Project 4 - Gamma-ray Astronomy

Basic Analysis of Gamma-ray Observations

Due date: April 11th - Last Day of Classes

Contact: matthew.lundy@mail.mcgill.ca

Office Hours: Scheduled on Request with min 18 hrs notice

If there is anything ambiguous about the assignment please send me an email! Please do **not** send emails to the VERITAS team before reaching out to me!

Submission format: Please include *.pdf* submission that contains the requested plots and the answers to the questions asked in the following parts. A jupyter notebook can also be substituted. Please include any scripts used.

Background

This will be a ***short*** project regarding the "best" wavelength regime; gamma-ray astronomy. You will be taking a high level list of parameterized events, performing gamma hadron separation, background estimation, and implementing some basic statistical methods. This may seem like a lot, but it will be fun, so time will just fly by!

Gamma-ray astronomy deals with the very low statistics and involves the merging of particle physics techniques and astronomical targets. We will be dealing with VHE (Very High Energy > 100 GeV) gamma-rays and using data from the VERITAS observatory.

Installation

Astropy and python are required. Please see the first assignment for guidance on FITS handling.

Data Exploration

In this section we will be loading the file into python and looking at the spatial distribution of the events. This is a non-standard file format with four columns. Real parametrized data will often include >20 parameters per event including notably excluded parameters like energy. We will only be dealing with 2 spatial parameters and 2 image parameters (i.e. information about the shape of cherenkov light pool).

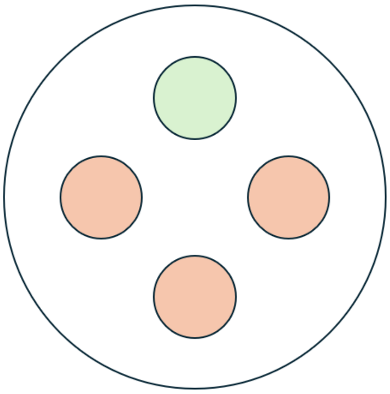
1. Load your .FITS file into python. Plot the sky position (right ascension and declination) of all events in the file. This can either be as a 2 dimensional histogram, or as points. Make sure that if you plot with points you have some way of determining if multiple points are overlapping (this can be achieved through a balance of point size and transparency).
2. You should be able to clearly identify a "hotspot" of events. Use *gamma-sky*(http://gamma-sky.net/) or Simbad (https://simbad.cds.unistra.fr/simbad/sim-fid) to identify your source. It should be a very famous supernova remnant with a photograph in the classroom. Record the coordinates of the source.
3. Also note the distribution of events towards the center and the edges. Do you see a trend? Qualitatively describe how the event density changes as function of distance from the center of the camera.
4. The pointing of the observation (which is the location of the center of the camera on the sky) is
   RA,Dec = 83.6329, 22.5258 deg
   Calculate the offset of this object from the pointing (right ascension and declination separately). Add the object as a SkyCoord object (https://docs.astropy.org/en/stable/api/astropy.coordinates.SkyCoord.html). You can use the default frame. You should note that one direction (i.e. RA and Dec) has a significant offset while the other is small. For the rest of the assignment you should treat the small offset as effectively 0 degrees.

Statistical Detection

We will now sum the number of events in a region around the source, and then estimate the background from several regions around the camera. We will then estimate the statistical significance of this measured excess. Make sure that you make the code in this section modular and easy to run with a modified event list as in the next section you will be running this part of the code ~1000 times.

1. Add each event into a Skycoord object and use the "separation" function to determine the absolute angular distance between each event and the source location. Sum the number of events that are 0.1 degrees away from the source location. This will be known as the number of N_{on} events.
2. Why would you use SkyCoord separation instead of just using the following:
   $dist = \sqrt{\Delta_{RA}^2 + \Delta_{Dec}^2}$
   with $\Delta_{RA} = RA_1 - RA_2$. There should be some part of the celestial sphere where it's inappropriate, identify one such region.
3. Repeat this summation process for 3 positions separated from the pointing of the telescope in the three cardinal directions from the center (i.e. sum the events in a 0.1 degree region centered on 3 other points). See the figure below which roughly illustrates the geometry of the regions. The total

number of events in the three regions together is known as the number of N_{off} events. We will use the events here to estimate the background in your source region.



4. Based on your answer to **Data Exploration: 3,** why is it important that these "Off regions" are the same distance from the camera center? What would happen if we were to select regions closer to the center or further away?
5. Apply the following formula, derived from the maximum likelihood ratio, to determine the significance:

$$S = \sqrt{-2 \ln \lambda} = \sqrt{2}\left\{N_{\mathrm{on}} \ln \left[\frac{1+\alpha}{\alpha}\left(\frac{N_{\mathrm{on}}}{N_{\mathrm{on}}+N_{\mathrm{off}}}\right)\right] + N_{\mathrm{off}} \ln \left[(1+\alpha)\left(\frac{N_{\mathrm{off}}}{N_{\mathrm{on}}+N_{\mathrm{off}}}\right)\right]\right\}^{1/2} .$$

   This is the famous Li & Ma (1983) formulation of the maximum likelihood. α is defined as the ratio of the area where N_{on} and N_{off} are collected (defined such that for a larger N_{off}, $\alpha < 1$). Determine α for your problem. From the paper (https://articles.adsabs.harvard.edu/pdf/1983ApJ...272..317L), determine what is the assumption about the statistical distribution of the signal and the background. Why is this appropriate for gamma-ray physics?
6. If you were to take a frequentist interpretation of the significance, what does this value that you calculated mean/represent?

Cut Optimization

In this section you will be taking the events and determining how you can get a higher significance by reducing the number of events. This is known as **cut optimization**. You will be using two parameters which correspond to the width, and the length of events. Your signal are gamma-rays and your background are hadrons. Our ability to seperate these relies on the fact that the light produced by gamma-rays has a distinct shape when compared to hadronic showers. Images that show this can be seen below.

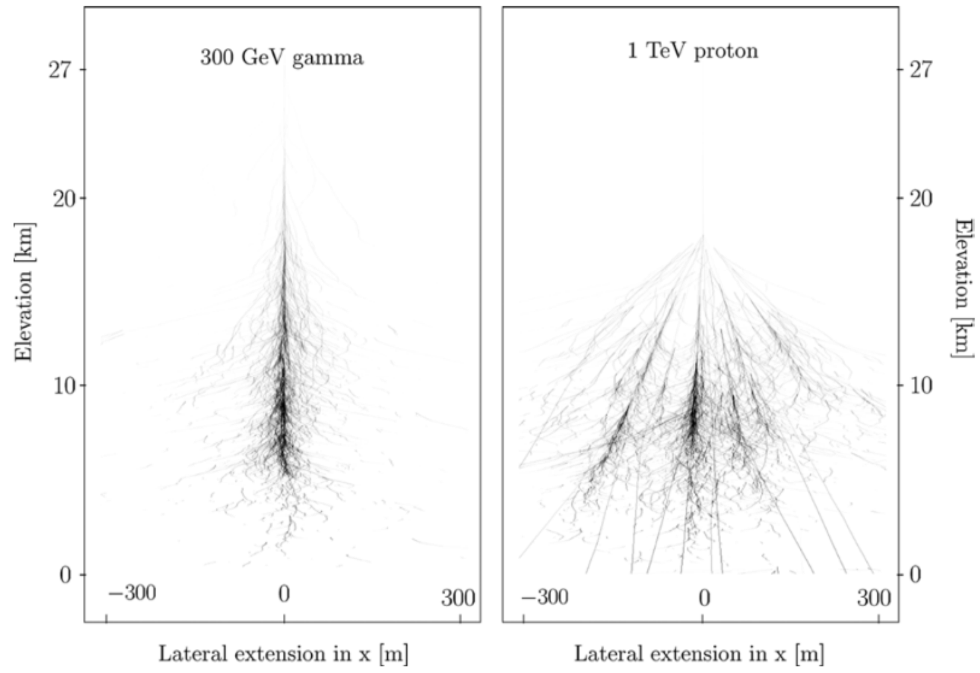


Fig. 1.3: *Comparison of the shower development for a leptonic and a hadronic shower (taken from [3]). While the electromagnetic shower (left side) shows only a small lateral spread compared to its longitudinal extension, the hadronic shower (right side) is quite a lot more extended and also more irregular in shape.*

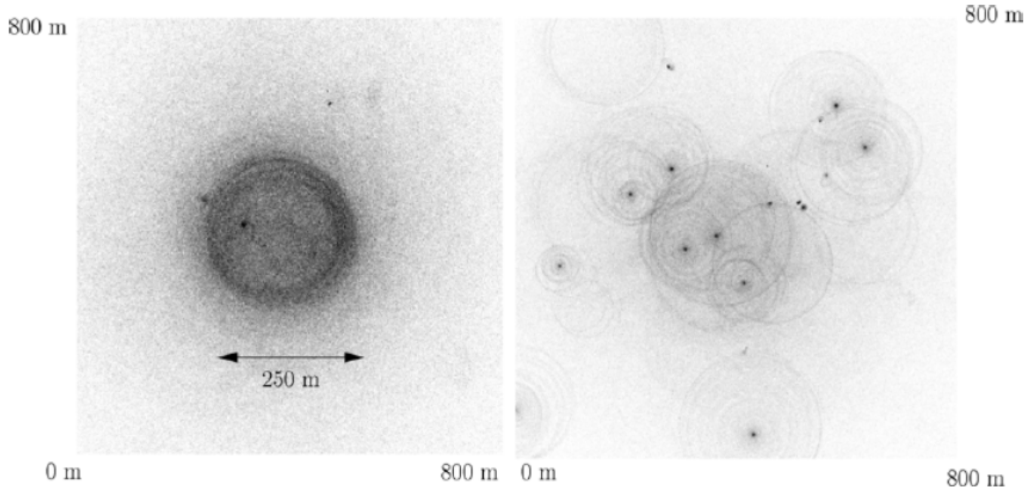


Fig. 1.5: *Distribution of Cherenkov photons on the ground level (taken from [3]). While an electromagnetic shower ($E = 300\,\mathrm{GeV}$) iluminates a circle with approximately 250 m diameter (left side) the distribution is far more irregular for a hadronic shower (right side, $E = 1\,\mathrm{TeV}$).*

Figures collected by Hupfer (2008).

1. Plot a histogram of the MSCWs (the mean scaled width, which corresponds to the average width of the telescope images), and separately of the MSCLs (the mean scaled length).
2. Repeat the determination of significance, as was performed in the previous section, but now with only events that fall below the following cuts. Note you will likely need something like a nested loop where for every MSCL cut you also check every MSCW cut:

MSCW: $0.9 \rightarrow 1.9$, steps of 0.1
MSCL: $0.9 \rightarrow 1.9$, steps of 0.1

3. Plot the 2D array of the significance for every cut combination. What is the optimal set of cuts on MSCW and MSCL that produce the most significant signal?
4. Suggest improvements that you could do to better choose these cuts. You can suggest a different method of determining the maximum of this parameter space or describe an entirely different way of probing the parameter space.

Marking

All sub parts are marked out of three. Comments are provided when appropriate.

1 point -> A substantial effort was put towards the question and there are clear steps taken towards the correct solution

2 points -> Some of the question has been done correctly and minor errors might be propagated throughout. This may also be given if there is a substantial misconception illustrated at some step that demonstrates a fundamental lack of understanding.

3 points -> The majority of question has been done correctly. Issues are minor and inconsequential.

In some cases half points may be given but the majority of times the divisions appear naturally in the subdivision of the question.

Progress Guidelines

Although there won't be strict progress markers you should roughly follow the time line:

- [ ] Tuesday April 8 : Finish Data Exploration
- [ ] Thursday April 10 : Finish Statistical Detection
- [ ] Friday April 11 : Finish Assignement

If you are falling beyond this timeline please reach out for help!