Optimization of the $B_s/\ B_d$ separation power in the ATLAS $B^0_{(s)} \to \mu^+\mu^-$ branching ratio measurement

Ondrej Kovanda

University of Sussex

ATLAS experiment

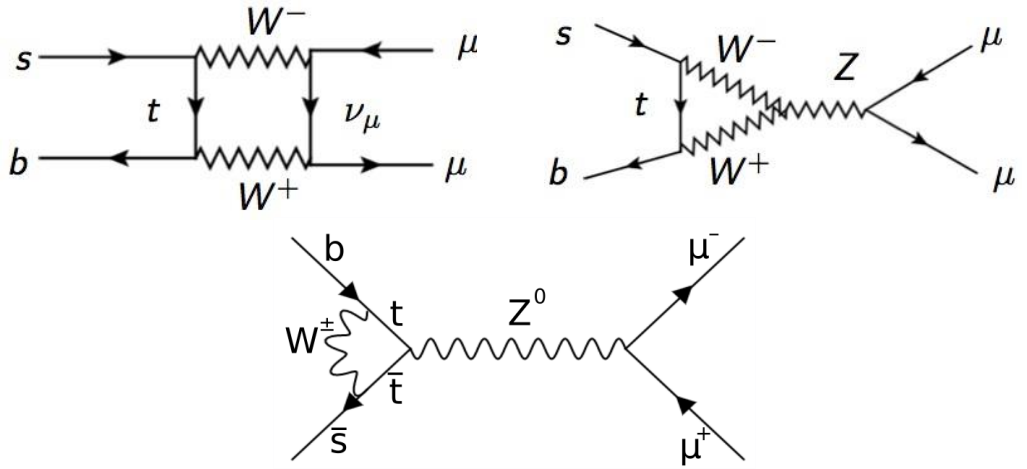
IoP HEPP & APP Annual Conference 2022

Why $B^0_{(s)} \to \mu^+\mu^-$?

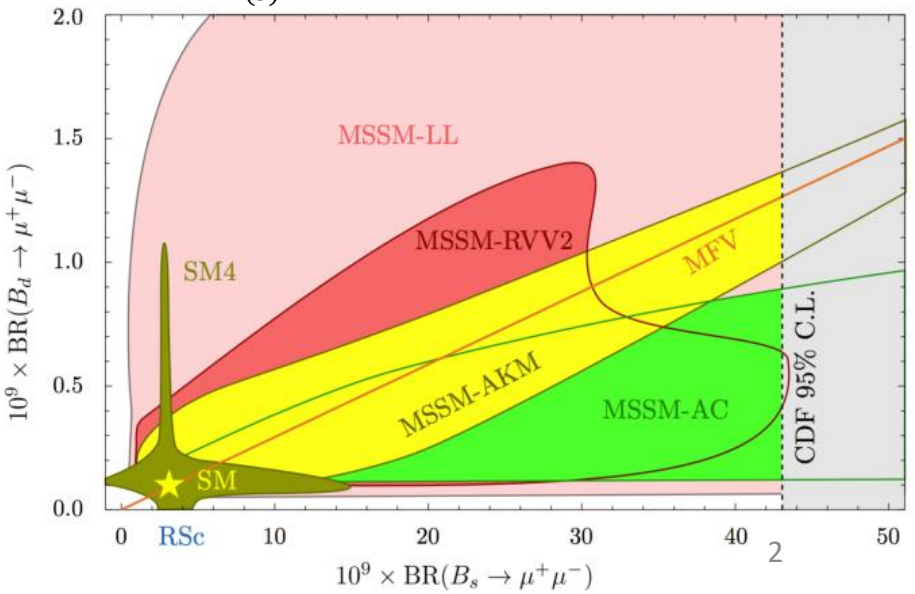
Flavour-changing neutral current process

- Rare in Standard Model
  - No tree level contribution
- Branching ratios precisely predicted in SM [1, 2, 3]
  - B_s : $(3.66 \pm 0.14) \times 10^{-9}$
  - B_d : $(1.03 \pm 0.05) \times 10^{-10}$
- Sensitive to NP
  - Rates altered in various BSM scenarios (LFV, MSSSM, ...)
- Clean observables
  - BR($B^0_{(s)} \to \mu^+\mu^-$), BR(B_d)/BR(B_s)

→ **Significant deviation from SM: hint of NP contribution**



$B(B_{(s)} \to \mu\mu)$ in various BSM scenarios [4]



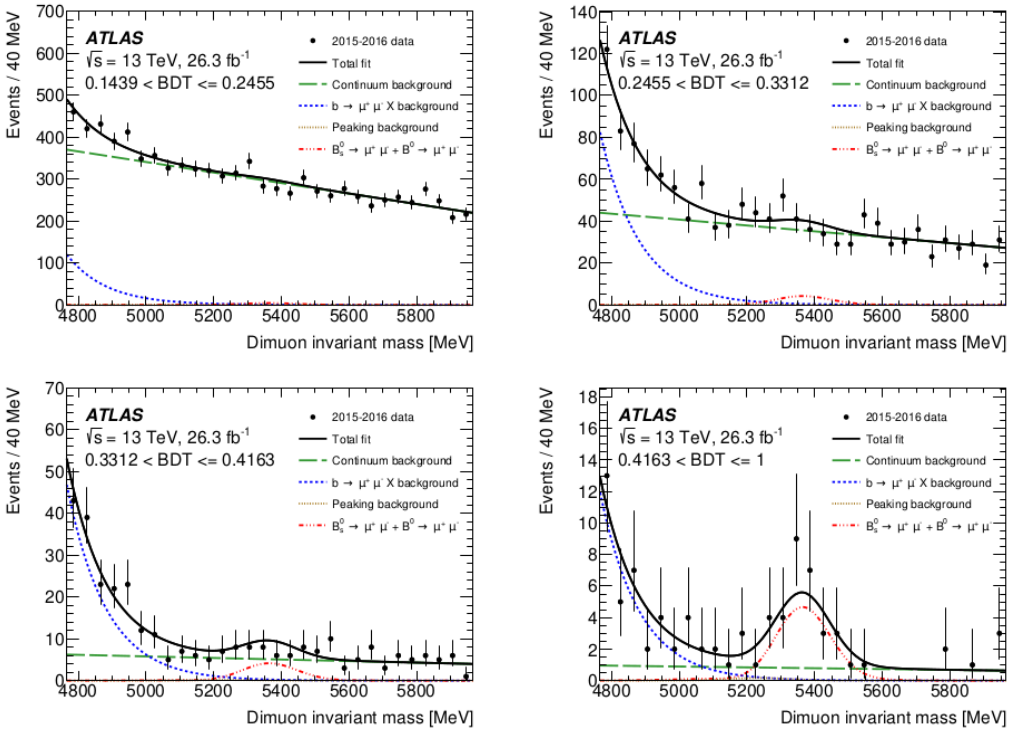
Measuring $B^0_{(s)} \to \mu^+\mu^-$ branching ratio

$$\mathcal{B}(B^0_{(s)} \to \mu^+\mu^-) = \frac{N_{d(s)}}{\varepsilon_{\mu^+\mu^-}} \times \left[\mathcal{B}(B^+ \to J/\psi\, K^+) \times \mathcal{B}(J/\psi \to \mu^+\mu^-)\right] \frac{\varepsilon_{J/\psi K^+}}{N_{J/\psi K^+}} \times \frac{f_u}{f_{d(s)}}$$

What are the main steps of $B^0_{(s)} \to \mu^+\mu^-$ BR measurement?

- Extracting signal yields
  - in fits to dimuon invariant mass distribution
  - Both yields extracted in a single fit
- Cancelling systematics
  - measurement wrt. abundant reference channel
  - In ATLAS we use: $B^+ \to J/\psi\,(\to \mu^+\mu^-)\, K^+$
- Suppressing background
  - MVA-based event selection
  - In ATLAS we use boosted decision tree (BDT)
- Maximizing sensitivity
  - Yield extraction in categories of BDT output with different S/B ratio

Fit to dimuon invariant mass distribution in categories of BDT output [5]



State of the art

- Results SM-consistent & statistics-dominated

	$B(B_s \to \mu^+\mu^-) \times 10^9$	$B(B_d \to \mu^+\mu^-) \times 10^{10}$ (95% CL)
ATLAS (2015/16) [5]	$3.2^{+1.1}_{-1.0}$ $(2.8^{+0.8}_{-0.7})$*	**< 4.3** (< 2.1)*
LHCb (Run 1 + Run 2) [1]	$3.09^{+0.46\,+0.15}_{-0.43\,-0.11}$	**< 2.6**
CMS (Run 1 + 2016) [7]	$2.9 \pm 0.7(exp) \pm 0.2(frag)$	**< 3.6**

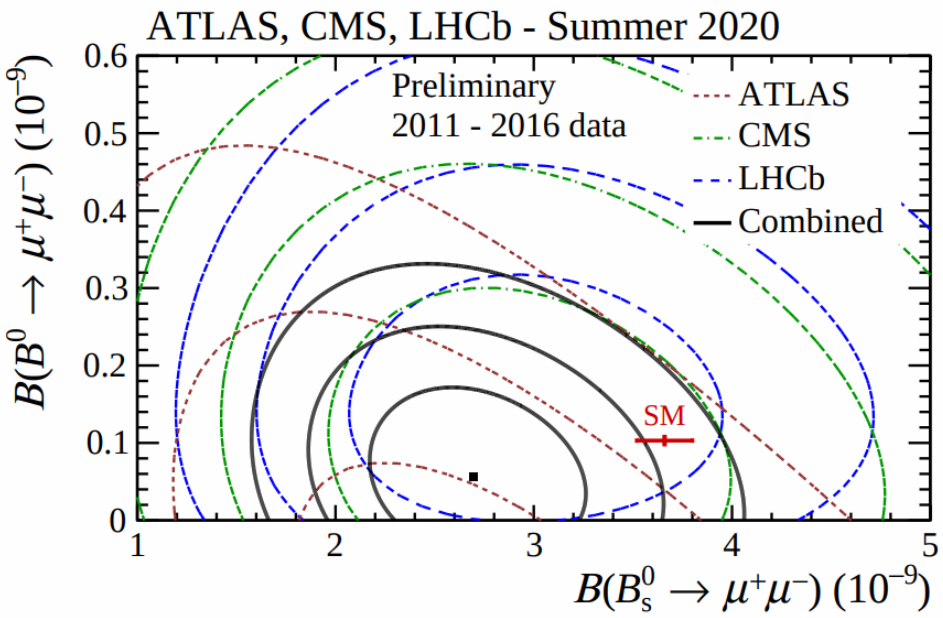
*combination with Run 1 results

Signals very close in mass with B_d yield ~10x smaller than B_s
→ signal yields correlated
→ B_d largely influenced by B_s

- CMS & LHCb:
  - lower BR correlation, better B_d sensitivity
- ATLAS: Run 1 (25 fb^{-1}), 2015/16 (26.3 fb^{-1}) [5]
  - Large $O(1)$ negative correlation of the BRs
  - Currently: Full-Run2 dataset
    - ~4x increase in statistics wrt. 2015/16

Can we minimize B_s / B_d correlation, thus increasing sensitivity to B_d?

Results of previous ATLAS, CMS & LHCb analyses [6]

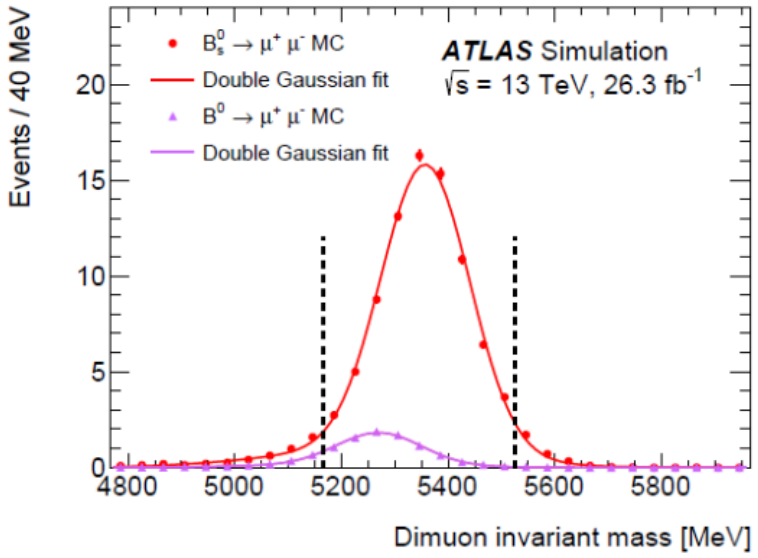


$B_s - B_d$ correlation in ATLAS

- $B_s - B_d$ mass difference: **87 MeV**
- ATLAS dimuon mass resolution roughly: **50 - 120 MeV**
  - One of the limiting factors in ATLAS' Bd sensitivity
- **Significant fraction of events recorded with small σ_m**
  →less overlap at low σ_m values

Peaks significantly overlap

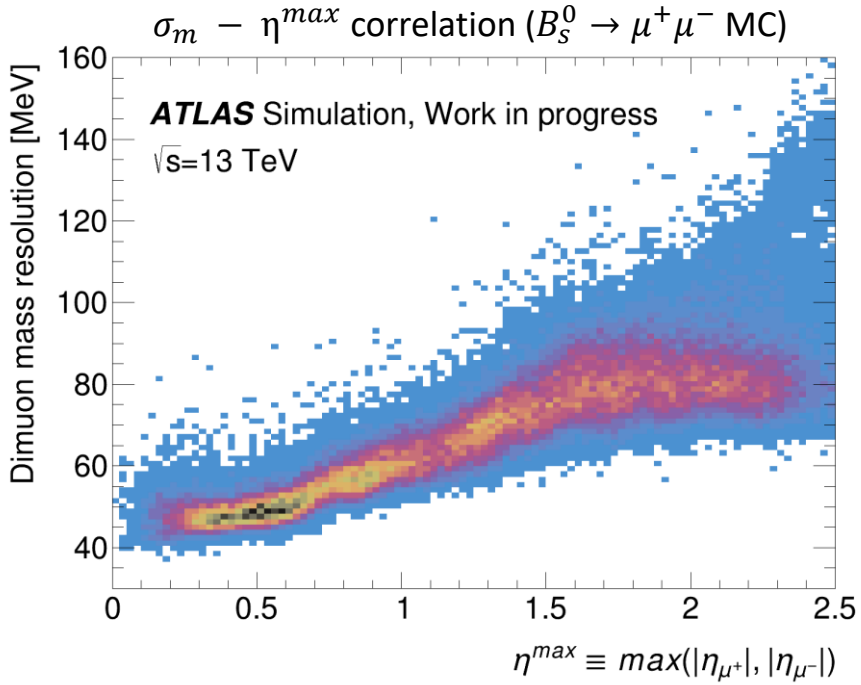
Overlapping B_d / B_s peaks [5]



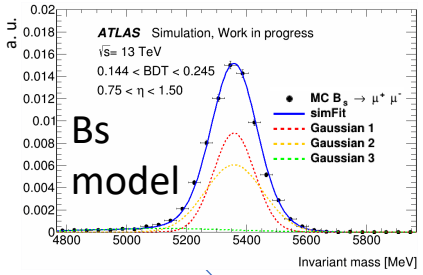
- Previously: no use of σ_m in yield extraction
  - Yields extracted in UML fit in categories of BDT output
- **Aiming for best separation of the B_s / B_d signals**
  - Full-Run2 goal: assess the gain of fitting events with different σ_m separately

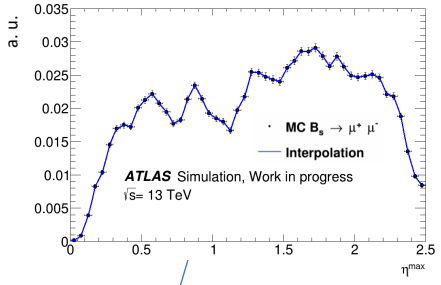
$B_s - B_d$ separation study

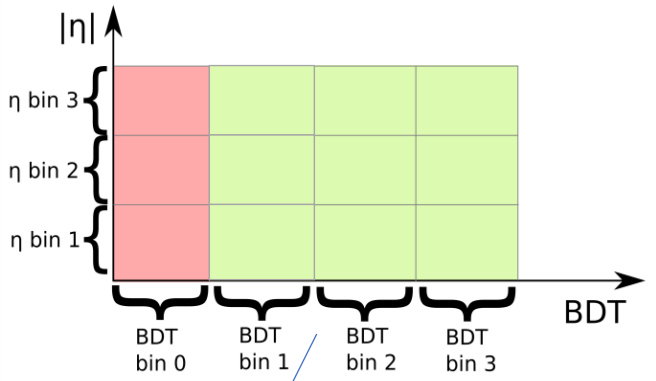
- Introducing $\eta^{max} \equiv \max(|\eta_{\mu^+}|, |\eta_{\mu^-}|)$:
  - Kinematical variable strongly correlated with σ_m
  - Better reproduced in MC than σ_m
  - Complicated shape and limited statistics
    → binning in η^{max} rather than fitting functional dependence
- What's the gain if we bin our sample in η^{max} on top of the BDT binning? What binning works best?
- Addressed in a toy-MC study



$B_s - B_d$ separation study







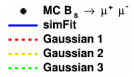
Build S & B mass models

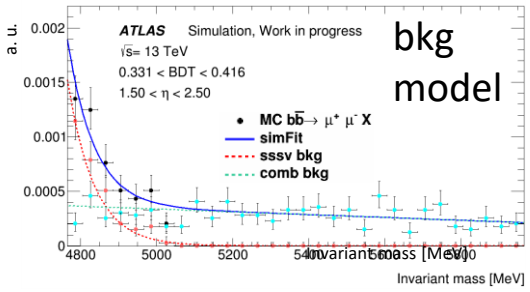
Generate toy-MCs in m, BDT, η^{max}

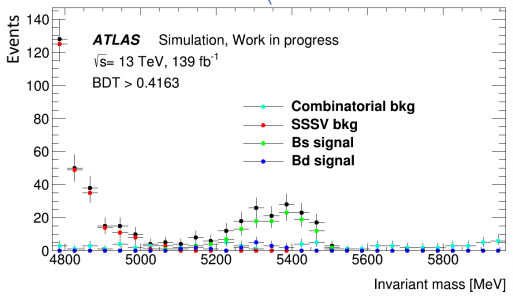
Design η^{max} & BDT dependent fit procedure

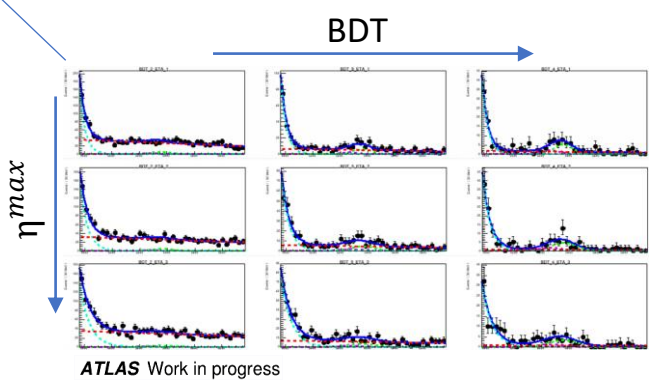
Apply various binnings, perform η^{max} dependent fit

Best-separating binning





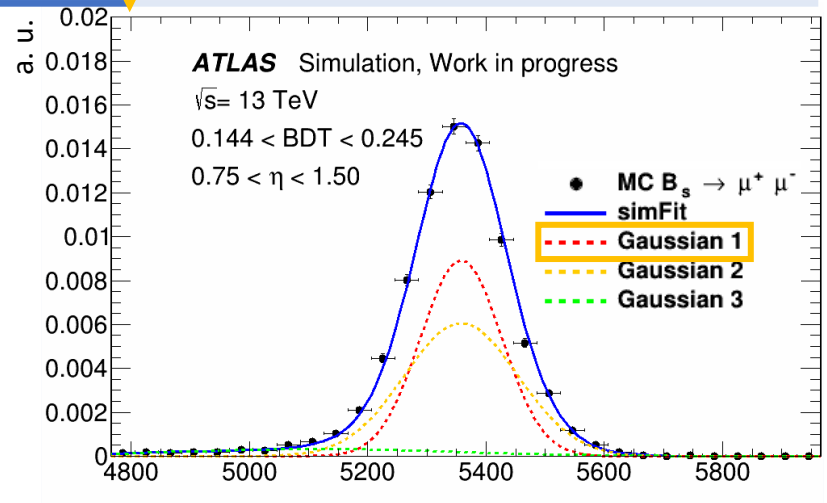


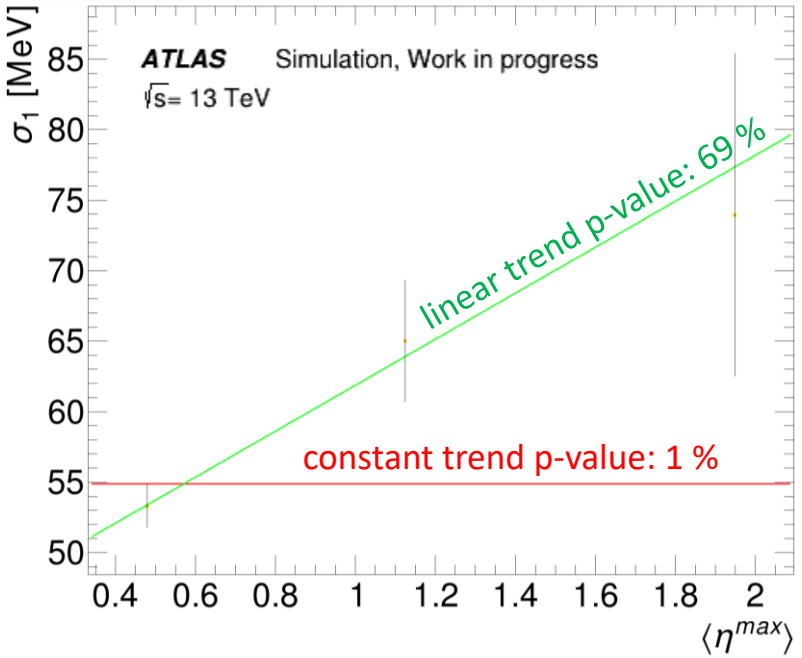


Building models for toy generation

- Toy MCs need to account for signal and background mass shape differences at different BDT & η^{max} values
  - Mass shape parameters dependence on BDT & η^{max} studied in MC
    - Constant or linear dependence found adequate
  - Dependences incorporated into a simultaneous fit over MC binned in BDT, η^{max}

Example of Bs mass shape, and one of its parameter dependences on η^{max}





Generating toys in mass, BDT, η^{max}

Full toy-generating model:

$$p(m, BDT, \eta_{max}) = g(m \mid BDT, \eta_{max}) \cdot h(BDT, \eta_{max})$$

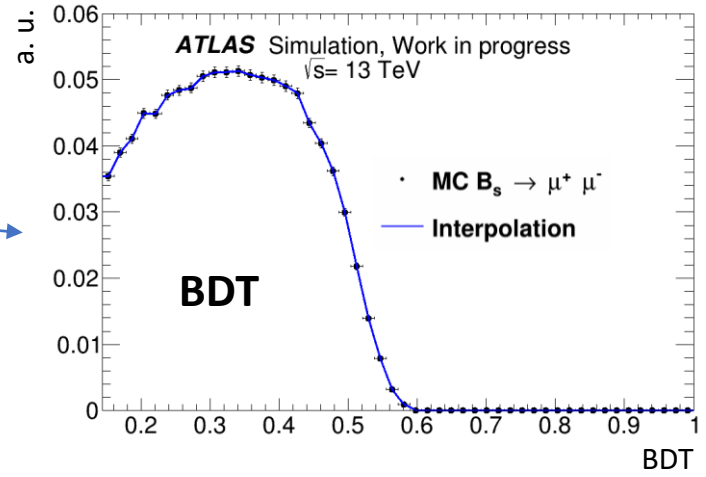
mass PDF

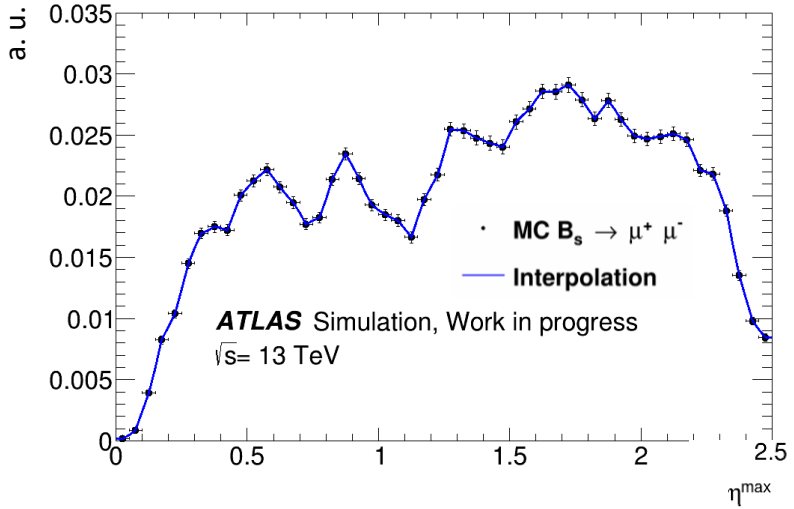
- Functional shapes & parameter depend on BDT, η^{max}

→conditional on BDT, η^{max}

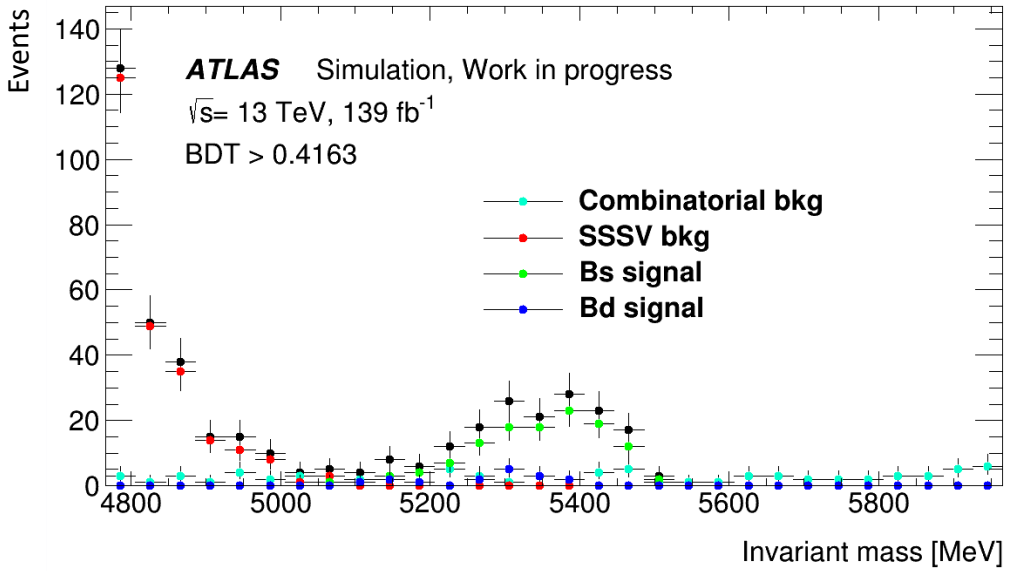
BDT & η^{max} pdfs

- interpolated histograms from MC



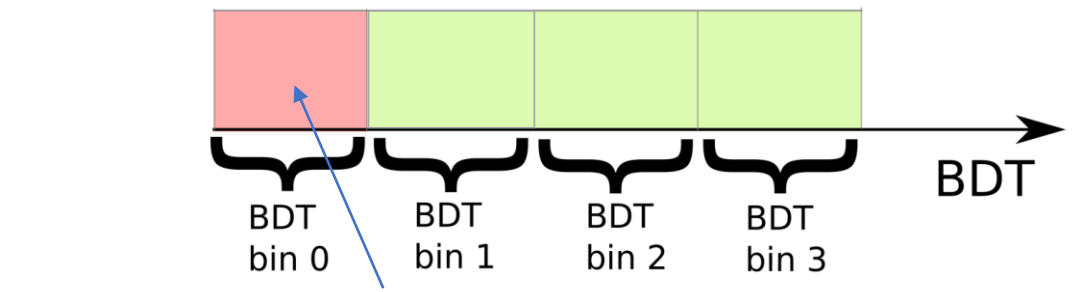


Example of generated toy events at hight BDT values

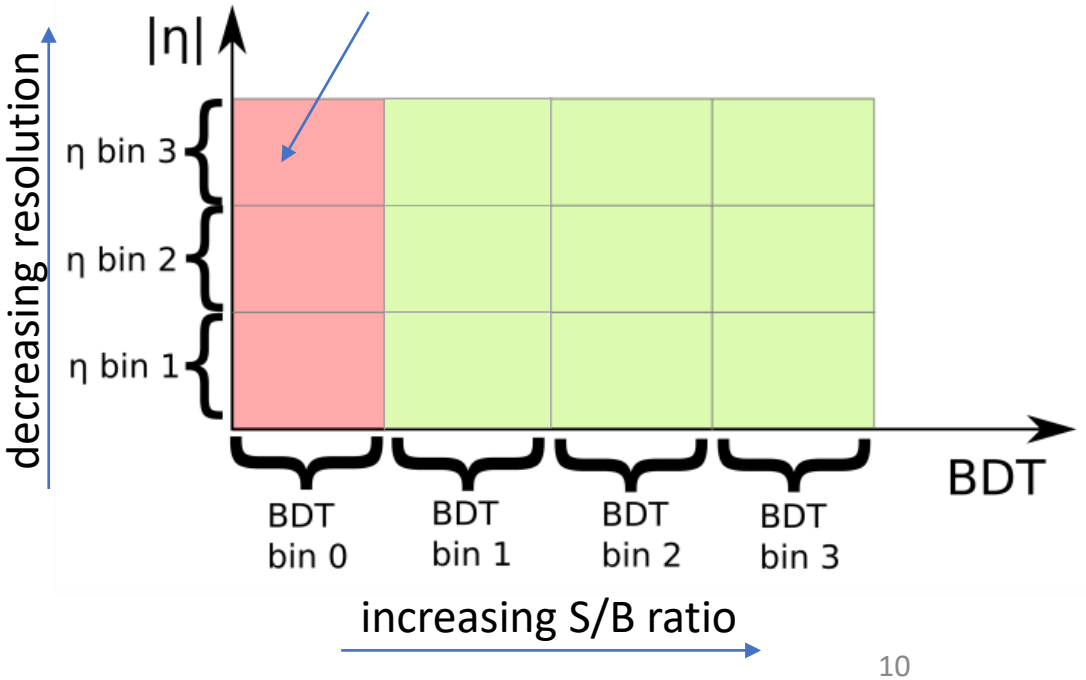


η^{max} - dependent fit procedure

- Starting point: 2015/16 analysis
  - Thoroughly studied and understood fit model
    - **reference for cross-checks**
  - No account for η^{max}
  - 3+1 BDT bins of equal signal efficiency
- Key objective: make use of narrower signal peaks at low η^{max}
  - Introduced dependence of signal widths on $< \eta^{max} >$
  - Extended the binning to 3+1 BDT bins x n η^{max} bins of equal signal efficiency
- Additional dependences in η^{max}
  - Accounting for bkg shape variations in η^{max}
  - Preliminary systematics due to η^{max} binning



BDT bin 0 is not for signal extraction, only constrains background shape

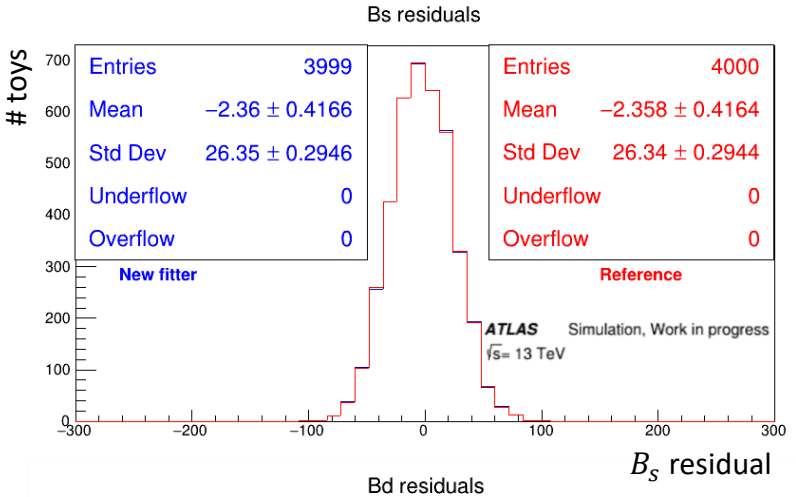


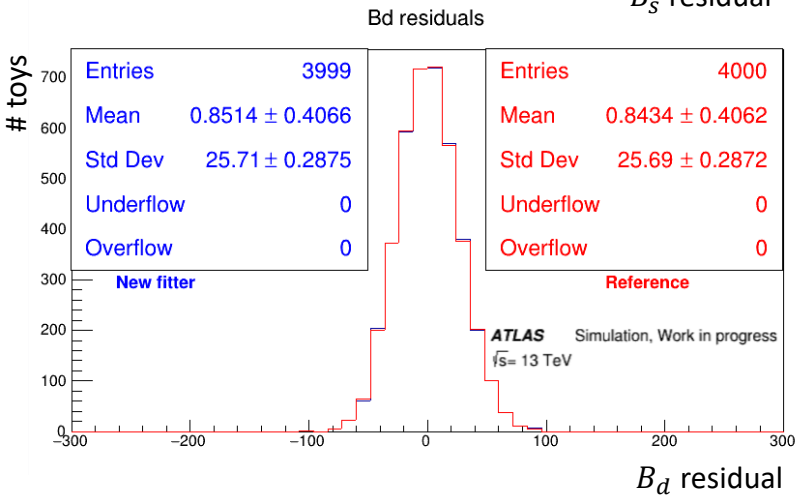
Validation against 2015/16 fit model

Validation against 2015/16 fit model to eliminate bugs in the procedure

- Mimic 2015/16 fitter with (BDT, η^{max}) dependent fitter
  - Freezing η^{max} dependence
  - Adjusting shape parameters to 2015/16 values
- Validation on 2015/16 data
  - The two fitters give identical result
- Validation on toys
  - Comparing pulls & residuals of the signal yields

Fitting tools are bug-free

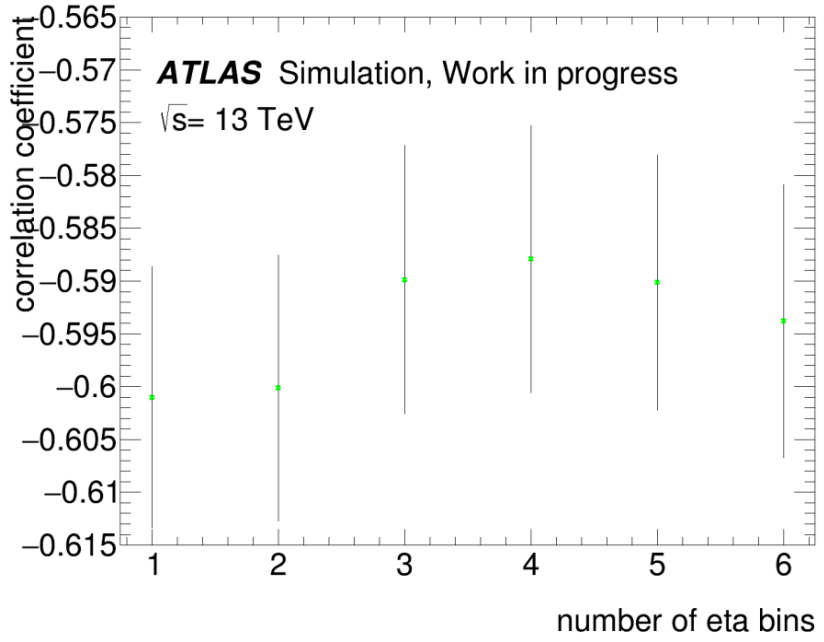


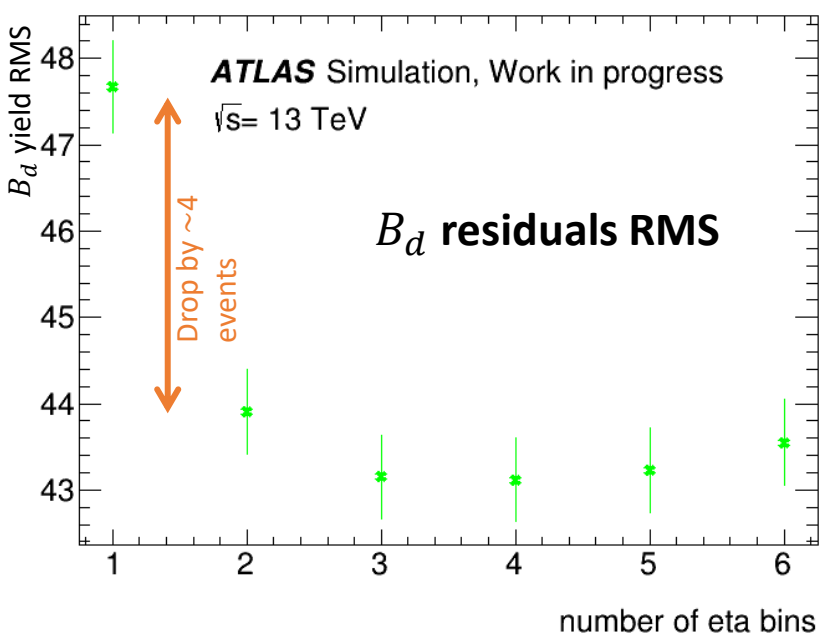


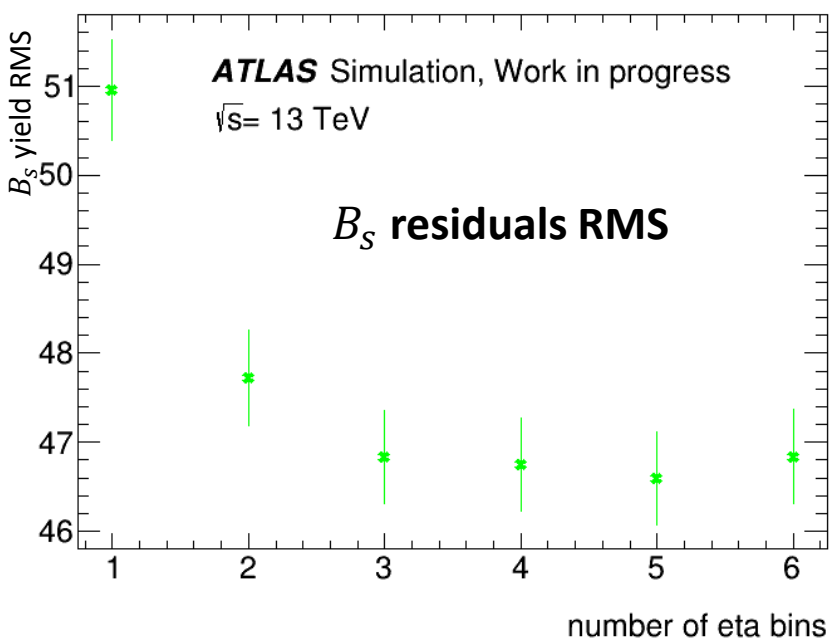
Preliminary results

Binning schemes with 1-6 η^{max} bins were explored

- Correlation of the signal yields stays very strong even with η^{max} binning
- O(10%) increase in B_d sensitivity once η^{max} binning is introduced
- Full-run 2 SM expectations:
  - B_s : 400 events
  - B_d : 44 events



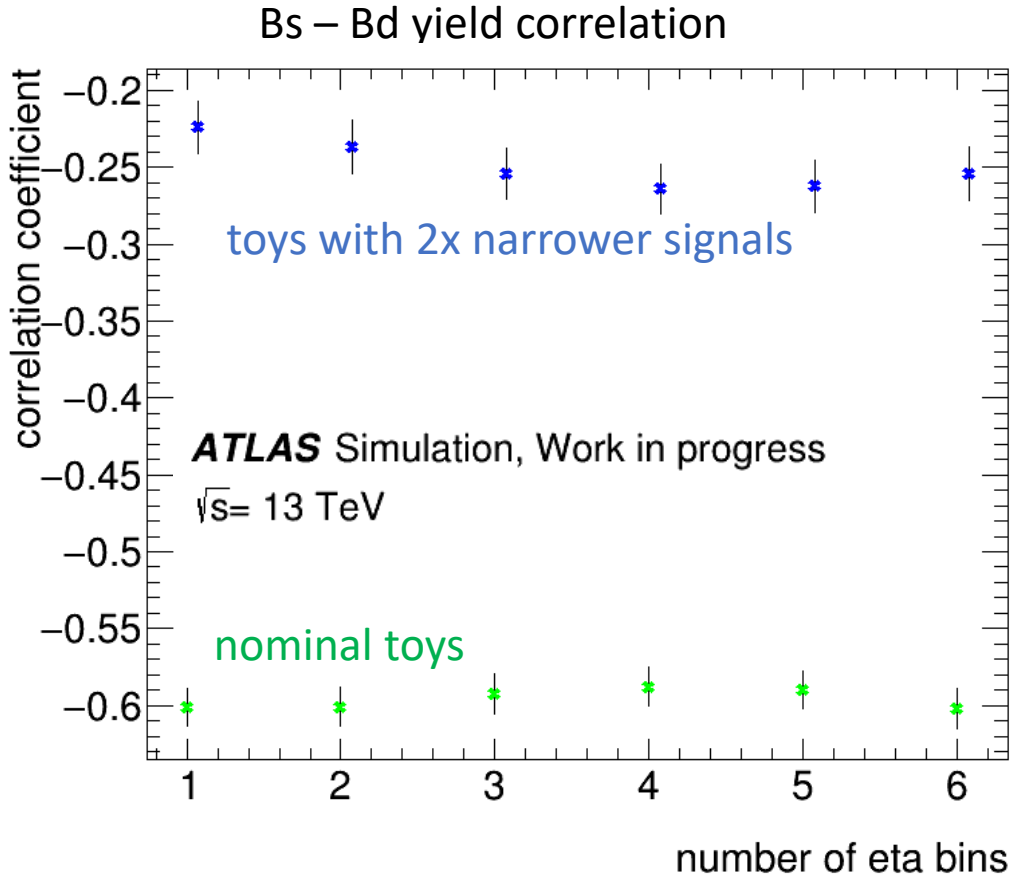




Additional studies

Besides overlap, the correlation can be caused by cross-talk through background components in the fit. Additional toy studies were performed:

- Check if the overlap is the driving force
  - Generate & fit toys with 2x narrower signals
    → Correlation significantly reduced
- See how important is the background cross-talk
  - 10 times reduced background yields in toys
    → Correlation unchanged

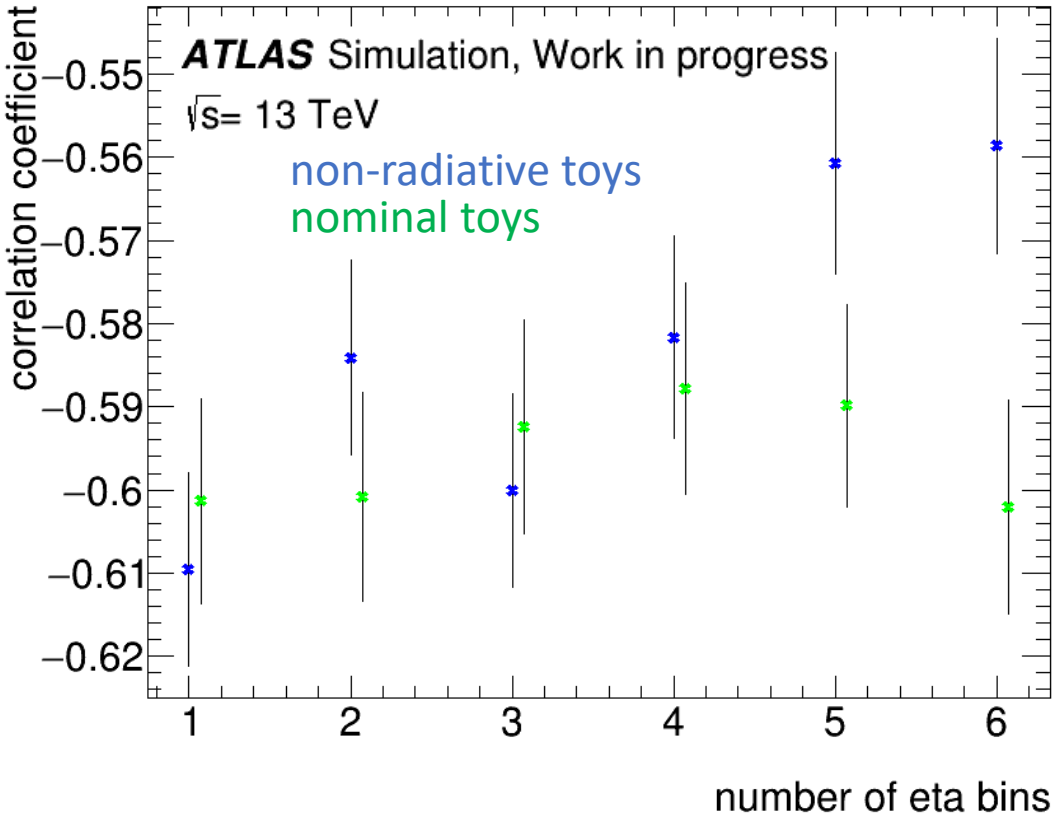


Additional studies

Some of the final state muons loose energy through photon radiation before being detected. These "radiative candidates" then accumulate in the low-mass tail. Part of the radiative Bs candidates is then found around the Bd mass.

- How much of the correlation is due to the radiative candidates?
  - Generate toys without the radiative component
    → Correlation largely unchanged
- Is the correlation statistics-dependent?
  - Generate toys at 2015/16 statistics
    → Correlation unchanged

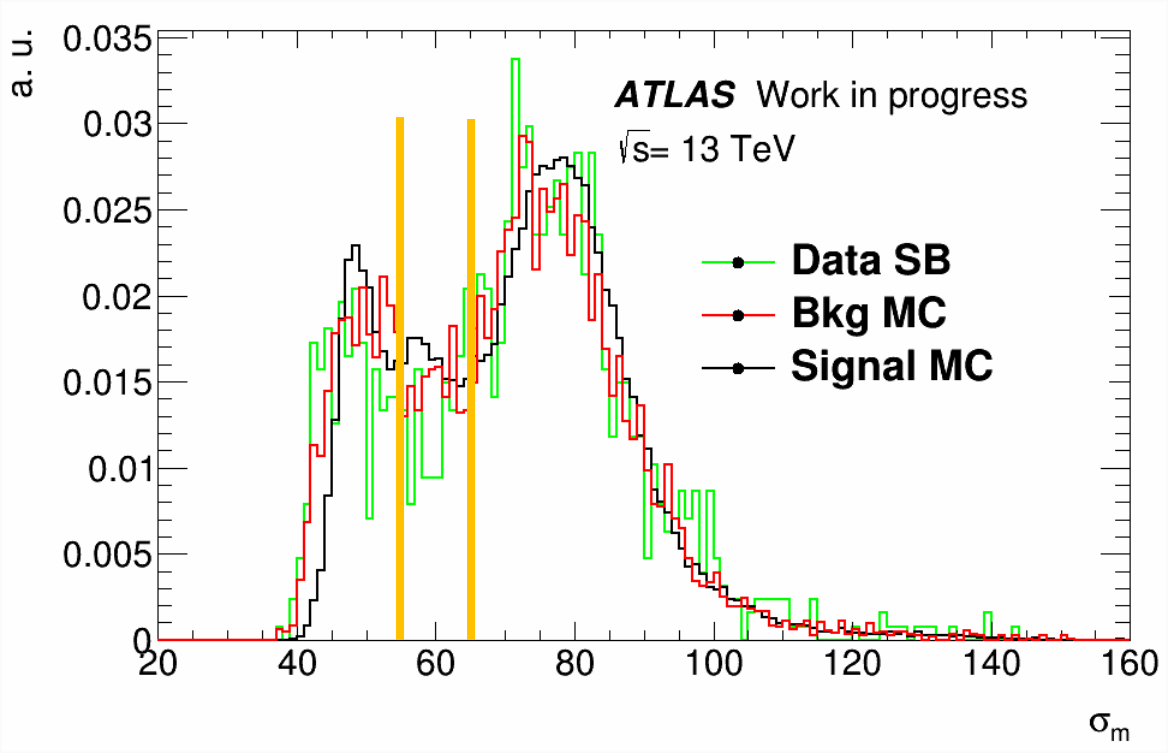
Bs – Bd yield correlation



Additional studies

An alternative approach was tested, introducing directly categories of σ_m rather than η^{max}.

- Categories cover three peaks observed in signal σ_m distribution
- Wrt. case with no η^{max}/ σ_m binning:
  - Correlation improves by ~10%
    - Better than η^{max} binning
  - Stat. uncertainty on Bd yield improves by ~10%
    - Smaller improvement on Bs yield
    - similar to η^{max} binning



Summary

- Full-Run2 $B^0_{(s)} \to \mu^+\mu^-$ ATLAS analysis is underway
- Toy study has been designed to assess the gain of introducing σ_m dependence into the B_s / B_d yield fit
  - Toy datasets were generated, including dependence of shapes on BDT & η^{max}
  - η^{max} - dependent procedure was designed & validated in the limiting case of no η^{max} binning
  - B_s / B_d yield correlation was studied as a function of η^{max} binning
- Additional toy studies were performed, pointing at peak overlap causing the correlation
- Binning in σ_m instead of η^{max} was tested
- We can reduce the correlation by about 10%, as well as the stat. uncertainty on the B_d yield
  - This gain must be assessed considering additional systematics introduced by resolution-dependent fit

Resources

[1] LHCb Collaboration, Analysis of neutral B-meson decays into two muons, Phys. Rev. Lett. 128 (2022) 041801, arXiv: 2108.0928 [hep-ph]
[2] C. Bobeth et al., $B_{s,d} \to l^+ l^-$ in the Standard Model with reduced theoretical uncertainty, Phys. Rev. Lett. 112 (2014) 101801, arXiv:1311.0903 [hep-ph]
[3] M. Beneke, C. Bobeth, and R. Szafron, Power-enhanced leading-logarithmic QED corrections to $B_q \to \mu^+ \mu^-$, JHEP 10 (2019) 232, arXiv:1908.07011 [hep-ph]
[4] D. Straub: New physics searches in flavour physics, Nuovo Cim.C 035N1 (2012), 249 arXiv:1107.0266 [hep-ph]
[5] ATLAS Collaboration, Study of the rare decays of B_s^0 and B^0 mesons into muon pairs using data collected during 2015 and 2016 with the ATLAS detector, JHEP 04 (2019) 098, arXiv: 1812.03017 [hep-ph]
[6] ATLAS, CMS, LHCb Collaborations, Combination of the ATLAS, CMS and LHCb results on the $B_{(s)}^0 \to \mu^+ \mu^-$ decays, ATLAS-CONF-2020-049 (2020)
[7] CMS collaboration, A. M. Sirunyan et al., Measurement of properties of $B_s^0 \to \mu^+ \mu^-$ decays and search for $B^0 \to \mu^+ \mu^-$ with the CMS experiment, JHEP 04 (2020) 188, arXiv:1910.12127 [hep-ph]