Chiquita Flux Calculation

Elina Brobeck

November 9, 2006

Original fit to the data (revised derivation)

The most basic fit one can use on the Chiquita data is a straight line in log space. To do this, we assume that the number flux of cosmic rays follows the power law

$$J_0(E_0) = \frac{dN_0}{dE_0} = CE_0^{-\gamma}$$

where $\gamma_0 = 3$ and $\log C \sim 24.5$. Here E_0 is the actual energy of the incoming particle (as opposed to the measured energy E).

For convenience in fitting the energy range measured by Chiquita, this function will be rewritten to be centered on $10^{17.8}$ eV.

$$\frac{dN_0}{dE_0} = C_{17.8} \left(\frac{E_0}{10^{17.8}} \right)^{-\gamma}.$$

The relationship between C and $C_{17.8}$ is $C = C_{17.8} \times 10^{17.8\gamma}$.

We are measuring the number flux per interval in observed energy, $\frac{dN}{dE}$. This is the convolution of the actual number flux with the function $R(E, E_0)$, which describes the distribution of reconstructed energies at each value of E_0.

$$J(E) = \frac{dN}{dE} = \int dE_0 A(E_0) R(E, E_0) \frac{dN_0}{dE_0}.$$

The function $A(E_0)$ is the acceptance of the array (the fraction of showers which pass all cuts on the data at a given energy).

The histogram of Chiquita data is binned in log space, with bin width $\log(E_{\text{upper}}) - \log(E_{\text{lower}}) = 0.2$. In terms of the central energy E_C of the bin, the bin width is

$$E_{\text{upper}} - E_{\text{lower}} = 10^{0.1} E_C - 10^{-0.1} E_C = 0.46 E_C.$$

Each data point in the histogram therefore contains a number of showers given by

$$N = \int_{10^{-0.1} E_C}^{10^{0.1} E_C} \frac{dN}{dE} dE.$$

To obtain the number of Chiquita showers in a given bin, we substitute the expression for $\frac{dN}{dE}$ above:

$$N(E_i) = \int_{\mathrm{E_i\ bin}} dE \int_{\mathrm{all\ E_0}} dE_0 A(E_0) R(E, E_0) \frac{dN_0}{dE_0}$$

where E_i is the central energy (in log space) of the ith bin.

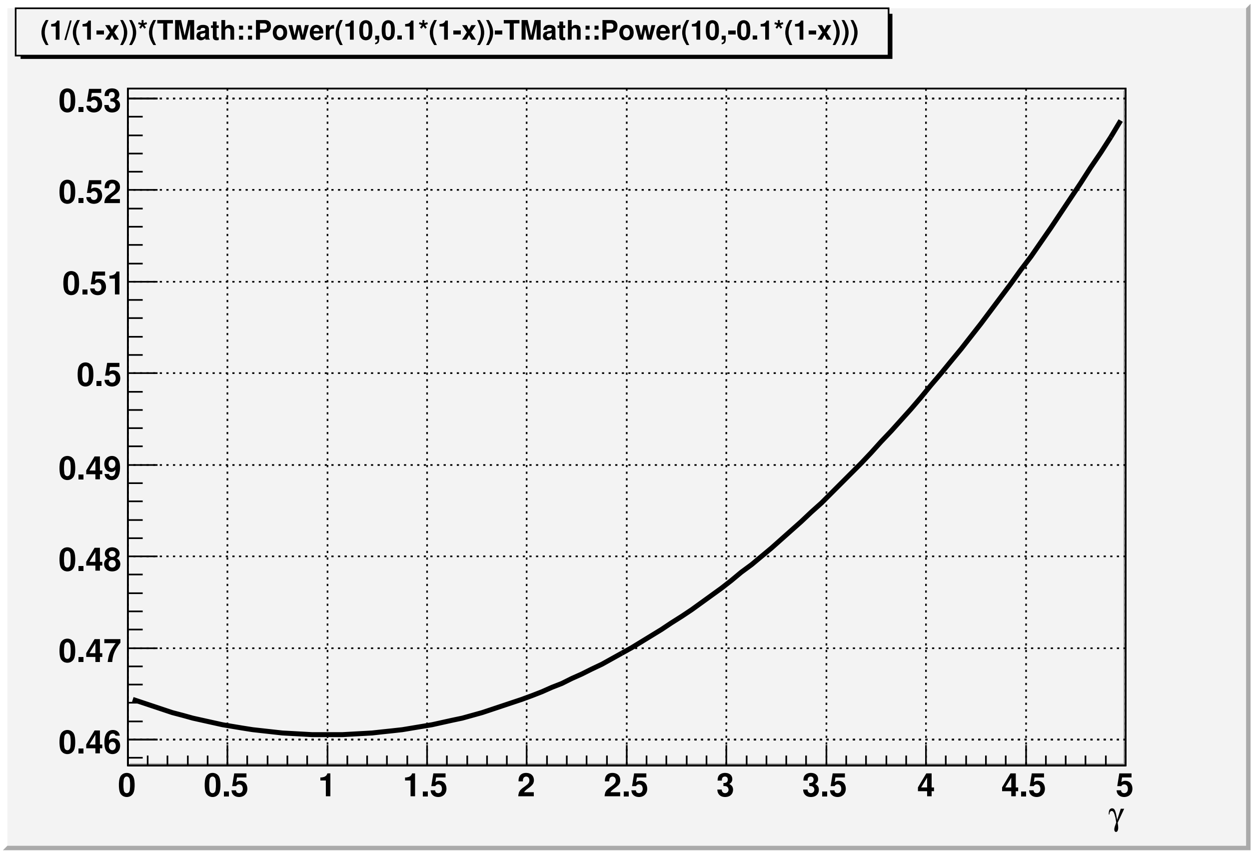
We have the function $R(E, E_0)$ in the form of a data table for a fixed set of values of E_0, so instead of a continuous integral, we need a sum. We make the approximation that the entire flux in the bin centered around E_0 is being transferred to the bin centered around the observed energy E with the efficiency $R(E,E_0)$.[1]

$$N(E_i) = \sum_{E_{0j}} A(E_{0j})R(E_i,E_{0j}) \int_{\mathrm{E_{0j}\ bin}} dE_0 \frac{dN_0}{dE_0}$$

The integral of the flux over the E_{0j} bin is given by

$$\begin{aligned}
\int_{\mathrm{E_{0j}\ bin}} CE_0^{-\gamma} dE_0 &= \int_{10^{-0.1E_{0j}}}^{10^{0.1E_{0j}}} CE_0^{-\gamma} dE_0 \\
&= \left.\frac{CE_0^{-\gamma+1}}{-\gamma+1}\right|_{10^{-0.1E_{0j}}}^{10^{0.1E_{0j}}} \\
&= \frac{C}{-\gamma+1}\left[\left(10^{0.1}\right)^{-\gamma+1} - \left(10^{-0.1}\right)^{-\gamma+1}\right] E_{0j}^{-\gamma+1}
\end{aligned}$$

This function is flat at $\gamma = 1$. At values of γ close to 3, it can be reasonably approximated by the value at $\gamma = 3$.



---

[1] The data table of values for $R(E,E_0)$ was calculated using simulated showers with an E^{-3} distribution over the width of the bin. It is not the efficiency for moving showers from the central energy of one bin to the central energy of another bin, but rather an approximation to the fraction of showers in one bin which get moved to another bin due to error in the reconstructed energies. This is why substituting $R(E_i,E_{0j})$ for $R(E,E_0)$ takes care of the integral over the E_i bin. Note, however, that the farther γ is from 3, the more error is introduced through R. (This also applies to $A(E)$, which is tabulated from the same set of simulated showers.)

For $\gamma = 3$,

$$\int_{\mathrm{E_0\ bin}} C_0 E_0^{-\gamma} dE_0 = 0.477 C_0 E_0^{-\gamma+1}.$$

The function which should be used to fit the data[2] is thus

$$\begin{aligned} N(E_i) &\simeq \sum_{E_{0j}} 0.477 A(E_{0j}) R(E_i, E_{0j}) C E_{0j}^{-\gamma+1} \\ &= \sum_{E_{0j}} 0.477 A(E_{0j}) R(E_i, E_{0j}) C'_{17.8} \left(\frac{E_{0j}}{10^{17.8}}\right)^{-\gamma'} \end{aligned}$$

where $\gamma' = \gamma - 1$.

We want to do the fit to C and γ in log space, so this is rewritten as

$$N(E_i) = \sum_{E_{0j}} 0.477 A(E_{0j}) R(E_i, E_{0j}) 10^{\log C'_{17.8} - \gamma' (\log E_{0j} - 17.8)}.$$

The factor $\log C_{17.8}$ has been replaced by $\log C'_{17.8}$, where

$$\log C = \log C_{17.8} + 17.8\gamma = \log C'_{17.8} + 17.8\gamma'.$$

What we want to plot is

$$F \times E^3 = \frac{dN}{dE} E^3 = C E^{-\gamma} E^3.$$

This is calculated as

$$F \times E^3 = \frac{dN}{dE} E^3 = C E^{-\gamma} E^3 = C E^{-(\gamma-1)} E^{-1} E^3 = C'_{17.8} \left(\frac{E}{10^{17.8}}\right)^{-\gamma'} E^2.$$

---

[2] In the original derivation, I had $\left[10^{0.1} - 10^{-0.1}\right]$ where I should have had $\frac{1}{-\gamma+1}\left[\left(10^{0.1}\right)^{-\gamma+1} - \left(10^{-0.1}\right)^{-\gamma+1}\right]$. This gave a constant factor of 0.465, which is still close to the value of the second expression when $\gamma = 3$. Ideally the entire expression should be included when fitting for γ. Alternatively, one could iterate on the value.

Iterative fit method

In this method, we start by defining the observed flux $J(E) = \frac{dN}{dE}$ as before:

$$J(E) = \frac{dN}{dE} = \int dE_0 A(E_0) R(E, E_0) \frac{dN_0}{dE_0} = \int dE_0 A(E_0) R(E, E_0) J_0(E_0).$$

As shown in the previous section, a reasonable approximation to the number of showers in the ith energy bin is

$$N(E_i) \simeq \sum_{E_{0j}} 0.477 A(E_{0j}) R(E_i, E_{0j}) C E_{0j}^{-\gamma+1}$$

We would like to get rid of the sum in order to turn this into an expression for C. To do this, we want to define an effective aperture $A'(E)$ such that

$$N(E) = 0.477 E\, A'(E)\, J(E).$$

$A'(E)$ is calculated from simulations by taking the number of simulated showers which reconstruct in bin E_i over the number of simulated showers thrown in bin E_i. (This will not be the correct ratio if the showers were not thrown with the correct $E^{-\gamma}$ spectrum.) The number of showers which reconstruct in bin E_i is given by

$$N_{\text{sim}}(E_i) = \int_{\mathrm{E_i\ bin}} dE \int_{\mathrm{all\ E_0}} dE_0 A(E_0) R(E, E_0) J_{\text{sim}}(E_0).$$

The effective aperture is therefore given by

$$A'(E_i) = \frac{N_{\text{sim}}(E_i)}{\int_{\mathrm{E_i\ bin}} J_{\text{sim}}(E_i) dE}.$$

Similarly, the measured number of showers in each bin is

$$\begin{aligned} N(E_i) &= A'(E_i) \int_{\mathrm{E_i\ bin}} J_0(E) dE \\ &= A'(E_i) \int_{\mathrm{E_i\ bin}} C E^{-\gamma} dE \\ &\simeq 0.477 A'(E_i) C E_i^{-\gamma+1} \\ N(E_i) &= 0.477 E_i A'(E_i) J(E_i) \end{aligned}$$

Note that this is an expression which relates the total number of showers in the bin, $N(E_i)$, to the flux at a single point, $J(E = E_i)$. The aperture, $A'(E_i)$ is an average over the bin, and depends on the shape of the spectrum.

We have assumed that the flux $J(E)$ has the form $J(E) = CE^{-\gamma}$. The function we ultimately want to plot is $J(E)E^3 = CE^{3-\gamma}$. The set of points given by this function should be approximately constant:

$$C_{\text{data}}(E_i) \equiv J(E_i) E_i^3 = C E_i^{3-\gamma}$$

Substituting in the expression for $N(E_i)$ above,

$$C_{\text{data}}(E_i) = \frac{N(E_i) E_i^2}{0.477 A'(E_i)}.$$

We can use this expression[3] to plot the points C_{data}, and then fit those points to the expression $C_{\text{data}}(E) = CE^{3-\gamma}$.

---

[3] In the code, I used an expression equivalent to $C_{\text{data}}(E_i) = \sum_{k=1}^{N} \frac{E_k^3}{0.477 E_i A'(E_k)}$, where the sum is over the set of k showers in the bin. I thought it would be more accurate to multiply by the energy E_k of the individual showers rather than the central energy of the bin; however, after going over the derivation again I'm not sure that it is correct to do that.

Call the best-fit values of these parameters C_2 and γ_2. We can use these values to generate a data set assuming an input spectrum of $J_2(E) = C_2 E^{-\gamma_2}$. (In other words, we work backward to see what the data set would look like if it were a perfect fit to the parameters C_2 and γ_2.) The generated data set is calculated from the function

$$N_{\text{data}}(E_i) = 0.477 E_i\, A'(E_i)\, C_2 E_i^{-\gamma_2}.$$

In the code, this is actually calculated from the earlier expression

$$N_{\text{data}}(E_i) = \sum_{E_{0j}} 0.477\, A(E_{0j}) R(E_i, E_{0j})\, C_2 E_{0j}^{-\gamma_2+1}.$$

As before, a new function $C'_{\text{data}}(E)$ is defined by the points

$$C'_{\text{data}}(E_i) = \frac{N_2(E_i) E_i^2}{0.46 A'(E_i)}.$$

Assuming no errors entered into the conversion from $C_2 E_i^{-\gamma_2}$ to N_{data} and back to C'_{data}, the C' points would lie on the line given by the best-fit values to C_{data}. (Error is introduced through the tabulated values of $A'(E)$, which were are based on simulations which assume an E^{-3} spectrum.)

The flux is now estimated[4] as

$$J(E_i) = \frac{C_{\text{data}}(E_i)}{C'_{\text{data}}(E_i)} C_2 E_i^{-\gamma_2}.$$

Questions:

Does it matter which expression is used to calculate N_{data}? I had originally chosen to use the sum assuming it was more accurate, but maybe using the expression with $A'(E)$ is more self-consistent?

Am I right to say that the reason for taking the ratio in the last step is to reduce the error coming from the fact that the simulations used $\gamma = 3$? Is there any other reason this helps?

When generating the data set N_{data}, would it be better to recalculate the constant using $\gamma = \gamma_2$ rather than using 0.477, which is the value for $\gamma = 3$?

---

[4] In my original notes, I had this expression as $J(E_i) = \frac{C'_{\text{data}}(E_i)}{C_{\text{data}}(E_i)} C_2 E_i^{-\gamma_2}$, but I think this must be an error. C'data should (almost) cancel with $C_2 E_i^{-\gamma_2}$. The expression in the main text is the one I used in the fit code.