Mathematical Problems of Computer Science 61, 15–26, 2024.
doi: 10.51408/1963-0112

UDC 519.24, 539.1, 004.4

Pythia8 MC Tuning Validation Using the Professor2 Package

Hazaravard M. Ghumaryan

A.I. Alikhanyan National Science Laboratory, Yerevan, Armenia
e-mail: hazar@yerphi.am

Abstract

This article presents a method for multi-parameter, simultaneous tuning of the Monte Carlo event generator. It is validated on the Pythia8 Monte Carlo event generator widely used in High Energy Physics (HEP). The obtained results show that the method can be used to constrain the free parameters of phenomenological models while properly taking into account the correlations existing among the parameters.
Keywords: Monte Carlo tuning, hadronization, Pythia8, Belle II experiment, Professor2 package.
Article info: Received 3 March 2024; sent for review 19 March 2024; accepted 21 May 2024.

1. Introduction

Historically, to constrain free parameters in multi-parameter models, a tuning procedure is applied based on the change-check method i.e., parameters are changed "one-by-one" and the impact of a particular parameter on the physics observable of interest is studied. Obviously, for any complex system with too many free parameters, this method becomes non-optimal in a practical sense as it also suffers not taking into account the correlations existing between parameters. The state-of-art method is to parameterize the generator behaviour in a simultaneous change of multi-parameter set by fitting the generator response with a polynomial for each physics observable entering the tuning list:

$$MC_b(p) \approx f^b(p) = \alpha_0^b + \sum_i \beta_i^{(b)} p_i' + \sum_{i \le j} \gamma_{ij}^{(b)} p_i' p_j'. \tag{1}$$

In formula (1), MC_b is the true response of the MC generator in a bin b, $f^b(p)$ is a set of functions that model the true MC response for each observable bin b when changing the parameter vector $\mathbf{p}$ and $\mathbf{p}' = \mathbf{p}$ - $\mathbf{p}_0$ is the parameter vector shifted from its nominal value $\mathbf{p}_0$. To get the optimal, i.e., "tuned" values of the parameters, the response function is minimized concerning the reference sample by performing the χ^2 minimization as shown in

formula (2):

$$\chi^2(p) = \sum_O \omega_O \sum_{b \in O} \frac{(f^b(p) - R_b)^2}{\Delta_b^2}, \tag{2}$$

where R_b is the reference value for bin b, Δ_b is the total uncertainty of the bin b and ω_O is the statistical weight for each observable.

Recent experimental results have inspired more modeling work in the theory community. Simultaneously, efforts to refine the existing models and parameters are ongoing. Although Pythia [1] has been extensively compared to LHC, Belle, and Belle II data, its constraints from e^+e^- colliders haven't been updated since 2009, relying on an undocumented tuning effort with the Professor [2] tool. In this work, we address this issue using the Pythia8 Monte Carlo event generator from the Belle II analysis framework.

2. Toolkits for Tuning

We created a tuning framework to model the continuum data [3] produced by the Pythia8 Monte Carlo event generator. This involves the utilization of a dedicated computing platform at AANL (belle2.yerphi.am) connected to the Worldwide GRID computing system [4]. To stay updated with the latest Belle II software releases, we established access to the CernVM File System (CVMFS) [5] repository on our local server. Being a member of the Belle II international collaboration, we have full access to KEK [6] and DESY [7] computing platforms (PC Farms), enabling the simultaneous distribution of jobs across different systems.

We have designed an automated framework capable of processing different parameter sets as input and deploying corresponding jobs to the GRID. Likewise, samples generated from the Worldwide computing system are collected and organized into folders corresponding to the simulation settings. On the local "belle2.yerphi.am" machine, the Belle II Analysis Software Framework (basf2) [8] environment has been configured for the author of this article. A set of jobs is then sent to the Worldwide GRID computing platform under the periodic control and monitoring of the DIRAC [9] project. Access to this platform is facilitated through the use of the GRID certificate.

3. Tuning Procedure

To address the challenges associated with multi-parameter optimization problems, a specialized package, i.e., Professor2 has been developed as an alternative to manual adjustment methods. The approach used by the Professor2 package is known as parametrization-based tuning. A notable advantage of this package is its capacity to handle correlations among different parameters, enabling multivariate minimization within the parameter space.

The primary objective of this method is to determine the correspondence function or characteristic polynomial between the generated Monte Carlo data and the reference data, ultimately minimizing the multi-dimensional χ^2 function. In our study, we utilize the Pythia8 model for generating physics events. We present a comparison and tuning of selected observables extracted from reference data and the Belle II off-resonance Monte Carlo event generator.

One of the important aspects of the tuning procedure is to choose the free parameters from

the models to which the obsevables of interest are sensitive. This is done by performing sensitivity checks explained in Section 3.2. Having the list of sensitive parameters the tuning procedure is performed by using the Professor2 toolkit. It requires samples to be generated with different sets of sensitive parameters. The Professor2 framework provides four types of parameter sampling: grid, uniform, sobol and latin hypercube sampling. The "sobol" and "latin" hypercube distributions aim at covering the space more evenly for low sample sizes. In this work, we used "uniform" sampling with $N = 1500$ parameter points as shown below:

prof2-sample -o output params.dat -n 1500

which creates 1500 sets of parameters within predefined intervals for each. These sets are used to generate Monte Carlo samples which are later used to extract the observables of interest and the model parameters to be tuned for. The observables of interest used in this work are described in detail in Section 2.1. The generated data sets were saved in ROOT file format, which was later analyzed to extract the distributions of interest in the binned (histogram) format. It is important to mention that a special care is taken to ensure that the histograms did not contain any empty bins, otherwise, the empty bins are encountered by setting the corresponding weights to be equal to zero.

To model the MC response, a characteristic polynomial from the generated Monte Carlo (MC) samples is constructed as shown below:

prof2-ipol runsdir ipolfile=ipol.dat –order=5

To consider the multiple correlations of parameters, the order of the polynomial can be changed accordingly. In this work, the order of polynomial is set to 5. Consequently, the output of the interpolation is the file "ipol.dat" , which contains the interpolation results essential for the tuning procedure, such as the MC response extracted for each observable in terms of bin content variation.

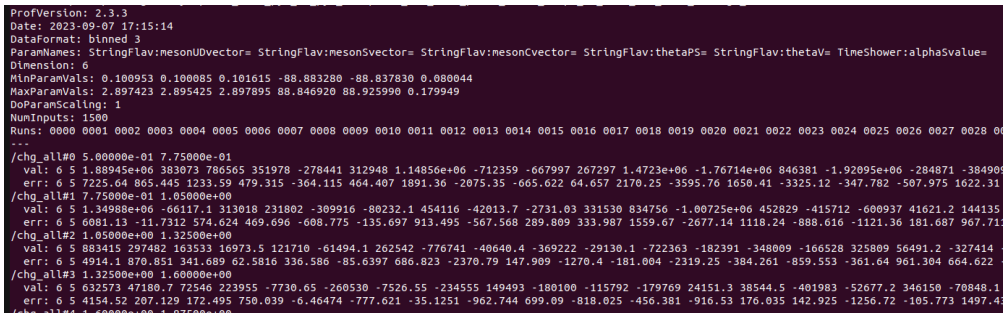
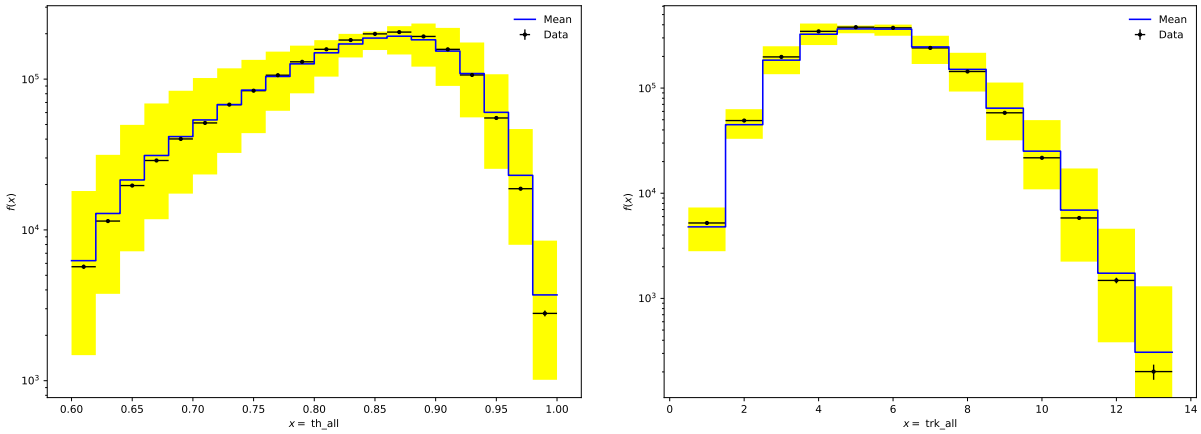


Fig. 1. "ipol.dat" file.

To ensure the full coverage of Monte Carlo samples generated with different parameter sets on the reference data, the "envelope" plots are used that effectively visualize the access of the generated data on the full spectra of the reference data:

prof2-envelopes mc -d refdir

The final step of the tuning procedure is performed using the prof2-tune tool. It minimizes the MC response function given in the form of the characteristic polynomial and accumulated in the "ipol.dat" file by comparing it with the reference data for each distribution obtained



(a) Envelope plot of Thrust

(b) Envelope plot of number of tracks in event

Fig. 2. The yellow band corresponds to the coverage of the generated MC samples by using different Pythia8 parameter settings and black points are from the reference sample.

from various MC runs with different parameter sets. These run combinations can be generated uniquely and randomly at runtime by prof2-tune, or they can be provided through a plain text file.

prof2-tune -d refdir ipolfiles=ipol.dat -r runsdir=refdir/../mc

3.1 Observables Used in Tuning

The variables (observables) used in the tuning procedure are tabulated in Table 1. and Table 2. The selection of a particular variable is motivated by the specifics of the tuning, namely, for what purpose the tuning of the generator is performed, thus, what model(s) entering the generator should be constrained. In this work, two-stage tuning was performed. First, three variables (see Table 1) of interest were selected. In the second stage, one more variable is added to the list (see Table 2).

Table 1. Observables used for the first stage tuning.

Thrust
Inclusive charge particle momentum spectra
Number of tracks in event

It is important to mention that adding more variables to the tuning list increases the possible correlation between the parameters in the tuning list thus making it more difficult to minimize multidimensional χ_2 for the MC response function.This study used three and four variables consequently, achieving a good agreement between the reference sample and MC simulations. In this work, the "Event" and "Event shape" variables are used to study the hadron production process later modeled in Pythia8, as well as to separate the events with different quark origins produced in the Belle II experiment. Specifically, the Thrust and

Table 2. Observables used for the second stage tuning.

Thrust
Inclusive charge particle momentum spectra
Number of tracks in event
visibleEnergyOfEventCMS

FoxWolframR2 [10] variables are crucial in distinguishing continuum events from BantiB events. At the same time, the use of event variables is essential to avoid possible issues related to Particle Identification (PID) inefficiency effects and background interference. The Thrust axis T is determined by the direction in which the sum of the longitudinal momenta of particles reaches its maximum. The thrust T is connected to the Thrust axis by

$$T = \frac{\sum |\mathbf{p_i} \cdot \mathbf{T}|}{\sum |\mathbf{p_i}|}, \tag{3}$$

where $\mathbf{p_i}$ represents the momentum of each particle.
The Fox-Wolfram moments are defined as follows:

$$H_x^l = \sum_{i,j=1}^{N} W_{ij}^x P_l(\cos\Omega_{ij}), \tag{4}$$

where W_{ij}^x is a weight factor, and $P_l(\cos\Omega_{ij})$ is the Legendre polynomial, and the FoxWolframR2 is given by the ratio

$$foxWolframR2 = \frac{H_2}{H_0}. \tag{5}$$

The inclusive charged particle momentum spectrum in high-energy particle physics refers to the distribution of momenta for all charged particles produced in a collision of beams. This observable provides insights into the overall behavior and characteristics of particle production within a given experiment.

Another important observable in the experiment is the "visibleEnergyOfEventCMS", which is defined as a sum of the energies of all particles that leave observable signals in the detector:

$$E_{\text{vis}} = \sum_{\text{visible particles}} E_i. \tag{6}$$

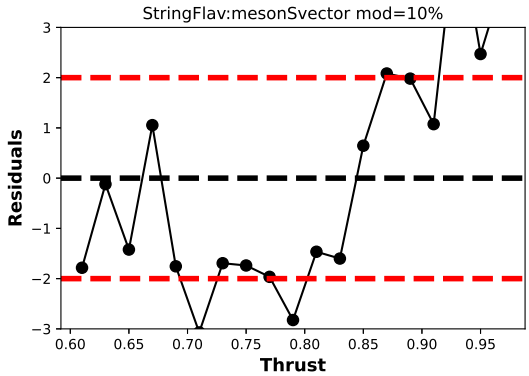
3.2 Sensitivity Checks

To reveal the sensitive parameters for the selected list of observables, the parameter sensitivity checks are perfomed using the normalized residuals extracted from two sample tests, as shown below :

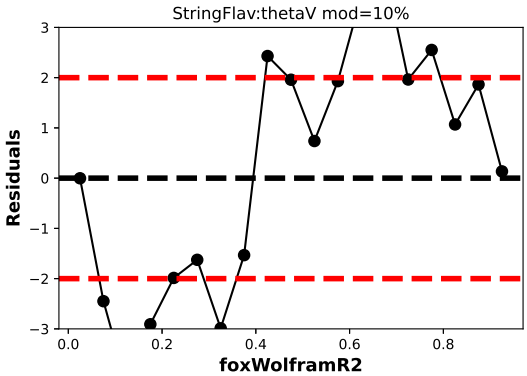
$$r_i = \frac{n_i - N\hat{p}_i}{\sqrt{N\hat{p}_i}\sqrt{(1 - N/(N+M))(1 - (n_i + m_i)/(N+M)}}. \tag{7}$$

Table 3. The list of six different parameters that are selected from the Pythia Monte-Carlo event generator for tuning the "Event" and "Event shape" variables.

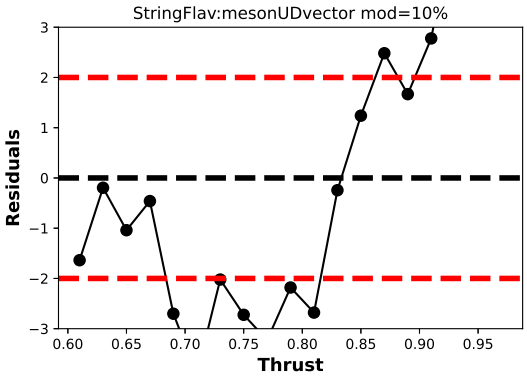
Parameters	Values
StringFlav:mesonUDvector	(default = 0.50; min = 0., max = 3.)
StringFlav:mesonSvector	(default = 0.55; min = 0., max = 3.)
StringFlav:mesonCvector	(default = 0.88; min = 0., max = 3.)
StringFlav:thetaPS	(default = -15.; min = -90., max = 90.)
StringFlav:thetaV	(default = 36.; min = -90., max = 90.)
TimeShower:alphaSvalue	(default = 36.; min = -90., max = 90.)



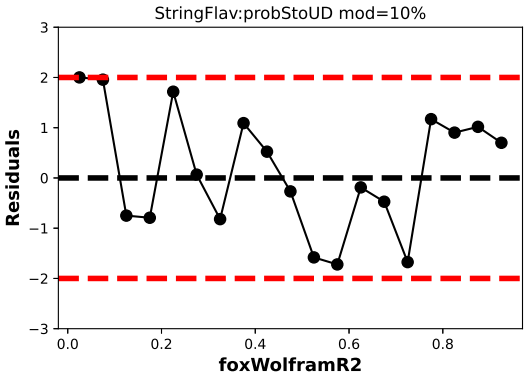
(a) The sensitivity plot of StringFlav:meson Svector Pythia parameter



(b) The sensitivity plot of StringFlav:thetaV Pythia parameter



(c) The sensitivity plot of StringFlav:meson UDvector Pythia parameter



(d) The sensitivity plot of StringFlav:probStoUD Pythia parameter

Fig. 3. The sensitivity check for parameters given in Table 2 based on Thrust and FoxWolfram distributions.

When normalized residuals exceed the 2 sigma level (see Fig. 3) for certain observables, we consider them to be sensitive to a particular parameter. Conversely, for other parameters, there is an absence of sensitivity, as indicated by residuals within the 2 sigma window. The most sensitive parameters concerning the "event" and "event shape" variables are listed in Table 3.

3.3 Validation Scheme

For the validation of the developed scheme, the Monte Carlo (MC) simulations provided by Belle II collaboration were used as reference samples.

4. Results

Our studies show that the validation of the Pythia8 MC tuning procedure using the Professor2 package is highly affected by the statistics and the number of generated samples, as well as the correlations between Pythia8 parameters. Controlling these factors is crucial for obtaining meaningful and reliable results from simulations using Pythia8. Meanwhile, by increasing the statistics and the number of generated MC samples for u,d,s,c quark combinations and using interpolation with a 5th order characteristic polynomial, we reproduce the distributions extracted from the reference sample as it can be seen in Figures 5-6. It is important to mention that due to correlations, the tuned values of parameters can differ from their default values although the spectra of observables of interest from tuned and reference samples are in a very good agreement.

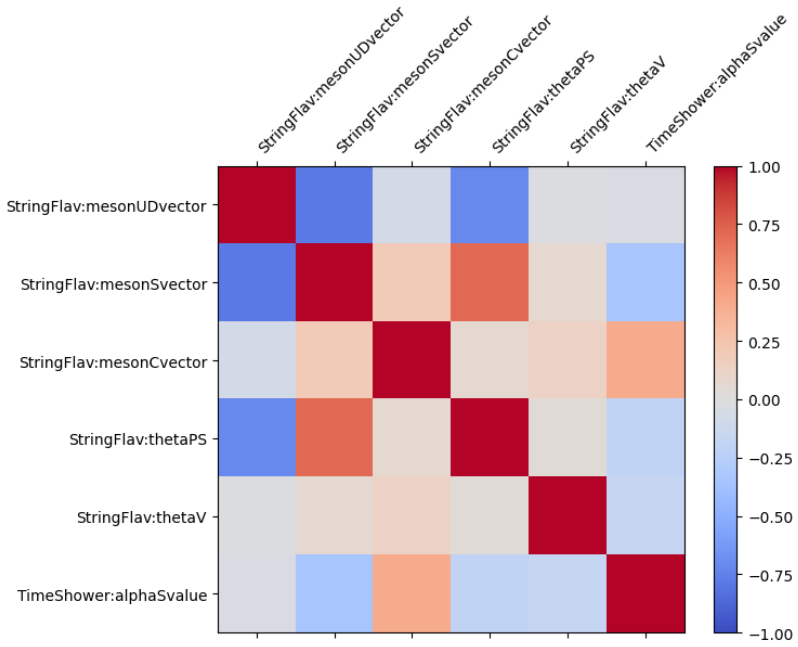


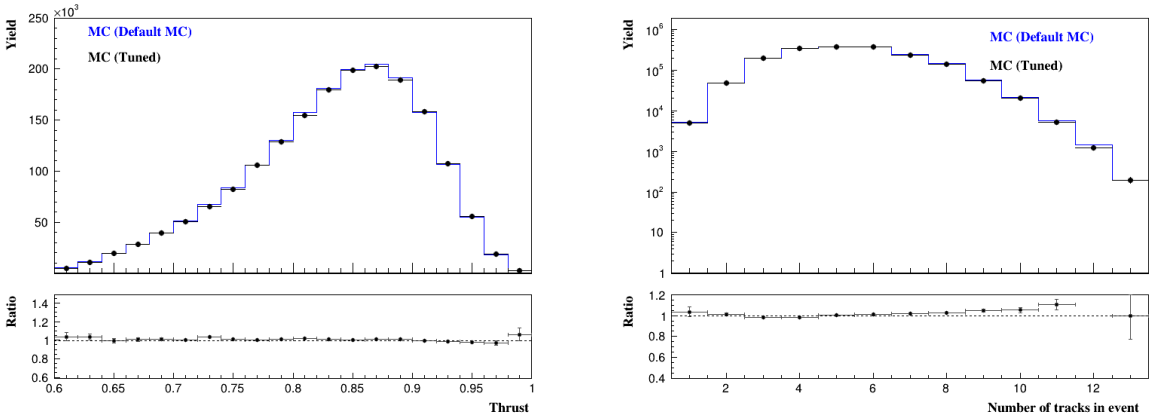
Fig. 4. Correlations values of the parameters for the first stage tune.

For the first tune we have chosen 3 variables, they are:

Table 4. Observables used for the first stage tuning.

Thrust
Inclusive charge particle momentum spectra
Number of tracks in event

Pythia8 parameters	Default	Tuned	Comment
StringFlav:mesonUDvector	=0.5	=0.444457	! Light-flavour vector suppression
StringFlav:mesonSvector	=0.55	=0.434957	! Strange vector suppression
StringFlav:mesonCvector	=2.8	=2.428458	! Charm vector suppression
StringFlav:thetaPS	=-15	=-13.500045	! Mixing angle *thetaPS*
StringFlav:thetaV	=36	=29.243306	! Mixing angle *thetaV*
TimeShower:alphaSvalue	=0.1365	=0.135795	! Effective alphaS(mZ) value



(a) (b)

Fig. 5. Comparing Thrust and Number of tracks in event distributions: Default Monte Carlo (MC) in blue vs. tuned in black.

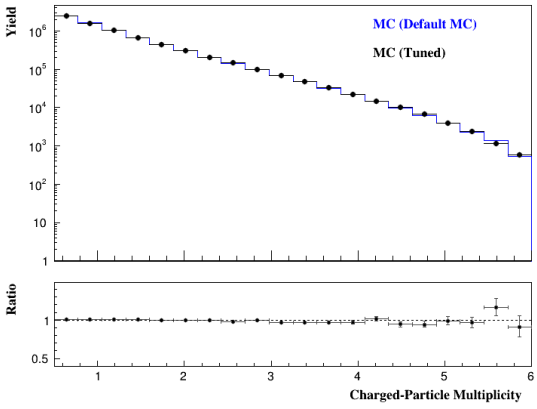


Fig. 6. Comparing Inclusive charge particle momentum distributions: Default Monte Carlo (MC) in blue vs. tuned in black.

The default and tuned values for the Pythia8 parameters.

For three variables, the parameters' default and tune values were quite similar, considering the errors.

Pythia8 parameters	MIGRAD errors
StringFlav:mesonUDvector	=4.685612e-02
StringFlav:mesonSvector	=8.267797e-02
StringFlav:mesonCvector	=1.690820e-01
StringFlav:thetaPS	=7.967455e+00
StringFlav:thetaV	=8.819869e+00
TimeShower:alphaSvalue	=4.741711e-04

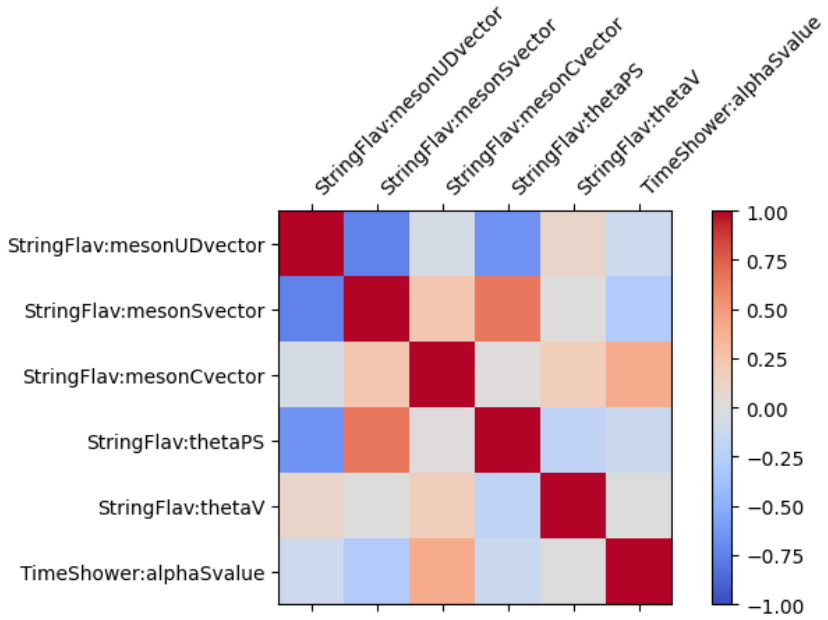


Fig. 7. Parameter correlation values for the second-stage tuning.

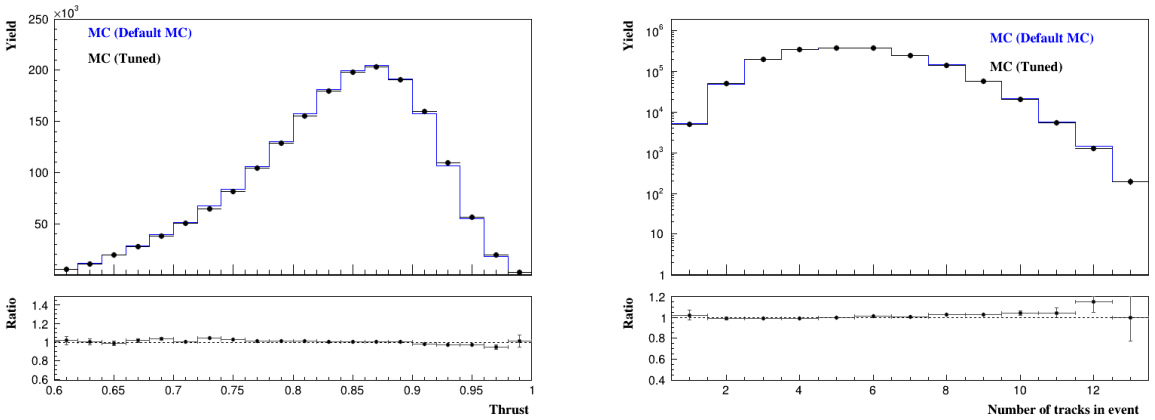
Next, we included the visibleEnergyOfEventCMS variable in our list of observables to study the impact of correlations when tuning with an increased number of variables. This tune was also performed with 5th-order polynomial to ensure the robustness for the minimization results.

As it can be seen from Fig. 8 (b), the comparison between tuned and reference samples for visibleEnergyOfEventsCMS observable is also a very good agreement.

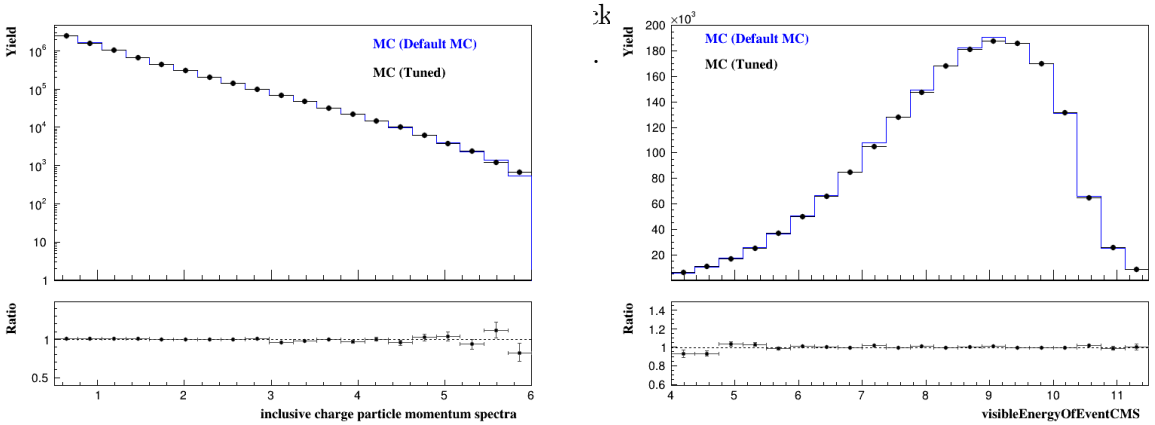
Pythia8 parameters	Default	Tuned	Comment
StringFlav:mesonUDvector	=0.5	=0.454673	! Light-flavour vector suppression
StringFlav:mesonSvector	=0.55	=0.432484	! Strange vector suppression
StringFlav:mesonCvector	=2.8	=2.388490	! Charm vector suppression
StringFlav:thetaPS	=-15	=-14.416715	! Mixing angle *thetaPS*
StringFlav:thetaV	=36	=32.127201	! Mixing angle *thetaV*
TimeShower:alphaSvalue	=0.1365	=0.135604	! Effective alphaS(mZ) value

The Default and tuned values for Pythia8 parameters. For four variables, the parameters default and tune values were quite similar, considering the errors.

Pythia8 parameters	MIGRAD errors
StringFlav:mesonUDvector	=4.270846e-02
StringFlav:mesonSvector	=7.325382e-02
StringFlav:mesonCvector	=1.705115e-01
StringFlav:thetaPS	=6.570061e+00
StringFlav:thetaV	=6.979057e+00
TimeShower:alphaSvalue	=4.257399e-04



(a) (b)



(a) (b)

Fig. 9. Comparing Inclusive charge particle momentum andvisibleEnergyOfEventCMS distributions: Default Monte Carlo (MC) in blue vs. tuned in black.

5. Conclusion

Validation of the tuning procedure developed to tune a certain set of parameters from the Pythia8 Monte Carlo event generator is performed using the Professor2 package. The tuning is done simultaneously for "event" and "event shape" variables using six parameters from Pythia8 MC. The obtained results show that the comparison between the reference and tuned spectra is in a very good agreement. The developed procedure can be extended to any model with free parameters constraining it by using real experimental data.

References

[1] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun.,2008, [Online]. Available: https://pythia.org/.

[2] Andy Buckley, Hendrik Hoeth, Heiko Lacker, Holger Schulz, and Jan Eike von Seggern, *Systematic event generator tuning for the LHC*,arXiv:0907.2973v1,2009, [Online]. Available: https://professor.hepforge.org/.

[3] Dennis Weyland, *Continuum Suppression with Deep Learning techniques for the Belle II Experiment*, 2017.

[4] [Online]. Available: https://wlcg-public.web.cern.ch/

[5] [Online]. Available: https://cernvm.cern.ch/fs/

[6] [Online]. Available: https://kekcc.kek.jp/service/kekcc/support/en/01/

[7] [Online]. Available: https://www.desy.de/

[8] [Online]. Available: https://software.belle2.org/development/sphinx/index.html

[9] [Online]. Available: https://dirac.readthedocs.io

[10] Geoffrey C . FOX and Stephen WOLFRAM,*Event Shapes in e^+e^+ annihilation*, North-Holland Publishing Company, 1978.

Pythia8 Մոնտե Կառլո գեներատորի կարգաբերում՝ Professor2 փաթեթի օգտագործմամբ

Հազարավարդ Մ. Ղումարյան

Ա. Ի. Ալիխանյանի անվան ազգային գիտական լաբորատորիա, Երևան, Հայաստան
e-mail: hazar@yerphi.am

Ամփոփում

Այս հոդվածում ներկայացված է Մոնտե Կառլո դեպքերի գեներատորի բազմապարամետրային, միաժամանակյա Թյունինգի մեթոդը: Այն հաստատվել է Pythia8 Մոնտե Կառլո իրադարձությունների գեներատորի վրա, որը լայնորեն օգտագործվում է բարձր էներգիայի ֆիզիկայում (HEP): Ստացված արդյունքները ցույց են տալիս, որ մեթոդը կարող է օգտագործվել ֆենոմենոլոգիական մոդելների ազատ պարամետրերը կարգաբերելու համար հաշվի առնելով պարամետրերի միջև առկա կոռելացիաները:

Բանալի բառեր՝ Մոնտե Կառլո թյունինգ, հադրոնիզացիա, Pythia8, Belle II գիտափորձ, professor2 փաթեթ:

Проверка настройки Pythia8 MC с использованием Professor2 пакета

Азаравард М. Гумарян

Национальная научная лаборатория имени А.И. Алиханяна, Ереван, Армения
e-mail: hazar@yerphi.am

Аннотация

В данной статье представлен метод для многопараметрической настройки генератора событий по методу Монте-Карло. Метод апробирован на настройке Монте Карло генератора Pythia8, широко используемого для физики высоких энергий (ФВЭ). Полученные результаты показывают, что метод может быть использован для определения свободных параметров феноменологических моделей, одновременно позволяя учесть корреляции существующие между параметрами.

Ключевые слова: Настройка Монте-Карло, адронизация, Pythia8, эксперимент Belle II, пакет профессор2.