

Noise Analysis of the Draco Galaxy Using C.A.C.T.U.S: Using Light to See the Dark

by Sasha Laundy, Swarthmore College
with professor Mani Tripathi
and Ph.D. candidate Peter Marleau
UC Davis Physics REU
June-August 2005

Introduction

Using C.A.C.T.U.S., a ground-based gamma ray telescope, we may have detected an excess of energetic photons in the nearby dwarf galaxy Draco. Because Draco is comprised mostly of dark matter, this could be an exciting result provided that the excess is not an artifact of our measuring device. One possible artifact we need to rule out is that the slight excess of noise in the same region directly leads to the excess of signal. This summer I wrote software to analyze this noise data and determine the likelihood that signal data are the result of noise. The results of my code are interesting but not completely understood at this point. They may lead the way to further analysis methods.

Background

The Earth is constantly bathed in a sea of cosmic radiation from all directions. It consists of electromagnetic radiation and (mostly charged) particles. Unfortunately, charged particles are deflected on their way to Earth by the magnetic fields of the galaxy, Solar System, and Earth. Particles with mass are deflected by the gravitational fields of the same bodies. This deflection means we have no information about the initial trajectory of charged or massive particles, and therefore no idea where they originated.

This provides a strong incentive for scientists to develop techniques that study neutral, massless particles. Electromagnetic radiation is a stellar (pun intended) candidate.

Astronomers have detected particles of energies of up to 10^{20} eV, but they don't yet have any idea how the most energetic were created. The shock waves of supernovae are thought to produce particles of energies up to 10^{12} eV, but there are no satisfactory models to explain higher energies even after years of searching. It would be convenient if we could just *look* in the direction of the highest of high-energy particles and study their production, but this is partially impossible because of the deflection discussed above. Gamma ray telescopes may be able to help. High-energy gamma rays are a particularly interesting band because the universe has been thoroughly mapped at every other energy, but there are still objects we can't see. As of 1998 there were no telescopes operating between 20 and 250 GeV. Now there are several, including C.A.C.T.U.S. which operates at 40 GeV and higher. As exploring other new bands have revealed previously unseen objects, we all hope to see some of this "dark matter" as we study the cosmos with these new eyes.

C.A.C.T.U.S. is a ground-based atmospheric Cherenkov detector. The obvious problem with using a telescope on the ground to look at things in the sky is that there is a troublesome atmosphere in the way. This type of telescope actually uses the atmosphere as its detection medium: when the high-energy cosmic radiation hits the atmosphere, it produces a spray of particles. These sprays are known as air showers, and have several useful properties that allow us to identify the type of incident particle and reconstruct its original trajectory.

Air Shower Formation

A high-energy particle enters the atmosphere and encounters an air molecule. The two exchange a photon to pair produce an electron and antielectron. Since they are traveling faster than the speed of light in air they produce Cherenkov radiation, as they travel through the atmosphere. They eventually produce new photons through bremsstrahlung as they encounter more air molecules, which pair produce new electrons and positrons and so on. The process repeats until the photon energy becomes less than that needed to pair produce. This is the shower's maximum because after this point particles are lost due to interactions with air molecules but none are gained. Luckily, the atmosphere is relatively transparent to the wavelength of the Cherenkov light so the shower at sea level consists primarily of these photons.

The showers have several characteristics that make them very useful detection tools. They have a very distinct pancake shape, on the order of 150m in diameter and 3m (10ns) thick, which helps discriminate them from background noise. Because of the relativistic incident energies, the initial particle's trajectory is preserved in the axis of the shower. The showers produced by gamma rays look very different from those of hadrons— because hadron showers produce many other types of particles they tend to have a more splotchy shape where gamma ray showers produce a neat ellipse. This again helps us discriminate between the signal we seek (gamma rays) and the background noise (charged particles).

Solar Two



Fig 1. An aerial view of the Solar Two site.

Solar Two was a proof-of-concept solar power plant. It ran successfully from 1982 to 1999, and has been succeeded by Solar Tres in southern Spain. Sunlight was collected by the field of mirrors, each called a heliostat, and focused onto the receiver at the top of the central tower to heat a material used to create steam and run a turbine to produce power for the California grid. In the late 1990's physicists got access to the facility and converted it into a gamma ray telescope, repairing and cleaning heliostats and installing a secondary mirror and camera on the tower. The camera consists of an array of photomultiplier tubes (PMTs) that take in photons and put out a current that is measured by our fast electronics. Since each PMT has its own channel and there are 80 channels, the electronics can also determine where in the heliostat field a particular photon originated.

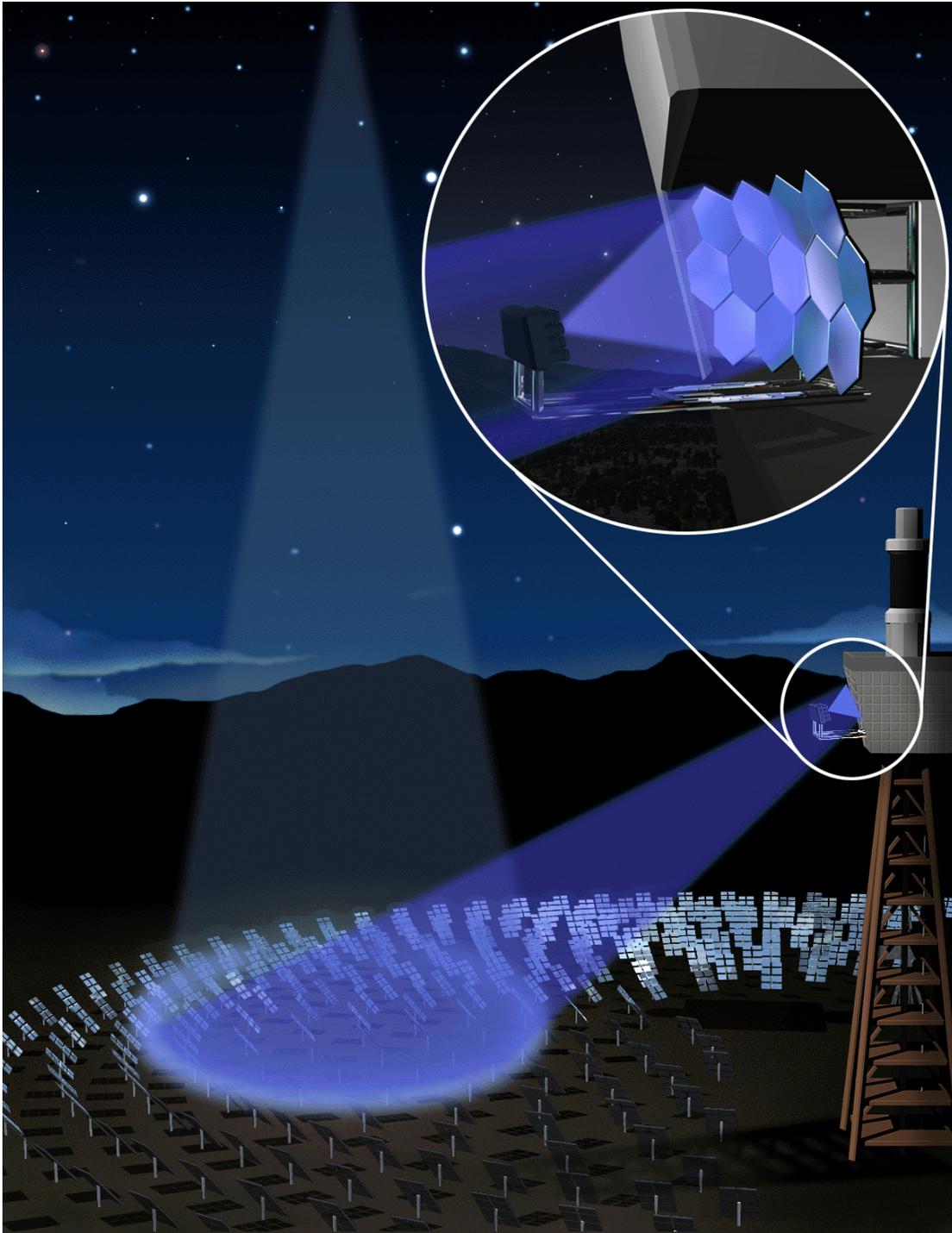


Fig 2. A schematic diagram of the C.A.C.T.U.S. telescope. A Cherenkov shower from overhead is focused by the heliostats onto the secondary mirror, which focuses it into the camera.

Probability Analysis

My part of the project was to begin to test one of the fundamental assumptions we make in data analysis— that the noise and signal data are uncorrelated, or that they are correlated in such a way that an increase in the noise will not produce a substantial increase in the signal. To do this, I want to calculate the probability of getting a trigger after taking the noise level into account. A “trigger” is a sufficient number of heliostats hit within a 10ns window, usually 13 or 15. This probability calculation is necessary because if we get a pulse height of 13, for example, it may be probable to have a noise count of 4 at any given time, which means that we have falsely registered a trigger. So let’s say that our threshold is 15. We want to find the probability for each time bin that we get *at least* 15, which is equal to 1 minus the probability of getting less than 15, since the total probability must be 1 and we have 80 channels. Mathematically,

$$P(15 \text{ or more}) = 1 - [P(0) + P(1) + \dots + P(14)]$$

To make matters more complicated, since it is impossible to tell directly whether a hit has come from noise or a shower, we have to consider all of the different ways we can get each number of hits. For example, the probability of getting a height of 2 is

$$P(2) = N(0)S(2) + N(1)S(1) + N(2)S(0)$$

where $N(1)$ is the probability of getting one hit from noise and $S(1)$ is the probability of getting one hit from signal. We calculate the signal probability by making a histogram of pulse height with the noise baseline subtracted. Because of the nature of statistical validity and the range of our electronics, the valid region of this histogram is the blue portion. The data in that region is fitted with a power law and extrapolated over all 80

channels. The extrapolation is then normalized such that it is now a probability function for the signal data.

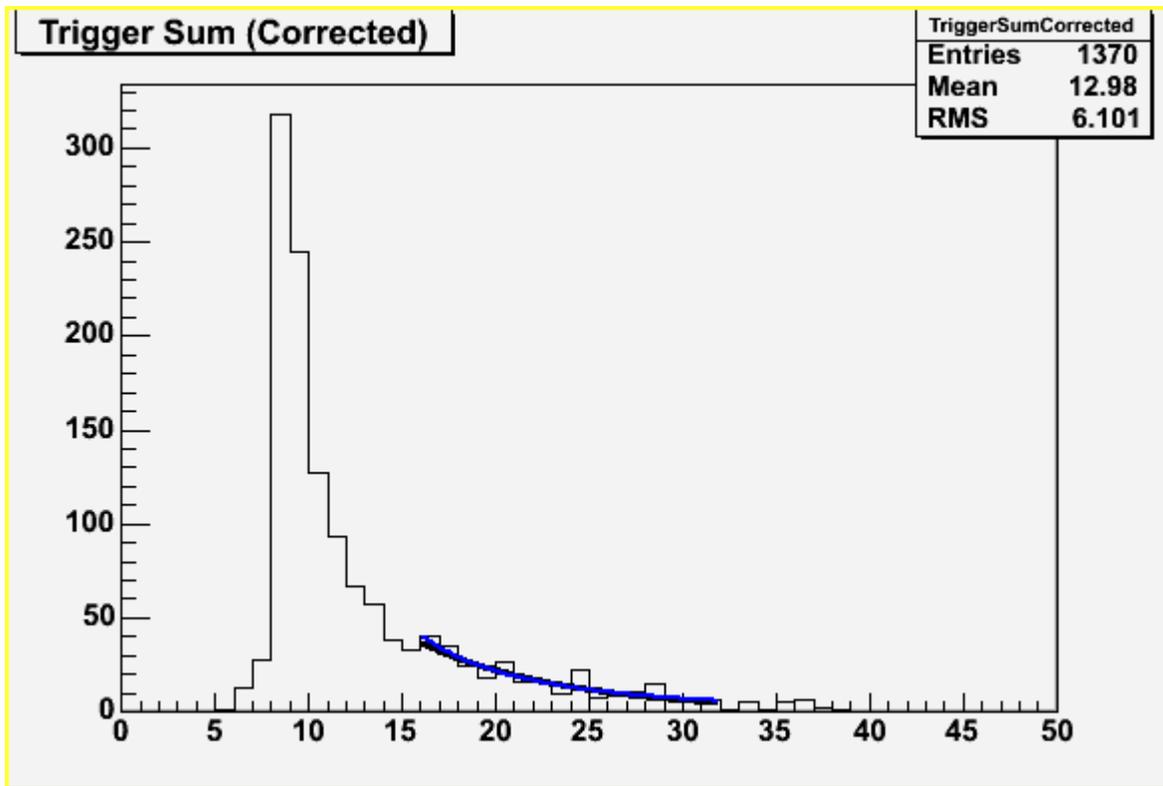


Fig 3. The original histogram, showing the fitting portion and fit.

Since the channel either has a hit or doesn't, the noise data is calculated using a binomial distribution. First we need the probability of a hit. The data stream can be thought of as a two dimensional graph with time on the x-axis. On the y-axis is the pulse height, which reflects the number of heliostats hit. If five heliostats are hit, the pulse height is five and so on. If the number of hits is above the threshold for that data run (11, 13, or 15 depending on the night) within a 10 ns window it is considered a trigger, and an air shower is inferred. The variables recorded by the electronics are the pulse height and the rise and fall times for the pulse within 0.5 ns. To find the probability of getting a hit from noise, we sample a portion of the data where we know there is no trigger and divide

the total time where we have a hit by the total time of the sample to find the average probability p that the channel has a hit. So then the probability of getting a hit on n out of 80 channels with a channel probability of p is

$$P_{\text{noise}}(n) = C(80,n) (p)^n (1-p)^{80-n}$$

Combining this noise probability distribution and the extrapolated signal probability distribution, we can calculate the total probability for each time bin getting a hit of at least a height of 15. For the purposes of graphing, this probability is multiplied by the width of the window to get a projected rate.

Preliminary Results

One immediately useful result of this probability can be found in the chart below, Figure 4. It is the plot of the projected trigger rate based on the probability for each bin versus the background rate, both in Hertz. The calculation was repeated several times with a different threshold value for each, producing a series of distinct curves. This would be useful to have tacked to the wall in the control room at Solar Two when we need to decide where to set the threshold for the evening's data. Sampling the background rate is simple and quick, so once we have that and a target trigger rate, we can quickly locate which threshold setting would best suit our needs, eliminating guesswork in choosing a threshold.

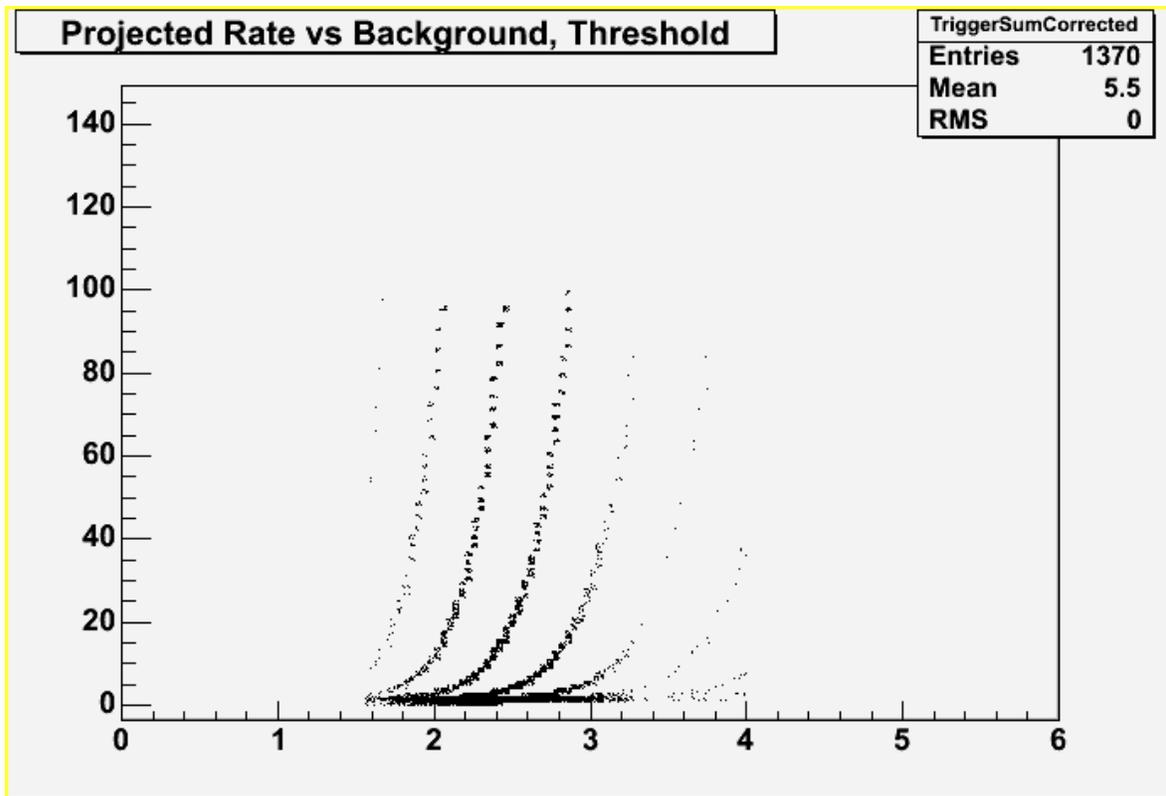


Fig 4. Plot of Projected Rate vs Background Rate for a number of Threshold values, ranging from 9 on the left to 25 on the bottom right.

One interesting and unexpected plot came from this study as well, shown below in Figure 5. This is the projected rate plotted versus the sigma value on the noise—essentially the fluctuation of the noise around its mean. The plot shows two distinct regions, one clear exponential in the $2 < x < 3$ range, and another distribution in the $1 < x < 2$ range. It is not yet clear to us what this means, but could indicate that there are two noise sources.

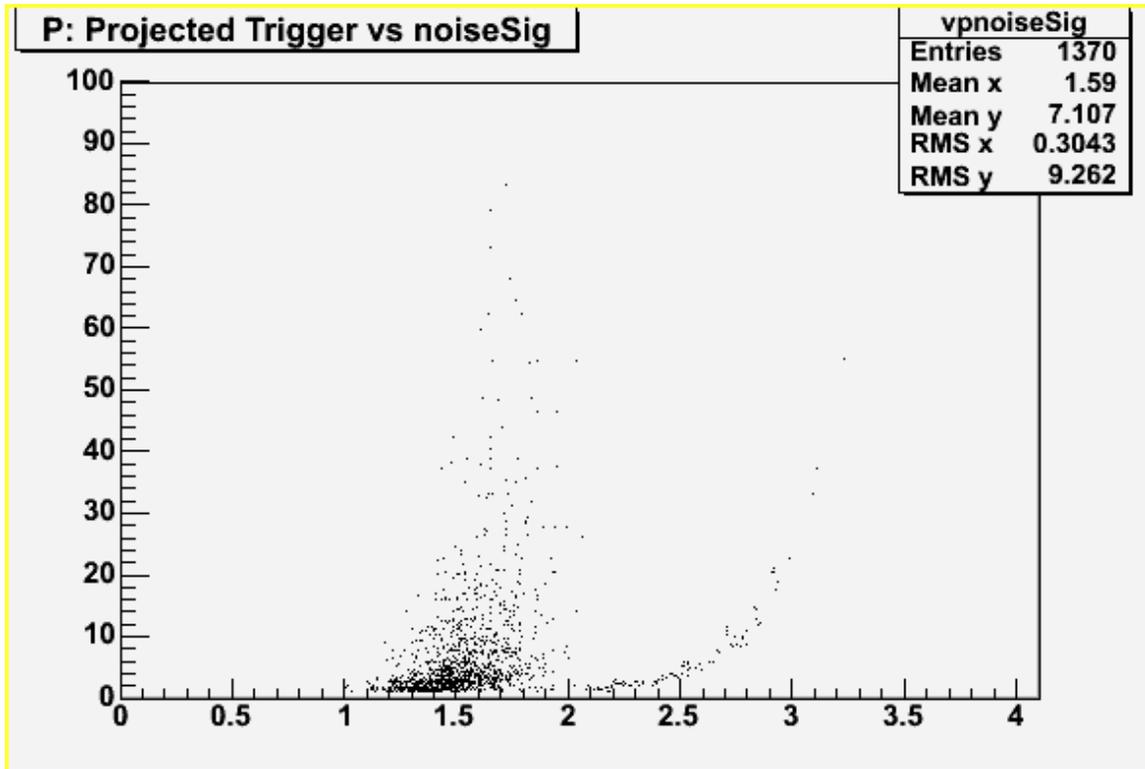


Fig 5. Projected Rate vs Noise's Sigma

Further Explorations

There is still a lot that can be done with this sort of analysis. This analysis was done assuming $\sigma=0$. In fact we have σ values for each data point, so we need to rerun this calculation to add error bars. We could also use an averaged value of the channel probability to cut down on the effects of fluctuation. And of course, we should investigate the source of the dual noise population.

Acknowledgements

Many thanks to Mani Tripathi for the chance to do this research, to NSF for making it financially possible, to Peter Marleau, David Isaacson, Max Chertok, John Stilley, and

John Felde for the physics insight, coding help and moral support. And last but not least, thanks to Rena Zieve for minimizing entropy.

More information, results, and pretty pictures can be found at the project's website that I redesigned in July, located at <http://ucdcms.ucdavis.edu/solar2/>