Fermi

Gamma-ray Space Telescope

Analysis Workshop
14 March 2010

Advanced Likelihood

E. Hays

Goals

- **Quality checks on spectral fitting of point sources**
  - **Major gotchas**
  - **Reminder of simple checks**
  - **Upper Limits**
- **Binned vs. Unbinned likelihood**

Major gotchas

- **Flux/Spectral analysis depends critically on calculating the proper exposure**

selection		livetime	response/exposure	minimization
gtselect	**gtmktime**	**gtlcube**	**gtexpmap**	**gtlike**
(gtbin)			**(gtsrcmaps)**	**(gtlike binned)**

- **Examples of things that can screw this up**
  - **fselect, fcopy**
    - **these do not update the header keywords used in the exposure calculation**
  - **Mismatch of data selection and IRF set**
    - **Use the diffuse class IRFs with the diffuse class event selection**
  - **Mismatch of ROI selection (*gtselect*) and data cube (*gtbin*) in binned likelihood analysis**

Major Gotchas II

- **Mismatch of calculated diffuse response and model diffuse components**
  - **Use the recommended diffuse models with the data (includes precalculated diffuse response values for each photon for those specific models)**
  - **Diffuse response for experts**
    - ***gtdiffrsp* calculates the diffuse response values**
    - **Use unique names in the input xml model for different diffuse model templates**
    - **Example: If you come up with a new version of the Galactic diffuse template, don't call it "GAL_v02"**
- **The currently recommended isotropic template is only appropriate for use with the diffuse event class**

Likelihood output - simple checks

Did the fit work and does it make sense?
Reading the tea leaves of gtlike output

- Did the minimization converge?
- Are the number of predicted photons reasonable?
- Do the parameter values make sense?
  - Are values hitting limits?
  - Is there a source with an extremely soft spectrum or hard spectrum?
- Do the parameter errors make sense?
  - Too small? Were enough parameters left free?
  - Larger than the parameter values? Is the source significant?
- Consider the above for the target source and field sources
- All of the above become more critical for faint sources, complex regions, time-binned flux light curves…

Fermi Gamma-ray Space Telescope

Gtlike bits

Convergence

```
Minuit did successfully converge.
# of function calls: 2401
minimum function Value: 2808753.9585
minimum edm: 0.74929079
```

Failed Convergence

```
WARNING: FunctionMinimum is invalid.

Exception encountered while minimizing objective function:
Minuit abnormal termination.  No convergence?
......................................!
Computing TS values for each source (40 total)
```

Example Source Results

Source A: 2 free parameters

```
Integral: 0.000706819 +/- 0.017999
Index: -4.87644 +/- 1.0237
LowerLimit: 100
UpperLimit: 100000
Npred: 0.116799
ROI distance: 10.1342
TS value: -0.00617604
```

Source B: 2 free parameters

```
Integral: 0.479765 +/- 0.156542
Index: -2.37132 +/- 0.12927
LowerLimit: 100
UpperLimit: 100000
Npred: 458.434
ROI distance: 3.26662
TS value: 25.9222
```

Source C: 0 free parameters

```
Integral: 6.28448
Index: -2.33404
LowerLimit: 100
UpperLimit: 100000
Npred: 315.177
```

Galactic Diffuse: 2 free parameters

```
Prefactor: 0.984611 +/- 0.00884279
Index: 0.0296784 +/- 0.00345898
Scale: 100
Npred: 175955
```

Crude Fit Quality Info

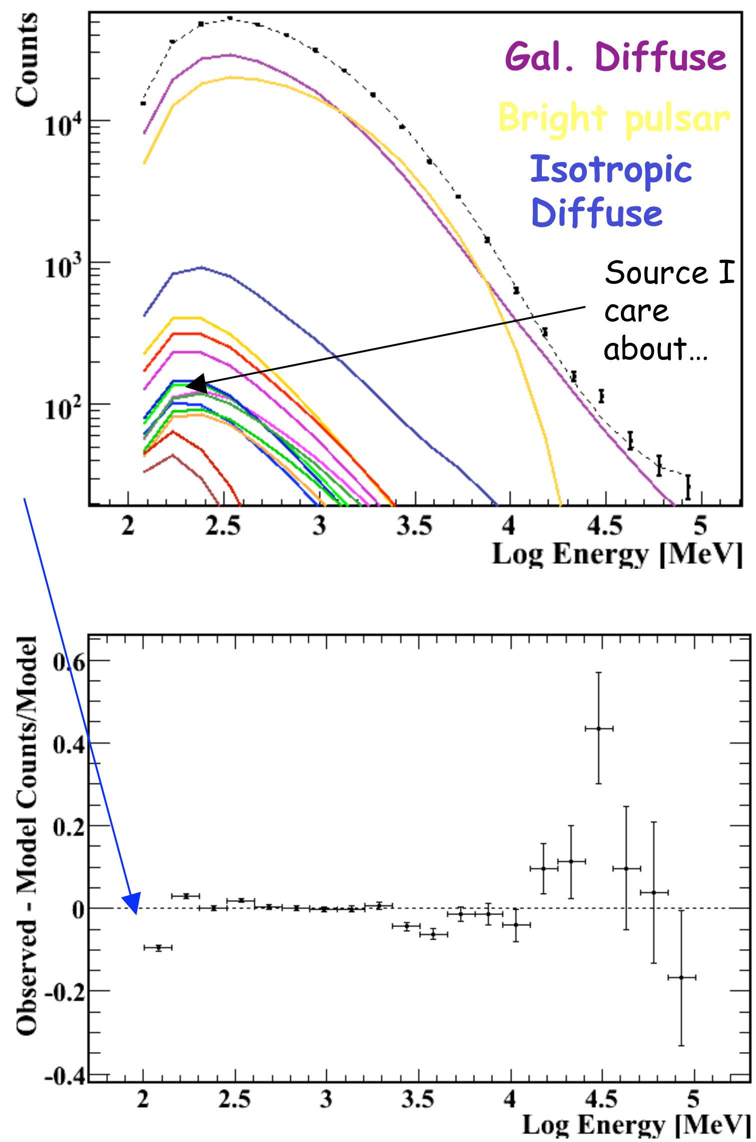
```
WARNING: Fit may be bad in range [100, 199.526] (MeV)
WARNING: Fit may be bad in range [281.838, 398.107] (MeV)
WARNING: Fit may be bad in range [2238.72, 4466.84] (MeV)
WARNING: Fit may be bad in range [25118.9, 35481.3] (MeV)

Total number of observed counts: 325124
Total number of model events: 325093

-log(Likelihood): 2808753.988
```

Spectral Residuals

- **Unbinned analysis produces predicted counts and residuals as a function of energy.**
  - Example: a long integration near the Galactic plane and a bright pulsar
- **Discrepancy at low energy common**
  - **Likelihood uses true energy**
- **Discrepancies strongly tied to diffuse model for most analysis**
  - **Diffuse mediates cross talk between target source and nearby neighbors**
  - **Consider relative source strength**
  - **Test impact of model choices and selections on target source**



Likelihood - ROI selection

How big?

- Big enough to constrain model components - source of interest, diffuse emission, nearby sources
- Small enough to avoid significant zenith cut loss to livetime
  - Practical advantage! less photons and less sources => less calculations for unbinned analysis
  - Analysis disadvantage! likelihood is an inclusive modeling strategy
- Recommendations
  - ~10 deg for isolated point source (E>100 MeV)
  - Larger regions (15-20 deg) benefit confused sources, aid in separating diffuse at low energy, improve error estimates
- Test it!
  - Are fit results reliable for different ROI radii?
  - What is the impact on GTIs?

Likelihood Model - sources

What should be included?

- **All sources that contribute photons to the selected region**
  - **Bright source list sources within ~10 deg of the ROI boundary - accommodates tail of low energy PSF**
  - **Same goes for 1FGL catalog sources once available**
- **Galactic diffuse model**
- **Isotropic diffuse model**
  - **Important for all parts of the sky…provides a home for residual instrument effects (cosmic rays)**

This is a starting point. Adapt to find what works best for your region and source.

Likelihood Model - spectra

What spectral shape?

- **Power laws are simple and well defined**
  - **For faint sources, difficult to justify more parameters**
- **BUT lots of LAT sources are not simple power laws… some tips to help motivate other spectral forms**
  - **Bright pulsars?**
    - **Try simple exponentially cutoff power laws to improve fits for the pulsar itself *and for nearby sources***
  - **Visually inspect energy-dependent ROI selections**
  - **Do power-law fit parameters vary significantly for different minimum energy selections or fits in separate energy bins?**
- **Confirm: Most accurate and unbiased way to determine spectral parameters and errors is by testing the hypothesis using the likelihood analysis**

Likelihood - reality checks

Is anything missing?

- **Visual inspection of count maps and residuals**
- **Test Statistic maps (for unbinned analysis)**
  - ***gttsmap*** **- Tests hypothesis of additional point source over a spatial grid**
  - **Very Calculation Intensive**
    - **try small regions (5 deg) and large grid spacing (0.5 deg)**
  - **Discrepancies may be additional source or component, or could be deficiencies in the diffuse model in some regions**
  - **Warning: *gttsmap* is not ideal for localization, use *gtfindsrc***
- **Predicted and residual count maps (for binned analysis)**
  - **Profiles, radial density, energy dependence**

Likelihood - checking results

Is the result consistent for a different analysis?

- Iteration
  - Consistent results for the best fit parameters?
    - Tip: *gtlike sfile=best_fit_model.xml*
- Data selection tests
  - Minimum energy selection?
  - ROI selection? (Keep in mind this also effects good time selection in combination with zenith cut)
  - Consistency in distinct energy bins (catalog analysis)
  - Agreement using front or back events (requires use of appropriate IRFs, diffuse response, and isotropic model for each)
  - Time selections?
- Fit and Minimization choices
  - Impact of starting parameter values in the model?
  - Fit tolerence? (converging to true minimum?)
  - Effects of optimizer?

Likelihood: Upper Limits I

- **For low Test Statistic (TS <~9), there is a python tool for determining the flux upper limit**
- **Profile method is used**
  – **Scan in flux to find value that gives Delta(Log Likelihood) = 1.35 (2.71/2)**

Upper Limits I

- **Set up the observation - same as for an unbinned analysis**
  - **For this example selecting energy 3000-10000 MeV**

```
>>> from UnbinnedAnalysis import *
>>> from UpperLimits import *
>>> infile='events_15_3000-10000_z105_gti.fits'
>>>scfile='L090823195543E0D2F37E30_SC00.fits'
>>> expmap='expmap.fits'
>>> expcube='ltcube.fits'
>>>obs=UnbinnedObs(infile,scfile,expMap=expmap,
expCube=expcube,irfs='P6_V3_DIFFUSE')
```

Upper Limits II

- **Make sure the xml model matches your selection**
  - **Adjust to fit power law in your energy band - use PowerLaw2 spectral model and adjust upper and lower limits to match your selection**

Model xml file

```
<source name="3C 273" type="PointSource">
<spectrum type="PowerLaw2">
<parameter free="1" max="1000.0" min="0.001" name="Integral"
scale="1e-09" value="10"/>
<parameter free="0" max="-1.0" min="-5.0" name="Index"
scale="1.0" value="-2.5"/>
<parameter free="0" max="300000" min="20" name="LowerLimit"
scale="1" value="3000"/>
<parameter free="0" max="300000" min="20" name="UpperLimit"
scale="1" value="10000"/>
</spectrum>
```

Upper Limits III

- **Run the unbinned likelihood analysis**

```
>>> analysisEBand=UnbinnedAnalysis(obs,'model_3000-
10000.xml',optimizer='MINUIT')

>>> loglikeEBand=analysisEBand.fit()

>>> print loglikeEBand
8151.66172988

>>> analysisEBand.Ts('3C 279')
92.896105960517161

>>> analysisEBand.Ts('3C 273')
24.10538584394817

>>> loglikeEBand.writeXml('fit_3000-10000.xml’)
```

Upper Limits IV

- **Generate the upper limit for selected energy range**

```
>>> ul=UpperLimit(analysisEBand,'3C 273')
>>> results=ul.compute(emin=3000,emax=10000)
0 0.85992510336 -4.87233419335e-05 8.60017840996e-101 1.0238844754
0.0739214054393 1.02399489505e-092 1.18784384745 0.264851439945
1.1879719491e-093 1.35180321949 0.544267079746 1.35194900315e-094
1.51576259153 0.893093107955 1.51592605721e-095 1.70298920163 1.35927962311
1.70317285854e-09(1.7014539204951863e-09, 1.7012704489365027)
>>> print results
1.70e-09 ph/cm^2/s for emin=3000.0, emax=10000.0,delta(logLike)=1.35
>>> print results.parvalues
[0.85992510335991024, 1.0238844754025653, 1.1878438474452204,
1.3518032194878755, 1.5157625915305306, 1.7029892016282362]
>>> print results.dlogLike
[-4.8723341933509801e-05, 0.073921405439250 52, 0.26485143994614191,
0.54426707974380406, 0.89309310795124475, 1.359279623112343]
```

Binned vs. Unbinned Likelihood

- **Unbinned: Treats each photon independently (position, energy)**
  - **Best theoretical performance**
  - **More sensitive - important for faint sources**
  - **Best option for low statistics scenarios (e.g. flux light curves)**
  - **Drawbacks:**
    - **Not for use with spatially extended sources**
    - **Difficult to diagnose problems in individual source fit**
- **Binned: Treats the data in bins of position and energy. Minimal criteria - photons > bins**
  - **Less computationally intensive than unbinned**
  - **Handles templates for extended sources**
  - **Allows nice diagnostics of fit (source maps, spatial profiles, energy dependent comparisons of prediction and model)**
  - **Drawback: At highest energies, can run into low statistics even for long integrations**

Use of both allows consistency check
(if both can be reasonably used)

Summary

- **Lots of ways to use the tools to evaluate spectral fitting and to validate results**
  - **Consistency is key**
  - **Analysis Cookbook provides basic starting points. Cicerone documentation provides deeper insight into into the likelihood technique**
  - **The First Catalog paper provides detailed examples of spectral fitting with the science tools**