Eur. Phys. J. Plus (2016) **131**: 191
DOI 10.1140/epjp/i2016-16191-6

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

On the potential of multivariate techniques for the determination of multidimensional efficiencies*

Benoit Viaud[a]

LAL, Univ. Paris-Sud, CNRS/IN2P3, Université Paris-Saclay, 91400 Orsay, France

Received: 8 April 2016
Published online: 9 June 2016 –

Abstract. Differential measurements of particle collisions or decays can provide stringent constraints on physics beyond the Standard Model of particle physics. In particular, the distributions of the kinematical and angular variables that characterise heavy meson multibody decays are non-trivial and can be used to probe this new physics. In the era of high luminosity opened by the advent of the Large Hadron Collider and of Flavor Factories, differential measurements are less and less dominated by statistical precision and require a precise determination of efficiencies that depend simultaneously on several variables and do not factorise in these variables. This article is a reflection on the potential of multivariate techniques for the determination of such multidimensional efficiencies. We carried out two case studies showing that multivariate techniques, such as neural networks, can determine and correct for the distortions introduced by reconstruction and selection criteria in the multidimensional phase space of the decays $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ and $D^0 \to K^-\pi^+\pi^+\pi^-$, at the price of a minimal analysis effort. We conclude that this method can already be used for measurements which statistical precision does not yet reach the percent level. With more sophisticated machine learning methods, the aforementioned potential is very promising.

1 Introduction

With the advent of the LHC and, a few years before, of Flavor Factories, High Energy Physics (HEP) entered an era of high statistical precision. More and more differential measurements of particle collisions or decays are now possible. For instance, stringent constraints can be imposed on the models predicting the dynamics of the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ by measuring the distribution of this decay in the four-dimensional space defined by q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ, a set of independent variables that provide a full description of the decay dynamics and that are defined in sect. 2.1. Deviations from the Standard Model (SM) predictions in specific regions of the phase space can be detected this way and sign the action of a physics beyond the SM. Loop-mediated rare decays like $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ are particularly sensitive to this new physics. Many models in which particles do not couple to weak interaction via only their left chiral component predict angular distributions that differ from the SM ones. More detail can be found, for example, in ref. [1].

In analyses of the kind introduced above, one has to account for the distortion of the phase space caused by reconstruction and selection criteria. In the example above, this is the distortion of the $(q^2, \cos\theta_l, \cos\theta_K, \phi)$ four-dimensional distribution. The most straightforward method would be to use a sample of simulated $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ phase-space decays. A four-dimensional binning could be defined, and the efficiency in each bin determined by the ratio between the yield of reconstructed and selected events and the yield of generated events. If this determination is made in terms of all the kinematic variables describing the decay and if the granularity of the binning is fine enough, the result does not depend on the distributions assumed for these variables by the simulation, which are not always realistic. However, this method would necessit to generate a huge sample. Even with only 10 bins per dimension, 10000 four-dimensional bins would have to be defined. It would take typically 10 million generated, reconstructed and selected events to determine the efficiency in all bins with less than a 5% uncertainty. Instead, sophisticated methods are available to account for efficiencies that depend simultaneously on several variables and that do not factorise in these variables (see sect. 2.1).

* Supplementary material in the form of a .pdf file available from the Journal web page at
http://dx.doi.org/10.1140/epjp/i2016-16191-6

[a] e-mail: viaud@lal.in2p3.fr

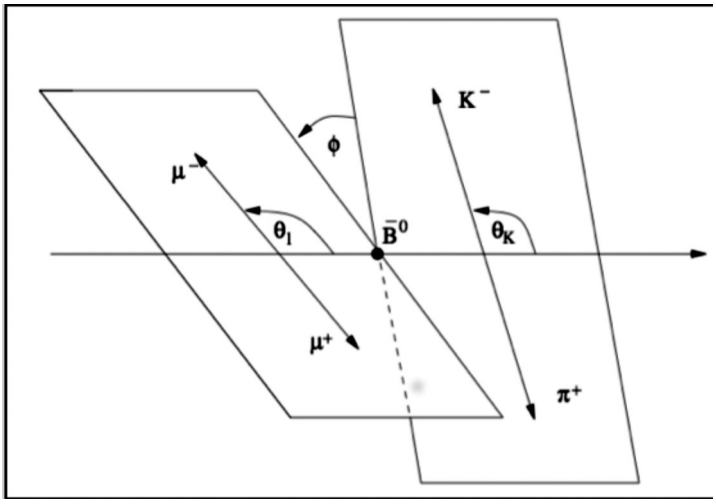

Fig. 1. Graphical representation of $\cos\theta_l$, $\cos\theta_K$ and ϕ. Their precise definitions are given in the text.

We propose in this article to explore the potential of another approach, suggested by rapid progresses in machine learning and multivariate analysis observed in the last decade. Techniques such as Neural Networks (NN) [2] or Boosted Decision Trees (BDT) [3], among others, can now detect with high sensitivity differences between two samples of events characterised by a large number of variables. They are routinely used nowdays in particle physics measurements to tell signal from background events. Said otherwise, these techniques are performant at comparing n-dimensional distributions to detect even subtle differences between signal and background samples (that traditional "by eye" studies would miss) and discriminate between event types via the NN or BDT score, a single variable incorporating all the information found in n dimensions. They should naturally be sensitive also to distortions introduced in the phase space of particle collisions or decays by reconstruction and selection criteria.

We do not take for granted that multivariate techniques reach the same level of precision as methods such as the principal moment analysis described in [1], nor that such a result can be obtained without a certain expertise in MVA techniques or without a time-consumming optimisation of the parameters that rule the technique's behavior and performance. However, the rythm at which MVA techniques have been progressing recently, their growing availability to basic users in the form of user-friendly packages, and the increasing typical expertise of high-energy physicists suggest MVAs might soon become very valuable tools, and easy to use, for multidimensional efficiency determination. It is therefore interesting to start exploring the potential of this approach.

This article only starts this exploration, with the ambition to answer the following question: what can be achieved if multivariate techniques are used in a very simple manner, using generic settings, like the default ones found in packages like [4]? In the case of analyses which do not require the same precision as the analysis of $B^0 \to K^{*0}\,(\to K^+\pi^-)\mu^+\mu^-$ introduced above, is this approach enough? Also, our goal is not to provide quantitative results, but an illustration of the potential of this approach.

In this article, we will first describe briefly a typical technique used to treat multidimensional efficiencies, and describe the approach we propose (sect. 2). Then, we will apply it to a first test case, involving the decay $D^0 \to K^-\pi^+\pi^+\pi^-$ (sect. 3). This decay can be used to improve our knowledge of D-meson mixing. For that purpose, one needs to determine the selection efficiency across the space defined by five independent variables in terms of which the decay dynamics can be expressed. Another test case will involve $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ (sect. 4). The results observed in these two cases will be summarized and discussed in the conclusion (sect. 5).

2 Techniques for multidimensional efficiencies

2.1 A classical technique

Methods of a certain complexity can be used to account for efficiencies that depend simultaneously on several variables and *that do not factorise in these variables*. The study of the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ provides a typical example [1]. The dymamics of this decay is fully described by four independent variables, q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ. They are defined as the invariant mass of the dimuon system squared, the cosine of the angle between the μ^+ (μ^-) and the direction opposite the B^0 ($\overline{B}^0$) in the rest frame of the dimuon system, the cosine of the angle between the direction of the K^+ (K^-) and the B^0 ($\overline{B}^0$) in the rest frame of the K^{*0} ($\overline{K}^{*0}$) system, and the angle between the plane defined by the μ^+ and μ^- and the plane defined by the kaon and pion in the B^0 ($\overline{B}^0$) rest frame, respectively. The definition of the three angles above is illustrated on fig. 1. In ref. [1], the technique introduced in ref. [5] is used

to determine the efficiency across the four-dimensional space defined by q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ. This involves a long sum of products of Legendre polynomials in the concerned variables:

$$\epsilon(\cos\theta_l, \cos\theta_K, \phi, q^2) = \sum_{\text{klmn}} c_{\text{klmn}} P_k(\cos\theta_l) P_l(\cos\theta_K) P_m(\phi) P_n(q^2), \tag{1}$$

where the terms $P_i(x)$ stand for Legendre polynomials of order i. The coefficients c_{klmn} are evaluated by performing a principal moment analysis of simulated $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ phase-space decays:

$$c_{\text{klmn}} = \frac{1}{N'} \sum_i^N \omega_i \left[\left(\frac{2k+1}{2}\right) \left(\frac{2l+1}{2}\right) \left(\frac{2m+1}{2}\right) \left(\frac{2n+1}{2}\right) P_k(\cos\theta_l) P_l(\cos\theta_K) P_m(\phi) P_n(q^2) \right], \tag{2}$$

where N is the total number of events and ω_i are weights imposed for instance to correct for known discrepancies between data and simulation. Their sum provides the total normalisation N'. For $\cos\theta_l$, $\cos\theta_K$, the angle ϕ and q^2, polynomials up to order 4, 5, 6 and 5 are used, respectively (see ref. [1]). Designing the method, understanding the properties of the Legendre polynomials and, more generally, of the sum in eq. (1), implementing the software to compare this parametrisation with the efficiencies observed for simulated decays so as to determine the highest orders to include until a proper description of the efficiency is obtained and to determine the c_{klmn} coefficients, and interpretating the results are time demanding tasks. In the end, hundreds of c_{klmn} are necessary. This might be worse in cases where more than four variables have to be dealt with, and it's not obvious the method will always work accurately.

2.2 A new approach

The approach we propose is based on the idea that if a BDT or a NN is very powerful at detecting differences between a signal and a background sample based on a given set of n variables, not by considering them individually, but by also exploiting their correlations, *i.e.* by comparing n-dimensional distributions, it should also be powerful at finding the differences between two samples differing only due to reconstruction and selection biases. In this case, instead of training the multivariate discriminator by comparing a *signal* and a *background* sample, one would compare samples of the same decay: an *original* sample made of generator level decays with a *distorted* sample made of decays that satisfied reconstruction and selection criteria. Instead of comparing discriminating variables, one would focus on the phase space of the decay, or in other words the n-dimensional distribution of events in the space defined by a set of independent variables that can describe fully the decay dynamics. In the example of the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$, these phase space variables are (q^2, $\cos\theta_l$, $\cos\theta_K$, ϕ).

In a four-stage approach, we first generate an original MC sample and a distorted one. The latter is generated in the same way as the former, save that reconstruction and selection cuts are applied. This stage is mandatory in most if not all physics analyses in HEP. The approaches like the ones in sect. 2.1 need that too. When high precision is necessary one generates samples containing up to a few million events. In most HEP collaborations, generating larger samples is challenging due to limited CPU and data storage capabilities. In the test cases presented in sect. 3 and 4, we use the `ROOT` package [6], and more specifically the `TGenPhaseSpace` class to generate these samples.

The second step is to train a multivariate analysis by comparing the samples generated above. In the case studies we carried out, we used the `TMVA` [4] package to train Multilayers Perceptron NNs. The only variables we provided the NNs with are the phase space variables.

The third stage is to parameterise the reconstruction and selection efficiency in terms of only a single variable, which makes the task of accounting for this multidimensional efficiency far more practical. For that purpose, we naturally use the NN score, which summarises into one single variable all the differences detected between the original and distorted phase spaces.

The fourth stage is to generate additional original and distorted samples, independent of those used above for training, in order to study the performance of the approach presented here. In practice, the parameterised efficiency determined in the third stage provides a per-candidate efficiency that is used to weight the distorted sample so as to reproduce the phase space (and the various characteristic distributions) of the original sample.

3 Application to the study of $D^0 \to K^-\pi^+\pi^+\pi^-$

Multibody D^0 decays provide examples where multidimensional efficiencies must be determined. These decays can be used, for instance, to improve our understanding of D^0 mixing [7]. This includes four-body decays. In this case,

a five-dimensional differential analysis must be carried out. Our case study in this article is the $D^0 \to K^-\pi^+\pi^+\pi^-$ decay, which dynamics can be described by the following set of masses:

$$\begin{aligned} m_{12} &= m(\pi_1^+\pi^-) \\ m_{23} &= m(\pi_2^+\pi^-) \\ m_{34} &= m(\pi_2^+K^-) \\ m_{123} &= m(\pi_1^+\pi^-\pi_2^+) \\ m_{234} &= m(\pi^-\pi_2^+K^-). \end{aligned}$$

3.1 Generation of the distorted and original samples

We generated an original sample and a distorted sample of $D^0 \to K^-\pi^+\pi^+\pi^-$ decays, both containing around 200000 decays. We tried to produce samples which features are typical of what could be observed at LHCb, the present world leading experiment in Flavor Physics [8]. For that purpose, an important element is to generate decays that have the same kinematics as in LHCb's laboratory frame. This is not possible with the `TGenPhaseSpace` class alone, which knows nothing of the physics of the proton-proton collision where D^0 mesons are produced. Using this class one can generate the four-momentum of each daughter particle given the four-momentum of the decaying particle, assuming a flat phase space (*i.e.* assuming that all the combinations of daughter four-momenta that respect energy and momentum conservation are equally allowed). To reproduce the kinematics of D^0 mesons produced in proton-proton collisions at a center-of-mass energy of 7 or 8 TeV, we combined several samples, each generated assuming a different value of the D^0 transverse momentum and rapidity. The relative contribution of each sample to the final one was decided according the D^0 production cross-sections measured in ref. [9] as a function of these quantities.

After having generated the original sample as described above, we produced the distorted sample by requiring all the decay products to be in the acceptance of the LHCb detector (*i.e.* the angle between their momentum and the nominal beam line should lie in the range 0.01–0.4 rad) and by applying the following kinematical cuts:

- $p_{\mathrm{T}}(D^0) > 3\,\mathrm{GeV}/c$;
- $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}/c$, $p > 3\,\mathrm{GeV}/c$ for all the decay products;
- $\max(p(K^-), p(\pi^+), p(\pi^+), p(\pi^-)) > 10\,\mathrm{GeV}/c$;
- $\max(p_{\mathrm{T}}(K^-), p_{\mathrm{T}}(\pi^+), p_{\mathrm{T}}(\pi^+), p_{\mathrm{T}}(\pi^-)) > 1.7\,\mathrm{GeV}/c$;
- $\max(p_{\mathrm{T}}(K^-), p_{\mathrm{T}}(\pi^+), p_{\mathrm{T}}(\pi^+), p_{\mathrm{T}}(\pi^-)) > 3.7\,\mathrm{GeV}/c$ with the hardware trigger reconstruction.

By construction, the generator-level p and p_{T}'s available in our samples do not account for the finite resolution of the reconstruction. Since in LHCb the momentum resolution of the offline reconstruction is very small (from 0.5% at low momentum to 1% at 200 GeV/c), we consider that we can safely neglect it. One exception is made for the first stage of LHCb's trigger system. It is a hardware trigger that searches for high p_{T} objects, based on a partial event reconstruction carried out by the front end electronics [10]. Several kinds of objects are searched for, involving several trigger *lines*: candidates are looked for in the muon system, while in the electromagnetic calorimeter clusters due to photons or electrons are sought. In the case of a decay like $D^0 \to K^-\pi^+\pi^+\pi^-$, the hardware trigger looks for high p_{T} clusters ($> 3.7\,\mathrm{GeV}/c$) in the hadron calorimeter, which resolution is $\sigma_{\mathrm{E}}/E = 80\%/\sqrt{E} \oplus 10\%$ with E expressed in GeV. We apply the trigger cut to particles which p_{T} has been smeared in order to reproduce this resolution. This trigger cut is applied to only a third of the events since LHCb events in which a $D^0 \to K^-\pi^+\pi^+\pi^-$ decay is produced are often triggered on independently, due to the decay products of the second charm hadron produced in the event (proton-proton collisions actually produce c-$\bar{c}$ *pairs*).

On fig. 2 we compare the distributions of m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the distorted and original samples. The selection efficiencies as a function of each of these variables (*i.e.*, the ratio between the distributions superimposed on fig. 2) are also shown on fig. 3. This illustrates the effect of the selection on the phase space.

3.2 Neural network training

We chose to use the `MLPBNN` NN provided by the `TMVA` package. This Multilayer Perceptron (MLP) NN is trained using the BFGS method instead of a simple back-propagation method. The definition of a MLP, and a description of the latter method can be found in ref. [4]. Also, a Bayesian regulation technique is employed. It is described in ref. [11]. We used the default configuration found in the example macro downloaded with the `TMVA` package. To the attention of the expert reader, we specify that in this case the MLP involves only one hidden layer, which comprises $N + 10$ neurons,

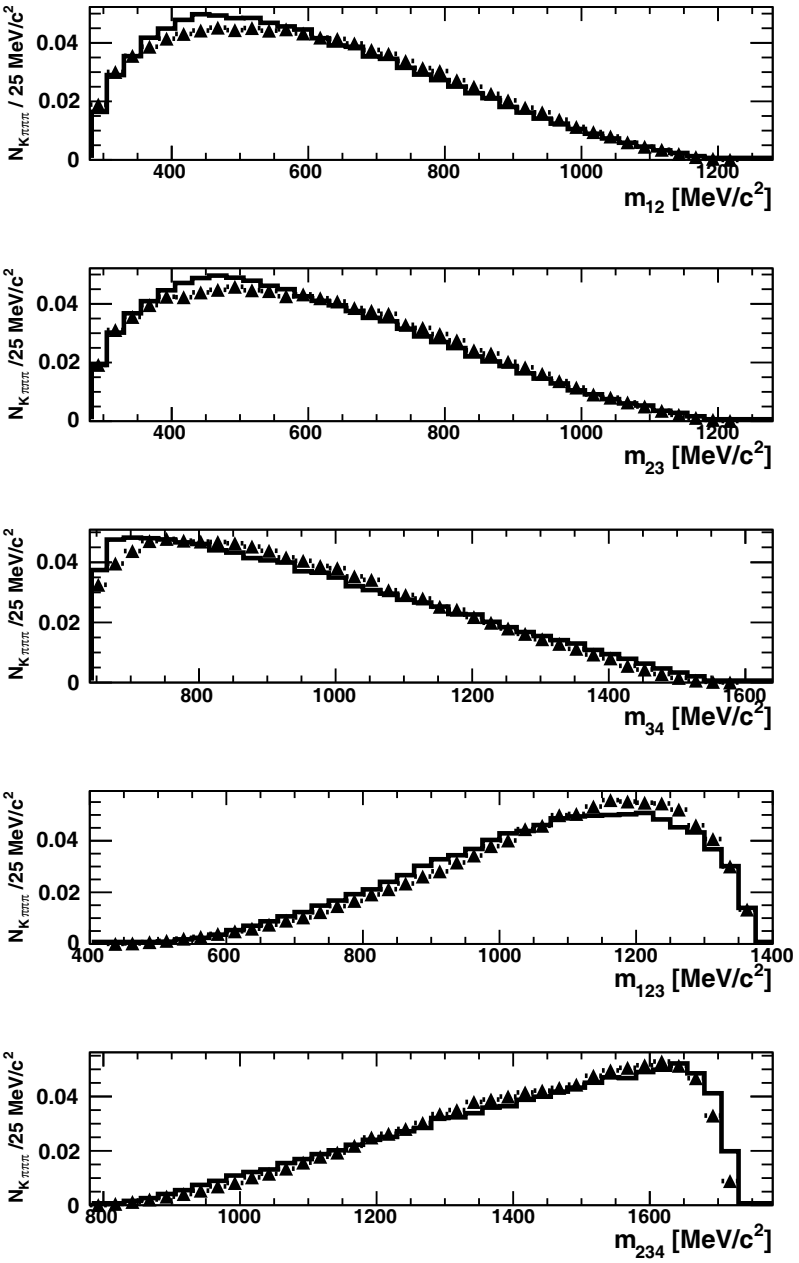

Fig. 2. Distributions of m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the original sample (histogram) and in the distorted one (full triangles). The normalisation is arbitrary.

where N is the number of input variables (m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in our case). The neuron activation function is *tanh*. Input variables are linearly scaled to lie with $[-1;1]$, and 600 training cycles are performed. An overtraining test is run every 5 cycles. All the other settings can be found in table 19 of ref. [4].

It took two hours to train this MLP to distinguish between decays from the distorted or original samples. We used a machine equipped with a Intel i7-640M 2.8 GHz 2-Core processor. The distribution of the NN score s in the original and distorted samples is shown on fig. 4. It also displays the ratio of these distributions, to which we fitted a parameterised efficiency, $\epsilon(s)$. We used for that a fifth order polynomial. The $\epsilon(s)$ function should match $\epsilon(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$, the multidimensional efficiency we aim at.

3.3 Results

To test the efficiency obtained above, we produced two additional distorted and original samples which size equals that of the training samples. We weighted each $D^0 \to K^-\pi^+\pi^+\pi^-$ decay i in the distorted test sample with $\omega_\mathrm{i} = 1/\epsilon(s_\mathrm{i})$.

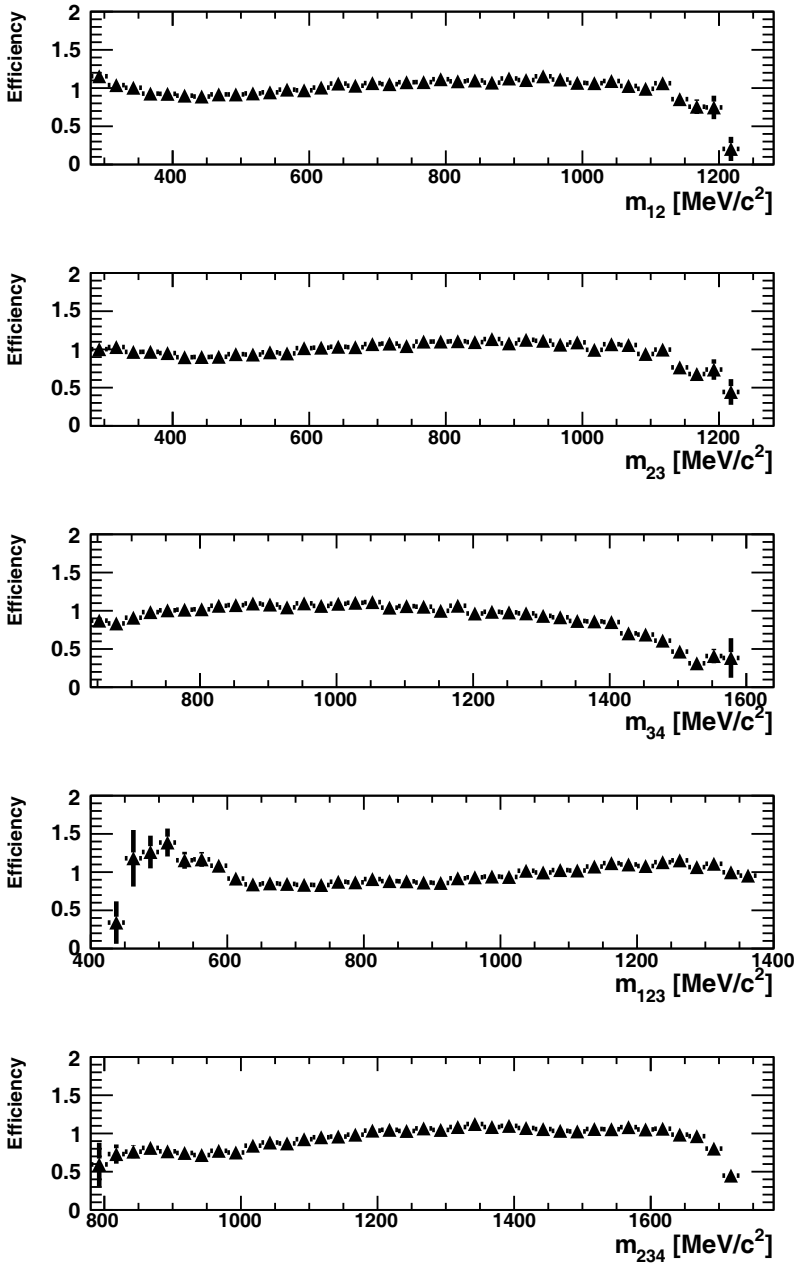

Fig. 3. Efficiency in m_{12}, m_{23}, m_{34}, m_{123} and m_{234} observed in the data generated for this study. Shown are the ratios of the distributions found in the distorted and original samples. The normalisation of these distributions was arbitrary.

The result can be found in fig. 5. It is the same as fig. 2 with distributions from the re-weighted distorted sample superimposed. One can see that these distributions match closely those observed in the original sample, before the distortions due to the selection were imposed. The ratios on fig. 6 are consistent with 1 all over the m_{12}, m_{23}, m_{34}, m_{123} and m_{234} spectra, unlike the un-weighted ratios on fig. 3, which show clear distortions. Consequently, the MLP we trained produces a single variable s to fully encompass the 5-dimensional information on the distortion of the $(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$ space. The $\epsilon(s)$ efficiency accounts simultaneously for the 5 individual efficiencies as a function of m_{12}, m_{23}, m_{34}, m_{123} and m_{234}. The yield in each bin of the corrected distributions in fig. 5 is calculated as the sum of the $\omega_{\rm i}$ weights over all the events that lie in this bin. In principle, this corrected yield can be correct even if $\epsilon(s)$ is not an accurate evaluation of $\epsilon(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$ everywhere in the $(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$ space. A compensation is possible in the sum. Also, the evaluation could be wrong in regions containing too little statistics for the effect to appear significantly on figs. 5 and 6. To constrain this possibility, we repeated the test of figs. 5 and 6 in various restricted regions of the $(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$ space. The results are on the figures provided in appendix A (in the supplementary material). The corrected distributions once more match the original ones.

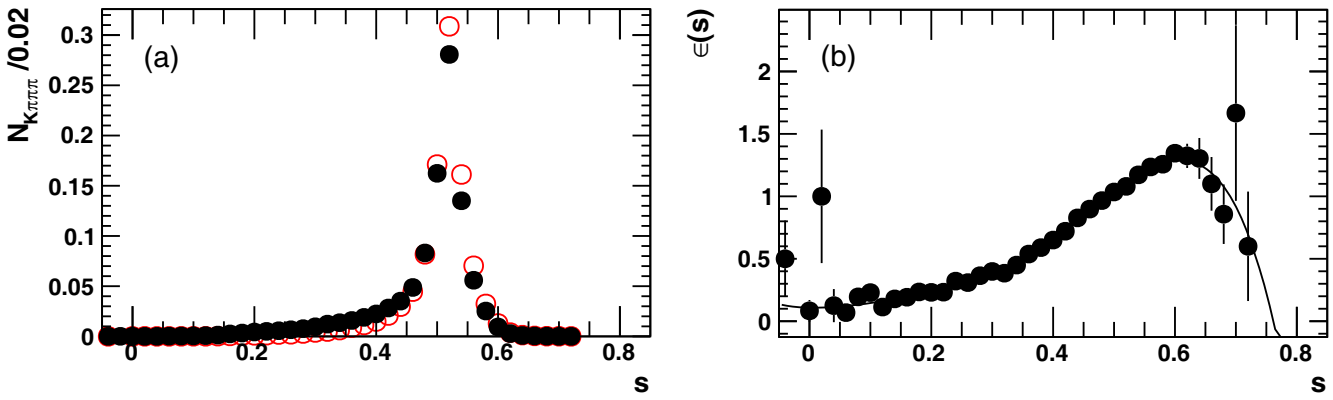

Fig. 4. Distributions and distribution ratio showing (a) the NN score s in the distorted (red open circles) and original (black full circles) $D^0 \to K^-\pi^+\pi^+\pi^-$ samples, and (b) the selection efficiency as a function of s (with an arbitrary normalisation). The fit leading to $\epsilon(s)$ is superimposed to the measured efficiencies.

Note that we still use the efficiency from fig. 4 for these corrections. The polynomial's parameters were not re-evaluated by performing in each region specific fits or new trainings. We conclude that $\epsilon(s) = \epsilon(m_{12}, m_{23}, m_{34}, m_{123}, m_{234})$, in the limit of the precision with which it can be verified with our data.

We reach the same conclusion when applying the same technique to correct an alternative distorted sample, obtained with tighter cuts and therefore showing stronger distortions in the phase space. We tightened one cut on the $D^0 \to K^-\pi^+\pi^+\pi^-$ decay products: $p_{\mathrm{T}} > 0.6\,\mathrm{GeV}/c$. Also, the hardware trigger selection is applied to all events instead of only a third. The results we obtained can be judged with the help of figs. 7 and 8. What we obtained in particular regions of the phase space is shown on figures provided in appendix A (in the supplementary material).

4 Application to the study of $\mathrm{B}^0 \to \mathrm{K}^{*0}(\to \mathrm{K}^+\pi^-)\mu^+\mu^-$

We performed a second case study to explore the potential of MVAs to treat multidimensional efficiencies. It is based on the decay $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$. A similar procedure to that described in sect. 3 has been carried out for that purpose.

4.1 Sample generation

We generated a distorted and an original sample, both containing about 2000000 decays, using once more the `ROOT` `TGenPhaseSpace` class. The kinematics of the B^0 meson is derived from the differential production cross-section as a function of transverse momentum and rapidity, measured in proton-proton collisions at a center-of-mass energy of 7 TeV [12].

To produce the distorted sample, we applied a selection as similar as possible to that in ref. [1]. We first impose all the decay products to be in the acceptance of the tracking system of the LHCb detector. The muons are also required to be in the acceptance of the muon system. To determine whether a given track verifies these criteria, its trajectory is extrapolated from the proton-proton interaction vertex to the position of the sub-detector under consideration as described in [13]. The impact of LHCb's dipole magnet is accounted for. The following other criteria are imposed:

- One of the muons should satisfy $p_{\mathrm{T}}(\mu) > 1.8\,\mathrm{GeV}/c$ in order to reproduce the cut used by the muon-specific line which dominates the hardware trigger in the case of decays such as $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$.
- $\max(p(K^+), p(\pi^-), p(\mu^+), p(\mu^-)) > 10\,\mathrm{GeV}/c$.
- $\min(p_{\mathrm{T}}(K^+), p_{\mathrm{T}}(\pi^-), p_{\mathrm{T}}(\mu^+), p_{\mathrm{T}}(\mu^-)) > 0.2\,\mathrm{GeV}/c$.
- $p_{\mathrm{T}}(B^0) > 4\,\mathrm{GeV}/c$.
- $p(B^0) > 40\,\mathrm{GeV}/c$.
- For all decay products $IP/\sigma_{\mathrm{IP}} > 3$, where IP stands for the minimum distance of a track to the proton-proton primary vertex.

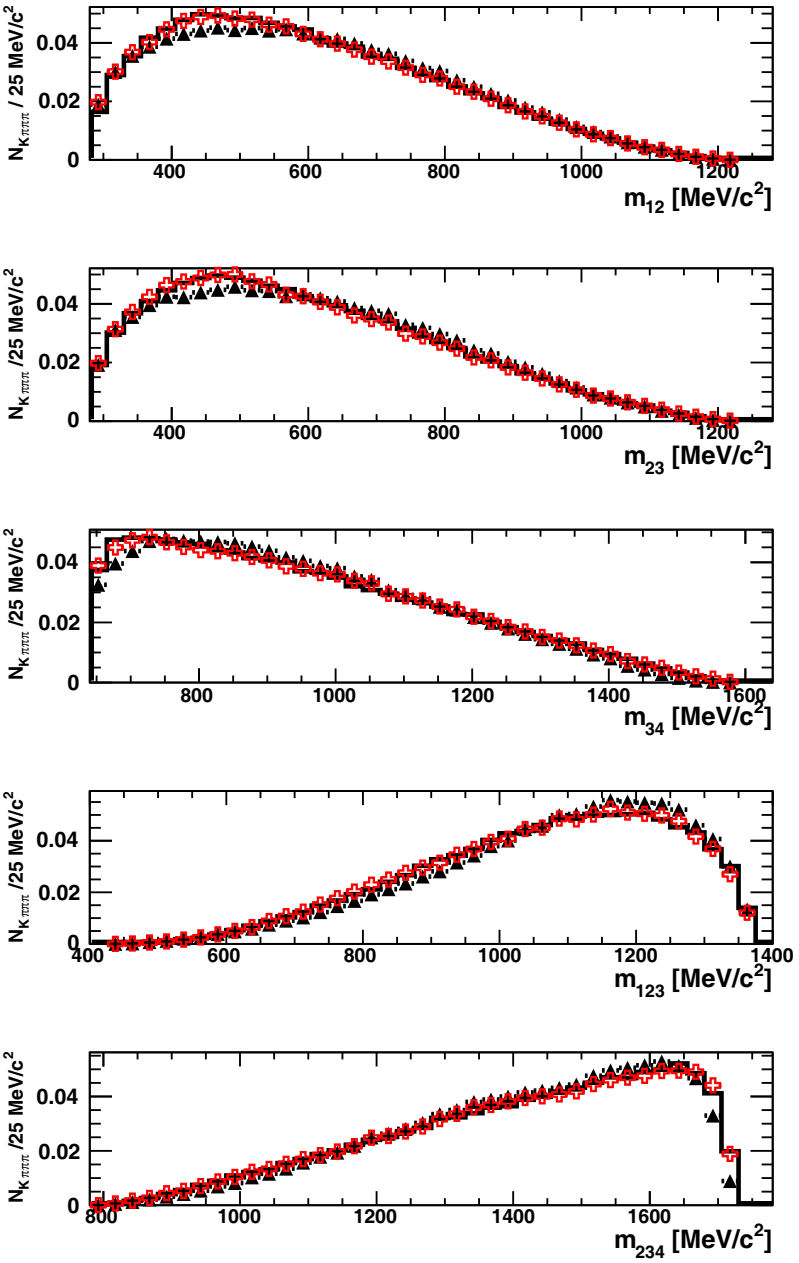

Fig. 5. Distributions of m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the original sample (histogram), in the distorted one (full black triangles) and in the distorted sample where the decays have been re-weighted using the $\omega_{\rm i}$ weights (open red crosses), as explained in the text. The absolute normalisation is arbitrary when the correction is not applied and natural when it is applied.

Only momenta and transverse momenta are directly accessible in the data generated with the `TGenPhaseSpace` class. The IP of a given track was therefore determined based on the angle between the $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decay products and the momentum of the B^0, assuming the latter travelled a typical 8 mm before its decay. The uncertainty on IP, $\sigma_{\rm IP}$, was computed based on the transverse momentum of the particle under consideration and on the typical uncertainty observed in LHCb. We used $\sigma_{\rm IP} = (15+29/p_{\rm T})\,\mu$m, with $p_{\rm T}$ in GeV/c, the uncertainty advertised by LHCb in recent publications (see for instance ref. [1]). With the set of criteria listed above, the distortion of the (q^2, $\cos\theta_l$, $\cos\theta_K$, ϕ) space, although not idenditical, is similar to that observed in ref. [1]. This can be seen on fig. 10, which shows the distortion in our generated data, in extreme regions of q^2: $0.1 < q^2 < 0.98\,{\rm GeV}^2/c^4$ and $18.0 < q^2 < 19.0\,{\rm GeV}^2/c^4$. This can be compared with fig. 9 to notice that the evolution of the distortion between these two extreme regions is reproduced.

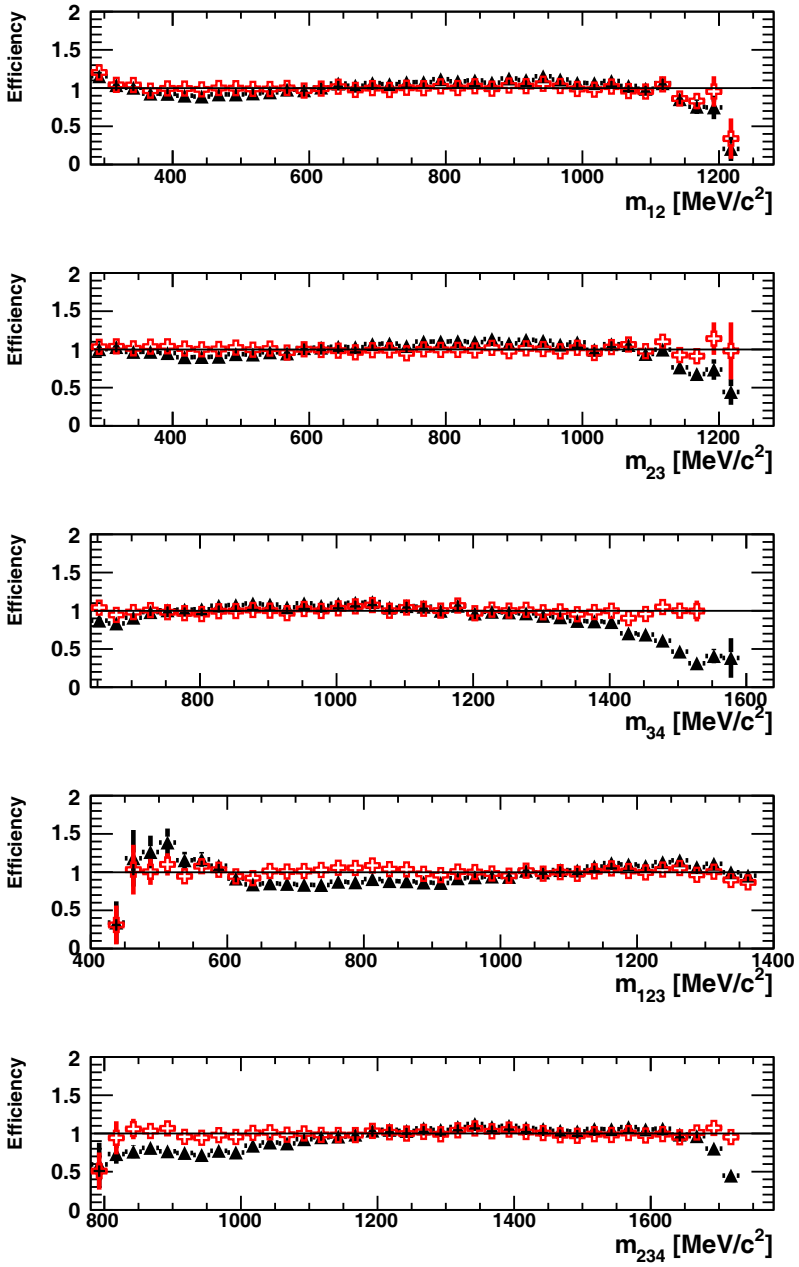

Fig. 6. Efficiency in m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the data generated for this study. Shown are the ratios of the distributions found in the distorted and original samples, with no correction (full black triangles) and for decays re-weighted using ω_i weights (red open crosses) as explained in the text. The absolute normalisation is arbitrary when the correction is not applied and natural when it is applied.

4.2 Neural network training

We trained the `MLP` NN provided by the `TMVA` package. It is the same NN as that described in sect. 3, modulo two differences: the classical back-propagation technique described in [4] is used instead of the BFGS method, and we use no Bayesian regulation technique.

The distortion of the (q^2, $\cos\theta_l$, $\cos\theta_K$, ϕ) space is more challenging to correct as that of the (m_{12}, m_{23}, m_{34}, m_{123}, m_{234}) space in the previous section. Indeed, as can be seen on figs. 9 and 10, the distortion the $\cos\theta_l$ and ϕ distributions is symmetric in these variables. As a consequence, the discriminative power of these variables is low: the distorted and original samples cannot be distinguished via a strong "preference" of one of them for the higher or the lower end of the $\cos\theta_l$ or ϕ distributions. In other words, the efficiencies with which distorted and original events would be selected by a cut on $\cos\theta_l$ or ϕ would not differ much, whatever the value of this cut. Another difficulty

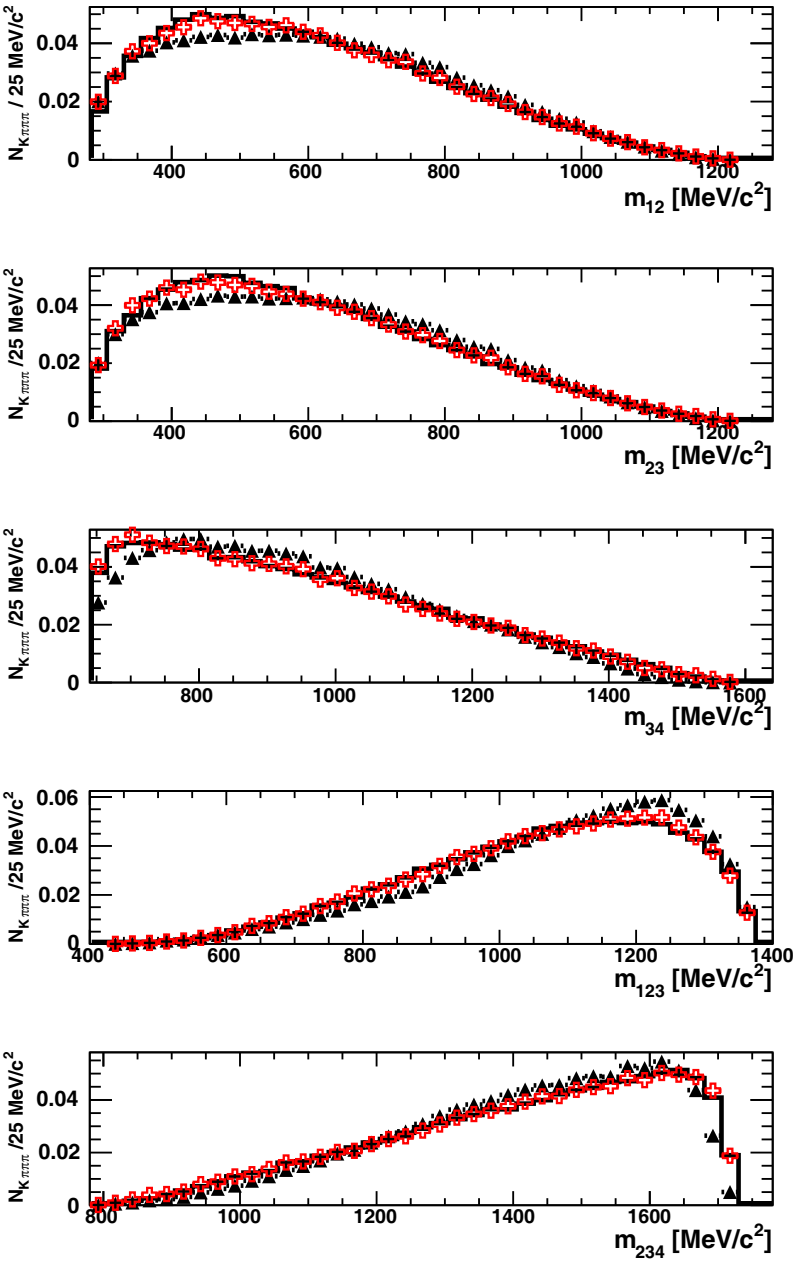

Fig. 7. Distributions of m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the original sample (histogram), in the distorted one obtained with a tighter selection (full black triangles) and in the same sample where the decays have been re-weighted using the $\omega_{\rm i}$ weights (red open crosses), as explained in the text. The absolute normalisation is arbitrary when the correction is not applied and natural when it is applied.

complicates the determination of the multidimensional efficiency $\epsilon(q^2, \cos\theta_l, \cos\theta_K, \phi)$ with the approach presented in this article. It stems from the fast variation as a function of q^2 of the distortion in the 3 other variables and the particular pattern it follows: at low q^2, the distortion is strong in $\cos\theta_l$ and very light in ϕ, while the opposite is observed at high q^2. There is therefore a more complicated structure to be understood by the NN. Also, the fact that in some q^2 regions only two variables can actually discrimate the distorted sample against the original one is a difficulty in itself. Moreover, it is not trivial for the NN to adapt to this varying behavior since it is trained using the whole distorted and original samples. It is not instructed to treat differently the various q^2 regions. This difficulty affects the determination of $\epsilon(q^2, \cos\theta_l, \cos\theta_K, \phi)$ in the whole sample and becomes more acute when it comes to determine the multidimensional efficiency in restricted regions of q^2.

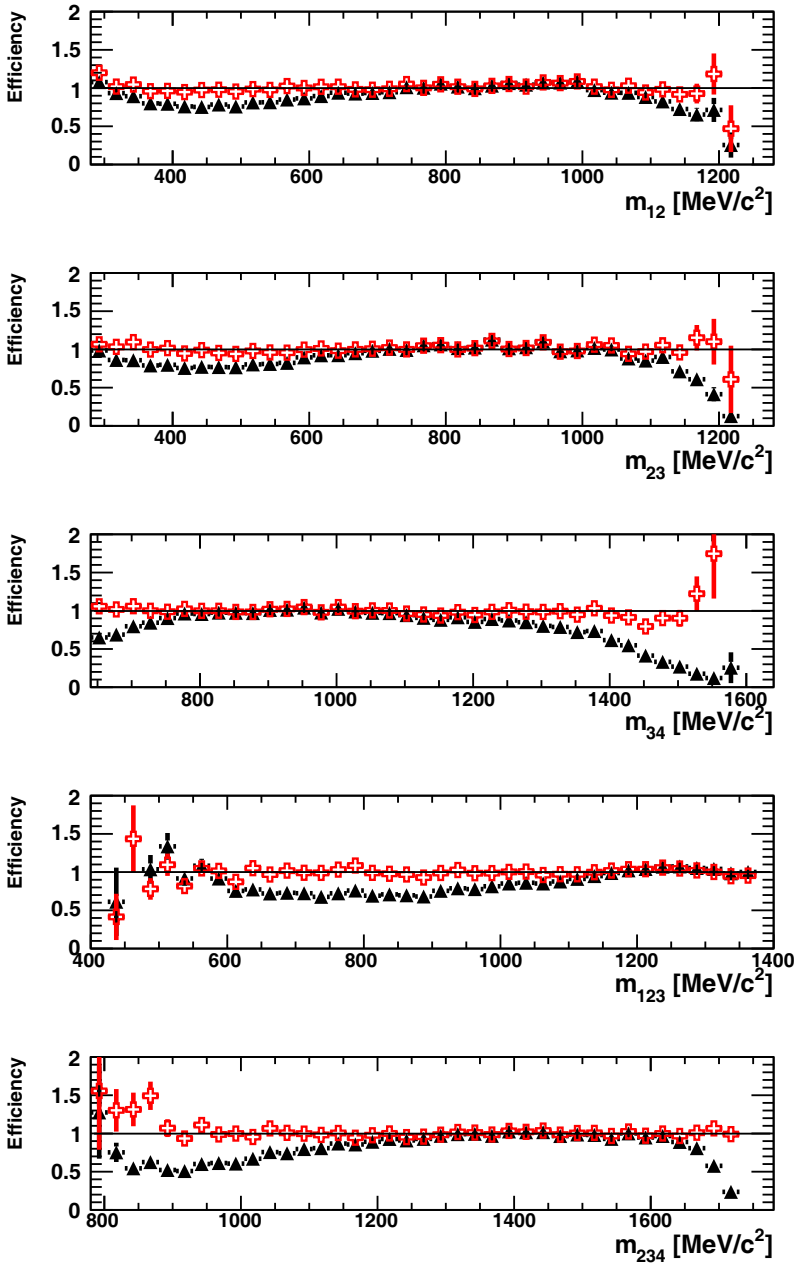

Fig. 8. Efficiency in m_{12}, m_{23}, m_{34}, m_{123} and m_{234} in the data generated for this study, with a tighter selection for the distorted sample. Shown are the ratios of the distributions found in the distorted and original samples, with no correction (full black triangles) and for decays re-weighted using $\omega_{\rm i}$ weights (red open crosses) as explained in the text. The absolute normalisation is arbitrary when the correction is not applied and natural when it is applied.

To overcome these difficulties, the input variables mapped to the NN's first layer are not directly q^2, $\cos\theta_l$, $\cos\theta_K$, and ϕ. We replace $\cos\theta_l$ and ϕ by $\exp\left(-\frac{\cos^2\theta_l}{4}\right)$ and $\exp\left(-\frac{\sin^2\phi}{4}\right)$, which distributions are not symmetric. Also, unlike in sect. 3, we cannot use blindly the MLP settings provided by default by `TMVA`. Instead, we devoted a limited effort (about a day of work) to a brute force optimisation. In practice:

- We further transformed the input variables so as to make their distributions gaussian, which helps the decorrelation algorithms used by the NN [4].
- We used three hidden layers instead of only one. The number of neurons constituting the first, second and third layers is 30, 8 and 4, respectively. It has to be compared with 4, the number of input variables.
- The number of training cycles was raised to 90000.

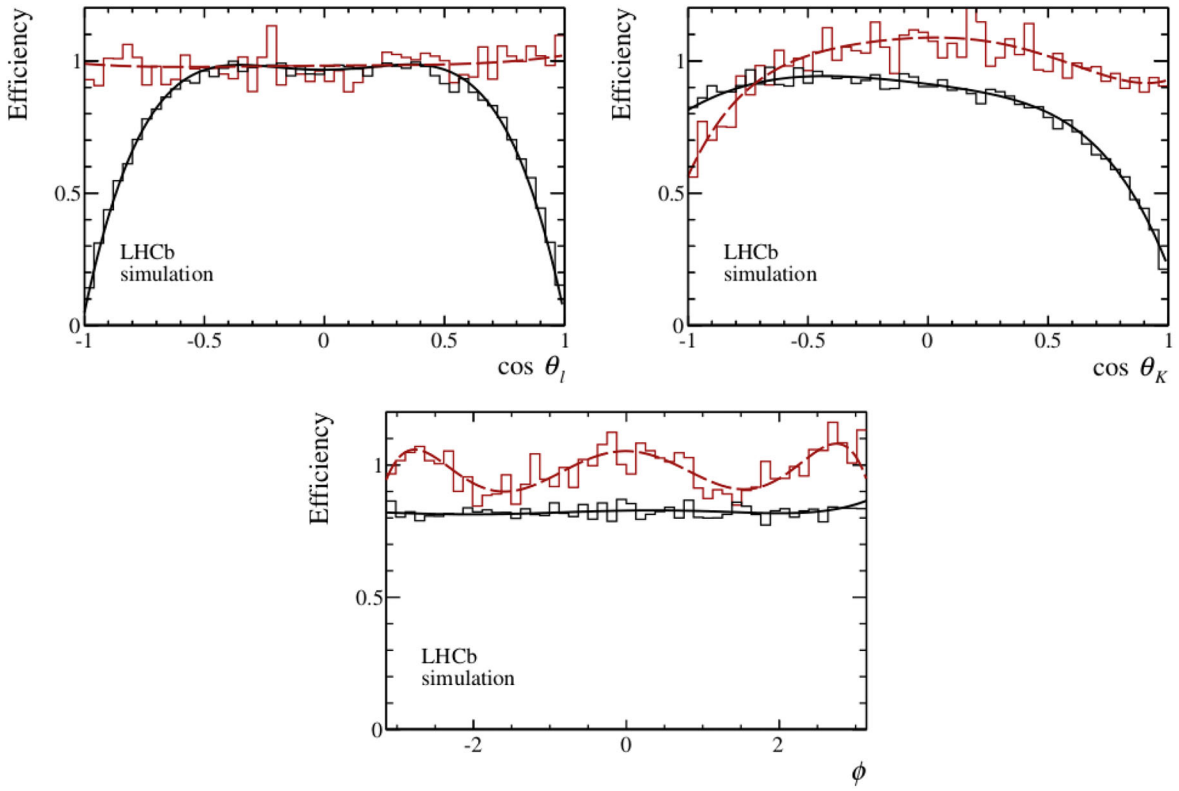

Fig. 9. Angular efficiency in $\cos\theta_l$, $\cos\theta_K$ and ϕ, as determined from a principal moment analysis of simulated three-body $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ phase-space decays (black solid and red long-dashed lines), and compared to simulated data (histograms). The efficiency is shown for the regions: $0.1 < q^2 < 0.98\,\mathrm{GeV}^2/c^4$ (black) and $18.0 < q^2 < 19.0\,\mathrm{GeV}^2/c^4$ (red). The absolute normalisation is arbitrary. This figure is reproduced from ref. [1].

– Overtraining tests were run every 10 cycles. Each time, the convergence is also tested. If 10 consecutive tests fail to observe an improvement of the error function, the training is considered optimal and stopped.

All the other settings can be found in table 19 of ref. [4]. With these settings, it took $\sim$ 12 hours to train the NN on the same machine as in sect. 3.

4.3 Results

We show on fig. 11 the distributions of the NN score s obtained in the original and distorted samples. This figure also shows $\epsilon(s)$, the parameterised efficiency fitted to the ratio of the original and distorted distributions.

We tested this efficiency in the same way as in sect. 3.3. Figure 12 shows the efficiency in q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ observed in test samples. Also shown are the efficiencies obtained with the corrected distorted test sample, in which each decay i is weigted by $\omega_\mathrm{i} = 1/\epsilon(s_\mathrm{i})$. We observe on this figure that the correction works precisely. The latter statement also holds in specific regions of the q^2 distribution, as can be seen on fig. 13, which shows this efficiency in the region $0.1 < q^2 < 0.98\,\mathrm{GeV}^2/c^4$, and on fig. 14, which focusses on the region $18.0 < q^2 < 19.0\,\mathrm{GeV}^2/c^4$. The corrected efficiencies in all three variables are consistent with a flat efficiency. When such an efficiency is fitted to the data points (horizontal line on these figures), it is consistent with 1. These results can be compared to what was obtained by the analysis reported in [1], with the principal moment analysis briefly described in sect. 2.1. The approach proposed here seems less precise statistically but of comparable accuracy. This is a promising result since it was obtained with very limited efforts and MVA-related competences. In ref. [1], LHCb determined angular observables characterising the $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ decay from an unbinned maximum likelihood fit performed in 10 regions of the q^2 spectrum. The corresponding corrected efficiencies (save in $0.1 < q^2 < 0.98\,\mathrm{GeV}^2/c^4$, already shown on fig. 13) can be seen on figures provided in appendix B (in the supplementary material). We remind that in each q^2 region the weights are still calculated from the $\epsilon(s)$ function fitted to the efficiency shown on fig. 11. No specific training and no specific evaluation of the polynomial is performed in individual regions.

As in sect. 3, our results suggest the $\epsilon(s)$ efficiency obtained with the approach proposed in this article can be used to evaluate the effiency at a given point of the decay phase space.

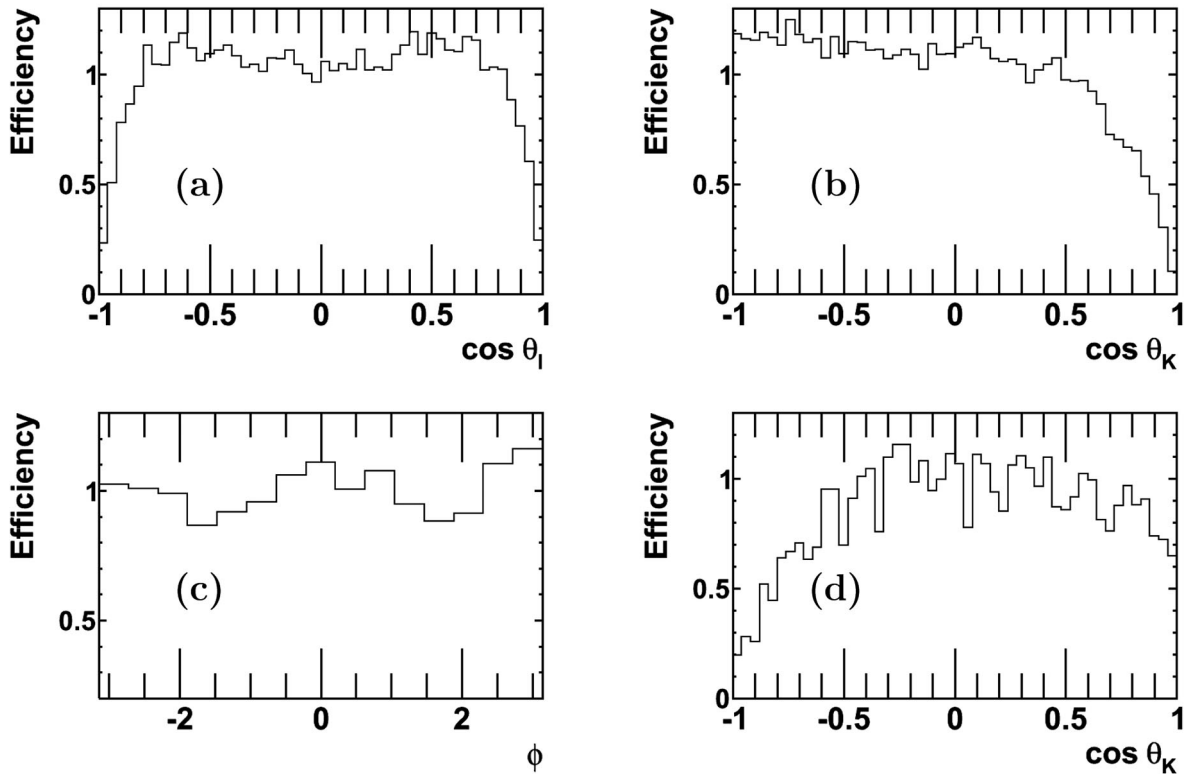

Fig. 10. Efficiency in $\cos\theta_l$, $\cos\theta_K$ and ϕ in the data generated for the present study, in (a), (b) the region $0.1 < q^2 < 0.98\,\mathrm{GeV}^2/c^4$ and (b), (c) the region $18 < q^2 < 19\,\mathrm{GeV}^2/c^4$. The absolute normalisation is arbitrary. The efficiency in $\cos\theta_l$ at high q^2 is flat and therefore not displayed here. For the same reason, the efficiency in ϕ at low q^2 is not shown.

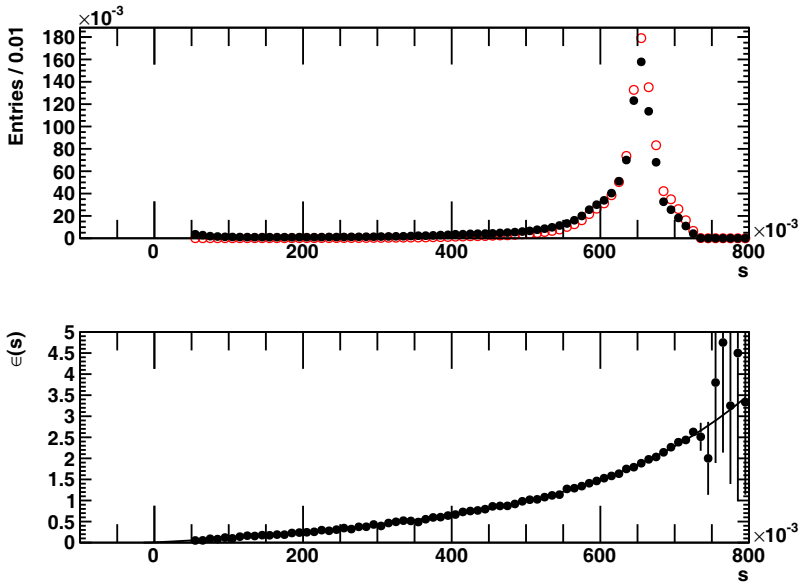

Fig. 11. Distributions and distribution ratio showing (a) the NN score s in the distorted (red open circles) and original (black full circles) $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ samples, and (b) the selection efficiency as a function of s (with an arbitrary normalisation). The fit providing $\epsilon(s)$ is superimposed to the measured efficiencies.

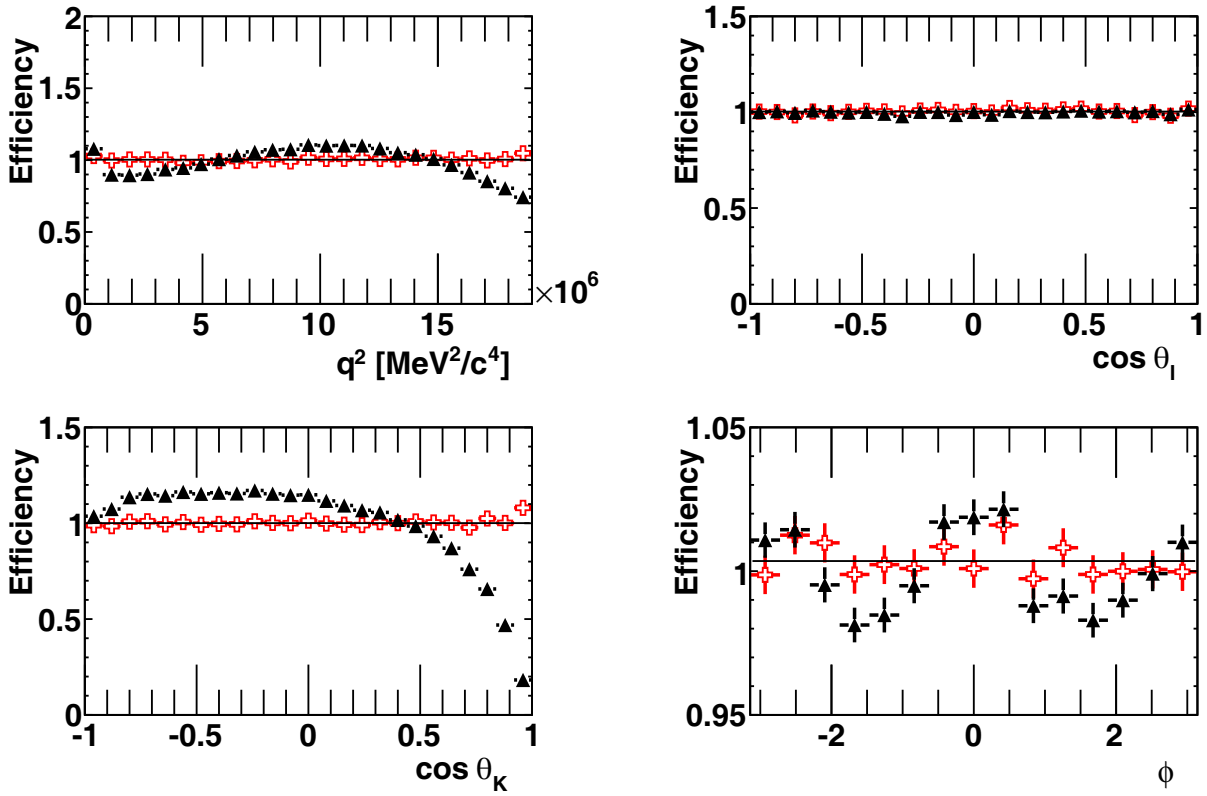

Fig. 12. Efficiency in q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ in the data generated for the present study. Shown are the ratios of the distributions found in the distorted and original samples, with no correction (black full triangles) and for decays re-weighted using ω_i weights (red open crosses) as explained in the text. The absolute normalisation is arbitrary when the correction is not applied, natural when it is.

5 Conclusion

We proposed a novel approach to the determination of multidimensional efficiencies and explored its potential with two examples which illustrate the needs of modern Heavy Flavor physics measurements: the studies of the decays $D^0 \to K^-\pi^+\pi^+\pi^-$ and $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$. We used neural networks to characterize the differences introduced in a 4- or 5-dimensional phase space by typical reconstruction and selection criteria. In both these test cases, the NN score allows to correct the selected samples in order to reproduce the phase space distributions observed in samples that have not undergone any selection. This is an evidence the approach developped here allows to evaluate the efficiency at any point of a multidimensional phase space. Compared to elaborate techniques like the principal moment method used in [1], this new approach seems less precise statistically although as accurate. It may suffice for measurements that do not require the same level of precision as the angular analysis of $B^0 \to K^{*0}(\to K^+\pi^-)\mu^+\mu^-$ reported in [1]. In such cases, it would represent a considerable gain of working time since satisfactory results can be achieved with minimum skills and knowledge of MVA techniques, using packages already routinely used within the HEP community, with no need of any elaborate optimisation nor of more than a few days of work and a few hours of CPU consumption. It also suggests that with more expertise and cutting-edge MVA techniques, a precise treatment of multidimensional efficiencies is possible and could be applied to measurements of primary importance.

Other applications of MVAs to HEP, besides signal *vs.* background discrimination, can be considered in the future. Selection efficiencies often rely on simulations that match real data only imperfectly and require systematic Data/MC comparisons to correct the simulation. When the correction must be applied to many variables, using a MVA to compare data with MC and derive "automatically" a unique number to re-weight the simulation would be a valuable tool.

In a given analysis, if one knows the efficiency at each point of the phase space, the signal events found in data can be corrected to obtained the distributions of interest before any selection bias, without having to use an imperfect simulation. This is not possible in cases where a rare decay is searched for. There, very few signal events are found, if any, and there is nothing to re-weight. The technique proposed in this article could be used to guide the selection design in order to obtain a flat $\epsilon(s)$. For that purpose, one could re-train the MVA developped for the signal selection with a weight applied to signal training events, derived from the $\epsilon(s)$ observed when the original selection is applied.

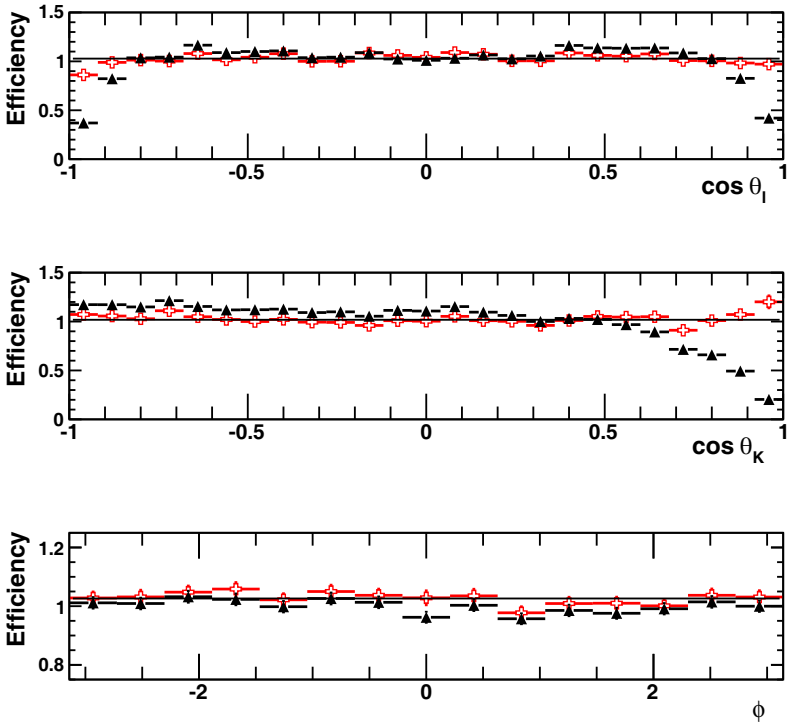

Fig. 13. Efficiency in $\cos\theta_l$, $\cos\theta_K$ and ϕ in the data generated for the present study, in the region $0.1 < q^2 < 0.98\,\mathrm{GeV}^2/c^4$. Shown are the ratios of the distributions found in the distorted and original samples, with no correction (black full triangles) and for decays re-weighted using ω_i weights (red open crosses) as explained in the text. The absolute normalisation is arbitrary when the correction is not applied, natural when it is.

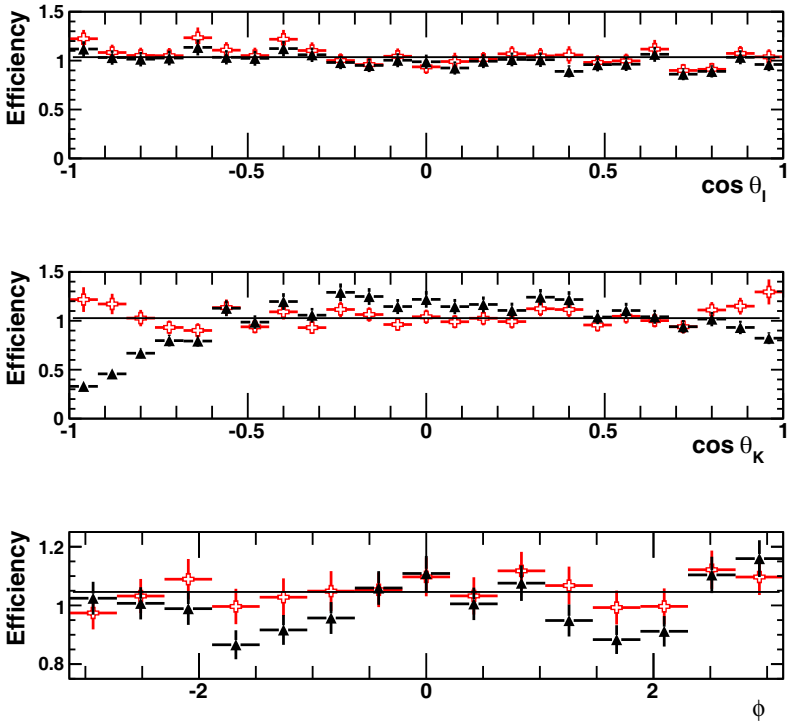

Fig. 14. Efficiency in q^2, $\cos\theta_l$, $\cos\theta_K$ and ϕ in the data generated for the present study, in the region $18.0 < q^2 < 19.0\,\mathrm{GeV}^2/c^4$. Shown are the ratios of the distributions found in the distorted and original samples, with no correction (black full triangles) and for decays re-weighted using ω_i weights (red open crosses) as explained in the text. The absolute normalisation is arbitrary when the correction is not applied, natural when it is.

A comparison with the results obtained with the technique developped in ref. [14] would be a benchmark here. If this goal is achieved, even a simulation that does not reproduce correctly the real phase space of the decay can be used to determine the efficiency. Upper limits are often normalised to the branching fraction of a well known non-suppressed decay which decay products are identical to the signal's. It differs from the signal only due to a different phase space. Ensuring for both modes a flat efficiency across the phase space would make their efficiency ratio closer to one and more robust against systematic uncertainties.

I thank Yasmine Amhis and Marie-Hélène Schune, as well as other members of the LHCb collaboration, Vincenzo Vagnoni, Carla Göbel, Angelo Di Canto, Patrick Spradlin, Tom Blake, Marc-Olivier Bettler and Albert Navarro Puig for their reading of the first version of this paper and their constructive comments. I also thank Marie-Hélène Schune for having guided me in the usage of the package that determines whether a track is in the acceptance of LHCb (see [13]).

References

1. LHCb Collaboration (R. Aaij *et al.*), JHEP **02**, 104 (2016) arXiv:1512.04442.
2. S. Haykin, *Neural Networks and Learning Machines*, 3rd edition (Prentice Hall, 2009).
3. B.P. Roe *et al.*, Nucl. Instrum. Methods A **543**, 577 (2005) arXiv:physics/0408124.
4. A. Hoecker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS ACAT (2007) 040, arXiv:physics/0703039.
5. F. Beaujean, M. Chrząszcz, N. Serra, D. van Dyk, Phys. Rev. D **91**, 114012 (2015) arXiv:1503.04100.
6. R. Brun, F. Rademakers, Nucl. Instrum. Methods A **389**, 81 (1997).
7. Particle Data Group (K.A. Olive *et al.*), Chin. Phys. C **38**, 090001 (2014).
8. LHCb Collaboration (R. Aaij *et al.*), Int. J. Mod. Phys. A **30**, 1530022 (2015) arXiv:1412.6352.
9. LHCb Collaboration (R. Aaij *et al.*), Nucl. Phys. B **871**, 1 (2013) arXiv:1302.2864.
10. R. Aaij *et al.*, JINST **8**, P04022 (2013) arXiv:1211.3055.
11. J.-H. Zhong *et al.*, Comput. Phys. Commun. **182**, 2655 (2011) arXiv:1103.2854.
12. LHCb Collaboration (R. Aaij *et al.*), JHEP **08**, 117 (2013) arXiv:1306.3663.
13. J. Lefrançois, M.H. Schune, *Measuring the photon polarization in $b \to s\gamma$ using the $B \to K^* e^+ e^-$ decay channel*, CERN-LHCb-PUB-2009-008 (2009).
14. J. Stevens, M. Williams, JINST **8**, P12013 (2013) arXiv:1305.7248.