Hilawe Semunegus
Updated: June 21, 2006
Original: May 24, 2006

SSM/I Data Quality Preliminary Analysis

The National Climatic Data Center's Remote Sensing and Applications Division has been working to improve the quality of SSM/I (Special Sensor Microwave Imager) data that has been archived since 1993. One such project focuses on the lack of reliable SSM/I data prior to early 1997. This report will attempt to identify all the data processing problems that have been identified with SSM/I data before February 4, 1997 (2/4/97). This report will also suggest a solution to deal with the array of data quality issues that have been found in the course of reading and displaying SSM/I data for the period of record (1993-present). Various users such as land-surface modelers and oceanic researchers rely on NCDC to provide accurate and quality controlled data for their studies.

Data Format

Comprehensive Large Array-Data Stewardship System (CLASS) has been serving SSM/I data starting from February 17, 1997. When the matter was investigated further, CLASS documentation indicated there was a format change during this period. With the help of Hanjun Ding at Office of Satellite Data Processing and Distribution (OSDPD), a search was made for any data format documentation before early 1997, so that any format changes can be found and addressed. It turned out that Axel Graumann had this documentation. After looking at the documentation for the earlier TDR (Temperature Data Record--Antenna temperature), SDR (Sensor Data Record--Brightness temperature), and EDR (Environmental Data Record—Geophysical parameters) data formats, it was determined that there has not been any changes in data format/structure according to the documentations. Jeff Haferman, civilian employee at the Navy Fleet Numerical Meteorology and Oceanography Center (FNMOC), who works with sending processed SSM/I data to CLASS, confirmed this.

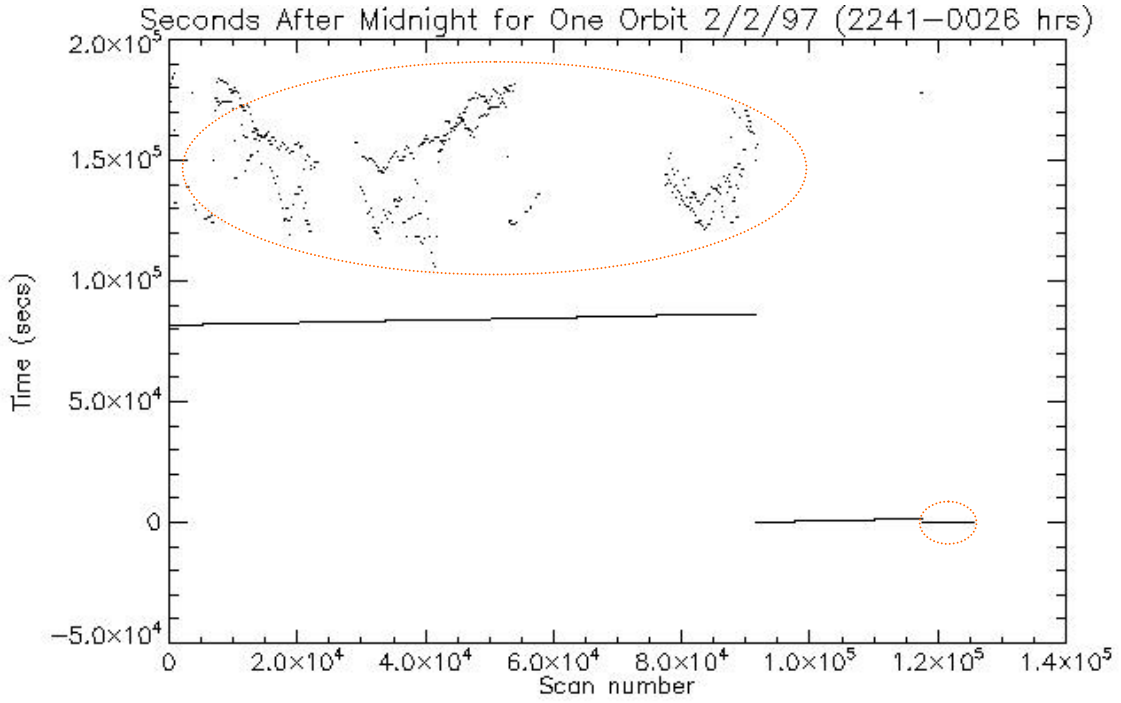
Data Processing

Haferman had indicated that the processing of SSM/I data has improved drastically starting in early 1997. He stated that FNMOC had a poor QA/QC system set up in the operational processing codes prior to 1997 and that these earlier data scans got

dropped on a regular basis and the remainder of the orbit either continued as normal, discarded, assigned zero values, or assigned unrealistic values that were one order of magnitude higher than expected. Also, FNMOC did not reship complete data to NESDIS when a short dataset was created, so only portions of an orbit were being sent, and the rest were discarded. Haferman indicated that they started to utilize rigorous operational data processing systems that are observed after early 1997. The exact date of this change to a more effective processing system was on and after February 4, 1997 (2/4/97).

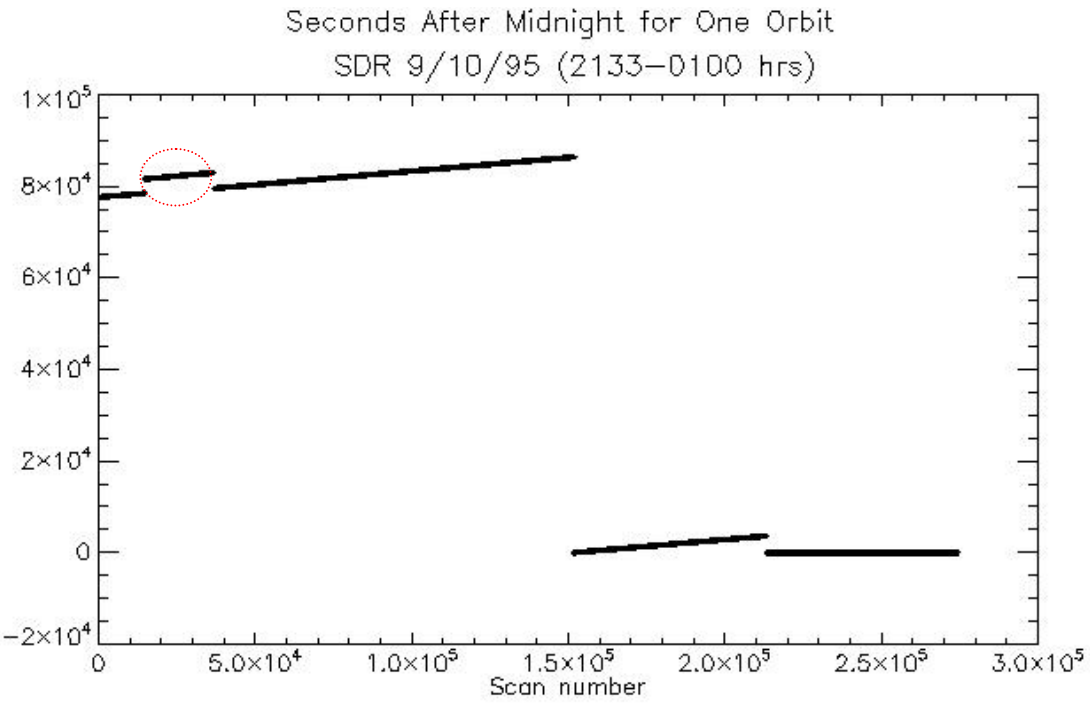
TDR/SDR Data Issues

The problems that were encountered with reading TDR/SDR data are explained here. The following graph will illustrate time in seconds after midnight (0-86399 seconds range) for one orbit for pre-2/4/97 data for the F13 platform:

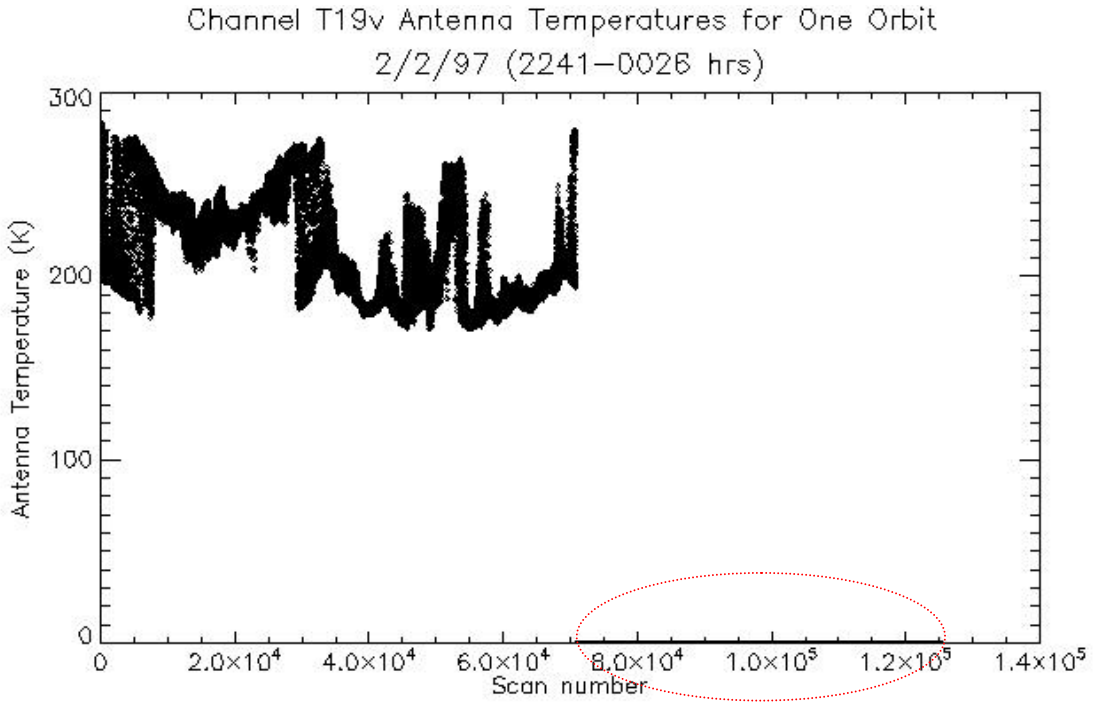


The upper portion of graph shows that time values exceed 86400 seconds (>100,000) or time values were assigned zero values as is circled on the bottom right of the graph. The larger red circled area (> 100,000 secs) was eventually eliminated and will be explained later in the report. This same trend can be found for data from all the platforms (F10, F11, F13) before 2/4/97.

Another time issue encountered for TDR and SDR data is the disconnected time values as is depicted in the top left corner of the plot:

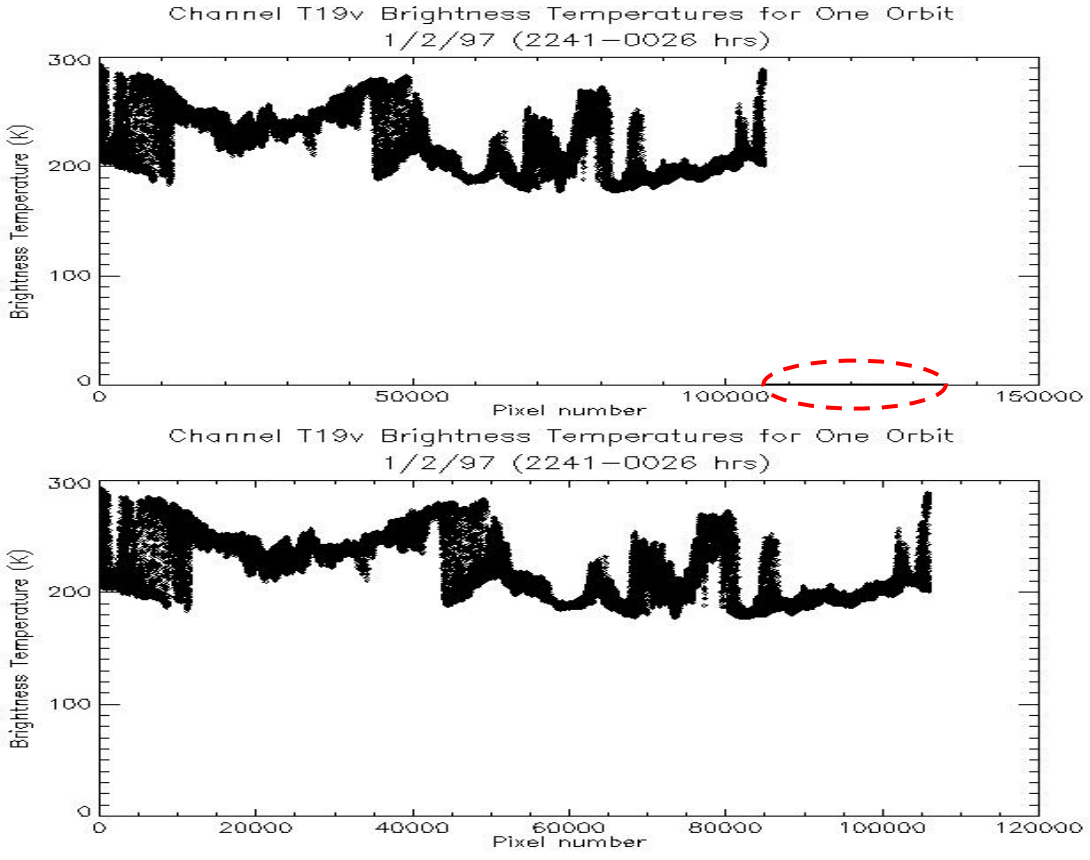


This graph shows the error in temperature values (Channel T19-Vertical) for one orbit:



Antenna temperature values toward the end of the scans were assigned zero values. Again, this trend is found across all platforms before 2/4/97. The way this problem was solved was by ignoring the actual number of scans that an orbit claims it holds (number of scans is calculated using either file size or by grabbing the defined number inside the

data structure) and to only read the number of scans that looked correct. For example, a given file before 2/4/97 will claim to have 1650 scans (for both file size calculated scan number and data structure defined scan number). But once the data is read and it is plotted, one notices that there are a lot of zeroes towards the end of the scan, as is shown in first plot below (circled red). However, if only the actual number of scans (calculated by counting the number of actual scans read) that were read by the read routine is used, then the second plot below will result or the zeroes disappear. As this example shows, most of the issues identified with TDR/SDR data before 2/4/97 are solved by changing how data is read in or it requires more code manipulation to figure out how to work around these issues. The read_TDR.pro IDL routine is heavily commented, so that a step-by-step explanation will help a user understand the types of manipulation needed to read earlier TDR/SDR data.



"End of File" and Other Data Structure Issues

For data before 2/4/97, whenever the scans were dropped, an "end of file" (EOF) block would be hit and then the same data file could be re-read after that point since there

were still data left over to read. This EOF block is supposed to only be found once in each file, as is found for data after 2/4/97. On average, the EOF block is hit three times per file read for pre-2/4/97 data. Some of these files have been found to have more than 10 EOF blocks. Here's a table describing the EOF block.

Word #	First Byte	Second Byte	
1		3	Number of two-byte words for the End of Product Block
2	001	002	Octal mode and submode
3	Checksum		

The method by which the various data blocks are read is by identifying the "Number of two-byte words (integer)" for the specified block. For data after 2/4/97, this specific integer was all that was necessary to identify which block was to be read. But after first attempting to read pre-2/4/97 data, it was necessary to also use the "Octal mode and submode" to identify data blocks. This eliminated all the non-random noise such as the scattered data that is seen in the top portion (time above 100,000 seconds) of the first figure of this report (TDR time vs. scan number plot).

Quality Control

After performing simple threshold quality control (QC) on both time and temperature values, it was discovered that this method would be grossly inadequate to deal with all the error characteristics that were found. After discussions with Wesley Berg, a system of quality flags was found to be the most accurate way of explaining to users the reasons why data were flagged and whether the data is reliable. Wesley Berg's quality flag classification shown below seems to be a sufficient method to flag SSM/I data.

0 = Good data in all channels in the swath
1 = Possible sun glint
2 = Climatology QC check warning
10 = Data is missing from input file or unreadable
20 = Geolocation check flagged pixel as bad
30 = Climatology check flagged pixel as bad
40 = Distance between pixels is nonphysical
50 = Antenna temperatures are < 50 K or > 325 K
60 = Lat/lon values are out of range
70 = Adjacent/cross-polarization pixel flagged as bad

Flag values 10 to 90 indicate that there are major issues resulting in missing values for at least one channel. Values from 1 to 9 indicate cautionary flags. These flag values would address erroneous data at the scan level.

Another quality control issue that is being addressed presently is the existence of SSM/I TDR and SDR NESDIS data filenames that claim to have 105 minutes of data, for example, but the physical size of those files indicate that there are only 30 kilobytes of data, when they should have file sizes that are more than 4 megabytes. Data files with such discrepancies are currently being grabbed by the read_TDR.pro program and placed in an “SSMI_bad_file” directory.

A suggestion was made by John Bates to check whether some of these errors resulted from a side lobe issue, where part of the SSM/I conical scanner sees the satellite platform itself. Analysis performed at the pixel level did not show such occurrences for the TDR or SDR data.