P203 - Updated diffuse response for P7REP

status: **Complete**
last update: 4 June 2014

Overview

This task re-creates the FT1 (and EXTENDEDFT1) files for Fermi's entire science mission. It does nothing more than recalculate the diffuse response columns in these data products after a problem was reported in January 2014. There are four diffuse response columns filled: galactic and isotropic for both source and clean events.

P203-FITS - This task reads MERIT and produces FT1 (photons) + EXTENDEDFT1

This task is identical with P202-FITS with the following exceptions:

1. The dataset reprocessed is extended to include all current (Level 1) data since the end of the P202 task.
2. New diffuse response
3. Only FT1 and EXTENDEDFT1 data products are produced
4. File naming and other bookkeeping replaces "P202" with "P203".

Refer to the P202 documentation for additional configuration details.

Refer to the Official LAT Datasets to see how these data fit into the big picture.

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=P203
   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

P203-FITS

This task generates FT1 and EXTENDEDFT1 data products.

Status chronology

- 2/20/2014 - Initial set-up of task
- 2/25/2014 - Begin trickleStream on Block 1 – entire P202 run list
- 2/28/2014 - Many problems: LSF seems to bunch up a large number of jobs before dispatching. This causes a shock wave of jobs causing a problem for AFS (reading the ~500 MB diffuse response file), and for the dataCatalog (the queries for FT2 file, and to determine the next file version for each data product). Therefore, reconfigure trickleStream to a mere dribble – no more than 300-400 jobs typically running simultaneously. This means a 1-day task has become a week-long effort.
- 3/3/2014 - Block 1 complete. Last week developed new scheme to move all dataCatalog queries from batch jobs into jython scriptlets. These changes will be integrated prior to processing any more data. Validity check: There are the same number of runs and the same number of events in the reprocessed P203 data as in the P202 data.
  310,326,817 events in 29158 files. The run range for block 1 (== entire range of P202) is 239557414 through 405329691.
- 3/4/2014 - Configure Block 2, consisting of Level 1 data since 5 Nov 2013 through 28 Feb 2014. Begin trickleStream.

#runs	30915	
#evts	67249594218	
start	239557417	2008-08-04 15:43:37
stop	415328595	2014-03-01 01:03:15

- 3/5/2014 - Block 2 complete.
- 5/19/2014 - Begin block 3, through run 422145472, adding 1198 new runs to the list. Total #runs = 32113.
- 5/20/2014 - Block 3 complete

- 6/3/2014 - Cut-over. Begin block 4, the final backfill, through run 423447612. Total #runs = 32342, an increase of 229 runs.
  (first Level 1 run after cut-over is 423453614.)
- 6/3/2014 16:05 - Block 4 complete. Summary from dataCatalog:

Name	Type	Files	Events	Size
EXTENDEDFT1	Group	32342	10,325,038,320	904.7 GB
FT1	Group	32342	360,437,538	32.4 GB

- 8/20/2014 - rollback run 27487 (run 395891323) with newly produced MERIT and FT2 files (see longer explanation in the P202 page).

Configuration

Identical with P202-FITS except:

Diffuse Model	based on contents of /afs/slac.stanford.edu/g/glast/ground/GLAST_EXT/diffuseModels /v4r1 (see https://confluence.slac.stanford.edu/display/SCIGRPS/Quick+Start+with+Pass+7)
Output Data Products	FT1, EXTENDEDFT1

Generation of output data products:

Data Product	destination	data content [1]	event selection [1]	makeFT1	gtselect	gtdiffrsp	gtmktime
EXTENDEDFT1	SLAC	FT1variables	((FT1EventClass & 0x00003EFF)!=0) pass7.6_Extended_cuts_L1	✅	❌	✅	✅
FT1	FSSC+SLAC	FT1variables	'source' and above EVENT_CLASS bits 2,3,4 evclass=2 filtered from EXTENDEDFT1	❌	✅	(inherited)	✅

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

Timing

- Each run requires approx 20-30 minutes of CPU time, depending on the machine-class being used. However, due to AFS and dataCatalog issues, block 1 running was restricted to ~500 or fewer jobs at a time. After ~30,000 trials, the mean CPU time for the mergeClump job step is 54 minutes.

DataCatalog query change (2/28/2014)

1) Refer to the modified files already in the DCtest task on /u38 which was used to prototype this change

2) update repTools.py with new version of getCurrentVersion(), and make a completely new release 00-01-05. Note that findFt2() is now obsolete.

3) In the P203-FITS/config directory, make these changes:

>> config.py - change pointer to new version of commonTools
>> setupRun.py - prepare list of output data product types
>> createClumps.jy - query for FT2 file name and store in pipeline var
>> processClump.py - fetch FT2 file name directly from pipeline var rather than via query
>> setupMerge.jy - query for latest file version of each output data type
>> mergeClumps.py - make pipeline vars -> env-vars

4) The usual git commit/tag/push