Probing new physics with rare B decays with tau leptons in the final state

Jacopo Cerasoli
CNRS - IPHC

03/06/2022, IPHC seminar, Strasbourg

Outline

1) Introduction to the Standard Model

2) Flavor anomalies and the search for new physics

3) The LHCb detector

4) Search for the rare $B^0 \to K^{*0}\tau^+\tau^-$ decay at LHCb (my thesis)

 - Event selection

 - Signal extraction

 - Sensitivity estimation

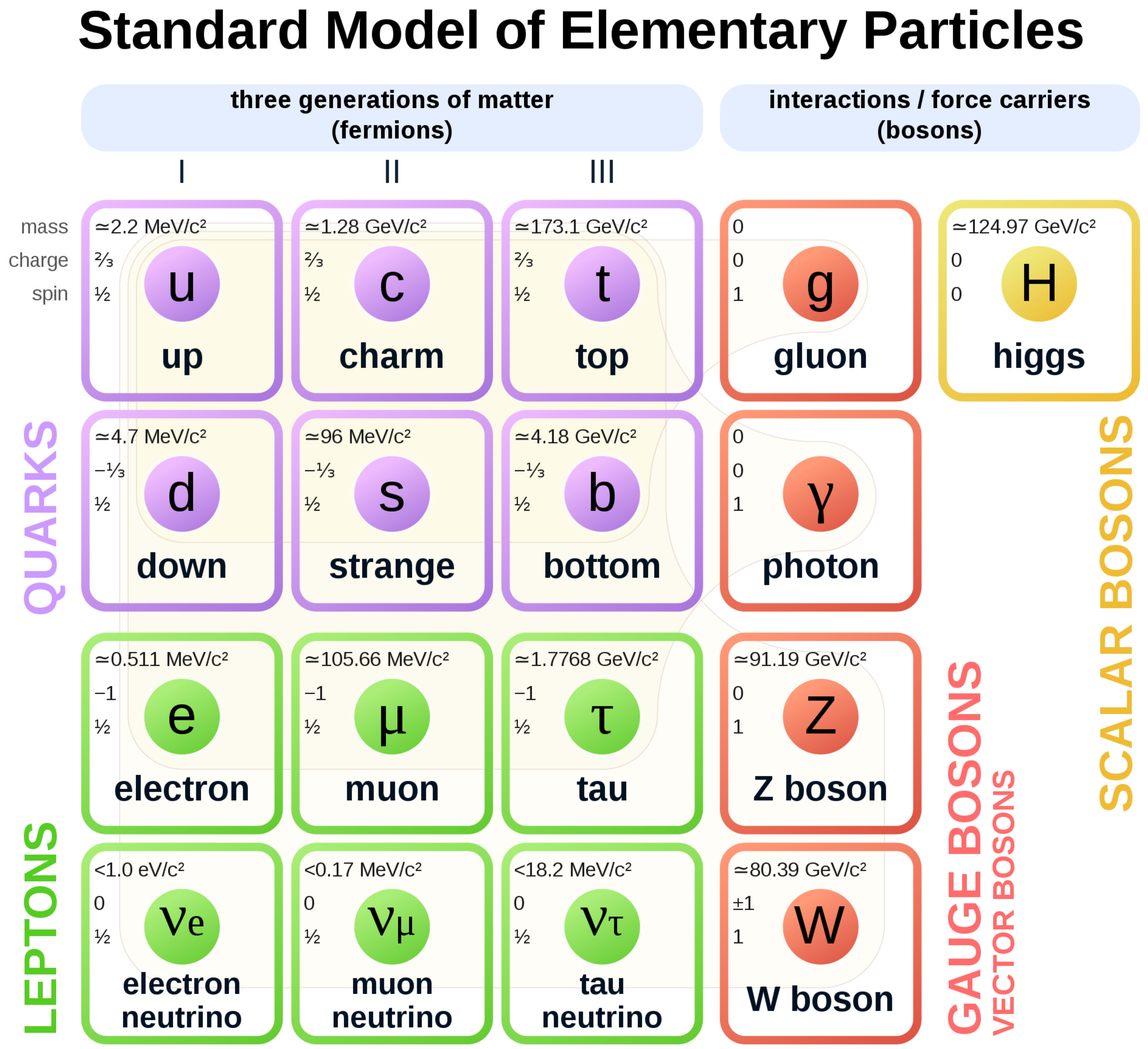
5) Future prospects and conclusions

From Facebook

The Standard Model and beyond

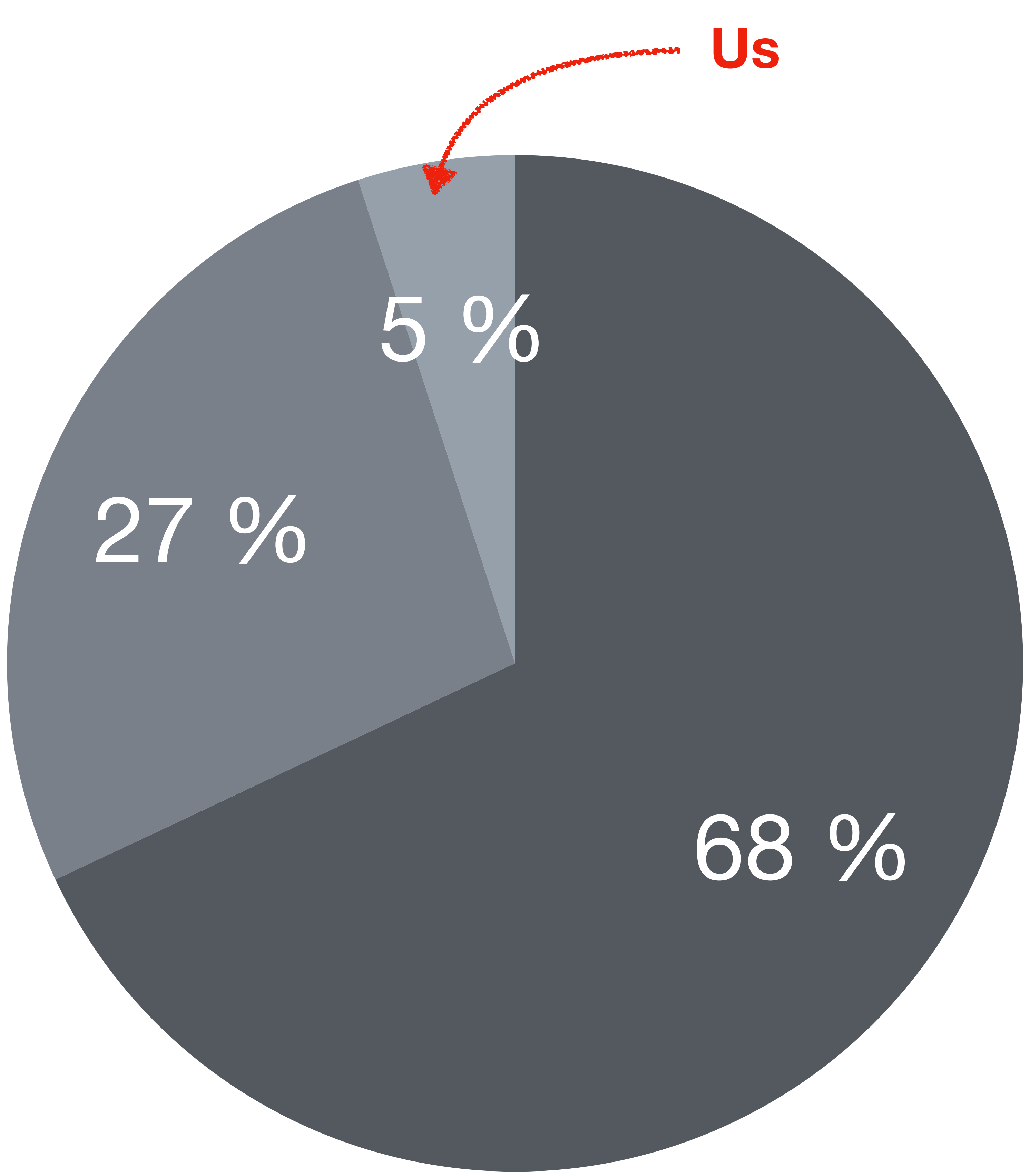
The Standard Model

- The **Standard Model** describes the behavior and interactions of the elementary particles
- Particles described as excitations of *dynamic fields* at a given point in space-time
- 6 leptons and 6 quarks divided in *three generations,* 4 gauge bosons + Higgs boson



Search for new physics

- MC shows impressive predictive power, but has several **shortcomings**:
  - Dark matter and dark energy (95 % of the universe) unexplained
  - Matter - anti-matter asymmetry not accounted for by the model
  - Neutrino non-zero mass terms not explained
  - Numerous theoretical prejudices (gravity, mass hierarchy, ...)



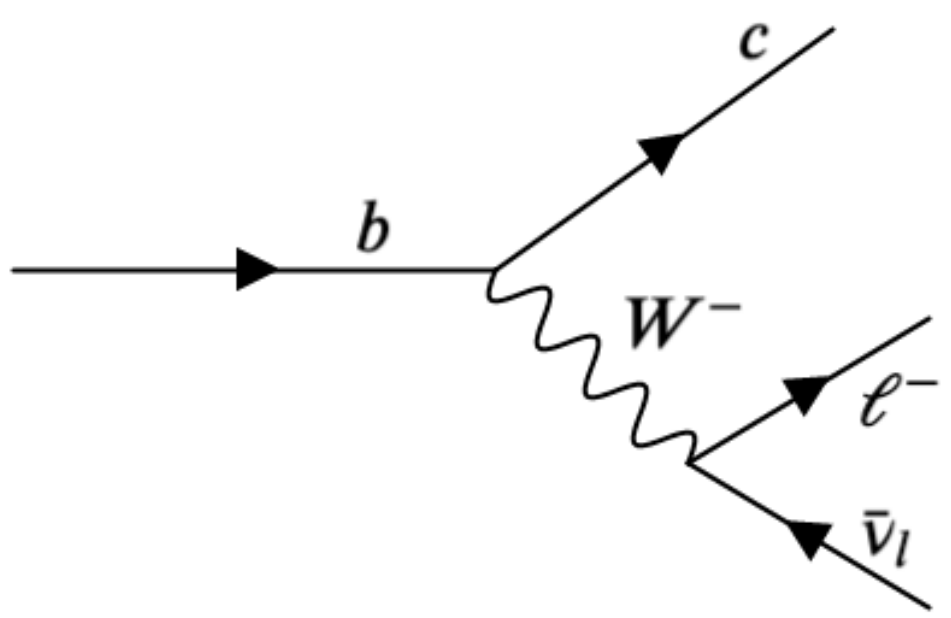
- Two main methods to search for **new physics beyond the SM**:
  - **Direct searches** of new particles
  - **Indirect searches** (this seminar): measurements of SM observables and comparison with theory predictions

Lepton Flavor Universality

- **Lepton Flavor Universality**: the three lepton generations have identical behaviors (except for differences due to their masses)
- **Accidental symmetry** experimentally well-established so far
- **Tested** by measuring **ratios of branching fractions**, examples:

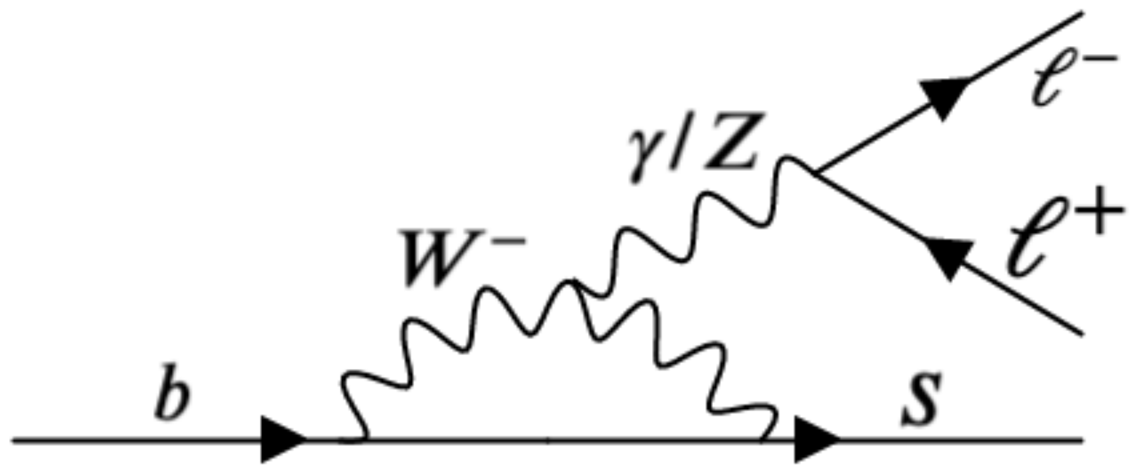
$$R_{D^{(*)}} = \frac{\mathscr{B}(B^0 \to D^{(*)-}\tau^+\nu_\tau)}{\mathscr{B}(B^0 \to D^{(*)-}\mu^+\nu_\mu)}$$

Tree-level $b \to c l^- \nu_l$ process



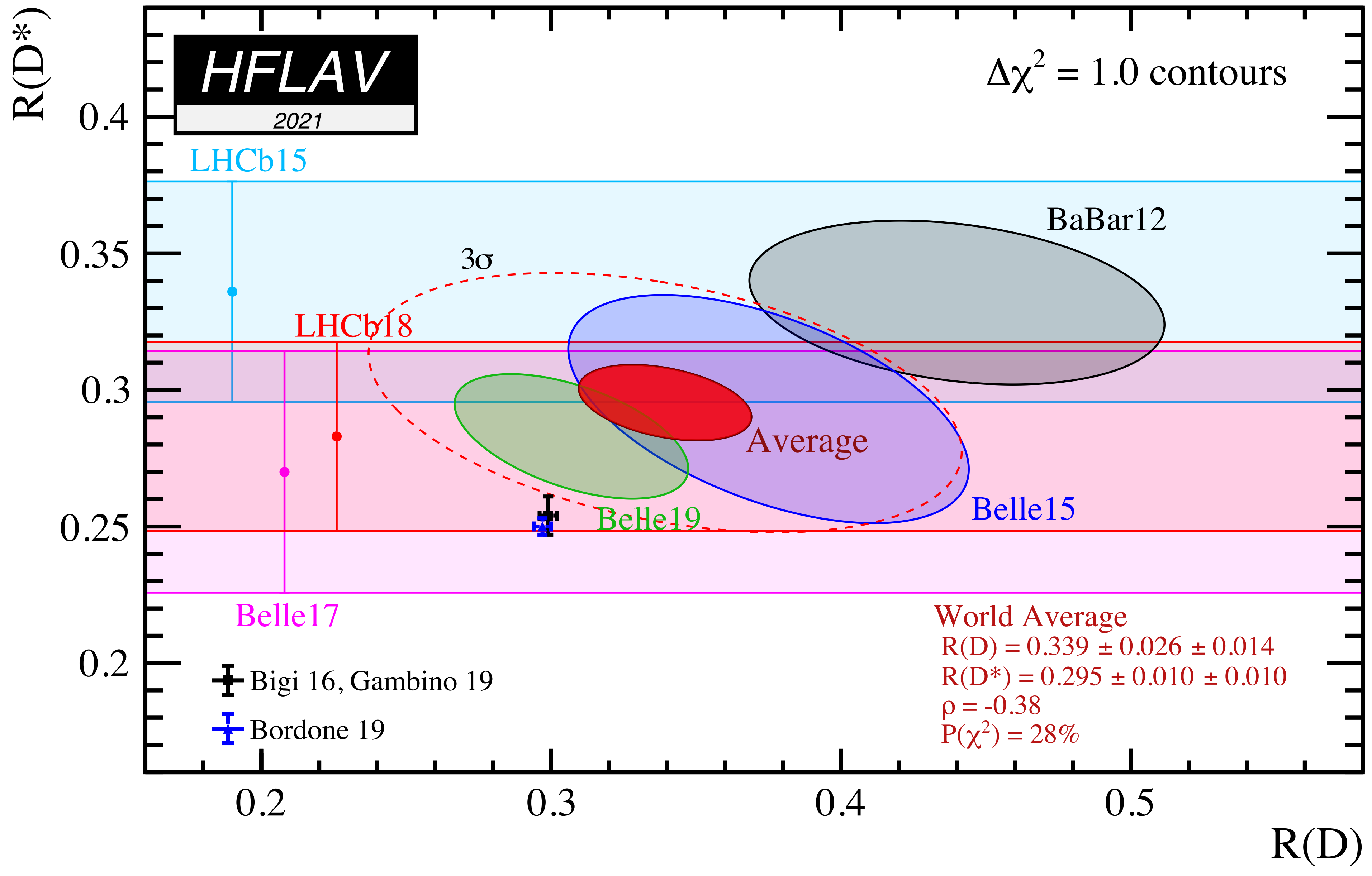
$$R_K = \frac{\mathscr{B}(B^+ \to K^+e^+e^-)}{\mathscr{B}(B^+ \to K^+\mu^+\mu^-)}$$

Higher-order $b \to s l^+ l^-$ process

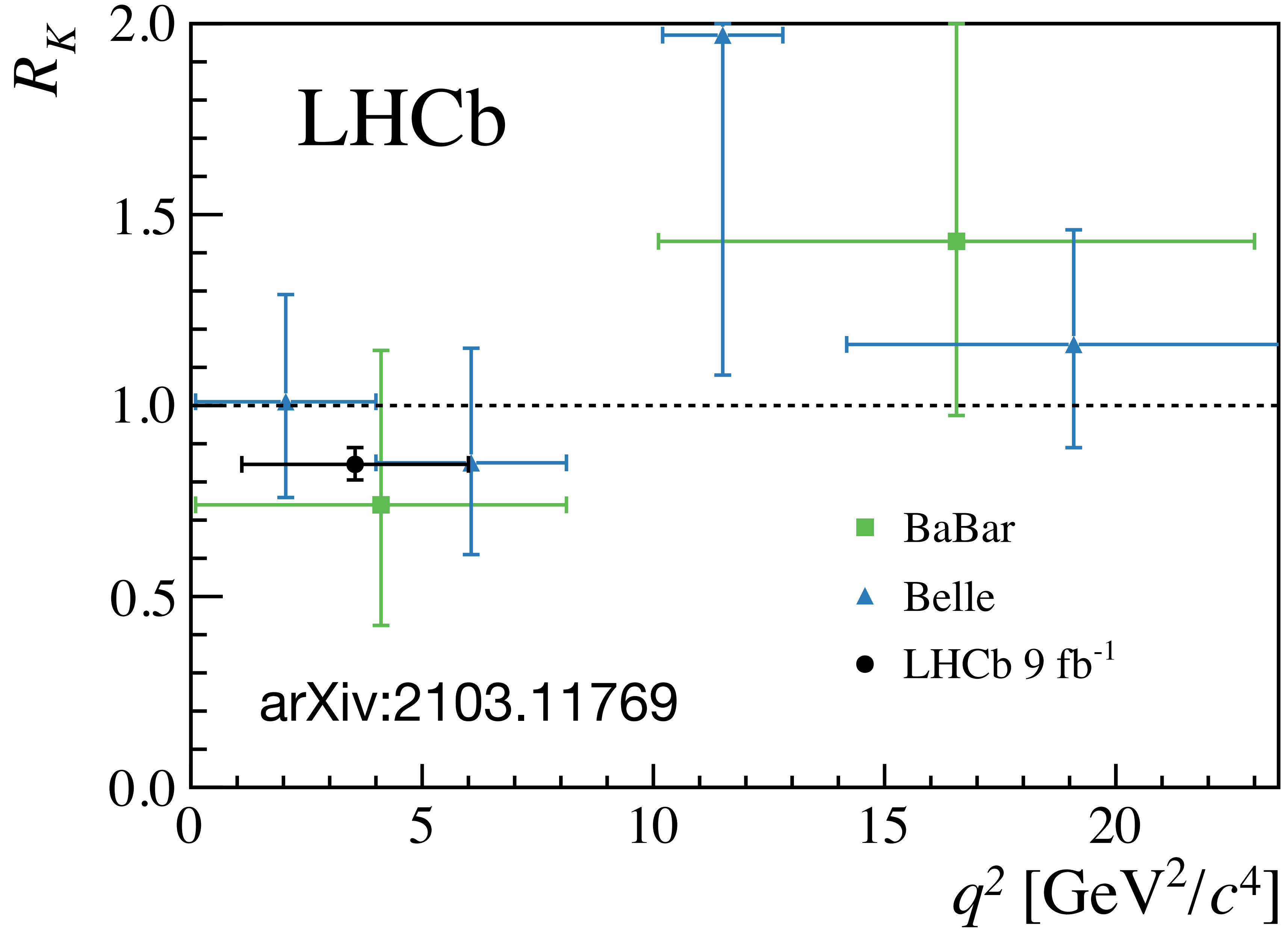


- Such measurements can reveal the presence of lepton flavor non-universal effects

Flavor anomalies



3.4 σ tension wrt the SM

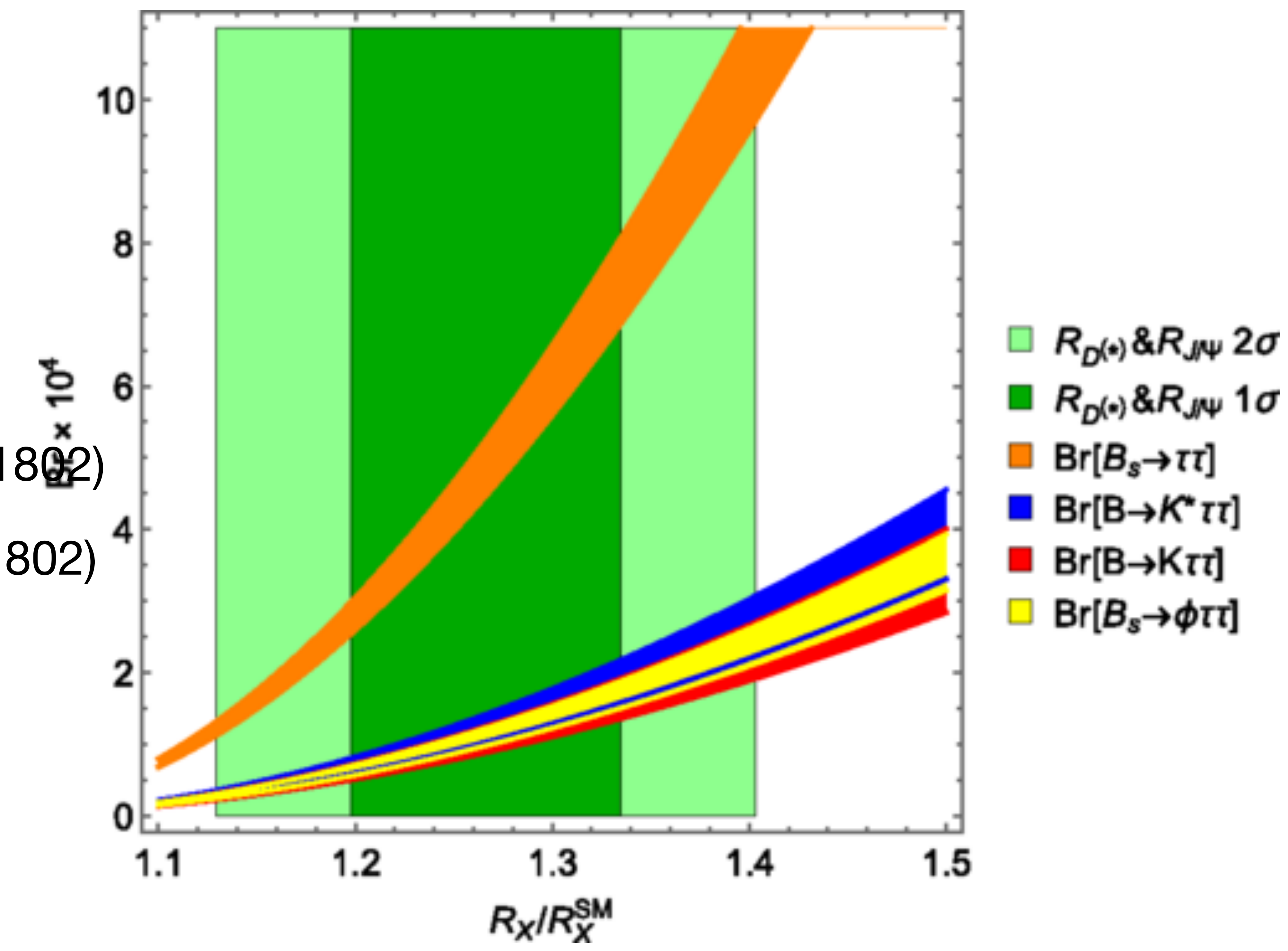


3.1 σ tension wrt the SM

- Additional anomalies (many more than those listed here) observed in **branching ratio measurements**, e.g. $B_s^0 \to \phi\mu^+\mu^-$ (Phys. Rev. Lett. 127, 151801), and **angular analyses**, e.g. P_5' in $B^0 \to K^{*0}\mu^+\mu^-$ (Phys. Rev. Lett. 125, 011802)

Rare B decays with τ leptons in the final state...

Phys. Rev. Lett. 120, 181802



- **Rare $b \to s\tau^+\tau^-$ transitions** are powerful probes for new physics:
  - m_τ ~ 17 m_μ ~ 3500 m_e, **taus could be the most sensitive to NP**
  - τ modes **still largely unexplored** (limits at 90% CL)

$\mathscr{B}(B^0_{(s)} \to \tau^+\tau^-) < 1.6\ (5.2) \cdot 10^{-3}$ (LHCb 2017, Phys. Rev. Lett. 118, 251802)

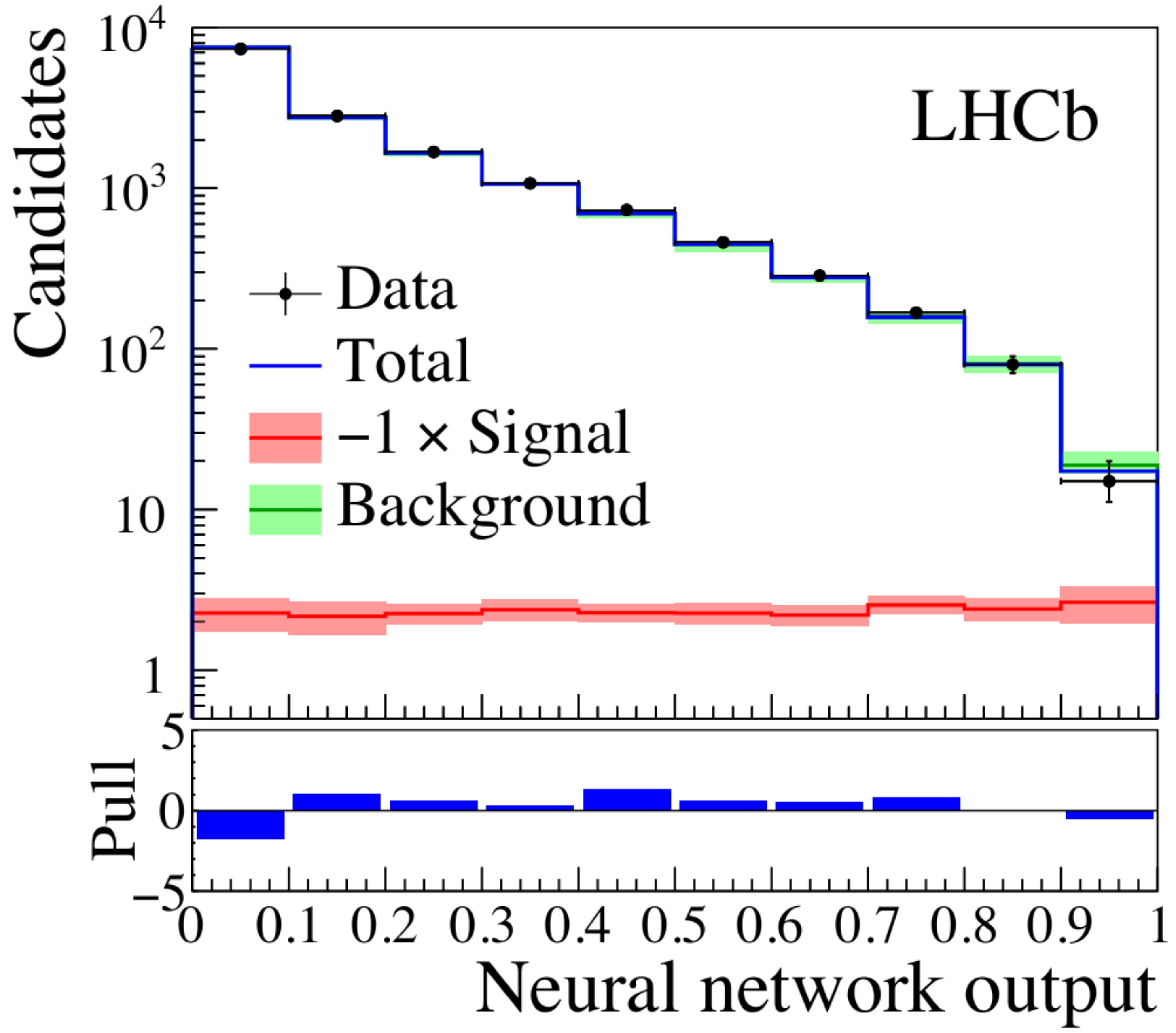
$\mathscr{B}(B^+ \to K^+\tau^+\tau^-) < 2.3 \cdot 10^{-3}$ (LHCb 2020, Phys. Rev. Lett. 118, 031802)

$\mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.0 \cdot 10^{-3}$ (Belle 2021, arXiv:2110.03871)

- **Complex experimentally**:
  - **τ decays before it can be detected**
  - Neutrinos in the final state: **missing energy**

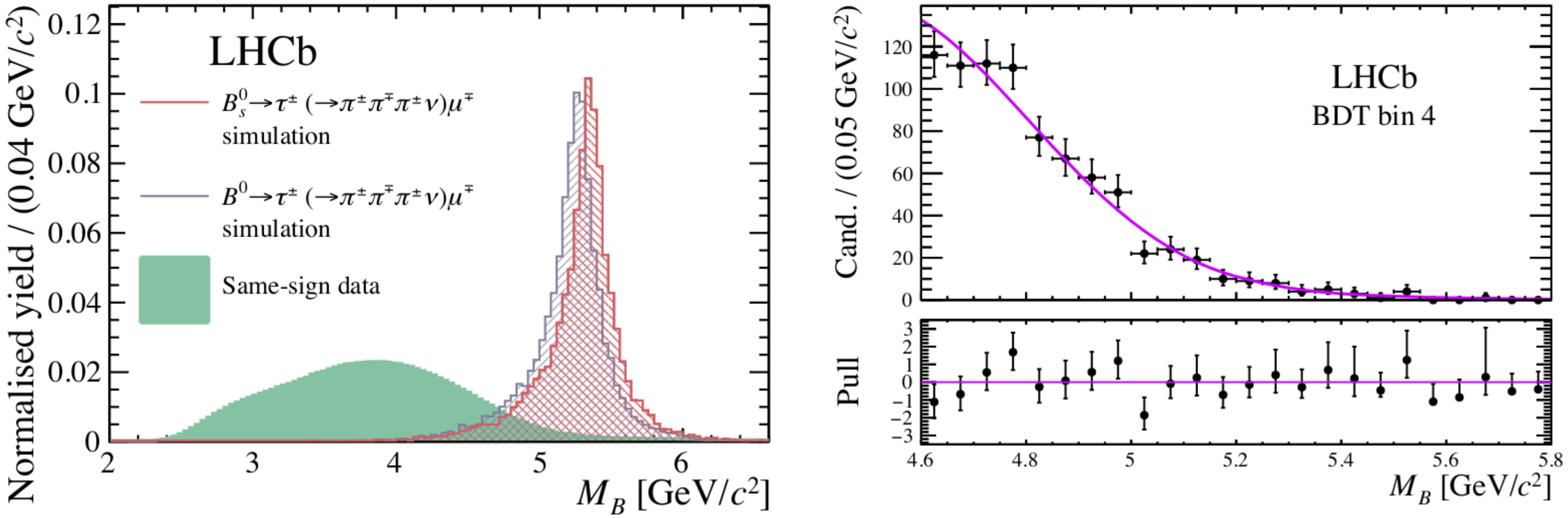
... at CPPM

- CPPM pioneered the study of **rare/forbidden B decays with tau's in the final state** at LHCb (95% CL limits):



$\mathscr{B}(B^0_{(s)} \to \tau^+\tau^-) < 2.1\,(6.8) \cdot 10^{-3}$

Phys. Rev. Lett. 118, 251802



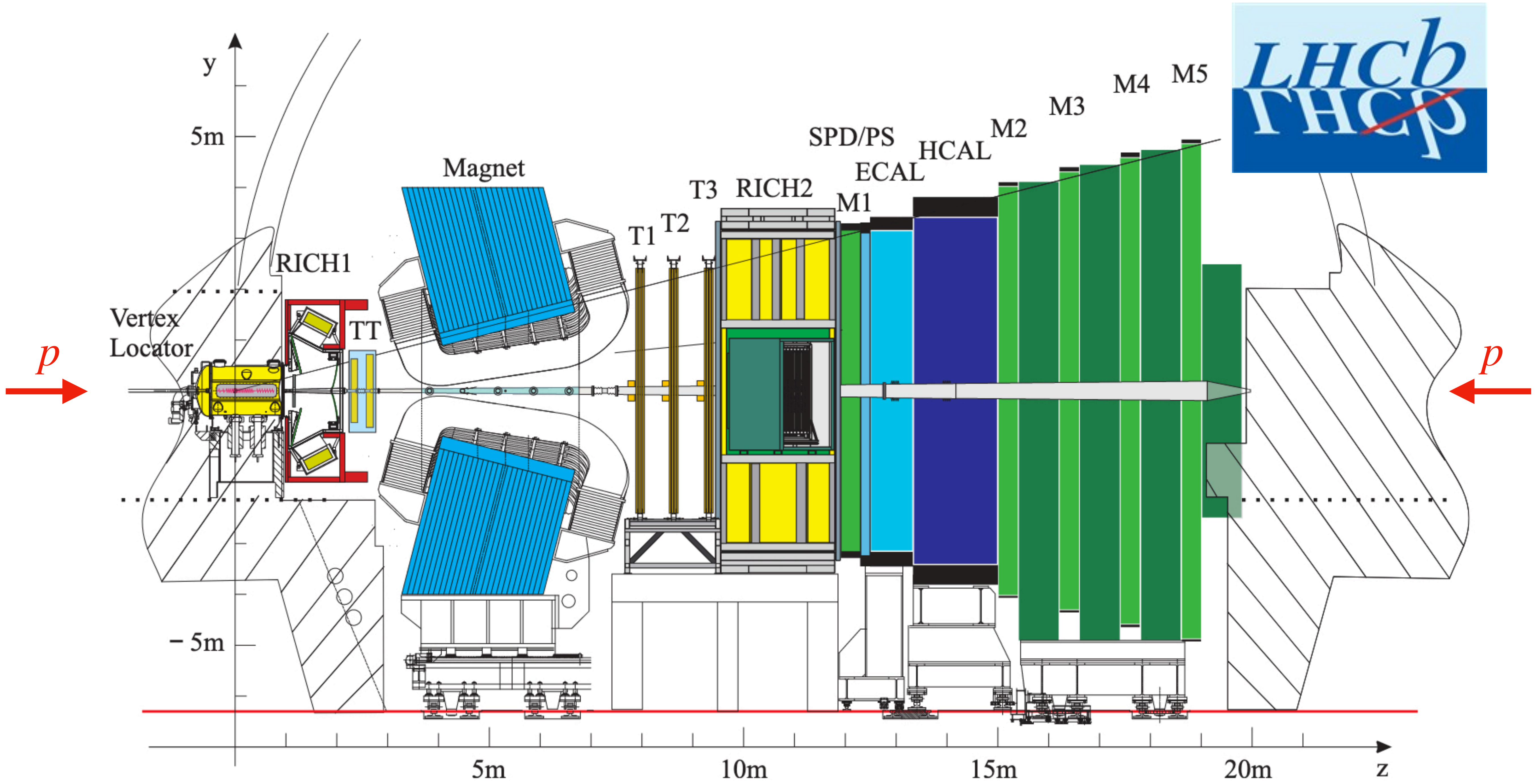
$\mathscr{B}(B^0_{(s)} \to \tau^\pm\mu^\mp) < 1.4\,(4.2) \cdot 10^{-5}$

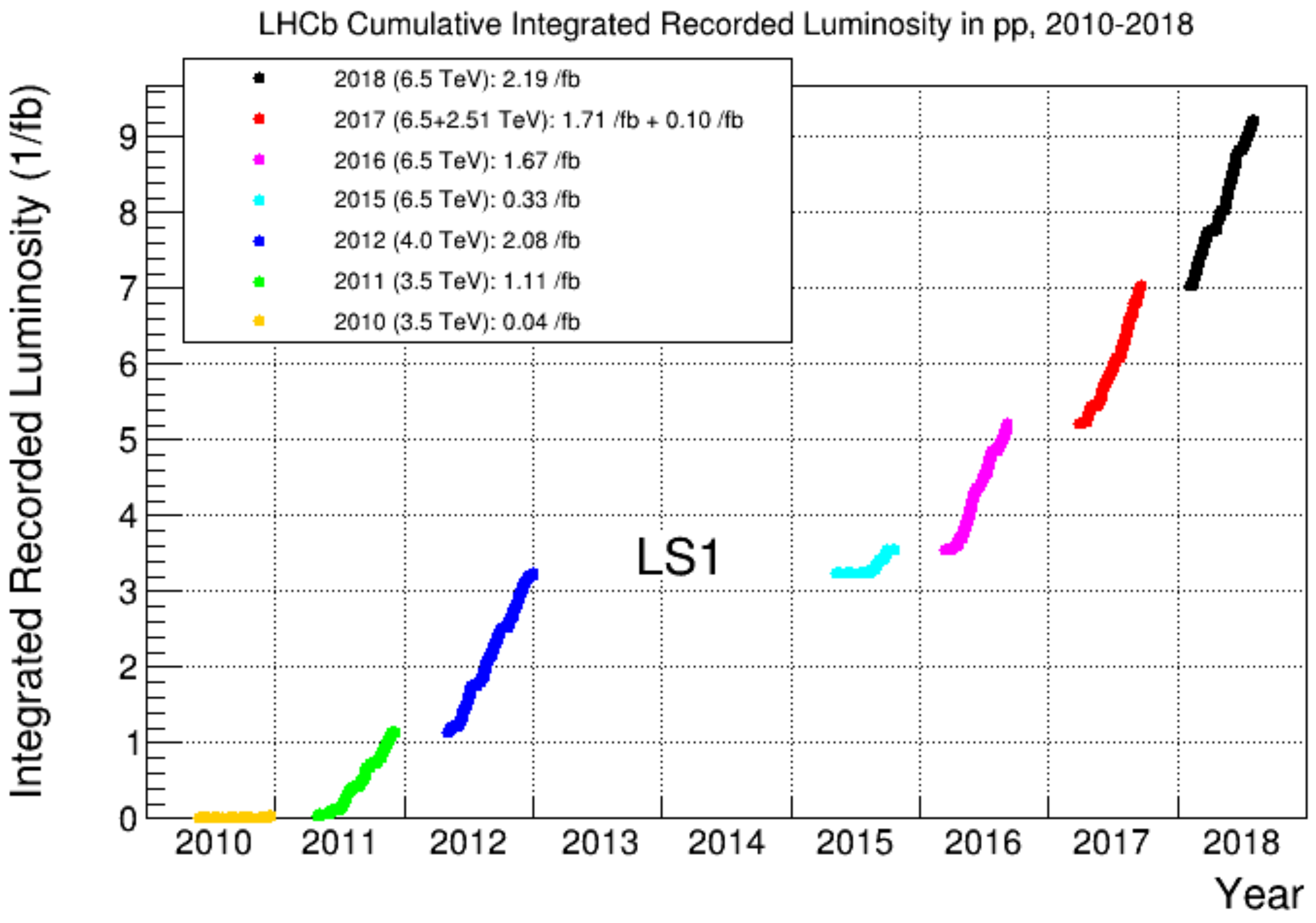
Phys. Rev. Lett. 123, 211801

- **Today's menu**: search for the $B^0 \to K^{*0}\tau^+\tau^-$ decay with 9 fb^{-1} pp collision data recorded at the LHCb experiment
- Rare $b \to sl^+l^-$ quark-level transition suppressed in the SM: $\mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-) \sim 10^{-7}$ (Phys. Rev. Lett. 120, 181802)

The LHCb detector

The LHCb detector (Int. J. Mod. Phys. A 30, 1530022)





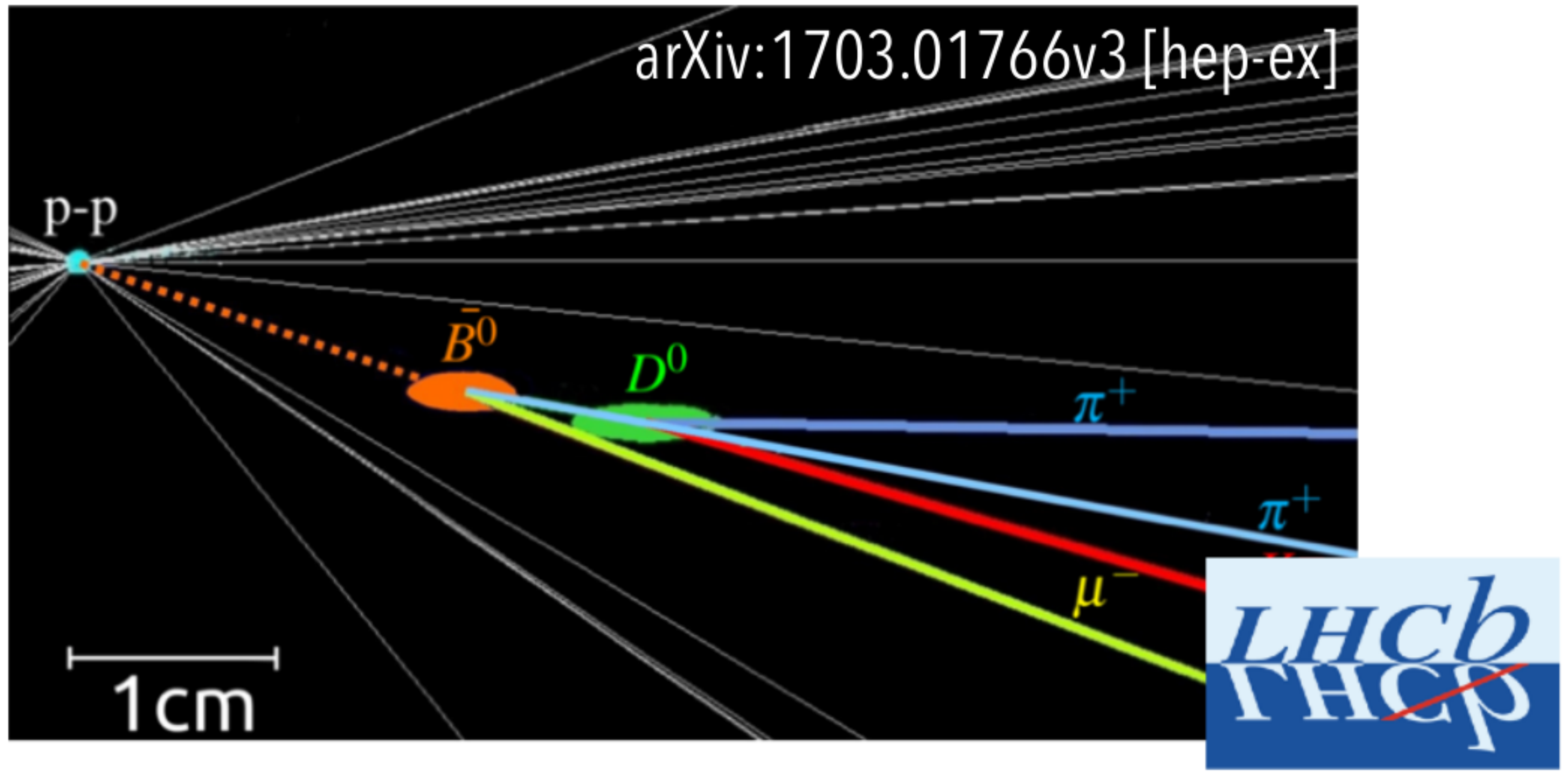
- $\sigma_{IP} = 15 + 29/p_T$ μm
- $\sigma_p/p = 0.5 - 1.0\,\%$, $p \in [2,200]$ GeV
- PID $\epsilon_\mu \sim 98\,\%$ with $\varepsilon_{\pi\to\mu} \sim 1\,\%$

LHCb vs Belle 2

LHCb

$pp \rightarrow B\bar{B}X$

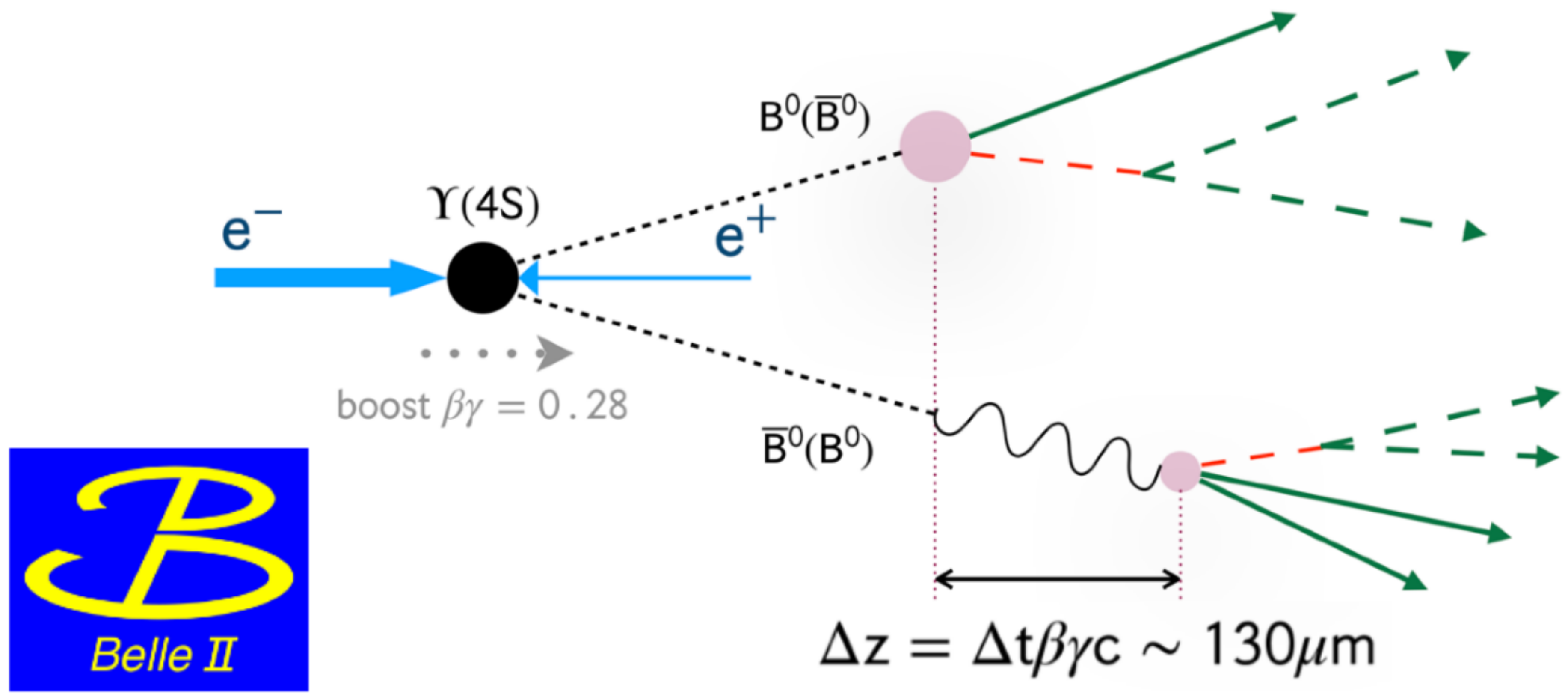
- σ_{bb} ~ hundreds of μb (B_s, B_c, Λ_b, ...)
- L ~ 10^{32} cm^{-2} s^{-1}
- Lower trigger efficiency
- Higher background
- B mesons decay length ~ 1 cm



Belle 2

$e^+e^- \rightarrow \Upsilon(4S) \rightarrow B\bar{B}$

- σ_{bb} ~ 1 nb
- L ~ 10^{34} cm^{-2} s^{-1}
- High trigger efficiency
- Clean environment, low background
- B mesons decay length ~ hundreds of μm



Search for the $B^0 \to K^{*0}\tau^+\tau^-$ decay

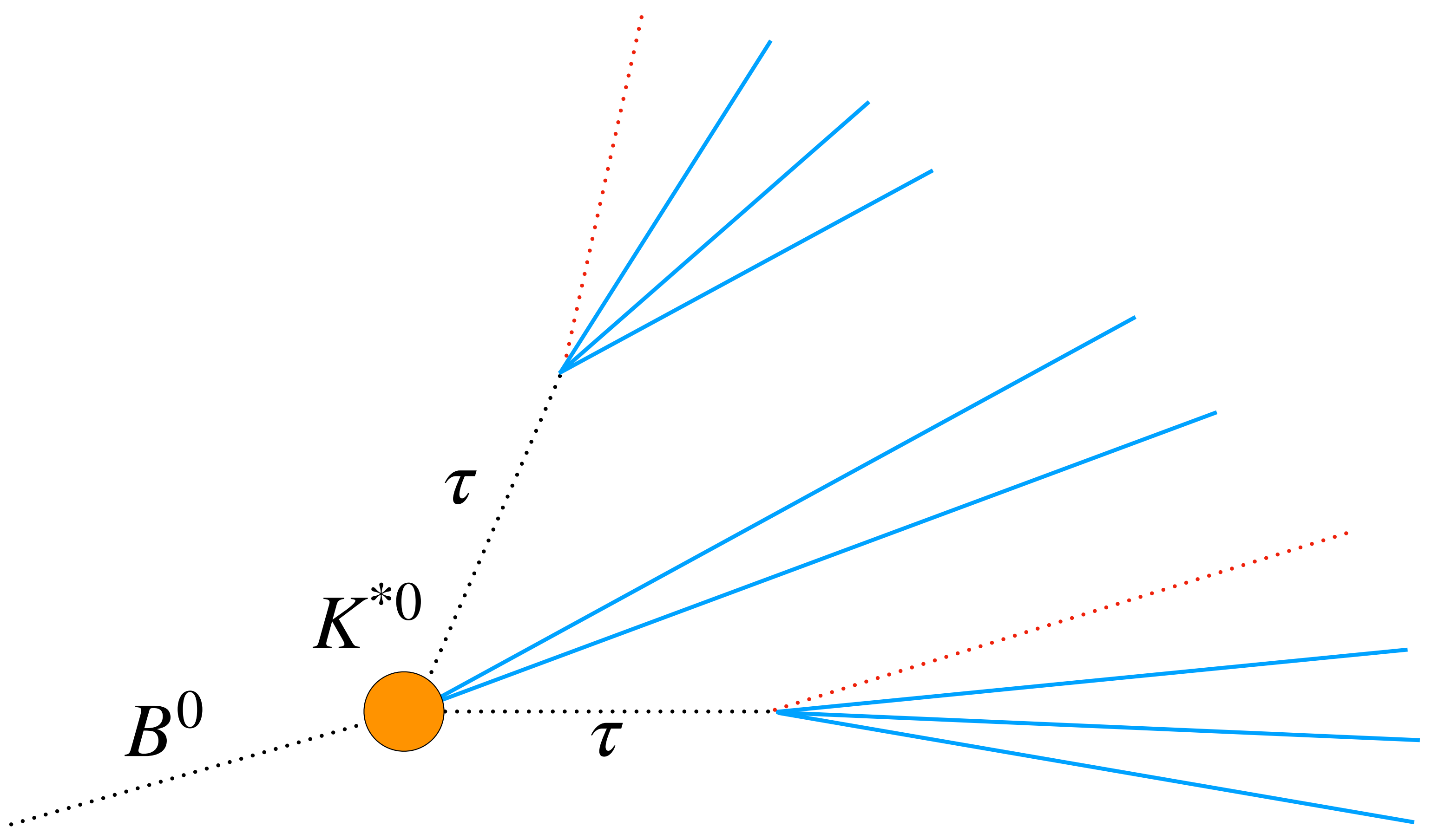
Introduction

- **First direct limit** set recently by the **Belle collaboration** with 711 fb^{-1} data:
  - τ reconstructed via one-prong decays: $\tau \to e\,\bar{\nu}_e\,\nu_\tau$, $\tau \to \mu\,\bar{\nu}_\mu\,\nu_\tau$ and $\tau \to \pi\nu_\tau$
  - $\mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-) < 2.0 \cdot 10^{-3}$ at $90\,\%$ CL (arXiv:2110.03871)

- I analyzed the **full LHCb dataset of ~9 fb^{-1},** decay reconstructed in **two final states**

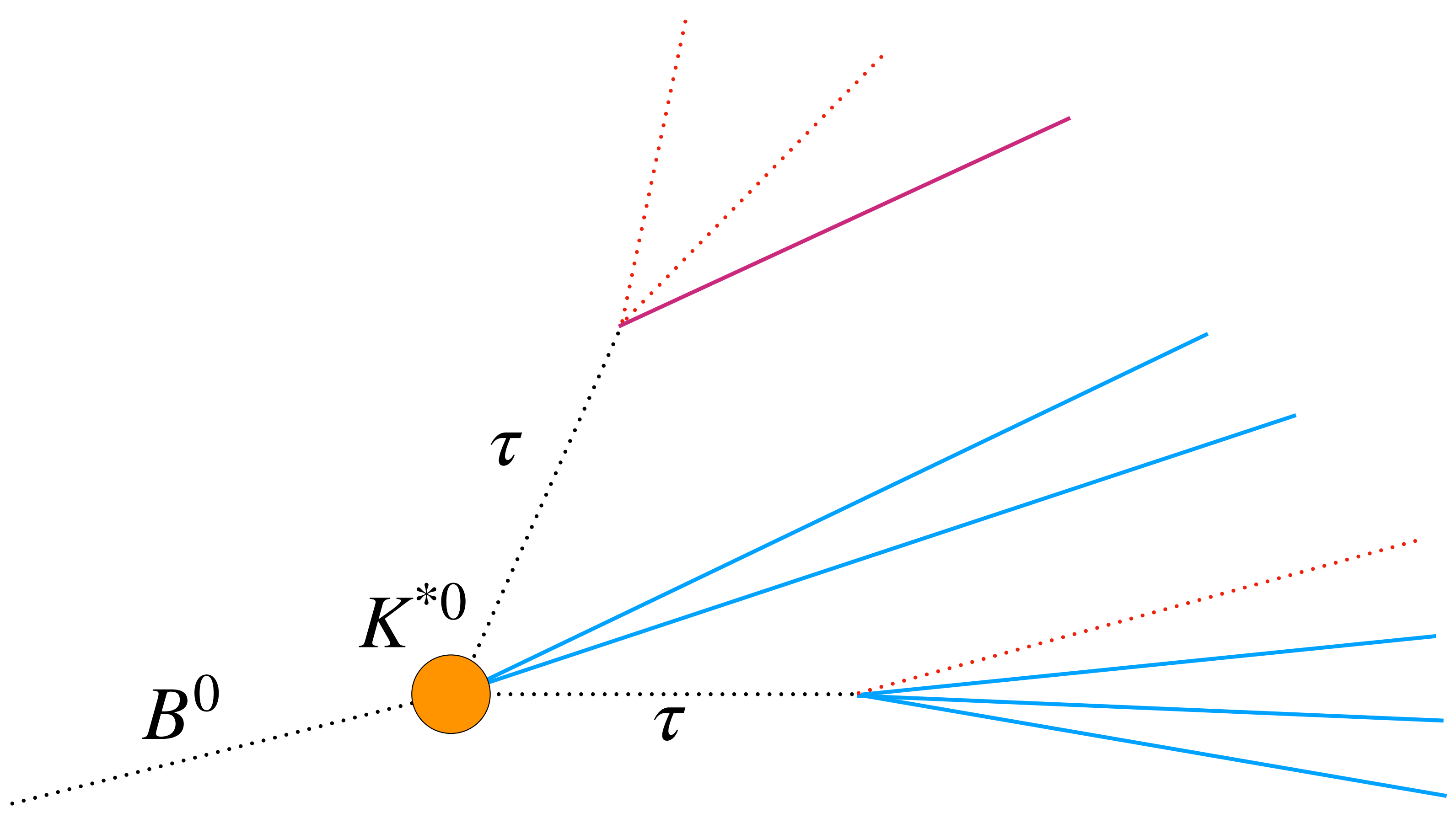
Fully hadronic final state

$B^0 \to K^{*0}(\to K^-\pi^+)\,\tau^+(\to \pi^+\pi^+\pi^-(\pi^0)\,\bar{\nu}_\tau)\,\tau^-(\to \pi^+\pi^-\pi^-(\pi^0)\,\nu_\tau)$



Mixed hadronic-leptonic final state

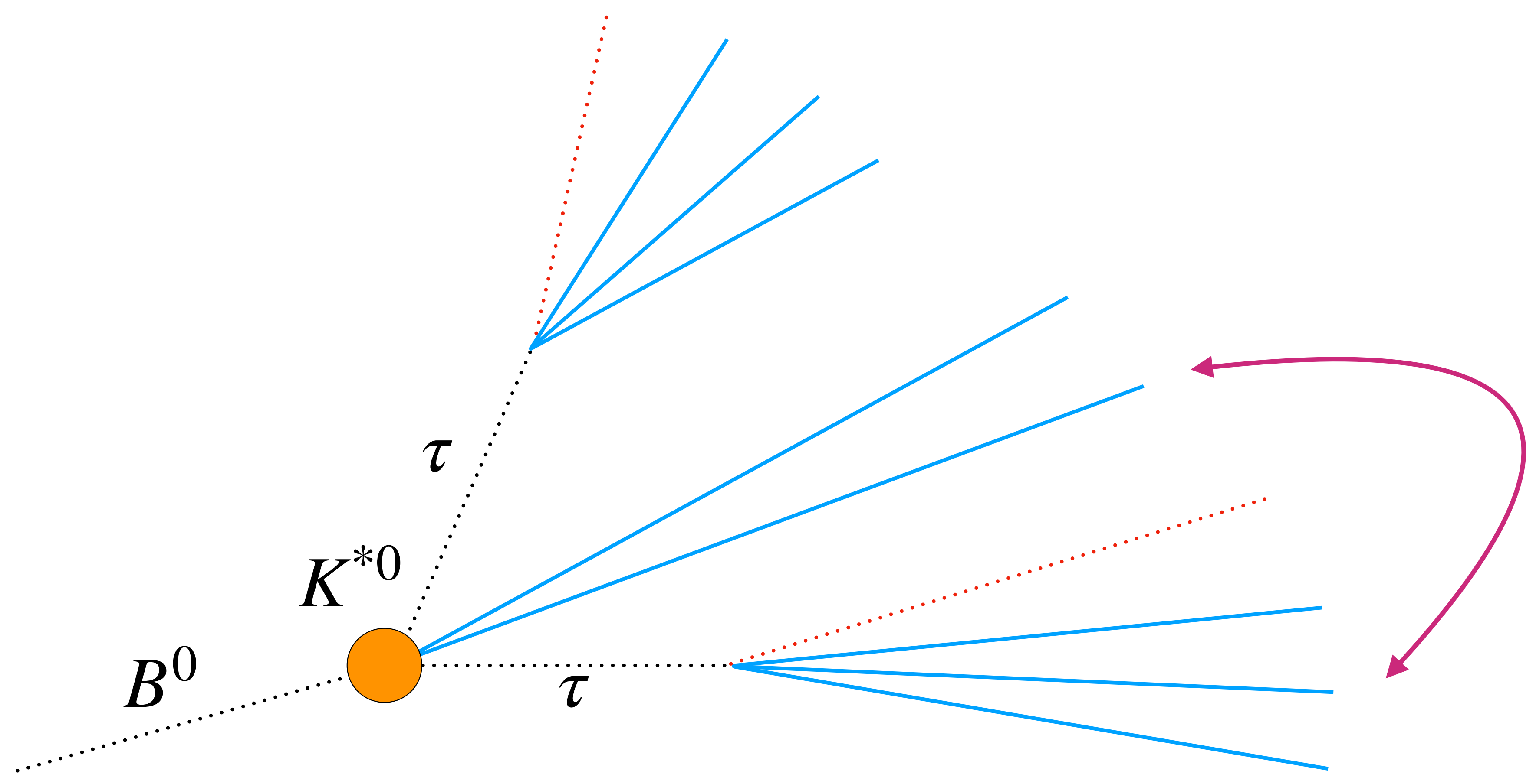
$B^0 \to K^{*0}(\to K^-\pi^+)\,\tau^+(\to \pi^+\pi^+\pi^-(\pi^0)\,\bar{\nu}_\tau)\,\tau^-(\to \mu^-\nu_\tau\bar{\nu}_\mu)$



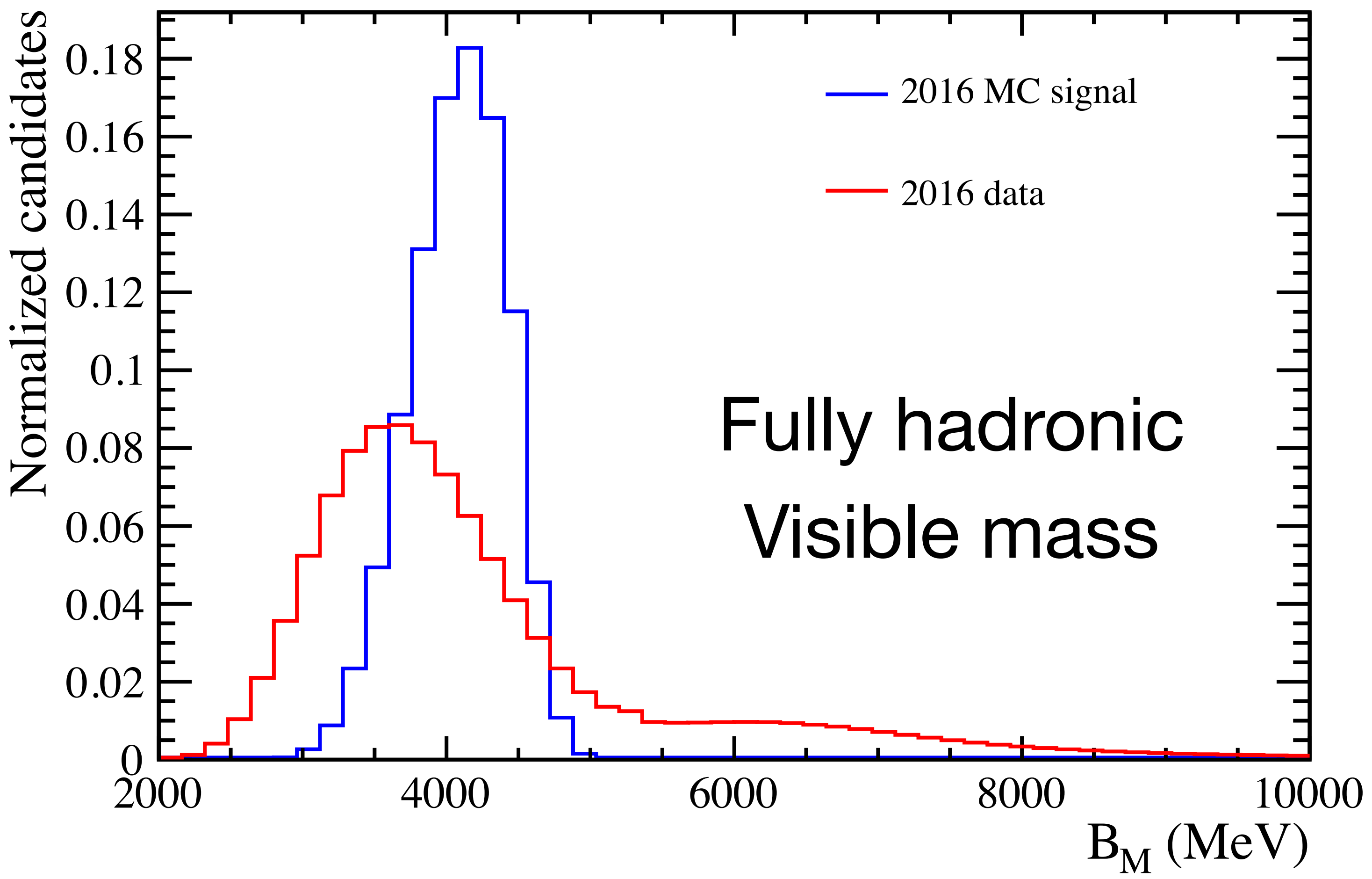
Challenge 1: background

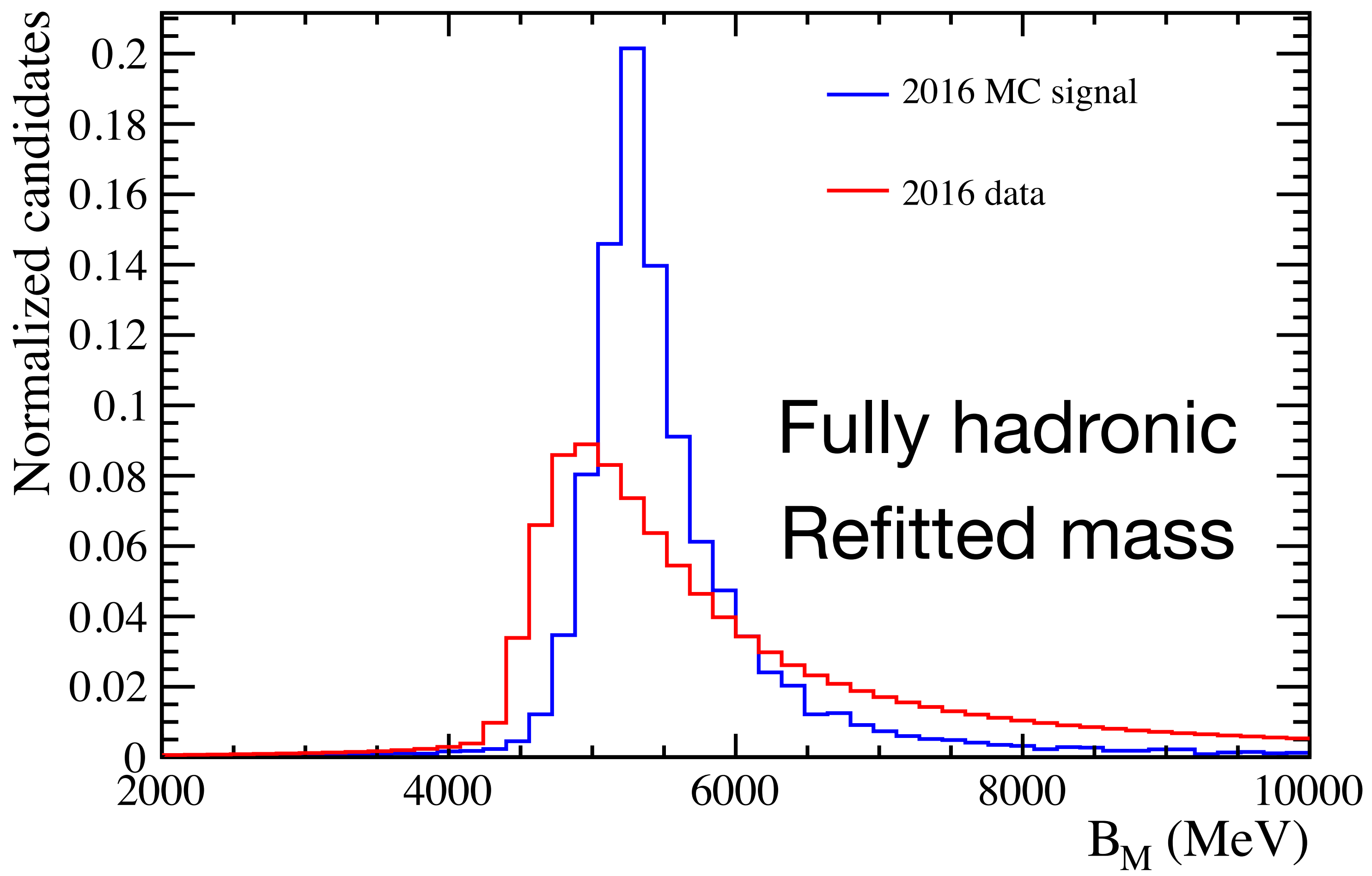
- ~ **2 (8) fully hadronic (mixed hadronic-leptonic) events** expected in the detector acceptance
- About **200 million processed events to be analyzed**! MVA techniques in order to suppress most of background

- **Many charged tracks in the final state** (8 for fully hadronic, 6 for mixed hadronic-leptonic)
  - Fully hadronic (mixed hadronic-leptonic) final state has on average ~10 (2) reconstructed candidates per event
  - Mainly due to **multiple signal candidates built with the same final state particles**

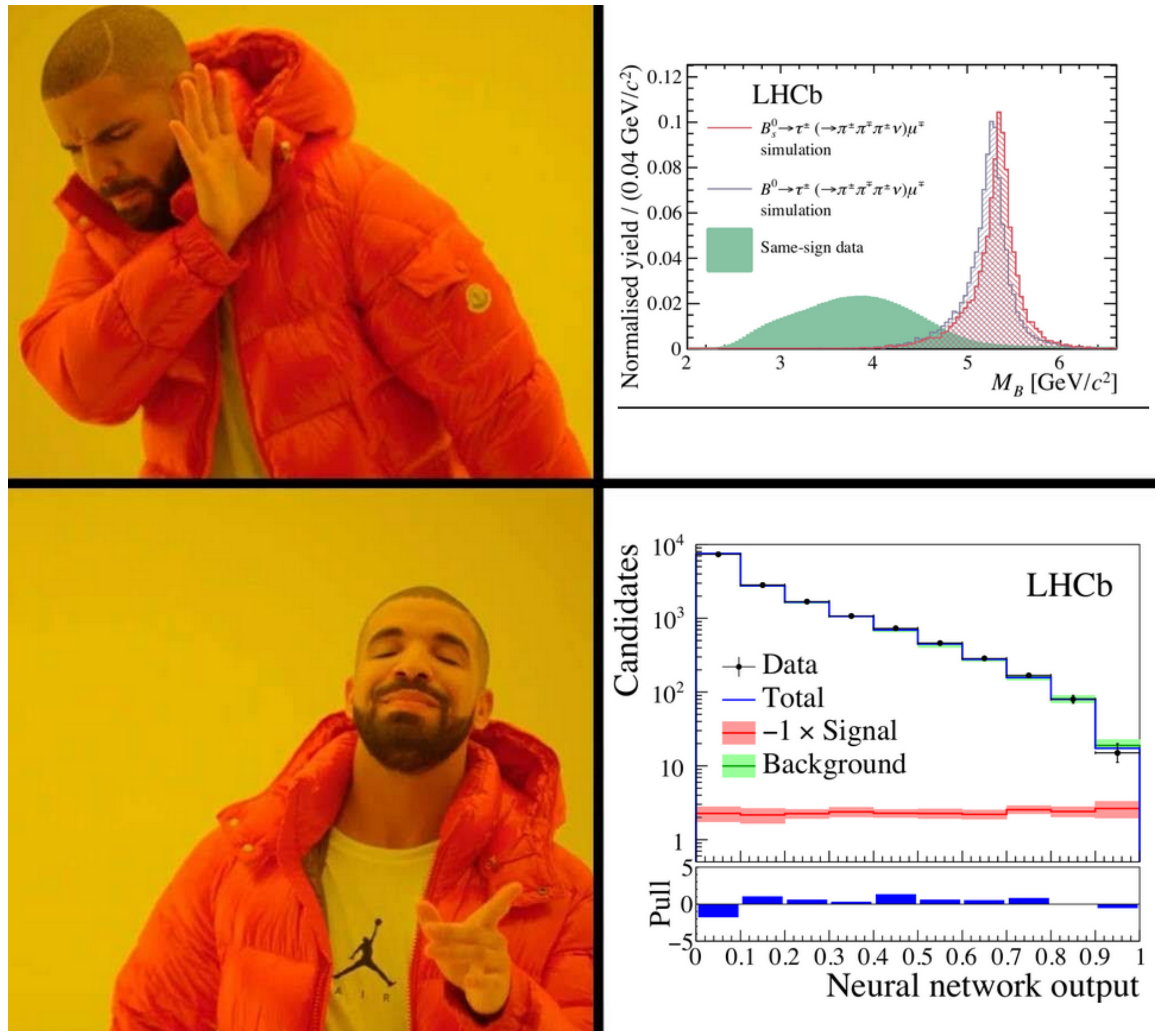


Challenge 2: missing energy





- Signal and background-dominated data peak at very close values: mass fit is not enough discriminating
- Likelihood fit performed on **boosted decision tree (BDT)**
  - Background description data-driven



Analysis strategy

1 Event selection

- Preselection: candidates reconstruction, particle identification and trigger
- Cut-based selection: isolation variables, used to reject most trivial background
- MVA selection: two BDTs in sequence used to suppress most of the background

2 Likelihood fit

- **Performed on output of a third BDT** trained after the full selection
- **Background modeled with data-driven method**, from events in control regions of the data

3 Branching ratio computation

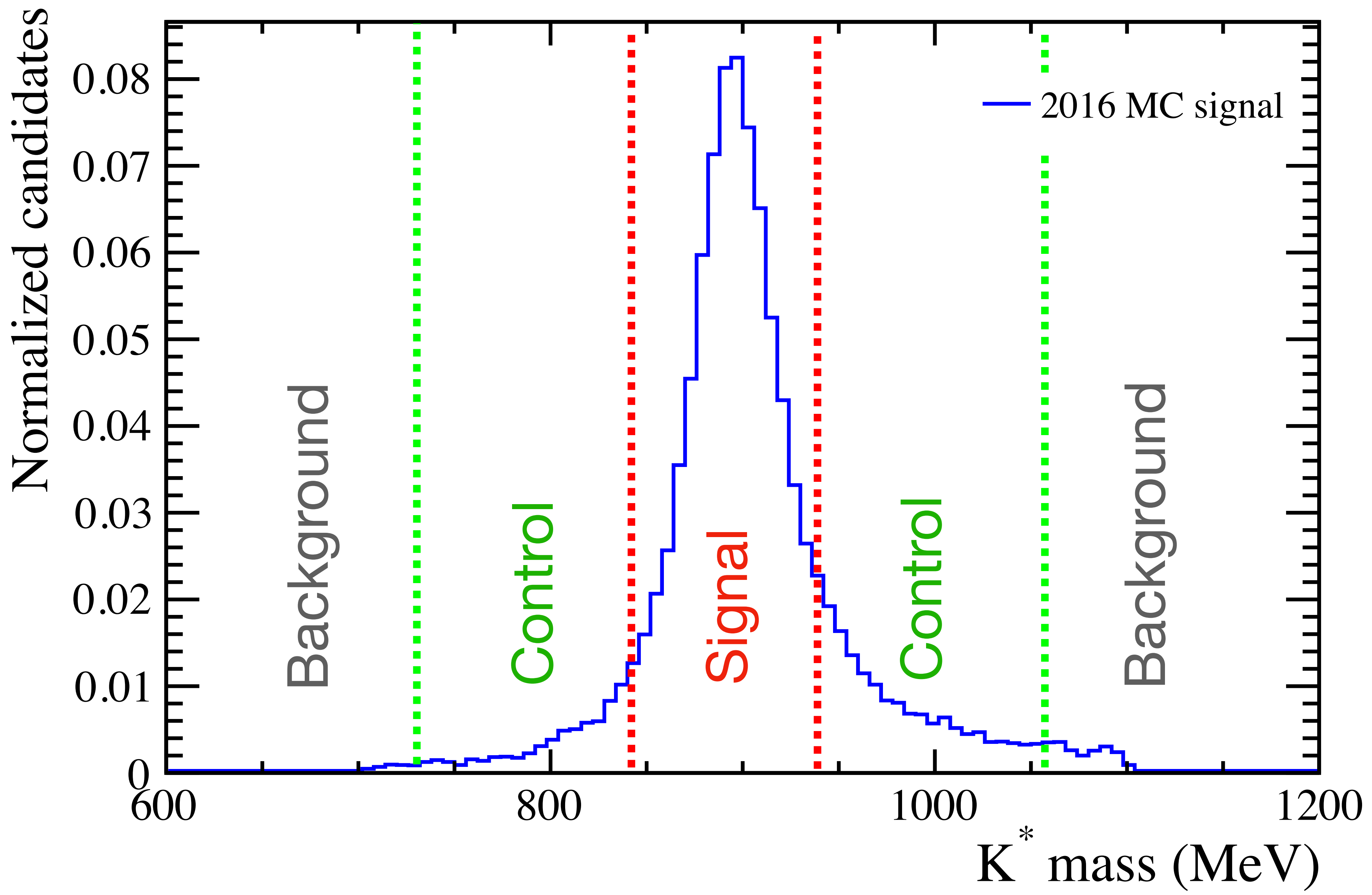
- Branching ratio measured relatively to **normalization channel** $B^0 \to D_s^+(\to K^+K^-\pi^+)\,D^-(\to \pi^-\pi^-K^+)$

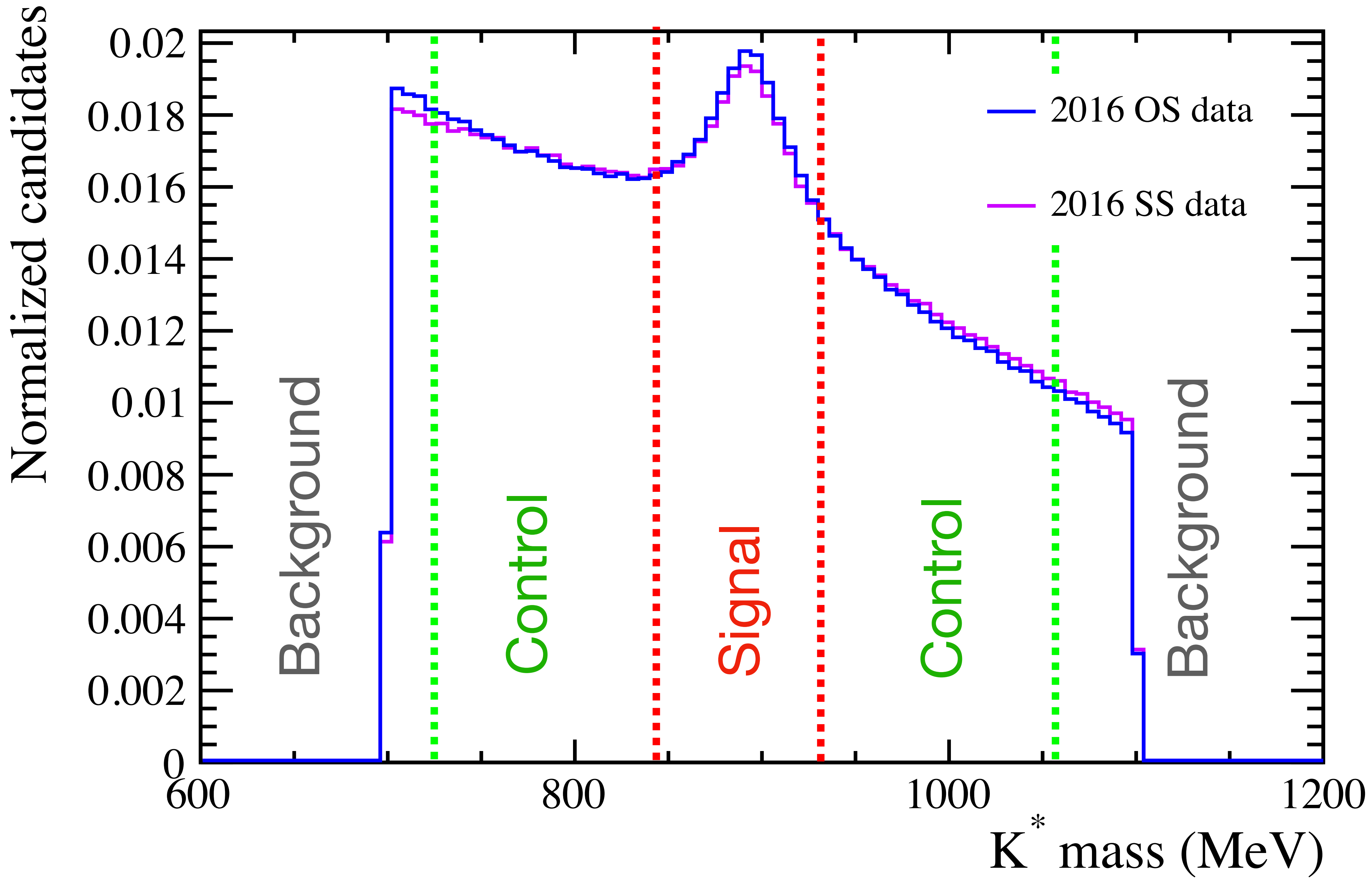
$$\mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-) = \frac{N_{sig}^{obs}}{\epsilon_{sig}\cdot\sigma L} \to \frac{\mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-)}{\mathscr{B}(B^0 \to D_s^+D^-)} = \frac{N_{sig}^{obs}}{\epsilon_{sig}\cdot\cancel{\sigma L}}\frac{\epsilon_{norm}\cdot\cancel{\sigma L}}{N_{norm}^{obs}} \to \mathscr{B}(B^0 \to K^{*0}\tau^+\tau^-) = \frac{N_{sig}^{obs}}{\epsilon_{sig}}\frac{\epsilon_{norm}}{N_{norm}^{obs}}\cdot\mathscr{B}(B^0 \to D_s^+D^-)$$

- If no signal is observed, a limit on the branching ratio is set

Analysis regions

- To describe the background with data-driven method, need to identify **background-dominated regions in data**
- Regions defined using K^{*0} **mass distribution**:
  - Signal region: $M_{K^{*0}} \in [846,938]$ MeV, **most signal-like** region, **likelihood fit** performed on data in it
  - Control region: $M_{K^{*0}} \in [724,846]$ or $M_{K^{*0}} \in [938,1053]$ MeV, used to extract **background template** for likelihood fit
  - Background region: $M_{K^{*0}} \in [700,724]$ or $M_{K^{*0}} \in [1053,1100]$ MeV, data for **background sample for BDT training**
- In order not to introduce any bias, the **fit BDT must be uncorrelated with** K^{*0} **mass**

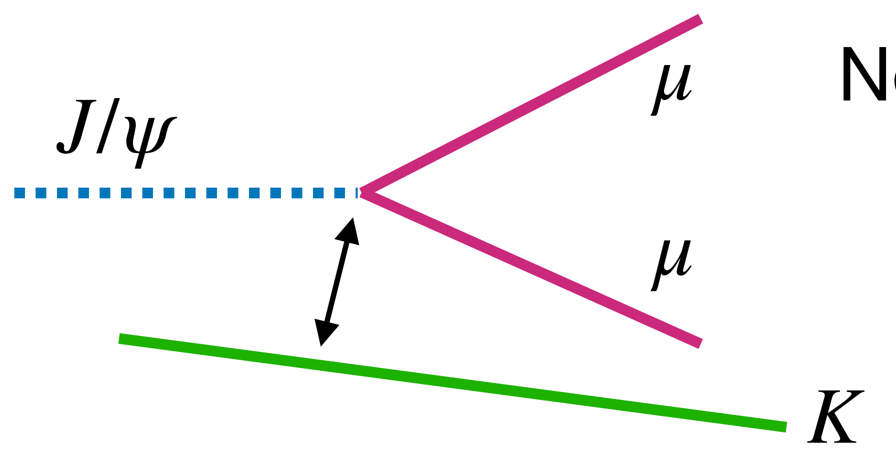




Event selection

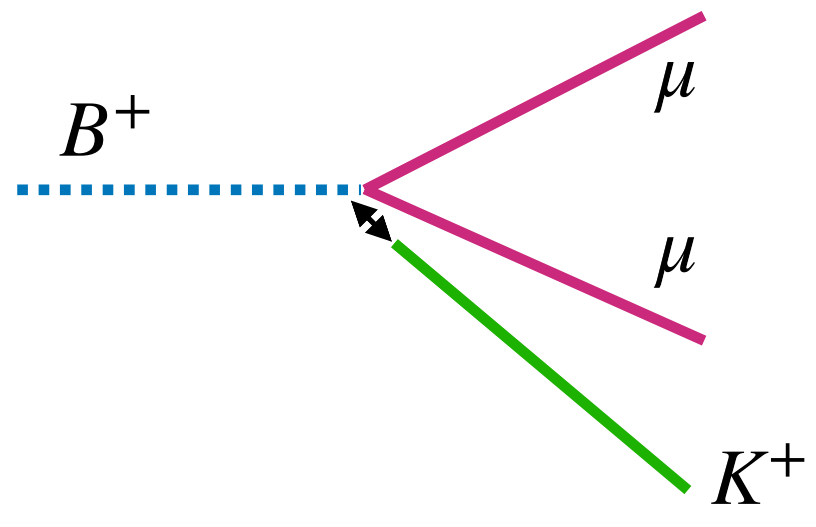
Cut-based selection

- **Isolation variables evaluate "activity" near reconstructed candidate**
- Ex: let's select $J/\psi \to \mu^+\mu^-$ and reject partially reco $B^+ \to K^+\mu^+\mu^-$ by adding a track to the muon vertex



New vertex χ^2 > original vertex χ^2

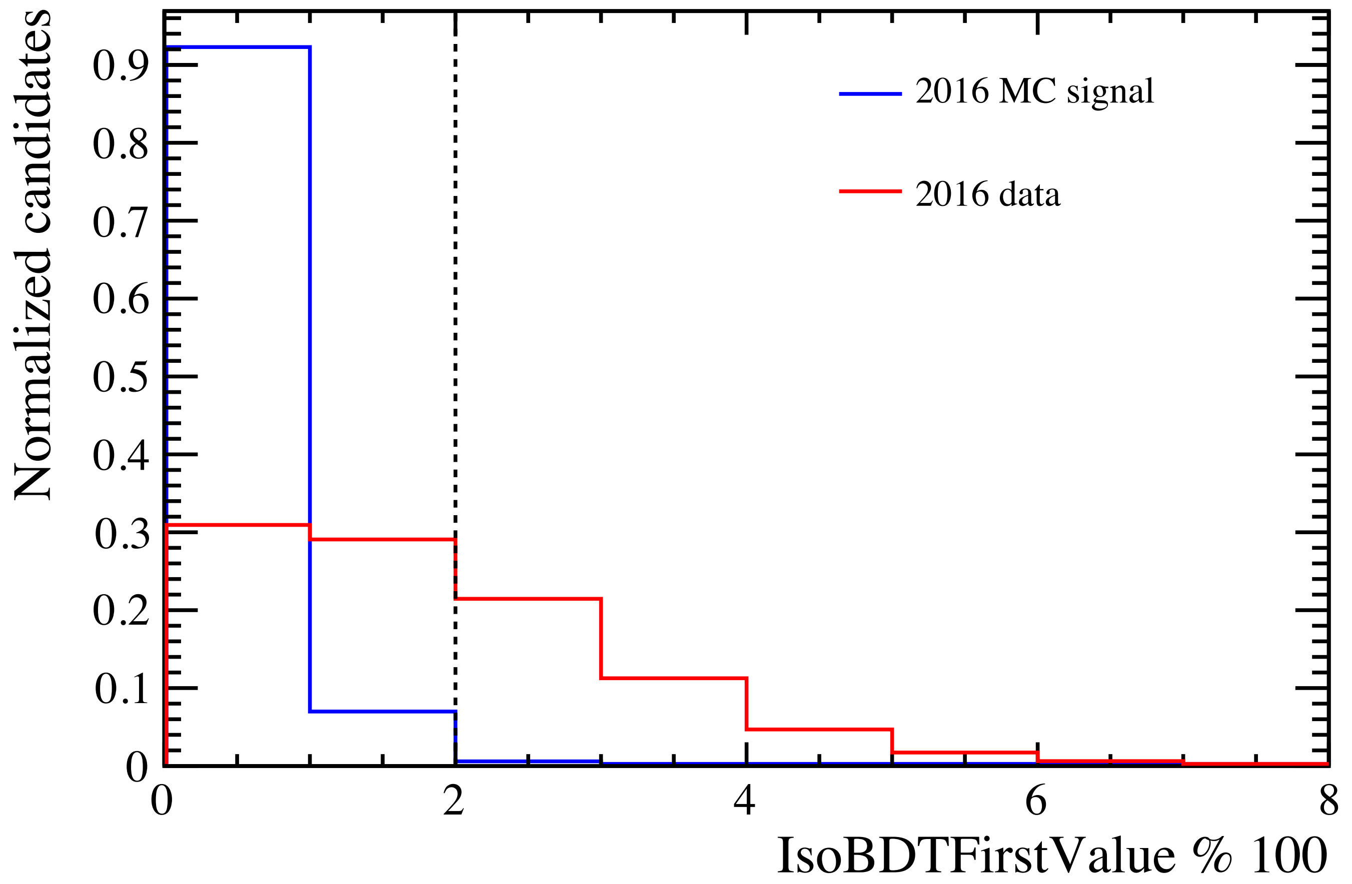
High level of isolation
Genuine $J/\psi \to \mu^+\mu^-$ decay

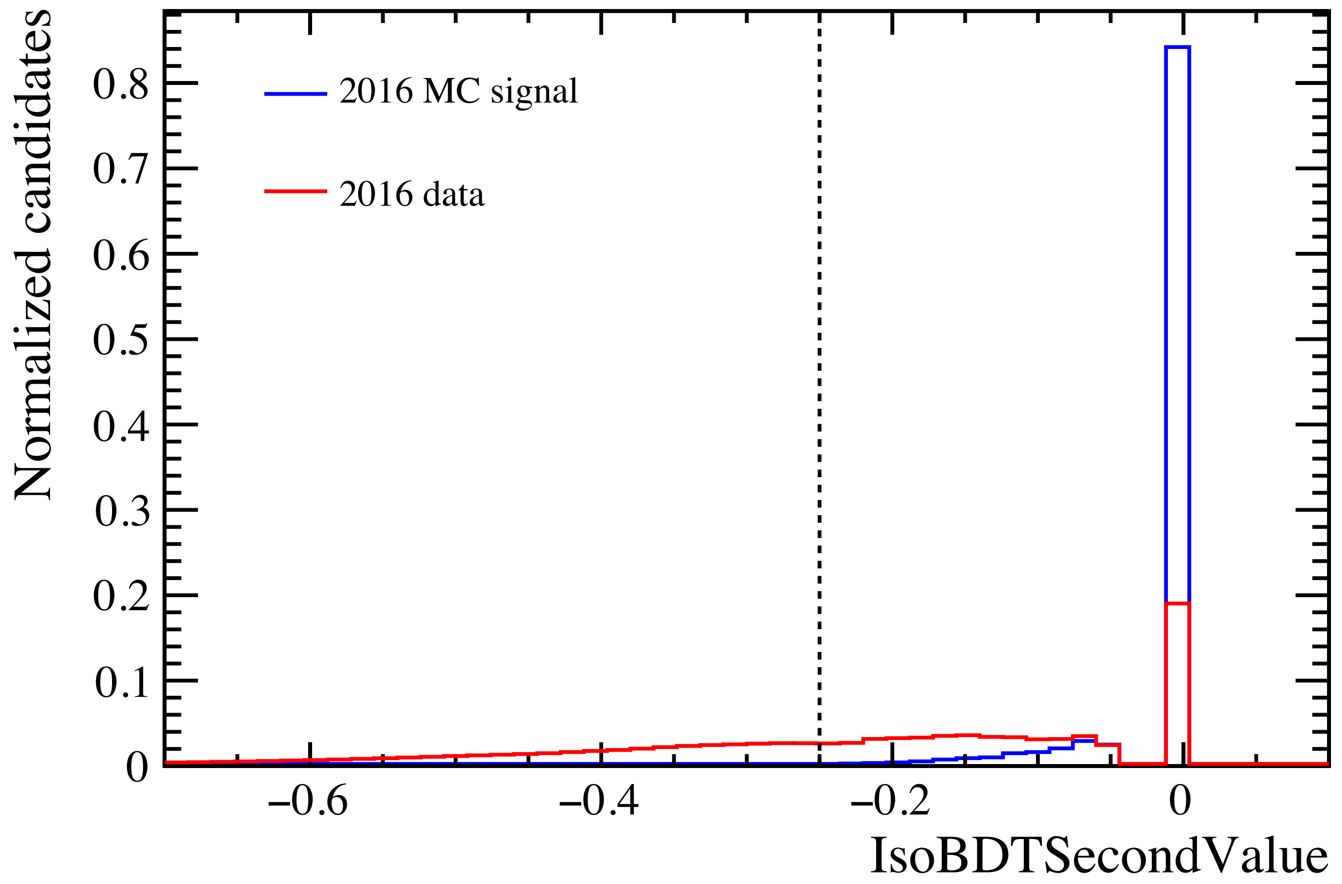


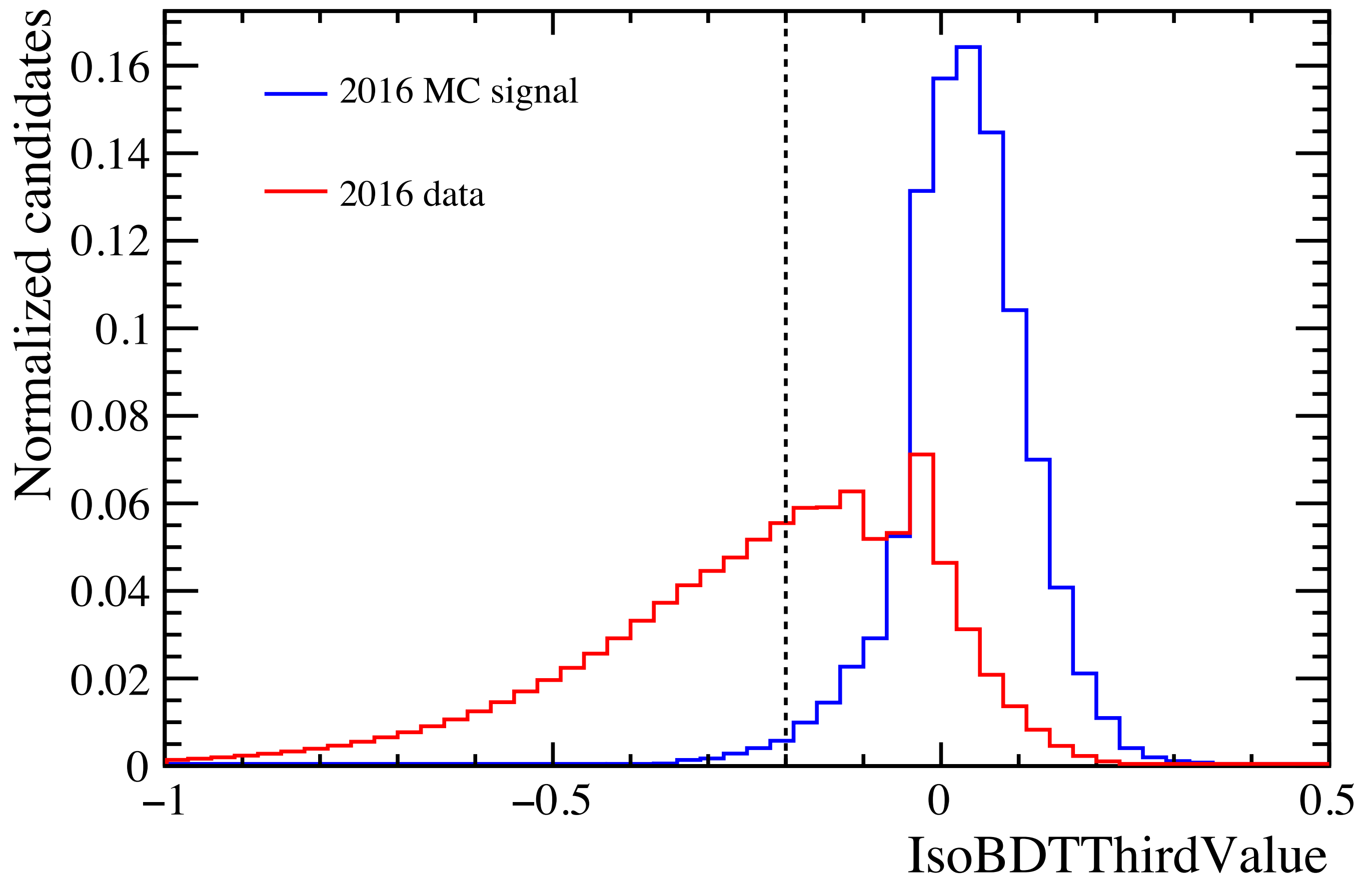
New vertex $\chi^2 \simeq$ original vertex χ^2

Low level of isolation
Partially reco background

- **Three track isolation variables** defined for each particle in the decay chain, isolation level derived from a BDT
- Cuts to reject most trivial background: $\epsilon_{sig} \sim 95\,\%$, $\epsilon_{bkg} \sim 50\,\%$

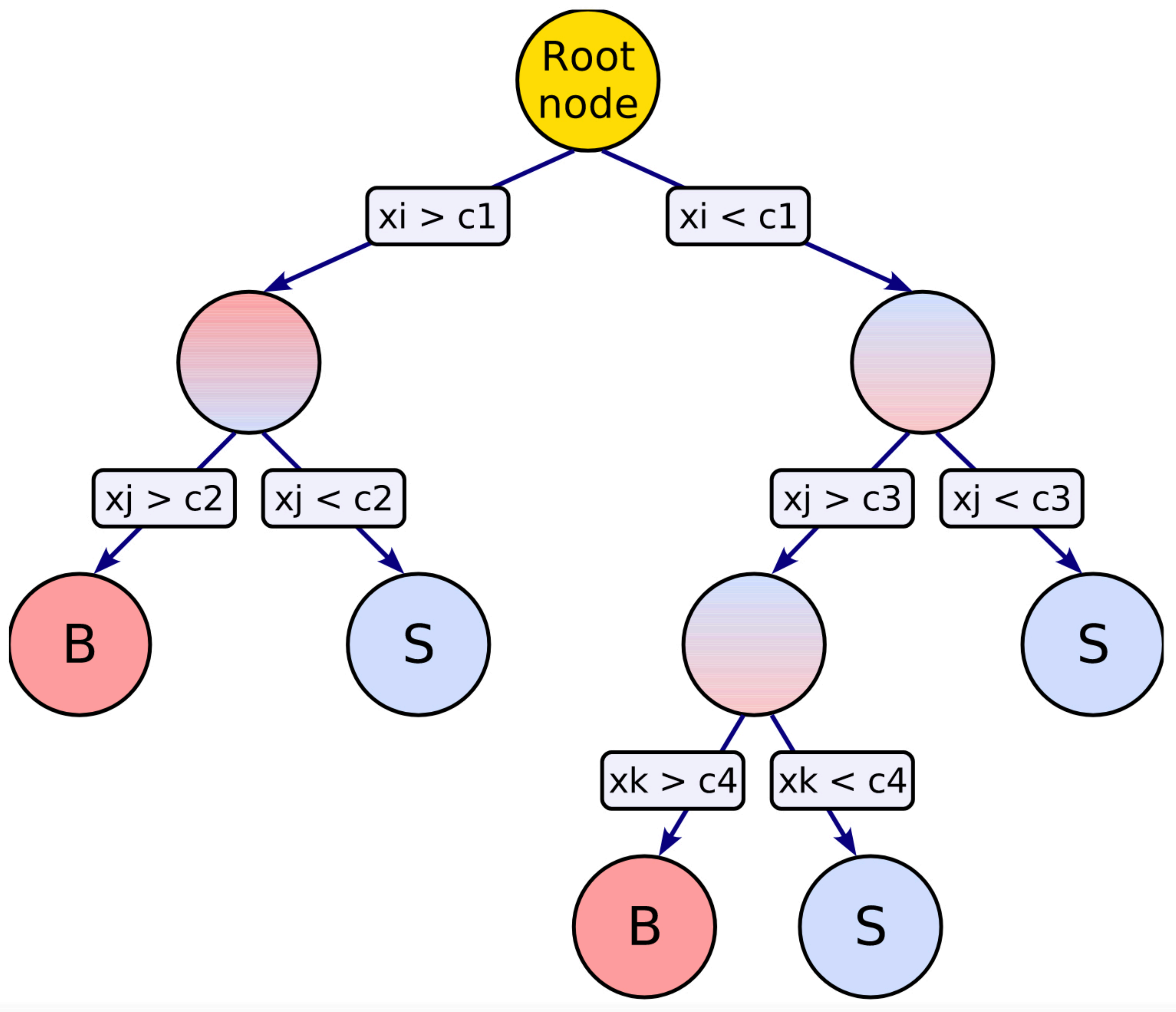






MVA selection: boosted decision trees

- BDTs are algorithms to **classify events in two (or more) categories** (e.g. signal and background)
- They use *training samples* and a list of variables to "learn" to distinguish between two categories of events
- Building block of a BDT is the *decision tree*



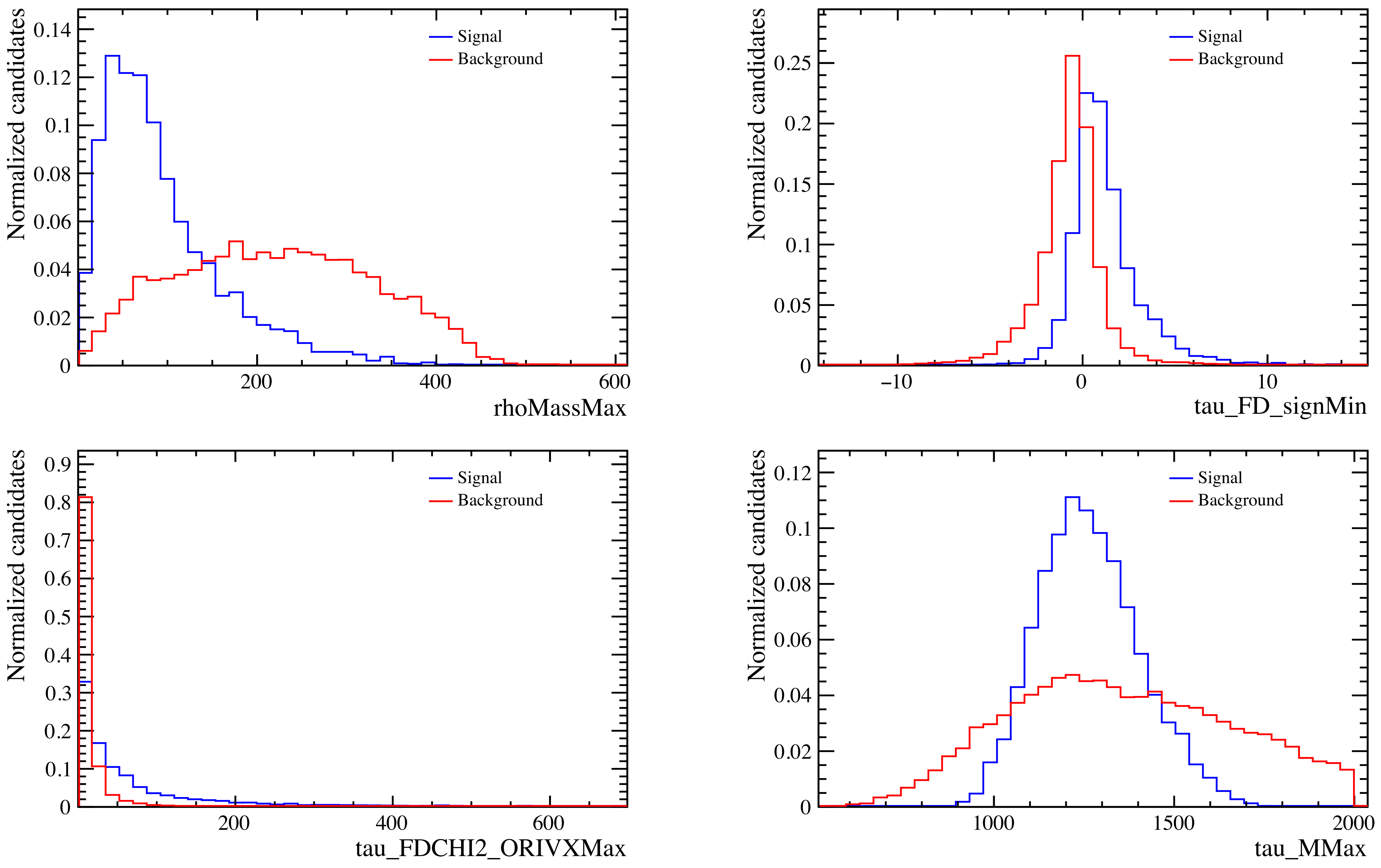
1. In a decision tree **samples are split using at each step the variable maximizing the discrimination**
2. Final nodes are labelled as "signal" or "background" depending on the majority of events belonging to them
3. **More decision trees are built**, at each iteration giving larger weights to misclassified events in order to improve the performance
4. The **decision trees are linearly combined** to form a BDT

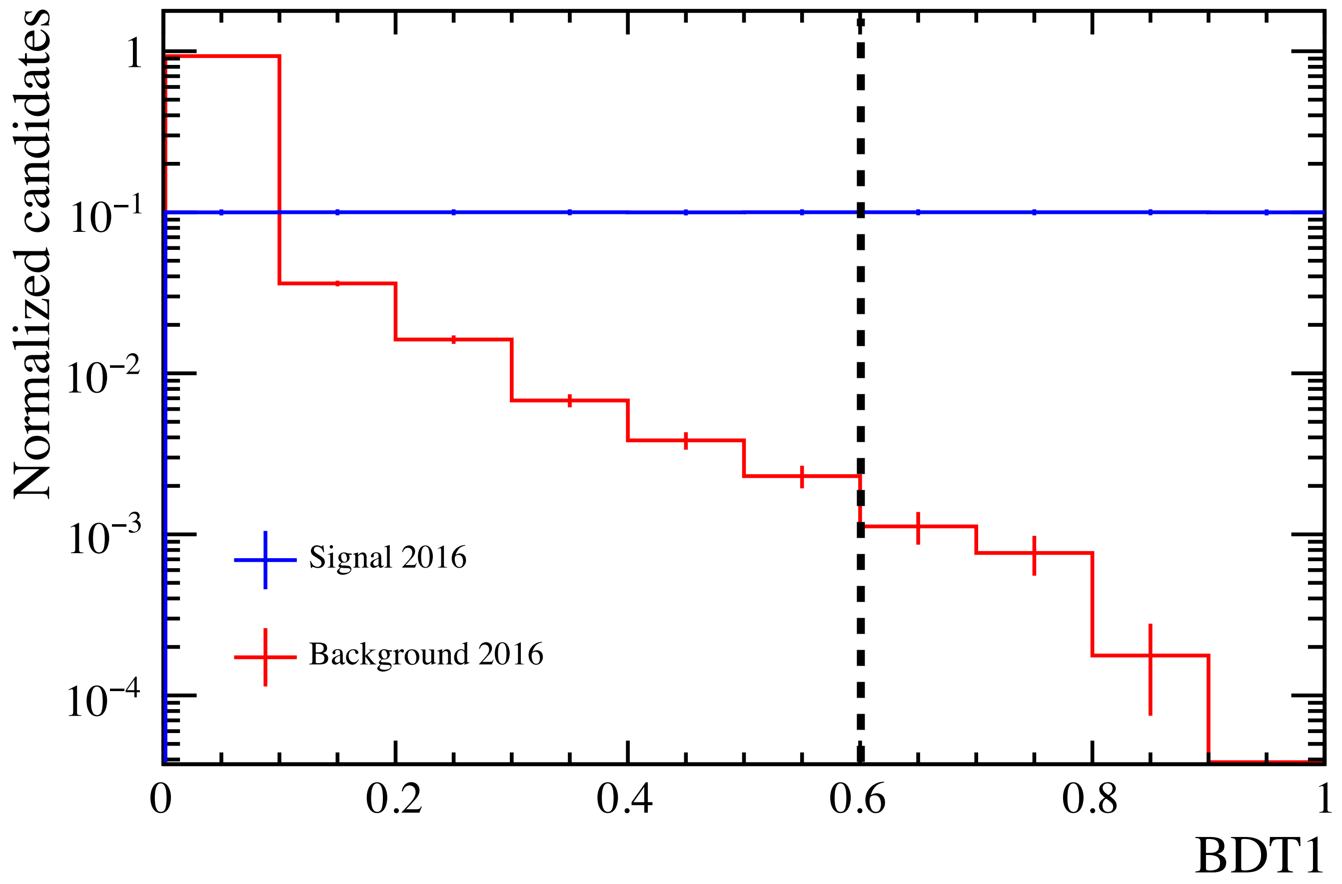
MVA selection

- **Two BDTs trained in sequence** used to suppress most of the background, **fit BDT** trained after full selection
  - Trained with MC signal events and data from the background region
- Background model for fit built from K^{*0} mass control region: **fit BDT must be uncorrelated with K^{*0} mass**
- **Variables chosen with iterative procedure** in order to maximize discriminating power
- BDT1 and BDT2 working points chosen to maximize sensitivity

Fully-hadronic: MVA selection and top-ranking variables

BDT1 (2016 example)





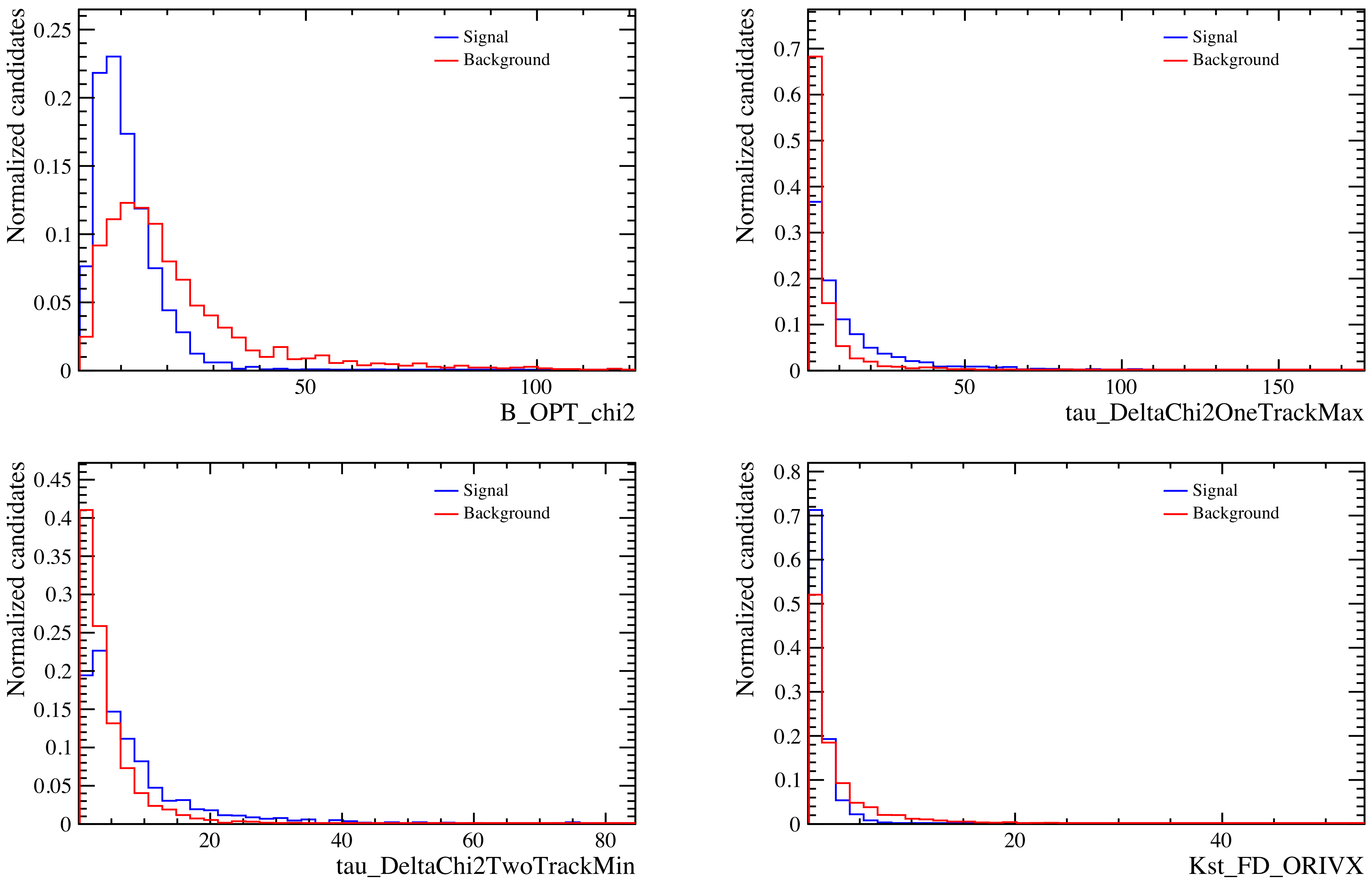
BDT1 > 0.6

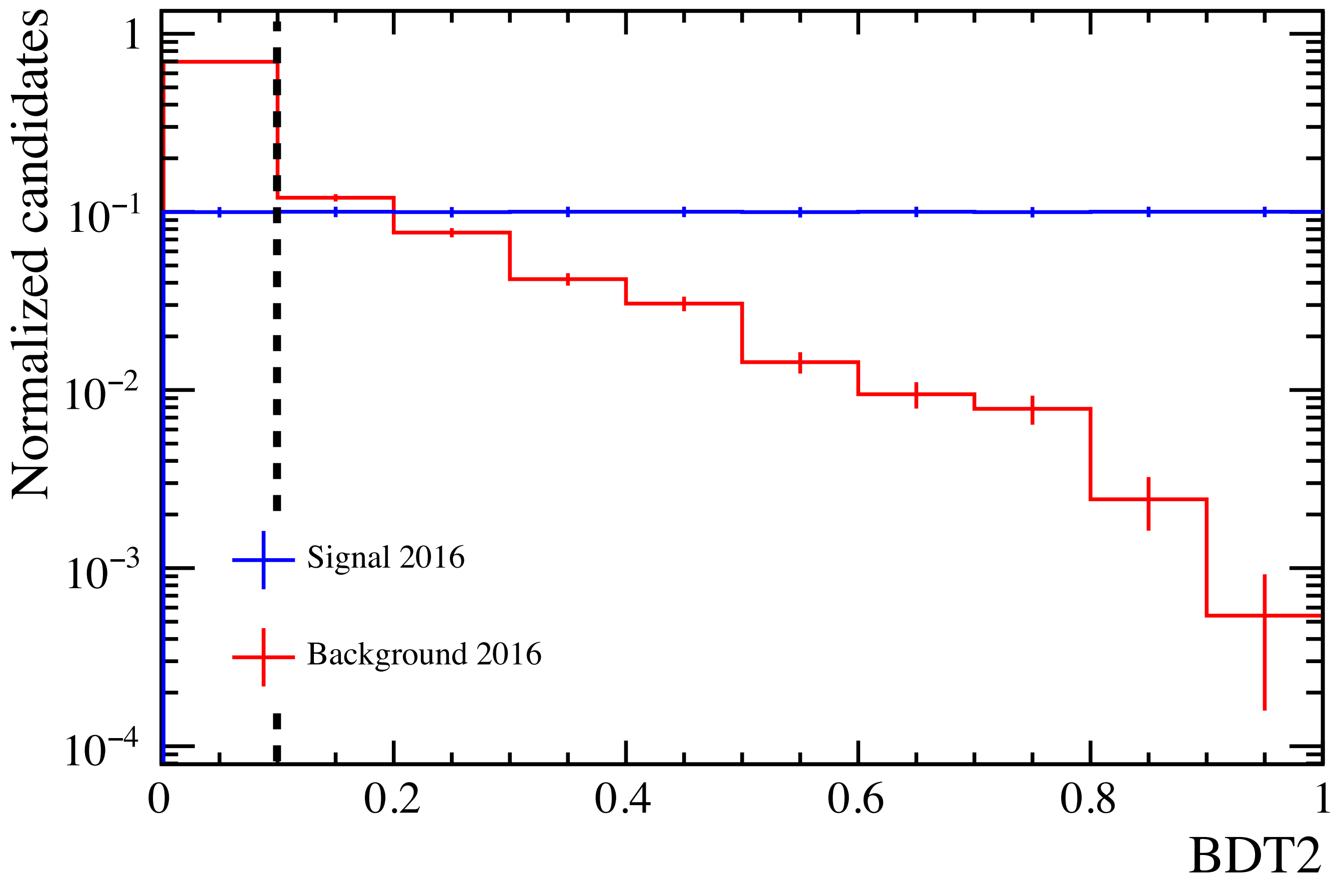
On training samples:

$\epsilon_{sig} \sim 40\,\%$

$\epsilon_{bkg} \sim 0.3\,\%$

BDT2 (2016 example)





BDT2 > 0.1

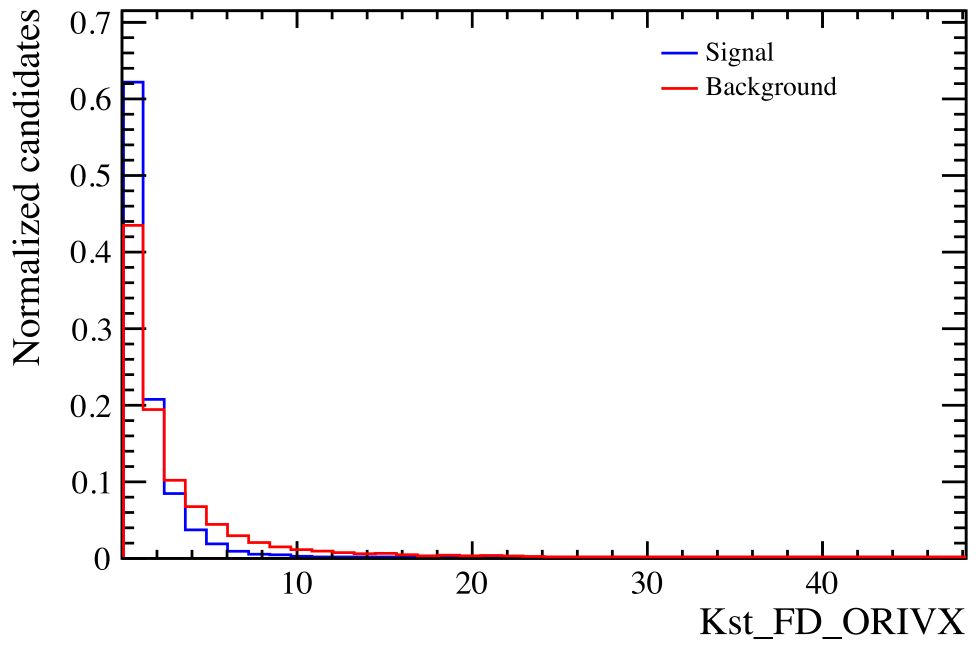
On training samples:

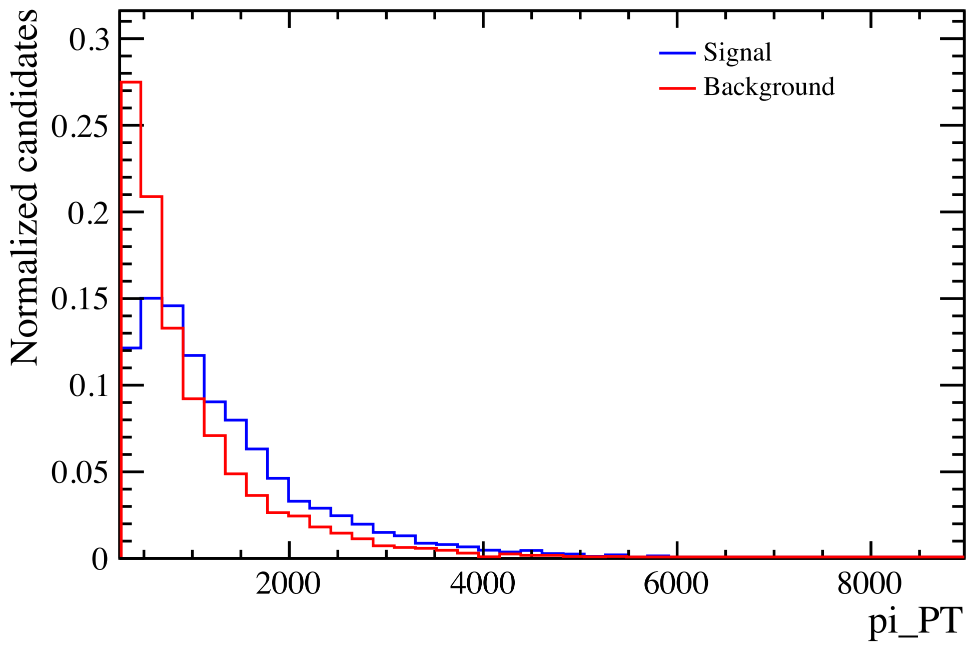
$\epsilon_{sig} \sim 90\,\%$

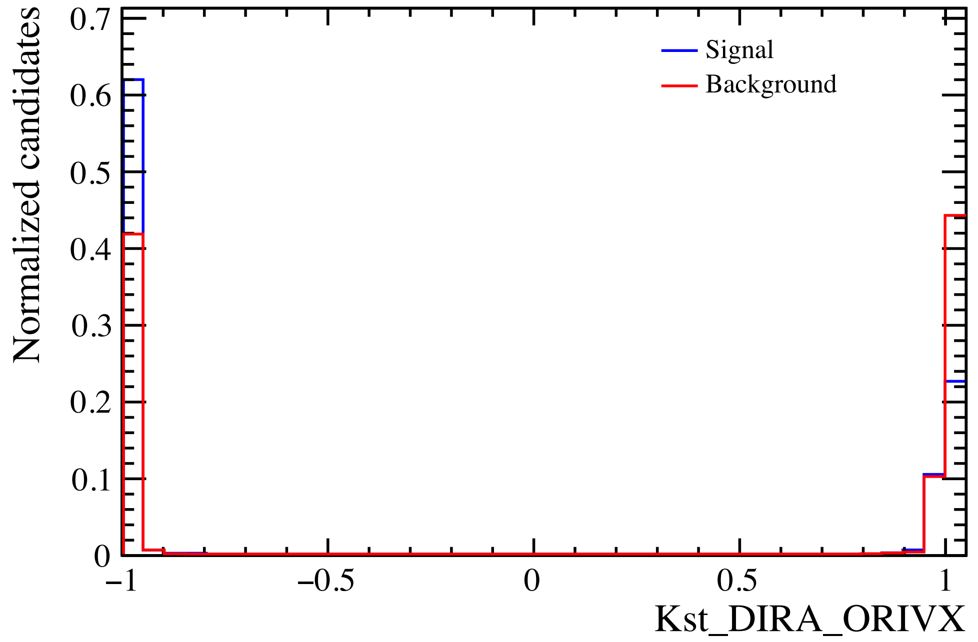
$\epsilon_{bkg} \sim 40\,\%$

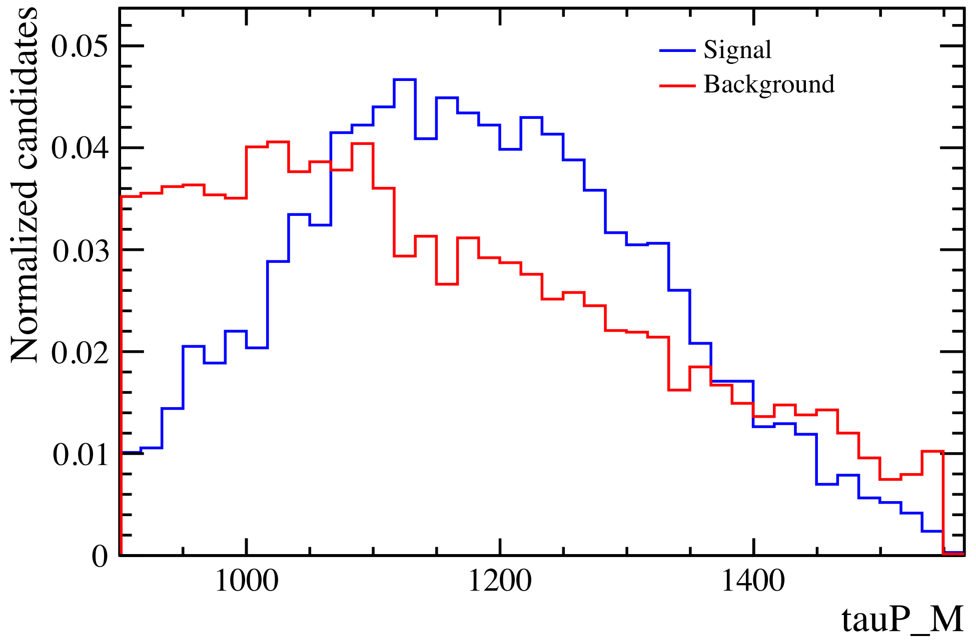
Mixed hadronic-leptonic: MVA selection and top-ranking variables

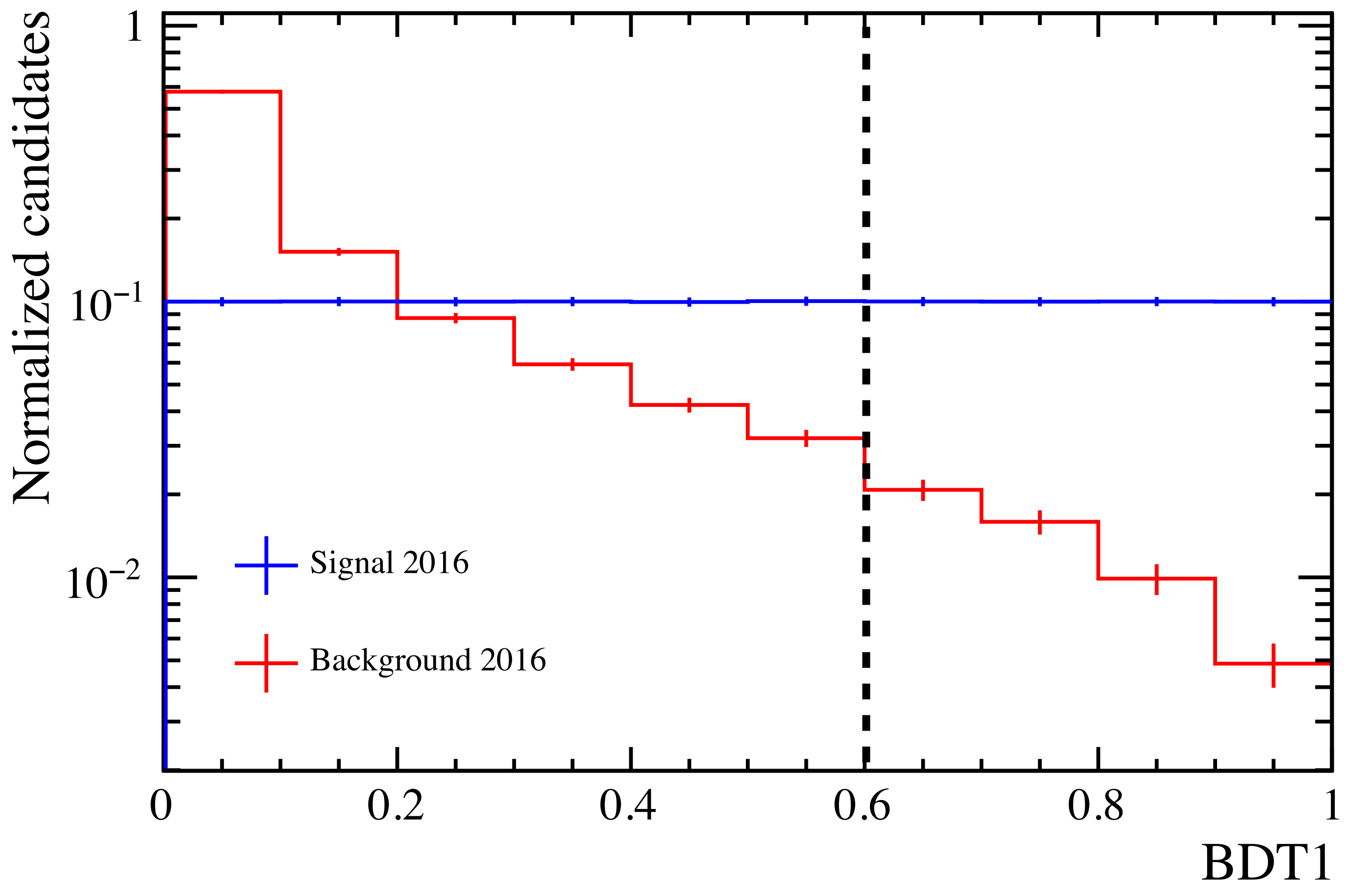
BDT1 (2016 example)











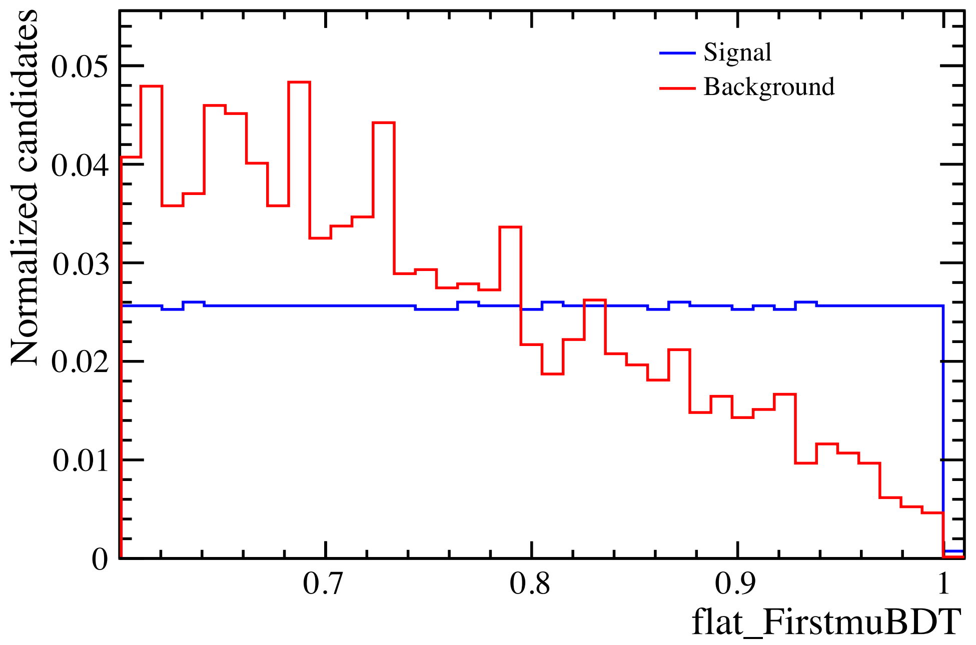
BDT1 > 0.6

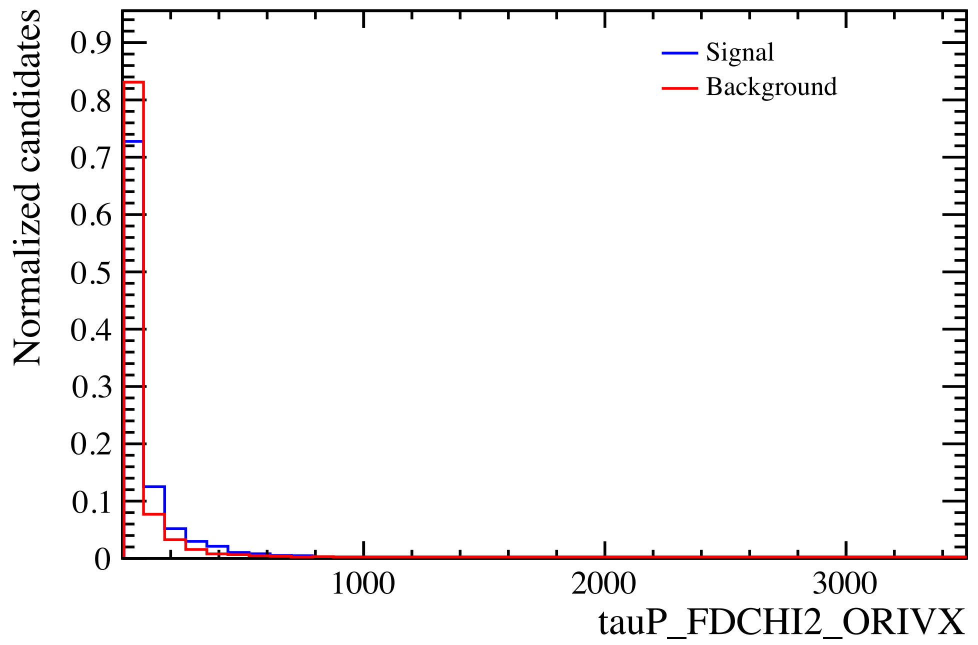
On training samples:

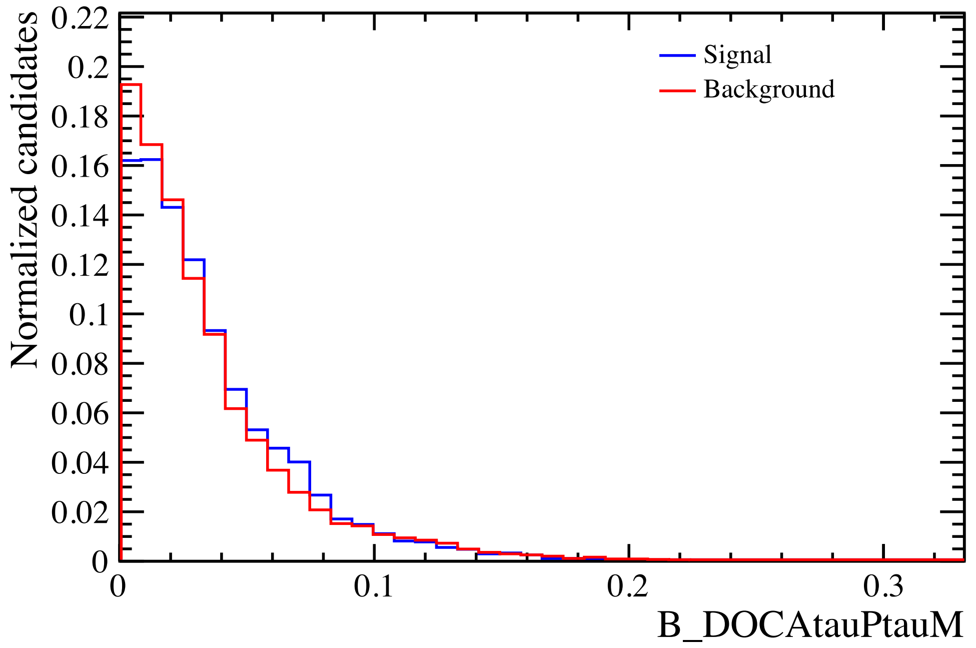
$\epsilon_{sig} \sim 40\,\%$

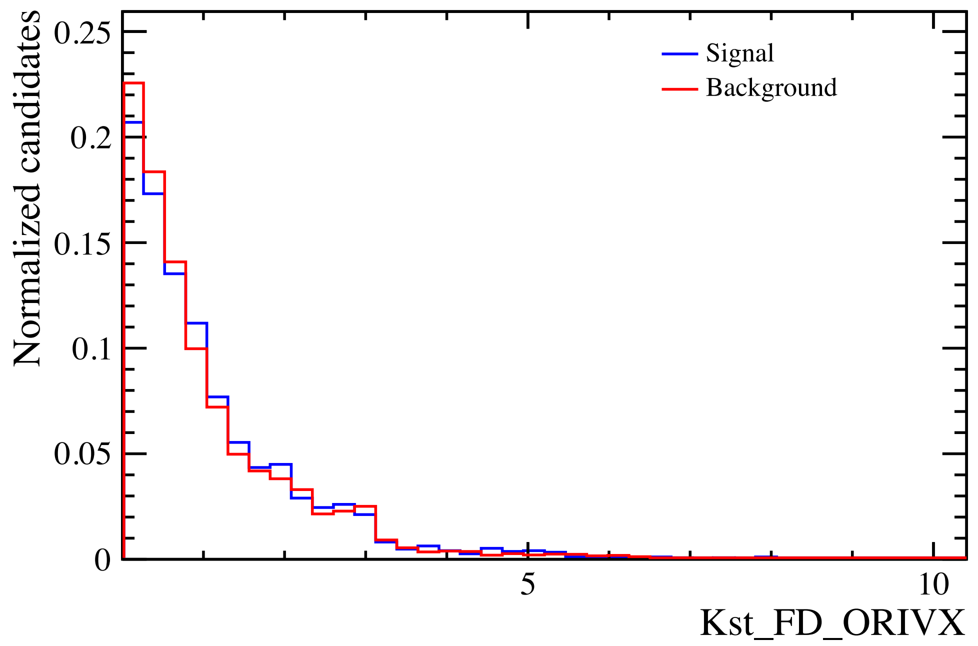
$\epsilon_{bkg} \sim 7\,\%$

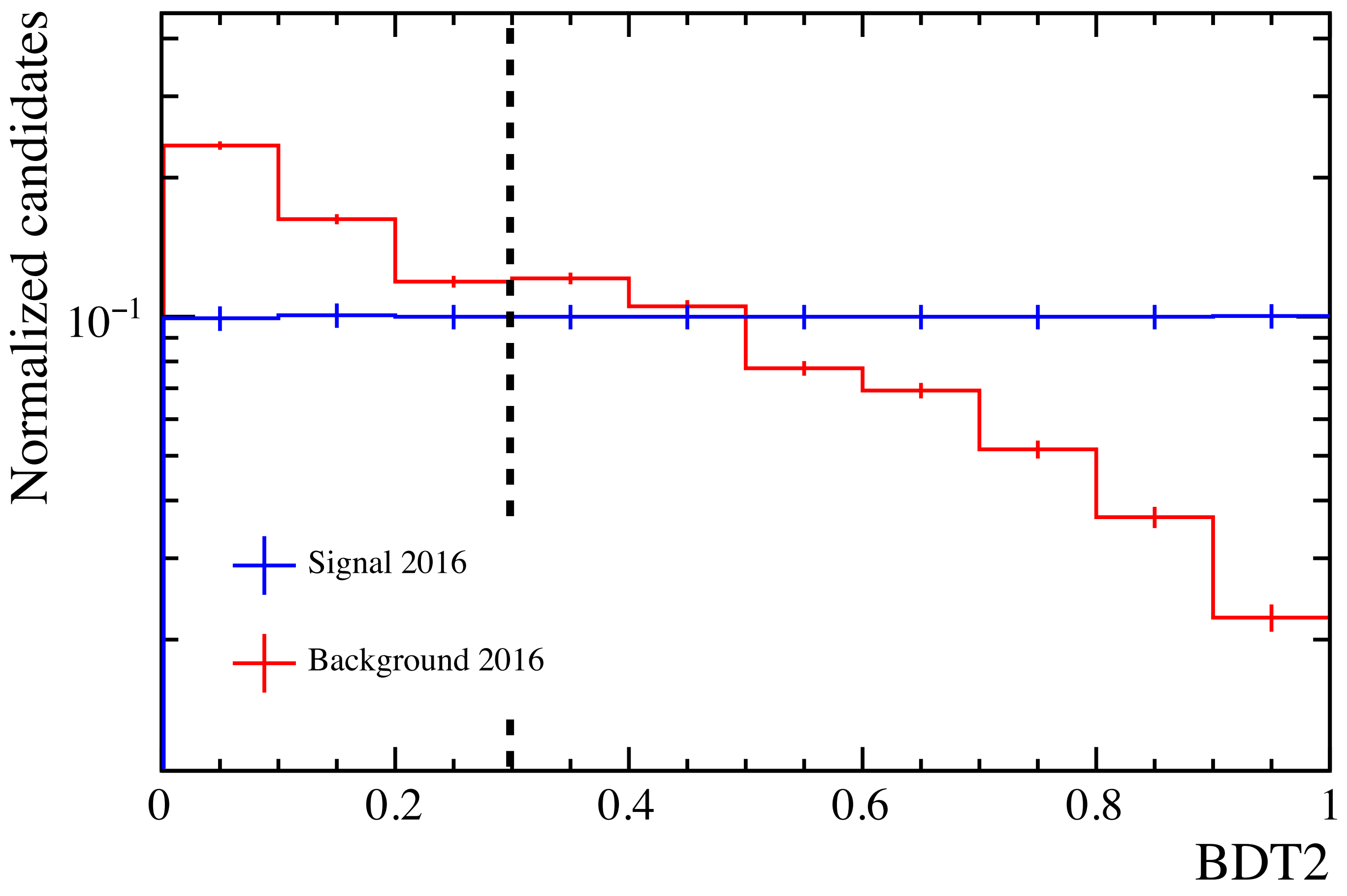
BDT2 (2016 example)











BDT2 > 0.3

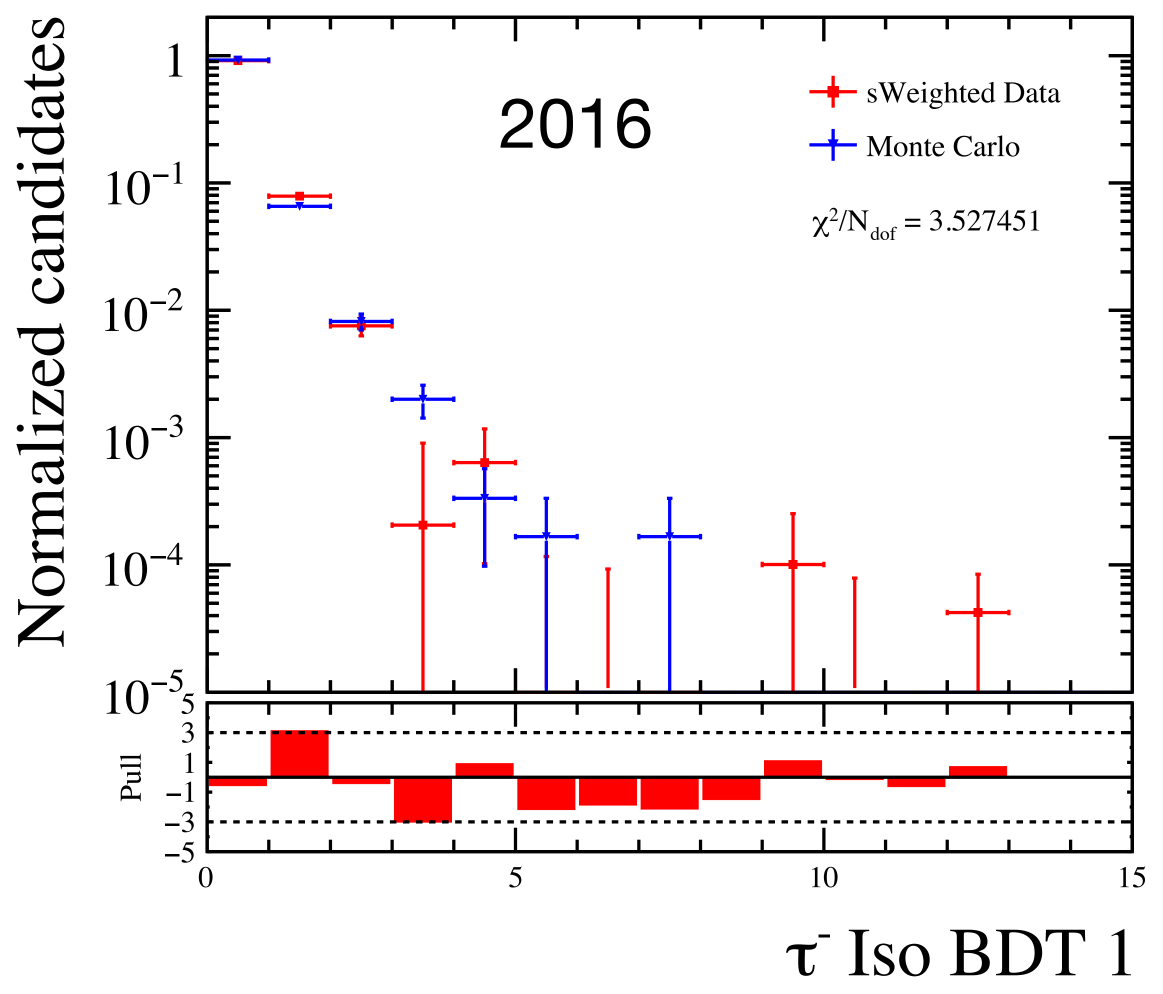
On training samples:

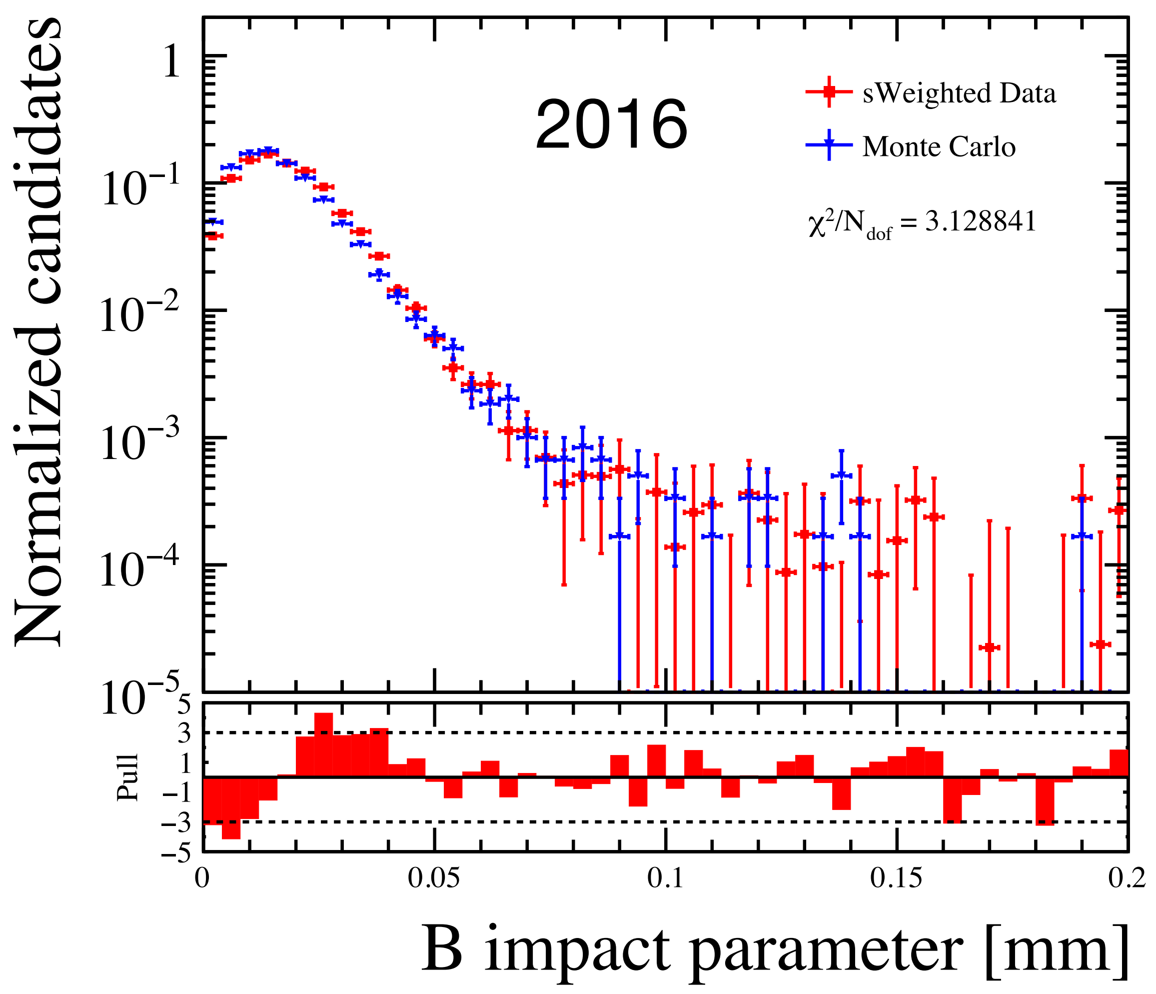
$\epsilon_{sig} \sim 70\,\%$

$\epsilon_{bkg} \sim 50\,\%$

Efficiency corrections: data-MC differences (1/2)

- **Agreement between data and simulation is checked** using the normalization channel





$$\chi^2/N_{dof} = \frac{1}{n_{bins}} \sum_{bins} \frac{(a_i^{MC} - a_i^{data})^2}{\delta a_i^{MC2} + \delta a_i^{data2}}$$

$$w_i = \frac{a_i^{data}}{a_i^{MC}} \pm \sqrt{(\frac{\delta a_i^{data}}{a_i^{MC}})^2 + (\delta a_i^{MC} \frac{a_i^{data}}{a_i^{MC2}})^2}$$

- Selection efficiencies corrected for data-MC differences using **iterative procedure**:

1) Weights w_i computed for the variable showing the worst χ^2/N_{dof}

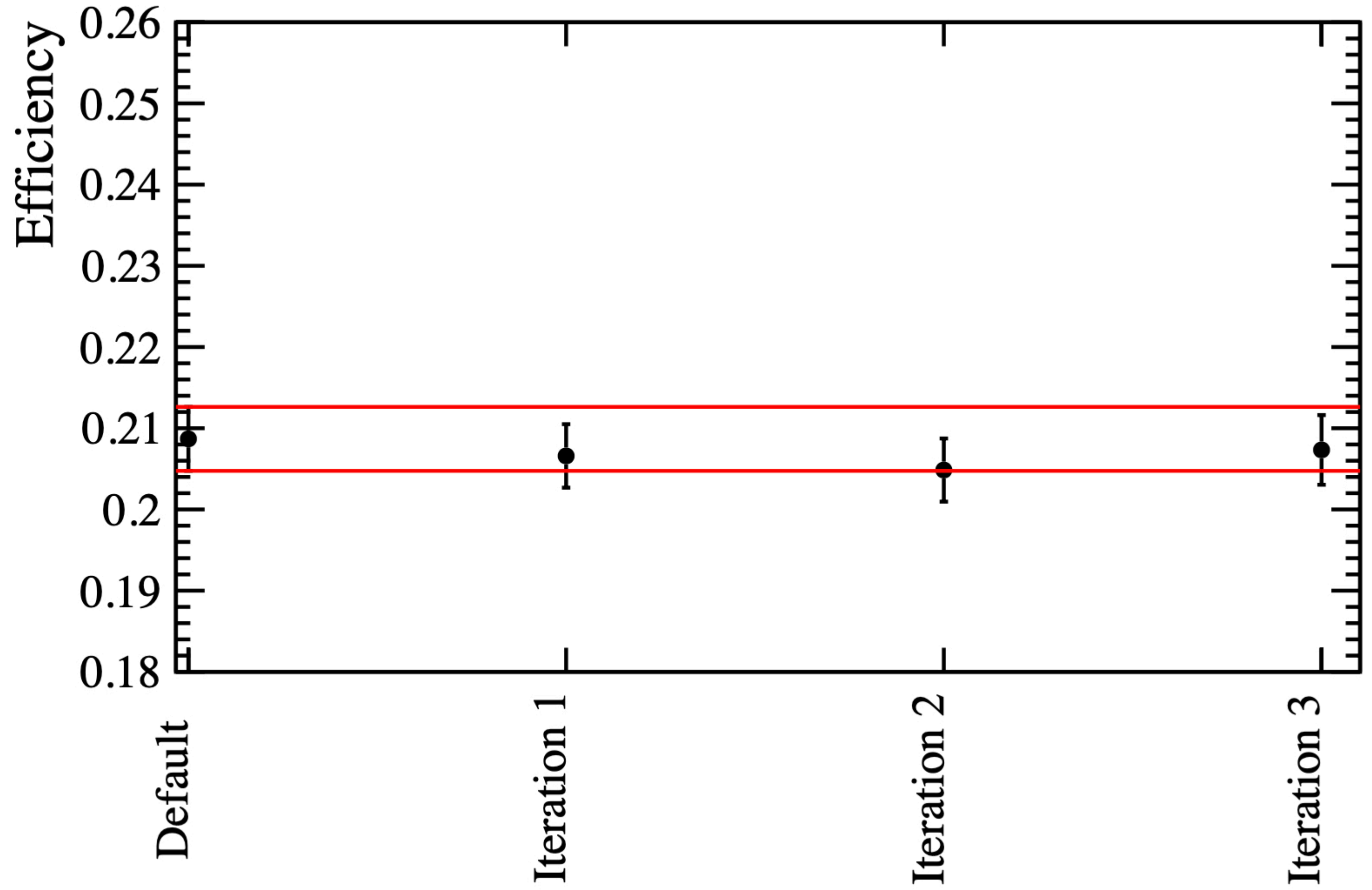
2) Corrected selection efficiency computed as $\epsilon^{corr} = \frac{\sum_i w_i n_i}{\sum_i w_i N_i}$ with n_i (N_i) number of events after BDT2 (trigger)

3) χ^2/N_{dof} recomputed on the re-weighted normalization channel

4) Steps 1-3 iterated. Weights multiplied at each step

Efficiency corrections: data-MC differences (2/2)

- **Three iterations** are performed, example for 2016 fully hadronic final state:



- The **relative difference between third iteration and default value** used to **assign systematic uncertainty**
- The obtained systematic uncertainty is of $O(1\%)$

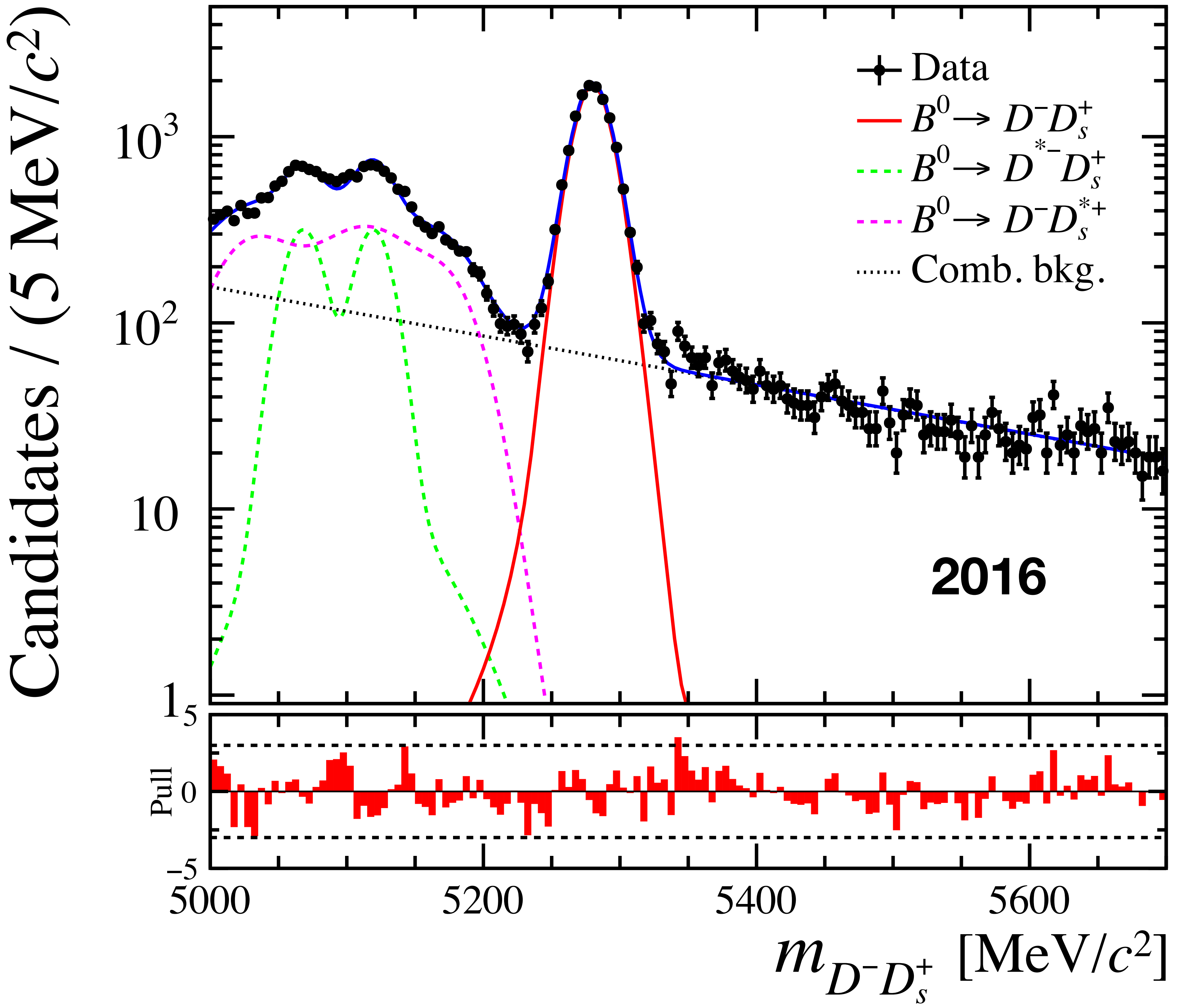
- Applied data-driven corrections also on PID ($O(1) \pm 0.1$ %) and trigger ($O(1) \pm 0.5$ %) efficiencies
- Final efficiency ~10^{-5} (~4×10^{-5}) for hadronic (mixed hadronic-leptonic) final state

Normalization channel yield

- $B^0 \to D_s^+ (\to K^+K^-\pi^+) D^- (\to \pi^-\pi^-K^+)$ selected with $m_D \in [1855,1885]$ MeV and $m_{D_s} \in [1955,1985]$
- Selection efficiency ~ $6 \cdot 10^{-4}$
- Backgrounds are combinatorial and partially reconstructed $B^0 \to D^{*-}D_s^+$, $B^0 \to D^- D_s^{*+}$

Channel	Branching ratio (%)
$B^0 \to D_s^+ D^-$	0.72 ± 0.08
$D^+ \to \pi^+\pi^+K^-$	9.38 ± 0.16
$D_s^+ \to \pi^+K^+K^-$	5.39 ± 0.15
Total	0.0036 ± 0.0004

Year	Normalization yield
2011	3274 ± 64
2012	7358 ± 95
2015	1926 ± 48
2016	11907 ± 119
2017	13170 ± 1245
2018	15438 ± 135

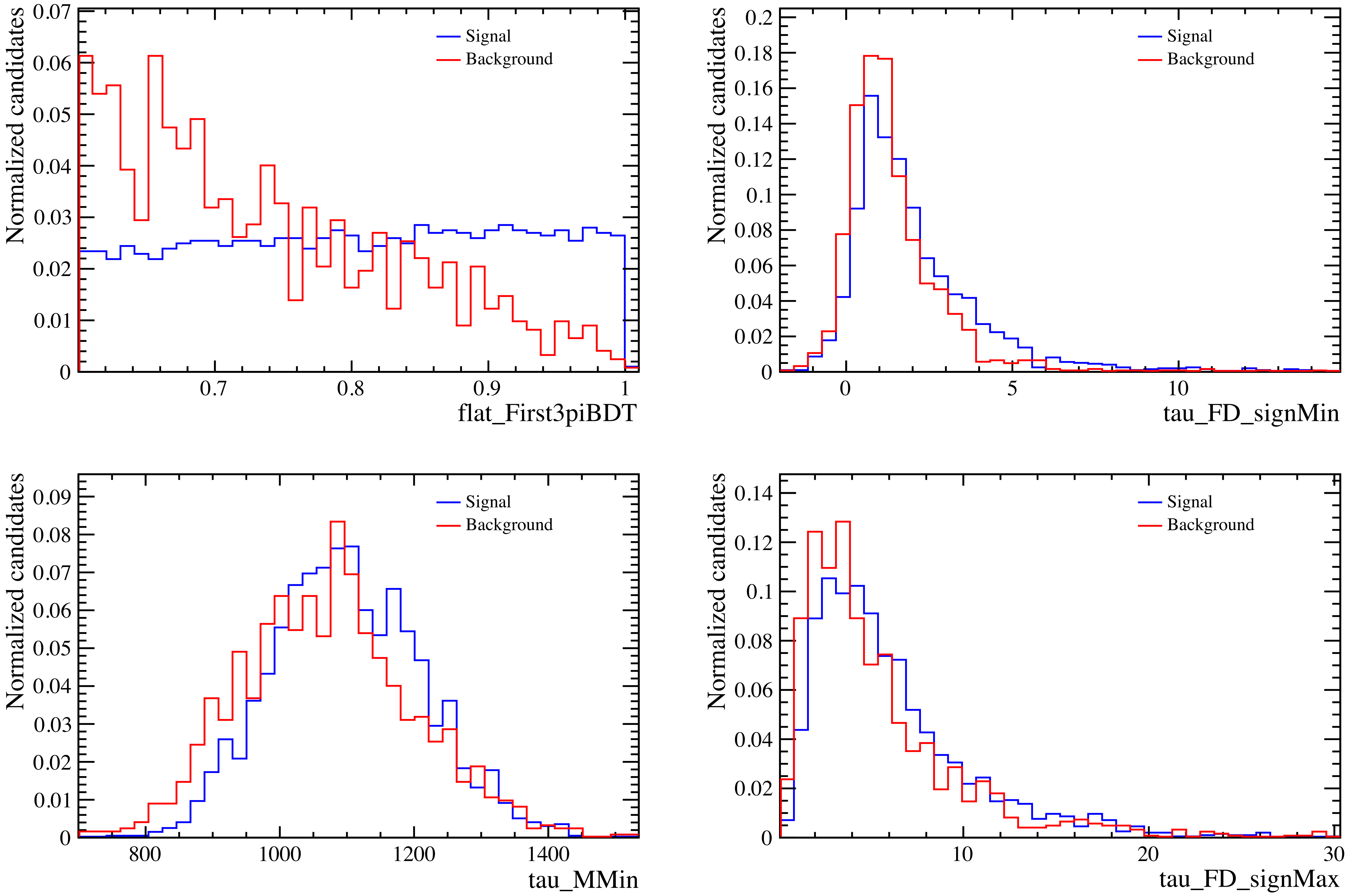


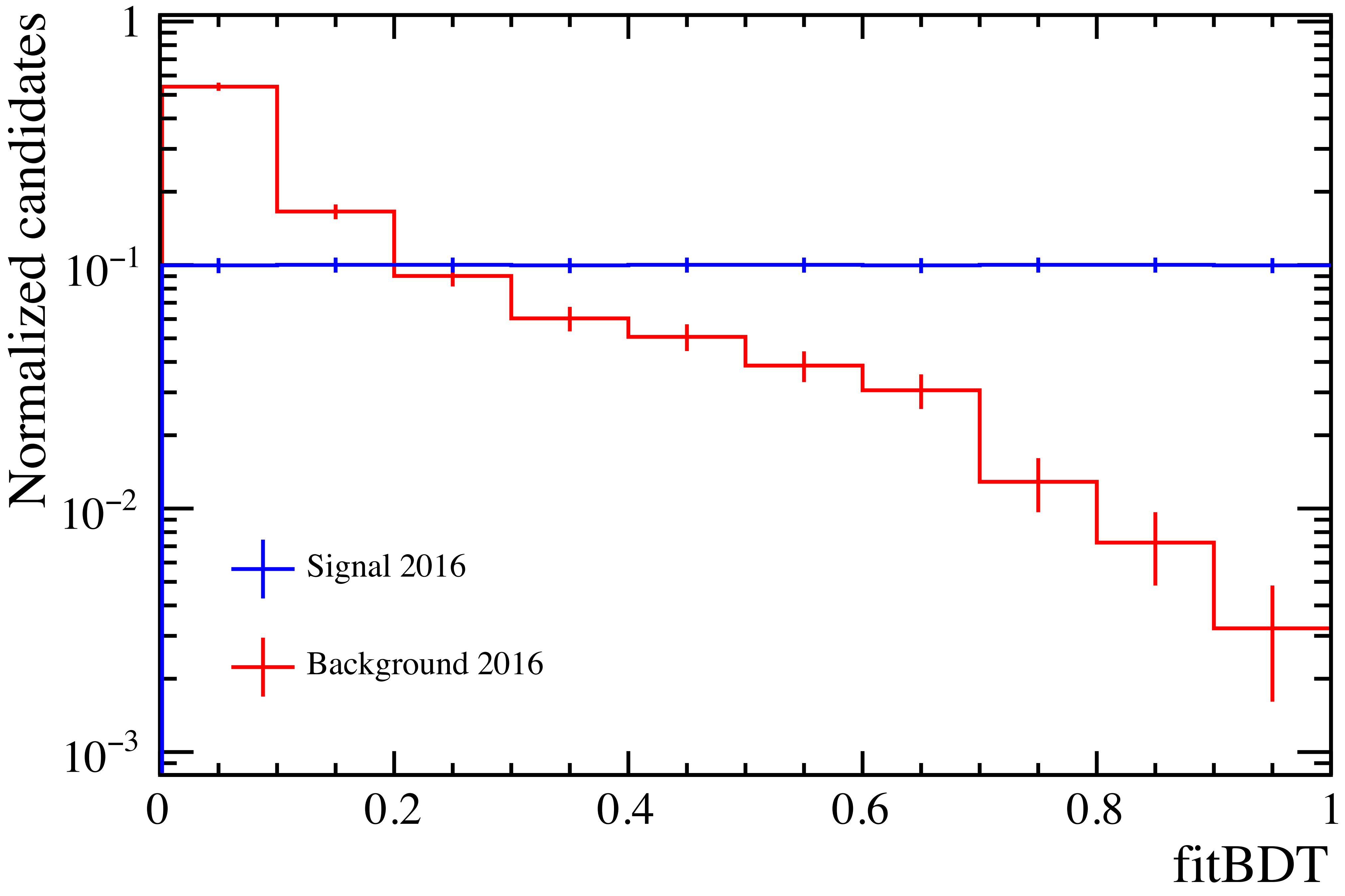
- Normalization yield and efficiency computed, as well as signal efficiency. **Signal yield is the only missing ingredient** to compute the final result

Likelihood fit and sensitivity

Fully-hadronic: fit BDT and top-ranking variables

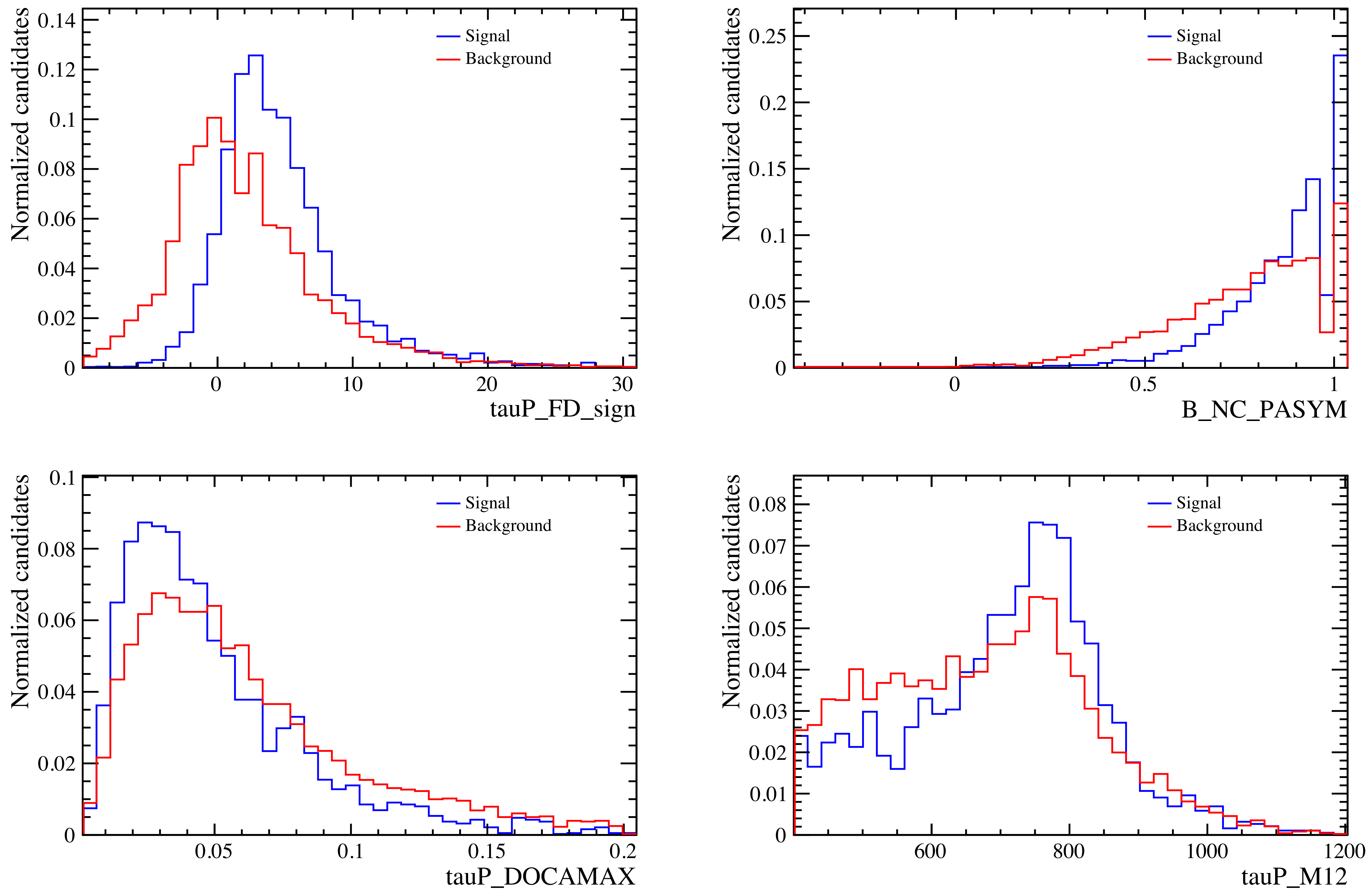
2016 example

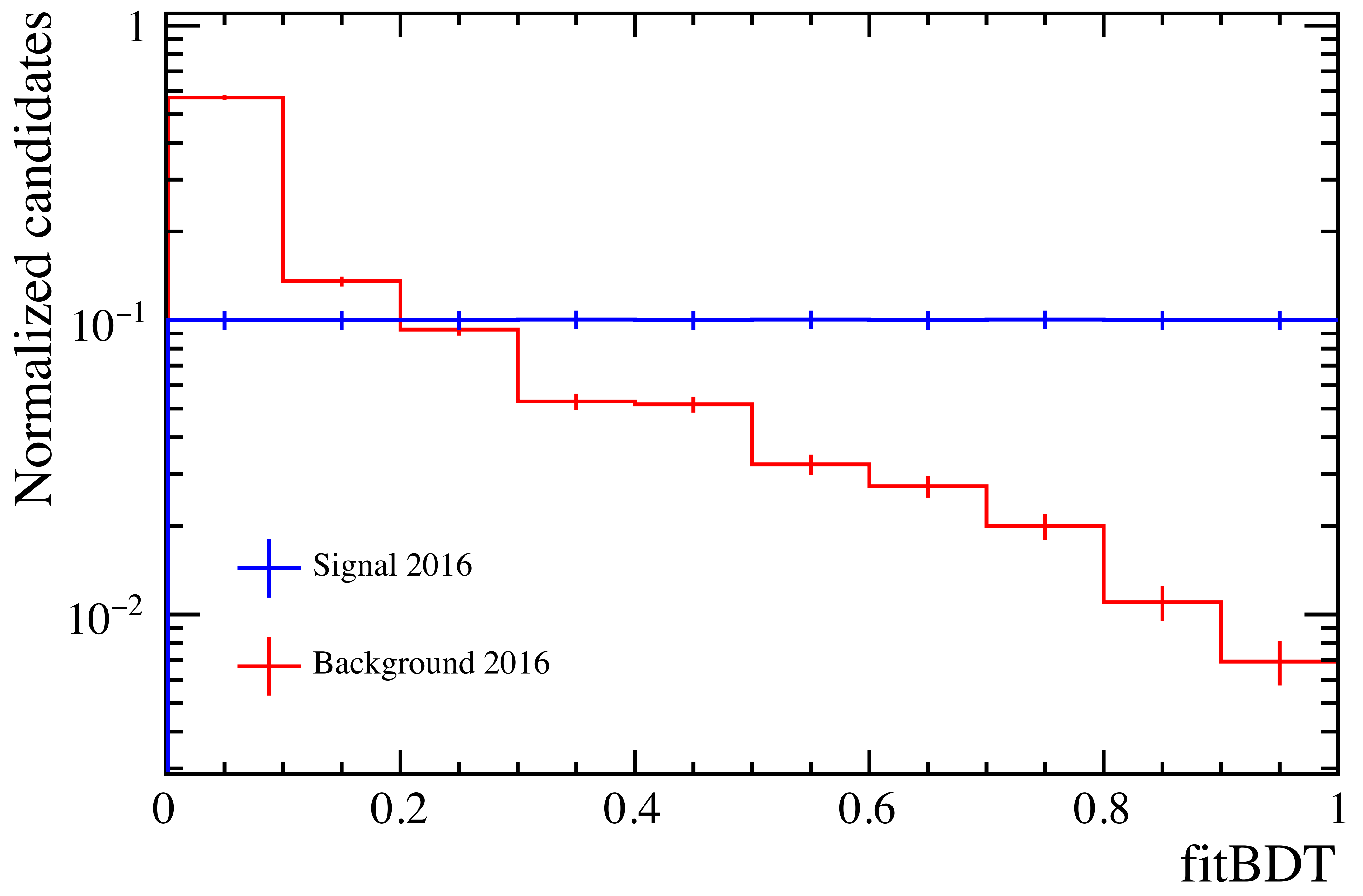




Mixed hadronic-leptonic: fit BDT and top-ranking variables

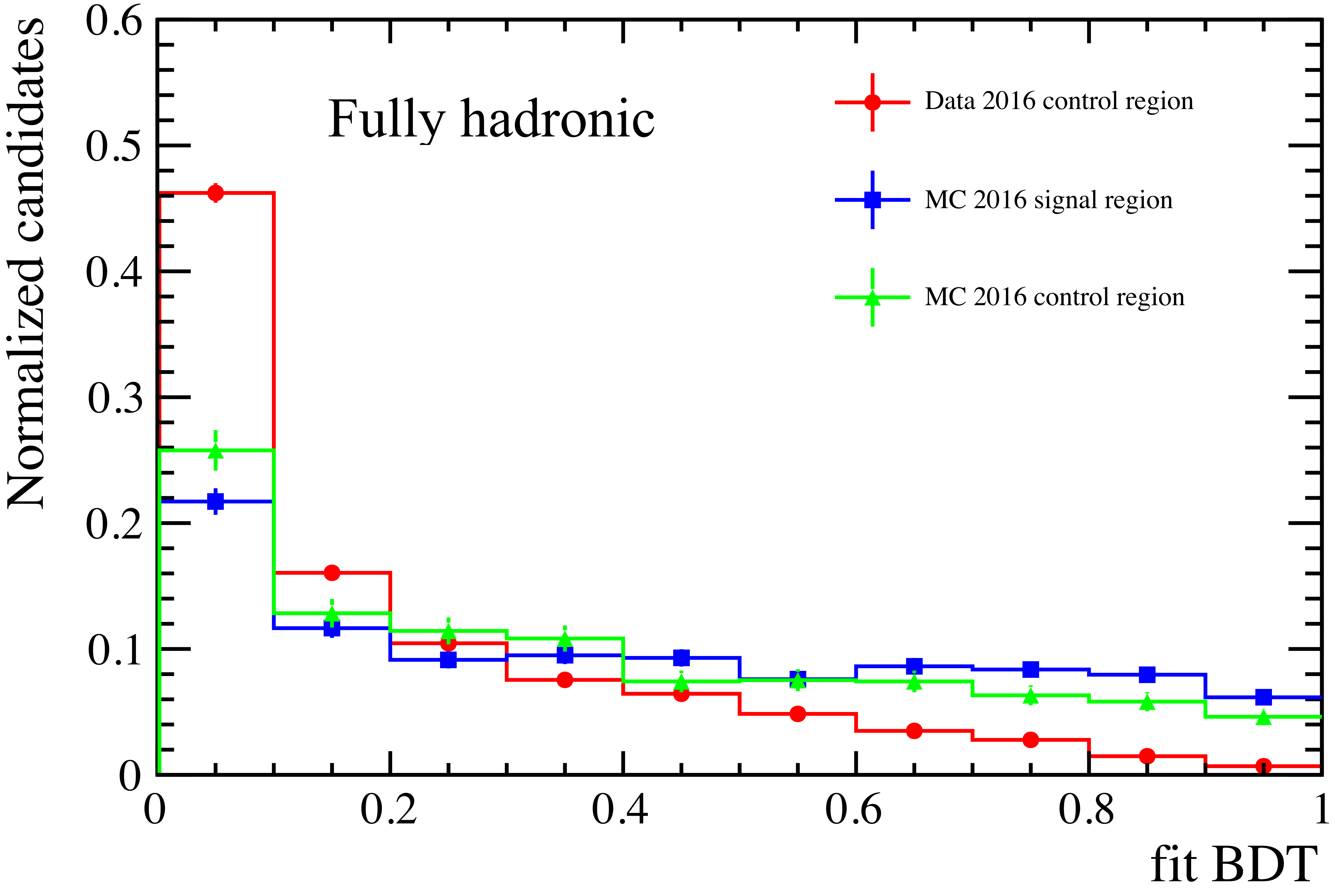
2016 example

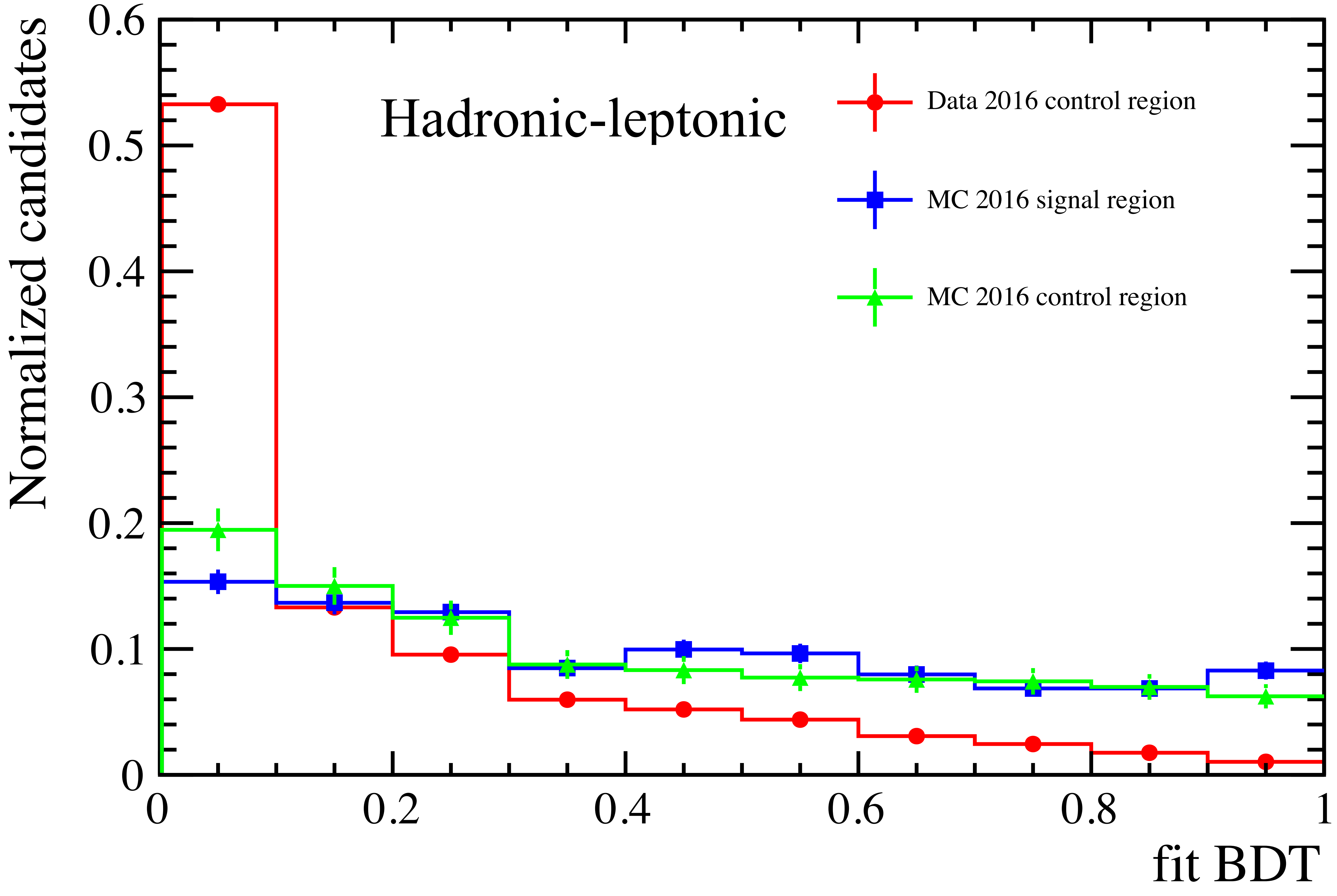




Fit components

$$\textbf{Data}_i = f_i\,\mu \cdot \textbf{Sig}_i + \frac{s_i}{c_i} n_i^{ctl} \cdot \textbf{Ctl}_i - f_i\,\mu \frac{\epsilon_i^{ctl}}{\epsilon_i^{sig}} \frac{s_i}{c_i} \cdot \textbf{Cont}_i$$

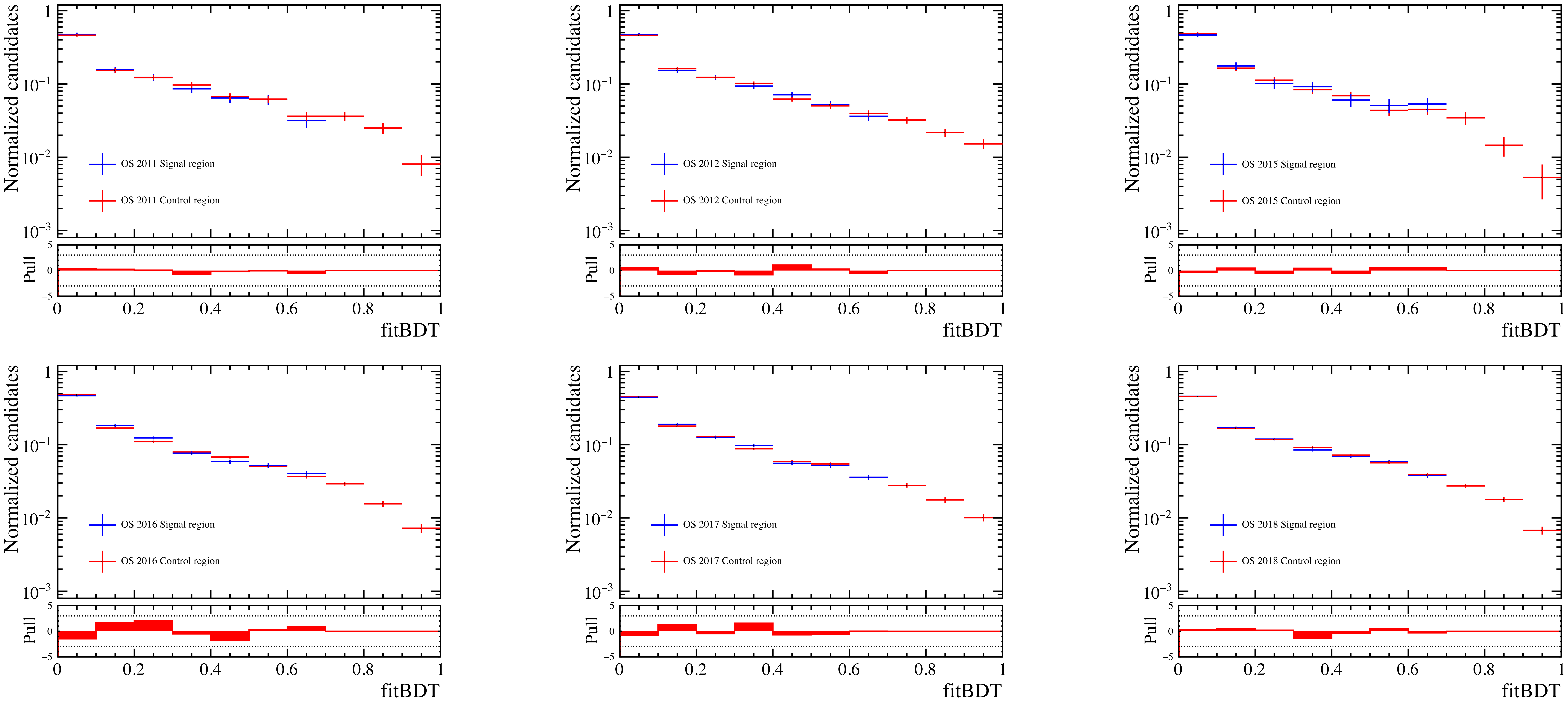




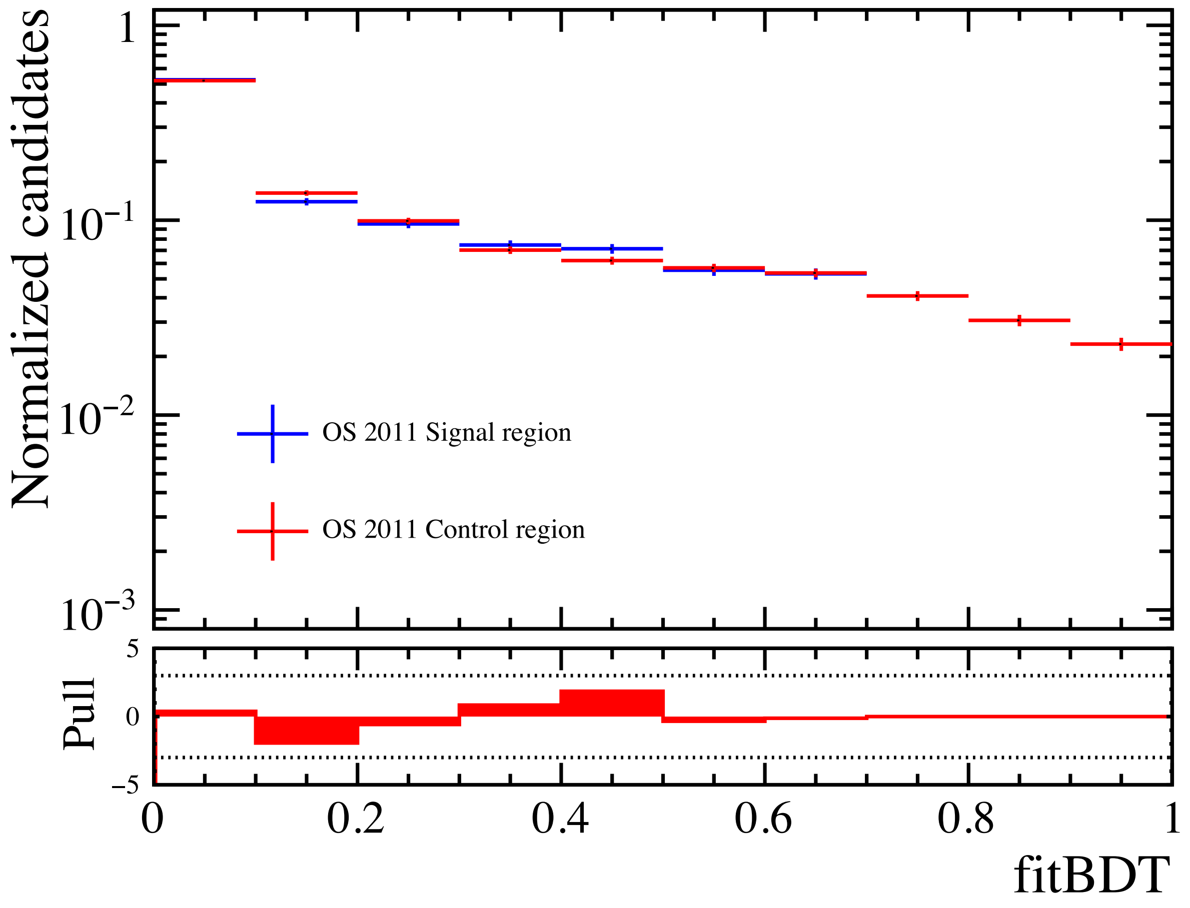
- MC distributions not totally flat due to the presence of neutral pion component
- **Bins close to fit BDT = 1 are the most sensitive** to the signal

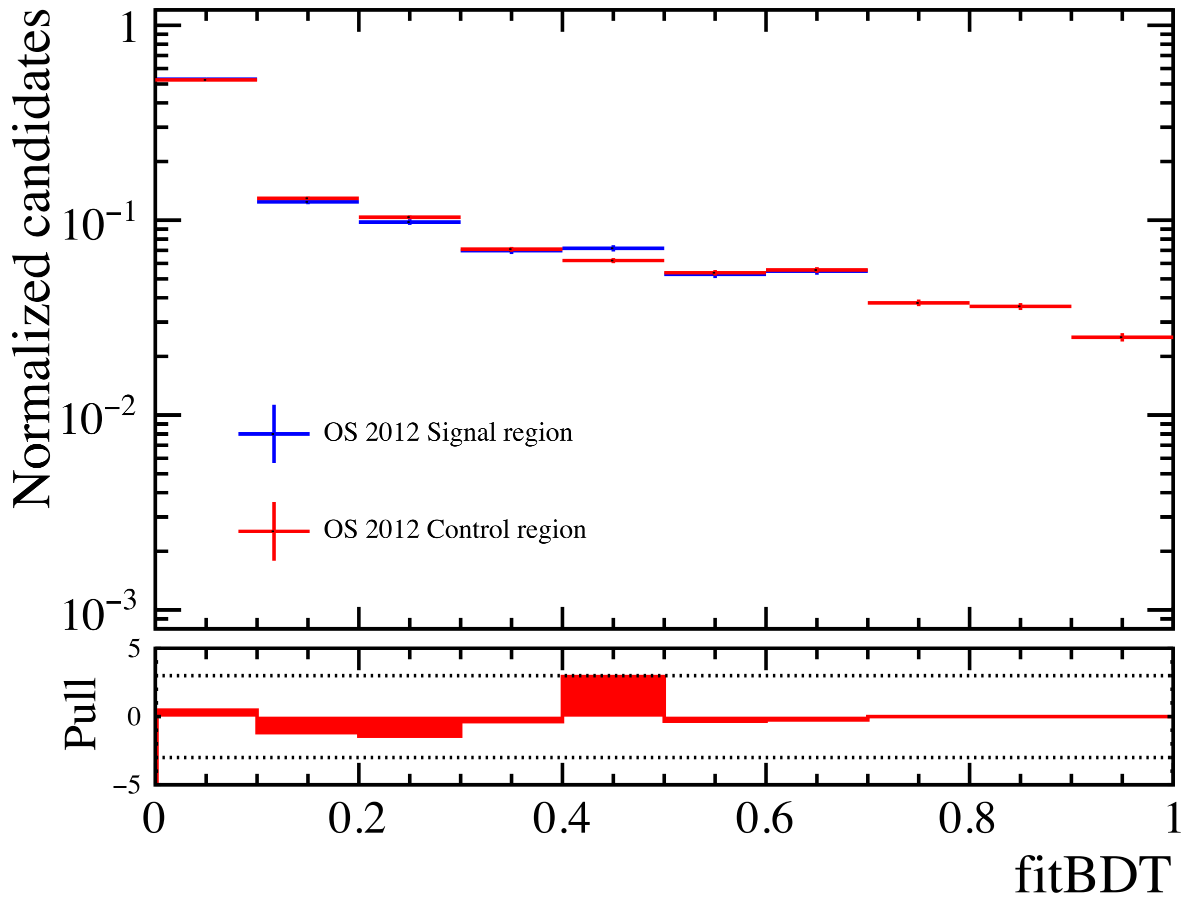
Background model validation: fully hadronic final state

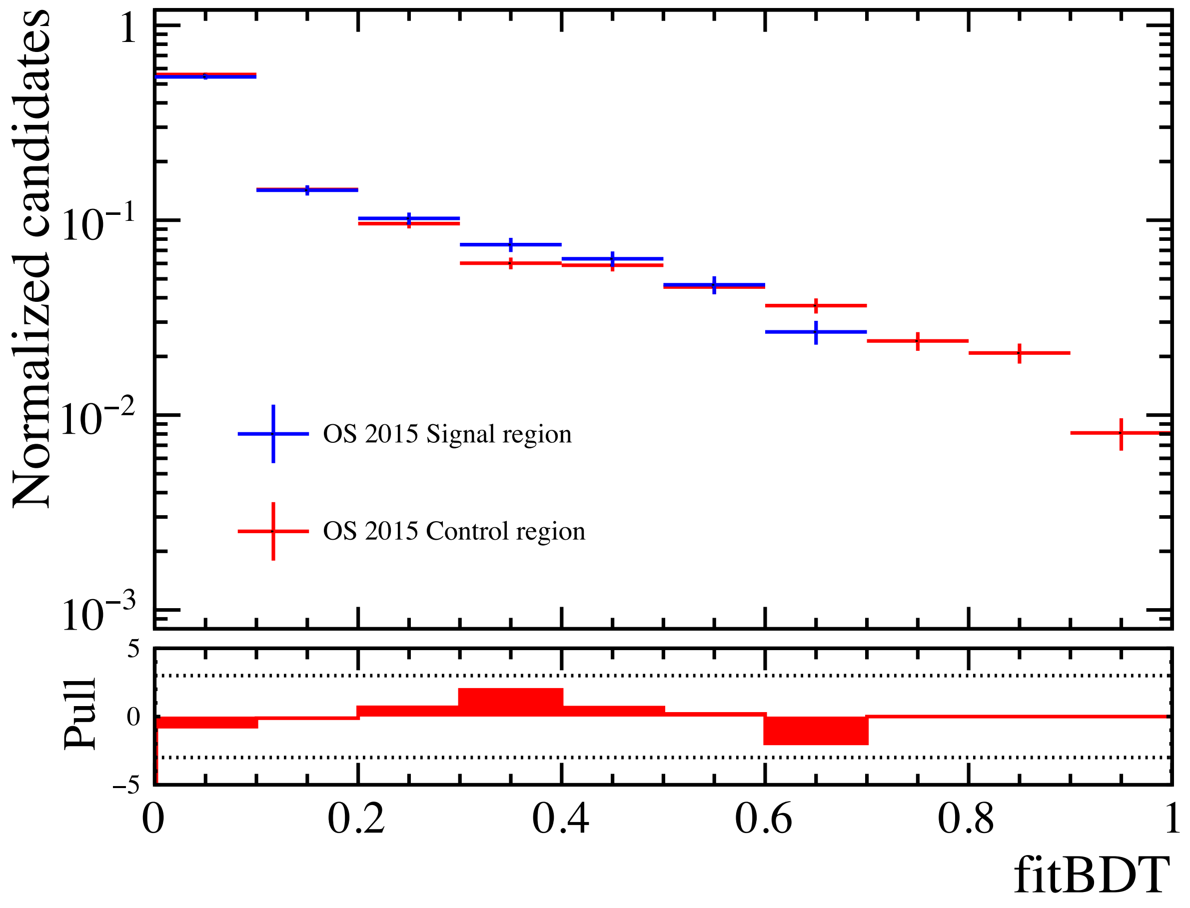
- Background model from **control region validated** by comparing it with distribution in signal region
- Last three bins of signal region blind

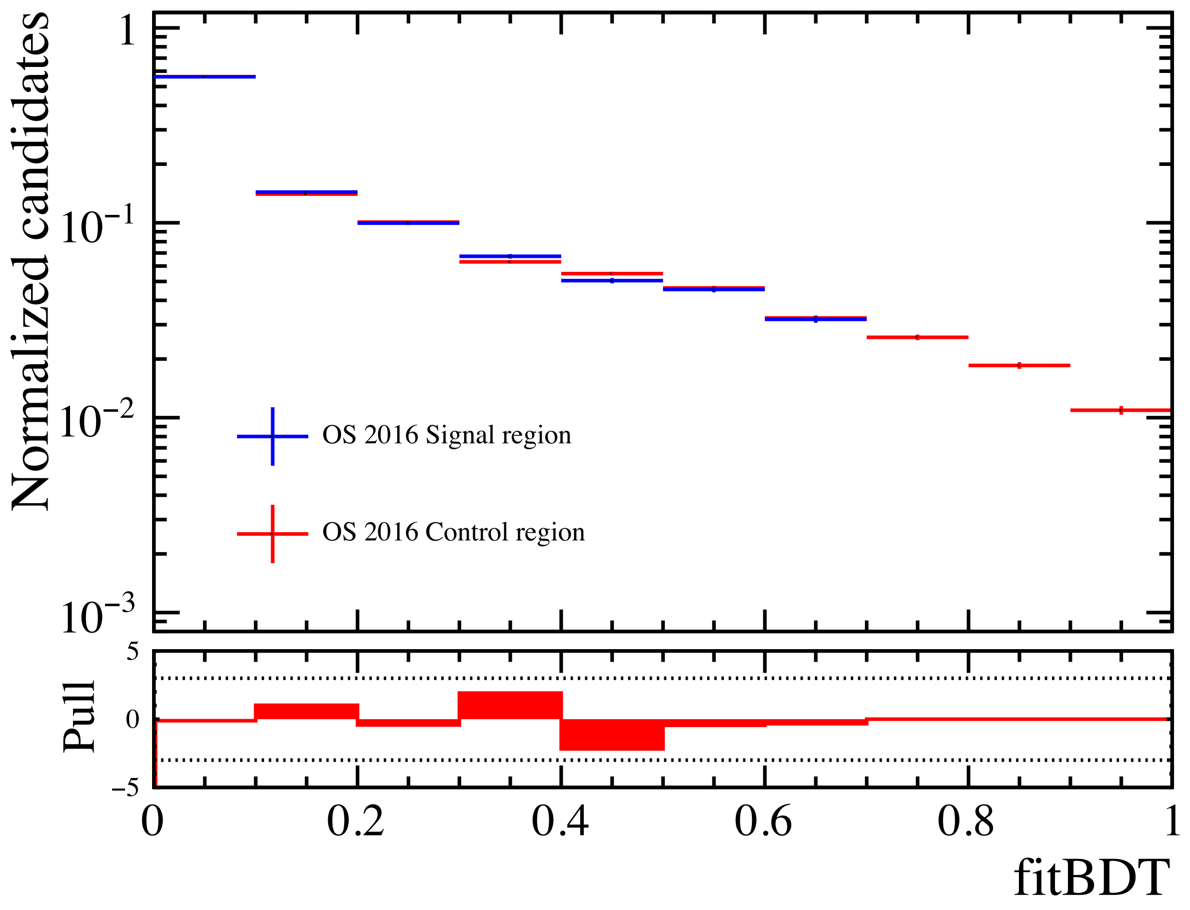


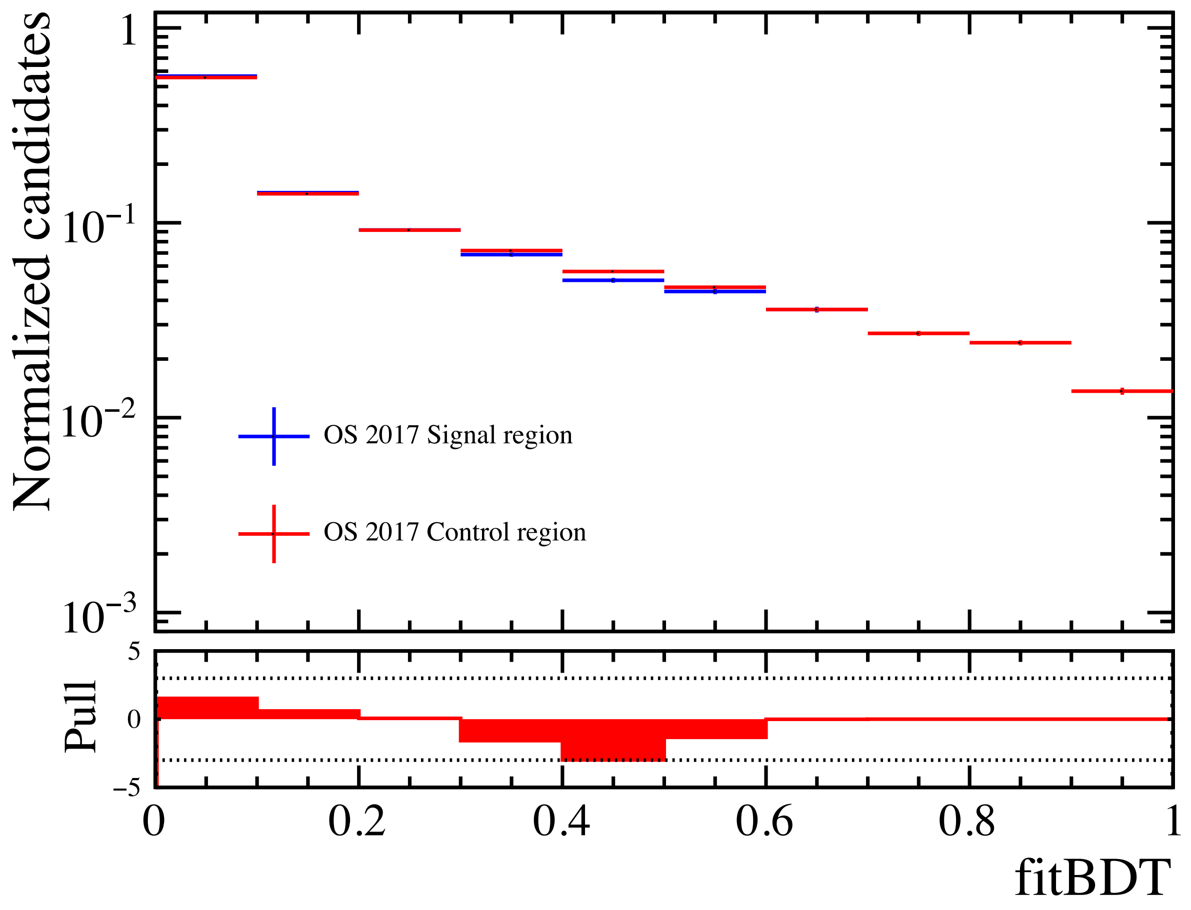
Background model validation: hadronic-leptonic final state

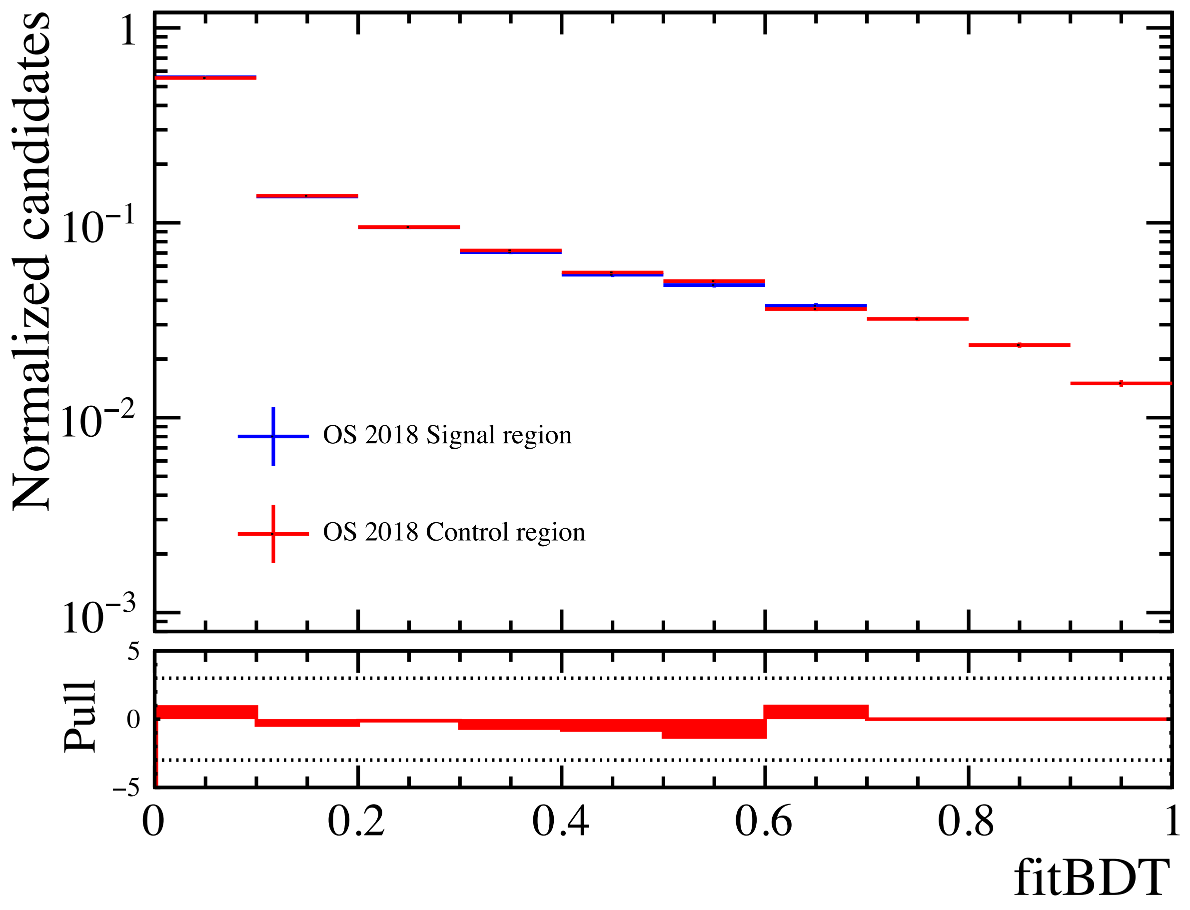






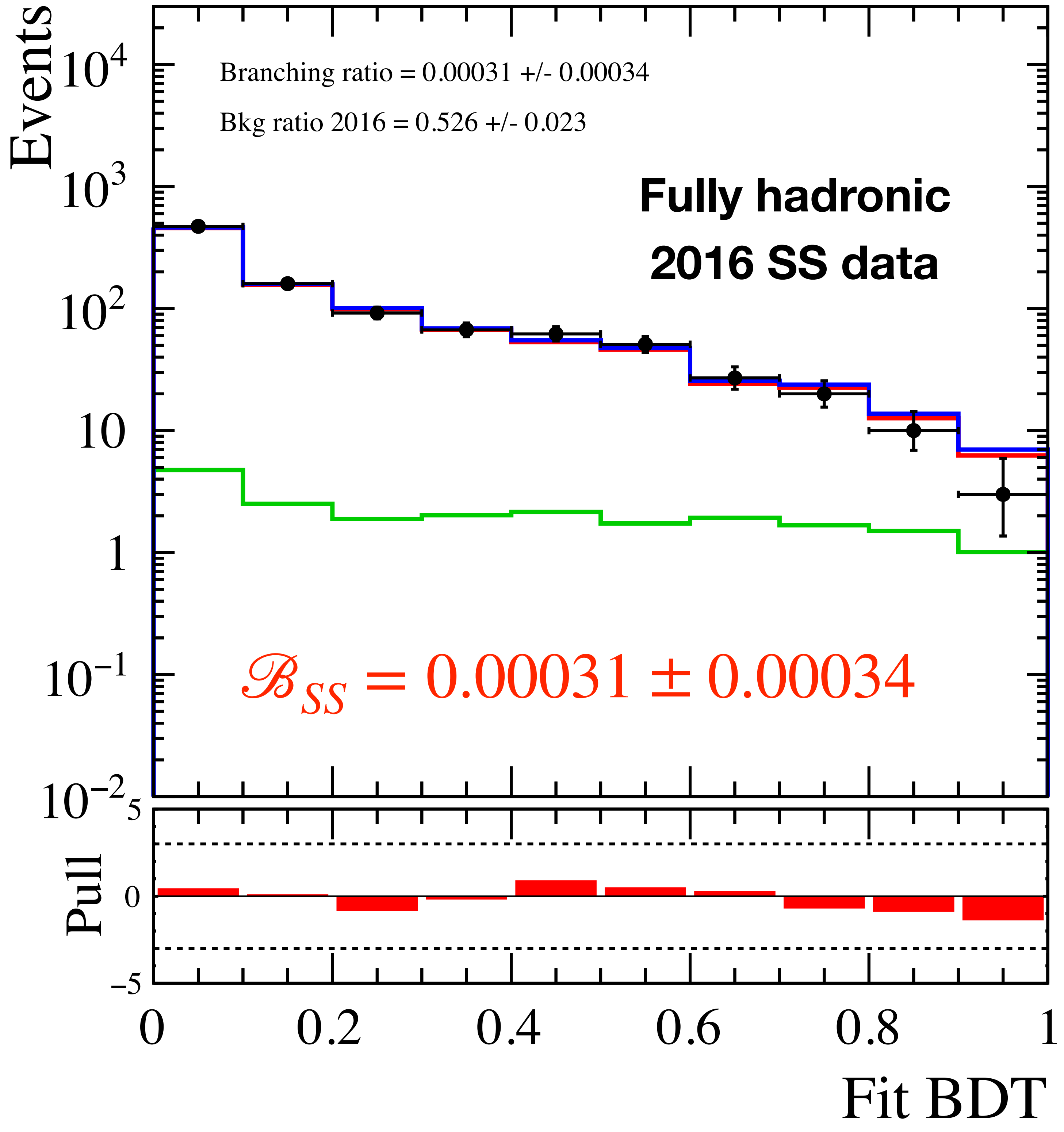


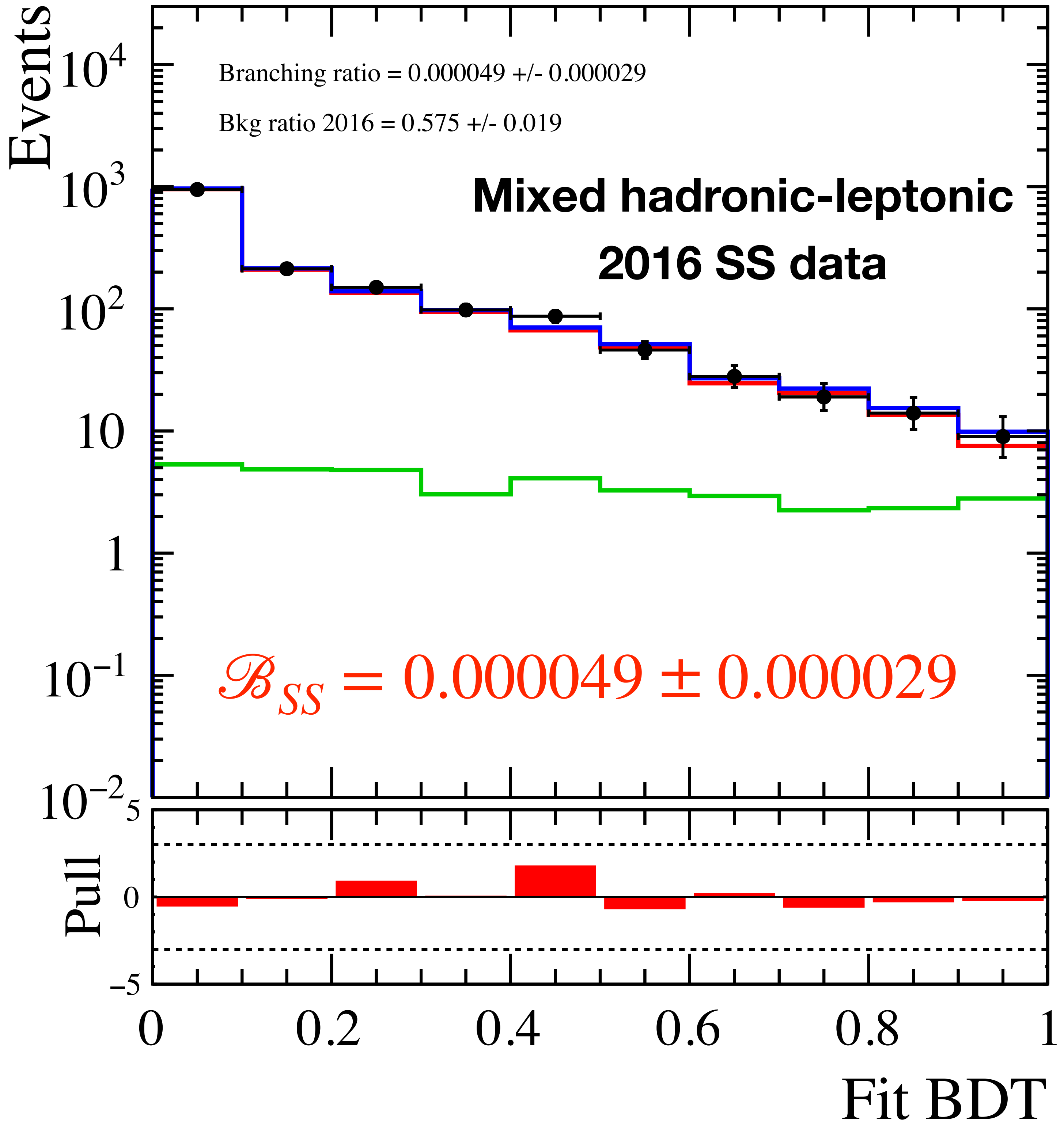




Cross-check on same-sign data

- Same-sign (SS) data are selected requiring **both tau leptons to have the same charge**
  - No signal is present in this dataset, **no need to blind the data**
  - **Fit result must be compatible with 0**





Expected upper limit

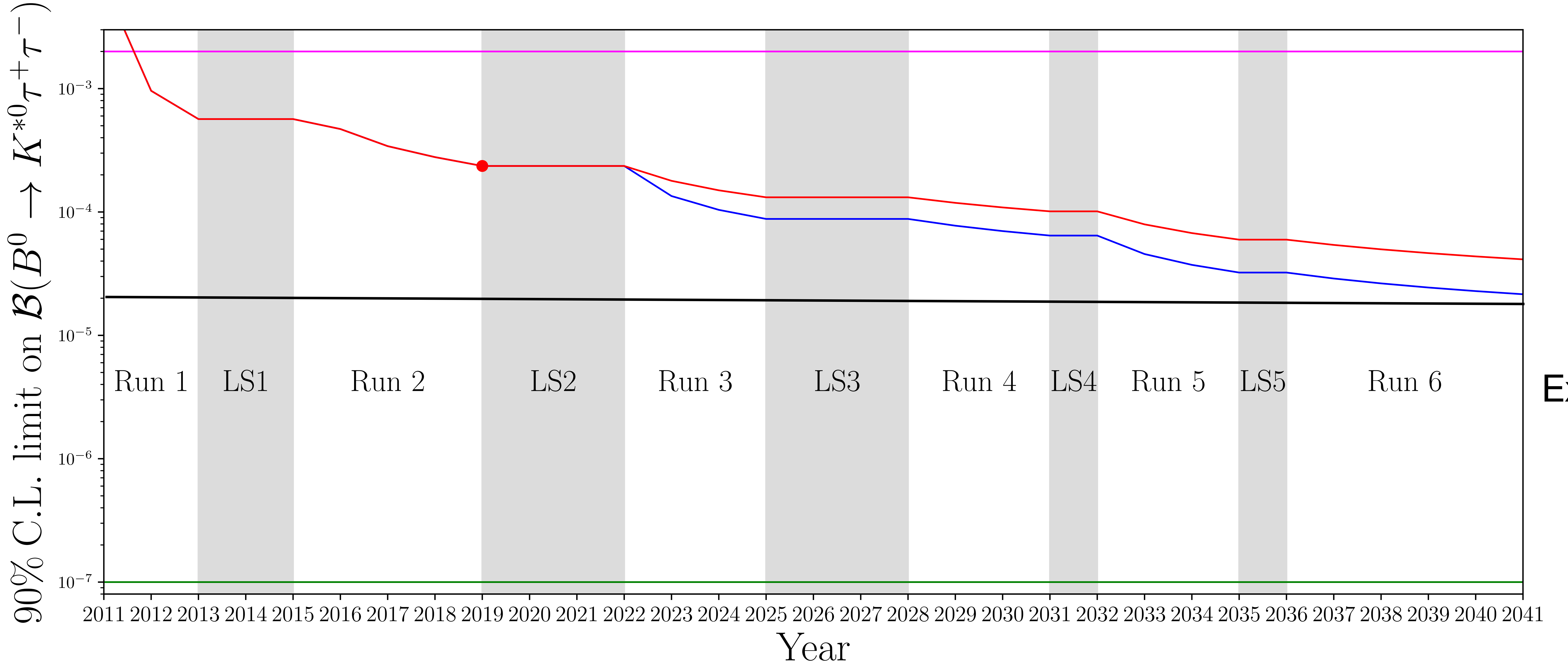
- **Data is still blind**
- **Expected upper limit** in the case where no signal is observed has been computed
- Systematic uncertainties taken into account:
  - Uncertainties on **input branching ratios**, **normalization yield**, normalization and signal (corrected) **efficiencies** (~ 1% limit increase)
  - Bin-by-bin statistical **fluctuations of signal and background templates** (~ 25 % limit increase)
  - Effect of **data-MC differences on the signal template** (negligible)

Final state	90% CL limit (10^{-4})	95% CL limit (10^{-4})
Fully hadronic	12.94	15.52
Mixed hadronic-leptonic	2.39	2.88
Combined	**2.36**	**2.82**

- **Sensitivity dominated by mixed hadronic-leptonic final state**
- **One order of magnitude improvement** with respect to current upper limit

Future prospects

- LHCb is undergoing a **major upgrade**, expected ~50 fb^{-1} by 2030 and ~300 fb^{-1} by the end of operations
- Two scenarios are envisaged for the search for the $B^0 \to K^{*0}\tau^+\tau^-$ decay:
  - Conservative: upper limit scaled by the **expected increase in recorded luminosity**
  - Optimistic: **increased luminosity + reasonable detector and analysis improvements**
- Upper limit **reduced by an order of magnitude with respect to present limit** and expected to be **further reduced by another order of magnitude** by LHCb and Belle 2 in the coming years



Legend (limits at 90% CL)

SM prediction: ~10^{-7}

Belle upper limit: 0.002

Expected upper limit (my thesis): 0.00024

Expected Belle 2 upper limit (50 ab^{-1}): ~2 x 10^{-5}


That's all Folks!