gPhoton: A Time-Tagged Database of GALEX Photon Events

Chase Million[a,*], Scott W. Fleming[b,c], Bernie Shiao[b], Mark Seibert[d], Parke Loyd[e], Michael Tucker[f], Myron Smith[b,1], Randy Thompson[b,c], Richard L. White[b], Karen Levay[b,c]

[a]*Million Concepts LLC, 2204 Mountain View Ave., State College, PA 16801, USA*
[b]*Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA*
[c]*Computer Sciences Corporation, 3700 San Martin Dr, Baltimore, MD 21218, USA*
[d]*The Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA*
[e]*Department of Astrophysics and Planetary Science, University of Colorado, Boulder CO*
[f]*Dept. of Physics and Astronomy, Appalachian State University, Boone, NC 28608, USA*

Abstract

We describe gPhoton[2], a new data product and software package from the Mikulski Archive for Space Telescopes (MAST), which enables analysis of GALEX ultraviolet data at the photon event level. The database is composed of approximately 1.1 trillion sky-projected photon events stored in a partitioned, 130 TB SQLServer backend. The accompanying python software allows users to generate calibrated light curves and images at user-defined temporal and spatial scales. The project's stand-alone, pure-Python calibration pipeline reproduces core functionality of the original mission pipeline, while command line tools enable reduction and use of the photon data for common classes of analyses, including the creation of calibrated light curves or images (as a singular coadd or temporally sliced into a data cube). The gPhoton software and source code are publicly available under a permissive license and will continue to be developed and maintained in the future. We describe the motivation, design, and implementation of the calibration pipeline, database, and tools, with emphasize on divergence from the mission pipeline and challenges of working with large volumes of photon-level data. We summarize the current astrometric and photometric performance compared against the original mission pipeline, and highlight an intra-visit science example studying the flares of CR Draconis.

1. Introduction

1.1. GALEX Overview

The Galaxy Evolution Explorer (GALEX; Martin et al., 2005) was a NASA Small Explorer (SMEX) telescope that surveyed the sky in the ultraviolet over ten years, between launch on 28 April 2003 and spacecraft termination on 28 June 2013. The spacecraft, instruments, data and calibration are well described in previous publications (Morrissey et al., 2005, 2007) and the missions technical documentation[2]. We will restrict discussion to topics that have not appeared elsewhere in the literature, are of particular importance to the gPhoton project, or are otherwise necessary for completeness.

GALEX carried two microchannel plate detectors (MCP), simultaneously exposed via a dichroic, and had a 1.25 degree field-of-view (FoV). It observed in two broad ultraviolet (UV) bands centered around 1528 Å (Far Ultraviolet or "FUV") and 2271 Å (Near Ultraviolet or "NUV"). The FUV detector failed in May of 2009, but the NUV detector continued to operate until the end of the mission. The detectors could observe in either direct imaging or slitless spectroscopic (grism) modes. Observations were conducted while the spacecraft was on the night side of each orbit (so-called "eclipses"), which lasted 1500-1800 seconds. To avoid detector burn-in or local gain sag effects caused by depletion of electrons in the multiplier plate, and to smooth over local irregularities in detector response, the telescope did not stare at a fixed location on the sky throughout an observation. Instead, it continuously moved the boresight relative to the target position. Several boresight patterns, or "modes", were used over the course of the mission, and these can impact the data in different ways that must be accounted for in the calibration procedures.

In the most basic "dither" mode, the spacecraft boresight would trace out a tight spiral pattern with a radius of $\sim 1'$. This mode was used most often for Deep or Medium Imaging Surveys (DIS, MIS) in which each full eclipse was spent observing a single region of the sky. In the Allsky Imaging Survey (AIS) mode, the spacecraft boresight would jump between 12 positions (or "legs") on the sky for short integrations of ~ 100 seconds each. Between each leg, the detector was set to a non-observing, low voltage state. This resulted in one independent observation (or "visit") per leg. Another mode, called "petal pattern", was used to distribute the flux from particularly bright targets across the detector. This was similar to the AIS

---

*Corresponding author

[1]Current Address: National Optical Astronomy Observatory, 950 N. Cherry Ave., Tucson, AZ 85719

[2]http://www.galex.caltech.edu/wiki/Public:Documentation

mode, except that there were a total of 8 legs, and they were tightly clustered within the approximate area of a single FoV. Unlike in the AIS survey, when in the "petal pattern" mode the detector continued to observe during the transition periods.

On 4 May 2010, an event referred to by the mission team as the "Course Sun Point" (CSP) anomaly, referring to the safe mode entered by the spacecraft at that time, resulted in image degradation of the NUV detector. This produces severe streaking in the detector's Y-direction, most probably due to a failed capacitor. Although the effect was largely corrected through subsequent calibration and onboard adjustments[3], observations taken between 4 May and 23 June 2010 have substantially worse point spread functions.

NASA support for the mission ended in February of 2011. At that time, ownership of the spacecraft was transferred to the California Institute of Technology for a mission phase called the "Complete the All-sky UV Survey Extension" (CAUSE), during which operating costs were solicited from individuals or institutions, and field location and source brightness constraints were relaxed. This enabled observations of bright regions of the sky not observed during the primary mission, and spacecraft slew rate limits were also relaxed, permitting a high-coverage "scan mode" that swept across several degrees of sky in a single integration. Ownership of the CAUSE-phase data resides with each of the primary investigators, and only a small fraction of it has been made available to the public at MAST at the time of writing, although much of the raw data has been copied to the MAST archive. Although the new calibration capabilities of gPhoton may be of particular value in using and interpreting CAUSE data, in particular, the scan mode observations or observations of very bright or dense fields, this paper (and the current gPhoton database) only includes the direct imaging data through the end of the NASA-supported mission, corresponding to General Release 7 (GR7). Future work may add gPhoton support for CAUSE phase, scan mode, or spectroscopic data collected throughout the mission. Through GR7, GALEX collected data over 34,389 (direct image) eclipses, which amount to ~ 18 TB of raw data and ~ 14 TB of reduced data, and covers 76.9% of the sky.

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 MAST. In Section 4 we describe the primary modules to use for constructing light curves and images. In Section 5 we highlight some of the implementation challenges we encountered, and our solutions to those problems, 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 an example science case enabled by gPhoton: stellar flares from CR Draconis.

2. Motivation

Of particular relevance to gPhoton, MCP detectors are non-integrating photon detectors that can record position and time information individually for every detected event. The GALEX detectors were capable of recording data with a time resolution of five microseconds, though the vast majority of observations were made in a "compressed" mode at five millisecond resolution. Due primarily to computer storage and processing constraints, calibrated GALEX data was only released and archived as per-visit or multi-visit ("coadd") integrated image maps, which have exposure depths on the order of hundreds to thousands of seconds. A minimum amount of documented analysis was performed by the mission team to understand the detector performance or calibration over timescales shorter than ~ 100 seconds.

Advances in storage and processing capabilities since the beginning of the mission now make storage, distribution, and analysis of the photon-level data technologically feasible. However, by the end of the mission, the GALEX calibration pipeline software (hereafter referred to as the "mission pipeline") had grown to sufficient complexity and dependence on its operating environment that attempting to run it outside its native environment proved unsuccessful. We have undertaken the gPhoton project, in part, to migrate key functionality of the mission pipeline into a stand-alone pipeline that can process raw spacecraft detector telemetry into calibrated lists of time-tagged photon events at five millisecond resolution. Another objective was to enable the creation of calibrated light curves and images at user-specified spatial and temporal scales, enabling in part, studies of short time-domain variability in the UV. Our design goals included the following key components:

- Creation of a database, containing (nearly) all photon events from the mission, that can be queried efficiently.

- Software that can perform the necessary calibrations (astrometric, photometric, exposure time, etc.), at least comparable to the original mission pipeline, over visit- and coadd-level timescales.

- Allow users to create an image (as a coadd over a specified integration time t) or an image cube (as a series of images, each with a specified integration time b).

- Allow users to create light curves at a specified time interval b.

- Lower the barrier to entry, such as: minimizing the number of primary (forward-facing) modules most

---

[3] http://www.galex.caltech.edu/wiki/Public:Documentation/Chapter_8

2

users will need, wrapping the SQL queries behind a python interface, creating commonly used output file formats like FITS and comma separated values (CSV), etc.

While gPhoton does reproduce much of the core functionality of the mission pipeline in an open source and user-friendly package, we would like to emphasize that it is not intended as either a full migration or a faithful port of the original mission pipeline. As will be described, archived output files from the mission pipeline have been utilized where deemed expedient, and some deviations from the original calibration and reduction methodology have been made in service to both computational efficiency and the unique properties of photon-level data.

3. Description of the Database

3.1. Description of Data Products from the Mission Pipeline Used By gPhoton

During the GALEX mission, data were downlinked from the spacecraft and assembled into monolithic telemetry files (-tlm). The ingest stage of the mission pipeline split these into various types of encoded raw detector event and spacecraft state (-scst) data, which included coarse aspect solutions from the onboard star tracker at one-second resolution, as well as spacecraft housekeeping records. The most important class of encoded raw detector data for gPhoton, containing nominal scientific observations, were the -raw6 files. The -raw6 were decoded with a sequence of bitwise manipulations into lists of raw detector positions (x and y) with timestamps for all detector events. These raw positions were further adjusted with "static" (detector-space) calibrations for wiggle, walk, nonlinearity and distortion, described more completely in Morrissey et al. (2007). For post-CSP data (after eclipse number 37460), the calibration was slightly modified because several onboard detector constants changed or were modified, and because an additional processing step was added for the YA value of the raw position data to correct for detector streaking caused by this anomaly. [Describe the post-CSP corrections better, especially what is the "YA" value?.]

3.2. Construction of the Database

The database is populated using photon list files produced by running event extraction on the full corpus of data, up to and including GR7. The first step in generating the database was to create aspect-corrected photon events as CSV files. Starting with the raw photon event (-raw6), refined spacecraft attitude (-asprta), and spacecraft state (-scst) files available at MAST, the celestial coordinates 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, the event positions on the detector, the aspect-corrected positions on the sky (as RA and DEC), and status flags used to track a variety of conditions related to the detector readout. We should add these as a table if they might AT ALL be relevant. We should err on the side of inclusion when it comes to details.

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. Therefore, a total of four CSV files are created: two CSV files for FUV and NUV data that contain aspect-corrected photon events, and two CSV files for FUV and NUV data containing photon events that were not aspect-corrected. The non-aspect-corrected photon events are still loaded into a database, because they are used for estimating dead time corrections (see Section 5.1 for details).

Null events can also arise from cases where events fall outside of the main detector FoV such as: erroneously data or stim events, events that are not aspect-correctable because they fall in a time period that is in a gap between refined aspect solutions [I don't understand this one (SWF)], events that occur between eclipses or visits [but why would these be outside the detector FoV?], or simply due to a failure of aspect refinement, they are not aspect-corrected. The right ascension and declination values for such events are assigned a value of null in the photon list file. The distinction between this "null data" and nominal or "non-null data" is important, as it is treated differently in both the database and high level calibration. We should also discuss masked pixels, how those are assigned currently (null or not null), a how those are treated.

For performance optimization purposes, the event-level data is distributed over multiple databases. The non-null (aspect corrected) data are spread across ten different databases organized by celestial declination, with bounding ranges selected to approximately balance the number of rows per database. The ten database tables are further divided into a total of 999 partitions. Each partition is then further divided into "zones" of $30''$. We make use of the fast zone matching algorithm described in Gray et al. (2006) for loading and querying the database. 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 is distributed evenly across that eclipse's footprint. Then, the cross-section of that eclipse's footprint against the zone boundaries 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 par-
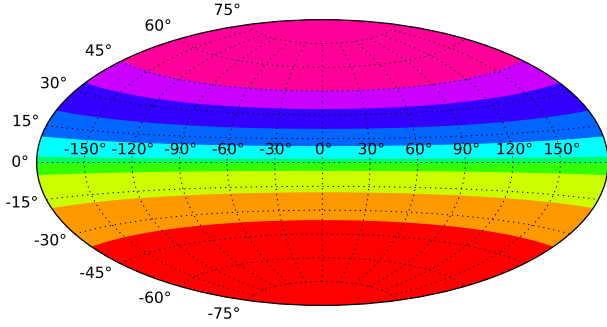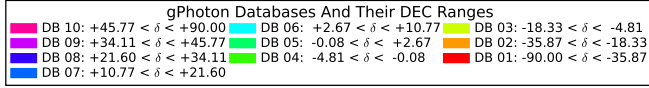
Figure 1: The DEC boundaries of the ten databases that comprise gPhoton. There are a total of 999 partitions across the ten databases, each with a variable number of $30''$ zones (stripes of DEC). The number of zones in each partition, and the number of partitions in each database, were assigned so that the size of the ten databases would be roughly equal to each other.

tition assignments by not requiring to actually compute the zone membership for all 1.1 trillion photon events beforehand. The distribution of the ten databases on the sky is shown in Fig. 1 Bernie: Please confirm boundaries are correct., along with a table summarizing the declination ranges and number of photon events in each database Scott will make this table once Bernie gets him the numbers. The assignment to a database or partition is strict, leaving open the possibility that a user's query spans two databases. The majority of normal queries access only a single database, but queries that do span multiple databases are handled on the server, transparent to both the software and end users. Null data which were not aspect corrected reside in a single database. What are the indexes of the databases?

4. Description of the Software Tools

Within gPhoton, in general, sky positions are reported as a two-element vector (right ascension and declination), both in J2000? decimal degrees. Time ranges (or "bins") are defined as two-element vectors where the first element is the start time and the second element is the end time. The gPhoton project defines timestamps in "GALEX time" throughout, where $t_{\mathrm{GALEX}} = t_{\mathrm{UNIX}} - 315964800$. Search ranges (in both space and time) are generally taken to be inclusive of the lower value and exclusive of the higher value, to eliminate duplicate counting of data at boundaries.

By default, the database tools define an effective detector FoV 1.1 degrees in diameter. This is to conservatively trim data observed on the edges of the GALEX MCPs, which suffer from poorly understood issues of sensitivity and spatial distortion. Data near the edges may be useful to cautious and knowledgeable investigators, though, so

Module	Function
gPhotonPipe	Generates aspect-corrected photon lists from a small set of user supplied input files. The input files are normally archived products from the original mission pipeline. Output from gPhotonPipe was used to populate the photon event database that the other modules query. Most users will not need to use this module.
gFind	Provides information on the available data coverage for a given area of the sky.
gAperture	Creates a light curve (returned as a table of times, calibrated fluxes, and additional parameters) for a given coordinate, time sampling, and aperture size.
gMap	Creates an image (in units of counts and/or calibrated fluxes) and/or image cubes (also in units of counts and/or calibrated fluxes), for a given area of the sky and (for cubes) time sampling.

Table 1: Summary of Primary gPhoton Modules

the effective detector size is adjustable from the command line. This conservative trimming of the effective FoV does not eliminate problems caused if there's a region of interest that extends towards the edge. These can include photometric apertures, background annuli, or image footprints that extend past the edge of the effective FoV. The best photometry will come from sources that are clear of both the physical and effective FoV boundaries.

There are four primary modules included in gPhoton and described in Table 1. These utilities are all written in Python and released under a permissive license. With the expectation of gPhotonPipe, the tools can be called either from the command line or imported as a module. If imported as a module, output is returned in a data structure. These command line utilities draw upon a large a large number of supporting libraries which will not be described in this paper, but are possibly of interest to users who want to perform advanced or specialized analyses with the gPhoton data. For more information, users are encouraged to consult the documentation available in the software repository, or at the MAST page for the project: https://archive.stsci.edu/prepds/gphoton/.

4.1. gPhotonPipe

The gPhotonPipe calibration implements a subset of the steps from the original mission pipeline in order to perform detector-level calibration and aspect correction of photon events. The module can be run through the command line, and accepts the raw scientific data file (-raw6), the spacecraft state file (-scst), and one or more refined aspect solution files (-asprta). It returns a "photon list

file" in CSV format, where each row corresponds to a detector event and records information such as the raw and calibrated detector positions, projected (de-dithered) sky positions, and a flag that encodes metadata on the photon event, such as the quality of the aspect solution used in the projection and whether the event falls in a known detector hotspot region. We need to add a table with the current columns and what they mean. The photon list files produced by gPhotonPipe are analogous (but not identical) to the extended photon list (-x)? files that were occasionally produced by the mission, but not archived at MAST Confirm this is true that MAST never archived these.

Note that all the inputs to the stand-alone calibration pipeline (-raw6, -scst, and -asprta files) are products from the original mission pipeline that are archived at MAST. By using these archived mission products directly, we avoid the need to recreate either the ingest or aspect correction stages of the mission pipeline. If an aspect file is not supplied, then gPhotonPipe will attempt to query the MAST database of -asprta data. Manually specifying the aspect file is possible for researchers who wish to further refine or modify the mission-provided aspect solutions.

4.2. gFind

This module is a search tool that allows the user to query the available coverage depth of a particular target. Given a target position, gFind returns the estimated (raw) exposure depth of available data over the whole mission, separated into time ranges corresponding roughly to discrete observations of the target. Rather than using the visit-based bookkeeping of the mission, which distinguished between observation modes and survey type, gFind uses the photon events themselves. A given position on the sky is considered to be observed if valid data exists in a time range where the position falls within one "effective FoV radius" of the spacecraft boresight, as given by the mission-provided refined aspect solution. The effective FoV can either be the default or a user-defined effective FoV from the command line. Distinct time ranges are identified based on a user-adjustable parameter that defines the maximum allowable gap between events for data to be considered contiguous. This parameter defaults to one second, the time resolution of the aspect solutions. Another user-defined parameter sets the minimum exposure depth (in seconds) required for a time range to be reported back.

4.3. gAperture

This module extracts and calibrates event-level data from the database to produce light curves, given user-specified parameters that can include target position, photometric aperture, background annuli sizes, desired integration depth (i.e., bin size), and any time range constraints. Rather than performing photometric measurements on pixelized and integrated images, as the mission pipeline did, gAperture conducts cone searches on the sky positions of individual photon events for the purposes of aperture photometry and background subtraction, enabling analysis at the native spatial resolution of the data. Each photon event is weighted by the detector flat value at the spot on the detector on which it occurred, using the mission-produced flats. All events within a given time bin are then weighted by the effective exposure time. At this point, gAperture does not apply aperture corrections to the fluxes, although look-up tables are available in the repository specify where, taken from cite the sources here. This would be a good spot to review each step in the process, i.e., a flow diagram.

The output from gAperture are CSV-formatted tables, with columns that include ...we should describe at least the primary columns, and how they are calculated where necessary. We would like to emphasize, though, that the output format and column definitions are subject to change as gPhoton is developed further, so users should always consult the documentation available in the repository.

4.4. gMap

This module creates integrated images and/or image cubes for targeted regions of the sky and specific time ranges, up to and including full depth coadds. Users can request either "count" images, which have not been corrected for exposure time or response (often useful for astrometry, diagnostics, or quick-looks), or "intensity" images, which are fully calibrated and suitable for photometric analysis. The images produced by gMap are analogous to the imaging data products produced by the mission pipeline, but with additional flexibility via user-tweakable parameters (e.g. dimensions, depth, edge trimming). If given a sequence of time ranges or a bin size, gMap will also produce "count" and/or "intensity" image cubes, which, at full resolution, were never available from the mission pipeline. All images are written in the Flexible Image Transport System (FITS) format that include headers populated using the World Coordinate System (WCS) standard. As with relative response correction in gAperture, rather than generating a relative response map, the individual events are simply weighted by the flat value assigned to the detector regions on which they fell. [Spatial exposure time correction is not implemented yet! No exposure time maps yet.]

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 MCP or photon-level data investigations. The details may also be of general interest to users of our software suite. We elaborate on several of these key components related to calibration and processing of the gPhoton data, noting that as the software is continually improved, some of these aspects may change. Whenever conflicting information exists, the documentation in the repository takes precedence.

5.1. Exposure Dead Time

Microchannel plates are subject to a global exposure "dead time" effect caused by the inability of the detector to process more than one event at a time. The effect therefore scales as a function of total global detector count rates: as count rates go up, the fraction of exposure lost to dead time likewise increases. The effect is linear as a function of "global count rate", up to approximately [add rates in NUV and FUV]. The global count rate is just the count rate across the entire FoV during an observation Confirm this definition of global count rate.. Other MCP instruments have used this global count rate relationship as a means to estimate the dead time correction factor. The GALEX detectors, however, were equipped with four built-in electrical pulsers ("stims") located off the main detection window that produced a known rate of events (between all four, 79 counts per second Confirm this number and that it's across all four detectors.). The mission pipeline extracted these stim events as a direct proxy for the exposure dead time effect. That is, the ratio of the measured stim count rate to the canonical stim count rate is equivalent to the ratio of the effective exposure time to the commanded exposure time. However, one must also consider the contribution of counting statistics in this estimate.

The shortest standard GALEX integrations on single targets were the ~ 100 second observations in the AIS survey. Assuming zero dead time, the 1-sigma error in the stim rate due simply to Gaussian counting statistics would be around 0.9 counts per second (cps) or $\sim 1\%$. This level of error is negligible as compared to other sources, and, indeed, was not even propagated by the mission pipeline. For shorter integrations, however, like those that can be produced by gPhoton, the errors in the exposure time produced by estimating dead time from stims become quite meaningful. For a ten second and one second observation, respectively, with zero dead time, the 1-sigma errors are ± 2.81 cps ($\sim 2.5\%$) and ± 8.9 cps ($\sim 10.1\%$). The problem compounds for real observations, during which the fractional dead time can range from 10% to 30%, which for a 10-second integration, produces 1-sigma errors of $\sim 3.8\%$ and $\sim 4.3\%$, respectively.

The solution used by gPhoton is to use the linear relationship between global count rate and dead time, which holds for reasonably low global count rates, to produce an empirical exposure time correction as a function of global count rate. GALEX has typical global count rates of 10,000 cps or more, making the 1-sigma error due to counting statistics truly negligible even for short integrations. The mission team did produce such an empirical dead time formula early on, but it suffered from a few problems that required us to revisit the problem and produce our own analysis. For one, while the result was recorded, the actual methodology was not, making it impossible to reproduce. Also, because it was early in the mission, extremely high count rate fields (in the non-linear response regime) of the detector had not been observed and were therefore not included in the analysis. Finally, the original analysis produced only a single model that, because their behavior was deemed so similar, simultaneously covered both the FUV and NUV detectors.

We plot global detector count rates against stim count rates. [Include this plot with best-fit slopes for both bands!] We used exposure times that had been corrected for shutter (see Section 5.2), but not dead time, when calculating these rates. In this analysis, we assume the foreground signal follows a linear model with two parameters (slope and intercept), and Gaussian counting errors in both x and y. For the background, we also used a linear model with unknown parameters. This "mixture model" was sampled by Markov Chain Monte Carlo (MCMC) against the data for ~ 1000 observations in each band to produce maximum likelihood model parameters for the stim count rate as a function of global count rate in both bands, which could then be converted to a fractional dead time by comparing the stim count rate against the reference rate. Rather than directly adopting the quoted stim reference rate of 79 cps for both bands, we used the maximum likelihood y-intercept value as the reference rate. This produced models that differed in each band by about 3%, which can be significant for GALEX. Thus, gPhoton treats the dead time correction slightly differently than the mission pipeline. Are we using the two different reference stim rates for FUV and NUV in the current version of the code?

5.2. Effective Exposure Time

The effective exposure time is computed as the raw exposure time (the difference between the observation start and end time), minus the amount of time considered "shuttered", then scaled by the global dead time. Add equation here. Shuttered periods are those periods of significant length, for our purposes, considered to be 0.05 seconds or more, during which no valid observational data is available. These might be periods during which the spacecraft was not actually observing the requested region of sky, but can also include data dropouts or periods during which a valid aspect solution is not available for any number of reasons, including being near the beginning or end of an observation or a failure of the aspect refinement stage of the mission pipeline. For aperture photometry, the effective exposure is computed at the requested sky position, and then applied uniformly across all events in both the aperture and background annulus. This approximation is more efficient than calculating the exposure across the whole region and fails only when the annulus or background contains a masked part of the detector (e.g. hotspots as described below) or crosses the edge of the effective FoV.

Prior GALEX documentation has suggested that users of the photon-level data assume that the observation-level exposure time is distributed evenly across the whole observation [CITE]. This assumption, though, is not true, and sometimes quite significantly so. Not only can shutter be distributed non-linearly over an observation[what does this mean? SWF], but the global count rate can change quite

substantially, even during an observation of a single target, because the detector is constantly moving on the sky, causing small changes in the global count rate due to field brightness changes. At the same time, the spacecraft is traveling through the shadow of the Earth, creating global brightness shifts due to airglow during the observation. If one plots global count rates as a function of time for a typical 1600 second MIS observation, there will be a characteristic, broad parabolic shape. This shape is inverted in a plot of effective exposure time as a function of the observing timestamps, because higher global count rates create more detector dead time, resulting in less effective exposure time. Researchers interested in the variability of targets within single observations should re-compute the exposure time for every bin in order to account for this bias. If using gAperture this is done automatically as part of the light curve creation. A plot demonstrating the smile/frown should be added here.

5.3. Background Correction

There are currently two methods gAperture uses to correct for background flux. The first method is a simple annulus estimation method where the total surface flux within an annulus surrounding the extraction aperture is scaled to the area of the aperture and subtracted. This method can be biased by the presence of relatively bright sources either within or near the annulus, although the bias can sometimes be mitigated by careful selection of annulus parameters (and assuming that the field is not crowded). The direct annulus method has the benefit of correcting for changing local background over an observation due to zodiacal light or airglow. It produces good agreement with MCAT [Make sure MCAT is explained.] values on average for integrations of AIS depth or longer, but with a larger dispersion, particularly when brighter stars that are in or near the annulus have caused an erroneously high background estimate.

The second method is to estimate the local sky background based upon backgrounds of nearby sources as reported in the official GALEX source catalog (MCAT). In this method, the coadd MCAT is searched for sources within a small region of the target position. The arithmetical mean of reported flux of all returned MCAT entries is then used as the estimated background at the target location. As one might expect, this method results in gAperture fluxes that are in good agreement with the MCAT on average, and with good dispersion, but does not correct for either inter- or intra-visit variations in the sky background.

We suggest that the first method be used for studies of intra-visit variability. The second method may be useful for researchers wishing to compare gPhoton results to previous work that used data from the GALEX catalogs. Note that the first method could also produce false variability if the source(s) in the aperture are not changing, but one or more sources in or near the annulus are. When this is a possibility, we suggest that researchers check the MCAT

as well as a gMap-produced image of the targeted region for nearby sources, and run an independent photometric analysis for those.

We also tested two variations of the annulus method that might mitigate the presence of stars in or near the background annulus, both of which were abandoned because they produced poor agreement with catalog fluxes and they were highly sensitive to somewhat arbitrary input parameters. The first method, which we called "swiss cheese", mimicked that applied by Morrissey in the analysis of the GALEX standard star LDS749B [CITE]: nearby bright stars from the MCAT out to several sigma, assuming Gaussian profiles, were masked out by excluding those data and sky regions from subsequent calculations. The second method was a direct port of the "sigma clipping" algorithm used by the mission pipeline [and described in Ted's white paper [CITE]], which iteratively excluded and interpolated over relatively bright regions, using full Poissonian statistics to account for low, diffuse UV background.

5.4. Relative Response Correction

Within the mission pipeline, the detector flat was projected into the sky according to the spacecraft aspect solution, and then integrated and interpolated into a high resolution relative response (-rrhr) map, which could then be applied to the integrated count maps to correct for detector response and effective exposure (as a function of sky position), as well as any static or dynamic masking (of hotspots, ghosts, or edge reflections). The gPhoton pipeline, instead, applies the detector flat at the detector level by weighting each individual photon event by the value of the flat at the detector location upon which it fell. Is this basically the work of sampling the flat at higher spatial resolutions, or does there still need to be work done on that front?

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

We compare the mean photon position within the aperture to the MCAT source position for XXXX randomly selected MCAT targets. One source of uncertainty here is that the center of brightness(?) was calculated by the mission pipeline on a pixel-based map, whereas gPhoton uses the photons directly. Another possible source of error is that one should ideally fit a two-dimensional Gaussian
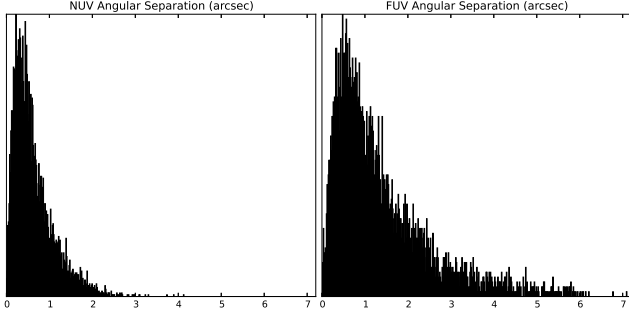
Figure 2: Angular separation between gPhoton and MCAT source positions for XXXX MCAT sources in both bands.
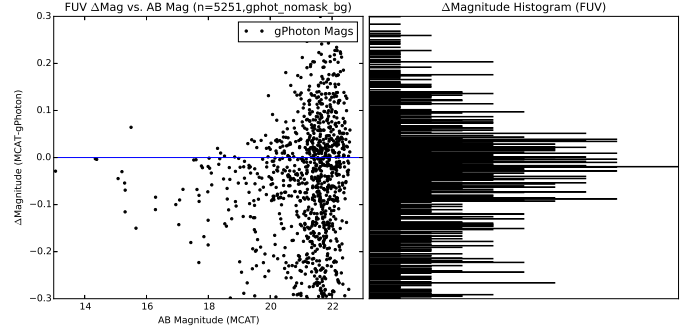

Figure 3: Comparison between MCAT and gPhoton FUV fluxes for XXXX randomly selected MCAT sources with exposure times $100 \leq t \leq 5000$.

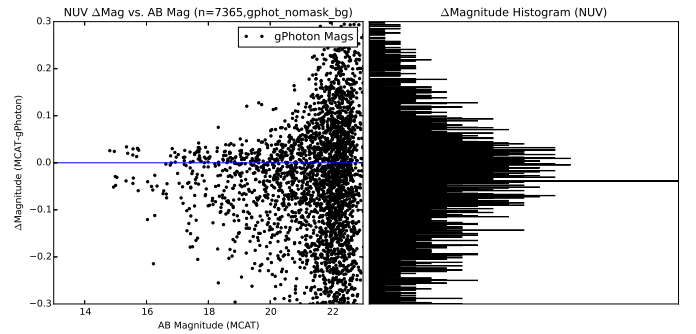Figure 4: Comparison between MCAT and gPhoton NUV fluxes for XXXX randomly selected MCAT sources with exposure times $100 \leq t \leq 5000$.

bounded by an off-center circle to find the center of brightness. Were the same sources selected from each band? Did the sources sample different observing modes, times, sky coordinates? We find that nearly all sources lie within one GALEX resolution element ($\sim 6''$), and that {XXXX, YYYY} lie within $XXXX''$ in the FUV and NUV bands, respectively (Fig. 2).

6.2. Relative Flux Precision

As a test of the relative photometric precision from gPhoton, we plot the difference in gAperture magnitude and MCAT magnitude against the MCAT magnitude for XXXX randomly selected MCAT sources. One of our goals is to be able to reproduce MCAT fluxes for most targets over visit- and coadd-timescales. All gAperture measurements were done at the MCAT source positions and over time ranges as reported in the MCAT, using an aperture radius of $0.1'$ (equivalent to "APER4" in the MCAT). To ensure sufficient signal to noise, only observations with depths greater than 100 seconds were used. To prevent any particularly deep fields or exposures from dominating the result, only fields with a total (estimated raw) exposure time of less than 5000 seconds were used.

We plot these comparisons for the FUV and NUV bands in Figs. 3 and 4, respectively. We find that XXXX% of sources agree to within 0.1 magnitudes. We need to discuss why there might be a difference between the two, and also why bright sources in the MCAT are fainter in gPhoton. The fluxes shown in the comparison plots make use of the aperture-method for background subtraction. We note that comparisons using the MCAT-values-method for background subtraction look quite similar.

6.3. Absolute Flux Precision

As described in Morrissey et al. (2007), the calibration standard star of choice for GALEX was the white dwarf LDS 749B, which is also one of the Hubble Space Telescope standards. A total of {259909, 595672} seconds of exposure depth in the FUV and NUV bands, respectively, were made on this source. We use gAperture to compute magnitudes for all of these observations in both bands and compare them against the standard magnitudes used by

GALEX (Fig. 5). We measure fluxes separated into visit levels to match the exposure times in the MCAT. The median difference in the FUV is XXXX and the median difference in the NUV is XXXX. The y-axis label is incorrect. What is causing the cloud of points above the main cluster in each, are those near edge, long/short exposures, post-CSP? Why is NUV fainter on average for gPhoton? There's no x-axis label.

7. Example Science Application - Stellar Flares From CR Draconis

CR Draconis (HIP 79796) is a fairly bright ($V \sim 10$) binary star system composed of two M dwarfs located ~ 20 pc away in a slightly eccentric orbit with a period of ~ 4 years (Tamazian et al., 2008), and has been known to exhibit flares for many decades now (Cristaldi and Rodonò, 1970). The system was identified as a high-amplitude variable in the second version of the GALEX Ultraviolet Variability (GUVV-2) Catalog (Wheatley et al., 2008), where a maximum NUV flux difference of two magnitudes was identified within the available visits at that time. Welsh et al. (2006) studied one CR Draconis' flare events with high temporal sampling by extracting light curves from sky-projected, "extended" (-x) photon list files, produced as non-standard products of the GALEX mission pipeline.
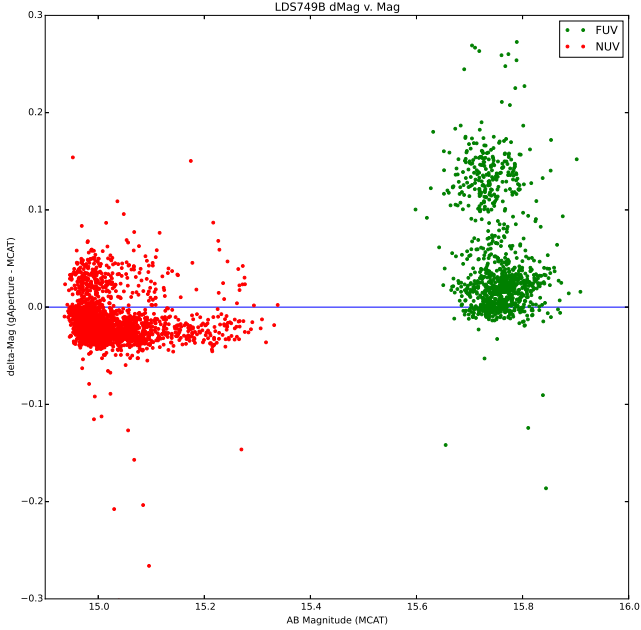
8

Figure 5: Comparison between MCAT and gPhoton fluxes for LDS 749B in the FUV (green) and NUV (red).

Using our gPhoton pipeline, we have searched for flares from the CR Dra system using all available GALEX data. The largest observed flare in GALEX is the one reported in Welsh et al. (2006), but we also see seven additional flares, spanning from 2003 through 2011 (nearly 2 full orbits of the binary). Fig. 6 shows each of the identified flares. When available, the FUV version of the light curves are shown in blue. Several of the flares are double-peaked, and some show elevated levels of flux before or after the flare event. There is also a range of amplitudes and durations, with some increasing in flux by less than a factor of two and lasting only a few minutes in duration. These short-duration flares have less an energy than the longer duration, stronger flares, but also can occur more frequently, and thus may still impact the habitability of exoplanets in those systems (e.g., Ramsay et al., 2013). While the total energy in a given flare event would be lower, there is also an increased chance for flare-planet alignment. There have also been studies of the flare rates in resolved M dwarf binaries as a function of orbital separation. The number of such binaries is small, but CR Draconis is one candidate, and given that the GALEX time baseline extends two full orbital periods, could improve the statistics from previous studies (Tamazian et al., 2008).

A detailed analysis of the flares is beyond the scope of this introductory paper, which serves to present the software itself, and is reserved for future papers that will focus on the astrophysics of flare stars with gPhoton. However, it is instructive to provide some short scripts demonstrating the basic work flow when creating the plot shown here. To define our photometric aperture, we used gMap to construct a deep coadd image, centered on CR Draconis, using
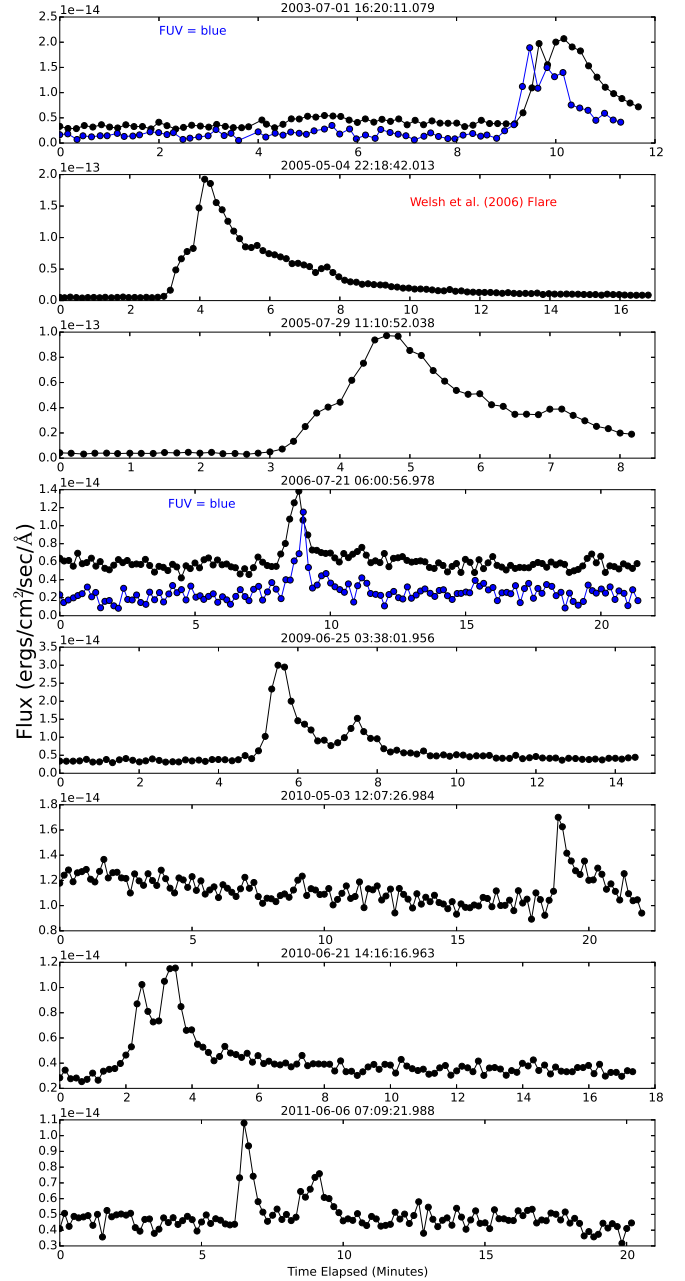

Figure 6: Flares detected on CR Draconis using gPhoton, across the lifetime of the mission. When available, FUV light curves are plotted (in blue) along with the NUV curves (in black). Fluxes have not been aperture corrected.
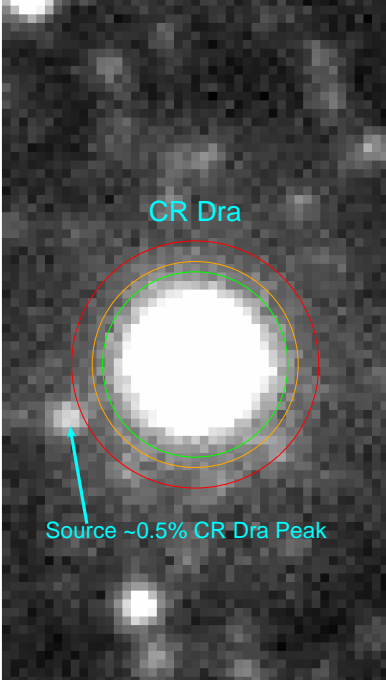
9

Figure 7: Deep coadd image of CR Draconis using all available NUV photon events. The image is in counts, since we are only looking to define our photometric aperture and search for possible contamination sources. The aperture, inner annulus, and outer annulus for photometry are represented by the green, orange, and red circles, respectively. There is a source to the lower left that is ∼ 0.5% the peak of CR Draconis itself. Although testing showed it did not have a major impact on the photometry significantly, we still define our apertures such that it is not included within them.

all available photon events (Fig. 7). Here's the example script that was used, which assumes you have installed gPhoton such that the software lies in your PYTHON-PATH environment variable, and thus can be imported as a package (see Inline Supplementary Code #1, available in the online version of the article):

With our apertures defined (we also adjusted the center of the apertures, due to the moderate proper motion of the star), we can construct our light curve file using gAperture. Here's the example script that was used to create the CSV file (see Inline Supplementary Code #2, available in the online version of the article):

After the CSV file is created, we wrote a separate script that reads in the CSV file, converts the t_{mean} timestamps from GALEX time to Julian Date, and then defined x-axis boundaries to center on each of the eight flare events of interest. Note that we did not apply aperture corrections to the fluxes shown in Fig. 6, but such corrections are available in Fig. 4 in Morrissey et al. (2007). A simple interpolation scheme is provided in gPhoton, using those values, called "apcorrect1" within the "galextools.py" module.

8. Conclusion

The GALEX mission has already proven to be extremely productive, and is likely to be one of the most influential ultraviolet astronomical surveys for the foreseeable future. Nevertheless, even when preparation of the higher level data is well documented and understood by future researchers, the priorities or interests of those users may not be the same as the data creators or archivists. In such cases, their normal recourse would be to go back to some minimally reduced version of the data and create their own procedure for further reducing the data. This can be onerous, time consuming, or impossible depending on the type of data, the quality of the documentation, and the availability of original team members to answer inevitable questions. The burgeoning practice of "reproducible research" seeks to address this, to some extent, by making the actual methods (as opposed to mere descriptions of the methods) available to researchers.

The gPhoton project is a trial into a new paradigm for data archiving, where the correlation between minimally reduced and high level data is explicitly laid bare through open-source software. That machinery is archived for future researchers to make modifications to the reduction methodology with minimal overhead or barriers to entry. Specific to GALEX, the gPhoton project simplifies, and in some cases enables, analyses using this data that were previously difficult-to-prohibitive, especially those related to short time domain photometry at the intra-visit level.

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 queries at 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 solvable 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.

9. Acknowledgments

Cristaldi, S., Rodonò, M., 1970. Photoelectric observations of flare stars. I. A&AS 2, 223.

Gray, J., Nieto-Santisteban, M.A., Szalay, A.S., 2006. The Zones Algorithm for Finding Points-Near-a-Point or Cross-Matching Spatial Datasets. Technical Report MSR-TR-2006-52. Microsoft Research. URL: http://research.microsoft.com/apps/pubs/default.aspx?id=64524.

Martin, D.C., Fanson, J., Schiminovich, D., Morrissey, P., Friedman, P.G., Barlow, T.A., Conrow, T., Grange, R., Jelinsky, P.N., Milliard, B., Siegmund, O.H.W., Bianchi, L., Byun, Y.I., Donas, J., Forster, K., Heckman, T.M., Lee, Y.W., Madore, B.F., Malina, R.F., Neff, S.G., Rich, R.M., Small, T., Surber, F., Szalay, A.S., Welsh, B., Wyder, T.K., 2005. The Galaxy Evolution Explorer: A Space Ultraviolet Survey Mission. ApJL 619, L1–L6. doi:10.1086/426387, arXiv:astro-ph/0411302.

Morrissey, P., Conrow, T., Barlow, T.A., Small, T., Seibert, M., Wyder, T.K., Budavári, T., Arnouts, S., Friedman, P.G., Forster, K., Martin, D.C., Neff, S.G., Schiminovich, D., Bianchi, L., Donas, J., Heckman, T.M., Lee, Y.W., Madore, B.F., Milliard, B., Rich, R.M., Szalay, A.S., Welsh, B.Y., Yi, S.K., 2007. The Calibration and Data Products of GALEX. ApJS 173, 682–697. doi:10.1086/520512.

Morrissey, P., Schiminovich, D., Barlow, T.A., Martin, D.C., Blakkolb, B., Conrow, T., Cooke, B., Erickson, K., Fanson, J., Friedman, P.G., Grange, R., Jelinsky, P.N., Lee, S.C., Liu, D., Mazer, A., McLean, R., Milliard, B., Randall, D., Schmitigal, W., Sen, A., Siegmund, O.H.W., Surber, F., Vaughan, A., Viton, M., Welsh, B.Y., Bianchi, L., Byun, Y.I., Donas, J., Forster, K., Heckman, T.M., Lee, Y.W., Madore, B.F., Malina, R.F., Neff, S.G., Rich, R.M., Small, T., Szalay, A.S., Wyder, T.K., 2005. The On-Orbit Performance of the Galaxy Evolution Explorer. ApJL 619, L7–L10. doi:10.1086/424734, arXiv:astro-ph/0411310.

Ramsay, G., Doyle, J.G., Hakala, P., Garcia-Alvarez, D., Brooks, A., Barclay, T., Still, M., 2013. Short-duration high-amplitude flares detected on the M dwarf star KIC 5474065. MNRAS 434, 2451–2457. doi:10.1093/mnras/stt1182, arXiv:1306.5938.

Tamazian, V.S., Docobo, J.A., Balega, Y.Y., Melikian, N.D., Maximov, A.F., Malogolovets, E.V., 2008. Preliminary Orbit and Differential Photometry of the Nearby Flare Star CR Dra. AJ 136, 974–979. doi:10.1088/0004-6256/136/3/974.

Welsh, B.Y., Wheatley, J., Browne, S.E., Siegmund, O.H.W., Doyle, J.G., O'Shea, E., Antonova, A., Forster, K., Seibert, M., Morrissey, P., Taroyan, Y., 2006. GALEX high time-resolution ultraviolet observations of dMe flare events. A&A 458, 921–930. doi:10.1051/0004-6361:20065304, arXiv:astro-ph/0608254.

Wheatley, J.M., Welsh, B.Y., Browne, S.E., 2008. The Second Galex Ultraviolet Variability (GUVV-2) Catalog. AJ 136, 259–266. doi:10.1088/0004-6256/136/1/259, arXiv:0805.2987.