

Fix the shared folder

- Close the virtual machine if you have it running, and wait for it to shut down
- Click on Settings
- Go to shared folders, double click on the shared folder and uncheck “automount”
- Click ok
- Run the VM
- Open a terminal
- You should be able to go into your shared folder and create a directory (try it)

Fix an issue with pyfits and the Virtual Box filesystem:

- Issue this command:
> `echo 'setenv PYFITS_USE_MEMMAP 0' >> ~/.tcshrc`
- Reboot the VM by issuing:
> `reboot`



Fermi Gamma-ray Space Telescope



Fermi/LAT data analysis, data products, and software

**Giacomo Vianello
Stanford University
(giacomov@slac.stanford.edu)**



- **Analysis framework and methods**
 - **Covered:**
 - **General concepts**
 - **Maximum Likelihood framework and implementation**
 - **Source detection**
 - **Not covered:**
 - **Instrument Response Functions description**
 - **Upper limits mathematical justification**
- **Software**
 - **Covered:**
 - **Science tools overview**
 - **Likelihood with gtburst**
 - **Non covered:**
 - **Production of LAT spectra/SED (but, for LLE data, you can figure it out easily in gtburst)**
 - **One by one description of science tools**
- **For what is not covered see the FSSC website (fermi.gsfc.nasa.gov/ssc) and the 1FGL catalog for the likelihood framework (available in ADS)**



- Terminology:

- Observation O : collection of detected events with measured energy and position in the sky

- Model M = mathematical description for the spatial and spectral distribution of the photon flux. M is a function of a set of parameters \mathbf{p}

- Realization $M^* = M(\mathbf{p}^*)$ where \mathbf{p}^* is a particular set of value for the parameters. From M^* you can get fluxes, photon indexes, etc...

- we aim to find the best M^* of M from which we can extract physical information, where M is a model good enough for our application

- Maximum likelihood approach** (“semi-bayesian”): **assume** a reasonable model M (based on prior knowledge), then find the set of parameters \mathbf{p}^* maximizing the probability (likelihood) that the observation is coming from M^*

Likelihood: formal approach



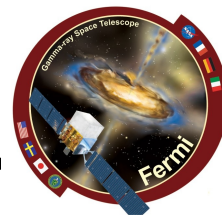
- Given a model M and a set of parameters \mathbf{p}^* the likelihood $L(\mathbf{p}^*)$ of M^* is the probability of obtaining the data from M^*
- The maximum likelihood estimation \mathbf{p}_{mle}^* of \mathbf{p} is the set of parameter for which $L(\mathbf{p}^*)$ is maximum (most probable set of parameter)
- How to estimate this?
 - Given an observation, we can bin events in multidimensional (energy, sky pixels...) bins.
 - N of obs. counts in each bin is Poisson distributed. L is the product of the probabilities of observing the detected counts in each bin, n_k , while m_k counts are predicted by the model ($N_{pred} = m_0 + m_1 + \dots + m_k$):

$$\mathcal{L} = \prod_k \frac{m_k^{n_k} e^{-m_k}}{n_k!} \quad \mathcal{L} = \prod_k e^{-m_k} \prod_k \frac{m_k^{n_k}}{n_k!} = e^{-N_{pred}} \prod_k \frac{m_k^{n_k}}{n_k!}$$

- This is called *binned likelihood*. We can maximize the function L (actually, $\log L$ which is easier to handle) with standard algorithms (Gradient descend, MIGRAD...) and find \mathbf{p}_{mle}^* .
- If we let the bin sizes get extremely small, then $n_k=0$ or 1 , and we are left with a product running over the number of photons (unbinned likelihood):

$$\mathcal{L} = e^{-N_{pred}} \prod_i m_i \quad \log \mathcal{L} = \sum_i \log(m_i) - N_{pred}$$

- Again we can maximize L and find \mathbf{p}_{mle}^*



The source model is considered as:

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

Point Sources

Galactic & EG Diffuse Sources

Other Sources

This model is folded with the Instrument Response Functions (IRFs) to obtain the predicted counts in the measured quantity space (E', p', t'):

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

Effective area
Energy dispersion
Point Spread Function

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 combined IRF. $\vec{L}(t)$ is the orientation vector of the spacecraft. The integral is performed over the Source Region, i.e. the sky region encompassing all sources contributing to the Region-of-Interest (ROI). In the standard analysis, only steady sources are considered

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



The function to maximize is:

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

where the sum is performed over photons in the ROI. The predicted number of counts is:

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

To save CPU time, a model-independent quantity, « exposure map (*cube*)» is precomputed:

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

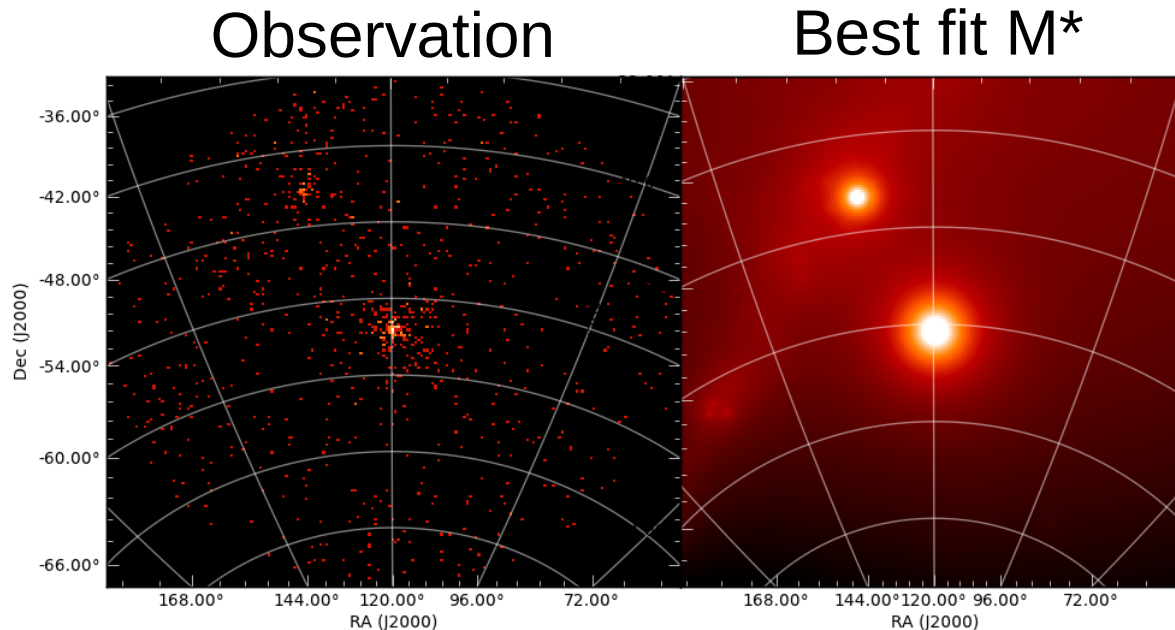
Then

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

Maximum likelihood example



- Observation: a 30x30 deg region in the sky (ROI)
- Building M:
 - Prior knowledge:
 - 2-year source catalog (2FGL): 2 sources in our ROI
 - Diffuse galactic emission (galactic gamma-ray background)
 - Extra-galactic diffuse and isotropic particle contamination
- Fit M to O → obtain M* with fluxes for the 2 sources, spectral parameters, etc...



Basic concepts 2



• **“Detection” problem:** is the source in green detected or not in LAT data?

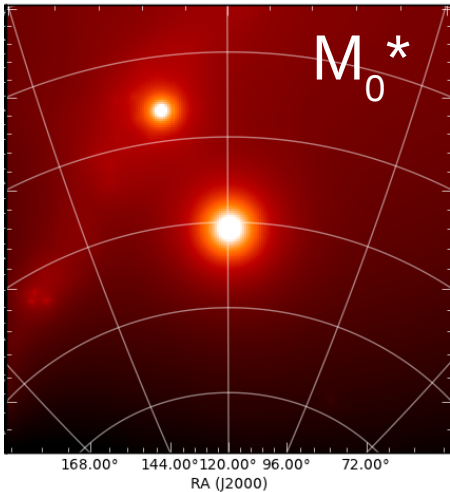
• **Likelihood ratio test idea:**

• find the maximum likelihood set of parameters p_0^* for a model M_0 without the point source (*null hypothesis*), and its likelihood L_0

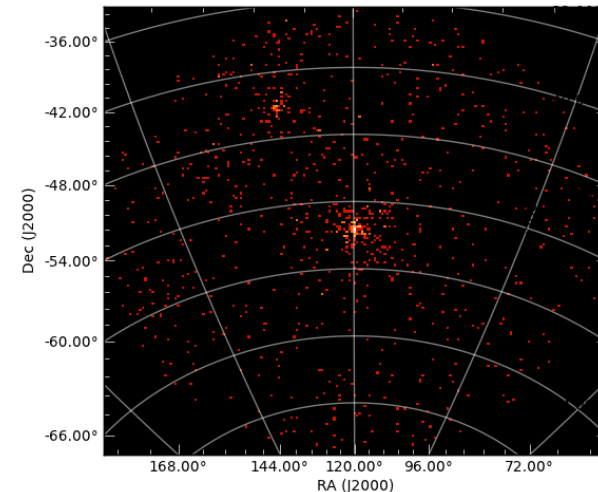
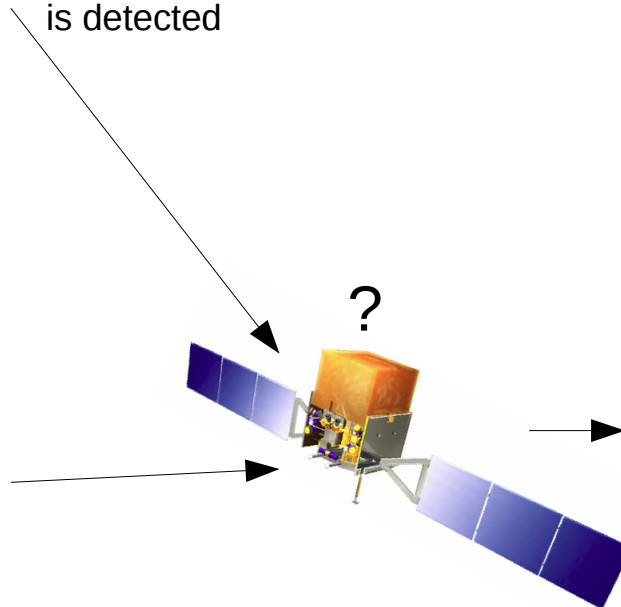
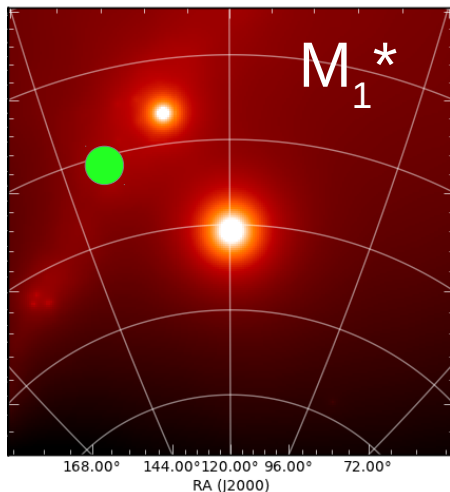
• find the maximum likelihood set of parameters p_1^* for the model $M_1 = M_0 +$ the point source component, and its likelihood L_1

• $L_1 \geq L_0$ by definition

• Using L_0 and L_1 quantify if the improvement is compatible or not with being a statistical fluctuation. If it is unlikely, then M_1^* is better than M_0^* and the source is detected



or





- Likelihood Ratio Test (LRT):
 - The Test Statistic for this test is defined as: $TS = -2 \log \left(\frac{\mathcal{L}_{\max,0}}{\mathcal{L}_{\max,1}} \right)$
 - In the limit of a large number of counts, TS is asymptotically distributed as a χ_n^2 where n is the number of parameters characterizing the additional source
 - Rule of thumb (NOT exactly true): \sqrt{TS} is approximately equal to the detection significance, i.e., $TS = 25 \rightarrow$ detection at ~ 5 sigma
 - Simulation showed that this is still approximately true even for small number of counts (\sim a few)
- NOTE: Due to computation constraints Fermi likelihood analysis assumes a position for the source (R.A. and Dec are fixed parameters) coming from other observatories or other algorithms (cluster finding...). There are some limited workarounds, and the position can be “optimized” (i.e., small changes). See `gttsmap` and `gtfindsrc` on the FSSC website (NOT covered today)

Software and data

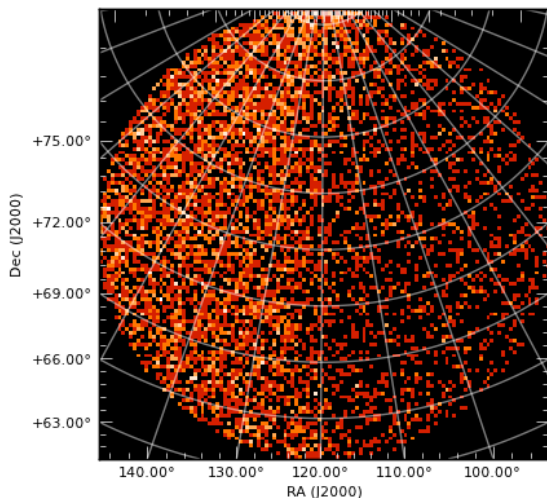


- A small percentage of the events detected by the LAT are actually photons
- A set of cuts is applied on the data to clean them:
 - The more stringent the cut, the cleaner the data, but the smaller the effective area
 - 2 classes of interest for us, both calibrated only above 100 MeV:
 - **Transient: loose selection, large effective area, dirty. Good for time intervals < 100-200 s**
 - **Source: much cleaner, smaller effective area, better PSF. Better than transient on time intervals > 100-200 s**
 - There is also the LLE class, featuring a very loose cut, very large PSF, but very large effective area from 20 MeV to 1 GeV. It is good for time and spectral analysis (NOT COVERED TODAY)
- Data comes in the form of list of events (FT1 file) plus a spacecraft file (FT2 file) containing housekeeping (attitude, orbital pos., livetime...)

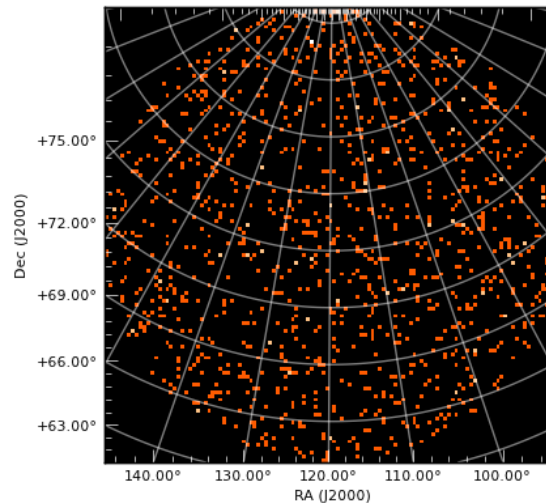


- The upper layer of the Earth atmosphere is a very bright source of gamma-rays (Limb)
- The angle between the Zenith of the spacecraft and the Limb is ~ 115 deg, but due to the large low-energy PSF photons from the Limb are present at lower Zenith angles
- Given our pointing strategy, the Limb is usually at the border of the FOV, and represents an issue for moderate to large off-axis angles ($> \sim 50$ deg)
- It is not easy to model in a likelihood analysis
- We use a Zenith cut: we exclude from the analysis all time intervals in which the border of our ROI has a Zenith angle larger than a certain value (z_{\max})
- When the source is off-axis, you have to find a trade off between the loss of exposure and the contamination

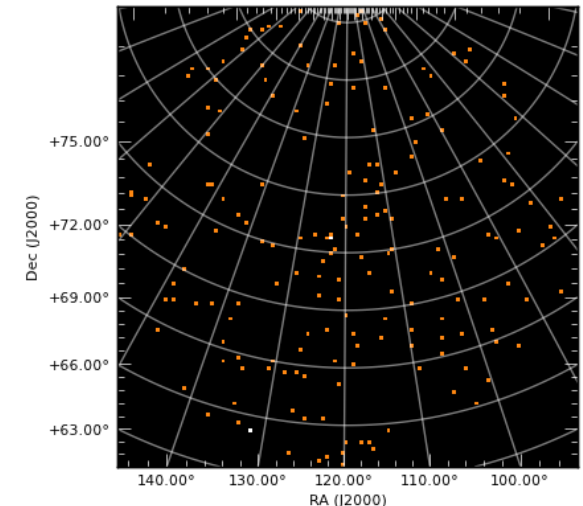
Transient class
(no Zenith cut)



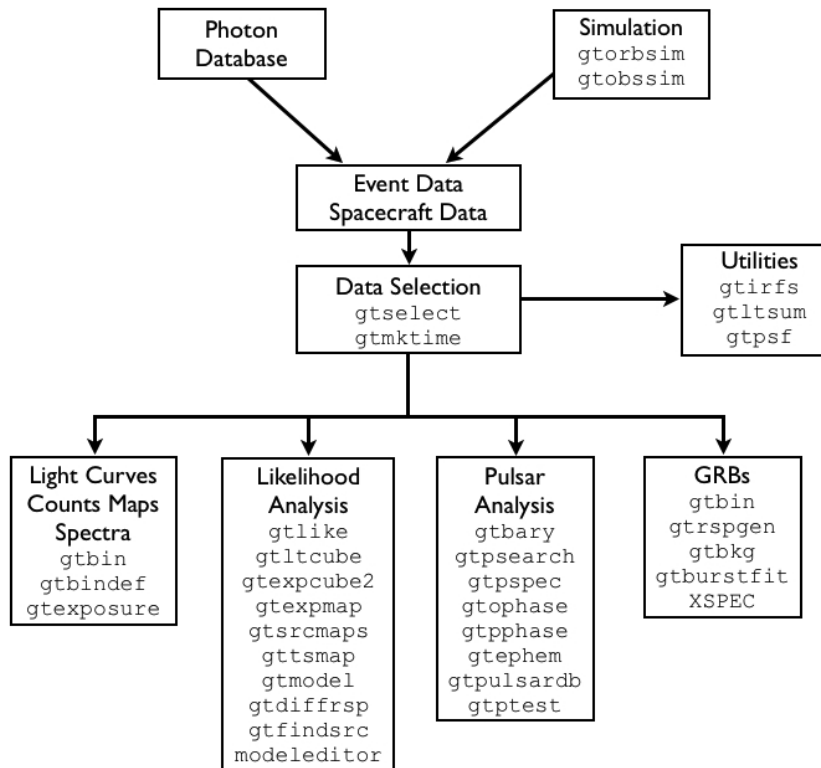
Transient class
($z_{\max} = 105$)



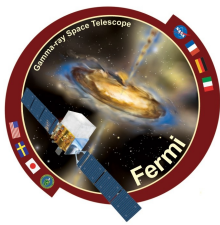
Source class
($z_{\max} = 105$)



(15 deg ROI, 10k seconds, 55 deg off-axis)



- Publicly available at the FSSC:
 - <http://fermi.gsfc.nasa.gov/ssc/data/analysis/>
- Already installed in your Virtual Machine!

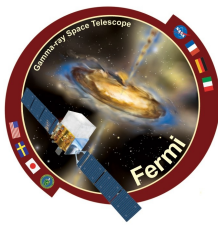


- There are two extensive tutorials on the FSSC website:
 - **Binned:**
http://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/binned_likelihood_tutorial.html
 - **Unbinned:**
http://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/likelihood_tutorial.html



- **Software suite with a Graphical user interface, specialized for analyzing transient sources (GRBs, Solar Flares...)**
- **Uses Fermi Science Tools, numpy and scipy**
- **Still in a very active development phase... not extremely stable! :-) be patient...**
- **Allow to:**
 - **Download GBM and LAT data (both LLE and standard)**
 - **LAT likelihood analysis, TS map, source localization...**
 - **LAT observation basic simulation (NOT COVERED today)**
 - **Produce GBM and LAT/LLE spectra (source, background and responses) to be analyzed in XSPEC (NOT COVERED TODAY)**
- **Pre-installed on your virtual machine. To run it:**
 - > `source ~/ferminit.csh`
 - (mount the shared folder, then create and go to a directory there, then:)
 - > `gtburst.py`

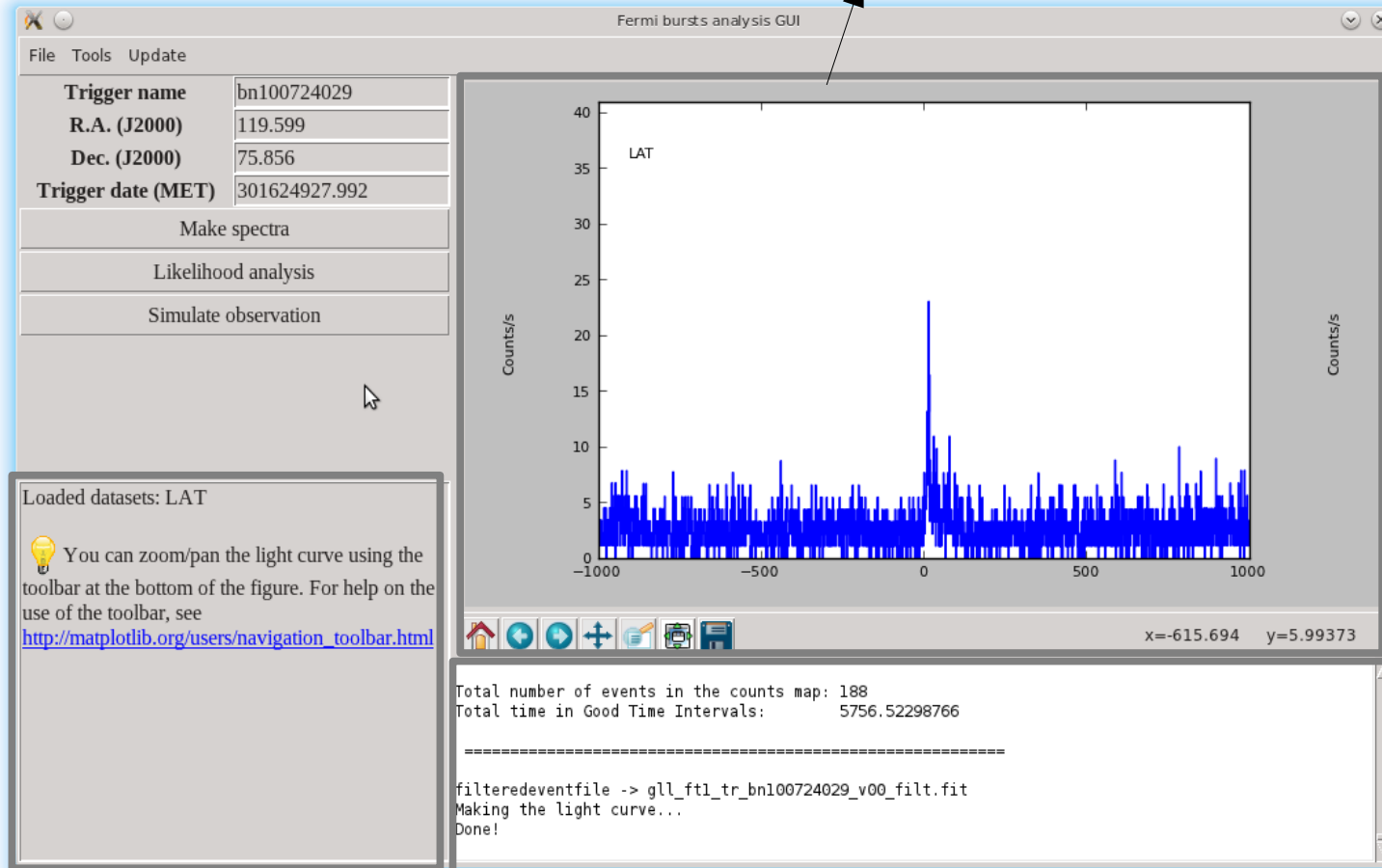
Gtburst: general considerations



- All output generated by the GUI is saved in the current directory, so **CREATE** a directory for a new analysis
- If you run gtburst twice in the same directory, the output files **WILL BE OVERWRITTEN**
- Go ahead and run it, by doing this:
 - > source ~/ferminit.csh
 - > sudo mount -a (provide your password, 'fermi')
 - > cd GRBWorkdir
 - > mkdir analysis
 - > cd analysis
 - > gtburst.py &



Figure canvas



Trigger name	bn100724029
R.A. (J2000)	119.599
Dec. (J2000)	75.856
Trigger date (MET)	301624927.992

Counts/s

LAT

Counts/s

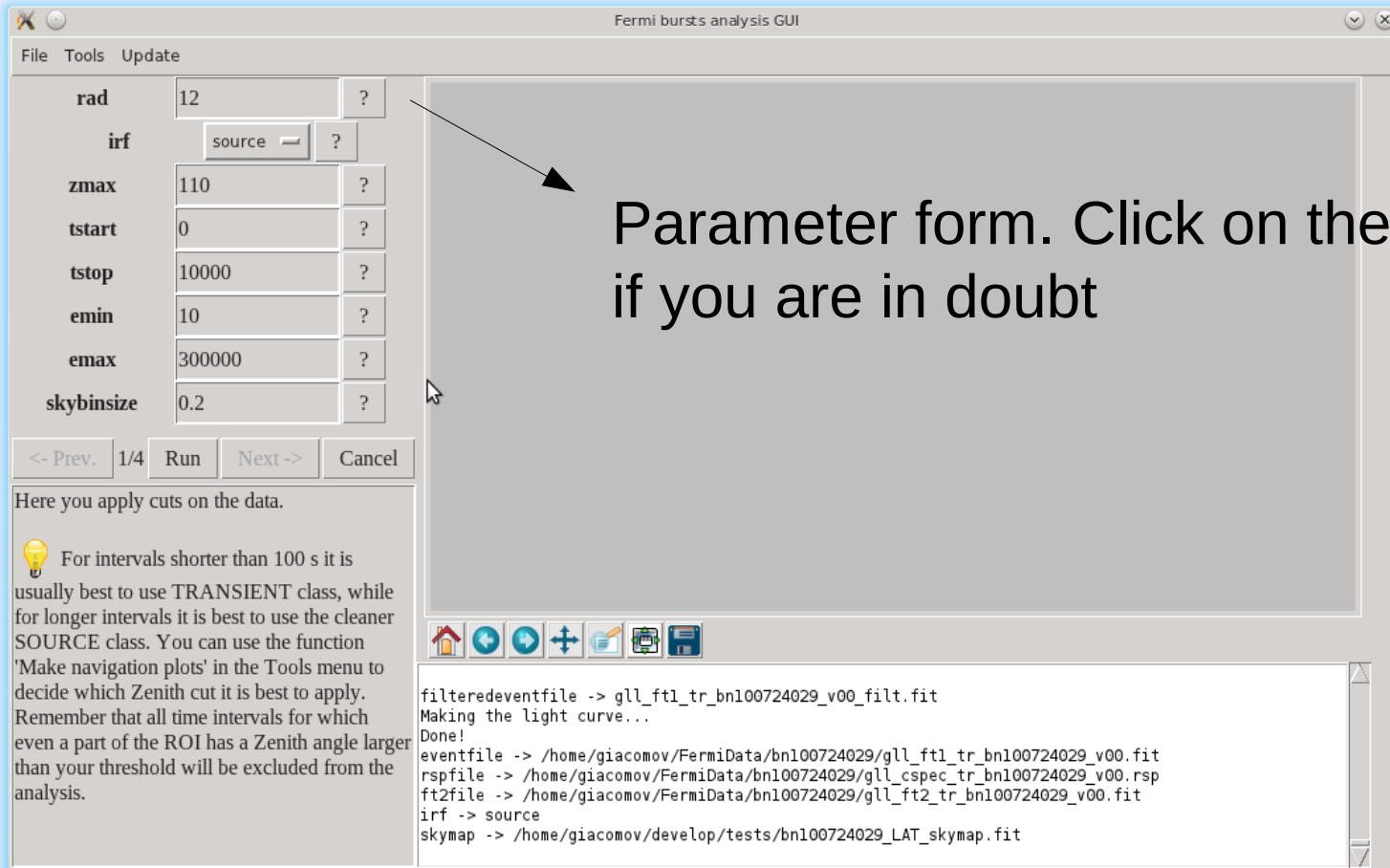
x=-615.694 y=5.99373

Total number of events in the counts map: 188
Total time in Good Time Intervals: 5756.52298766

filteredeventfile -> gll_ft1_tr_bn100724029_v00_filt.fit
Making the light curve...
Done!

Console: output from the commands

Help area: **ALWAYS** read here before doing anything

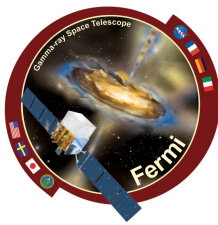


Parameter form. Click on the ? if you are in doubt

Here you apply cuts on the data.

💡 For intervals shorter than 100 s it is usually best to use TRANSIENT class, while for longer intervals it is best to use the cleaner SOURCE class. You can use the function 'Make navigation plots' in the Tools menu to decide which Zenith cut it is best to apply. Remember that all time intervals for which even a part of the ROI has a Zenith angle larger than your threshold will be excluded from the analysis.

```
filteredeventfile -> gll_ft1_tr_bn100724029_v00_filt.fit
Making the light curve...
Done!
eventfile -> /home/giacomov/FermiData/bn100724029/gll_ft1_tr_bn100724029_v00.fit
rspfile -> /home/giacomov/FermiData/bn100724029/gll_cspec_tr_bn100724029_v00.rsp
ft2file -> /home/giacomov/FermiData/bn100724029/gll_ft2_tr_bn100724029_v00.fit
irf -> source
skymap -> /home/giacomov/develop/tests/bn100724029_LAT_skymap.fit
```



- **All the output displayed in the Console is also saved in a file gtburst.log in the current directory**
- **Each step of each analysis thread corresponds to a python script gt*.py, which you can run from the command line**
- **If you look in the gtburst.log file or in the Console you can figure out how to reproduce your analysis with the python scripts, and possibly repeat the same analysis for other transients**



- **From the log file gtburst.log:**

```
===== gtdocountsmap =====
```

- Running gtdocountsmap on dataset LAT with this parameters:

```
rad                = 12
eventfile          =
/home/giacomov/FermiData/bn130325203/gll_ft1_tr_bn130325203_v00.fit
zmax               = 100
emin               = 100
emax               = 300000
skybinsize         = 0.2
skymap             = /home/giacomov/develop/tests/bn130325203_LAT_skymap.fit
rspfile            =
/home/giacomov/FermiData/bn130325203/gll_cspectr_bn130325203_v00.rsp
ft2file            =
/home/giacomov/FermiData/bn130325203/gll_ft2_tr_bn130325203_v00.fit
tstart             = 0
tstop              = 100
dec                = -21.1
irf                = transient
ra                 = 130.7
```

- **This corresponds to running:**

```
> gtdocountsmap.py rad=12
eventfile=/home/giacomov/FermiData/bn130325203/gll_ft1_tr_bn130325203_v00.fit zmax=100
emin=100 emax=300000 skybinsize=0.2 ..[etc]
```

Update gtburst



- **Go the Update menu and select “Update to the last development version”**
- **Wait for the download of the new file, and click “Yes” to the question**
- **You're done!**

Download data



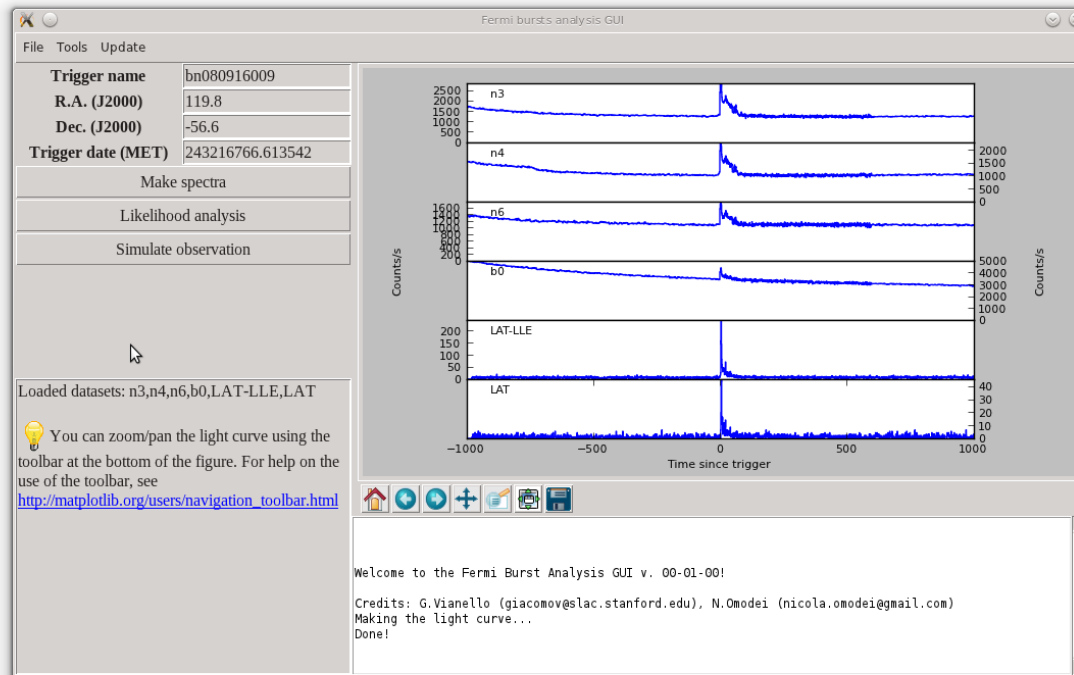
- Select “Download datasets” from the File menu
- Click on “Browse triggers”: the sw will download the list of all Fermi triggers from the Fermi Trigger catalog (available at heasarc)
- Click on the scroll down menu and select “GRB” (only GRBs will be displayed)
- Select a trigger and double click on it (for example, 080916009)
- You will go back to the previous window. Select which data you want to download by checking the appropriate boxes
- Click “Download data” and wait...
- Your data will be downloaded in `~/FermiData/[name of the trigger]` and automatically loaded
- **NOTE:** the download feature is **EXPERIMENTAL** and kind of error prone, especially on slow or buggy connections. You might need to retry.



Load existing data



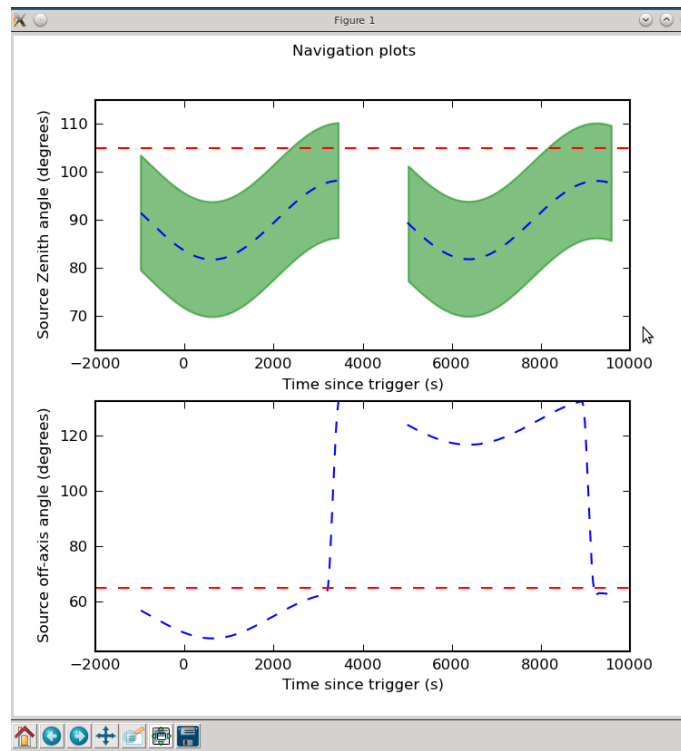
- Click on the File menu
- Choose “Load data from a directory”
- Navigate to the directory named after the trigger of interest, and enter the directory (double click on it)
- Click Ok
- Check the detectors for which you want to load the data. The number in parenthesis is the angle between each detector and the trigger
- Click Ok
- After a few seconds light curves of your data will appear

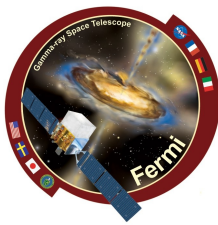


Navigation plots



- At any time you can click on the Tools menu and choose “Make navigation plots”
- You will see the Zenith angle (upper panel) and the off-axis angle (lower panel) of the source.
- The green band marks the limits of a 12 deg ROI
- The red lines are respectively Zenith 105 (normal zmax cut), and off-axis = 65 deg (LAT nominal field of view for standard classes)





- Click on the button “Likelihood analysis” in the main window
- First step: data selection
 - You have to define your cuts: radius of the region of interest (it is always centered on the position taken from the Trigger catalog), source class, zenith cut (help yourself with the navigation plots)...
 - Then click run, and you will see a sky map of the events surviving your cuts. Experiment with the cuts, when you are happy, click Next
- Second step: Model definition.
 - You have to choose one model for the particle and residual background and one for the galactic (read the text in the Help space!)
 - Click run
- Third step: edit your model
 - Here you can double click on a parameter and edit it. Advanced... skip for now. Just click on Done and then on “next”
- Fourth step: perform the likelihood!
 - Just click on Run and wait (5-10 minutes)
 - You will see the original (observed) image, the best fit model, and a table with the flux of the GRB (if its TS > 20) or an upper limit otherwise
- Done! Take note of your results and click Finish