LHCb-ANA-20XX-YYY
October 14, 2021

First observation of $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$ decays

Wojciech Krupa[1]

[1] *AGH University of Science and Technology, Krakow, Poland*

Abstract

The first observation of the $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$ decays is reported using 6.0 fb^-1 of proton-proton collision data collected by the LHCb experiment. The D_s^{*+} mesons are reconstructed through the decay chain $D_s^{*+} \to \gamma$ ($D_s^+ \to K^+ K^- \pi^+$). The branching fraction relative to that for XXXX decays is measured to be:

$$\frac{\mathcal{B}(B_s^0 \to D_s^{*+} K^{*-})}{\mathcal{B}(B_s^0 \to XX)} = XXXX \pm XXX \pm XXXX$$
$$\frac{\mathcal{B}(B_s^0 \to D_s^{*+} K_S^0 \pi^-)}{\mathcal{B}(B_s^0 \to XX)} = XXXX \pm XXX \pm XXXX$$

where the first uncertainty is statistical, and the second is systematic. Using a recent measurement of $B_s^0 \to$ XX the absolute branching fraction of $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$ are measured as

$$\mathcal{B}(B_s^0 \to D_s^{*+} K_S^0 \pi^-) = (XX \pm XX \pm XX(sys) \pm XX(norm)) \times 10^{-5}$$
$$\mathcal{B}(B_s^0 \to D_s^{*+} K^{*-}) = (XX \pm XX \pm XX(sys) \pm XX(norm)) \times 10^{-5}$$

where the third uncertainty is due to the uncertainty on the branching fraction of the normalisation channel.

Contents

1 Introduction

The precise measurement of the angle γ of the Cabibbo-Kobayashi-Maskawa (CKM) unitarity triangle is on the top of the flavour physics program of LHCb experiments. The CKM angle γ is one of the least well-determined CKM parameters and may be extracted using time-independent rates of decays such as $B^+ \to \overline{D}^0 K^+$ or by the time-dependent study of the $B_s^0 \to D_s^{(*)\pm} K^\mp$ decays or the analogous $B_s^0 \to D_s^{(*\pm)} \pi^\mp$. In the time-dependent measurements with the family of $B_{s,d} \to D_{s,d} h$ decays the sensitivity to results from the interference between the $b\overline{q} \to (c\overline{q})(\overline{u}q)$ and $b\overline{q} \to (u\overline{q})(\overline{c}q)$ transitions that occurs through the $B^0\overline{B}^0$ and $B_s^0\overline{B}_s^0$ mixing.

Determination of the CKM angle γ in tree-level open-charm b-hadron decays are theoretically clean. It is a good opportunity to measure sensitive to New Physics beyond the Standard Model effects. The precision obtained at LHCb experiment allows the angle γ to be determined with an uncertainty of around 5°. This precision is obtained by combining various measurements performed at LHCb as the most accurate measurements have an accuracy of 10° to 20°.

The analysis of $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$ decay is believed to be challenging because it requires the selection of soft photon from $D_s^{*+} \to D_s^+ \gamma$ and ($B_s^0 \to D_s^{*+} K^{*-}$ analysis) reconstruction of resonance state K^{*-} which decay to K_S^0 and another pion. Since the K_S^0 mesons are reconstructed at LHCb using a different type of track, the selection of candidates is divided to events with K_S^0 reconstructed using downstream and long track respectively. The D_s^+ meson is reconstructed only $D_s^{*+} \to K^- K^+ \pi^+$ mode.

This decay has not been observed before. In this note, we describe the reconstruction of, $B_s^0 \to D_s^{*+} K^{*-}$, $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and XXXX using the full Run 2 data sample. The XXXX has already been observed, and its branching fraction is. Here we report the first observation of $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$. This is the first step towards of the future measurement of time-dependent CP asymmetry in $B_s^0 \to D_s^{*+} K^{*-}$ decay.

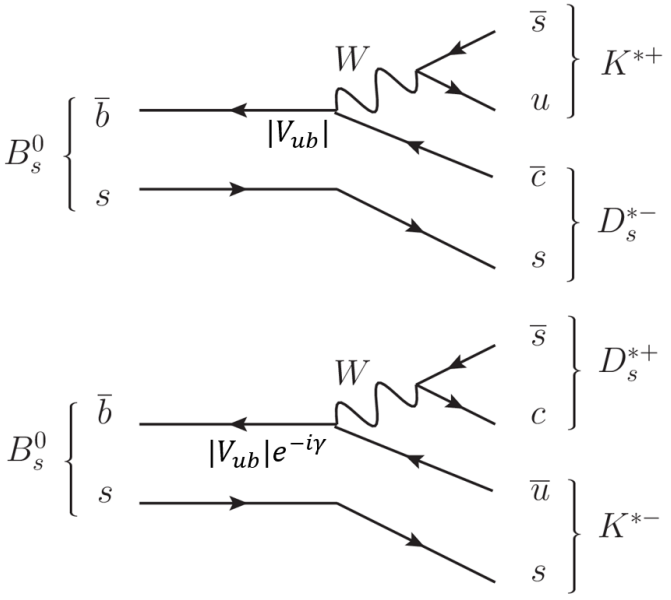

Figure 1: Feynman diagram for $B_s^0 \to D_s^{*+} K^{*-}$ decay.

2 Analysis strategy

This section presents an analysis strategy for branching fraction calculation of $B_s^0 \to D_s^{*+} K_S^0 \pi^-$ and $B_s^0 \to D_s^{*+} K^{*-}$ decays and analysis of reference channel XX . It is organised as follows: Chapter 3 describe data sample, stripping selection, trigger lines used in this analysis and simulated samples. Chapter 4 describes preliminary selection D_s^+, K_S^0, andD^*, multivariate analysis using BDT and post-MVA selection.

3 Data Samples

This analysis is based on a data sample over LHCb Runs 2, collected between 2015-2018. The total integrated luminosity for LHCb during this time is approximately $6.0\,\text{fb}^{-1}$. Data is taken from different stripping lines for different years (Tab.1).

Year	Stripping Used	Integrated luminosity (fb^{-1})	Energy (TeV)
2015	`Stripping24r1`	0.328	6.5
2016	`Stripping28r1`	1.665	6.5
2017	`Stripping29r2`	1.609	6.5
2018	`Stripping34`	X.XX	6.5

Table 1: Summary of the data used in the analysis, where $\sqrt{s}$ represents the center of the mass-energy of the proton-proton collisions

Data is taken from `B02DsstKsPiLLDsst2DGammaD2HHHBeauty2CharmLine` and `B02DsstKsPiDDDsst2DGammaD2HHHBeauty2CharmLine` line in the BHADRON stripping stream which produce two sets of candidates, one set of LL candidates and one set of DD candidates. Below, in the Tab.2-Tab.10 there are listed detailes of stripping selection.

KS0DDDInputsBeauty2CharmFilter	
PT	> 250 MeV
M	$\in (467, 527)$ MeV

Table 2: K_S^0 downstream stripping selection.

PiInputsBeauty2CharmFilter	
Track χ^2	< 4
PT	> 100 MeV
P	> 1000 MeV
Min. IP	> 4
TRGHP	< 0.4

Table 3: π from D_s^+ stripping selection.

PiPIDTopoInputsBeauty2CharmFilter	
PT	> 500 MeV
P	> 5000 MeV
PiPIDInputsBeauty2CharmFilter	
PIDK	< 20

Table 4: Additional π stripping selection.

GammaBeauty2CharmFilter	
PT	> 90 MeV
CL	> 0.25

Table 5: γ stripping selection.

PiInputsBeauty2CharmFilter	
χ^2 śladu	< 4
PT	> 100 MeV
P	> 1000 MeV
Min. IP	> 4
TRGHP	< 0.4

KInputsBeauty2CharmFilter	
χ^2 śladu	< 4
PT	> 100 MeV
P	> 1000 MeV
Min. IP	> 4 mm
TRGHP	< 0.4

D2HHHPIDBeauty2CharmFilter	
Liczba π: PIDp < 10	$= 0$
Liczba π: PIDK > 20	$= 0$
Liczba K: PIDK < -10	$= 0$

Table 6: π and K from D_s^+ stripping selection.

Dsst2DGammaCPVD2HHHBeauty2Charm	
$\lvert\Delta_M\rvert$	$\in (80, 250)$ MeV

Table 7: D_s^{*+} stripping selection.

B02DsstKsPiDDDsst2DGammaD2HHHBeauty2Charm	
χ^2 śladu	< 4
DIRA	> 0.999
τ	> 0.2 ps
IP χ^2	< 25
PT	> 500 MeV
P	> 5000 MeV
BPVVDCHI2	> 1000
VFASPF(VCHI2/VDOF)	< 10

Table 8: B_s^0 stripping selection.

First B_s^0 daughter	
P	> 10000 MeV
PT	> 1700 MeV
χ^2 śladu	< 4
Min. IP	< 16
Min. IP PV	> 0.1 mm

Table 9: B_s^0 daughter stripping selection.

Second B_s^0 daughter	
χ^2 śladu	< 4
PT	> 500 MeV
P	> 5000 MeV
or	
ID	K_s^0
PT	> 500 MeV
P	> 5000 MeV

Table 10: B_s^0 daughter stripping selection.

The additional requirement was made in $B_s^0 \to D_s^{*+} K^{*-}$ analysis. Since there is no stripping selection for the K^{*-} the preselection of K^* base on the cut on invariant mass of $K_S^0\pi^-$: $m(K_S^0\pi^-) < 1400$ MeV. The Fig. 2 present invariant mass of $K_S^0\pi^-$ candidates after $m(K_S^0\pi^-) < 1400$ MeV for 2015 data sample. The red colour indicates 75 MeV mass windows around the nominal mass of K^{*-}. Only events with K^{*-} mass within this mass window will be considered in the Bs$\to D_s^{*+} K^{*-}$ analysis.

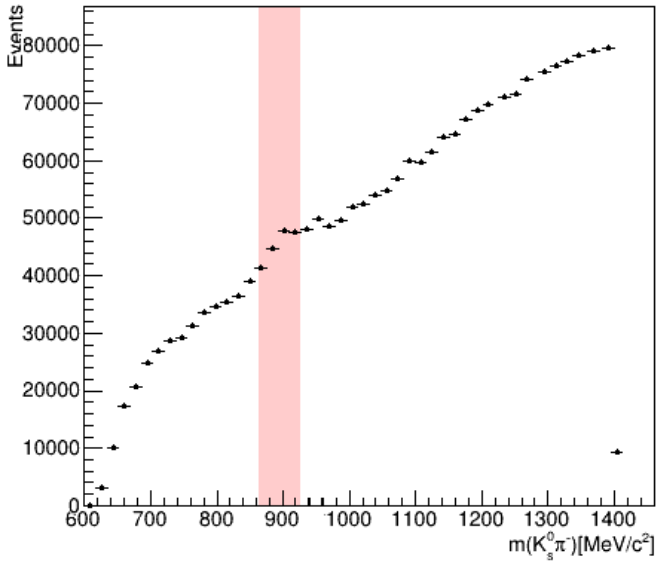

Figure 2: Invariant mass of $K_S^0\pi^-$ candidates for 2015 data sample with $m(K_S^0\pi^-) < 1400$ MeV cut. The red color indicate 75 MeV mass windows around nominal mass of K^{*-} (892)

3.1 Trigger

All candidates used in this analysis are required to have passed the following triggers:

- `L0HadronDecision`
- `L0MuonDecision`
- `L0DiMuonDecision`
- `L0ElectronDecision`
- `L0PhotonDecision`
- `Hlt1TrackMVADecision`
- `Hlt1TwoTrackMVADecision`
- `Hlt2IncPhiDecision`
- `Hlt2Topo2BodyDecision`
- `Hlt2Topo3BodyDecision`
- `Hlt2Topo4BodyDecision`

3.2 Simulated events

Several Monte Carlo simulated samples were used in several parts of this analysis. These samples have been divided due to the method of production. MC is generated for both the signal mode and all considered background modes.

3.2.1 Full simulated events

Full simulation samples have been generated for signal and background modes. Both magnet polarities were generated, the 2016 sample was generated with Pythia8 and Sim09, the 2020 sample was generates with PythiaX and SimX. All samples were generated with all daughters in LHCb acceptance. The signal sample is produced with the D_s^+ decay to $D_s^+ \to K^+K^-\pi^+$. The specific MC data samples requested may be found in Tab.11.

Event type number	Year	Decay	Options	Type
13266301	2016	$B_s^0 \to D_s^{*+} K^{*-}$	`sqDalitz, DecProdCut`	DST
XXXXXXX	2020	$B_s^0 \to D_s^{*+} K_S^0 \pi^-$	`sqDalitz, DecProdCut`	
XXXXXXX	2020	$B_s^0 \to D^+ K_S^0 \pi^-$	`sqDalitz, DecProdCut`	
13264221	2018	$B_s^0 \to D_s^{*+} \pi^-$	`sqDalitz, DecProdCut`	DST

Table 11: Full MC samples used in analysis

3.2.2 Simplify simulated events

The simplify simulated events have been generated using RapidSim simulation package. RapidSim allows for the fast simulation of phase space decays of beauty and charm quark hadrons, allowing quick studies of the properties of signal and background decays in physics analysis. RapidSim features generations of events using different identification hypotheses, different range of momentum, and angular acceptance of the detector. The specific MC data samples generated using RapidSim may be found in Tab.12.

Sample Number	Decay	Options
1	XXXXXXX	`option`

Table 12: Simplify MC samples used in analysis

3.2.3 MC application in MVA

As the simulated events pass the stripping line condition, they represent the decay signal and can be used in training of multivariate analysis method. As the number of events useful in Multivariate Analysis' method training was low (what can result in overtraining the method ..... , *multiplication?*).

3.2.4 MC application in efficiency calculation

Another application of simulated events is a calculation of efficiency of both selection criteria and BDT method. Efficiency is calculated for each criterion as the number of MC events that pass the requirements to the total number of simulated events. Since the expected number of data events is low, both efficiencies should be as big as possible. Selection efficiency is used in the calculation of total efficiency, which is taken into account during relative branching fraction calculation.

4 Offline Selection

The offline selection focuses on selecting intermediate states D_s^+ and K_S^0 and resonance states D_s^{*+} and K^{*-}. Tables and summarize all selection requirements, which are described in the following paragraph. Tables 3.1 and 3.2 summarize all selection requirements. The high number of proton-proton interactions during data talking resulted in a large fraction of events having more than one reconstructed PV. These events are called multiple events. There are no multiply events after all selection stages in this analysis, and we didn't perform the best candidates selection. More information can be found in section 4.4. The preliminary selection of candidates is divided into a selection of each type of particle in the full decay chain.

4.1 Selection of K_S^0

In LHCb K_S^0 mesons can be reconstructed using a different type of tracks. The type of track depends on detectors, which particle pass afters proton-proton collision in LHC. Long tracks are reconstructed from hits recorded in VELO, TT detector, and T stations. Downstream are a track of particle which did not pass VELO detector and are reconstructed only from TT and T stations hits. As a result, this particle's mass resolution is worse than for the long track, but the number of candidates is significantly bigger. The number and quality of candidates imply separate selection for long (LL) tracks and downstream (DD).

	Variable	Cut	Justification
DD	$m(\pi\pi)$	$= m_{K_S^0(PDG)} \pm$ 20 MeV	Selection genuine K_S^0
	K_S^0 FD_{sig}	> 7	Removing random combinations
LL	$m(\pi\pi)$	$= m_{K_S^0(PDG)} \pm$ 20 MeV	Selection genuine K_S^0

Table 13: Offline selection requirements for K_S^0 candidates.

The FD_{sig} stands for fight distance significance defind as:

$$FD_{sig} = \frac{z_{K_S^0} - z_B}{\sqrt{\sigma_{K_S^0}^2 - \sigma_B^2}} \tag{1}$$

where $\sigma_{K_S^0}$ and σ_B are the errors in the z position of the K_S^0 and B decay vertex. Flight distance significance is defined as the z distance between the particle and B_s^0 end vertices divided by the sum in the quadrature of this position's uncertainties. The same requirement is also made for D_s^+ mesons to separate prompt D_s^+ mesons created in proton-proton collisions from that one from B mesons decays.

The efficiency of selection criteria is calculated using simulated events. Table 14 show efficiency of each selection requirement for K_S^0 candidates.

4.2 Selection of D_s^+

The D_s^+ meson can be reconstructed using 3 decay modes: $D_s^+ \to K^+K^-\pi^+$, $D_s^+ \to K^+\pi^-\pi^+$ and $D_s^+ \to \pi^+\pi^-\pi^+$ since the branching fraction for $D_s^+ \to K^+K^-\pi^+$ is around 5 - 10

	Variable	Cut	Signal Efficiency (%)
DD	$m(\pi\pi)$	$= m_{K^0_S(PDG)} \pm$ 20 MeV	96.6
	$K^0_S\ FD_{sig}$	> 7	92.7
LL	$m(\pi\pi)$	$= m_{K^0_S(PDG)} \pm$ 20 MeV	96.3

Table 14: Efficiency of cuts (K^0_S candidates).

times bigger than for $K^+\pi^-\pi^+$ and $\pi^+\pi^-\pi^+$ combination respectively only $K^+K^-\pi^+$ combination is considered in this analysis.

Mode	BR
$D_s^+ \to K^+K^-\pi^+$	(5.45 ± 0.17) %
$D_s^+ \to K^+\pi^-\pi^+$	(0.66 ± 0.04) %
$D_s^+ \to \pi^+\pi^-\pi^+$	(1.09 ± 0.05) %

Table 15: Branching fraction of D_s^+ meson decay to $3h$ final states

Following requirements was made to remove the combinatorial background from misidentifying of kaons or pions and physical background from misidentification of D^+ or Λ_c^+ meson or a random combination of K or π with D^0. This combination can mimic signal final state.

1. $D^+ \to K^+\pi^+\pi^-$
   Misidentification of kaon from $D_s^+ \to K^+K^-\pi^+$ with $D^+ \to K^+\pi^-\pi^+$. Removed if the mass of combination under π hypothesis is within $D^+ \pm 20$ MeV mass window unless the combination under K hypothesis is within $D_s^+ \pm 20$ MeV mass window and kaon fulfill the stronger requirement of PIDK(K).

2. $D^0 \to K^+\pi^-$ or $D^0 \to K^+K^-$
   Combination of $D^0 \to K^+K^-$ or $D^0 \to K^+K^-$ meson with random K^+ or π^+. removed if a combination of $K^+\pi^-$ or $K^+K^- > 1850$.

3. $\Lambda_c^+ \to K^+p\pi^-$
   Misidentification of kaon from $D_s^+ \to K^+K^-\pi^+$ with $\Lambda_c^+ \to K^+p\pi^+$. Removed if the mass of combination under p hypothesis is within $\Lambda_c^+ \pm 20$ MeV window unless the combination under K hypothesis is within $D_s^+ \pm 20$ MeV mass window and kaon fulfill the stronger requirement of PIDK(K).

The Fig. 5 show contributions from Λ_c^+ decay where the p is misidentified as K, contributions from D^0 decay where the π is misidentified as K, $K\pi$ and KK mass distribution with visible background contribution of D^0 meson. All of them were removed by applying vetos. Histograms were created using the 2015 data sample.

Since the D_s^+ meson decay to the combination of kaons and pions, it can decay by intermediate state $D_s^+ \to (\phi \to K^+K^-)\pi^+$, $D_s^+ \to K^+$ $(K^{*0} \to K^-\pi^+)$ or by no-resonance mode to $K^+\pi^-\pi^+$ final state. ϕ and K^{*0} has been investigated, but no additional requirements were made (Fig. 5).

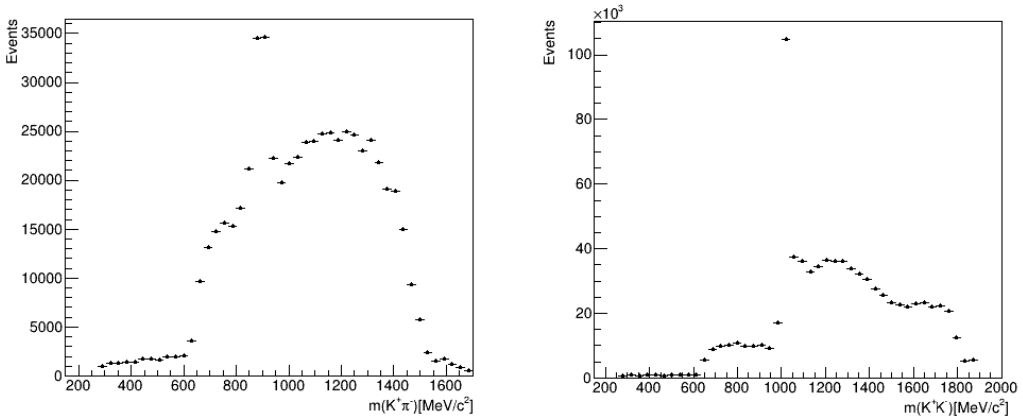

Figure 3: Mass distribution of K^+K^- (left) and $K^+\pi^-$ (right) candidates from $D_s^+ \to K^+K^-\pi^-$ (the 2015 data sample).

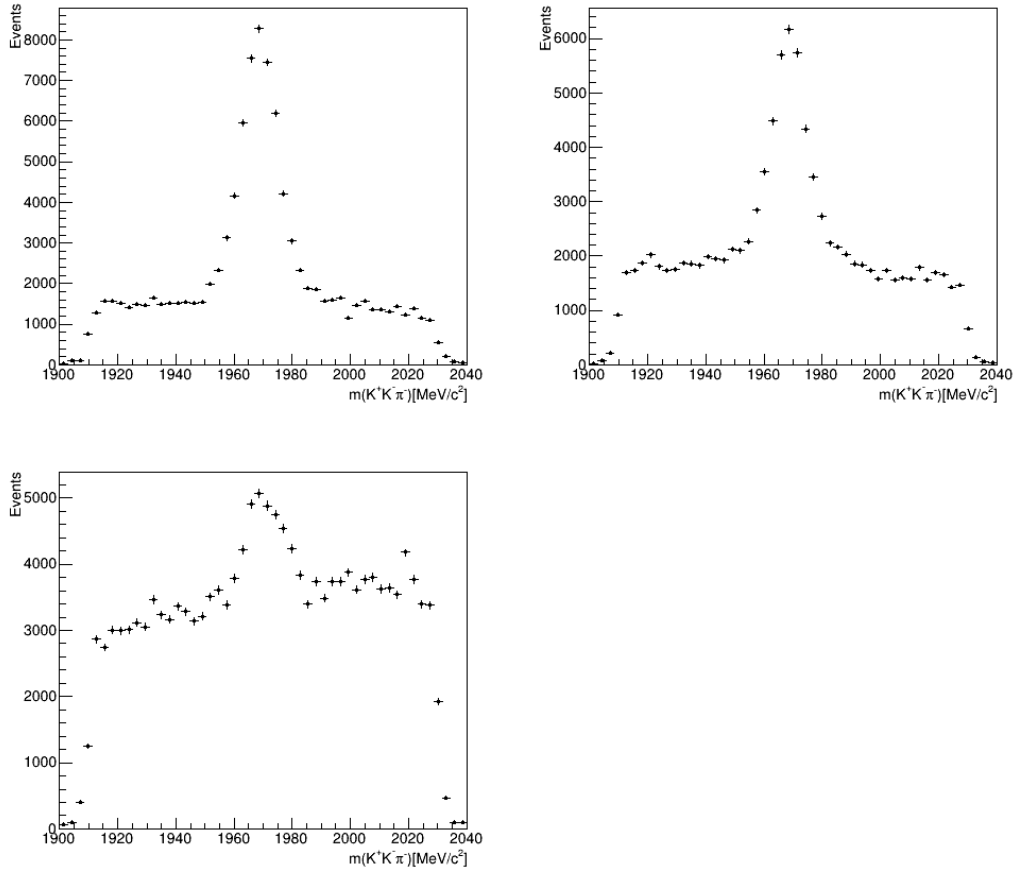

Figure 4: Mass distribution $D_s^+ \to (\phi \to K^+K^-)\pi^+$ (left), $D_s^+ \to K^+$ $(K^{*0} \to K^-\pi^+)$ (right) and no-resonance $K^+\pi^-\pi^+$ candidates mass distribution (bottom, the 2015 data sample).

The Fig. 4 show $D_s^+ \to (\phi \to K^+K^-)\pi^+$ (left), $D_s^+ \to K^+$ $(K^{*0} \to K^-\pi^+)$ (right) and no-resonance $K^+\pi^-\pi^+$ (bottom) candidates mass distribution. Visible is difference

between signal to background ratio for each distribution.

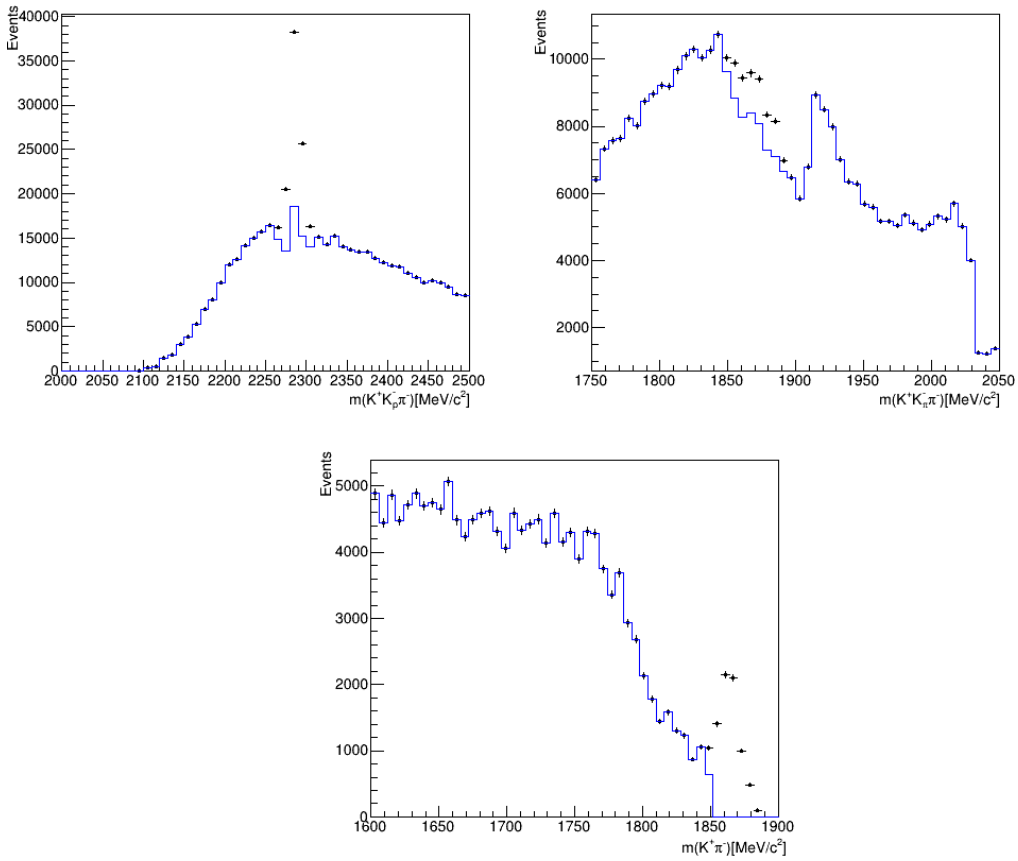

Figure 5: Background contributions from Λ_c^+ decay where the p is misidentified as K The D_s^+ invariant mass is recalculated under proton mass hypothesis for the kaon, black dots are the mass distribution of $K^+p\pi^+$ before removing Λ_c^+ contribution, blue line - after (top left). Background contributions from D^0 decay where the π is misidentified as K The D_s^+ invariant mass is recalculated under pion mass hypothesis for the kaon, black dots are the mass distribution of $K^+\pi^-\pi^+$ before removing D^0 contribution, blue line - after (top right). $K\pi$ and KK mass distribution with visible background contributions from D^0 meson before (black dots) and after (blue line) removing D^0 meson contribution (bottom).

Since the D^+ meson contribution is removed by stripping, we do not expect D^+ mass peak to be visible on $K^+K^-{}_{\pi^-}\pi^+$ mass distribution.

The efficiency of selection criteria was calculated for kinematic cut of D_s^+ candidates (Tab. 17).

4.3 Selection of and D_s^{*+}

Following paragraph describe the selection of resonance states D_s^{*+}. The offline selection of D_s^{*+} base on the selection of D_s^+ mesons and γ. The selection of γ is done by BDT described in the following paragraph. For D_s^{*+} resonances, an additional requirement is made. The Δ_M variable describes the mass difference between D_s^{*+} and D_s^+. The mass of

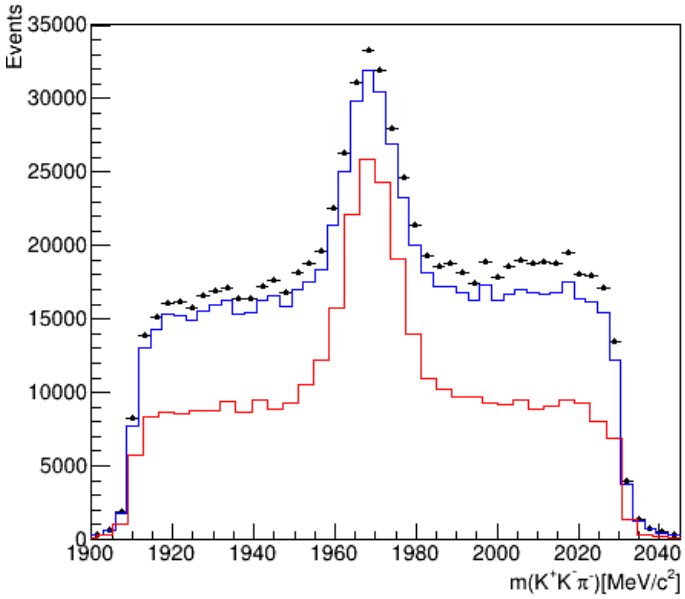

Figure 6: The $K^+K^-\pi^+$ candidates mass distribution after stripping and PID cut (black dot), after applying vetos (blue line), after stripping, PID cut, applying vetos and kinematic cuts (red line, the 2015 data sample).

the particle is well-known, and the difference of these masses should be around 144 MeV so, Δ_M requirement is $\Delta_M \in (104, 184)$ MeV.

4.4 Multiple candidates selection

Reconstruction of events may result in more than one signal candidate for the same B_s^0 events. It is generally associated with a number of candidates for a photon and combinations of D_s^+ meson daughters. The multiplication factor was calculated after each stage of selection to ensure that at the end, no more than one candidate per event left. In the table Tab. 26-33 there are multiple candidates rate for stripping and preselected, Run 2 sample (Appendix A). Since no more than 1 or 2 events with no more than 2 candidates per event left after all selection criteria (Tab. 20), no additional action has been taken.

4.5 Selection of K^{*-} (construction works)

In the analysis of $B_s^0 \to D_s^{*+}K^{*-}$ the non-resonant background is reduced by removing events in which the $K_S^0\pi$ invariant mass is greater than 75 MeV from the nominal K^{*-} mass. The $B_s^0 \to D_s^{*+}K^{*-}$ is a pseudoscalar to vector-vector, decay, which means the K^{*-} must be londitudially polarised, so the K_S^0 helicity angle follows a $cos^2\theta$ distribution *** Removing events with an absolute value of $cos^2\theta$ less than 0.3 improves the

	Description	Requirement
$D_s^+ \to K^+K^-\pi^+$	$m(K^+K^-\pi^+)$	$= m_{D_s^+} \pm 25$ MeV
	PIDK(K)	> 5
	PIDK(π)	< 5
	hasRich($K/\ \pi$)	$= 1$
	isMuon($K/\ \pi$)	$= 0$
	FD_{sig}	> 0
D^0 veto	$m(K^+K^-)$	< 1850 MeV
	$m(K^+\pi^-)$	< 1850 MeV
Λ_c^+ veto	$m(K^+K^-{}_p\pi^+)$	$\neq m(\Lambda_c^+) \pm 40$ MeV || PIDK(K^-) - PIDp(K^-) > 5
D^+ veto	$m(K^+K^-{}_{\pi^-}\pi^+)$	$\neq m(D^+) \pm 20$ MeV || PIDK(K^-) > 7
$D_s^+ \to K^+\pi^-\pi^+$	$m(K^+\pi^-\pi^+)$	$= m_{D_s^+} \pm 25$ MeV
	PIDK(K)	> 5
	PIDK(π)	< 5
	hasRich($K/\ \pi$)	$= 1$
	isMuon($K/\ \pi$)	$= 0$
	FD_{sig}	> 0
D^0 veto	$m(K^+\pi^-)$	< 1850 MeV
$D_s^+ \to \pi^+\pi^-\pi^+$	$m(\pi^+\pi^-\pi^+)$	$= m_{D_s^+} \pm 25$ MeV
	PIDK(π)	< 0
	hasRich($K/\ \pi$)	$= 1$
	isMuon($K/\ \pi$)	$= 0$
	FD_{sig}	> 0

Table 16: Offline selection requirements for D_s^+ candidates.

Variable	Cut	Signal Efficiency (%)
$m(K^+K^-\pi^+)$	$= m_{D_s^+} \pm 25$ MeV	83.9
hasRich($K/\ \pi$)	$= 1$	99.1
isMuon($K/\ \pi$)	$= 0$	96.3
$D_s^+\ FD_{sig}$	> 0	92.7

Table 17: Efficiency of cuts (D_s^+ candidates).

	Description	Requirement
D_s^{*+}	Δ_M	$\in (104, 184)$

Table 18: Offline selection requirements for D_s^{*+} candidates.

Variable	Cut	Signal Efficiency (%)
Δ_M	$\in (104, 184)$	82.6

Table 19: Efficiency of cuts (D_s^{*+} candidates).

Multip.	Number of events	%
Run Downstream		
1	1638	99,9
2	2	0,01

Multip.	Number of events	%
Run 2 Long		
1	1195	99,9
2	2	0,01

Table 20: Multiple candidate rate for Run 2 data sample after selection.

	Description	Requirement
K^*	PIDK(π)	< 0

Table 21: Offline selection requirements for K^* candidates.

Variable	Cut	Signal Efficiency (%)
PIDK(π)	< 0	86.6

Table 22: Efficiency of cuts (K^* candidates).

4.6 Boosted Decision Tree

Several MVA algorithms have been investigated, and a BDT algorithm with the gradient boost method using the XGBoost framework [1] has been used to reduce the contribution of combinatorial background. The XGBoost library offers implementation of gradient boosting algorithms,

Separate algoriths have been developed to reduce combinatorial background (B_s^0 BDT) and another to improve selection of K^{*-} (892) candidates (K^{*-} BDT). Since the downstream and long tracks have different properties, a separate algorithm was trained for both types of tracks. This apporach is applied for both algorithms: B_s^0 BDT and K^{*-} BDT.

4.6.1 Configuration of the algorthm

In order to reduce this dangerous over-fitting, some hyperparameters of the classifier are adjusted:

- *learning rate* - η is used to prevents overfitting. η Learnig rate after each boosting step, shrinks the feature weights to make the boosting process more conservative.
- *maximum depth* of a tree. Higher maximum depth make the model more complex and more likely to overfit.
- *minimum child weight* is a minimum sum of instance weight needed in a child. If the leaf node with the sum of instance weight is less than minimum child weight then the building process will give up further partitioning.
- *gamma* is a minimum loss reduction required to make a further partition on a leaf node of the tree. The larger gamma is, the more conservative the algorithm will be.
- *subsample:* is a ratio of the training sample. As a default is set to 0.5 what means that algorithm would randomly sample half of the training data prior to growing trees. and this will prevent overfitting. Subsampling will occur once in every boosting iteration.

The following parametrisation has been choosed in order to increase efficiency of algorithm:

4.6.2 Description of physics variables

Besides a number of standard variables, a feature that estimates the imbalance of p_T around the B_s^0 candidate momentum vector is also used. It is defined as

$$I_{p_T} = \frac{p_T(B_s^0) - \sum p_T}{p_T(B_s^0) + \sum p_T} \tag{2}$$

where the sum is taken over all other charged tracks inconsistent with B_s^0 candidate. This cone is defined by a circle with a radius of 1.5 units in the plane of pseudorapidity and azimuthal angle (both defined below and expressed in radians). The I_{p_T} enhance the selection of B_s^0 candidates.

Hyperparamter	Value
learning rate	1638
maximum depth	1638
minimum child weighte	1638
gamma	1638
subsample	1638

Table 23: kk

Another variables related to resonance states is Δ_R:

$$\Delta_R(D_s^+) = \sqrt{(\eta(D_s) - \eta(\gamma))^2 + (\phi(D_S) - \phi(\gamma))^2} \tag{3}$$

is also defined similarly for K^{*-} resonance, where η and ϕ are the pseudorapidity and azimuthal angle in the LHCb lab. frame (the right-handed Cartesian system is being used, with the z axis pointing along the beams direction). The polar angle - θ, is measured with respect to the z axis, the azimuthal angle - ϕ, is the angle with respect to the x axis of the projection in the xy plane. The rapidity and pseudorapidity are defined as :

$$y = \frac{1}{2} ln(\frac{E + p_z}{E - p_z}) \tag{4}$$

$$\eta = -ln(tan\frac{\theta}{2}) \tag{5}$$

The Δ_R is known to be invariant under boost along the beams axis. As the typical radius of a hadron jet in the $\eta - \phi$ plane has been measured to be about 1, expected value of Δ_R should be below 1.

The radial flight distance is defined as:

$$RFD(D_s^+) = \frac{Vx(D_s^+, z) - Vx(B_s^0, z)}{\sigma^2_{Vx(D_s^+,z)} + \sigma^2_{Vx(B_s^0,z)}} \tag{6}$$

where σ_z stands for error in z vertex position of particle. The RFD variables is also used in selection of K^0_S mesons.

4.6.3 Description of B_s^0 algorithm

For B_s^0 BDT, a simulated sample was employed with generator level cuts and true matching applied. The background sample is data high mass sideband region of the B_s^0 mass (above 5600 MeV). For K^{*-} BDT, the same simulated sample with the same true matching selection and K^{*-} low and high mass sideband region (belove 847 MeV and above 937 MeV).

4.6.4 Description of K^* algorithm

4.6.5 B_s^0 algorithm

Initial set of variables was investigated:

1. D_s^{*+} and D_s^{*+} daughters
   The Δ_R distance for D_s^{*+}: $\Delta_R(D_s^{*+})$, the D_s^+ impact parameter χ^2 w.r.t PV: D_s^+ IPχ^2, the D_s^+ radial flight distance, RFD(D_s^+), the D_s^+ χ^2 of the vertex: D_s^+ Vx χ^2, the γ transverse momentum: γ PT, the γ condence level, γ CL, condence level is the probability that the Calo object tagged as a photon is indeed a photon. For true photons this variable peaks at 1, for fake photons - at 0.

2. D_s^+ daughters
   The ProbNN variables built using multivariate techniques by combining tracking and PID information from each sub-system into a single probability value for each particle hypothesis (kaon, pion, proton): K^+, K^- ProbNNK and π^+ ProbNNpi.

3. B_s^0 meson
   The impact parameter χ^2 w.r.t PV: B_s^0 IPχ^2, the P_T asymmetry: I_{p_T}, the χ^2/ndf of the lifetime: B_s^0 τ χ^2/ndf, the χ^2/ndf of the end vertex: B_s^0 Vx χ^2

Track type	Variable
	B_s^0 IPχ^2
	I_{p_T}
	B_s^0 τ χ^2/ndf
	B_s^0 Vx χ^2/ndf
	γ P_T
	γ CL
	h1 ProbNNk
	h2 ProbNNk(pi)
	h3 ProbNNk(pi)
	Δ_R D_s^{*+}
	D_s^+ IPχ^2
	RFD(D_s^+)
	D_s^+ Vxχ^2
	Δ_R D_s^{*+}

Table 24: BDT input variables

4.6.6 K^{*-} algorithm

Again, initial set of variables was investigated:

1. K^{*-} and K^{*-} daughters
   The Δ_R distance for K^{*-}: $\Delta_R(K^{*-})$

Track type	Variable
	Δ_R K^*
	K^0_S FDχ^2
DD	K^0_S daughter 1 P_T
DD	K^0_S daughter 2 P_T
LL	K^0_S daughter 1 IPχ^2_{max}
LL	K^0_S daughter 2 IPχ^2_{max}
LL	K^0_S daughter 1 IPχ^2_{min}
LL	K^0_S daughter 2 IPχ^2_{min}
LL	K^0_S χ^2_{IP}

Table 25: BDT input variables

Classifier for long tracks: minimum and maximum impact parameter χ^2 w.r.t PV of K^0_S daughters: π^+, π^- IPχ^2_{max}, IPχ^2_{min}, the K^0_S flight distance χ^2: K^0_S FDχ^2, the K^0_S impact parameter χ^2 w.r.t PV: K^0_S IPχ^2

Classifier for downstream tracks: the K^0_S transverse momentum: K^0_S P_T

5 MC and data comparison

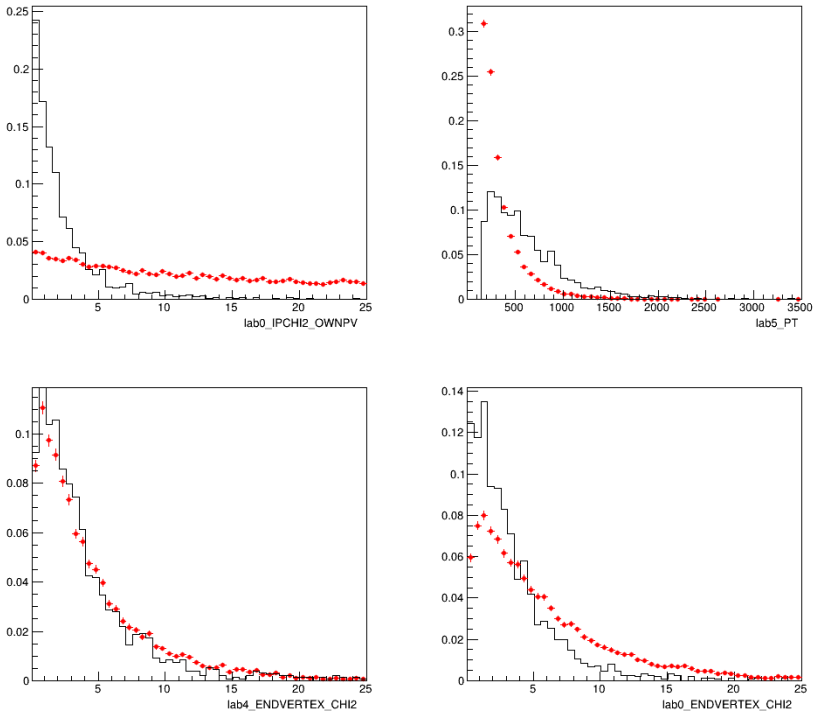

Figure 7: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection.

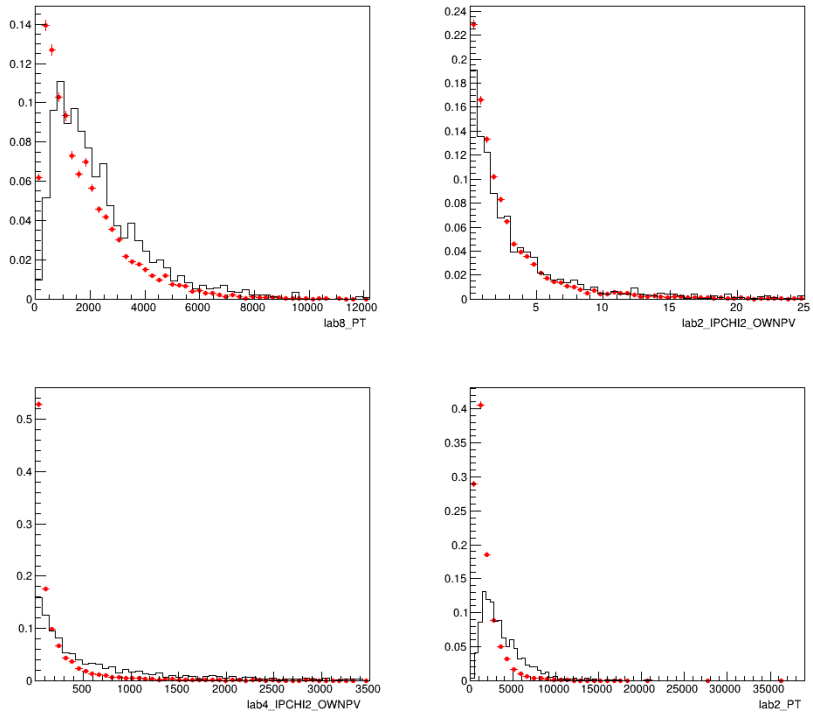

Figure 8: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection.

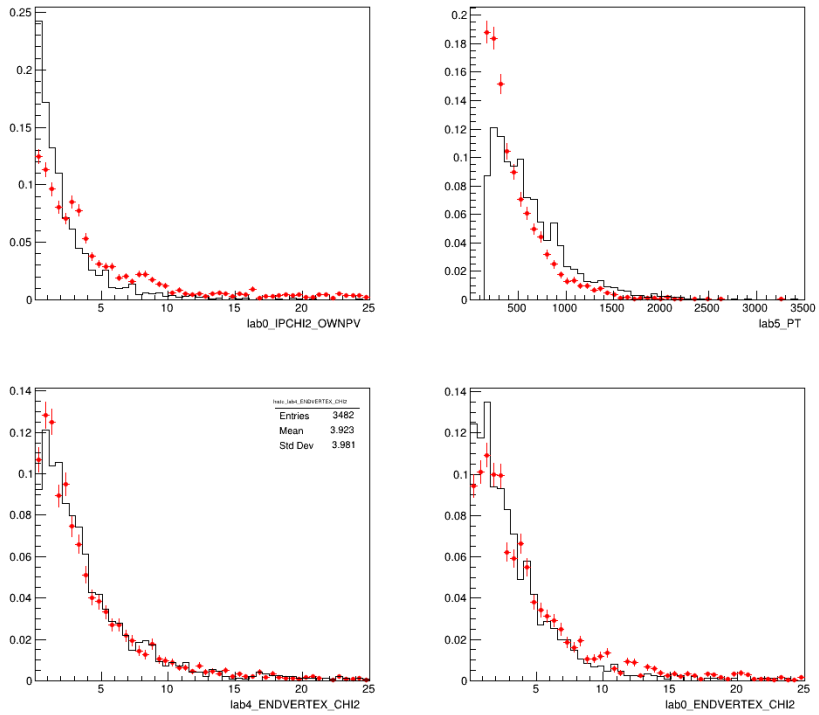

Figure 9: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection and BDT cut: BDT>0.1

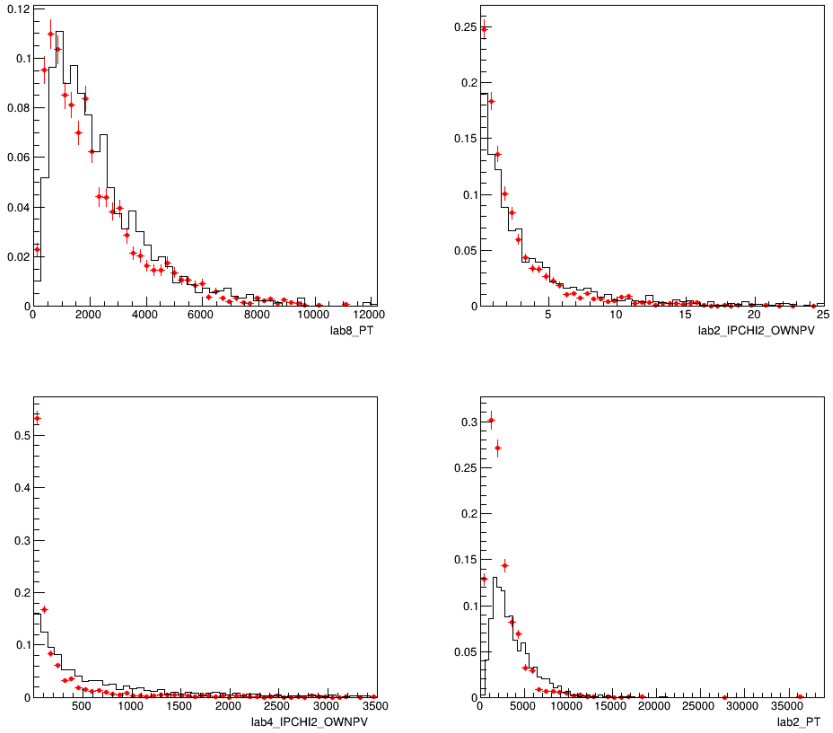

Figure 10: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection and BDT cut: BDT>0.1

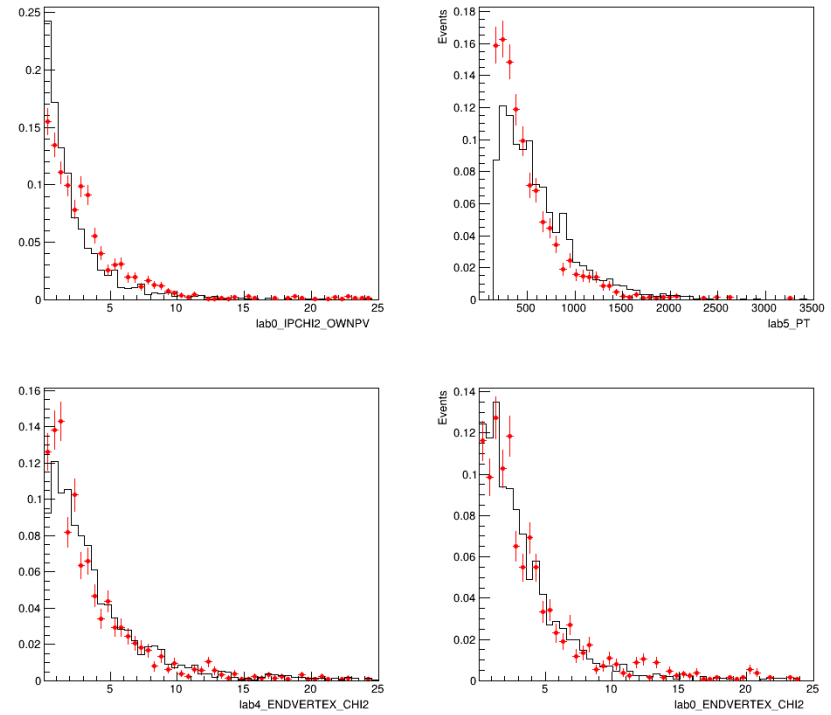

Figure 11: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection and BDT cut: BDT>0.8

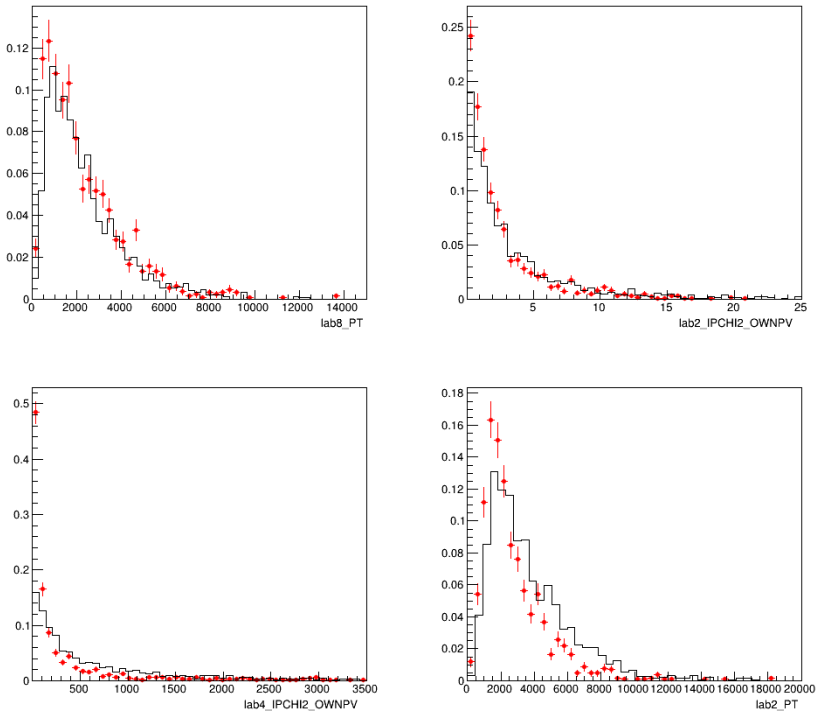

Figure 12: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, downstream tracks, data after preselection and BDT cut: BDT>0.8

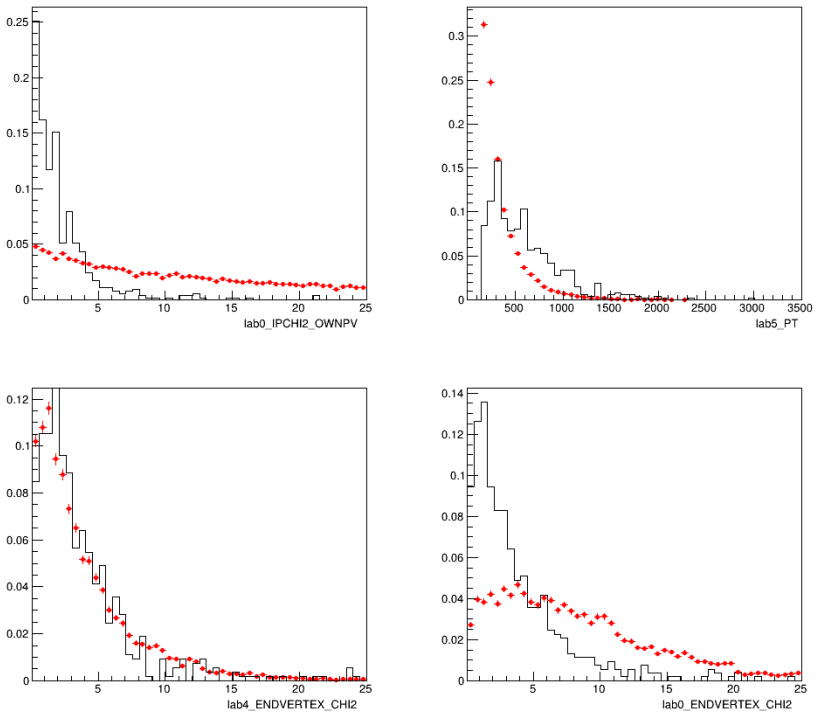

Figure 13: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection.

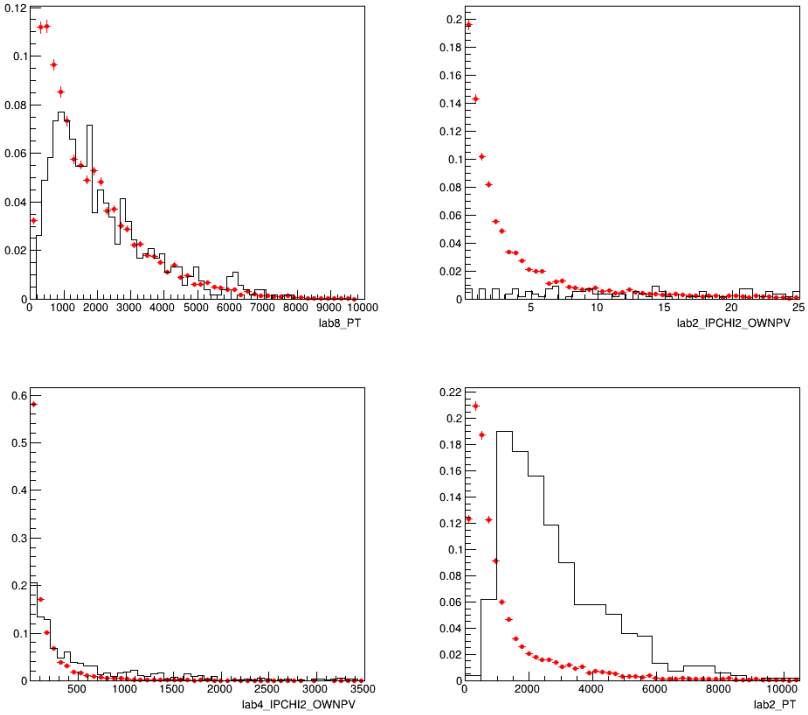

Figure 14: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection.

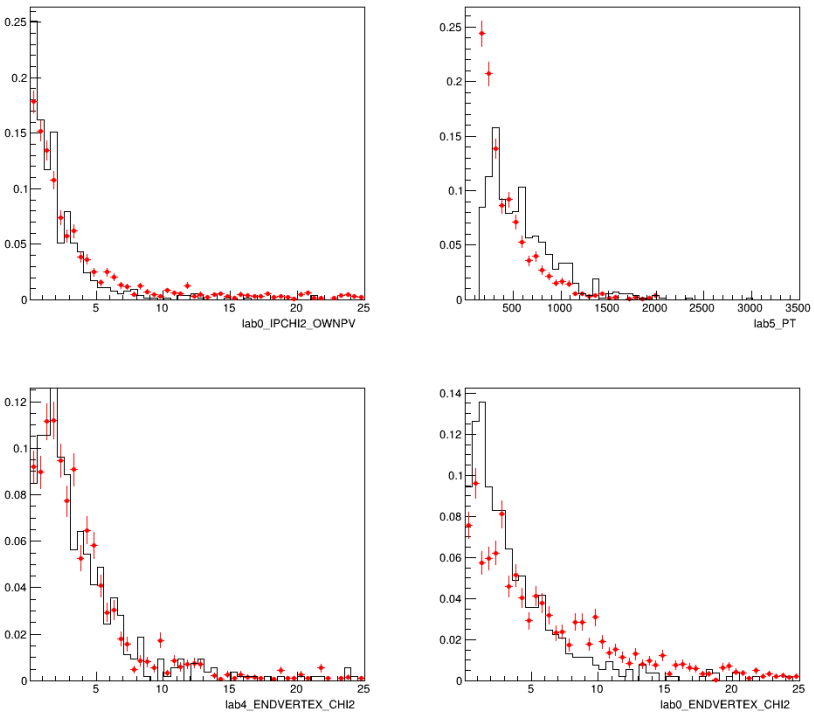

Figure 15: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection and BDT cut: BDT>0.1

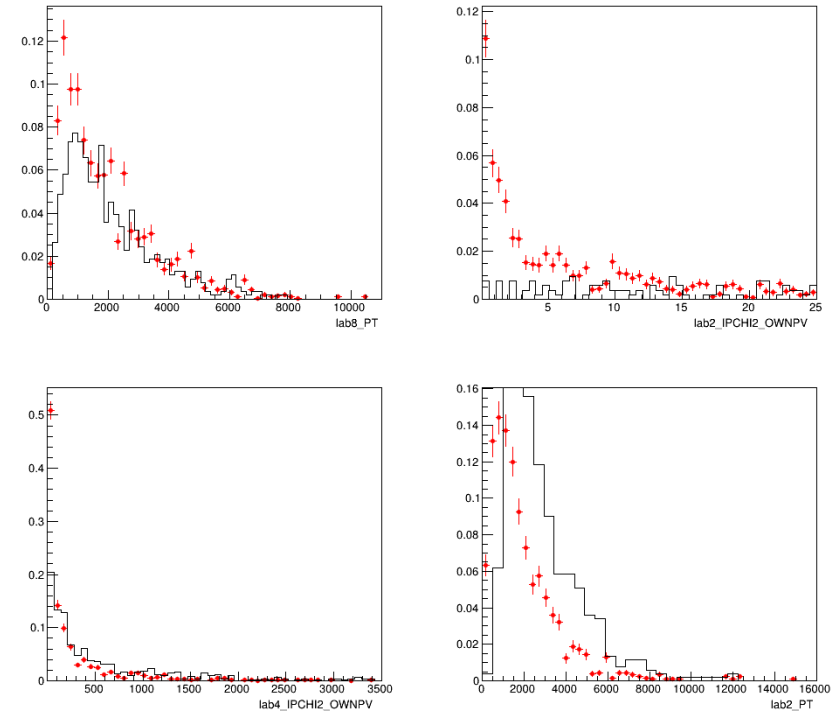

Figure 16: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection and BDT cut: BDT>0.1

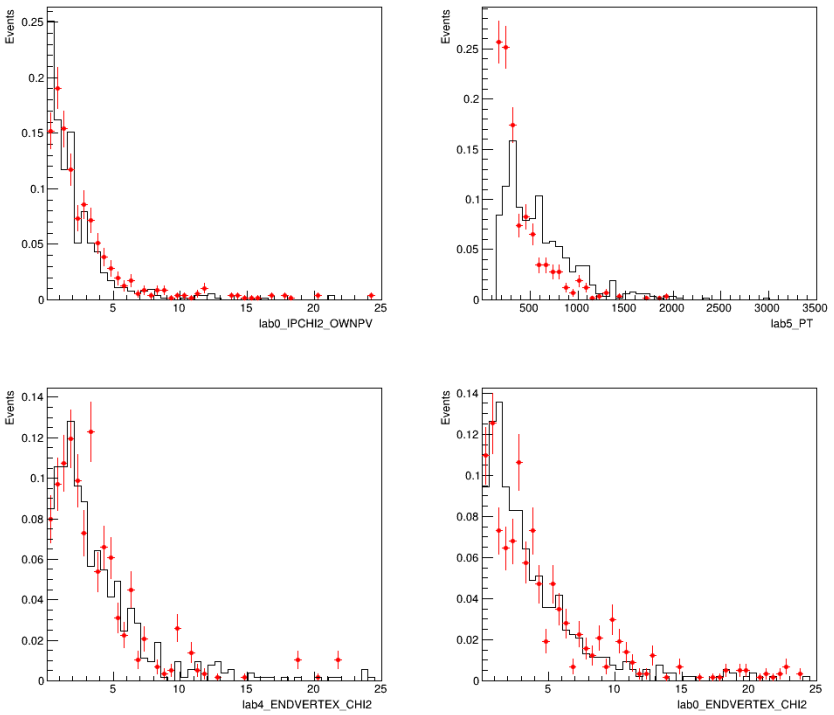

Figure 17: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection and BDT cut: BDT>0.8

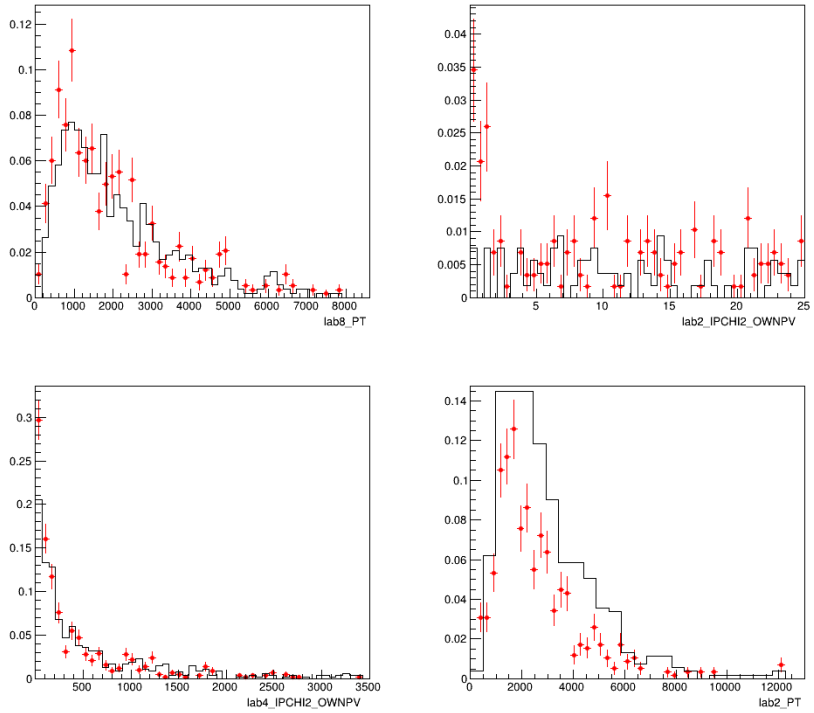

Figure 18: The comparison of simulated data (black lines) and Run 2 data (red dots) for selected variable, long tracks, data after preselection and BDT cut: BDT>0.8

A Appendix A

Below, can be found summary of multiple candidates after stripping and preseletcion for 2015, 2016, 2017 and 2018.

Multip.	Number of events	%
2015 Downstream		
1	436053	37,5
2	334000	28,7
3	213991	18,4
4	213991	15,1
5	175635	9,1
6	106540	7,8
7	91041	4,3
8	50915	3,9
...	...	...
678	1	-
1072	1	-

Multip.	Number of events	%
2015 Long		
1	464700	26,7
2	353080	20,3
3	228491	13,1
4	190532	11,0
5	116185	6.6
6	99605	5.7
7	57310	3,2
8	51504	2,9
...	...	...
732	1	-
1020	1	-

Table 26: Multiple candidate rate for 2015 stripping data.

Multip.	Number of events	%
2016 Downstream		
1	3787227	27,5
2	2869520	20,9
3	1835820	13,4
4	1490482	10,9
5	898392	6,5
6	757719	5,5
7	428491	3,1
8	380089	2,7
...	...	...
640	1	-
735	1	-

Multip.	Number of events	%
2016 Long		
1	464700	26,7
2	353080	20,3
3	228491	13,1
4	190532	11,0
5	116185	6.6
6	99605	5.7
7	57310	3,2
8	51504	2,9
...	...	...
1523	1	-
1384	1	-

Table 27: Multiple candidate rate for 2016 stripping data.

Multip.	Number of events	%
2017 Downstream		
1	4639174	27,1
2	3498859	20,4
3	2235461	13,0
4	1814863	10,6
5	1091647	6,3
6	921116	5,3
7	519122	3,0
8	460521	2,7
...	...	...
519	1	-
638	1	-

Multip.	Number of events	%
2017 Long		
1	3248619	26,4
2	2409999	20,6
3	1557358	13,7
4	1271315	10,3
5	782624	6,4
6	660226	5,4
7	379905	3,0
8	338375	3,7
...	...	...
1205	1	-
1902	1	-

Table 28: Multiple candidate rate for 2017 stripping data.

Multip.	Number of events	%
2018 Downstream		
1	3071762	25,5
2	2326876	17,1
3	1488581	16,3
4	1209598	14,0
5	729515	8,5
6	614482	6,1
7	347626	3,1
8	308082	3,o
...	...	...
778	1	-
1077	1	-

Multip.	Number of events	%
2018 Long		
1	3765491	25,1
2	2830697	18,8
3	1826930	12,1
4	1504148	10,0
5	919360	6,1
6	785668	5,2
7	448144	2,9
8	405995	2,7
...	...	...
1416	1	-
1524	1	-

Table 29: Multiple candidate rate for 2018 stripping data.

Multip.	Number of events	%
2015 Downstream		
1	963	73.065
2	283	21.47
3	53	4.021
4	9	6.83
5	8	6.07
6	2	1.52

Multip.	Number of events	%
2015 Long		
1	763	68,1
2	203	18,1
3	119	10,6
4	22	01,9
5	7	0,6
6	5	0,5

Table 30: Multiple candidate rate for 2015 preselected data.

Multip.	Number of events	%
2016 Downstream		
1	17873	72,7
2	5230	21,3
3	1019	10,4
4	348	1,4
5	51	2,1
6	39	1,6
7	12	0,1

Multip.	Number of events	%
2016 Long		
1	6503	72,8
2	1872	20,9
3	364	4,1
4	140	1,6
5	38	0,4
6	17	0,2

Table 31: Multiple candidate rate for 2016 preselected data.

Multip.	Number of events	%
2017 Downstream		
1	17873	72,7
2	5230	21,3
3	1019	10,4
4	348	1,4
5	51	2,1
6	39	1,6
7	12	0,1

Multip.	Number of events	%
2017 Long		
1	6503	72,8
2	1872	20,9
3	364	4,1
4	140	1,6
5	38	0,4
6	17	0,2
6		

Table 32: Multiple candidate rate for 2017 preselected data.

Multip.	Number of events	%
2018 Downstream		
1	13691	73,2
2	3853	20,6
3	809	4,3
4	269	1,4
5	38	0,2
6	24	0,1
7	14	0,1
8	4	0,1

Multip.	Number of events	%
2018 Long		
1	8501	71,9
2	2549	21,5
3	525	4,4
4	192	1,6
5	30	0,2
6	21	0,1

Table 33: Multiple candidate rate for 2018 preselected data

References

[1] *Xgboost documentation*, `https://xgboost.readthedocs.io/en/latest/`. Accessed: 2021-01-01.