

gPhoton: A Time-Tagged Database of GALEX Photon Events

Chase Million¹, Scott W. Fleming^{2,3}, Bernie Shiao², Mark Seibert⁴, Randy Thompson^{2,3}, Myron Smith⁵, Richard L. White², Karen Levay^{2,3}

chase.million@gmail.com

ABSTRACT

We describe the gPhoton project, an effort to archive the complete corpus of approximately 1.1 trillion photon events recorded by the GALEX spacecraft. The project includes a standalone calibration pipeline that reproduces core functionalities of the GALEX mission’s original calibration pipeline, but also generates lists of aspect corrected photon level data. A suite of software tools simplifies 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. We highlight some of the science enabled by gPhoton, including stellar flares, high cadence sampling of known variables, and cross-mission analyses (exemplified through simultaneous GALEX observations of Kepler targets). The database is housed at the MAST archives, and extends the utility and impact of GALEX data 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 surveying the sky in the ultraviolet. GALEX had a

¹Million Concepts LLC, 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

⁴The Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA

⁵NOAO...need full address from Myron.

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

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, and had both a direct imaging mode and a slitless spectroscopic mode via a CaF₂ grism. 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 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 have reduced point spread functions. NASA support for the mission ended in February 2011. Spacecraft ownership 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, have changed in the intervening years, or apply only to the gPhoton project. 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 ~ 1.1 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 along with our solutions, some of which may be applicable to other photon event databases. In Section 6 we present 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 sampling 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 timescales 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 approximately ~ 1.1 trillion GALEX photon events accumulated over the course of the whole mission, and a suite of software tools to use those photons to construct lightcurves and images at user-specified spatial and temporal scales. While gPhoton does reproduce the core functionality of the GALEX mission calibration pipeline in an open source and simple to use package, it is not intended to fully reproduce the mission pipeline. Rather, the intention is to create of a set of tools that serve as a starting point for researchers to perform their own scientific analyses on the reduced data, which might include refining the GALEX calibrations entirely.

The spacecraft observed during stretches of time when the Sun was eclipsed by the Earth, known as “eclipses”, which typically lasted 1600-2000 seconds. To avoid detector burn-in or local gain sag effects caused by depletion of electrons in the multiplier plate, the GALEX spacecraft did not statically stare at locations on the sky, but continued to move over the course of each observation. Several observing patterns or modes were used over the course of the mission with potential consequences for the nature of the underlying data. The most basic mode was referred to as the dither pattern in which the spacecraft boresight would trace out a tight spiral (or dither) with a radius of a few arcminutes. The dither pattern was used most often for deep or medium imaging surveys in which each full eclipse of approximately 1600 seconds was spent observing a single region of the sky. In the All-sky Imaging Survey (AIS) observing mode, the spacecraft would jump between multiple positions (typically 12 legs) on the sky for short integrations of 100 seconds each with the detector set to a non-observing state during the transition between each position, producing one independent GALEX observation (“visit”) per position. The petal pattern mode, used to observe brighter targets, was similar to the AIS mode except that the positions (typically 8 legs) were tightly clustered within the approximate area of a single FOV and the detector continued to observe in the transition periods, spreading the light of any single source more widely across the detector. Finally, a scan mode was implemented late in the mission—after detector count rate safety limits were relaxed, and used frequently during the CAUSE phase—in which the spacecraft boresight would rapidly traverse many degrees of sky at a time. Scan mode data presents unique challenges for analysis over the other modes, including uncertain calibration and frequent failure of aspect refinement; while gPhoton can use scan mode data, results in these cases should be used with additional care and caution.

3. Description of the Database

The construction of a trillion-row SQL Server database requires advanced planning in order to make queries more efficient. The first step in generating the gPhoton database was to create the aspect-corrected photon events as comma-separated values (CSV) files. Starting with the

raw photon event (-raw6), refined spacecraft attitude (-asprta), and spacecraft state (-scst) files available at MAST, the celestial positions of the photon events are calculated and then exported to the CSV file. At this stage, the CSV file contains the time of the photon event, event positions on the detector, the aspect-corrected positions on the sky of those events (RA,DEC), and status flags used to track a variety of conditions related to the detector readout. Only a subset of the photon events from a given eclipse are able to have aspect-corrected positions calculated (most often because only a subset of photon events in an eclipse correspond to an observation, others happen during dead time, slewing, etc. **I don't think this explanation is quite correct, Chase please revise.**). Therefore two CSV files for each band are created: a CSV file for FUV, NUV data that contain aspect-corrected photon events, and a CSV file for FUV, NUV data containing those photon events that were not aspect-corrected. The non-aspect-corrected photon events are used for estimating deadtime corrections, and so are uploaded into the database, but are otherwise not used beyond this point.

We divide the sky into ten different databases, based on declination. These ten databases are then further sub-divided into a total of 999 partitions. Each of these 999 partitions are composed of a number of “zones”, which are stripes of declination 30'' in height. Photon events from the aspect-corrected CSV file for an eclipse are then assigned into one of these ten databases, taking into account that some eclipses may span more than one database/partition/zone. In order to quickly assign a given photon event to the correct database, we make use of the zone match algorithm of Gray et al. (2006), both for loading and querying the database. The distribution of the ten databases are shown in Fig. 1, along with the DEC range of each. The gPhoton software operates seamlessly across these database boundaries when queries span multiple databases, but most queries only need to operate within a single database, which helps improve performance.

Both the database boundaries and the number of 30'' zones assigned to each partition were defined so that the databases are of similar sizes. This was accomplished by assuming the total number of photon events in a given eclipse are distributed evenly across that eclipse's footprint. Then, the cross section of that eclipse's footprint against the zones is calculated to determine which zones that eclipse overlaps. The number of photons in each zone from this eclipse is estimated based on the cross-sectional area, e.g., if a given eclipse spans two zones, but only 10% of the eclipse's footprint is in one of the zones, 90% of its total photon events would be considered to belong to the first zone, and 10% to the other. This method of estimating photon events in each zone is not precise, since it assumes photon events are evenly distributed across the eclipse, but it does serve as a quick approximation to define the database boundaries and partition assignments, since it is not necessary to actually compute the zone membership for all the 1.1 trillion photon events beforehand.

4. Description of the Software Tools

There are four utilities functions included in gPhoton and described in Table 1. These utilities are all written in Python and released under a permissive license. With the exception of gPhoton, the tools can be called either from the command line or from within the Python interpreter prompt. If called from within the interpreter, output is returned in a Python data structure. These command line utilities draw upon a large a large number of supporting libraries which will not be described but are possibly of interest to researchers who wish to work more deeply with the GALEX data.

Utility	Function
gPhoton	Generates aspect corrected photon list files from a small set of user supplied input files which are nominally archived products of the mission pipeline. Output from gPhoton is used to populate the master photon event database hosted at MAST.
gFind	Provides information on the available data coverage for a given coordinate.
gAperture	Creates lightcurves for a requested target (tables of times, calibrated fluxes, and additional parameters, with various calibrations applied).
gMap	Creates calibrated count, response, or intensity maps (images) or data cubes (movies).

Table 1.

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 are products of the mission calibration pipeline that are archived at MAST **Is it true that all of these are already archived?**. The raw6 and scst files are products of the ingest stage of the mission pipeline and simply contain parsed subsets of data that was originally in the raw spacecraft telemetry files (tlms). While the scst files contain many parameters related to spacecraft status, including the coarse estimated boresight pointing from the star tracker, only a single temperature value from the scst is used for the purpose of applying a temperature dependent plate scale correction. The asprta files, products of the aspect correction stage of the mission pipeline, contain refined boresight pointing estimates on a per-second basis, created 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 do not need to recreate the ingest or aspect correction stages of the mission calibration

pipeline.

In most cases, a flag column entries which is not equal to zero indicates that an event could not be aspect corrected for some reason. One exception is events which are flagged because they fall within a gap in the refined aspect solution during which the detector continue to observe (for example, between legs of a petal pattern observation); in such cases, the events are aspect corrected but flagged as having unreliable aspect solutions. Another exception is events which fall within regions of the detector known to be sources of hotspots which can be aspect corrected but should not be naively trusted. Events which are not aspect correctable for some reason (for example, those arising from the detector stims) are referred to as null data because they contain null values in the right ascension and declination columns of the photon list file. Null data can be useful for detector calibration and, in particular, exposure time correction based upon the stim or global count rates. Other than hotspots, the gPhoton project does not implement any artifact flagging or masking.

4.2. gFind

This utility provides a mechanism for the user to inspect the available data coverage for a particular target. Given a target position, it will return the estimated (raw) depth of available data over the whole mission, as well as the time ranges for specific observations in which these data are contained. Rather than using the visit-based bookkeeping of the mission logs 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 (*asprta*). The contiguity of data for the purposes of grouping them into discrete time ranges is defined by a user adjustable parameter that defines the maximum gap allowed for data to be considered part of the same observation. There is a similar user-adjustable parameter that defines the minimum exposure depth considered to be a valid exposure.

4.3. gAperture

This utility performs photometric analysis given a target position, aperture radius, optional inner and outer radii of a background annulus, and optional time ranges. Simple aperture photometry is used, but the fact that the aspect corrected photon-level data is available allows us to perform a cone search on the data to precisely count the number of photons within the aperture rather than binning the photons into pixels—sacrificing spatial precision in the process—and then performing interpolations as would be done for normal integrated image data. The background counts are similarly computed. Rather than integrating a relative response map, the photon events can be individually weighted by using their detector positions to sample the flat directly. Effective exposure times (raw exposure time modified by effects of data dropouts and detector deadtime) are

estimated from the count rates over the entire detector. The photometry can be computed within any time bin, up to and including an integration over the entire mission. The output format is a comma separated value format file containing a large number of columns which include raw counts, background rates, estimated errors, mean response over the aperture, and mean time of arrival for photon events within the time bin. When gPhoton is called from within the Python interpreter, the returned data structure contains a large number of additional parameters including the complete photon list data for the targeted region of the sky, permitting detailed analysis.

4.4. gMap

This utility generates integrated count, response, and intensity image maps for targeted regions and time ranges. All image data is written in the Flexible Image Transport System (FITS) with headers populated using World Coordinate System standards. Count (cnt) images are integrated images which have not been scaled by the detector response, generated by performing a box search for events in the requested region and then binning the events into pixels with the same spatial dimensions as products generated by the mission. High resolution relative response (rrhr) maps are created by adding successive projections of the detector flat onto the sky, using the boresight position, at one second intervals, scaled by effective exposure time. Intensity (int) maps are the "calibrated" image products from which photometric analyses may be performed, are simply count maps divided by relative response maps.

The images produced by gMap are analogous to the imaging data products produced by the mission pipeline and archived at MAST with the exceptions that users are not limited to images with specific dimensions and coadds can be created arbitrarily across visits and surveys. Also, each of these types of images can also be generated as data cubes or movie files with user specified time bins whereas full resolution and calibrated movie files were never available from the mission. Note that because the gnomonic projection used by GALEX increasingly distorts images near 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.

5. Implementation Challenges and Solutions

Numerous challenges have arisen during the development of this project, and our solutions to some of these may be helpful to other microchannel plate or photon level data investigations. The details may be of general interest to users of our software suite as well.

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. Background Correction

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.

5.4. 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.

5.5. 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

- Gray, J, Nieto-Santisteban, M. A., & Szalay, A. S. 2006, *Microsoft Technical Report*, MSR-TR-2006-52
- 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

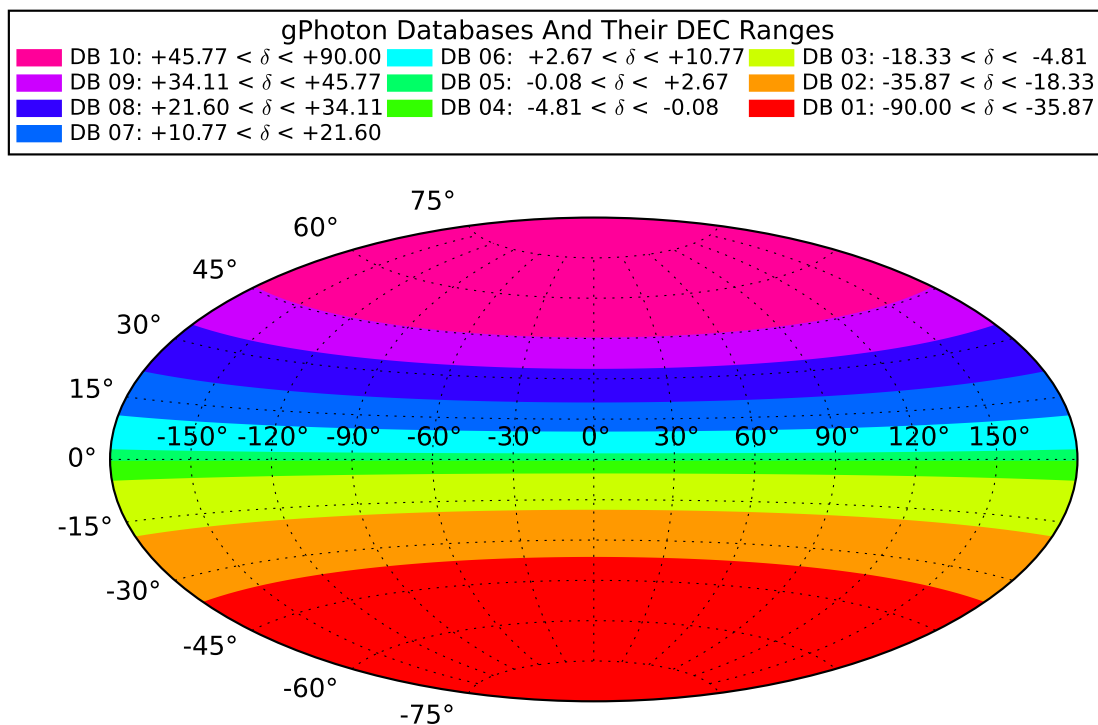


Fig. 1.— The DEC boundaries of the ten databases that comprise gPhoton. There are a total of 999 partitions across the ten databases, each with a variable number of $30''$ zones (stripes of DEC). The number of zones in each partition, and the number of partitions in each database, were assigned so that the size of the ten databases would be roughly equal to each other.