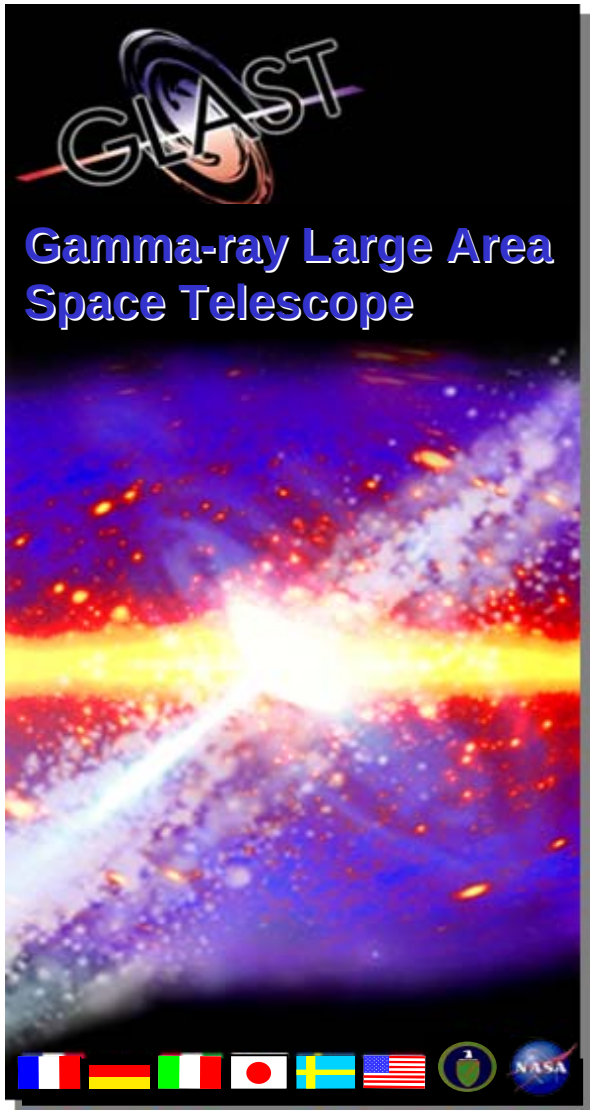


Diffuse + Catalog Session

Source Detection Performance

Analysis by J. Ballet (CEA Saclay), presented by S. Digel (SLAC)

Introduction: Studies of source detection algorithms for LAT data

- This is a follow-up to presentations at successive collaboration meetings
  - For details see Jean's posting at https://confluence.slac.stanford.edu/display/SCIGRPS/Improved+source+detection+on+obssim2
- Reminder: *Why do we want to investigate source detection algorithms?*
  - The catalog pipeline needs an algorithm for the source detection step
  - And for Automated Science Processing (Quick Look)
- Intuitively, a fast algorithm for source detection ought to be possible – either parametric (ie., it knows what the PSF is and/or has a model for diffuse emission) or more likely not
  - Likelihood analysis intrinsically is model fitting – it won't answer any questions that you are not asking, so source searches are brute force

Source detection methods under study

Currently used in the Catalog pipeline

- **MRF – Multi Resolution Filter (Ballet) based on application of MR_FILTER (developed by J.-L. Starck, Saclay)**
  - **Wavelet algorithm for de-noising images**
  - **Important details of its application for LAT source detection include running on separate bands and merging the results**
  - **Source extraction from filtered image is done with SExtractor***
- **PGW (Tosti) – PGWave**
  - **End-to-end source detection algorithm that uses FT1 & FT2 files**
  - **Background images are estimated by smoothing the input – filtering is in wavelet space, and the source extraction from the filtered images is via a custom algorithm**
- **UW – Binned likelihood for isolated point source on uniform background* (Burnett)**
  - **Works on 8 successively finer grids of HEALPix, with finer gridding used for higher energies; obviously does not accept images as input. Detections are merged across bands**
  - **An overall significance TS; fluxes are not provided – but could be calculated**
  - *** Now iterates once – adds bright sources to the 'background'**

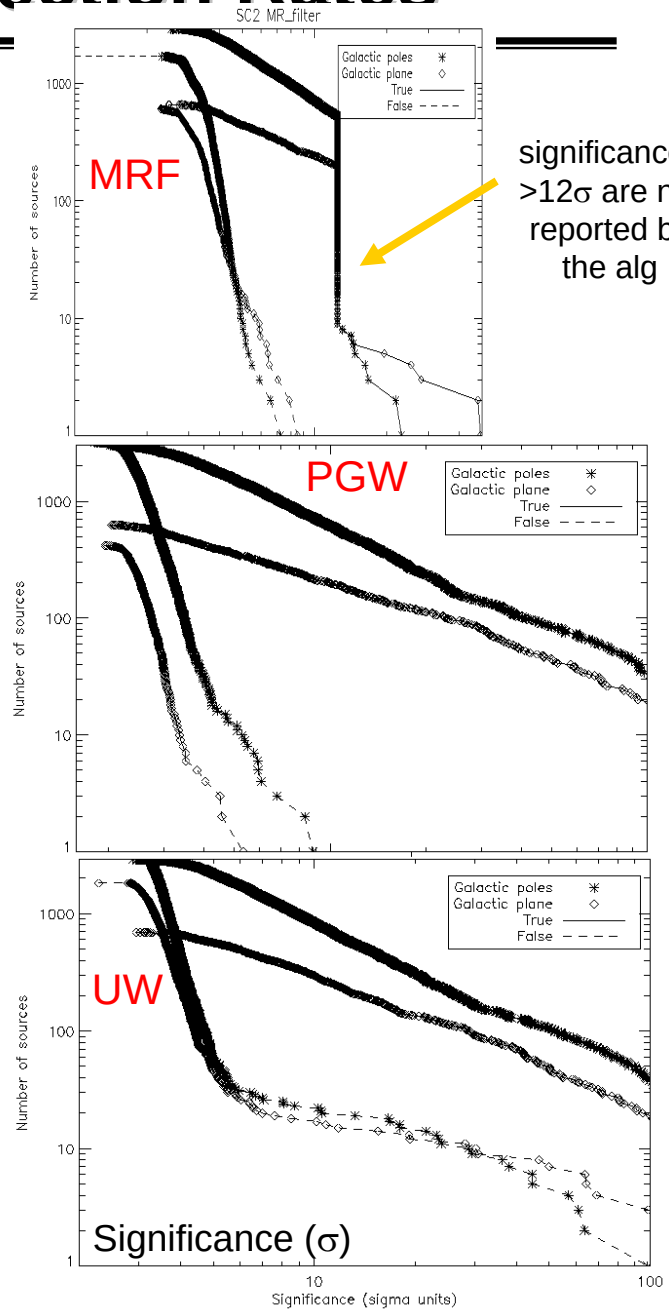
* Widely used in astronomical image analysis (http://terapix.iap.fr/rubrique.php?id_rubrique=91/)

Status of methods

- Each has been under continued refinement; updates presented at Catalog EVO meetings have helped each of the algorithms advance
- Getting beyond the test pattern studies, the algorithms are now generally compared against the obssim2 [1-year gtobssim simulation] data
  - The true source list has been (painstakingly) constructed by Toby
  - The detectability of each has also been evaluated by Jean in the Catalog pipeline Likelihood analysis
  - Down side is no residual backgrounds
- Here is an update on the obssim2 comparison for the current versions of the algorithms

Numbers and False Detection Rates

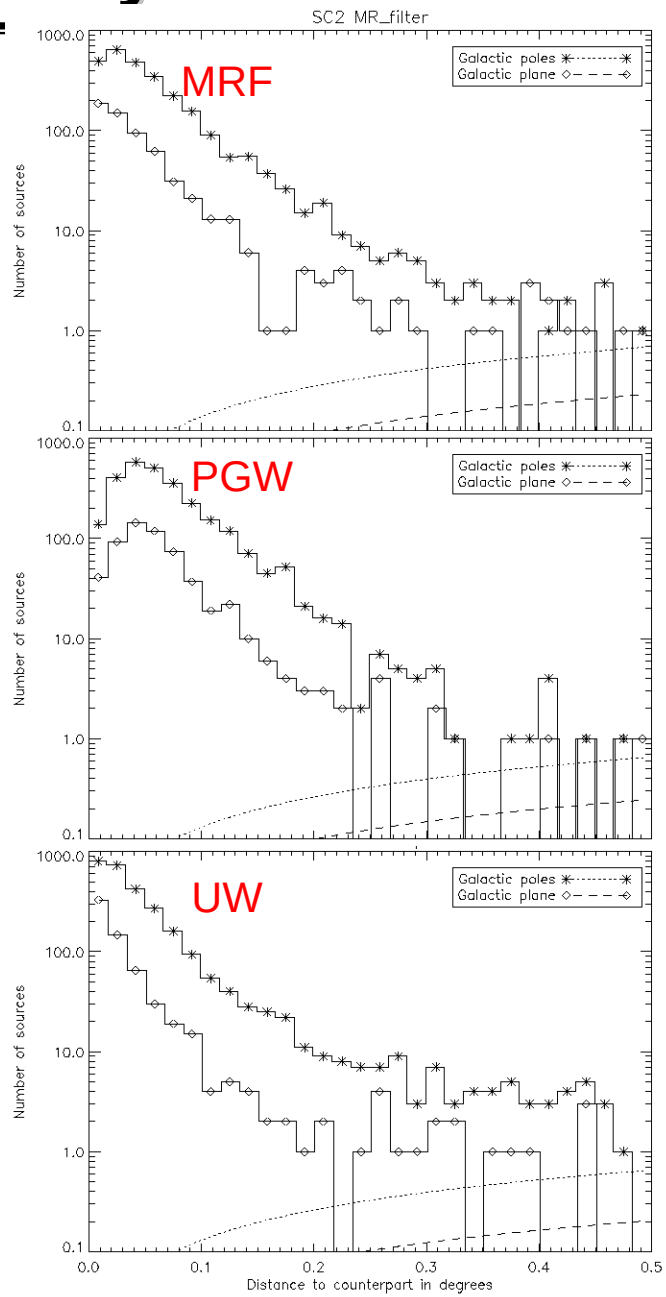
- **Cumulative distributions of True and spurious source detections vs. source significance**
- **Spurious is defined by 0.5° matching radius only**
- **Top curve (stars) is high latitudes (|*b*| > 10°), bottom (diamonds) is low latitudes**
- **Numbers sources are comparable among the algorithms (see later)**
- **UW algorithm has larger false detection rate (not a problem for catalog analysis)**



significances >12σ are not reported by the alg

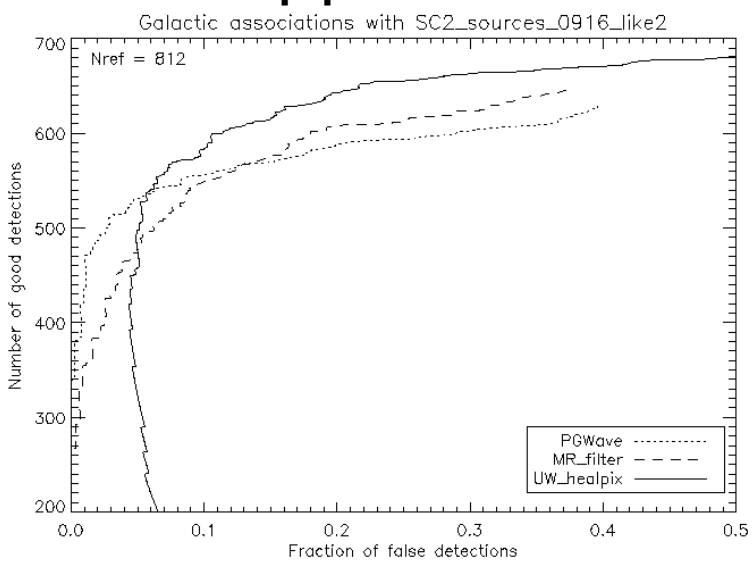
Positional Accuracy

- **<0.5° match with an obssim2 source was used here; TS > 25 required for the obssim2 sources**
  - **Curves are for low and high latitude sources**
  - **Expected numbers of chance associations are shown**
- **PGW is limited here by the pixel size in the wavelet analysis; MRF somewhat less so**
- **UW has the most accurate positions**

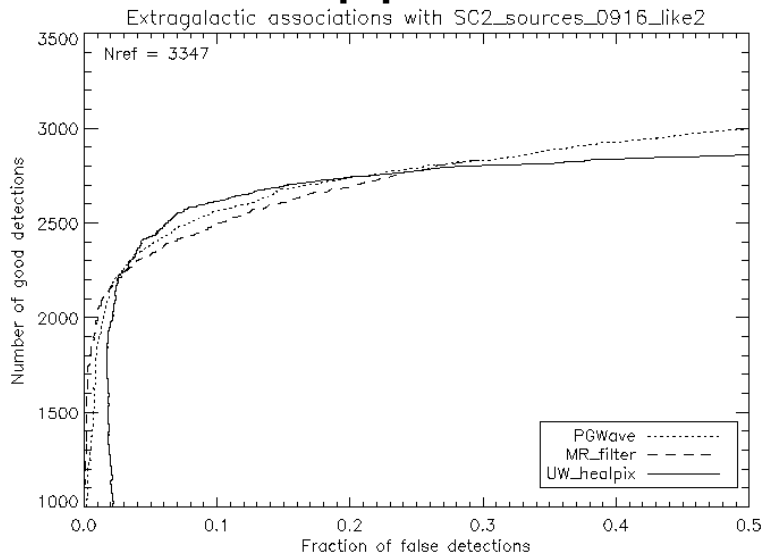


True detections vs. FDR

|*b*| < 10°



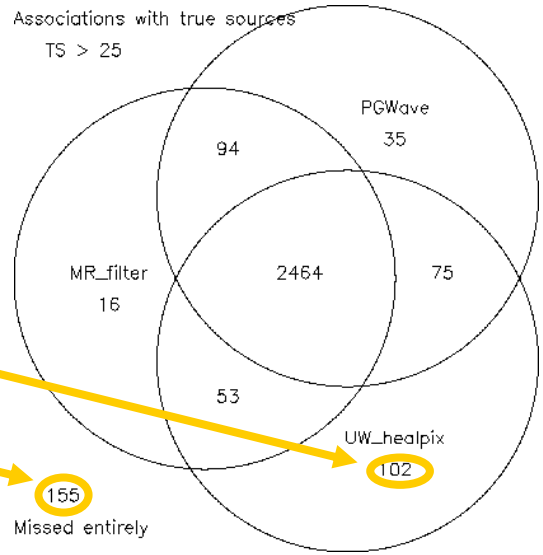
|*b*| > 10°



- **A fairly high False Detection Rate is not a problem in the Catalog pipeline; here a limit of TS = 10 is used for comparison with obssim2 sources**
- **UW has a distinct advantage at low latitudes; performances are comparable at high latitudes, although PGW has advantage at high FDR**

Sources found by 0, 1, or >1 method

- **UW finds more sources 'uniquely' than the other methods; the improvement appears to be due to the iterative procedure in the UW algorithm for finding fainter sources near brighter ones**
- **The significant sources that are missed entirely tend to be faint and soft**



Current Status*

- The UW algorithm has a measurable performance edge, especially at low latitudes
- Jean is currently running a check that the Catalog Pipeline can reject the high-significance false positives that the UW algorithm finds
- If that does not turn out to be a stumbling block, UW will be implemented in the Pipeline

* or ‘Plan Forward’