



OPEN ACCESS

Pushing the Limits of the Cosmic Origin Spectrograph (COS) with an Optimized Background Correction

Svea Hernandez¹ , Andrei Igoshev² , Jo Taylor³ , David Sahnou³, and Logan Jones³

¹ AURA for ESA, Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA; sveash@stsci.edu

² Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, UK

³ Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA

Received 2024 April 21; accepted 2024 June 13; published 2024 July 8

Abstract

Observations utilizing the ultraviolet capabilities of the Cosmic Origin Spectrograph (COS) onboard the Hubble Space Telescope are of unique value to the astronomy community. Spectroscopy down to 900 Å with COS has enabled new science areas. However, contrary to the situation at longer wavelengths, these observations are limited by detector background noise. The background correction currently applied by the standard calibration pipeline (Ca1COS) is not optimized for faint targets, limiting the scientific value of low signal-to-noise ratio (S/N) observations. In this work we investigate a possible dependence of the variations of the dark rate in both segments of the COS far-ultraviolet detector on time, detector high voltage (HV), and solar activity. Through our analysis we identified a number of detector states (on a configuration basis, e.g., HV and segment) characterizing the spatial distribution of dark counts, and created superdarks to be used in an optimized two-dimensional (2D) background correction. We have developed and tested Another COS Dark Correction (ACDC), a dedicated pipeline to perform a 2D background correction based on statistical methods, producing background-corrected and flux-calibrated spectra. While our testing of ACDC showed an average improvement in S/N values of ~10%, in a few cases the improvements in S/N reached 60% across the whole wavelength range of individual segments.

Unified Astronomy Thesaurus concepts: [Calibration \(2179\)](#); [Hubble Space Telescope \(761\)](#); [Ultraviolet astronomy \(1736\)](#); [Publicly available software \(1864\)](#)

1. Introduction

The Space Telescope Imaging Spectrograph and the Cosmic Origins Spectrograph (COS) onboard the Hubble Space Telescope (HST) have demonstrated the significant impact of ultraviolet (UV) observations on modern astronomy. Until the full-scale development of the next generation UV/Optical/Infrared observatories, e.g., Habitable Worlds Observatory,⁴ there are no prospects for large space missions with this integral coverage of the electromagnetic spectrum. Therefore, observations taken with these two instruments are of legacy value. We must continue to push the limits of these HST spectrographs, not only with future observations, but also through the usage of archival data.

At wavelengths <1130 Å the sensitivity of COS declines by a factor of 100 (compared to that at higher wavelengths), and is comparable to that of FUSE (McCandliss et al. 2010;

Hirschauer 2023). As a result, the COS coverage of such blue wavelengths has played an important role in scientific studies involving He II re-ionization (Worseck et al. 2011, 2016), Lyman continuum escape fraction from low-*z* galaxies (Borthakur et al. 2014; Izotov et al. 2016a, 2016b, 2018a, 2018b, 2021; Leitherer et al. 2016; Hernandez et al. 2018; Flury et al. 2022a, 2022b), circumgalactic medium (Tumlinson et al. 2013; McCabe et al. 2021), and cold ISM (James et al. 2014; Hernandez et al. 2020), to name a few. For many of these studies, observations are taken in the background-limited regime where an optimal detector background subtraction is critical. One example is the starburst galaxy Mrk 54 observed as part of HST program ID 13325, with a count rate at ~910 Å (rest frame) of $\sim 1.3 \times 10^{-6}$ counts s⁻¹ pix⁻¹, compared to the dark rate of COS at the time, $\sim 4.8 \times 10^{-6}$ counts s⁻¹ pix⁻¹.

We refer to counts not originating from photons incident on the detector as dark counts. These counts are a combination of radioactive decay of atoms in the microchannel plates; charged particles in the environment of HST; and other sources internal to the instrument. It has been previously reported that the COS far-ultraviolet (far-UV) background is somewhat correlated with solar activity (Dashtamirova et al. 2019), detector gain, high voltage (HV), and time (Figure 1). The exact behavior of

⁴ <https://science.nasa.gov/astrophysics/programs/habitable-worlds-observatory/>

these background variations is not yet fully understood. The first five panels in Figure 1 clearly highlight the variations in dark rate depending on the region of the detector and the time of the observations. Given the complex dependencies of the COS background with each of these parameters, a robust and well-characterized background correction as a function of time is required for an accurate analysis of observations of faint far-UV targets.

The standard COS calibration pipeline, CalCOS,⁵ determines the background contribution to science spectra by computing the average counts in pre-defined regions in that particular exposure. For COS far-UV observations there are two pre-defined background regions (typically below and above the science extraction region). The dark rate at the target location can differ from that at the pre-defined regions used for computing the background contributions to the target spectrum. Additionally, there are pixel-to-pixel variations in the spectral extraction region possibly differing from those at the background regions, which can lead to the over- or under-subtraction of the dark counts in the final spectrum. Performing a robust Pulse Height Amplitude (PHA; see Section 3) screening (Ely et al. 2014) along with optimizing the size of the extraction box can help reduce the background contribution. However, these practices do not guarantee an accurate background calibration. In addition to these methods, and given the importance of a reliable background subtraction, different research collaborations have independently worked to improve the calibration of low signal-to-noise ratio (S/N) far-UV observations (e.g., Leitherer et al. 2016; Worseck et al. 2016; Makan et al. 2021, FaintCOS).

The COS far-UV G140L/800 setting (central wavelength at 800 Å) was made available to the community in 2018, and was primarily designed for background-limited observations (Redwine et al. 2016). This new configuration decreases the background levels by a factor of two, and increases the S/N of background-limited data sets, with the caveat that it is restricted to low resolution observations ($R \sim 2000$). To further optimize the calibration of COS data products and especially benefit users observing in the limits of Poisson noise, here we present recently developed software to improve low S/N ($\lesssim 5$) COS far-UV observations by applying a more accurate and tailored background correction. This new background correction is performed on the two-dimensional (2D) COS science exposures, accurately modeling and subtracting the expected dark counts in the science extraction region itself.

This paper is organized as follows. In Section 2 we describe the COS dark observations used to characterize the background spatial and temporal variations. We present a brief summary of our in-depth investigation on the COS background noise in Section 3. Section 4 provides an overview of the parametric background correction and the adopted data-driven approach.

And lastly, in Sections 5 and 6 we introduce our newly developed software Another COS Dark Correction (ACDC) and present our concluding remarks, respectively. We note that a much more detailed technical report describing our investigation into the background properties of the instrument can be found in our official documentation page.⁶

2. Description of Observations

The spatial characterization of the behavior of the dark counts on the COS far-UV detector relied on the compilation of all the far-UV dark exposures stored in the STScI archive, Mikulski Archive for Space Telescopes (MAST⁷). We retrieved from MAST a total of 7190 COS far-UV dark exposures (including both segments A and B) available as of 2023 September. The observations cover a time period between 2009 June 6 and 2023 September 11. The majority of the dark exposures were taken as part of COS far-UV detector dark monitoring programs, which nominally collect five individual exposures every week, each with an exposure time of 22 minutes (1330 s). This set of 7190 exposures were taken at different HVs, with the most widely used being HV = 163, 167, 169, 171, 173, and 178 digital units (du). As detailed in Dashtamirova et al. (2019), the conversion between du and volts follows

$$\text{HV} = -(\text{du} \times 15.69 + 2500). \quad (1)$$

To facilitate the analysis and characterization of the COS background, once the dark exposures were retrieved from the archive, these were ingested into (1) an SQLite database, and (2) a MySQL database, each with different formats. The SQLite database was designed to track the dark counts on different regions of the far-UV detector over 25 s intervals. For each dark exposure, the dark counts are measured in five separate and standard regions of the detector (comparable to those defined in Figure 1; see Figure 9 in Dashtamirova et al. 2019), and inserted into the SQLite database. The parameters used to define these regions, as adopted by the COS team, are detailed in Table 1. The second database uses MySQL, and rather than tracking dark counts in time segments, each row stores information for each individual dark event recorded in the suite of COS dark exposures. The database contains 8 tables, one for each of the typically used HV settings (163, 167, 169, 171, 173, 175, 178) and one for all other HVs, often used in experimental or calibration observations.

3. Characterization of the COS Background

As part of this work we preformed an in-depth investigation on the factors influencing the background levels both globally and spatially, with a special focus on the PHA from each event.

⁵ <https://hst-docs.stsci.edu/cosdhh/>

⁶ <https://github.com/jotaylor/acdc-hst>

⁷ <https://mast.stsci.edu>

Table 1
COS Detector Regions used in the Monitoring of the Dark Rates

Region	Min x	Max x	Min y	Max y
Segment A				
Inner	1260	15,119	375	660
Left	1060	1260	296	734
Right	15,119	15,250	296	734
Bottom	1060	15,250	296	375
Top	1060	15,250	660	734
Segment B				
Inner	1000	14,990	405	740
Left	809	1000	360	785
Right	14,990	15,182	360	785
Bottom	809	15,182	360	405
Top	809	15,182	740	785

In brief, for each photon that lands on the COS far-UV detectors, a cascade of electrons is created which is characterized by a PHA that is then registered by the electronics (see Section 4.1.7 in Hirschauer 2023). It has been generally understood that external events are distinguished from background noise events by their PHA values, where dark events typically have low PHAs (PHA = 0–2) and real/external events register higher PHA values (PHA \geq 3). As noted in Sahnou et al. (2011), background counts are expected to display a negative exponential pulse height distribution (PHD). In contrast, the PHD observed for photon events is more commonly quasi-Gaussian.

For a detailed description of our work we refer to the technical document in our public page.⁸ Briefly summarized, we found that the PHDs from dark counts do not necessarily follow a simple power-law trend, instead they show a more complex distribution. While dark counts do show some correlation with the solar flux, this correlation is weak and only statistically significant in segment A. Additionally, we found that the spatial structure and temporal distribution of the dark counts are neither easily nor accurately described by a simple model. Therefore, our approach for modeling the dark counts is primarily data-driven.

4. Parametric Background Correction

The issue of accurately correcting the background contributions to the science spectrum can be described as follows: in a given pixel we observe a total number of counts defined as $N = n_s + n_d$ in each science exposure with exposure time of t_s . This number consists of n_s counts from the astronomical source and n_d counts from other sources, already identified as dark counts. We assume that both of these counts follow a Poisson

distribution $n_s \sim \text{Pois}(\psi_s)$ and $n_d \sim \text{Pois}(\psi_d)$ where ψ_s and ψ_d are the true fluxes for the target signal and dark counts, respectively. It is important to note that the total number of counts, N , also follows a Poisson distribution, namely $N \sim \text{Pois}(\psi_s + \psi_d)$. In a general case, the probability mass function for the Poisson distribution is described as

$$f(k, \psi) = \frac{\psi^k e^{-\psi}}{k!}, \quad (2)$$

where k is the total number of events, in this case N .

In a generic COS far-UV observation, it is likely that the scientific source is not precisely well-characterized, i.e., n_s . Instead, our approach is to analyze detector regions outside of both the science extraction region and the Wavelength Calibration Aperture (WCA) region, where the registered counts consist only of dark signal, therefore, $N = n_d$. Using this approach, we focus on accurately modeling the 2D spatial dark variations by using a combination of empirical superdarks. This allows us to parameterize the background subtraction, fixing the value ψ_d using either the empirical superdarks or background model. We adopt a data-driven modeling approach where we assume a finite number of detector states describing the dark behavior: Q, P_1, \dots, P_n , where Q is a quiescent state and P_1, \dots, P_n are peculiar states of the detector. Furthermore, we assume that at any given time the dark count distribution can be described as a linear combination of a quiescent and peculiar states:

$$W = \alpha_1 Q + \sum_{i=1}^N \alpha_{i+1} P_i. \quad (3)$$

The optimal values for the coefficients α_i are obtained through likelihood maximization. This maximization approach is done through the analysis of the 2D science exposures, intentionally excluding the counts registered in the science extraction and WCA regions. In an effort to obtain the best background model, our optimized software iterates over each pixel to estimate the dark count distribution, $W(x, y)$, and minimizes C statistic (Cash 1979) as follows:

$$C = 2 \sum_{i=0}^N W_i - N_i \log(W_i) + \log(N_i!) \quad (4)$$

where W_i is the predicted number of dark counts in a pixel with index i and N_i is the recorded number of photons in the same pixel during the science exposure.

4.1. Background Models: Multi-states

Given the complex correlations between the spatial and temporal trends with respect to solar activity and other factors, we adopted a systematic approach that identifies a relatively small number (2–10) of separate states of the detector for each segment and HV combination.

⁸ <https://acdc-hst.readthedocs.io/en/latest/about.html#the-finer-details>

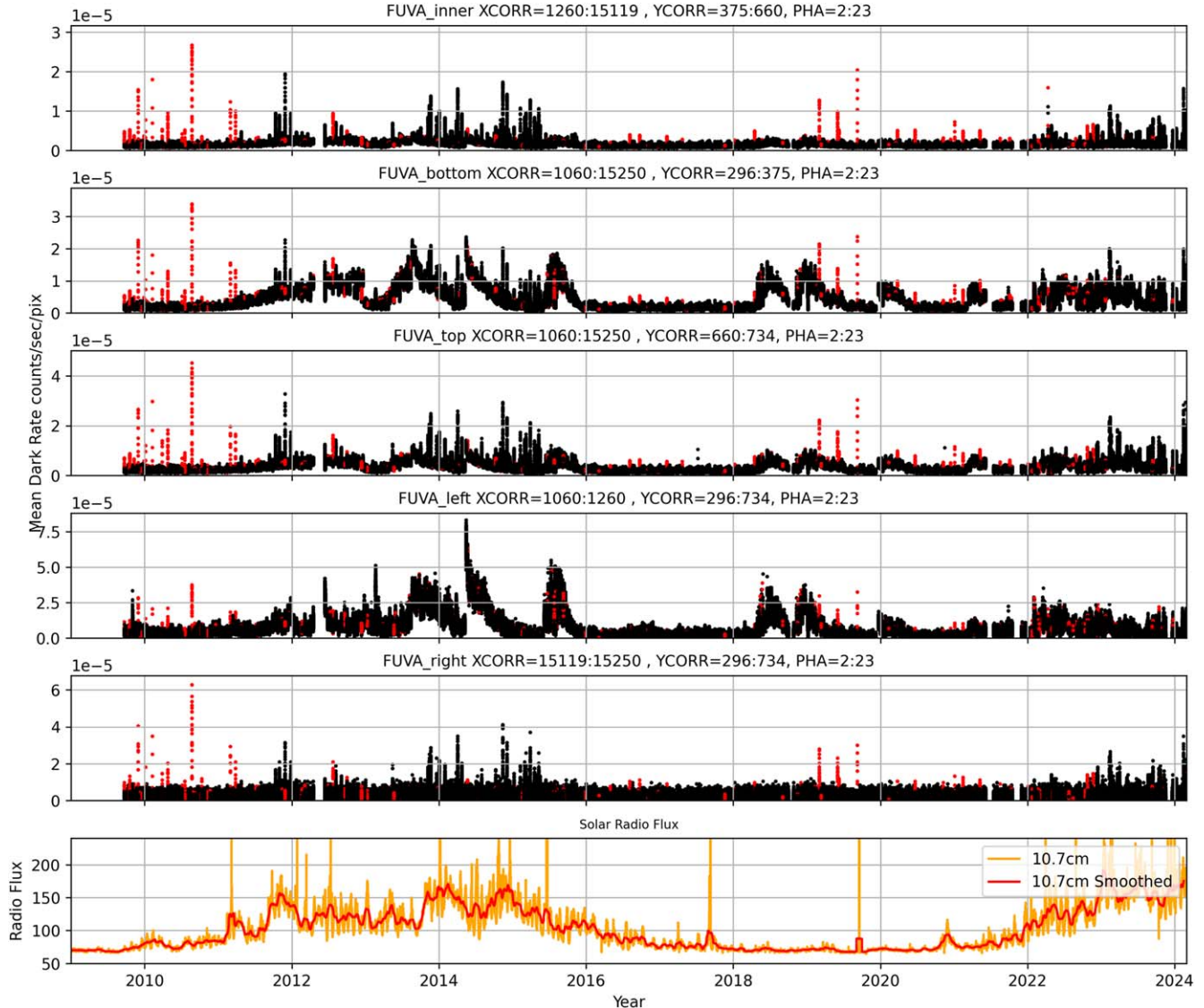


Figure 1. COS far-UV (segment A) dark rate as a function of time as measured in the COS far-UV Detector Dark Monitor (<https://www.stsci.edu/hst/instrumentation/cos/performance/monitoring>). Each panel represents a different region on the detector. The dark rate estimates are obtained from COS dark exposures stored in the MAST archive. The red points show dark rate values observed close to the South Atlantic Anomaly. The trends in the different panels highlight how the dark rate varies spatially. The bottom panel shows the solar radio emission at 10.7 cm (yellow line, in units of $10^{-22} \text{ W m}^{-2} \text{ Hz}^{-1}$) by the National Oceanic and Atmospheric Administration, NOAA (<https://www.swpc.noaa.gov/products/solar-cycle-progression>), with a smoothed version overplotted (red line).

For this approach we used the MySQL database described in Section 2, storing all dark counts extracted from the existing dark exposures observed over the years. This database includes time of arrival for each event, its corrected location on the detector (x , y), event PHA, detector segment and HV, and the name and proposal ID of the data set that recorded each of these counts. For each HV and segment, we first filtered these photons by adopting the following criteria: (1) we included only events recorded within the pre-defined inner region (see Section 2), and (2) we selected only those dark counts with

PHA values between 2 and 27 as `CalCOS` by default excludes anything outside of these values. With a final list of events per setting (i.e., HV and segment), we then sorted the counts by their time of arrival. We divided all of the selected events into suites of 100,000 counts, and identified separate data suites. The dark counts from multiple files were combined to create individual data suites. We note that using a fixed number of dark counts (100,000) per data suite guaranteed that we probed the spatial distribution of photons on the detector evenly and consistently in each data suite.

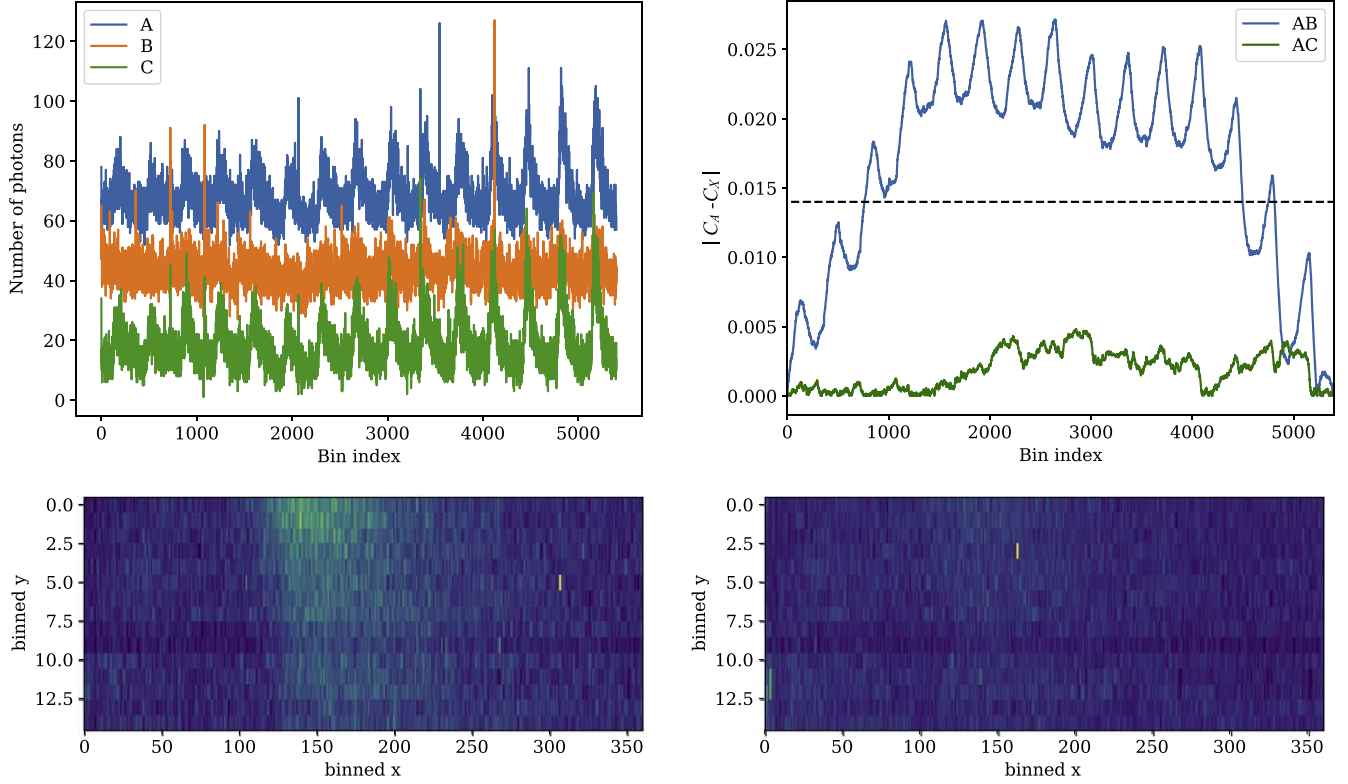


Figure 2. Example for segment A, HV = 167. Top left: reshaped data for data suites A, B, and C. For visualization effects, B and C are shifted downward by 25 and 50 respectively in the vertical direction. Top right: values of $|C_A - C_x|$ as a function of data index. In our adopted approach we have chosen a threshold of $D \geq 0.014$, shown with a dashed line. Any data suites above this threshold are considered as a new detector state. On the bottom panels we show the 2D binned data suites for two different detector states identified by our approach. The bottom left shows data suite A and the bottom right shows data suite B. We highlight the present structure (“glow”) in this particular state for segment A, which was entirely absent in any of the segment B states.

For each data suite we created 2D arrays which were binned into 360 bins in the x -direction and 15 bins in the y -direction. We reshaped these arrays into one-dimensional arrays with length $360 \cdot 15 = 5400$ (see top left panel in Figure 2). At this stage we compared different data suites to each other applying a test similar to the Kolmogorov–Smirnov (Rosenblad 2011) test. Namely, we computed the cumulative distribution of counts along the data axis (along the index ranging from 0 to 5399), and normalized them. To compare data suite A with data suite B we computed:

$$D = \max|C_A - C_B|, \quad (5)$$

where C_A and C_B are normalized cumulative distributions. We illustrate this particular step in our analysis in the top right panel in Figure 2 for segment A, HV = 167. Every time a data suite was identified as different from the previous ones based on its estimated D value, it was added to the basis vector. This identification was done by comparing each data suite against all elements of the basis vector and adding individual new data suites to the vector only if they were significantly different based on their D value (Equation (5)); for our analysis we

Table 2
Number of Detector States Identified as Part of our Analysis

Segment	HV	Number of States	Number of Exposures
A	163	3	608
A	167	10	1515
A	169	5	479
A	173	3	252
A	178	3	65
B	163	3	1215
B	167	3	589
B	169	4	1155
B	175	2	624

choose $D = 0.014$ as the threshold for selecting new detector states (shown as a dashed line in the top right panel of Figure 2). This particular threshold was chosen based on the original two states, quiescent and high-activity superdarks. Visually, we can clearly confirm that data suite A and C in the top left panel in Figure 2 are relatively similar to each other, describing a similar detector state. Once we compute D for

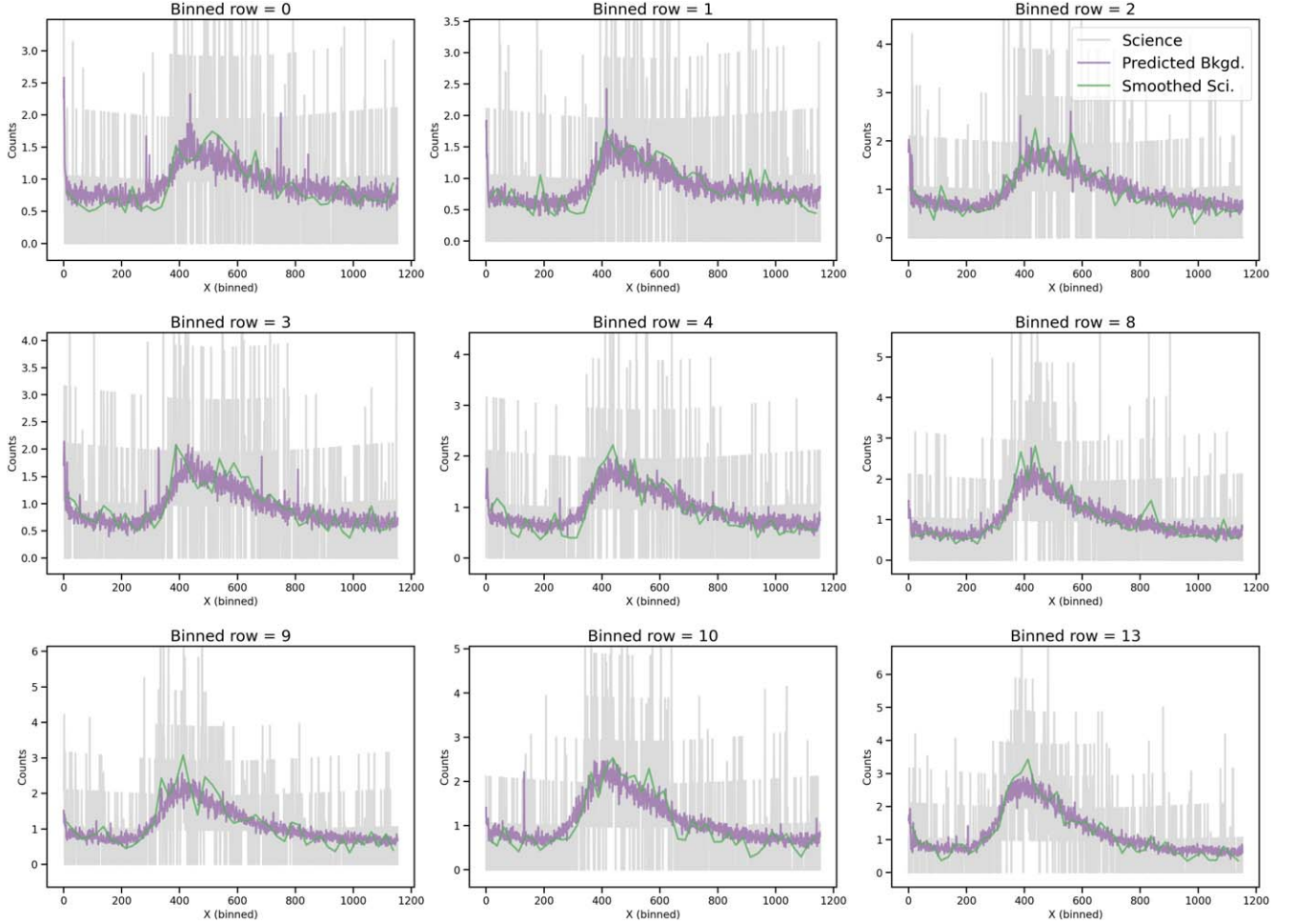


Figure 3. Diagnostic plot for one of the data sets for target SDSS-J145735.13+223201.8 (LCAG02TPQ). We display the output from the optimized background correction code (ACDC) which uses the suite of multi-state superdark. Subpanels show the dark count profiles in each binned row, excluding the science extraction and WCA regions. The dark profiles observed in the science exposures are shown in gray, the smoothed science exposures in green, and the best superdark model in purple. We highlight the agreement observed between the green and purple curves, showcasing the effectiveness of our adopted background correction approach.

these two data suites, it is clear that their D estimate is below the established threshold.

On the other hand, comparing the data suites A and B shows clear differences in the top left panel in Figure 2. The inferred D estimate for these sets clearly reaches a value higher than the established threshold of $D = 0.014$ (see the maximum value in the blue curve in the top right panel of Figure 2), indicating that the two states sampled by data suites A and B are drastically different. Overall, the bottom panels in Figure 2 show how our approach identifies detector states where the spatial distribution of counts is clearly different, in this particular case for data suites A and B.

Once we identified a final number of detector states, we then created superdark for each of these states by coadding the corresponding dark exposures as tracked in each of the data suites. To avoid having individual exposures included in multiple data suites, we treated single exposures as non-

divisible units. Lastly, we note that three separate states for segment A, at three different HVs (167, 173 and 178), were ultimately discarded from our final list of characterized states due to their extremely low number of events; even after binning both spatially and across PHAs, $>5\%$ of superpixels contained zero events. In Table 2 we show the final number of detector states identified by our approach for each detector configuration, HV and segment.

5. ACDC: Another COS Dark Correction

The optimal background correction algorithm was incorporated as the main component in our new software, ACDC. Additionally, with the adopted approach described in Section 4.1, we characterized different states of the detector for each configuration and created superdark (or empirical models) for each of them.

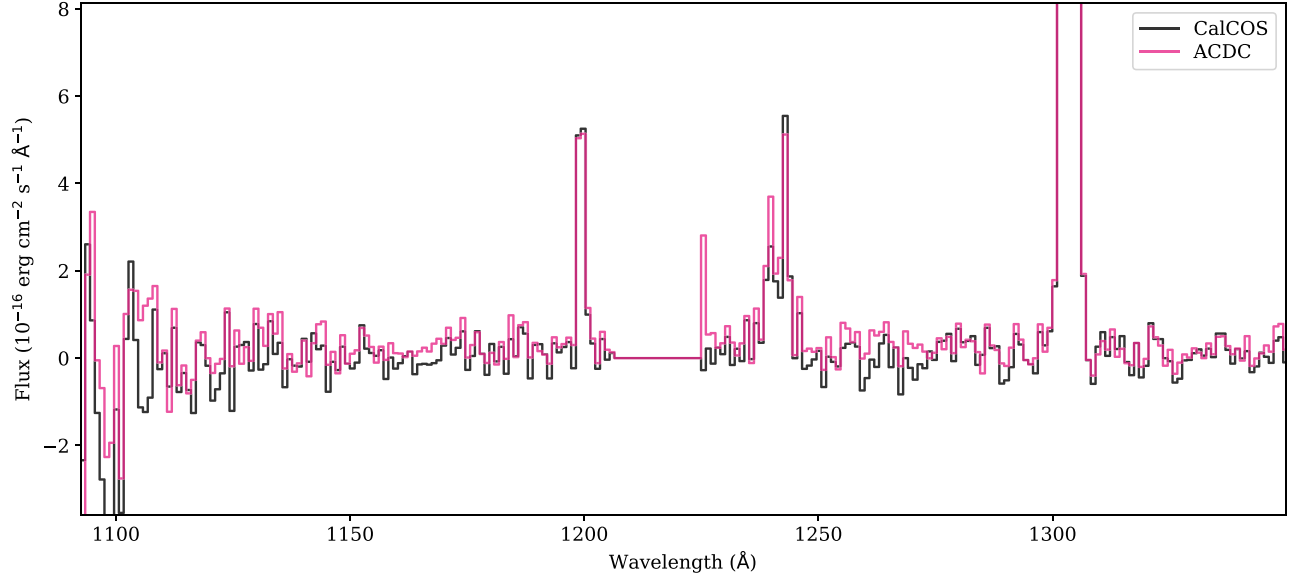


Figure 4. HCG92-7 (PID: 13321, LCAY06020) spectra, taken with the COS far-UV G130M/1222 setting, binned to 15 resolution elements for better visualization of the flux differences. The data set was taken on 2014 August 13. In black we show the standard CalCOS product; in magenta we show the ACDC product.

Table 3
Low S/N Testing Suite

Target	PID	Dataset	Date	Grating	Cenwave	Median S/N	HV _{SegA}	HV _{SegB}
KISSR1637	11522	LB6206010	10/24/10	G130M	1291	2	169	167
SDSS-J145735.13+223201.8	13293	LCAG02010	8/29/13	G160M	1577	2	167	169
IC-1586	13481	LCDR01010	11/28/13	G140L	1105	2	167	...
HCG92-7	13321	LCAY06020	8/13/14	G130M	1222	1	167	175
J1152+3400	13744	LCM802030	5/8/15	G140L	1280	1	167	163
J1333+6246	13744	LCM803030	7/6/15	G140L	1105	1	167	...
SDSS-J103020.91+611549.3	13654	LCOX08010	9/12/15	G130M	1327	2	167	163
GP1205+2620	14201	LCXR16010	2/28/16	G160M	1577	1	167	169
J1107+4528	14079	LCTD19010	01/23/17	G130M	1291	1	167	175
2MASS-J15570234-1950419	15310	LDMP24010	8/2/18	G130M	1222	1	163	167
J081112+414146	15626	LDXE11010	9/17/19	G140L	800	1	163	...
J003601+003307	15626	LDXE08010	9/25/19	G140L	800	1	163	...
J124423+021540	15626	LDXE43010	3/25/20	G140L	800	1	163	...
SDSS154714.35+175153.1	17115	LF1411010	4/29/23	G130M	1291	1	167	169
NGC-1313-P2	17180	LEVH02010	5/31/23	G130M	1291	2	167	169
BD-10-47	16701	LEOO11010	6/18/23	G130M	1222	1	173	175

ACDC was developed as a Python tool, easily installable and released for public use,⁹ along with the suite of superdarks. Given that the background correction is an intermediate step in the COS calibration pipeline, to be able to replace the nominal CalCOS background correction with the optimized background method, we took the `corrtag`¹⁰ files and processed them with ACDC. ACDC was designed to (1) determine the best

dark model as described in Section 4, (2) use this model to subtract the estimated background counts from the science image, and (3) finally run the remaining CalCOS calibration steps to produce `x1d` (extracted 1D spectra) products. We note that ACDC identifies individual best-model superdarks for every single input exposure. As a final step, users can then coadd all of the corrected `x1d` files (for a given configuration) into a single spectrum by calling the flux-weighted `coadd` function developed as part of the ULLYSES initiative (Roman-Duval et al. 2020) and installed as part of ACDC.

⁹ <https://github.com/jotaylor/acdc-hst>

¹⁰ `corrtag` files are intermediate CalCOS products. These files are binary tables containing corrected event lists.

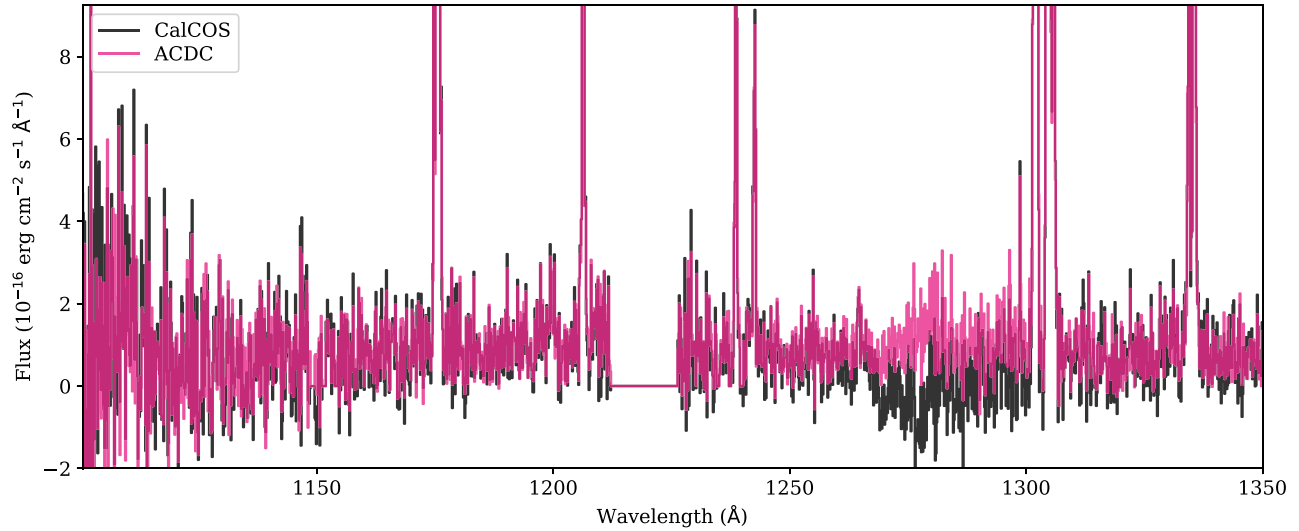


Figure 5. 2MASS-J15570234-1950419 (PID: 15310, LDMP24010) spectra, taken with the COS far-UV G130M/1222 setting, binned to 2 resolution elements (12 native pixels). The data set was taken on 2018 August 2. In black we show the standard CalCOS product; in magenta we show the ACDC product. We highlight the over-subtraction in the CalCOS spectrum around 1275 Å.

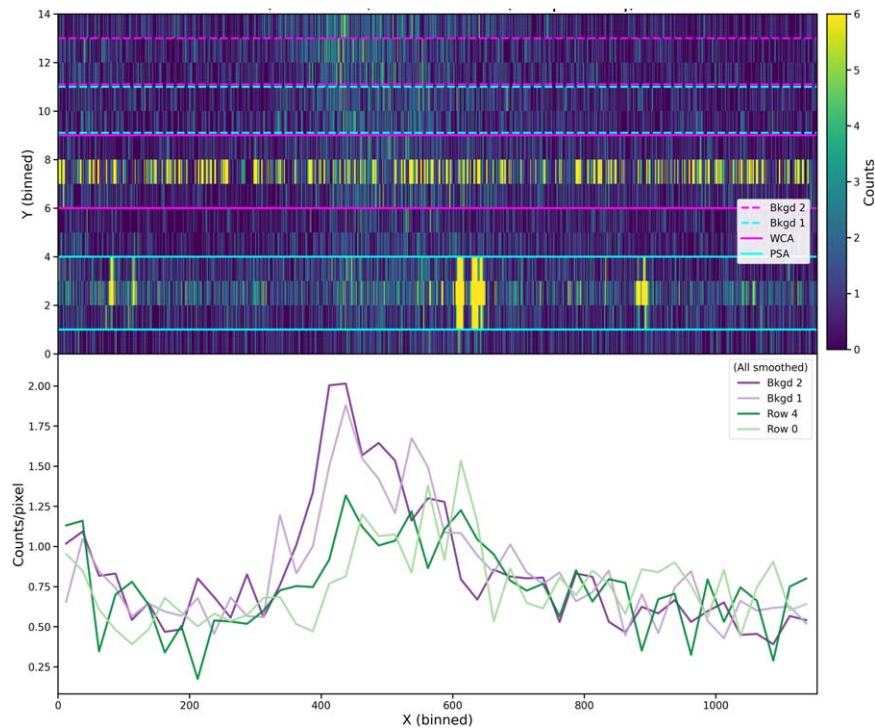


Figure 6. Top: 2D binned image of one of the science exposures (LDMP24UVQ) for target 2MASS-J15570234-1950419, taken at lifetime position 4 using segment A. The location of the science extraction region is shown with cyan solid lines. The location of the WCA is shown with magenta solid lines. The pre-defined background regions used by CalCOS to remove the background contributions to the science spectrum are shown with dashed lines, on the upper part of the detector. Bottom: Smoothed 1D profiles for the two pre-defined background regions on the upper part of the detector, compared to the profiles of the rows below and above the science extraction region (binned rows 0 and 4). Note the drastic difference between the profiles.

A few extra considerations were taken to further improve the performance of ACDC. One of these considerations included the integration of sigma-clipping when creating each superdark best model, to avoid including pixels or detector regions

affected by hot or bad pixels. Additionally, some of the resulting superdark displayed obvious gain sagged regions, particularly at the LP1 locations. To avoid obtaining a biased/incorrect superdark model for a given science exposure, ACDC

makes use of the information in the GSAGTAB¹¹ reference table appropriate for the science exposures and identifies heavily sagged regions (or the absence of them) intrinsic to the science frame, as these need to be present (or removed) in the best model superdark. Before initiating the likelihood maximization to select the best combination of superdarks for the science data, the software integrates the information on the gain sagged regions from the science exposures and when needed, interpolates over heavily sagged regions to match the expected detector state as observed in the science frames. This component of the code is particularly critical for accurately estimating the background contributions to the science spectra taken at LPI, before the gain sagged regions developed.

Given the complexity of the optimized background correction in ACDC, to facilitate a quick inspection of the accuracy of the best superdark model, the software automatically generates diagnostic figures. These figures include plots of the dark profiles in the science exposures from each individual binned row (outside of the science and WCA regions) against those from the estimated best dark model. An example for one of the data sets (LCAG02TPQ) for target SDSS-J145735.13+223201.8, specifically for segment A, is shown in Figure 3. The science extraction region for this exposure taken at lifetime position 5 is located in binned rows 6–8, and the WCA region is found in binned rows 11–12. Each subpanel shows, for each binned row: the counts in the `corrtag` file (gray), the mean binned `corrtag` counts over 25 superpixels (green), and the dark counts based on the best model superdark (purple). Overall, this figure highlights the accuracy of ACDC when creating the best superdark model, which matches the spatial dark variations observed in the 2D science exposure. Particularly notable is the presence of the “glow” feature seen only in segment A superdarks.

5.1. ACDC versus CalCOS

We tested the new software and superdarks on a suite of low S/N observations (Table 3). In general, we find overall improvements performing a 2D background correction over the standard CalCOS pipeline (Figure 4). These improvements are more accentuated in certain cases, specifically for segment A exposures that exhibit the presence of a glow-like structure (lower left panel in Figure 2). To highlight this point, we show in Figure 5 the comparison of the calibrated products using CalCOS (in black) against the data corrected using ACDC (in magenta) for target 2MASS-J15570234-19504. This particular data set, LDMP24010, was taken as part of PID 15310 using the G130M/1222 setting on 2018 August 2. We confirmed that the over-subtraction observed in the final calibrated CalCOS product originates from an overestimate

in the background contribution. The observations were taken at lifetime position 4, where the science spectrum falls at $y \sim 425$, on the lower part of the detector (binned row 2; see Figure 6). The pre-defined background regions are instead located at $y \sim 565$ and 630, on the upper part of segment A, exactly at the location of the “glow” (binned rows 9–12). The bottom panel in Figure 6 clearly highlights the differences in the 1D profiles of the background on the upper part of the detector compared to those closer to the science extraction region toward the bottom of the detector. Our improved software, along with the multi-state superdarks, more accurately account and correct for the spatial background contributions at the location of the science spectrum.

We also note that part of the testing and validation of our software involved the inspection of the final S/N values for the two different calibration methods: ACDC and the default CalCOS calibration. Using the 2D optimized background correction approach, we found that on average the S/N values in the final 1D spectra are increased by $\sim 10\%$ on both segments compared to the exposures calibrated with the nominal CalCOS pipeline. However, we highlight that in a few cases the improvements in the S/N were as high as $\sim 60\%$, both in localized wavelength regions (Figure 5), or across the full wavelength range of individual segments.

6. Conclusion

In this work we investigated methods for better characterizing the background on the COS far-UV segments in order to perform a more-accurate dark correction, which is particularly critical for observations of faint targets. We carefully analyzed all dark exposures available in the STScI MAST archive as of 2023 September, and confirmed that the distribution of dark counts on the COS detector varies spatially with time. We adopted a systematic approach that identifies between 2 and 10 different detector states per setting (segment and HV). We then created superdarks for each state and used these superdarks to perform a 2D background correction, accurately predicting and subtracting the expected dark counts in a given COS far-UV science exposure. A much more detailed technical report describing our investigation into the background properties of the instrument can be found in our official documentation page.¹²

Our optimal background correction algorithm is incorporated and released as the Python tool, ACDC,¹³ also available in Python Package Index (PyPI¹⁴). This software, along with the publicly available superdarks we have created, is able to perform a more accurate background correction than the default CalCOS calibration, specifically for low S/N data sets; our testing of ACDC showed that in several cases, the standard

¹¹ This reference file provides the locations of rectangular regions for portions of the far-UV detector that have very low pulse height amplitude, known as gain sagged regions.

¹² <https://github.com/jotaylor/acdc-hst>

¹³ <https://github.com/jotaylor/acdc-hst>

¹⁴ <https://pypi.org/project/acdc-hst/>

background correction incorrectly over-subtracted the dark counts due to the location of the static and pre-defined background regions. Additionally, we confirmed that in general ACDC increases the S/N of the final spectra by $\sim 10\%$ compared to that of the CalCOS products. We also note that implementing the approach adopted by ACDC into the standard CalCOS software would require extensive modifications to the flow of the pipeline; therefore for the time being ACDC is better used as a stand alone code. Overall, ACDC benefits COS far-UV background-limited observations irrespective of the configuration, and promises to fully exploit the scientific potential of both archival and future data of faint targets.

Acknowledgments

We thank the anonymous referee for helpful suggestions to improve the contents of this manuscript. The data analyzed in this paper were obtained from the Mikulski Archive at the Space Telescope Science Institute (MAST).

ORCID iDs

Svea Hernandez  <https://orcid.org/0000-0003-4857-8699>

Andrei Igoshev  <https://orcid.org/0000-0003-2145-1022>

Jo Taylor  <https://orcid.org/0000-0003-4068-5545>

References

- Borthakur, S., Heckman, T. M., Leitherer, C., & Overzier, R. A. 2014, *Sci*, **346**, 216
 Cash, W. 1979, *ApJ*, **228**, 939
 Dashtamirova, D., White, J., & Sahnou, D. 2019, Changes in the COS/FUV Dark Rate: Impact on the Monitoring Program and Background Extraction Regions, Instrument Science Report, **COS 2019-11**
 Ely, J., Bostroem, K. A., Debes, J. H., et al. 2014, AAS Meeting #224, **122.07**
 Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022a, *ApJS*, **260**, 1
 Flury, S. R., Jaskot, A. E., Ferguson, H. C., et al. 2022b, *ApJ*, **930**, 126
 Hernandez, S., Aloisi, A., James, B. L., et al. 2020, *ApJ*, **892**, 19
 Hernandez, S., Leitherer, C., Boquien, M., et al. 2018, *MNRAS*, **478**, 1292
 Hirschauer, A. S. 2023, COS Instrument Handbook v. 16.0, 16
 Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016a, *Natur*, **529**, 178
 Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016b, *MNRAS*, **461**, 3683
 Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, *MNRAS*, **474**, 4514
 Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018b, *MNRAS*, **478**, 4851
 Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2021, *MNRAS*, **503**, 1734
 James, B. L., Aloisi, A., Heckman, T., Sohn, S. T., & Wolfe, M. A. 2014, *ApJ*, **795**, 109
 Leitherer, C., Hernandez, S., Lee, J. C., & Oey, M. S. 2016, *ApJ*, **823**, 64
 Makan, K., Worseck, G., Davies, F. B., et al. 2021, *ApJ*, **912**, 38
 McCabe, T., Borthakur, S., Heckman, T., et al. 2021, *ApJ*, **923**, 189
 McCandliss, S. R., France, K., Osterman, S., et al. 2010, *ApJL*, **709**, L183
 Redwine, K., McCandliss, S. R., Zheng, W., et al. 2016, *PASP*, **128**, 105006
 Roman-Duval, J., Proffitt, C. R., Taylor, J. M., et al. 2020, *RNAAS*, **4**, 205
 Rosenblad, A. 2011, *JApSt*, **38**, 867
 Sahnou, D. J., Oliveira, C., Aloisi, A., et al. 2011, *Proc. SPIE*, **8145**, 81450Q
 Tumlinson, J., Thom, C., Werk, J. K., et al. 2013, *ApJ*, **777**, 59
 Worseck, G., Prochaska, J. X., Hennawi, J. F., & McQuinn, M. 2016, *ApJ*, **825**, 144
 Worseck, G., Prochaska, J. X., McQuinn, M., et al. 2011, *ApJL*, **733**, L24