

# P120 reprocessing

## P120 Reprocessing

status: **Complete**

last update: 17 August 2011

This page is a record of the configuration for the P120 reprocessing project, event reclassification using Pass 7.3/7.4/7.6. This project involves reprocessing with [Pass7 classification trees](#) and (ultimately) new IRFs. The name "P120" derives from the word "processing" and the **initial** file version to be used for the output data products, e.g., r0123456789\_v120\_merit.root.

- [P120-MERIT](#) - this [task](#) reads DIGI+RECON+MERIT and produces reprocessed MERIT + FILTEREDMERIT (photons) + ELECTRONMERIT
- [P120-FT1](#)
  - [V1.0](#) of this task reads MERIT and produces FT1 (photons) + LS1 (merit-like FITS file for photons) + electron FITS file
  - [V2.1](#) of this task is modified such that FT1 and LS1 files are filtered for FSSC, while new EXTENDEDFT1 and EXTENDEDLS1 files are produced containing all photon event classes.
- [P120-LEO-MERIT](#) - this [task](#) reads DIGI+RECON+MERIT and produces reprocessed MERIT + FILTEREDMERIT (photons) + ELECTRONMERIT for 200 runs of earth limb (L&EO) data

## Datafile names, versions and locations

Data file version numbers for this reprocessing will begin with v120.

### XROOT location and file naming

Location template:

```
/glast/Data/Flight/Reprocess/<reprocessName>/<dataType>
```

Locations for P120:

```
/glast/Data/Flight/Reprocess/P120/merit
/glast/Data/Flight/Reprocess/P120/filteredmerit
/glast/Data/Flight/Reprocess/P120/electronmerit
/glast/Data/Flight/Reprocess/P120/ft1
/glast/Data/Flight/Reprocess/P120/extendedft1
/glast/Data/Flight/Reprocess/P120/electronft1
/glast/Data/Flight/Reprocess/P120/ls1
/glast/Data/Flight/Reprocess/P120/extendedls1
```

File naming:

Data Type	aka	Send to FSSC	Naming template
MERIT		No	r<run#>_<version>_<dataType>.root
FILTEREDMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONMERIT		No	r<run#>_<version>_<dataType>.root
ELECTRONFT1		No	gll_el_p<procVer>_r<run#>_<version>.fit
EXTENDEDFT1		No	gll_xp_p<procVer>_r<run#>_<version>.fit
FT1	LS-002	Yes	gll_ph_p<procVer>_r<run#>_<version>.fit
EXTENDEDLS1		No	gll_xe_p<procVer>_r<run#>_<version>.fit
LS1	LS-001	Yes	gll_ev_p<procVer>_r<run#>_<version>.fit

Note: 'procVer' is a field added to the file name (and the keyword "PROC\_VER" in the primary header) added to the FFD 5/12/2010. Ref: [http://fermi.gsfc.nasa.gov/ssc/dev/current\\_documents/Science\\_DP\\_FFD\\_RevA.pdf](http://fermi.gsfc.nasa.gov/ssc/dev/current_documents/Science_DP_FFD_RevA.pdf)

Example:

```
/glst/Data/Flight/Reprocess/P120/merit/r0239557414_v120_merit.root
/glst/Data/Flight/Reprocess/P120/filteredmerit/r0239557414_v120_filteredmerit.root
/glst/Data/Flight/Reprocess/P120/electronmerit/r0239557414_v120_electronmerit.root
/glst/Data/Flight/Reprocess/P120/extendedft1/gll_xp_p120_r0239559565_v120.fit
/glst/Data/Flight/Reprocess/P120/ft1/gll_ph_p120_r0239559565_v120.fit
/glst/Data/Flight/Reprocess/P120/electronft1/gll_el_p120_r0239559565_v120.fit
/glst/Data/Flight/Reprocess/P120/extendedls1/gll_xe_p120_r0239559565_v120.fit
/glst/Data/Flight/Reprocess/P120/ls1/gll_ev_p120_r0239559565_v120.fit
```

## DataCatalog location and naming

Logical directory and group template:

```
Data/Flight/Reprocess/<reprocessName>:<dataType>
```

Note that the <dataType> field (following the colon) is a DataCatalog 'group' name, and file names are of the form r<run#>.

Naming examples:

```
Data/Flight/Reprocess/P120:MERIT r0239557414
Data/Flight/Reprocess/P120:FILTEREDMERIT r0239557414
Data/Flight/Reprocess/P120:EXTENDEDFT1 r0239557414
Data/Flight/Reprocess/P120:FT1 r0239557414
Data/Flight/Reprocess/P120:ELECTRONFT1 r0239557414
Data/Flight/Reprocess/P120:EXTENDEDLS1 r0239557414
Data/Flight/Reprocess/P120:LS1 r0239557414
```

## Data Sample

The currently defined data sample for P120 reprocessing includes:

First run	239557414 (MET), 2008-08-04 15:43:34 (UTC)	beginning of Science
Last run	333880535 (MET), 2011-08-01 08:35:33 (UTC)	Official Pass7 release
Total runs	16,459	
Total MERIT events	35,921,666,747	all "events"
Total FILTEREDMERIT/EXTENDEDFT1/LS1 events	5,035,929,409	all photon event classes
Total ELECTRONMERIT/ELECTRONFT1 events	68,055,849	
Total LS1 (FSSC selection) events	1,025,359,231	event classes (bits) 0,2,3,4 (transient, source, clean, ultraclean)
Total FT1 (FSSC selection) events	142,042,060	event classes (bits) 2,3,4 (source, clean, ultraclean)
Total disk space used	33.9 TB	

Summary from [DataCatalog](#) as of 8/2/2011.

Name	Files	Events	Size
ELECTRONFT1	16459	68,055,849	6.4 GB
ELECTRONMERIT	16459	68,055,849	147.4 GB
EXTENDEDFT1	16459	5,035,929,409	441.3 GB
EXTENDEDLS1	16459	5,035,929,409	816.5 GB
FILTEREDMERIT	16459	5,035,929,409	4.0 TB
FT1	16459	142,042,060	12.9 GB
LS1	16459	1,025,359,231	166.6 GB

MERIT	16459	35,921,666,74 7	28.3 TB
-------	-------	--------------------	---------

NOTE: One run, 242429468, of type TrigTest was declared 'good for science' but long after this task got started, so it has been intentionally omitted.

8/17/2011 update: Four orphaned runs are being reprocessed, including one TrigTest run and four nadir-pointed runs.

Stream	run	type
16459	242429468	(TrigTest)
16460	333355876	(nadirOps)
16461	333358500	(nadirOps)
16462	333365716	(nadirOps)

## Bookkeeping

- (This page): Define ingredients of reprocessing (processing code/configuration changes)
- Processing History database: <http://glast-ground.slac.stanford.edu/HistoryProcessing/HProcessingRuns.jsp?processingname=P120>
  - List of all reprocessings
  - List of all data runs reprocessed
  - Pointers to all input data files (-> dataCatalog)
  - Pointers to associated task processes (-> Pipeline II status)
- Data Catalog database: <http://glast-ground.slac.stanford.edu/DataCatalog/folder.jsp>
  - Lists of and pointers to all output data files
  - Meta data associated with each output data product

## P120-MERIT

### Status chronology

- 8/17/2011 - Begin reprocessing four orphan runs.
- 8/1/2011 - Begin and complete final backfill through run 333880535 (2011-08-01 08:35:33 UTC). For now, the three nadirOps runs are represented by dummy place-holder entries in the runFile.txt – their pipeline streams will fail.
- 7/29/2011 - Modified selection criteria for reprocessing run selection (findRunsRepro.py) to allow nadir-pointed data. This means adding '|| slIntent=="nadirOps"' to the dataCatalog selection string. See <https://confluence.slac.stanford.edu/display/ISOC/Nadir+Obs+Test+-+26+July+2011> for a list of runs affected by the nadir-pointed test.
- 7/26/2011 - Recovered three missing runs (below).
- 7/21/2011 - Begin and (mostly) complete reprocessing block 13, through run 332930755 (2011-07-21 08:45:53 UTC), 528 new runs - special GRB request. Note three runs failed skim/merge and are being investigated:

stream	run	UTC	Crash location
16136	332054399	2011-07-11 05:19:57	Filtered Merit skim failure
16151	332140182	2011-07-12 05:09:40	Electron Merit skim failure
16156	332169056	2011-07-12 13:10:54	Filtered Merit skim failure

- 6/17/2011 - Begin and complete reprocessing block 11, through run 329923889 (2011-06-16 13:31:27 UTC), 889 new runs.
- 4/19/2011 - Begin and complete reprocessing block 10, through run 324849509 (2011-04-18 19:58:27 UTC), 52 new runs.
- 4/15/2011 - Begin and complete reprocessing block 9, through run 324551768 (2011-04-15 10:51:27 UTC)
- 4/13/2011 - Begin and complete reprocessing block 8, through run 324368491 (2011-04-13 06:21:29 UTC)
- 4/3/2011 - The three missing runs have been produced by Level 1. Runlist recreated and those runs rolled back. There are no missing runs at this point.
- 3/29/2011 - Due to some hidden I/O problems, two changes have been made to this task, neither of which should affect data content. Note that this change will take effect on runs **after** 321756673, or any runs rolled back after this date.

Package	Old Version	New Version	Reason for change
GPLtools	v1r15p1	GPLtools-02-00-00	Checks size of files before and after move between disk and xroot
skimmer	07-07-00	08-01-00	detects failure to open input file

- 3/17/2011 - Last block complete (one instance of skimmer failure in mergeClumps)
- 3/15/2011 - Expand P120 to present (last run 321756673, 2011-03-14 00:51:11 UTC). Three runs continue to be dummied out. Once those runs have proper RECON files, their streams can be rolled back.

run	task stream	# subStreams
-----	-------------	--------------

306353950	11624	10
316611240	13431	8
320850543	14171	7

- 2/4/2011 - Expand P120 to present (last run 318211122, 2011-01-31 23:58:40 UTC). This include three runs for which there are no Recon files which have, for the moment, been supplied with 'dummy' entries in the runFile.:

306353950	'bad chunk' (known previously) - awaiting GR update
313483912	missing 700s, being worked on
316611240	'bad chunk' - awaiting GR update

- 1/12/2011 - Expand P120 coverage to include Crab ToO
  - Crab ToO 9/23/2010 15:50:50 to 9/27/2010 19:49:38, corresponding to MET r0306949696-r0307308940
  - reconfig through end of Sep 2010: 10916 -> 11841 runs, increase of 925 runs
  - No recon file for run 306353950 (being worked on, for the moment, place dummy file in runFile.txt)
  - First run in new block completed, awaiting checkered flag to continue...
- 8/29/2010 - Discovered three merge steps that silently failed (xroot file access). TASK complete.
- 8/28/2010 - processing formally complete (10916 runs), but some discrepancy in # of events
- 8/26/2010 - serious xroot problems. See initial [distribution](#) of files across xroot servers. From this report (courtesy Wilko) it is easy to see where problems are likely to arise - when the number of servers involved is small, e.g. two or three.
- 8/19/2010 - production continues at a crawl due to xroot server difficulties
- 8/16/2010 - resume full production, but at a slow trickle (max 350 simultaneous processClump jobs)
- 8/8/2010 - block 2 reprocessing complete. Many xroot server problems. (5 days to process 2084 runs)
- 8/3/2010 - begin block 2 reprocessing (through 255132033 MET), bringing the total runs reprocessed to 2721, about 5-1/2 months of data.
- 7/28/2010 - block 1 re-reprocessing complete
- 7/27/2010 - New GlastRelease (v17r35p10) containing new evtUtils, "to make the FT1EventClass bits compatible with the ScienceTools". Cleanup, including removing all files created last week during the first attempt.
- 7/21/2010 - block 1 reprocessing complete
- 7/20/2010 - agree upon 'pilot block' of runs (239557417 - 243220241), 637 runs. Begin...
- 7/19/2010 - submit first test run. success. await feedback

## Configuration

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P120-MERIT
Task Status	<a href="http://glast-ground.slac.stanford.edu/Pipeline-II/exp/Fermi/task.jsp?task=41146114">http://glast-ground.slac.stanford.edu/Pipeline-II/exp/Fermi/task.jsp?task=41146114</a>
GlastRelease	<del>v17r35p8</del> v17r35p10
Input Data Selection	"standard" from  <a href="https://confluence.slac.stanford.edu/display/SCIGRPS/LAT+Dataset+Definitions">https://confluence.slac.stanford.edu/display/SCIGRPS/LAT+Dataset+Definitions</a>  along with "&& (RunQuality != "Bad"    is_null ( RunQuality )"
s/c data	FT2 from P105 (runs 239557414 - 271844560), then from current Level 1 production
Input Run List	<a href="ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-MERIT/config/runFile.txt">ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-MERIT/config/runFile.txt</a>
photonFilter	CTBParticleType==0 && CTBClassLevel>0
electronFilter	CTBParticleType==1
jobOpts	<a href="ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-MERIT/config/reClassify.txt">ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-MERIT/config/reClassify.txt</a>
Output Data Products	MERIT, FILTEREDMERIT, ELECTRONMERIT

## Timing and Scaling

- (beyond block 2 results) Due to xroot problems (overstressing a small number of machines) the processing throughput dropped to 25-30 runs /hour (190-225 jobs/hour)
  - Wilko begins redistributing files around the xroot system in order to balance the load. This is only partially done by task completion.
  - Logs of job submission can be found [here](#)
- (block 1 results) The processClump step is taking ~40 hequ-minutes (or ~65 fell-minutes). With >500 simultaneous jobs running, there is little noticeable strain on xroot. There are five servers in the yellow-orange load range and they are claiming ~110-130 MB/s I/O rate.
- .
- The mergeClumps step is taking ~5 hequ-minutes
- It was observed that submitting 70 runs at once put a strain on /u30, home of GlastRelease. Some 93 of ~540 jobs failed with I/O error, but succeeded upon rollback.

## Load balancing

Introduce new trickleStreams.py script to (partially) assess pipeline activity and only the number of jobs considered appropriate based on available data. (block 1)

```
maxProcessClumps = 600      ## prevent overload of xroot
maxMergeClumps = 20        ## prevent overload of xroot (inactive)
maxStreamsPerCycle = 20    ## prevent overload of /u30 on startup
timePerCycle = 900         ## 15 minutes: allow time for dust to settle
```

With these parameters, it took ~ 5 hours to reach a point where fewer than 20 jobs per cycle were regularly submitted. Another 4.5 hours for the task to complete. On average, one run generated 7.5 processClump batch jobs.

For subsequent data (beyond block 2), xroot displayed such stress, that the maxProcessClumps limit was reduced to 250 or 300.

## P120-FT1

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

### Status chronology

- 8/1/2011 - Begin and complete final block of Pass7 reprocessing
- 7/26/2011 - Recovered the three missing runs (see P120-MERIT chronology), and reran stream 1018 (run 245403855), which had a bogus tstart time in the datacatalog – leap second issue, and recovered 27 events in that run.
- 7/22/2011 - Begin and (mostly) reprocessing block 13, through run 332930755 (2011-07-21 08:45:53 UTC), 525 new runs (+ 3 'dummy' runs due to skim crashes, see above) - special GRB request.
- 7/3/2011 - Task complete through run 329923889 (2011-06-16 13:31:27 UTC), 15,763 runs
- 6/30/2011 - New ST 09-24-00 (with gtdiffsp fix), restart trials with task version 2.1
- 6/14/2011 - Begin trials. Concern that gtdiffsp is crashing often (20-25% of time)
- 6/7/2011 - **IMPORTANT UPDATE:** a decision was made to rollback entire task with these changes:
  - calculate diffuse response for 'source' and 'clean' event classes
  - Produce new subset photon files for FSSC (FT1 with source and above, LS1 with transient and above)
  - Update various configurations (ScienceTools, evtClassDefs, etc.)This is being done by creating a whole new task, version 2.0, which from the pipeline perspective will overlay the older version 1.0.
- 4/19/2011 - Begin and complete reprocessing block 10, through run 324849509 (2011-04-18 19:58:27 UTC), 52 new runs.
- 4/15/2011 - Begin and complete reprocessing block 8, through run 324551768 (2011-04-15 10:51:27 UTC)
- 4/14/2011 - Begin and complete reprocessing block 7, through run 324368491 (2011-04-13 06:21:29 UTC)
- 4/3/2011 - The three missing runs have now been reprocessed. There are no missing runs at this point.
- 3/17/2011 - Catch up with P120-MERIT (last run 321756673, 2011-03-14 00:51:11 UTC)
- 3/15/2011 - Due to missing run, rollback runs 11624-11841. Bookkeeping is now correct.
- 2/4/2011 - Catch up with merit production (through 30 Sep 2010), but with one missing run/stream
- 1/28/2011 - Pass 7.4 reincarnation of this task complete through 31 Jul 2010
- 1/24/2011 - Entire task, xroot files, dataCat entries deleted. Prepare to reprocess as Pass 7.4
- 8/31/2010 - Pass 1 of this task is complete (through 31 July 2010)
- 8/30/2010 - Problem with makeFT1 stressing /u38 (very large temporary file needed when using xml representation of event classes was being written to \$PWD). Jim makes update to fitsGenApps => ST 09-18-03, put into production at stream 1400.
- 8/29/2010 - Begin Pass 1 of task...

### Configuration (version 2)

Task Location	/nfs/farm/g/glast/u38/Reprocess-tasks/P120-FT1
Task Status	<a href="http://glast-ground.slac.stanford.edu/Pipeline-II/exp/Fermi/task.jsp?task=65047878">http://glast-ground.slac.stanford.edu/Pipeline-II/exp/Fermi/task.jsp?task=65047878</a>
Input Data	MERIT (from P120-MERIT)
spacecraft data	FT2 from P105 (runs 239557414 - 271844560), then from current Level 1 production
Input Run List	<a href="ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-FT1/config/runFile.txt">ftp://ftp-glast.slac.stanford.edu/glast.u38/Reprocess-tasks/P120-FT1/config/runFile.txt</a>
Reprocessing Mode	reFT1
meritFilter	FT1EventClass!=0
evtClassDefs	00-19-01
eventClassMap	EvtClassDefs_P7V6.xml
ScienceTools	09-24-00
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 <a href="https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7">https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7</a> )
Diffuse Response	'source' using P7SOURCE_V6 IRF 'clean' using P7CLEAN_V6 IRF
IRFs	P6V7, contained within ScienceTools release
Output Data Products	<a href="#">FT1</a> , <a href="#">LS1</a> , <a href="#">EXTENDEDFT1</a> , <a href="#">EXTENDEDLS1</a> , <a href="#">ELECTRONFT1</a>

Processing chain for FITS data products

Data Product	selection	makeFT1	gtdiffrsp	gtmtime	gttcube
FT1	'source' and above EVENT_CLASS bits 2,3,4	true	true	true	false
LS1	'transient' and above EVENT_CLASS bits 0,2,3,4	true	true	true	false
FT1EXTENDED	FT1EventClass!=0	true	true	true	false
LS1EXTENDED	FT1EventClass!=0	true	true	true	false
ELECTRONFT1	CTBParticleType==1	true	false	true	false

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

- 1/28/2011 - Without diffuse response, the mergeClumps jobs are taking about 10 hequ-minutes of CPU to complete.
- 8/31/2010 - The primary batch job, mergeClumps, took a (mean) time of 42 cpu minutes (primarily a mixture of hequ and fell class machines). With P120-MERIT files nicely distributed across xroot servers, there were no xroot limitations to the processing. After the update to makeFT1, there was no longer an issue with overloading /u38 (\$PWD). The next bottleneck was the pipeline processing itself. This task consists of three batch jobs and four scriptlets; it was observed that the pipeline allowed hundreds of jobs to dwell in the READY state for extended periods of time, thus making it impossible to keep LSF saturated. Nevertheless, the maximum number of simultaneous jobs approached 2000. The task essentially completed in 8 hours, although some lingerers kept 'running' for another nine hours (mostly in SSUSP). A profile of job processing rate appears in this plot:

.

# P120-LEO-MERIT

## Status chronology

- 8/16/2010 - Task complete (199 runs)
- 8/13/2010 - Create task

## Configuration

Identical to the P120-MERIT task, except use FT2 files from P110 reprocessing.