Readme: Measurements of the Ultraviolet Spectral Characteristics of Low-mass Exoplanetary Systems (MUSCLES)

Survey Description

MUSCLES is a spectral survey of 11 low-mass, planet-hosting stars, 7 M and 4 K dwarfs. The spectra cover wavelengths from 5 Angstroms to 5.5 microns, with emphasis on high-energy radiation. Data sources for the various regions of the spectra are

- **X-rays**: *Chandra/XMM-Newton* and APEC models (Smith+ 2001; ApJ 556:L91)
- **FUV - blue visible**: *HST*
- **Lyα**: Reconstructed spectrum from model fits to Lyα line profiles accounting for the stellar flux and the interstellar medium (Youngblood+ 2016; in press)
- **Visible - IR**: synthetic photospheric spectra from PHOENIX atmosphere models (Husser+ 2013; A&A 553:A6)

Proxima Centauri

In response to popular demand after the announcement of the discovery of a habitable-zone planet orbiting Prox Cen, we added Prox Cen to the MUSCLES sample and created an SED from archival data using methodology consistent with the other MUSCLES SEDs. We highly encourage users of the Prox Cen spectra to read these notes on the reduction of the Prox Cen spectra, as they address an important issue regarding the stellar effective temperature and bolometric flux.

References

The following papers provide details on various aspects of the MUSCLES treasury dataset and its creation. When using the MUSCLES spectra, please cite all or a selection of these works as appropriate to your application, along with the version number of the spectra used.

- **France+ 2016**: ApJ 820:89F

Contact

Any issues/questions with the data product itself should be addressed to Parke Loyd, robert.loyd@colorado.edu.

Issues/questions about downloading the data product from the MAST database should be addressed to STScI, archive@stsci.edu.
Version History

Version 0.0 (v00) 2015/11/13

- Initial internal release

Version 0.1 (v01) 2015/01/04

- corrected a mistake in the normalization of Gj 176 STIS G230L data (~10% change)
- Lya profile for GJ 1214 revised to match a new Lya - Mg II empirical relationship (see Youngblood+ 2016 for details)
- improved artifact correction in the X-ray spectra
- added metadata to headers of CXO component spectra

Version 0.2 (v02) 2015/01/07

- corrected a gap at ~50-100 Angstroms between the APEC model spectrum and EUV estimates accidentally introduced in version 0.1

Version 1.0 (v10) 2015/01/28

- official public release
- fixed error in the comment in header of HD 97658 X-ray component spectrum that said the spectrum was scaled from HD 85512 according to the two stars’ distances when really it was their bolometric flux
- adaptive resolution spectra added 2015/04

Version 2.0 (v20) 2016/10/28

- Revised noise estimation for photometry. This did not directly affect the normalization values computed, but indirectly did so by potentially affecting which measurements were identified as outliers.
- Improved the filtering of photometry gathered by the Vizier photometry viewer. The photometry table this service returns includes duplicate entries, and the previous pipeline version was overly-conservative in rejecting potential duplicates. The new version does a better job of identifying and retaining unique measurements. This and the above change resulted in changes to the normalization of the PHOENIX and G430L spectra (and thus of the bolometric flux) of <1% in all cases except Gj667C, for which the normalization decreased by 10.1%.
- Corrected error in the propagation of exposure times through the coaddition of Echelle spectrum orders.
- Revised the coaddition of Echelle spectra from coadding all orders from all exposures using 1/error^2 weighting to coadding orders from the same exposure with 1/error^2 weighting then coadding separate exposures using exposure time weighting.
- Moved to double precision for all FITS columns.
- Corrected for the accidental use of an earlier draft of the EUV reconstruction in the SEDs, changing EUV levels in the broadband EUV bins by 0.1-3% (except for Gj1214, where the change
was 13-20%).

- Revised optimal splicing algorithm to account for serious data quality flags when evaluating which data to choose within spectral overlaps when assembling stitched SEDs.
- Added Proxima Centauri to the library of spectra.

Version 2.1 (v21) 2017/01/25

- Changed the method for splicing the reconstructed Lya flux into the data. The previous splice range was overly extensive (10 Å), eliminating the FUV continuum and broad extended wings of the Lya line (see Fig 7 of Youngblood+ 2016). The new method retains the COS data up to ±400 km/s of the Lya rest wavelength. A fit to the Lorentzian wings of the Lya profile, detected in the COS data but below the noise level of the STIS data used for the Lya reconstruction, is used to fill the gap between ±400 km/s and the region where the reconstructed Lya profile begins to dominate. The Lya reconstructions do not themselves include Lorentzian wings because the STIS data could not constrain the amplitude of these wings. This change results in increases of order 1% to the integrated Lya flux, on par with expectations based on the Youngblood+ 2016 work. The fit parameters are saved in header keywords (described later in this readme).
- Fixed bug wherein the 'BOLOFLUX' keyword was not being saved in the SED headers.

Version 2.2 (v22) 2018/01/11

- Tweaked the "adaptive binning" spectra: (1) Major emission lines are now excluded from adaptive rebinning and (2) bins that have grown in width to larger than 1.0 AA are resampled to 1.0 AA bins.

File Naming Scheme

```
hlsp_muscles_TELESCOPE_INSTRUMENT_TARGET_GRATING-or-FILTER_VERSION_PRODUCT.fits
```

Values of na indicate "not applicable" and multi indicates "multiple." PRODUCT can be one of -
var-res-sed: panchromatic SED that retains the native instrument resolutions where possible -
const-res_sed: panchromatic SED binned to a constant 1 Angstrom resolution - component-spec:
one of the individual spectra used to construct the composite SEDs, including cropped wavelengths. The HST STIS E140M/G140M spectra used for Lya reconstruction are also provided -
adapt-var-res-sed: panchromatic SED that retains the native instrument resolutions where possible, downsamples in low signal-to-noise regions to avoid negative fluxes

Data Product Description

All the spectra have primary extension with a header that gives information on the observation(s) that make up the spectrum. The spectrum itself is in the first extension, named 'SPECTRUM'. For all header values, 'MULTI' indicates multiple values for the field, which will be followed by a numbered list of the various values (e.g. a MULTI value for INSTRUME will be followed by INSTRU00, INSTRU01, ... to specify the various instruments that contributed data to the spectrum).
Primary Header Keywords

Those keywords that are not self-explanatory are described below. Some keywords are omitted for data products to which they do not apply or are not well-defined (e.g. EXPTIME in the panchromatic SEDs, DATEOBS for model spectra).

- **TARGNAME**: name of the target star
- **RA_TARG**: right ascension coordinate of the target (decimal degrees)
- **DEC_TARG**: declination coordinate of the target (decimal degrees)
- **PROPOSID**: HST proposal number for MUSCLES (always 13650)
- **HLSPNAME**: name of this HLSP (always 'Measurements of the Ultraviolet Spectral Characteristics of Low-mass Exoplanet Host Stars')
- **HLSPACRN**: acronym for the HLSP (always 'MUSCLES')
- **HLSPLEAD**: lead preparer of the data product (always R. O. Parke Loyd)
- **PR_INV_L**: principle investigator last name (always France)
- **PR_INV_F**: principle investigator first name (always Kevin)
- **DATE-OBS**: date of the start of the first observation in YYYY-MM-DDTHH:MM:SS.SSS
- **EXPSTART**: modified julian date of the start of the first observation
- **EXPEND**: modified julian date of the end of the last observation
- **EXPTIME**: characteristic exposure time in s, determined using the method specified in EXPDEFN
- **EXPDEFN**: method used to determine EXPTIME value
- **EXPMAX**: maximum exposure time for a spectral element in s
- **EXPMIN**: minimum nonzero exposure time for a spectral element in s
- **EXPMED**: median exposure time in s
- **NORMFAC**: normalization factor applied to the spectrum before merging into the composite spectrum and/or before use for the Lya reconstruction (see Loyd+ 2106; ApJ in press)
- **WAVEMIN**: minimum wavelength in the spectrum in Angstroms
- **WAVEMAX**: maximum wavelength in the spectrum in Angstroms
- **AIRORVAC**: wavelengths specified in their vacuum or air values (always 'vac')
- **SPECRES**: wavelength at which WAVERES us specified in Angstroms
- **WAVERES**: wavelength resolution at SPECRES in Angstroms
- **FLUXMIN**: minimum flux in the spectrum in erg s-1 cm-2 Angstroms-1
- **FLUXMAX**: maximum flux in the spectrum in erg s-1 cm-2 Angstroms-1
- **BOLOFLUX**: bolometric flux measurement (integral of flux >5 Angstroms) for the star, including a blackbody fit for flux >5.5 microns that contributes a few percent to the integral in most cases. Because of the included tail, this is NOT the same value as one will find by integrating the full extend of the SED.

Spectrum Header Keywords

The spectrum extension of the FITS files contains just the default required keywords for the FITS standard, with the addition of a TDESC keyword associated with each data column (along with the standard TTYPE, TFORM, and TUNIT keywords) that describes the data in that column.
Spectrum Data Columns

The spectrum extension of the FITS files is a data table that contains these columns (according to the type of spectrum):

All spectra:

- **WAVELENGTH** : midpoint of the wavelength bin in Angstroms
- **WAVELENGTH0**: left (blue) edge of the wavelength bin in Angstroms
- **WAVELENGTH1**: right (red) edge of the wavelength bin in Angstroms
- **FLUX** : average flux density in the wavelength bin in erg s-1 cm-2 Angstroms-1

All except model spectra:

- **DQ** : data quality flags propagated from the original observation (defined by the instrument, NOT defined by MUSCLES -- see docs for that instrument for DQ flag definitions)
- **ERROR** : error on the flux density in erg s-1 cm-2 Angstroms-1
- **EXPEND** : modified julian date of the end of the last exposure contributing data to the bin
- **EXPSTART** : modified julian date of the start of the first exposure contributing data to the bin
- **EXPTIME** : total exposure time of observations contributing to that bin, averaged according to bin widths when two or more elements have been rebinned to a coarser resolution

SEDs:

- **BOLOFLUX** : flux density normalized by the bolometric integral of the flux given in the primary header as BOLOFLUX, in Angstroms-1
- **BOLOERR** : error on BOLOFLUX
- **INSTRUMENT** : binary flag(s) identifying the source spectrum or spectra for the data in the bin. These values are defined in the second FITS extension, INSTLGNd. When rebinning, some bins acquire data from multiple sources. In these cases, the binary flags are combined in a bitwise OR sense (e.g. 010 (2) and 100 (4) becomes 110 (6)).
- **NORMFAC** : normalization factor applied to the source spectrum when merging into the composite SED
- **LNZ_GAM, LNZ_NORM** : the width and scale factor the best fit to the Lorentzian wings of the Lya profile as measured in the COS data (see the notes for Version 2.1 earlier in this readme). The equation of the fit is then \( \text{flux\_density} = \text{lnz\_norm} \times \frac{\text{lnz\_gam}/2}{(w - w0) + (\text{lnz\_gam}/2)^2} \) where \( w \) is wavelength and \( w \) is the center of the reconstructed Lya profile.

Component Spectrum SRCSPCECS Extension

The second extension of the component spectrum source files, named SRCSPCECS, gives identifiers that can be used to locate the observatory-level data used to create the spectrum. For **HST** observations, both DATASET_ID values and ROOTNAME are given, since sometimes multiple files with different ROOTNAMEs are associated with a single observation.
**SED INSTLGND Extension**

The SED FITS files have a second extension named INSTLGND that provides the information needed to associate component spectra with the binary keys given in the INSTRUMENT column of the SPECTRUM extension. The included data columns are:

- **BITVALUE**: value corresponding to the INSTRUMENT column of the SPECTRUM extension
- **GRATING**: grating used for observation
- **HLSP_FILE**: name of the corresponding HLSP file
- **INSTRUMENT**: instrument used for observation
- **TELESCOPE**: telescope used for observation

**Negative Flux Bins**

Data in regions of low flux contain some bins with negative flux values, as expected due to statistical noise in the signal and subtracted background levels. The physical impossibility of negative flux poses a problem for some applications of these spectra. *We caution the user against simply setting these bins to zero flux. This introduces an upward bias in the flux over the broadband FUV continuum of several percent to over 10% in the worst case.*

We have created "adaptive resolution" SEDs as a solution that relies only on the data (as opposed to further models or interpolations) for applications where negative-flux bins are undesirable. For these spectra, we iteratively averaged negative-flux bins with their neighbors until no negative-flux bins remain. This results in spectra that are effectively downsampled in low-flux regions to improve S/N to a point where the flux over the larger wavelength range can be measured. As with the unaltered SEDs, we provide both var-res (variable resolution) and const-res (constant 1 Angstrom resolution) versions of these spectra. The var-res spectra retain the (often very) coarse resolution in the low flux regions of the SED. In contrast, the 1 Angstrom binning of the const-res SEDs implies that areas of the spectrum that have been downsampled to a resolution coarser than 1 Angstrom will then have been upsampled. We caution the user that these regions will appear to have a fidelity higher than their true fidelity.

Users can also implement their own solutions to the problem of negative-flux bins. An (incomplete) list of solutions that users could implement include replacing the low-flux regions of the SEDs with:

- PHOENIX model output (included in this HLSP). This could serve as a lower limit (i.e. photosphere-only) flux estimate in these regions.
- Data from one of the brighter targets (appropriately renormalized).
- A custom model of stellar upper-atmosphere emission, such as that of Fontenla+ 2016 (ApJ submitted) in the case of Gj 832.

**Other Caveats and Special Cases**

Users of this dataset should be aware of the following issues, described in greater detail in Section 2.6 of Loyd+ 2016 (ApJ in press).
• All: The PHOENIX spectra obtained from Husser+ 2013 grid truncate at 5.5 microns. To compute a more accurate bolometric flux, we fit a blackbody spectrum to the data near this edge when computing the full integral of the spectrum. This value is included in the SED header and is used when computing the boloflux (flux normalized by the bolometric integral) values.

• All: *HST* STIS data showed systematically lower absolute fluxes and were normalized upward to match corresponding *HST* COS data where statistically justifiable. Note that the Eps Eri *HST* STIS E230M and E230H data are suspect, but no *HST* COS data were available to normalize against.

• All (esp. GJ 832, GJ 876): Flares were observed in some of the data. Data from the flaring times was *not* discarded. Particularly large flares that occurred during the GJ 832 and GJ 876 *HST* COS G130M observations.

• All but Eps Eri: A miscalibration in the wavelength solution of the *HST* COS G230L data by as much as 4 Angstroms (~600 km s⁻¹) appears to be present near the short wavelength edge (~1800 Angstroms).

• GJ 1214, GJ 876, GJ 436, GJ 581: Wavelengths of the *HST* STIS E230H and G230L data appear miscalibrated at the Mg II 2796,2803 lines by 1-2 Angstroms (165-330 km s⁻¹).

• GJ 1214, GJ 667C: STIS G430L data blueward of ~3850 Angstroms was culled in preference for PHOENIX model because of low S/N. (In all other cases, data was always given preference over model.)

• GJ 667C: slit aperture was significantly off-center in *HST* STIS observations. Normalization factors to match STIS data to COS data are consequently very high.

• GJ 1214, GJ 832, GJ 581, GJ 436: *HST* pipeline extraction of the STIS G140M data failed to locate the spectrum on the detector. The spectrum was “manually” extracted from the two dimensional data (x2d). This was also necessary for the STIS G230L data for GJ 1214.

• GJ 1214: *Chandra* X-ray data provided only upper limits. We used a model fit to archival *XMM-Newton* data instead.

• HD 97658: No X-ray data were collected. We used the HD 85512 data scaled according to the ratio of bolometric flux between the two stars because of their similar levels of Fe XII emission relative to the bolometric flux.

• GJ 581, GJ 876: Two markedly different levels of X-ray activity were observed. We only use data from the more quiescent levels.

• GJ 436: Evidence of a faint second source is present in the *HST* STIS G230L data and may contribute flux to all observations.