

gPhoton: A Time-Tagged Database of GALEX Photon Events

Chase Million¹, Scott W. Fleming^{2,3}, Bernie Shiao², Randy Thompson^{2,3}, Richard L. White²,
Karen Levay^{2,3}

`chase.million@gmail.com`

ABSTRACT

We describe the gPhoton project, an effort to calibrate, archive, and make available the complete corpus of approximately 1.8 trillion photon events individually recorded by the GALEX spacecraft. The project includes a standalone calibration pipeline which reproduces core functionalities of the GALEX mission’s original calibration pipeline, and a small suite of tools to simplify use of the photon level data with particular emphasis on short time domain analyses. The software is freely available as an open-source application¹ and will continue to be updated in the future. We describe the set of software tools that are available, highlight differences between the gPhoton standalone calibration compared to the original mission’s pipeline, and demonstrate the performance of gPhoton’s calibration by analyzing standard star fluxes and measuring the pipeline’s astrometric precision. Finally, we provide a few examples of science enabled by our pipeline in the form of M flare stars, known variables, and cross-mission analysis using Kepler targets that happened to be observed simultaneously by GALEX. The database will be hosted at the MAST archives, and will further extend the legacy of GALEX in the absence of dedicated, large-scale, space-based, UV-survey missions in the near future.

1. Introduction

The Galaxy Evolution Explorer (GALEX; Martin et al. 2005) was a NASA Small Explorer telescope that spent approximately ten years performing an all sky survey with a 1.25 degree field-of-view in two ultraviolet (UV) bands, centered around 1528 Å (FUV) and 2271 Å (NUV). Each band used its own microchannel detector, while a CaF₂ grism provided slitless spectroscopy across the entire GALEX field-of-view. The spacecraft was launched on 28 April 2003 and operated until 28 June 2013, although the FUV detector failed in May 2009. On 4 May 2010, an event occurred

¹Million Concepts, 2204 Mountain View Ave., State College, PA 16801, USA

²Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD, 21218, USA

³Computer Sciences Corporation, 3700 San Martin Dr, Baltimore, MD, 21218, USA

¹<https://github.com/cmillion/gPhoton>

that is referred to as “the Course Sun Point” (CSP), which generated substantial streaking in the NUV detector’s Y-direction. Although the effect is largely correctable through subsequent calibration and onboard adjustments, observations taken between 4 May and 23 June 2010 all have reduced point spread functions. NASA support for the mission ended in February 2011. The spacecraft was then transferred to the California Institute of Technology for the Complete the All-sky UV Survey Extension (CAUSE) phase of the mission, during which operating costs were solicited from individuals or institutions. The GALEX mission collected data over **some number of eclipses**, observing **some fraction of the sky**, and accumulating **some amount of** approximate total raw data volume and **some other amount of** approximate reduced data volume.

The spacecraft, detectors, and calibration are well described in Morrissey et al. (2005, 2007). We will attempt to restrict discussion to elements that are not discussed in those papers, or that have changed in the intervening years, or that apply only to gPhoton. In Section 2 we describe the motivation behind constructing the gPhoton database and software suite. In Section 3 we describe the design and content of the ~ 1.8 trillion row database hosted at the Mikulski Archive for Space Telescopes (MAST). In Section 4 we describe the main programs available to use with the database for constructing lightcurves, images, and animated data cubes. In Section 5 we highlight some of the implementation challenges we came across, and our solutions, some of which may be applicable to other photon event databases. In Section 6 we conduct some initial tests of the calibration precision by studying the astrometric, relative, and absolute fluxes produced by the software. Finally, in Section 7, we highlight a few example science cases enabled by gPhoton, including the study of stellar flares, high cadence lightcurves of known variables, and cross-mission science that can be conducted by combining gPhoton with data from other missions.

2. Motivation

Microchannel plates are non-integrating photon detectors. That is, rather than accumulating detector events in integrated bins, the microchannel plates record position and time information for every event individually. Due to computer storage and processing constraints, the GALEX mission team only released integrated maps of these data on per-observation time scales of hundreds to thousands of seconds. While the GALEX detectors were capable of recording events with a time resolution of five thousandths of a second, and a spatial resolution of **something** on the detector, the high time and spatial resolution photon level data was not released to the community except in isolated cases by special request, and no serious attempts were made to understand the detector performance or calibration parameters for integrations shorter than about 100 seconds. Furthermore, by the end of the mission, the GALEX calibration pipeline software had grown to sufficient complexity that no attempts to get it to run outside of the GALEX network of computers at Caltech were successful. With advances in storage capacity and processing capability, we have undertaken this project to build a standalone GALEX calibration pipeline, using it to generate and archive aspect-corrected positions for the ~ 1.8 trillion GALEX photon events accumulated over

the course of the whole mission, and a suite of software tools to make those photons usable at arbitrary spatial and temporal scales.

Our project is an effort to reproduce the core functionality of the GALEX mission calibration pipeline in an open source and simple to use package. The intention of the gPhoton project was never to fully reproduce the calibration capabilities of the GALEX mission pipeline, nor to anticipate and accommodate all possible science uses of the data. Rather, the intention was the create a minimal set of tools that will serve as a jumping off point for researchers to perform their own scientific analyses on the reduced data, which might include refining the GALEX calibrations entirely.

3. Description of the Database

This section will include Bernie’s description of the database table, architecture, and current status of data that’s loaded.

4. Description of the Software Tools

We have created a software suite of command-line utilities, written in python, to interact with the photon event database and enable analysis of the data contained within. There are four main functions included in gPhoton:

gPhoton Used to take data products created by the original mission pipeline and output tables of photon event positions, times, and energy levels. This is also used to populate the master photon event database hosted at MAST.

gFind Used to determine the available data coverage for a given coordinate.

gAperture Used to create lightcurves (text tables of time and fluxes, with various calibrations applied).

gMap Used to create calibrated intensity maps (images) or animated data cubes (movies).

We describe these main functions in more detail below.

4.1. gPhoton

The gPhoton standalone calibration pipeline implements only a minimum set of features from the full mission pipeline. The command line function, *gPhoton.py*, accepts the raw scientific data file (-raw6), the spacecraft state file (-scst), and a refined aspect file (-asprta), and returns a table

in comma-separated-value (csv) format, in which each row corresponds to a detector event and contains both raw and aspect-corrected detector positions, as well as sky positions and a number of intermediate detector parameters. The -raw6, -scst, and -asprta files were all products of the mission calibration products that are archived at MAST **Is it true that all of these are already archived?**. The -raw6 and -scst files simply contain parsed subsets of data that was originally in the raw spacecraft telemetry files (-tlms). While the -scst files contain a great many parameters related to spacecraft status, including the coarse estimated boresight pointing from the star tracker, only a single temperature is used for the purpose of applying a temperature dependent plate scale correction. The -asprta files contain refined boresight pointing estimates on a per-second basis and were created by the mission pipeline by iteratively comparing 15-second depth intensity maps generated from the best-available aspect solution against star catalogs. By using these mission products directly, we did not have to recreate the “ingest” or “aspect correction” stages of the mission calibration pipeline. The gPhoton standalone calibration pipeline also does not implement any artifact flagging or masking, except for hotspots. Events which fall inside of the masked regions of the detector hotspot masks are given a flag value (in the “flag” column of the .csv output) indicating as such. Hotspots were known to come and go over the course of the mission, and the hotspot masks were defined to be static and conservative, so some data flagged as “hotspot” may be perfectly valid scientific data; in general, though, the masked regions represent such a small fraction of the available detector that they should be left alone. For short time domain analysis, it could be an issue if your source traverses a hotspot, as this will manifest as a marked dip in the observed count rate.

4.2. gFind

This utility allows the user to inspect the available data coverage for a particular target. Given a source position, it will return the estimated (raw) depth of available data over the whole mission, as well as the contiguous time ranges (roughly, “observations”) in which these data are contained. Rather than using the mission logs or mean visit centers to determine the available coverage of a specific location on the sky, *gFind* attempts to use the data itself. A position on the sky is considered to be covered during time ranges in which it falls within half a detector diameter (an adjustable parameter with a default of 1.25 degrees) of the boresight center, as defined by the refined aspect solution. The contiguity of data for the purposes of grouping them into discrete time ranges is determined by a (user adjustable) parameter that defines the maximum gap between data for it to be considered contiguous. The default value is one second, which is also the time resolution of the aspect file. There is, similarly, a (user adjustable) parameter that defines the minimum exposure depth considered to be a valid exposure (the default is one second).

4.3. gAperture

This utility generates aperture photometry for requested target positions and time ranges. Simple aperture photometry is used. The user provides a target position, aperture radius, optional inner and outer radii for a background annulus, and optional time ranges. The fact that we have the aspect corrected photon-level data allows us to do some innovative things in terms of computing the photometry. Rather than binning the photons into pixels, losing spatial precision in the process, and then performing complex interpolations over the pixels in order to estimate the content of a circular aperture, the counts within the aperture can be computed directly by performing a cone search on the data. The background counts are similarly computed. In order to increase the reliability of the background correction, weve implemented a “swiss cheese” algorithm that masks out known sources within the annulus. The count rate within the effective area of the annulus following the “swiss cheese” masking is then scaled to the area of the aperture. Rather than integrating a relative response map, the photon events can be individually weighted by using their detector positions to directly sample the flat. Effective exposure times are computed directly from the global count rates (to estimate both shutter and deadtime corrections). The photometry can be computed within any time bin, up to and including an integration over the entire mission. The output format is a csv file containing a large number of columns including raw counts, background rates, estimated errors, mean response over the aperture, and mean time of arrival for photon events within the time bin.

4.4. gMap

This utility generates integrated image maps for targeted regions and time ranges. The format of all output is in the Flexible Image Transport System (FITS). Images are generated by performing a box search for events in the requested region and then binning the events into a histogram that has been properly proportioned based upon the World Coordinate System (WCS) in that region. Note that because the gnomonic projection used by GALEX increasingly distorts images as you approach the poles, you can expect images generated by *gMap* to, in general, not have the same aspect ratio in pixels as they have in degrees. **Need to discuss animated data cubes here as well.**

5. Implementation Challenges and Solutions

During the process of developing gPhoton we have anticipated numerous implementation challenges, some of which may be applicable to other photon event-based databases, or the details of which may be of general interest to users of our software suite. We detail some of those here.

5.1. Photon Level Analysis

5.2. Effective Exposure Time Calculation

The exposure time correction and, in particular, the global dead time correction, was not reliable for short time ranges. This is because the dead time correction was computed in the original mission pipeline by comparing the measured stim rate against the nominal stim rate of 79 counts per second. At short time ranges (on the order of seconds), the variance in this rate due simply to Poissonian counting statistics made the error in the exposure time calculation sufficiently high as to make flux calculations meaningless. The solution was to use the empirical dead time correction, a linear fitted relationship of dead time as a function of the global detector count rate (which can be thousands of counts per second). Our tests have shown that this is a reliable estimator of the dead time correction and leads to stable/reliable effective exposure times at short time ranges.

5.3. Relative Response Correction

The relative response (i.e. flat field) was not characterized at spatial resolutions that make it meaningful for short time domains. That is, because the detector is constantly in motion with respect to the sky during any observation, any given source samples a region of the detector. The resolution chosen for the flat field (approximately equal to the width of a single “dither”) was chosen with the assumption that an integrated observation would smear out any sub-resolution detector response effects, and the uncertainties would average out. This assumption does not hold for short exposure times. We have no solution for this, and it continues to be a problem. but we know of efforts by other investigators to characterize the flat at higher spatial resolutions. This is one area that may be improved with future versions of our software.

5.4. Flux Uncertainties

We are going to need to explain the current way the pipeline estimates uncertainty, some of the challenges with deriving an accurate uncertainty estimate, and perhaps some ideas for improvements in the future. Note: Include a plot of LDS 749B vs. effective exposure time, like Chase has made, and comment.

6. Calibration Tests

6.1. Astrometric Solutions

I’m not sure what the best tests here are, other than to trace the photocenters of objects across time?

6.2. Relative Flux Precision

Here is where we will take a bunch of plates and study the variability as a function of distance from plate center, working under the assumption that most non-extended sources should be constant within their uncertainties.

6.3. Absolute Flux Precision

We have made use of the same white dwarf standard stars that were used by the GALEX mission pipeline to help calibrate and test our absolute flux precision. Specifically, we use the white dwarf stars LDS 749B What are the other ones?.

Note: Include as many white dwarfs from Camarota & Holberg 2014 as possible to compare their final fluxes and ours.

7. Example Science Applications

7.1. Flare Stars

7.2. Cross-Mission Overlap

8. Conclusion

The GALEX data set has already proven to be extremely productive and is likely to be one of the most influential ultraviolet astronomical surveys for the foreseeable future. The gPhoton project simplifies, and in some cases enables, analyses using this data that were previously difficult or impossible, especially those related to short time domain photometry. While the gPhoton project is an effort to calibrate and make available the GALEX photon level data specifically, some of the techniques described here can be applicable to other observational databases that make use of non-integrating detectors, particularly microchannel plates. The fact that spatial analyses can be performed by making direct cuts on the photon level data rather than artificially degrading the spatial components of the data by integrating and interpolating onto image maps offers potential advantages in terms of both the flexibility of the data archive and the computational cost of analysis. The corresponding data management and volume issues associated with storing and retrieving massive amounts of photon level data is non-trivial, but also entirely soluble with appropriate use of existing database and storage technology. The behavior of the GALEX detector during

very short timespans (which correspond to very small spatial sampling of the detector) is not well characterized, and further work on improving the resolution of the detector response, as well as correctly propagating flux uncertainties, will be required to derive the maximum utility from the photon level data.

Place acknowledgements here.

REFERENCES

- Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, *ApJ*, 619, L1
- Morrissey, P., Schiminovich, D., Barlow, T. A., et al. 2005, *ApJ*, 619, L7
- Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, *ApJS*, 173, 682