Search for Point Sources and Emission from the Galactic Plane in Milagrito

Andrew J. Smith

July 9, 1999

1 Background Subtraction via "Direct Integration"

To date, all searches for point source DC emission in Milagro data have used the method of "time sloshing" to estimate the background. This method is well described in Alexandraes et al. In this method, a few hours of data are buffered in computer memory. Then a background map is generated by selecting the events one by one and combining them with times from a random event in the buffer. The local coordinates of the event (HA-Dec) are combined with the new time to obtain a new RA-Dec. Typically 10 or more random events are constructed for every real event. An alternative method has been developed and used by Gus in his search for 10s bursts. In this method, for each real event time, every event in the buffer is added to the background map in a manner equivalent to time sloshing an infinite number of times. For time intervals (t/24h < binsize/360. deg), the earth's rotation is smaller than the binsize, so one can simply overlay a map in HA-Dec coordinates onto an RA-Dec map where the RA for each event is just the HA+Sidereal time.

I use this method to search for DC sources using the following algorithm. A few hours of data are read. A signal map is generated along with a local coordinate map with HA-Dec coordinates. Additionally, the number of events that are collected in each 23.9344696 sec interval is recorded. This period is the amount of time that it takes the earth to rotate 0.1 degrees. After each time sloshing period, typically hours or more, the local coordinate map is normalized, then a background map is generated by adding the number of events collected in each 24s interval times the local coordinate map to the RA-DEC background map offsetting the local coordinate map appropriately by the sidereal time of each 24s time interval.

This method has the advantage that it removes a statistical uncertainty from the estimation of the background that plagues the traditional time sloshing method, while retaining the insensitivity to systematic errors in the estimation of the background that make time sloshing such an attractive technique. Furthermore, for an increased rate, unlike times sloshing where the time to compute the background map increases linearly with rate, direct integration takes the same amount of time. A higher rate manifests itself as a larger number of events in each 24s time interval.

2 Application of LiMa to Milagro Data

In 1980-something Li and Ma wrote a paper describing how to estimate the significance of an excess of events given a number of measured events on and off source. The complication that makes their method paper worthy is that they properly account for how fluctuations in the measurement of the background affect the significance. While looking back at their paper, I came to the realization that we, or at least I, have been using their formalism for estimating significance wrongly, in two ways actually.

In the LiMa formalism, let S be the number of events measured on source, B is the number of events measured off source, and α is the ratio of exposure on to off. When estimating the background with time sloshing, I have used the factor α in the LiMa equation as the number of signal events divided by the number of background events randomly generated, typically $\frac{1}{10}$. This is wrong. In our background estimation method, signal events can be, and in general are, included in the background estimation. If there is an excess in the signal region, the background will be over estimated. Another problem with this approach is that the value chosen for α is in fact wrong. This can be easily seen in the limit of a an infinite number of time sloshing or equivalently, the direct integration method described above. With this formalism $\alpha \rightarrow 0.$ and the error in the background is negligible. This is clearly never the case as in the time sloshing method, there are a limited number of background events that can be sloshed into the signal bin. In general, α for the case of infinite time sloshing is given by the ratio of the binsize in RA to the time sloshing period. For example, for a 1.5 deg signal bin and a 2 hour time sloshing period, $\alpha = \frac{1.5}{30} = 0.05$. Where 30 is the number of degrees the earth rotates in 2 hours.

These two effects, the inclusion of signal in the background and the mis-estimation of α, work in opposite directions. The former leads to under estimation of the significance by a factor of $\frac{1}{1+\alpha}$ arewhile the latter leads to an over estimation by a factor of $\sqrt{1+\alpha}$. Figure 1 shows a plot of 6197 independent 1.5×1.5 degree bins between declination 10 and 60 degrees for the entire Milagrito data set. The background was estimated by direct integration. In each case the mean of the distributions is consistent with 0 and the distributions fit well to a Gaussian, the widths however vary depending on how the significance is calculated. In plot 1a, the LiMa significance is estimated assuming that α is zero and the contribution of the signal to the background is not removed. The resulting RMS for this distribution, 0.966, while numerically close to the expected value of 1.0, deviates from unity by more than 3σ. In plot 1b, the correct value for α is used, but no correction is made for the contribution of signal events to the background. This results in a RMS $<< 1.0$. In plot 1c, the signal is removed from the background, but alpha is underestimated, resulting in an RMS that is too large. The final plot uses a correct value for α and accounts for the presence of signal events in the background estimation. The final distribution has an RMS consistent with unity, as a gaussian statistical distribution must.

In this analysis, a time sloshing period of 2 hours was used. For point source searches, the choice of a time sloshing period is a compromise between the increased statistical error associated with a short interval, and the increasing probability of introducing a systematic error associated with gradual changes in the local coordinate event distribution for longer intervals. As shown above, the LiMa parameter α is determined by the time sloshing interval. Although 2 hours is the chosen time sloshing period, in general, the interval can be shorter. In this analysis, a time sloshing interval ended when 2 hours passes, the run changes, or the experiment is turned off. Short time sloshing periods do have poorly estimated backgrounds, i but as the number of events is also proportional to the time interval, they have a reduced contribution to the final result. The sum of all time sloshing periods times the number of events collected during the period for this analysis is equal to 6975 sec, only a few percent smaller than the maximum period, 7200 sec. The shorter periods do not significantly affect the results of this analysis.

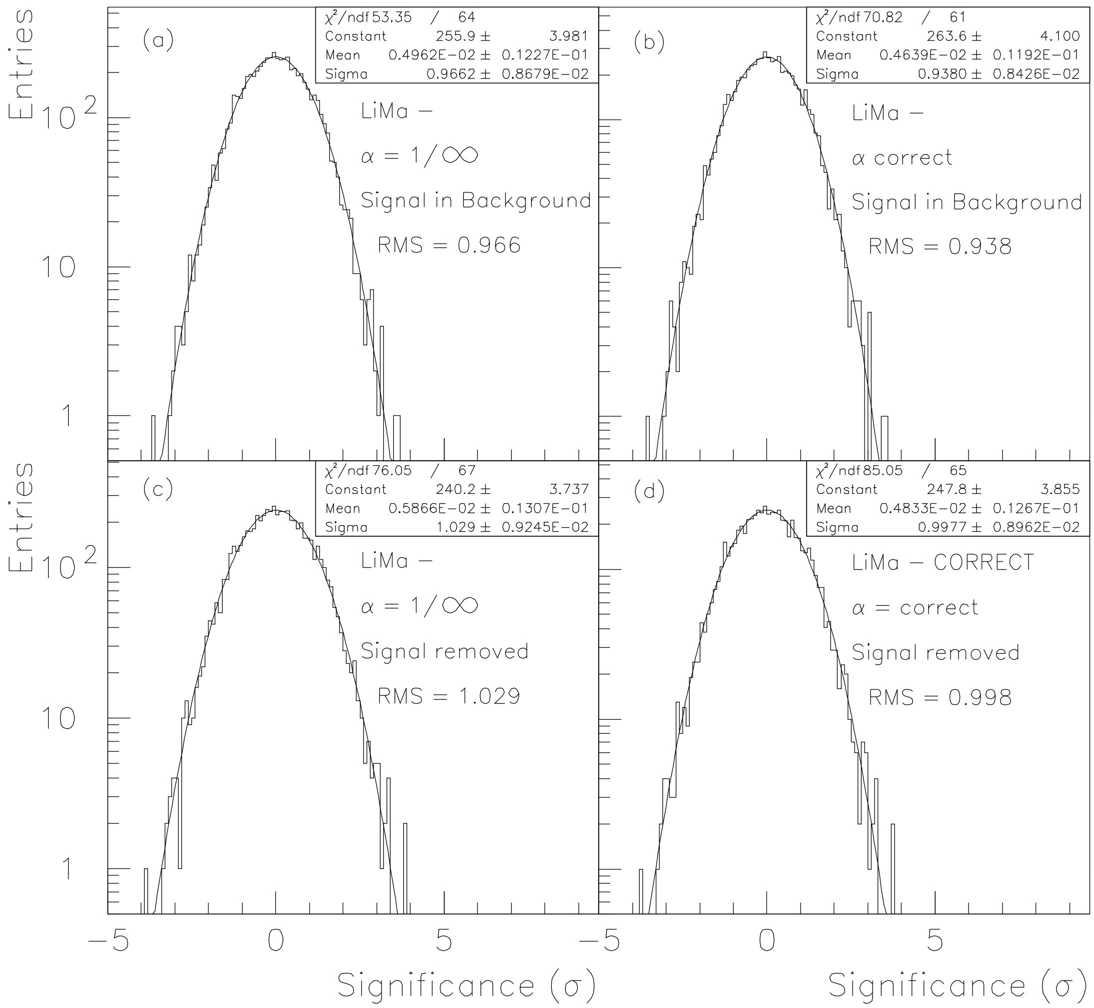

Figure 1: Comparison of results when the LiMa significance algorithm is applied in four different ways.

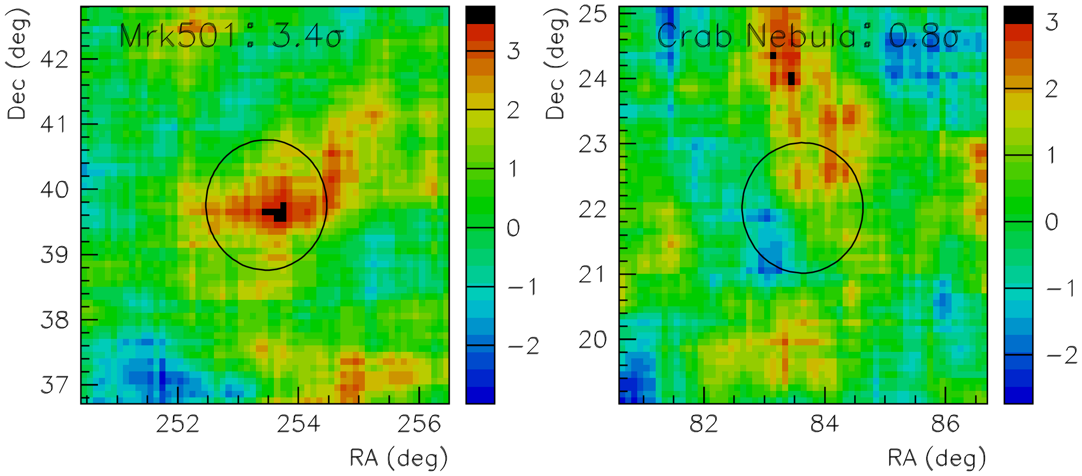

Figure 2: Significance maps in the vicinity of Mrk 501 and the Crab Nebula.

3 Search for DC point source emission

A map of the sky for declination -60 to +90 was created along with a corresponding background map using the direct integration method. An NFIT cut of >= 50 PMTs was applied to the data. This cut was found to be more or less optimal by both Stefan and Joe. The bin size of the maps is 0.1 deg. The sky was searched for gamma-ray point sources by combining the $.1 \times .1$ deg bins to form 1.5×1.5 deg search bins. A 1.5 deg square bin has equal area as a 0.85 deg radius round bin. Again, this binsize has been found to be more or less optimal for point source searches with the corresponding NFIT cut.

To test consistency with previous analyses, I looked at the vicinity of Mrk 501 and the Crab. Figures 2a and 2b are significance maps for the regions in the vicinity of Mrk 501 and the Crab Nebula respectively. I observe a 3.4σ excess at the position of Mrk501 and a 0.8σ excess at the Crab. Both results are consistent with previous analyses.

Figure 3a shows the distribution of significance for 6197 non-overlapping bins with declinations between 10 deg and 60 deg. Does this plot look familiar? Figure 3b is the same data, but heavily oversampled with an entry for every 0.1 deg bin. Plots 3c and 3d are the same as plots as 3a and 3b, but the region of the high significance tail was blown up and the histogram statistics are displayed. The most significant point for the case of non-overlapping bins is 3.8σ which has a pre-trials probability of 7.2×10^{-5} and a post trials probability of .45. For the oversampled distribution, the number of effective trials is difficult to evaluate. It is clearly more than the 6197 independent trials in plot 3a, but clearly less than the 18 million entries. Conservatively, we might estimate the number of trials at $100,000$. For this trials factor, the recognized 5σ new detection threshold for gamma-ray sources is raised to about 6.9σ. There is no evidence of new gamma-ray point sources in this data set.

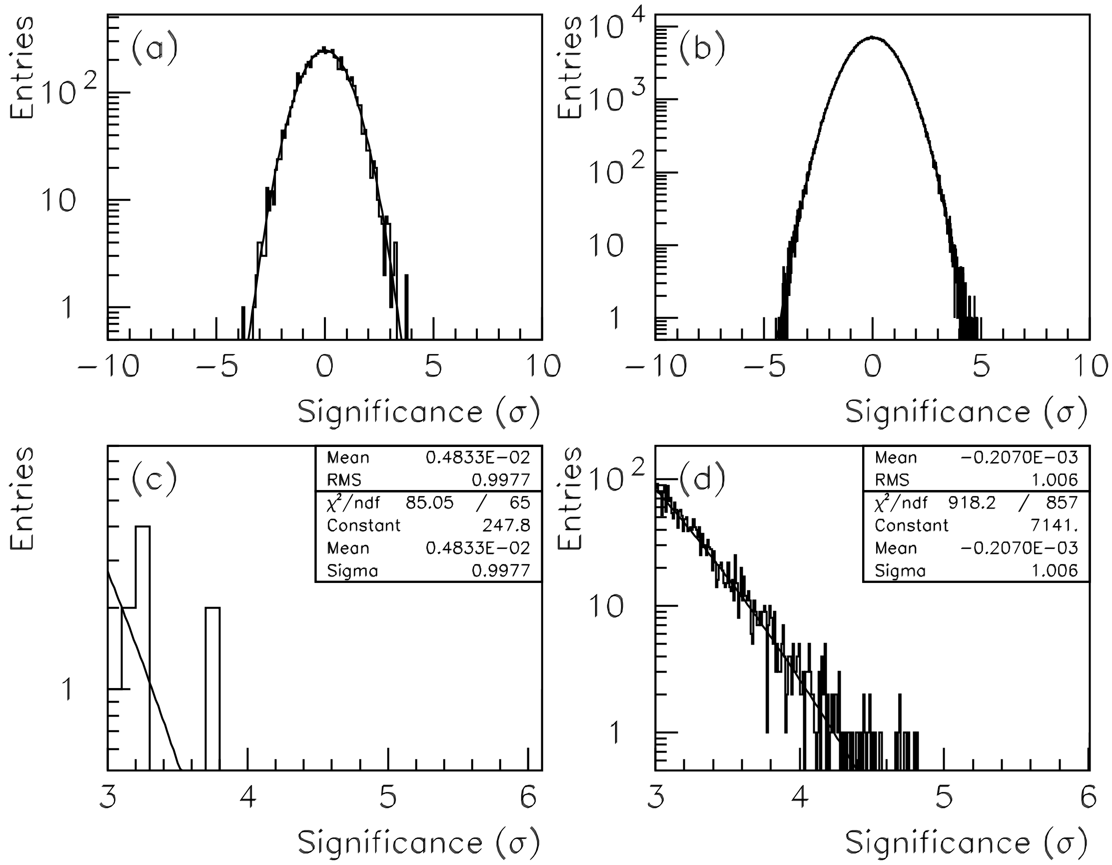

Figure 3: Distribution of significances for the sky over Milagrito.

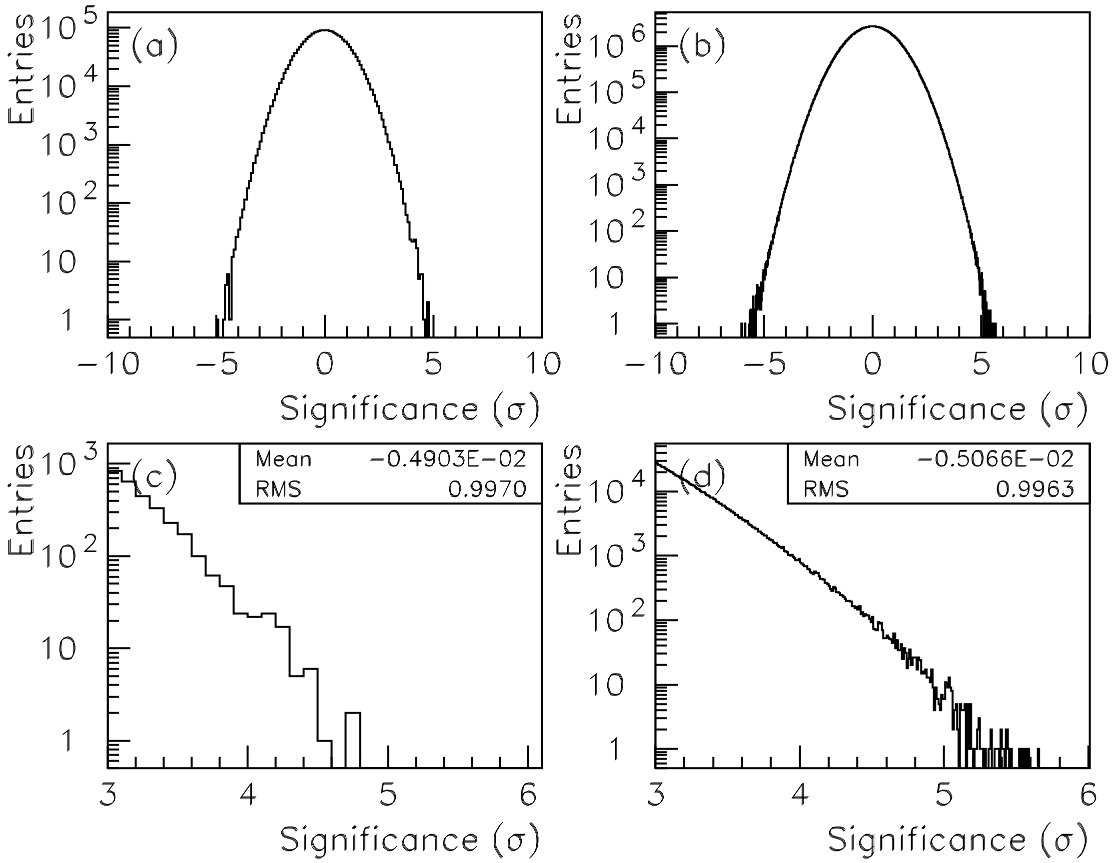

Figure 4: Distribution of significances for the sky over Milagrito for each operational day.

4 Search for Day Scale Point Source Emission

The same analysis performed in the previous section on the entire Milagrito data set was applied to the data collected for each Julian day that the experiment was operational. The significance distribution for all days combined is shown in figures 4a and 4b. In these plots, only bins with estimated backgrounds greater than 100 events are considered. The number of trials, of course, increases by a factor of about 350, the number of days that Milagrito was operational. This increases the threshold for a post trials 5σ detection to 7.6σ. No point in excess of 7.6 sigma is observed, so no detection can be claimed.

5 Search for Diffuse Emission from the Galactic Plane

Again using the direct integration method to estimate background, a second map of the entire sky was created with the NFIT cut reduced to >= 10 and and the time sloshing interval extended to 5

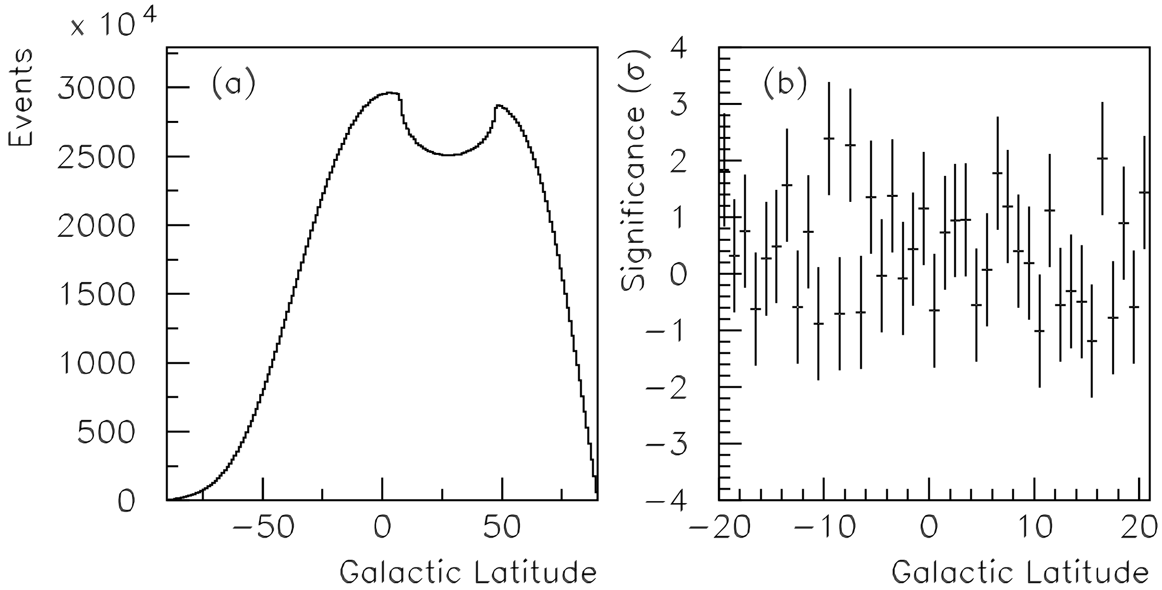

Figure 5: Plot of event distribution and significance of the excess as a function of galactic latitude.

hours. This map is optimized to study diffuse emission over regions large compared to Milagrito's angular resolution. Using this map the data were divided into bins of galactic latitude. Plot 5a shows how our data are distributed as a function of galactic latitude. The "bite" taken out of the top of the otherwise smooth distribution is the excluded region with equatorial declination greater than 70 deg. Near the poles background estimation becomes difficult. Plot 5b shows the significance of excess as a function of galactic latitude binned in 1 deg bins. There is no evidence in our data for detection of diffuse emission from the galactic plane. Table 5 lists the number of signal and background events detected within 2 and 5 degrees of the galactic plane and the derived 99.9% CL upper limit as a ratio to the cosmic ray proton flux. For the calculation of the upper limit, the effective area for protons is assumed to be equal to that for gammas, and 77% of reconstructed trigger are assumed to come from cosmic-ray protons, with the remainder due to He and heavy nuclei.

Cut	Signal	Background	Significance	99.9% CL Upper Limit
$\|b\| < 2$ deg	87094832	87093908	0.02σ	4.23×10^{-4}
$\|b\| < 5$ deg	217309904	217300854	0.53σ	2.68×10^{-4}

As the diffuse galactic gamma-ray flux is known to vary as a function of galactic longitude, it makes sense to break up the data into longitudinal bins as well. Plot 6a and 6b show the significance of the excess as a function of galactic longitude for events with latitude less than 2 deg and 5 deg respectively. There is no evidence of a significant excess attributable to diffuse emission in either plot. Figure 7 is a plot of our upper limit for emission from the galactic plane plotted as a function of galactic longitude. The same assumptions are made as before about the effective areas and the fraction of triggers induced by cosmic-ray protons. Also shown on the plot is a much less sensitive measurement made by Whipple. We kick their butt.

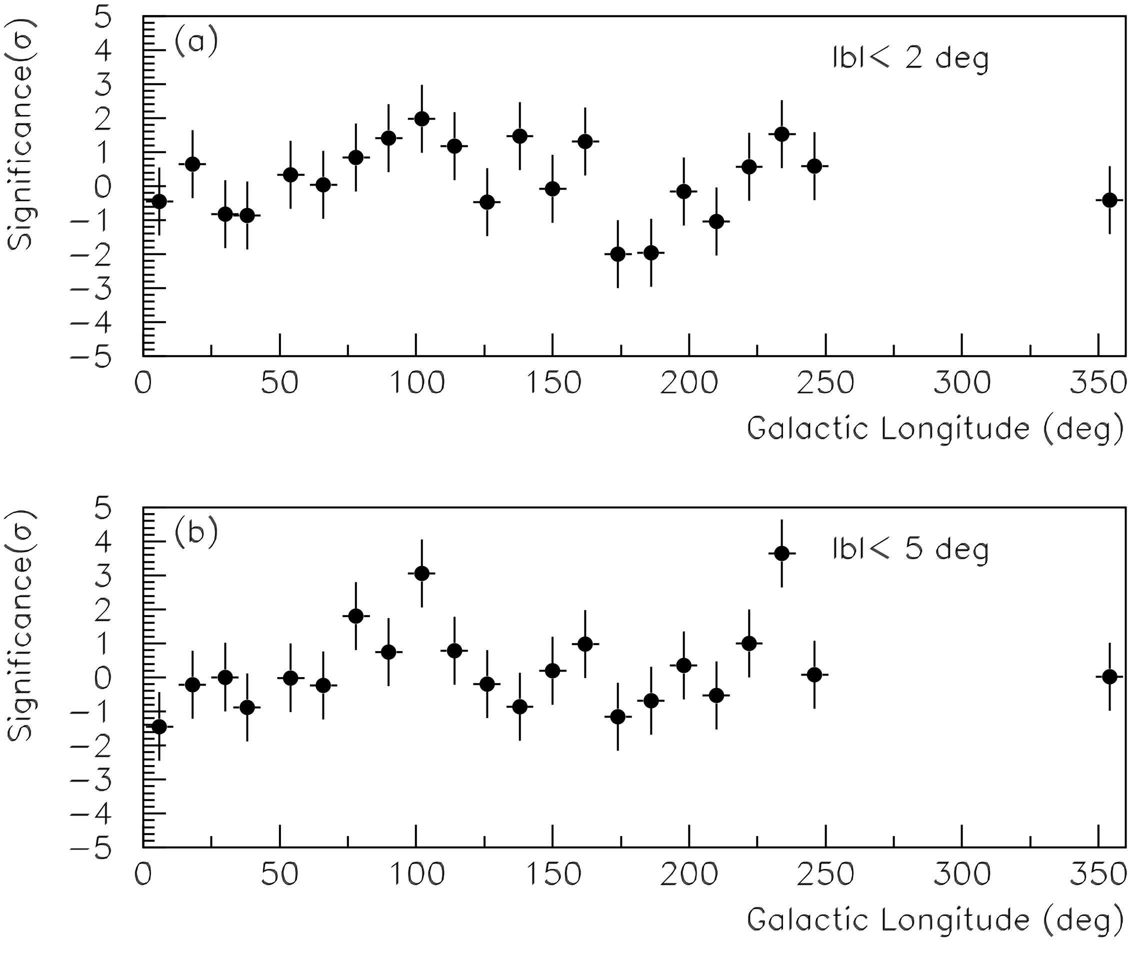

Figure 6: Significance of excess as a function of galactic longitude.

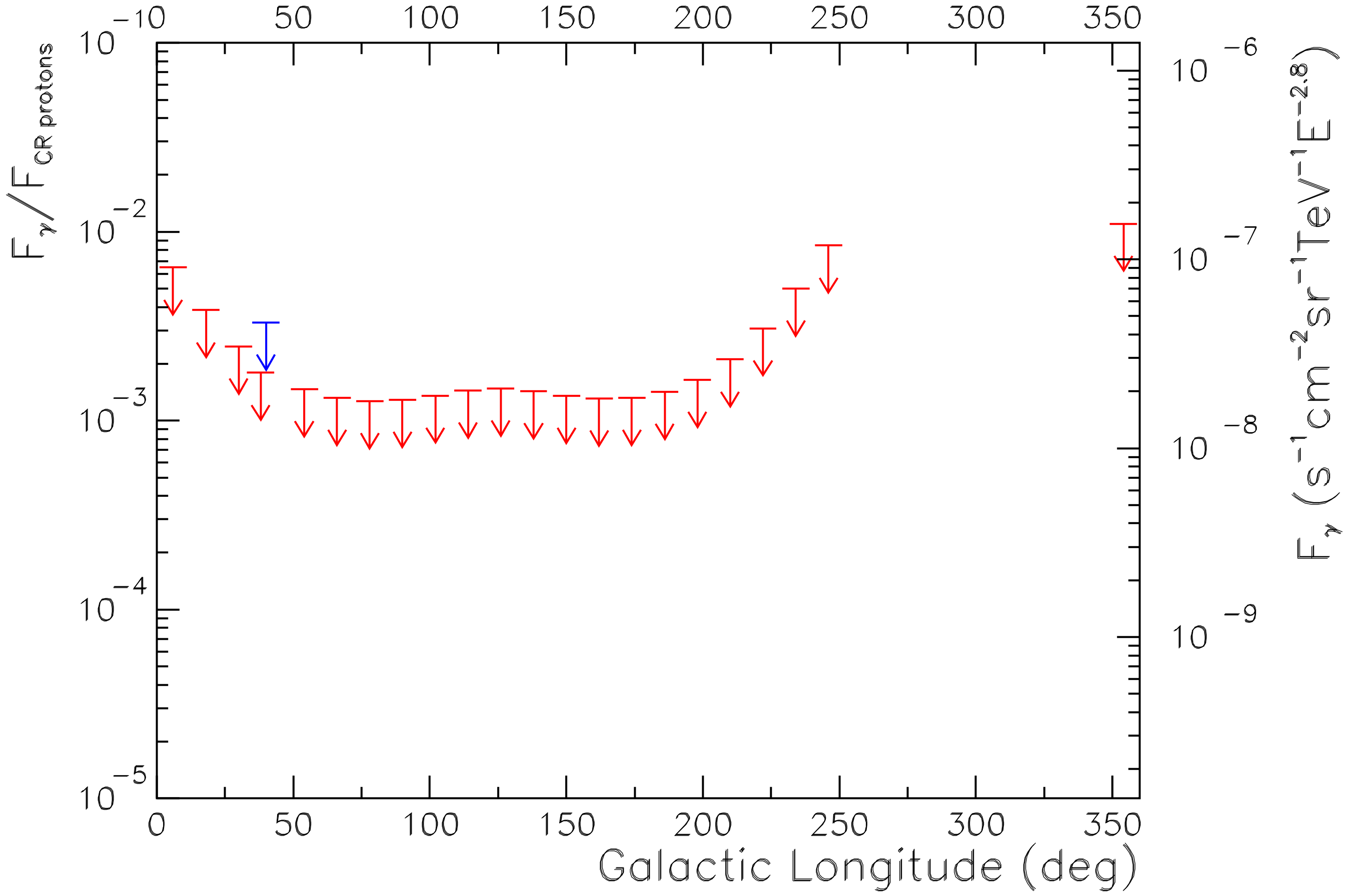

Figure 7: Upper limits (99.9% confidence limit) on diffuse gamma-ray flux from the galactic plane ($|b < 2\,\text{deg}\,|$) plotted as a function of galactic longitude plotted as a function the cosmic ray proton flux.