LAT source catalog

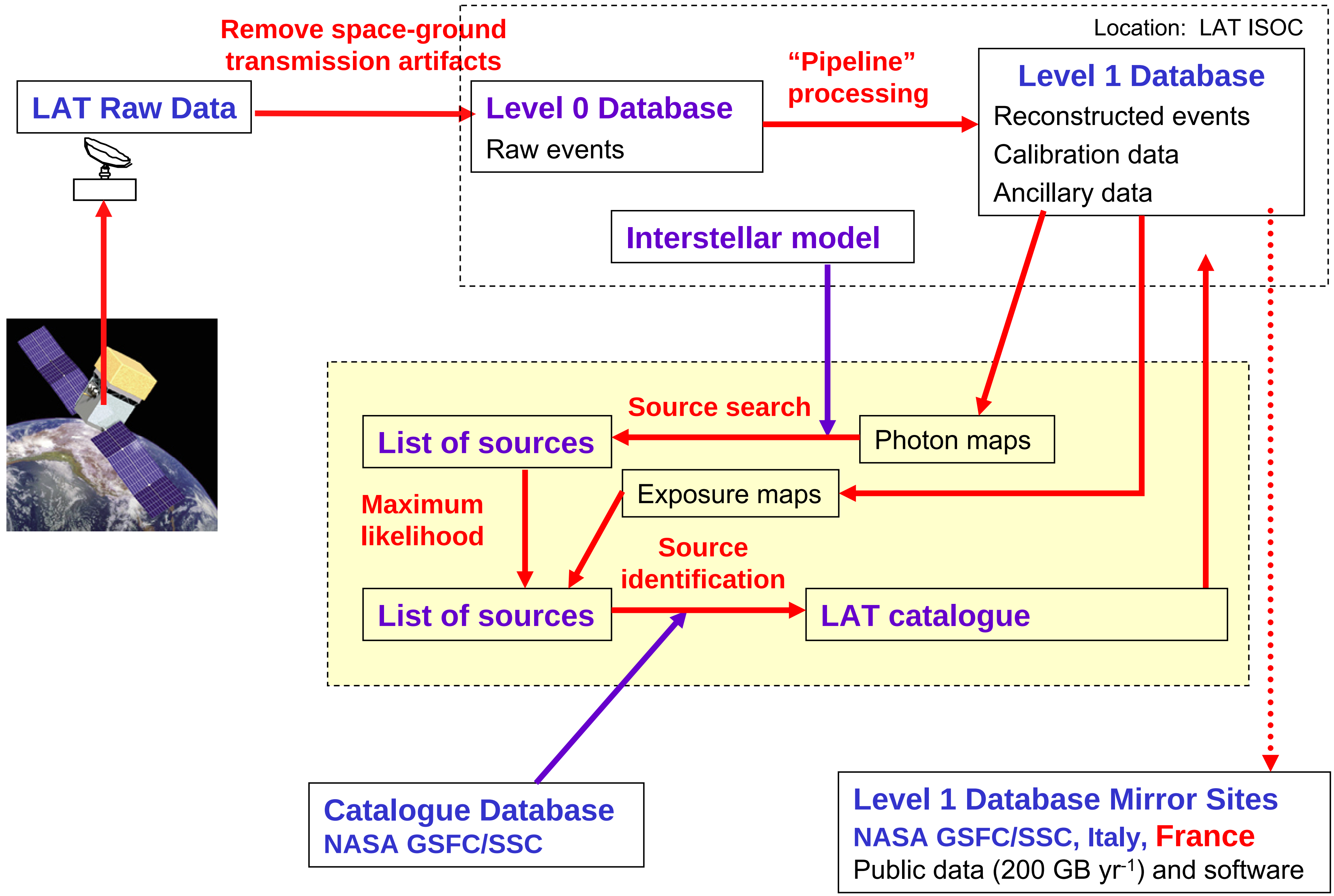
Thousands of γ-ray sources are expected at the LAT sensitivity.
Most will be extragalactic (blazars) but we also expect many Galactic sources (pulsars in the first place).
What were EGRET's unidentified sources ?
What will GLAST find below EGRET's sensitivity level ?

CEA/SAp (I. Grenier) is responsible for putting together the LAT source catalog. Obtained support from CNES.

Main aims :

1. Official LAT product (after 1 year, 3 years, 5 years)
2. Provide the basis for statistical studies (population)
3. Provide the starting point for detailed studies (individual sources)
4. Useful also inside the LAT collaboration (before one year)

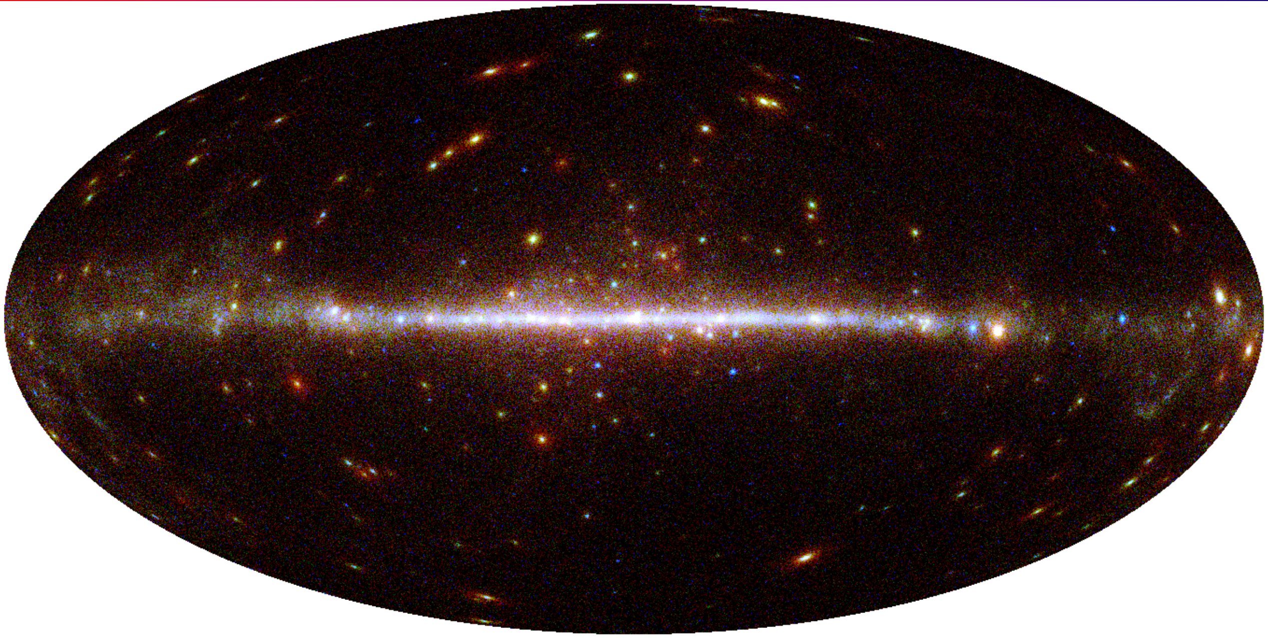
Data flow



Remove space-ground transmission artifacts
LAT Raw Data
Location: LAT ISOC
Level 0 Database
Raw events
"Pipeline" processing
Level 1 Database
Reconstructed events
Calibration data
Ancillary data
Interstellar model
Source search
List of sources
Photon maps
Exposure maps
Maximum likelihood
Source identification
List of sources
LAT catalogue
Catalogue Database
NASA GSFC/SSC
Level 1 Database Mirror Sites
NASA GSFC/SSC, Italy, France
Public data (200 GB yr-1) and software

Source detection

DC2 simulation
(image S. Digel)



Goal:

Detect γ-ray sources over the entire sky

List of seed positions for validation step

Specificities (difficulties):

- Very structured Galactic diffuse emission (interstellar medium). Model using surveys at other wavelengths (HI, CO, dust, …)
- PSF improves with energy (3° at 0.1 GeV, 0.1° at 10 GeV)

Model:

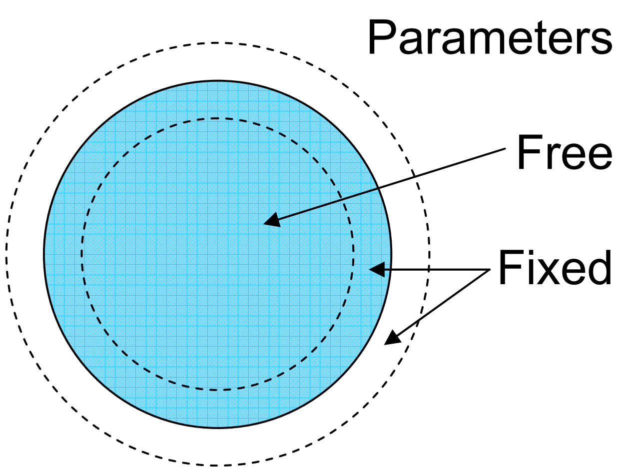
- Cover the sky with a number of images, split into several energy bands
- Use wavelets to find concentrations of photons
- Official detection algorithm not yet chosen, several good contenders

Source validation

Maximum likelihood: provide final list of sources

- ✓ Restricted to events above 100 MeV (faster, and more reliable)
- ✓ Pave the sky with Regions of Interest centered on no more than 10 sources
- ✓ Adjust flux and spectral slope of all sources in center of region
- ✓ 5 iterations; Use fast optimizer, then final call to MINUIT to get errors
- ✓ Threshold at Test Statistic = 25

- ✓ Standard region of interest (20° radius) contains many sources (several tens). Optimizers have trouble converging over so many parameters.
- ✓ Reduce area over which sources are fitted, taking nearby sources (fixed) from run on nearby regions of interest.

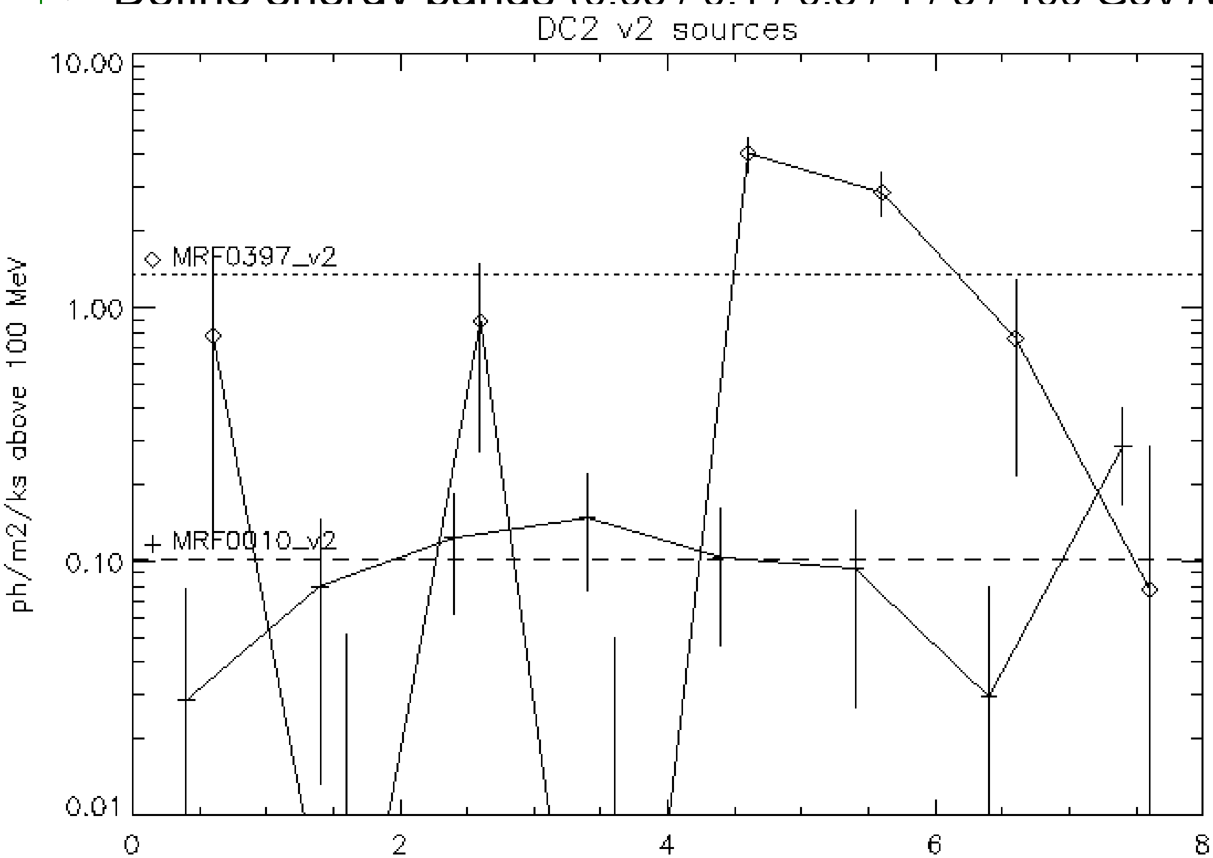


Source characterization

Aim: Provide more energy information than a simple power-law index, and crude light curves (allowing to select interesting sources)

- ✓ The list of sources, and their spectral index, are taken from the global run
- ✓ Define energy bands (0.03 / 0.1 / 0.3 / 1 / 3 / 100 GeV) and run maximum

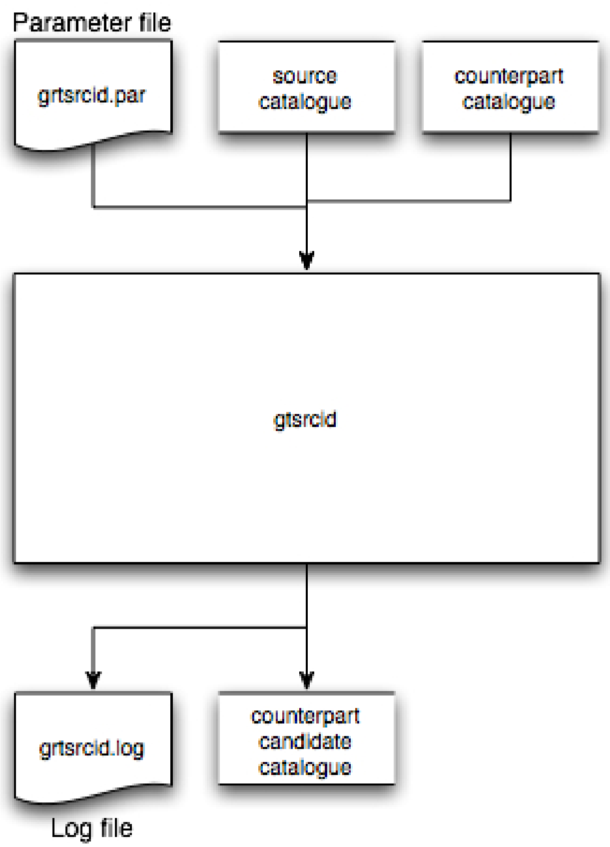
for DC2) and run



- ✓ Dashed and dotted lines show the flux returned on the full interval
- ✓ MRF0010 is nicely constant
- ✓ MRF0397 is clearly variable

Automatic source identification

gtsrcid, provided by CESR Toulouse (J. Knödlseder):
Find geometric coincidences to γ-ray sources from a catalogue of potential counterparts (blazars, pulsars, …)
Figure of Merit calculation from quantities available in input catalogues (spectral energy distribution)



DC2 simulation

- ✓ First time we ran the catalog pipeline on representative data
- ✓ 534 sources possibly detectable over 100 MeV (Test Statistic > 10) among the 1719 simulated ones
- ✓ 423 at TS > 25 (same cut as in the catalog) vs 380 in the catalog.
- ✓ 62 missed at TS > 25. Most are soft sources confused with harder ones, or faint sources confused with brighter ones.

- ✓ 16 catalog sources have no true source in their error box (false positives)
- ✓ 4 have a true source not far outside their error box.
- ✓ Most of the rest are soft, probably due to imperfect modelling of the structured diffuse emission.
- ✓ 200 sources identified, 85% correct extragalactic, 44% correct Galactic

Computing resources

At DC2, used 10 local independent CPUs accessing data directly on central NFS disk. Pipeline managed by OPUS scheduler.

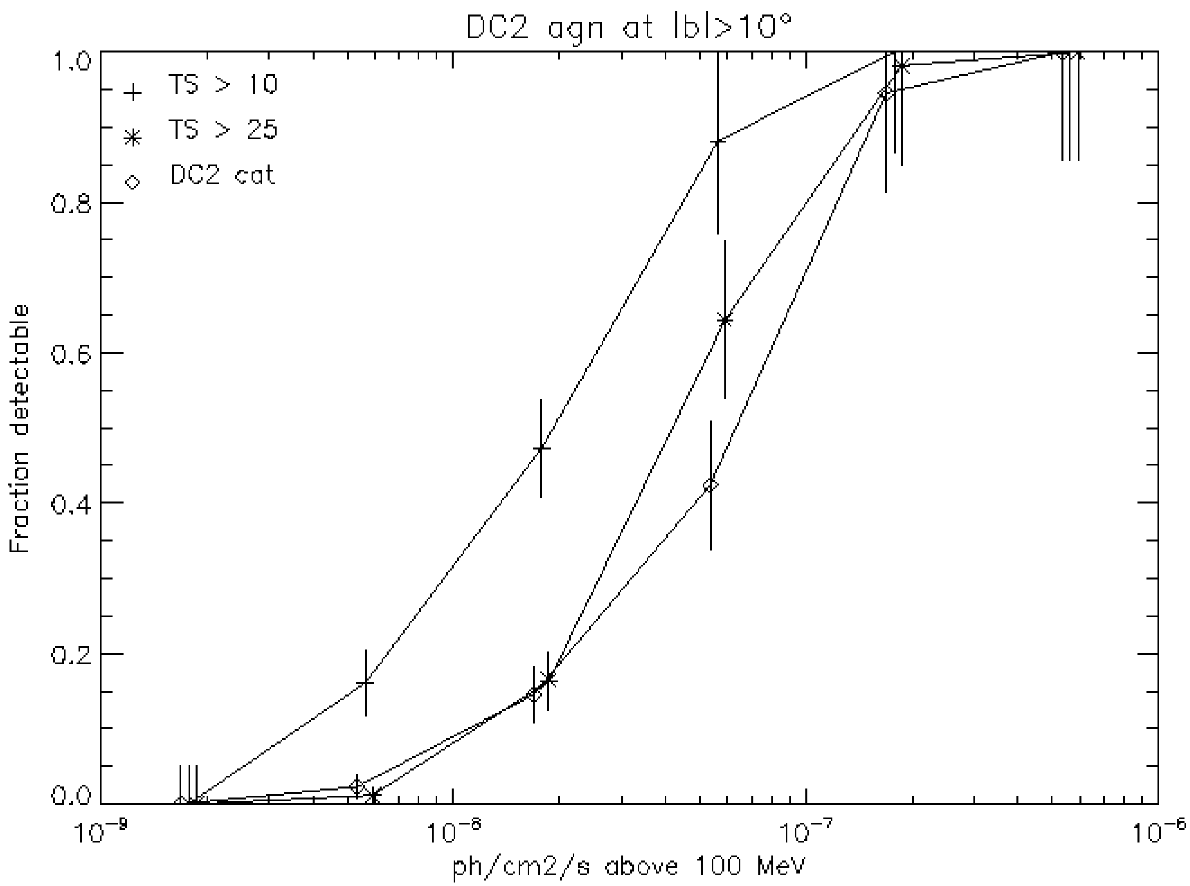
Total computing time 90 CPU days (mostly the maximum likelihood validation).

For 1 year, expect from 500 to 1000 CPU days if no speed up.

Since DC2:

- ✓ D. Landriu has extended the concept to work on local PC farm (15 CPUs available). No direct access to our disk, needs to transfer the data back and forth.
- ✓ Next step is big PC farms like Lyon. Cannot use OPUS to manage the pipeline there (no monitoring). A simpler scheduling system using qsub and a master script can work. Easily adaptable to other such PC farms, or Grid.

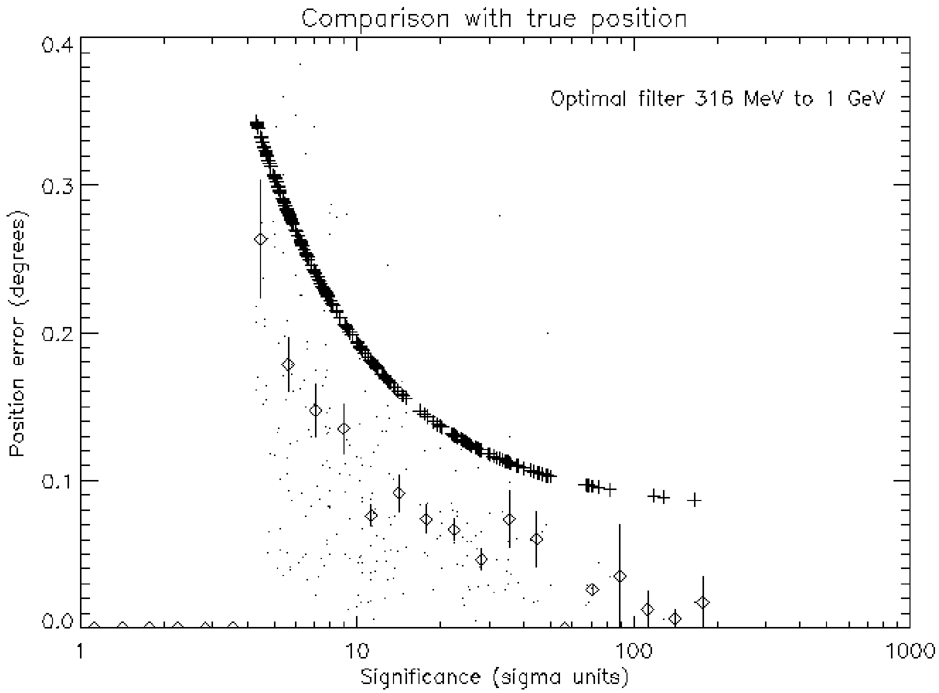
55 days sensitivity

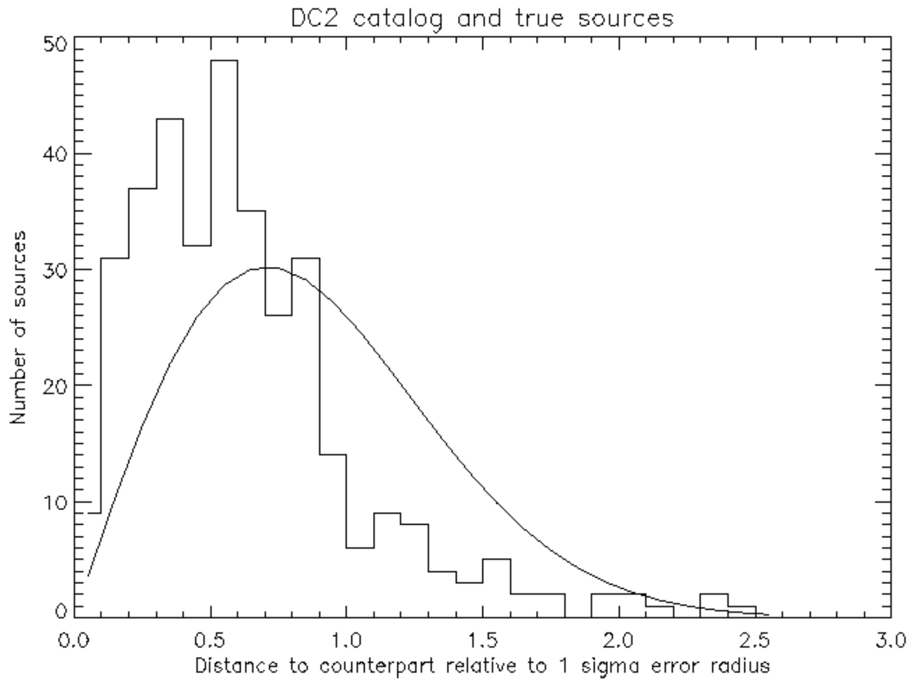


- ✓ Select all AGN outside the Galactic plane. Bin in log(flux). Compute ratio of detected sources to total number
- ✓ TS > x: likelihood using true source position
- ✓ DC2 cat: full detection process

- ✓ DC2 cat is very close to TS > 25. 50% detection rate 5 10^{-8} ph/cm^2/s in 55 days, extrapolates to 1.5 10^{-8} ph/cm^2/s in 1 year
- ✓ All spectral indices (soft sources more difficult)

DC2 position accuracy

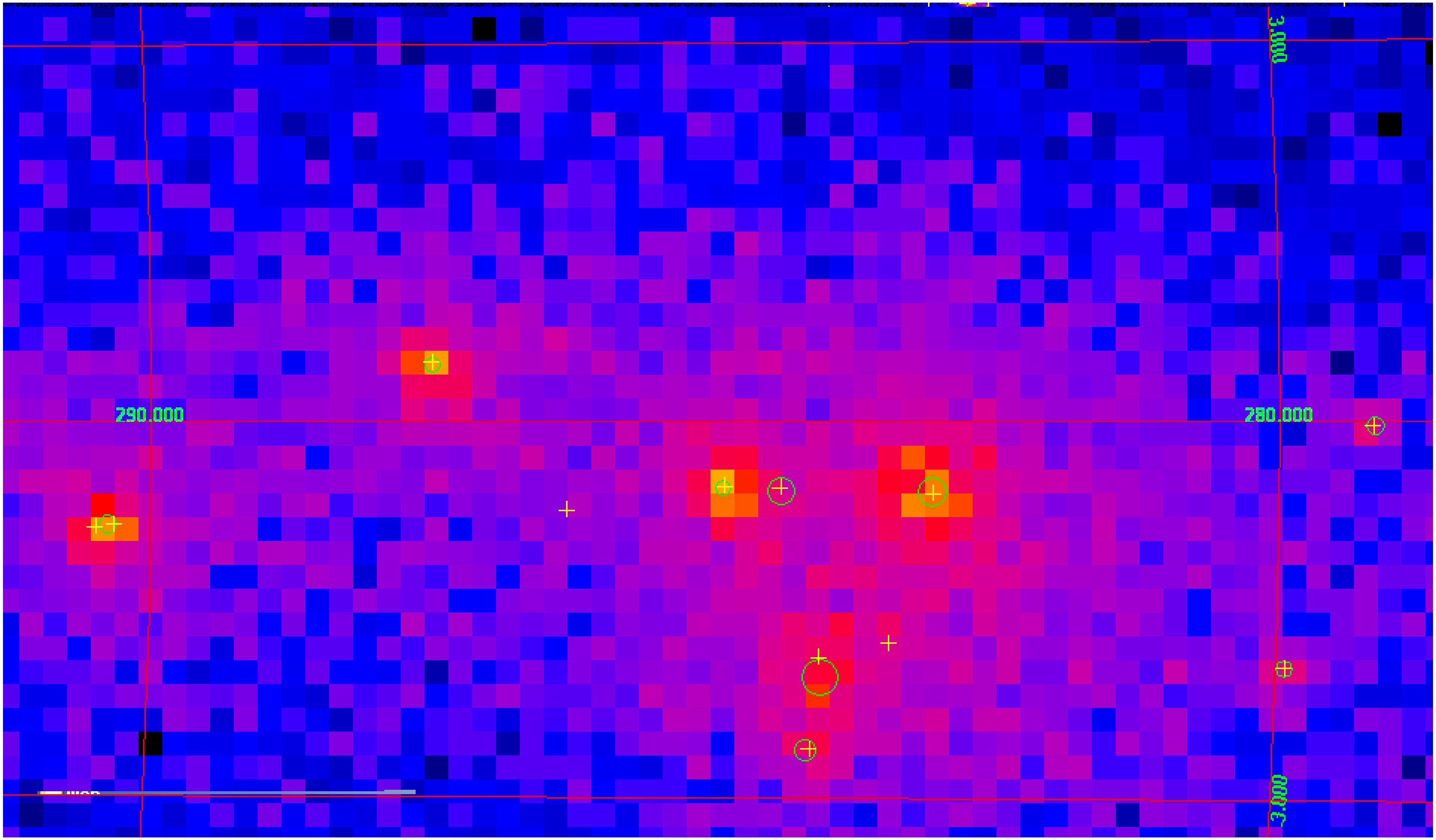




With a method that provides source significance, one can:

- ✓ Calibrate position error (by simulation) as a function of PSF width, significance (in sigma units) and pixel size (done on ST checkout 3)
- ✓ Use this for all sources in all bands
- ✓ With optimal filter, position error = 1.8 HWHM / significance + pixel / 2.5
- ✓ Use combined (smaller) position error when merging bands (only if association is not ambiguous)

DC2 sources



Areas for improvement

The current catalog still has a number of serious limitations:

- ✓ Initial source detection is still a place holder (but not so bad already).
- ✓ The likelihood program is very slow as soon as many sources are considered together, even though we are still doing only very simple things (simple power law spectrum, fixed position, no energy redistribution).
- ✓ The source position is not that bad, but can be improved by combining energy bands. Must compare with likelihood optimization.
- ✓ The position error was extremely rough for DC2. Deducing it as a function of source significance is better, but round. Ideal would be to get it from likelihood.
- ✓ The errors on flux and spectral index (from MINUIT) need to be assessed. Certainly very bad for not significant source fluxes in energy bands and time intervals.
- ✓ How many free diffuse emission parameters in each region of interest (local analysis) ? 2 spatial parameters but no spectral one for DC2.

Speeding up the likelihood algorithm is key to adding new functionality.

LAT source catalog

A first catalog was produced (and used) for DC2

Outlook:

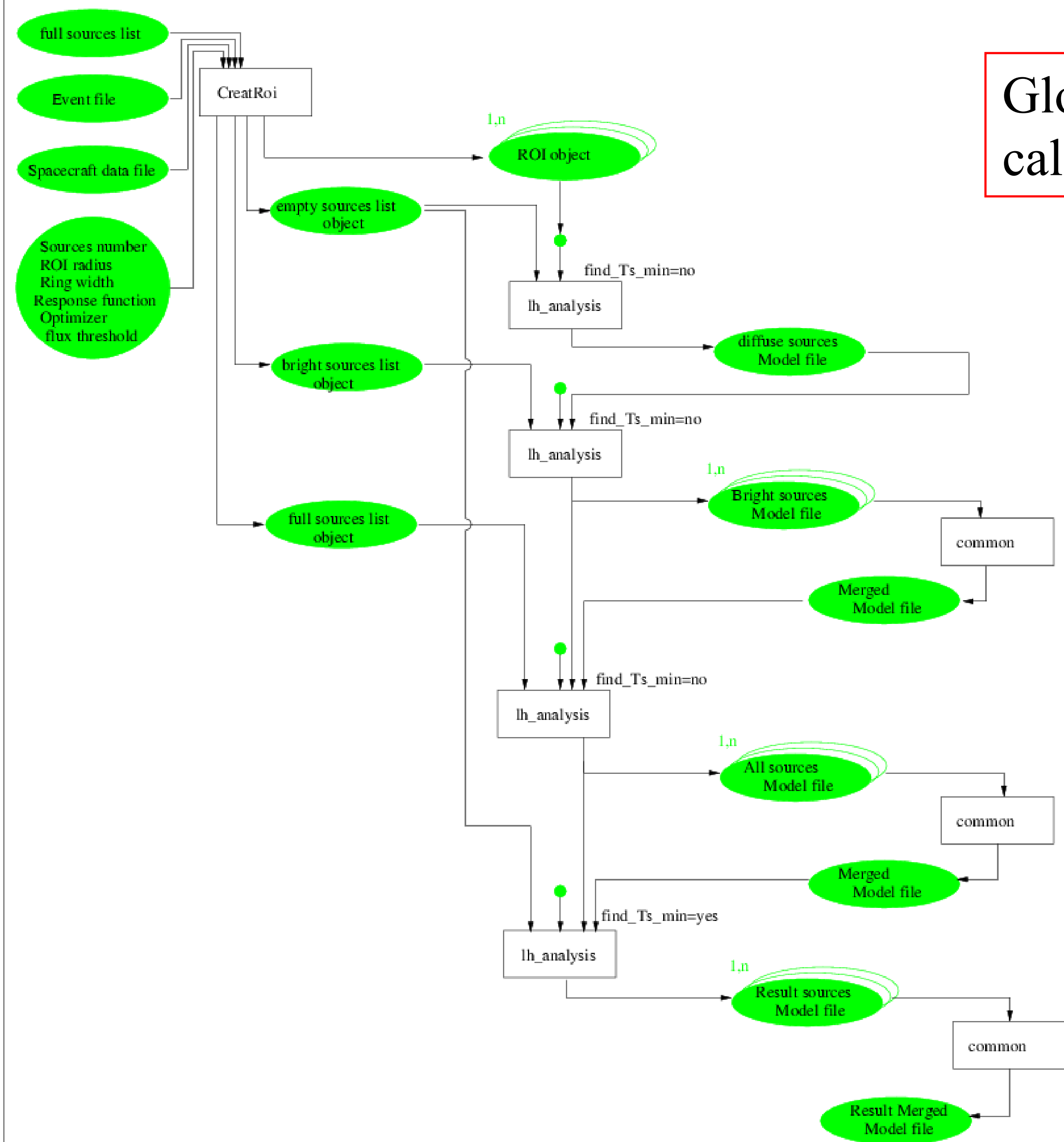
1. Select initial source detection method
2. Optimize source localization and error box determination
3. Integrate source identification
4. Build data base to support catalog work
5. Test the catalog pipeline on one year's worth of data
6. Think of adapting the catalog pipeline to transient source detection

Source catalog contents

The official catalog format is given at https://confluence.slac.stanford.edu/display/ST/LAT+Source+Catalog+Contents

- ✓ Source_Name (1GL + a number for internal reference), RA, DEC, Conf_95_SemiMajor (95% error radius) from all-sky source detection
- ✓ Flux100 (100 MeV to 200 GeV) and error, Spectral_Index and error, Signif_Avg (in sigma units) from likelihood
- ✓ Flux30_100, Flux100_300, Flux300_1000, Flux1000_3000, Flux3000_100000 (0.03 / 0.1 / 0.3 / 1 / 3 / 100 GeV) and errors from likelihood in bands
- ✓ Flux_History(n), Unc_Flux_History(n) and History_Start(n+1) from likelihood in one month intervals

Global scheme for calling likelihood



Idea. Run likelihood in several steps to facilitate convergence

Input: full list of sources, split into bright/faint

✓ At each step, use the same RoIs, events lists, exposure maps, but a different XML source file

✓ After each step, merge XML source files into a global (all-sky) one, keeping parameters from RoI to which source is closest (if several)