© Indian Academy of Sciences

DATA PIPELINE

AstroSat/SSM data pipeline

B. T. RAVISHANKAR[1,*] , S. VAISHALI[1], D. BHATTACHARYA[2], M. C. RAMADEVI[1], ABHILASH SARWADE[1] and S. SEETHA[3]

[1]Space Astronomy Group, ISITE Campus, U.R. Rao Satellite Centre, Outer Ring Road, Marathahalli, Bengaluru 560 037, India.
[2]Inter-University Centre for Astronomy and Astrophysics, Ganeshkhind, Pune 411 007, India.
[3]Department of Astronomy and Astrophysics, Raman Research Institute, Sadashivanagar, Bengaluru 560 080, India.
*Corresponding Author. E-mail: ravibt@ursc.gov.in

Abstract. The data pipeline at the Payload Operation Centre (POC) of the Scanning Sky Monitor (SSM) onboard AstroSat involves: (i) fetching the Level-0 data from the Indian Space Science Data Centre (ISSDC), (ii) Level-0 to Level-1 data processing followed by Level-2 data generation, and (iii) transfer of the Level-1 and Level-2 data back to ISSDC for dissemination of the re-packaged Level-2 data products. The major tasks involved in the generation of Level-1 and Level-2 data products are: (a) quality checks; time, alignment corrections, (b) temporal-HK plots generation, and, (c) image processing; light curve generation. The typical turn around time for this fully automated pipeline is about 25 min for one orbit data. In this paper, details of all the stages of this data pipeline are discussed.

Keywords. Fully automated data pipeline—X-ray astronomy payload—coded mask data processing.

1. Introduction

Scanning Sky Monitor (SSM) (Ramadevi *et al.* 2017) aboard the multi-wavelength astronomy mission AstroSat (Agrawal 2006; Singh *et al.* 2014) is an assembly of three coded mask cameras, monitoring the X-ray sky in 2.5–10 keV energy band (see Fig. 1). Each of the SSM cameras has one-dimensional Coded Mask optics and gas-filled proportional counter detector. The two edge cameras (SSM1 and SSM2) are canted away by 45° from the base plane of the central camera (SSM3), and further, they are also inclined in the canted plane by +12° and −12° in order to look at different parts of the sky with some overlap in the fields of view. The whole assembly is capable of rotation on an axis near-parallel to that of the SSM3 camera as shown, and the assembly is

mounted such that this is also along the +Yaw axis of the spacecraft. The stepper motor based rotation mechanism is designed to undertake one complete rotation from 5° to 355° and reverse in stare-and-step mode in order to observe different parts of the sky, with a typical stare time of 10 min and step-angle of 10°.

Among the three SSM cameras, the two edge cameras, SSM1 and SSM2 are of the same dimension and the central SSM3 camera has different dimensions (for accommodating them on the spacecraft structure). Therefore the imaging properties of the fields of view, angular resolution, etc., of SSM3 camera are also different. All these are listed in Table 1.

SSM being an X-ray sky monitor the data down-linked at the Ground Station in every visible orbit is subjected to immediate processing. The highest data rate available for processing is once every ~ 1.5 h, the orbital duration, and the data-sets are transferred from the data server at ISSDC to the POC of SSM located at the Space Astronomy Group of U. R. Rao Satellite

---

This article is part of the Special Issue on "AstroSat: Five Years in Orbit".
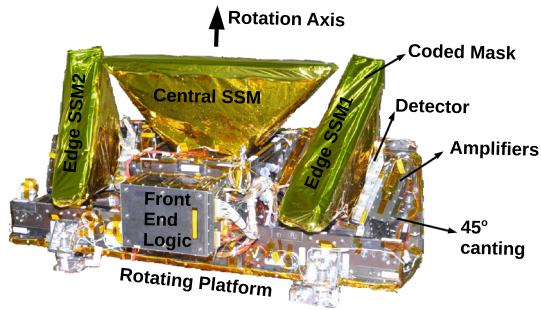
Figure 1. SSM assembly flight model of the three SSM cameras with their respective electronics packages mounted on a rotating platform.

Centre (URSC). Therefore complete automation of the entire processing is essential.

This paper describes the data levels and the processing undertaken to generate each level.

2. Data architecture and data pipeline

Three different levels are identified for SSM data, namely, Level-0, Level-1 and Level-2. The data-flow and the interfaces for the three levels of data are depicted in the overall architecture in Fig. 2. The Level-0 data consist of the binary payload data along with the necessary auxilliary data for the same UT range. The payload HBT (high bitrate telemetry) data, including health (also termed housekeeping or HK parameters) in binary format within the Data Handling package's envelope are generated for the three SSM cameras separately. The Level-0 datasets received at the POC are processed to generate higher level products – Level-1 data followed by Level-2 data. The Level-1 data consist of intermediate products including raw event lists and HK parameters – all UT-tagged for different energy bands and for different identified stare-durations for each of the three SSM cameras.

The Level-2 data consist of flux estimates obtained after image reconstruction with calibrated, anode-response corrected, clean event-lists per stare per energy-band. The Level-1 data as well as the incremental Level-2 data consisting of observational updates to the sources detected, are transferred from the POC to ISSDC. At ISSDC, the incremental Level-2 dataset is loaded in a database to generate the final data products via a custom built web application.

2.1 *Data transfer and pipeline automation*

The data transfer operations and the pipeline processing are controlled by two sets of shell-scripts which function in tandem. The data transfer script is programmed to handle transfers of different kinds of files and in either direction between the ISSDC and the POC.

The data fetching cron jobs running at the POC periodically poll the Data server at ISSDC for the availability of new Level-0 data files, and transfer the files to the POC over a dedicated Virtual Routing and Forwarding (VRF) meant for AstroSat over National Knowledge Network (NKN). Likewise separate directories are identified at the ISSDC data server to post the Level-1 and Level-2 data products. The data transfer scripts utilise *Expect* (which in turn is based on *TCL/Tk*, (Libes & O'Reilly 1995) for automating the transfer dialogue. In order to address possible link and server issues upon unsuccessful transfers, additional transfer attempts are built in the design. All Level-0 data files transferred are also verified for MD5SUM tags for possible transmission bit-errors. The cron jobs monitoring the ISSDC areas for different types of files to be transferred to the POC make sure that multiple instances are not run if by chance a previously initiated transfer has not ended owing to network delays or server issues.

Table 1. Specifications of the three SSM cameras.

Parameter	SSM1/SSM2	SSM3
Mask plate width (x), mm	60	60
Mask plate length (y), mm	500	634
Height, mm	251.4	306.4
Mask plate thickness, mm	3	3
Inter-pattern gap, mm	4	4
Window-support rod diameter, mm	2	2
Calibration wire diameter, mm	1	1
Field-of-view (FoV) ($x \times y$)	$26.8° \times 100°$	$22.1° \times 100°$
Angular resolution (x, y)	$10' - 13'$, $2.5°$	$10' - 13'$, $2.5°$
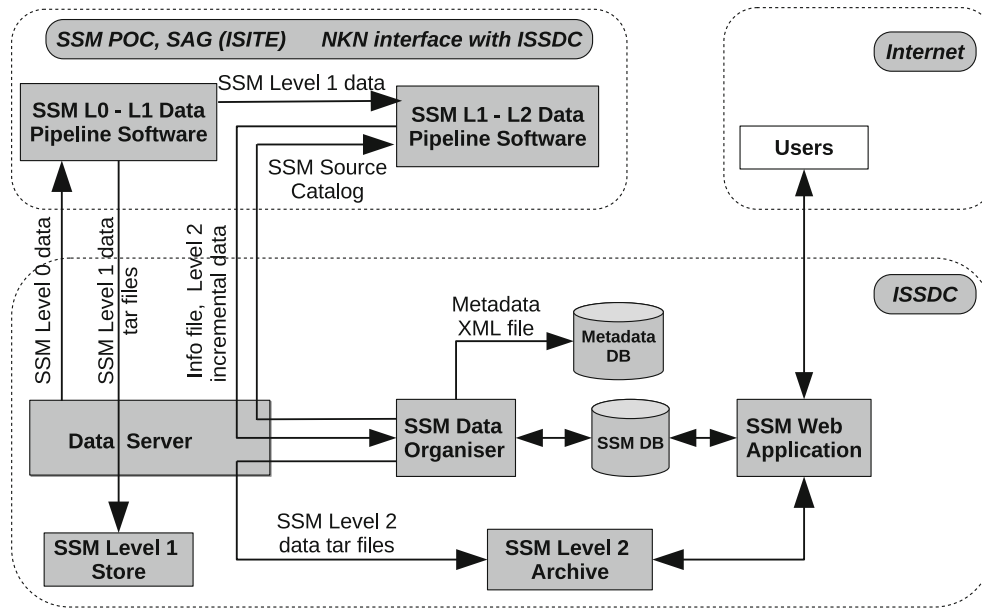
Figure 2. Overall architecture of the SSM data pipeline with interface details between the data servers at ISSDC and the SSM POC at URSC.

The Level-0 to Level-2 pipeline is developed as a shell-script calling individual modules in two broad stages to produce Level-1 output first and use that to subsequently generate Level-2 data products. The modules are developed primarily in *ANSI-C* and *java*. For some other different stages, *awk*, *sed*, *pgplot*, *gnuplot* are also employed in the automated pipeline.

The data transfer and data processing daemons interact indirectly via the *inotify* (Love 2005) API which monitors specified directories to trigger Level-0 data processing, and Level-1 and Level-2 data transfer, after attaching MD5SUM hashkeys.

For all received Level-0 data, along with the data products a quality report is also generated based on the following criteria: (a) whether all necessary files are present, (b) possible time-lag between the instrument data and attitude data, (c) the RS-decoding parameters during the dump of the data of each of the three cameras, (d) presence of outliers among instrument house-keeping parameters, and, (e) any pipeline errors reported during the course of execution.
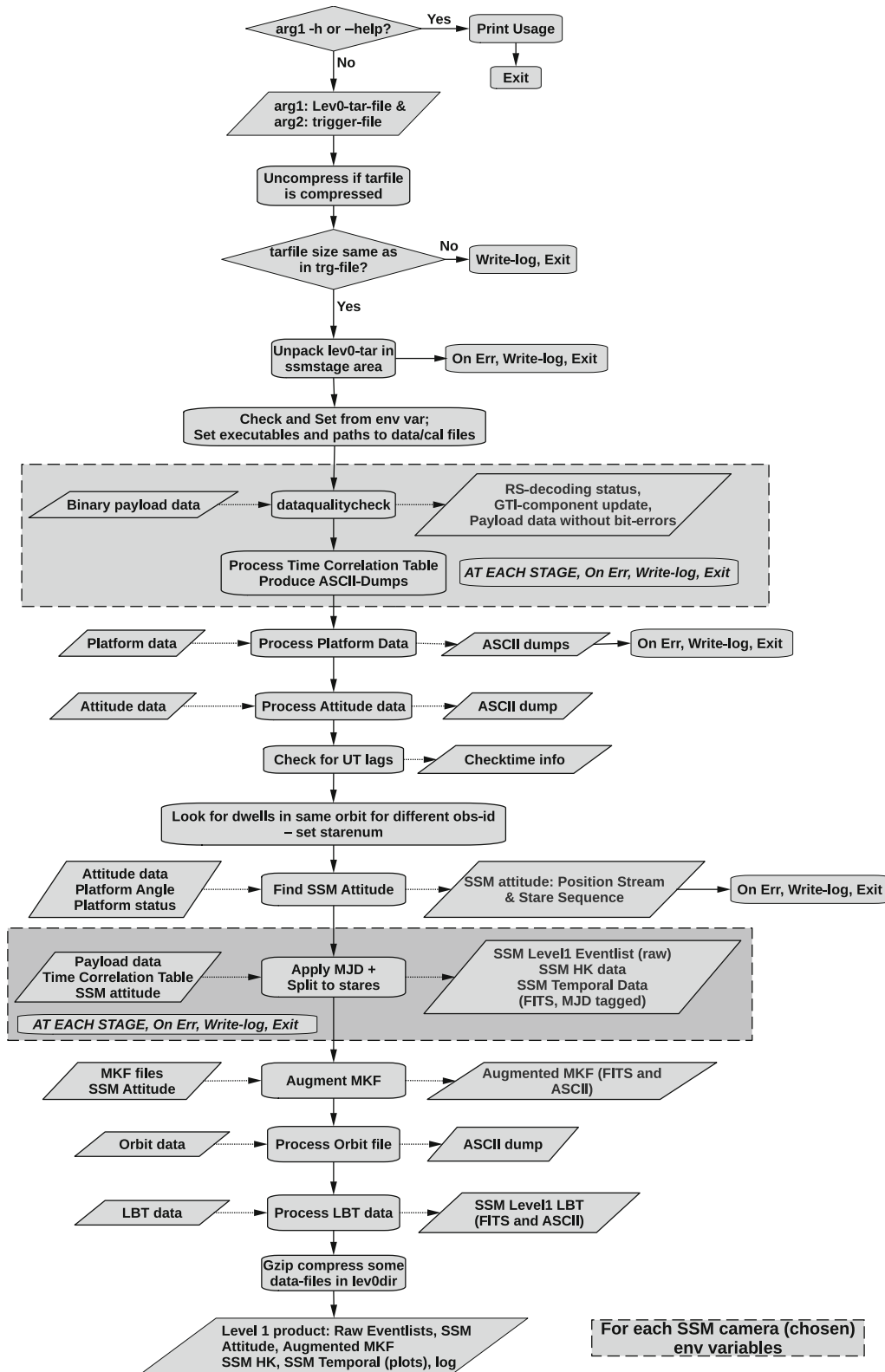
3. SSM Level-0 data

The SSM instrument dataset is written in segments/pages each of 2048 bytes size by the onboard Processing Electronics subsystem of the respective SSM camera. Each 2k page consists of instrument house keeping parameters, time-tags of the instrument clock latched with corresponding Onboard Computer (OBC) clock at a rate of 1.024 s, Temporal parameters (such as count-rates) and individual photon strike events recorded in the detector (with information about anode-ID and the voltages recorded at either end of the anode wire). The 2k pages are further embedded within an envelope of the Baseband Data Handling subsystem of the spacecraft which encodes the data with RS-encoding to check for transmission errors. This forms one of the main components of the Level-0 data. Along with this are packed auxiliary files including the following:

(a) the SSM platform data, with information about the stare-and-step operation including the platform angle and status whether platform is rotating or is stationary, etc, at intervals of 1.024 s.

(b) Attitude data, with the inertial coordinates in units of quaternions at intervals of 172 ms.

(c) Time Correlation Table, correlating the SSM clock to corresponding UT, produced using the latched OBC time.

(d) Make filter (MKF) file with samples at a rate ~ 100 ms of all health and orbital parameters (like South Atlantic Anamoly (SAA) region) that will be included to obtain the good time intervals (GTIs).

## 4. Level-0 to Level-1 data processing

The data flow diagram of the SSM Level-0 to Level-1 data processing is shown in Fig. 3. The Level-0 tar file along with the corresponding trigger-file form the
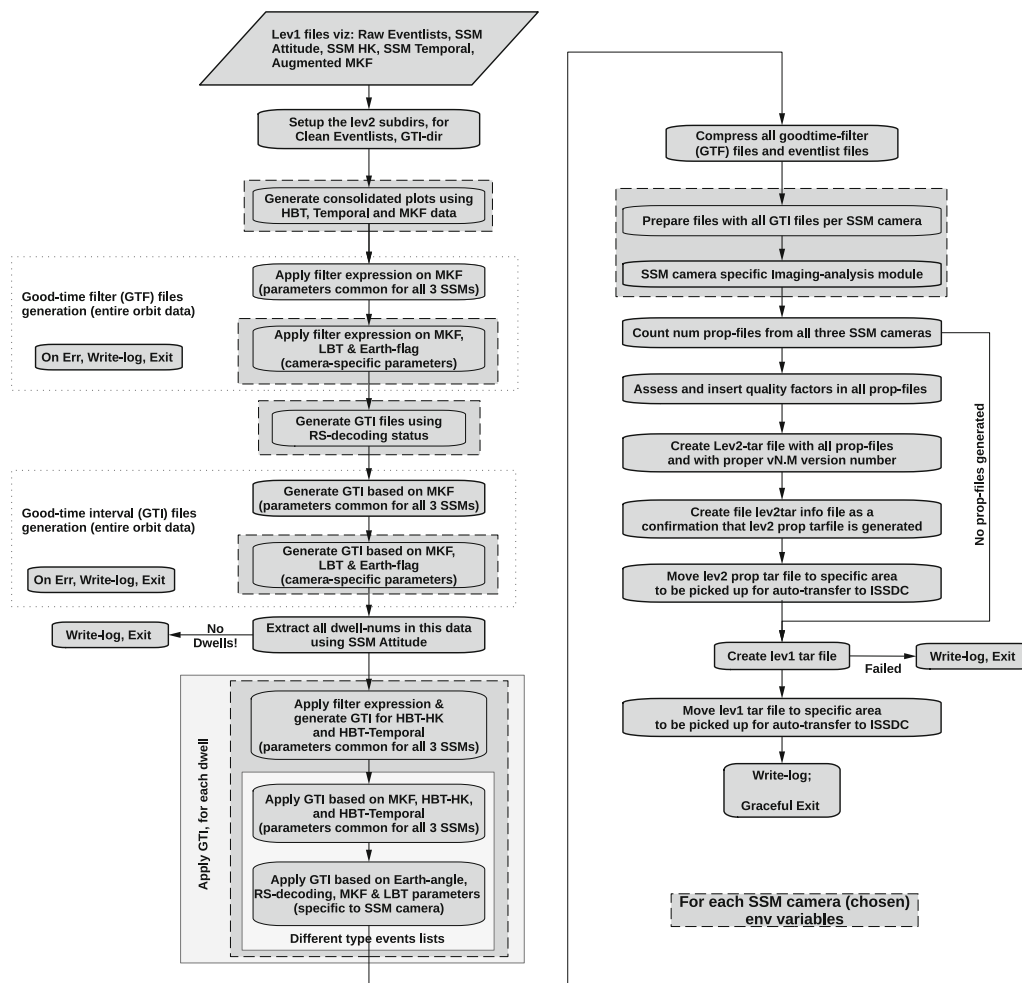
Figure 3. Data flow diagram: SSM Level-0 to Level-1 processing undertaken at the POC as part of the automated pipeline.

input. The trigger file consists of the size of tar file in bytes populated in the ISSDC data server and used at the POC for verification. After setting up the analysis environment, the very first task is to assess the quality of the data based on the RS-decoding parameters attached to every payload 2k page. All 2k
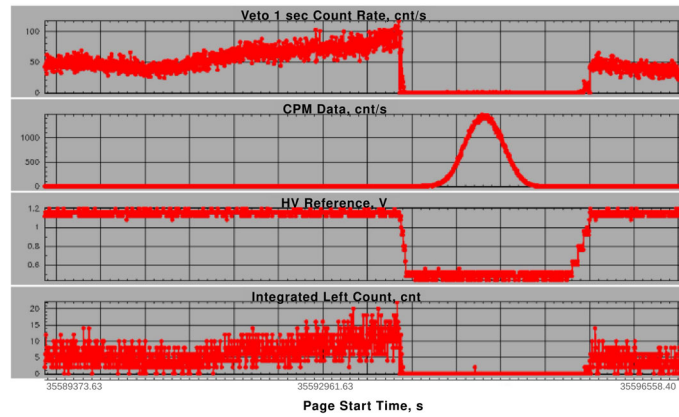
Figure 4. One sample plot of some of the housekeeping parameters produced using *pgplot* at SSM Level-1 stage; here for orbit 28198 and for SSM3 camera; The High Voltage (HV) reference is lowered during SAA passages (as also observed by the increase in Charged Particle Monitor (CPM) count-rate); Since HV is lowered during this duration SSM Veto and integrated left count rates are reduced to zero.

Figure 5. Data flow diagram: SSM Level-1 to Level-2 processing undertaken at the POC as part of the automated pipeline.
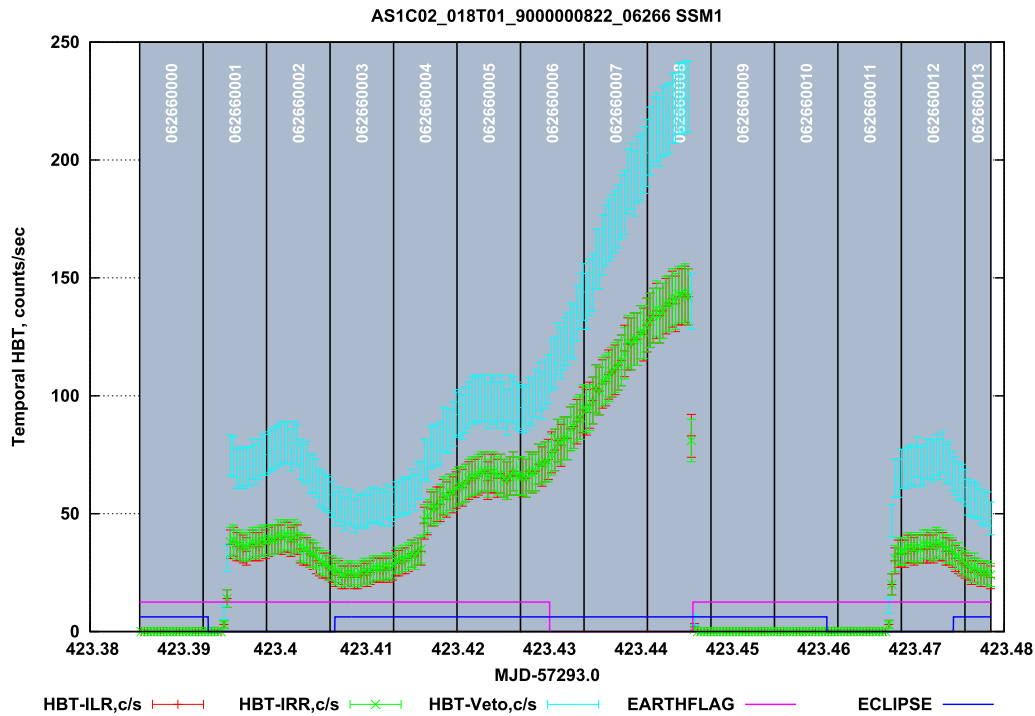
Figure 6. Consolidated plot with some HBT temporal parameters, here for orbit 6266 and for SSM1 camera for a scheduled calibration observation with the Crab source; grey strips indicate data binned at 10 min with stare-sequence number marked in each; ILR and IRR are respectively integrated left and right count rates from all anodes; also plotted is Veto count rate, and the dip in all the count rates corresponds to the SAA region; also indicated are spans of eclipse and Earth in the FoV.

pages with uncorrectable bit-errors are dropped and all these contribute to bad-time to be accounted for in the GTI intervals. The RS-decoding parmeters are also used later to prepare a quality report to be sent back to ISSDC. The Time Correlation Table (TCT) files, the SSM platorm data with information about the stare-and-step operation and the attitude data are processed, and ASCII dumps of the same are produced. The attitude information is available in the form of quaternions and they are converted to the Right Ascension (RA) and Declination (Dec) of each of the three spacecraft body axes of Yaw, Roll and Pitch. Possible time-offsets between the instrument data and the auxilliary data for a given dump are checked. The RA and Dec of the SSM pointing axes are then computed using the attitude data of the spacecraft pointing, SSM platform angle, the status of its rotation, and the ground-measured alignment angles of the flight module namely, the Cant-angle ($\sim 45°$), the inclination angles of the two edge cameras, SSM1 and SSM2 ($\sim 12°$), and the reference position-angle for the platform rotation mechanism.

The output SSM attitude data consist of Position stream and Stare sequence, with the inertial coordinates of the three SSM cameras. The Stare sequence is determined by picking all instances when the SSM platform is in stare-mode of typically 10 min. With the available Stare sequence, the instrument event data is UT-tagged and split into individual stares, and saved as different extensions of a FITS binary table (using *cfitsio* library, Pence 1999). The events are also segregated into two kinds – based on whether anode-IDs at either ends match, or if one of the ends is marked 8 if recorded only at one end. Also segregated into separate FITS files – Stare-wise – are the HK parameters and temporal parameters. Plots are generated for all these parameters using the *pgplot* library (Fig. 4). The MKF file made available with spacecraft position, attitude parameters is augmented with some SSM specific parameters. The orbit and Low Bitrate Telemetry (LBT) files providing spacecraft parameters are also processed. The Level-1 output produced consists of the raw event-lists for each of the SSM camera, SSM attitude, augmented MKF files and SSM HK and temporal parameters, and their plots.
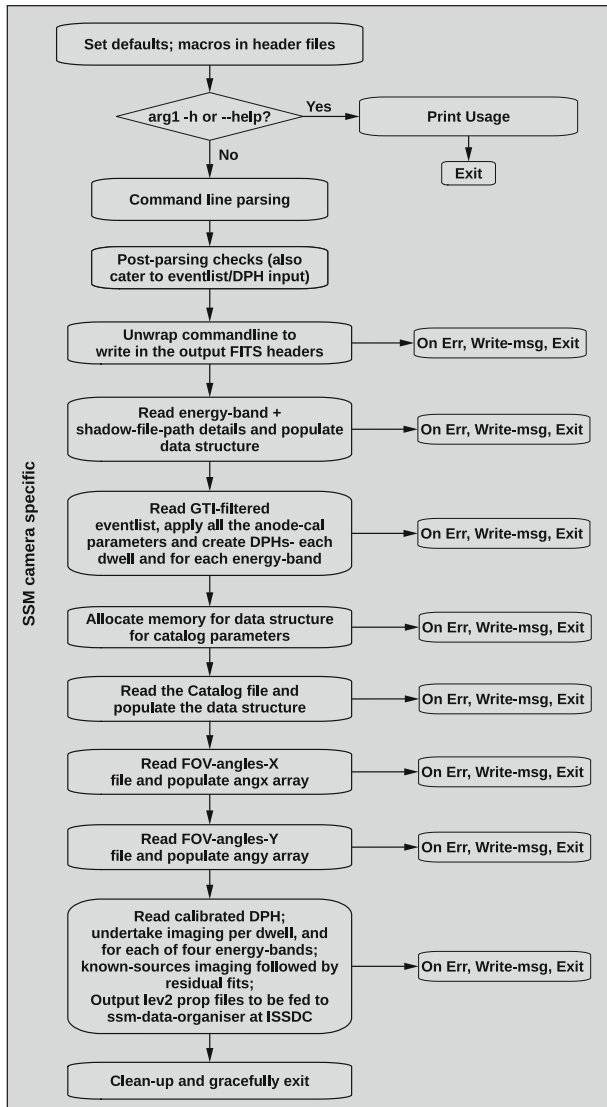
Figure 7. Data flow diagram: SSM Level-2 imaging details; corresponds to the 'SSM camera specific imaging-analysis module' block of the data flow diagram in Fig. 5.

## 5. Level-1 to Level-2 data processing

The full Level-1 to Level-2 data flow diagram of SSM is as shown in Fig. 5. Level-1 data products form the input, mainly the event-lists (especially of the type in which both anode-IDs per event match), augmented MKF file, SSM attitude, SSM HK and Temporal data. After setting up the necessary output directories, plots marking individual stares with different temporal count-rates are generated using *gnuplot* for quick checks of the data received (Fig. 6). The next major stage is the generation of good time filter files based on the filter expressions set for individual parameters, for entire orbit data – this, for parameters common for

all three SSM cameras (such as the SAA flag, CPM data), as well as those specific to individual cameras (such as Sun angle, RS decoding error count, electronic subsystem status). The good time filter files are used to consolidate and generate good time intervals. After extracting individual stare data, some more stare/dwell specific filtering is included to the GTIs (such as different voltage and temperature monitoring values). All these GTI files are applied on the raw event lists to produce clean Level-2 event lists. The consolidated GTI per camera per stare is collected to correct for the exposure time. The GTI-cleaned event lists of unambiguous anode-ID type form one of the main inputs to a camera specific imaging module. The modules leading to the imaging analysis are shown in the data flow diagram in Fig. 7.

The main tasks undertaken as part of the imaging (Fig. 7) are to read energy-band details and corresponding shadow response library, mean free path values which are then populated in different data structures. Anode calibration parameters are read and applied on the GTI-filtered event-lists, producing energy-band specific Detector Plane Histogram (DPH) for every stare of each camera. The SSM catalog and the resolution elements in terms of camera coordinates of θ_x and θ_y are read. For the computed SSM attitude, the list of known sources in the FoV is determined and populated in another data structure. With the band-specific DPHs and the list of known sources in the field for every stare, the imaging analysis is intiated. To describe here briefly, it is designed based on Bayesian Richardson-Lucy technique (see Ravishankar & Bhattacharya 2003) as well as *svdfit* based forward-fitting method. The energy-bands considered are: (a) 2.5–4 keV, (b) 4–6 keV, (c) 6–10 keV, and (d) 2.5–10 keV. Contributions of the known sources are fit using the response library, their response removed from the observed DPH, and in the residual, new sources are looked for. This exercise is undertaken separately for each of the four energy bands and for every camera, and, for every stare/dwell.

Figure 8 shows the lightcurve for the Crab source from early 2016 for observations with SSM3 camera; here it is also compared with the MAXI (Matsuoka *et al.* 2009) lightcurve.[1] In the inset are zoomed in plots for clarity. The data of all sources will be made public after efforts undertaken to address the dispersion in the data is applied for all observations in a data

---

[1]The non-standard MAXI product of 2.5–10 keV light curve is generated using the *MAXI on-demand Process* at http://maxi.riken.jp/mxondem/.
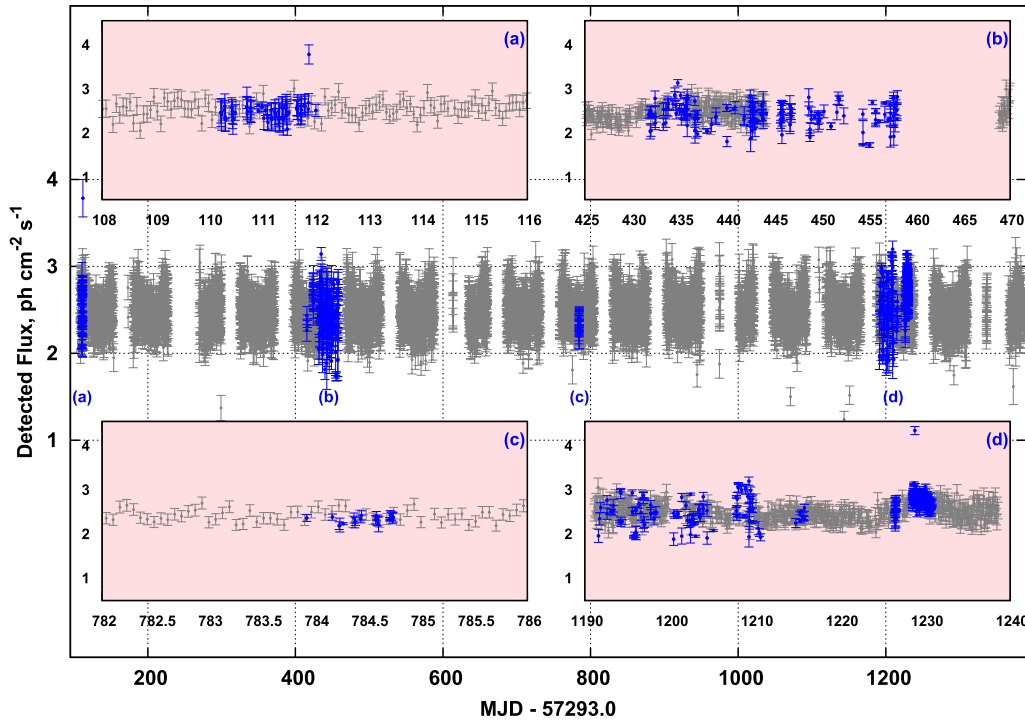
Figure 8. SSM3 Crab lightcurve from some set of observations across a few years (in blue). Plotted in grey is the MAXI lightcurve of Crab (the non-standard MAXI product of 2.5–10 keV light curve is generated using the *MAXI on-demand Process* at http://maxi.riken.jp/mxondem/) for comparison. The inset plots (a)–(d) are zoomed versions of the SSM observation slots as marked in the main panel.

regeneration campaign. An account of these efforts and details of the image processing will be provided in a future publication.

The output of the imaging analysis will be *prop-files* (Fig. 5) which are incremental Level-2 product consisting of updates to the light curves of every detected source in four different energy bands with associated parameters of goodness of fit, earth angle, background count rate, etc. The prop-files are updated with the quality factors determined. Though the imaging is undertaken stare-wise for each energy-band, the *prop-files* for each SSM camera are populated detected source-wise. That is, during imaging, observations per stare/dwell, per camera, per energy-band are considered and the flux values in each energy band, goodness of fit, background rate, and other parameters like earth angle, total number of sources in the stare, etc, are noted in a data structure for every detected source. In the *prop-files*, however, for every detected source, all the parameters determined are populated for each stare/dwell they are detected in, in another data structure. The *prop-files* are also marked separately in the data structure in three categories: (a) as incremental flux update to the light curves of known sources,

(b) alerts on outbursts in known sources, and, (c) alerts about newly detected sources. Level-2 tar packages are generated with these files. These, and the Level-1 tar-files are moved to a designated area from which the automated file-transfer daemon will pick the files and transfer them to the appropriate areas at ISSDC.

6. Turn around time

The turn around time for the SSM product generation for the data corresponding to one orbit duration is about 25 min. The time it takes for the Level-0 to Level-1 processing is ~ 5–8 min, and that for Level-1 to Level-2 processing is ~ 12–16 min. The duration varies based on how crowded the FoV is with strong sources, and the background rate. The data transfer time itself varies based on the network bandwidth and typically takes less than a few minutes, considering the additional checks introduced as explained in Section 2.1 The POC system employed which provides this processing time involves a workstation with six core 3 GHz Intel Xeon processor and 6 GB RAM, and RHEL operating system.
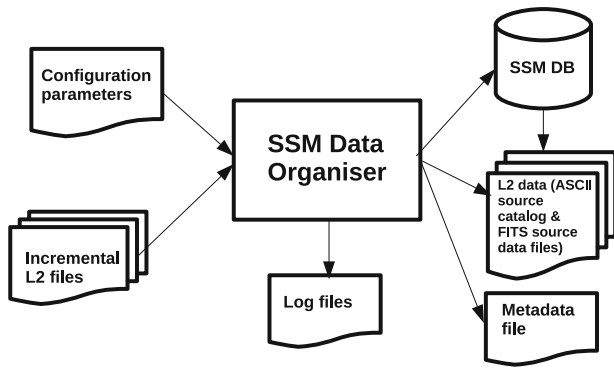
Figure 9. SSM Data Organiser module of the automated pipeline at the ISSDC which ingests the incremental updates to the lightcurves of the sources detected into a database and manages the disseminated data products.
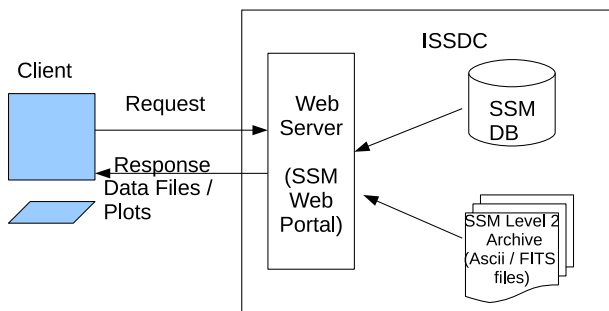
Figure 10. SSM Data Dissemination web portal at ISSDC provides an user interface to the SSM's Level-2 data products.

7. Level-2 data reorganisation and dissemination

SSM Data Organiser (SSMDO) is a command line based Java application to create SSM Level-2 data products (see Fig. 9). This application is a part of SSM Data Pipeline (DP)/SSM Level-1 to Level-2 data processing chain and is set up to run at ISSDC in an automated fashion.

The Level-2 incremental data files need to be organised in such a way that for every source observed by SSM upto a given point of time, there is one file (FITS format) containing up-to-date observation details.

SSMDO makes use of a MariaDB database to store the observation details of sources observed by SSM. One table for every source observed by SSM is maintained in the database, containing all the observation parameters pertaining to that source. Tables to store the source catalog details and alert information are also available.

Once the incremental Level-2 data files are available at the designated location at ISSDC, execution of

SSMDO application is triggered. A configuration file is used to specify parameters at runtime. Each incremetal Level-2 data file is processed and the required information/observation parameters are extracted and stored in the appropriate database tables corresponding to the source observed. Using data from the tables, the Level-2 data files (FITS) for each source is updated. The source catalog file (ASCII) is also updated. The Level-2 data products – source catalog and source data files – along with metadata file are then provided to the archival system for long term archival.

The output messages and error messages at every stage are logged for debugging purposes and also to provide a report on the execution of the module.

Provision is made for users to view the observed source catalog of SSM and also browse through the light curves of all sources observed by SSM through a website as shown in Fig. 10. In addition, provision is made for downloading plots of light curves and hardness ratios in PNG or JPG formats, and to download the data in ASCII, FITS, and VOTable formats. All these are process validated.

8. Conclusion

This paper gives an account of the data pipeline implemented for the Scanning Sky Monitor on AstroSat. The data pipeline as a fully automated software, has been tested and validated.

References

Agrawal P. C. 2006, Adv. Space Res., 38, 2989

Libes D., O'Reilly T. 1995, Exploring Expect: A Tcl-based Toolkit for Automating Interactive Programs (O'Reilly Media Inc.)

Love R. 2005, Linux J., 2005, 8

Matsuoka M., Kawasaki K., Ueno S. *et al.* 2009, Publ. Astronom. Soc. Japan, 61, 999

Pence W. 1999, in Mehringer D. M., Plante R. L., Roberts D. A., eds, Astronomical Society of the Pacific Conference Series, Vol. 172, Astronomical Data Analysis Software and Systems VIII, 487

Ramadevi M. C., Seetha S., Bhattacharya D. *et al.* 2017, Exp. Astron., 44, 11

Ravishankar B. T., Bhattacharya D. 2003, BASI, 31, 491

Singh K. P., Tandon S. N., Agrawal P. C. *et al.* 2014, in Takahashi T., den Herder J.-W.A., Bautz M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9144, Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray, 91441S