Clouds and the Earth's Radiant Energy System
(CERES)

Configuration Code Number Assignment Strategy - Edition 3 (Terra and Aqua) and Beyond

Primary Author

Lisa Coleman

Science Systems Applications Inc. (SSAI)
One Enterprise Parkway, Suite 200
Hampton, VA 23666

NASA Langley Research Center
Climate Science Branch
Science Directorate
21 Langley Boulevard
Hampton, VA 23681-2199

TABLE OF CONTENTS

Section Page

TABLE OF CONTENTS

Section Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

**CERES Configuration Code Number Assignment Strategy -
Edition 3 (Terra and Aqua) and Beyond**

1.0 Introduction

The numerous data products generated by the Clouds and the Earth's Radiant Energy System (CERES) science data processing system are produced by a series of software packages referred to as Program Generation Executives (PGE). Each PGE produces a unique data product, and each execution of a CERES PGE produces a unique instance of a data product.

The CERES science data processing system is a complex system. This system consists of PGEs that process measurement-level data, PGEs that process regionally and/or temporally averaged data, PGEs that ingest data from the CERES instruments, and PGEs that ingest data from other sources external to the CERES project. The diagram in Reference 1 is a high-level depiction of the flow of CERES data through the different software packages that comprise the CERES Data Processing System.

Since the onset of CERES production processing, the science software for the PGEs has been replaced multiple times, externally provided input data products have been reprocessed, and new PGEs have been added to the science data processing system. Code changes have been driven by hardware and software upgrades, better understanding of the instrument operations and calibration, and refinement of the scientific algorithms executed by the various CERES PGEs. As these events occurred, the CERES data products were reprocessed, beginning with the first data collected from the instruments after their launch up to the present, resulting in multiple versions of the CERES products for the same time frame. The CERES science data processing system includes PGEs that are capable of processing multiple versions of a product, as well as PGEs that are only capable of processing a single version of a product. A naming convention with the flexibility to span decades is needed to meet the provenance requirements for critical climate science data records.

Prior to the launch of TRMM, the first satellite carrying a CERES instrument, the CERES Systems Engineering Committee defined the standards for the CERES product file names. The standard file naming convention is as follows:

"CER"_ProdID_SamplingStrategy_ProductionStrategy_ConfigurationCode.Instance
where:

- "CER" is a constant text string indicating the product is a CERES product
- ProdID indicates the CERES product name, such as "BDS" or "SSF"
- SamplingStrategy (SS) indicates the instrument(s) and/or satellite source of the input data, e.g., "Terra-FM1-MODIS" indicates the input data were collected from the CERES FM1 instrument, as well as the MODIS instrument on board the Terra satellite
- ProductionStrategy (PS) indicates the version of the output product, such as "Edition3"
- Configuration Code (CC) is a six-digit number that acts as a unique identifier for the software version and version of the input data used to produce a CERES data product.

- Instance identifies the temporal span and, if applicable, the geographical area included within a particular granule of the data product

The purpose of the CC number is to provide a provenance or lineage mechanism for indicating that a change occurred in the CERES data processing system that potentially results in changes within a CERES data set. These changes include, but are not limited to, software modifications, changes in the operational environment, or changes to the input data used to produce the product. If a CERES instance is produced multiple times, a different CC number is assigned for each generation of the instance. Once a CERES product is publicly available, it is not permissible to remove it from the CERES archives, and thus multiple versions of a single instance may exist. The different CC numbers distinguish between the different versions.

1.1 CC Numbers Within the Scope of This Document

The information in this document pertains to the CC numbers affiliated with CERES Terra and Aqua Edition3 and later data sets, as well as NPP Edition1 and later data sets. For information on the assignment of CC numbers for products produced prior to Terra and Aqua Edition3, see Reference 2.

1.2 Definitions for This Document

The following are definitions for selected terms used in this document for the purpose of describing the strategies behind CERES CC number assignments.

1.2.1 Definition: Data Record

For purposes of this document, the term “data record” refers to the time frame during which data are recorded from a CERES instrument and then processed through the CERES data processing system. The data record begins at launch and continues through the operational life of the instrument.

1.2.2 Definition: Data Set

For purposes of this document, the term “data set” refers to a collection of instances of a data product with the same product identifier, sampling strategy, and production strategy. For example, instances of the SSF product that contain the string “CER_SSF_Terra-FM1-MODIS_Edition3A” in the filename comprise a data set.

1.2.3 Definition: Target PGE

For purposes of this document, the term “target PGE” refers to any PGE that uses the output of another PGE as input. One PGE can have several target PGEs, and a single PGE can be a target PGE of multiple PGEs. Target PGE information is found in the Operator’s Manual associated with each CERES PGE (see Reference 3).

In the example illustrated in Figure 1-1, there are two target PGEs for PGE 4.5-6.5P4 (block A): PGEs 9.2P2 and 5.0P2 (blocks B and C, respectively). For PGE 5.0P2, there is one target PGE, PGE 7.2.1P1 (Block D). PGE 7.2.1P1 is a target for multiple PGEs, 5.0P2 and 7.3.1P1 (block E).

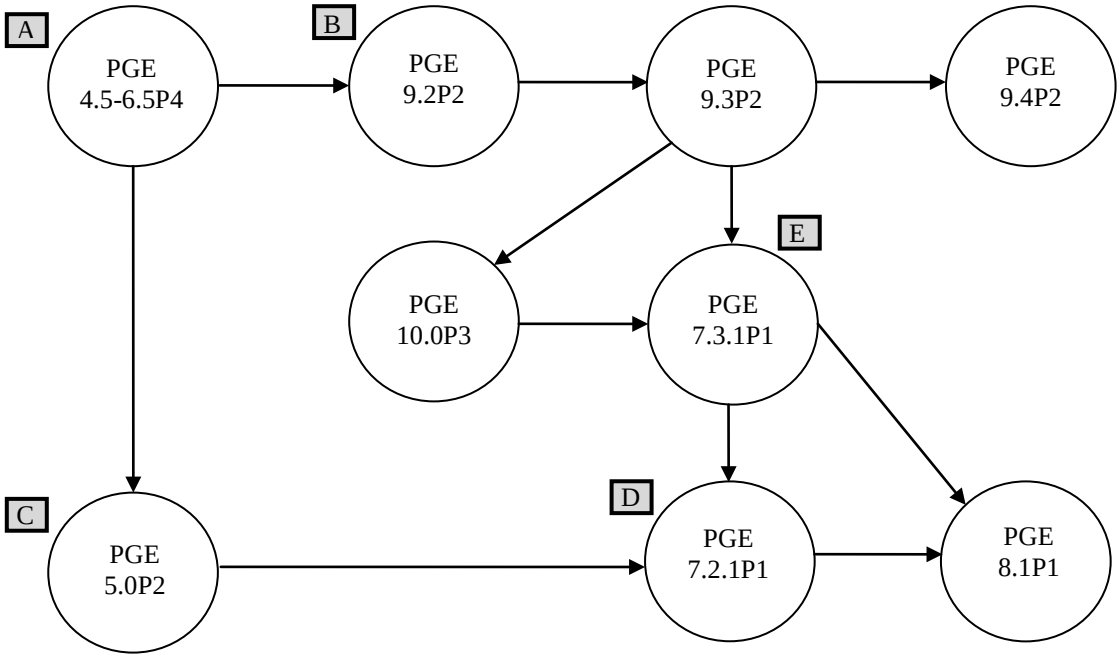

Figure 1-1. Target PGE Examples from Edition3 Processing

1.2.4 Definition: Downstream PGE

For purposes of this document, the term "downstream PGE" refers to any PGE that follows another PGE in the processing flow. For a given PGE, the downstream PGEs include the target PGEs that immediately follow that PGE, as well as the target PGEs of the target PGEs, and so on. For example, in Figure 1-1, all of the PGEs that process after PGE 4.5-6.5P4 are considered to be downstream from 4.5-6.5P4. The last PGE in the stream will be the last PGE to process, and will have no target PGEs.

1.3 Changes in CC Number Assignment Strategy for Terra and Aqua Edition3 Data Sets

1.3.1 Initial Values for Edition3

Prior to Edition3 processing, CC number modifications were always based on the previous CC number value assigned to a given PGE. With Edition3, the values of the CC numbers for the majority of the CERES products were initially set to "300300" regardless of previous values. The value of "3" was chosen for the first and forth positions to reflect the version of the data set, Edition3. The first, or leftmost, three digits were designated to indicate the source of a major input data set. The second, or rightmost, three digits were designated to indicate an algorithmic change to the CERES software that resulted in the generation of a new executable.

1.3.2 Implementation of Product Suites

The first CC numbers assigned to CERES products were 001001. The value of the CC number for a given PGE was incremented when the criteria were met, and were incremented independently of increments for other PGEs. Thus, the assignment of a CC number to any one of the CERES products was not correlated with the CC number for any other CERES product. This approach continued throughout the generation of the Edition2 version of the CERES data products. See Reference 2 for details regarding the assignment of CC number values from the TRMM launch through the Edition2 versions of the CERES data products.

The approach of grouping products together into suites of products with a common CC number began with Edition3 processing. Which PGEs are included in each suite is decided on by the CERES team strictly for convenience when assigning CC numbers. A CERES product suite is a grouping of CERES products based on similarities in product level, i.e., Level 2 or Level 3, sampling strategy, production strategy, and major input data. The products in a suite tend to process serially with little lag time in between. Information regarding which products are included in which suites is maintained in Reference 4. The benefits of this grouping approach are:

- The commonality across CC numbers helps link the products together in the minds of the various CERES teams responsible for producing the data products.
- There are fewer CC numbers for these team members to track for the CERES processing system, which increases efficiency in preparing for processing the CERES products and reduces the risk of error.

1.3.3 CC Number Value Increment Criteria – Software Changes

Prior to processing the Edition3 data sets for Terra and Aqua, the value of the CC number for the output from any given PGE was incremented with every update to the PGE software. This included updates to the science software as well as the scripts that managed the execution of the science software. Beginning with Edition3, the CC number value is only incremented for changes to the science software that result in a recompilation, i.e., a new version of the executable. Minor changes to the scripts that manage the execution of the science software no longer result in a modification to the CC number.

1.3.4 CC Number Value Increment Criteria – Processing Environment Upgrades

Prior to processing the Edition3 data sets for Terra and Aqua, the value of the CC number for the output from all CERES PGEs was incremented whenever an update to the processing environment occurred. Beginning with Edition3, if the executables could still produce the same results after the upgrade as before, the need for new executables did not exist and therefore the CC number will not be incremented.

2.0 Strategy of Assignment of CC Numbers

The strategy for the assignment of CC numbers for CERES products begins with assigning an initial value for each CERES data set, as discussed in Section 2.1. The strategy for modifying the CC number within a data set is discussed in Section 2.2.

The CC number for a CERES product is a 6-digit code that is expressed as two triads, such as "300300," and is shown in Table 2-1. CC1 and CC4 reflect the version (Edition) of the software. The first three digits, or triad T1, reflects the change and denotes the lineage of the input data sets to the PGE. The second three digits, or triad T2, captures software changes and identifies the lineage of the code used to produce the product.

Table 2-1. CC Code Layout

Triad 1 (T1) Input Data			Triad 2 (T2) Software Version		
CC1-T1	**CC2-T1**	**CC3-T1**	**CC4-T2**	**CC5-T2**	**CC6-T2**
CERES Dataset version "Edition"	Input data version MSB	Input data version LSB	CERES Dataset version "Edition"	Software version MSB	Software version LSB

2.1 Initial Values

As discussed in Section 1.3.1, the initial CC number values assigned to a data set reflect the version of the data set. The first and forth digits of the CC number match the version of the data set, e.g., the initial CC numbers for Edition3 data sets were "300300." Likewise, the initial values of the CC numbers used for Aqua and Terra Edition4 data sets include a value of "4" in the first and forth positions of the CC number, and the initial values for the NPP Edition1 data sets include a "1" in the first and forth positions.

It is possible that a PGE is capable of producing results for multiple data sets, such as the Instrument Subsystem PGEs that produce the Terra and Aqua Edition3 and Edition4 data sets, as well as the NPP Edition1 data set. In this case, the Terra and Aqua Edition4 and NPP Edition1 processing began after more than 13 years of the Terra and Aqua Edition3 data sets were generated. During the processing of these 13 years, the PGE software used to produce the Terra and Aqua Edition3 data sets was replaced in the production environment multiple times, and the CC number was incremented with each replacement. To correlate the version of the software used to produce the different data sets, the last two digits of the CC number are consistent across all the data sets, i.e., the initial CC number for NPP Edition1 was "100104," indicating that the data products were generated by the same software as the Terra and Aqua Edition3 data products with the CC number "300304."

2.2 Value Changes

The strategy for modifying the value of the first three (leftmost Triad 1) digits of the CC number within a CERES data set are discussed in Section 2.2.1. The strategy for changing the value of the last (rightmost Triad 2) three digits is discussed in Section 2.2.2.

2.2.1 Changes to the First Three Digits (Triad 1)

The value of the first three digits is incremented whenever a change occurs to a primary input data set. The value of the CC number for at least the output product generated by the PGE that ingests the modified input data will be incremented, along with the CC number for the products produced by other PGEs in the same suite. If the modified input data also affects the results of products generated by downstream PGEs not included in the same suite, the values of the CC numbers for those PGEs also need to be incremented. Changes to secondary input data sets, such as ancillary input data, may not be significant enough to warrant a change in the value of Triad 1, however. In such cases, the change may be indicated in Triad 2. An example of this scenario is discussed in Section 2.2.1.3.

The typical increment value of the first three digits is one.

2.2.1.1 Changes in the First Three Digits for Multiple PGEs

It is critical to indicate the impact of updated input products throughout the end-to-end CERES science data processing flow. If the change to the input data is expected to influence the scientific results produced by a PGE, then it is assumed that the input data change will also affect the scientific results produced by each downstream PGE (see Section 1.2.4). The value of the first three digits for the products produced by each downstream PGE, therefore, is also incremented. For example, within the data record processed for Edition3, both the MODIS cloud property and GMAO meteorological input data sets were significantly modified at different points in the data record. The influence of these changes not only affected the results produced by the PGEs that ingested these data and those PGEs within the same suite, but also all the CERES products produced by all of the downstream PGEs (see Reference 1). The first three digits of the CC number for all of the affected products were incremented to maintain the same value with each change to the input products. Table 2-2 shows specific examples of increments in Triad 1 for Edition 3 based on changes to major input data sets. The first digit indicates that this is a CERES Edition 3 item. The second two digits indicate the version of the input data. The resulting concatenated value for the first three digits of the CC number associated with the affected CERES products is explained below.

Table 2-2. Edition3 CC Number Changes – First 3 Digits Triad 1

First 3 CC Number Digits	Dates	MODIS Collection ID	GMAO Version ID	Clouds Subsystem Version Input to SCC Generation
300	Launch through April 2006	4	4	Edition2
301	May 2006 through November 2007	5	4	Edition2
302	December 2007 through June 2010	5	5	Edition2
303	July 2010 forward	5	5	Edition1-CV

2.2.1.2 Changes in the First Three Digits for a Single PGE Suite Only

If the change in the input data is not expected to influence any of the CERES products generated by target PGEs that are included in different suites, then the first three digits for only the products generated by the affected suite need to be incremented. For example, a December 2012

change in an external provider resulted in a reorganization of the input data to the Clouds Subsystem Snow and Ice Pre-Processor. This change affected the PGE software, requiring a change in the CC number of the ESNOW and EICE output products. As this change did not affect the results, no change was necessary to the CC number for the output products generated by the downstream PGEs in different suites that used the ESNOW and EICE products as input.

2.2.1.3 Input Data Changes Affecting Second CC Number Triad Value

It is possible that a change to an input data set results in a change to the second CC number triad instead of the first triad. This situation can occur if the following criteria exist:

- The input data set is not the primary input data set to the PGE
- A recompilation of the PGE software is not required
- The change in the input data set does not scientifically impact the results produced by the PGE.

An example of this occurred with SCCR 1002 for the Synoptic SARB Subsystem PGE 7.2.1P1. In this case the naming convention for the snow and ice ancillary input data sets changed due to the source of the data. As the format of the input data did not change, and there was no scientific impact on the product generated by the PGE, a recompilation of the PGE software was not necessary. In this situation, the value of the second triad was incremented by one for all the PGEs in the associated suite.

2.2.2 Changes to the Last Three Digits (Triad 2)

The last three digits, the rightmost three digits, indicate the software version used to generate the data product.

2.2.2.1 Criteria for Incrementing Value of Last Three Digits of the CC Number

The last three digits of a CERES CC number are incremented whenever at least one of the following criteria are met:

- There is a change in the PGE executable that generates the product
- There is a change in an executable for a PGE that generates a product within the same product suite
- There is a change to a secondary input data set that does not scientifically impact the results.

An example of the last item is discussed in Section 2.2.1.3.

2.2.2.2 Value of Increment

When an increment of the value of the second triad is required, the typical value of the increment is one. There are exceptions, however, such as during the generation the Edition3 SSF data sets. Different PGEs, or versions, of the Inversion Subsystem Main-Processor, were used to process the Terra and Aqua data products. While processing the data for the first years of the data record separate modifications were required by each PGE at different points in the data record, resulting in different CC numbers between the Terra and Aqua Edition3 SSF data products for several months. Table 2-3 provides an example of the various increments for the SSF Edition3 product for the Terra satellite, and Table 2-4 provides an example of the various increments for the Aqua satellite product.

Table 2-3. Example of Increment to Second Triad Using Edition3A Terra SSF CC Numbers

SCCR ID	CC Number	CC Number Date Range	Change Description
811	300301	03/2000 – 10/2005	First software correction
786	300300	11/2005 – 07/2006	Original Edition3A Terra software
786	301300	07/2006 – 11/2007	
786	302300	12/2007 – 02/2010	
811	302301	03/2010 – 06/2010	
811	303301	07/2010 – 12/2010	
870	303302	01/2011 – 12/2011	Second software correction
939	303303	07/2010 – 12/2012	Third software correction

Table 2-4. Example of Increment to Second Triad Using Edition3A Aqua SSF CC Numbers

SCCR ID	CC Number	CC Number Date Range	Change Description
871	300301	07/2002 – 03/2005	First software correction
812	300300	04/2005 – 04/2006	Original Edition3A Aqua software
812	301300	05/2006 – 11/2007	
812	302300	12/2007 – 06/2010	
812	303301	07/2010 – 12/2010	
871	303301	01/2011 – 12/2011	
939	303303	07/2010 – 12/2012	Second software correction

An important item to note regarding the Edition3 SSF processing is that the first data dates to be processed were not at the beginning of the data record for either satellite. This was due to input data availability on the computing platform. After the onset of Terra processing with an executable generated under SCCR 786, a software error was identified that only affected data dates before October 31, 2005 and required an update to the executable. While the updates to the software were prepared and tested, processing of the data record continued for unaffected dates with the executable generated under SCCR 786. As indicated in Table 2-3, the last three digits of the CC number for the SSF products produced under SCCR 786 are 300. The updated executable was generated under SCCR 811, and the associated last three digits of the CC number for the SSF products are 301. Likewise, the first dates of the Aqua Edition3A SSF product to be generated were also not at the beginning of the Aqua data record.

While processing the dates up through July 2010, the Terra PGE was modified an additional time, but the Aqua PGE was not. As a result of SCCR 939, however, the executables for both the Terra and Aqua versions were replaced. At this point, it was reasonable to increment the CC number by one for Terra, and by two for Aqua. With this approach, the CC numbers for the data sets from both satellites were once again the same.

While it is highly beneficial for the CC number to be the same value for all products within a product suite for a given month, it is not always possible. For example, if an unplanned

reprocessing of a portion of a data set begins “mid-suite,” the CC numbers within a product suite will get out of sync each other. An example of this occurred with the need to reprocess the SYNI product for May and June 2012 with a new executable. The SYNI product is generated by the third PGE to process within a product suite. The CC numbers for the SYNI product were incremented, as well as for the SYN1deg products generated by a target PGE that were therefore also reprocessed. As it may be preferable to not reprocess products simply to synchronize CC numbers, the CC number associated with the existing TSI products generated by the first and second PGEs to process within the suite was now out of sync with the SYNI and SYN1deg products for those months. The TSI CC number was incremented, however, to match the SYNI and SYN1deg products before the product for further months was generated.

3.0 Communication of CC Number Assignments

A Production Request (PR) is the formal communication between the DMT and the ASDC requesting that CERES data be processed. Each PR indicates the data products that are to be generated, the data dates, the PGE to use, and the value of any processing options. The CC numbers for both input and output products are also specified in the PRs. The Software Configuration Change Request (SCCR) identification number under which the PGE executable was compiled is also included in the PR. The information provided in the PRs is used by the ASDC Operations Team to generate the requested CERES data products.

3.1 PR Tool Resources – Historical CC Number Information

Prior to March 1, 2013, the CERES PRs were distributed to the CERES Data Management Team and the ASDC SIT and Operations teams as a PDF file or as an Excel spreadsheet. Since March 1, 2013, the CERES PRs are stored in the PR Data Base (PRDB) and accessed through the PR Tool web interface. This interface, behind the Langley firewall, is accessed through the following URL specified in Reference 5.

A web interface with the CERES Production Monitoring Data Base (PMDB) provides the capability to search for the CC number assigned to CERES products. Using this tool, the user can search on any of the following: product, satellite, instrument, production strategy, and beginning and ending dates. The display of the results of the search lists the CC numbers and associated date ranges for the criteria selected. This tool is accessible from the PR Tool. Users within the firewall with access to the AMI operational system within ASDC may also search the CERES product listings on the DPO to view the product filenames, which include the CC number.

3.2 Updates to CC Number Assignments

Effective February 10, 2015, a spreadsheet file containing the current CC number assignments for each active CERES PGE is included in Reference 4. Whenever this file is updated entries are made to the included revision history and the CERES DMT and ASDC are notified via email regarding the updates.

REFERENCES

1. CERES Data Flow Diagram
   http://ceres.larc.nasa.gov/documents/ATBD/pdf/atbd_linked_dfd_12_04_2013.pdf

2. Sampling Strategy, Production Strategy, and Configuration Code Implementation at the Langley TRMM and Terra Information System (LATIS), by Maria Vallas Mitchum and Sandra Nolan. http://ceres.larc.nasa.gov/documents/intern_docs/pdf/ss.ps.cc.pdf

3. CERES Data Management System Operator's Manuals
   http://ceres.larc.nasa.gov/ops_man.php

4. CERES Configuration Code Number Assignments Organized by PGE
   http://ceres.larc.nasa.gov/prod_cc_code_doc.php

5. Ref_PR-Tool_URL. : http://ceres-prtool.larc.nasa.gov/pr_search.php.

Appendix A - ACRONYMS

AMI	ASDC Modernization through Integration
ASDC	Atmospheric Sciences Data Center
BDS	CERES BiDirectional Scan archival product
CC	Configuration Code
CERES	Clouds and the Earth's Radiant Energy System
DMT	Data Management Team
DPO	Data Products Online
EICE	CERES Ice internal ancillary input product
ESNOW	CERES Snow internal ancillary input product
FM1	CERES Flight Model 1
GMAO	Global Modeling and Assimilation Office
IES	CERES Instrument Earth Scans internal product
MODIS	Moderate-Resolution Imaging Spectroradiometer
PGE	Product Generation Executives
PMDB	Production Monitoring Data Base
PR	Production Request
PRDB	Production Request Data Base
PS	Production Strategy
SIT	System Integration and Test team at the ASDC
SS	Sampling Strategy
SSF	CERES Single Scanner Footprint, TOA/Surface Fluxes and Clouds product
SYNI	CERES Synoptic Radiative Fluxes and Clouds Intermediate product
SYN1deg	CERES Synoptic Radiative Fluxes and Clouds 1-degree product suite
TRMM	Tropical Rainfall Measuring Mission
TSI	CERES Time Space Interpolate product