

Experience with using unfolding procedures in ATLAS

Silvia Biondi On Behalf of the ATLAS Collaboration^{1,a}

¹University of Bologna and INFN

Abstract. In the ATLAS experiment, several unfolding methods are used to correct experimental measurements for detector effects, like acceptance and resolution. These methods use as input the raw experimental distributions, as well as Monte Carlo simulation for the description of the detector effects. The systematic uncertainties associated to the various unfolding methods are evaluated. The statistical and systematic uncertainties affecting the raw measurements and/or the simulation are propagated through the unfolding procedure. The resulting corrected measurements with their uncertainties can be directly compared with the corresponding theoretical predictions.

1 Introduction

Unfolding is the procedure of correcting a measured quantity, which in many cases is the distribution of an observable, for distortions due to the measurement process (for example, detector effects).

Many analyses of data collected with the ATLAS detector [1] use this technique in order to infer an unknown distribution from a measured one. Different methods are available and implemented in the RooUnfold [2] framework. In spite of this a standard procedure can be delineated that is totally independent from the chosen method. The treatment of the statistical and systematic uncertainties is also a crucial technical aspect when the unfolding is applied to an analysis. Since the treatment can change and there is not a general guide line for it, it is important to mention some of the main differences of the used methods.

Section 2 describes the general unfolding procedure, the motivation of applying the unfolding and the wide range of different methods. Section 3 outlines the needs in the preparation of an analysis for the unfolding and the standard steps to be followed. Finally, section 4 shows practical applications and issues in three different analyses.

2 The unfolding procedure

Unfolding allows to infer the true distribution from the reconstructed one, distorted by detector effects, as limited acceptance, reconstruction efficiency and finite resolution.

Consider t as the truth value of an observable with its probability density function (pdf) $f(t)$. The measurement of such an observable is typically blurred by the finite resolution of the detector and is expected to result in reconstructed values r distributed according to a pdf $g(r)$. The probability to

^ae-mail: silvia.biondi@cern.ch

observe a reconstructed value r given the corresponding truth value t is defined as $p(r|t)$. The total number of truth and reconstructed values are denoted by μ and ν , respectively.

Since in most cases the values are given in binned form, with N_t (N_r) bins, Δt_i (Δr_j) bin width and t_j (r_j) bin centre, the *response matrix* R [3] can be defined as

$$R_{ij} = p(r \in (\Delta r)_i | t \in (\Delta t)_j) \quad (1)$$

that represents the probability to reconstruct truth values from the interval $(\Delta t)_j = [t_j - \Delta t/2, t_j + \Delta t/2]$ in the interval $(\Delta r)_i = [r_i - \Delta r/2, r_i + \Delta r/2]$. The response matrix can be seen as a mathematical representation of the distortion of the detector.

The number of truth values in bin j is defined as

$$\mu_j = \mu \cdot \int_{(\Delta t)_j} f(t) dt, \quad (2)$$

and the expected number of reconstructed values in bin i is defined as

$$\nu_i = \nu \cdot \int_{(\Delta r)_i} g(r) dr. \quad (3)$$

The Response Matrix can be calculated from Monte Carlo events for which both the truth and reconstructed values are available. All events for which exist both truth and corresponding reconstructed values fill a matrix or two-dimensional histogram.

The Response Matrix and the vector of the reconstructed events ν are the input for the unfolding procedure, while the outputs are the estimate of the vector of the true events μ and the corresponding *covariance matrix*. Indeed, the μ and ν vectors are connected via the R matrix,

$$\nu_i = \sum_j R_{ij} \mu_j. \quad (4)$$

This relation is valid for just one process, but different processes (background processes for example) can also contribute to the number of the events in each bin. Including the estimated contribution of the background processes in each reconstructed bin, eq. 4 becomes

$$\nu_i = \sum_j R_{ij} \mu_j + \beta_i \quad (5)$$

where β_i is the number of expected background events in bin i . This relation can also be written in matrix form:

$$\nu = R\mu + \beta. \quad (6)$$

The process of predicting the expected reconstructed spectrum from a spectrum of truth values is called *folding*, because each truth value is folded with the detector resolution. The reverse problem that allow to infer the truth spectrum from an observed one is consequently named *unfolding*.

The *migration matrix* M gives the conditional probability that a selected event, generated in a bin i , is reconstructed in a bin j . Formally, it is related to the response matrix R by the relation $R = M\epsilon$ where ϵ is a diagonal matrix whose elements represent the probabilities of reconstructing an event generated in a given bin.

2.1 Advantages of the unfolding

The use of the unfolding technique in physics analyses is widespread and important for many reasons. First of all, through the unfolding process it is possible to obtain results which are independent from detector and reconstruction effects. Consequently, the unfolded differential distributions can be easily compared both between different experiments (like ATLAS and CMS) and with theoretical predictions. Last but not least, unfolding allows to tune the Monte Carlo simulations. Indeed, with the final true distribution a clear idea of the process has been obtained, without any dependence on detector effects. This allows to better understand the process and to improve the simulations used in the physics analyses.

There are some cases in which unfolding is not needed. For example in case of the measurement of an integrated observable, (e.g. an inclusive cross section)¹ or when the parameters of interest can be estimated directly by a fit on the reconstructed spectrum.

Unfolding is the method of choice if one is interested in an actual estimate of the true spectrum. In particular, this is the case, as mentioned above, if one plans to compare the estimated spectra to those obtained by other experiments. These comparisons are typically only meaningful if detector effects have been removed from the data. A rather particular problem is the inference on parameters which can only be derived from corrected spectra, e.g. asymmetry variables.

On the other hand, fitting reconstructed spectra might be the easier option if one is only interested in parameter values within a theoretical framework, e.g. masses, cross sections, etc. For those fits, the truth distribution is implicitly folded via the Monte Carlo simulation and it includes all detector and physics effects.

Certainly, the results are affected by the statistical and systematics uncertainties due to the unfolding procedure and detector effects. There are many different methods to evaluate them and the choice often depends on the particular problem the analyzers deal with. The uncertainties treatment is explained in details in section 4.

3 The unfolding procedure in the analysis

Before performing unfolding in an analysis, one should consider several aspects. The first one concerns the interpretation of the results, in particular the level (full or fiducial phase space) to which one unfolds. It is very analysis-dependent, but generally, the unfolding should be applied after the calibration and reconstruction of all the objects needed for the analysis. Moreover, it is often recommended to unfold after the definition of the fiducial phase space, which should be defined closely following the event selection in data, including both kinematic cuts on the objects and event topology cuts. The aim is to minimize the Monte Carlo input in the measurement and for this reason unfolding to a fiducial phase space reduces the theoretical uncertainties contribution.

In a nutshell, the standard procedure to be followed is the following:

- choice of observable(s);
- choice of binning;
- building of the Migration Matrix;
- the unfolding process itself, testing different methods and performing validation tests;
- choice of the method, looking at the best performances;

¹In this case a sort of unfolding is still performed in order to correct for migrations across thresholds, detector effects and acceptance.

- the optimization, through dedicated parameters and binning.

First of all, an observable must be chosen for the unfolded differential distribution. The second step is the choice of the binning and usually this is a compromise between the statistics and the observable resolution in each bin of the distribution. Once the observable and the binning have been chosen, the migration matrix can be built. If it is very far from being diagonal and symmetric (without any physical motivation), calibrations, event selections and binning have to be revised again. Indeed, having large non-diagonal terms makes the problem ill-posed and the results less stable. The aim of this step is to avoid this case and to have a matrix which is the most diagonal as possible.

Once the migration matrix is satisfactory, the fundamental step is the application of the unfolding itself. It is crucial to test, at this level, different methods and to perform different validation tests. This is needed because the unfolding process, as any inverse problem, rarely admits exact solutions and the approximate solutions are subject to potential instabilities. In order to avoid this problem, different techniques are needed to calculate the approximate solutions and some regularization techniques are required to overcome the potential instabilities. If there are problems with this step, e.g. the validation tests give bad results for the studied methods, it is recommended to come back to the binning choice. Otherwise, the preferred method can be chosen. This step is very important too, because many different methods are available, like:

- Singular Value Decomposition (SVD), using Tikhonov regularization [5];
- Iterative Bayes-inspired regularized unfolding [7];
- Iterative Dynamically Stabilized method (IDS) [8];
- Full Bayesian Unfolding (FBU) [9].

In general, the choice of the method strongly depends on the analysis and the problem the analysis aims to solve (statistics, etc.).

Fortunately, there are some recommendations to follow and there are some key factors to look at in order to make the right choice. The choice of the unfolding method depends on the conditions of your analysis, e.g. data and Monte Carlo statistics, background fraction and statistics, etc. One should estimate the MC shape uncertainties associated to different unfolding methods and the method with the smallest a priori uncertainty and bias should in general be preferred. This is the main criteria for the choice of an unfolding method. Other factors must however also be taken into account:

- the unfolding methods affect the size and the correlations of the statistical uncertainties (mainly through the regularization, but also through transfers of events). A method introducing less modifications and correlations in the statistical uncertainties is to be preferred;
- the behaviour of an unfolding method (mainly its systematic uncertainty) might also strongly depend on the binning that is used. Methods allowing the use of a finer binning, without introducing a too large MC shape systematic, are to be preferred. The binning is in some sense a regularization parameter. Going to finer binning increases the correlations between the bins of the unfolding result.

There could be cases in which two different methods have a similar performance. In this cases, the most computationally efficient method has to be preferred.

4 Application of the unfolding in the analyses in ATLAS

In this document, different analyses are shown in order to discuss the choice of the unfolding method and to examine how the statistical and systematic uncertainties are treated, according to the chosen method and the strategy of the specific analysis.

4.1 Measurement of the differential cross section of a top quark pair production

The first example concerns the measurement of the differential cross section of the production of a pair of top quarks with high transverse momentum (p_T) at $\sqrt{s} = 8$ TeV. The analysis considers the *lepton+jets* channel where one top (anti-top) decays leptonically and the other anti-top (top) decays hadronically ($t\bar{t} \rightarrow W^+ b W^- \bar{b} \rightarrow qq' blv\bar{b}$).

A combination of a bin-by-bin acceptance and efficiency corrections and the Singular Value Decomposition (SVD) unfolding method are used to correct the measured top quark p_T distribution to the particle and the parton-level unfolding. The differential cross-section measurement of interest is a function of the hadronically decaying top p_T and the procedure to unfold the distribution of $p_{T,\text{reco}}$ from the true one is composed of several steps, outlined in:

$$\frac{d\sigma_{t\bar{t},i}}{dp_{T,\text{ptcl}}^i} = \frac{N_{\text{ptcl}}^i}{\Delta p_{T,\text{ptcl}}^i \mathcal{L}} = \frac{1}{\Delta p_{T,\text{ptcl}}^i \mathcal{L} f_{\text{ptcl}|\text{reco}}^i} \cdot \sum_j M_{ij}^{-1} f_{\text{reco}|\text{ptcl}}^j f_{t\bar{t},l+jets} (N_{\text{reco}}^j - N_{\text{reco,bkg}}^j), \quad (7)$$

where N_{reco}^j is the number of observed events in bin j of $p_{T,\text{reco}}$ with the detector-level selection applied, N_{ptcl}^i is the total number of events in bin i of $p_{T,\text{ptcl}}$ that meet the fiducial region selection, $\Delta p_{T,\text{ptcl}}^i$ is the size of bin i of $p_{T,\text{ptcl}}$ and \mathcal{L} is the integrated luminosity of the data sample.

First, the non- $t\bar{t}$ background contamination, $N_{\text{reco,bkg}}^j$, is subtracted from the observed number of events in each $p_{T,\text{reco}}$ bin. The contribution of $t\bar{t}$ events from the other decay channels is taken into account by the multiplicative correction $f_{t\bar{t},l+jets}$, which represents the fraction of $l+jets$ $t\bar{t}$ events extracted from the nominal $t\bar{t}$ sample.

In a second step the factor $f_{\text{reco}|\text{ptcl}}^i$, also referred to as acceptance correction, corrects the $p_{T,\text{reco}}$ spectrum for the $t\bar{t}$ events that pass the detector-level selection but fail the particle-level selection. For each $p_{T,\text{reco}}$ bin j , $f_{\text{reco}|\text{ptcl}}^j$ is defined as the ratio of the number of events that meet both the detector-level and particle-level selections to the number of events that satisfy the detector-level selection. The distribution of the acceptance correction $f_{\text{reco}|\text{ptcl}}^j$ is shown in Figure 1.

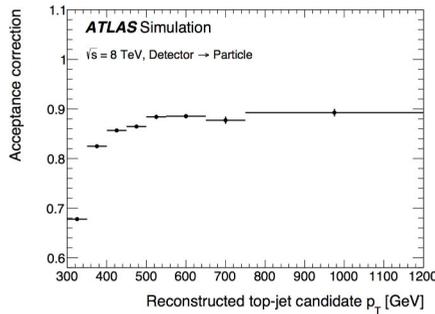


Figure 1. Distribution of the correction factor $f_{\text{reco}|\text{ptcl}}^j$ as a function of $p_{T,\text{reco}}$. It represents the ratio of the number of events that meet both the detector-level and particle-level to the number of events that satisfy the detector-level selection requirements. [4]

The third step corrects for detector resolution effects. A migration matrix is constructed to correlate the $p_{T,\text{reco}}$ -binned distribution to the $p_{T,\text{ptcl}}$ distribution. The matrix M_{ij} represents the probability for an event with $p_{T,\text{ptcl}}$ in bin i to have a $p_{T,\text{reco}}$ in bin j . This matrix is shown in Figure 2. It shows that approximately 50% to 85% of events have values of $p_{T,\text{ptcl}}$ and of $p_{T,\text{reco}}$ that fall in the same p_T bin.

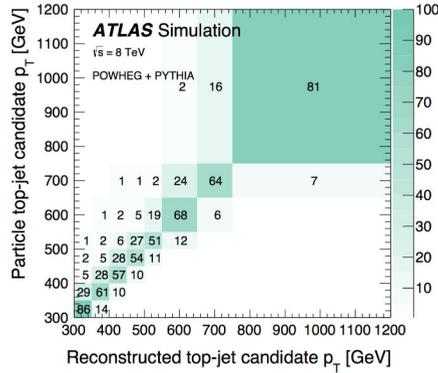


Figure 2. Migration matrix between the particle-level $p_{T,ptcl}$ and reconstructed detector-level $p_{T, reco}$. [4]

The inversion of the migration matrix to correct $p_{T, reco}$ to the particle level is carried out by an unfolding scheme based on Tikhonov regularization which is implemented through the singular value decomposition (SVD) of the matrix [5]. This scheme is chosen to reduce sizable statistical fluctuations that are introduced by instabilities in the inversion procedure. In the SVD method, the inversion of the matrix is easier thanks to its decomposition into two orthogonal matrices U and V and a diagonal matrix S :

$$M = USV^T \tag{8}$$

for which only the diagonal S matrix can be easily inverted:

$$M^{-1} = (USV^T)^{-1} = VS^{-1}U^T. \tag{9}$$

The fourth step is to apply a bin-by-bin correction factor $f_{ptcl!reco}^i$, also referred to as efficiency correction, which restores the contribution of $t\bar{t}$ events that fulfil the particle-level selection but not the detector-level selection. This factor is defined as the ratio of the number of events that satisfy both the particle-level and detector-level selections to the number that meet the selection at particle level only. The distribution of the efficiency correction $f_{ptcl!reco}$ is shown in Figure 3.

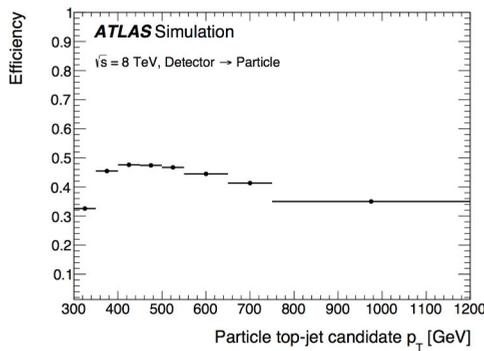


Figure 3. Distribution of the correction factor $f_{ptcl!reco}$ as a function of $p_{T,ptcl}$. It represents the ratio of events that meet both the particle-level and detector-level to those that satisfy the particle-level selection requirements. [4]

After these steps, some validation tests have been performed:

- the MC closure test aims to check if a reconstructed MC distribution, that is unfolded using a statistically independent sample from the same generator, provides a result that is compatible with the truth distribution of the unfolded sample;
- the data-driven closure test, also called stress-test, is needed to check how much the data/MC disagreement affects the unfolding procedure.

The measured differential cross section is shown in Figure 4. MC-based pseudo-experiments are

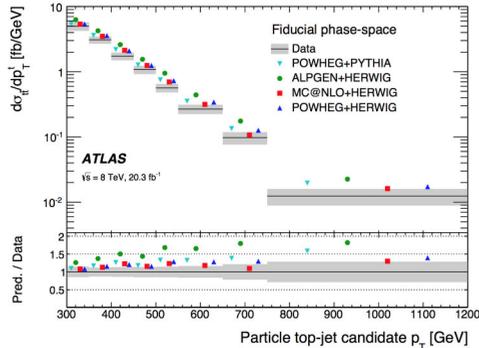


Figure 4. Fiducial particle-level differential cross section as a function of the hadronic top-jet candidate p_T . Different MC predictions are compared with the final results. The lower part of the figure shows the ratio of the MC prediction to the data. The shaded area includes the total statistical plus systematic uncertainties. The points of the various predictions are spaced along the horizontal axis for presentation only; they correspond to the same p_T range. [4]

used in order to evaluate the uncertainties of the unfolding procedure. For each pseudo-experiment, a smearing is performed in each bin of the reconstructed distribution by a Poisson fluctuation for the statistical uncertainties and by a Gaussian fluctuation for each systematic uncertainty. The next step is to perform the unfolding for each smeared distribution. The covariance matrix obtained from the ensemble of the unfolded distribution takes into account the bin-by-bin correlations due to the migration effects and different systematics sources (detector and background modeling). The complete covariance matrix, instead, takes into account other sources as well, like the $t\bar{t}$ generator, the parton shower, ISR/FSR and the PDF uncertainties.

An analogue $t\bar{t}$ measurement has been performed at 13 TeV [6]. The main differences with respect to the 8 TeV analysis are the following:

- the differential cross section has been studied as a function of the hadronic top p_T and of the absolute value of its rapidity;
- the unfolding method is the Iterative Bayes-inspired [7], explained in detail in section 4.2, instead of the SVD.

A different unfolding method was used since it provides smaller long-range correlations and better performances in terms of data processing.

4.2 Measurement of the cross section of a $W^\pm Z$ boson pair production

The third example concerns the measurement of the integrated and differential cross section, as a function of the jet multiplicity, of a pair of $W^\pm Z$ bosons production at $\sqrt{s} = 13$ TeV [10]. The leptonic decay of the $W^\pm Z$ pair ($W^\pm Z \rightarrow l^\pm \nu l^+ l^-$, $l = e$ or μ) is considered.

The Iterative Bayes-inspired method is used in order to correct for detector effects, such as finite resolution, limited acceptance and imperfect efficiency. In this method, the unfolding is performed more than once in such a way that the input distribution to be unfolded in an iteration is the unfolded output of the previous one. This method has been chosen because the computation of the uncertainties is very fast, there is only one regularisation parameter (that is the number of iterations) and, consequently, the regularisation process is straight-forward.

The procedure of the unfolding can be represented as in Figure 5 and it is very similar to the process described for the $t\bar{t}$ differential cross section analysis, in section 4.1.

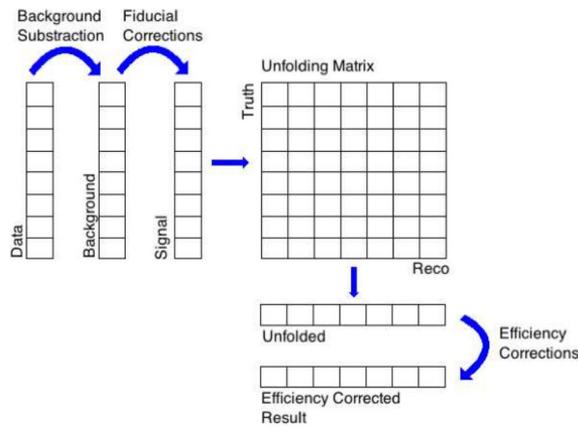


Figure 5. Schematic view of the unfolding procedure. Fiducial corrections are applied to the signal yield before the actual unfolding process. The result is corrected with efficiency factors.

Initially, the estimated background is subtracted from data, and fiducial factors are applied in order to correct for the number of events that fall out of the detector phase-space. Then the unfolding process accounts for migrations among different bins while in the final step, efficiency factors are applied to account for the detector imperfect efficiency.

In this analysis, significant migrations are observed. This is due to the chosen binning, but since this is done in number of jets, it is as fine as possible. A stability check has been performed, in which the stability is defined as the ratio between the reconstructed jets multiplicity and the generated jets multiplicity. It has been verified that the ratio is always larger than 60% in each bin.

Also in this case, after the steps shown in Figure 5, some optimization and validation (closure) tests are required. First, the optimization of the number of iterations of the method is necessary. Looking at the statistical uncertainty and at the bias due to the method, the best compromise has been found in the three iterations case.

The measured differential cross section is shown in Figure 6.

In this analysis, differently from the previous one, MC-based pseudo-experiments are used only for the statistical uncertainty evaluation. For each systematic uncertainty, instead, the migration matrix is varied, while the input distribution remains unchanged. This is unfolded with the different matrices and at the end a covariance matrix is obtained for each systematics.

An important difference with respect to the $t\bar{t}$ analysis is that now the unfolding uncertainty due to the shape chosen as prior is estimated. The method is similar to a stress test, in which the MC distribution is reweighted, taking into account the difference between data and MC shapes. The MC

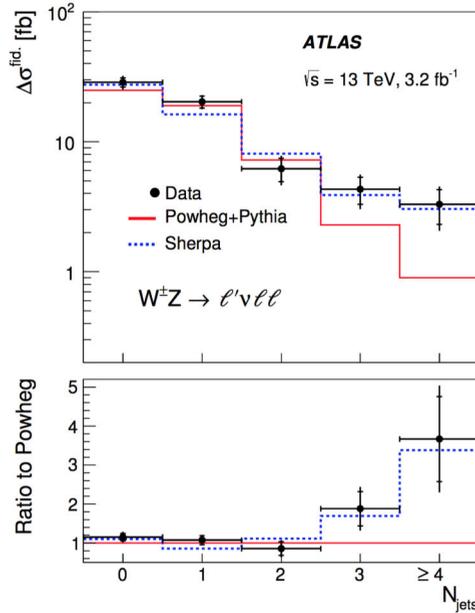


Figure 6. The measured $W^\pm Z$ differential cross section in the fiducial phase space as a function of the exclusive jet multiplicity of jets with $p_T > 25$ GeV. The inner and outer error bars on the data points represent the statistical and total uncertainties, respectively. The measurements are compared to the prediction from Powheg+Pythia (red line) and Sherpa (dashed blue line). The lower part of the figure shows the ratios of the MC prediction from Sherpa and the data to the MC prediction from Powheg+Pythia. [10]

plays the role of the data and it is unfolded. The unfolding uncertainty is taken into account in the final result, together with all the other uncertainties.

4.3 Measurement of the inclusive jet cross-section analysis

The last example concerns the inclusive jet cross-section analysis [11] and in particular the measurement of the double-differential cross sections as a function of the jet p_T and jet rapidity, shown in Figure 7.

This analysis uses the Iterative, Dynamically Stabilized method [8] which is an innovative method that uses a data-driven regularization in order to dynamically reduce the statistical fluctuations.

5 Conclusion

In this report it has been described in some details how the unfolding process has been implemented in few selected ATLAS analyses. One important aspect to be outlined is the fact that the more appropriate unfolding method can depend on the accumulated statistics (see Sec. 4.1) or on the particular problem to be solved within a specific analysis (see Sec. 4.2).

These cases also show that there is not a “better than others” method because the choice is very problem dependent (statistics, systematics evaluation, etc.). Nevertheless, the ATLAS collaboration

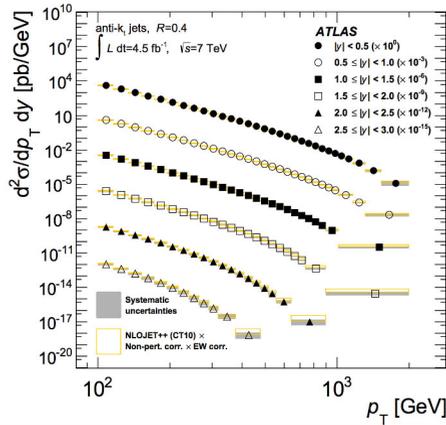


Figure 7. Double-differential inclusive jet cross sections as a function of the jet p_T in bins of rapidity, for anti- k_t jets with $R = 0.4$. For presentation, the cross sections are multiplied by the factors indicated in the legend. The statistical uncertainties are smaller than the size of the symbols used to plot the cross-section values. [11]

adopts a standard unfolding procedure, described in Sec 3, which gives a reference framework independent from the chosen method and some recommendations to be followed.

References

- [1] ATLAS Collaboration, JINST 3 (2008) S08003.
- [2] *RooUnfold: ROOT Unfolding Framework*, <http://hepunix.rl.ac.uk/~adye/software/unfold/RooUnfold.html>.
- [3] A. Armbruster, K. Kroninger, B. Malaescu, F. Spano, *Practical considerations for unfolding*, ATL-COM-PHYS-2014-277.
- [4] ATLAS Collaboration, *Measurement of the differential cross-section of highly boosted top quarks as a function of their transverse momentum in $\sqrt{s} = 8$ TeV proton-proton collisions using the ATLAS detector*, Phys. Rev. **D93** (2016) 032009.
- [5] A. Hoecker, V. Kartvelishvili, *SVD Approach to Data Unfolding*, Nucl. Instrum. Meth. **A372** (1996) 469-481.
- [6] ATLAS Collaboration, *Measurements of top-quark pair differential cross-sections in the lepton+jets channel in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector*, ATLAS-CONF-2016-040.
- [7] G. D’Agostini, *A multidimensional unfolding method based on Bayes’ theorem*, Nucl. Instrum. Meth. **A362** (1995) 487.
- [8] B. Malaescu, *An iterative, dynamically stabilized method of data unfolding*, arXiv:0907.3791.
- [9] G. Choudalakis, *Fully Bayesian Unfolding*, arXiv:1201.4612 [physics.data-an].
- [10] ATLAS Collaboration, *Measurement of the $W^{\pm}Z$ boson pair-production cross-section in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Phys. Lett. **B762** (2016) 1.
- [11] ATLAS Collaboration, *Measurement of the inclusive jet cross-section in proton-proton collisions at $\sqrt{s} = 7$ TeV using 4.5 fb $^{-1}$ of data with the ATLAS detector*, JHEP **02** (2015) 153.