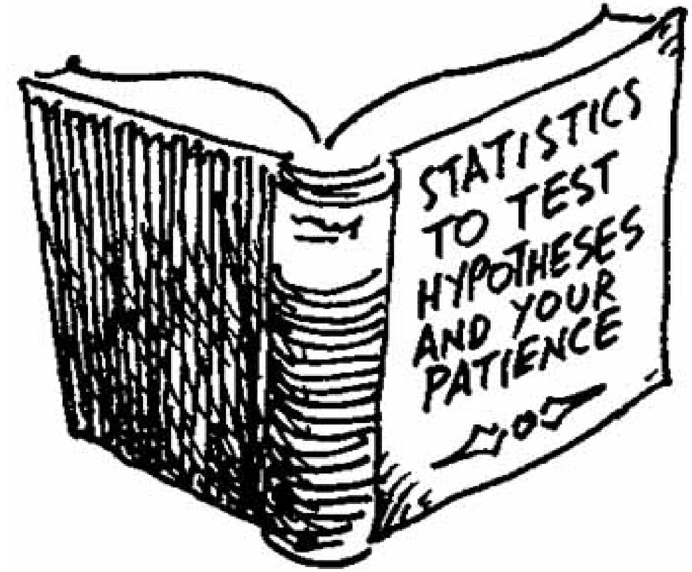


Maximum Likelihood Estimation & Gamma-ray Spectral Analysis

David C. Y. Hui

Department of Astronomy & Space Science
Chungnam National University

- Basic Concept of Maximum Likelihood (ML) Estimation
- ML estimation in Gamma-ray Astronomy
- Instrumental Response Function (IRF)
- Source Model & Likelihood Function
- Binned / Unbinned Analysis
- Significance (TS value)
- Practical Procedures with Fermi Science Tools

Data Modeling

Question: “**What is the prob. that a particular set of fitted parameters is correct?**”

Data Modeling

Question: "**What is the prob. that a particular set of fitted parameters is correct?**"

- **There is only one correct model!!!**

Data Modeling

Question: "**What is the prob. that a particular set of fitted parameters is correct?**"

- **There is only one correct model!!!**

The correct question should be:

"For a given set of model parameters, what is the prob. that the observed data could be drawn?"

Data Modeling

Question: "**~~What is the prob. that a particular set of fitted parameters is correct?~~**"

- **There is only one correct model!!!**

<u>The correct question should be:</u>

"For a given set of model parameters, what is the prob. that the observed data could be drawn?"

This probability is called -- *Likelihood*

Maximum Likelihood Estimation

For estimating the **"true"** values of the parameters, we find those values that maximize the likelihood.

This form of parameter estimation is called:

Maximum Likelihood (ML) Estimation

Definition:

- Suppose the likelihood function *L* depends on *k* parameters, θ_1 , θ_2.....θ_k, choose as estimates those values of the parameters that maximize *L* for a single observation.

Intuitive example:

- Suppose we survey 20 individuals in a city to ask each if they support the new government policy. If in our sample, 6 favored the new policy, find the estimate for *p*, the true but unknown fraction of the population that favor the new policy.

$$P(Y=6)=\binom{20}{6} p^6 (1-p)^{14}$$

Find *p* maximize P(Y=6):

$$\frac{d\,[6 \ln(p) + 14 \ln(1-p)]}{dp} = \frac{6}{p} - \frac{14}{1-p} = 0$$

$$p = 6/20$$

Not-so Intuitive example:

- Suppose we interview successive individuals in the city and stop interviewing when we find the person who favor the policy. If the fifth person is the first one who likes the policy, find the estimate for *p*, the true but unknown fraction of the population that favor the new policy.

$$P(X=5) = (1 - p)^4\, p$$

Find *p* maximize P(X=5):

$$\frac{d\,[4 \ln (1 - p)] + \ln (p)}{dp} = \frac{1}{p} - \frac{4}{1 - p} = 0$$

$$p = 1/5$$

ML in γ-ray Astronomy

1. **Deconvolution** tends to non-unique and unstable to the presence of Possion noise.

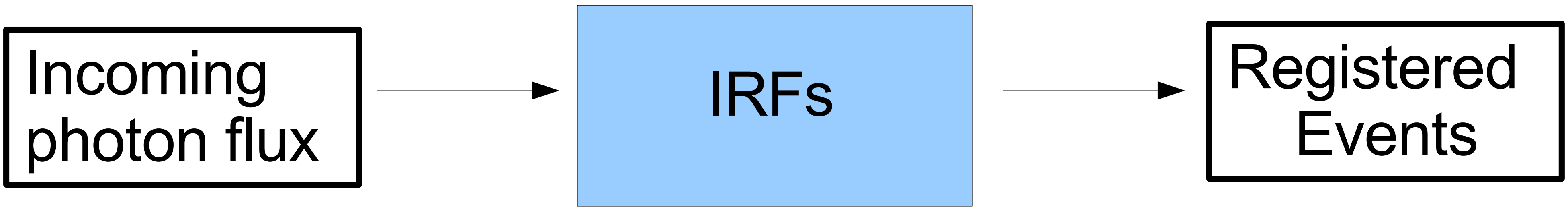
2. Usual χ^2**-fitting** is also a kind of ML estimation. It assumes that the number of counts in each bin will have a Gaussian distribution with an expectation value equal to the model counts.

In general, this condition is not satisfy in LAT data with reasonable binning.

3. **Source confusion –** The PSF of LAT is board. Special care has to be taken to account for the contribution of the PSF wings of surrounding bright sources in our **R**egion-**O**f-**I**nterest (**ROI**).

=> ***Multi-dimensional analysis is required for LAT data.***

Instrumental Response Functions



1. Hardware Responses

2. Assigning probability of a event is resulted from an astrophysics photon.
**** IRFs and the selected events have to be matched.****

<u>Parametrized IRF:</u> E – Energy p - Position

$$\boldsymbol{R}(E',p' ; E,p,t) = \boldsymbol{A}(E,p,t)\, \boldsymbol{P}(p' ; E,p,t)\, \boldsymbol{D}(E' ; E,p,t)$$

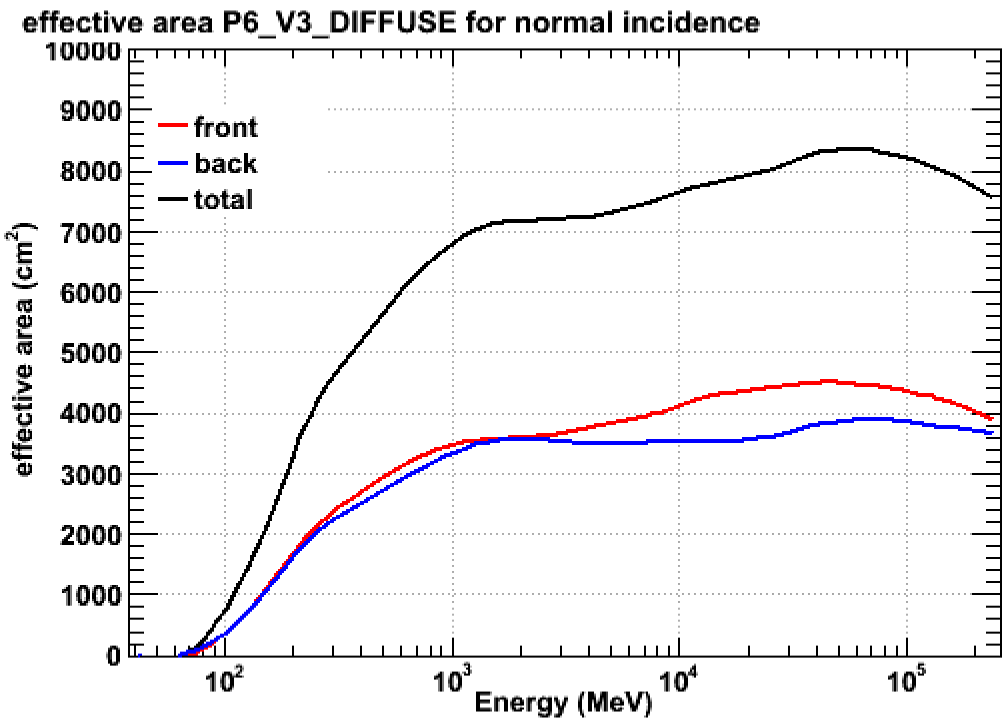
Effective Area

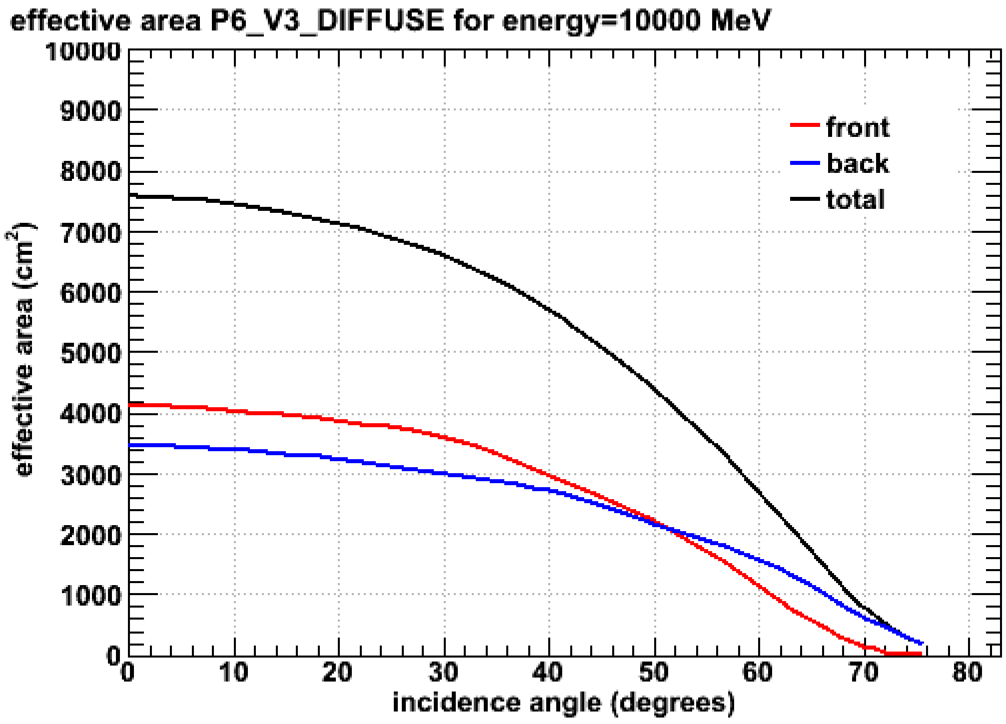
PSF

Energy dispersion
(Ignored in LAT analysis)

Effective Area

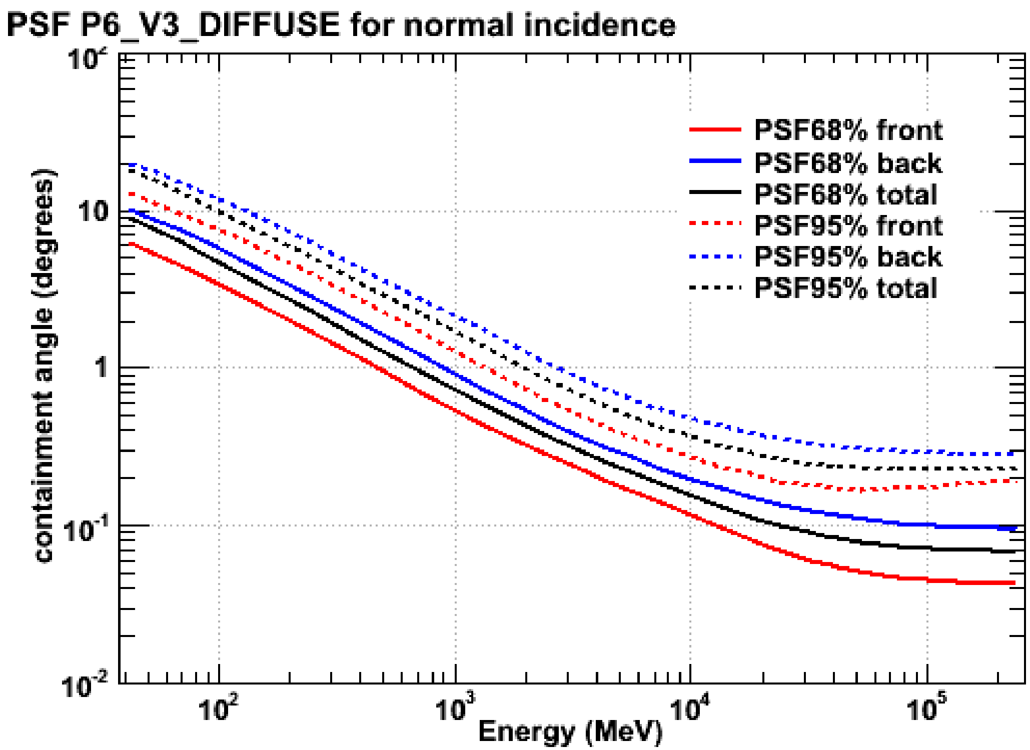
The ability to detect a photon of given energy *E* and position *p.*

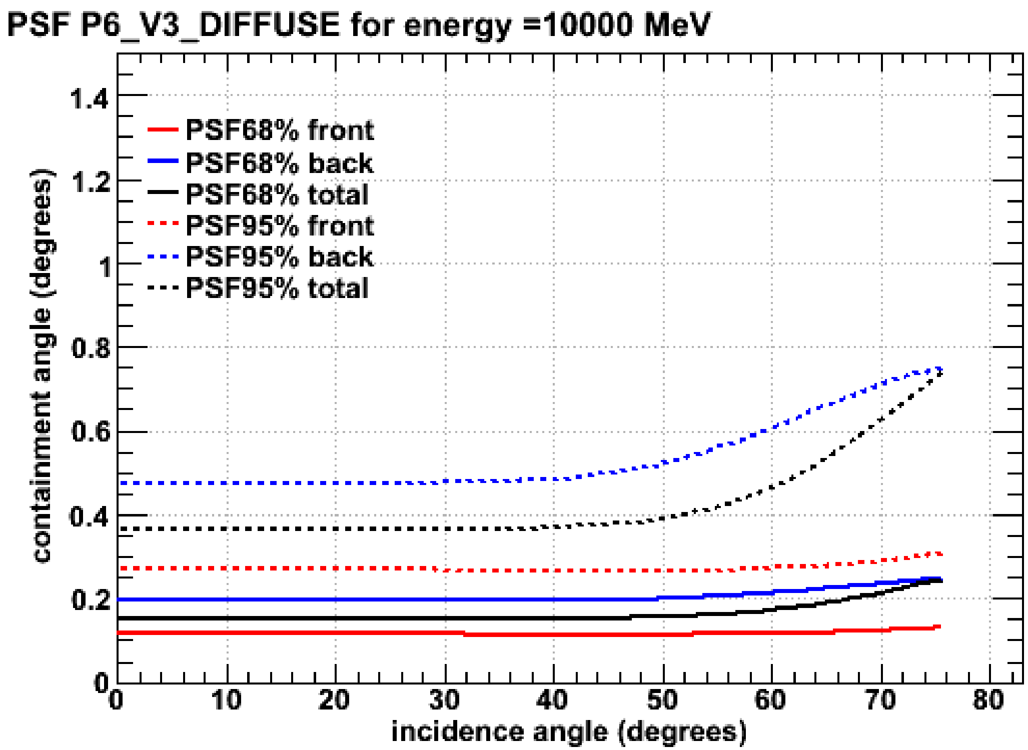




Point Spread Function

The probability that a photon of given energy *E* and position *p* is registered on a given position on the detector at *p'*.





Source Model

For a single point source:

$$S_i(E,p) = \epsilon(E)\,\delta(p-p_i)$$

$*$ $\epsilon(E)$ — Spectral model (e.g. Power law)

In view of source confusion, we must also consider nearby sources:

Complete source model:

$$S(E,p) = \sum_i S_i(E,p)$$

Likelihood Function

Assume the LAT data are binned in to many energy bins, each bin will contain a small number of counts:

No. of counts in each bin is characterized by the Poisson distribution

Probability of detecting n_j **counts in the** j^{th} **bin:**

$$P_j = m_j^{n_j} \exp(-m_j) / n_j!$$

* m_j - expected number of counts in the j^{th} bin which is the **convolution of source model and IRF**.

Likelihood Function

Assume the LAT data are binned in to many energy bins, each bin will contain a small number of counts:

No. of counts in each bin is characterized by the Poisson distribution

Probability of detecting n_j counts in the j^{th} bin:

$$P_j = m_j^{n_j} \exp(-m_j) / n_j!$$

* m_j - expected number of counts in the j^{th} bin which is the **convolution of source model and IRF**.

<u>Likelihood function *L*:</u>

$$L = \prod_j P_j$$

Likelihood Function

Assume the LAT data are binned in to many energy bins, each bin will contain a small number of counts:

No. of counts in each bin is characterized by the Poisson distribution

Probability of detecting n_j counts in the j^{th} bin:

$$P_j = m_j^{n_j} \exp(-m_j) / n_j!$$

* m_j - expected number of counts in the j^{th} bin which is the **convolution of source model and IRF**.

<u>Likelihood function *L*:</u>

$$L = \prod_j P_j$$

Now the task is: **Find out the set of spectral parameters of the adopted model so as to maximize *L***

To Bin or Not to Bin ?

To Bin or Not to Bin ?

Likelihood function for finite bin size

$$L = \exp(-N_{predict}) \prod_j m_j^{n_j} / n_j!$$

To Bin or Not to Bin ?

Likelihood function for finite bin size

$$L = \exp(-N_{predict}) \prod_j m_j^{n_j} / n_j!$$

*__Trade-off:__ Accuracy vs. bin-size

To Bin or Not to Bin ?

Likelihood function for finite bin size

$$L = \exp(-N_{predict}) \prod_j m_j^{n_j} / n_j!$$

***Trade-off:** Accuracy vs. bin-size

When the bin-size tends to zero: $n_j = 0, 1$

Unbinned **Likelihood function**

$$L_{unbin} = \exp(-N_{predict}) \prod_j m_j$$

To Bin or Not to Bin ?

Likelihood function for finite bin size

$$L = \exp(-N_{predict}) \prod_j m_j^{n_j} / n_j!$$

***Trade-off:** Accuracy vs. bin-size

When the bin-size tends to zero: $n_j = 0, 1$

***Unbinned* Likelihood function**

$$L_{unbin} = \exp(-N_{predict}) \prod_j m_j$$

L_{unbin} allows the most accurate spectral analysis!

To Bin or Not to Bin ?

Likelihood function for finite bin size

$$L = \exp(-N_{predict}) \prod_j m_j^{n_j} / n_j!$$

***Trade-off:** Accuracy vs. bin-size

When the bin-size tends to zero: $n_j = 0, 1$

***Unbinned* Likelihood function**

$$L_{unbin} = \exp(-N_{predict}) \prod_j m_j$$

L_{unbin} allows the most accurate spectral analysis!

** Practical consideration: *Computation Time!!!*

Test Statistic (TS) & Significance

***Test statistic* (TS):** -2 times the logarithm of the ratio of the likelihood for the model without the additional source (the null hypothesis) to the likelihood for the model with the additional source.

Wilks' Theorem: For sufficiently large no. of photon,
If there is no additional source then the TS should be drawn from a χ^2_ν distribution, where ν is the difference in the degree of freedom between the models with and without the additional source.

Test Statistic (TS) & Significance

***Test statistic* (TS):** -2 times the logarithm of the ratio of the likelihood for the model without the additional source (the null hypothesis) to the likelihood for the model with the additional source.

Wilks' Theorem: For sufficiently large no. of photon,
If there is no additional source then the TS should be drawn from a χ^2_ν distribution, where ν is the difference in the degree of freedom between the models with and without the additional source.

Integrating χ^2_ν from the observed TS value to infinity gives the probability that the apparent source is a fluctuation.

Test Statistic (TS) & Significance

***Test statistic* (TS):** -2 times the logarithm of the ratio of the likelihood for the model without the additional source (the null hypothesis) to the likelihood for the model with the additional source.

Wilks' Theorem: For sufficiently large no. of photon,
If there is no additional source then the TS should be drawn from a χ^2_ν distribution, where ν is the difference in the degree of freedom between the models with and without the additional source.

Integrating χ^2_ν from the observed TS value to infinity gives the probability that the apparent source is a fluctuation.

Significance for the additional source: ~$(\mathrm{TS})^{1/2}\,\sigma$

Practical Procedures with Tools

Revision for Albert's Lecture

1. Filtering Data

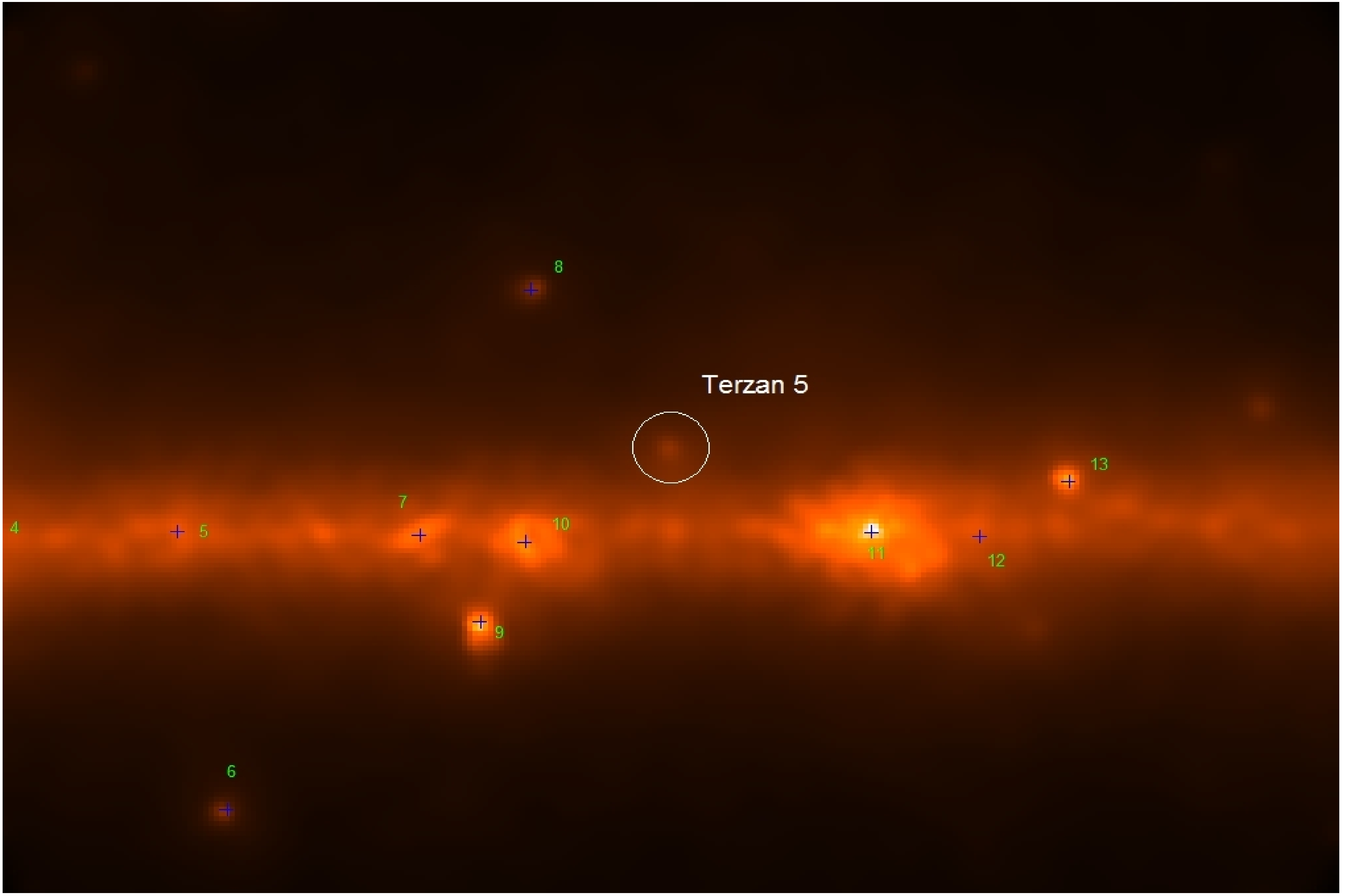
```
gtselect evclsmin=3 evclsmax=3 infile=@evt.txt outfile=ter5_filtered_5deg.fits \
  ra=267.02 dec=-24.7792 rad=5 tmin=239557417 tmax=282529415 emin=200 \
  emax=300000 zmax=105

gtmktime scfile=L091218021133E0D2F37E85_SC00.fits \
filter="IN_SAA!=T&&DATA_QUAL==1" \
roicut=yes  evfile=ter5_filtered_5deg.fits \
outfile=ter5_filtered_5deg_gti.fits
```

2. Creating Image

```
gtbin evfile=ter5_filtered_5deg_gti.fits  \
  scfile=NONE outfile=ter5_filtered_5deg_gti_img_gal.fits \
  algorithm=CMAP nxpix=100 nypix=100 binsz=0.1 coordsys=GAL \
  xref=3.8392487 yref=1.6869037 axisrot=0 proj=AIT
```

Practical Procedures with Tools



Practical Procedures with Tools

3. Generating an exposure map

3a. Calculating integrated livetime as a function of sky position and off-axis angle:

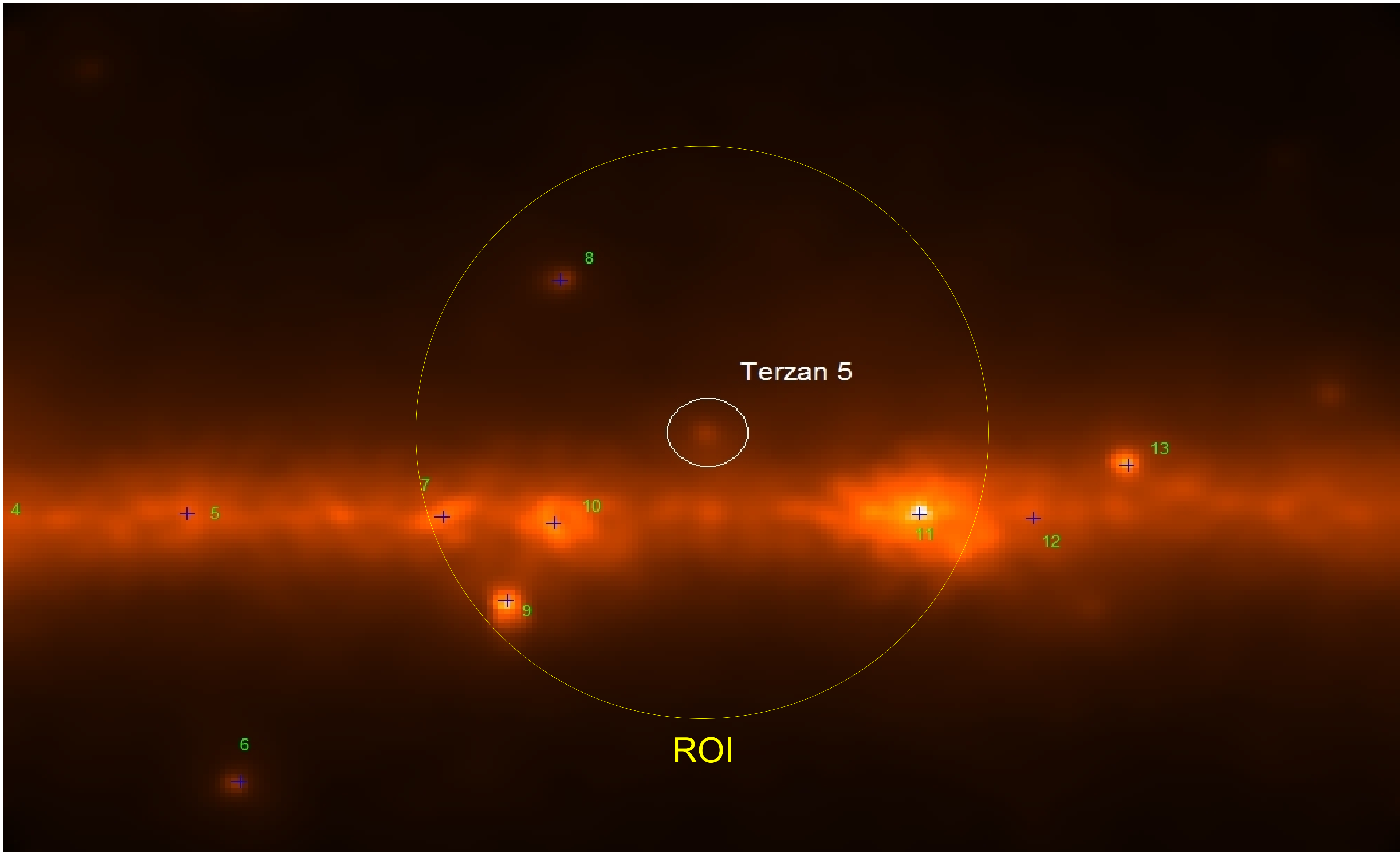
gtltcube evfile=ter5_filtered_5deg_gti.fits scfile=L091218021133E0D2F37E85_SC00.fits \
outfile=expcube.fits dcostheta=0.025 binsz=1

3b. Calculating exposure map

gtexpmap evfile=ter5_filtered_5deg_gti.fits scfile=L091218021133E0D2F37E85_SC00.fits \
expcube=expcube.fits outfile=expmap_10deg.fits irfs=P6_V3_DIFFUSE \
srcrad=10 nlong=120 nlat=120 nenergies=20

Practical Procedures with Tools

4. Defining Source Model



Practical Procedures with Tools

4. Defining Source Model

Two types of model:

a. Diffuse Source

i) Galactic diffuse: gll_iem_v02.fit

ii) Extragalactic diffuse: isotropic_iem_v02.txt

b. Point Source

Examples:

Power-law

$$\frac{dN}{dE} = N_0 \left(\frac{E}{E_0} \right)^{\gamma}$$

N_0	γ	E_0
Prefactor	Photon Index	Scale

Practical Procedures with Tools

4. Defining Source Model

Two types of model:

a. **Diffuse Source**

i) Galactic diffuse: gll_iem_v02.fit

ii) Extragalactic diffuse: isotropic_iem_v02.txt

b. **Point Source**

Examples:

Super-exponential Cutoff

$$\frac{dN}{dE} = N_0 \left(\frac{E}{E_0}\right)^{\gamma_1} \exp\left(-\left(\frac{E}{E_c}\right)^{\gamma_2}\right)$$

N_0	E_0	γ_1	γ_2	E_c
Prefactor	Scale	Index1	Index2	Cutoff-energy

Practical Procedures with Tools

4. Defining Source Model (xml file)

4a. Galactic & Extragalactic diffuse gamma-rays

```
<?xml version="1.0" ?>
<source_library title="source library">
<source name="EG_v02" type="DiffuseSource">
<spectrum file="../isotropic_iem_v02.txt" type="FileFunction">
<parameter free="1" max="1000" min="1e-05" name="Normalization" scale="1" value="5.39258" />
</spectrum>
<spatialModel type="ConstantValue">
<parameter free="0" max="10.0" min="0.0" name="Value" scale="1.0" value="1.0"/>
</spatialModel>
</source>

<source name="GAL_v02" type="DiffuseSource">
<!-- diffuse source units are cm^-2 s^-1 MeV^-1 sr^-1 -->
<spectrum type="ConstantValue">
<parameter free="1" max="10.0" min="0.0" name="Value" scale="1.0" value= "0.981163"/>
</spectrum>
<spatialModel file="../gll_iem_v02.fit" type="MapCubeFunction">
<parameter free="0" max="1000.0" min="0.001" name="Normalization" scale= "1.0" value="1.0"/>
</spatialModel>
</source>
```

Practical Procedures with Tools

4. Defining Source Model (xml file)

4b. Source-of-interest

```
<source name="Terzan 5" type="PointSource">
<!-- point source units are cm^-2 s^-1 MeV^-1 -->
<spectrum type="PLSuperExpCutoff">
<parameter free="1" max="1000" min="1e-05" name="Prefactor" scale="1e-08" value="0.125358"/>
<parameter free="1" max="0" min="-5" name="Index1" scale="1" value="-1.70573"/>
<parameter free="0" max="1000" min="50" name="Scale" scale="1" value="100"/>
<parameter free="1" max="30000" min="500" name="Cutoff" scale="1" value="3000.18"/>
<parameter free="0" max="5" min="0" name="Index2" scale="1" value="1"/>
</spectrum>
<spatialModel type="SkyDirFunction">
<parameter free="0" max="360." min="-360." name="RA" scale="1.0" value="267.02"/>
<parameter free="0" max="90." min="-90." name="DEC" scale="1.0" value="-24.7792"/>
</spatialModel>
</source>
```

Practical Procedures with Tools

4. Defining Source Model (xml file)

4c. Other sources in ROI

```
</source>
<source name="SRC 7" type="PointSource">
<spectrum type="PowerLaw">
<parameter free="1" max="1000.0" min="0.001" name="Prefactor" scale="1e-09"
value="4.22578"/>
<parameter free="1" max="-1.0" min="-5.0" name="Index" scale="1.0" value="-2.05213"/>
<parameter free="0" max="2000.0" min="30.0" name="Scale" scale="1.0" value="100.0"/>
</spectrum>
<spatialModel type="SkyDirFunction">
<parameter free="0" max="360" min="-360" name="RA" scale="1.0" value="271.329"/>
<parameter free="0" max="90" min="-90" name="DEC" scale="1.0" value="-21.649"/>
</spatialModel>
</source>
                              .
                              .
                              .
                              .
                              .

</source_library>
```

Practical Procedures with Tools

5. Likelihood Analysis

Computation time is a concern!

Do a quick BUT rough estimate first!

gtlike plot=yes irfs=P6_V3_DIFFUSE expcube=expcube.fits \
srcmdl=src_model_5deg.xml statistic=**UNBINNED** optimizer=**DRMNFB** \
evfile=ter5_filtered_5deg_gti.fits scfile=L091218021133E0D2F37E85_SC00.fits \
expmap=expmap_10deg.fits

```
Computing TS values for each source (8 total)
........!
EG_v02:
Normalization: 5.39258 +/- 0.157725
Npred: 25412

GAL_v02:
Value: 0.981163 +/- 0.00501332
Npred: 188131

SRC 11:
Prefactor: 29.5296 +/- 0.878905
Index: -2.39692 +/- 0.0129244
Scale: 100
Npred: 16712
ROI distance: 4.25057
TS value: 7785.28
                    .
                    .
                    .
Terzan 5:
Prefactor: 0.125358 +/- 0.0338302
Index1: -1.70573 +/- 0.135371
Scale: 100
Cutoff: 3000.18 +/- 565.976
Index2: 1
Npred: 2267.18
ROI distance: 0
TS value: 568.854

Total number of observed counts: 252263
Total number of model events: 252341
```

Practical Procedures with Tools

5. Likelihood Analysis

Computation time is a concern!

Do a quick BUT rough estimate first!

gtlike plot=yes irfs=P6_V3_DIFFUSE expcube=expcube.fits \
srcmdl=src_model_5deg.xml statistic=UNBINNED optimizer=**DRMNFB** \
evfile=ter5_filtered_5deg_gti.fits scfile=L091218021133E0D2F37E85_SC00.fits \
expmap=expmap_10deg.fits

Then go for a more detailed analysis

gtlike plot=yes irfs=P6_V3_DIFFUSE expcube=expcube.fits \
srcmdl=src_model_5deg.xml statistic=UNBINNED optimizer=**NEWMINUIT** \
evfile=ter5_filtered_5deg_gti.fits scfile=L091218021133E0D2F37E85_SC00.fits \
expmap=expmap_10deg.fits

Practical Procedures with Tools

6. Visual Inspection

i) Bin the data into spectrum

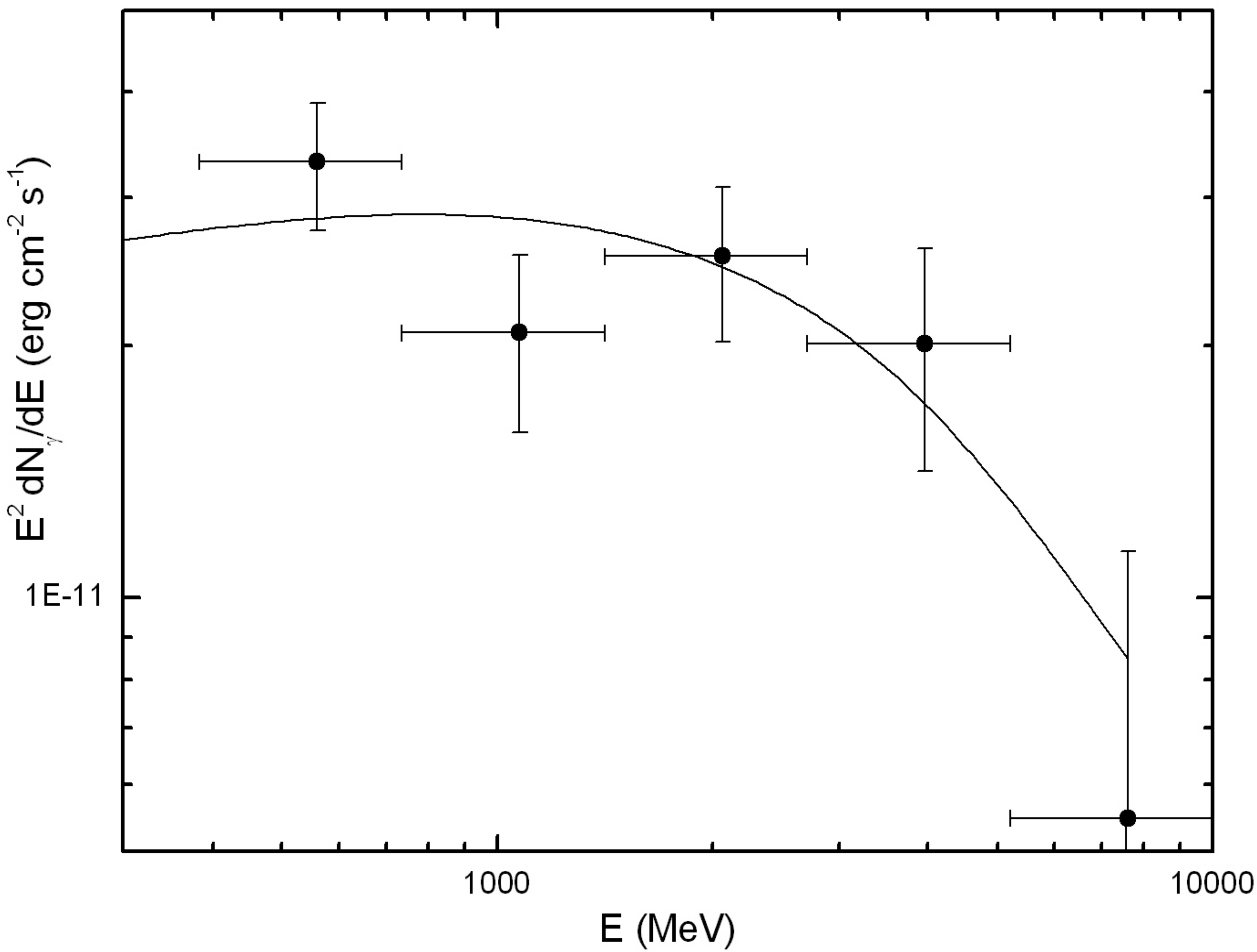
gtbin evfile=ter5_filtered_1deg_gti.fits \
scfile=NONE **outfile=ter5_1deg_spec_bin5.fits** \
algorithm=PHA1 ebinalg=**LOG** emin=200 emax=10000\
enumbins=5

ii) Generate response matrix

gtrspgen respalg=PS specfile=ter5_1deg_spec_bin5.fits
scfile=L100104092057E0D2F37E74_SC00.fits \
outfile=ter5_1deg_spec_bin5.rsp thetacut=60 dcostheta=0.05 irfs=P6_V3_DIFFUSE \
ebinalg=LOG emin=50 emax=20000 enumbins=100 \

iii) Calculate the background spectrum

gtbkg phafile=ter5_1deg_spec_bin5.fits **outfile=bg_spec_bin5.fits** \
scfile=L100104092057E0D2F37E74_SC00.fits expcube=expcube.fits \
expmap=expmap_5deg.fits irfs=P6_V3_DIFFUSE **srcmdl=outputcutoff0.5-20.xml** \
target=Terzan5 \



E^2 dN$_\gamma$/dE (erg cm^{-2} s^{-1})
1E-11
1000
10000
E (MeV)