VERITAS: Gamma Ray Astronomy for Unknown Sources

William John-Pierre Duhe

Department of Physics, Columbia University
Nevis Laboratories
Department of Physics, Loyola University,

Abstract

Here I will give an overview of the science and techniques used at Nevis Laboratories, as part of the VERTIAS collaboration, to analyze gamma ray sources. We will offer an account of the analysis of two of these sources, UFOs (Unidentified Fermi Objects) 2FGL-J1115.0-0701 and 2FGL-J2004.6+7004. I will then move into how this analysis prompted an investigation into the algorithmic methods of solving the significance of such events.

\[1\text{wjdube@loyno.edu}\]
1 Introduction

We must first begin with an introduction to the objectives and nature of the research conducted within Nevis pertaining to gamma ray Astronomy. To do this let us begin with short answers to a few elementary questions sequentially.

- What is VERTIAS?
- What purpose does Gamma Ray astronomy serve in modern physics?
- What software do we utilize to study these Gamma Rays?

1.1 What is VERITAS:

VERITAS (Very Energetic Radiation Imaging Telescope Array System) is a ground-based gamma-ray instrument operating at the Fred Lawrence Whipple Observatory.
(FLWO) in southern Arizona, USA. It is an array of four 12m optical reflectors for gamma-ray astronomy in the GeV - TeV energy range. These imaging Cherenkov telescopes are deployed such that they have the highest sensitivity in the VHE energy band (50 GeV - 50 TeV), with maximum sensitivity from 100 GeV to 10 TeV. This VHE observatory effectively complements the NASA Fermi mission.[1]

1.2 What purpose does gamma ray astronomy serve in modern physics?:

Currently gamma rays are the least studied and least understood of the radiation spectrum. This is not due to their lack of importance, as they are created by some of the most powerful and important processes that our universe can behold, rather is is a consequence of our lack of technological capabilities to study them. One could spend an entire paper, or perhaps a book, discussing the different magnificent sources that emit gamma rays and the possible phenomenological nature of these sources. Thus, we shall narrow the scope of this discussion to one such source as the topic of this paper, the infamous Dark Matter (DM ). I will assume the reader has a sufficient understanding of the current discussion on DM, if one requires a review of this topic refer to [2].

It has been suggested in recent years that DM could self annihilate or decay to produce gamma radiation[3]. If we can isolate regions of of this DM then one can look at these regions and expect to see a certain amount of Gamma Rays being emitted at a specific spectrum. It is important to note that one must not only look at regions excessively dense with DM but must also ensure that there are no other known sources of gamma ray emitters to minimize the noise associated with our measurements and assure that we are observing rays produced from the hypothetical processes. Once we have seen a region of space that fits the following criteria;

1. a high count of gamma ray emission in GeV band
2. a hopeful quantity of DM
3. no observable sources of emission

then we can move onto trying to study whether or not this region could be producing gamma rays via the self interaction mechanisms mentioned above. If we see a strong source with a spectrum and fluxes associated with known models we can offer indirect proof of the existence of DM particles. If we see a strong source with spectrum and fluxes that can not be matched to any known models then we can likewise provide constraints on these models. There is a catalog of sources that meet the criteria above, according the the recent Fermi data, and they are categorized as Unidentified Fermi Objects or UFOs.
1.3 What software do we utilize to study these gamma rays?:

The main software packages used to analyze these sources are EventDisplay and VEGAS. Since the analysis provided is done in VEGAS we will focus on this software package here. The VERITAS Collaboration has developed VEGAS, the VERITAS Gamma-ray Analysis Suite, a data-analysis software package for the processing of single- and multiple telescope data produced by the array. The package consists of a core of six stages as well as visualization and diagnostic components. It has been developed in C++ using modern object-oriented design patterns to be highly flexible, configurable and extendable. VEGAS utilizes CERNs ROOT data-analysis framework and runs on Linux and Mac OS X systems.[4]

By analyzing a standard set of data pertaining to the Crab Nebula and comparing it with known results from other VEGAS users and other software suites we can assure our analysis methods are coherent.

2 2FGL-J1115.0+0701 and 2FGL-J2004.6+7004

Here I will offer up the analysis results for the two UFOs 2FGL-J1115.0-0701 and 2FGL-J2004.6+7004. 2FGL-J1115.0-0701 falls directly into the Fermi catalog of UFO’s. In order to narrow the origin of these enigmatic sources, multiwavelength data is essential. X-ray and radio data are available for most of the 2FGL sources, while a lack of good optical imaging and spectroscopic coverage demands dedicated observations. While 2FGL J2004.6+7004 shares in some of the characteristics associated with UFO’s it has also been suggested to be thought of as a radio-quiet AGN. Along with an optical characterization of the source, these observations could determine the nature of this interesting object aside from it’s relationship to DM.

<table>
<thead>
<tr>
<th></th>
<th>Sigma</th>
<th></th>
<th>Sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft Cuts</td>
<td>-1.4949</td>
<td>Soft Cuts</td>
<td>0.2071</td>
</tr>
<tr>
<td>Medium Cuts</td>
<td>-1.3597</td>
<td>Medium Cuts</td>
<td>-0.0326</td>
</tr>
<tr>
<td>Hard Cuts</td>
<td>-0.4049</td>
<td>Hard Cuts</td>
<td>0.5536</td>
</tr>
</tbody>
</table>

Table 1:The calculated Ring Background Model sigma values for each source considering soft, medium, and hard cuts.

<table>
<thead>
<tr>
<th></th>
<th>Sigma</th>
<th></th>
<th>Sigma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soft Cuts</td>
<td>-0.8232</td>
<td>Soft Cuts</td>
<td>0.6855</td>
</tr>
<tr>
<td>Medium Cuts</td>
<td>-1.8067</td>
<td>Medium Cuts</td>
<td>-0.6091</td>
</tr>
<tr>
<td>Hard Cuts</td>
<td>0.3312</td>
<td>Hard Cuts</td>
<td>0.2122</td>
</tr>
</tbody>
</table>

Table 2:The calculated Wobble Analysis sigma values for each source considering soft, medium, and hard cuts. See attached cuts files for details.
While a indicative sigma appeared for soft cuts in the EventDisplay analysis there was no counter-part to this anomaly in the VEGAS. This suggest statistical fluctuations as the source of the anomaly.

\subsection{Differential Flux Upper Limits with VEGAS:}

At one point inquired whether there any way I could control the energy binning associated with the VAUpper Limit output in vaStage6. I was attempting to have the VAUpper limit output to give multiple differential flux upper limit calculations for a particular set of bins that could be set by the user similar to the methods employed in EventDisplay. From this inquierey arose a bit of a debate. The explanation for the potential need for upper limits associated with individual bins can be stated as such:

The normalization factor that you can retrieve with the automatic settings in VEGAS is computed for the entire energy range, i.e. the statistical upper limit to the number of excess events that you use is computed from Non and Noff between your Emin and Emax, and thus your normalization factor depends on your overall sigma. The differential upper limit that one gets by multiplying that normalization factor by your assumed spectral shape evaluated at your chosen energy is, consequently, following a power law. I believe that this approach might be missing some relevant energy-dependent information, since the differential flux UL that you would expect from a given bin should depend on the statistical upper limit to the number of excess events in that specific bin, not on the overall upper limit to the number of excess events (e.g. if your overall significance is 0, and 1.5 for a certain bin, the differential flux upper limit VEGAS can provide would be over-constraining). Daniel Nieto - VEGAS Users eLog

One retort to such an assertion offered in defence of VEGAS’ methods was:

Since we don’t detect the source I wouldn’t worry too much about losing information. When we calculate our UL we use the full sensitivity of the instrument, i.e. all of the energies where we get have a non-zero effective area. Since we assume a source spectrum (as we must) when calculating an upper limit, once we have our upper limit, we can quote it at any energy and often we choose the threshold. Of course ”any” energy is probably not reasonable - but any energy at which VERITAS has reasonable sensitivity could be the energy where you quote a UL. This single-bin UL is the best we can do. -Taylor -VEGAS Users eLog

But which way yields a more accurate analysis? Are they essential the same? It appears there is no integrated way to calculate differential upper limits in multiple energy bins in VEGAS. My best understanding is that I would have to run vaStage6
multiple times, specifying the integral flux upper limit energy range each time. One must also adjust the base energy range, which is oddly in different units than the integral flux upper limit. You can then convert these results into a differential flux for particular energy bins. Then the two methods could be compared. When I underwent this task I arrived at the below values for the differential flux per energy bin.

<table>
<thead>
<tr>
<th>Bins (GeV)</th>
<th>( N_{on} )</th>
<th>( N_{off} )</th>
<th>StatisticalUL</th>
<th>Diff Flux</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 400 - 630 )</td>
<td>22</td>
<td>178</td>
<td>20.5</td>
<td>( 1.71 \times 10^{-10} )</td>
</tr>
<tr>
<td>( 630 - 1000 )</td>
<td>7</td>
<td>67</td>
<td>10.2</td>
<td>( 7.43 \times 10^{-13} )</td>
</tr>
<tr>
<td>( 1000 - 1580 )</td>
<td>2</td>
<td>46</td>
<td>4.07</td>
<td>( 5.17 \times 10^{-14} )</td>
</tr>
<tr>
<td>( 1580 - 2510 )</td>
<td>2</td>
<td>43</td>
<td>4.43</td>
<td>( 1.00 \times 10^{-14} )</td>
</tr>
<tr>
<td>( 2510 - 3980 )</td>
<td>3</td>
<td>25</td>
<td>7.7</td>
<td>( 3.07 \times 10^{-13} )</td>
</tr>
<tr>
<td>( 3980 - 6310 )</td>
<td>0</td>
<td>10</td>
<td>3.27</td>
<td>( 2.13 \times 10^{-13} )</td>
</tr>
<tr>
<td>( 6310 - 10000 )</td>
<td>0</td>
<td>2</td>
<td>3.97</td>
<td>( 1.19 \times 10^{-12} )</td>
</tr>
</tbody>
</table>

Table 3: Example of key values obtained for the various bins of 704 utilizing Vegas using Medium cuts.

Figure 1: Shown is a comparison between the single differential flux calculated by Vegas (blue) and the several differential flux values for different energy bins calculated by our macro (red). Left: Medium Cuts Right: Soft Cuts

One can see that the value obtained from the generic VEGAS algorithm is giving a higher differential flux upper limit than is available from the data. The designed macro also yields a more broad profile of how these limits are behaving at different energies. It seems one could obtain better upper limits by utilizing the macro’s developed however I would suggest further testing against known quantities. If testing confirms what we have seen initially it might be worth while to alter the VEGAS source code to incorporate these methods.
3 Determining Significance with Low Event Counts

Due to the sensitive nature of the VERTIAS telescope run time is often limited due to atmospheric or celestial conditions as well as the transient nature of Gamma Ray sources. A problem one can encountered when trying to analyze sources with relatively little run time is that very little event count data, $N_{on}N_{off}$, is made available. This is a problem due to the underlying mathematical methods we use to calculate the significance of a source. This can be somewhat mitigated by applying particular cuts to to your data runs.

Figure 2: Default settings: We are lacking in background data we see a heavy peak in the negative regime of our significance distribution.
Figure 3: On/Off Event Cuts at a 7/7 Ratio. By applying cuts to the on/off event count requirements for the bins we can smooth out this anomaly by eliminating the over counting of events on the boundary of our visual area where our statistics become incoherent.

Figure 4: On/Off Event Cuts at a 1/20 Ratio. One must realize though that these cuts must be applied to an appropriate ratio or they will falter as demonstrated above. If $\alpha$ is 0.05 then we would expect the ratio of on to off counts to be 1/20.
Figure 5: Tolerance Radius Cuts to 1.5*. Likewise we can filter out some of our boundary events that are polluting our statistics by reeling back the tolerance radius from 3* to 1.5* trimming the amount of events we take from our boundary.

Figure 6: Tolerance Radius Cuts to 1.5* and On/Off Event Cuts at a 1/20 Ratio. When we combine both of these cuts together we get the above corrections.
### Table 4: Key values obtained for the various cuts.

<table>
<thead>
<tr>
<th>Cuts</th>
<th>Mean Value</th>
<th>Sigma</th>
<th>$\chi^2/ndf$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Mods</td>
<td>$-0.048 \pm 0.014$</td>
<td>$0.848 \pm 0.008$</td>
<td>700.2/115</td>
</tr>
<tr>
<td>$7 - 7$</td>
<td>$0.152 \pm 0.032$</td>
<td>$0.834 \pm 0.010$</td>
<td>101.1/112</td>
</tr>
<tr>
<td>$1 - 20$</td>
<td>$0.846 \pm 0.032$</td>
<td>$0.609 \pm 0.031$</td>
<td>67.08/62</td>
</tr>
<tr>
<td>$TR1.5*$</td>
<td>$-0.095 \pm 0.017$</td>
<td>$0.871 \pm 0.011$</td>
<td>152.8/105</td>
</tr>
<tr>
<td>$TR1.5*1 - 20$</td>
<td>$-0.043 \pm 0.017$</td>
<td>$0.843 \pm 0.012$</td>
<td>124.7/102</td>
</tr>
</tbody>
</table>

### 3.1 Li and Ma’s Method

Aside from having to make these cuts, which can ”kind of” fix the problem, perhaps we could delve into the fundamental problem by elaborating on the methods that were developed by Li and Ma in 1983 [5] to calculate the significance. In this paper they study the effectiveness of three equations they have derived 5, 9, 17.

In order to test the significance of a source against possible background interference of each they preform Monte Carlo simulations and test the reliability of each of these formulas with the Null Hypothesis. They conclude that Eq.17 serves to most accurately model such a system with a range of alpha values and projected off counts. There is a small caveat to this derivation though as it is explicitly stated in their paper that it only holds accurate as long as on/off counts are not ”too few”. What does this mean and what exactly is ”too few”? Here we make an attempt to answer these questions by first recreating Li and Ma’s Monte Carlo simulations, in Mathematica, and exploring lower event count ranges. The left graph shows the points drawn from the equation with the Poission distribution. The middle graph shows these points set into a probability histogram and then plotted against a Normal distribution. The right most graph shows the Normal probability plot compared to an estimated normal distribution using the cumulative distribution function.

![Figure 7: Li and Ma’s equation 5 $N_{off} = 140, \alpha = .4$](image)


Here we see that each equation models the data relatively well, but this is because we have a lot of events! Much more than than the tentative minimum set by Li and Ma, $N_{\text{event}} > 10$. If we instead choose lets say $N_{off} = 8$ then we should expect to see much more deviation from the Normal Distribution. Since we have already validated our simulation for all three formulas we will now simply look at Equation 17.
Figure 11: Li and Ma’s equation 17 $N_{off} = 4, \alpha = .8$

We can see even for alternative $\alpha$ values we get a degregation of coherence when we drop in event counts. While several combinations of $\alpha$ and $N_{off}$ are plotted here one can manipulate the simple Mathematica script provided in the appendix to explore more values. I would also expect a similiar effect to occur if we were to similiarly test the Rolke method used to genereate the differential flux upper limits in the previous section. This would be the cause of some decoherance we have observed in our analysis.

### 3.2 Alternatives

It has been suggested and argued for in [6] and [7] that there are alternatives to the equation derived by Li and Ma to determine the significance of a Gamma Ray sources, particularly in cases where event count data is limited. Provided with the source code for Mathematica one could incorporate these formalisms as well and compare them with the results from Li and Ma to determine whether they might be appropriate as alternative methods to analyze sources which contain low event count information.

### 4 Conclusion

I have offered a motivation for the study of Gamma Rays as well as a surface level introduction to the VEGAS software package used for analysis.

Then I extended this into a demonstration of these motivations and methods via the analysis of UFOs 2FGL-J1115.0-0701 and 2FGL-J2004.6+7004. Sadly, we have seen neither of these sources are likely candidates for Dark Matter produced Gamma Radiation. Although they have not offered indirect proof as to the existence of Dark Matter particles their analysis has yielded useful information regarding some limits of the techniques employed by VEGAS.

Noted is the discrepancy between the methods used in EventDisplay and VEGAS to calculate the differential upper limits of weak sources and one possible approach to rectify such. I have also shown that for low enough event counts one can not
trust the significance calculations produced by VEGAS. While this can be somewhat mitigated by applying a particular set of cuts to our data analysis I have attempted to motivate an investigation into the use of new alternative algorithms to determine the significance of such sources. This could become ever more useful as the efficiency of our instruments increases and we begin to capture ever more transient events.

5 Appendix

5.1 Mathematica Script

\[
\text{off} = 4; (*\text{Expected Number of Off Counts}*) \\
\text{\textalpha} = .8; (*\text{Ratio of On/Off Counts}*) \\
\text{\textbeta} = 100000; \\
\text{\textbeta} = \text{RandomVariate[}\text{PoissonDistribution[}\text{\textalpha}\times\text{off}], 100000]; \\
\text{Non} = \text{RandomVariate[}\text{PoissonDistribution[}\text{off}], 100000]; \\
\]

(*\text{Normal Distribution}*)

\[
\text{h2} = \text{Plot[}\text{PDF[NormalDistribution[]}, \text{x}, \{\text{x}, -6, 6\}, \\
\text{PlotStyle} \rightarrow \text{Directive[Red, Thick]}]; \\
\text{Attributes /@ \{Greater, Sign\}} \\
\text{S1[off_, \text{\textalpha}_\_]} = \\
\text{Sign[Non - \text{\textalpha}\times\text{Noff}]\times \text{Sqrt}[2]\ (\text{Non}\times\text{Log}\left[\frac{1 + \text{\textalpha}}{\text{\textalpha}}\ \frac{\text{Non}}{\text{Non} + \text{Noff}}\right] + \\
\text{Noff}\times\text{Log}\left[\frac{1 + \text{\textalpha}}{\text{\textalpha}}\ \frac{\text{Noff}}{\text{Non} + \text{Noff}}\right]\}^{1/2}; \\
\]

\text{ListPlot[S1[off, \text{\textalpha}], Frame \rightarrow \text{True}, Axes \rightarrow \text{False}, \\
\text{PlotLabel} \rightarrow \text{"Formula 17"}]} \\
\]

\text{hist22 = \text{Histogram[S1[off, \text{\textalpha}], 26, "ProbabilityDensity", \\
\text{PlotLabel} \rightarrow \text{"Formula 17", AxesLabel} \rightarrow \text{"Probability"}; \\
\]

\text{Show[hist22, h2]} \\
\]

\text{ProbabilityScalePlot[S1[off, \text{\textalpha}], "Normal", \\
\text{PlotLabel} \rightarrow \text{"Formula 17"}]}
\[ S_2[\text{off}_\_, ~\alpha] = \frac{(\text{Non} - \alpha \times \text{Noff})}{(\text{Non} + \alpha^2 \times \text{Noff})^{1/2}}; \]

\begin{verbatim}
ListPlot[S2[off, \[Alpha]], PlotLabel -> "Formula 9"]
hist2 = Histogram[S2[off, \[Alpha]], 26, "ProbabilityDensity",
   PlotLabel -> "Formula 9", AxesLabel -> "Probability"];
Show[hist2, h2]
ProbabilityScalePlot[S2[off, \[Alpha]], "Normal",
   PlotLabel -> "Formula 9"]
\end{verbatim}

\[ S_3[\text{off}_\_, ~\alpha] = \frac{(\text{Non} - \alpha \times \text{Noff})}{(\alpha \times (\text{Non} + \text{Noff})^{1/2}; \]

\begin{verbatim}
ListPlot[S3[off, \[Alpha]], PlotLabel -> "Formula 5"]
hist3 = Histogram[S3[off, \[Alpha]], 26, "ProbabilityDensity",
   PlotLabel -> "Formula 5", AxesLabel -> "Probability"];
Show[hist3, h2]
ProbabilityScalePlot[S3[off, \[Alpha]], "Normal",
   PlotLabel -> "Formula 5"]
\end{verbatim}

5.2 Macro for Extracting Differential Flux Upper Limits

* \file templateMacro.C
* \ingroup macros
* \brief Prints stored config file from any stage
* Original Authors: William Duhe and Daniel Nieto

\begin{verbatim}
void diffUL()
{
   cout<<"Generates differential flux upper limit"<<endl;
   cout<<"--------"<<endl;
   cout<<"The macro header looks like this: "<<endl;
   cout<<"templateMacro(const string &filename, const ostream& out = cout)"<<endl;
   cout<<endl;
   cout<<"or you can also use this if you already have an active loaded instance of VARootIO";
   cout<<"templateMacro(VARootIO &io, const ostream& out = cout)"<<endl;
   cout<<endl;
\end{verbatim}
cout<<"Example:"<<endl;
cout<<"---------"<<endl;
cout<<"templateMacro("myfile.root");"<<endl;
cout<<"or if you wanted the output to a text file you could do"<<endl;
cout<<"ofstream outfile("output.txt")"<<endl;
cout<<"templateMacro("myfile.root", outfile);"<<endl;
}

void diffUL(const char *filename, const ofstream& out = cout)
{
    // VARootIO io(filename, true); // the boolean here refers to the file being read
    // io.loadTheRootFile();
    /* TFile io(filename);
       if(io.IsOpen())
         diffUL(io,out);
       else
         {
             out<<"The root file "<<filename<<" could not be opened"<<endl;
             return;
         }
    */

    TFile f(filename);
    if (f.IsZombie()) {
        cout << "Error opening file" << endl;
        exit(-1);
    }
    f.ls();

    VAUpperLimit *upperLimit = new VAUpperLimit();
    upperLimit->GetObject("VAUpperLimit", upperLimit);
    Double_t statisticalUL = upperLimit->GetUL();
    cout << "Statitical UL: " << statisticalUL << endl;
    Double_t PI = upperLimit->GetPhotonIndex();
    cout << "PHOTONINDEX: " << PI << endl;
    Double_t RunT = upperLimit->GetLiveTime();
    cout << "LiveTime: " << RunT << endl;

    Double_t E_0 = 1;
    TF1 *sshape = new TF1("sshape","pow(pow(10,x)/[0],[1]),-3,3");
    TF1 *DeltaE = new TF1("DeltaE","pow(10,x+.02)-pow(10,x-.02)",-3,3);
sshape->SetParameter(0,E_0);
sshape->SetParameter(1,PI);

TGraphAsymmErrors *EA = (TGraphAsymmErrors*) upperLimit->GetEffectiveArea();
EA->Draw("AL");
//TH1F *EA_hist = (TH1F*) EA->GetHistogram();
//EA_hist->Draw();
// TGraphAsymmErrors EA = upperLimit->GetEffectiveArea();
cout << "EffectiveArea: "<<EA<<endl;
Int_t Npoints = EA -> GetN();
Double_t xval = 0;
Double_t yval = 0;
Double_t Emin = 6.31; // energy in TeV
Double_t Emax =10; // energy in TeV
Double_t EASDE =0;
cout << Npoints << endl;
TGraph *EAS = new TGraph(Npoints);

for(Int_t i =0;i < Npoints; i++)
{
    EA->GetPoint(i,xval,yval);
    //cout << i << " " << xval << " " << yval << " " << sshape(xval) << " " << DeltaE(xval) <<endl;
    EAS->SetPoint(i,xval,yval*sshape(xval));

    if(xval>log(Emin) && xval<log(Emax))
    {
        EASDE +=DeltaE(xval)*yval*sshape(xval);
        // cout << DeltaE(xval)*yval*sshape(xval) << " " << EASDE <<endl;
    }
}

EAS->Draw("same");

AnalysisSummary->cd();
Double_t mean = (log10(Emax)+log10(Emin))/2;
cout << (statisticalUL)/(RunT*EASDE*pow(10,4))<<endl;
cout << ((statisticalUL)/(RunT*EASDE*pow(10,4)))*sshape(mean)<< endl;
cout << mean << endl;
cout << "Differential UL = " << ((statisticalUL)/(RunT*EASDE*pow(10,4)))*sshape(mean) << " @ " << pow(10,mean) << " TeV";
}

// Extract Differential Flux

void diffUL(VARootIO &io, const ostream& out = cout)
{
    io.loadTheRootFile();
    if(!io.IsOpen())
    {
        out<<"The root file "<<filename<<" could not be opened"<<endl;
        return;
    }

    io.closeTheRootFile();
}
References


