF adopted to fit the MC samples is:

$$\mathcal{P}_{sig}(m) = f_G \times G(\mu, \sigma_G) + (1 - f_G) \times ADCB(\mu, R_1\sigma_G, R_2\sigma_G, \alpha_1, N_1, \alpha_2, N_2). \tag{4.4}$$

Here, μ and σ_G represent the mean value and width of the Gaussian component, respectively. The Asymmetric Double-sided Crystal Ball [68] (ADCB) PDF is defined by parameters α_1, N_1, α_2 and N_2 which characterize the left and right tails of the Crystal Ball function. Additionally, $R_1\sigma_G$ and $R_2\sigma_G$ represent the widths of the ADCB. Gaussian mean. The parameter f_G is the fraction of the Gaussian component with respect to the ADCB.

The results of the fit for MC samples for all years and polarities are shown in Fig. 37, with fit parameters listed in Tables 10 and 11. The parameters given in these tables are fixed in the fits performed to the data, apart from the Gaussian mean μ and width σ.

In the fits to data, a background component is included, described by a third-order Bernstein polynomial (n=3),

$$\mathcal{P}_{bkg}(m) = \sum_{i=0}^{n} a_i \binom{n}{i} m^i \cdot (1-m)^{n-1} \tag{4.5}$$

with a_3 fixed to 1. This background component accounts for the combinatorial background present in the data after the BDT selection. The total PDF is given by

$$\mathcal{P}^{\pm\frac{U}{D}} = N_{sig}^{\pm\frac{U}{D}} \mathcal{P}_{sig}^{\pm\frac{U}{D}} + N_{bkg}^{\pm\frac{U}{D}} \mathcal{P}_{bkg}^{\pm\frac{U}{D}} \tag{4.6}$$

Parameter	2016	2017	2018
α_1	1.89 ± 0.02	1.90 ± 0.02	1.86 ± 0.02
α_2	-2.25 ± 0.03	-2.33 ± 0.03	-2.28 ± 0.04
N_1	1.10 ± 0.04	1.07 ± 0.04	1.18 ± 0.04
N_2	2.55 ± 0.15	2.08 ± 0.15	2.61 ± 0.23
$R(\sigma_G)_1$	1.88 ± 0.02	1.92 ± 0.02	1.90 ± 0.02
$R(\sigma_G)_2$	1.72 ± 0.01	1.75 ± 0.01	1.74 ± 0.01
f_G	0.62 ± 0.01	0.63 ± 0.01	0.63 ± 0.01
μ	1969.41 ± 0.01	1969.40 ± 0.01	1969.40 ± 0.01
σ	5.90 ± 0.04	5.94 ± 0.05	5.90 ± 0.04

Tabela 10 – Parameters extracted from the MC MagUp sample fit.

Parameter	2016	2017	2018
α_1	1.90 ± 0.02	1.86 ± 0.02	1.85 ± 0.02
α_2	-2.30 ± 0.03	-2.31 ± 0.03	-2.25 ± 0.06
N_1	1.05 ± 0.04	1.12 ± 0.05	1.16 ± 0.05
N_2	2.26 ± 0.16	2.16 ± 0.13	2.56 ± 0.34
$R(\sigma_G)_1$	1.91 ± 0.02	1.92 ± 0.02	1.92 ± 0.05
$R(\sigma_G)_2$	1.74 ± 0.01	1.74 ± 0.01	1.74 ± 0.02
f_G	0.64 ± 0.01	0.65 ± 0.021	0.65 ± 0.02
μ	1969.40 ± 0.01	1969.40 ± 0.01	1969.40 ± 0.01
σ	5.94 ± 0.04	5.99 ± 0.05	5.99 ± 0.04

Tabela 11 – Parameters extracted from the MC MagDown sample fit.

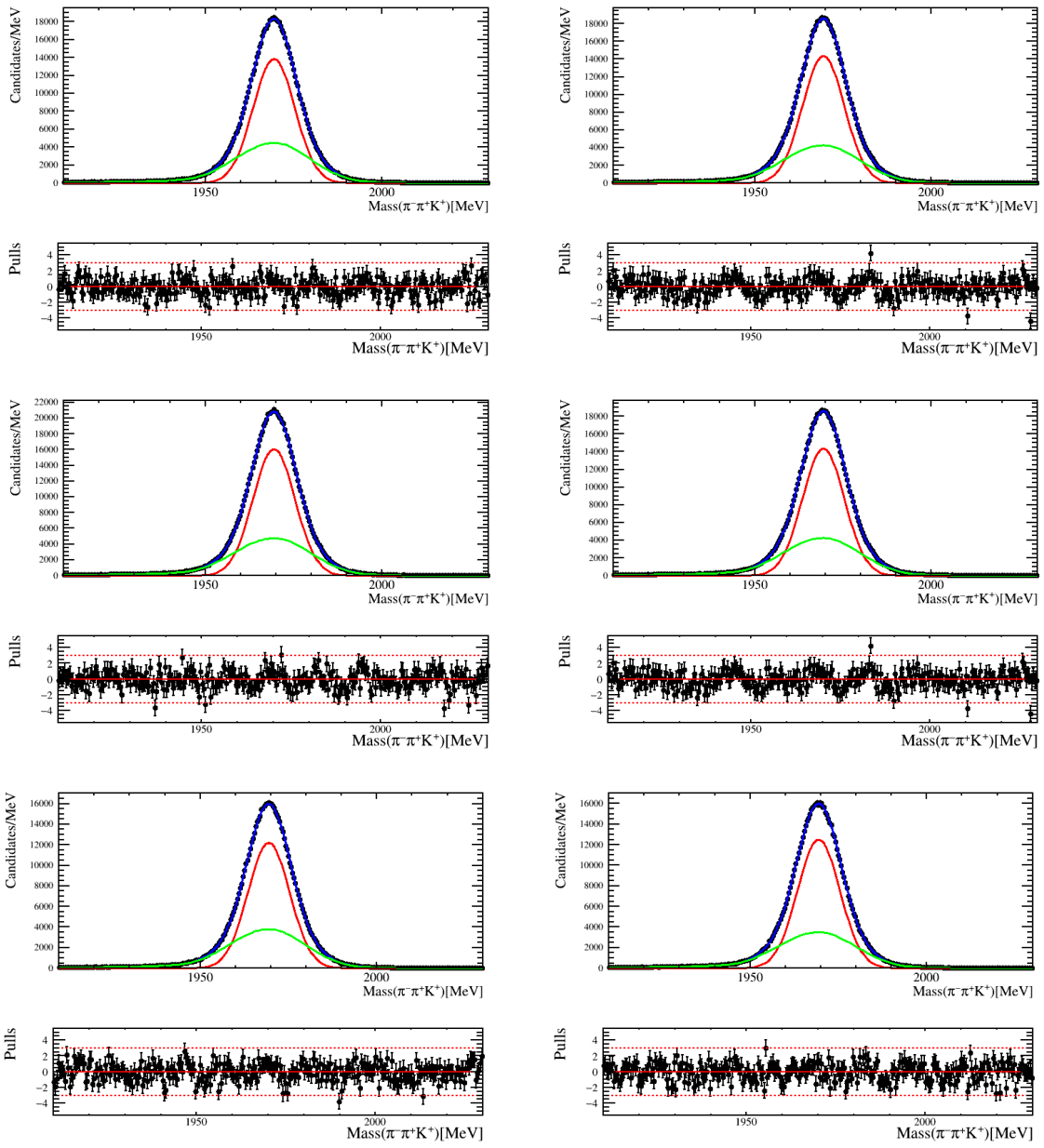

Figura 37 – Invariant mass fits for $D_s^+ \to \pi^-\pi^+K^+$decays for (left) MagUp and (right) MagDown from (top) 2016, (middle) 2017 and (bottom) 2018 MC samples. The signal components of the fit are a Gaussian (in red) Gaussian and Asymmetric Double Crystal Balls (in green); the combined PDF is displayed in blue.

The resulting data fits are presented in Fig. 38 and the fit parameters are shown in Tables 12 and 13.

Parameter	2016	2017	2018
μ	1969.08 ± 0.01	1969.01 ± 0.01	1969.25 ± 0.01
σ	6.344 ± 0.005	6.175 ± 0.004	6.146 ± 0.004
a_0	1.106 ± 0.002	1.107 ± 0.003	1.108 ± 0.003
a_1	1.015 ± 0.004	1.028 ± 0.004	1.027 ± 0.004
a_2	1.053 ± 0.005	1.050 ± 0.005	1.041 ± 0.005

Tabela 12 – Parameters extracted from the data MagUp sample fits

Parameter	2016	2017	2018
μ	1968.87 ± 0.01	1968.77 ± 0.01	1969.08 ± 0.01
σ	6.391 ± 0.005	6.244 ± 0.005	6.213 ± 0.05
a_0	1.102 ± 0.003	1.104 ± 0.003	1.106 ± 0.003
a_1	1.022 ± 0.004	1.029 ± 0.004	1.028 ± 0.004
a_2	1.048 ± 0.005	1.047 ± 0.005	1.042 ± 0.004

Tabela 13 – Parameters extracted from the data MagDown sample fits

The signal yields, their uncertainties, as well as the purity are summarized in Table 14. The Dalitz plot for this channel after the selection is shown in Fig. 39, this Dalitz plot is plotted with the squared invariant masses calculated with the momenta corrected by Decay Tree Fittter with mass constraint (DTF). In the Fig. 39 we can identify some resonances expected for this channel like $K^*(892)$, $K^*(1410)$ and/or $K^*(1430)$ on S($\pi^- K^+$) and $\rho(770)$ and $f^0(980)$ on S($\pi^-\pi^+$).

Sample	Signal yields	Purity
2016 Up	$(4.691 \pm 0.004) \times 10^6$	68.00%
2016 Down	$(4.837 \pm 0.004) \times 10^6$	67.63%
2017 Up	$(5.002 \pm 0.004) \times 10^6$	70.45%
2017 Down	$(5.166 \pm 0.004) \times 10^6$	70.36%
2018 Up	$(5.886 \pm 0.004) \times 10^6$	71.02%
2018 Down	$(5.297 \pm 0.004) \times 10^6$	69.70%

Tabela 14 – Signal yields, their uncertainties and the purity of all the samples after the selection criteria.

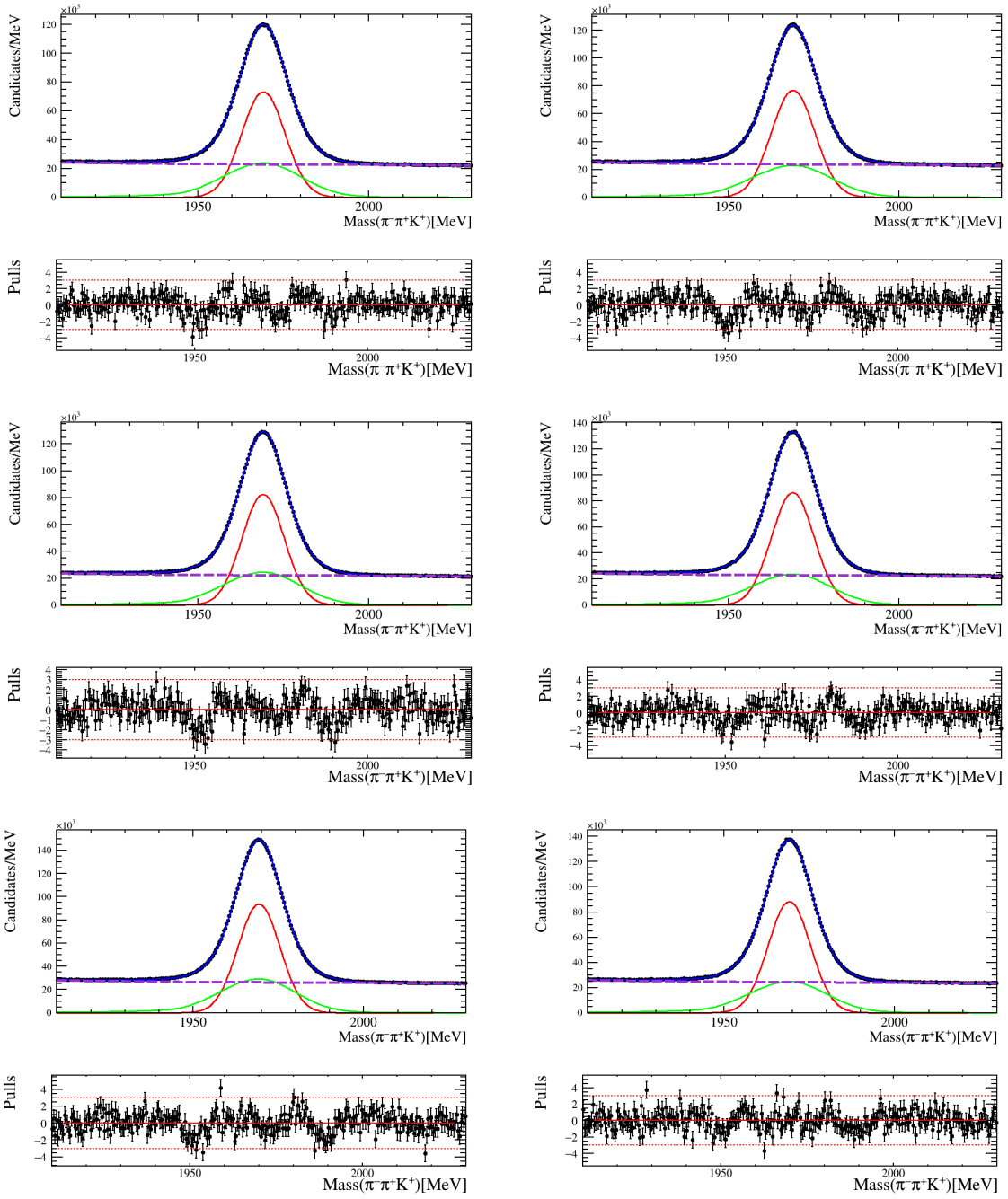

Figura 38 – Invariant mass fits for $D_s^+ \to \pi^-\pi^+K^+$candidates for (left) MagUp and (right) MagDown from (top) 2016, (middle) 2017 and (bottom) 2018 data samples. The signal components of the fit are a Gaussian (in red) Gaussian and Asymmetric Double Crystal Balls (in green); the background is represented in purple, and the combined PDF is displayed in blue.

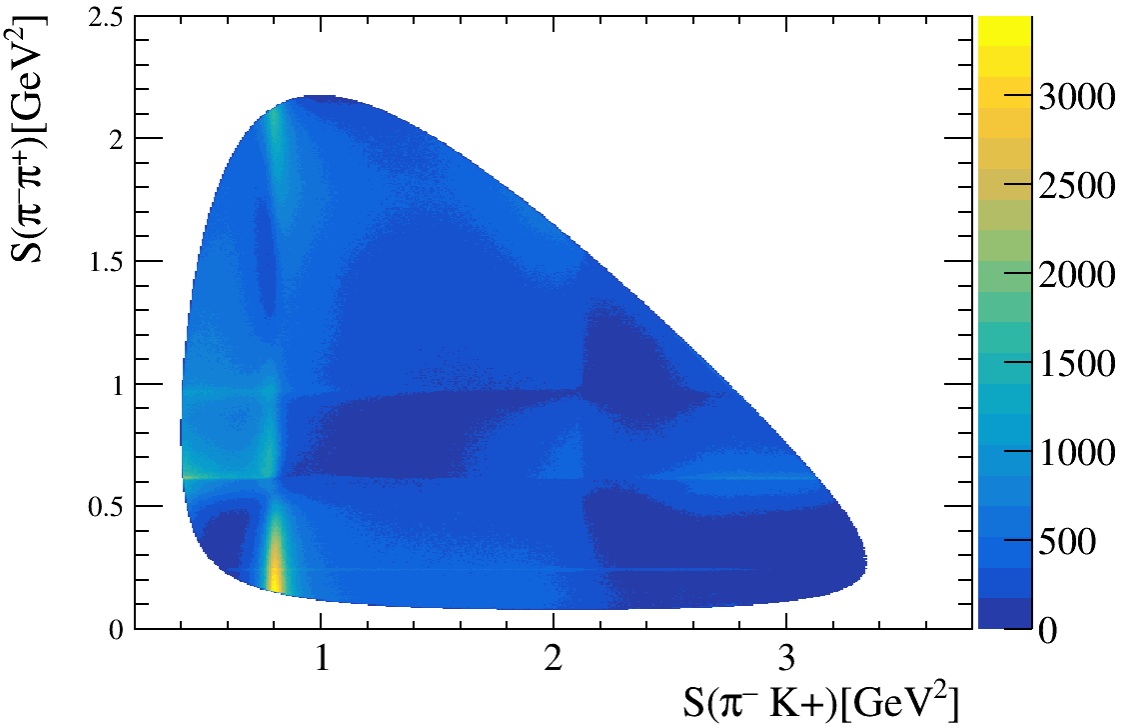

Figura 39 – Dalitz plot for $D_s^+ \to \pi^-\pi^+K^+$ after the selection.

5 Search for CP Violation in $D_s \to \pi\pi K$

5.1 Analysis strategy: Miranda Technique

Once the candidates have been selected, this sample can be used for analysis. It is desirable that the chosen technique operates in the Dalitz plot, since studies of CP violation in three-hadron decays of B mesons indicate that asymmetries in phase space tend to be more pronounced than those obtained from integrated observables [69, 36]. In addition, the interference between resonances that populate and overlap across the Dalitz plot can provide additional sources of strong-phase differences, enhancing the CP-violation effects in three-body decays.

In this analysis, the mirandizing method is employed, which was developed by researchers from CBPF in collaboration with the University of Notre Dame, USA [70, 71, 72]. This technique is a model-independent method that allows mapping the significance of the asymmetry distribution across the Dalitz Plot (DP). The combined DP of $D_s^{\pm}$ is divided into bins, and the observable S^i_{CP}, which represents the statistical significance of the difference for the population D_s^+ and D_s^- in each bin i, is calculated as:

$$S^i_{CP} = \frac{N_i^+ - \alpha N_i^-}{\sqrt{(\delta N_i^+)^2 + \alpha(\delta N_i^-)^2 - 2\alpha\sigma_{N_i^+ N_i^-}}}, \quad \alpha = \frac{N_{tot}(D_s^+)}{N_{tot}(D_s^-)} \ . \tag{5.1}$$

Here, N_i^+ and N_i^- represent the number of signal candidates of D_s^+ and D_s^- in the i-th bin, respectively. The quantities δN_i^+ and δN_i^- denote the uncertainties associated with N_i^+ and N_i^-, respectively. The constant α is introduced to remove global net asymmetries, while $N_{tot}(D_s^{\pm})$ represent the total number of $D_s^{\pm}$ integrated over the DP. Additionally, $\sigma_{N_i^+ N_i^-}$ accounts for the correlation between $N^i(D_s^+)$ and $N^i(D_s^-)$, and $N_{tot}(D_s^{\pm})$.

Once the S^i_{CP} is computed in each bin, we can define a χ^2 test as:

$$\chi^2 = \sum_{N_{bins}} (S^i_{CP})^2$$

From this test statistic, a p-value is obtained considering the number of degrees of freedom ndof $= N_{bins} - 1$. In the absence of CP violation, the S^i_{cp} values are expected to follow a standard normal distribution. Deviations from this behavior may indicate signs of CP violation. A p-value of 3×10^{-7} or less corresponds to an observation of CP violation with at least 5σ confidence level, rejecting the CP symmetry hypothesis.

As can be observed in CP-violation search publications from Run 1 the approach traditionally used in the mirandizing method consists of simply counting the candidate

populations in each Dalitz-plot bin for particles and antiparticles to compute the S_{CP} values[1].

The main drawback of this approach is that it also counts background events present in each bin. These background contributions can introduce asymmetry effects in those regions and additionally increase the uncertainties of the analysis. This counting-based method may not be sufficiently sensitive to measure such an effect, and could even confuse it with possible asymmetries originating from the background.

Therefore, to ensure that this work is not affected by these issues, the particle and antiparticle populations used to compute the S_{CP} values are obtained through invariant-mass fits to the $\pi^-\pi^+K^+$ spectrum in each Dalitz-plot bin. The advantage of this approach is that it minimizes background effects, allowing the results obtained with this method to be significantly more reliable than those based solely on event counting.

As the goal of this work is to identify and map asymmetries across the Dalitz plot, the fit model includes the raw asymmetry and the charge combined yield as fit parameters. The raw asymmetry, A_{raw}, and the total, charge-combined, number of events, N_{tot}, are defined as:

$$\begin{aligned} A_{raw} &= \tfrac{N_i^+ - N_i^-}{N_i^+ + N_i^-} \\ N_{tot} &= N_i^+ + N_i^- \end{aligned} \tag{5.2}$$

The choice of these parameters, as well as the fitting procedure itself, will become clearer later when the per-bin mass fits are discussed.

However, adopting these parameters in the PDF model causes the correlation term obtained from the fit to no longer represent the correlation between N^+ and N^-, but instead the correlation between the raw asymmetry, A_{raw}, and the total number of events, N_{tot}. For this reason, it is necessary to adapt the expression for S_{CP} so that it can be computed in terms of the raw asymmetry and the total number of events:

$$S_{CP}^i = \frac{\frac{N_{tot}}{2}[\bar{A}^i]}{\sqrt{[\bar{A}^i]^2\frac{(\delta N_{tot}^i)^2}{4} + [\bar{B}^i]^2\frac{(\delta A_{raw}^i)^2}{4} + [\,\bar{A}^i\,][\bar{B}^i\,]\frac{\sigma(N_{tot}^i, A_{raw}^i)}{2}}}, \tag{5.3}$$

where

$$\begin{aligned} \bar{A}^i &= [(1 + A_{raw}^i) - \alpha(1 - A_{raw}^i)] \\ \bar{B}^i &= N_{tot} * (1 + \alpha) \end{aligned} \tag{5.4}$$

with δN_{tot}^i and δA_{raw}^i being the uncertainties of A_{raw}^i) and N_{tot}^i and $\sigma(N_{tot}^i, A_{raw}^i)$ the covariance between A_{raw}^i and N_{tot}^i.

[1] In this case, the S_{CP} observable takes a simpler form, since the correlation term in this approach is zero.

The observable defined in Eq. 5.3 is the one that will be used in the CP-violation search strategies, relying exclusively on the fit results obtained in the signal region. For studies in the background region and for investigations of asymmetries, a different approach will be adopted; this will be described and motivated together with the studies presented in chapter 6.

5.1.1 Binning scheme

Since the mirandizing method is a binned analysis technique, the choice of the binning scheme is of crucial importance. The most basic binning models one might consider are uniform binning and adaptive binning, which divide the Dalitz plot into bins of equal size and bins with equal population, respectively.

However, neither of these binning schemes adapts well to the geometry assumed by resonant structures in the Dalitz plot. Therefore, this work employs a third binning strategy, referred to as physical binning. For this analysis, the physical binning adopted was constructed by taking into account two main aspects:

1. Exploit the resonant structure of the $D_s^{\pm} \to \pi^-\pi^+K^+$ decay, in particularly starting the bin division by splitting the main spin-1 resonances ($K^*(892)^0$ and $\rho(770)^0$) at the nominal mass and at the node of the angular distribution;

2. Define the bins to have similar signal yields, not varying more than a factor of two.

To define the number of bins used in the analysis, this work was guided by previously published [73, 74] results indicating that the best sensitivity is achieved with approximately 20 bins. Using fewer bins may lead to excessive integration over localized asymmetries, thereby diluting such effects. On the other hand, adopting significantly more than 20 bins reduces the population in each region, decreasing the statistical precision per bin and, consequently, diminishing the overall sensitivity.

The size of each bin was defined using fits to the invariant mass $\pi^-\pi^+K^+$performed in each bin to estimate the signal yields while varying the bin boundaries, until a configuration was found in which all bins have a similar signal content. To perform these fits, the 2016–2018 samples were combined, and the PDF adopted for this procedure is the same as that described in Eq. 4.4–4.6.

The resulting baseline binning scheme, referred to as the physics-motivated binning, is shown in Fig. 40.

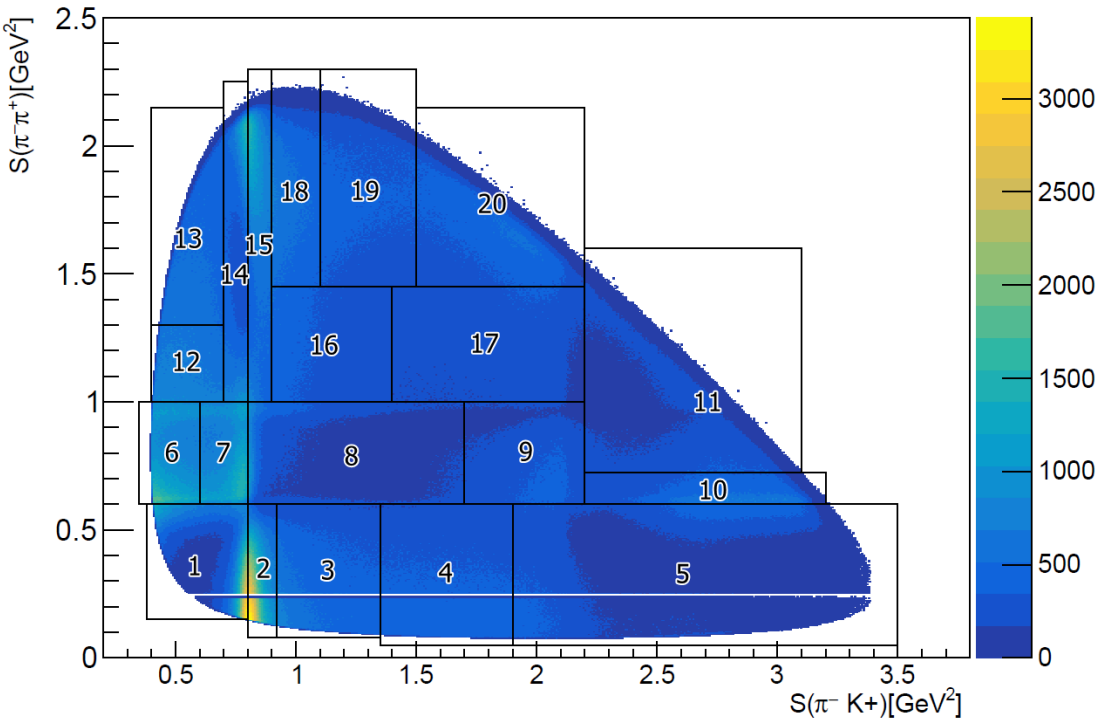

Figura 40 – Binning scheme physics motivated for $D_s^+ \to \pi^-\pi^+K^+$.

Fig. 41 presents examples of fits in three different bins of the Dalitz plot, illustrating how the background varies across distinct regions of the DP. all the others fits used to estimate number of events in the binning scheme can be found on appendix A.

5.2 Fit procedure

The main strategy to obtain the raw asymmetries and the total number of events used to compute S_{CP} is based on extracting these parameters separately for the MagUp and MagDown samples, using a simultaneous of the binned $\pi^-\pi^+K^+$ invariant-mass distributions for D_s^+ and D_s^-.

For the fitting, a simultaneous fit to four subsamples is adopted: MagUp D_s^+, MagUp D_s^-, MagDown D_s^+ and MagDown D_s^-. This approach allows some of the fit parameters to be shared among the subsamples, providing global consistency when compared to independent fits. Another important advantage of this strategy is that the raw asymmetry and the total number of events can be directly used as fit parameters between the D_s^+ and D_s^- samples.

For the fitting procedure, we first perform the simultaneous fit on the Monte Carlo subsamples to extract the signal PDF shape parameters. Both the MC and data fits are

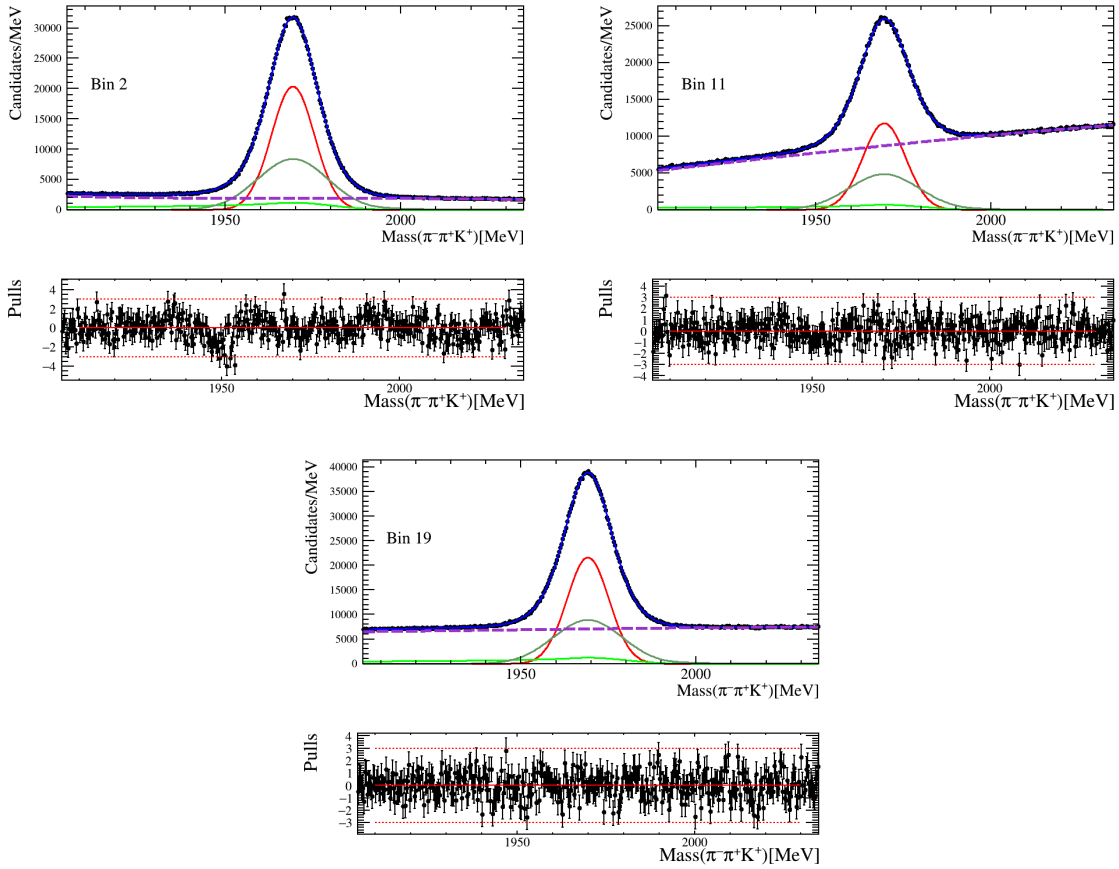

Figura 41 – $\pi^-\pi^+ + K^+$ invariant-mass fit for bins numbered (top left) 2, (top right) 11, and (bottom) 19. The components of the fit are displayed as follows: the Gaussian in red, the Crystal Balls in green, the Bernstein polynomial in purple, and the combined PDF in blue.

conducted for each bin in the physics-motivated binning.

The combined PDF is expressed as:

$$\mathcal{P}^{\pm\frac{U}{D}} = \frac{N_{tot}^{sig}}{2}(1 \pm A_{Raw}^{sig})^{\frac{U}{D}}\mathcal{P}_{sig}^{\pm\frac{U}{D}} + \frac{N_{tot}^{bkg}}{2}(1 \pm A_{Raw}^{bkg})^{\frac{U}{D}}\mathcal{P}_{bkg}^{\pm\frac{U}{D}} \ . \tag{5.5}$$

where $N_{tot}^{sig/bkg}$ is the total number of signal/background candidates and $A_{raw}^{sig/bkg}$ the raw asymmetry for the signal/background. The signal PDF, $\mathcal{P}_{sig}$, is defined as the sum of a Gaussian (G) PDF and an Asymmetric Double Crystal Ball (ADCB) PDF.

$$\mathcal{P}_{sig}(m) = f_G \times G(\mu, \sigma_G) + (1 - f_G) \times ADCB(\mu + \Delta\mu, R_1\sigma_G, R_2\sigma_G, \alpha_1, N_1, \alpha_2, N_2). \tag{5.6}$$

Here, μ and σ_G represent the mean value and width of the Gaussian component, respectively. The Asymmetric Double Crystal Ball (ADCB) PDF is defined by parameters α_1, N_1, α_2 and N_2 which characterize the left and right tails of the Crystal Ball function. Additionally, $R_1\sigma_G$ and $R_2\sigma_G$ represent the widths of the ADCB. An offset $\Delta\mu$ is introduced to account

for the shift in the most probable mean of the ADCB relative to the Gaussian mean. The fractions of each PDF component are f_G for the Gaussian, $(1 - f_G)$ for the ADCB.

The combinatorial background PDF is described by a third-order Bernstein polynomial (n=3), expressed in eq. 5.7, with a_3 fixed in 1.

$$\mathcal{P}_{bkg}(m) = \sum_{i=0}^{n} a_i \binom{n}{i} m^i \cdot (1-m)^{n-1} \tag{5.7}$$

5.2.1 MC fits

The Monte Carlo fit is performed using only the signal PDF, since Monte Carlo samples do not contain background contributions. The goal is to extract the signal shape for each of the four subsamples in order to study which shape parameters can be fixed to achieve the best description of the data in the four corresponding data subsamples.

Some parameters are fixed for each bin and year based on the results from the MC fits. Specifically, the tail parameters of the ADCB, the relative fraction between the Gaussian and the ADCB, and the ratio of widths between the Gaussian and the ADCB are kept fixed.

The parameters allowed to vary during the fit include the Gaussian mean, its sigma, the raw asymmetry, and the offset (with its range adjusted to align with the uncertainties observed in the MC fit for each bin and year), as well as the Bernstein polynomial parameters.

All fits performed on the Monte Carlo samples follow strict quality criteria. The numerical minimizer adopted is Minuit, and all fits in this analysis are required to have status = 0 and covQual = 3. The first condition is a convergence criterion of the fit: a status value of 0 indicates that the minimizer has successfully converged mathematically. The second is an indicator of the quality of the covariance matrix and reflects whether Minuit was able to correctly estimate the uncertainties and correlations among the fit parameters; a value of 3 indicates that the covariance matrix was fully determined and is positive definite.

In addition to these criteria, it is required that the fit pulls lie within the range $[-3, 3]$, ensuring statistical compatibility between the fitted model and the data.

Figure 42 shows the fit result for bin 11 of the 2016 data sample, which occupies a region of the DP without any resonant structures. The fit for bin 2, in the region of interference between $K^*(892)$ and $\rho^0(770)$, is shown in Fig. 43, for the 2018 sample.

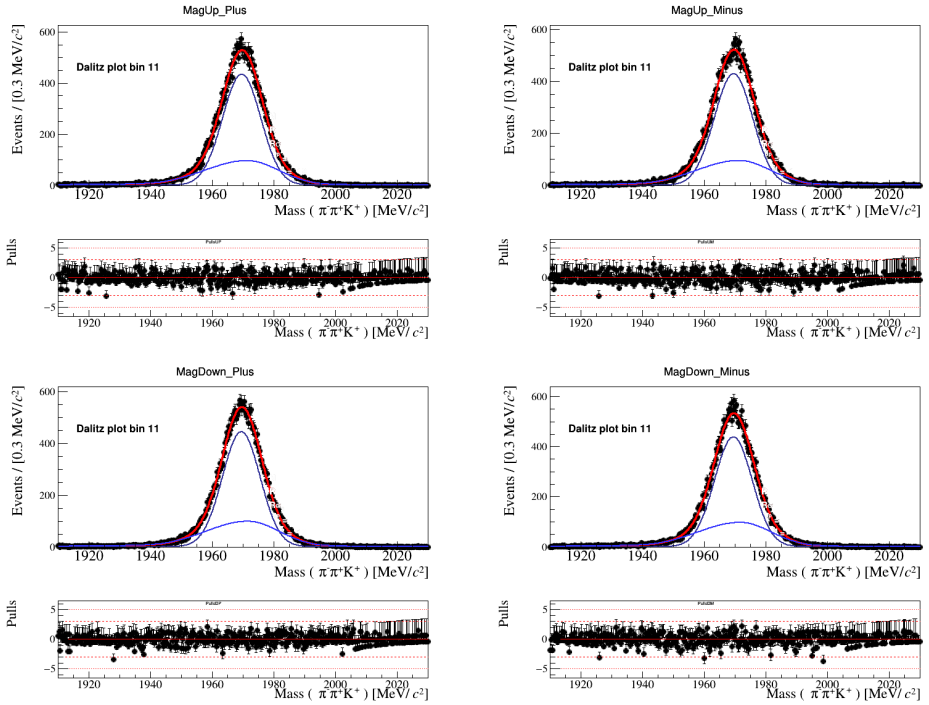

Figura 42 – Invariant mass fit of bin 11 for the 2016 $D_s^+ \to \pi^-\pi^+K^+$ MC sample, showing for projections: D_s^+ Magnet Up (top left), D_s^- Magnet Up (top right), D_s^+ Magnet Down (bottom left) and D_s^- Magnet Down (bottom right).

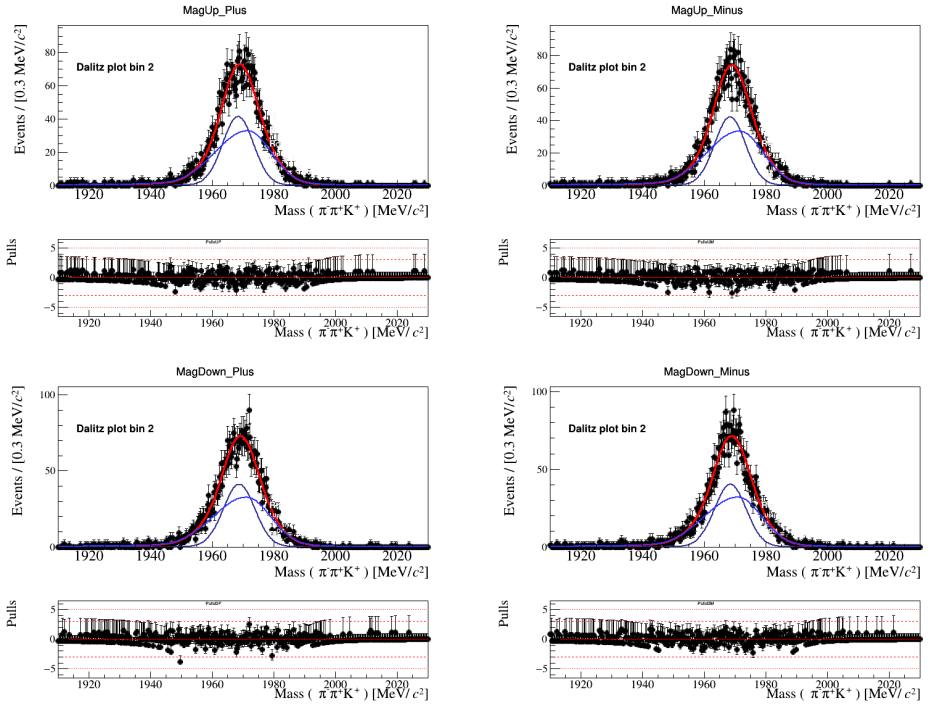

Figura 43 – Fit of bin 2 (2018 sample) for the $D_s^+ \to \pi^-\pi^+K^+$ MC sample, near the interference region between $K^*(892)$ and $\rho^0(770)$.

All the parameters obtained from MC used on data fits can be found on Appendix B. A comprehensive collection of all MC fits is available in Appendix C.

5.2.2 Data fits

For the data fit, the process follows a similar methodology, incorporating the background model described at the beginning of this section. At this stage, the analysis remains blinded, and the raw asymmetry and signal yields per charge are not disclosed at the conclusion of the procedure.

As explained earlier, the parameters fixed in each bin and year are based on results obtained from the Monte Carlo fits, adopting the model that showed the best performance among the different possible combinations of fixed parameters. In particular, the tail parameters of the ADCB, the relative fraction between the Gaussian and the ADCB components, and the ratios between the Gaussian and ADCB widths were fixed.

The parameters left free to vary during the fit include the Gaussian mean, its width, the raw asymmetry, the offset between the Gaussian and the ADCB components, as well as the parameters of the Bernstein polynomial. The offset between the Gaussian and the ADCB, in particular, had its allowed range constrained to be compatible with the uncertainties obtained from the Monte Carlo fits for each bin and year.

Once again, Fig. 44 (bin 1 for 2016 sample), 45 (bin 11 for 2017 sample) and 46 (bin 19 for 2018 sample) present three examples of these fits. These examples illustrate how background levels and shapes vary across different regions of the Dalitz plot, as well as how effectively the Bernstein polynomial captures and adapts to these variations.

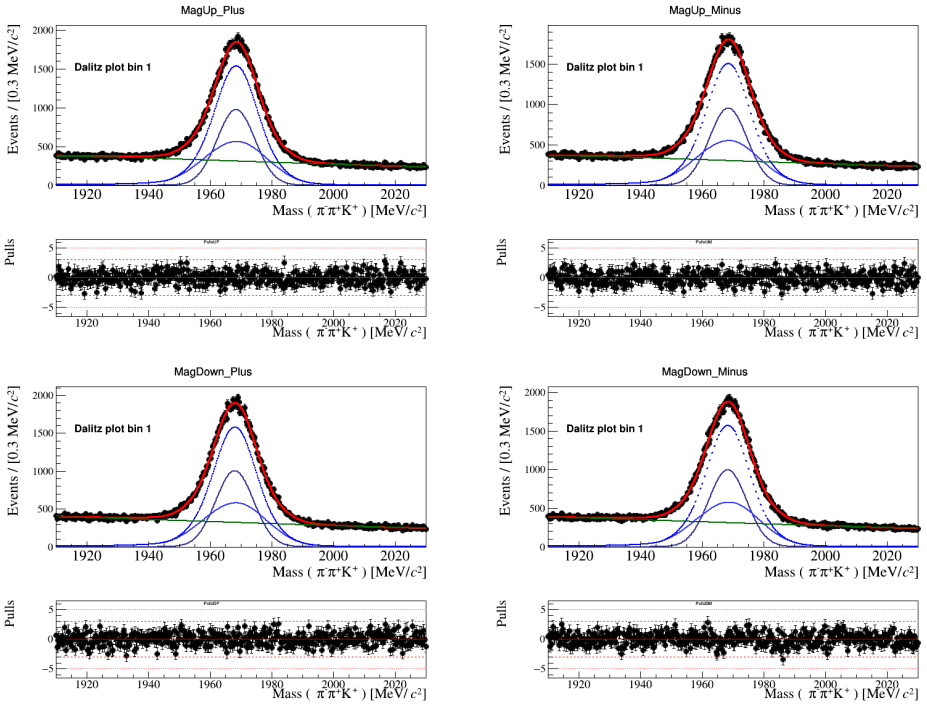

Figura 44 – Fit of bin 1 (2016 sample) for the $D_s^+ \to \pi^-\pi^+K^+$ Data sample, near the interference region between $K^*(892)$ and $\rho^0(770)$.

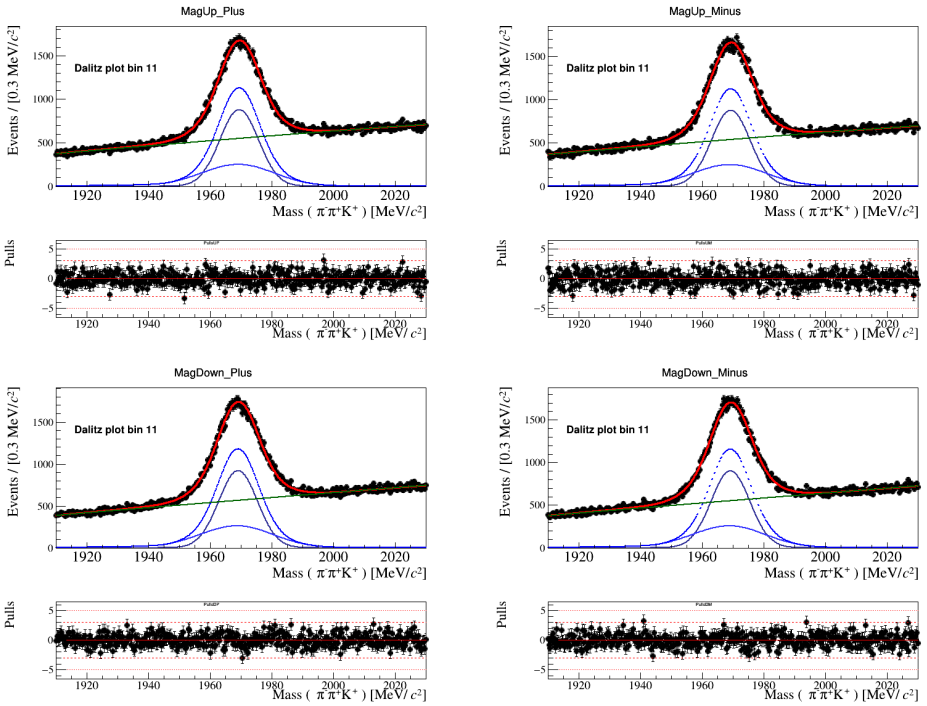

Figura 45 – Fit of bin 11 (2017 sample) for the $D_s^+ \to \pi^-\pi^+K^+$ Data sample in a region without any structure.

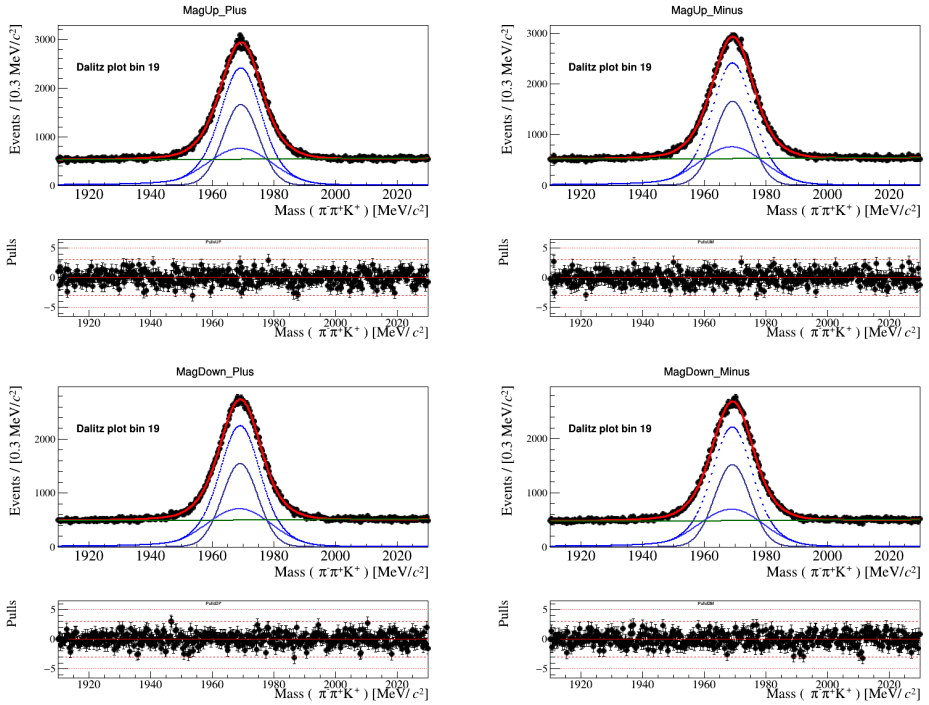

Figura 46 – Fit of bin 19 (2018 sample) for the $D_s^+ \to \pi^-\pi^+K^+$ Data sample, near the region close $K^*(892)$ lobe.

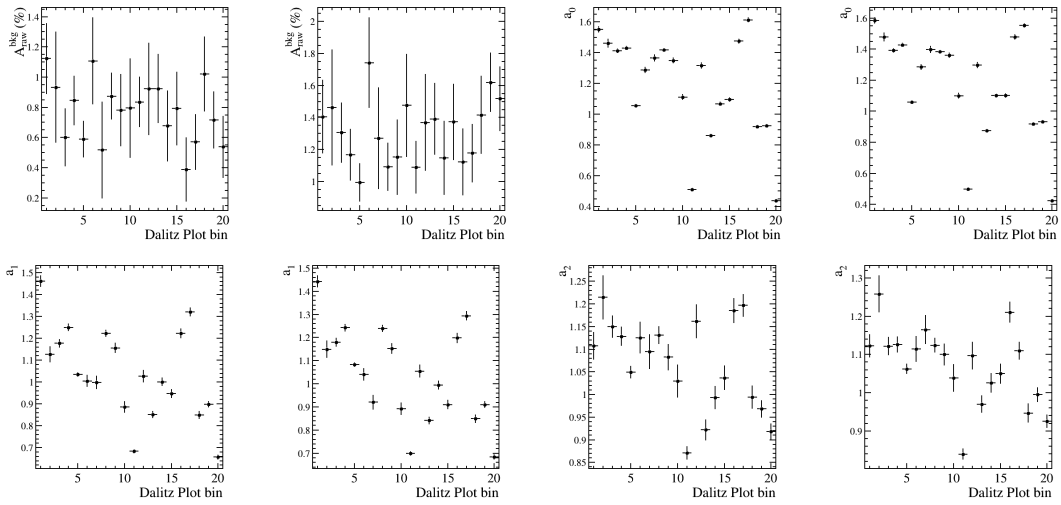

Figura 47 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2016 Data samples. In the first row, we present the A_{RAW}^{bkg} and the offset between the Gaussian and the ADCB. The second and third rows contain the Bernstein polynomial coefficients. The right plots correspond to the results for the MagUp samples, while the left plots correspond to the MagDown samples.

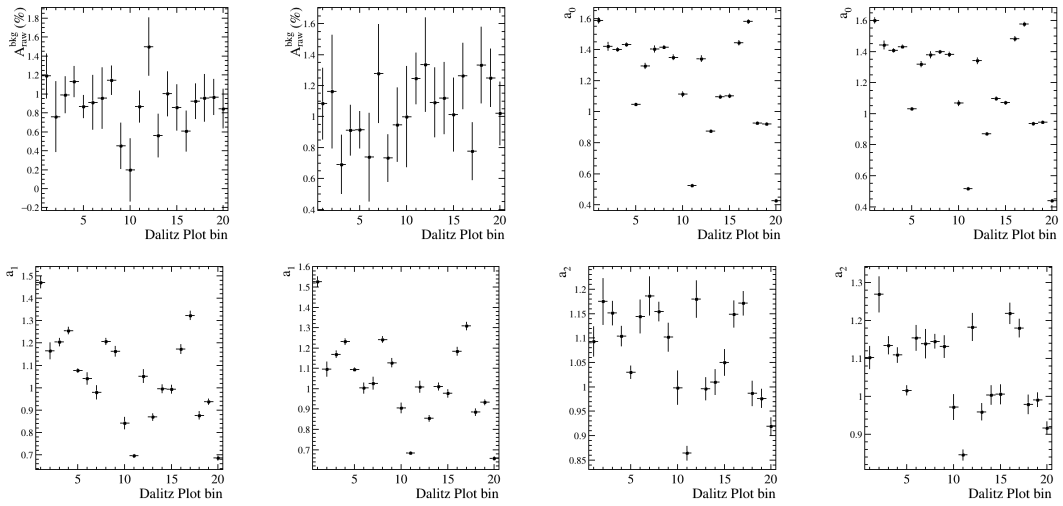

Figura 48 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2017 Data samples. In the first row, we present the A_{RAW}^{bkg} and the offset between the Gaussian and the ADCB. The second and third rows contain the Bernstein polynomial coefficients. The right plots correspond to the results for the MagUp samples, while the left plots correspond to the MagDown samples.

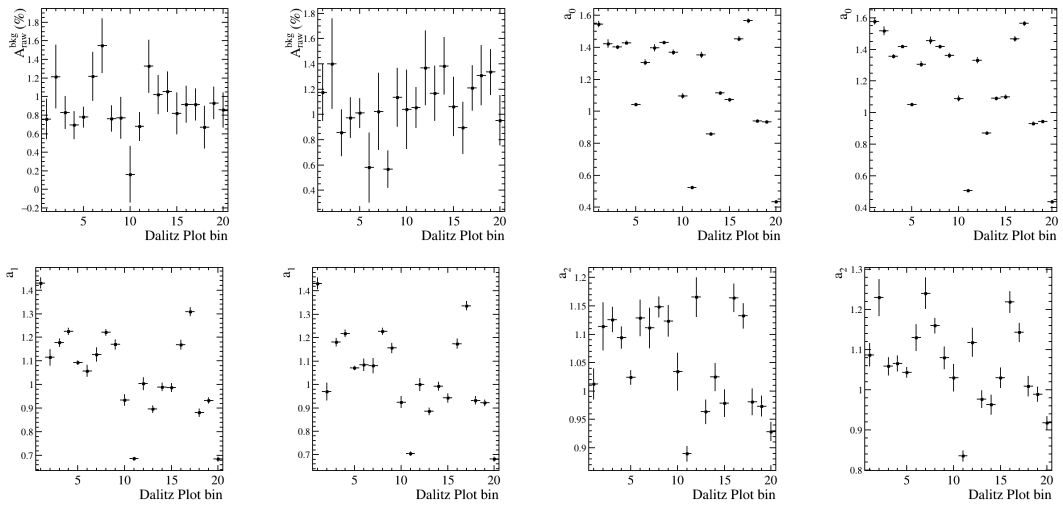

Figura 49 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2018 Data samples. In the first row, we present the A_{RAW}^{bkg} and the offset between the Gaussian and the ADCB. The second and third rows contain the Bernstein polynomial coefficients. The right plots correspond to the results for the MagUp samples, while the left plots correspond to the MagDown samples.

The distributions of all parameters obtained from our fits are displayed in the Fig. (47 -49). All fits conducted in this analysis are provided in the appendix C of this note.

5.3 Validations

5.3.1 Fit validation

Since the same PDF model is adopted for the fits in all bins, it is necessary to verify that these models are consistent and free from biases. Ideally, this validation would be performed by testing the fit model on multiple independent data samples with the same characteristics as those used to construct it, in order to identify possible deviations from the expected behavior. However, there are no independent data sets that reproduce the structure used to build the fit model in each bin. Therefore, the model validation is performed using pseudo-experiments (toys) generated from the fitted model itself, with the goal of checking for biases in the extracted yields and assessing whether the uncertainties associated with the fit parameters are correctly estimated.

To generate these pseudo-experiments, distributions are produced based on the original data set, using the information from the original sample together with the fitted model. The only difference with respect to the original distributions is that the number of events used to populate them is allowed to fluctuate according to a Poisson distribution. In total, 500 toy experiments are generated and fitted for each bin and each year, in order to test the consistency of the fit model across all distributions.

These toy samples are generated using all fit parameters, with the exception of the signal and background raw asymmetries, which are fixed to zero. This choice is motivated by the fact that the analysis is still in a blinded stage, and the signal asymmetry in particular is not an observable that can be accessed at this point.

To assess the presence of possible biases in these parameters, the pull distributions of the signal and background raw asymmetries are computed for each bin.As shown in Fig. 50, the pull distributions of A_{raw} and A_{raw}^{bkg} are well-behaved and can be accurately fitted with Gaussian functions. The Gaussian mean and width are displayed in Figures 51 and 52 as a function of the Dalitz plot bin for all subsamples. These values are consistent with the expected results for all bins and years analyzed, 0 for the mean and 1 for sigma.

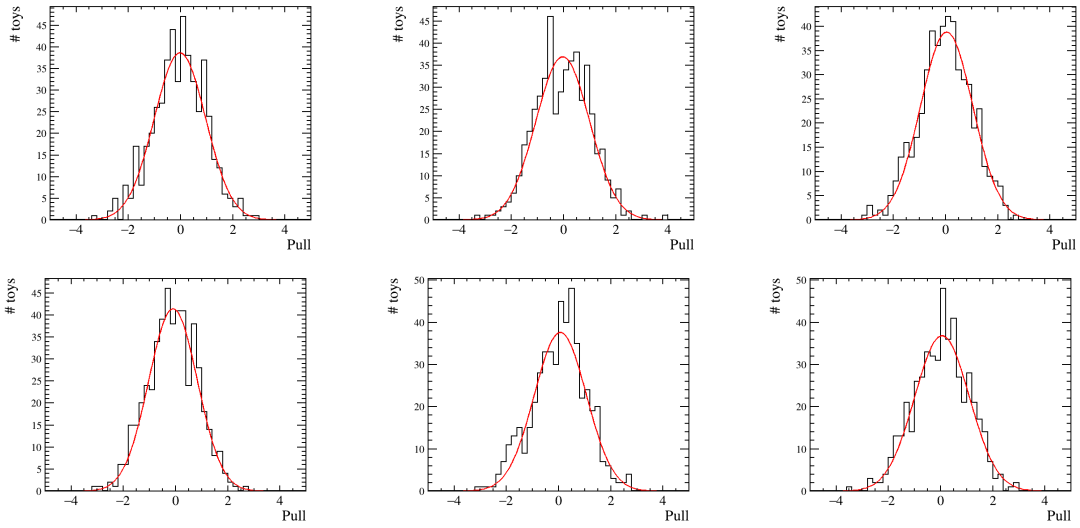

Figura 50 – Distribution of pulls for A_{raw} (top plots) and A_{raw}^{bkg} (botton plots) for 2016, 2017 and 2018 Data for the bin 1.

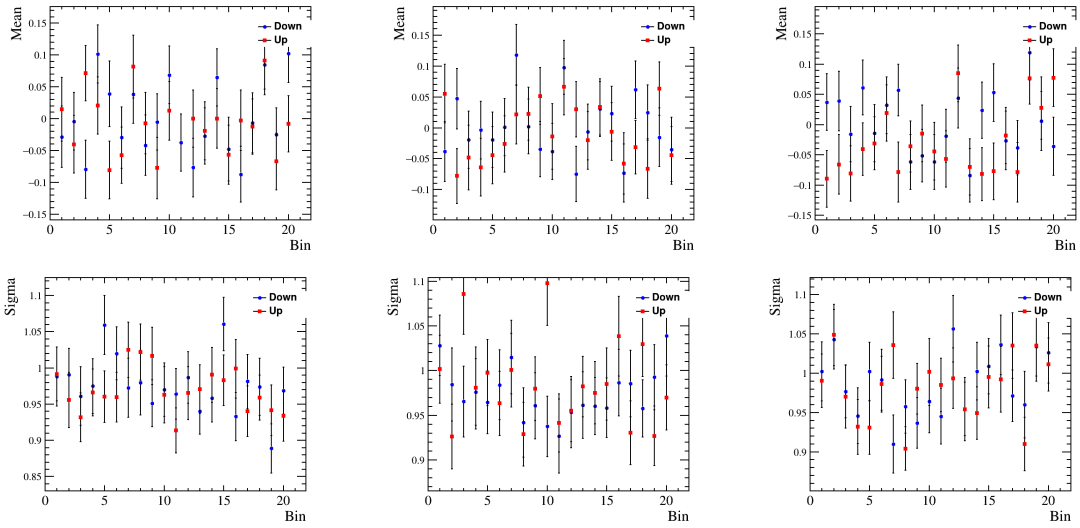

Figura 51 – The Gaussian mean (top plots) and width (botton plots) for the A_{raw} as a function of the Dalitz plot bin for up (blue) and Down (black) for 16, 17 and 18 samples.

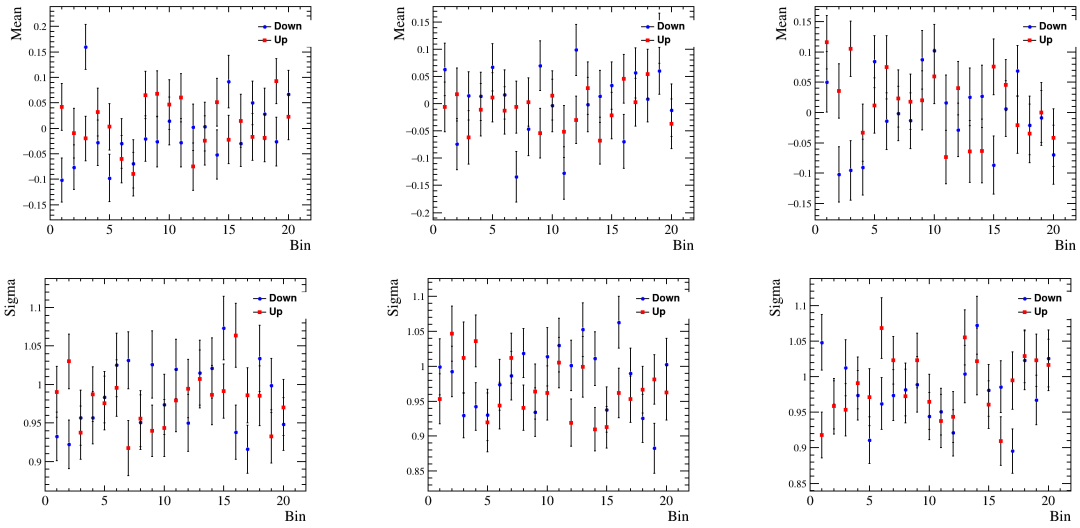

Figura 52 – The Gaussian mean (top plots) and width (botton plots) for the A_{raw}^{bkg} as a function of the Dalitz plot bin for up (blue) and Down (black) for 16, o17 and 18 samples

From this study, we conclude that there is no evidence of bias in our fits or any inconsistencies in the uncertainty estimations. This demonstrates that the fits performed on this sample are robust and reliable.

5.3.2 observable validation

Another important aspect to be validated is whether the observable defined in Eq. 5.3 is also unbiased. For this purpose, the study makes use of the results previously obtained from the toys used to validate the PDF employed in the fits. The information from each toy is used to compute the S_{CP} observable and its corresponding p-value on a toy-by-toy basis.

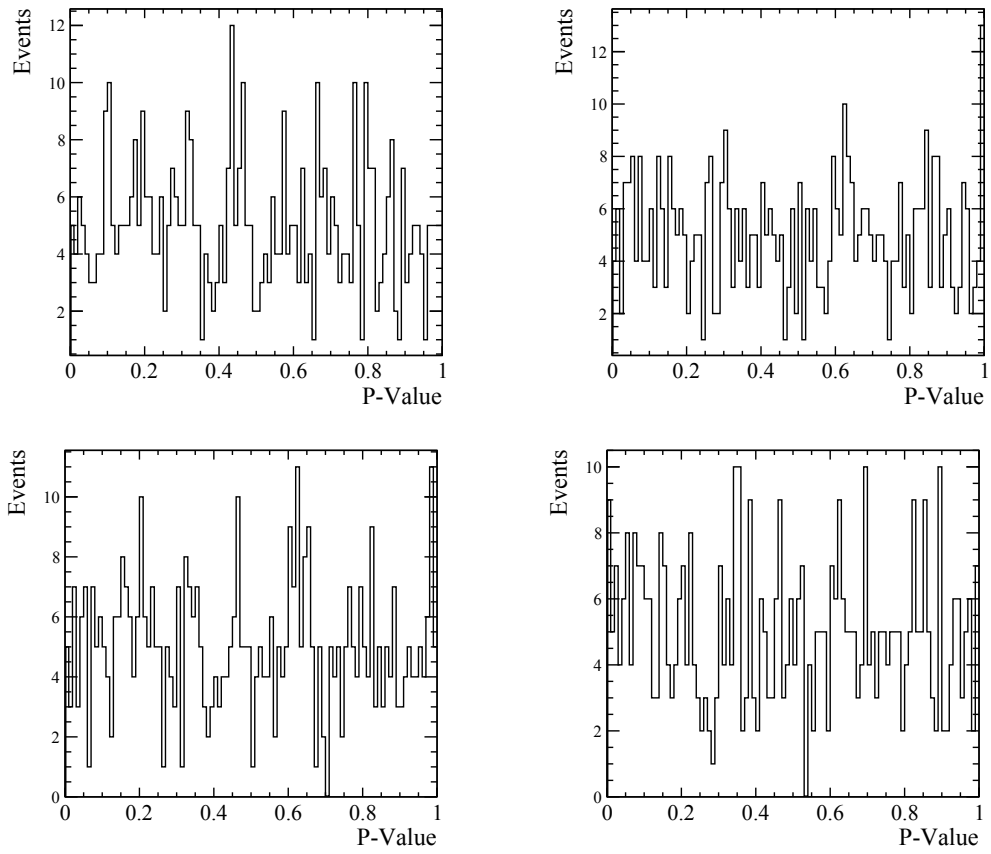

Figura 53 – P-value distribution for the observable validation with 500 toys. On top the resutls for 2016 and for 2017 sample, on botton the reesult for 2018 and for all sample .$Mean_{16} = 0.49 \pm 0.01$, $Mean_{17} = 0.50 \pm 0.01$, $Mean_{18} = 0.50 \pm 0.01$ and $Mean_{all} = 0.49 \pm 0.01$

In the absence of asymmetries, as imposed in the generation of these toys, the p-values are expected to fluctuate uniformly between 0 and 1 across all toys. The resulting p-value distribution obtained from this study is shown in the figure 53. This behavior is consistent with the original hypothesis, leading to the conclusion that the observable is not biased.

5.4 Background Studies

Once the strategy for studying the Dalitz plot of the signal candidate events has been defined, it is possible to proceed with the study of the background region and to investigate one of the potential sources of asymmetry in this analysis. As described in Chapter 4, the background may still contain a residual contribution from contamination of the $\bar{D}^0 \to \pi^0\pi^-K^+$ meson after the selection, in addition to several other effects that may bias the measurement, even within the adopted technique.

In this way, an additional understanding of the presence of asymmetries in the background region and of their impact on the analysis technique is important to ensure that all possible sources of asymmetry do not interfere with the final result of this analysis.

A study aimed at testing these conditions is performed using the sidebands of the $\pi^-\pi^+K^+$invariant mass distribution. The purpose is to probe this region using the mirandizing technique in order to assess its sensitivity to possible asymmetries. Since these regions do not require signal–background separation and are composed purely of background events, it is not necessary to apply the observable as defined in eq. 5.3 on last Chapter. That observable was designed to compute S_{CP} through background subtraction obtained from invariant mass fits. Consequently, this study allows the use of a simplified version of the S_{CP} observable:

$$S^i_{CP} = \frac{N_i^+ - \alpha N_i^-}{\sqrt{\alpha((\delta N_i^+)^2 + (\delta N_i^-)^2)}}, \quad \alpha = \frac{N_{tot}(D_s^+)}{N_{tot}(D_s^-)}, \tag{5.8}$$

Here, N_i^+ and N_i^- represent the number of entries of $\pi^-\pi^+K^+$ and $\pi^+\pi^-K^-$ candidates in the sidebands, respectively, and the corresponding uncertainties are taken as their square roots. The factor α is, once again, employed to remove global net asymmetries. As before, the distribution of S^i_{CP} can be obtained and the p-value evaluated. In this approach the correlation term become 0.

The invariant mass ranges of the $\pi^-\pi^+K^+$ system adopted in this study are 1910–1930 MeV for the left sideband and 2010–2030 MeV for the right sideband. The mirandizing procedure is computed separately for each sideband and also for the combined sideband sample. And this asymmetry check is conducted for the combined samples from 2016 to 2018 as well as separately for each year.

As shown Figs. 54 and 55, no evidence of charge asymmetry is observed; all p-values are consistent with the null hypothesis and S_{CP} distributions are well behaved across the Dalitz plot. Thus, we can conclude that there is no sensitivity to any asymmetry coming from background candidates.

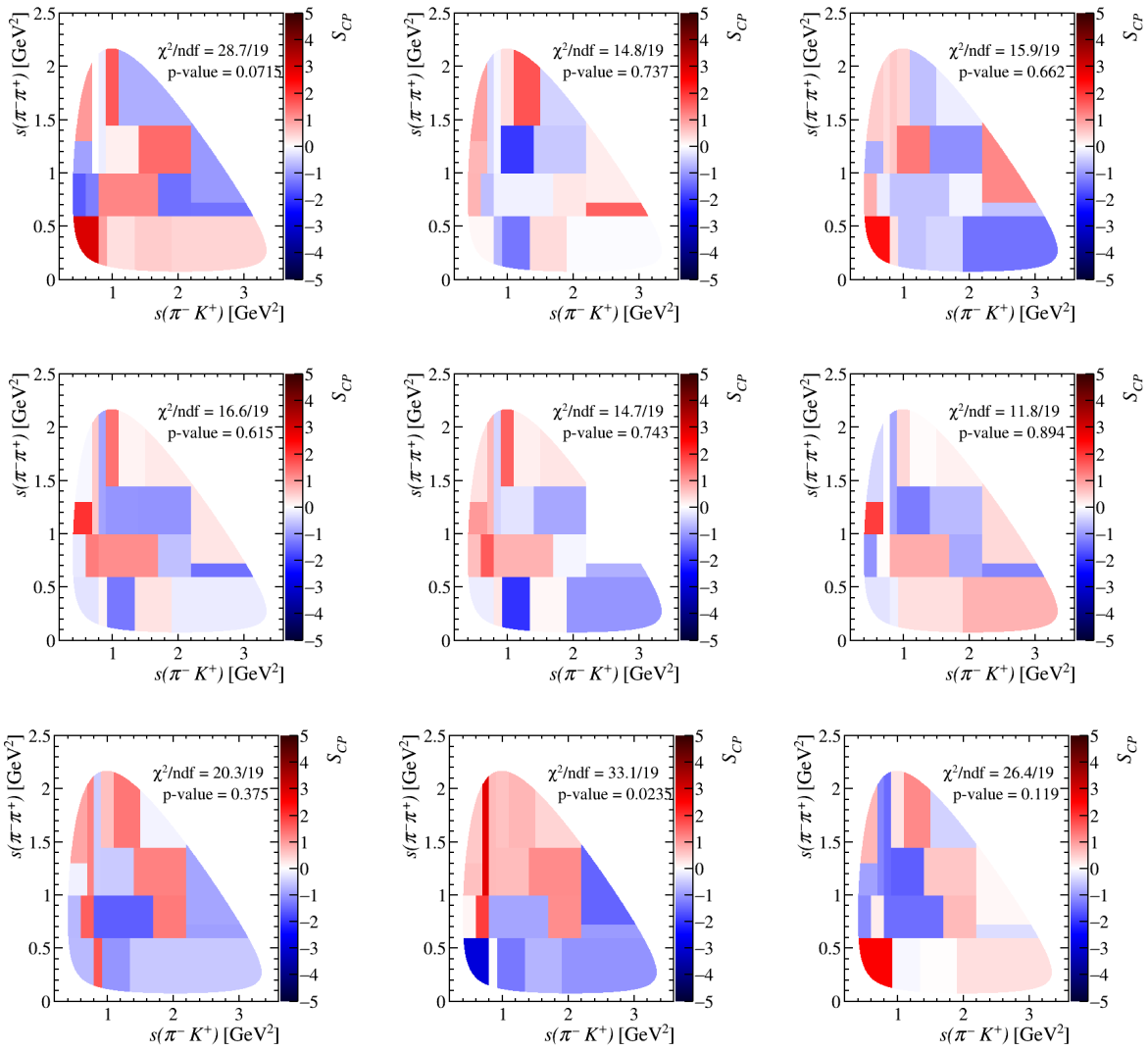

Figura 54 – S_{CP} distributions in the Dalitz plot for the mass sidebands of the (1^{st} row) 2016, (2^{nd} row) 2017 and (3^{rd} row) 2018 samples. Left plots show both sidebands combined, middle plots correspond to left sideband and right plots to right sideband.

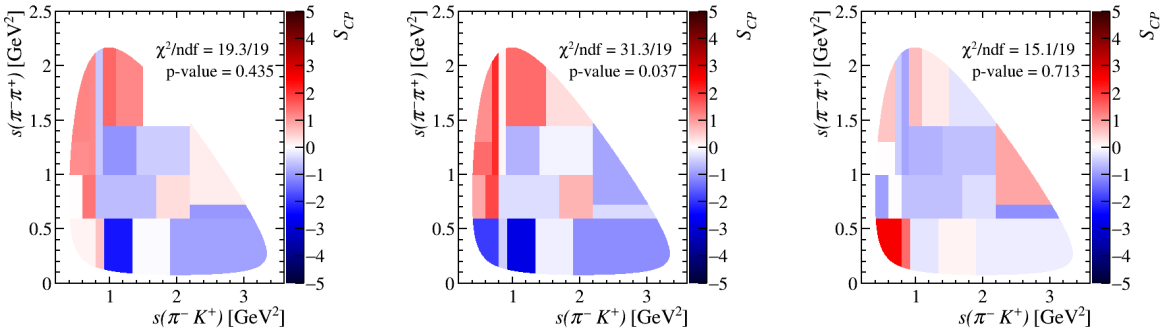

Figura 55 – S_{CP} distributions in the Dalitz plot for the mass sidebands of the combined samples 2016 – 2018. Left plots show both sidebands combined, middle plots correspond to left sideband and right plots to right sideband.

6 Systematic checks

Once the background studies indicate no asymmetries in that region that could influence the signal-region analysis, it remains necessary to investigate the possibility that the data sample may already contain asymmetries from other sources. These asymmetries are related to production and detection effects. A production asymmetry occurs when D_s^+ and D_s^- mesons are produced at different rates in the proton-proton collisions, mainly as a consequence of strong-interaction effects in the hadronization process. A detection asymmetry, on the other hand, arises when final-state particles of a given charge are reconstructed with higher efficiency than those of the opposite charge. In general, such asymmetries originate from systematic effects inherent to experiments of this type.

In general, studies of this kind are performed using control channels, i.e. channels in which the effect under investigation is not expected to be present. In the case of CP violation searches, one would use a Cabibbo-favoured decay with similar dynamics and the same final state. However, this analysis does not rely on a control channel to study these instrumental asymmetry effects.

In order to overcome this limitation, this analysis adopts two complementary strategies to study such effects and assess their impact on the analysis:

1. **Simulation-based studies:** A detailed study using RapidSim is conducted to evaluate the potential impact of detection, production, and PID charge asymmetries on our signal sample. These simulations allow for a controlled injection of known asymmetries to assess their effect on the S_{CP} distribution.

2. **Cross-checks with similar channels:** Studies of charge asymmetries are performed using Cabibbo-favored 3-body charm decays, for which no CPV effects are expected. Although not sharing the same final state, any potential effect appearing there would point to possible effects in our sample too.

6.1 Charge asymmetries studies with simulations

A first approach to study the possible impact of instrumental asymmetries on the$D_s^+ \to \pi^-\pi^+K^+$ channel consists of using simulated events generated with RapidSim. In addition, this study also includes investigations of potential effects arising from asymmetries induced by the PID cuts applied during the selection procedure.

In this study, RapidSim is employed to generate simulated samples of the $D_s^+ \to \pi^-\pi^+K^+$ decay. The events are generated using the EvtGen DDalitz model, which is

based on information from the FOCUS collaboration. In the present analysis, however, the final simulated samples are scaled to be compatible with the signal yields obtained in the final data sample. The set of variables available in RapidSim is rather limited when compared to the information available in data events. Nevertheless, a selection as close as possible to that applied to data is imposed on the RapidSim samples, within the constraints of the variables available in these simulations.

In order to ensure a good level of agreement between the RapidSim samples and data, the simulated events are reweighted to reproduce the kinematic and phase-space distributions observed in data. This reweighting procedure is performed using the Gradient Boost Reweighter (GBReweighter) algorithm, which is used to align the RapidSim samples with the sPlotted data distributions.

The variables included in the reweighting procedure comprise the momentum, transverse momentum, and pseudorapidity of both the initial- and final-state particles, as well as the squared invariant masses $S_{\pi\pi}$ and $S_{\pi K}$ that define the Dalitz plot. The hyperparameters adopted in this reweighting procedure are summarised in Table 15

Hyper-parameter	Value
n_estimator	500
learning_rate	0.1
MaxDepth	6

Tabela 15 – Hyper-Parameters for Reweithing process of RapidSim samples.

The outcome of the reweighting procedure can be assessed by comparing Figs. 56 and 57, which show the distributions of the relevant variables in simulation and data before and after the reweighting, respectively. As can be observed, after the reweighting the RapidSim samples provide a very faithful representation of the sPlotted data. This conclusion is further supported by the comparison based on the ROC curve shown in Fig. 58. After the reweighting procedure, the ROC curve is found to be approximately diagonal, indicating that the reweighted simulation and data samples are essentially indistinguishable by the classifier.

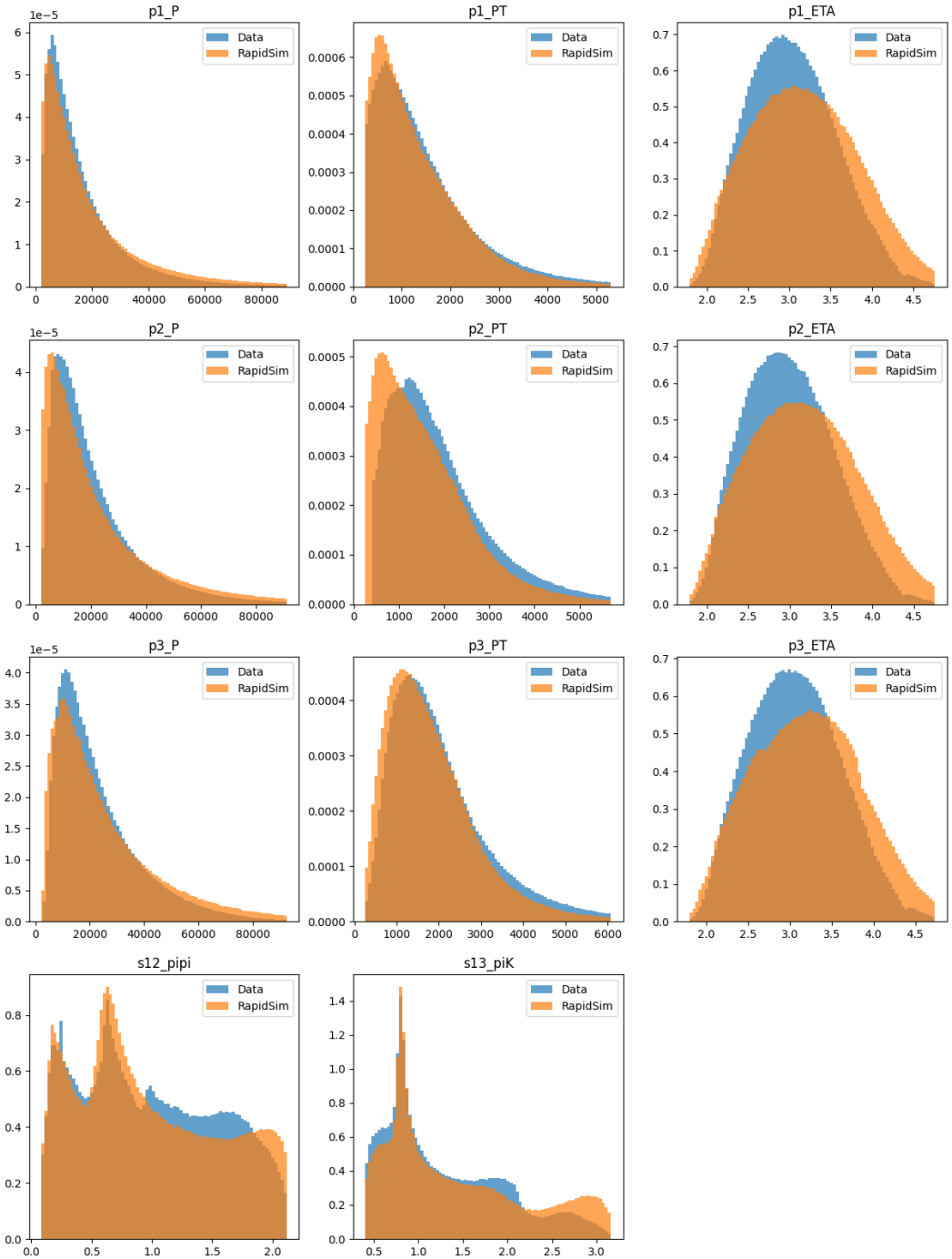

Figura 56 – Comparison between the sPlotted data (blue) and RapidSim distributions (orange) before the reweighting process for $D_s^+ \to \pi^-\pi^+K^+$: kinematic variables of the final state particles.

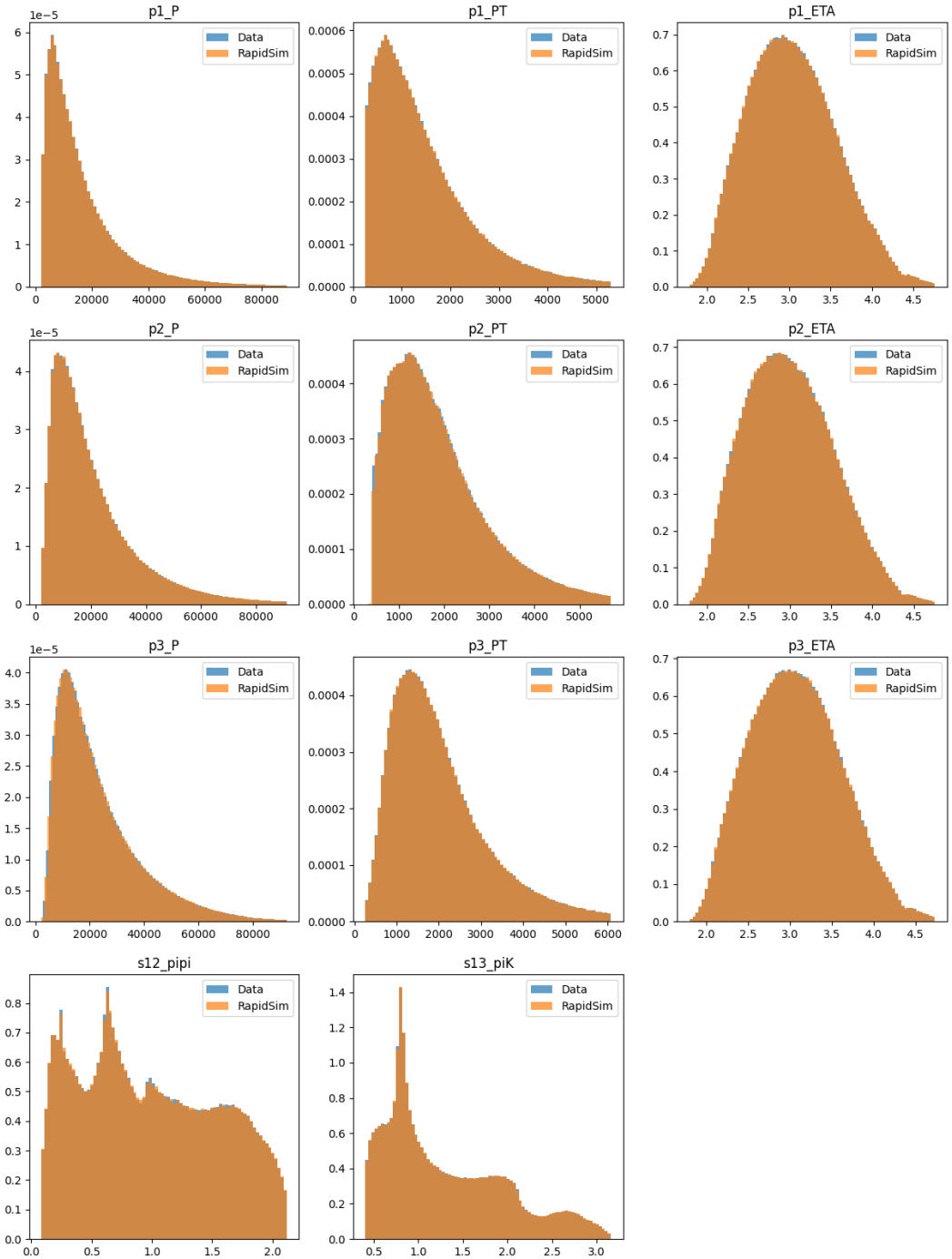

Figura 57 – Comparison between the sPlotted data (blue) and RapidSim distributions (orange) after the reweighting process for $D_s^+ \to \pi^-\pi^+K^+$: kinematic variables of the final state particles.

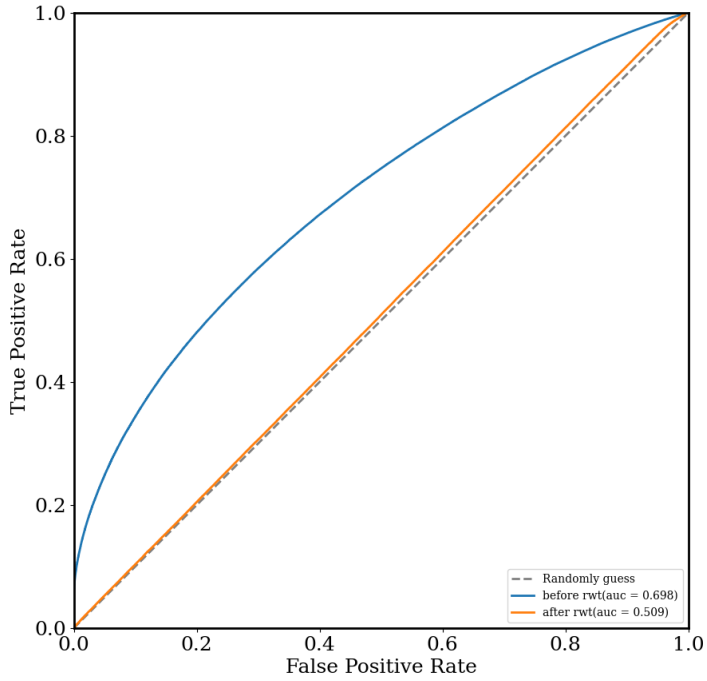

Figura 58 – ROC curve before (blue) and after (orange) the reweighting procedure for $D_s^+ \to \pi^-\pi^+K^+$. The similarity between the sPlotted data and the weighted RapidSim samples is quantified by the AUC (Area Under the Curve). An AUC score closest to 0.5 indicates the samples are nearly indistinguishable, demonstrating excellent agreement.

Once the RapidSim samples are shown to provide a good representation of the data, the next step is to introduce in these samples the asymmetries that are intended to be studied. Under the assumption that these asymmetries are small, the raw asymmetry can be factorised as:

$$A_{raw} \approx A_{prod} + A_{det} + A_{PID} + A_{CP} \tag{6.1}$$

with A_{prod}, A_{det} and A_{pid} representing the production, detection, and PID asymmetries.

The idea is to inject detection, production, and PID asymmetries into the simulated samples one at a time, in order to study their individual impact as well as the effect of their combination. The injection of these asymmetries is performed using an algorithm that assigns an effective asymmetry based on the kinematic properties of each event, such as momentum, pseudorapidity, or rapidity, depending on the asymmetry model under consideration. From the assigned asymmetry, a ratio R is computed, defined as the ratio between the number of positively charged events and the number of negatively charged events , as describe in the fig. 59. This ratio is then used to probabilistically reject

events according to the value of R and the charge of the reconstructed D_s meson.The full procedure is illustrated in Fig. 59.

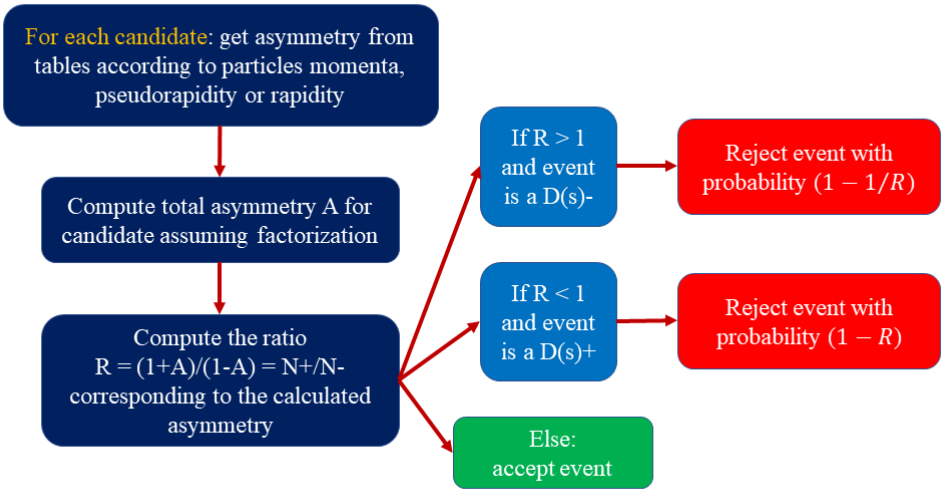

Figura 59 – Diagram illustrating the algorithm used to inject production, detection, and PID asymmetries into RapidSim samples through a rejection sampling procedure.

The detection asymmetry models adopted for pions and kaons are taken from [6, 7], as shown in Fig. 60, where the final bin is extended to 100 GeV for this study. PID asymmetries are derived from the PIDCalib package efficiency tables, parameterized in terms of $p_i = |\vec{p}|, \eta$. For production asymmetries, two different models were considered based on [8]: the Run 1 measurement and a model mimicking the Pythia 8 prediction, adjusted to align with the Run 1 measurement, as illustrated in Fig. 61.

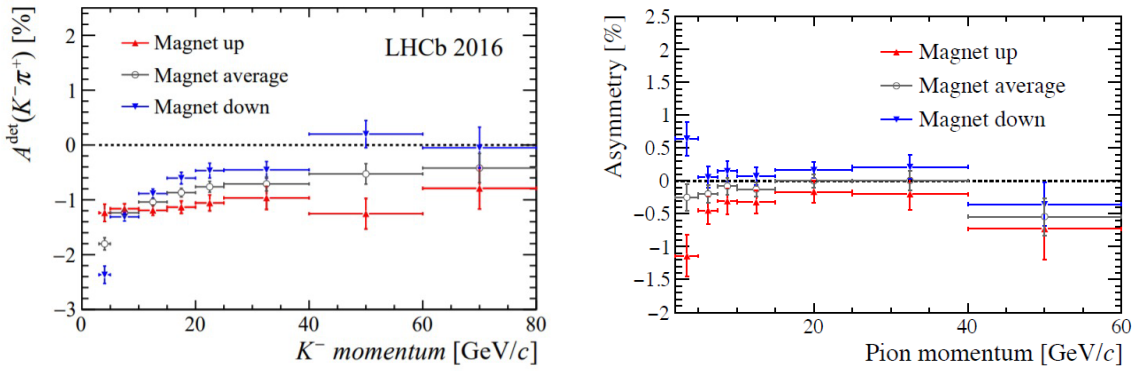

Figura 60 – Detection asymmetries measured for kaon (plot on left), coming from [6] and pions (plot on right) with 2016 data, from [7].

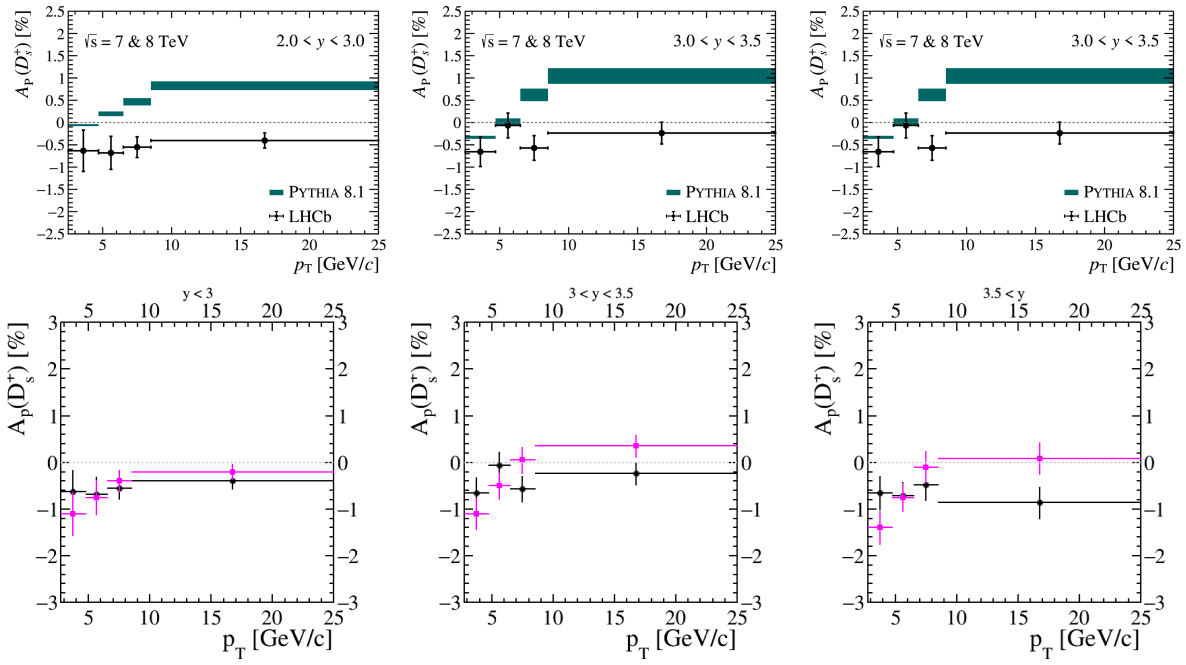

Figura 61 – Production asymmetries for D_s^+ compared to the Pythia model, as reported in [8] (top plots). Production asymmetry models used for the studies: the Run 1 measurement (black dots) and an alternative model (magenta dots), which mimics the shape of the Pythia prediction but is shifted to have a better compatibility with the Run 1 measurement (bottom plots).

To conduct our study, S_{CP} is computed following the same approach adopted in the background, is computed following a stepwise process under different scenarios:

- Without any asymmetry (null tests);
- With only detection asymmetries ;
- With only PID asymmetries;
- With only production asymmetries ;
- With both PID and detection asymmetries;
- With PID, detection, and production asymmetries;

The number of generated events corresponds to approximately three times the statistical size of the data sample obtained after the final selection. In order to make optimal use of this available statistics, these events are employed in a bootstrapping procedure to resample the simulated dataset, producing ten independent samples at a stage prior to the injection of asymmetries. This approach allows the impact of the injected asymmetries to be studied across multiple pseudo–data samples.

The results of these studies for one set are presented in Fig. 62, which displays the S_{CP} distribution across the Dalitz plot, as well as the corresponding p-values as a example. The inclusion of asymmetries in these simulations does not show any significant impact on the values S_{CP} observed in our channel. All the results can be seen in appendix D.

Finally, the distribution of p-values obtained from these samples is shown in the figure 63. From this figure, as well as from the corresponding distributions presented in the Appendix D, it can be observed that all p-values obtained in all tests lie within the range between 1% and 100%. As a result, it can be stated that, under the conditions in which this study was conducted, no sensitivity to any of the asymmetry models considered is observed, given the statistical sample of 32 million events obtained after the final selection.

However, a total of ten samples is still insufficient to study possible trends in the p-value distribution. Additional simulations are therefore required and are currently being generated in order to evaluate the effects using a larger number of toy samples, enabling a more comprehensive study aimed at the publication of this work.

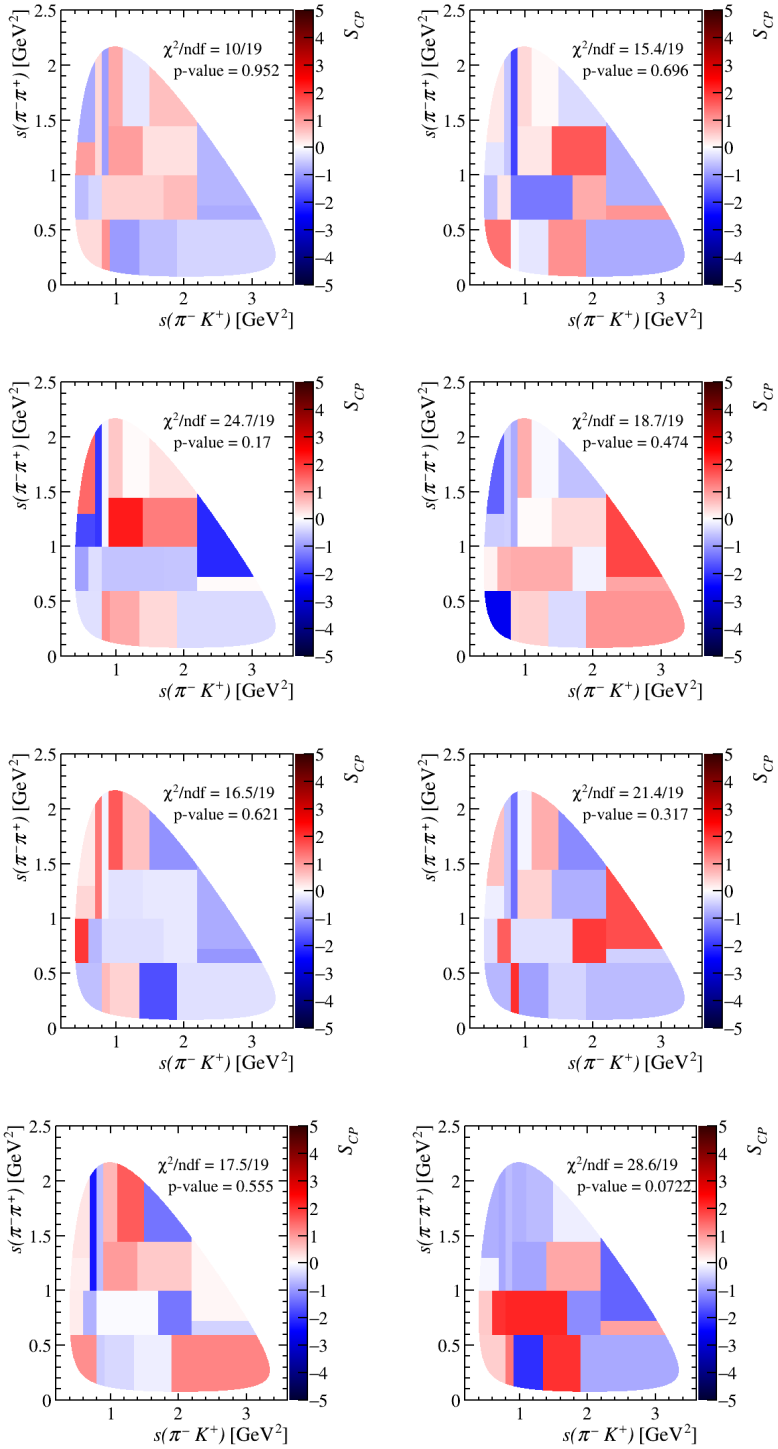

Figura 62 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described in the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

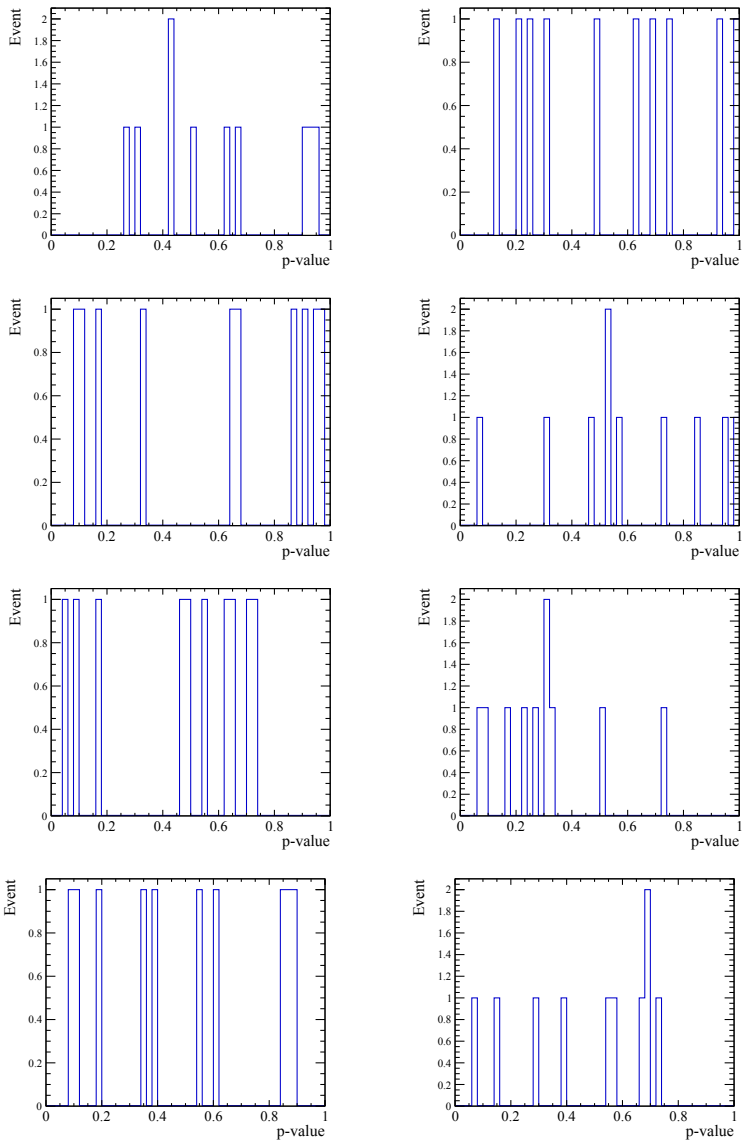

Figura 63 – Results obtained for p-value with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described in the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

6.2 Cross-checks with Cabibbo-favored decays.

Although the simulation studies make use of asymmetry models derived from measurements performed by the LHCb experiment, the results cannot be considered fully equivalent to a direct evaluation of such asymmetries using a proper control channel.

Asymmetries arising from other processes may still be present in Run 2 data samples. With this in mind, studies were carried out using Cabibbo-favoured decay channels.

The channels adopted were $D_s^+ \to K^-K^+\pi^+$ and $D^+ \to K^-\pi^+\pi^+$. Owing to the significant differences in the corresponding Dalitz plots, these channels cannot be regarded as genuine control channels. Nevertheless, they can be used to investigate the presence of asymmetries, thereby acting as complementary studies to those performed with RapidSim simulations. These studies can provide valuable information on potential sources of nuisance asymmetries in data samples with sizes comparable to those obtained for $D_s^+ \to \pi^-\pi^+K^+$ decay.

The decay $D_s^+ \to K^-K^+\pi^+$ shares the same initial state and has two final-state particles identical to those of the signal channel, while $D^+ \to K^-\pi^+\pi^+$ has a similar three-body topology but with a different charge configuration.

Both channels were subjected to a selection procedure that is essentially identical to that adopted for the signal channel, with the exception of the cuts used to remove $\bar{D}^0$ contaminations and the PID requirements, which were adjusted to account for the different particle content in the final state. For these samples, the decision function obtained during the MVA training was applied such that the final selection is equivalent to that adopted for the signal channel. The optimal working point chosen for the output variable is the same as that used in the signal analysis. The invariant mass distributions after the full selection can be seen in Figures 64, as well as the corresponding Dalitz plots shown in Figure 65.

Specifically for the $D^+ \to K^-\pi^+\pi^+$ channel, the Dalitz plot shown in the figure is not represented in terms of the invariant masses s_{12} and s_{13}, but rather using the observables s_{high} and s_{low}, constructed from these invariants. This choice is motivated by the symmetric structure of the final-state phase space, since the two pions are identical particles. In channels where particles 2 and 3 are identical, the upper and lower lobes of the Dalitz plot are related by a mirror symmetry with respect to the diagonal $s_{12} = s_{13}$ Therefore, for CP-violation studies, it is advantageous to analyses only one side of the Dalitz plot (folded representation), combining the statistics of equivalent regions and allowing a direct comparison between CP-conjugate regions.

To ensure statistical precision comparable to that of the $D_s^+ \to \pi^-\pi^+K^+$ signal sample when applying the asymmetry tests, each control mode is divided into subsamples. The subsamples are constructed such that the number of candidates within a 30 MeV window around the parent-particle mass is of the same order as in the signal sample. The splitting is performed randomly.

Owing to the high purity achieved in these channels, as can be seen in the figure 64, this study proceeded by computing the populations of $D_{(s)}^+$ and $D_{(s)}^-$ candidates in each

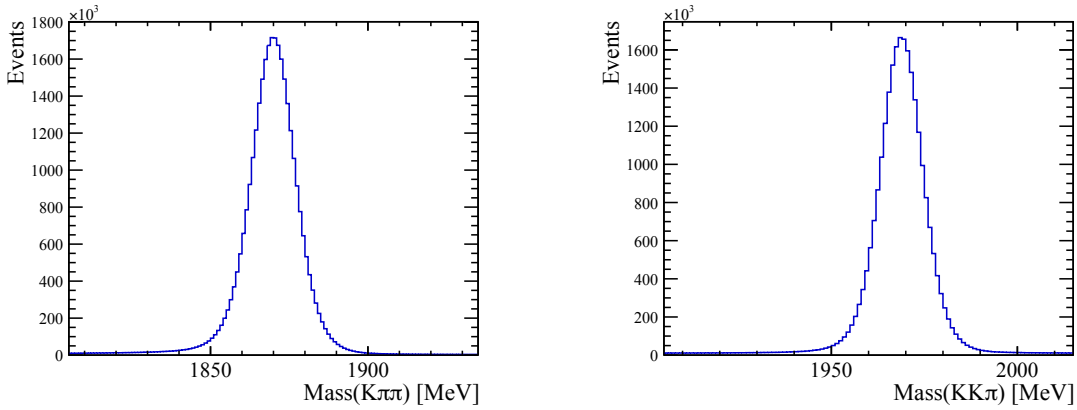

Figura 64 – invariant mass for $D^+ \to K^-\pi^+\pi^+$ (left) and for $D_s^+ \to K^-K^+\pi^+$ (right) after the selection.

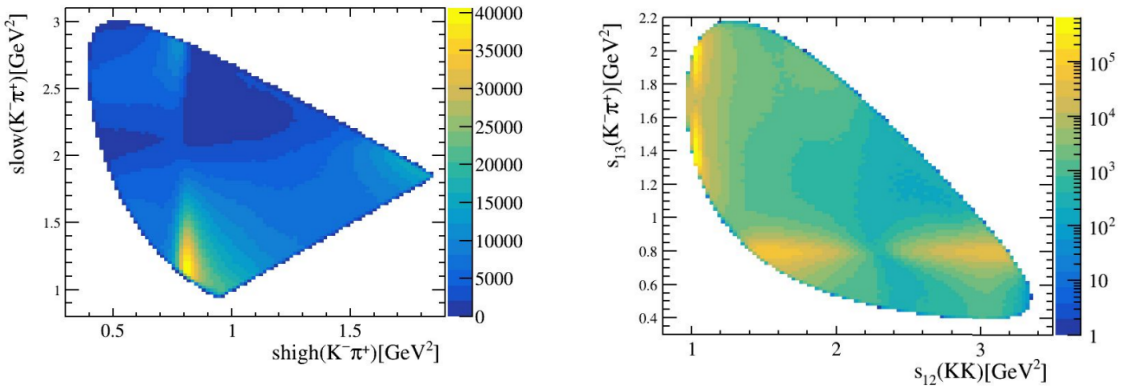

Figura 65 – Folded Dalitz plot for $D^+ \to K^-\pi^+\pi^+$ (left) and Dalitz plot for $D_s^+ \to K^-K^+\pi^+$ in a log scale (right). In the folded DP, *shigh* and *slow* are the maximum and minimum values of the pair of $m(K^+\pi^+)$ invariant masses, respectively.

bin through simple event counting, using the same expression and methodology adopted for the background studies performed in the invariant-mass sidebands. As in the sideband studies, the uncertainties were taken as the square root of the population in each bin.

For this study, two binning schemes were considered. First, physics-motivated binnings were constructed, consisting of 22 bins for $D_s^+ \to K^-K^+\pi^+$ and 19 bins for $D^+ \to K^-\pi^+\pi^+$. In addition, an alternative uniform binning based on a 5×5 grid was adopted. This results in 18 bins for $D_s^+ \to K^-K^+\pi^+$ and 19 bins for $D^+ \to K^-\pi^+\pi^+$.

Representative results for the Mirandizing tests in both binning schemes are shown in Figs. 66 and 67 for the combined 2016–2018 dataset, with all additional subsamples provided in Appendix E. No discernible structure or pattern is observed in any of the binnings. The p-values obtained per year and for all subsamples across both binning

schemes are reported in Tabs. 17–19.

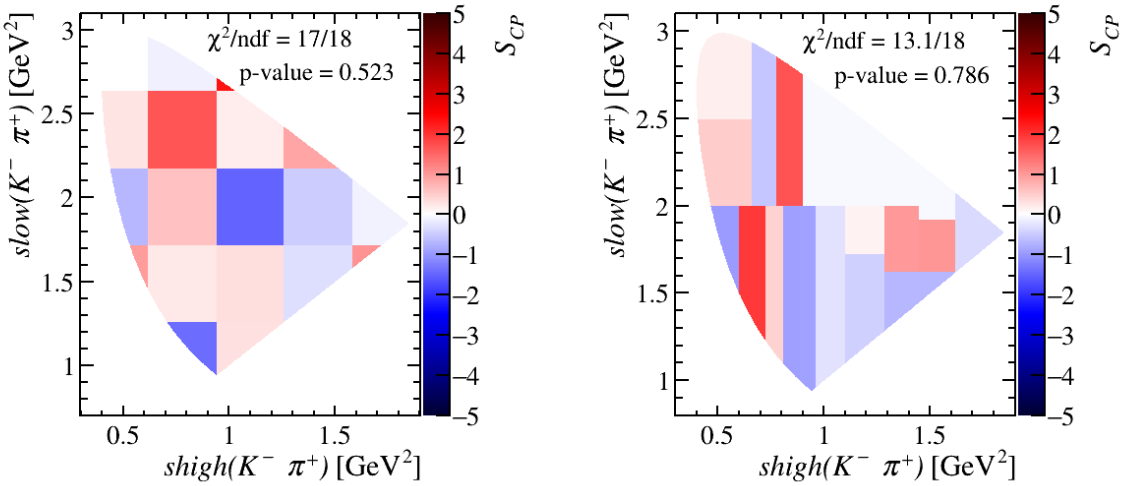

Figura 66 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. On right we have the uniform binning and on left the physics binning.

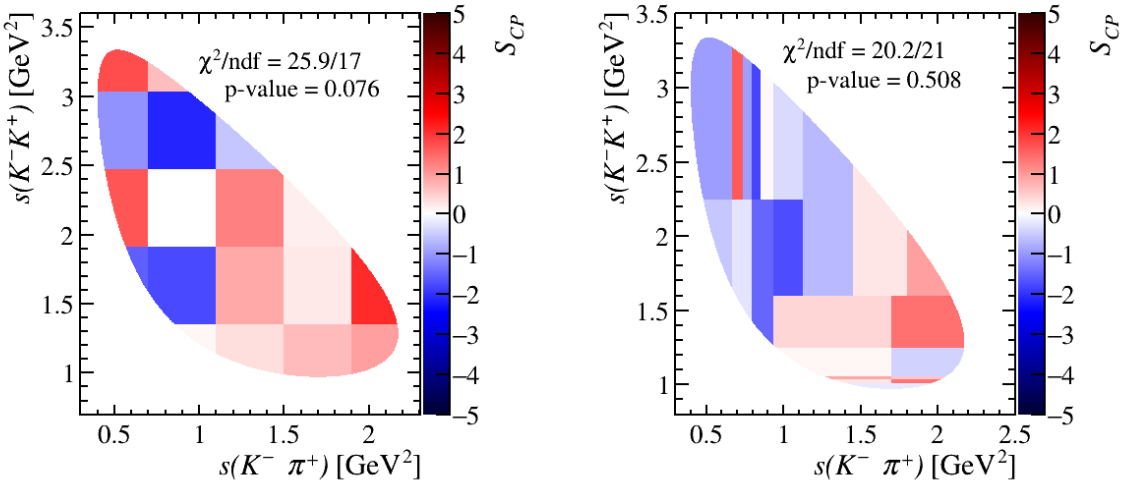

Figura 67 – The Mirandizing method was applied to the decays of $D_s^+ \to K^-K^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. On right we have the uniform binning and on left the physics binning.

	16	**17**	**18**	**All**
Set 1	18%	53%	10%	51%
Set 2	2%	10%	19%	0.1%
Set 3	29%	2%	22%	1%
Set 4	21%	48%	32%	27%

Tabela 16 – P-values for $D_s^+ \to K^-K^+\pi^+$ for physics binning.

	16	**17**	**18**	**All**
Set 1	56%	92%	77%	73%
Set 2	69%	74%	80%	71%
Set 3	40%	17%	47%	23%
Set 4	0.3%	5%	87%	12%
Set 5	36%	62%	15%	47%

Tabela 17 – P-values for $D^+ \to K^-\pi^+\pi^+$ for physics binning.

	16	**17**	**18**	**All**
Set 1	7%	8%	32%	8%
Set 2	13%	6%	93%	5%
Set 3	2%	7%	1%	0.1%
Set 4	65%	54%	75%	75%

Tabela 18 – P-values for $D_s^+ \to K^-K^+\pi^+$ for uniform binning.

	16	**17**	**18**	**All**
Set 1	48%	94%	33%	52%
Set 2	53%	8%	93%	34%
Set 3	67%	22%	42%	25%
Set 4	19%	37%	87%	22%
Set 5	80%	35%	37%	96%

Tabela 19 – P-values for $D^+ \to K^-\pi^+\pi^+$ for uniform binning.

The p-value results reported in the tables 16 - 19, as well as the distributions of the S_{CP} values across the Dalitz plot presents on appendix E, do not reveal large asymmetries that could compromise the present analysis.

In the specific case of the $D_s^+ \to K^-K^+\pi^+$ channel, datasets with p-values of the order of 0.1% were found for both the physics-motivated and the uniform binnings. However, this channel has already been published, as it was used as a control channel for the $D^+ \to K^-K^+\pi^+$ analysis. The asymmetries observed in that study are therefore already known and motivated the development of a modified observable designed to handle such effects (see Ref. [] for details).

The purpose of the present study is instead to evaluate whether these effects would still be visible in a dataset with a statistical power approximately four to five

times smaller than that used in the $D^+ \to K^-K^+\pi^+$ analysis. From this perspective, the results obtained here are considered acceptable: three out of the four sets exhibit p-values compatible with the hypothesis of charge symmetry, while the remaining channel still shows p-values far from those expected for a 5σ observation of an asymmetry using this technique.

7 The contribution for the run 3

This chapter is dedicated to the contributions made to the LHCb upgrade, focusing in particular on the work performed to finalize the code responsible for the calibration procedure on Electromagnetic calorimeter (Ecal). It reports the Ecal calibrations carried out in 2024, as well as the major upgrade implemented in preparation for data taking in 2025.

This project was carried out in person at CERN during the years 2024 (for two months) and 2025 (for three months), under the supervision of Jean-François Merchand (LAPP – Centre National de la Recherche Scientifique, France), Rosen Matev (CERN), Florian Reiss (University of Freiburg, Germany), and Marianna Fontana (INFN Bologna, Italy). It represents a contribution from RTA-Rio to the LHCb RTA.

7.1 Ecal calibration

During Run 3, the ECAL calibration was performed using two distinct procedures. The so-called absolute calibration is obtained through the calibration using neutral pions (π^0). The data collected after the absolute calibration are then used as a reference to compute a relative calibration, commonly referred to as the LED calibration. The contributions made during the upgrade and data-taking periods described in this work are primarily related to the absolute calibration procedure.

The main difference in the electromagnetic calorimeter calibration procedure between Run 2 and Run 3 is that the process is no longer performed using minimum-bias samples, but instead relies on collected data recorded by LHCb. This update was made possible by the implementation of a dedicated HLT1 trigger line for Run 3, specifically designed to collect events associated with the $\pi^0 \to \gamma\gamma$ decay. This decay channel is chosen due to its very large branching fraction, approximately 98%. For the absolute calibration, about 100–120 million events are collected at the beginning of each LHC fill by this dedicated trigger line. The requirements applied in this trigger line are summarized in Table 20.

HLT1 requirements	
Mass (MeV)	50 - 300
$Pt_{(\gamma\gamma)}$ (MeV)	>1000
Et_{γ} (MeV)	> 200
$Et_{(\gamma\gamma)}$ (MeV)	> 400
E_{γ}^{19}	0.7 - 1.0
nPvs	1

Tabela 20 – The HLT1 trigger requirements of $\pi^0 \rightarrow \gamma\gamma$ trigger line.

Photons are defined as 3×3 clusters of Ecal cells. To reconstruct the $\gamma\gamma$ final state, pairs of clusters are combined according to their spatial positions and energies. In order to reduce the large number of possible combinations, selection requirements are applied on the transverse energy of each cluster (Et_{γ}) and on the energy of the two-cluster system ($Et_{\gamma\gamma}$). An additional selection is imposed on the E_{γ}^{19} variable of each cluster, defined as the ratio between the energy deposited in the most energetic cell of the cluster and the total cluster energy, as well as on the number of allowed primary vertices.

During Run 2, about 300 million minimum-bias events were used for each calibration procedure. Producing and processing these samples required a significant amount of time and computational resources, which limited the execution of the calibration to an average frequency of about once per month. In addition, further time and resources were needed to derive and validate the corresponding sets of calibration constants. By using collected LHCb data, the calibration procedure can instead be executed whenever needed, in principle even between consecutive fills, allowing for a more frequent update of the calibration constants and, consequently, improved calibration precision.

Concerning the calibration procedure itself, the raw data collected for this purpose are fully reconstructed at the HLT2 level using the LHCb Moore and DaVinci frameworks. The selection criteria applied at the HLT2 stage during the reconstruction process are detailed in Table 21.

HLT2 requirements	
$isnotH_{\gamma}$	> 0.7
E_{γ}^{19}	>0.7
Pt_{π^0} (MeV)	> (200 (MeV) $\times(7 - \eta_{\pi^0})$)
Pt_{γ} (MeV)	>300
Additional requirements	
nPVs	$\geq$1
Pt_{π^0} (MeV)	> 1700 (Only on tigher region)

Tabela 21 – The HLT2 requirements and additional requirements adopted for π^0 calibration [9].

Table 21 also includes additional selection requirements applied after the reconstruction process. $IsNotH_\gamma$ is a variable based on a neural network of the multilayer perceptron (MLP) type that provides the probability for a given candidate to be a neutral particle or an anti-hadron. The variable Pt_{π^0} denotes the transverse momentum of the pion, and η_{π^0} its pseudorapidity, while nPVs corresponds to the number of reconstructed primary vertices. These cuts are designed to ensure that only low pile-up samples are used and to maintain control over the background levels in calorimeter cells located in the lateral regions of the middle and outer sections of the electromagnetic calorimeter. The definition of these regions is illustrated in Fig. 68.

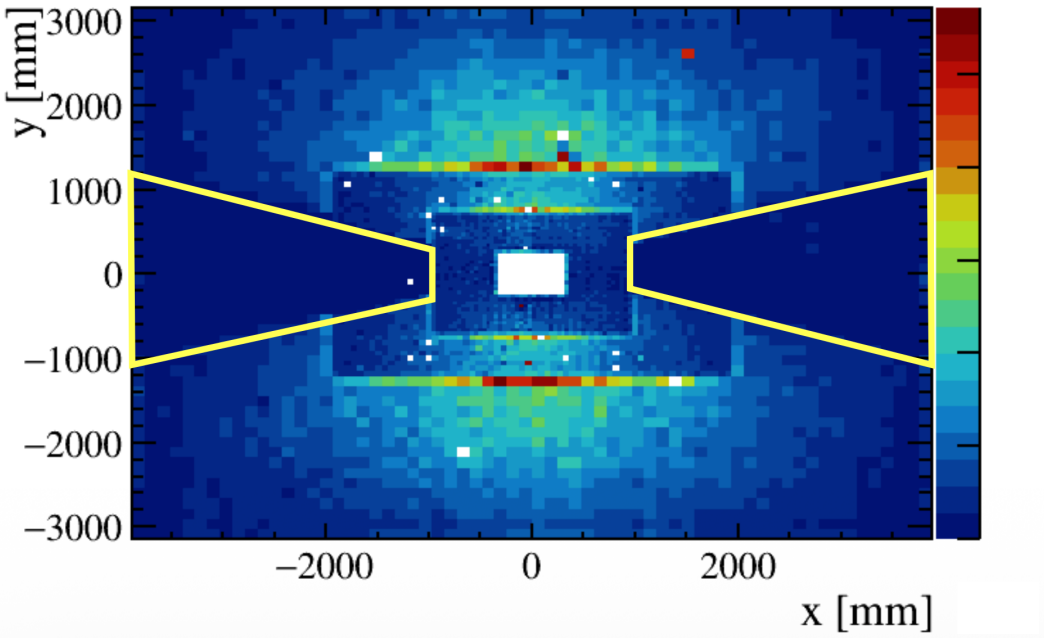

Figura 68 – Front view of the electromagnetic calorimeter. The yellow region indicates the boundaries of the area where the transverse momentum cut is applied in order to reduce background levels.

The HLT2 selection and the additional cuts were studied and defined by Yang Gao (Central China Normal University, CCNU) and precede the contributions of the RTA-Rio group to the absolute calibration of the electromagnetic calorimeter.

After the samples are reconstructed, invariant mass histograms of the $\gamma\gamma$ system are filled separately for each of the 6016 cells of the electromagnetic calorimeter, allowing these distributions to be fitted with a probability density function (PDF). The PDF used in the fit is composed of a Gaussian function to model the signal component and a second-order polynomial to describe the background. The fits are performed using the RooFit framework independently for each of the 6016 calorimeter cells, providing an estimate of the π^0 mass peak position in each cell. This information is then used to compute a correction factor, denoted as λ, defined as:

$$\lambda = \frac{M^{\pi^0}_{fit}}{M^{\pi^0}_{PDG}} \tag{7.1}$$

where $M^{\pi^0}_{fit}$ is the π^0 mass obtained from the fit in each calorimeter cell, and $M^{\pi^0}_{PDG}$ s the nominal π^0 mass taken from the Particle Data Group (PDG). The correction factors λ computed in this way are then used to correct the position of the π^0 mass peak in a subsequent reconstruction step, after which the fits are performed once again to update the mass value obtained for each cell.

After seven iterations of this procedure, the resulting set of correction factors is stored and labelled as λ_1. This set is then used as input to reprocess the same data sample, which is subsequently subjected to an additional seven iterations of the fitting procedure. At the end of this second cycle, a new set of correction factors, denoted as λ_2, is obtained. The corrections used to calibrate the calorimeter cells are defined as $\Lambda = \lambda_1 \times \lambda_2$.

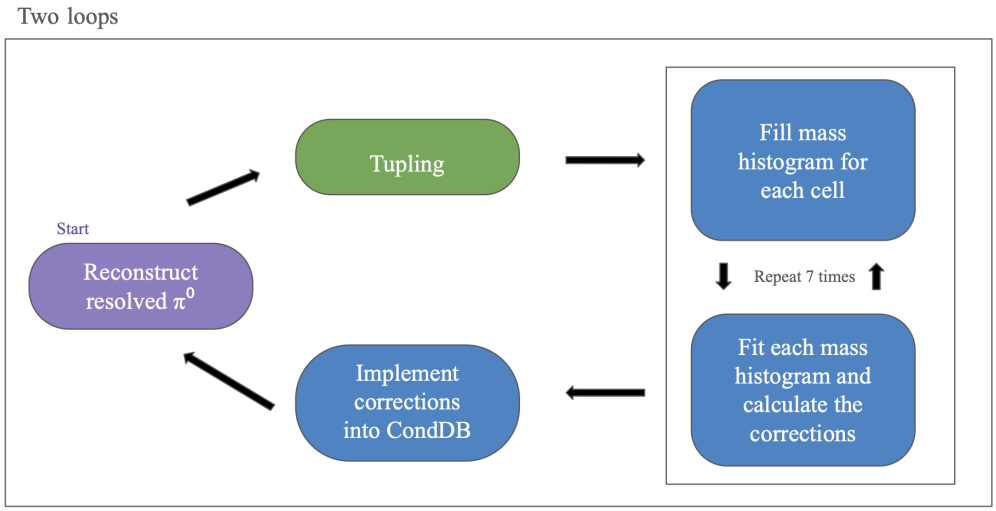

Figura 69 – The π^0 calibration flow.

In order to ensure the quality of the fits, strict criteria are applied to classify fits as good or bad. Given the large number of fits performed in a single calibration, these criteria are based on quantities extracted from the fits like the fitted mass position. The criteria are described in Table 22.

Fit status	0 or 1
covQual	3
π^0_{mean} (MeV)	105 - 155
π^0_{σ} (MeV)	4 - 25

Tabela 22 – Criteria used to define good fits by the calibration code

Fits that do not satisfy the criteria described above are not used to compute the lambda corrections. Owing to the way the events are reconstructed, it is expected that corrections applied to cells in the vicinity of those not corrected can help stabilize the fit in these cells, thereby allowing corrections to be determined for them.

Once the constants are calculated, they can be properly used to update the reference for the high-voltage (HV) setting of each cell according to:

$$HV^{New} = HV^{old} \times \Lambda^{\frac{1}{\alpha}} \tag{7.2}$$

where α is an intrinsic parameter of the photomultiplier tubes.

Among the contributions made to this code, the following can be highlighted:

- Finalization, testing, and bug fixing of the preliminary version of the code.
- Update of the fitting code, including the formulation of new criteria for good fits.
- Parallelization of the fitting stage of the code.

At the beginning of the project, the code was not yet finalized, which required extensive testing to verify the presence of bugs. In addition, a complete update of the fitting code was performed in order to take advantage of more recent RooFit tools, as well as to define new criteria for good fits that are more compatible with the behavior observed in the 2024 data. Finally, only a limited portion of the code was parallelized at the start of the project, causing the fits to be processed sequentially, one at a time. The parallelization alone was responsible for reducing the total fit processing time from approximately four hours to about fifteen minutes.

The full integration of the π^0 calibration code into the LHCb environment was only completed in 2025. This prevented the code from being used in the various EFF computing modes and from being executed by the Experiment Control System (ECS) during the 2024 data-taking period. Throughout 2024, the entire calorimeter calibration was therefore produced offline using the LHCb swdev machines, which limited the data processing capacity and significantly increased the time required for a calibration, to approximately 24 hours per calibration.

In order to illustrate the results obtained with this calibration, the figures 70 compare the values of the reconstructed π^0 mass on a cell-by-cell basis in each of the three regions of the calorimeter. In addition, Fig. 70 shows the one-dimensional distributions of the reconstructed masses for each calorimeter region in a single calibration.

In Fig. 71 it is possible to compare the integrated result obtained from data collected in 2025 exclusively for this purpose, before and after the calibration of this sample, demonstrating the effectiveness of the method.

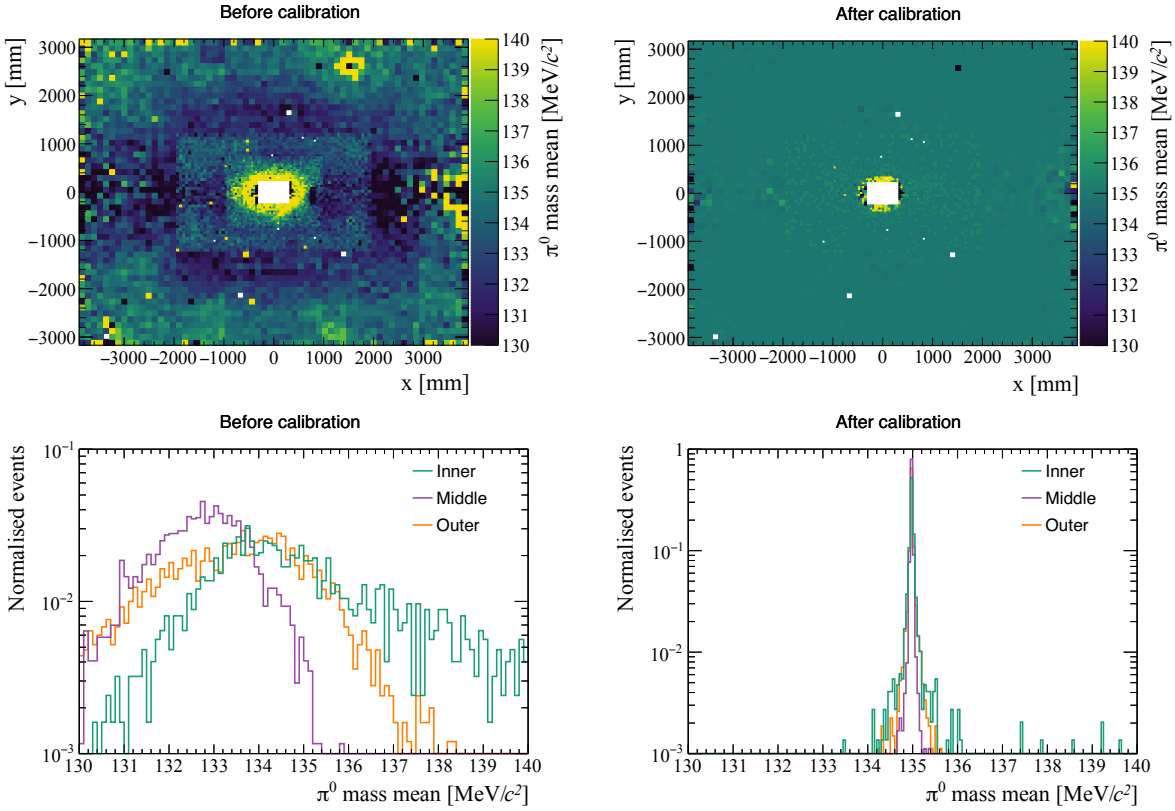

Figura 70 – Fitted $\gamma\gamma$ mass mean before (left) and after (right) performing the calibration of September 2024, for each ECAL cell (top) and per ECAL region (bottom).

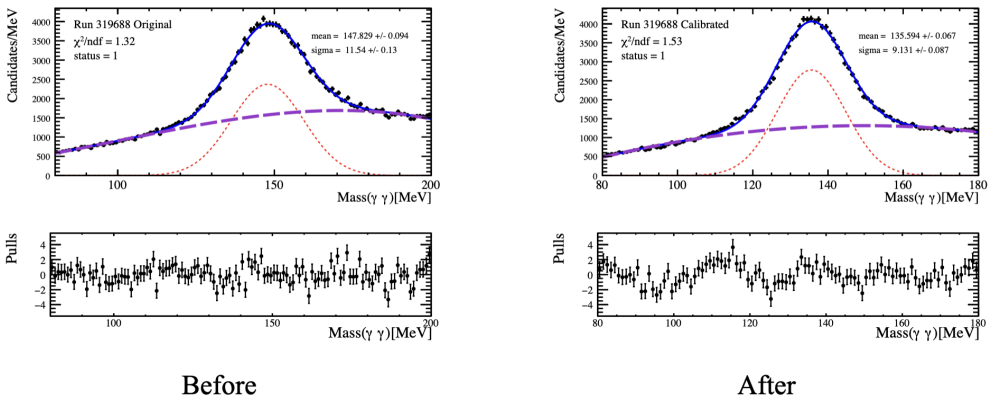

Figura 71 – Invariant mass fit of the $\gamma\gamma$ system before (left) and after (right) the calibration using LHCb data.

8 Conclusion

The charm-quark sector is a particularly interesting environment in which to study CP violation, both because only a single CP-violation measurement is currently known in this sector and because it represents a fertile ground for searches for physics beyond the Standard Model.

This work presents a search for CP violation in the decay channel $D_s^+ \to \pi^-\pi^+K^+$. The search is performed in the phase space of the decay (Dalitz plot) using a model-independent technique that aims to map asymmetries across the phase space through an observable, S_{CP}, which quantifies the difference in statistical significance between the D_s^+ and D_s^- populations. From this statistical test, a p-value is defined based on the S_{CP} distribution and the number of degrees of freedom. If successful, this study would constitute the first measurement of CP violation in a charged charmed meson.

The analysis is performed using LHCb data collected during Run 2, produced in proton–proton collisions at a center-of-mass energy of 13 TeV and corresponding to an integrated luminosity of 5.6 fb^{-1}. The final selected sample contains approximately 32 million D_s^+ candidates, a statistical precision comparable to that used in the CP-violation measurement in the D^0 meson published by LHCb in 2019.

The strategy to compute the S_{CP} values in each bin of the physical binning relies on performing simultaneous fits to the invariant mass distribution of the $\pi^-\pi^+K^+$ system in each bin, allowing the extraction of the asymmetries, total yields, and their associated uncertainties. The fits are first performed on Monte Carlo samples in order to determine the signal shape parameters, which are then fixed in the fits to data. This procedure ensures consistent fits to the data while avoiding potential biases.

In the absence of a Cabibbo-favored control channel with dynamics similar to those of the $D_s^+ \to \pi^-\pi^+K^+$ decay, this work investigates the presence of spurious production and detection asymmetries through studies based on RapidSim simulations, complemented by cross-checks using Cabibbo-favored channels also collected during Run 2. These studies, together with the analysis of the sidebands of the $D_s^+ \to \pi^-\pi^+K^+$ channel, reveal no indication of asymmetries.

At the time of writing, results in the signal region are not yet available. Following LHCb policy, analyses are conducted in a blinded manner, and this analysis is still under review by the corresponding working group. As a result, authorization to unblind and analyze the signal region has not yet been granted. Nevertheless, given the advanced stage of the review, the results from the signal region are expected to be published in the near future.

In addition to the CP-violation analysis, this work also presents a contribution to the LHCb Upgrade I and to Run 3 through the development and operation of the electromagnetic calorimeter calibration. This contribution represents an effort by RTA-Rio within the LHCb RTA. Among the main contributions of this project are the completion of a new code responsible for the electromagnetic calorimeter calibration, further upgrades to optimize this code, the production of all electromagnetic calorimeter calibrations throughout 2024, and the final integration of the calibration code with the LHCb Experiment Control System in 2025.

A Fits for the binning scheme

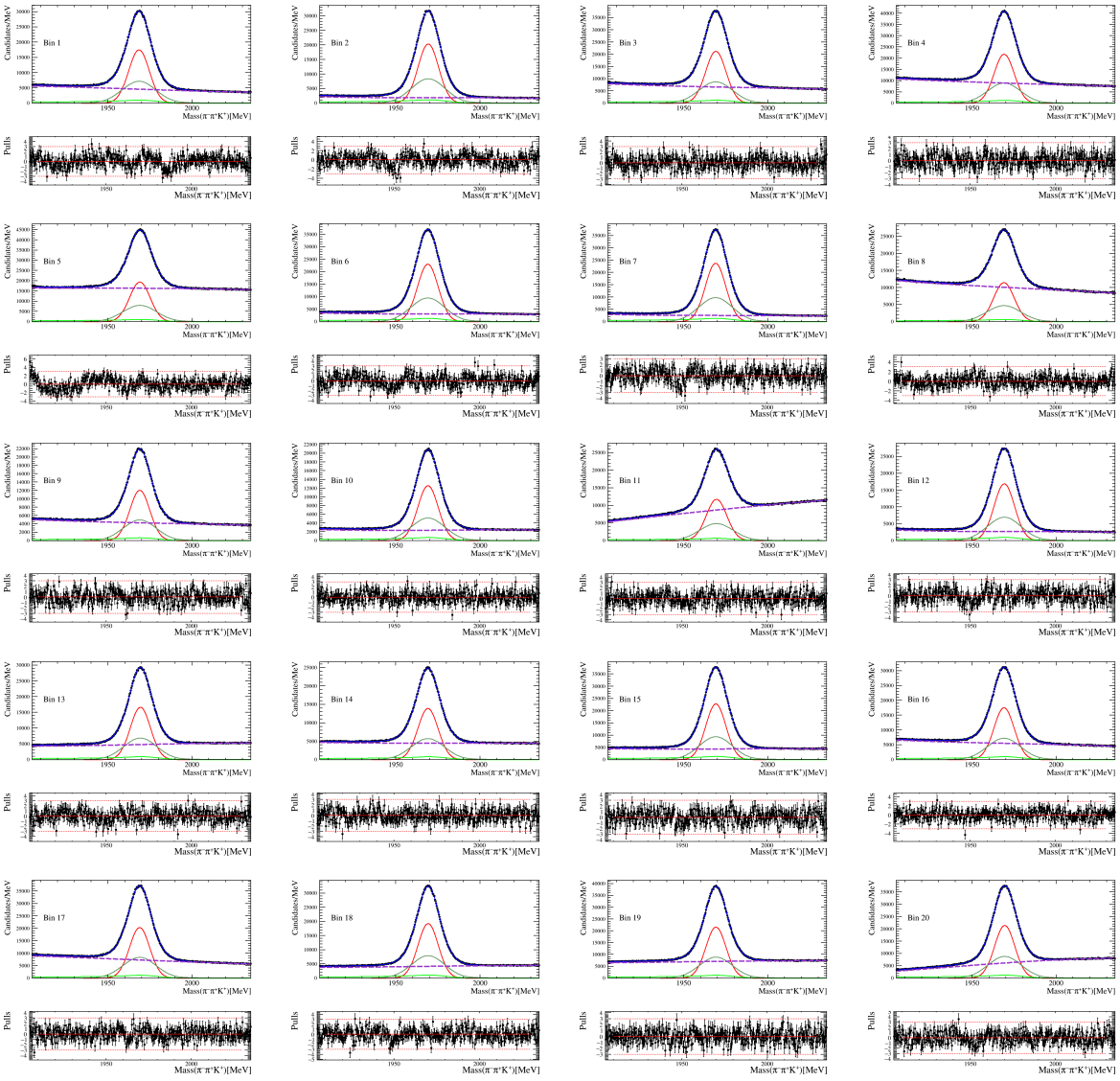

Figura 72 – $\pi^-\pi^+ + K^+$ invariant-mass fit for bins numbered The components of the fit are displayed as follows: the Gaussian in red, the Crystal Balls in green, the Bernstein polynomial in purple, and the combined PDF in blue.

B Fit parameters

B.0.1 Monte Carlo parameters

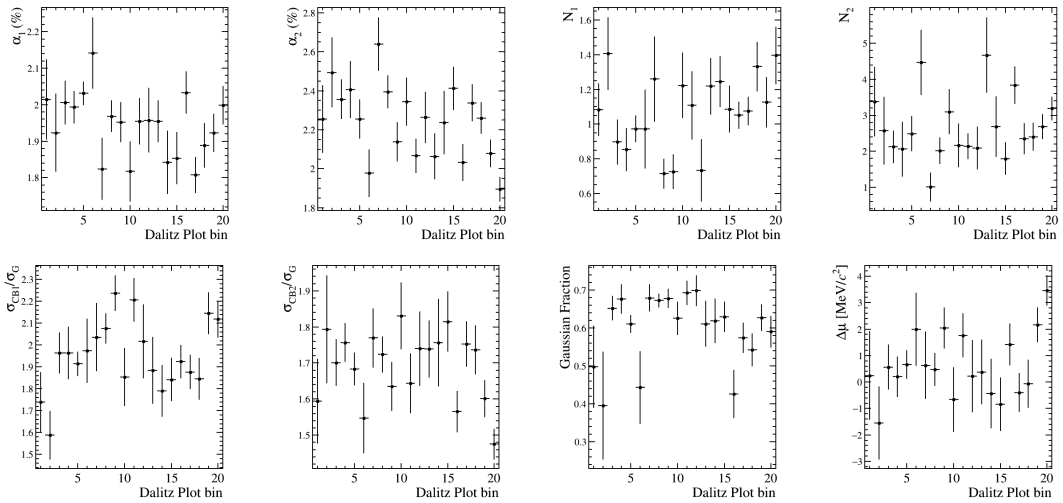

Figura 73 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2016 MC samples.

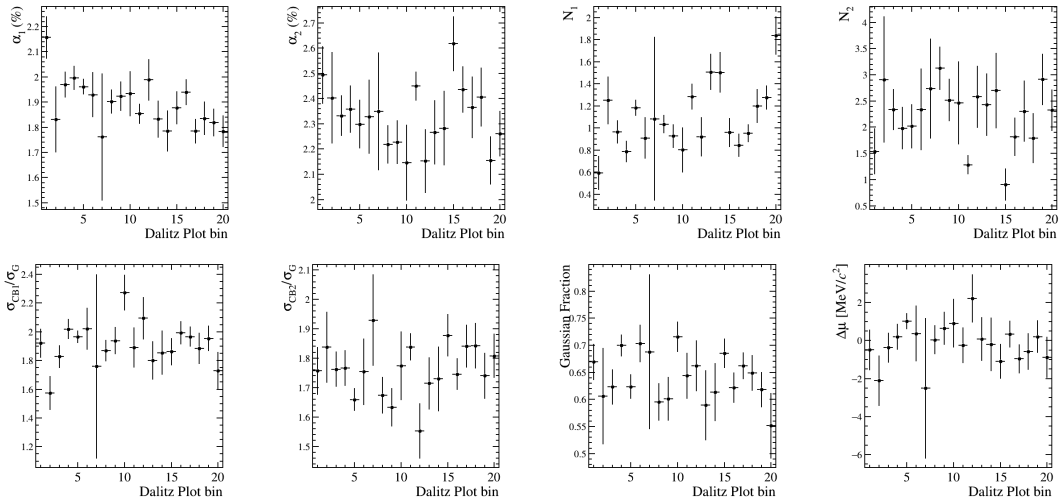

Figura 74 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2017 MC samples.

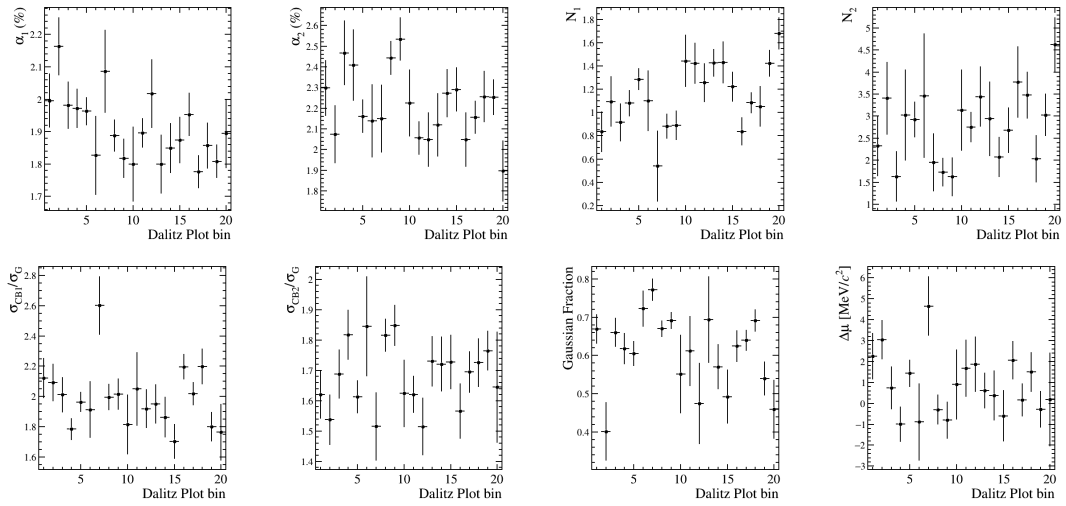

Figura 75 – $D_s^+ \to \pi^-\pi^+K^+$ signal PDF shape parameters as a function of the Dalitz plot bins for the fit results of the 2018 MC samples.

C Fit Results

C.0.1 Fits to $D_s^+ \to \pi^-\pi^+K^+$ Monte Carlo Sample

Figura 76 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2016 MC sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

Figura 77 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2017 MC sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

Figura 78 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2018 MC sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

C.0.2 Fits to $D_s^+ \to \pi^-\pi^+K^+$ Data Sample

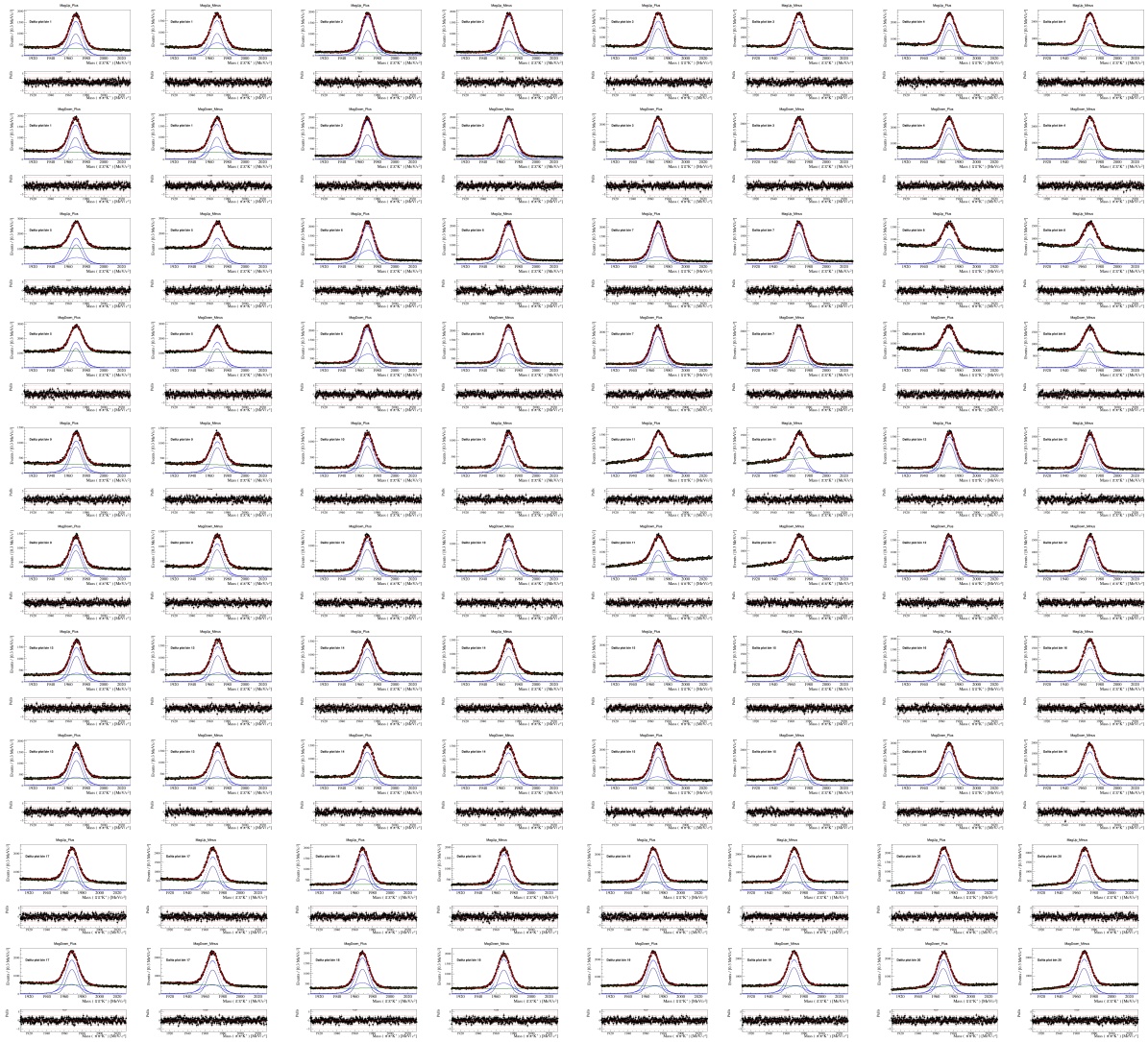

Figura 79 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2016 data sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

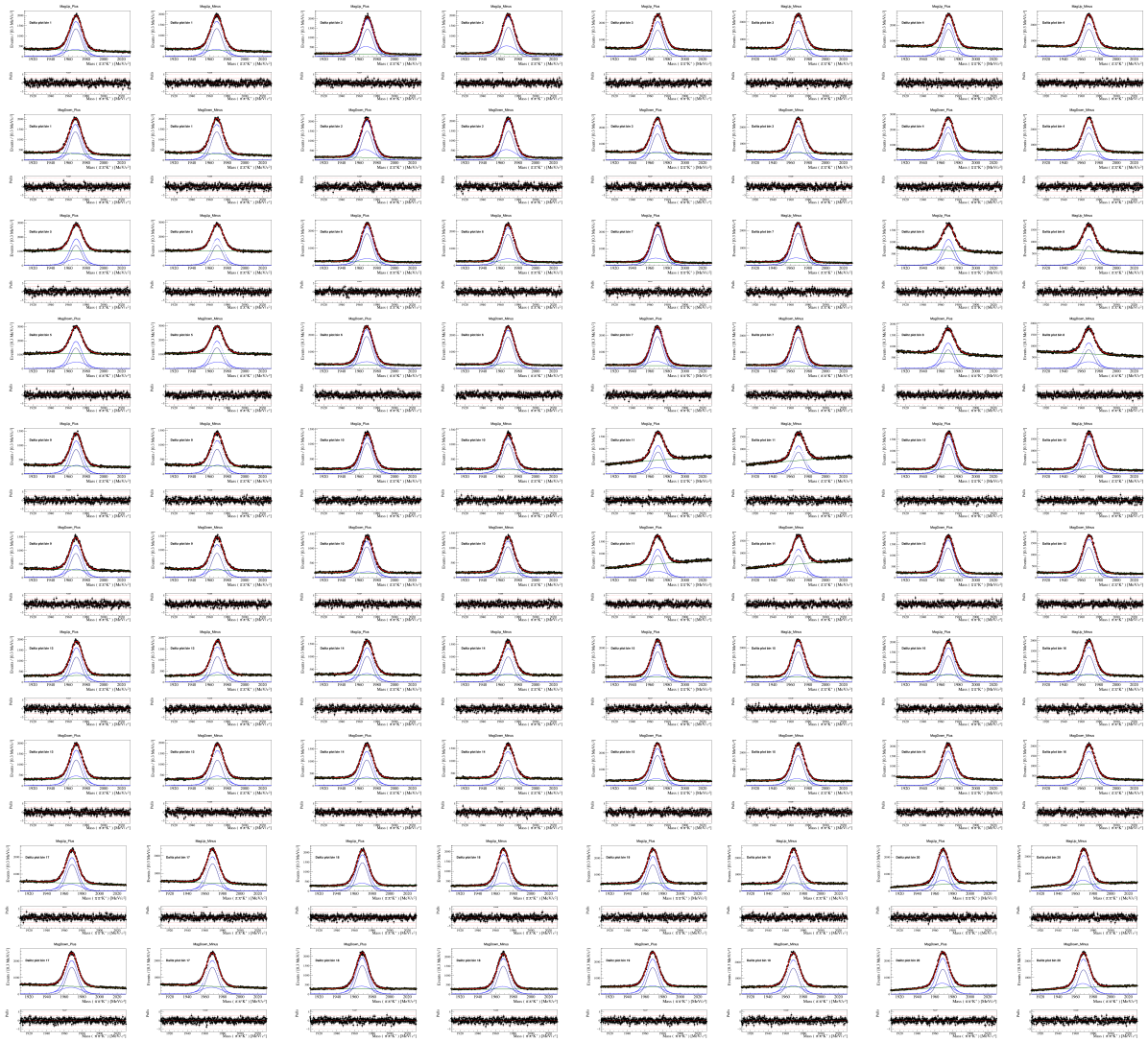

Figura 80 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2017 data sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

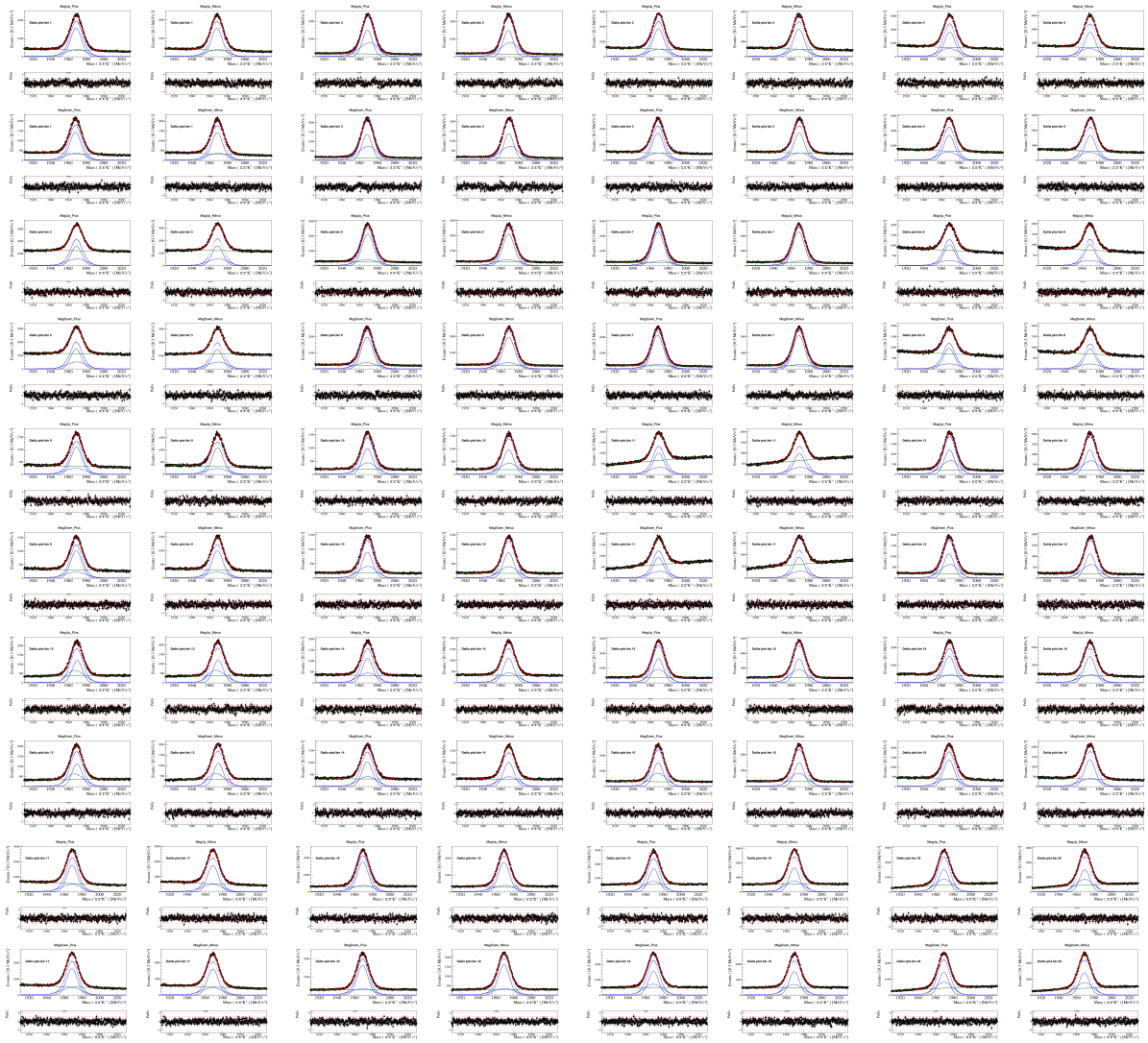

Figura 81 – Nominal fits to $D_s^+ \to \pi^-\pi^{-+}K^+$ invariant mass distribution per Dalitz plot bin for 2018 data sample. From top left for bin 1 to the bottom right position for bin 20, 4 plots are shown, corresponding to the fit results for MagUp on top, and magDown on the bottom (D_s^+ on the left and D_s^- on the right).

D Mirandizing of the RapidSim samples

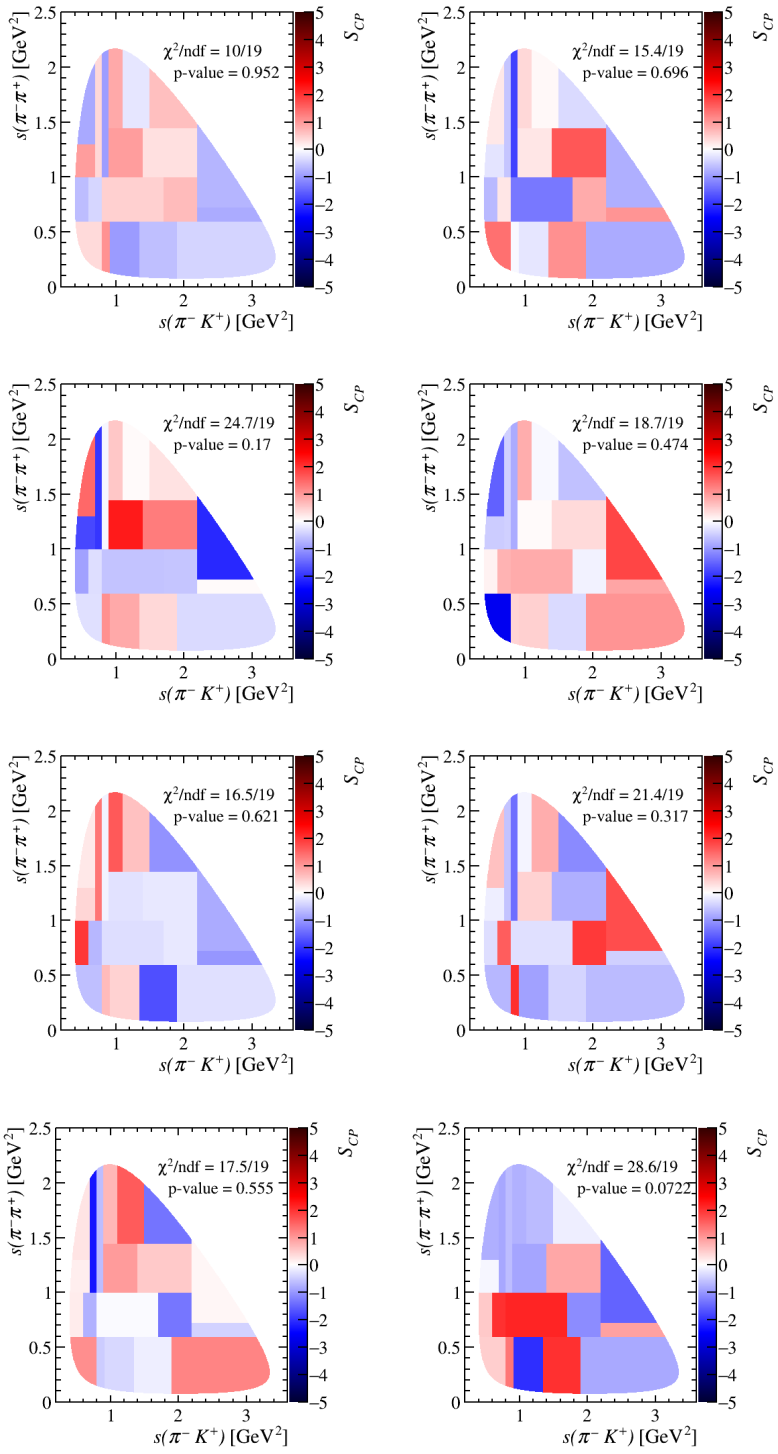

Figura 82 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ ffor the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

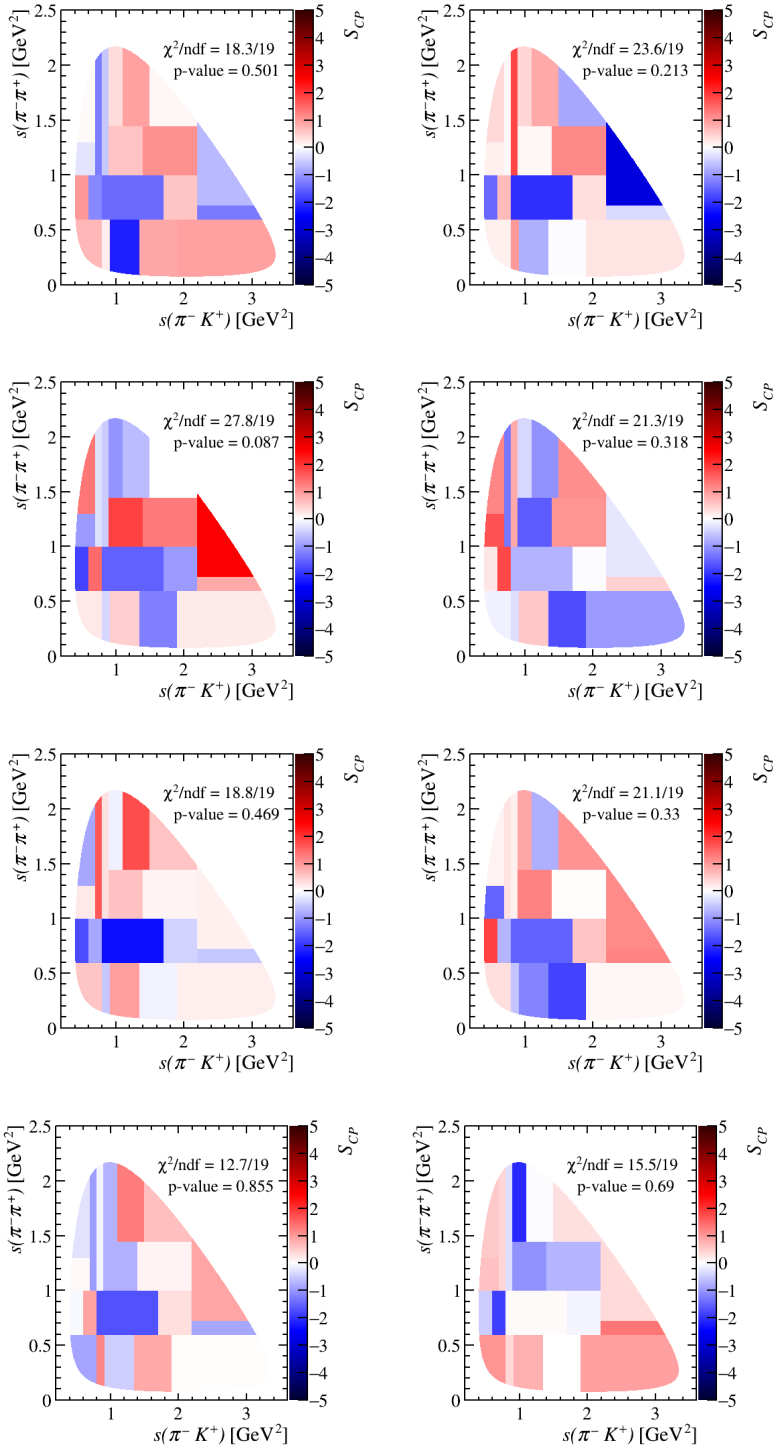

Figura 83 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

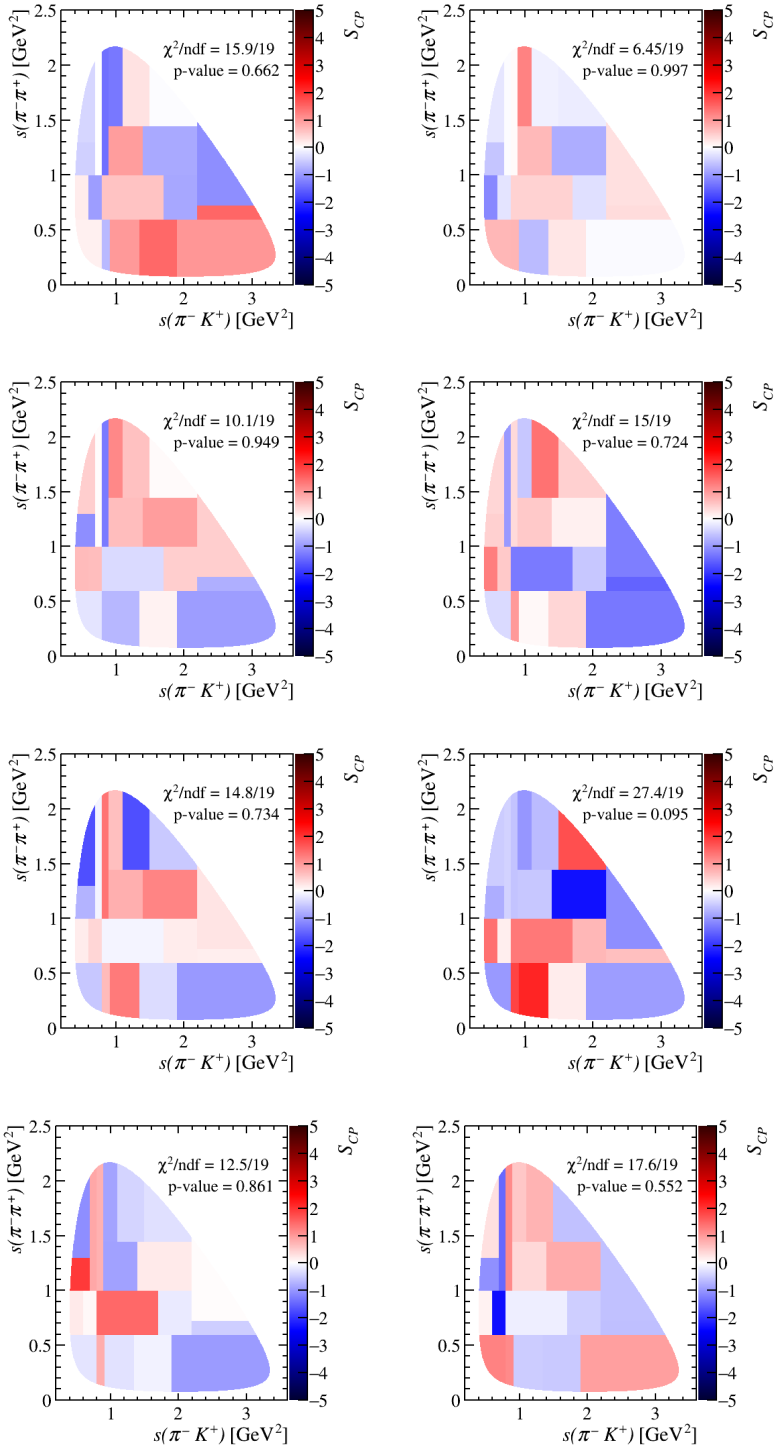

Figura 84 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

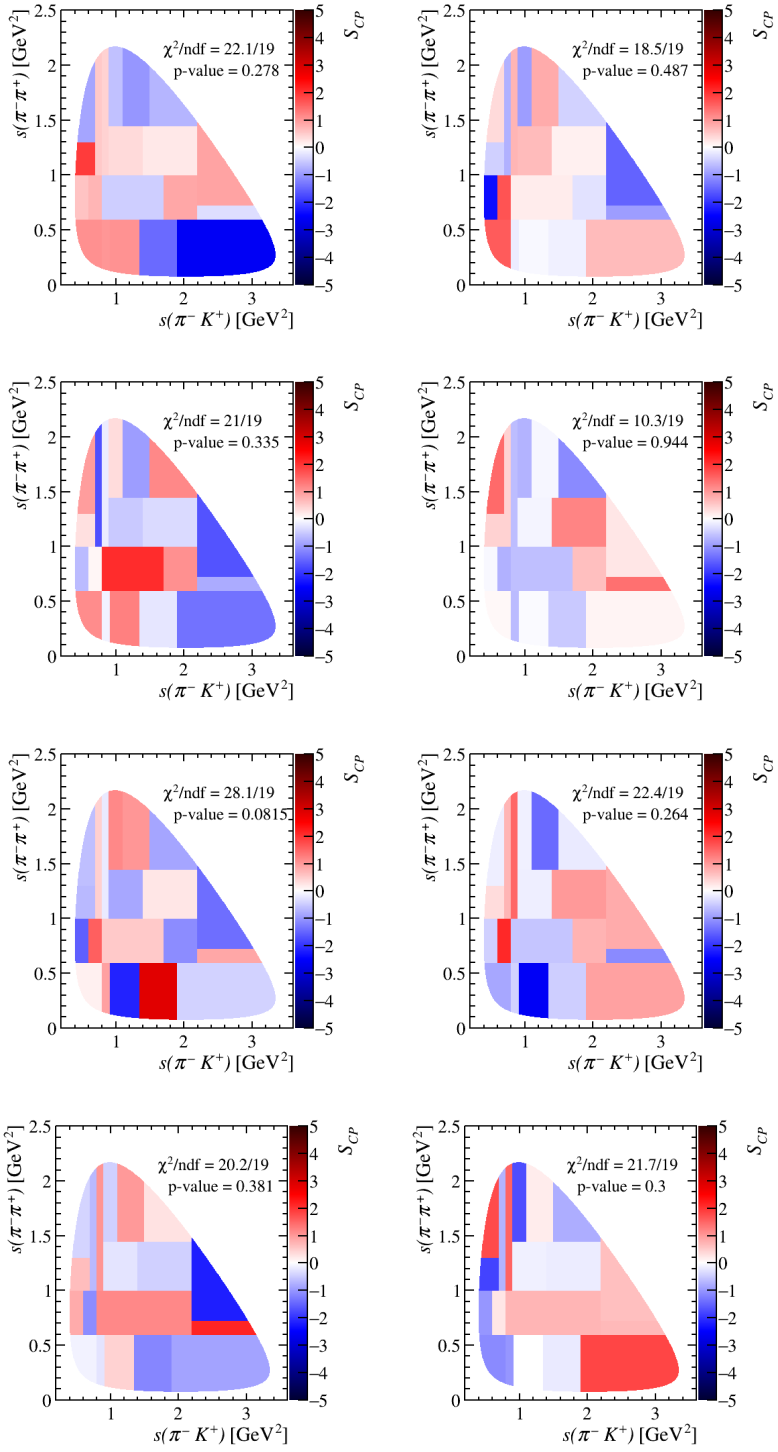

Figura 85 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

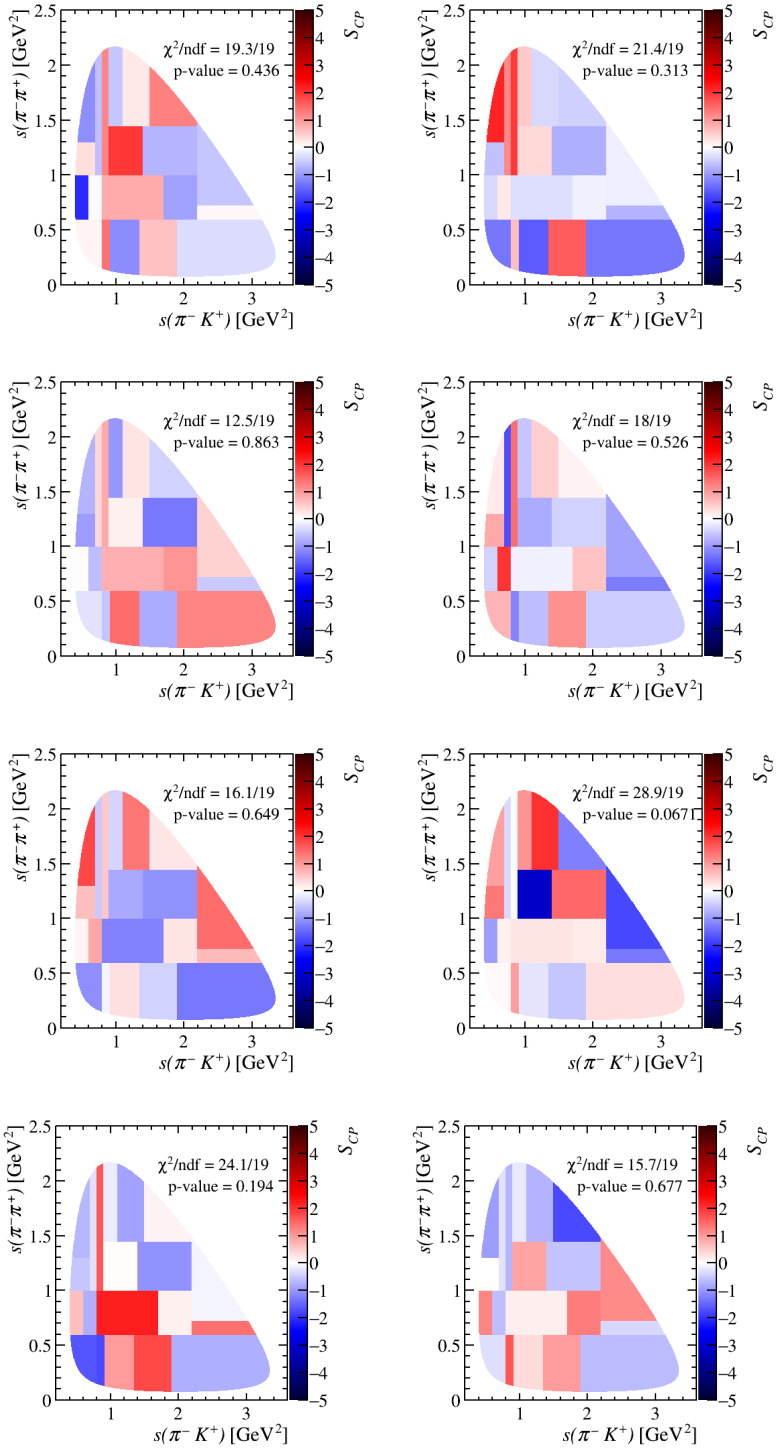

Figura 86 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

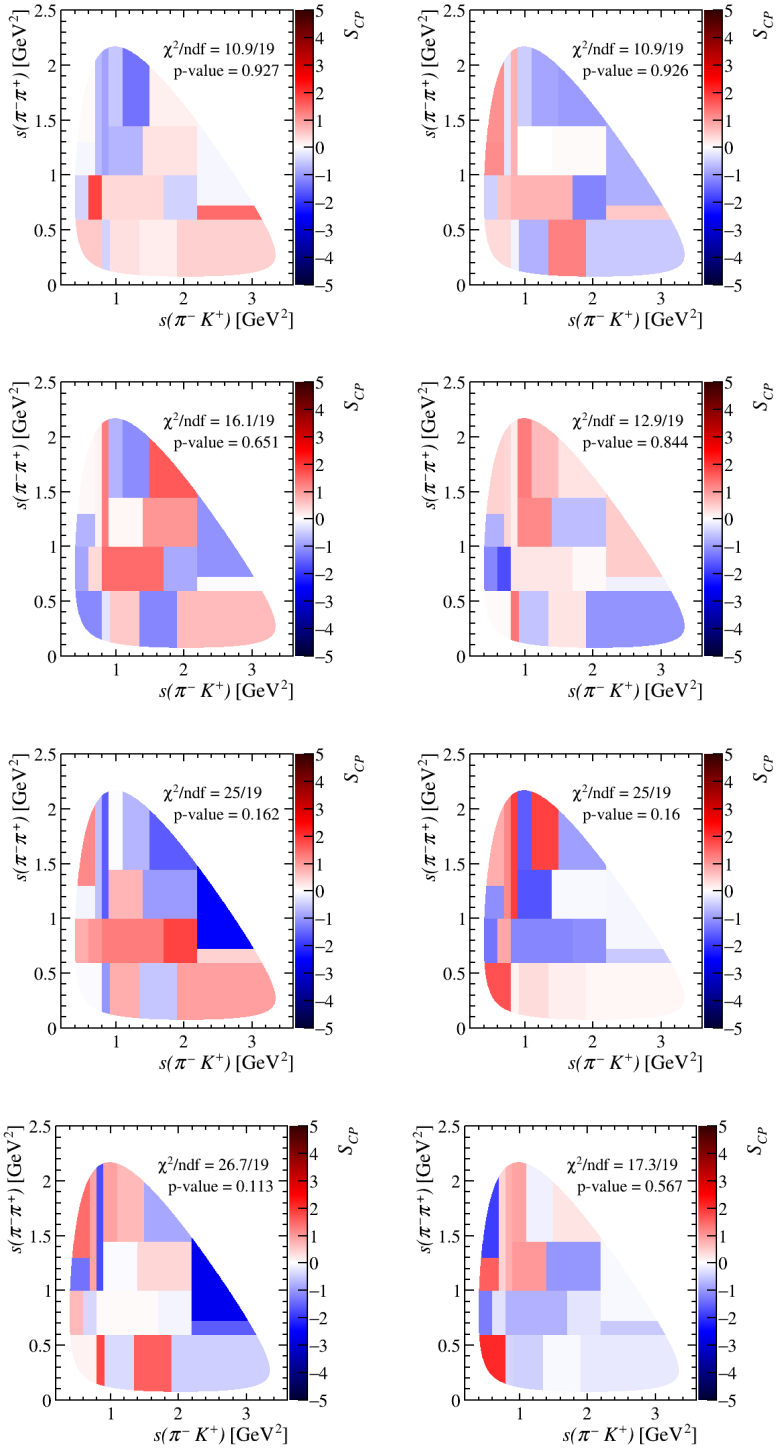

Figura 87 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

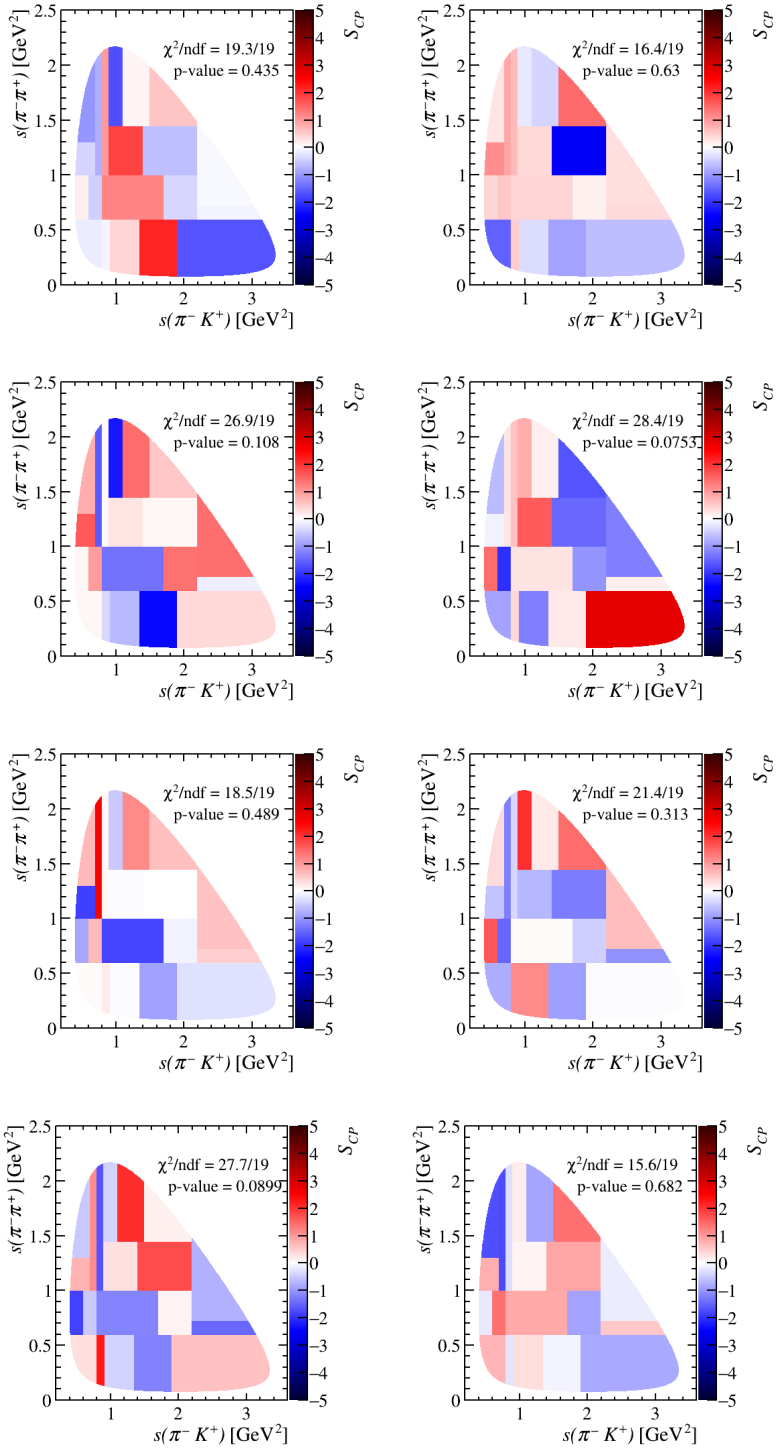

Figura 88 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

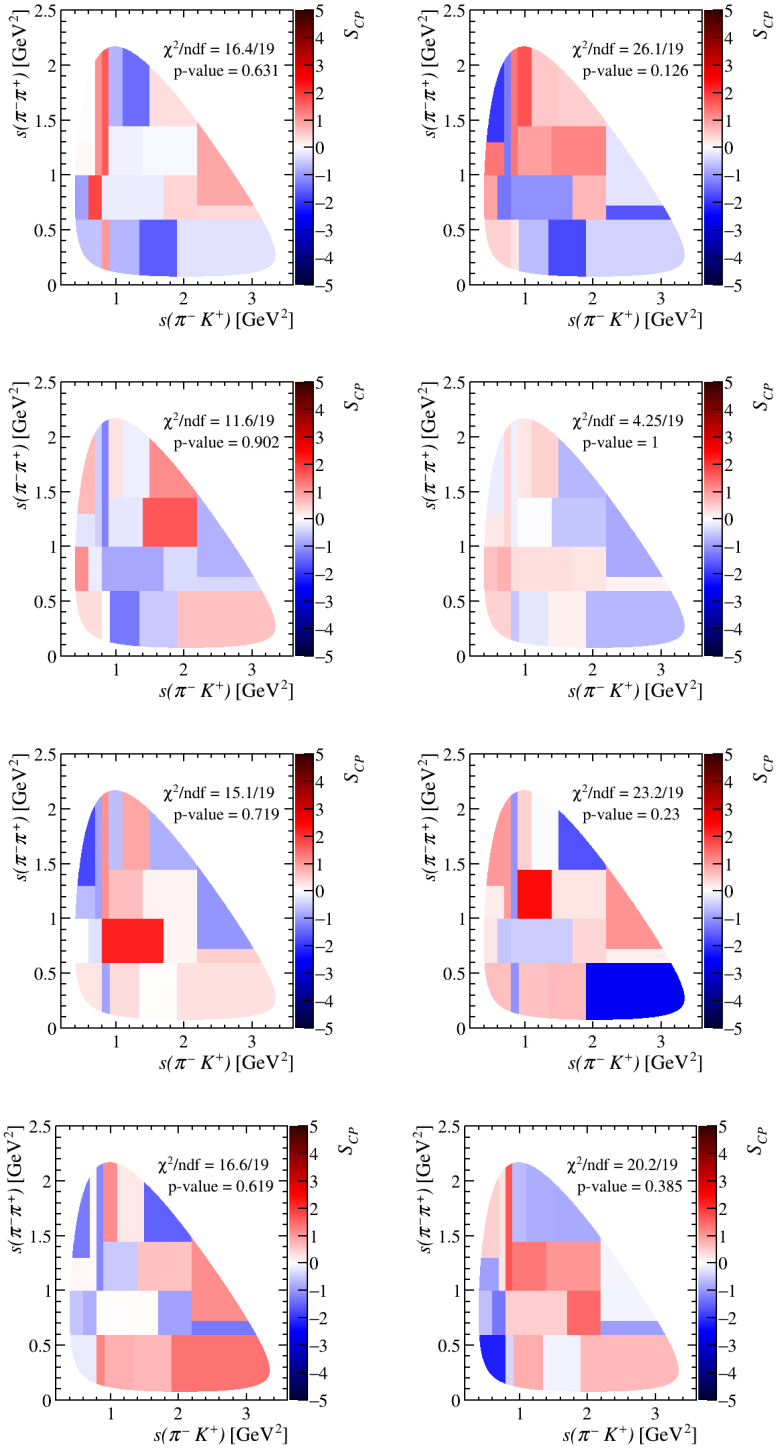

Figura 89 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

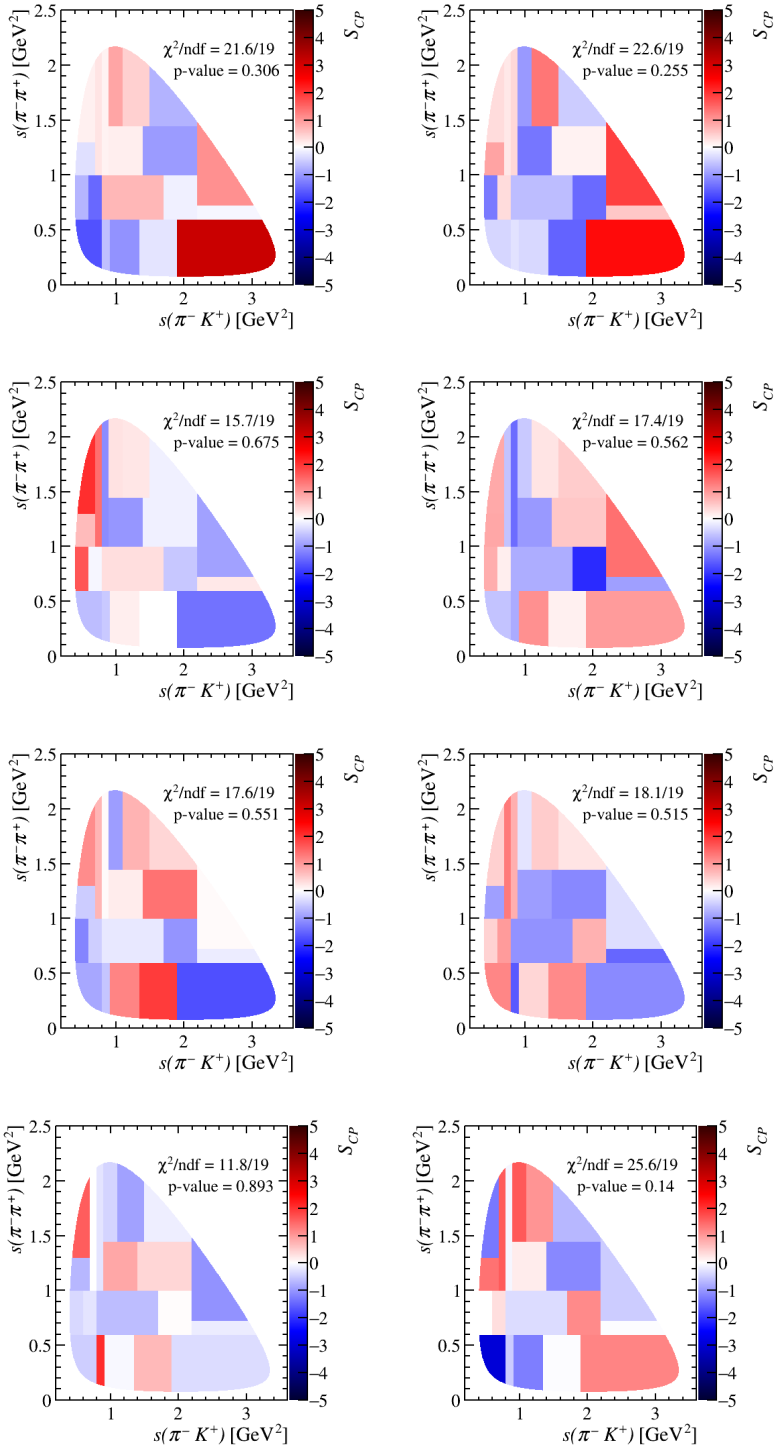

Figura 90 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

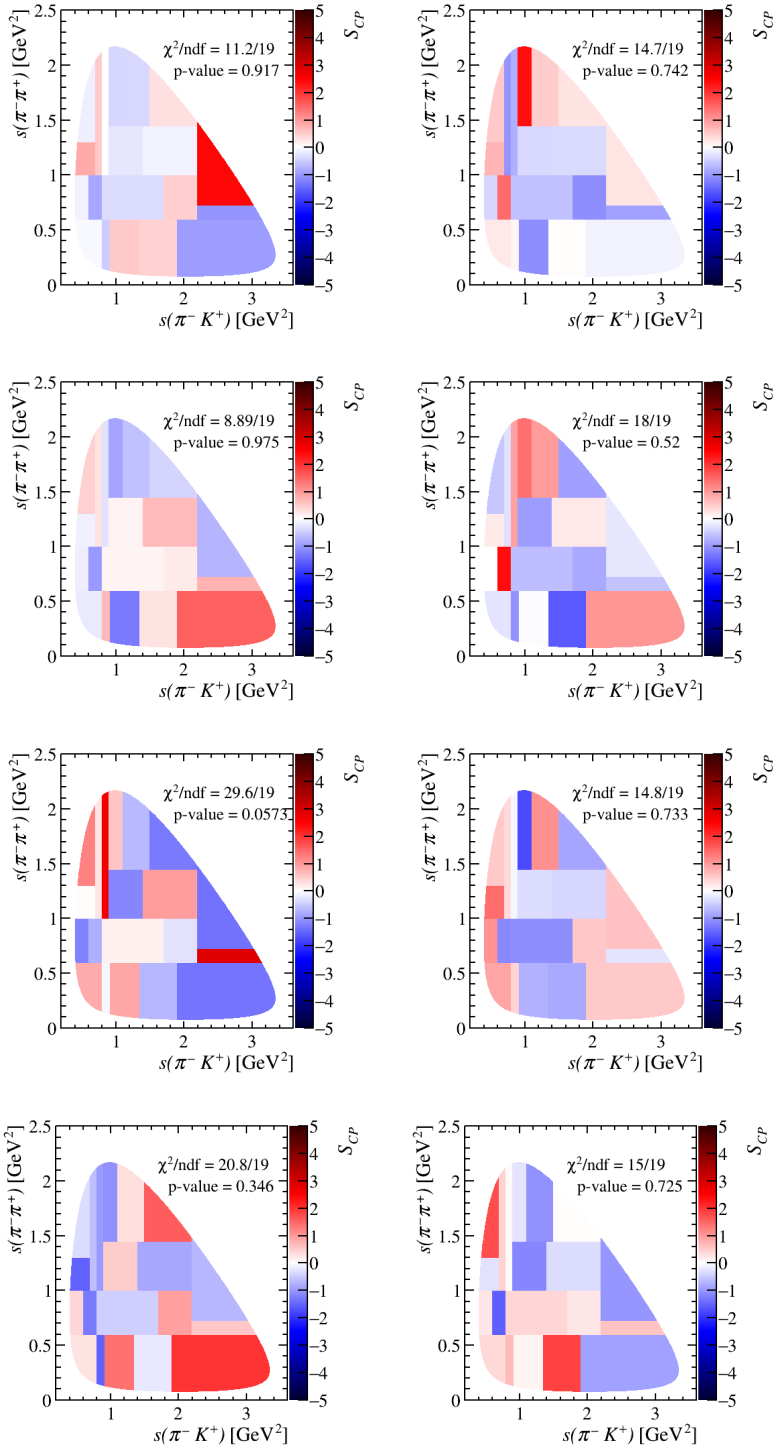

Figura 91 – Results obtained for S_{CP} with RapidSim samples for our channel $D_s^+ \to \pi^-\pi^+K^+$ for the asymmetry models described. In the first line we have the null test (left) and the result for detection (right). In the second line we have PID (left) and the production Run 1 model (right). In the third line we have the production model that mimics Pythia (left) and a result that combines PID and detection (right). In the last line we have a combination of PID+Detection+production Run 1 (left) and a result with PID+detection+production with Pythia model (right).

E Cross-Checks with other channels figures

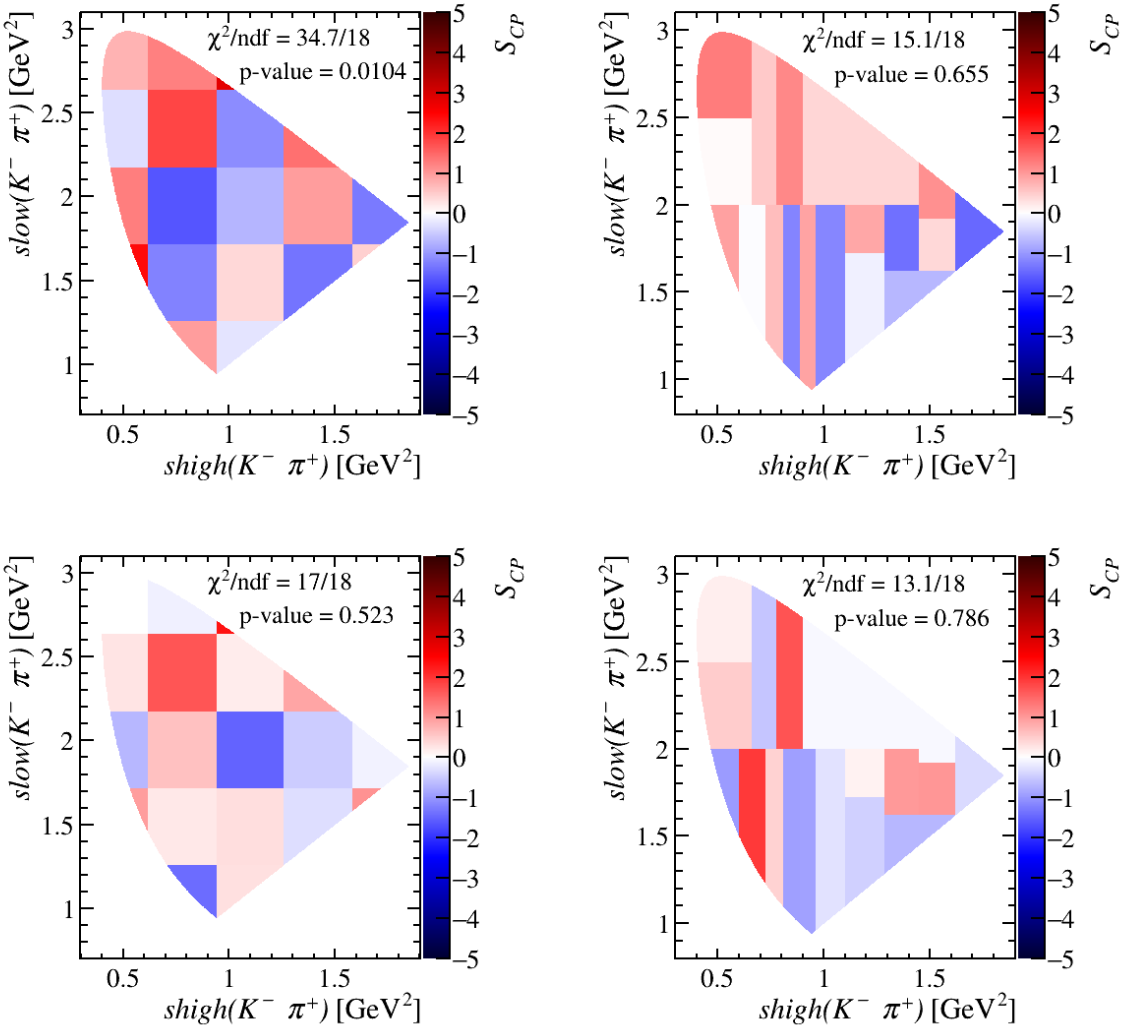

Figura 92 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

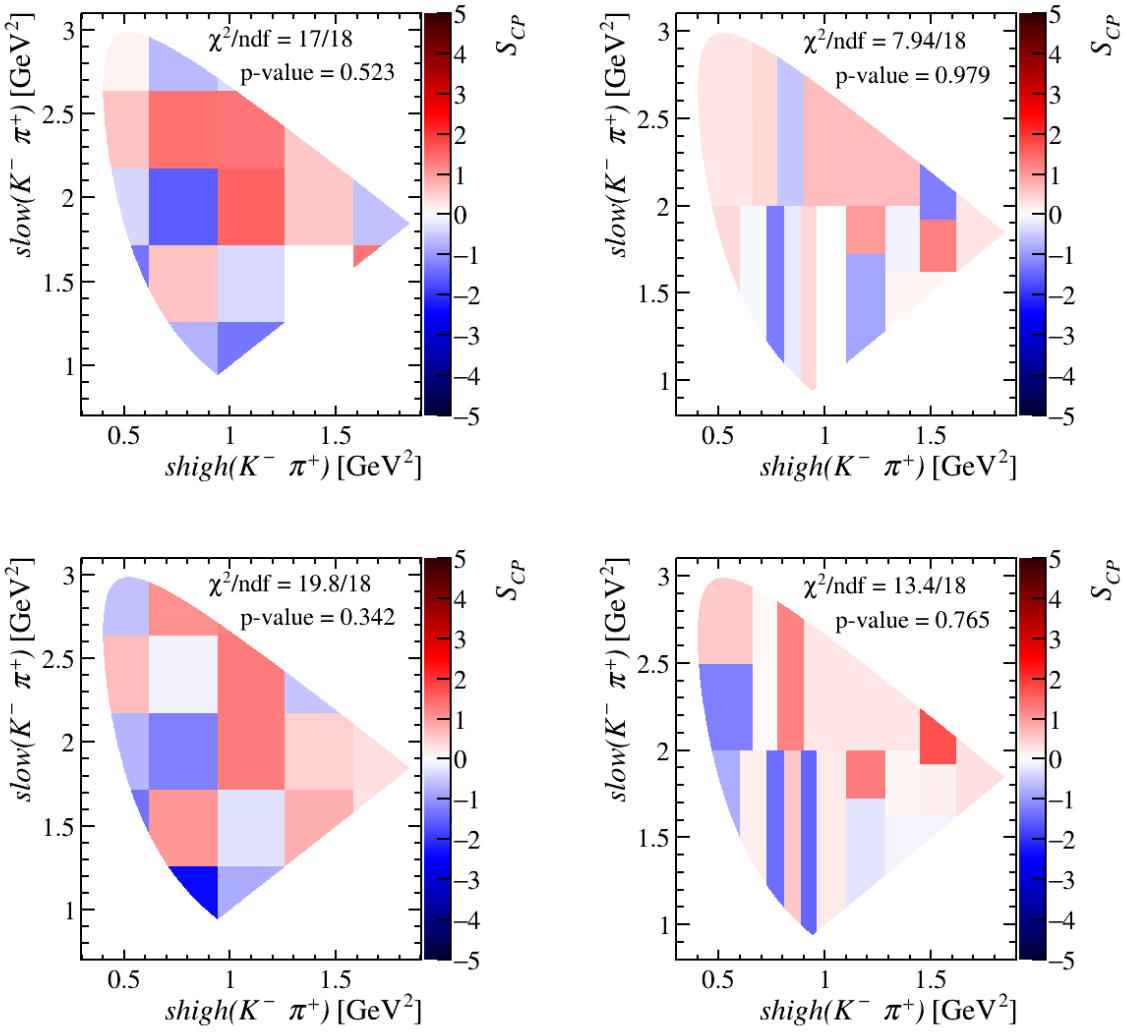

Figura 93 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

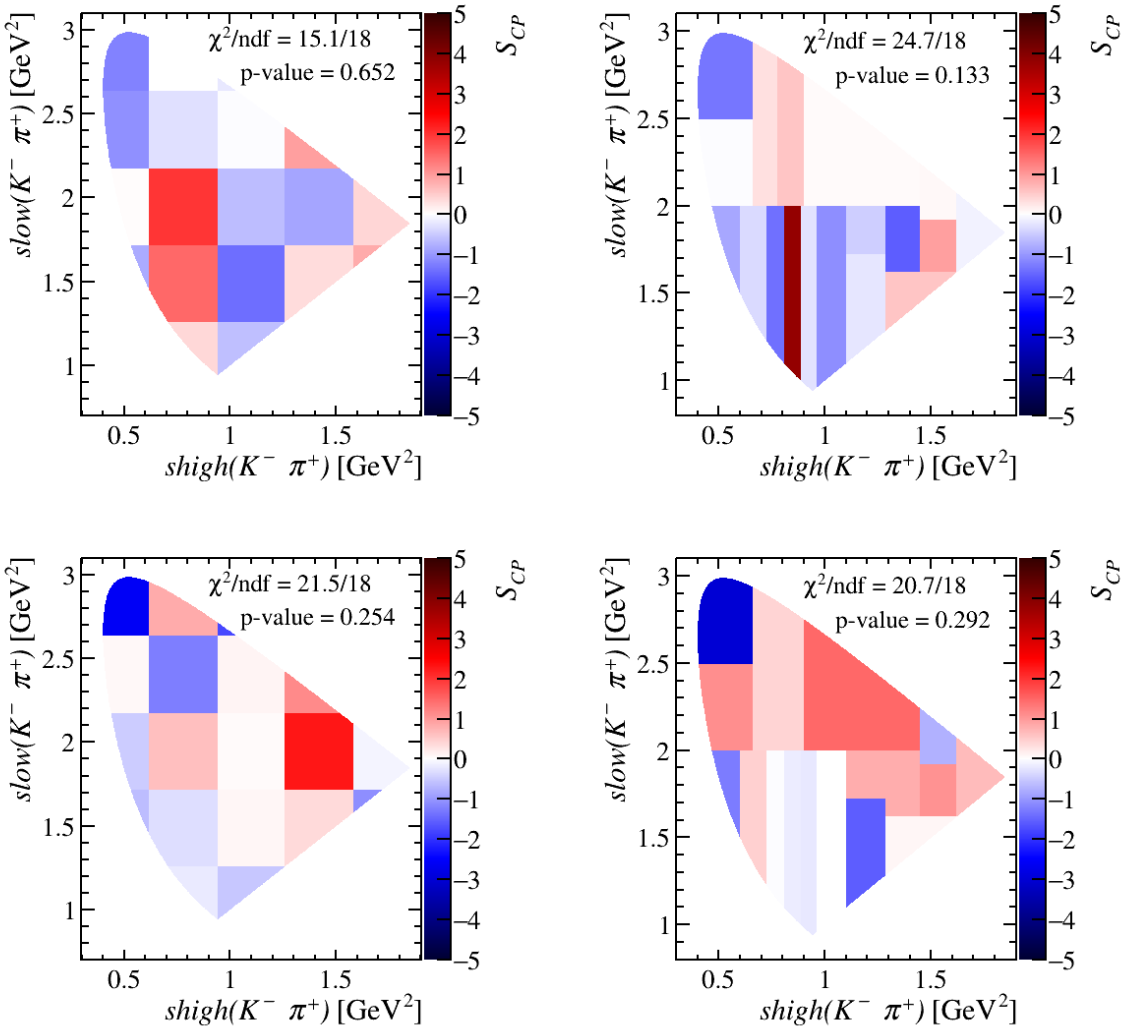

Figura 94 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

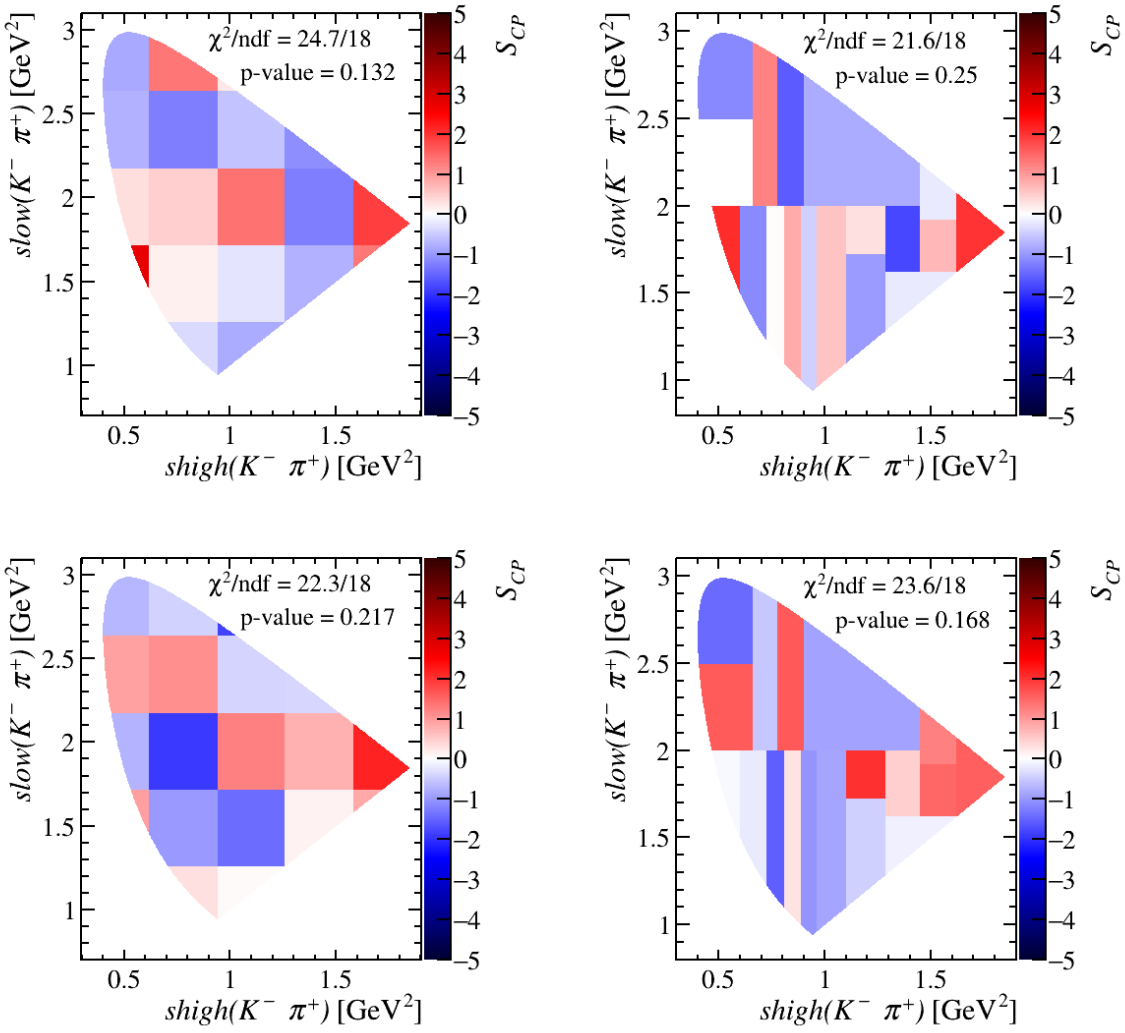

Figura 95 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

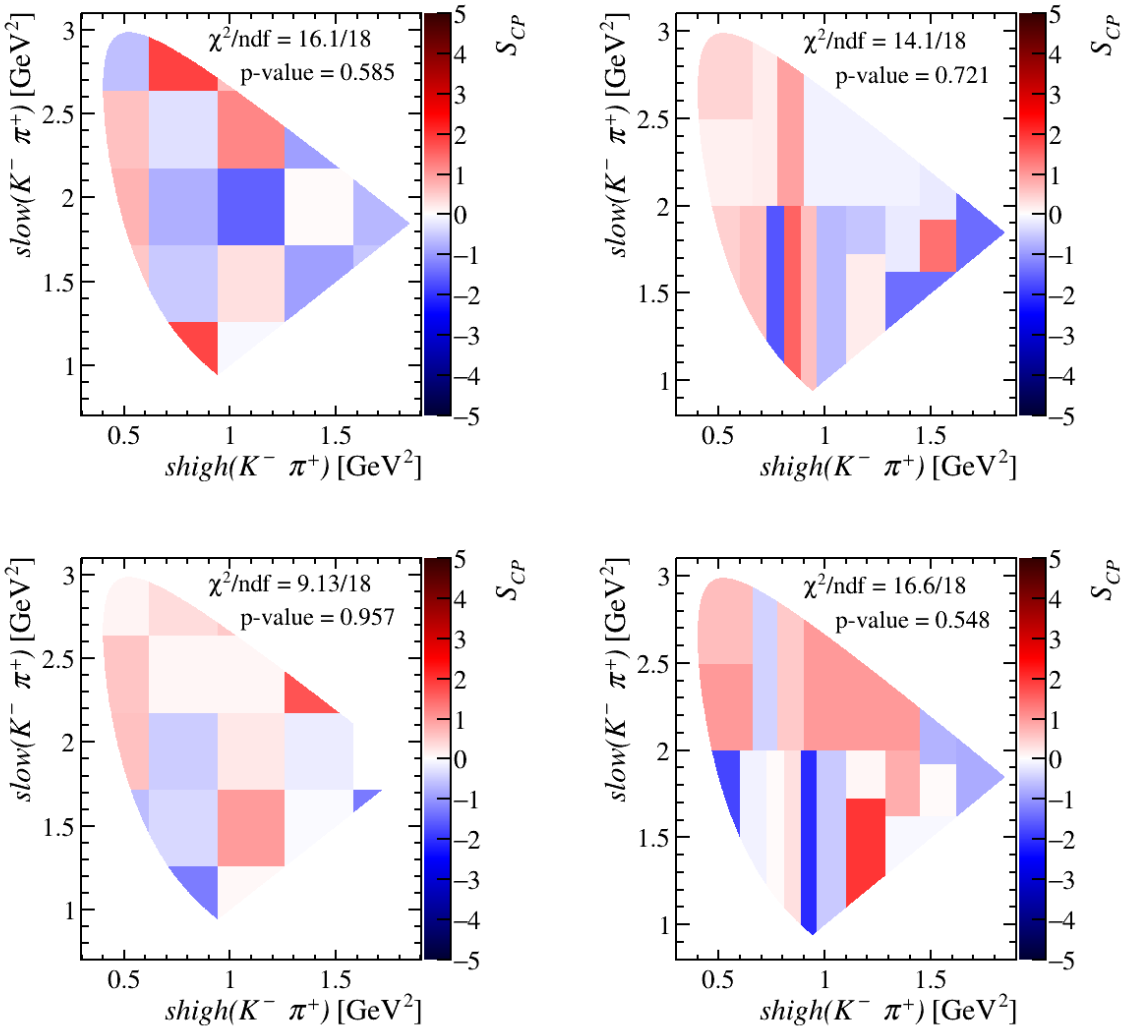

Figura 96 – The Mirandizing method was applied to the decays of $D^+ \to K^-\pi^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

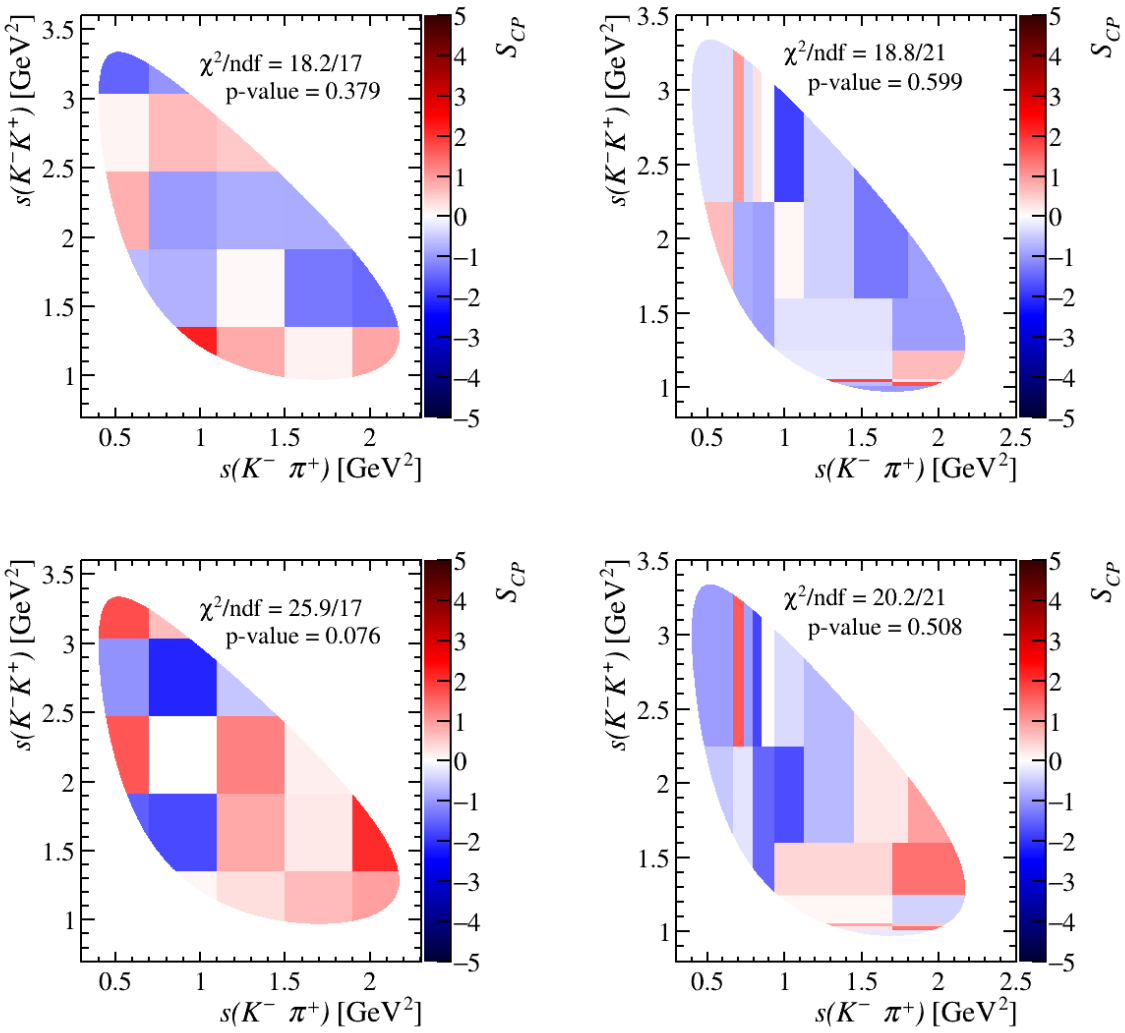

Figura 97 – The Mirandizing method was applied to the decays of $D_s^+ \to K^-K^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

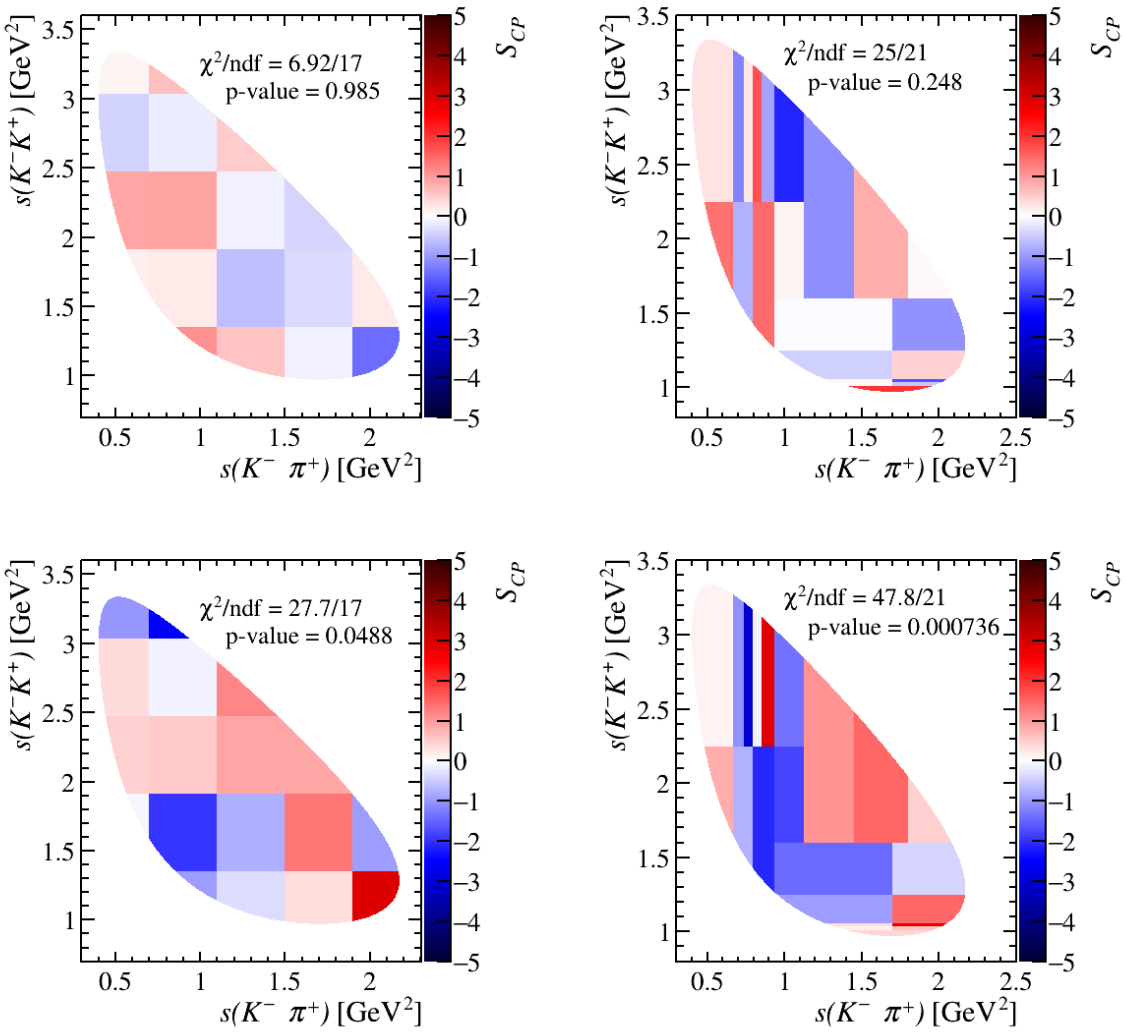

Figura 98 – The Mirandizing method was applied to the decays of $D_s^+ \to K^-K^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

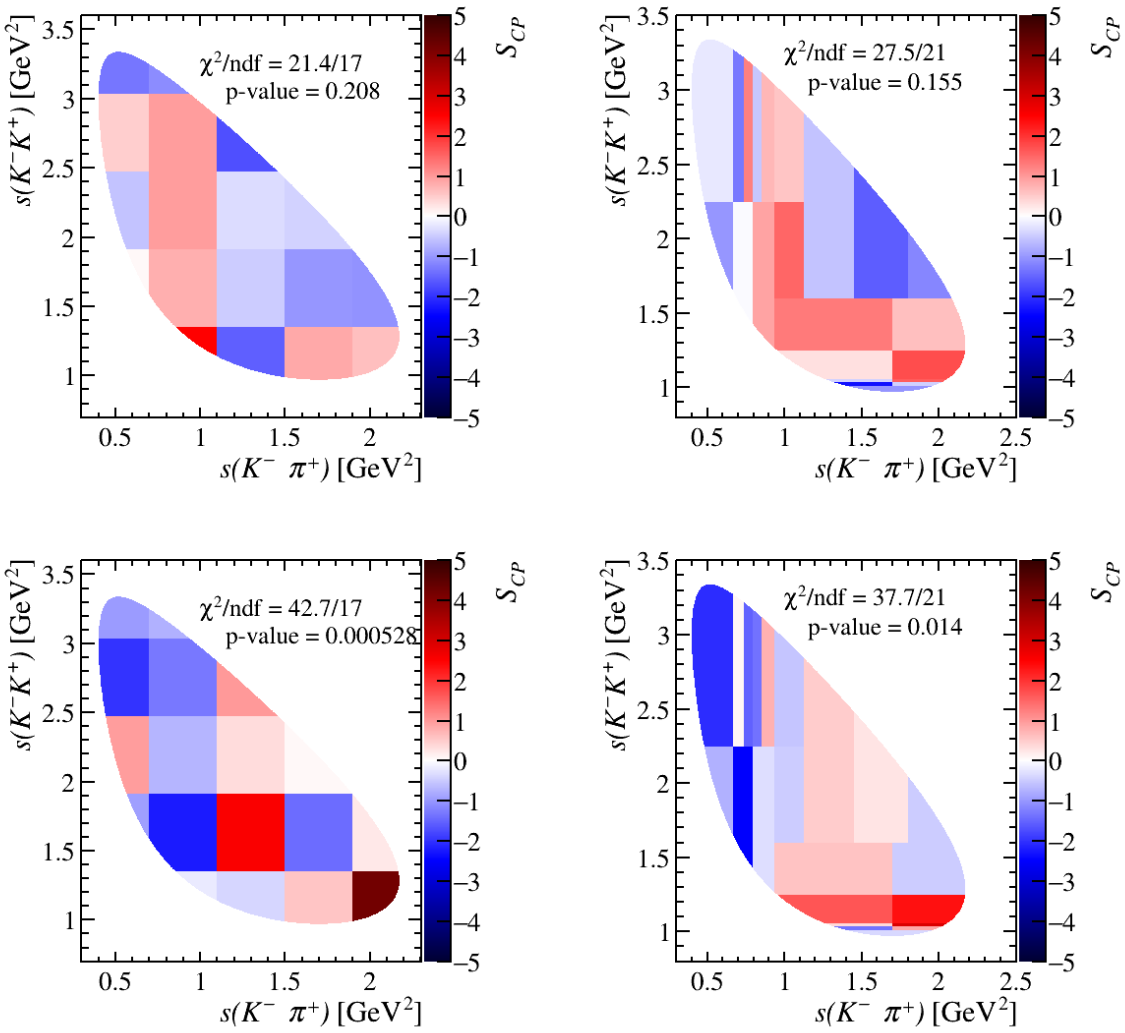

Figura 99 – The Mirandizing method was applied to the decays of $D_s^+ \to K^-K^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

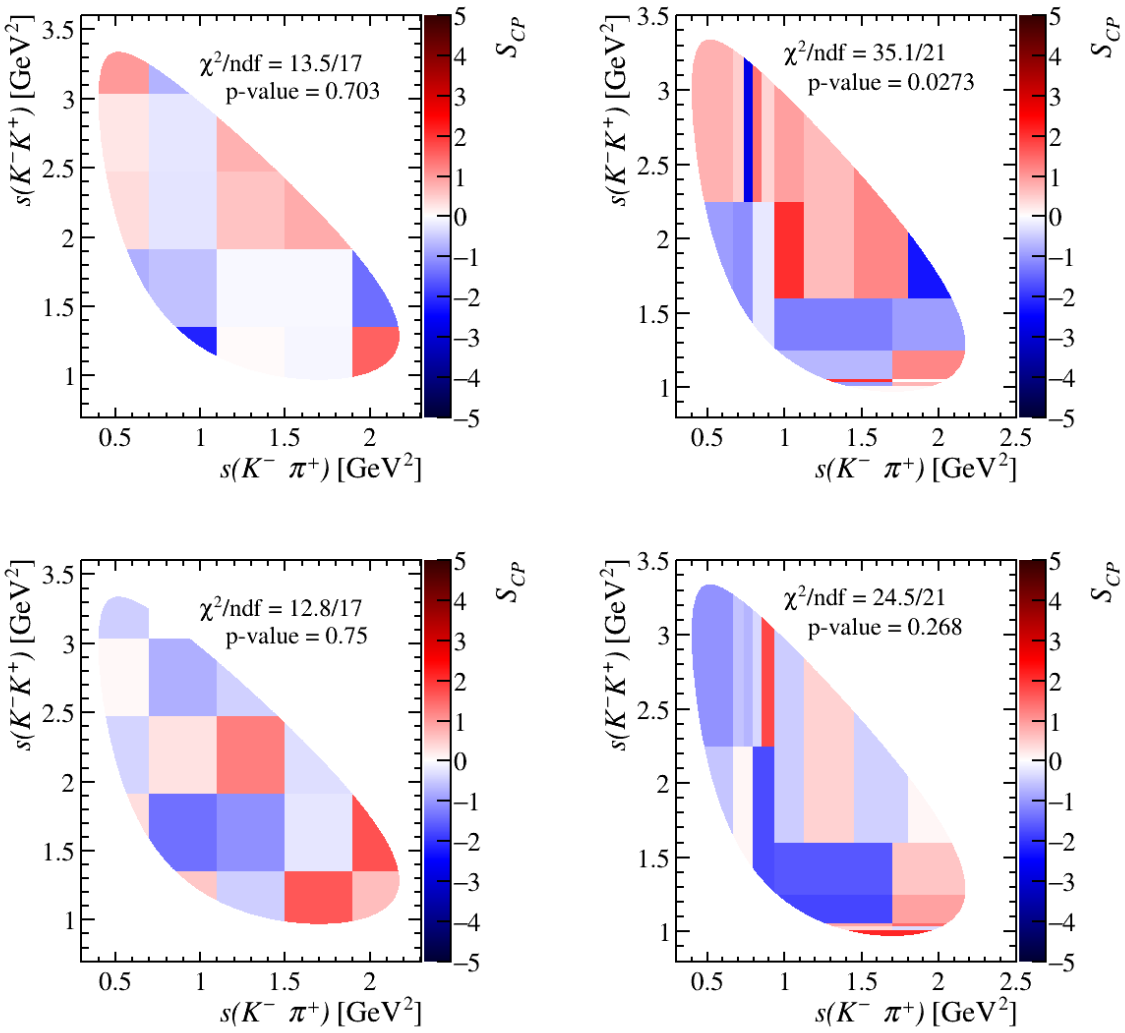

Figura 100 – The Mirandizing method was applied to the decays of $D_s^+ \to K^-K^+\pi^+$. The results were analyzed using uniform and physics binning for all sample. The top two plots represent the null test, while the bottom two plots show the results when the data is separated by charge. On right we have the uniform binning and on left the physics binning.

Referências

[1] Wikipedia contributors, "Standard model," 2026. Accessed: 2026-01-13.

[2] C. Elsasser, "$\bar{b}b$ production angle plots."

[3] A. A. Alves Jr, L. Andrade Filho, A. Barbosa, I. Bediaga, G. Cernicchiaro, G. Guerrer, H. Lima Jr, A. Machado, J. Magnin, F. Marujo, *et al.*, "The LHCb Detector at the LHC," *Journal of instrumentation*, vol. 3, no. 08, p. S08005, 2008.

[4] E. P. Olloqui, L. Collaboration, *et al.*, "LHCb preshower (ps) and scintillating pad detector (SPD): Commissioning, calibration, and monitoring," in *Journal of Physics: Conference Series*, vol. 160, p. 012046, IOP Publishing, 2009.

[5] C. Abellan Beteta, A. Albero, Y. Amhis, S. Barsuk, C. Beigbeder-Beau, I. Belyaev, R. Bonnefoy, D. Breton, O. Callot, M. Gomez, A. Camboni, H. Chanal, D. Charlet, M. Chefdeville, V. Coco, E. Cogneras, A. Comerma-Montells, S. Coquereau, O. Deschamps, and A. Zhokhov, "Calibration and performance of the lhcb calorimeters in run 1 and 2 at the lhc," 08 2020.

[6] L. Dufour, "High-precision measurements of charge asymmetries at LHCb," 2019. Presented 13 Sep 2019.

[7] A. Davis, L. Dufour, L. Dufour, F. Ferrari, S. Stahl, J. Van Tilburg, and M. A. Vesterinen, "Measurement of the instrumental asymmetry for $K^-\pi^+$-pairs at LHCb in Run 2," tech. rep., CERN, Geneva, 2018.

[8] R. e. a. LHCb collaboration, "Measurement of D_s production asymmetry in pp collisions at $\sqrt{s} = 7$ and 8 TeV," *Journal of High Energy Physics*, vol. 2018, Aug. 2018.

[9] LHCb / MooreOnline contributors, "Kali_produceMDF.py — script python do pacote calopi0/pi0_run3_kalipackage." https://gitlab.cern.ch/lhcb/MooreOnline/-/blob/master/MooreOnlineConf/options/CaloPi0/Pi0_Run3_KaliPackage/Kali_produceMDF.py, 2026. Acesso em: 19 fev 2026.

[10] et. al (LHCb Collaboration), "Observation of CP violation in charm decays," *Physical review letters*, vol. 122, no. 21, p. 211803, 2019.

[11] et. al (FOCUS Collaboration), "Study of the doubly and singly cabibbo suppressed decays $D^+ \to K^+ \pi^+ \pi^-$ and $D_s^+ \to K^+ \pi^+ \pi^-$," *Physics Letters B*, vol. 601, no. 1-2, pp. 10–19, 2004.

[12] S. H. Neddermeyer and C. D. Anderson, "Note on the nature of cosmic-ray particles," *Physical Review*, vol. 51, no. 10, p. 884, 1937.

[13] C. M. LATTES, H. Muirhead, G. P. Occhialini, and C. F. Powell, "Processes involving charged mesons," *Nature*, vol. 159, no. 4047, pp. 694–697, 1947.

[14] G. ROCHESTERDr and C. BUTLERDr, "Evidence for the existence of new unstable elementary particles," *Nature*, vol. 160, no. 4077, pp. 855–857, 1947.

[15] Y.-R. Liu, H.-X. Chen, W. Chen, X. Liu, and S.-L. Zhu, "Pentaquark and tetraquark states," *Progress in Particle and Nuclear Physics*, vol. 107, pp. 237–320, 2019.

[16] R. P. Feynman, "Space-time approach to quantum electrodynamics," *Physical Review*, vol. 76, no. 6, p. 769, 1949.

[17] S.-I. Tomonaga and J. R. Oppenheimer, "On infinite field reactions in quantum field theory," *Physical Review*, vol. 74, no. 2, p. 224, 1948.

[18] J. Schwinger, "Quantum electrodynamics. i. a covariant formulation," *Physical Review*, vol. 74, no. 10, p. 1439, 1948.

[19] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson, "Experimental test of parity conservation in beta decay," *Phys. Rev.*, vol. 105, pp. 1413–1415, Feb 1957.

[20] T. D. Lee, R. Oehme, and C. N. Yang, "Remarks on possible noninvariance under time reversal and charge conjugation," *Phys. Rev.*, vol. 106, pp. 340–345, Apr 1957.

[21] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turlay, "Evidence for the 2π Decay of the K_2^0 Meson," *Phys. Rev. Lett.*, vol. 13, pp. 138–140, Jul 1964.

[22] M. Kobayashi and T. Maskawa, "CP violation in the renormalizable theory of weak interaction," *Prog. Theor. Phys.*, vol. 49, pp. 652–657, 1973.

[23] N. Cabibbo and R. Gatto, "Structure of weak interactions and unwanted processes," *Physical Review*, vol. 116, no. 5, p. 1334, 1959.

[24] S. L. Glashow, J. Iliopoulos, and L. Maiani, "Weak interactions with lepton-hadron symmetry," *Phys. Rev. D*, vol. 2, pp. 1285–1292, Oct 1970.

[25] J. Beringer *et al.*, "Particle data group," *Phys. Rev. D*, vol. 86, no. 010001, 2012.

[26] L. Wolfenstein, "Parametrization of the Kobayashi-Maskawa matrix," *Phys. Rev. Lett.*, vol. 51, pp. 1945–1947, Nov 1983.

[27] M. Golden and B. Grinstein, "Enhanced CP violations in hadronic charm decays," *Physics Letters B*, vol. 222, no. 3-4, pp. 501–506, 1989.

[28] Y. Grossman, A. L. Kagan, and Y. Nir, "New physics and CP violation in singly cabibbo suppressed D decays," *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, vol. 75, no. 3, p. 036008, 2007.

[29] H.-Y. Cheng and C.-W. Chiang, "Revisiting CP violation in d→ pp and vp decays," *Physical Review D*, vol. 100, no. 9, p. 093002, 2019.

[30] A. Lenz and G. Wilkinson, "Mixing and CP violation in the charm system," *Annual Review of Nuclear and Particle Science*, vol. 71, no. 1, pp. 59–85, 2021.

[31] R. Aaij, A. Abdelmotteleb, C. Abellan Beteta, F. Abudinén, T. Ackernley, B. Adeva, M. Adinolfi, P. Adlarson, H. Afsharnia, C. Agapopoulou, *et al.*, "Measurement of the time-integrated cp asymmetry in d 0→ k-k+ decays," *Physical review letters*, vol. 131, no. 9, p. 091802, 2023.

[32] P. D. Group, "Review of particle physics," *Prog. Theor. Exp. Phys.*, vol. 2024, p. 083C01, 2024.

[33] L. Collaboration *et al.*, "Search for *cp* violation in $D^+_{(s)} \to K^-K^+K^+$ decays," *arXiv preprint arXiv:2303.04062*, 2023.

[34] E. Byckling and K. Kajantie, *Particle kinematics*. Wiley London, New York, 1973.

[35] R. Aaij, A. Abdelmotteleb, C. Abellan Beteta, F. Abudinén, T. Ackernley, B. Adeva, M. Adinolfi, H. Afsharnia, C. Agapopoulou, C. Aidala, *et al.*, "Direct cp violation in charmless three-body decays of b±mesons," *Physical Review D*, vol. 108, no. 1, p. 012008, 2023.

[36] R. Aaij, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, S. Ali, *et al.*, "Measurement of c p violation in the phase space of b±→ k+ k- π±and b±→ π+ π- π±decays," *Physical Review Letters*, vol. 112, no. 1, p. 011801, 2014.

[37] R. Aaij, B. Adeva, M. Adinolfi, A. Affolder, Z. Ajaltouni, S. Akar, J. Albrecht, F. Alessio, M. Alexander, S. Ali, *et al.*, "Measurements of cp violation in the three-body phase space of charmless b±decays," *Physical Review D*, vol. 90, no. 11, p. 112004, 2014.

[38] N. Kalantar-Nayestanaki, R. Kappert, M. Kavatsyuk, J. Messchendorp, V. Rodin, B. Collaboration, *et al.*, "Amplitude analysis and branching fraction measurement of the decay ds+→ k+ π+ π-," *Journal of High Energy Physics*, vol. 2022, no. 8, p. 196, 2022.

[39] A. Collaboration, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, pp. 1–29, 2012.

[40] G. Aad and *et al.* (ATLAS Collaboration), "The ATLAS experiment at the CERN large hadron collider," *Journal of Instrumentation*, vol. 3, pp. S08003–S08003, aug 2008.

[41] S. *et al.* (CMS Collaboration). Chatrchyan, "The CMS Experiment at the CERN LHC," *JINST*, vol. 3, p. S08004, 2008.

[42] C. Collaboration *et al.*, "A new boson with a mass of 125 gev observed with the cms experiment at the large hadron collider," American Association for the Advancement of Science, 2012.

[43] K. . t. t. Aamodt, "The ALICE experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, pp. S08002–S08002, aug 2008.

[44] F. C. Marcastel, "CERN's accelerator complex," tech. rep., 2013.

[45] L. Evans and P. Bryant, "Lhc machine," *Journal of instrumentation*, vol. 3, no. 08, p. S08001, 2008.

[46] S. Holmes, R. S. Moore, and V. Shiltsev, "Overview of the tevatron collider complex: goals, operations and performance," *Journal of Instrumentation*, vol. 6, no. 08, p. T08001, 2011.

[47] R. t. Aaij, "Performance of the lhcb vertex locator," *Journal of Instrumentation*, vol. 9, Sept. 2014.

[48] R. t. L. C. Aaij, "Measurement of the track reconstruction efficiency at LHCb," *Journal of Instrumentation*, vol. 10, p. 23. P02007, 2015.

[49] LHCB COLLABORATION, "LHCb: Inner tracker technical design report," 2002. CERN-LHCC-2002-029.

[50] LHCB COLLABORATION, "LHCb: Outer tracker technical design report," 2001. CERN-LHCC-2001-024.

[51] LHCB COLLABORATION, "LHCb magnet: Technical design report," 2000. CERN-LHCC-2000-007.

[52] LHCB COLLABORATION, "LHCb: RICH technical design report," 2000. CERN-LHCC-2000-037.

[53] LHCB COLLABORATION, "LHCb calorimeters: Technical design report," 2000. CERN-LHCC-2000-036.

[54] LHCB COLLABORATION, "LHCb muon system technical design report," 2001. CERN-LHCC-2001-010.

[55] "The upgrade i of lhcb velo—towards an intelligent monitoring platform," *Journal of Instrumentation*, vol. 15, p. C06009, jun 2020.

[56] C. Abellan Beteta, "The commissioning of the upstream tracker for the lhcb upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1079, p. 170619, 2025.

[57] L. Gruber, "Lhcb scifi — upgrading lhcb with a scintillating fibre tracker," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 958, p. 162025, 2020. Proceedings of the Vienna Conference on Instrumentation 2019.

[58] LHCb Collaboration, "Design and performance of the LHCb trigger and full real-time reconstruction in Run2 of the LHC," *Journal of Instrumentation*, vol. 14, no. 04, p. P04013, 2019.

[59] R. Brun and F. Rademakers, "ROOT: An object oriented data analysis framework," *Nuclear instruments and methods in physics research section A: accelerators, spectrometers, detectors and associated equipment*, vol. 389, no. 1-2, pp. 81–86, 1997.

[60] https://gitlab.cern.ch/lhcb-datapkg/WG/CharmWGProd/-/blob/master/productions/Run2_D2HHH/D2hhh_Run2Tuples-WithCuts.py?ref_type=heads,. last update July , 2020.

[61] LHCb Collaboration, "Performance numbers - lhcb speakers bureau." https://lhcb.web.cern.ch/speakersbureau/html/PerformanceNumbers.html, 2025. Accessed: 2025-01-28.

[62] G. A. Cowan, D. C. Craik, and M. D. Needham, "Rapidsim: an application for the fast simulation of heavy-quark hadron decays," *Computer Physics Communications*, vol. 214, pp. 239–246, 2017.

[63] J. C. Aaij, Anjos and Appel, "Dalitz plot analysis of $d \to k\pi\pi$ decays," *Phys. Rev. D*, vol. 48, pp. 56–62, Jul 1993.

[64] A. t. Höcker, "TMVA - Toolkit for Multivariate Data Analysis," 2007. http://hal.in2p3.fr/in2p3-00158246, CERN-OPEN-2007-007, LAL 07-75.

[65] M. Pivk and F. R. Le Diberder, "SPlot: A Statistical tool to unfold data distributions," *Nucl. Instrum. Meth. A*, vol. 555, pp. 356–369, 2005.

[66] M. Pivk and F. Diberder, "A statistical tool to unfold data distributions," *arXiv preprint physics/0402083*.

[67] W. Verkerke and D. Kirkby, "The roofit toolkit for data modeling," in *Statistical problems in particle physics, astrophysics and cosmology*, pp. 186–189, World Scientific, 2006.

[68] C. Jarlskog, "Commutator of the quark mass matrices in the standard electroweak model and a measure of maximal CP nonconservation," *Phys. Rev. Lett.*, vol. 55, pp. 1039–1042, Sep 1985.

[69] R. Aaij, B. Adeva, M. Adinolfi, C. Adrover, A. Affolder, Z. Ajaltouni, J. Albrecht, F. Alessio, M. Alexander, S. Ali, *et al.*, "Measurement of c p violation in the phase space of b±→ k±π+ π- and b±→ k±k+ k- decays," *Physical Review Letters*, vol. 111, no. 10, p. 101801, 2013.

[70] I. Bediaga, I. Bigi, A. Gomes, G. Guerrer, J. Miranda, and A. Dos Reis, "On a c p anisotropy measurement in the dalitz plot," *Physical Review D*, vol. 80, no. 9, p. 096006, 2009.

[71] I. Bediaga, J. Miranda, A. dos Reis, I. Bigi, A. Gomes, J. O. Goicochea, and A. Veiga, "Second generation of "miranda procedure" for c p violation in dalitz studies of b (and d and τ) decays," *Physical Review D*, vol. 86, no. 3, p. 036005, 2012.

[72] I. Bediaga, I. Bigi, J. Miranda, and A. dos Reis, "C p asymmetries in three-body final states in charged d decays and c p t invariance," *Physical Review D*, vol. 89, no. 7, p. 074024, 2014.

[73] R. Aaij *et al.*, "Search for CP violation in $D^+ \to K^-K^+\pi^+$ decays," *Phys. Rev. D*, vol. 84, p. 112008, 2011.

[74] R. Aaij *et al.*, "Search for CP violation in the decay $D^+ \to \pi^-\pi^+\pi^+$," *Phys. Lett. B*, vol. 728, pp. 585–595, 2014.