ISSN 1068-3372, Journal of Contemporary Physics (Armenian Academy of Sciences), 2021, Vol. 56, No. 2, pp. 85–90. published in Izvestiya Natsional'noi Akademii Nauk Armenii, Fizika, 2021, Vol. 56, No. 2, pp. 136–143.

"Professor2" Package for Tuning the Parameters of Fragmentation Process in e^+e^- Annihilation

V. A. Muradyan[a], *, N. Z. Akopov[a], H. M. Ghumaryan[a], and G. A. Karyan[a]

[a] *Alikhanyan National Science Laboratory, Yerevan, Armenia*

**e-mail: v.muradyan@yerphi.am*

Received March 3, 2021; revised March 19, 2021; accepted March 26, 2021

Abstract—The verification scheme in the package "Professor2" for setting parameters of the most used physics generator PYTHIA8.2 for simulating the processes in high energy physics is described. The goal is to tune the PYTHIA8.2 parameter list for the fragmentation process below $b\bar{b}$ the production threshold. As an initial parameter list, we use the Belle (predecessor of the Belle II experiment) default parameter settings for fragmentation. For the MC pseudo-data, we use the Belle II software package i.e. basf2, and generator script developed by the main author of this paper. To validate the whole procedure, we use momentum distribution of charged pions and kaons to tune two free parameters of the fragmentation model in PYTHIA8 i.e. LUND fragmentation model. We demonstrate that the Professor2 package is a handful to perform multi-parameter tunes simultaneously.

Keywords: Pythia, Monte Carlo, Professor2, fragmentation, simulation

DOI: 10.3103/S1068337221020134

1. INTRODUCTION

The tuning of the MC generators is the most important problem for modern high-energy physics [1, 2]. The Belle II experiment at the KEK accelerator (Japan) [3] plans to collect $50ab^{-1}$ data for physics research beyond the Standard Model [4]. Such studies require not only the availability of large statistics but also an accurate description of the MC data by generators, for adequate subtraction of background events. In this context, the description of the background continuum plays an important role in the study of rare B meson decays.

Before the creation of the "Professor2" [5] package in traditional methods of manual (also coarse) adjustment, there were always problems with finding free parameters in models, especially with an increase in the number of model parameters. The method implemented in the 'Professor2' package is called parameterization-based tuning.

The approach implemented in the "Professor2" package is the parameterization of the generator behavior using polynomials of varying degrees. Then the polynomial is fitted following the generator responses in each bin according to the observed variable from the data array owing to changes in the components of the P-dimensional vector of $p = (p_1, \ldots, p_P)$ parameters. The method then uses the goodness-of-fit test for the fitted function and minimizes it. As a result, a set of the best parameters $p_{(\mathrm{tune})}$ is obtained that describes the experimental data. One of the main advantages of the "Professor2" package is its ability to handle cross-correlations between different parameters, performing multivariate minimization in the PYTHIA [6] parameter space (see Fig. 1), which does not apply to the aforementioned 'manual' and 'coarse' tuning methods.

This method requires a beefy computer. At the same time, the collection of the necessary statistics for a large number of parameters may take several days, and only a few minutes, to find the best parameter. Then you should run the generator with a set of the best parameters and compare the simulation results with experimental data. In the following sections, we will describe in more detail how to use the "Professor2".

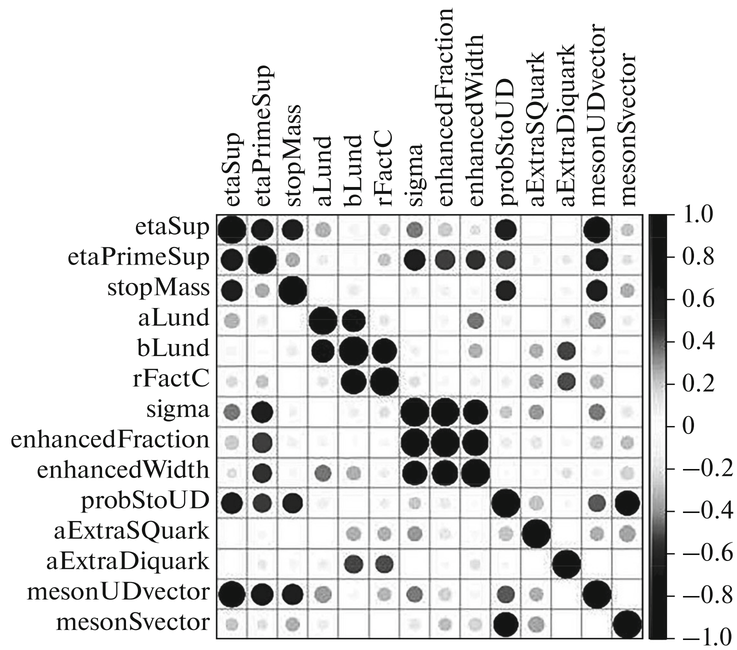

Fig. 1. Correlation between some Pythia8 parameters.

2. DESCRIPTION OF THE "PROFESSOR2" METHOD

The main goal of the method is to determine the correspondence function between the generated and experimental data and then to minimize it. The following describes the steps in general and the implementation of the method for the process $e^+ + e^-$ of annihilation into quarks, where then each of the quarks with some probabilities turns into hadrons (the process of hadronization or fragmentation).

2.1. Generator Response

As mentioned, there is a generator that must be run for many parameter sets. To select tuning parameters, the sensitivity of the parameters is checked first. For this, the parameter is shifted from its nominal (default) value and the effect of the shift on the momentum distributions of formed particles is considered.

As an example of a sensitivity test, Fig. 2 and Fig. 3 show a two-model statistical test performed for parameters *a* (StringZ: aLund) and *b* (StringZ: bLund), showing statistically significant (Fig. 2), not significant (Fig. 3) the effect on the spectra of various hadrons in the final state.

The Normalized Residuals check formula is expressed as:

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

where $\hat{p}_i = (n_i + m_i)/(N + M)$, n_i/m_i is the number of events of the i-th bin in the reference/modified samples, N/M is the total number of events in the reference/modified samples.

The method provides a polynomial match for the content of many generated bins for each set of parameters. The peculiarity of the approach used in this work is that the contents of the bins for the set of parameters are compared not with the real data from the accelerators, but with the generator responses generated with the default parameter values. To select a generator, as mentioned above, we are interested in the fragmentation process in events $e^+ + e^-$, which is studied in the Belle2 experiment at the KEK accelerator, so it is obvious to use the Belle2: basf2 [7] environment (Fig. 4). The basis for the generation is the PYTHIA8 event generator. Fragmentation in PYHTIA8 is implemented based on the LUND [8] string model and is represented as:

$$f(z) = (1/z)(1 - z)^a \exp(-bm_T^2/z), \tag{2}$$

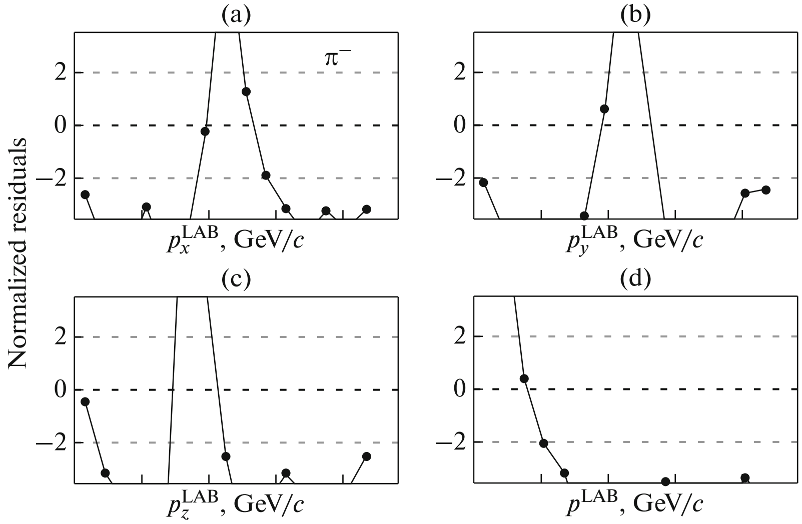

Fig. 2. Sensitivity test to parameters *a* and *b* for the pulse components of negatively charged pions (p_x, p_y, p_z, p,) respectively, (a), (b), (c) and (d).

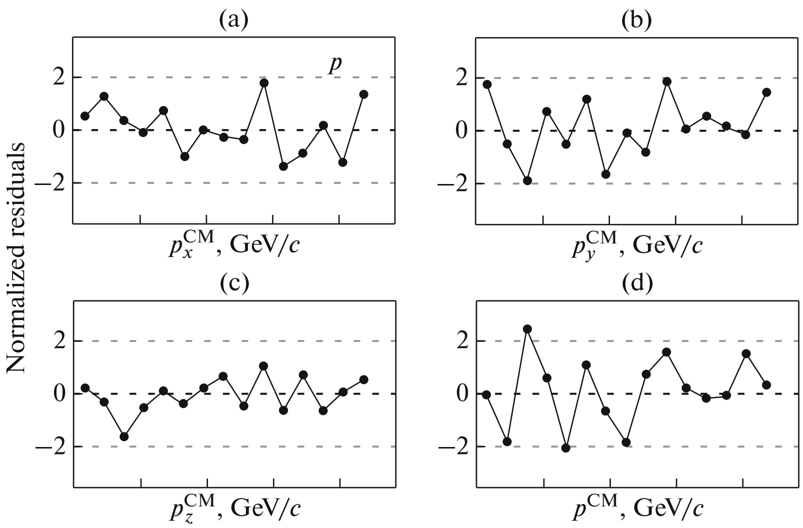

Fig. 3. Sensitivity test to parameters *a* and *b* for the pulse components of negatively charged pions (p_x, p_y, p_z, p,) respectively, (a), (b), (c) and (d).

where *a* and *b* are free parameters varying in the intervals: [0, 2] and [0.2, 2] (with default values in Belle: 0.32, 0.62), m_T is the transverse mass, z is the relative energy of the hadron. The generation program is written in Python and adapted for the environment: Belle2, basf2.

At the first step of the tuning process, one should define a set of parameters with the values of which the spectra of various observables will be generated. For this, the "Professor2" package has a function that allows one to generate the random values of parameters in a given range—prof2—sample (Fig. 5). This function generates 100 (default) parameter sets with different values.

The minimum required number of parameter sets is implemented by the prof2—ncoeffs command (Fig. 6). In our case, with 2 parameters, and, say, in a second-order polynomial, this minimum number is 6 (in fact, in practice, much more is needed than this minimum); in this work, we took 100.

The next step is to start the generator with all the parameters obtained. In this part, we will not use "Professor2", but our code that allows us to generate events: $e^+ + e^- \to q\overline{q}$, for $u\overline{u}$, $d\overline{d}$, $s\overline{s}$, $c\overline{c}$ quark combinations (continuum data). The pseudo-data to be used are saved in ROOT [9] files. After generating the MC events, we will have 4 ROOT files (Fig. 7) for each combination of quarks, and all this is repeated 100 times (according to the number of parameter sets).

Fig. 4. Medium interface basf2.

```
prof2-sample -o output params.dat
```

Fig. 5. This command will generate 100 sets and store them in the "output" file, and in the "params.dat" file the names of the parameters and their maximum and minimum values are given.

```
prof2-ncoeffs 2
```

```
Polynomial order     Minimum samples
        0                   1
        1                   3
        2                   6
        3                  10
        4                  15
        5                  21
        6                  28
        7                  36
        8                  45
        9                  55
       10                  66
```

Fig. 6. The minimum required number of parameter sets using the prof2–ncoeffs command are given.

```
ccbar.root  ddbar.root  ssbar.root  uubar.root
```

Fig. 7. ROOT files.

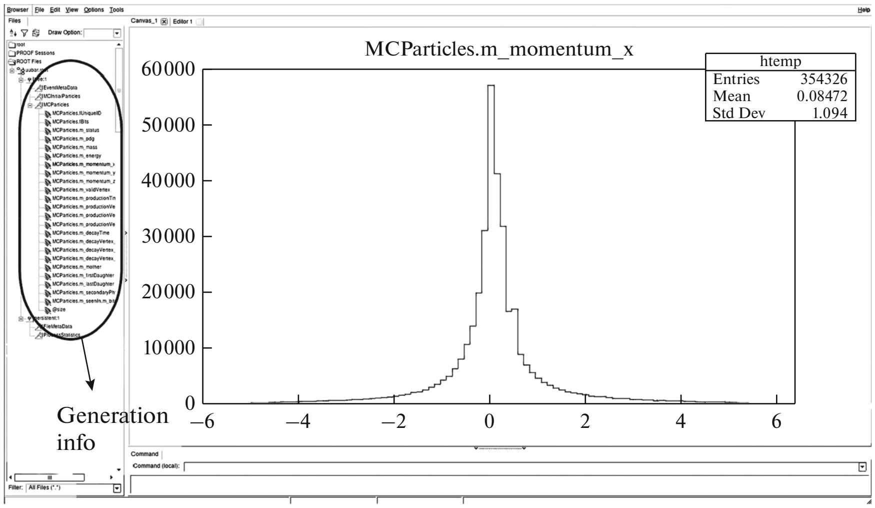

Fig. 8. Distribution over the x-component of the pulse obtained from the sample continuum.

2.2. "Computations"

After determining the content of all bins for each individual set of parameters, we denote the value of the bin content as $MC_b(p)$ (obviously, this will be a function of the parameters), the method uses a second-order polynomial as a basis for parameterization $MC_b(p)$:

$$MC_b(p) \approx \alpha_0 + \sum_i \beta_i^{(b)} p_i' + \sum_{i<j} \gamma_{ij}^{(b)} p_i' p_j', \tag{3}$$

where the shifted parameter vector is: $p' \equiv p - p^0$. For a given polynomial form, one needs to define the coefficients to best describe the behavior of the generator in each bin.

To check the agreement criterion, the "Professor2" computes the functional and minimizes it. The functional χ^2 is determined as:

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

where R_b is the reference value for the bin b, and the error Δ_b is the total uncertainty of the reference value for the bin b, weights ω_{O} for each O observable are also included ("observable", physical quantities: pulse projections, energy, ...). Thus, statistical weights can be assigned to increase (decrease) the significance of various distributions.

3. PROCESS SETUP

Having 100 different sets of parameters, that is, 100 starts of the generator for each combination of quarks, it is necessary to combine the samples obtained as a result of the generation of 4 combinations of quarks to form a continuum of the sample (Fig. 8).

The next step is to obtain the distributions of momentum components ($p_x p_y p_z$) for charged pions and kaons (π^+, π^-, K^+, K^-) from the sample continuum. These distributions are also saved in ROOT files. Each ROOT file contains 12 different distributions in the form of histograms. As a result, there are 1200 different histograms with different parameters.

An important note is that histograms should not contain zero bins or their weights should be zero. Before one can find the best combinations of parameters that will describe the reference data, one must run the command prof2–ipol.

As mentioned above, the generator starts with parameter values: $a = 0.32$ and $b = 0.62$ is used as reference data. In the final step, the prof2–tune command is run to get the specific a and b that describe the

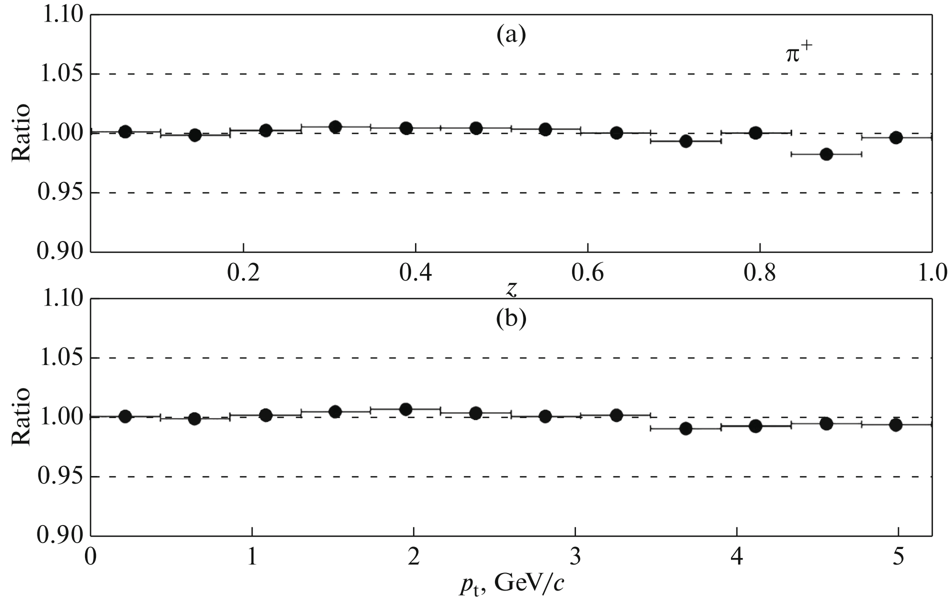

Fig. 9. The ratios of the bin contents of the tuned and reference samples for a positively charged pion as a function of relative energy (top) and transverse momentum (bottom).

reference data. Then the generated samples with the tuned parameters are compared with the reference samples. The comparison shows very good agreement between the reference and tuned samples (Fig. 9), which confirms the correctness of the procedure described in this work.

4. CONCLUSION

The main goal of this article was to verify the correctness of the "Professor2" package used for the fragmentation process of quarks that were produced in collisions. The achievement of this goal is demonstrated in this work. Such verification of the Professor package has not yet been implemented for reaction $e^+ + e^-$. The next step is to test the fragmentation model using experimental data.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Sirunyan, A.M., Tumasyan, A.R., Khachatryan, V.A., et al., *J. Contemp. Phys.*, 2017, vol. 52, p. 311.
2. Sirunyan, A.M., Tumasyan, A.R., Khachatryan, V.A., and Petrosyan, A.G., *J. Contemp. Phys.*, 2016, vol. 51, p. 229.
3. Abe, T. et al., *Belle II Technical Design Report*, e-Print: 1011.0352 [physics.ins-det] (2010).
4. Kou, E. et al., *PTEP*, 2019, vol. 2019, p. 123C01; *PTEP*, 2020, vol. 2020, p. 029201, (erratum).
5. Buckley, A., Hoeth, H., Lacker, H., Schulz, H., and von Seggern, J.E., *The European Physical Journal C*, 2010, vol. 65, p. 331.
6. Sjöstrand, T. et al., *Comput. Phys. Commun.*, 191, 159 (2015).
7. Kuhr, T., Pulvermacher, C., Ritter, M., et al., *Comput Softw Big Sci*, 2019. vol. 3, p. 1.
8. Andersson, B.-G., *Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol.*, 1997, vol. 7, p. 1.
9. Brun, R. and Rademakers, F., *Nucl. Instrum. Meth. A*, 1997, vol. 389 p. 81.

Translated by V. Musakhanyan