Calibration Reference Data Recommendation Method Postprint

Authors: Zhao Bowei, Wei Shoulin, Juanjuan Ren, Wang Feng, Lingchen Xiaoji, Liu Chao

Date: 2025-07-02T00:00:00+00:00

Abstract

The efficient operation of automated astronomical data processing pipelines relies on precise calibration reference data recommendation mechanisms. Consequently, comprehensive calibration reference data systems have emerged. This study systematically reviews the calibration reference data recommendation methods employed by major international telescopes, and conducts an in-depth analysis of the advantages and disadvantages of each approach. It focuses on introducing a novel recommendation strategy based on textual rules and its accompanying calibration reference data system, as well as their flexibility and efficiency in automated data processing. Furthermore, it elaborates on the critical role and potential application value of this system in the scientific data processing of the Chinese Space Station Survey Telescope (CSST), and outlines its future development prospects. This research provides new ideas and methods for the recommendation of calibration reference data in astronomical data processing, holding significant theoretical and practical importance.

Full Text

Preamble

Vol. 43, No. 2

June 2025

PROGRESS IN ASTRONOMY

Calibration Reference Data Recommendation Methods

ZHAO Bowei[1,2], WEI Shoulin[3], REN Juanjuan[1], WANG Feng[4], LING Chenxiaoji[1], LIU Chao[1,2]

[1] Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[4] Center for Astrophysics and Great Bay Center of National Astronomical Data Center, Guangzhou University, Guangzhou 510006, China

Abstract

The efficient operation of automated astronomical data processing pipelines relies on precise calibration reference data recommendation mechanisms, which has led to the development of comprehensive calibration reference data systems. This paper systematically reviews the calibration reference data recommendation methods employed by major international telescopes and provides an in-depth analysis of their respective advantages and disadvantages. We focus on introducing a novel text-based rule recommendation strategy and its accompanying calibration reference data system, highlighting its flexibility and efficiency in automated data processing. Furthermore, we elaborate on the critical role and potential applications of this system in the scientific data processing of the China Space Station Survey Telescope (CSST), and discuss its future development prospects. This research provides new insights and methodologies for calibration reference data recommendation in astronomical data processing, holding significant theoretical and practical importance.

Keywords: Astronomical data processing; Calibration reference data; Space astronomical telescopes

CLC number: P171
Document code: C

1 Introduction

Astronomy is the discipline that studies various celestial objects and phenomena in the universe, and its research is highly dependent on observations. Through observations, abundant data can be obtained, which must undergo necessary processing before being used for measurements and analysis to reveal fundamental physical properties of celestial objects such as brightness, temperature, spectral type, radial velocity, chemical composition, and structure. These key data not only provide crucial support for in-depth exploration of the universe and the formation and evolution of various celestial objects, but also establish the core position of astrophysics in modern astronomy.

Astronomical observation techniques have undergone leapfrog development from naked-eye observations and hand-drawn sky maps to modern astronomical telescopes, gradually forming a comprehensive observation system that includes astrophotography, photometry, spectroscopy, full-band astronomy, and digital sky surveys. Meanwhile, astronomical telescopes have achieved tremendous progress from small to large apertures, from single to multi-mirror systems, and from ground-based to space-based platforms. Today, modern astronomical observations primarily rely on large-aperture telescopes equipped with advanced terminal instruments, providing unprecedented detection capabilities for astronomical research.

The application of astronomical telescopes extends beyond simply collecting observational data; their core objective is to advance scientific research and create new knowledge. Therefore, transforming raw observational data into scientific knowledge is a crucial process, in which data processing and analysis software serve as a bridge. Typically, raw astronomical observational data not only contains genuine signals from celestial objects but also includes atmospheric and instrumental effects, mixed with various sources of noise. Consequently, the initially acquired raw data must undergo complex processing procedures including instrumental effect correction, cosmic ray removal, and various calibrations before being transformed into scientific data suitable for astronomers' analysis.

However, as the complexity of modern astronomical data processing continues to increase, a single terminal instrument often involves multiple types of data, including scientific observational data and calibration data. For example, the raw data processing for X-shooter, the second-generation instrument on the European Southern Observatory's Very Large Telescope (VLT), involves over a hundred data types. Additionally, different astronomical telescopes are designed and built around specific scientific objectives from the outset. To achieve these objectives and provide scientists with high-quality scientific data, precise calibration of terminal equipment and scientific observational data is the most fundamental and critical step.

A complete calibration workflow includes calibration observations, calibration data processing (i.e., generating calibration reference files), and selecting the optimal calibration reference files for scientific data processing. Since obtaining ideal calibration observational data typically consumes significant observation time, and the allocation strategy for observation time directly affects the selection rules for calibration reference data in subsequent scientific data processing, a trade-off is inevitably required between conducting more calibration observations and dedicating more time to scientific observations. For ground-based astronomical observations, ideally, calibration data obtained on the same night should be used to calibrate scientific observational data from that night. However, if the terminal instrument exhibits high stability, calibration data from previous or subsequent nights can also be used to calibrate scientific data from the current night to improve observation efficiency.

Space astronomical telescopes operate outside Earth's atmosphere, enabling full-band, high-precision, long-duration observations that yield high-quality data.

However, the complex space environment can cause radiation damage to detectors, leading to changes in instrument performance that affect data processing and the quality of scientific products. Therefore, in-orbit calibration is crucial for space telescopes. Renowned international space telescopes such as the Hubble Space Telescope (HST), James Webb Space Telescope (JWST), and Nancy Grace Roman Space Telescope all have periodic in-orbit calibration observations. Furthermore, astronomical telescopes are typically equipped with multiple terminal instruments, each with unique observation modes requiring specific types of calibration observations. Different types of calibration observations usually have different monitoring cycles due to their varying objectives, further increasing the complexity of calibration reference data.

With the explosive growth of astronomical observational data volume, building automated scientific data processing pipelines has become an inevitable trend. To ensure accurate processing of scientific data, automated pipelines must not only automatically identify and select the best calibration reference files but also comprehensively consider factors such as scientific observation strategies and calibration schemes. Consequently, how to efficiently and accurately select the optimal calibration reference data during scientific data processing has become a complex and critical issue. In response to this need, several international telescope projects have developed specialized systems for recommending the best calibration reference data, which play a vital role, especially in more complex space telescope missions. Among them, JWST' s Calibration Reference Data System (CRDS) is particularly notable for its uniqueness and advancement. Based on HST' s long-term experience and to meet JWST' s data processing requirements, a calibration reference data system was developed with the core functions of building calibration reference file recommendation rules and automating the recommendation of the best calibration reference files, while also considering storage and management functions. This system serves as the core link between calibration reference files and scientific data processing pipelines, automatically recommending the best options from numerous in-orbit calibration reference files of multiple types and versions according to established rules for use by scientific data processing pipelines. Moreover, the text-based rule calibration reference data system offers extremely high flexibility and scalability, applicable not only to the selection of calibration reference files for astronomical telescope observational data but also to other scenarios requiring automated rule-based recommendations.

Section 2 of this paper systematically reviews and summarizes the main calibration reference data recommendation methods adopted by various international telescopes. Section 3 provides a detailed introduction to the currently more flexible and advanced text-based rule calibration reference data system. Section 4 focuses on the application of this system to the China Space Station Survey Telescope (CSST) scientific data processing. Section 5 concludes with a summary and outlook.

2 Historical Review

Based on the requirements of scientific data processing, during actual astronomical observations, telescopes periodically interleave various necessary calibration observations for different terminal instruments and observation modes. These periodically acquired calibration observation data, after processing, generate different versions of various calibration reference files. Therefore, during subsequent large-scale automated processing of astronomical observational data, how to automatically select the best files from numerous different versions of various calibration reference files for calibrating scientific observational data—i.e., how to automatically assign the best calibration reference files to scientific data processing pipelines—significantly affects data processing results and precision, requiring focused research.

Regarding calibration reference data recommendation methods, various telescope projects have adopted diverse solutions. A relatively simple method involves retrieving calibration reference files closest in observation time to the scientific data from a database using Structured Query Language (SQL) for use by scientific data processing pipelines. However, this method has numerous drawbacks, such as difficulty in reproducing historical processing results, excessive dependence on databases, inability to propagate for use, and difficulty in customizing reference files according to special requirements. Consequently, international astronomical telescopes have adopted different calibration reference data recommendation methods based on their actual needs.

The Chandra X-ray Observatory (CXO), launched by NASA in 1999 as the third satellite in the Great Observatories program, uses a method based on the "Chandra Calibration Database" (CALDB). This method is built according to the standard model of the CALDB system developed by NASA' s High Energy Astrophysics Science Archive Research Center (HEASARC), which is used by most high-energy astronomical telescopes. Although the name CALDB seems to imply a database, it does not essentially involve a real database. The core of the CALDB method is that the rules for selecting calibration reference files are stored in index files, which list all available calibration reference files and their usage conditions. While not simple to construct, this method has advantages such as reproducibility of past recommendations and software portability.

The Spitzer Space Telescope (SST), NASA' s final space telescope in the Great Observatories program launched in 2003, uses "caltrans" (short for calibration transfer) software for retrieving and selecting required calibration reference files during scientific data processing. Developed by the Spitzer Science Center (SSC) and implemented in C language, this software interacts with an Informix database. SST uses what is perhaps the purest database-based method, where selection rules for calibration reference files are essentially expressed as SQL queries on the database. Although this system supports simple update and rollback operations, it is difficult to reproduce previous recommendation results (reference files), not easily propagated for use, and cannot be used in other environments.

The Gemini Telescope, consisting of two 8-meter-class ground-based optical astronomical telescopes implemented by the Association of Universities for Research in Astronomy (AURA) and operational since 2000, employs a hybrid method. While most information is stored in a database, the selection of calibration reference files is performed through SQL queries generated by Python code. Simple SQL queries often cannot meet requirements in cases with complex selection rules, whereas this hybrid method can construct arbitrarily complex rules. Although it provides a web service for external users to select calibration reference files, the system does not allow propagation, does not support multiple versions of rules, and is not easily customizable by users.

The Hubble Space Telescope (HST), a large space optical telescope jointly developed by the United States and Europe and named after American astronomer Edwin Hubble, was launched in 1990 aboard the Space Shuttle and operates in low Earth orbit. To better support scientific data processing and analysis, particularly the operation of scientific data calibration pipelines, HST began planning and designing the Calibration Database System (CDBS) before launch. This system has been successfully used by HST for over 20 years, during which its working methods have been improved and upgraded while the overall design framework has remained essentially unchanged.

The basic design of CDBS is database-centric, with the database containing information on all types of calibration reference files. For each calibration reference file of all HST terminal devices, CDBS adds a file description containing terminal device observation mode parameters that define the exposure type used by the calibration reference file. This descriptive information is stored in separate database tables for each HST terminal device, with each table containing selectable parameters for all observation modes of the calibration reference file set for that terminal device. The calibration reference file description also includes a “UseAfter” timestamp, marking the earliest time the reference file should be used. Additionally, all selection rules for the best calibration reference files are stored in XML (eXtensible Markup Language) files, which specify the search conditions used when recommending reference files for scientific data.

The main function of CDBS is to query the database for various types of calibration reference files required for each new scientific dataset being processed (or that needs reprocessing) and return the filename of the best reference file, which is then updated to the value of a specific keyword in the scientific dataset header using the “bestref” program. The scientific data processing pipeline retrieves the required best reference files from the scientific dataset header keywords for calibrating the scientific data.

Although CDBS was successfully applied to HST for over 20 years, numerous limitations were identified during its long-term use. For example, it is difficult to test and inspect new reference files, hard to rollback submitted reference files containing errors, does not support remote use, does not meet users’personalized customization needs for reference file selection rules, and makes it difficult or nearly impossible to reproduce historical results. Since selection rules are stored in the database, the truly effective selection rules are difficult to understand. The database-centric design philosophy also limits the expansion of selection rule types, making it difficult to implement new types of selection rules. These limitations ultimately stem from CDBS' s dependence on databases, i.e., storing selection rules in databases.

Given these potential limitations of CDBS, the Space Telescope Science Institute (STScI) developed a new system for JWST called the Calibration Reference Data System (CRDS). This system uses text to encompass reference file selection rules, effectively solving the database dependency problem and achieving system portability. After successful trials on JWST, HST gradually transitioned from CDBS to CRDS. NASA' s next-generation flagship space mission, the Roman Space Telescope, will also use this calibration reference data system. Unlike general-purpose space astronomical telescopes such as JWST and HST, the Roman Space Telescope also has survey capabilities.

Table 1 summarizes the calibration reference data recommendation methods used by several international telescopes.

Telescope Name	Operational Period	Calibration Reference Data Recommendation Method
Chandra X-ray Observatory	July 1999–present	Calibration Database (CALDB)
Spitzer Space Telescope	August 2003–January 2020	Calibration Transfer (caltrans)
Gemini Telescope	2000–present	Hybrid method of database + Python queries
Hubble Space Telescope	April 1990–present	CDBS for first 24 years, CRDS currently
James Webb Space Telescope	Launched December 2021, planned 10-year operation	Calibration Reference Data System (CRDS)
Nancy Grace Roman Space Telescope	Expected launch 2026 or 2027, 5-year operation	Calibration Reference Data System (CRDS)

Section 3 will provide a detailed introduction to the new calibration reference data recommendation system, CRDS.

3.1 Overview

With the development of space astronomy, scientists have designed the Calibration Reference Data System (CRDS) for three major space telescopes—JWST, HST, and Roman—to better serve automated astronomical data processing. CRDS is a new system for managing and distributing calibration reference files. For scientific data processing pipelines, CRDS manages two categories of reference files: "data reference files" that include calibration data or information, and "parameter reference files" that contain various configuration parameter information used in scientific data processing pipelines.

Figure 1 shows the relationship between the Roman Space Telescope' s Wide Field Instrument (WFI) Level 2 data processing pipeline (used to generate calibrated single-exposure data) and CRDS. CRDS is called a "system" because, to support its core function of "best reference file recommendation," a complete service support platform must be built, ultimately making it a comprehensive "system" that includes Python libraries, command-line programs, web servers, and databases for managing and distributing all types of calibration reference files for all terminal instruments of an astronomical telescope, efficiently and flexibly serving automated scientific data processing.

CRDS was newly designed, abandoning the traditional solution of using databases to store calibration reference file selection rules. Instead, it adopts a simpler and more flexible approach: encompassing all selection rules for a version in a simple text file. This method offers several advantages: First, the textual format makes selection rules for various types of calibration reference files clear and easy to understand. Second, these rule files are version-controlled, with different versions not conflicting or causing confusion, making them easy to manage and enabling reproducibility of historical processing results. Third, text-format rule files are easy to propagate, allowing remote users to use them easily without installing complex software. Finally, the content of rule files is relatively simple, and users can conveniently customize and modify rules according to their actual needs when necessary, making it flexible and easy to operate.

To implement CRDS' s core function of "best reference file recommendation," the prerequisite is to build text-based rule files for various types of calibration reference data, i.e., to achieve best reference file recommendation based on these rule files. As a complete system, to support this requirement, CRDS needs to have multiple functions, including submission of new reference files while simultaneously reconstructing new rule files, best reference file recommendation, scientific data reprocessing, and various auxiliary functions such as reference file and rule file verification (see Table 2). Figure 2 shows an overall schematic of CRDS functions. Different functions involve different users. CRDS involves three main types of users: calibration scientists, scientific data processing pipeline operation scientists, and CRDS system managers. Calibration scientists primarily produce and generate various calibration products (i.e., calibration reference files) and submit them to CRDS, providing input for the system. Scientific data processing pipeline operation scientists are the users of CRDS' s best reference file recommendation function. CRDS system managers are responsible for maintaining and managing the entire CRDS system.

In addition to some web interactive tools, CRDS is primarily developed based on Python. To support the above functions, it implements script command-line tools, Python API (Application Programming Interface) interfaces, and simple visual web interfaces. CRDS software has client and server versions; the version released on GitHub is its client version, while the server-side software (mainly controlling file submission, management, and network services) is not publicly released for security reasons. Through CRDS client software or websites, users can access functions including reference file submission and inspection, best reference file recommendation, scientific data reprocessing, and various auxiliary functions (see Table 2).

Table 2 Detailed Introduction to CRDS Functions

Function	Description
bestrefs	Recommends the best calibration reference files for scientific datasets, with modes including file mode, reprocessing mode, rule file test mode, and comparison mode
certify	Performs verification checks on reference/rule files, including basic document format, semantics, parameter constraints, etc.
checksum	Adds, deletes, or verifies hash values (sha1sum) of rule files, or checksums and datasums of reference files
diff/rowdiff	Compares different versions of rule or reference files and returns differences; diff compares FITS table data column by column, while rowdiff specifically compares row by row
get_{synphot}	Downloads reference files mappable by a certain pmap version from CRDS server to local
query_{affected}	Automatically determines scientific datasets requiring reprocessing due to rule or reference file updates
list	Lists or queries various CRDS information, such as local CRDS configuration, reference/rule files on server or local cache, etc.

Function	Description
matches	Explains CRDS rules to find rule content matching a reference file
newcontext	Automatically generates new version context files (imap or pmap files)
refactor	Automatically reconstructs new version rmap files
submit/rc_{submit}	Submits reference files to CRDS server (requires CRDS account permission)
sync	Synchronizes reference/rule files from server to local
uniqname	Renames reference files with HST' s unique official CRDS names
uses	Views/lists dependent files of a reference/rule file

3.2 Selection Rule Files

The CRDS system revolves around selection rule files, which describe how to select or assign the most appropriate calibration reference files for scientific data. CRDS rule files are hierarchically designed text files. The hierarchical mapping relationship between rule files helps scientific observational data simply and clearly select the required best calibration reference files through step-by-step mapping. The rule hierarchy has three levels (see Table 3): pmap files (corresponding to the telescope pipeline level, one pmap per telescope), imap files (corresponding to the terminal instrument level of the telescope, one imap per instrument), and rmap files (corresponding to the calibration data type level of the instrument, one rmap per reference file type). These three rule files have hierarchical mapping/selection relationships and are therefore all called mapping files, with imap and pmap files also known as context files. Based on these rule files, CRDS implements a nested organizational structure corresponding to different levels of functionality.

Figure 3 shows the hierarchical structure of CRDS rule files. All observation mode reference files for a certain calibration type of a telescope' s instrument are included in versioned rmap files (see columns 3 and 4 in Figure 3). An instrument' s imap file collects a certain version of rmap files for all calibration types of that instrument, which constitute the effective recommendation version of the imap file (see columns 2 and 3 in Figure 3). The telescope pipeline-level pmap file collects a certain version of imap files for all instruments of that telescope (see columns 1 and 2 in Figure 3). In summary, CRDS primarily operates based on four types of files: three hierarchical mapping/rule files and the final calibration reference files recommended by mapping (see Figure 3).

Table 3 Introduction to CRDS Rule Files at Different Levels

Level	File Type	Description
Telescope pipeline level	pmap (pipeline context)	Manages all instruments of a telescope, used to map imap files
Instrument level	imap (instrument context)	Manages all reference file types of an instrument, used to map rmap files
Reference file type level	rmap (reference type mapping)	Manages all versions of reference files of a certain type, used to map specific reference files

The three types of hierarchical mapping/rule files in CRDS (pmap/rmap/imap) have the same structure and syntax, mainly consisting of two parts: a "header" section and a "selector" section (see Figure 4). The header section provides descriptive information about the mapping file, including file source, filename, telescope name, mapping file type, and information for verifying file integrity (see header content in Figure 4). A very important field is "parkey," which represents the dataset parameters used by the selector to find the best reference file (usually keywords from FITS file headers or JWST/Roman data model names). The selector includes matching rules for finding mapping results, typically in nested structures that can include high-level selectors and nested sub-selectors. Additionally, comments can be added between the header and selector sections as needed.

Regarding the "selector"in rule files, the selectors for pmap and imap rule files are relatively simple (see column 1 in Figure 4). The "parkey" field in pmap rule files is typically "INSTRUME," representing the instrument keyword, and its selector section simply matches and maps to imap filenames corresponding to different instruments. The "parkey" field in imap rule files is typically "REFTYPE," representing the calibration reference file type keyword, and its selector section simply matches and maps to rmap filenames corresponding to different reference file types of that instrument. In contrast, rmap rule files are more complex (see column 2 in Figure 4). The "parkey" field settings include multiple keywords depending on the situation, and its selector is typically a nested hierarchy to match various header file keywords.

CRDS rule file selectors support various selection rules including exact matching, various expression matching (Match), UseAfter (start usage time) selection, software version selection, closest time selection, geometrically nearest selection, and bracket selection, as detailed in Table 4. Among them, Match and UseAfter are the most commonly used selectors (see selector parts in Figure 4). Depending on scientific objectives, observation strategies, calibration schemes, and other factors, different telescopes' scientific data processing pipelines may have different selection rules for reference files. Additionally, if scientists or other users have more complex reference file selection needs, they can add and use new selection rules by modifying the content of rule files in the locally installed CRDS path.

Table 4 Introduction to Selector Rules in CRDS Rule Files

Rule Type	Description
Match	Finds tuples that best match scientific data header information, supporting exact matching, enumeration lists, wildcards, regular expressions, literal expressions, relational expressions, range expressions, exclusion expressions, etc.
UseAfter	Finds the reference file with the most recent time earlier than the scientific data observation time
SelectVersion	Uses pipeline software versions and various relational expressions for selection; different versions of scientific data processing software may use different versions of reference files
ClosestTime	Selects the most recent time
GeometricallyNearest	Selects the nearest distance
Bracket	Returns two selection results

3.3 Reference File Submission and Verification

The prerequisite for CRDS to recommend the best reference files is to properly store and manage submitted calibration reference files and reconstruct corresponding rule files when submitting reference files (see CRDS submission function in Figure 2). Since space telescopes produce large amounts of multi-version, multi-type reference files updated over time during in-orbit calibration—especially daily monitoring items such as bias and dark current with daily update frequencies—and because data processing algorithms are updated during in-orbit operations, reference file generation is also continuously updated, reference file submission is frequent and may even involve batch submissions.

Reference file submission generally falls into two categories: adding a new type of reference file that did not exist before, or replacing an existing reference file (e.g., having a better flat-field reference file); and adding a new version of an existing type of reference file for processing scientific observational data from a specific time interval.

Since a problematic single reference file can affect the processing of large batches of scientific data, reference file quality verification is crucial. Reference files submitted to CRDS must be valid and absolutely error-free. Therefore, when submitting calibration reference files to CRDS, a series of strict verifications/checks are performed (see CRDS verification function in Table 2). Only reference files that pass verification can be successfully submitted to the CRDS server and trigger reconstruction of new version rule files. Additionally, newly reconstructed rule files must also pass verification before being stored on the CRDS server for scientific data use.

Reference file and rule file verification checks are primarily based on constraint files with the .tpn suffix (.tpn constraint definition files) defined in CRDS. CRDS' s unique tpn constraint files define various checks required for reference/rule files (rmap files). For HST, tpn files define almost all CRDS checks, inherited from the original CDBS checks, which is also the source of CRDS tpn file syntax. For JWST/Roman, tpn checks are used to further expand JWST/Roman data model checks, for example, by adding "required" check options, checking matrix dimensions, checking relationships between keywords, and distinguishing acceptable values for different backend modules. The tpn file syntax supports explicit, synthetic, include, replace, and constraint instructions. Currently, the constraint instruction is most commonly used, with each line defining a constraint condition check typically including five fields: , , , (whether it must exist), and (a series of values separated by commas, using Python expressions), with fields separated by spaces. Through CRDS' s internally defined tpn files, comprehensive checks can be performed on reference files, from header keywords to data matrices and tables.

In addition to CRDS-specific tpn format checks, CRDS verification (the "certify" function) also includes: JWST/Roman data model checks, fitsverify checks, table row checks, FITS/ADSF/JSON/YAML format checks, rmap rule file update checks, and file hash value verification. When submitting reference or rule files, warnings or errors are issued if problems are found that prevent passing checks. Users can modify reference/rule files based on warning or error information to complete submission. Only reference files that pass checks are version-renamed and used to build hierarchical rule files, which are stored on the CRDS server for scientific users.

3.4 Best Reference File Recommendation

The core function of CRDS is to recommend the required best calibration reference files for scientific data processing based on constructed rule files, i.e., the "bestrefs" function (see Figure 2 and Table 2). Calibration of a single scientific observation image typically requires many types of calibration reference files, which vary depending on the terminal instrument, observation mode, etc., used to acquire the scientific data. Therefore, as calibration data products, calibration reference file header information contains keywords recording relevant information about the reference file, including the actual observed terminal instrument, configuration parameters, reference file type, usage time, production time, and keywords specific to different data types. CRDS uses this reference file keyword information to reconstruct and generate new version rule files, then assigns the best calibration reference files through hierarchical mapping according to the rule hierarchy structure.

Pmap rule file updates set the corresponding CRDS configuration for scientific data processing pipelines at a certain time point (i.e., the currently effective rule file version). CRDS reads the header information of the scientific data to be processed and, based on the selected pmap file: first matches and maps to the imap file (included in the selector of this pmap file) according to the value of the "INSTRUME" keyword (the content of the "parkey" field in the pmap file); then matches and maps to the rmap file (included in the selector of this imap file) based on the required reference file type "REFTYPE" (the content of the "parkey" field in the imap file) (if "REFTYPE" is not set, it matches all types of rmap files included in this imap file); finally recommends the best calibration reference file based on the selection rules and "parkey" field in the rmap file (see Figures 3 and 4).

For HST, calling CRDS' s "bestrefs" function recommends the best calibration reference files and updates the reference filenames in the scientific data headers for subsequent scientific data processing. For JWST and Roman, CRDS is fully integrated into their scientific data processing pipelines, which automatically map and select the best calibration reference files at runtime based on the selected pmap file. The pmap file used defaults to the operational pmap file, but other older versions can also be selected as needed. Therefore, when publishing articles using JWST data, in addition to specifying the version of the JWST scientific data processing pipeline used, the pmap file version and CRDS client software version must also be stated, as these determine the calibration reference files used for data processing. Additionally, for traceability, the used pmap file version and best calibration reference filenames are finally recorded in the header keywords of the processed scientific data.

Using the best reference file function requires network connection to the CRDS server and configuration of key environment variables, including CRDS_{{{SERVER}}*{{URL}}} (CRDS server URL) and CRDS*{PATH} (local path for CRDS cache), to download the required CRDS rule files and reference files to local for data processing. This also allows users to conveniently view and flexibly modify rule files to meet customized needs for calibration reference file usage in data processing.

3.5 Scientific Data Reprocessing Function

Calibration reference files and raw scientific observational images together serve as important inputs to scientific data processing software, determining the quality of scientific data products. Scientific data inevitably requires reprocessing, typically due to updates in data processing algorithms or calibration reference files. For space telescopes, instrument degradation caused by the environment requires periodic in-orbit monitoring, which generates new version calibration reference files that change over time. Additionally, updates to calibration schemes and calibration product generation algorithms during in-orbit operations also lead to updates in calibration reference files. Therefore, calibration reference files need frequent updates and submission to CRDS, and corresponding rule files also need synchronous updates.

Since CRDS recommends the best calibration reference files for scientific data based on rule files, once selection rules are updated or new reference files are added, the recommended reference files for the same scientific data will inevitably change, meaning previously processed scientific data may need reprocessing.

The reprocessing function is also an important and novel feature of CRDS. CRDS' s reprocessing function typically works with archived databases to determine scientific datasets requiring reprocessing due to new version reference file submissions. When a scientific data processing pipeline selects a new version pmap file, CRDS automatically calculates the scientific data needing reprocessing due to rule file changes, thereby automatically triggering scientific data reprocessing (see CRDS reprocessing function in Figure 2). The reprocessing function calculates affected scientific datasets by comparing reference files assigned by old and new rule files. CRDS can then automatically determine which processed scientific datasets need reprocessing. After confirmation, the reprocessing system stores logs and recommended reprocessing datasets, which can be obtained via e-mail, client programs, or web interfaces. Finally, scientific data processing pipeline operation scientists or scientific data users decide whether to reprocess the affected scientific data and execute the reprocessing operations.

3.6 Other Auxiliary Functions

In addition to building selection rule files and assigning best reference files based on hierarchical rule files, CRDS provides diverse auxiliary tools to facilitate rule file and reference file management. These auxiliary functions include: downloading and management of CRDS rule files/reference files/status information, constraint checks and format checks, file comparisons, content viewing, dependency viewing, etc., as detailed in Table 2. These functions and tools enhance the usability and manageability of the CRDS system.

Additionally, as a complete system, CRDS has a dedicated network server side to facilitate the implementation of various functions. The CRDS server side primarily manages CRDS rule files, reference files, and metadata. Different telescopes have their own CRDS server URLs. The CRDS server side provides various network services through JSONRPC and HTTP interfaces, including reference file/rule file submission and archiving, best reference file recommendation, reprocessing functions, and file propagation downloads, to support various functions of client software. Figure 5 shows the process of scientific data processing pipelines obtaining the best calibration reference files from the CRDS server side.

4.1 CSST Introduction

The China Space Station Survey Telescope (CSST) is China’ s first large-scale space optical astronomical telescope, with a primary mirror aperture of 2 meters and an effective field of view of 1.72 square degrees. It will operate in low Earth orbit, co-orbiting with the space station for maintenance and upgrades, with a planned 10-year operational lifetime. The survey telescope consists of five backend modules (multicolor imaging and slitless spectroscopy survey camera, multichannel imager, integral field spectrograph, exoplanet imaging coronagraph, and terahertz spectrometer) (see Figure 6), featuring large-field survey observations, various precision astronomical measurements, and deep field/ultra-deep field observations. CSST is expected to achieve major breakthroughs in multiple research fields including cosmology, galaxies and active galactic nuclei, the Milky Way and nearby galaxies, stellar science, exoplanets and solar system objects, astrometry, and transient and variable sources, making it China’ s flagship space astronomy project.

Note: Image from Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences.

To achieve “large field of view, high image quality, and multi-band” capabilities, CSST’ s five first-generation observation terminal instruments are all highly distinctive. The main survey camera (MSC) has a focal plane composed of 30 CCDs (charge-coupled devices) of 9,000$\times 9,000 pixels, with a total of 2.5 billion pixels, of which 18 detectors are used for m
wave infrared module has 8 detectors of 640\times$512 pixels, with 4 used for near-infrared imaging and 4 for slitless spectroscopy. The multichannel imager (MCI) can conduct simultaneous observations in three channels—ultraviolet, visible, and near-infrared—with each channel equipped with filter wheel components that can subdivide bands and detectors consistent with the survey module. The integral field spectrograph (IFS) adopts an improved image slicer-based field integral unit. Since space observations are not limited by atmospheric seeing, this instrument can achieve high spatial resolution imaging spectroscopy detection across the entire optical band. The exoplanet imaging coronagraph (CPI-C) uses a technique called coronagraphy, which suppresses stellar brightness by introducing an occulter or mask in the field of view

around the star, enabling the faint light of surrounding planets to be observed. This coronagraph is mainly used for observing exoplanets in the visible to near-infrared bands. The high-sensitivity terahertz detection module (HSTDM) detects in the terahertz frequency band between infrared and microwave, used for wide-band molecular line searches, formation and evolution of Milky Way molecular clouds, fine observations of neutral carbon in extragalactic galaxies, and studies of cold gas in the nearby universe.

4.2 CSST Calibration Data Recommendation System

The CSST scientific data processing system is a specialized system for survey telescope data processing, with main tasks including observational data simulation, scientific observation requirement planning, and data processing. Among these, data processing is the core task of the CSST scientific data processing system, requiring processing of raw data (Level 0 data) from all terminal instruments to generate Level 1 and Level 2 data that meet scientific user needs, ultimately providing important observational basic data for scientific research. CSST' s many characteristics—including multiple backend modules, numerous observation modes, simultaneous multicolor imaging and slitless spectroscopy for the survey module, large field of view and mosaic CCDs, and diverse scientific objectives—combined with complex space environment factors, pose great difficulties for scientific data calibration, resulting in extremely complex calibration data and challenges for calibration reference data management and allocation. Therefore, developing a calibration reference data recommendation system specifically for CSST (CSST Calibration Data System, CCDS) is one of the core tasks during the development of the scientific data processing system. This system must not only address the complexity of calibration data during CSST' s in-orbit operation but also be able to orderly manage large numbers of various calibration reference files that change over time. By establishing reasonable recommendation rules, the system will ensure that the best calibration reference files can be efficiently and accurately recommended to scientific data processing pipelines under the complex in-orbit working modes of the space telescope.

Based on CSST' s actual requirements, calibration products can be obtained through two methods: calibration product generation toolkits and calibration product generation pipelines. The calibration product generation toolkits mainly include generation of calibration products for the survey module where monitoring frequency is not very high, and various calibration products for precision measurement modules. These generated calibration products are submitted to CCDS after manual inspection to ensure they are error-free, i.e., through manual methods via ordinary web pages or Python APIs. However, daily monitoring items for survey module detectors (such as bias, dark current, and internal flat fields) require high monitoring frequency and involve large data volumes, necessitating development into calibration product generation pipelines to achieve automatic generation, inspection, and automatic submission to CCDS, which must be performed in parallel—posing new challenges for CCDS. Additionally, generation of some products in the calibration product generation pipeline requires calling and depending on other calibration products. For example, dark current products depend on combined bias for generation, while internal flat-field products depend on both combined dark current and combined bias, making the related dependency logic increase the complexity of the CCDS system. Figure 7 shows the current interaction schematic between the developing calibration product generation pipeline and CCDS. It is evident that meeting the needs of large-scale automated calibration product generation and submission places higher demands on CCDS. Moreover, considering the diversity of CSST terminal instruments and data, calibration product data types will also be very diverse. Therefore, in addition to supporting reference data formats such as FITS/ADSF/JSON/YAML, CCDS will also expand to support more reference file formats, including TOML/PICKLE format files. As the CSST scientific data processing system develops and calibration product generation toolkits/pipelines are developed, requirements for CCDS will continuously upgrade, making CCDS more expandable.

---

5 Summary and Outlook

Calibration reference data recommendation systems play an important role in automated astronomical data processing, automatically assigning the best calibration reference files for scientific data from different periods, terminal instruments, and observation modes. Through decades of international research and practice, comprehensive calibration reference data systems integrating calibration reference data management, rule formulation and reconstruction, and automated reference file recommendation have gradually developed. These systems have been implemented on internationally renowned space telescopes such as JWST/HST and the planned Roman Space Telescope, and can be widely and flexibly applied to calibration reference data management and allocation for survey and general-purpose telescopes, as well as support for automated scientific data processing pipelines.

With the continuous development of observational astronomy, various survey and medium-to-large aperture telescopes are emerging, and the need for automated processing of massive astronomical observational data makes research and construction of comprehensive calibration reference data management and allocation systems critically important. Particularly for telescopes like CSST, which not only feature a survey module capable of high-resolution multicolor imaging and slitless spectroscopy surveys over large sky areas, but also integrate multiple precision measurement modules for various fine astronomical observations, the complexity makes calibration reference data management and allocation extremely challenging, urgently requiring a comprehensive calibration reference data system. With continuous development in computer science and interdisciplinary fields, calibration reference data systems can be continuously optimized and improved with more advanced network technologies and artificial intelligence technologies, supporting rapid upload and management of large batches and ultra-large volumes of reference files, intelligent construction of selection rules and reference file recommendation, making the entire astronomical data processing workflow increasingly efficient and intelligent, forming an intelligent link from observational data to scientific objectives, and ultimately achieving the goal of optimizing telescope scientific output.

This research is supported by the China Manned Space Engineering CSST Special Scientific Research Fund, project number CMS-CSST-2025-A19. We thank the China Manned Space Station Engineering CSST Scientific Data Processing System for supporting this work, and we appreciate the valuable comments from the reviewers.

References

[1] Harris D. *Astronomy: A Study of Celestial Objects*. New York: Murphy & Moore Publishing, 2022: 10
[2] "China Discipline and Frontier Field Development Strategy Research (2021–2035)" Project Team. *China Astronomy 2035 Development Strategy*. Beijing: Science Press, 2023: 8
[3] Huang Youran, Xu Aiao, Tang Yuhua, et al. *Observational Astrophysics*. Beijing: Science Press, 1987: 3
[4] Xiao Naiyuan, Xuan Huancan. *Illustrated History of Astronomy*. Nanjing: Nanjing University Press, 2012: 5
[5] Cheng Jingquan. *The Era of Giant Telescopes: Modern Optical Astronomical Telescopes*. Nanjing: Nanjing University Press, 2023: 1
[6] Djorgovski S G, Mahabal A A, Drake A J, et al. *Planets, Stars and Stellar Systems*. Dordrecht: Springer Reference, 2013: 223
[7] Ferguson H C, Greenfield P, Axelrod T, et al. *astro2010: The Astronomy and Astrophysics Decadal Survey*. Washington: The National Academies Press, 2010: 15
[8] Léna P, Rouan D, Lebrun F, et al. *Observational Astrophysics*. 3rd ed. Heidelberg: Springer-Verlag, 2012:
[9] Freudling W, Romaniello M, Bramich D M, et al. *A&A*, 2013, 559: A96
[10] Vernet J, Dekker H, D' Odorico S, et al. *A&A*, 2011, 536: A105
[11] Zombeck M V. *Handbook of Space Astronomy and Astrophysics*. Third Edition. Cambridge: Cambridge University Press, 2007: 597
[12] Zhan H. *ChSBu*, 2021, 66: 1290
[13] Gu Y D. *ChSBu*, 2022, 37: 1031
[14] Hands A D P, Ryden K A, Meredith N P, et al. *Space Weather*, 2018, 16: 1216
[15] Doxsey R. *SpaceOps 2006 Conference*. Rome: AIAA, 2006: 5936
[16] Yoshioka K, Miyoshi Y, Kurita S, et al. *Space Weather*, 2021, 19: e02611

[17] Bohlin R C. *AJ*, 2016, 152: 60
[18] https://www.stsci.edu/hst/instrumentation/wfc3/calibration-plan, 2025
[19] https://www.stsci.edu/hst/instrumentation/acs/calibration/calibration-plans, 2025
[20] Menzel M, Davis M, Parrish K, et al. *PASP*, 2023, 135: 058002
[21] Rigby J, Perrin M, McElwain M, et al. *PASP*, 2023, 135: 048001
[22] https://www.stsci.edu/jwst/science-execution/approved-programs/calibration, 2025
[23] Freudling W, Zampieri S, Coccato L, et al. *A&A*, 2024, 681: A93
[24] Gong Q, Bergkoetter M, Berrier J, et al. *Journal of Astronomical Telescopes, Instruments, and Systems*, 2020, 6: 045008
[25] Bailey V P, Bendek E, Monacelli B, et al. *Proc. SPIE*, 2023, 12680: 126800T
[26] Scaramella R, Amiaux J, Mellier Y, et al. *A&A*, 2022, 662: A112
[27] Euclid Collaboration, Paterson K, Schirmer M, et al. *A&A*, 2023, 674: A172
[28] Greenfield P, Miller T. *Astronomy and Computing*, 2016, 16: 41
[29] Graessle D E, Evans I N, Glotfelty K, et al. *SPIE*, 2006, 6270: 62701X-1
[30] Lee W, Laher R, Fowler J W, et al. *ASPC*, 2005, 347: 594
[31] https://www.stsci.edu/stsci/meetings/calhst/home.html, 2025
[32] Li J, Cui C Z, He B L, et al. *Progress in Astronomy*, 2013, 31: 1
[33] Werner M W, Roelling T T, Low F J, et al. *ApJS*, 2004, 154: 1
[34] Fazio G G, Hora J L, Allen L E, et al. *ApJS*, 2004, 154: 10
[35] Hirst P, Cardenes R. *Proc. SPIE*, 2016, 9913: 99131E
[36] Jenkner H. *European Southern Observatory Conference and Workshop Proceedings*, 1988, 28: 357
[37] Cox C R, Tullos C. *Proc. SPIE*, 1993, 1945: 69
[38] Cox C R, Lubow S, Tullos C. *Observatory Operations to Optimize Scientific Return*, 1998, 3349: 218
[39] Swam M S, Lubow S, Hurt L. *Astronomical Data Analysis Software and Systems*, 2004, 314: 824
[40] Spergel D, Gehrels N, Baltay C, et al. https://arxiv.org/abs/1503.03757, 2025
[41] https://readthedocs.org/projects/jwst-pipeline/downloads/pdf/latest/, 2025
[42] https://roman.gsfc.nasa.gov/science/roses/Roman_{{{WFI}}*{{Processing}}*}{v6}.pdf, 2025
[43] https://hst-crds.stsci.edu/static/users_{guide}/index.html, *CRDS User Manual*, 2025
[44] Boyett K, Mascia S, Pentericci L, et al. *ApJ*, 2022, 940: L52
[45] Withers S, Muzzin A, Ravindranath S, et al. *ApJ*, 2023, 958: L14
[46] Gao M, Zhao G H, Gu Y D. *Bull Chin Acad Sci*, 2015, 30: 721
[47] Zhan H. *Sci Sin-Phys Mech Astron*, 2011, 41: 1441
[48] CSST Science Team. *CSST Science White Paper*, 2024: 16
[49] Manned Space Station Engineering CSST Scientific Data Processing System. *Software System Design Description*, 2022: 5