A Markov Chain Monte Carlo for Galactic Cosmic Ray physics

A. Putze[a,1], L. Derome[1], D. Maurin[2], L. Perotto[1], R. Taillet[3]

***Abstract*— Propagation of charged Cosmic Rays in the Galaxy depends on the transport parameters, which number can be large depending on the propagation model under scrutiny. Yet, a standard approach to determine these parameters is a manual scan, leading to an inefficient and incomplete coverage of the parameter space. The awaited results of forthcoming experiments call for a better strategy. An automated statistical tool is required for a full coverage and a sound determination of the transport and source parameters. We implement a Markov Chain Monte Carlo (MCMC), which is well suited for multi-parameter determination. Its capabilities and performances are explored on the phenomenologically well understood Leaky Box Model. A trial function based on binary space partitioning proves to be very efficient, allowing a simultaneous determination of up to nine parameters, including transport and source parameters (slope, abundances). The best model includes both a low energy cut-off and reacceleration, which values are compatible with those found in diffusion models. A Kolmogorov spectrum for the diffusion slope ($\delta = 1/3$) is excluded. The marginalised probability density function for δ and α (the slope of the source spectra) are $\delta \approx 0.55 - 0.60$ and $\alpha \approx 2.14 - 2.17$, depending on the dataset used and the number of free parameters in the fit.**

I. Introduction

One issue of cosmic-ray (CR) physics is the determination of the transport parameters in the Galaxy. Such a determination is based on the analysis of the secondary-to-primary ratio (e.g. B/C, sub-Fe/Fe), in which the dependence on the source spectra factors remain mostly sensitive to the propagation processes (e.g., [20] and references therein). For nearly 20 years, the determination of these parameters relied mostly on the most constraining data, namely the HEAO-3 data, taken in 1979, which cover the $\sim 1 - 35$ GeV/n range ([11]).

For the first time since HEAO-3, several satellite or balloon borne experiments (see ICRC 2007 reporter's talk [4]) have obtained better data in the same energy range, i.e. 1 TeV/n$-$PeV/n, or covered a yet scarcely explored range in terms of nucleus: the ATIC collaboration has presented some results for the B/C ratio at $0.5 - 50$ GeV/n ([25]), and for H to Fe fluxes at 100 GeV$-$100 TeV ([24]). The TRACER collaboration has recently published spectra for O up to Fe in the GeV/n-TeV/n range ([1], [5]). The CREAM experiment ([26]) flew a cumulative duration of 70 days in December 2004 and December 2005 ([27] and preliminary results in [19], [31]), and again in December 2007. A fourth flight is schedule for December 2008[1]. Finally, some exciting data will soon come from the PAMELA satellite.

It is relevant to question the method to extract the propagation parameters in order to take the best advantage of this wealth of new data. This determination is an important issue for many theoretical and astrophysical questions, as it is linked, amongst others, to the transport in turbulent magnetic fields, sources of CRs, γ-ray diffuse emission (see [30] for a recent review and references). It also proves to be crucial for indirect dark matter detection studies (e.g. [8], [9]). The usage in the past has been mostly based on a manual or semi-automated—hence partial—coverage of the parameter space (e.g., [14], [29], [33]). More complete scans have been performed in [16], [20], [21], although in an inefficient manner: the addition of a single new free parameter (as done for example in [21] compared to [20]) remains prohibitive in terms of computer time. Such a scan is generally further complicated by the observation of large degeneracies in the parameter space [20].

Therefore it is necessary to use an efficient and sound numerical tool to i) cover efficiently the parameter space and ii) enable the enlargement of the parameter space at a minimal computing time cost. The Markov Chain Monte Carlo (MCMC) algorithm, widely used for cosmological parameter estimates (see e.g. [6], [10], [15] and references therein), meets these demands.

II. Markov Chain Monte Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) techniques have proven particularly well-defined for Bayesian parameter estimates ([18], [22]). The Bayesian approach aims at assessing to which extent an experimental dataset improves our knowledge of a given theoretical model. Considering a model depending on m parameters

$$\boldsymbol{\theta} \equiv \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}\}, \tag{1}$$

we wish to determine the conditional probability density function (PDF) of the parameters given the data, $P(\boldsymbol{\theta}|\mathrm{data})$. This so-called *posterior* probability quantifies the change in the degree of belief one can have in the m parameters of the model in the light of the data. Applied to the parameter inference, Bayes theorem reads

$$P(\boldsymbol{\theta}|\mathrm{data}) = \frac{P(\mathrm{data}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})}{P(\mathrm{data})}, \tag{2}$$

[1]Laboratoire de Physique Subatomique et de Cosmologie LPSC, Grenoble, France
[2]Laboratoire de Physique Nucléaire et des Hautes Energies LPNHE, Paris, France
[3]Laboratoire de Physique Théorique LAPTH, Annecy-le-Vieux, France
[a]Corresponding author: putze@lpsc.in2p3.fr

[1]http://cosmicray.umd.edu/cream/cream.html

where $P(\mathrm{data})$ is the data probability (the latter does not depend on the parameters and hence, can be thought as a normalisation factor). This theorem links the posterior probability to the likelihood of the data $\mathcal{L}(\boldsymbol{\theta}) \equiv P(\mathrm{data}|\boldsymbol{\theta})$ and the so-called *prior* probability, $P(\boldsymbol{\theta})$, giving the degree of belief one has *before* observing the data. The technically difficult point of Bayesian parameter estimates lies in the determination of the individual posterior PDF, which requires an (high-dimensional) integration of the overall posterior density. Thus an efficient sampling method for the posterior PDF is mandatory. For models having more than a few parameters, regular grid-sampling approaches break down and statistical techniques are required [7]. Among the latter, MCMC algorithms have been fully tried and tested for Bayesian parameter inference [18], [22].

MCMC methods aim at exploring any *target* distribution of a vector of parameters $p(\boldsymbol{\theta})$, by generating a sequence of n points (hereafter a chain)

$$\{\vec{\theta}_i\}_{i=1,\ldots,n} \equiv \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n\}. \tag{3}$$

Each $\boldsymbol{\theta}_i$ is a vector of m components [e.g., as defined in Eq. (1)]. In addition, the chain is Markovian in the sense that the distribution of $\boldsymbol{\theta}_{n+1}$ is entirely influenced by the value of $\boldsymbol{\theta}_n$. MCMC algorithms are built such as the time spent by the Markov chain in a region of the parameter space is proportional to the target PDF value in this region. Hence, from such a chain, one can obtain an independent sampling of the PDF. The target PDF as well as all marginalised PDF are estimated by counting the number of samples within the related region of parameter space.

A. The algorithm

The prescription we use to generate the Markov chains from the unknown target distribution is the so-called Metropolis-Hastings algorithm (the interested reader is referred to [18], [22, chapter 29] for further details and references.). The Markov chain grows in jumping from the current point in the parameter space $\boldsymbol{\theta}_i$ to the following $\boldsymbol{\theta}_{i+1}$. As said before, the PDF of the new point only depends on the current point, i.e. $\mathcal{T}(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_i) = \mathcal{T}(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i)$. This quantity defines the transition probability for state $\boldsymbol{\theta}_{i+1}$ from the state $\boldsymbol{\theta}_i$. The Metropolis-Hastings algorithm gives a prescription on the transition probability to ensure that the stationary distribution of the chain asymptotically tends to the target PDF one wishes to sample from.

At each step i (corresponding to a state $\boldsymbol{\theta}_i$), a trial state $\boldsymbol{\theta}_{\mathrm{trial}}$ is generated from a *proposal* density $q(\boldsymbol{\theta}_{\mathrm{trial}}|\boldsymbol{\theta}_i)$. This proposal density is chosen so that samples can be easily generated (e.g. a Gaussian distribution centred on the current state). The state $\boldsymbol{\theta}_{\mathrm{trial}}$ is accepted or rejected depending on the following criterion: forming the quantity

$$a(\boldsymbol{\theta}_{\mathrm{trial}}|\boldsymbol{\theta}_i) = \min\left(1, \frac{p(\boldsymbol{\theta}_{\mathrm{trial}})}{p(\boldsymbol{\theta}_i)} \frac{q(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{\mathrm{trial}})}{q(\boldsymbol{\theta}_{\mathrm{trial}}|\boldsymbol{\theta}_i)}\right), \tag{4}$$

the trial state is accepted as a new state with a probability a (rejected with probability $1-a$). The transition probability is then

$$\mathcal{T}(\boldsymbol{\theta}_{i+1}|\boldsymbol{\theta}_i) = a(\boldsymbol{\theta}_{\mathrm{trial}}|\boldsymbol{\theta}_i) q(\boldsymbol{\theta}_{\mathrm{trial}}|\boldsymbol{\theta}_i). \tag{5}$$

If accepted, $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_{\mathrm{trial}}$, whereas if rejected, the new state is equal to the current state, $\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i$. This criterion ensures that once at its equilibrium, the chain samples the target distribution $p(\boldsymbol{\theta})$.

B. Chain analysis

The starting point $\boldsymbol{\theta}_0$ of a chain is chosen randomly. The time needed to reach a state uncorrelated to $\boldsymbol{\theta}_0$, i.e. to "forget" the starting point, is called the burn-in length b. The b first samples $\{\boldsymbol{\theta}_i\}_{i=1,\ldots,b}$ of the chain have to be discarded for a further analysis. Each step depends on the previous one, which makes the samples of the chain correlated. To obtain independent samples, thinning the chain is mandatory. The correlation length l is defined as the smallest j for which the states θ_i and θ_{i+j} of the chain are considered uncorrelated.

C. Choice of the target and trial functions

1) Target function: As already said, we wish to sample the target function $p(\boldsymbol{\theta}) = P(\boldsymbol{\theta}|\mathrm{data})$. Using Eq. (2) and the fact that the algorithm is not sensitive to the normalisation factor, this amounts to sample the product $P(\mathrm{data}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})$. Assuming a flat prior $P(\boldsymbol{\theta}) = \mathrm{cst}$, the target distribution reduces to

$$p(\boldsymbol{\theta}) = P(\mathrm{data}|\boldsymbol{\theta}) \equiv \mathcal{L}(\boldsymbol{\theta}). \tag{6}$$

Here, the likelihood function is taken as

$$\mathcal{L}(\boldsymbol{\theta}) = \exp\left(-\frac{\chi^2(\boldsymbol{\theta})}{2}\right). \tag{7}$$

The $\chi^2(\boldsymbol{\theta})$ function for n_{data} data is

$$\chi^2(\boldsymbol{\theta}) = \sum_{k=1}^{n_{\mathrm{data}}} \frac{(y_k^{\mathrm{exp}} - y_k^{\mathrm{theo}}(\boldsymbol{\theta}))^2}{\sigma_k^2}, \tag{8}$$

where y_k^{exp} is the measured value, y_k^{theo} is the hypothesised value for a certain model and the parameters $\boldsymbol{\theta}$, and σ_k is the known variance of the measurement. For example, y_k^{exp} and y_k^{theo} represent the measured and calculated B/C ratios.

2) Trial function: Despite the effectiveness of the Metropolis-Hastings algorithm, in order to optimise the efficiency of the MCMC and minimise the number of chains to be run, trial functions should be as close as possible to the true distributions. We use below a sequence of three trial functions to explore the parameter space. The first step is a coarse determination of the parameter PDF using m independent Gaussian distributions centred on $\boldsymbol{\theta}_i$. This allows to calculate the covariance matrix used by the second proposal density, an N-dimensional Gaussian of covariance matrix V. This step provides a better coverage of the parameter space by taking into account possible correlations between the m parameters. The last step takes advantage of a binary space partitioning (BSP) algorithm. The results of the covariance matrix run are used to subdivide the parameter space into boxes for each of which a given probability is affected. The proposal density is defined as a uniform function equal to the assigned probability.

In comparison to the other two proposal densities, this on is not symmetric, because it is only dependent on the proposal state $q(\boldsymbol{\theta}_{\rm trial})$.

III. Implementation in the propagation model

The MCMC with the three above methods are implemented in the USINE package[2], that computes the propagation of galactic CR nuclei and anti-nuclei for several propagation models (LBM, 1D and 2D diffusion models). The reader is referred to [20] for a detailed description for the nuclear parameters (fragmentation and absorption cross sections), energy losses (ionisation and Coulomb) and solar modulation (force-field) used.

In this algorithm, it is up to the user to decide i) which data to use, ii) which observable is retained to calculate the likelihood, and iii) the number of free parameters m (of the vector $\boldsymbol{\theta}$, depending on the propagation model chosen) for which we seek the posterior PDF.

The framework we use is the Leaky Box Model (LBM), a simple propagation model widely used in the past decades. This model contains most of the CR phenomenology and is well adapted for a first implementation of the MCMC tool.

A. Leaky Box Model (LBM)

The LBM assumes that all cosmic-ray species are confined in the Galaxy, with an escape rate equal to $N/\tau_{\rm esc}$, where the escape time $\tau_{\rm esc}$ is rigidity-dependent, and is written as $\tau_{\rm esc}(R)$. This escape time has two origins. First, cosmic rays can leak out the confinement volume and leave the Galaxy. Second, they can be destructed by spallation on interstellar matter nuclei. This latter effect is parameterised by the grammage x (usually expressed in g/cm^2), defined as the column density of interstellar matter encountered by a path followed by a cosmic ray. The cosmic rays that reach the Earth have followed different paths, so that they can be described by a grammage distribution $N(x) \equiv dN/dx$. The LBM assumes that

$$N(x) \propto \exp^{-\lambda_{\rm esc}(R)x} \,, \tag{9}$$

where the mean grammage $\lambda_{\rm esc}(R) = \langle x \rangle$ is related to the mass m, velocity v and escape time $\tau_{\rm esc}(R)$ through $\lambda_{\rm esc}(R) = \bar{m} n v \tau_{\rm esc}(R)$.

The function $\lambda_{\rm esc}(R)$ determines the amount of spallations undergone by a primary species, and thus determines the secondary-to-primary ratios, for instance B/C. The grammage $\lambda_{\rm esc}(R)$ is known to provide an effective description of diffusion models [2]: it can be related to the efficiency of confinement (which is determined by the diffusion coefficient and to the size and geometry of the diffusion volume), spallative destruction (which tend to shorten the average lifetime of a cosmic ray and thus to lower $\lambda_{\rm esc}$), and a mixture of other processes (such as convection, energy gain and losses).

In this paper, we compute the fluxes in the framework of the LBM with minimal reacceleration by the interstellar turbulence, as described in [23], [28]. The grammage $\lambda_{\rm esc}(R)$ is parameterised as

$$\lambda_{\rm esc}(R) = \begin{cases} \lambda_0 \beta R_0^{-(\delta-\delta_0)} R^{-\delta_0} & \text{when } R < R_0, \\ \lambda_0 \beta R^{-\delta} & \text{otherwise;} \end{cases} \tag{10}$$

where we allow for a break, i.e. a different slope below and above a critical rigidity R_0. The standard form used in the literature is recovered setting $\delta_0 = 0$.

To summarise, our LBM with reacceleration may involve up to five free parameters, i.e. the normalisation λ_0, the slopes δ_0 and δ below or above the cut-off rigidity R_0, and a pseudo-Alfvén velocity $\mathcal{V}_a$ related to the reacceleration strength.

B. Source spectra

We assume that the primary source spectrum $Q_j(E)$ for each nuclear species j considered are given by ($\beta = v/c$)

$$Q_j(E) \equiv dQ_j/dE = q_j \beta^{\eta_j} R^{-\alpha_j} , \tag{11}$$

where q_j is the source abundance, α_j is the slope of species j and the term β^{η_j} encodes our ignorance about the low energy spectral shape. We assume that $\alpha_j \equiv \alpha$ for all j, and unless stated otherwise, $\eta_j \equiv \eta = -1$ in order to recover $dQ/dp \propto p^{-\alpha}$, as obtained from acceleration models (e.g., [13]).

In this work, the source abundances are initialised to the solar system abundances ([17]) times the first ionisation potential (FIP) taken from [3]. During the run, the elemental abundances are then rescaled—keeping fixed the relative isotopic abundances—to match experimental data (see Fig. 1 in [21] for further details), so the result is not sensitive to the input values.

The measurement of all propagated isotopic fluxes should completely characterise all source spectra parameters, i.e. the q_j and α_j parameters should be free. However, only elemental fluxes are available, which motivates the above *rescaling* approach. In Sec. IV-B, a few runs are undertaken to determine self-consistently, along with the propagation parameters, i) α and η, and ii) the source abundances for the primary species C, O and the mixed N elements (main contributors to the boron flux).

IV. Results

In particular, the sequential use of the three sampling methods (Gaussian step, then covariance matrix step, then binary space partitioning) is found to be the most efficient: the results presented hereafter all rely on this sequence.

A. Fitting the B/C ratio

1) HEAO-3 data alone: In this section the model parameters are constrained by HEAO-3 data only ([11]). These data are the most precise data available at the present day for the stable nuclei ratio B/C in the energy range of 0.62 to 35 GeV/n.

The values obtained for our Model I $= \{\lambda_0,\, R_0,\, \delta\} = \{54,\, 4.2,\, 0.7\}$ are in fair agreement with those derived by [32], who found $\{\lambda_0,\, R_0,\, \delta\} = \{38.27,\, 3.6,\, 0.7\}$. The difference for λ_0 could be related to the fact that [32] rely on a mere

[2]A public version will be released soon (Maurin, in preparation).

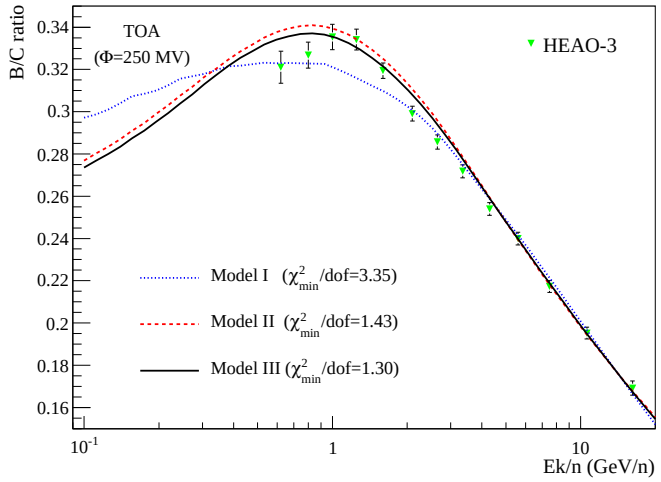

Fig. 1. Best fit curves for Model I (blue dotted line), II (red dashed line), and Model III (black solid line) using only the HEAO-3 data (green symbols). The curves are modulated with $\Phi = 250$ GV.

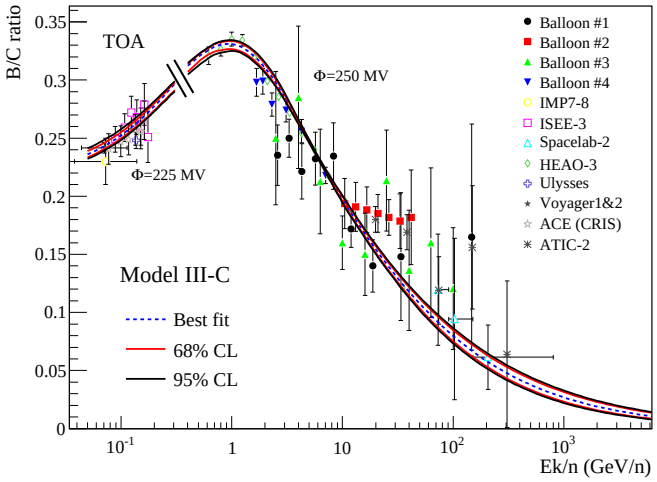

Fig. 2. Confidence regions of the B/C ratio for Model III-C. The blue dashed line is the best fit, red solid line is 68% CL and black solid line 95% CL. Two modulation parameters are used: $\Phi = 225$ MV below 0.4 GeV/n (adapted for ACE+Voyager1&2+IMP7-8 data) and $\Phi = 250$ MV above (adapted for HEAO-3 data).

eye inspection to extract the best fit or/and use a different set of data.

The reacceleration mechanism was invoked in the literature in order to decrease the spectral index δ toward the preferred value 1/3 given by a Kolmogorov spectrum of turbulence. The estimated propagation parameter values for the models II ($\{\lambda_0, \delta, \mathcal{V}_a\} = \{26, 0.52, 88\}$) and III ($\{\lambda_0, R_0, \delta, \mathcal{V}_a\} = \{30, 2.8, 0.58, 75\}$) are indeed slightly smaller than for Model I, but the Kolmogorov spectral index is excluded for all this three cases. This result agrees with the findings of [20], in which a more realistic two dimensional diffusion model with reacceleration and convection was used. Note that the values for $\mathcal{V}_a \sim 80$ km s^{-1} kpc^{-1}, would lead to a true speed $V_a = \mathcal{V}_a \times \sqrt{hL} \sim 80$ km s^{-1} in a diffusion model for which the thin disk half-height is $h = 0.1$ kpc and the halo size is $L = 10$ kpc: this is consistent with values found in [21].

The best χ^2 value per degree of freedom, $\chi^2_{\rm min}$/dof, for each model allows to compare the relative merit of the models. LB models with reacceleration better fit HEAO-3 data: χ^2/dof of 1.43 and 1.30 respectively for the Models II and III compared to χ^2/dof = 4.35 for Model I. The best fit B/C fluxes are shown along the B/C HEAO-3 data modulated at $\Phi = 250$ MV in Fig. 1. Physically, the origin of a cutoff R_0 in $\lambda_{\rm esc}$ at low energy can be related to convection in diffusion models [12]. Hence, it is a distinct process as reacceleration. The fact that Model III performs better than Model II hints at the fact that both processes are significant, as found in [20].

2) Confidence levels for the B/C ratio: Taking advantage of the knowledge of the χ^2 distribution, we can extract a list of configurations, i.e. a list of parameter sets, based on CLs of the χ^2 PDF allowing to derive e.g. fluxes[3].

The B/C best fit curve (dashed blue), the 68% (red solid) and 95% (black solid) CL envelopes are shown in Fig. 2. This demonstrates that, for the specific case of the LBM, current data already strongly constrain the B/C flux (as contained in the very good value $\chi^2_{\rm min} = 1.06$), even at high energy. This leads to good prospects for the discriminating power of forthcoming data. However, such a conclusion has to be confirmed by an analysis in a more refined model (e.g. diffusion model), where the situation might not be as simple.

B. Adding free parameters related to the source spectra

In all previous studies (e.g. [14]), the source parameters are investigated once the propagation parameters have been determined from the B/C ratio (or other secondary to primary ratio). We propose a more general approach, where we simultaneously fit all the parameters. With the current data, it already provides strong constraints on the CR source slope α and source abundances (CNO). Awaiting for better data to take full advantage of the method, we show how this approach can also help uncover inconsistencies in measured fluxes.

For all models below, taking advantage of the results obtained in Sec. IV-A, we retain Model III-C. The roman number refers to the free transport parameters of the model (III $= \{\lambda_0, R_0, \delta, \mathcal{V}_a\}$) and the capital refers to the choice of the B/C dataset (C=HEAO-3+Voyager1&2+IMP7-8). This is supplemented by source spectra parameters and additional data.

1) Source shape α and η from Eq. (11): We first add, as a free parameter, the universal source slope α (Model III-C+1). We then let η, parameterising a universal low energy shape of all spectra, as a second free parameter (Model III-C+2). In addition to B/C which constrains the transport parameters, some primary species must be added to constrain α and η. We restrict ourselves to O, the most abundant boron progenitor, as it has been measured by the HEAO-3 experiment [11], but

[3]For instance, it can be used to predict the $\overline{p}$ or $\overline{d}$ background flux to look for a dark matter annihilating contribution, as done, e.g. in [9].

Model-Data	λ_0 g cm^{-2}	R_0 GV	δ	$\mathcal{V}_a$ km s^{-1}kpc^{-1}	α	η	$10^{20} \times (q_C\|q_N\|q_O)^\dagger$ (m^3 s GeV/n)$^{-1}$
III-C‡	27^{+2}_{-2}	$2.6^{+0.4}_{-0.7}$	$0.53^{+0.02}_{-0.03}$	86^{+9}_{-5}	-	-	-
III-C+1	37^{+2}_{-2}	$4.4^{+0.1}_{-0.2}$	$0.61^{+0.01}_{-0.01}$	64^{+4}_{-4}	$2.124^{+0.005}_{-0.007}$	-	-
III-C+2	29^{+2}_{-2}	$2.7^{+0.3}_{-0.4}$	$0.55^{+0.01}_{-0.02}$	84^{+4}_{-7}	$2.16^{+0.01}_{-0.01}$	$0.3^{+0.1}_{-0.2}$	-
III-C+4	40^{+3}_{-1}	$4.6^{+0.2}_{-0.1}$	$0.64^{+0.01}_{-0.02}$	58^{+2}_{-5}	$2.13^{+0.01}_{-0.01}$	-	$1.93^{+0.04}_{-0.004}\|0.089^{+0.007}_{-0.005}\|2.42^{+0.04}_{-0.05}$
III-C+5	38^{+1}_{-2}	$4.4^{+0.1}_{-0.3}$	$0.60^{+0.02}_{-0.01}$	81^{+4}_{-1}	$2.17^{+0.02}_{-0.02}$	$-0.4^{+1.2}_{-0.1}$	$2.2^{+0.2}_{-0.1}\|0.107^{+0.01}_{-0.006}\|2.7^{+0.3}_{-0.1}$

‡ III-C: propagation parameters are $\{\lambda_0,\ R_0,\ \delta,\ \mathcal{V}_a\}$ and the B/C dataset is HEAO-3+Voyager1&2+IMP7-8.
† Abundances are $1.65|0.10|2.04$ for HEAO-3 ([11]).

TABLE I

MOST PROBABLE VALUES OF THE PROPAGATION PARAMETERS (AFTER MARGINALISING OVER THE OTHER PARAMETERS) FOR MODELS III-C+_,. THE ADDITIONAL FREE PARAMETERS CORRESPOND TO: 1 = $\{\alpha\}$, 2 = $\{\alpha,\ \eta\}$, 4 = $\{\alpha,\ q_C,\ qN,\ qO\}$, 5 = $\{\alpha,\ \eta,\ q_C,\ qN,\ qO\}$ WITH O DATA FROM HEAO-3. THE UNCERTAINTY ON THE PARAMETERS CORRESPOND TO 68% CL OF THE MARGINALISED PDF.

also very recently by the TRACER experiment [1]. During our study it became apparent that these two datasets are incompatible at low energies and that the HEAO-3 dataset seems to be more reliable. The most probable parameters are gathered in Tab. I, where, to provide a comparison, the first line reports the values found for Model III-C (i.e. with $\gamma \equiv \alpha + \delta$ is fixed to 2.65). We remark that adding HEAO-3 oxygen data in the fit, the propagation parameters λ_0, R_0 and δ overshoot Model III-C's results. The parameter $\mathcal{V}_a$ undershoots for this model, as it is anti-correlated to the former parameters. As a consequence, the fit to B/C is worsened, especially at low energy (not shown).

To remedy this, we let η as a free parameter (model III+2). The net effect is to absorb whatever uncertainty that comes from either the modulation level or the source spectrum low energy shape. The value $\eta \simeq 0.3$ might give a reasonable guess of the low energy shape of the source spectrum, but might be as well a consequence of systematics of the experiment.

Whereas it is premature to draw any firm conclusion on the low energy shape, we can turn the argument around as to serve as a diagnosis of the low energy data quality. For instance, assuming that the shape of all light to heavy elements is the same, extracting and comparing η_i for each of these i elements may diagnose some systematics remaining in the data. It would be worth fitting the H and He species which are the best measured fluxes to date, and this is left for a future study carried in the diffusion models.

2) α, η and source normalisation q_i: The last two models add, as free parameters, the CNO elemental source abundances (relative isotopic abundances are fixed to SS ones). The data used in the fit are B/C, C, N and O, all measured by HEAO-3 (TRACER data for C and N have not been published yet). The models, which are denoted for short 4 and 5 in the text below, are:

- III-C+4: $\{\lambda_0,\ R_0,\ \delta,\ \mathcal{V}_a\} + \{\alpha, q_C,\ q_N,\ q_O\}$;
- III-C+5: $\{\lambda_0,\ R_0,\ \delta,\ \mathcal{V}_a\} + \{\alpha,\ \eta, q_C,\ q_N,\ q_O\}$.

The most probable values are gathered in Tab. I. Compared with the respective Models 1 and 2, leaving the source abundances q_C, q_N and q_O free in 4 and 5 does not significantly change the conclusions. Again, adding η (2 and 5) as a free parameter allows to absorb the low energy uncertainties on the data, so that we obtain $\alpha = 2.17$ (5) instead of the biased value 2.13 (4). The same conclusions hold for other propagation parameters. On the derived source abundances, the impact of adding the parameter η is to increase them. The relative C:N:O abundances (O $\equiv$ 1) are respectively $0.78 : 0.36 : 1$ (4) and $0.82 : 0.40 : 1$ (5), the second model providing values slightly closer to those derived from HEAO-3 data $0.81 : 0.49 : 1$.

All source elemental abundances, when rescaling them to match the data or including them in the MCMC, are roughly in agreement, however our approach underlines the importance of taking properly into account the correlations between the parameters to extract unbiased estimates of the propagation and source parameters.

The goodness of fit on the models on the B/C, C, N and O data are no better, but no worse than those with $q_{C,N,O}$ fixed. As soon as primary fluxes are included in the fit (compared to Model III-C), the $\chi^2_{\rm min}$ is worsened. This is due to a combination of an imperfect fit on the primary fluxes and, as already said, a worsened B/C fit because the propagation parameters are optimised to match the former rather than the latter. The best fit CNO fluxes are plotted in Fig. 3 as an illustration.

3) Perspective on source spectrum parameters: Other primary species may have been combined in the χ^2 calculation to i) constrain further α, and/or ii) to check the hypothesis $\alpha_i \neq \alpha_j$ for different species, and/or iii) diagnose some problems in the data if we believe the slope should be universal. However, using a few primary species (O or CNO) already affects the goodness of fit of B/C. As there are many more measured primary fluxes than secondary ones, taking too many primaries would weight too much in the χ^2 compared to B/C, and this would drive the MCMC in regions of the parameter space such as to fit these fluxes rather than B/C. As systematics are known to be larger in flux measurements than in ratio, this may lead to biased estimates of the propagation parameters. Letting η

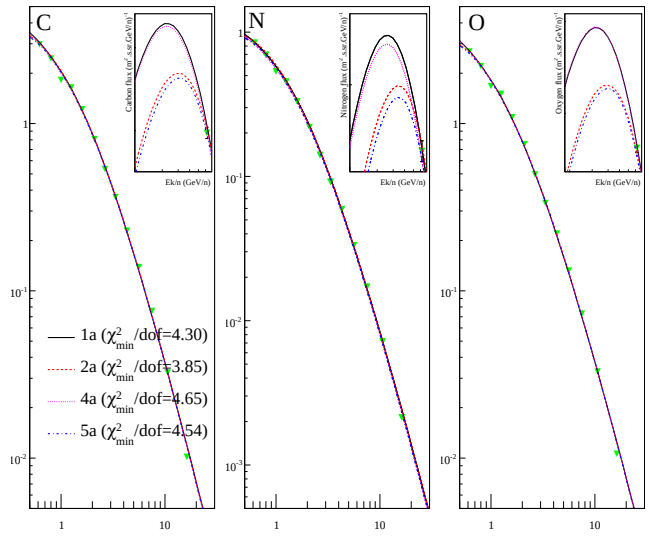

Fig. 3. Carbon, nitrogen and oxygen fluxes from best fit models. A very good fit to HEAO-3 data is achieved for all models and all three elements: the "large" $\chi^2_{\rm min} \sim 4$ found for all models is partly related to the fact that for some energies, the error bars on HEAO-3 data are likely to be underestimated. Models with η set to -1 (1 and 4) or with η converging to $\sim 0.3 - 0.4$ (2 and 5) differ only at very low energy.

as a free parameter allows to some extent to settle this issue for the low energy part of the spectrum.

To illustrate the difficulty on data systematics, suffices to say that for the most abundant species H and He, all experiments were giving mutually incompatible measurements, until AMS and BESS experiments flew ten years ago. We then cannot expect HEAO-3 data, taken in 1979, to be completely free of such drawbacks. Again, we await the publication of several forthcoming new data before pursuing further along this line.

V. Conclusion

We implemented a Markov Chain Monte Carlo to extract the posterior distribution functions of the propagation parameters in a Leaky Box Model. Three trial functions were used, namely a standard Gaussian step, an N-dimensional Gaussian step and its covariance matrix, and a binary space partitioning. For each method, a large number of chains can be run in parallel to speed up the PDF calculations. The three trial functions were used sequentially, each method providing some inputs for the next: while the first one is very good at zoning the gross range of the propagation parameters, it is not as efficient to provide a fine description of the PDF. This is somehow the reverse for the two others, so that the sequential use ending with the binary space partitioning provides the best description of the PDF.

In this first paper, we focused on the phenomenologically well understood Leaky Box Model, to ease and simplify the discussion and implementation of the MCMC. In agreement with previous studies, we confirm that a model with a rigidity cutoff performs better than without and that reacceleration is preferred over no reacceleration. Such a model can be associated to a diffusion model with wind and reacceleration. As found in [20], the best fit models demand both a rigidity cutoff (wind) and reacceleration, but do not allow to reconcile the diffusion slope with a Kolmogorov spectrum for turbulence. In a last stage, we let free the abundance and slope of the source spectra as well as the elemental abundances of C, N and O. This shows some correlation between the propagation and source parameters, potentially biasing the estimates of these parameters. The best fit for the slope of the source abundances is $\alpha \approx 2.17$ using HEAO-3 data. The MCMC approach allows to draw confidence intervals for the propagation parameters, the source parameters, and also for all fluxes.

A wealth of new data on galactic cosmic ray fluxes are expected soon. As illustrated on the LBM, the MCMC is a robust tool to handle the complexity of data and model parameters, where one should fit at the same time all source and propagation parameters to avoid bias. The next step is to apply it to more realistic diffusion models and on larger datasets including more nuclear species.

References

[1] M. Ave *et al. Astrophysical Journal*, 678:262–273, 2008.
[2] V. S. Berezinskii, S. V. Bulanov, V. A. Dogiel, and V. S. Ptuskin. *Astrophysics of cosmic rays*. Amsterdam: North-Holland, 1990, edited by Ginzburg, V.L., 1990.
[3] W. R. Binns *et al. Astrophysical Journal*, 346:997–1009, 1989.
[4] P. Blasi. *ArXiv e-prints*, 0801.4534, 2008.
[5] P. J. Boyle *et al. ArXiv Astrophysics e-prints*, 2007.
[6] N. Christensen, R. Meyer, L. Knox, and B. Luey. *Classical and Quantum Gravity*, 18:2677–2688, 2001.
[7] G. Cowan. *Statistical data analysis*. Publisher: Oxford: Clarendon Press, 1998 Series: Oxford science publications. ISBN: 0198501560, 1997.
[8] T. Delahaye *et al. ArXiv e-prints*, 712.2312, 2007.
[9] F. Donato *et al. Physical Review D*, 69(6):063501, 2004.
[10] J. Dunkley *et al. Monthly Notices of the Royal Astronomical Society*, 356:925–936, 2005.
[11] J. J. Engelmann *et al. Astronomy & Astrophysics*, 233:96–111, 1990.
[12] F. C. Jones. *Astrophysical Journal*, 229:747–752, 1979.
[13] F. C. Jones. *Astrophysical Journal Supplement*, 90:561–565, 1994.
[14] F. C. Jones *et al. Astrophysical Journal*, 547:264–271, 2001.
[15] A. Lewis and S. Bridle. *Physical Review D*, 66(10):103511, 2002.
[16] A. M. Lionetto *et al. Journal of Cosmology and Astro-Particle Physics*, 9:10, 2005.
[17] K. Lodders. *Astrophysical Journal*, 591:1220–1247, 2003.
[18] David MacKay. *Information Theory, Inference, and Learning Algorithms*. Publisher: Cambridge University Press. ISBN: 0521642981, 2003.
[19] P. S. Marrocchesi *et al.* 36:3129, 2006.
[20] D. Maurin *et al. Astrophysical Journal*, 555:585–596, 2001.
[21] D. Maurin *et al. Astronomy and Astrophys.*, 394:1039–1056, 2002.
[22] Redford M. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
[23] J. L. Osborne and V. S. Ptuskin. *Soviet Astronomy Letters*, 14:132, 1988.
[24] A. D. Panov *et al. ArXiv Astrophysics e-prints*, 0612377, 2006.
[25] A. D. Panov *et al. ICRC2007, ArXiv e-prints*, 0707.4415, 2007.
[26] E. S. Seo *et al. Advances in Space Research*, 33:1777–1785, 2004.
[27] E. S. Seo *et al. COSPAR, Plenary Meeting*, 36:1846, 2006.
[28] E. S. Seo and V. S. Ptuskin. *Astrophysical Journal*, 431:705–714, 1994.
[29] A. W. Strong and I. V. Moskalenko. *Astrophysical Journal*, 509:212–228, 1998.
[30] A. W. Strong *et al. Annual Review of Nuclear and Particle Science*, 57:285–327, 2007.
[31] S. P. Wakely *et al. COSPAR, Plenary Meeting*, 36:3231, 2006.
[32] W. R. Webber *et al. Astrophysical Journal*, 508:940–948, 1998.
[33] W. R. Webber, M. A. Lee, and M. Gupta. *Astrophysical Journal*, 390:96–104, 1992.