Iterative-Bayesian unfolding of cosmic-ray isotope fluxes measured by AMS-02

Bueno, E. F.; Barão, F.; Vecchi, M.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in University of Groningen/UMCG research database](#)

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

Iterative-Bayesian unfolding of cosmic-ray isotope fluxes measured by AMS-02

E.F. Bueno [a,*], F. Barão [b,c], M. Vecchi [a]

[a] *Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD, Groningen, The Netherlands*
[b] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP), 1649-003 Lisboa, Portugal*
[c] *Departamento de Física , Instituto Superior Técnico - IST, Universidade de Lisboa - UL, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal*

ARTICLE INFO

ABSTRACT

The measurement of the isotopic composition of cosmic rays (CRs) provides essential insights into the understanding of the origin and propagation of these particles, namely the CR source spectra, the propagation processes and the galactic halo size. The Alpha Magnetic Spectrometer (AMS-02), a CR experiment operating aboard the International Space Station since May 2011, has the capability of performing these measurements due to its precise determination of the velocity provided by its Time of Flight (TOF) and Ring Imaging Cherenkov (RICH) detectors. The correct interpretation of the data requires the measurements to be deconvoluted from the instrumental effects. The unique design of AMS-02, with more than one detector being used to measure the same particle flux, requires a novel approach to unfold the measured fluxes. In this work, we apply for the first time the so-called iterative-bayesian approach in the context of the direct measurement of the cosmic-ray fluxes. Moreover, we introduce a non-parametric regularization method for the detector response functions and a single, smooth flux prior covering the range of measurements from both detectors, TOF and RICH. The accuracy of the method is assessed using a simulated flux based on previous AMS-02 measurements and taking into account the full detector response. In addition, the estimation of the errors and a discussion about the performance of the method are also shown, demonstrating that the method is fast and reliable, allowing for the recovery of the true particle fluxes in the whole energy range.

1. Introduction

The elemental abundance of light isotopes such as ^{2}H, ^{6}Li and ^{10}Be in the Solar System is low if compared to their abundance in CRs, which indicates these are particles of secondary origin: they are produced during the interaction of primary CRs, those produced and accelerated at the sources, with the interstellar medium (ISM) [1]. Therefore, the measurement of the flux of such isotopes provides essential information about the CR composition at the source and the propagation processes in our Galaxy [2]. Moreover, radioactive species such as ^{10}Be also provide important constraints on the galactic halo size [3]. Magnetic spectrometer experiments such as AMS-01[4], PAMELA [5] and IMAX [6] have measured the fluxes of some of these isotopes. The mass identification is performed by combining the measurements of the relativistic particle velocity $\beta = \frac{v}{c}$ (v being the velocity and c being the speed of light), measured either directly or through energy deposition, and the rigidity measured by the tracker and given by $R = pc/Ze$, where p is the particle momentum, Z the atomic number and e the elementary charge. The mass, is thus given by the equation

$$m\,c^2 = \frac{R\,Z\,e}{\beta\,\gamma}\,, \tag{1}$$

where $\gamma = \frac{1}{\sqrt{1-\beta^2}}$ is the Lorentz factor. The measurements before the Alpha Magnetic Spectrometer (AMS-02), reaching at most a few GeV per nucleon (GeV/n), are limited by detector velocity and rigidity resolutions. In AMS-02, isotopes are identified either in the velocity and rigidity phase space [7], or through the use of mass templates in events binned in velocity [8].

Additionally to particle identification, one of the main challenges in particle physics experiments is the reconstruction of the true properties of the particles. The finite resolution of the detector, energy losses and other instrumental effects lead to biasing and smearing of quantities such as velocity and rigidity. In the case of the measurement of isotopic spectra, typically obtained as a function of the kinetic energy per nucleon or rigidity, one must recover the particles' spectra before interacting with the detector. Generally speaking, the true value of a variable in a given bin, x_t^j, is related to its measured value, x_m^i, through the response function of the detector, $R(x_m^i|x_t^j)$:

$$x_m^i = \sum_j R(x_m^i|x_t^j)\,x_t^j\,. \tag{2}$$

The problem is then to invert the previous equation, obtaining the values of x_t^j. This is a typical ill-posed inverse problem [9]. Naively

---

* Corresponding author.
E-mail addresses: e.ferronato.bueno@rug.nl (E.F. Bueno), barao@lip.pt (F. Barão), m.vecchi@rug.nl (M. Vecchi).

inverting Eq. (2) leads to unstable results because of the nature of the problem: as the response function smears and introduces biases on the measured values, statistical fluctuations are amplified by the inversion procedure [10]. Hence, other methods have to be used to obtain *estimators* of x_t. This estimation procedure is frequently called unfolding, and it has become commonplace in particle and astroparticle physics.

Several approaches have been developed throughout the years to unfold the data [10,11]. Techniques such as Tikhonov regularization [12] and singular value decomposition (SVD) [13] are used in unfolding to solve Eq. (2) adding regularization terms, resulting in solutions with desirable properties. Others rely upon alternative techniques, which do not solve Eq. (2), as is the case of the forward unfolding [14] and the fully bayesian approach [15].

In this paper, we will focus on the application of an iterative approach, based on Bayes' theorem, similar to the works of D'Agostini [16], Richardson [17], Lucy [18], Mülthei and Schorr [19]. This approach has been successfully applied in many contexts, from collider experiments [20,21] to the study of extensive air showers [22] and the direct measurement of CR nuclei fluxes [5]. We have applied this method for the first time in the context of the measurement of isotopic fluxes with AMS-02, where the events are counted as a function of the velocity measured by two different detectors, TOF and RICH. We introduce regularization in the prior, which combines the measurements in overlapping regions, and in the detector's response function.

This paper is organized as follows: in Section 2 we give an overview of AMS-02, describing the TOF and RICH in more detail. In Section 3 we recall the ingredients for the calculation of the flux, followed by the development of the flux unfolding expressions based on Bayes' theorem. In Section 4 we describe in detail the application of the proposed unfolding technique in mock fluxes in AMS-02, from the construction of the response function to the error estimation. We conclude in Section 5.

2. The Alpha Magnetic Spectrometer

AMS-02 is a particle physics detector operating aboard the International Space Station since May 2011, having collected over 200 billion events so far. AMS-02 consists of different detectors [23]: the nine-layer silicon tracker, combined with the 0.15 T permanent magnet, measures the magnitude and sign of the charge, and the rigidity ($R = pc/Ze$) of the particle; the Transition Radiation Detector, used to separate leptons from hadrons; the Time of Flight (TOF), responsible for measuring the velocity of the particle, also acting as the main trigger of AMS-02; the Ring Imaging Cherenkov Detector (RICH), which measures the magnitude of the charge, as well as the velocity; the Anti-Coincidence Counter (ACC), that rejects particles with high-incidence angle; and the Electromagnetic Calorimeter (ECAL), located right below the RICH, used to measure the energy of the particle, as well as to separate leptons from hadrons.

AMS-02 can measure the fluxes of isotopes in CRs in a wide energy range because of the accuracy and complementarity of the velocity measurements provided by the TOF and the RICH. The TOF consists of two pairs of layers of scintillator planes, one above and the other below the magnet. The time taken for a particle to travel between the pairs of scintillators while traversing the magnetic field, Δt, is registered. The length of the path taken by the particle, L, is obtained with the track reconstruction made by the silicon tracker. In possession of these quantities, the velocity β is simply $\beta = L/c\Delta t$, and has a resolution $\Delta\beta/\beta \approx 4\%$ for single-charged particles and $\beta \approx 1$ [24]. The RICH is a proximity-focusing detector which measures β using the Cherenkov effect. It includes a radiator plane in which particles traveling with a velocity above the speed of light in the medium emit electromagnetic radiation peaked in the ultraviolet region [25]. The radiation emitted is detected below in a plane of 4×4 pixelized photomultiplier tubes (PMTs). The PMTs of the detection plane are arranged to be outside the ECAL acceptance, thus reducing the amount of matter traversed

by the particles before they reach the calorimeter. Considering this arrangement, the radiator plane is made of two materials; its center consists of sodium fluoride (RICH-NaF), with a refraction index $n = 1.33$ and threshold of emission $\beta = 0.75$. The NaF produces broad Cherenkov rings which avoid the ECAL hole, increasing the acceptance of the detector. Surrounding this material there is the aerogel (RICH-AGL) with a refraction index $n = 1.05$ and threshold of emission $\beta = 0.95$. The velocity resolution in the NaF and in the AGL are $\Delta\beta/\beta \approx 0.35\%$ and $\Delta\beta/\beta \approx 0.12\%$, respectively, for particles with $Z = 1$ and $\beta \approx 1$ [23].

Monte Carlo (MC) simulations of protons were also used in this study. These simulations were produced by the AMS-02 collaboration through a dedicated software based on the GEANT4 package [26]. The software simulates the interactions of particles with the AMS-02 material. It produces detector responses, which are then used to reconstruct the desired event properties in the same way as in data.

3. Flux unfolding

The calculation of the particle flux is made by binning the events in velocity, kinetic energy per nucleon or rigidity (depending on the experiment) and requires the knowledge of the detector surface and solid angle (geometrical acceptance) together with the detector exposure time and event selection efficiencies. Hence, the particle flux can be expressed as,

$$\Phi_m(x) = \frac{N(x)}{A\,\varepsilon(x)\,T(x)\,\Delta x}\,, \tag{3}$$

where $N(x)$ is the number of detected events in a finite interval of x, $\varepsilon(x)$ the trigger and selection efficiencies estimated from Monte Carlo simulation, A the constant geometric acceptance of the detector, $T(x)$ the detector exposure time for particles with rigidity higher than the maximal geomagnetic cutoff computed with IGRF [27], within the detector acceptance; and Δx the interval width.

Since $N(x)$ includes the instrumental effects, the obtention of the true number of events requires the application of an unfolding method. The following expression relates the measured number of events in a given bin, $N(x_m^i)$, to its true values, $N(x_t^j)$,

$$N(x_m^i) = \sum_j R(x_m^i|x_t^j)\,\varepsilon(x_t^j)\,\eta_0(x_t^j)\,N(x_t^j)\,, \tag{4}$$

where the factor $R(x_m^i|x_t^j)$ is the response function of the detector, and $\eta_0(x_t^i)$ is an efficiency obtained from the response function that translates the probability of a true value being observed inside the detection range, that is, $\eta_0(x_t^j) = \sum_i R(x_m^i|x_t^j)$. The selection efficiency, $\varepsilon(x_t^i)$, although estimated from simulated events as a function of the true variable, x_t, needs to be corrected for eventual discrepancies between data and simulations on the efficiencies of selection criteria, that is

$$\varepsilon(x_t) \rightarrow \varepsilon(x_t) \underbrace{\prod_k \frac{\varepsilon_{Data}(x_m^k)}{\varepsilon_{MC}(x_m^k)}}_{\delta(x_m)}\,, \tag{5}$$

where ε_{MC} and ε_{Data} are, respectively, efficiencies as estimated from simulations and experimental data for the set (k) of data selection criteria. This ensures that the obtained efficiency values agree with those from the actual experiment. Hence, Eq. (4) becomes

$$\frac{N(x_m^i)}{\delta(x_m^i)} = \sum_j R(x_m^i|x_t^j)\,\varepsilon(x_t^j)\,\eta_0(x_t^j)N(x_t^j)\,. \tag{6}$$

The goal of the unfolding procedure is to obtain the true flux, $\Phi_t(x_t)$, which corresponds to the incident flux purged from instrumental effects. To this end, the true number of events, $N(x_t)$ must be obtained. The inversion of Eq. (6) is intrinsically unstable due to the properties of $R(x_m^i|x_t^j)$ [11], and therefore it cannot be done directly. Using Bayes' theorem, one can estimate the probability of the true value being x_t in the bin i, when x_m was measured

$$P(x_t^i|x_m^j) = \frac{P(x_m^j|x_t^i)\,P(x_t^i)}{\sum_k P(x_m^j|x_t^k)\,P(x_t^k)}\,, \tag{7}$$

where the probability $P(x_m^j|x_t^i)$ represents the *likelihood* that x_m in bin j was produced by x_t in bin i, while $P(x_t^i)$ represents the *prior* knowledge of x_t. As a result, the probability $P(x_t^i|x_m^j)$ is the *posterior*, or the updated knowledge of x_t. In the context of this problem, the likelihood is the response function of the detector $R(x_m|x_t)$, and the prior is the previous knowledge about $N(x_t)$. With an estimate of the response function, and with some prior knowledge of $N(x_t)$, it is possible to use Eq. (7) to estimate the true value from the measured one. This is done in the following equation, where the true number of events in the bin i is derived from the accumulated number of events, the efficiency corrections and the posterior probability:

$$\hat{N}_t(x_t^i) = \frac{1}{\varepsilon(x_t^i)\,\eta_0(x_t^i)} \sum_j \frac{N_m(x_m^j)}{\delta(x_m^j)} \, P(x_t^i|x_m^j) , \qquad (8)$$

where $\hat{N}_t(x_t^i)$ represents the *estimator* of $N(x_t^i)$, and is called the unfolded number of events. The terms outside the sum represent the efficiency corrections. Since the posterior is naturally dependent on the prior, a convergent iterative process has to be established, with $\hat{N}_t(x_t^i)$ being used to define a prior for the next iteration. In each step, the prior is regularized by means of a fit to the data. The convergence of the method is evaluated through the use of the ξ estimator:

$$\xi_{k+1,k} = \sum_i \left(\frac{\hat{N}_t^{k+1}(x_t^i) - \hat{N}_t^k(x_t^i)}{\hat{N}_t^k(x_t^i)} \right)^2 , \qquad (9)$$

where k is the number of the iteration.

In the next section, the method will be applied in the context of the AMS-02 experiment using a simulated flux that mimics the actual conditions of the experiment.

4. Application to mock AMS-02 fluxes

In order to show the method at work in the context of single-charged isotopes flux measurements in AMS-02, it has been applied to mock fluxes. Considering the precise measurements of the velocity provided by the TOF and the RICH detectors, the fluxes are measured as a function of the velocity in three overlapping ranges: $0.5 < \beta < 0.9$ in the TOF; $0.78 < \beta < 0.99$ in the RICH-NaF; and $0.96 < \beta < 0.996$ in the RICH-AGL. The events were generated following the proton spectrum parametrization given in [28], that was adjusted to the AMS-02 proton flux [29] obtained with data collected from May 2011 to November 2013. The events were then smeared and biased according to the three detector response functions. In order to reproduce the effect of data selection efficiencies, generated events were weighted according to the selection efficiencies in every range. Fig. 1 shows the measured and generated fluxes as a function of kinetic energy per nucleon, $E_k = (\gamma - 1)m_p c^2$, where γ is the Lorentz factor and m_p is the proton mass. The noticeable differences between the true and measured fluxes emphasize the necessity of using a method to correct the measurements for the detector response.

In the following subsections, we describe the construction of the posterior probability for the unfolding procedure.

4.1. Detector response function regularization

The detector response functions, often referred to as migration matrices, are obtained from the Monte Carlo simulations of the detector. As in the isotopic flux measurements, the events are counted in bins of velocity (or, equivalently, kinetic energy per nucleon). These matrices correspond to binned distributions of β_t, the velocity generated in the simulation, versus β_m, the velocity as reconstructed in the detector. Fig. 2 shows these response functions for the TOF (a), RICH-NaF (b) and RICH-AGL (c). The TOF has a larger velocity bias at low energies due to the energy loss undergone by the particles. Both RICH-NaF and RICH-AGL are less penalized by this effect.

Given that the simulations have a finite number of events, regularizing the matrices is a crucial step to avoiding fluctuations caused
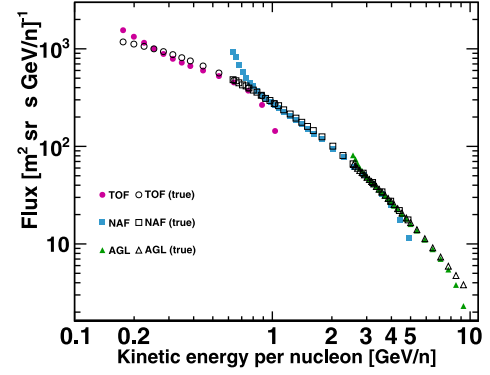

Fig. 1. True (open symbols) and measured (solid symbols) mock proton fluxes versus the kinetic energy per nucleon for the TOF (circles), RICH-NaF (squares) and RICH-AGL (triangles).

by the statistical uncertainties. To achieve this goal, we use the Kernel Density Estimation (KDE) [30], a non-parametric method which eases the smoothing of features that would need complex parametrizations to be described, albeit at a higher computational cost.

To this end, the one-dimensional projections used for calculating the likelihood are regularized to obtain smooth distributions. Fig. 3 shows examples of these projections for the TOF, RICH-NaF, and RICH-AGL in panels (a), (b), and (c), respectively, for three velocity intervals. The β_m distribution in the TOF is composed of a Gaussian core and a tail corresponding to events whose velocity reconstruction was heavily affected by the energy loss. For the two RICH regions, tails on the measured velocity are visible for two main reasons: the low number of Cherenkov photons in the ring produced by single-charged particles and the presence of additional hits in the RICH PMT plane caused by secondary particles from protons interacting with the AMS-02 material.

The obtained parametrizations, $f(\beta_m|\beta_t)$, were then used to calculate the probabilities $P(\beta_m^j|\beta_t^i)$ by integrating them over the bins of measured velocity, β_m^j, for a given bin of true velocity, β_t^i:

$$P(\beta_m^j|\beta_t^i) = \int_{\beta_m^j} f(\beta_m|\beta_t)\, d\beta_m . \qquad (10)$$

This process is repeated for every bin, allowing for the construction of a smooth migration matrix.

4.2. Prior construction

The prior, representing the degree of belief in the true variable, is a critical component in calculating the posterior defined in Eq. (7). Aiming to demonstrate our method's validity even in the most conservative case, we chose to adopt a prior built from a flat flux in which we have essentially no information about the shape of the true flux. Despite our conservative choice, our method converges after few iterations. It is also worth noting that due to the iterative nature of the method, the choice of the prior used at the very first step of iteration has a minimal impact on the final result.

After every iteration, the unfolded number of events, $\hat{N}(E_k(\beta)_t)$, is calculated using Eq. (8) for every detector range, allowing for the calculation of the corresponding flux according to the equation

$$\Phi(E_k)_t = \frac{\hat{N}(E_k)_t}{A\,T(E_k)_t\,\Delta E_k} . \qquad (11)$$

These fluxes are regularized through a fit of the model described in [28], performed on the three ranges simultaneously as shown in Fig. 4. This parametrization represents an updated degree of belief in the flux as a function of the true variable and is used to obtain the prior for the next iteration through Eq. (11): the updated flux is used
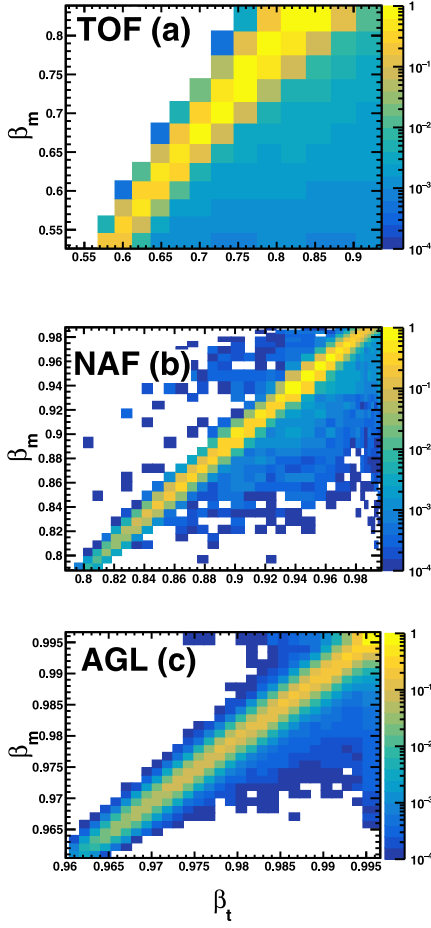
Fig. 2. Migration matrices of the TOF (a), RICH-NaF (b) and RICH-AGL (c). See text for discussion.

Fig. 3. The resolution functions for the TOF (a), RICH-NaF (b) and RICH-AGL (c) in different true velocity bins. See text for discussion.

Fig. 4. Model (black dashed line) adjusted to the results of the mock proton flux after two unfolding iterations in the TOF (pink circles), RICH-NaF (blue squares) and RICH-AGL (green triangles).

together with the acceptance, exposure time and bin width to compute the new prior probability of $N(E_k)_t$. The advantage of such an approach is that it describes the data well, providing a smooth description of the flux, even in transition zones between the ranges, where resolution and threshold effects may cause the points to fluctuate, as seen in Fig. 4. This guarantees that the prior for the next iteration is a smooth, continuous function. Since the fit is performed to all points simultaneously, it implicitly combines the measurements by choosing the curve that best describes the data, ensuring that the unfolded fluxes match in the transition zones between the detectors. It also prevents the fluctuations that may come from the unfolding procedure from propagating to the next iteration, ensuring the method converges.

4.3. Error estimation

The calculation of the unfolded number of events depends directly on the detector response function, as well as on the number of measured events. These quantities have errors associated with them, which are propagated to the result of the unfolding. In order to estimate the contributions from each of these quantities, a bootstrapping method was implemented. This method consists of generating n randomized sets of data based on the original one and performing the unfolding procedure in all sets, resulting in a distribution of unfolding results with n entries per bin. This allows for calculating the covariance matrices, which contain the total uncertainties in each bin in their diagonals. Considering the iterative nature of the procedure, where the results of one iteration are used as the prior for the next, the propagation of the uncertainty is done by sampling the prior.
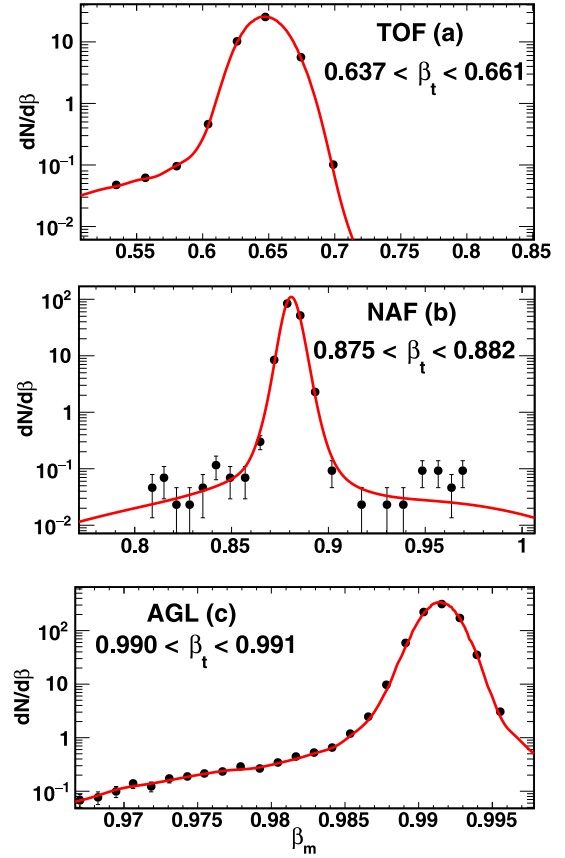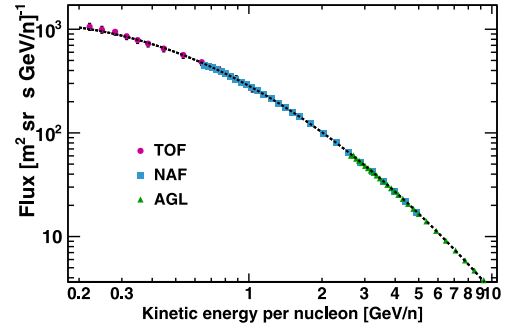
Assuming that the uncertainties on the number of measured events and on the migration matrices are uncorrelated, they are treated separately. For the contribution from the number of measured events, 10^3 sets of measurements were generated by sampling new values $N(E(\beta)^i_m)$ in every bin according to a Gaussian distribution. That is, the number of events in each bin i of each new set, $N'(E(\beta)^i_m)$, is obtained by sampling a Gaussian distribution G centered at $N(E(\beta)^i_m)$, with its root mean squared (RMS) given by $\sqrt{N(E(\beta)^i_m)}$, that is

$$N'(E(\beta)^i_m) = G(N(E(\beta)^i_m); \sigma_N). \qquad (12)$$

The unfolding procedure was applied to the complete set of measurements, yielding a distribution of unfolded counts. Using the distribution
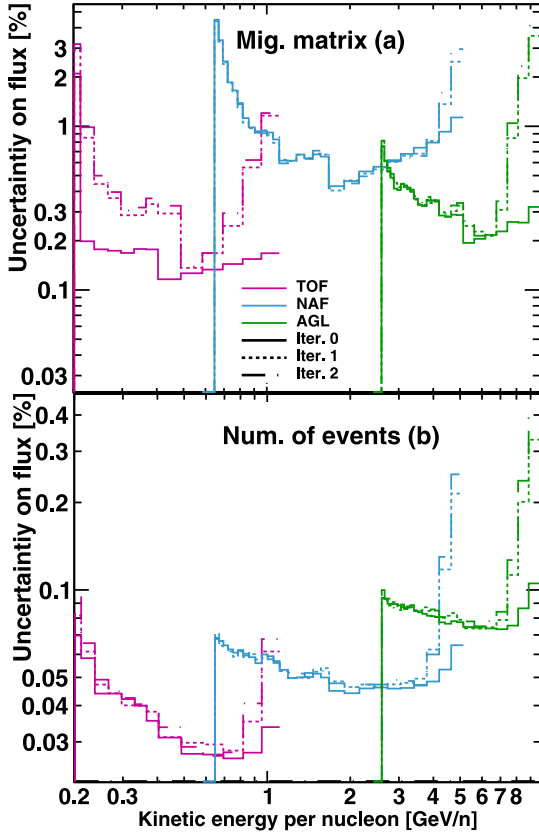
Fig. 5. Relative error on unfolded flux for the migration matrix contribution (top) and on the number of measured events (bottom). Pink lines represent the TOF, blue the NAF, and green the AGL. Solid lines are the errors from iteration 0 (flat prior), while dotted and dot-dashed correspond to the errors for iterations 1 and 2, respectively.

Fig. 6. Residuals and ξ calculated for consecutive iterations in the three velocity ranges. The top panel shows the residuals between iterations 0 and 1, the mid panel shows between 1 and 2, and the bottom shows between 3 and 2. The different ranges are represented by different markers and colors: pink circles (TOF), blue squares (NAF) and green triangles (AGL). See text for discussion.

of the results, the covariance matrix of the unfolded number of events was obtained.

After every iteration, the prior is updated with the previously obtained unfolded counts. We associate with the prior the uncertainty obtained in the previous iteration and propagate it by sampling the unfolded counts according to their uncertainty, analogously to what is stated in Eq. (12).

A similar exercise was performed to obtain the errors coming from the uncertainties on the migration matrices. This time, the measured number of events was kept fixed, and 10^3 migration matrices were generated according to

$$P(E(\beta)_m^j | E(\beta)_t^i) = G(P(E(\beta)_m^j | E(\beta)_t^i); \sigma_P), \qquad (13)$$

where G is a Gaussian distribution centered at the original $P(E(\beta)_m^j | E(\beta)_t^i)$, and width given by σ_P, that represents the uncertainty on each migration matrix bin. Analogously to the previous case, the unfolding is performed with the 10^3 different matrices, resulting in a covariance matrix unfolded counts per velocity range, from where the uncertainties were obtained.

The relative errors for both contributions, coming from the diagonal of the covariance matrices, are shown for the different ranges and iterations in Fig. 5. Panel (a) shows the uncertainty coming from the migration matrices. In all cases, the errors are larger in the extremes of each range. The explanation for the increase at the end is the same for the three detectors: at the upper edges of each velocity range, the velocity resolution degrades, hence increasing the migration of events. Therefore, fluctuations in the migration matrices lead to fluctuations in the result. The increase at the beginning of each range is due to
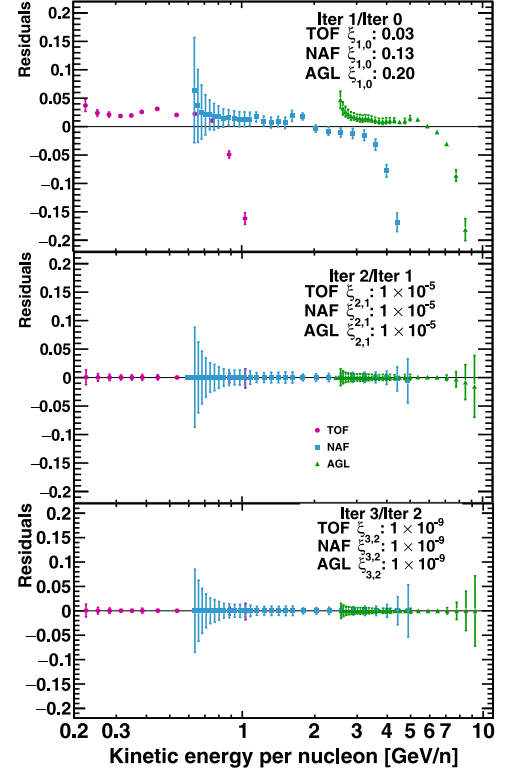
different physical phenomena: in the TOF, it is due to the energy loss, which produces a strong bias in the response function at the lowest energies. Hence any response fluctuations in that region of the matrix produce significant changes in the unfolded results. The errors on the RICH follow a similar trend, but in this case this behavior is explained by the Cherenkov emission kinetic energy per nucleon threshold. Events close to the threshold have a smaller probability of being reconstructed. Hence the migration matrix is less populated in this region. As a result, the dispersion of the unfolding results increases, increasing the error. The same behavior is seen in panel (b), where the uncertainty coming from the number of events is shown.

The errors from both contributions increase progressively with the number of iterations, as expected from the error propagation done via the prior sampling. The error increases faster from iteration 0 to 1 than from 1 to 2. This is due to the change in the prior: while iteration 0 has a prior which is flat in the entire energy range, iterations 1 and 2 use a model which is fitted to the results, leading to larger uncertainties.

Once both contributions are estimated, the total error of the unfolding procedure is given by the sum in quadrature of the uncertainties from the migration matrix and the number of events.

4.4. Number of iterations

Considering the iterative nature of this method, the number of iterations is an important parameter. Too few iterations lead to a result far from the true value, while a large number of iterations leads to a larger error due to the error propagation. Therefore, one must stop iterating as soon as the differences between two consecutive iterations

are small enough. Fig. 6 shows the residuals between consecutive iterations, defined as

$$R_i^{k+1,k} = \frac{N_i^{k+1} - N_i^k}{N_i^{k+1}}, \qquad (14)$$

where k is the number of the iteration and i is the bin index. In addition, the value of ξ defined in Eq. (9) is shown for every range in each of the panels. The largest changes are between iteration 0 and 1 due to the use of a flat prior in iteration 0. Iterations 1 and 2 already show signs of convergence, with the residuals displaying small fluctuations in the last points of the AGL range. The bottom panel shows that iterations 3 and 2 yield essentially the same results, with $\xi_{3,2} \approx 10^{-9}$ meaning that it is enough to stop at iteration 2.

The similarity between iterations 3 and 2 highlights the importance of regularizing the prior in each iteration. The spline fit removes eventual fluctuation that appear in the calculations and ensures convergence, whereas the methods with no regularization have oscillations due to positive feedback of these fluctuations [16].

4.5. Comparison with true flux

The top panel of Fig. 7 shows the unfolded flux after three iterations, calculated with Eq. (11), for the different ranges, while in the bottom, the residuals comparing the results with the true fluxes are shown. Generally, the agreement between the unfolded and true values is very good, being below 1% in the entire measurement range. Minor fluctuations are seen at the beginning of the RICH-NaF and RICH-AGL ranges. These are related to steep drops in the number of measured events due to the efficiency of the detector, which makes the true distribution harder to obtain in these regions. In the case of the RICH the efficiencies drop rapidly below 0.8 and 3 GeV/n, respectively, due to the emission threshold of Cherenkov radiation. Still, the method can recover the true distribution quite well in these regions. There is also a good agreement between the detectors in the overlapping regions.

5. Conclusion

The study of the isotopic composition of CRs will provide important information about the propagation of CRs in our own Galaxy, as well as about the CR source composition and the Galactic halo size. These measurements are intrinsically challenging due to several factors: large backgrounds, as in the case of single-charged isotopes, and small mass difference, as is the case of Be isotopes. As a consequence, the available measurements are limited to the kinetic energy per nucleon range from sub-GeV/n to a few GeV/n. The AMS-02 experiment, with its precise measurements of the rigidity and the velocity, especially provided by the Cherenkov detector, has the capability of measuring the fluxes of isotopes reaching energies up to 10 GeV/n. The correct interpretation of these data depends on correcting the results for the detector response. The unique design of AMS-02, with three different measurements of the velocity, requires special treatment. In this work, we applied an iterative unfolding method based on Bayes' theorem in the context of the isotopic flux measurements in AMS-02. The general mathematical aspects of the method were discussed, emphasizing the importance of data/simulation and efficiency corrections. In order to demonstrate the applicability and performance of the method, a mock flux was generated according to the measured proton spectrum, which was then smeared and biased according to the response function of the different detectors included in the analysis. Regularization was also introduced in two stages of the method. The response function was regularized with KDEs, which could readily describe the non-Gaussian features of the measurements. In addition, the unfolding of the flux in the three different ranges was connected via a common prior, obtained through the fit of a model using all the available points. This fit leads to the smoothing of features in the flux, such as the transition between ranges, and prevents fluctuations from leading the method to calculation to
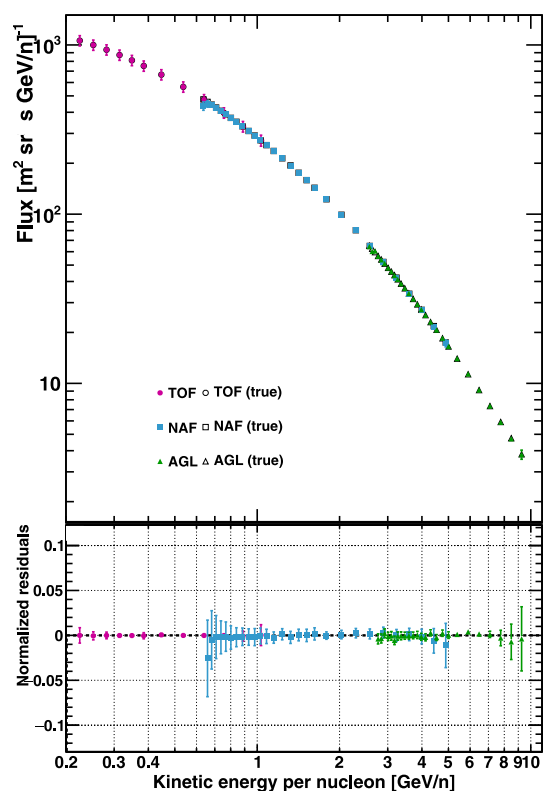

Fig. 7. True (open symbols) and unfolded (solid symbols) mock proton fluxes after three iterations (top) and the normalized residuals (bottom). See text for discussion.

diverge. The error, estimated through a bootstrapping procedure, has shown strong dependence on the migration matrix, growing rapidly in regions with high bias or low statistics. Finally, the results have shown that only a few iterations of this procedure are enough to recover the true spectrum within the uncertainties, enabling the true flux to be estimated up to 10 GeV/n. The fast convergence of this method also makes it suitable for studying the time-variability of isotopic fluxes, where a large number of flux unfolding calculations are necessary.

CRediT authorship contribution statement

E.F. Bueno: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. **F. Barão:** Conceptualization, Methodology, Software, Writing – review & editing. **M. Vecchi:** Writing – review & editing, Funding acquisition, Visualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] T.K. Gaisser, R. Engel, E. Resconi, Cosmic Rays and Particle Physics, second ed., Cambridge University Press, 2016, http://dx.doi.org/10.1017/CBO9781139192194.

[2] B. Coste, et al., Constraining Galactic cosmic-ray parameters with $Z \leq 2$ nuclei, Astron. Astrophys. 539 (2012) A88, http://dx.doi.org/10.1051/0004-6361/201117927.

[3] S. Hayakawa, K. Ito, Y. Terashima, Origin of cosmic rays, Progr. Theoret. Phys. Suppl. 6 (1958) 1–92, http://dx.doi.org/10.1143/PTPS.6.1, arXiv:https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.6.1/5269796/6-1.pdf.

[4] M. Aguilar, et al., The alpha magnetic spectrometer (AMS) on the international space station: Part I – results from the test flight on the space shuttle, Phys. Rep. 366 (6) (2002) 331–405, http://dx.doi.org/10.1016/S0370-1573(02)00013-3, URL https://www.sciencedirect.com/science/article/pii/S0370157302000133.

[5] O. Adriani, et al., Measurements of cosmic-ray hydrogen and helium isotopes with the PAMELA experiment, 818, (1) 2016, p. 68, http://dx.doi.org/10.3847/0004-637x/818/1/68.

[6] G.A. de Nolfo, et al., A measurement of cosmic ray deuterium from 0.5–2.9 GeV/nucleon, AIP Conf. Proc. 528 (1) (2000) 425–428, http://dx.doi.org/10.1063/1.1324352, arXiv:https://aip.scitation.org/doi/pdf/10.1063/1.1324352, URL https://aip.scitation.org/doi/abs/10.1063/1.1324352.

[7] M. Aguilar, et al., AMS Collaboration, Properties of cosmic helium isotopes measured by the Alpha Magnetic Spectrometer, Phys. Rev. Lett. 123 (2019) 181102, http://dx.doi.org/10.1103/PhysRevLett.123.181102, URL https://link.aps.org/doi/10.1103/PhysRevLett.123.181102.

[8] E. Bueno, F. Barão, M. Vecchi, A parametric approach for the identification of single-charged isotopes with AMS-02, Nucl. Instrum. Methods Phys. Res. A 1031 (2022) 166564, http://dx.doi.org/10.1016/j.nima.2022.166564, URL https://www.sciencedirect.com/science/article/pii/S016890022200167X.

[9] V. Blobel, Unfolding methods in particle physics, in: PHYSTAT 2011, CERN, Geneva, 2011, pp. 240–251, http://dx.doi.org/10.5170/CERN-2011-006.240.

[10] C.A. Pruneau, Data Analysis Techniques for Physical Scientists, Cambridge University Press, 2017, http://dx.doi.org/10.1017/9781108241922.

[11] G. Cowan, A survey of unfolding methods for particle physics, in: M.R. Whalley, L. Lyons (Eds.), Conf. Proc. C 0203181 (2002) 248–257.

[12] A.N. Tikhonov, Solution of incorrectly formulated problems and the regularization method, Soviet Math. Dokl. 4 (1963) 1035–1038.

[13] A. Höcker, V. Kartvelishvili, SVD approach to data unfolding, Nucl. Instrum. Methods Phys. Res. A 372 (3) (1996) 469–481, http://dx.doi.org/10.1016/0168-9002(95)01478-0, URL https://www.sciencedirect.com/science/article/pii/0168900295014780.

[14] J. Albert, et al., Unfolding of differential energy spectra in the MAGIC experiment, Nucl. Instrum. Methods Phys. Res. A 583 (2) (2007) 494–506, http://dx.doi.org/10.1016/j.nima.2007.09.048, URL https://www.sciencedirect.com/science/article/pii/S0168900207020840.

[15] G. Choudalakis, Fully Bayesian unfolding, 2012, http://dx.doi.org/10.48550/ARXIV.1201.4612.

[16] G. D'Agostini, Improved iterative Bayesian unfolding, in: Alliance Workshop on Unfolding and Data Correction, 2010, arXiv:1010.0632.

[17] W.H. Richardson, Bayesian-based iterative method of image restoration*, J. Opt. Soc. Amer. 62 (1) (1972) 55–59, http://dx.doi.org/10.1364/JOSA.62.000055, URL http://opg.optica.org/abstract.cfm?URI=josa-62-1-55.

[18] L.B. Lucy, An iterative technique for the rectification of observed distributions, Astron. J. 79 (1974) 745, http://dx.doi.org/10.1086/111605.

[19] H. Mülthei, B. Schorr, On an iterative method for the unfolding of spectra, Nucl. Instrum. Methods Phys. Res. A 257 (2) (1987) 371–377, http://dx.doi.org/10.1016/0168-9002(87)90759-5, URL https://www.sciencedirect.com/science/article/pii/0168900287907595.

[20] G. Aad, et al., ATLAS Collaboration, Measurements of $W\gamma$ and $Z\gamma$ production in pp collisions at $\sqrt{s}$ = 7TeV with the ATLAS detector at the LHC, Phys. Rev. D 87 (2013) 112003, http://dx.doi.org/10.1103/PhysRevD.87.112003, URL https://link.aps.org/doi/10.1103/PhysRevD.87.112003.

[21] G. Aad, et al., ATLAS Collaboration, Measurement of the transverse momentum distribution of W bosons in pp collisions at $\sqrt{s}$ = 7TeV with the ATLAS detector, Phys. Rev. D 85 (2012) 012005, http://dx.doi.org/10.1103/PhysRevD.85.012005, URL https://link.aps.org/doi/10.1103/PhysRevD.85.012005.

[22] A. Albert, et al., Cosmic ray spectrum of protons plus helium nuclei between 6 and 158 TeV from HAWC data, Phys. Rev. D 105 (2022) 063021, http://dx.doi.org/10.1103/PhysRevD.105.063021, URL https://link.aps.org/doi/10.1103/PhysRevD.105.063021.

[23] M. Aguilar, et al., The alpha magnetic spectrometer (AMS) on the international space station: Part II — Results from the first seven years, Phys. Rep. 894 (2021) 1–116, http://dx.doi.org/10.1016/j.physrep.2020.09.003, URL https://www.sciencedirect.com/science/article/pii/S0370157320303434.

[24] V. Bindi, et al., Calibration and performance of the AMS-02 time of flight detector in space, Nucl. Instrum. Methods Phys. Res. A 743 (2014) 22–29, http://dx.doi.org/10.1016/j.nima.2014.01.002, URL https://www.sciencedirect.com/science/article/pii/S0168900214000102.

[25] F. Giovacchini, J. Casaus, A. Oliva, The AMS-02 RICH detector: Status and physics results, Nucl. Instrum. Methods Phys. Res. A 952 (2020) 161797, http://dx.doi.org/10.1016/j.nima.2019.01.024.

[26] J. Allison, et al., Recent developments in Geant4, Nucl. Instrum. Methods Phys. Res. A 835 (2016) 186–225, http://dx.doi.org/10.1016/j.nima.2016.06.125, URL https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[27] C.C. Finlay, et al., International Geomagnetic Reference Field: the eleventh generation, Geophys. J. Int. 183 (3) (2010) 1216–1230, http://dx.doi.org/10.1111/j.1365-246X.2010.04804.x, arXiv:https://academic.oup.com/gji/article-pdf/183/3/1216/1785065/183-3-1216.pdf.

[28] M. Boudaud, et al., The pinching method for Galactic cosmic ray positrons: Implications in the light of precision measurements, Astron. Astrophys. 605 (2017) A17, http://dx.doi.org/10.1051/0004-6361/201630321.

[29] M. Aguilar, et al., AMS Collaboration, Precision measurement of the proton flux in primary cosmic rays from rigidity 1 GV to 1.8 TV with the alpha magnetic spectrometer on the international space station, Phys. Rev. Lett. 114 (2015) 171103, http://dx.doi.org/10.1103/PhysRevLett.114.171103, URL https://link.aps.org/doi/10.1103/PhysRevLett.114.171103.

[30] K. Cranmer, Kernel estimation in high-energy physics, Comput. Phys. Comm. 136 (3) (2001) 198–207, http://dx.doi.org/10.1016/S0010-4655(00)00243-5, URL https://www.sciencedirect.com/science/article/pii/S0010465500002435.