Statistical properties of cosmic ray fluxes and anisotropy predictions

Guilhem Bernard[1], Timur Delahaye[1,2] Pierre Salati[1], Richard Taillet[1]

[1] *LAPTh. Université de Savoie. CNRS. BP 110, Annecy-le-Vieux F-74941. France.*
[2] *Institut d'Astrophysique de Paris, UMR 7095 - CNRS, Université Pierre & Marie Curie, 98 bis boulevard Arago, 75014, Paris, France*

bernard@lapth.cnrs.fr

Abstract: Spectral anomalies detected by the PAMELA and CREAM experiments could be due to the presence of nearby and young cosmic ray sources. This can be studied in the myriad model, in which cosmic rays diffuse from point-like instantaneous sources located randomly throughout the Galaxy. The computation of erros bars associated to the flux (and anisotropy) turns out not to be as straightforward as it seems in the myriad model, as the standard deviation is infinite when computed for the most general statistical ensemble. The goals of this poster is to describe a method to associate error bars to the flux measurements which has a clear statistical meaning. We show that the quantiles (68% confidence levels, for instance) of the flux distribution are well-defined, even though the standard deviation is infinite. We also use the fact that local sources have known positions and ages to reduce the statistical ensemble from which random sources are drawn in the myriad model. In this context, we also discuss the status of the spectral features observed in the proton flux by CREAM and PAMELA.

Keywords: icrc2013, cosmic rays, variance, Monte Carlo, statistics

1 Introduction

Cosmic-ray nuclei observed in the Solar System were accelerated in discrete sources distributed in the Galactic Disk. Many theoretical predictions of the fluxes and anisotropy rely on Monte Carlo simulations in which the positions and ages of the sources are randomly drawn in some statistical ensemble. The mean values and the variances or the outcome of many realizations is then used to infer the expected quantities and their uncertainties, respectively. This procedure may be flawed when the variance is infinite, which happens even in very common situations. In this poster, we first describe the problem more precisely, and then we present and discuss several solutions.

2 Statement of the problem

The cosmic ray density at location $\mathbf{x}_o$ due to a unique source of age t_s located at $\mathbf{x}_s$ is given by the propagator $\mathscr{G}_1(\mathbf{x}_s, t_s \to \mathbf{x}_o)$. For instance, in a propagation model with no Galactic wind, no spallation, and a diffusion volume not limited by boundaries, and with O chosen as the origin, it is simply given by

$$\mathscr{G}_1(r_s) = \frac{1}{(4\pi D t_s)^{3/2}} e^{-r_s^2/4Dt} \tag{1}$$

but it may be computed in more complex models. If the location and age of the source are drawn randomly from a given distribution $d^4n(\mathbf{r}_s)/d^3\mathbf{r}_s dt$, the flux $\mathscr{F}_1$ is a random variable. Its mean value is given by

$$\langle \phi_1 \rangle \propto \int dt \iiint \mathscr{G}_1(r_s) \frac{d^4n(\mathbf{r}_s)}{d^3\mathbf{r}_s\, dt}\, d^3\mathbf{r}_s \tag{2}$$

and its variance

$$\sigma_1^2 = \langle \phi_1^2 \rangle - \langle \phi_1 \rangle^2 \tag{3}$$

where

$$\langle \phi_1^2 \rangle \propto \int dt \iiint \mathscr{G}_1^2(r_s) \frac{d^4n(\mathbf{r}_s)}{d^3\mathbf{r}_s dt}\, d^3\mathbf{r}_s \tag{4}$$

When N independant sources are present, the corresponding moments of the total flux are simply given by

$$\langle \phi_N \rangle = N \langle \phi_1 \rangle \quad \text{and} \quad \sigma_N^2 = N\sigma_1^2 \tag{5}$$

As expected, the mean value $\langle \phi_N \rangle$ is equal to the steady-state model with a continuous distribution of sources. However, for a large class of propagators, including the form given in Eq. 1, the integral giving the variance σ_1^2 (and thus σ_N^2) is divergent: the variance is infinite. Physically, the divergence is due to rare but strong events, at the lower end of the integration over r_s and t_s variables.

The standard deviation σ is commonly interpreted as the typical spread of the random values around the mean, and a high standard deviation could be interpreted as if the actual value of the flux had a disturbingly high probability to be very far from the mean value. We will see later that this may not always be the case. Actually, this confusing situation, in which some rare events have a very small contribution to the mean, but give rise to a very high standard deviation, is not uncommon in physics (Levy flights, Cauchy distribution).

In this proceeding, the presentation of the methods and results focuses on the cosmic ray flux, but it can be generalized to any statistical quantity, such as the anisotropy amplitude and direction.

3 Regularization by cut-off

One could argue that the problem we considered is physically irrelevant, because we know for sure that there is no supernova remnant with zero age and null distance to the Earth. One can impose a lower cut-off in ages and distances, based on observations. However, even with reasonable values for the cut-off, the variance indeed is finite but still can have large values. From observations of the Solar neighbourhood, we have a good idea of the distribution of sources that are young and close. Given the age $t_{\min}$ of the youngest local supernova remnant, one can compute the mean value and standard deviation of the total flux ϕ, by applying that cut-off $t_{\min}$ to the age distribution.

For the sake of illustration, Fig. 3 features the mean value $\langle\phi\rangle$ and the standard deviation σ_ϕ of the total proton flux using a lower cut-off of $t_{\min} = 100$ yr in the source distribution, for a realistic set of propagation parameters. In the same figure we have also plotted the data points from the CREAM and PAMELA experiments. With the chosen value for the cut-off, the standard deviation at high energies remains of the same order of magnitude as the flux. The relative value of the standard deviation, $\sigma_\phi/\langle\phi\rangle$, is fairly independent of the energy. This trend can be shown analytically to hold for sources located in a thin disk. In the framework of a purely diffusive model (no wind/spallation), the relative dispersion can be approximated by

$$\frac{\sigma_\phi}{\langle\phi\rangle} \sim \frac{R}{4L\sqrt{2\nu t_{\min}}} \quad . \tag{6}$$

This ratio does not depend on the diffusion coefficient D, hence it does not depend on energy. It is of the order of unity for $t_{\min} \sim 100$ yr.

For sources distributed in a disk with a finite thickness h, one obtains the same result as before as long as $h \ll \sqrt{Dt_{\min}}$. In the opposite limit, one finds

$$\frac{\sigma_\phi}{\langle\phi\rangle} \propto D^{1/4} \quad . \tag{7}$$

In both cases, the relative standard deviation does not vary much with energy.

One can also adjust the cut-off on a more theoretical way, by choosing to eliminate events that make the standard deviation very high, without contributing significantly to the the mean value. This is the approach adopted by [1]. These authors used a cut-off given by $t_{\min} = R_{\max}/\sqrt{4\nu D(E)}$, where ν is the rate of Galactic source explosions. It is chosen as "the time over which one source goes of within a distance from Earth such that CR from that source reach us within a time $t_{\min}$". Said differently, the expected number of Galactic sources of age $t_{\min}$ contributing significantly to the flux in the Solar neighbourhood is just equal to unity. Note that this condition is now explicitly energy-dependent: some sources will be discarded at high energy but not at low energy. With this condition, the standard deviation gives a fair order of magnitude of the spread of the values around the mean. It is difficult though to interpret it in rigourous statistical terms. Indeed, the value of the variance depends quite strongly on the exact value chosen for the cut-off. Choosing $t_{\min}/2$ or $2\times t_{\min}$ rather than $t_{\min}$, for instance, has a small effect on the mean but a drastic effect on the standard deviation. The value chosen by [1] is indicated by an arrow in the figure. The mean obtained with this cut-off is about 10 % lower than the true mean. A lower cut-off would give a more precise mean value, but a much larger variance.

4 Catalog

The situation described above occurs as long as we are ignorant of the positions and ages of the CR sources. The young and nearby objects are responsible for the divergence of the flux variance and potentially lead to the problems encountered above in the statistical analysis. However, we do have data concerning the distribution of nearby sources, for which catalogues are available. A natural way to regularise the variance is then to separate the sources into two sets. The first set contains the young and local sources, which can be extracted from the catalogues. The second group, about which we have little information, includes the old or distant sources and will be treated in the same way as in the statistical analysis of Sec. 3. This procedure allows us to regularise the variance of the flux in the most natural way while reducing its uncertainties. Following [2], we have used two catalogues.

(i) The Green survey [3] compiles various informations on supernova remnants, but fails to systematically provide their ages or the precision with which their distances from the Earth have been determined. A quite thorough bibliographic work has been summarised in the appendix of [2], which we have borrowed as a complement to the Green catalogue. In total, we have collected 27 local SNR with their ages, distances from the Sun and, when possible, the corresponding observational uncertainties.

(ii) Pulsars are not expected to be sources of primary CR nuclei. As residues of supernova explosions, they are nevertheless a good tracer of old SNR that are too old to be directly detected in radio waves. Moreover, because they are point-like objects, their distances from the Sun is much easier to measure. Their ages can also be estimated precisely through spin-down. After removing millisecond pulsars from the ATNF catalogue [4] (by selecting $\dot{P} > 5\times 10^{-18}$) and objects associated with known SNR, we are left with 157 objects with their ages and distances.

These 27+157 sources are found to be representative of the local environment: the number of objects found in the catalogues agrees well with what can be inferred from various Galactic distributions found in the literature, provided that the supernova explosion rate is, on average, approximately equal to 3 per century. At least, this is true within the 2 kpc nearby the Sun, for sources younger than 30,000 years. At later ages, SNR become too dim to be detected and the Green survey cannot be trusted anymore. However, the catalogues agree fairly well with theoretical expectations and do not suffer from major biases, at least not more than the theoretical models. In this work, we define the "local" region as the domain extending 2 kpc around the Sun, with sources younger than 50,000 years. A rate of one supernova explosion per century in our Galaxy is also quoted in the literature [2]. If this is the case, we would be in a locally high-density region of sources. We have plotted the flux considering this low rate, but will discuss the validity of this assumption in a forthcoming letter. The CR proton flux produced by the SNR and pulsars of our catalogue is presented in Fig. 1 for the "min" CR propagation benchmark model. The sources mainly contributing to the flux at high energy are given in [2]. In the PAMELA energy region, the flux is dominated by pulsars, whereas SNR come into play at the energies of CREAM.

5 Meaning of infinite variance: Toy model

The variance of the flux being infinite does not necessarily imply that the random values are typically very different from the mean. To illustrate this affirmation, consider the case of a unique point-like steady-state source (no time dependence) located in the Galactic disk, with cosmic ray diffusion taking place in a boundless space. The solution

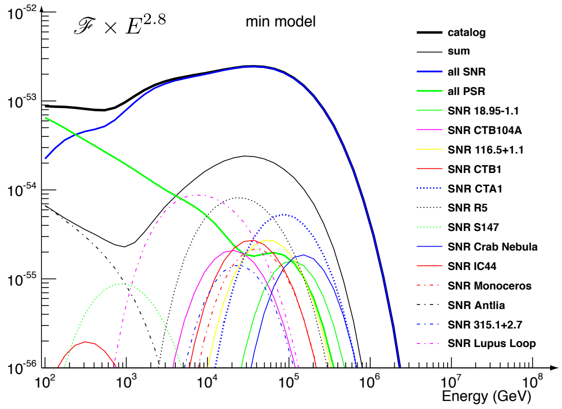

Figure 1: CR proton flux plotted as a function of energy for the SNR and pulsars which dominate over the other objects of our catalogue.

of the diffusion equation is given by $\varphi = a/r$ where a is a constant. Assuming that this source is uniformly distributed inside a disk of radius R leads to the probability distribution function

$$dp = \frac{2\pi r\,dr}{\pi R^2} = \frac{2r\,dr}{R^2} \ . \tag{8}$$

We can readily infer the mean flux

$$\langle\varphi\rangle = \int_0^R \frac{a}{r}\,\frac{2r\,dr}{R^2} = \frac{2a}{R} \ , \tag{9}$$

and the average value of the flux squared

$$\langle\varphi^2\rangle = \int_\varepsilon^R \frac{a^2}{r^2}\,\frac{2r\,dr}{R^2} = \frac{a^2}{R^2}\ln\left(\frac{R}{\varepsilon}\right) \ , \tag{10}$$

where we have introduced a cut-off value ε at the lower end of the radial distribution to exhibit the divergence of $\langle\varphi^2\rangle$. The variance of φ goes to infinity as $\varepsilon \to 0$.

However, the distribution of φ (which is what we are really interested in) is well-behaved. From the relation between r and φ, we can write $dr = a\,d\varphi/\varphi^2$ so that

$$dp(\varphi) = \frac{2r\,dr}{R^2} = \frac{2a^2\,d\varphi}{R^2\,\varphi^3} \ . \tag{11}$$

The probability that the flux is lower than a given value Φ may be expressed as

$$P(<\Phi) = \int_{\varphi(r)}^{\varphi(R)} dp(\varphi) = 1 - \frac{a^2}{R^2\Phi^2} \ , \tag{12}$$

provided that $\Phi > \Phi_0 \equiv a/R$. Introducing the Heavyside distribution Θ leads to

$$P(>\Phi) = \frac{a^2}{R^2\Phi^2}\,\Theta\left(\Phi - \frac{a}{R}\right) = \frac{\langle\varphi\rangle^2}{4\,\Phi^2}\,\Theta\left(\Phi - \frac{\langle\varphi\rangle}{2}\right) \ . \tag{13}$$

The probability that $\Phi > 10\,\langle\varphi\rangle$ is only 1/400, even though the variance is infinite. Indeed, the flux is more likely to be lower than the mean value, whereas one might have guessed the opposite, considering the divergence of the variance.

When N sources are considered, the mean flux value and the variance are both just multiplied by N. The probability distribution $p_N(\phi)$ for the flux can be obtained by recurrence from

$$p_N(\phi) = \int p(\varphi)\,p_{N-1}(\phi - \varphi)\,d\varphi \ . \tag{14}$$

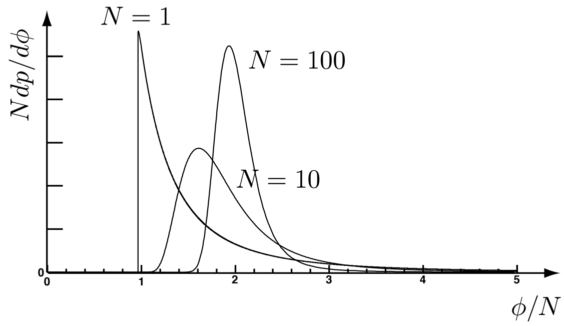

Figure 2: Probability distribution of ϕ/N for $N = 1$, 10 and 100 sources.

	2D	3D
steady-state	$\langle\varphi\rangle$ finite	$\langle\varphi\rangle$ finite
	$\langle\varphi^2\rangle \to \infty$	$\langle\varphi^2\rangle$ finite
time-dependent	$\langle\varphi\rangle$ finite	$\langle\varphi\rangle$ finite
	$\langle\varphi^2\rangle \to \infty$	$\langle\varphi^2\rangle \to \infty$

Table 1: Divergence of the variance

These are displayed in Fig. 2. The variance still diverges. In the high-ϕ region, the flux is dominated by the contribution of a single source and the probability distribution is given by

$$\frac{dp_N}{d\phi} = N\,\frac{dp}{d\varphi} = \frac{2Na^2}{R^2\phi^3} \ . \tag{15}$$

For $N = 100$ sources, the probability that $\phi > 2\,\langle\phi\rangle$ is 2.5×10^{-3} and $P(\phi > 10\,\langle\phi\rangle)$ is vanishingly small.

If we now consider time-dependent sources spread homogeneously inside an infinite diffusive halo (DH) with pure diffusion, the variance is given by the integral

$$\sigma_\varphi^2 \propto \int dt \int 4\pi r^2\,dr\,\frac{1}{(4\pi Dt)^{3/2}}\,e^{-r^2/2Dt} \ , \tag{16}$$

which diverges with the lower cut-off in ages as $1/\sqrt{t_{\rm min}}$. For a 3D homogeneous distribution of steady-state sources, σ_φ does not diverge (see Table 1).

6 Quantiles are finite

For any given probability density $p(\phi)$ that behaves as $p(\phi) \sim \phi^{-\alpha}$ as $\phi \to \infty$, the mean value

$$\langle\phi\rangle \equiv \int \phi\,p(\phi)\,d\phi \tag{17}$$

is finite and the variance, related to

$$\langle\phi^2\rangle \equiv \int \phi^2\,p(\phi)\,d\phi \tag{18}$$

is infinite, whenever $2 < \alpha < 3$. This is not an exotic condition, as for the CR flux due to diffusion from sources

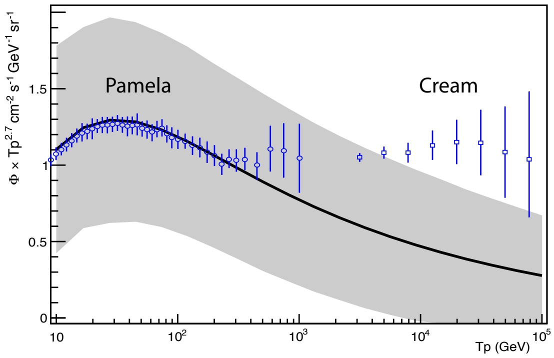

Figure 3: Mean value (solid line) and standard deviation (grey band) of the distribution of flux, for plausible propagation parameters.

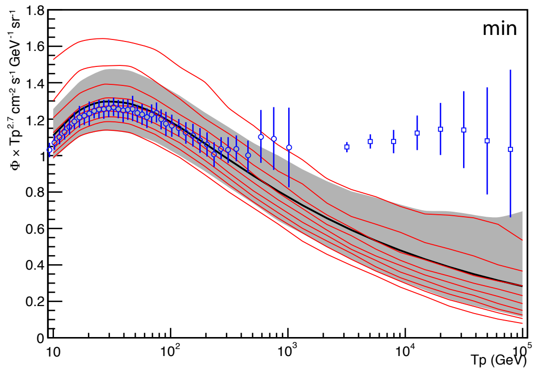

Figure 4: Mean flux (red curves) and envelopes (grey band) representing the standard deviation of the flux for the min, med, and max propagation model for $\nu = 3$ century^{-1}.

having a uniform distribution in a thin disk (2D distribution) or a heavy disk (3D distribution), we have $p(\phi) \sim \phi^{-7/3}$ and $p(\phi) \sim \phi^{-8/3}$ respectively, as shown in [5].

The divergence of the variance is due to the high-ϕ tail of the $p(\phi)$ distribution, *i.e.* to the very local sources (small r and t). It will not be removed by taking into account spallation, Galactic wind, energy losses or escape through the boundaries of the diffusion volume, as these processes have a very small effect on the propagator at low r and t. We will only present the simplest case of pure diffusion, which captures the essence of the problem.

What we are really interested in when we perform Monte Carlo simulations, is to know which values are likely to be observed. One feels that one way or another, the very rare but strong events than make the variance diverge could be safely discarded. As discussed above, a cut-off in the distribution is not the safest way to do it. It turns out that when $\alpha > 1$, the quantiles are well-defined. These are defined as ϕ-interval in which the random variable has a given probability q to be found ($q = 0.1$ for deciles, for instance). Formally, they are obtained by adjusting the limits ϕ_n and ϕ_{n+1} in the following integral to have

$$\int_{\phi_n}^{\phi_{n+1}} p(\phi)\, d\phi = q \qquad (19)$$

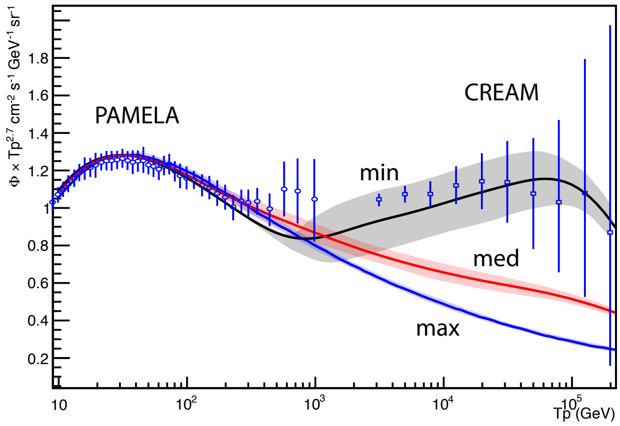

Figure 5: Blue, red, and black curves feature the total flux ϕ computed as the sum of the mean external flux $\langle \phi_{\rm ext} \rangle$ and the contribution $\phi_{\rm cat}$ from the catalogue. They correspond to the max, med and min CR propagation benchmark models, respectively. The bands that extend around the curves have the same meaning as in Fig. 4. They indicate the standard deviation of the flux associated to the observational errors on the ages and distances of the SNR of the catalogue.

They can also be obtained from a very large set of random realization, by sorting the outcome by values and dividing the sorted list into $1/q$ bins.

Figure 4 displays the deciles obtained for the proton flux at high energy, obtained in the case of diffusion in a cylindrical volume of radius $R_{\rm max}$ and half-height L, diffusion coefficient $D(E) \propto D_0 (E/1\ {\rm GeV})^{\delta}$, where the numerical values for the diffusion parameters L, D_0 and δ have been chosen as be reproduce the observed B/C ratio (see [6]).

7 Conclusions

The methods presented here can be applied to the computation of the expected proton spectrum, due to the local sources (catalog) and the rest of the sources (Monte Carlo or continuous source distribution). An example is shown in Fig. 5, for three benchmark sets of parameters (min, med and max). For a complete study, see [7].

Acknowledgment: The ICRC 2013 is funded by FAPERJ, CNPq, FAPESP, CAPES and IUPAP.

References

[1] P. Blasi and E. Amato, Journal of Cosmology and Astroparticle Physics, Issue 01 (2012) id. 010
[2] T. Delahaye, J. Lavalle, R. Lineros, F. Donato, N. Fornengo, Astronomy & Astrophysics 524 (2010) 51
[3] D. A. Green, A Catalogue of Galactic Supernova Remnants, 2009, `http://www.mrao.cam.ac.uk/surveys/snrs/`
[4] R. N. Manchester, G. B. Hobbs, A. Teoh and M. Hobbs, The Astronomical Journal 129 (2005) 1993–2006
[5] G. Bernard, T. Delahaye, P. Salati, R. Taillet, Astronomy & Astrophysics 544 (2012) 92
[6] D. Maurin, F. Donato, R. Taillet, P. Salati, The Astrophysical Journal 555 (2001) 585–596
[7] G. Bernard, T. Delahaye, Y.-Y. Keum, W. Liu, P. Salati, R. Taillet, accepted for publication in Astronomy and Astrophysics, 2013.