Likelihood analysis general tutorial

(including source specifications)

G. Tosti
INFN - Perugia

REFERENCES

DC2 code releases:
https://confluence.slac.stanford.edu/display/DC2/DC2+code+releases

GLAST Data Access:
http://glast.gsfc.nasa.gov/ssc/dev/databases/DC2/

GLAST Science Tools Workbook:
http://glast-ground.slac.stanford.edu/workbook/

Binned Likelihood Tutorial:
http://glast-ground.slac.stanford.edu/workbook/science-tools/pages/sciTools_binnedLikelihoodTutorial/binnedLikelihood.htm

Unbinned Likelihood Tutorial:
http://glast-ground.slac.stanford.edu/workbook/science-tools/pages/sciTools_likelihoodTutorial/likelihoodTutorial.htm

LAT Likelihood Algorithm description (Jim's presentation):
http://glast.gsfc.nasa.gov/ssc/resources/guc/040809/Likelihood_jc.ppt

Likelihood Analysis: Introduction

The final aim of any data analysis work is to derive the best possible estimate for the caracteristics of a source.

The Maximum Likelihood Analysis (MLA) has been successfully used in the analysis of gamma-ray data and it has also a central role in the LAT Data analysis.

The GLAST Science Analysis Software provides a tool to perform:

- **Unbinned** **Maximum Likelihood Analysis** (this talk)
- **Binned** **Maximum Likelihood Analysis**

Likelihood Analysis: Introduction

The Maximum Likelihood is used to compare source measured counts with the predicted counts derived from a source model.
In the Unbinned version of MLA the source Model considered is:

$$S(E,\hat{p},t) = \sum_i s_i(E,t)\delta(\hat{p}-\hat{p}_i) + S_G(E,\hat{p}) + S_{\rm eg}(E,\hat{p}) + \sum_l S_l(E,\hat{p},t),$$

Point Sources | Galactic & EG Diffuse Sources | Other Sources

This model is folded with the Instrument Response Funcion (IRF) to obtain the predicted counts in the measured quantities space (E',p',t'):

$$M(E',\hat{p}',t) = \int_{\rm SR} dE d\hat{p} R(E',\hat{p}',t;E,\hat{p}) S(E,\hat{p},t)$$

where:

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

is the total IRF and the integral is done over the Source Region (SR), that is the region of the sky containing all the sources that can contribute to the Region of Interest (ROI). In the standard Analysis only "steady" sources are considered:

$$S(E,\hat{p},t) \rightarrow S(E,\hat{p})$$

Likelihood Analysis: Introduction

The function that is maximized is

$$\log \mathcal{L} = \sum_j \log M(E'_j, \hat{p}'_j, t_j) - N_{\rm pred}$$

Where the sum is performed over the ROI. The predicted counts are given by:

$$N_{\rm pred} = \int_{\rm ROI} dE' d\hat{p}'\, dt M(E', \hat{p}', t)$$

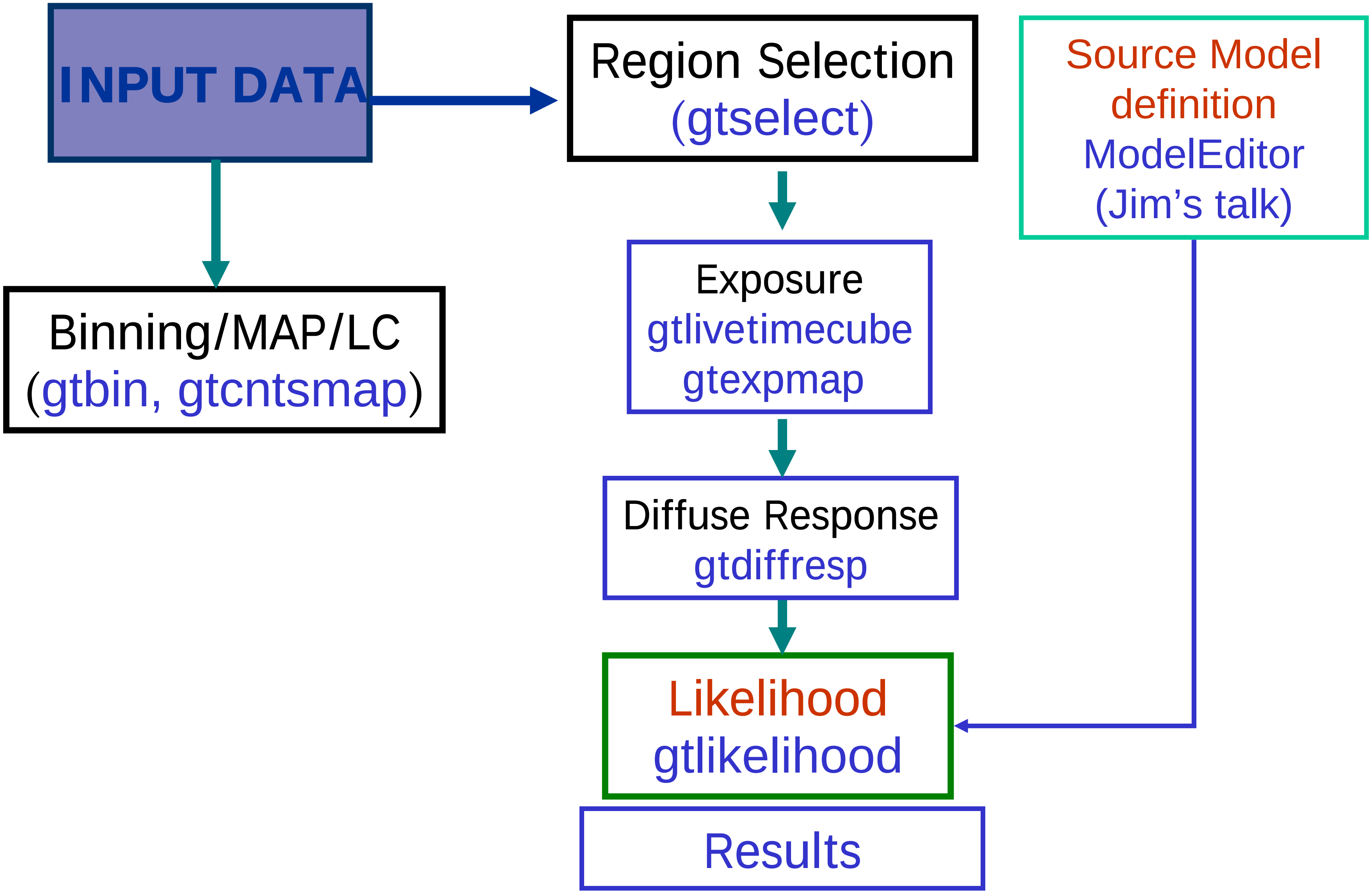
To reduce the CPU time, a quantity, indipendent from any source model and similar to an exposure map, is precomputed:

$$\varepsilon(E, \hat{p}) \equiv \int_{\rm ROI} dE'\, d\hat{p}'\, dt\, R(E', \hat{p}', t; E, \hat{p})$$

Then

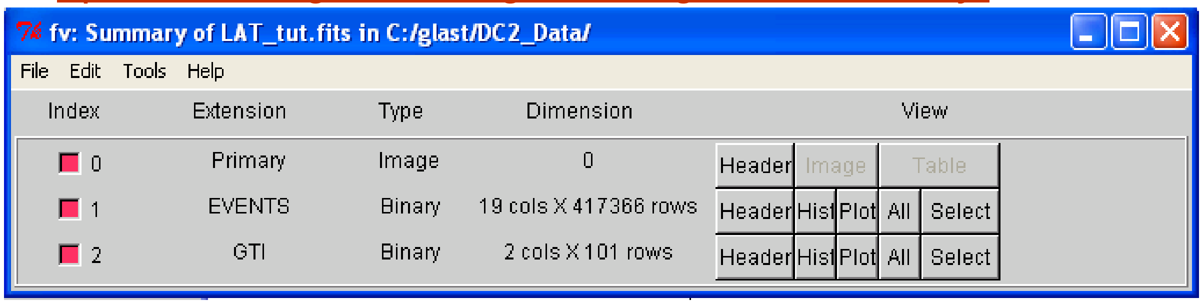
$$N_{\rm pred} = \int_{\rm SR} dE d\hat{p}\, S(E, \hat{p})\, \varepsilon(E, \hat{p})$$

The Unbinned Likelihood Analysis Diagram

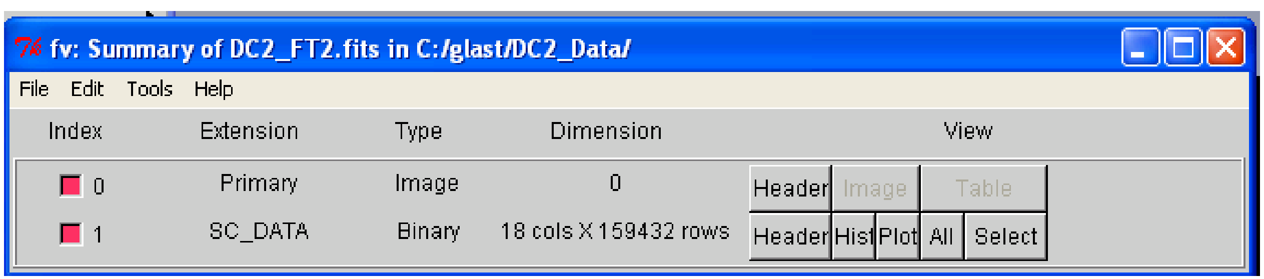


INPUT DATA

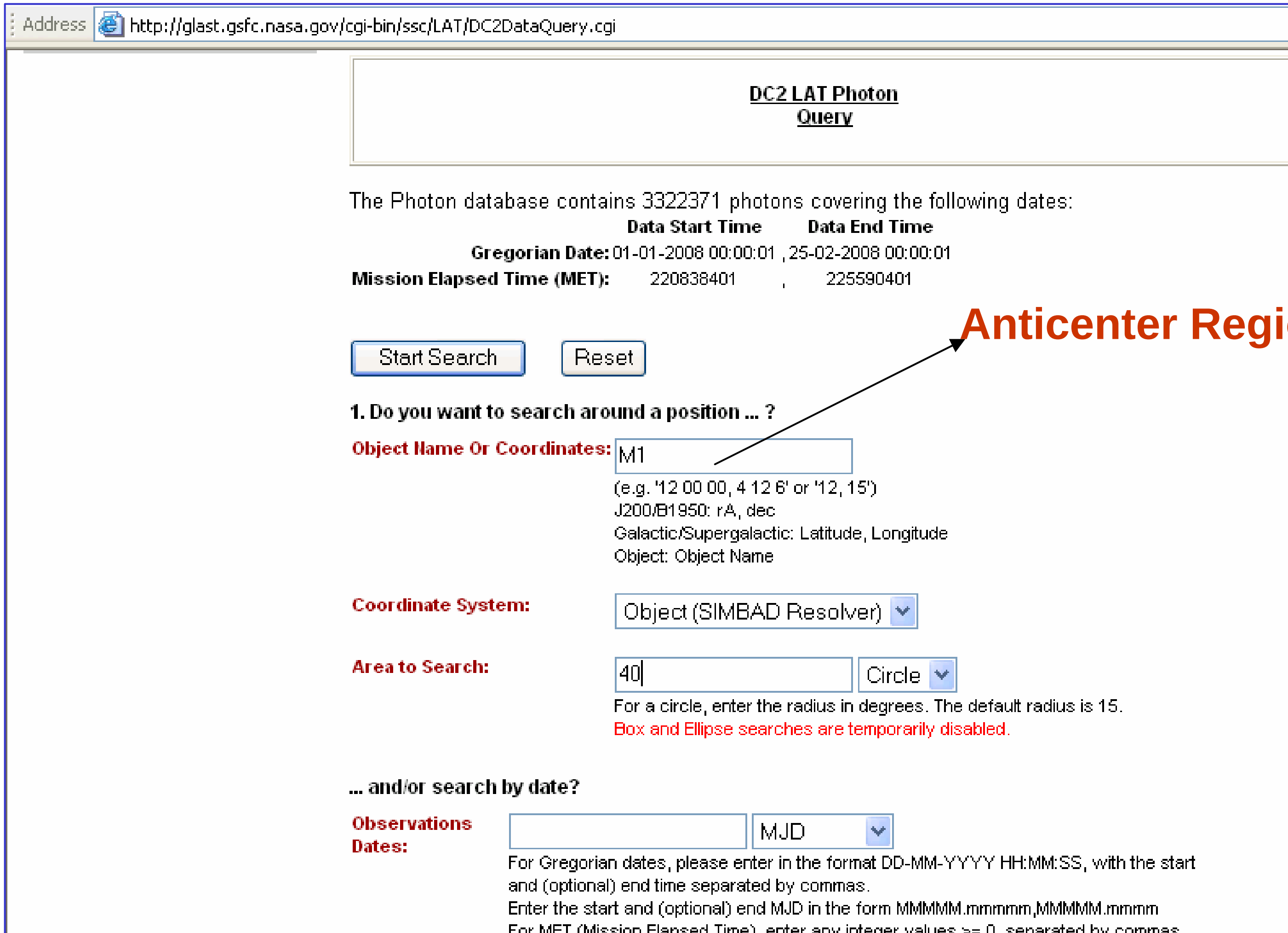
- The **photon FT1 fits** FT1 file:
  - LAT_allsky_220838401.126_V01.fits (1 week of data)
  - ftp://heasarc.gsfc.nasa.gov/FTP/glast/DC2/allsky/



- and the **pointing and livetime history FT2** files.
  - **FT2 file** :DC2_FT2.fits
  - ftp://heasarc.gsfc.nasa.gov/FTP/glast/DC2/misc_data/DC2_FT2.fits



INPUT DATA


Address http://glast.gsfc.nasa.gov/cgi-bin/ssc/LAT/DC2DataQuery.cgi

DC2 LAT Photon Query

The Photon database contains 3322371 photons covering the following dates:

	Data Start Time		Data End Time
Gregorian Date:	01-01-2008 00:00:01	,	25-02-2008 00:00:01
Mission Elapsed Time (MET):	220838401	,	225590401

Start Search Reset

1. Do you want to search around a position ... ?

Object Name Or Coordinates: M1
(e.g. '12 00 00, 4 12 6' or '12, 15')
J200/B1950: rA, dec
Galactic/Supergalactic: Latitude, Longitude
Object: Object Name

Coordinate System: Object (SIMBAD Resolver)

Area to Search: 40 Circle
For a circle, enter the radius in degrees. The default radius is 15.
Box and Ellipse searches are temporarily disabled.

... and/or search by date?

Observations Dates: MJD
For Gregorian dates, please enter in the format DD-MM-YYYY HH:MM:SS, with the start and (optional) end time separated by commas.
Enter the start and (optional) end MJD in the form MMMMM.mmmmm,MMMMM.mmmm
For MET (Mission Elapsed Time), enter any integer values >= 0, separated by commas.

Anticenter Region - Crab

http://glast.gsfc.nasa.gov/cgi-bin/ssc/LAT/DC2DataQuery.cgi

Step 1-Event Selection

$$$> gtselect ←gtselect.par

Input FT1 file [test_events_0000.fits] : LAT_tut.fits
Output FT1 file [filtered_events_0000.fits] : LAT_tut_filtered.fits
RA for new search center (degrees) <0 - 360> [86.4] : 86.57
Dec for new search center (degrees) <-90 - 90> [28.9] : 22.01
radius of new search region (degrees) <0 - 180> [20] :
start time (MET in s) [0] : 220838401
end time (MET in s) [0] : 221616000
lower energy limit (MeV) [30] :
upper energy limit (MeV) [200000] :
Event classes (-1=all, 0=FrontA, 1=BackA, 2=FrontB, 3=BackB, 4=class A) <-1 – 4 [-1] :

Scriptable form of the command:
gtselect infile="LAT_tut.fits" outfile="LAT_tut_filtered.fits" ra=86.57 dec=22.01 rad=20 coordSys="CEL" tmin=220838401 tmax=221616000 emin=30 emax=200000 eventClass=-1

Class A: highest quality→Best PSF & Lowest background contamination.
Class B: More effective area, but more background

Step 2: Counts Map

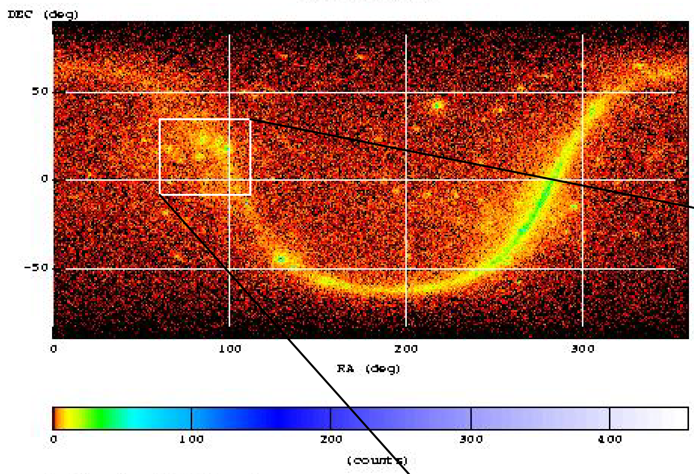
$$$> gtcntsmap ← gtcntsmap.par

Event data file [test_events_0000.fits] : LAT_tut_filtered.fits
Spacecraft data file [test_scData_0000.fits] : DC2_FT2.fits
Output file name [countsMap.fits] : LAT_tut_countsMap.fits
Minimum energy (MeV) <20 - 3e5> [30] :
Maximum energy (MeV) <20 - 3e5> [200000] :
Number of energy bounds <2 - 100> [21] : 2
Right Ascension of map center (degrees) <0 - 360> [86.4] : 86.57
Declination of map center (degrees) <-90 - 90> [28.9] : 22.01
Number of longitude pixels [160] :
Number of latitude pixels [160] :
Pixel size (degrees) <1e-2 - 2> [0.25] :
Projection method <AIT|ARC|CAR|GLS|MER|NCP|SIN|STG|TAN> [STG] : AIT

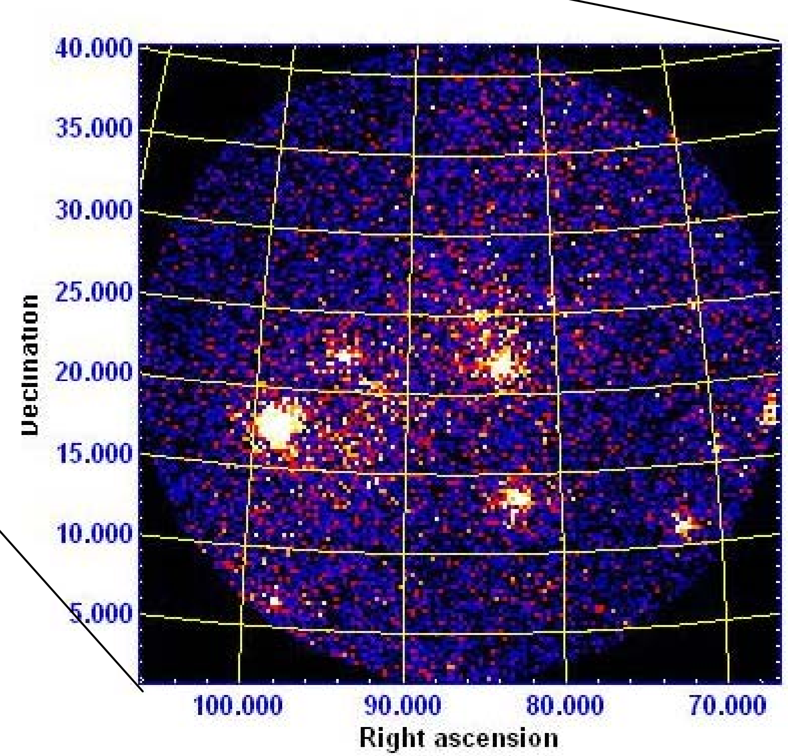
Scriptable form of the command:

gtcntsmap evfile="ac_dc2_filtered.fits" scfile="DC2_FT2.fits" outfile="ac_dc2_cmap.fits" emin=30 emax=2000 nenergies=10 ra=86.4 dec=28.9 nra=160 ndec=160 pixel_size=0.25 proj="AIT"

Step 2-Counts Map



LAT_allsky_220838401.126_V01.fits



Step 3:The Source Model

EGRET Sources in the Crab field

Result of VizieR Search within **20°** of **M1** (J2000=05:34:32.0+22:00:52) (no other constraint specified)
ordered by increasing **_r**

Max. Entries: 200 | Output layout: ascii table | **ALL** columns | Modify the Quer | ReSubmit

J/ApJS/123/79/3eg — Third EGRET catalog (3EG) (Hartman+, 1999) (ReadMe)
Third EGRET Source Catalog (table 4)

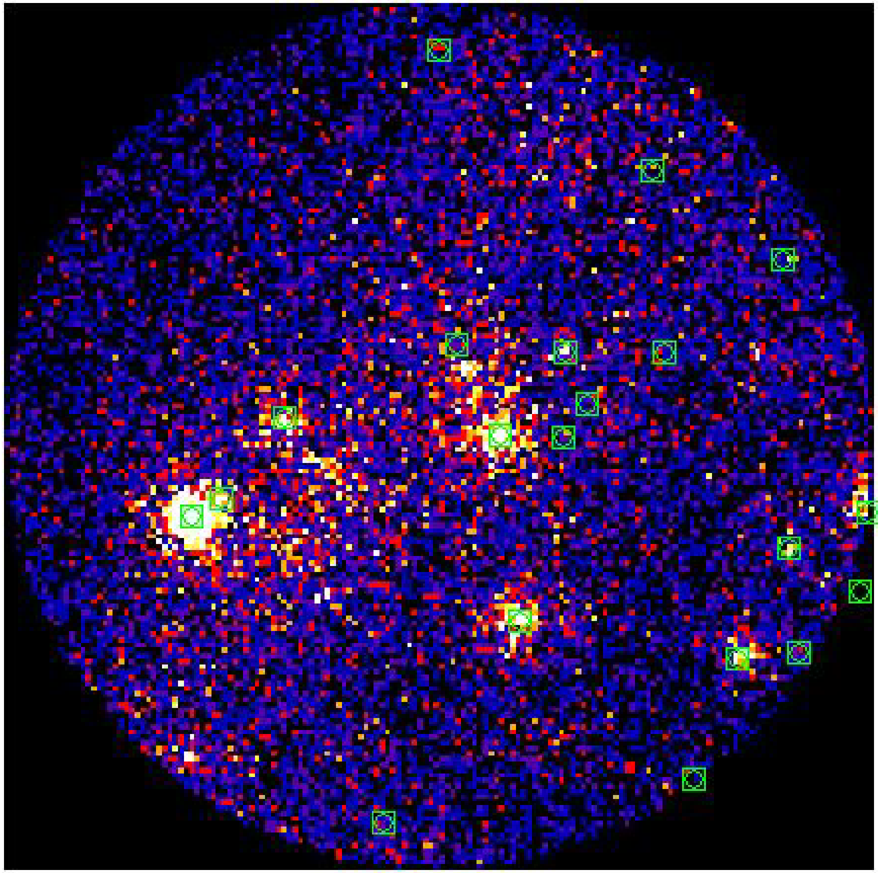
To get all details for a row, just click on the row number in the leftmost `Full' column.

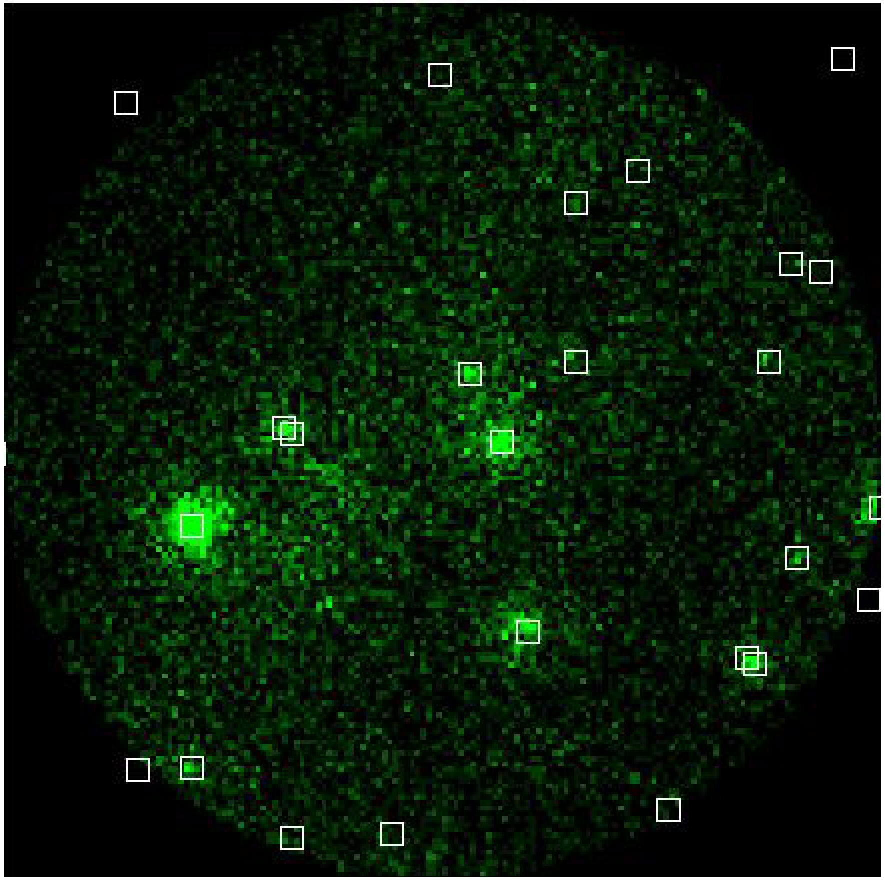
Full	3EG	RAJ2000 deg	DEJ2000 deg	F 10-8ph/cm2/s	Sp-Index	OtherNames
1	J0534+2200	083.57	+22.01	226.2	2.19	2EG J0534+2158, GEV J0534+2159, PSR B0531+21, Crab pulsar
2	J0521+2147	080.38	+21.80	20.7	2.48	2EG J0521+2206
3	J0516+2320	079.11	+23.35	168.9	2.67	91/06/11 Sol.Fl.
4	J0542+2610	085.69	+26.17	19.1	2.67	GRO J0542+26, S147 SNR?
5	J0520+2556	080.14	+25.75	15.7	2.83	2EG J0520+2626
6	J0500+2529	075.07	+25.49	11.3	2.52	
7	J0530+1323	082.74	+13.38	93.5	2.46	2EG J0531+1324, GEV J0530+1340, 0528+134
8	J0617+2238	094.30	+22.63	51.4	2.01	2EG J0618+2234,GEV J0617+2237, IC 443 SNR?
9	J0628+1847	097.18	+18.79	23.9	2.30	
10	J0459+3352	074.78	+33.87	18.6	2.54	2EG J0506+3424
11	J0439+1555	069.81	+15.93	42.9	2.27	2EG J0437+1524
12	J0633+1751	098.49	+17.86	352.9		2EG J0633+1745, GEV J0634+1746, PSR J0633+1746, Geminga pulsar
13	J0450+1105	072.61	+11.09	109.5	2.27	0446+112, 2EG J0450+1122
14	J0433+2908	068.40	+29.14	22.0	1.90	2EG J0432+2910, GEV J0433+2907, 0430+2859
15	J0439+1105	069.81	+11.09	9.4	2.44	
16	J0423+1707	065.92	+17.13	15.8	2.43	2EGS J0426+1636
17	J0546+3948	086.55	+39.81	13.7	2.85	2EG J0545+3943
18	J0426+1333	066.67	+13.56	14.0	2.17	2EG J0422+1414
19	J0459+0544	074.93	+05.75	12.1	2.36	0459+060
20	J0556+0409	089.06	+04.15	16.9	2.45	2EGS J0555+0408

http://vizier.u-strasbg.fr/

Step 3:The Source Model

The 3EG & Cat_v0 sources in the field





DC2 - Source Catalog

ftp://heasarc.gsfc.nasa.gov/FTP/glast/DC2/misc_data/LATSourceCatalog_v0.fits

Step 3:The Source Model

- **Source models should be written in an XML file:**

Source Model description

```
<?xml version="1.0" ?>
<!DOCTYPE source_library
 SYSTEM '$(LIKELIHOODROOT)/xml/A1_Sources.dtd'>
<source_library title="source library">

   <source name="Extragalactic Diffuse" type="DiffuseSource">
    <spectrum type="PowerLaw">
     <parameter free="1" max="100" min="1e-05" name="Prefactor" scale="1e-07" value="1.32"/>
     <parameter free="0" max="-1" min="-3.5" name="Index" scale="1" value="-2.1"/>
     <parameter free="0" max="200" min="50" name="Scale" scale="1" value="100"/>
    </spectrum>
    <spatialModel type="ConstantValue">
     <parameter free="0" max="10" min="0" name="Value" scale="1" value="1"/>
    </spatialModel>
   </source>

<source name="PKS 0528+134" type="PointSource">
    <spectrum type="PowerLaw">
     <parameter free="1" max="1000.0" min="0.001" name="Prefactor" scale="1e-009" value="13.65"/>
     <parameter free="1" max="-1.0" min="-3.5" name="Index" scale="1.0" value="-2.46"/>
     <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="3.40282e+038" min="-3.40282e+038" name="RA" scale="1.0" value="82.74"/>
     <parameter free="0" max="3.40282e+038" min="-3.40282e+038" name="DEC" scale="1.0" value="13.38"/>
    </spatialModel>
   </source>
```

Step 4-Livetimes

$$$> gtlivetimecube ← gtlivetimecube.par

Event data file [test_events_0000.fits] : LAT_tut_filtered.fits
Spacecraft data file [test_scData_0000.fits] : DC2_FT2.fits
Output file [expCube.fits] : LAT_tut_expCube.fits
Step size in cos(theta) <0. - 1.> [0.025] : 0.05
Pixel size (degrees) [1] :
Working on file DC2_FT2.fits
....!

More info on the Sky pixelization used by gtlivetimecube can be found here HEALPix (http://healpix.jpl.nasa.gov/)

To add two livetimeCube you can use:

$$$>gtaddlivetime

Livetime cube 1 [expCube_00.fits] : expCube0.fits
Livetime cube 2 [expCube_01.fits] : expCube1.fits
Output file [expCube.fits] : expcube_01.fits

Scriptable form of the command:
gtlivetimecube evfile="ac_dc2_filtered.fits" scfile="DC2_FT2.fits" outfile="myexpCube.fits" cos_theta_step=0.05 pixel_size=1

This step is not necessary for the DC2, the expcube files are already avalilabe at: http://glast.gsfc.nasa.gov/ssc/dev/databases/DC2/livetimecubes.html

Step 5-Exposure Map

$$$> gtexpmap ← gtexpmap.par

Event data file [test_events_0000.fits] : LAT_tut_filtered.fits
Spacecraft data file [test_scData_0000.fits] : DC2_FT2.fits
Exposure hypercube file [expCube.fits] : LAT_tut_expCube.fits
output file name [expMap.fits] :
Response functions <DC2|DC1A|G25> [DC2] :
Radius of the source region (in degrees) [30] :
Number of longitude points <2 - 1000> [120] :
Number of latitude points <2 - 1000> [120] :
Number of energies <2 - 100> [20] :
Computing the ExposureMap.

Scriptable form of the command:

gtexpmap evfile="ac_dc2_filtered.fits" scfile="DC2_FT2.fits" exposure_cube_file="myexpCube.fits" outfile="exp_map.fits" rspfunc="DC2" source_region_radius=30 number_of_longitude_points=120 number_of_latitude_points=120 number_of_energies=20

Step 6-Diffuse Source Responses

$$$> gtdiffresp ← gtdiffresp.par

Event data file [test_events_0000.fits] : LAT_tut_filterd.fits
Event data file [test_events_0000.fits] : LAT_tut_filtered.fits
Spacecraft data file [test_scData_0000.fits] : DC2_FT2.fits
Source model file [my_source_model.xml] : ac_source_model.xml
Response functions to use <DC2|DC1A|G25> [DC2] :

```
Computing exposure at (82.74, 13.38)....................!
Computing exposure at (83.57, 22.01)....................!
Computing exposure at (98.49, 17.86)....................!
adding source Crab Pulsar
adding source Extragalactic Diffuse
adding source Galactic Diffuse
adding source Geminga
adding source PKS 0528+134
Working on...
```

Scriptable form of the command:

```
gtdiffresp  evfile="ac_dc2_filtered.fits" scfile="DC2_FT2.fits"
source_model_file="ac_source_model.xml"  rspfunc="DC2"
```

This step is not necessary for the DC2, the FT1 file already has pre-computed diffuse response columns!!!

Step 7: Unbinned Likelihood

$$$> gtlikelihood ← gtlikelihood.par

Statistic to use <BINNED|UNBINNED> [UNBINNED] :
Spacecraft file [] : DC2_FT2.fits
Event file [] : LAT_tut_filtered.fits
Unbinned exposure map [LAT_tut_expMap.fits] : LAT_tut_expMap.fits
Exposure hypercube file [LAT_tut_expCube.fits] : LAT_tut_expCube.fits
Source model file [ac_source_model.xml] : ac_source_model.xml
Response functions to use [DC2] :
Optimizer <LBFGS|MINUIT|DRMNGB> [DRMNGB] :

The results are stored in the file: result.dat

Scriptable form of the command:

```
gtlikelihood statistic="UNBINNED" scfile="DC2_FT2.fits" evfile="ac_dc2_filtered.fits"
exposure_map_file="none"  exposure_cube_file="myexpCube.fits"
source_model_file="ac_source_model.xml"
rspfunc="DC2" optimizer="MINUIT" flux_style_model_file="my_flux_out.xml"
```

Step 7: Unbinned Likelihood
Results

DRMNGB

Crab Pulsar:
Prefactor: 28.6678 +/- 1.08105
Index: -2.24443 +/- 0.0365408
Scale: 100
Npred: 1227.55
ROI distance: 2.78104
TS value: 1631.35

Extragalactic Diffuse:
Prefactor: 9.7619 +/- 0.150878
Index: -2.1
Scale: 100
Npred: 18199.7

Galactic Diffuse:
Prefactor: 0.001 +/- 1.45158
Index: -2.1
Scale: 100
Npred: 0.0976448

Geminga:
Prefactor: 33.4275 +/- 1.03978
Index: -1.89 +/- 0.0205917
Scale: 100
Npred: 2060.16
ROI distance: 11.9441
TS value: 7254.42

PKS 0528+134:
Prefactor: 15.2024 +/- 0.0588233
Index: -2.57212 +/- 3.16165
Scale: 100
Npred: 521.55
ROI distance: 9.35745
TS value: 454.352

Total number of observed counts: 22047
Total number of model events: 22009

-log(Likelihood): 226339.8276

MINUIT

Crab Pulsar:
Prefactor: 31.7406 +/- 1.83759
Index: -2.3361 +/- 0.0366084
Scale: 100
Npred: 1266.67
ROI distance: 2.78104
TS value: 1659.21

Extragalactic Diffuse:
Prefactor: 9.60841 +/- 0.078923
Index: -2.1
Scale: 100
Npred: 17913.5

Galactic Diffuse:
Prefactor: 0.00122981 +/- 0.00228569
Index: -2.1
Scale: 100
Npred: 0.120084

Geminga:
Prefactor: 37.6889 +/- 1.55683
Index: -1.91693 +/- 0.0195631
Scale: 100
Npred: 2239
ROI distance: 11.9441
TS value: 7282.72

PKS 0528+134:
Prefactor: 15.6705 +/- 1.43737
Index: -2.51653 +/- 0.0638708
Scale: 100
Npred: 553.776
ROI distance: 9.35745
TS value: 461.376

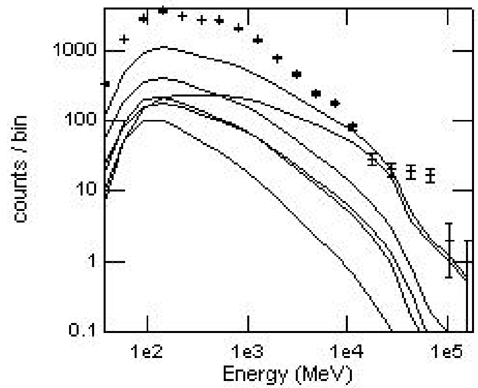
Total number of observed counts: 22047
Total number of model events: 21973.1

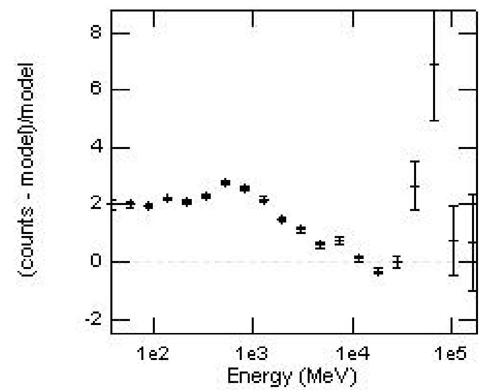
-log(Likelihood): 226331.4097

Writing flux-style xml model file to ac_out_flux.xml

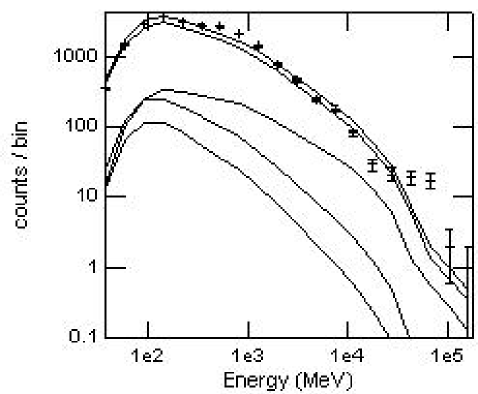
Results

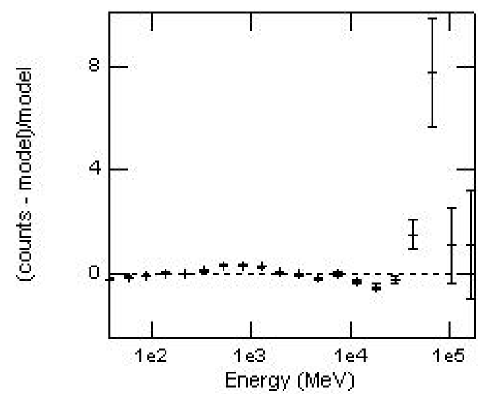
Before the fit





After the fit





(see Jim’s talk on the interactive use of the Likelihood)

Results

The flux-style xml file can be then used as an input file for gtobssim or Gleam

Address C:\glast\DC2_Data\ac_out_flux.xml


```
- <source_library title="Likelihood_model">
  - <source flux="0.118835" name="Crab_Pulsar">
    - <spectrum escale="MeV">
      - <particle name="gamma">
          <power_law emax="200000" emin="30" gamma="2.3361" />
        </particle>
        <celestial_dir dec="22.01" ra="83.57" />
      </spectrum>
    </source>
  + <source flux="3.28745" name="Extragalactic_Diffuse">
  + <source flux="0.124046" name="Geminga">
  - <source flux="0.0642419" name="PKS_0528+134">
    - <spectrum escale="MeV">
      - <particle name="gamma">
          <power_law emax="200000" emin="30" gamma="2.51653" />
        </particle>
        <celestial_dir dec="13.38" ra="82.74" />
      </spectrum>
    </source>
  + <source name="all_in_ac_out_flux.xml">
  </source_library>
```

Conclusion (1)

In the analysis of the DC2 data, taking into account that:

- One can select the photons collected ROI from the DATA Server (FT1 file);
- The livetimecube files were preconputed;
- The Diffuse response was precomputed, and the supplementary columns included in the FT1 file you download from the server

the only steps you need to perform a likelihood analysis are:

- Build a Source Model for your ROI
- Calculate the ExposureMap (gtexpmap)
- Run gtlikelihood

Conclusion (2)

To analyze the time & spectral variations of a source you have to run iteratively the likelihood tool, using procedures similar to the following

- Step 1 -Download data for your ROI from the DATA server over the entire time interval (T) where you want to study variability;
- Step 2 -Build a Source Model for your ROI
- Step 3 -Divede T in N time bins, and for each bin obatein a FT1 file with gtselect;
- Step 4 -Compute the livetimecube (gtlivetimecube) for each time
- Step 5 -Calculate the ExposureMap (gtexpmap)
- Step 6 - Run gtlikelihood and rename the output result file

Repeat steps 3-6 for each time bin……write a script to do this work……

THAT'S ALL……

MORE ABOUT LIKELIHOOD IN THE NEXT TALKS…….