P130 Reprocessing

P130 Reprocessing

status: **Complete**
last update: 6 Aug 2012

This page is a record of the configuration for the P130 reprocessing project, recalculating the diffuse response (see here for a description of the change). This task is identical with P120-FT1, v2.1 [please see for details], run in July 2011, except for the following details.

1. New version of ScienceTools (09-26-02) which includes new integration scheme for gtdiffrsp
2. New run list (updated for data processed since the end of P120)
3. Name of Pipeline task (P120->P130) as well as in file naming

Data Sample

The data sample for P130 reprocessing includes:

First run	239557414 (MET), 2008-08-04 15:43:34 (UTC)	beginning of Science
Last run	356429251 (MET), 2012-04-18 08:07:29 (UTC)	
Total runs	20,385	
Total input MERIT events	44,398,334,636	all "events"
Total disk space used	33.9 TB	

Summary from DataCatalog as of 4/17/2012 and updated 8/6/2012 (after new evt selection filter).

Name	Files	Events	Size
FT1	20385	184,008,254	16.6 GB
LS1	20385	1,295,164,924	210.5 GB
ELECTRONFT1	20385	85,057,599	8.0 GB
EXTENDEDFT1	20385	~~11,240,869,507~~ 6,280,448,356	~~984.6~~ 550.3GB
EXTENDEDLS1	20385	~~11,240,869,507~~ 6,280,448,356	~~1.8~~ 1.0 TB

Bookkeeping

1. (This page): Define ingredients of reprocessing (processing code/configuration changes)
2. Processing History database: http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P130
    a. List of all reprocessings
    b. List of all data runs reprocessed
    c. Pointers to all input data files (-> dataCatalog)
    d. Pointers to associated task processes (-> Pipeline II status)
3. Data Catalog database: http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp
    a. Lists of and pointers to all output data files
    b. Meta data associated with each output data product

P130-FT1

This task generates all desired FITS data products. An example of the code processing chain appears on a child page.

Status chronology

- 1/26/2012 - Initial task configured, 1000 runs reprocessed. On hold for validation.
- 2/1/2012 - Okay to continue. Begin block 1 reprocessing, containing 19133 runs:

	MET (sec)	MJD(days)	UNX(sec)	UTC
1st run	333886531	55774.427442129629	1312193731.000	2011-08-01 10:15:29
last run	349266293	55952.433946759258	1327573493.000	2012-01-26 10:24:51

- 2/3/2012 - Block 1 reprocessing complete
- 4/2/2012 - Begin block 2, raising the total # runs to 20126, through run 354926369 2012-03-31 22:39:29 (UTC)
- 4/18/2012 - Final reprocessing block: additional 259 runs through 356429251, 2012-04-18 08:07:29 (UTC)
- 4/19/2012 - Complete.
- 5/1/2012 - Reconfigure start time - reducing by 1.9 seconds - for run 245403855 due to old leap second problem - to regain 27 events at the start of the run. This is stream 1019.
- 8/3/2012 - A request to update the event selection filter on data in the period 1 Aug 2011 and 19 Apr 2012 was made to bring EXTENDED FITS and FILTEREDMERIT files into synch. See this page for details. The P130-FT1 task has been modified to run a new event filter; mergeClumps job steps for the following runs rolled back:

	Run	Task Stream
Start	333886531	16460
End	356429251	20384

  This action will recreate **only** the EXTENDEDFT1 and EXTENDEDLS1 data products. These will automatically be ingested by the astroserver. Note that the file original-config.py contains the former configuration for this task, while config.py contains the changes discussed above.
- 8/6/2012 - Re-reprocessing complete

Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P130-FT1
Task Status	http://glast-ground.slac.stanford.edu/Pipeline-II/task.jsp?task=74765664
Input Data	MERIT from P120-MERIT (runs 239557414 - 333880535) and Level 1 (runs 333886531 - present)
spacecraft data	FT2 from P105 (runs 239557414 - 271844560), then from current Level 1 production
Input Run List	ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P130-FT1/config/runFile.txt
Reprocessing Mode	reFT1
meritFilter	FT1EventClass!=0 for P120-reprocessing, and (FT1EventClass& 0x00003EFF)!=0 for subsequent data
evtClassDefs	00-19-01
eventClassMap	EvtClassDefs_P7V6.xml
ScienceTools	09-26-02
Code Variants	redhat5-i686-32bit-gcc41 (Optimized)
Diffuse Model	based on contents of /afs/slac.stanford.edu/g/glast/ground/GLAST_EXT/diffuseModels/v2r0 (see https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7)
Diffuse Response	'source' using P7SOURCE_V6 IRF 'clean' using P7CLEAN_V6 IRF
IRFs	P7V6, contained within ScienceTools release
Output Data Products	FT1, LS1, EXTENDEDFT1, EXTENDEDLS1, ELECTRONFT1

Processing chain for FITS data products

Data Product	selection	makeFT1	gtdiffrsp	gtmktime	gtltcube
FT1 (for FSSC)	'source' and above EVENT_CLASS bits 2,3,4	⊕	⊕	⊕	⊖
LS1 (for FSSC)	'transient' and above EVENT_CLASS bits 0,2,3,4	⊕	⊖	⊕	⊖
FT1EXTENDED	FT1EventClass!=0	⊕	⊕	⊕	⊖
LS1EXTENDED	FT1EventClass!=0	⊕	⊖	⊕	⊖
ELECTRONFT1	CTBParticleType==1	⊕	⊖	⊕	⊖

Note that diffuse response is calculated for 'source' and 'clean' event classes only.

Note on 'Code Variant': The SLAC batch farm contains a mixture of architectures , both hardware (Intel/AMD 64-bit) and software (RHEL5-64, gcc v4.1, etc.).

Timing and Performance

With a sample of the first 8128 runs, the CPU-intensive part of the task (mergeClumps) is consuming ~42 CPU-min/job (differences in machine class are small). TrickleStream configured to allow 1500 simultaneous mergeClumps jobs, and up to 150 job starts every 200 second cycle.