DOI: 10.18721/JPM.12310
УДК 53.088.6

AN ANALYSIS OF UNFOLDING METHODS FOR MEASUREMENT OF HADRON DIJET PRODUCTION CROSS SECTIONS

A.Yu. Egorov, Ya.A. Berdnikov, V.A. Bakaev, I.M. Nikittsina

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russian Federation

The comparison results of different methods of detector's distortion elimination have been presented. The following methods were taken: bin-by-bin correction method, migration matrix inversion one, the one of maximal likelihood with Tikhonov regularisation (TUnfold), the one of singular value decomposition of the migration matrix (SVD), the one of D'Agostini iterations. The comparison of selected methods was performed through Monte Carlo simulation of hadron dijet production in proton-proton collisions at center-of-mass energy of 2.76 TeV and the simulation of a response of the CMS detector at Large Hadron Collider. The optimal scheme of unfolding was chosen for the measurement under study. Practical recommendations for building of unfolding procedure were given.

Keywords: detector effect, unfolding, TUnfold, SVD unfolding, D'Agostini iterations

Citation: Egorov A.Yu., Berdnikov Ya.A., Bakaev V.A., Nikittsina I.M., An analysis of unfolding methods for measurement of hadron dijet production cross sections, St. Petersburg Polytechnical State University Journal. Physics and Mathematics. 12 (3) (2019) 113–119. DOI: 10.18721/JPM.12310

АНАЛИЗ МЕТОДОВ ОБРАТНОЙ СВЕРТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ПРИ ИЗМЕРЕНИИ СЕЧЕНИЙ РОЖДЕНИЯ ПАР АДРОННЫХ СТРУЙ

А.Ю. Егоров, Я.А. Бердников, В.А. Бакаев, И.М. Никитцина

Санкт-Петербургский политехнический университет Петра Великого,

Санкт-Петербург, Российская Федерация

Представлены результаты сравнения методов устранения детекторных искажений. Рассмотрены следующие методы: поправочных коэффициентов, обращения матрицы миграции, максимизации функции правдоподобия с регуляризацией Тихонова (TUnfold), сингулярного разложения матрицы миграции (SVD), итераций Д'Агостини. Сравнение выбранных методов осуществлялось на примере обработки результатов измерения сечений рождения пар адронных струй в протон-протонных столкновениях при энергии системы центра масс протонов 2,76 ТэВ на детекторе CMS (Compact Muon Solenoid) на Большом адронном коллайдере. В результате сравнения методов выбрана оптимальная схема устранения детекторных искажений в данном измерении и даны рекомендации по построению процедуры коррекции детекторных искажений.

Ключевые слова: детекторное искажение, обратная свертка, TUnfold, SVD, итерации Д'Агостини

Ссылка при цитировании: Егоров А.Ю., Бердников Я.А., Бакаев В.А., Никитцина И.М. Анализ методов обратной свертки экспериментальных данных при измерении сечений рождения пар адронных струй // Научно-технические ведомости СПбГПУ. Физико-математические науки. 2019. Т. 12. № 3. С. 123–130. DOI: 10.18721/JPM.12310

Introduction

A physical device (e.g., a detector) measuring a physical quantity often presented as a distribution or a histogram inevitably introduces distortions related to finite resolution, reconstruction efficiency and systematic effects. Detector effects is a general term describing all possible distortions introduced during measurements.

The detectors currently used in high-energy physics are multi-level systems that include thousands of sensors, complex electronics and millions of channels. Examples of such detectors are ATLAS (A Toroidal LHC ApparatuS) [1] and CMS (Compact Muon Solenoid) [2] experiments at the Large Hadron Collider. Detector effects are in these experiments result from a large number of stochastic processes and can lead to significant distortions.

Unfolding is the common name for a class of procedures aimed at correcting for the distortions introduced during measurements.

The measured distribution distorted by the detector is called the reconstructed data. On the other hand, the distribution that could be generated by an ideal detector is called the true distribution. The true distributions for the experimental data are unknown. Unfolding is intended to restore true distributions based on reconstructed data. Unfolding is necessary during analysis of reconstructed data for obtaining measurement results that are independent of the details of the experiment and can be used outside the laboratory.

There are different unfolding methods, each with their own advantages and limitations. The specific method can be chosen depending on whether it is applicable to the given measurement task. The reason for this is that mathematically speaking, unfolding is essentially an ill-posed problem. A unique solution cannot be found. Moreover, the problem is often unstable, i.e., small fluctuations in the reconstructed data can lead to large fluctuations in the reconstructed distribution. It is especially important for the solution to be stable because the measured distribution has statistical fluctuations that can be amplified during unfolding. Different regularization techniques can be applied to stabilize the solution.

Even though unfolding is a mathematically ill-posed problem and the optimal method depends heavily on the specific measurements, it is still possible to develop an algorithm for analyzing the applicability of different unfolding methods to the given measurements, so that an optimal unfolding scheme can be constructed.

We have analyzed the applicability of different unfolding methods to measuring the cross-sections for dijet production in proton collisions using a CMS detector at the LHC [2] at center-of-mass energy of $\sqrt{s} = 2.76$ TeV.

The goal of this study consisted in constructing an optimal unfolding scheme for these specific measurements. An additional goal was to develop practical recommendations for analyzing the applicability of different unfolding methods in measurements.

Measured physical quantities

The accuracy of different methods was assessed by measuring the differential cross sections for dijet production

$$\sigma^{incl},\ \sigma^{\mathrm{MN}},\ \sigma^{excl} \text{ and } \sigma^{excl\ veto},$$

where σ^{incl} is the inclusive cross section for dijet production, σ^{MN} is the cross section for the production of Mueller–Navelet dijets, σ^{excl} is the cross section for the production of “exclusive” dijets, $\sigma^{excl\ veto}$ is the “exclusive” cross section with a jet veto (their characteristics are given below).

The given differential cross sections are measured as functions of rapidity separation between the jets

$$\Delta y = |y_1 - y_2|,$$

where y_1, y_2 are the rapidities of the first and second jet in the dijet, and y is the rapidity expressed as

$$y = (1/2)\ \ln[(E + p_z)/(E - p_z)].$$

Here E is the energy of the jet, p_z is the momentum along the beam of colliding hadrons.

A collision of protons producing two or more jets with a momentum $p_\perp$ transverse to the beam exceeding the threshold of 35 GeV/c is called an inclusive event. All pairwise combinations of hadron jets with a transverse momentum $p_\perp \geq 35$ GeV/c in an inclusive event contribute to the inclusive production cross section for hadron jets σ^{incl}.

The cross section for the production of Muller–Navelet (MN) dijet describes a pair consisting of a jet with the maximum rapidity and a jet with the minimum rapidity in an inclusive event among jets with the transverse momentum $p_\perp \geq 35$ GeV/c. Thus, MN events belong to the inclusive subclass.

A subclass of inclusive events where strictly two jets with the transverse momentum $p_\perp \geq 35$ GeV/c are produced is called exclusive (σ^{excl} is the production cross section). A pair of jets produced in an exclusive event is called an exclusive dijet.

Exclusive vetoed events are a subclass of exclusive events where jets additional to the exclusive dijets, with the transverse momentum $p_{\perp} \geq p_{\perp\ veto} = 20$ GeV/*c* ($s^{excl\ veto}$ is the exclusive production cross-section with veto).

The width of the rapidity range Δy depends on kinematic constraints and varies from 0 to 8.0.

The definitions given above allow to assess possible distortions.

Firstly, jets can migrate at transverse momentum thresholds of 35 and 20 GeV/*c* because the resolution of the detector is limited by this parameter. Such migration affects the number of events in different classes. Decreased number of events in a class is interpreted as limited acceptance for this class. Increased number of events in a class is interpreted as the background. These distortions in turn affect the shape and size of the measured cross section.

Secondly, jets can migrate in rapidity because the resolution of the detector is limited by this parameter. As a result of such migration, events from one cell of the true distribution over the rapidity range contribute to other cells of the reconstructed distribution. This affects the shape of the measured cross section.

Unfolding procedure

Hadron collision samples obtained with Monte Carlo (MC) generators are used for unfolding. The distributions obtained from MC collision samples before detector simulation correspond to the true distributions. These MC distributions are also called generator distributions. Reconstructed MC distributions are obtained after detector simulation.

MC simulation allows to estimate the loss of events due to efficiency and limited acceptance of the detector, the contribution from background events (events simulating the signal), and the migration of contributions between histogram cells due to limited resolution of the detector. Migration of events between histogram cells is described by a migration matrix. The distribution at the generator level, the reconstructed MC distribution and the migration matrix are the input data necessary for performing the unfolding procedure.

We used two MC models to analyze the accuracy of different unfolding methods: PYTHIA8 (8135) [3] with Tune 4C [4] and HERWIG++ (2.7.1) [5] with Tune UE-EE-3C [6]. Generator events are processed with a CMS detector model built in the GEANT4 software package [7]. By using two different MC models, we can make a cross-check, i.e., unfolding the simulation results of the first MC generator with the second one. The results obtained using the first generator act as the reconstructed data. Cross-checking allows to compare the result of unfolding with the generator distribution. The accuracy with which the generator distribution is reproduce depends both on the unfolding algorithm and on the adequacy of the physical models embedded in the generators.

The following convolution methods were compared in this study:

bin-by-bin correction factors;

inversions of migration matrix;

maximizing the likelihood function using Tikhonov's regularization (TUnfold) [8];

singular value decomposition of the migration matrix (SVD) [9]

D'Agostini iterations [10].

Regularization can be used with the TUnfold, SVD, and D'Agostini iteration methods.

Regularization type (by absolute value, regularization of the first and second derivative) and the value of the regularization parameter τ can be chosen in the TUnfold method. The optimal value of this parameter can be selected automatically by minimizing the global correlation $\rho_{\max}$ or using the *l*-curve method [8].

Regularization by the SVD method is done by discarding some singular values of the migration matrix. The singular values to be discarded can also be done automatically using the parameter d_i [9].

Finally, limiting the number of iterations acts as regularization in the D'Agostini method.

Notably, increasing the size of histogram cells can also serve as regularization of the problem. However, this can introduce a bias into the unfolded distribution compared to the generator level. Performing unfolding in this study, we considered the accuracy with which the generator distribution is reproduced depending on the cell size selected.

There are different methods for including background events (i.e., events mistakenly identified as signal) and reconstruction efficiency (acceptance).

The first method is called expanding the migration matrix. Background events and events that were not detected due to limited efficiency and acceptance are added to additional rows and columns of the migration matrix in this method.

The second method involves subtracting the background. Acceptance is taken into account using correction coefficients. This method is used for uncorrelated background, when background events occur due to independent processes with the final state imitating a signal.

The third method involves correction factors for taking into account background and acceptance. It is used when background and acceptance are correlated with the signal.

Results and discussion

As noted above, we have analyzed the applicability of unfolding for different methods.

Different values of regularization parameters were set manually and using methods for optimal value search. Each of the methods for taking into account background and acceptance was used for each of the unfolding methods.

We have obtained cross-checked results for all cross section and all methods. The proportion of background events increases with extending rapidity range from 40% at $\Delta y = 0$ to 90% at $\Delta y = 8$. The proportion of events not included as a result of limited efficiency and acceptance is from 20% at $\Delta y = 0$ to 40 % at $\Delta y = 8$. Event migration between the cells in the rapidity range Δy does not exceed 10 %, remaining almost constant in the entire rapidity range. Events mostly migrate to neighboring cells, which makes the migration matrix almost diagonal.

Cross-checking indicates that introducing correction factors for taking into account background and acceptance is the only method allowing to reconstruct the cross sections at the generator level. Other methods for taking into account background and acceptance generate a strongly biased unfolded distribution. Background and acceptance in the given measurement appear due to the migration of hadron jets relative to the transverse momentum thresholds, equal to 20 and 35 GeV/*c*. Thus, background and acceptance depend on the number of events in the signal and are, therefore, correlated with the signal level.

Unfolding methods yield the same results

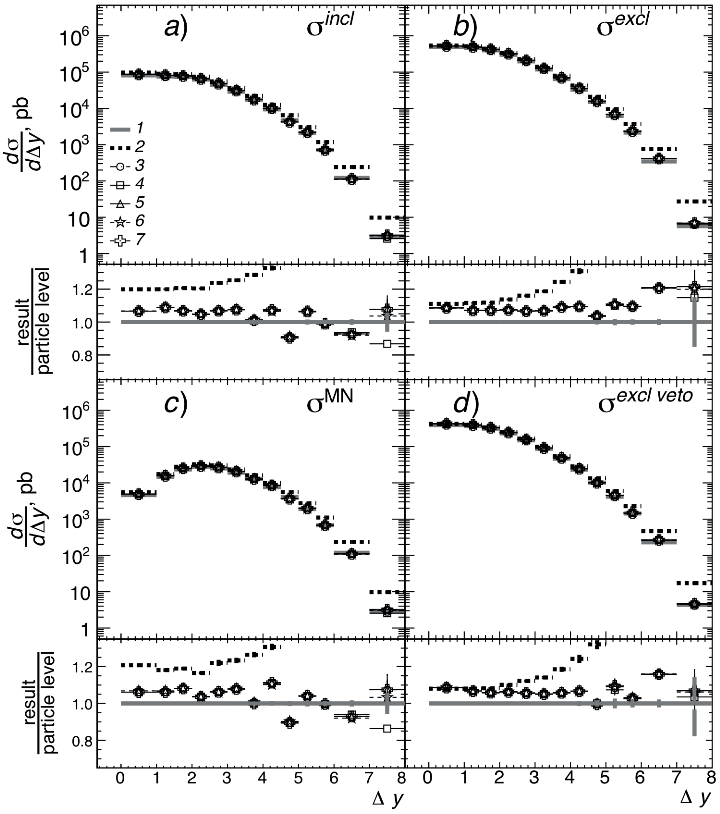

Fig. 1. Cross-check for differential cross sections σ^{incl} (*a*), σ^{excl} (*b*), σ^{MN} (*c*) and $\sigma^{excl\ veto}$ (*d*).
Cross sections (*2*) at the detector level were obtained using the HERWIG++ generator. The results obtained using TUnfold (*3*), SVD (*4*), D'Agostini iterative method (*5*), correction factors (*6*), migration matrix inversion (*7*) were compared with the cross section for particle level obtained with HERWIG++ (*1*). Unfolding was performed using the PYTHIA8 generator. The ratios of the unfolding results to the particle-level cross sections (result/particle level), i.e., at the generator level, are shown.

after background and acceptance have been taken into account by the method of correction factors. Using different regularization parameters, we found that the optimal solution has either minimal or no regularization, which follows from the fact that the migration matrix is close to diagonal.

Fig. 1 shows the cross-checking results for different unfolding methods with correction of background and acceptance using correction factors and optimally selected regularization parameters.

As follows from the figure, unfolding reproduces the distribution at the generator level with an accuracy of 20%. Analyzing the accuracy of unfolding as a function of the size of histogram cells, we can see that the result does not depend on the cell size chosen. Additional studies indicate that unfolding results deviate from the generator distribution because the detector model has different responses to hadron jets obtained using HERWIG++ and PYTHIA8 generators. Apparently, the reconstructed experimental data cannot be unfolded with a single MC generator; two should be used, and the difference between the results can serve as an estimate of the systematic error resulting from the specific MC model chosen.

The analysis carried out led us to choose the TUnfold method. Acceptance and background are taken into account using correction factors. The method was selected because its implementation in the ROOT framework [11] seems better developed than implementations of other methods; it is well documented and has the greatest flexibility in adjusting and controlling the unfolding process.

Practical recommendations

Unfolding should be applied after calibrating the detector and reconstructing all objects necessary for analysis (i.e., charged particle trajectories, calorimeter towers, particles, jets). Additional corrections, for example, for energy and resolution can be introduced (preferably, before unfolding) in experiments even after reconstructing the objects. Phase space and selections should be determined and applied identically for experimental data and MC simulation. Since the unfolding problem is ill-conditioned from a mathematical standpoint, the applicability of different unfolding methods should be checked. The approaches described in this study are not the only ones currently available.

After completing all preparatory operations, namely,

reconstructing and correcting data;
running MC simulation in several models;
reconstructing MC objects;
selecting the experimental data and MC models;
selecting the unfolding methods, methods for including background, efficiency and acceptance

the applicability of different methods is analyzed in the following order.

1. Select a physical quantity for applying unfolding (not necessarily coinciding with the measured physical quantity).
2. Select the size of histogram cells.
3. Construct distribution at the generator level, the reconstructed level and the migration matrix for MC models.
4. Test different unfolding methods using cross-checking with MC samples of events.
5. Choose best method.
6. Analyze residual bias of the unfolding results relative to the generator level.
7. Determine the effect of the histogram cell size selected on the unfolding results. If a new cell size is selected, the procedure is repeated for this new size. If the residual bias is due to the MC models used for unfolding, the residual bias is used to estimate the model-dependent systematic uncertainty of unfolding.
8. Perform unfolding using all MC models.
9. Calculate (using different MC models) the reconstructed distribution, statistical and systematic uncertainties based on unfolding results.

Conclusion

Analyzing different unfolding methods for measuring the cross-sections for dijet production in proton-proton collisions at $\sqrt{s} = 2.76$ TeV using the CMS detector, we have selected the optimal unfolding scheme for processing the experimental data obtained by the given measurements. We have provided practical recommendations for constructing an algorithm for analyzing the applicability of different unfolding methods.

The results of this study were obtained within the framework of State Task of the Ministry of Education and Science of Russian Federation 3.1498.2017/4.6.

REFERENCES

1. **Aad G., Abat E., Abdallah J., et al.** (The ATLAS Collaboration), The ATLAS experiment at the CERN Large Hadron Collider, Journal of Instrumentation. 3 (August) (2008) S08003.

2. **Chatrchyan S., Hmayakyan G., Khachatryan V., et al.** (The CMS Collaboration), The CMS experiment at the CERN LHC, Journal of Instrumentation. 3 (August) (2008) S08004.

3. **Sjöstrand T., Mrenna S., Skands P.**, A brief introduction to PYTHIA 8.1, Computer Physics Communications. 178 (11) (2008) 852– 867.

4. **Corke R., Sjöstrand T.**, Interleaved parton showers and tuning prospects, Journal of High Energy Physics. 2011 (3) (2011) 32, 1–51.

5. **Вдhr M., Gieseke S., Gigg M. A. et al.**, Herwig++ physics and manual, The European Physical Journal C. 58 (4) (2008) 639–707.

6. **Gieseke S., Grellscheid D., Hamilton K. et al.**, Herwig++ 2.5 release note, arXiv preprint arXiv:1102.1672 (2011).

7. **Allison J., Amako K., Apostolakis J. et al.**, Recent developments in Geant4, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 835 (November) (2016) 186–225.

8. **Schmitt S.**, TUnfold, an algorithm for correcting migration effects in high energy physics, Journal of Instrumentation. 7 (10) (2012) T10003.

9. **Hoecker A., Kartvelishvili V.**, SVD approach to data unfolding, Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 372 (3) (1996) 469–481.

10. **D'Agostini G.**, Improved iterative Bayesian unfolding, arXiv preprint arXiv: 1010.0632. (2010).

11. **Antcheva I., Ballintijn M., Bellenot B. et al.**, ROOT – a C++ framework for petabyte data storage, statistical analysis and visualization, Computer Physics Communications. 180 (12) (2009) 2499–2512.

Received 27.03.2019, accepted 25.04.2019.

THE AUTHORS

EGOROV Anatoliy Yu.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
egorov.a@spbstu.ru

BERDNIKOV Yaroslav A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
berdnikov@spbstu.ru

BAKAEV Victor A.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
v_bakaev@phmf.spbstu.ru

NIKITTSINA Inna M.
Peter the Great St. Petersburg Polytechnic University
29 Politechnicheskaya St., St. Petersburg, 195251, Russian Federation
nikitsinna@yandex.ru

СПИСОК ЛИТЕРАТУРЫ

1. **Aad G., Abat E., Abdallah J., et al.** (The ATLAS Collaboration). The ATLAS experiment at the CERN Large Hadron Collider //Journal of Instrumentation. 2008. Vol. 3. August. P. S08003.

2. **Chatrchyan S., Hmayakyan G., Khachatryan V., et al.** (The CMS Collaboration). The CMS experiment at the CERN LHC // Journal of Instrumentation. 2008. Vol. 3. August. P. S08004.

3. **Sjöstrand T., Mrenna S., Skands P**. A brief introduction to PYTHIA 8.1 // Computer

Physics Communications. 2008. Vol. 178. No. 11. Pp. 852–867.

4. **Corke R., Sjцstrand T.** Interleaved parton showers and tuning prospects // Journal of High Energy Physics. 2011. Vol. 2011. No. 3. P. 32. Pp. 1–51.

5. **Вдhr M., Gieseke S., Gigg M. A. et al.** Herwig++ physics and manual // The European Physical Journal C. 2008. Vol. 58. No. 4. Pp. 639–707.

6. **Gieseke S., Grellscheid D., Hamilton K. et al.** Herwig++ 2.5 release note //arXiv preprint arXiv:1102.1672. 2011.

7. **Allison J., Amako K., Apostolakis J. et al.** Recent developments in Geant4 //Nuclear Instruments and Methods in Physics Research. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2016. Vol. 835. November. Pp. 186–225.

8. **Schmitt S.** TUnfold, an algorithm for correcting migration effects in high energy physics // Journal of Instrumentation. 2012. Vol. 7. No. 10. P. T10003.

9. **Hoecker A., Kartvelishvili V.** SVD approach to data unfolding //Nuclear Instruments and Methods in Physics Research. Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 1996. Vol. 372. No. 3. Pp. 469–481.

10. **D'Agostini G.** Improved iterative Bayesian unfolding // arXiv preprint arXiv:1010.0632. 2010.

11. **Antcheva I., Ballintijn M., Bellenot B. et al.** Root – a C++ framework for petabyte data storage, statistical analysis and visualization // Computer Physics Communications. 2009. Vol. 180. No. 12. Pp. 2499–2512.

Статья поступила в редакцию 27.03.2019, принята к публикации 25.04.2019.

СВЕДЕНИЯ ОБ АВТОРАХ

ЕГОРОВ Анатолий Юрьевич – *ассистент кафедры экспериментальной ядерной физики Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
egorov.a@spbstu.ru

БЕРДНИКОВ Ярослав Александрович – *доктор физико-математических наук, профессор, заведующий кафедрой экспериментальной ядерной физики Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
berdnikov@spbstu.ru

БАКАЕВ Виктор Анатольевич – *кандидат физико-математических наук, доцент кафедры экспериментальной ядерной физики Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
v_bakaev@phmf.spbstu.ru

НИКИТЦИНА Инна Михайловна – *магистрант кафедры экспериментальной ядерной физики Санкт-Петербургского политехнического университета Петра Великого.*

195251, Российская Федерация, г. Санкт-Петербург, Политехническая ул., 29
nikitsinna@yandex.ru