

PAPER • OPEN ACCESS

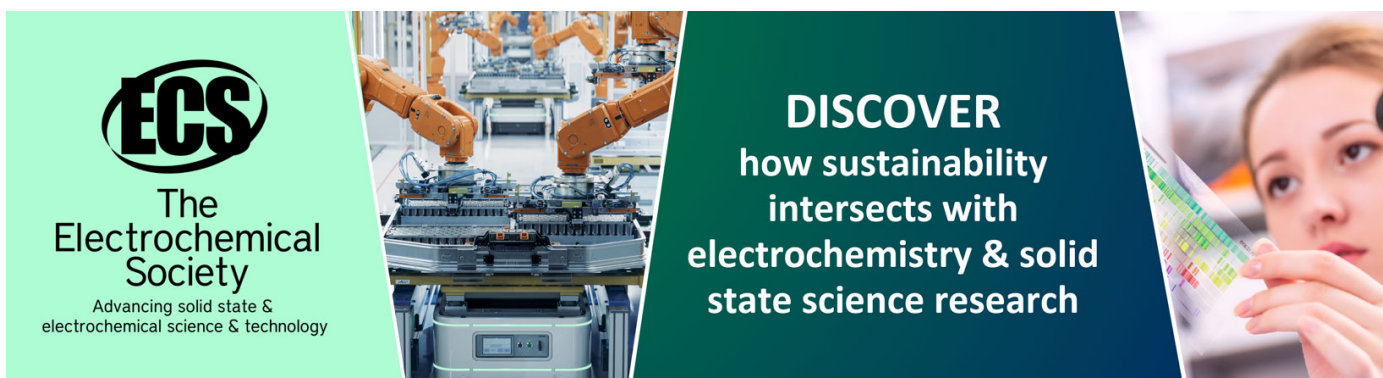
Corrections of fluctuation observables with the Unfolding techniques

To cite this article: D S Prokhorova and E V Andronov 2020 *J. Phys.: Conf. Ser.* **1690** 012134

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

You may also like

- [On the problem of laser error correction of multi-axis influence of loading on laser accuracy correction of multi-axis systems](#)
V I Teleshevsky, V A Sokolov and Ya I Pimushkin
- [A review: the application of generative adversarial network for mechanical fault diagnosis](#)
Weiqing Liao, Ke Yang, Wenlong Fu et al.
- [Heat exchanger design based on minimum entropy generation](#)
S Bucsa, D Dima, A Serban et al.



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Corrections of fluctuation observables with the Unfolding techniques

D S Prokhorova and E V Andronov

Saint-Petersburg State University, 7/9 Universitetskaya Emb., 199034, Saint Petersburg, Russia

E-mail: daria.prokhorova@cern.ch

Abstract. The paper studies the possibility to correct strongly intensive quantities $\Delta[P_T, N]$, $\Sigma[P_T, N]$, and $\Sigma[N_F, N_B]$ for detector inefficiencies by means of RooUnfold package. Several tests of applying Unfolding technique to 1D- and 2D-dimensional distributions were done on Monte-Carlo generated data in NA61/SHINE acceptance. Results reveal that several aspects of this analysis have to be improved.

1. Introduction

In the data analysis we base our predictions on the measured quantities that are in general strongly biased by the detector, which mixes the true physics results with the experimental influence. Therefore, in order to obtain the results independent of the detector by which they were measured, one has to perform the corrections of the experimental data for the detector inefficiencies.

This paper studies the possibility to correct strongly intensive quantities (SIQs) [1], [2] that are widely used in the analysis of event-by-event fluctuations [3], [4]. To be independent of the volume and event-by-event volume fluctuations (under some conditions) SIQs are constructed as combinations of the first and second moments of the extensive event quantities distributions. There are several possible ways to correct them: bin-by-bin method (used in [4]), straightforward correction of the moments [5] or deconvolution of the distributions whose moments are used. The latter one can be done with the Unfolding method by RooUnfold [6]. In this approach Monte-Carlo (MC) simulations are used to define detector inefficiencies by comparison of the pure MC data and MC events that were reconstructed as data in the experiment, meaning they have the same detector bias. This information is used to fill detector Response Matrix (RM) whose elements reflect the probabilities to measure X_{rec} in an event with X_{sim} produced. Further, the “inverted” RM has to be applied to the measured distribution, which gives us the true unsmeared one. The drawback of the bin-by-bin correction method, where each bin of the distribution is corrected by the multiplication by a constant factor, is that it can be applied only if the detector response matrix is diagonal, which means there is no event migration. The second method requires the manual fitting of distributions which gets the more difficult the higher the dimension of distributions to correct. Unfolding in its turn handles all these.

The paper shows the preliminary tests that were done to get familiar with the unfolding technique. As a first step RM was applied to another sub-set of the same MC that formed pseudo-data: pure and biased. Only after the study of all the features of the procedure and its



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

successful cross-validation on the second MC (to exclude the model dependence) this method can be applied to correct NA61/SHINE data.

The paper is organised as follows. The section “Quantities of interest” introduces the particular strongly intensive quantities and distributions that have to be unsmeared. In the section “Analysis details” some technical info is presented. “Results” section shows the comparison of the pure, biased and unfolded data and includes some discussion of the outcome. Conclusions are mentioned in the corresponding section.

2. Quantities of interest

In the study of event-by-event multiplicity and transverse momentum fluctuations one can use strongly intensive quantities (1) and (2), with P_T being a scalar sum of the event charged particles transverse momentum and N being an event charged particles multiplicity:

$$\Delta[P_T, N] = \frac{1}{C_\Delta} [\langle N \rangle \omega[P_T] - \langle P_T \rangle \omega[N]], \quad (1)$$

$$\Sigma[P_T, N] = \frac{1}{C_\Sigma} [\langle N \rangle \omega[P_T] + \langle P_T \rangle \omega[N] - 2(\langle P_T \cdot N \rangle - \langle P_T \rangle \langle N \rangle)], \quad (2)$$

where $\langle \dots \rangle$ means the averaging over all events. The intensive quantities that are used in the formulas above are scaled variances $\omega[N]$ and $\omega[P_T]$, $P_T^{(k)}$ is a scalar sum of the transverse momentum of particles in k -th event (M events in total):

$$\omega[N] = \frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle}, \quad \omega[P_T] = \frac{\langle P_T^2 \rangle - \langle P_T \rangle^2}{\langle P_T \rangle}, \quad \langle P_T \rangle = \frac{\sum_{k=1}^M P_T^{(k)}}{M} \quad (3)$$

The normalisation factors C_Δ and C_Σ of (1) and (2) include the scaled variance of the particle spectra p_T , with $\overline{p_T}$ meaning averaging over all particles in all events and $N^{(k)}$ being a charged particles multiplicity in k -th event (M events in total):

$$C_\Delta = C_\Sigma = \langle N \rangle \omega(p_T), \quad \omega(p_T) = \frac{\overline{p_T^2} - \overline{p_T}^2}{\overline{p_T}}, \quad \overline{p_T} = \frac{\sum_{k=1}^M P_T^{(k)}}{\sum_{k=1}^M N^{(k)}} \quad (4)$$

Another strongly intensive quantity $\Sigma[N_F, N_B]$ reflects joint fluctuations of multiplicities N_F and N_B calculated in some Forward and Backward pseudorapidity intervals [7]:

$$\Sigma[N_F, N_B] = \frac{1}{\langle N_B \rangle + \langle N_F \rangle} [\langle N_B \rangle \omega[N_F] + \langle N_F \rangle \omega[N_B] - 2(\langle N_F N_B \rangle - \langle N_F \rangle \langle N_B \rangle)] \quad (5)$$

Due to complexity and time consumption we avoid 3D Unfolding, and, providing that SIQs are centrality independent, in this analysis we perform 2D Unfolding of $P_T - N$ and $N_F - N_B$ joint distributions and 1D Unfolding of N distribution (for comparison) and particle p_T spectra.

3. Analysis details

The tests were performed for p+p@158GeV/ c Monte-Carlo data by EPOS1.99 generator [8] in the NA61/SHINE acceptance [9]. Event and track selection criteria were chosen as in [4]. Results for $\Delta[P_T, N]$ and $\Sigma[P_T, N]$ are obtained in η region [2.9, 5.8] in the lab frame. Forward window for $\Sigma[N_F, N_B]$ is in pseudorapidity range [2.9, 5.8], Backward - is in range [0, 2.9], all in lab frame.

Response matrix (RM) was built on the 85% of MC statistics, pseudo data - on the rest 15%: true - pure generator, biased - MC reconstructed in the same chain as data at NA61/SHINE.

Statistical uncertainties were calculated using the bootstrap method: the same RM was applied to sampled biased pseudo-data to obtain sampled unfolded distributions to be compared with the sampled true pseudo-data ones. In future, one can also apply bootstrap for RM construction. The iterative Bayesian method RooUnfoldBayes was used in this analysis.

4. Results

This section shows preliminary results that were obtained with the unfolding procedure within the attempt to restore the value corresponding to the pure pseudo-data from the one that corresponds to the biased pseudo-data (table 1). Therefore, one should compare the fourth column to the second one to test performance of the considered method. The strongly intensive measures $\Delta[P_T, N]$, $\Sigma[P_T, N]$ and $\Sigma[N_F, N_B]$ are listed at the bottom of the table, while all the quantities above are parts of their definitions.

Table 1. Preliminary results of the unfolding of fluctuation measures in EPOS1.99 pseudo-data in NA61/SHINE experimental acceptance. True values are obtained from pure generator data, biased values are from the reconstructed MC, unfolded values are the results from the distribution unsmeared by RooUnfoldBayes. $\overline{p_T}$ is the averaging of p_T over all particles in all events, $\langle \dots \rangle$ is the averaging over events.

Fluctuation measures	True value	Biased value	Unfolded value
$\langle N \rangle$ from 1D N distr.	2.92 +- 0.05	3.47 +- 0.05	2.92 +- 0.06
$\langle N \rangle$ from 2D $P_T - N$ distr.	2.92 +- 0.07	3.47 +- 0.09	2.90 +- 0.08
$\omega[N]$ from 1D N distr.	1.99 +- 0.04	1.72 +- 0.03	2.02 +- 0.04
$\omega[N]$ from 2D $P_T - N$ distr.	1.99 +- 0.05	1.72 +- 0.04	1.96 +- 0.05
$\langle P_T \rangle$	0.96 +- 0.09	1.12 +- 0.05	0.94 +- 0.03
$\omega[P_T]$	0.78 +- 0.07	0.66 +- 0.02	0.75 +- 0.05
$\overline{p_T}$	0.327 +- 0.004	0.329 +- 0.003	0.328 +- 0.003
$\omega(p_T)$	0.134 +- 0.002	0.132 +- 0.004	0.135 +- 0.004
$\Delta[P_T, N]$	0.99 +- 0.06	0.81 +- 0.07	0.85 +- 0.09
$\Sigma[P_T, N]$	1.03 +- 0.03	1.07 +- 0.06	0.98 +- 0.07
$\Sigma[N_F, N_B]$	1.09 +- 0.06	0.90 +- 0.09	1.05 +- 0.05

In the table the difference between line 1 and line 2 (as well as between line 3 and 4) is that the former was obtained from the straightforward 1D unfolding of the multiplicity N distribution, while the latter was calculated from the projection of the unfolded by 2D unfolding $P_T - N$ distribution. One can see that the results of 1D unfolding of N are consistent with the true value. However, the discrepancy between true and unfolded values in the 2nd and 4th lines is larger. Several calculations were done with the sequential decrease of the P_T bin size, which caused the improvement of the unfolded value: for example, the decrease of P_T bin from 100 MeV/ c to 50 MeV/ c to 25 MeV/ c caused the positive changes of $\langle P_T \rangle$ from 1.05 to 0.99 to 0.94 GeV/ c . However, this significantly increases the program execution time. The further study will be held.

The idea of applying Unfolding technique comes from the fact that event migration strongly influences the measured quantities. Namely, figure 1 shows examples of the detector Response Matrix that are clearly non-diagonal. One can see (figure 1(a)) that, for example, different values of N_{rec} (what would be measured) correspond to one N_{sim} (what was really produced). To plot figure 1(b) the matching between simulated and reconstructed charged particles was performed

(with some threshold on the matching efficiency), but still, there are off-diagonal elements that again come from the detector disturbances during measurements.

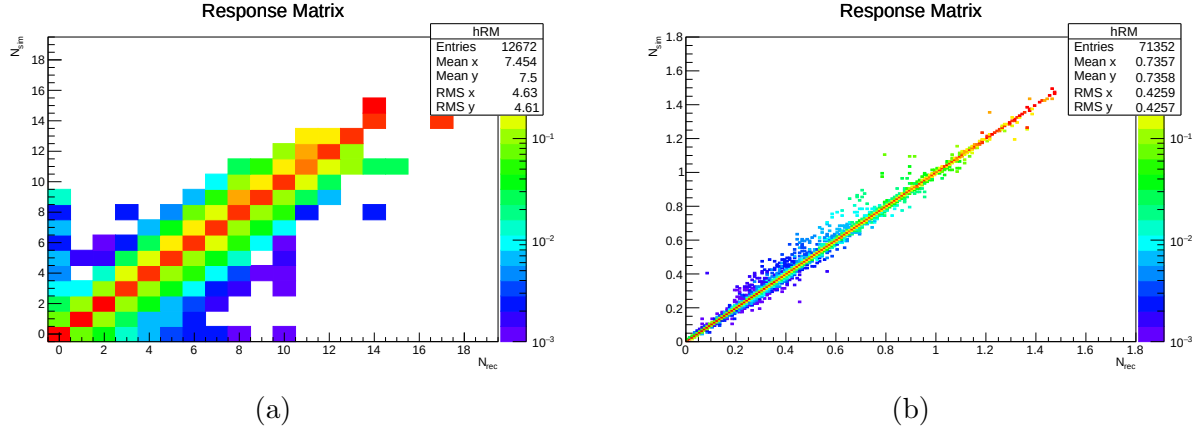


Figure 1. Response matrix of the detector biasing (a) charged particle multiplicity: N_{rec} vs N_{sim} (b) particle transverse momentum: p_T^{rec} vs p_T^{sim} in GeV/c (with matching).

5. Conclusions

The paper shows preliminary tests of the Unfolding procedure applied to 1D- and 2D-distributions in order to correct fluctuation observables in NA61/SHINE acceptance. The detector response matrix (figure 1) is not diagonal, which excludes the use of bin-by-bin method of corrections. Table 1 reveals that the unfolded results are close to the pure generator values. However, the discrepancies are unacceptable, which do not allow us to claim that at the current stage of this analysis the used unfolding procedure is working well in the conditions of NA61/SHINE experimental setup. The discrepancy mainly comes from the P_T unfolding which implies the need of finer P_T binning. Therefore, the next steps will be to increase Monte-Carlo statistics to build detector response matrix, find the optimal values for the bin sizes of non-integer quantities (p_T and P_T) and finally perform the cross-validation of the procedure on the second Monte-Carlo with the simulation of trigger.

Acknowledgments

This work is supported by the Russian Science Foundation under grant 17-72-20045. We thank members of the CERN NA61/SHINE Collaboration for their support and help.

References

- [1] Gorenstein M I and Gazdzicki M 2011 *Phys. Rev. C* **84** 014904
- [2] Gazdzicki M, Gorenstein M and Mackowiak-Pawlowska M 2013 *Phys. Rev. C* **88** 024907
- [3] Prokhorova D 2019 *EPJ Web of Conferences* **204** 07013
- [4] Aduszkiewicz A *et al.* 2016 *Eur. Phys. J. C* **76** 635
- [5] Nonaka T, Kitazawa M and Esumi S 2018 *Nucl. Instrum. Methods. Phys. Res. A* **906** 10
- [6] URL <https://gitlab.cern.ch/RooUnfold/RooUnfold>
- [7] Andronov E V 2015 *Theor. Math. Phys.* **185** 1383
- [8] Werner K and Pierog T 2006 *Phys. Rev. C* **74** 044902
- [9] Czopowicz T URL <https://edms.cern.ch/document/1549298/1>