1997 HST Calibration Workshop
Space Telescope Science Institute, 1997
S. Casertano, et al., eds.

New Calibration Systems Projects at STScI

Stephen Lubow

Space Telescope Science Institute, Baltimore, MD 21218

Abstract. STScI has continued the development of systems for HST data calibration. The calibration database system (CDBS) tracks information about all HST calibration files. CDBS has been rebuilt to be more automated and provide some additional capabilities for users. We have also begun investigating the requirements for an on-the-fly calibration system. This system would allow users to retrieve data that are automatically calibrated at STScI with the most recently recommended calibration files and software.

1. Introduction

HST raw data are initially calibrated in a pipeline system at STScI. The raw and calibrated data are currently stored in the DADS archive and can be accessed remotely by GOs through the Starview user interface. In addition, users can obtain datasets plus newly recommended calibration files via Starview, and then recalibrate the data at their own sites using STSDAS software. The current system has met many user needs. However, some improvements are underway. The new calibration systems projects will provide better capabilities for managing calibration data at STScI and will offer more calibration services to GOs.

We describe below two projects for improving the STScI calibration systems: the rebuilt calibration database system (CDBS) and the on-the-fly calibration (OTFC) system.

2. CDBS

2.1. Description

The calibration database system (CDBS) plays a key role both in the initial calibration and in data recalibration by GOs. CDBS is a separate database system that maintains a record of all calibration files ever delivered for HST. It provides information to the calibration pipeline about which files should be used to calibrate an observation. In addition, it provides information to users via the Bestref system in Starview about the reference files that are currently recommended for each observation. (The Bestref screen is found by selecting "HST Instrument Searches" followed by "Reference" in Starview.) The CDBS system was recently completely redesigned and rebuilt. The new system is currently operational.

2.2. Data Flow

Calibration data files are prepared by instrument groups at STScI. The instrument groups also create load files that describe the contents of the calibration files. The data and load files are delivered to CDBS by the instrument groups. CDBS processes the load files and stores this information on a Sybase database system. Information about all calibration files ever used for HST is stored in this database. CDBS also undertakes several consistency checks (about a dozen) for possible errors in creating the load file or in processing the data.

CDBS generates code for updating a database in the calibration pipeline. The pipeline database contains a list of the currently recommended files to be used in the STScI initial calibration of observations. The data and code files are delivered to the pipeline system. The pipeline informs CDBS of when those data were loaded into the pipeline and into its database.

CDBS delivers the calibration data files to the DADS system for archival storage. The DADS system informs CDBS of when the data are archived.

The Bestref system in Starview provides GOs with information about which reference files are the currently optimal for calibrating a given observation. The Bestref system uses the CDBS database as the source of information for determining the recommended files. This selection involves considerations of the observation mode, the date of the observation, and the date of use (useafter date) for each possible calibration file.

2.3. New Capabilities

CDBS was redesigned and rebuilt in part to handle the greater complexity of the NICMOS and STIS instruments (Lubow et al. 1997). The CDBS data processing has been fully automated to minimize operator errors. The new database design is simpler and makes system maintenance and software development much easier than had been the case with the old system.

The new CDBS automatically determines whether a new file being delivered replaces some existing file currently in use. If so, the existing file is marked as out of use in the CDBS database and code is generated to remove that entry from the pipeline database. The information about the old file is kept in the CDBS database for tracking purposes.

Bestref was redesigned to use a new algorithm for determining the recommended files for each HST observation. The previous algorithm recomputed the recommended files for all HST observations every night. The computational demands of the system were rapidly exceeding the capacity of the hardware and the SM2 (second servicing mission) instruments would considerably worsen the situation. The new Bestref system each night recomputes the recommended files for only a small subset of HST observations, namely, those that can be affected by the newly delivered calibration files. As a result, the Bestref performance has been radically improved. The NICMOS and STIS data are being easily processed by Bestref.

The new CDBS database stores comment and level of change information about each data file. The following levels of change can appear on the Bestref screen

- N/A - No original or recommended file
- NO CHANGE - exactly the same files
- UNKNOWN - level of change not determined (only for preSM2)
- TRIVIAL - not important
- MODERATE - maybe important
- SEVERE - definitely important (recalibrate)

Levels TRIVIAL, MODERATE, and SEVERE are determined by CDBS software, based on information supplied by STScI instrument groups. Specifically, for each calibration file delivered, the instrument groups designate in a accompanying load file its level of change relative to some previously delivered related file, called the comparison file. The comparison file name is determined by CDBS. In the case of calibration files that are images, the level of change is specified by the instrument groups for the entire file. However, in the case of calibration files that are tables, the level of change is specified for each row of the

table. This row-level specification permits Bestref to determine whether a newly delivered calibration table contains of change of importance to an observation that uses a particular row of that table. In the case of deliveries of large tables with a small number of significant row changes, this capability becomes quite useful.

CDBS stores these change levels internally in its database by using *equivalence classes*. This method permits a rapid level of change determination between any two calibration files without following a chain of change levels among related files (i.e., transitivity problem, see Lubow et al. 1997).

2.4. New Features for Users

As a result of the new capabilities described above, some new features have been provided in the Bestref screens of Starview. Bestref now indicates the level of change for each recommended calibration file, relative to the file used to carry out the original calibration. In general, if one or more indications of SEVERE changes are listed, then recalibration is warranted. In practice, the level of change information is highly incomplete for pre-SM2 (before the second HST servicing mission of Feb '97) calibration files. A default change level of UNKNOWN is applied to such cases where incomplete change level information is available. Post SM2 calibration files, including all those for NICMOS and STIS, should have proper level of change determinations. The MODERATE change level suggests that recalibration may be needed, while the TRIVIAL change level usually indicates that recalibration is not necessary.

Important information about calibration files is now being stored as calibration comments in the CDBS database. These comments can be accessed through Starview. In general, complete comments are found in calibration files delivered after SM2. Currently, calibration file comments are available through Starview by selecting the "HST Instrument Searches", followed by "CALIBRATION" and by specifying the name of the calibration file. This feature is useful, though awkward for the purposes of most users, who would want a comprehensive historical list of comments for reference files that are relevant to some particular observation dataset. Such a capability is under development and should soon be available in the Starview Bestref screen. For a given observation, a user will be able to obtain all comments that were provided by instrument groups. These comments are selected from the database based on the date of observation, the types of files used to calibrate the observation, and the relevant mode of observation. The comments listed will include those for the original calibration files and those for the currently recommended files, as well as comments that may be found in several intervening files.

3. On-The-Fly Calibration

An on-the-fly calibration (OTFC) system at STScI would allow users to obtain data from the DADS archive which are calibrated with the most up-to-date calibration files and software. In most cases, this capability requires that data be automatically calibrated at STScI when the data are requested by a user. An OTFC system has been successfully developed for HST data by the CADC/ST-ECF (Crabtree et al. 1995). The requirements for an OTFC system at STScI are somewhat different from the CADC/ST-ECF, but their pioneering efforts provide STScI with a basis for development.

To date, we have conducted a preliminary investigation of the design issues and requirements for an OTFC system at STScI. Some of the main points are described below.

3.1. Advantages of OTFC

The HARP report (Hanisch et al. 1997) recommended that STScI pursue development of an OTFC system to reduce the level of data storage in the DADS archive. It may be possible to store only uncalibrated data in DADS and thereby not need the storage space

that would have been required for the calibrated data. WFPC2 calibrated data are a factor of ten larger than the uncalibrated data. Furthermore, the uncalibrated WFPC2 data compress very well (by a factor of ten), while the calibrated data do not compress well. The HARP report recommended that uncalibrated data be stored in compressed form and that calibrated data not be stored. The storage requirements for DADS are rapidly increasing with STIS and NICMOS and will become even greater with the Advanced Camera (ACS). The large advantages of OTFC for WFPC2 would likely be realized for ACS.

Another motivation comes from the scientific need to recalibrate data. About 2/3 of GOs recalibrate their data prior to analysis. Cameras such as WFPC2 produce dark files that typically should be retroactively used to calibrate observations taken in the prior week. Unfortunately, this means that the original calibration carried out in the STScI pipeline is not optimal. Furthermore, this less than optimal calibration is stored in the DADS archive. An OTFC system would help solve this problem by providing users with the calibration having the appropriate dark, as soon as it is available.

Currently, instruments that undergo evolution of calibration files or calibration software require users to carry out their own recalibrations at their home sites. It may well be easier for both the users and the STScI support groups to have STScI carry out the recalibration.

3.2. Design

The OTFC system should work closely along the lines of the current STScI calibration pipeline, so that both systems will produce nearly the same results. Users should be able to request OTF calibrated data electronically, much in the manner that data requests are currently carried out.

The OTFC system would handle each request for calibrated data by obtaining the raw data from the DADS archive. The raw data could then be calibrated using the appropriate STSDAS routines. Headers in the calibrated data will be updated to reflect the processing. The calibrated data will then be sent to the user.

To coordinate the processing activities, the OTFC system should employ the OPUS blackboard technology (Rose 1996). This system allows for parallel processing of tasks distributed over multiple cpus.

The system will need to function in a highly automated manner. The number of datasets calibrated per day in the OTFC system will greatly exceed the number of datasets calibrated by the current calibration pipeline system. The OTFC system will need to be robust and be able to automatically recover from problems with calibration processing. Possible problem datasets will be identified in advance of OTF calibration so that the system will be able to process them as best as is possible. In some cases, no calibration may take place, but the user will need to be informed of that outcome. The system will also need to handle cases where the calibration procedures have evolved over time and data might be missing from the headers that is needed to carry out processing with current software.

There are several approaches that can be taken to identify and resolve the above problems and special situations. One possibility is to store information about all problem observation datasets in a database, which the OTFC system would check before carrying out its calibration. This approach would require that appropriate parameters be identified to describe all such problems. Another approach would be to use a set of rules in a decision tree, as has been adopted by the CADC/ST-ECF (Crabtree et al. 1995). This approach works well when the rules can be simply stated with few exceptions.

3.3. Plans

In the near-term we expect to develop a prototype which will be heavily tested. We believe that it would to beneficial to have an OTFC system in place for the ACS, before SM3.

Our initial goal will be to produce a calibration system that will provide calibrations that are as good as would be obtained by users today who recalibrate manually. Beyond that, various improvements and customizations are possible.

Acknowledgments. I thank Colin Cox, Lee Hurt, Shaw-Hong Kao, Bernie Simon, Rick Whitman, for their assistance in CDBS development. I greatly benefited from the OTFC Working Group discussions with Faith Abney, Andy Fruchter, Lee Hurt, Robert Jedrzejewski, Marc Postman, Daryl Swade, Jay Travisano, and Rick White.

References

Crabtree, D. R., Durand, D., Gaudet, S., & Hill, N., 1995, *Astronomical Data Analysis Software and Systems V*, ASP Conf. Ser., Vol. 101, eds. G. H. Jacobi & J. Barnes (San Francisco: ASP), 505

Hanisch, R. J., Abney, F., Donahue, M., Gardner, L., Hopkins, E., Kennedy, H., Kyprianou, M., Pollizzi, J., Postman, M., Richon, J., Swade, D., Travisano, J., & White, R., 1996, *HARP: The Hubble Archive Re-Engineering Project Summary Report* (Baltimore: STScI)

Lubow, S., Cox, C., Hurt, L., & Simon, B., 1997, *Redesign of CDBS* (Baltimore: STScI)

Rose, J. F., 1996, *Astronomical Data Analysis Software and Systems VI*, A.S.P. Conf. Ser., Vol. 125, eds. G. Hunt & H. E. Payne (San Francisco: ASP), 38