photometric effects of RICE compression on GALEX full-depth images

Bekah Albach, Michael St. Clair, Chase Million (Million Concepts)

Most methods of compressing FITS images – for instance, gzipping entire FITS files – make "subsetting" or other types of random reads from those images impossible. FITS tile compression is a useful technique for reducing stored and transferred data volumes while still permitting random reads. In most cases, the RICE algorithm offers the best performance in tile compression. However, it is slightly lossy on floating-point arrays (lossless on integer arrays), which raises concerns about its effects on scientific analysis. To investigate this, we conducted photometric comparisons between lossless and RICE-compressed versions of full-depth images generated from all valid "MIS-like" GALEX eclipses (visits) in both NUV and FUV (when available; not all eclipses with NUV data also have FUV data). This document gives a brief summary of our most important findings. The bottom line is that we found that lossy RICE compression (parameters: qlevel=10, 100x100 tilesize, subtractive dither type 2) did not meaningfully affect photometry on these images. Our full data and code are available upon request. We hope this will serve as a jumping-off-point for similar analyses by data architects considering tile compression.

photometric reproducibility

The simplest way to look at photometric reproducibility is to compare source brightness. We executed DAOStarFinder (an implementation of the DAOPHOT algorithm in the *photutils* Python package) on lossless and RICE-compressed versions of each image, matched detected sources, and performed aperture photometry on these sources using portions of the gPhoton 2 pipeline. We found that the maximum and mean AB magnitude offsets for matched sources in all images are very small (Table 1). In fact, they are below the modeled threshold of photometric repeatability for GALEX (Million et al., 2016). Therefore, RICE compression does not have a meaningful deleterious effect on our GALEX photometry.

	Mean	Std	Max	Min	P25%	P50%	P75%
Max AB Mag Offset, NUV	0.004765	0.001285	0.016261	0.002449	0.004031	0.004434	0.005009
Mean AB Mag Offset. NUV	0.000821	0.000211	0.002608	0.000446	0.000718	0.00077	0.000833
Max AB Mag Offset, FUV	0.008303	0.002809	0.023514	0.00046	0.006146	0.007275	0.010159
Mean AB Mag Offset, FUV	0.001539	0.000578	0.00468	0.00046	0.001135	0.001258	0.001813

Table 1. Summary statistics for photometric differences for matched source detections between lossless and RICE-compressed images. The dataset includes 20,853 NUV eclipses and 10,667 FUV eclipses. Note that, because GALEX NUV images are much brighter, more sources are generally detected in the NUV than in FUV — the average number of source matches was 24,520 in the NUV and 4,551 in the FUV.

The largest AB magnitude offsets across all eclipses for sources matched between lossless and RICE-compressed images were 0.0235 in the FUV and 0.0162 in the NUV (Table 1). There is a strong positive correlation between low exposure time eclipses and larger mean AB magnitude offsets. For example, in the NUV, 100% of the eclipses with mean AB magnitude offsets greater than 0.0015 also have effective exposure times of less than 300 seconds. We believe this is basically a signal-to-noise ratio effect: eclipses with lower exposure times have lower photon count values (which are then scaled by exposure time to produce physical flux values), but the quantization noise introduced by RICE compression scales little with photon count values, so it has a larger relative effect in eclipses with lower exposure times.

Similarly, greater AB magnitude offsets are observed in sources of lower AB magnitude (e.g. Fig. 2). This is intuitive: the fainter a source is, the more likely that even small amounts of noise in the background or source will introduce photometric differences. These offsets in AB magnitude are roughly symmetrical: neither raw nor RICE compressed images are consistently greater in AB magnitude than the other (Fig. 2). This suggests that the subtractive dither used in the RICE algorithm is working as intended.

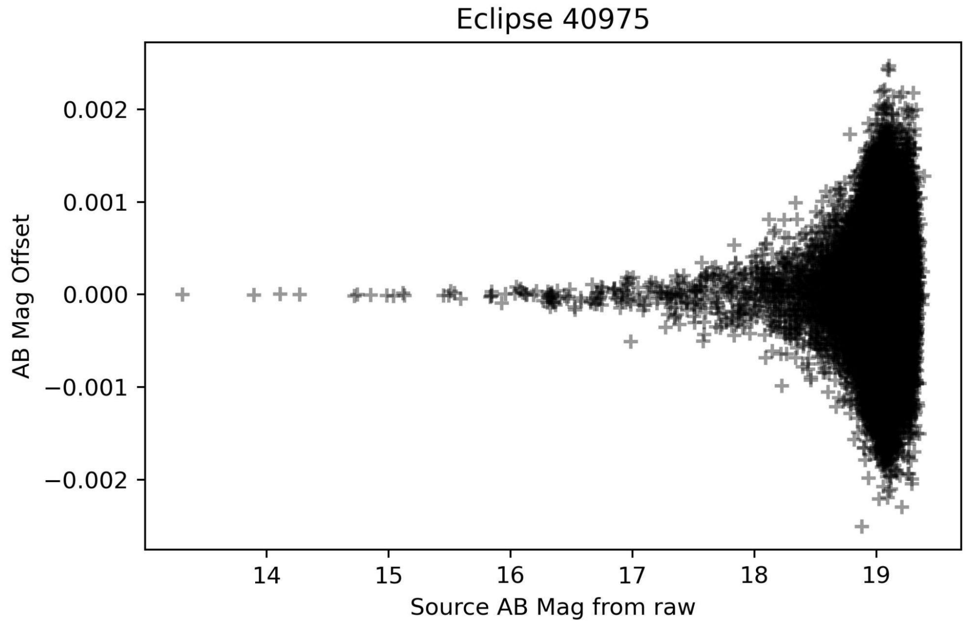

Fig. 2. Photometric offsets between lossless and RICE-compressed versions of individual sources (crosses) from eclipse 40975 in the NUV. The AB magnitude of each source on the x-axis is derived from the raw image.

source-finding reproducibility

Another way to study the effects of compression is to determine whether DAOStarFinder-detected sources we could *not* match between both the lossless and RICE-compressed images are valid source detections. DAOStarFinder identifies sources by searching for relative maxima above a specified threshold. For each eclipse, a number of additional sources in both raw and RICE were identified which were not within the one pixel threshold for being considered a “matching” source. In both the FUV and NUV, the number of total mismatched sources is approximately one tenth the number of matched sources. If RICE compression was introducing additional spurious noise, we might expect that the number of mismatches detected would be higher in the RICE-compressed images than in the lossless images. However, the number of unmatched sources detected in both RICE and lossless images is roughly equal.

Qualitative examination of images indicates that the majority of unmatched sources are either background noise that DAOStarFinder misidentified as a source due to a low threshold value or are stochastic point detections within an extended source. While RICE compression may ‘shuffle’ spurious source detections by DAOStarFinder, it does not seem to actually exacerbate this problem or interfere with valid source detections.

references

Bradley, L., Sipőcz, B., Robitaille, T., et al. (2022). astropy/photutils: 1.5.0 (1.5.0). Zenodo. https://doi.org/10.5281/zenodo.6825092

Stetson, Peter B. (1987). DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry. *Publications of the Astronomical Society of the Pacific* (99). https://ui.adsabs.harvard.edu/link_gateway/1987PASP...99..191S/doi:10.1086/131977