Pythia generator parameters tuning with Professor2 package oriented for Belle2 physics

Hazaravard Ghumaryan

(on behalf of the Yerevan Belle2 group)

hazar@yerphi.am

Oct 21 – 25, 2024 Yerevan, Armenia

A. ALIKHANYAN National Laboratory

MMCP

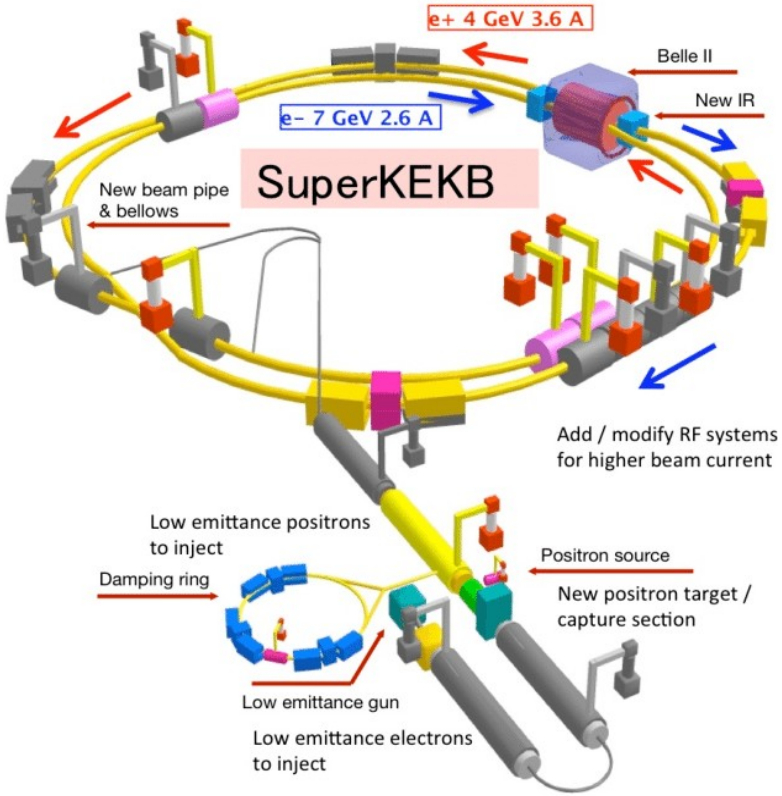
Belle II

Motivation

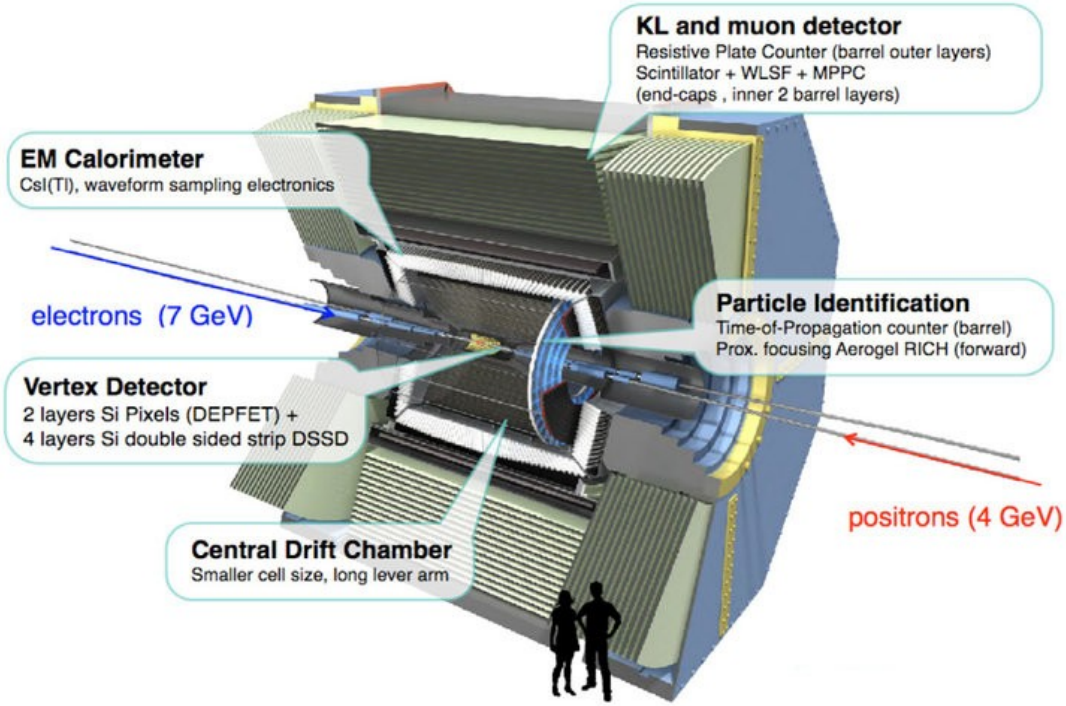
The goal of this study is to tune the Belle II off-resonance MC on the Belle II continuum data for the event variables. In this stage no PID efficiencies are involved. To achieve that goal the following studies have been performed :

- Parameter sensitivity checks to reveal the parameters for the tuning list
- Generation of the Monte Carlo samples with different Pythia8 settings sampled by the Professor2 package *(https://professor.hepforge.org/)*
- *Extraction of the physics observables from both Data & MC*
- *Tuning the off-resonance MC on the Belle II continuum data by using Professor2 package*

The Belle II experiment at the SuperKEKB accelerator in Japan



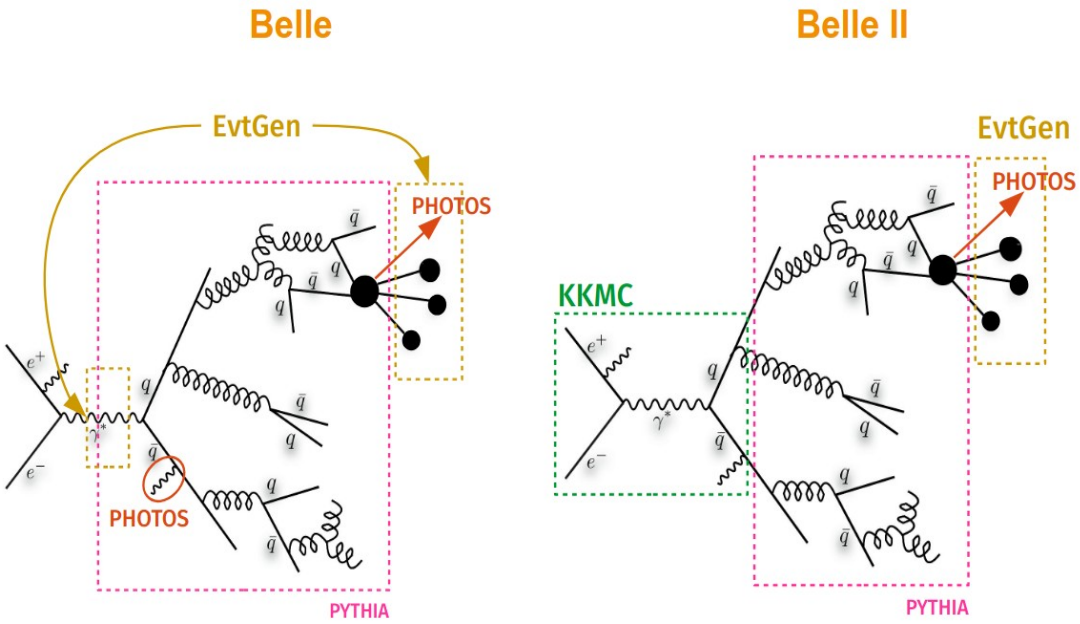
The SuperKEKB collider



Schematic view of the Belle II detector

Monte Carlo generators

PYTHIA is a program for the generation of high-energy physics collision events, i.e. for the description of collisions at high energies between electrons, protons, photons and heavy nuclei. It contains theory and models for a number of physics aspects, including hard and soft interactions, parton distributions, initial- and final-state parton showers, multiparton interactions, fragmentation and decay.

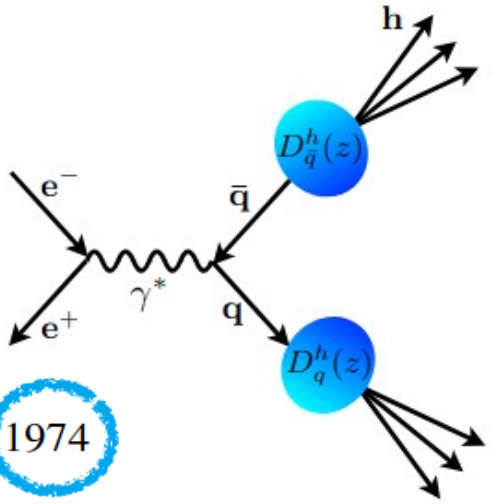


- The fragmentation process has yet to be understood from first principles
- Commonly used phenomenological models of hadronisaion:
  - **String model**
  - Cluster model
  - Independent model
  - …
- String fragmentation model was proposed by X. Artru and G. Mennessier as early as 1974
- Later was developed by **Lund** group
- And still is widely in use in Monte Carlo generators
  - **PYTHIA**
  - Lepto
  - …
- Lund symmetric fragmentation functions

$$f(z) = \frac{1}{z}\ (1-z)^{\text{StringZ:aLund}}\ e^{-\text{StringZ:bLund}\, m_T^2/z}$$

  - parameters modifying **s-quark, diquark and c-quark fragmentation**

- e^+e^- annihilation



Data and MC productions

MC15rd_b, 4S_offrs	MC luminosity (fb^{-1})
$u\bar{u}$	3.32
$d\bar{d}$	3.32
$s\bar{s}$	3.32
$c\bar{c}$	3.32
$\tau^+\tau^-$	3.32
$\mu^+\mu^-$	3.32
$\gamma\gamma$	1.66
e^+e^-	0.083
$e^+e^-e^+e^-$	0.803
$e^+e^-\mu^+\mu^-$	0.803
l^+l^-XX	0.803
h^+h^-ISR	0.830

MC Sample (4S_offres)

/belle/MC/release-06-01-08/DB00002678/
MC15rd_b/prod000*/s00/e0008/4S_offres/
//mdst/sub00

Data Sample (4S_offres)

/belle/Data/release-06-00-12/DB00002392/
proc13/prod00025341/e0008/4S_offres/*/all/
mdst/

Integrated luminosity for data sample:
Continuum (exp8 / proc13): 0.806 fb^{-1}

List of observables & event selection

Initial set (only event variables)

Thrust

Fox-Wolfram moment R2

VisibleEnergyOfEventCMS

MissingEnergyOfEventCMS

MissingMomentumOfEvent

Number of tracks in an event

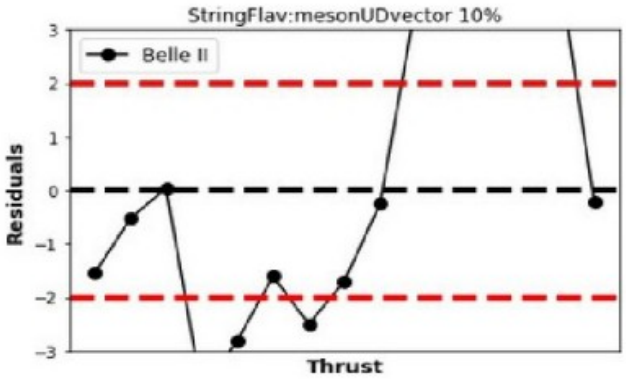
Final set (track variables added)

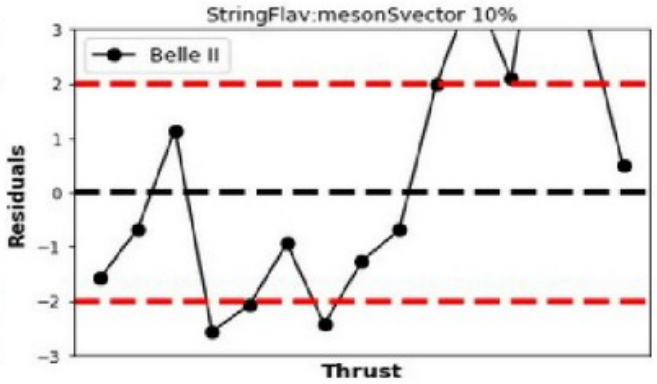
Charged particle multiplicity

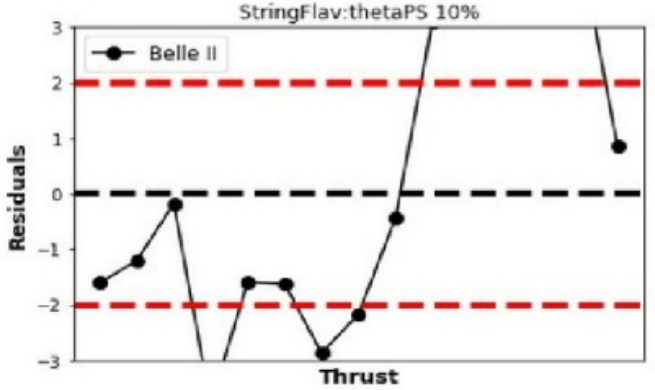
D^0, K_s invariant mass spectra near the resonance peak

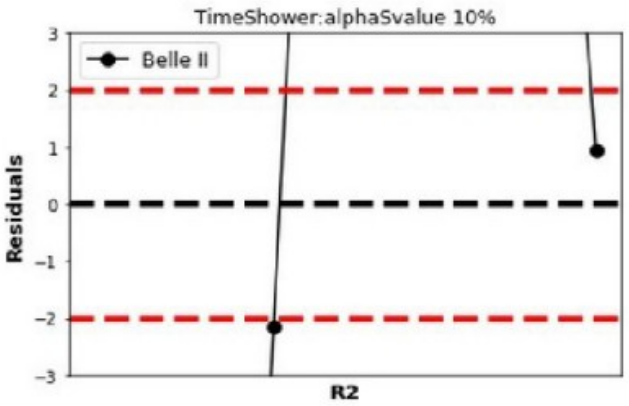
- *Good tracks : $p_\perp$ [GeV] > 0.1 , |d0| [cm] < 1 , |z0| [cm] < 2*
- *Good clusters : clusters not associated to charged track deposits, E_{lab} [GeV] > 0.1 and cluster in the CDC acceptance (0.296706 < θ [rad] < 2.61799)*
- *Number of good tracks ≥ 3*
- *Number of good clusters ≥ 3*
- *E_{vis} [GeV] > 4 (sum of good track momenta and good photon energy)*
- *$2 \leq E_{ECL}$ [GeV] < 7 (total energy in ECL)*
- *Momentum balance 0 < p_z [GeV] < 5 (sum of z component of momenta of good tracks and good photons)*

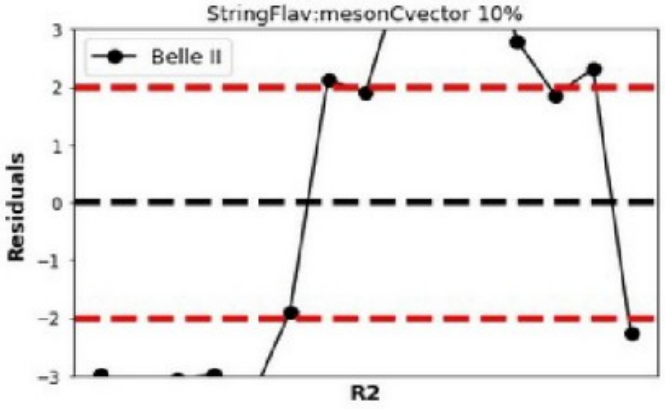
Sensitivity checks
(varying parameters from their default values)

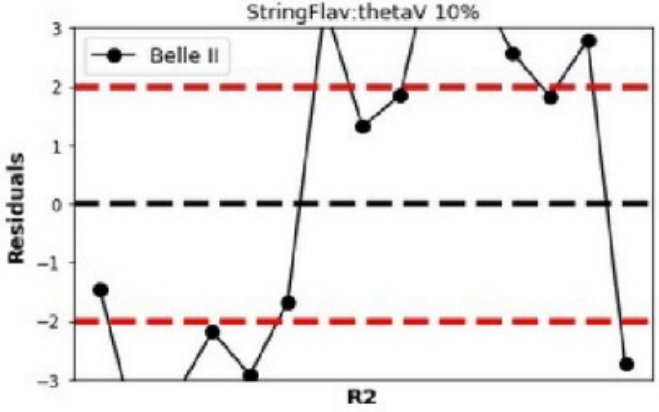












Based on the sensitivity checks the parameters below have been selected for the tuning.

- **StringFlav:mesonUDvector** (default = 0.50; minimum = 0.; maximum = 3.) the relative production ratio vector/pseudoscalar for light (u, d) mesons.
- **StringFlav:mesonSvector** (default = 0.55; minimum = 0.; maximum = 3.) the relative production ratio vector/pseudoscalar for strange mesons.
- **StringFlav:mesonCvector** (default = 0.88; minimum = 0.; maximum = 3.) the relative production ratio vector/pseudoscalar for charm mesons.
- **StringFlav:thetaPS** (default = -15.; minimum = -90.; maximum = 90.) gives the mixing angle theta_PS in the pseudoscalar meson sector (which is rather poorly determined), expressed in degrees.
- **StringFlav:thetaV** (default = 36.; minimum = -90.; maximum = 90.) gives the mixing angle theta_V in the vector meson sector (which is somewhat better determined), expressed in degrees.
- **TimeShower:alphaSvalue** (default = 0.1365; minimum = 0.06; maximum = 0.25) The alpha_strong value at scale M_Z^2. The default value corresponds to a crude tuning to LEP data, to be improved.

Extraction of observables from both Data & MC

- *Good tracks : $p_{\perp}$ [GeV] > 0.1 , |d0| [cm] < 1 , |z0| [cm] < 2*
- *Good clusters : clusters not associated to charged track deposits, E_{lab} [GeV] > 0.1 and cluster in the CDC acceptance (0.296706 < θ [rad] < 2.61799)*
- *Number of good tracks ≥ 3*
- *Number of good clusters ≥ 3*
- *E_{vis} [GeV] > 4 (sum of good track momenta and good photon energy)*
- ~~*2 ≤ E_{ECL} [GeV] < 7 (total energy in ECL)*~~
- *Momentum balance 0 < p_z [GeV] < 5 (sum of z component of momenta of good tracks and good photons)*

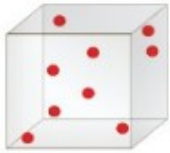
Professor2 tuning toolkit

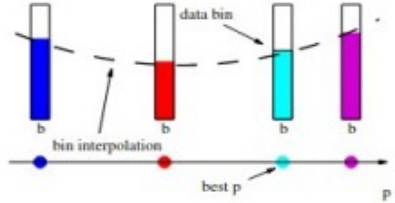
(*https://professor.hepforge.org*)

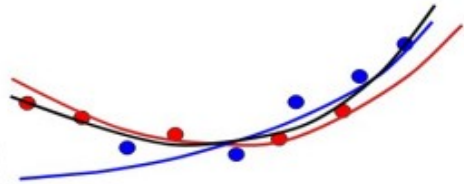
- Open source tuning toolkit for the MC event generators
- Automated tuning approach
- Provides parameter grid for tuning lists parameters (*prof2-sample*)
- Build interpolations by using given order characteristic polynomial (*prof2-ipol*)
- Selecting the best polynomial by fitting it to the experimental data (*prof2-tune*)

Tuning procedure in Professor (1D, 1Bin)

1. Random sampling: N parameter points in n-dimensional space
2. Run generator and fill histograms
3. For each bin: use N points to fit interpolation (2^{nd} or 3^{rd} order polynomial): ***5th order***
4. Construct overall (now trivial) $\chi^2 \approx \sum_{bins} \frac{(interpolation-data)^2}{error^2}$
5. and Numerically *minimize* `pyMinuit, SciPy`







Num params, P	$N_2^{(P)}$ (2nd order)	$N_3^{(P)}$ (3rd order)
1	3	4
2	6	10
4	15	35
6	28	84
8	45	165
10	66	286

$N_5^{(P=6)}$

462

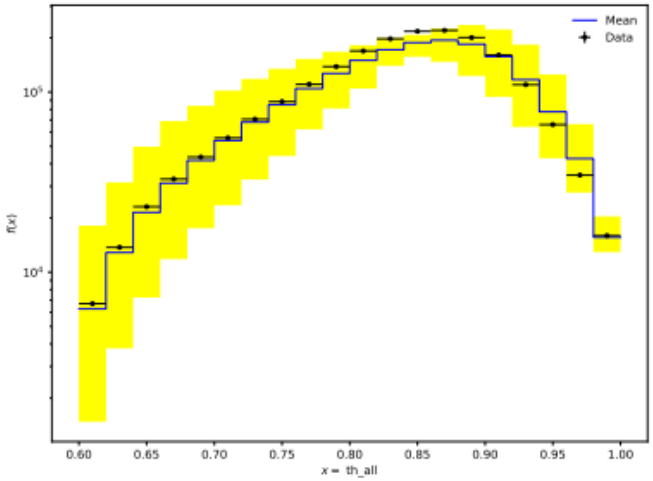
$$\mathrm{MC}_b(\boldsymbol{p}) \approx f^{(b)}(\boldsymbol{p}) = \alpha_0^{(b)} + \sum_i \beta_i^{(b)}\, p_i' + \sum_{i \le j} \gamma_{ij}^{(b)}\, p_i'\, p_j'$$

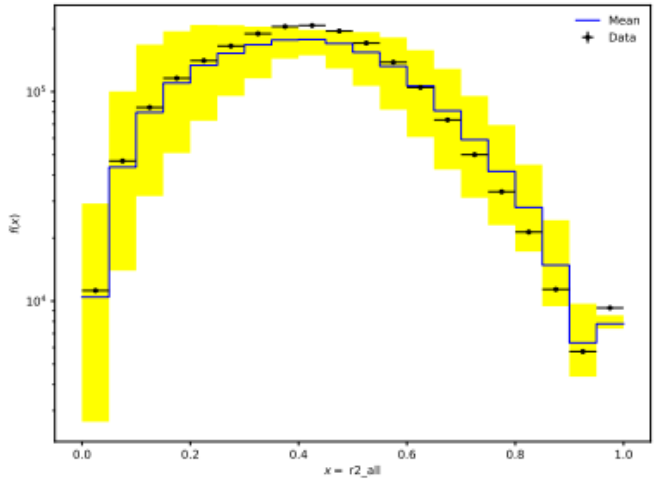
$$N_n^{(P)} = 1 + \sum_{i=1}^{n} \frac{1}{i\,!} \prod_{j=0}^{i-1} (P + j)$$

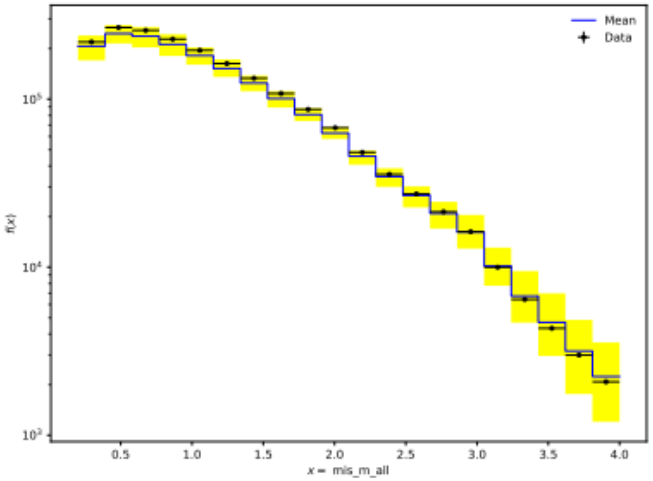
for this study 1500 samples have been generated

Sufficient number of generated samples (prof2-envelopes)

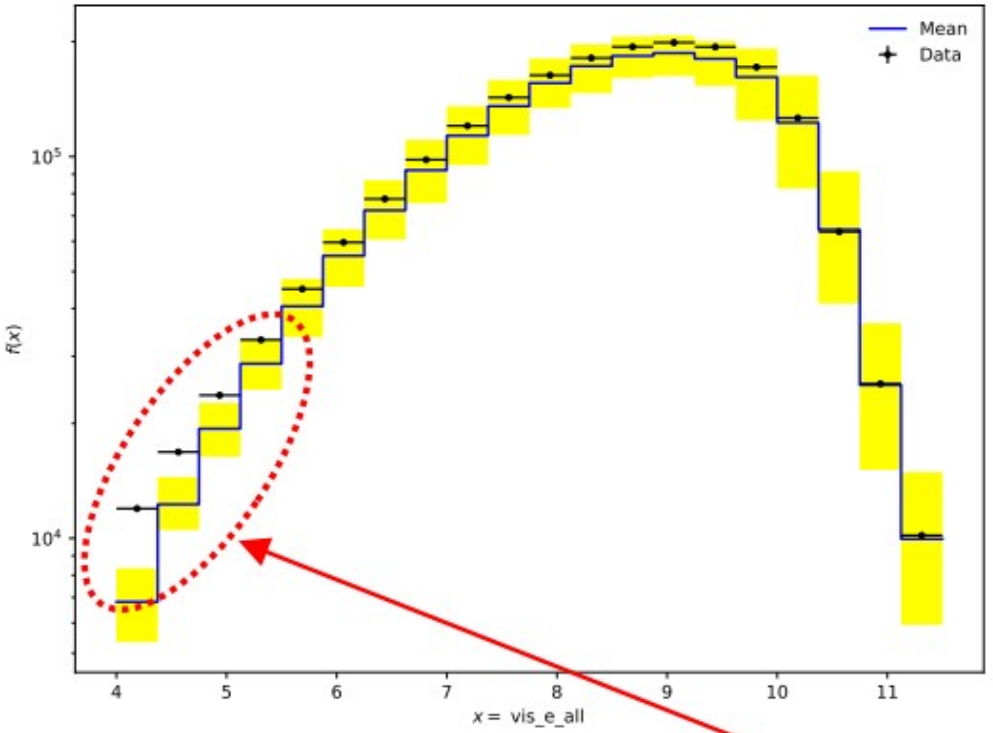
The yellow bands show the coverage of the generated MC samples by using different Pythia8 parameter settings and black points are the experimental data from the contunuum sample.

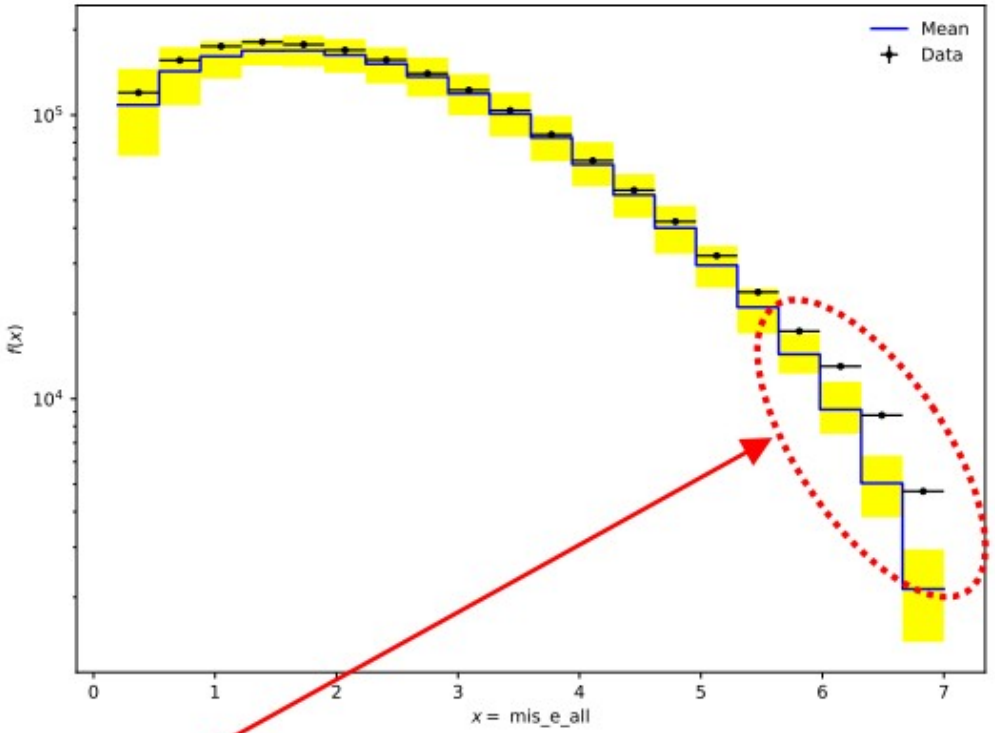






MC grid coverage (prof2-envelopes)





for the visible energy and missing energy observables there are some data points out of the MC grid coverage thus providing very small weights during the tune

Professor2 tune

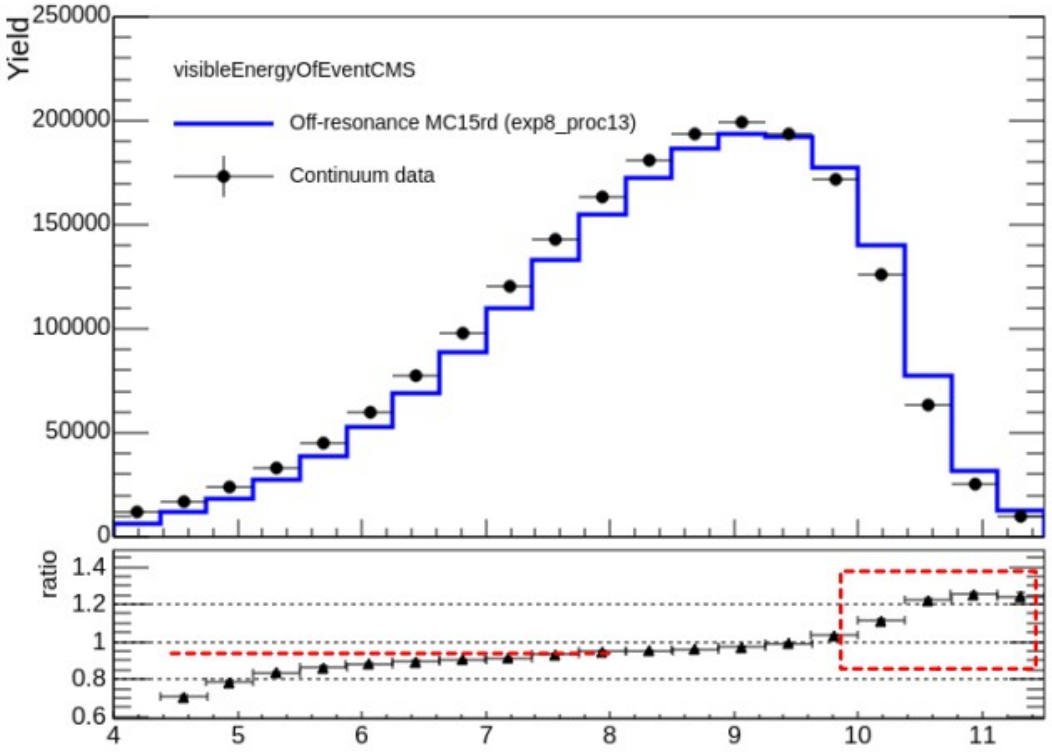
5-th order polynomial has been used for the interpolation and the parameter space is scanned up to 100000 starting points in order to ensure the robustness for the minimization results. The tuned parameters are extracted by using 1500 MC samples (anchor points).

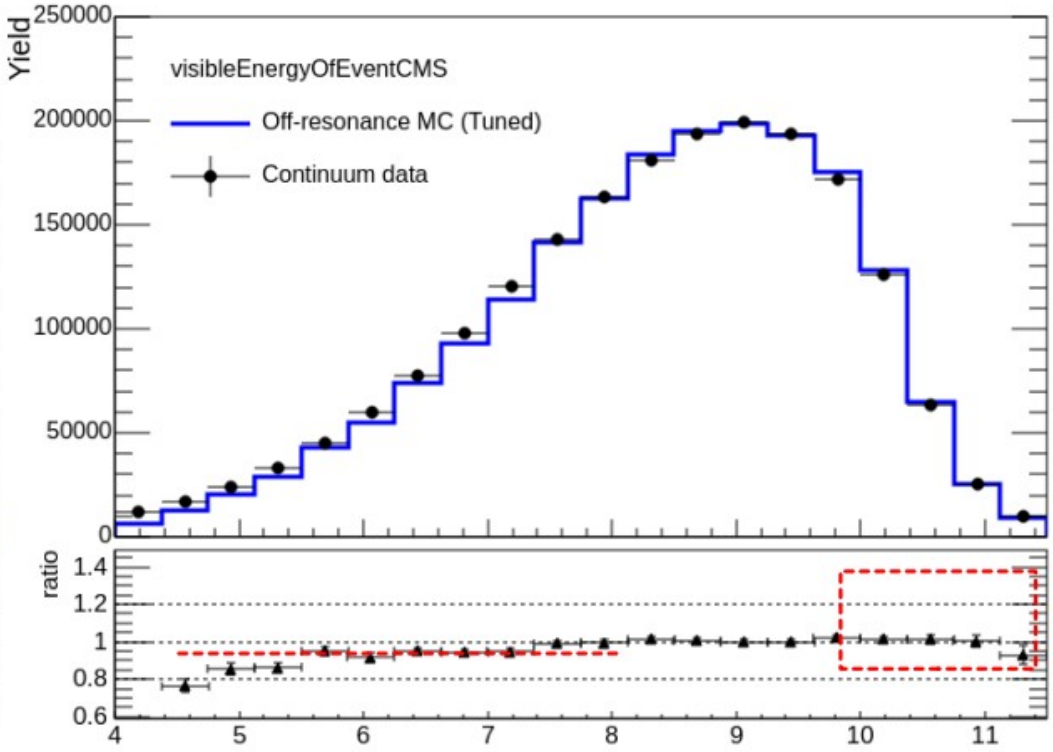
	Pythia8 default value	Tuned value
StringFlav:mesonUDvector	0.5	0.62
StringFlav:mesonSvector	0.55	2.47
StringFlav:mesonCvector	0.88	2.86
StringFlav:thetaPS	-15	-59.88
StringFlav:thetaV	36	3.55
TimeShower:alphaSvalue	0.1365	0.139

```
#                                   #
# Limits:                           # Weights used
#                                   #
# Fixed:                            # /chg_all      1.0
#                                   # /d0_all       1.0
# Minimisation result:              # /ks_all       1.0
#                                   # /mis_e_all    1.0
# GOF 2049.408260                   # /mis_m_all    1.0
# UNITGOF 2049.408260               # /r2_all       1.0
# NDOF 151.000000                   # /th_all       1.0
                                    # /trk_all      1.0
                                    # /vis_e_all    1.0
                                    #
StringFlav:mesonUDvector        0.621664
StringFlav:mesonSvector         2.446557
StringFlav:mesonCvector         2.863498
StringFlav:thetaPS              -59.882703
StringFlav:thetaV               3.547790
TimeShower:alphaSvalue          0.139261
#
# MIGRAD errors:
#
# StringFlav:mesonUDvector      3.498581e-02
# StringFlav:mesonSvector       4.554261e-02
# StringFlav:mesonCvector       2.291809e-02
# StringFlav:thetaPS            2.575822e+00
# StringFlav:thetaV             1.769148e+00
# TimeShower:alphaSvalue        2.423326e-04
#
```

Parameter correlation matrix

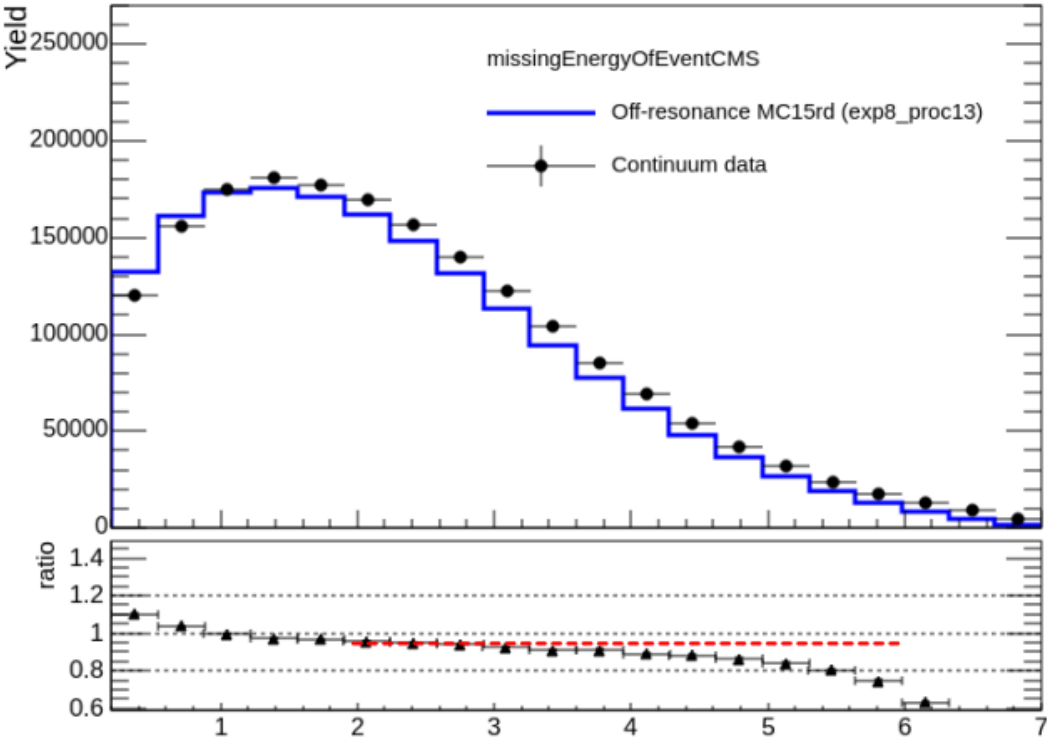
Results (before vs. after tune)

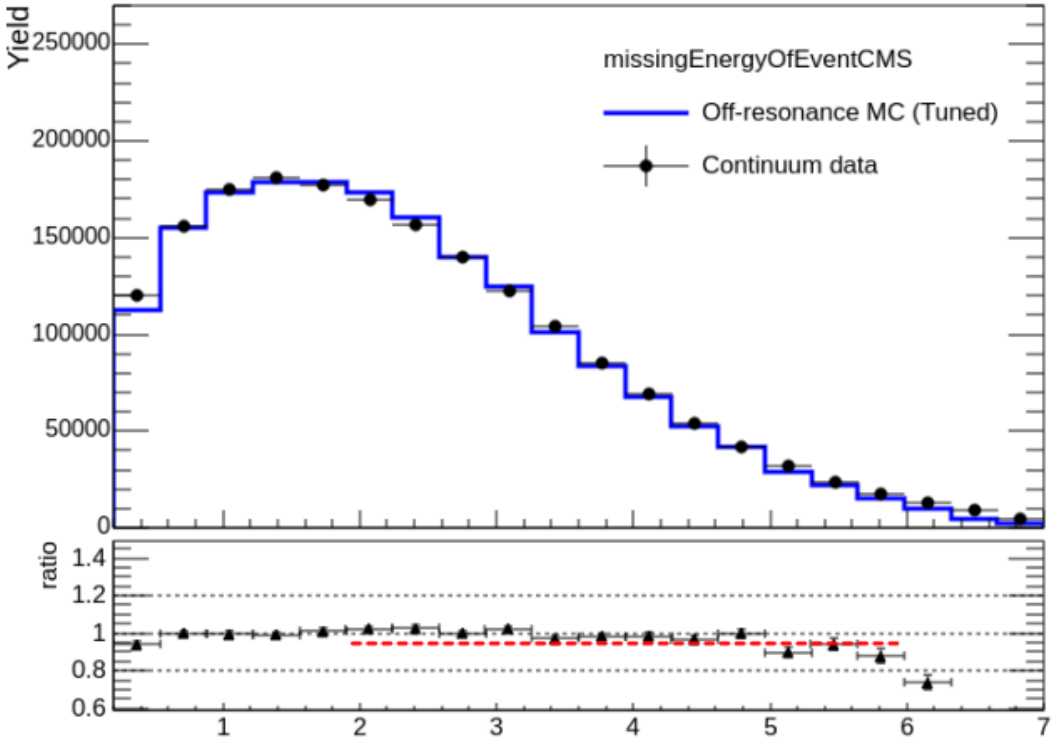




substantial improvement at the high energy tail & slight improvement of the shape in the intermediate range

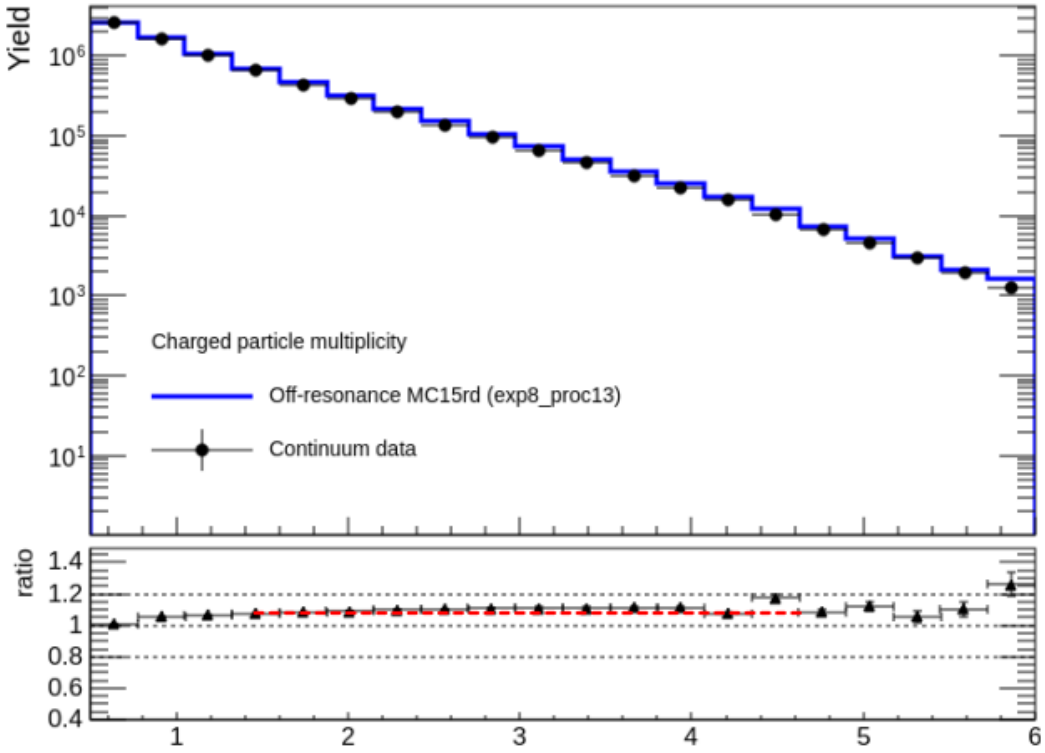
Results (before vs. after tune)

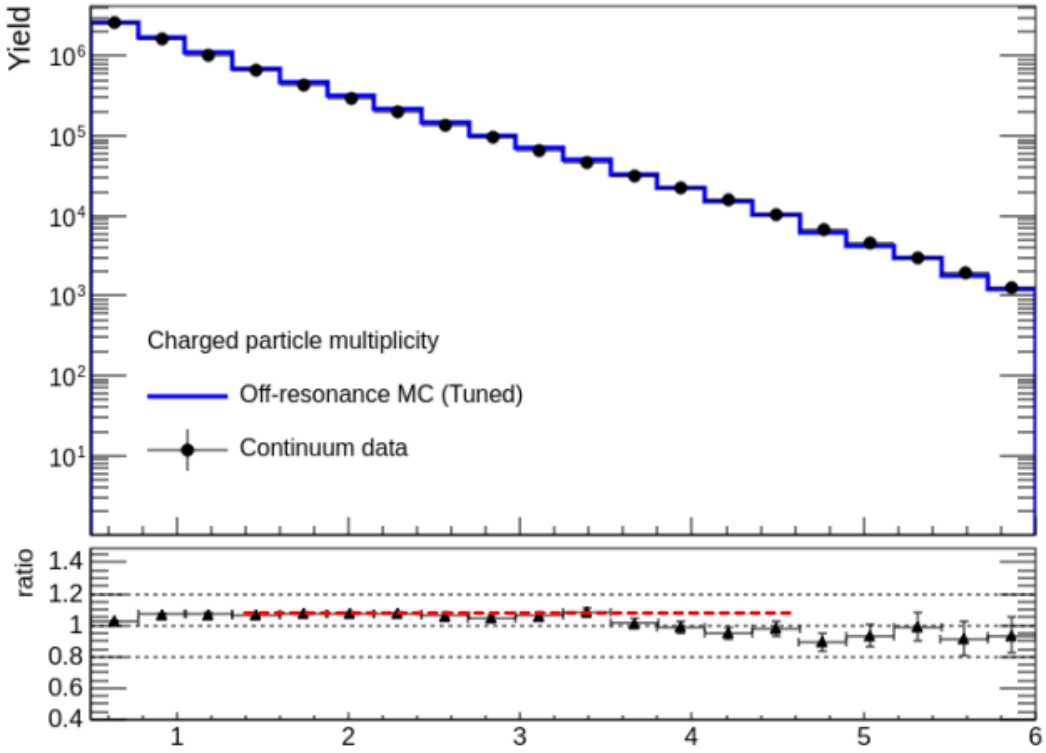




visible improvement of the shape in the intermediate range

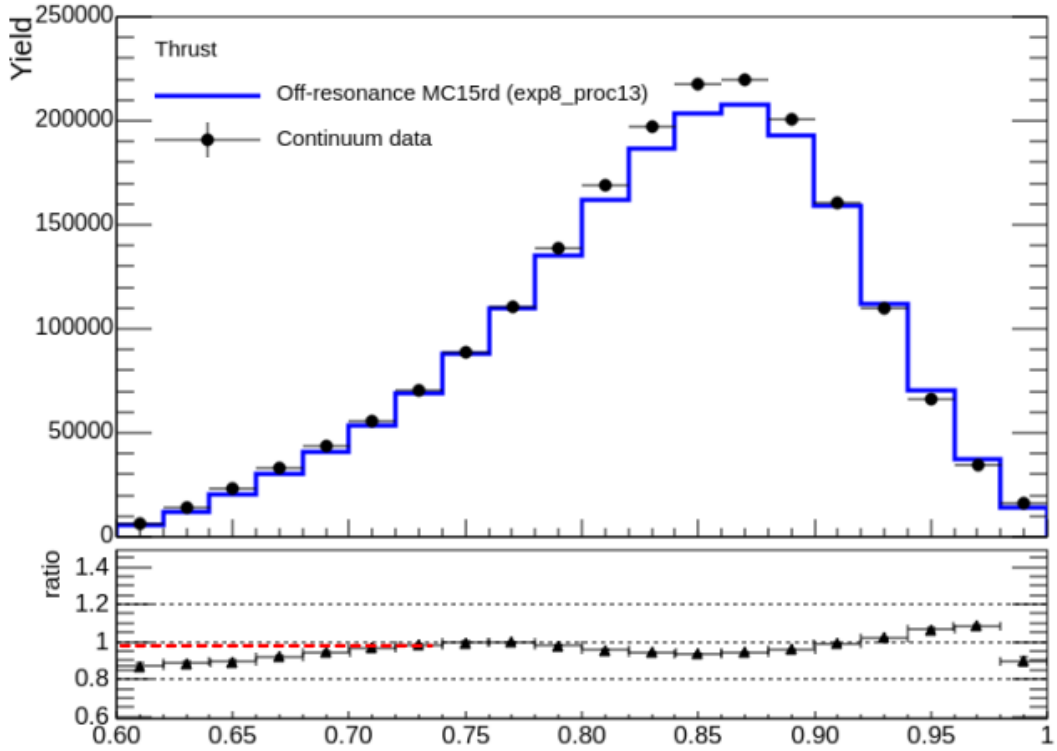
Results (before vs. after tune)

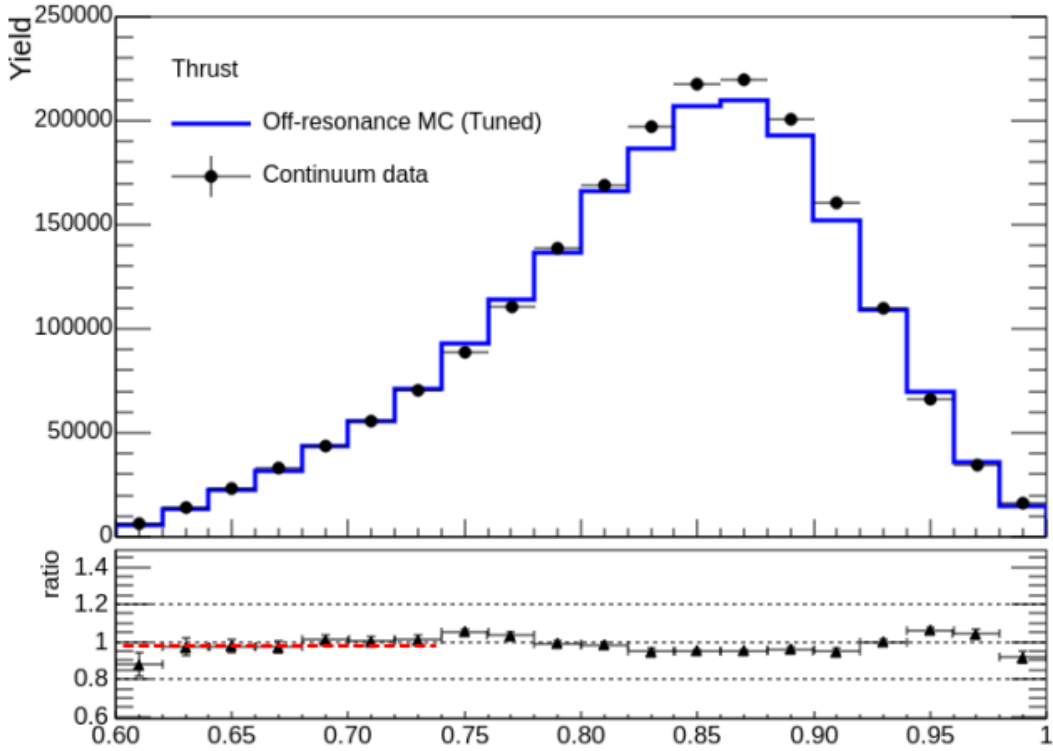




slight improvement of the shape in the intermediate range

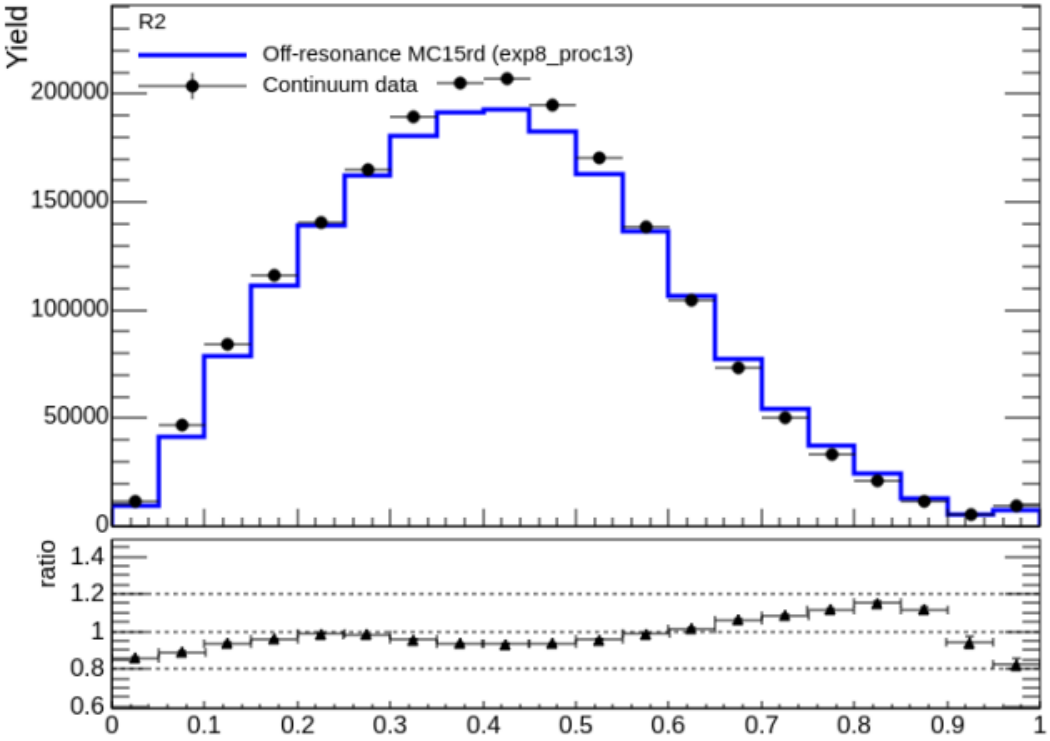
Results (before vs. after tune)

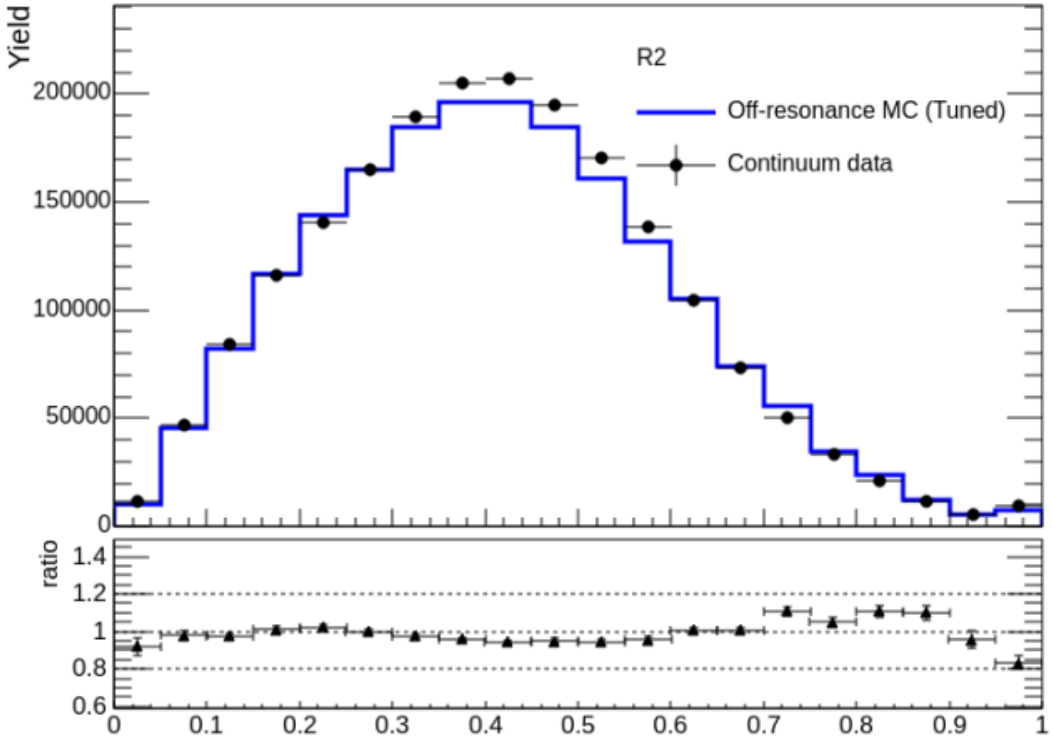




slight improvement of the shape in the lowest range

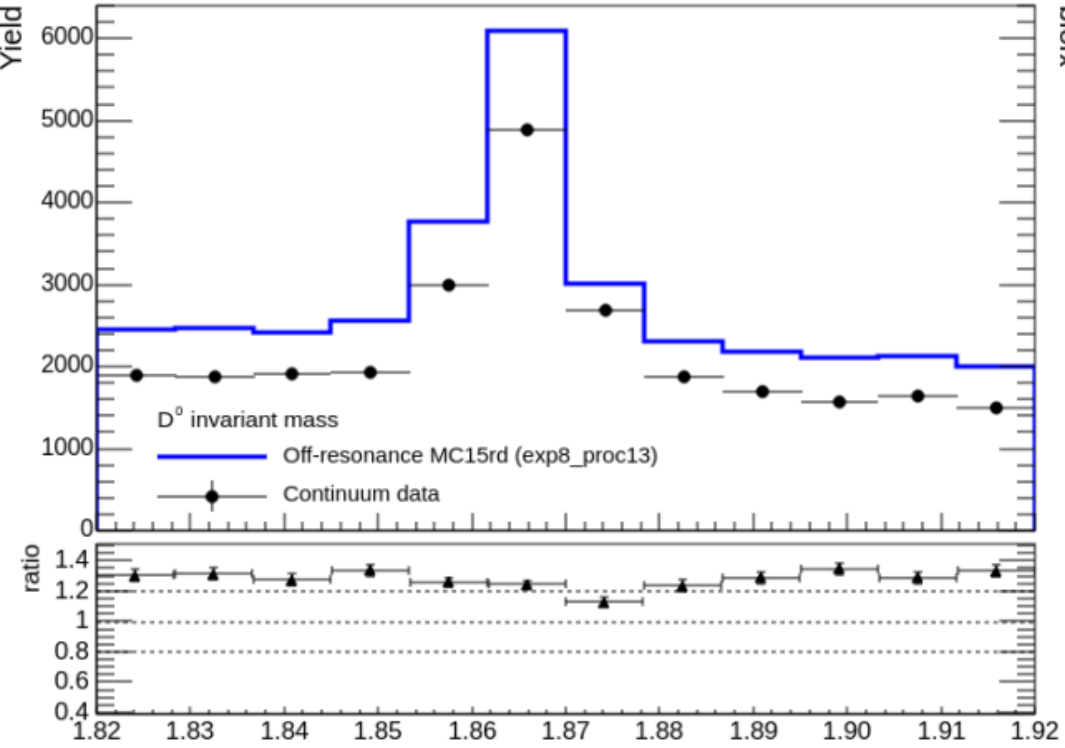
Results (before vs. after tune)

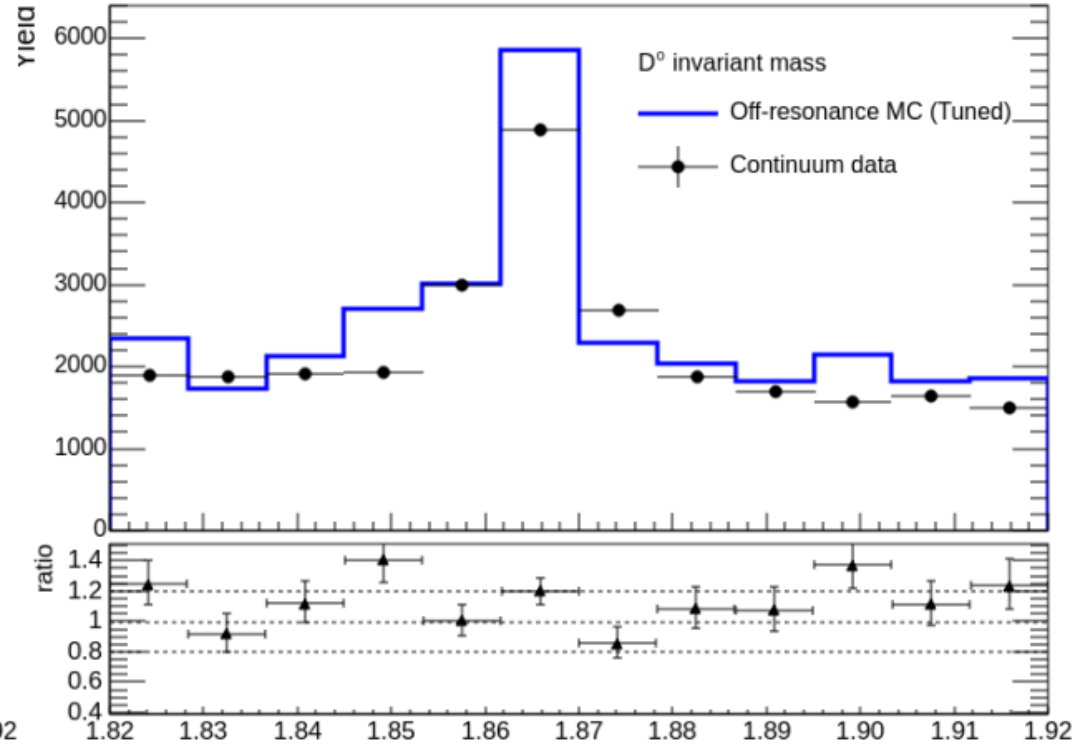




slight improvement of the shape in the whole range

Results (before vs. after tune)





visible improvement of the shape in the whole range, although limited statistics

Conlusion

We improved substantially the Data vs. MC comparison for visibleEnergyOfEventCMS and missingEnergyOfEventCMS observables as well as visible improvement is seen for track variables while no degradation observed for others.

```
#
# Limits:
#
# Fixed:
#
# Minimisation result:
#
# GOF 2049.408260
# UNITGOF 2049.408260
# NDOF 151.000000
StringFlav:mesonUDvector          0.621664
StringFlav:mesonSvector           2.446557
StringFlav:mesonCvector           2.863498
StringFlav:thetaPS                -59.882703
StringFlav:thetaV                 3.547790
TimeShower:alphaSvalue            0.139261
#
# MIGRAD errors:
#
# StringFlav:mesonUDvector        3.498581e-02
# StringFlav:mesonSvector         4.554261e-02
# StringFlav:mesonCvector         2.291809e-02
# StringFlav:thetaPS              2.575822e+00
# StringFlav:thetaV               1.769148e+00
# TimeShower:alphaSvalue          2.423326e-04
```

```
#
# Weights used
#
# /chg_all        1.0
# /d0_all         1.0
# /ks_all         1.0
# /mis_e_all      1.0
# /mis_m_all      1.0
# /r2_all         1.0
# /th_all         1.0
# /trk_all        1.0
# /vis_e_all      1.0
#
```