DC1 lesson learned

Francesco Longo
University and INFN, Trieste, Italy
francesco.longo@ts.infn.it

Outline

- DC1 purposes and scope
- DC1 preparation
- LAT data analysis needs
  - Background rejection and OnBoardFilter
  - Gamma-ray Analysis
- DC1 data
- Science Tools for DC1
  - Architecture
  - Distribution
  - Documentation
- DC1 “community”
  - The Wiki page
  - DC1 mailing list
- What’s next?
  - Analysis of DC1 data

Purposes of the Data Challenges S.Ritz

- "End-to-end" testing of analysis software.
- Familiarize team with data content, formats, tools and realistic details of analysis issues (both instrumental and astrophysical).
- If needed, develop additional methods for analyzing LAT data, encouraging alternatives that fit within the existing framework.
- Provide feedback to the SAS group on what works and what is missing from the data formats and tools.
- Uncover systematic effects in reconstruction and analysis.

Support readiness by launch time to do all first-year science.

Lessons from Data Challenge 1

- "End-to-end" testing of analysis software.
  - First usage of "key tools" (Likelihood, GRB tools)
- Familiarize team with data content, formats, tools and realistic details of analysis issues (both instrumental and astrophysical).
  - FITS, FTOOLS stuff, Exposure, TS maps, Binning, Spectral Analysis, Count rate triggers …
  - Background rejection and StdCut events
  - Galactic Plane Modeling
- If needed, develop additional methods for analyzing LAT data, encouraging alternatives that fit within the existing framework.
  - Tools development
- Provide feedback to the SAS group on what works and what is missing from the data formats and tools.
  - User feedback
- Uncover systematic effects in reconstruction and analysis.
  - DC1 preparation

First attempt to do LAT science for many users

Data Challenge Planning Approach S.Ritz

- Walk before running: design a progression of studies.
- DC1. Modest goals. Contains most essential features of a data challenge.
  - 1 simulated day all-sky survey simulation
    - find GRB
    - a few physics surprises
    - exercise:
      - exposure, orbit/attitude handling, data processing pipeline components, analysis tools
- DC2, start end of CY04. More ambitious goals. Encourage further development, based on lessons from DC1. One simulated month.
- DC3. Support for flight science production.

What's been done: preparatory work S.Ritz

- Very large effort during the past ~9 months by many people.
- Instrument analysis:
  - done previously with earlier tools for AO, PDR, etc., demonstrating LAT meets requirements.
  - Now done again with new tools. More to do, but more than adequate for DC1.
- Fluxes
- Data formats, processing
- Science tools

Already a great success!

DC1

R.Dubois

- **Focal point for many threads**
  - **Orbit, rocking, celestial coordinates, pointing history**
  - **Plausible model of the sky**
  - **Background rejection and event selection**
  - **Instrument Response Functions**
  - **Data formats for input to high level tools**
  - **First look at major science tools – Likelihood, Observation Simulator**
  - **Generation of datasets**
  - **Populate and exercise data server at SSC & LAT**
  - **Code distribution on windows and linux**
- **Involve new users**
- **Teamwork!**

Backgrounds

S.Ritz

- **DC1 is an approximate modeling of one day of LAT data.**
  - **at face value, this means 400M background triggers (4kHz). [This is not the number generated, which is larger, since many miss the instrument.]**
  - **using an updated version of Bill Atwood's background rejection analysis shown in Rome, residual contamination of photon sample would be about ~6%.**
- **One-day science is generally NOT background limited. Several purposes to generating background for DC1:**
  - **exercise the machine, find the problems (already done!)**
  - **generate amounts of background needed anyway to complete the analysis.**
- **We therefore decided to unhook the background generation for DC1**
  - **Rejection analysis already at a sufficient level to estimate instrument performance for gammas (Aeff).**
  - **Used these cuts on the photon sample for DC1. Provides a good description of impacts of background rejection.**
    - at normal incidence, Aeff asymptotes to 10,000 cm^2. At 100 MeV, ~4,500 cm^2
    - small fall-off in area for E>10 GeV, will be improved soon. Not a background rejection issue.
  - **Already at a sufficient level to make background a non-issue for DC1 science analyses.**
  - **This allows the background rejection analysis to proceed at its own pace.**
- **At end of DC1, both background rejection and signal analyses will be completed to the levels planned.**

Std Cut files

- Why from dataServers not the stdCut file as default and the other data only as option?
- Use as default the events that passed the background, PSF, and energy resolution filters (i.e., the events for which the DC1 response functions apply)
- More kinds of response functions (e.g., a set that applies to the events that don't pass the PSF filter), but even so the event flags should be accessible in the data servers.
- Bias on Analysis!

DC1 lesson (already) learned

Cf S.Ritz

- Lots of hard work by many people on the machinery up until the last minute.
  - Richard's talk
- Sky model is fairly rich and accurate. Some details could be better
  - no flaring sources implemented (though some variable sources might have day-long fluxes different from your expectation…)
  - 2 GRB "models" available
- No (intentional) hardware problems implemented.
  - decided to postpone to DC2, when ISOC is up and running
- No onboard filter in data path yet
  - similar to background rejection situation. The incremental loss of area (after other cuts) is now expected to be very small.
- Instrument response functions are not really mature.
  - some problems still. certainly good enough for DC1 science goals, however!

DC1 data

- FITS versus ROOT data
- Spacecraft pointing info only in FITS?
- Pointing history available as .dat or as separate tree in ROOT file?
- ROOT data server?

DC1 Minimum Results

S.Ritz

- The existence of the data sets and the volume of data generated for background analyses already meets one of the success criteria.
- A minimum set of plots and tables that we must collectively produce:
  - TABLE 1: found sources, ranked by flux (E>100 MeV). Table has the following columns
    - reconstructed location and error circle
    - flux (E>100 MeV) and error
    - significance
    - 3EG identification (yes or no) [note: DON'T assume DC1 sky is the 3EG catalog!]
    - extra credit:
      - » include flux below 100 MeV
      - » spectral indices of brightest sources
      - » comparison of 3EG position and flux characteristics with GLAST analysis
  - FIGURE 1: LogN-logs plot of TABLE1
  - TABLE 2: list of transients detected. Columns are
    - location and error circle
    - flux (E>100 MeV) and error
    - significance
    - duration
  - FIGURE 2: light curve
    - Extra credit: FIGURE 2a: spectra.
  - PLUS: reports of any physics surprises found.

MC Truth unveiled

- Scripting for Catalog generation
- How to analyze new sources? E.g. the Galactic Center
- Connection among different tools e.g. blind searching tool and likelihood analysis
- Take into account instrument behaviour in different ranges
- Success of GRB trigger algorithms

Source Detection

- Many methods developed (wavelets, voronoi tesselation..)
- Need compare the results
- Likelihood result on blind search detected sources
- Generation of catalog with different methods significance?
- Study of Interstellar emission

GRB analysis

- Lot of work in Italy and US
- 5 different trigger criteria
- Need a comparison
- Need to study onboard LAT trigger performances (e.g. for on ground follow up)
- Need to study weaker GRB near the detection limit and optimize trigger with realistic background
- Dividing the sky in several spatial bin is an effective way to reduce background
- Floating threshold and trigger window selections (time, events, …) to be refined with variable background rates
- On board buffer with localization and timing info available for unbinned search on board
- Effects of CR on localization accuracy could be solved by weighting the events by their distance to accumulated centroid
- Need refinements in Spectral analysis

Beyond the minimum

S.Ritz

- Here are a few suggestions:
  - we may generate and release more days of data
    - better exercise tools and infrastructure
    - more transients
  - spectral analyses
  - localization studies
    - one-day localization of Vela is particularly interesting
  - analysis improvements

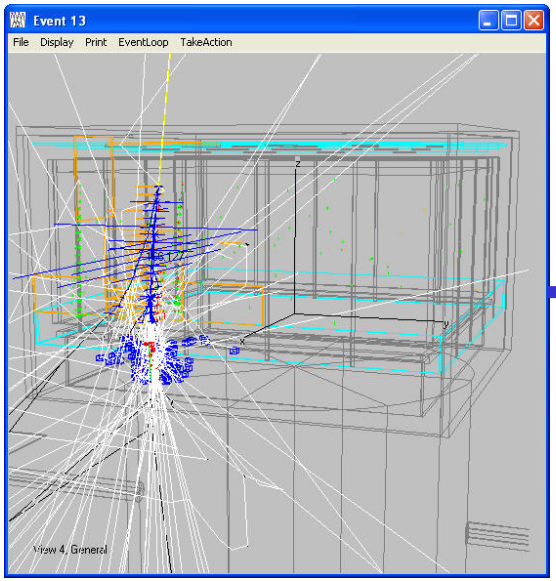
- But don't let this list limit you. The sky is the limit!

Beyond minimum results

- Source detection methods
- New analysis methods and/or languages
- Little usage of Science Tools?

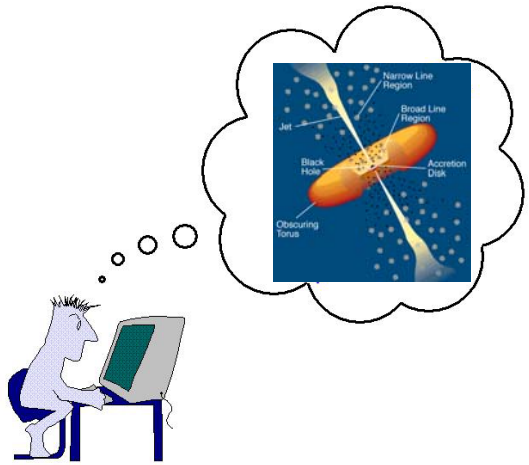
High-Level Analysis

S.Digel and P.Nolan



(T. Usher)

Science Tools

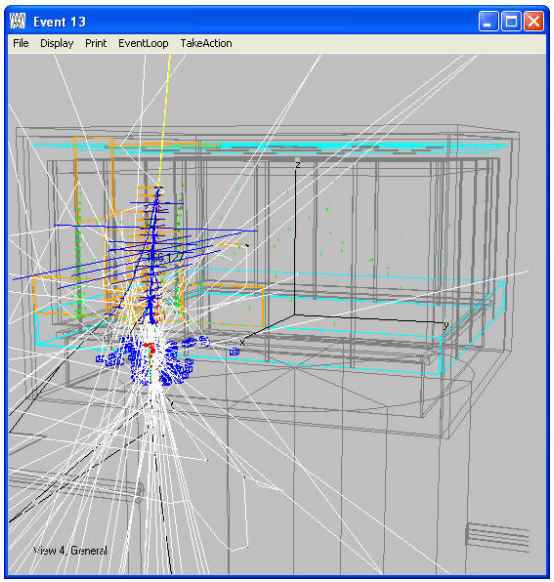


(Apologies to R. Dubois, D. Flath, M. Urry, P. Padovani)

Seth Digel & Patrick Nolan

HEPL/Stanford Univ.

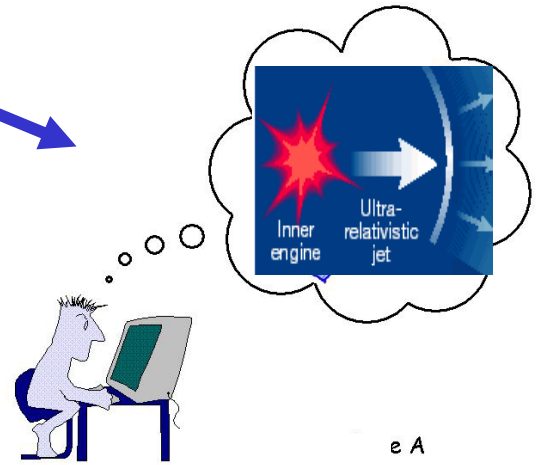
DC1 Analysis



(T. Usher)

Science Tools

Other Tools(?)



(Apologies to R. Dubois, D. Flath, R.Roy Britt)

DC1 Tools

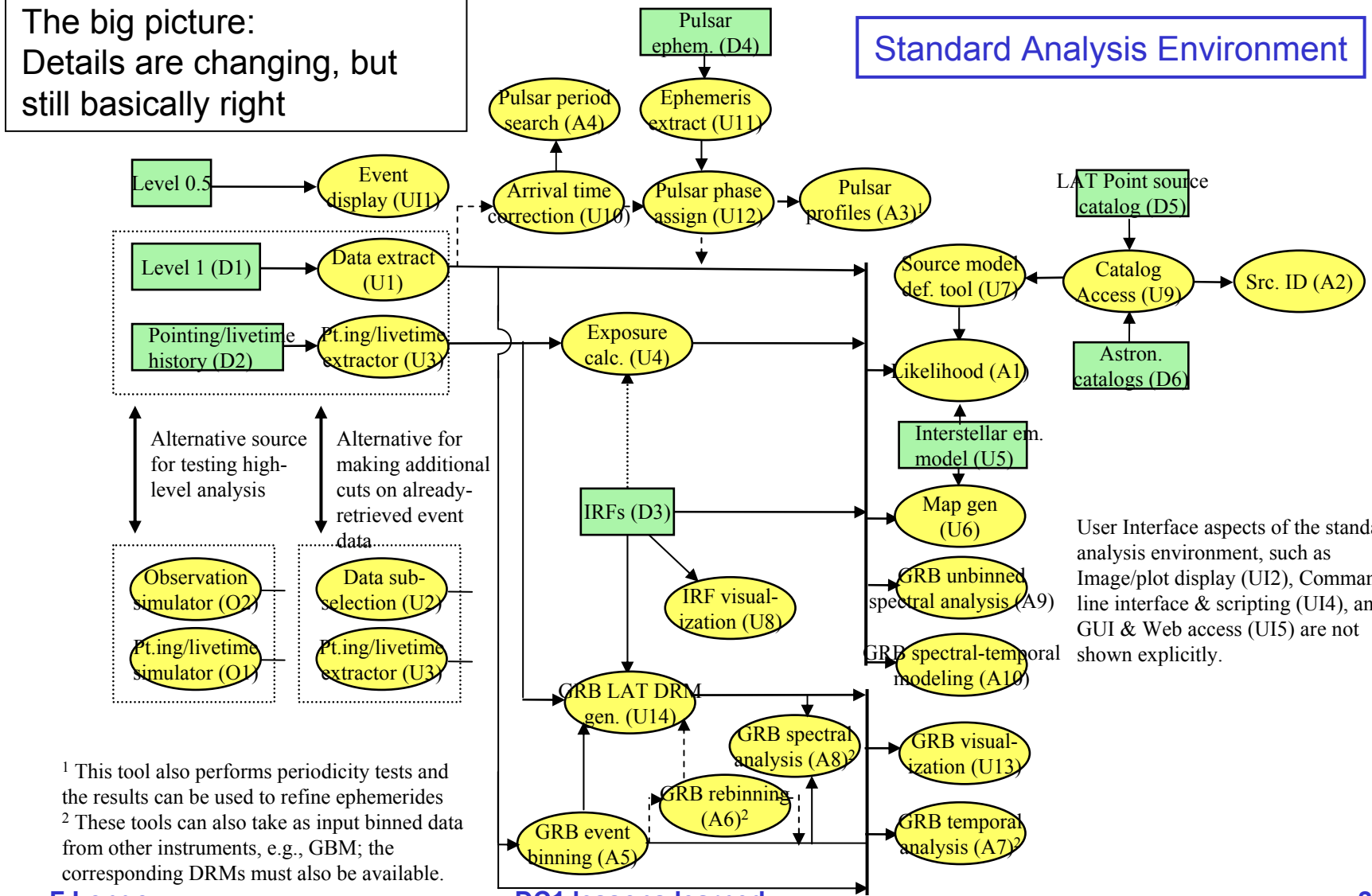
S.Digel and P.Nolan

- **Data Challenge 1 is not about science**
  - **The emphasis is on the analysis techniques themselves and the tools to implement them**
- **What are the DC1 goals for the science tools?**
  - **Briefly, an end-to-end test of the system, going back to instrument simulation, event classification and generation of response functions, through high-level processing, and managing the flow of data in between**
  - **So we are at the high-level processing end of the chain now (more or less – will be some iteration, e.g., on event classes). We want to**
    - **shake test the science tools; actually they won't take a lot right now**
    - **introduce the analysis methods for gamma-ray astronomy with the LAT**
    - **get feedback on functionality – from users and from developers**

S.Digel and P.Nolan

Science Tools in ~~DC1~~ DC3

The big picture:
Details are changing, but still basically right



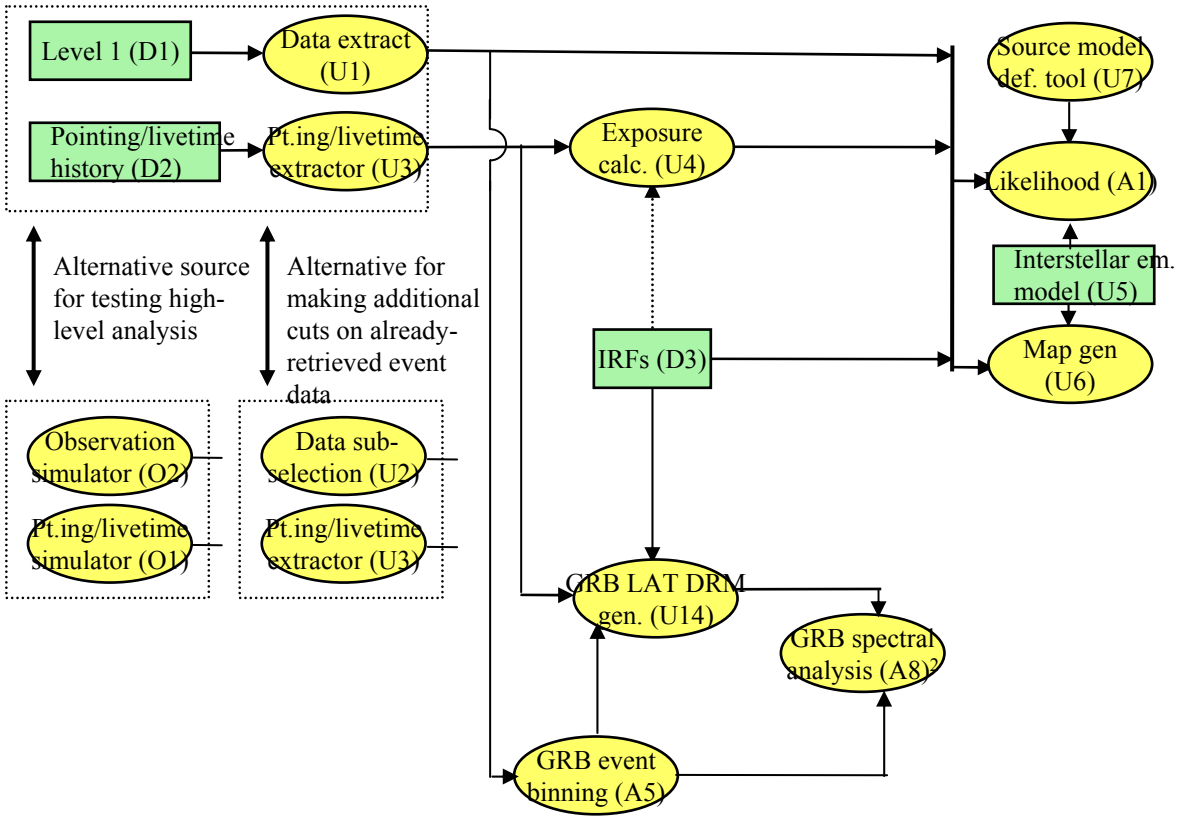
User Interface aspects of the standard analysis environment, such as Image/plot display (UI2), Command line interface & scripting (UI4), and GUI & Web access (UI5) are not shown explicitly.

[1] This tool also performs periodicity tests and the results can be used to refine ephemerides
[2] These tools can also take as input binned data from other instruments, e.g., GBM; the corresponding DRMs must also be available.

S.Digel and P.Nolan

Science Tools in DC1

- **All components are still prototypes**



The DC1 functionality is
Data extraction
Limited visualization
Model definition
Model fitting
Observation simulation

Implementation of Science Tools

S.Digel and P.Nolan

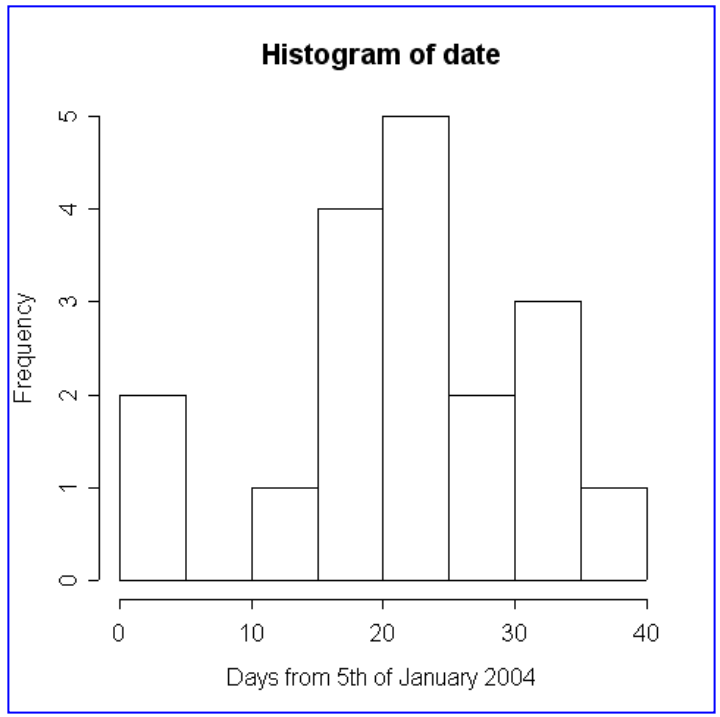
- Reminder: the tools are implemented as FTOOLS
  - HEASARC convention across missions
- You will notice that this defines much of the 'look and feel'
  - Provides a uniform interface
  - HOOPS for prompting at the command line
  - GUI is coming
- FITS files for data

Prototype Science Tools in DC1

- Many tools planned to be included in SAE not yet available
- GRB Tools
  - Trigger methods
  - GRB visualisation
  - GRB spectral analysis
  - GRB temporal analysis
- Exposure and Map generation
- Alternative source detection methods
- ROOT tools

DC1 communication issues

- Usage of Wiki page
  - a “cascade” model
- DC1 mailing list
  - ScienceTools usage mailing lists?
- cvs repository for DC1 tools?
  - Use the /users/ tree



Wiki usage (Riccardo)

DC1 feedback

- Astrophysical data analysis
- Software usage and reliability
- Documentation
- Data access and data server usage
- UI stuff
- Software installation and release
- Software infrastructure & framework
- Communication and Time frame

DC1 feedback

- **Analysis issues**
  - First analysis on Astro data
  - Transients:
    - Notification for deep observation
    - Not requested to be run by every user
    - Need of a Quick Look tool for DC2 study
  - Likelihood tool:
    - How to check fit quality?
    - Which statistical minimum for reasonable results?
    - Help from experienced users
  - In some cases incorrect FITS headers
  - Definition of Galactic diffuse emission (model, units …)
  - SLAC installation of common Astro tools (FTOOLS, ds9, XSPEC..)

DC1 feedback

- **Software:**
  - Clear documentation for most tools
  - More files greater the possibility to do mistakes (could be FT2 information embedded in FT1?)
  - FT1 data and Exposure in different directions tests?
  - Need for Sanity checking on the inputs. (e.g. Likelihood tool with 'BACK' response functions without event filtering)
  - TsMap crashes
  - Different optimizers different results?
  - Most self-contained not help for developing own tools that fit within framework (e.g. read/visualize data, manipulate multi-dimensional arrays, get background model density at a given point, access the IRFs etc etc). Need more documentation on existing “base” tools
  - How to generate different Energy Bin fits file?
  - Plot tools within the SAE and exposure maps generation missing
  - ROOT format for exposure map or more general exposure tool
  - Coordinates specification (e.g. for modeldef)
  - CPU time

DC1 feedback

- **Software reliability**
  - **Statistical interpretation of Likelihood**
  - **Need for comparison among different analysis**

DC1 feedback

- **Documentation:**
  - Wiki pages good idea
  - Fine documentation of the science tools (for Linux users?)
  - Not updated to new versions
  - Need for dynamic documentation (e.g. galactic diffuse for obsSim)
  - Weekly status report?
  - Needs for tips on developing own models
  - Info on tool contact and updates on each web page

DC1 feedback

- **Access to Data:**
  - **Some failures**
  - **Documentation on SSC better than SLAC data server (e.g. time selection units?)**
  - **Cut on tree variables directly?**
  - **Full data set available not from DataServers**

DC1 feedback

- **Coherent User Interface**
  - Who used the Likelihood GUI?
  - At least command line with arguments from arbitrary directory
  - Unique settings startup command for analysis session
  - Simple usage explanation (e.g. “– help”)
  - Tab completion or command history available
  - User should not worry about data and executable location
  - High priority for DC2 for ST to be distributed to the general community.

DC1 feedback

- **Software Installation**
  - Binary distributions (v1r0p1) very nice but which Linux version? (7.3 vs 9.0?)
  - New releases not publicly distributed as binary… Only "core" developers
  - Necessity for knowing how to include new features in newer version of packages
  - Windows vs Linux compilation and distribution

DC1 feedback

- **Software infrastructure**
  - **Bookkeeping Info (events N, area, Energy spectrum parameters, energy range for events)**
  - **Messages from ScienceTools s/w (Info, Warning, etc)**
    - **Level control, name of algorithm, service, ..**
    - **Std message format**
  - **Setup of environment by automated way .(c)sh scripts not easily maintainable**
  - **PFILES definition for unique parameter file directory**
  - **Coding rules for avoiding compilation warnings**

DC1 feedback

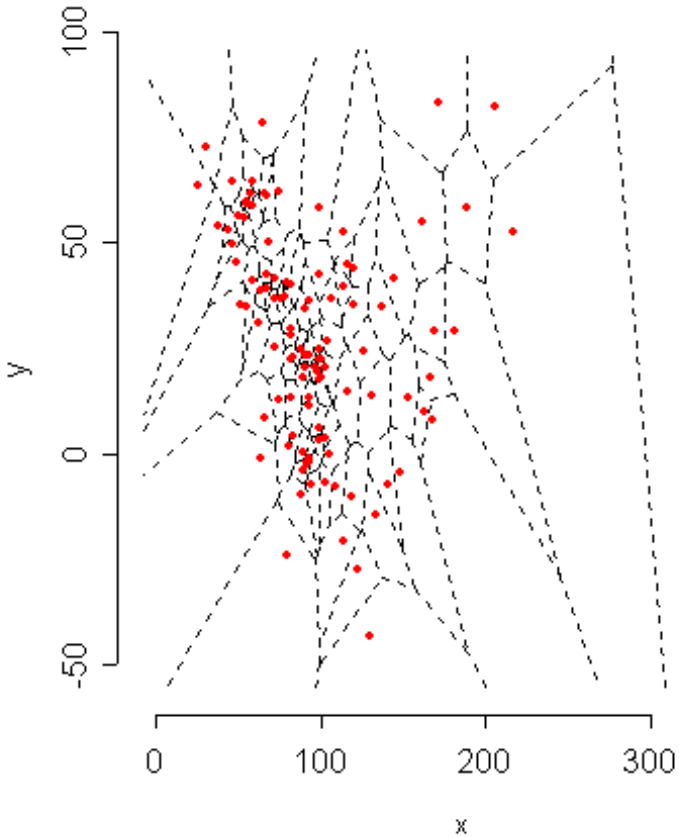
- **Framework**
  - **FTOOLS structure for distributed software**
  - **How to extend or use other tools?**
  - **How to include new tools?**

DC1 feedback

- **General Things:**
  - **Communication issues: (FT1Energy fix, updates to data servers, new ST releases?)**
  - **Before holiday work caused net loss of time**

Proposal

- **Science Tools user meetings?**
- **Continue to analyse DC1 data before DC2!**



Voronoi tesselation of GRB

Example of usefulness of communication

Conclusions

- DC1 a success!
- First analysis of astrophysical data for many of us
- Science Tools developments
- Work on Background and Filter
- DC1 data generation and storage
- Team collaboration

Acknowledgements

- Thanks to those who sent me comments
- Steve, Julie, Dirk, Toby, Jay, Jerry, Traudl, Seth, Benoit, Nicola, Gino, Claudia, Riccardo, Alessandro, Michael, Francesca, Monica, Luca

Science Tools in DC1

S.Digel and P.Nolan

- Details of the contents of the event summaries are still converging
- The interstellar emission model is still the model used by EGRET team (Bertsch et al. 1993, Hunter et al. 1997)
- The instrument response functions are defined only for the events that pass the filters presented by Bill Atwood at the collaboration meeting in Rome
  - Eventually we expect to have more than one event class
  - PSF and energy resolution are being defined on a grid of energies and inclination angles, with a analytic function fit. No interpolation of the parameters between grid points is attempted
  - The IRFs are not yet in CALDB, although this switch should be completely transparent to the user
- The interface to EGRET data and pointing/livetime history is not complete (and was not planned for DC1)

Science Tools in DC1

S.Digel and P.Nolan

- The DRM generator RspGen understands only circular cutout regions so far; this is not a limitation at all for analyzing bright GRBs, but the intent is to make it understand custom shapes for crowded fields
- The map generation tool does not exist yet
  - EventBin (GRB event binning) can make counts maps
  - Exposure maps can be generated with [what]
- No visualization is integrated with the tools yet
- The orbit and attitude simulation is still idealized and not yet a standalone tool