

PAPER • OPEN ACCESS

Reconstruction of particle's energy spectrum in experiment with Unfolding technique

To cite this article: V V Alekseev *et al* 2019 *J. Phys.: Conf. Ser.* **1390** 012071

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Reconstruction of particle's energy spectrum in experiment with Unfolding technique

V V Alekseev¹, Yu V Bogomolov¹ O A Dunaeva^{1,3}, A G Mayorov²,
V V Malakhov² and S A Rodenko²

¹ Yaroslavl State P.G. Demidov University, 150000 Sovetskaya st., 14, Yaroslavl, Russia

² National Research Nuclear University MEPhI, 115449 Kashirskoe shosse 31, Moscow, Russia

E-mail: olaydy@gmail.com

Abstract. In elementary particle physics, the characteristics of the events can be distorted to varying degrees due to instrumental effects, which leads to inaccuracies in experimental distributions. To solve such difficulties and restore distributions close to the true, a various mathematical approaches are used. In this paper, we examine the quality of unfolding methods for solving this problem. Different variants of unfolding are tested for the quality of restoring the true distribution by simulating the passage of cosmic ray particles through the PAMELA magnetic spectrometer. The following algorithms were analyzed: D'Agostini, SVD and TUnfold.

1. Introduction

Reconstruction of the true particle characteristics and their distributions is one of the serious problems in data processing of the particle physics experiments. The uncertainties, noises and other instrumental effects lead to the particle characteristics such as velocity, rigidity, energy etc to be subjected to different distortions. Today this kind of problems is solved with the unfolding technique which allows to obtain the true distribution of a characteristic from the measured one based on the known migration matrix [9, 2]. The migration matrix relates to the probability distribution of the "true" characteristic value to be measured as a different value and can be obtained with the numerical simulation of scientific equipment response. Unfolding technique has a many realizations based on the various mathematical approaches each of which has its advantages and drawbacks. In the presented work they were tested for stability against varying the data sample and the corresponding algorithm parameters. The data sample has been taken from the PAMELA experiment [8] simulation dataset, where the propagation of a particles through the PAMELA instrument had been performed and processed into standard format approved in the collaboration and used in the experimental data processing.

2. Unfolding methods

In this section the main targets of unfolding procedure, its known problems and the common solutions are reviewed. A law, describing dependence of one physical characteristic value on another, often is a continuous function, but when it is measured with the physical device due to naturally limited resolution and uncertainty of the measured value, due to a limited statistics an

³ To whom any correspondence should be addressed.



analytical form of the law is represented as a discrete distribution. In this analysis the measuring physical characteristic is a cosmic ray particles flux spectra which is represented as the particle flux distribution over its rigidity or energy. The accuracy of measured energy distribution (flux) is depends on the events statistic accumulated in each energy interval, and therefore how the whole range of the measured characteristic is split into intervals. But, splitting the range into intervals results in inevitable errors: the measured characteristic value in this case due to its natural uncertainty can be attributed to the wrong interval. If the original "true" law is not uniform and intervals thus have different "weights" the more "weighted" intervals will tend to accumulate more wrongly attributed event, which in turn will lead in distortion of the measured distribution. The unfolding algorithm is served to correct the measured distribution to get desired "true" one. Lets assume that we have a discrete distribution over a physical characteristic which is split into N intervals which we will call further bins: $\Delta = (\Delta_1, \Delta_2, \dots, \Delta_N)$. Let $\tau = (\tau_1, \tau_2, \dots, \tau_N)$ be the unknown true probabilities of the variable value to lie within the corresponding bin Δ and $m = (m_1, m_2, \dots, m_N)$ be the experimental number of events found in the corresponding bin. Later τ will be referred as the discretisation of a continuous true spectrum or simply true distribution. Experimentally measured frequencies of variable values to be within each bin. Uncertainties (migrations) are described as the response (migration) matrix $R = (r_{ij})$, where r_{ij} is the probability to find the variable value in j -th bin as long as its true value is in the i -th bin. The expected number of events in each bin for the observed variable $\nu = (\nu_1, \nu_2, \dots, \nu_N)$ can be calculated in the following way: $\nu_i = \sum_j R_{ij}\tau_j$ or $\nu = R\tau$. The unfolding purpose is to restore the true distribution τ with all the features basing on the measured data μ and the known migration matrix R .

The direct unfolding method is based on assumption that the distribution over the number of events in bins obeys the Poisson statistics[2]. The probability to find μ_i events in i -th bin is expressed as $P(m_i, \nu_i) = e^{-\nu_i} \nu_i^{\mu_i} / \mu_i!$ (here $\nu = R\tau$). The desired distribution τ can be estimated by the maximum likelihood method as $\hat{\tau} = R^{-1}\mu$. It is equivalent to the naive idea to solve the equation $\mu = R\tau$ by inverting the matrix R . This method gives the consistent and unbiased estimation of the τ but with sufficiently high uncertainties[3], turned out to be very sensitive to the measured characteristic features and eliminate the possibility to use it in practice[9, 2].

Therefore one may build the biased estimation of τ or to use other methods of unfolding. The most popular approaches are based on the following ideas:

- (i) Tikhonov's regularisation and its modifications (SVD method[6] by Kartvelishvili and Hoecker, regularisation based on the entropy[3]),
- (ii) bayesian methods (iterative[4] and improved iterative[5] methods by D'Agostini, methods by Kuusela and Panaretos[7]),
- (iii) bin-by-bin correction factors and systematic biases[3] (model dependent).

This article contains the analysis of three unfolding algorithms in application to the simulation data of the PAMELA magnetic spectrometer.

Bayesian unfolding

This method has been developed by G.D'Agostini. True particle characteristics values are seen as causes, and their detections in different bins - as effects. The migration matrix is interpreted as the set of conditional probabilities to get the effects due to considered causes. The Bayes theorem gives the possibility to estimate the unknown posterior probabilities for values to be found in each bin of the distribution of the measured values. This approach requires the unidentified true prior probabilities for the events distribution. An initial "true" distribution should be chosen as the first step of the successive approximation to achieve the true distribution. The important part of the work is to analyze the unfolding quality as function of the number of iterations.

SVD unfolding

The basic idea of this method consists in solving the unfolding equation ($\mu = R\tau$) as the minimization problem $L(\tau) = (R\tau - \mu)^T(R\tau - \mu) \rightarrow \min_{\tau}$, adding a special regularization term to this function and solving the resulting linear system with singular value decomposition of the extended matrix. The regularization term $S(\tau)$ may provide continuity or smoothness of the desired distribution:

$$L(\tau) = (R\tau - \mu)^T(R\tau - \mu) + \alpha S(\tau) \rightarrow \min_{\tau}. \quad (1)$$

Authors of SVD method (V.Kartvelishvili, A.Hoecker) propose to normalise equations of the system, extend this system (the matrix of this system is based on the normalised migration matrix and the matrix corresponding on the regularisation term), perform the singular value decomposition of the extended matrix and cut-off small singular values before the solving the system. Also authors propose to choose the regularisation coefficient as the square of the “last big” singular value of matrix decomposition (hereinafter the number of this value is referred as k) but it can be choosed some other way. So we have to investigate the dependance of unfolding quality on the regularisation coefficient α and binnig features.

TUnfold procedure

This algorithm is also based on the regularisation approach performs the choosing of the regularisation coefficient values by L-curve analysis. The TUnfold algorithm (S.Schmitt) is implemented in ROOT package (TUnfold method) and is regarded as the basic method of unfolding in data unfolding in high energy physics problems. Because the optimal regularisation parameters are choosed automatically we can only consider the impact of binning features and responce matrix structure on the unfolding quality.

3. Data base and results

The considered unfolding methods were tested on the Monte-Carlo numerical simulation of the particles propagation through the PAMELA magnetic spectrometer [8]. One of the main objective of this instrument is in precise measurements of the cosmic ray (CR) energy spectrum (distribution). For this purpose the PAMELA consists of special detector systems allows to identify the type of particles and their general characteristics. The basic detector within PAMELA is a tracking system which measures the particle rigidity (it can be converted to the kinetic energy because of known particle type). Due to the finite spatial resolution of the detector and available physical processes for the particle, the measured rigidity is differ from the true value. As a result the reconstructed energy spectrum will be distorted and does not corresponds to a true energy distribution. In this case the unfolding methods might reconstruct the true energy spectrum.

Several unfolding techniques were considered above and here we evaluate the quality of them. The simulation gives a nice possibility because of for each particle the true and measured energies are known and comparison of the true energy distribution with reconstructed will be good probe for these methods, it will be possible to note the advantages and disadvantages of each of them.

In a general case we need two data sets: one is for migration matrix construction and the second for the tests and analysis. This is doing because of the true spectrum in the real life is unknown (assumedly it's has a power low energy shape with -2.7 degree) and migration matrix created a priori according to other data. In this work spectra with different shapes for train and tests was used. We construct the migration matrix based on the power low energy spectrum with -2 degree and use it in the different unfolding method to reconstruct the power low energy distribution with -2.7 degree.

Before testing methods the energy range have to be splited to intervals (bins). The considered energy range [0.5GV, 1.55GV] has been splitted to 10 bins. In real task it's no way to limit



Figure 1. Migration matrix for wide bins (left) and small (right).

energy range, and particles with rigidity out of range can be detected in considered energy range and effects to reconstructed distribution. Additional bins (left and right one) have been used to account for this effect. The addition bins are most affected so after analysis its can be ignored as noisy. For some methods there is recommended to use diagonal response matrix ($R_{ii} > 0.5$) [2]. So we tried to make a binning for get diagonal response matrix like in left side of figure 1 (wide binning). Another binning has been made for testing binning stability of methods (“small” bins).

The considered methods are implemented in the following packages: ROOT (TUnfold), RooUnfold 1.1.1 (TUnfold, SVD, Bayes). Bayes method has the parameter — the number of iterations. However in the articles is not theoretical advices how to choose this parameter. There is the same problem in SVD method to choose a regularization parameter. Strong regularisation (large α value) corresponds to strong smoothing, and small value of α equivalent to naive method.

Thus, issues for which testing has made:

- Possibility to use an another event distribution for migration matrix construction
- Influence of binning on the unfolding quality
- Resistance to method parameter selection, how to choose it

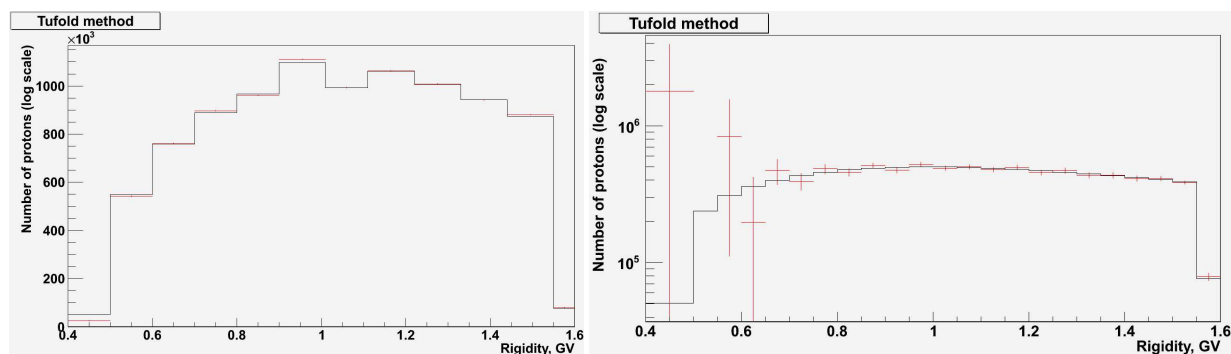


Figure 2. TUnfold method results for wide (left) and small (right) bins.

Lets look to the unfolding methods results, the true distribution is marked by black colour, and other colours are used for reconstructed distributions. In figure 2 the TUnfold method results is shown. There are present all bins, additional bins have been excluded from others methods for better presentation. TUnfold gives satisfactory results for wide bins but shows unacceptable behavior for small bins.

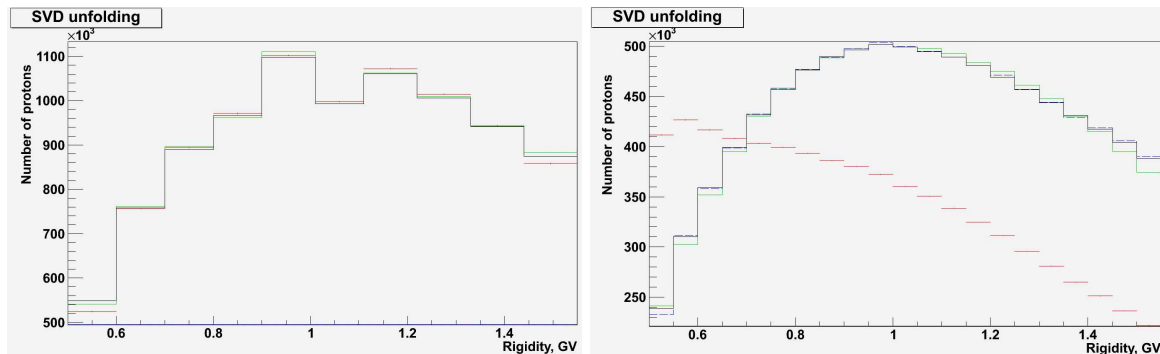


Figure 3. SVD method results for wide (left) and small (right) bins.

Figure 3 presents the SVD method results, for small bins SVD has been tested for $k = 2$ (red), $k = 5$ (green), and maximum $k = 12$ (blue). Notice that this values correspond to strong, medium and small regularisation coefficients α respectively. Strong regularisation is smoothing much the distribution and makes it like distribution which used for migration matrix construction (red one). SVD with small regularisation has been given better results. SVD method for wide bins has been tested for $k = 5$ (red), $k = 12$ (green) and it shows good results (for $k = 2$ the behavior is the same as for small bins).

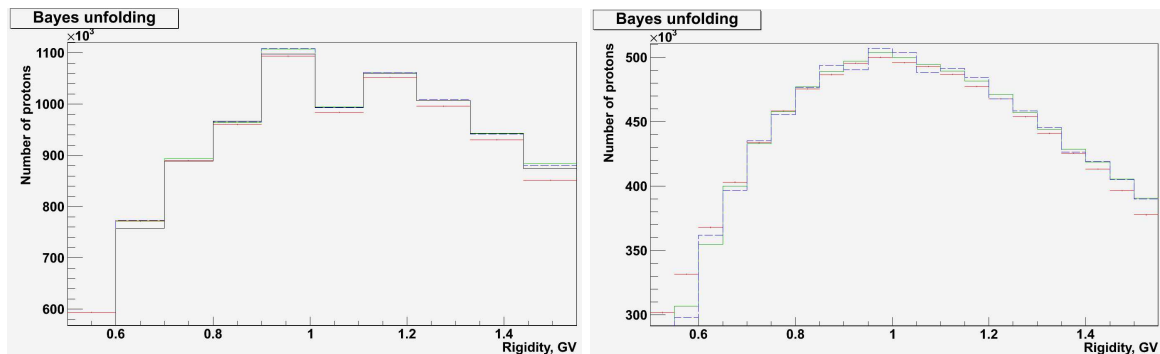


Figure 4. Bayes method results for wide (left) and small (right) bins.

Bayes method has been tested for iterations number $n = 1$ (red), $n = 5$ (green), and $n = 100$ (blue) 4. Results after first iteration have a problems in few first and a little bit in last bins, for wide bins no other problems.

For small bins with the iterations number increase the quality becomes better for some bins, same time for some bins it goes down. It is expected that with the iterations number increase the quality will not be worse. Thus, for small bins iterative Bayes method shows not stable results and choosing number of iterations is a critical.

So, each method and its parameters has been discussed, but the quality of methods need to be compared. In figure 5 the distribution of relative error are shown for all unfolding methods. For Bayes method (red colour) iterations number has been choosen as 5, and for SVD method (green) parameter k has been choosen as 5. For wide bins no way to choose best method, SVD method (green colour) gives bigger error in few last bins, Bayes (red one) — in second bin, and TUnfold method (blue) in first bin has too much big error (0.512). For small bins TUnfold result is unacceptable, and SVD method shows better quality than Bayes method which has bigger error in few first bins. Remind that for accurate conclusions about method quality the mathematical metric is needed.

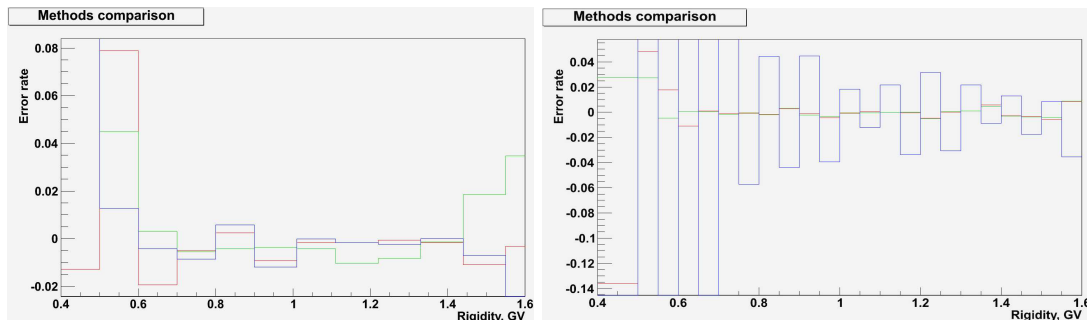


Figure 5. The error rate for all unfolding methods for wide (left) and small (right) bins.

Table 1. Methods comparison.

	TUnfold	SVD	Bayesian
resistance to parameter	Yes	Partly	No
resistance to binning	No	Yes	Partly
resistance to distribution type	Yes, for good binning	Yes, except small k	Yes

4. Conclusion

The comparison of the different unfolding procedures is shown in the table 1. The Bayesian method shows the perceptible dependence on iterations number and binning. SVD method has a regularization coefficient choosing problem. It's known that a big regularization corresponds to the much smoothing of data and the method will reconstruct the distribution which was used for the migration matrix construction. TUnfold method is sensitive to a binning also. Thus, our tests shows that the unfolding procedures can be used in the applications only in the case of “good binning”: as a result of the binning the migration matrix R should be diagonal with $R_{ii} > 0.5$.

Acknowledgments

A.G. Mayorov, V.V. Malakhov and S.A. Rodenko acknowledges support from MEPHI Academic Excellence Project (contract 02.a03.21.0005).

References

- [1] Adye T Unfolding algorithms and tests using RooUnfold *Proceedings of the PHYSTAT 2011 Workshop (CERN, Geneva, Switzerland)* (2011)006 pp 313–318
- [2] Blobel V Unfolding methods in high energy physics experiments *in Proceedings of the 1984 CERN School of Computing* (1985)09 pp 84–114
- [3] Blobel V 2002 An unfolding method for high energy physics experiments *Contribution to the Conference on Advanced Statistical Techniques in Particle Physics (Durham)*
- [4] Cowan G A 2002 Survey of Unfolding Methods for Particle Physics *Proc. Advanced Statistical Techniques in Particle Physics (Durham)*
- [5] DAgostini G 1995 A multidimensional unfolding method based on Bayes theorem *Nucl. Instrum. Methods Phys. Res. A* **362** 23 pp 487–498
- [6] DAgostini G 2010 Improved iterative Bayesian unfolding *arXiv:1010.0632*
- [7] Hoecker A and Kartvelishvili V 1996 SVD Approach to Data Unfolding *Nucl. Instrum. Meth. A* **372** 3 pp 469–481
- [8] Kuusela M and Panaretos V 2015 Statistical unfolding of elementary particle spectra: Empirical Bayes estimation and bias-corrected uncertainty quantification *Annals of Applied Statistics* **9** 3 pp 1671–1705
- [9] Picozza P et al 2007 PAMELA - A payload for antimatter matter exploration and light-nuclei astrophysics *Astropart. Phys* **27** 4 pp 296–315