

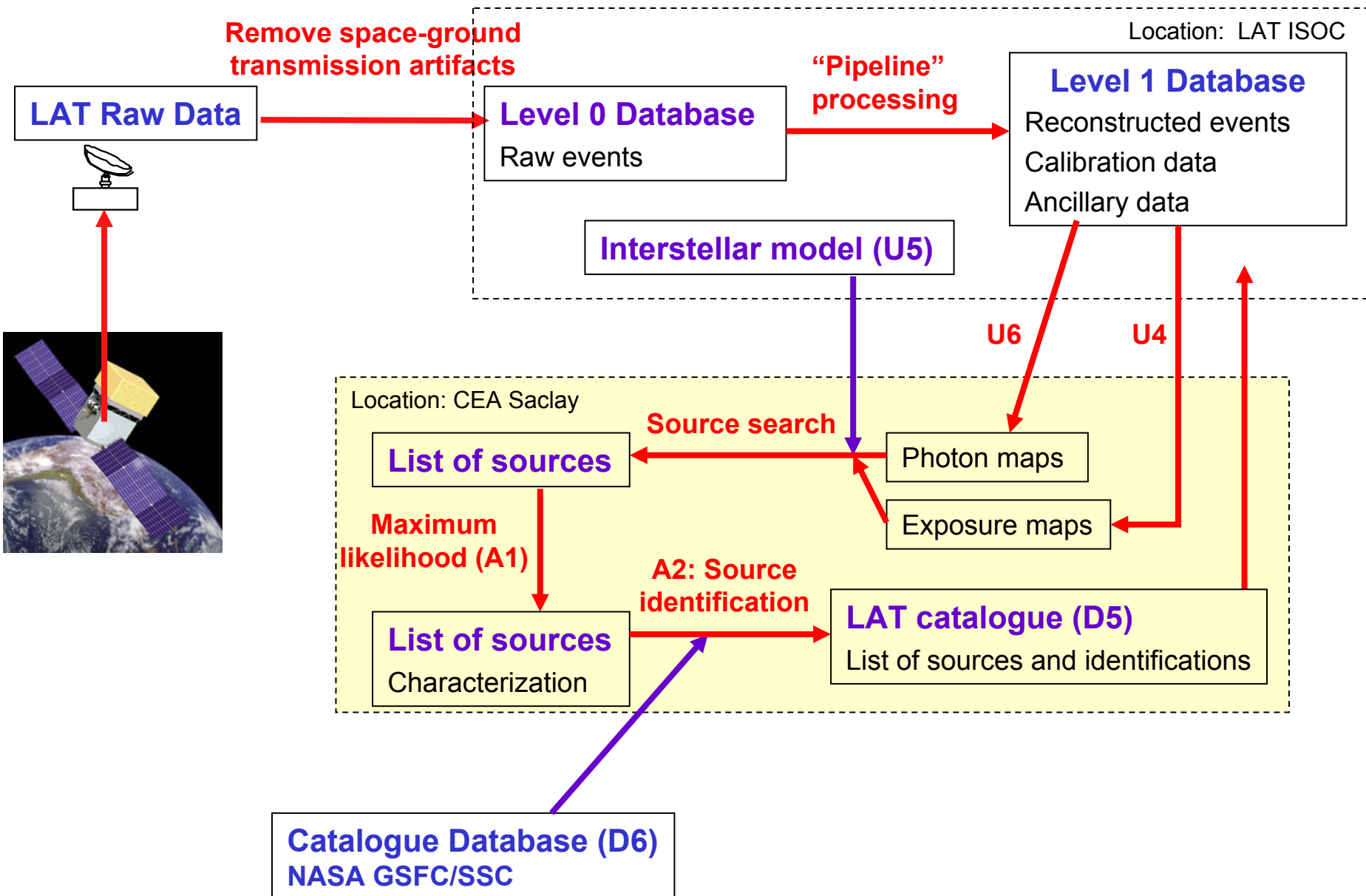
Source catalogue generation

Aim: Build the LAT source catalogue (1, 3, 5 years)

Three main functions:

- ✓ Find unknown sources over the whole sky (output: list of positions). This is the main purpose of this presentation.
- ✓ Localize sources. Output: list of precise positions (and uncertainties)
- ✓ Characterize sources (significance, flux, spectrum, variability, counterparts). This is no different from studying already known sources, and is well covered by the existing (or foreseen) Science Tools. Output: final catalogue

Catalogue production pipeline



Catalogue pipeline. Sequence

Aim: Implement automatic loop to find and characterize the sources

Minimal features:

1. Detect bright sources with current diffuse model (several energy bands)
2. Adjust diffuse emission model (global), possibly iterate to 1
3. Add bright sources to diffuse model
4. Detect faint sources over that diffuse model (several energy bands)
5. Localize faint sources
6. Run Likelihood at that position to get precise flux, spectrum and significance

- ✓ Task scheduling tool (like OPUS) for distributing work over CPUs
- ✓ Simple database for bookkeeping and for the source lists

Catalogue pipeline. Schedule

1. Identify candidate source search algorithms. Done (next talks)
2. Define evaluation criteria. Started. Done by November 2004
3. Build pipeline prototype. By end 2004.
4. Evaluate candidate algorithms. Beginning of 2005
5. First selection of source search algorithm. Mid 2005 (before DC2).
Reconsider mid 2006.
6. Define processing database. By end 2005
7. Integrate pipeline elements (including flux history, identification). 2006.
8. Ready: end 2006

All-sky source search

Aim: Look for a fast method to find sources over the whole sky

Algorithm will be the subject of following talks. Here I discuss several issues generic to all image-based methods:

- ✓ Selection criteria
- ✓ Coordinates, projection, sky partition, pixel size
- ✓ Energy information
- ✓ Threshold, false detections
- ✓ Simulations
- ✓ Variable and extended sources

All-sky source search. Selection

Aim: Define criteria for selecting a source search algorithm

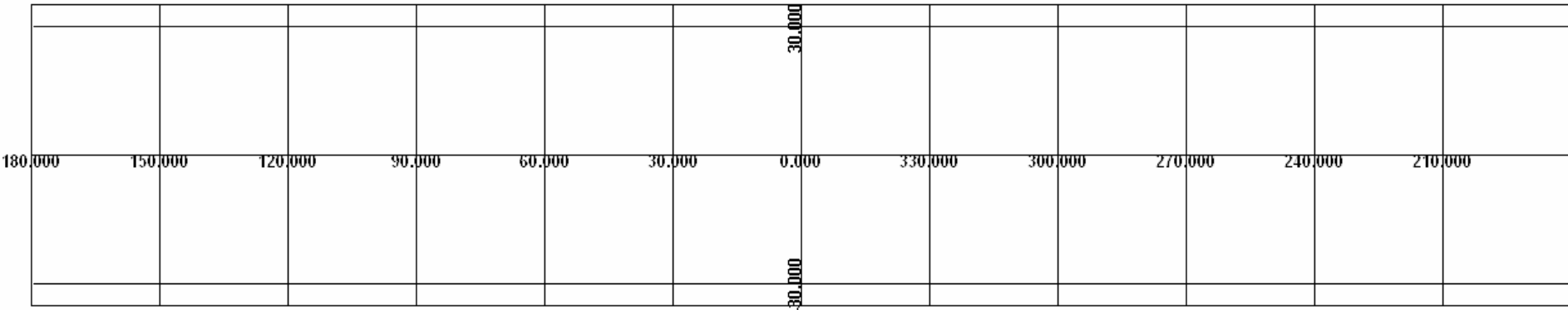
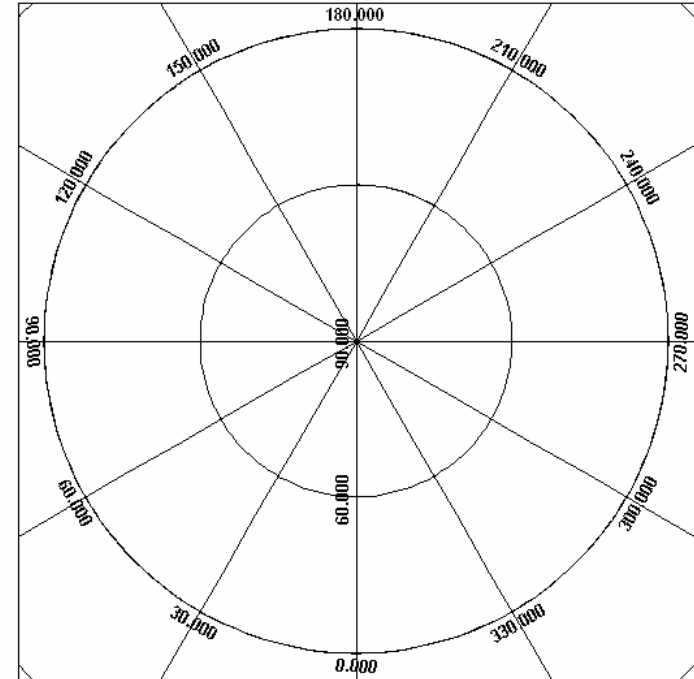
Difficulty is that criteria are best applied on final output (after maximum likelihood). Because of computing requirements, can be done only after most parameters of a given method have been optimized (projection, pixel size, energy bands, ...)

Criteria in approximate order of importance for source search itself:

1. Detection power (no new source will be found after that)
2. Resolving power (sources close to each other)
3. Flux and position estimate (to start later steps)
4. False detection rate (can be controlled later)
5. Computing time (of entire pipeline. Many false detections bear on this)

All-sky source search. Coordinates

- ✓ Coordinates: Galactic most logical choice (on which source density and diffuse emission have simplest distribution)
- ✓ Cartesian projection (CAR) around the Galactic plane. Wraps around in longitude. Source elongation: $1 / \cos b$
- ✓ Polar projection (ARC: $r = \pi/2 - b$ or $\pi/2 + b$, $\theta = l$) around the poles. Extend to square map. Source elongation: $r / \sin r$
- ✓ Best limit where $r = 1$ rad ($|b| = 32.7^\circ$), where elongation is 1.19.
- ✓ Need to extend both projections in latitude by about 1 PSF
- ✓ Can also use polar projection much more locally, and move around. More overhead.



All-sky source search. Energy bands

- ✓ Background limits source detectability by its Poisson fluctuations, even if the diffuse emission model is accurate. Approximate signal to noise (for weak sources) is S / \sqrt{B} , where S and B are taken over 1 PSF.
- ✓ All sources do not have the same spectrum. Soft sources will be better seen above the diffuse emission at low energy, hard sources at high energy.
- ✓ PSF varies enormously from low energy ($> 4^\circ$ below 100 MeV) to high energy ($< 0.2^\circ$ above 1 GeV). This means that low energy photons from bright sources act as background to nearby fainter sources.

- ✓ Splitting into several energy bands is better than summing everything. Example for optimal filter method (just from source lists): 105 sources in 0.1-1 GeV band, 109 in 0.1-0.316 (51) + 0.316-1 (89).
- ✓ 4 energy bands (30 MeV / 100 MeV / 316 MeV / 1 GeV / 10 GeV) should be all right. Cannot split indefinitely (more degrees of freedom)
- ✓ Pixel size should be adapted to PSF in each band
- ✓ Merge source lists is simplest, add likelihood images (before applying threshold) should be better

All-sky source search. Threshold

- ✓ Trade-off between detection power and false detection rate. Better limit the number of false detections or their fraction (wrt true sources) ?
- ✓ Plot both as a function of significance
- ✓ Maximum likelihood will normally provide final significance

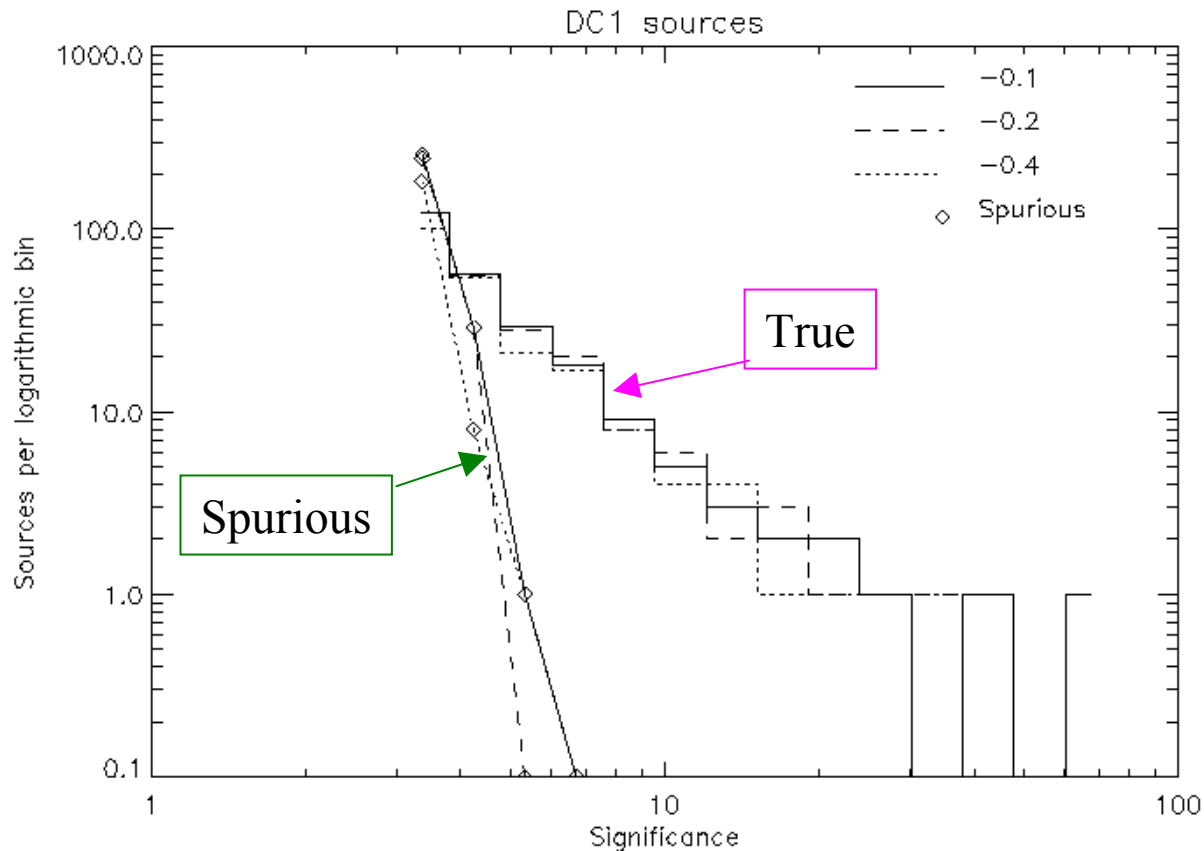
Example for optimal filter method as a function of bin size (all-sky).

316 MeV – 1 GeV band

PSF width in that band is about 0.67° (at 30% maximum).

This kind of diagram (or cumulative equivalent) is the tool of choice to select the best threshold for source detection.

Can be advantageous to keep a rather low threshold and let maximum likelihood cut.



All-sky source search. Simulations

The DC1 data has provided a rich resource for testing the algorithms. It has however several limitations :

1. It is only a particular trial. It is useful to be able to simulate the same model many times, to test robustness.
2. The point spread function that the sources follow is representative, but not well reproduced by the Science Analysis Environment (not enough data to follow energy and off-axis variations, imperfect functional form).
3. It does not allow to study in detail particular aspects, like sources close to each other at varying flux ratio.

We don't actually need a full instrument simulation for testing source detection algorithms. It is more important to control carefully the conditions of the simulation.

- ✓ Points 1 and 2 are best addressed by obsSim, with a smoothly varying PSF as a function of energy and off-axis angle. Need to incorporate elongation when off-axis.
- ✓ Point 3 is best addressed by local simulations (with known PSF).

All-sky source search. Special cases

Variable sources (blazars mostly). This covers two different things:

- ✓ Identify as variable sources which have been detected over the entire time period. Do we have a Science Tool for that ?
- ✓ Detect variable sources which have been missed over the entire time period (because of dilution). Can be done by repeating the source search over shorter time intervals (like one week), or by a specific algorithm (like looking for variability systematically in sky 'pixels').

Extended sources (external galaxies, supernova remnants, interstellar structures). This covers two different things:

- ✓ Identify as extended sources which have been detected by the point-source algorithm. Can be done by comparing source shape with PSF convolved with a Gaussian of variable width.
- ✓ Detect extended sources which have been missed by the point-source algorithm. Can be done by wavelet algorithms, or simply by looking for excesses in residual photon map (sources and diffuse emission subtracted).

Source localization

- ✓ Done locally (for each source in turn)
- ✓ Typical algorithm (like SExtractor) uses a smoothed map as input, and interpolates to find the maximum.

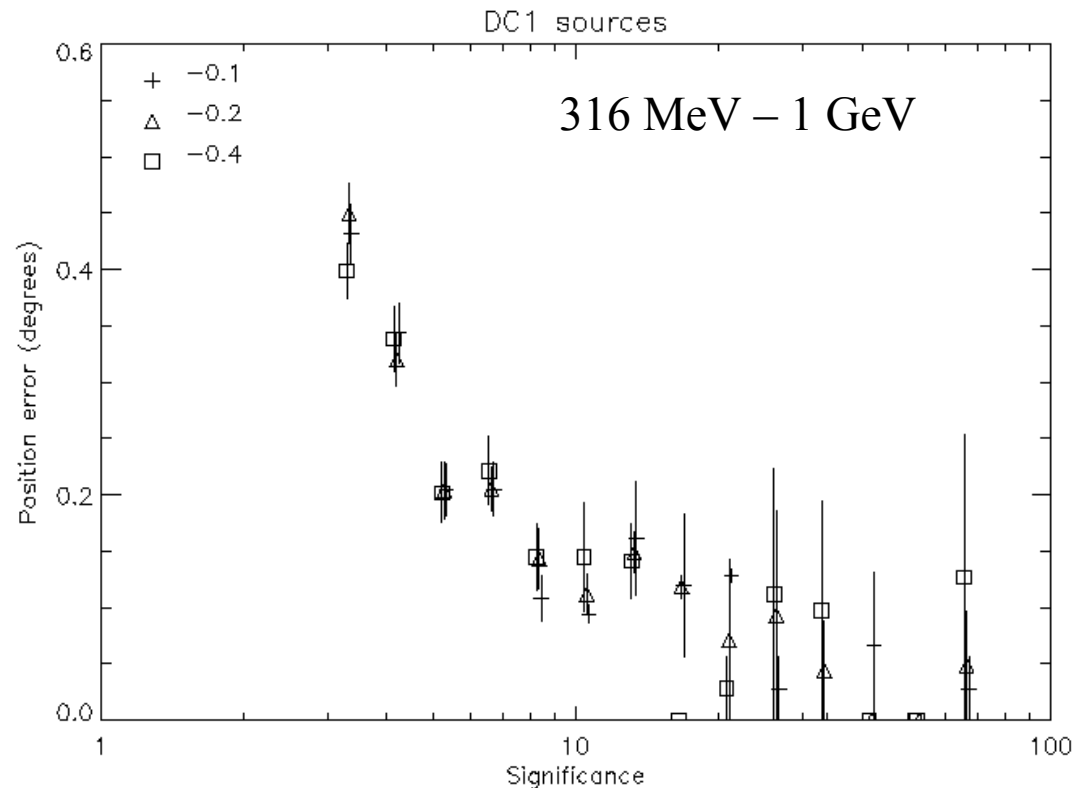
Example for optimal filter method as a function of bin size.

PSF width in that band is about 0.67° (at 30% maximum).

This kind of diagram (as a function of significance) gives a reasonable value for the position error, if the algorithm does not estimate it source by source.

TS maps from Likelihood can be used as input instead. Is it significantly better (much more CPU intensive) ?

Bin size about 0.3 PSF width is fine (for detectability also). 0.05° above 1 GeV



Tools required

Either used in pipeline itself, or as a help to adjust parameters

- ✓ Tool to provide an all-sky exposure map ($\text{cm}^2 \text{ s}$) averaged over any energy band (power-law weights) as a FITS image, from the pointing history and effective area calibration, accounting for any cuts on the data (like cut on off-axis angle). This now exists in the Science Tools.
- ✓ Tool to provide the point spread function averaged over off-axis angle and over any energy band (power-law weights) as a FITS image, both at a given position (using the pointing history) and on average over the sky (assuming off-axis angle is homogeneously covered).

Source catalogue generation

Catalogue generation is on the way !

Several open points:

1. Is cartesian geometry all right (paving the sky with 3 or more large pieces) ? Should we investigate convolution in spherical geometry ?
2. How to deal best with the energy information ?
3. How should we set the detection threshold ? Low enough and let likelihood reject the false detections, or high enough and use likelihood for characterisation only ?
4. Should we implement additional cuts on the data (e.g. on off-axis angle) ? Is source elongation off-axis an issue ?
5. How best to detect variable sources ?
6. Extended sources. Not an immediate priority.