PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

The Son-Of-X-Shooter (SOXS) datareduction pipeline

David R. Young, Marco Landoni, Stephen J. Smartt, Sergio Campana, Paolo D'Avanzo, et al.

David R. Young, Marco Landoni, Stephen J. Smartt, Sergio Campana, Paolo D'Avanzo, Riccardo Claudi, Pietro Schipani, Matteo Aliverti, Andrea Baruffolo, Sagi Ben-Ami, Giulio Capasso, Rosario Cosentino, Francesco D'Alessio, Ofir Hershko, Hanindyo Kuncarayakti, Matteo Munari, Giuliano Pignata, Kalyan Radhakrishnan, Adam Rubin, Salvatore Scuderi, Fabrizio Vitali, Jani Achrén, José Antonio Araiza-Durán, Iair Arcavi, Federico Battaini, Anna Brucalassi, Rachel Bruch, Enrico Cappellaro, Mirko Colapietro, Massimo Della Valle, Rosario Di Benedetto, Sergio D'Orsi, Avishay Gal-Yam, Matteo Genoni, Marcos Hernandez, Jari Kotilainen, Gianluca Li Causi, Laurent Marty, Seppo Mattila, Michael Rappaport, Davide Ricci, Marco Riva, Bernardo Salasnich, Ricardo Zanmar Sanchez, Maximilian Stritzinger, Hector Pérez Ventura, "The Son-Of-X-Shooter (SOXS) data-reduction pipeline," Proc. SPIE 12189, Software and Cyberinfrastructure for Astronomy VII, 121891I (29 August 2022); doi: 10.1117/12.2629282



Event: SPIE Astronomical Telescopes + Instrumentation, 2022, Montréal, Québec, Canada

The Son-Of-X-shooter (SOXS) Data-Reduction Pipeline

- David R. Young^a, Marco Landoni^b, Stephen J. Smartt^a, Sergio Campana^b, Paolo D'Avanzo^b, Riccardo Claudi^s, Pietro Schipani^d, Matteo Aliverti^b, Andrea Baruffolo^s, Sagi Ben-Ami^{h,e}, Giulio Capasso^d, Rosario Cosentino^{f,k}, Francesco D'Alessio^g, Ofir Hershko^h, Hanindyo Kuncarayakti^{j,q}, Matteo Munari^k, Giuliano Pignata^{m,t}, Kalyan Radhakrishnan^s, Adam Rubin^c, Salvatore Scuderi^{v,k}, Fabrizio Vitali^g, Jani Achrén^l, José Antonio Araiza-Duran^w, Iair Arcaviⁿ, Federico Battaini^s, Anna Brucalassi^w, Rachel Bruch^h, Enrico Cappellaro^s, Mirko Colapietro^d, Massimo Della Valle^d, Rosario Di Benedetto^k, Sergio D'Orsi^d, Avishay Gal-Yam^h, Matteo Genoni^b, Marcos Hernandez^f, Jari Kotilainen^{j,q}, Gianluca Li Causi^r, Laurent Marty^d, Seppo Mattila^q, Michael Rappaport^h, Davide Ricci^s, Marco Riva^b, Bernardo Salasnich^s, Ricardo Zanmar Sanchez^k, Maximilian Stritzinger^u, and Hector Ventura^f
 - ^aAstrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK
- bINAF Osservatorio Astronomico di Brera, Via Bianchi 46, I-23807, Merate, Italy
 cESO, Karl Schwarzschild Strasse 2, D-85748, Garching bei München, Germany
 dINAF Osservatorio Astronomico di Capodimonte, Sal. Moiariello 16, I-80131, Naples, Italy
 eHarvard-Smithsonian Center for Astrophysics, Cambridge, USA
 fFGG-INAF, TNG, Rambla J.A. Fernández Pérez 7, E-38712 Breña Baja (TF), Spain
 gINAF Osservatorio Astronomico di Roma, Via Frascati 33, I-00078 M. Porzio Catone, Italy
- ^hWeizmann Institute of Science, Herzl St 234, Rehovot, 7610001, Israel ⁱMax-Planck-Institut für Extraterrestrische Physik, Giessenbachstr. 1, D-85748 Garching, Germany
 - ^jFinnish Centre for Astronomy with ESO (FINCA), FI-20014 University of Turku, Finland ^kINAF Osservatorio Astrofisico di Catania, Via S. Sofia 78 30, I-95123 Catania, Italy ^lIncident Angle Oy, Capsiankatu 4 A 29, FI-20320 Turku, Finland ^mUniversidad Andres Bello, Avda. Republica 252, Santiago, Chile ⁿTel Aviv University, Department of Astrophysics, 69978 Tel Aviv, Israel
 - °Dark Cosmology Centre, Juliane Maries Vej 30, DK-2100 Copenhagen, Denmark PAboa Space Research Oy, Tierankatu 4B, FI-20520 Turku, Finland
- ^qTuorla Observatory, Dept. of Physics and Astronomy, FI-20014 University of Turku, Finland ^rINAF Istituto di Astrofisica e Planetologia Spaziali, Rome, Italy
- ^sINAF Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122, Padua, Italy ^tMillennium Institute of Astrophysics (MAS)
- ^uAarhus University, Ny Munkegade 120, D-8000 Aarhus, Denmark ^vINAF - Istituto di Astrofisica Spaziale e Fisica Cosmica, Via Corti 12, I-20133, Milano, Italy ^wINAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125, Firenze, Italy

Further author information: (Send correspondence to David Young)

David Young: E-mail: d.r.young@qub.ac.uk Marco Landoni: E-mail: marco.landoni@inaf.it Stephen Smartt: E-mail: s.smartt@qub.ac.uk

Software and Cyberinfrastructure for Astronomy VII, edited by Jorge Ibsen, Gianluca Chiozzi, Proc. of SPIE Vol. 12189, 121891I · ⊚ 2022 SPIE 0277-786X · doi: 10.1117/12.2629282

ABSTRACT

The Son-Of-XShooter (SOXS) is a single object spectrograph (UV-VIS & NIR) and acquisition camera scheduled to be mounted on the European Southern Observatory (ESO) 3.58-m New Technology Telescope at the La Silla Observatory. Although the underlying data reduction processes to convert raw detector data to fully-reduced science ready data are complex and multi-stepped, we have designed the SOXS Data Reduction pipeline with the core aims of providing end-users with a simple-to-use, well-documented command-line interface while also allowing the pipeline to be run in a fully automated state; streaming reduced data into the ESO Science Archive Facility (SAF) without need for human intervention. To keep up with the stream of data coming from the instrument, there is the requirement to optimise the software to reduce each observation block of data well within the typical observation exposure time. The pipeline is written in Python 3 and has been built with an agile development philosophy that includes CI and adaptive planning.

Keywords: SOXS, Pipeline, Data Reduction, Spectroscopy, Imaging

1. INTRODUCTION

The SOXS (Son Of X-Shooter) instrument is a new medium resolution spectrograph ($R \simeq 4500$) capable of simultaneously observing 350-2000nm (U- to H-band) to a limiting magnitude of R ~ 20 (3600sec, S/N ~ 10). It shall be hosted at the Nasmyth focus of the New Technology Telescope (NTT) at La Silla Observatory, Chile (see¹ for an overview). This paper describes the design of the SOXS data-reduction pipeline and data-flow system. Details of each of the other SOXS subsystems can be found in a set of related.^{1–28}

Details of the three detectors included in the SOXS instrument, that the pipeline shall receive data from, are given in Section 2. Section 3 explains the aims and goals of the pipeline and Section 4 describes the pipeline software architecture and development environment. Finally, Section 5 outlines the data-products to be expected from the pipeline and the planned data-flow from raw data coming off the telescope through to the data owners collecting reduced data from the ESO SAF.

2. THE 3 SOXS DETECTORS

SOXS comprises three instruments; the UV-VIS and NIR spectrographs and an Acquisition and Imaging Camera (AC). The instruments are to be mounted on the NTT's Nasmyth focus rotator flange. It is the role of the SOXS data reduction pipeline to reduce the pixel data collected by each of these instruments into science-ready data products.

2.1 The NIR Spectrograph

The SOXS NIR spectrograph is a cross-dispersed echelle, employing "4C' (Collimator Correction of Camera Chromatism) to image spectra in 800-2000nm wavelength range, in 15 orders, onto a 2kx2k 18 micron pixel Teledyne H2RG TM array (see Figure 1). It will achieve a spectral resolution R $\simeq 5000$ (1 arcsec slit).

2.2 The UV-VIS Spectrograph

The UV-VIS spectrograph employs the novel design of 4 ion-etched transmission gratings in the first order (m=1) to obtain spectra in the 350-850nm wavelength range (providing an overlap of 50nm with the NIR arm for cross-calibration). The spectral band is split into four poly-chromatic channels and sent to their specific grating.¹⁹ Unlike the NIR arm, the UV-VIS arm will include an Atmospheric Dispersion Corrector (ADC). Each of the four dispersion orders are imaged to separate areas of the e2V CCD and aligned linearly along the direction of the CCD columns (see Figure 2).

2.3 The Acquisition and Imaging Camera

Although the primary use of the SOXS acquisition camera is to acquire spectral targets to allow for their centring on the slit, the camera's $3.5' \times 3.5'$ FOV and 0.205 arcsec/px scale will also allow for science-grade, multi-band imaging. Observers will be able to select from 7 filters; the LSST u, g, r, i, z, y set and Johnson V.

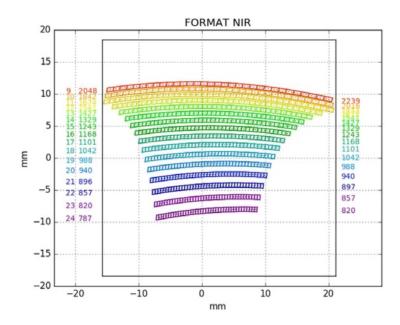


Figure 1: The SOXS NIR spectral format, a reproduction of Figure 5 of Vitali et al. (2018).²⁴

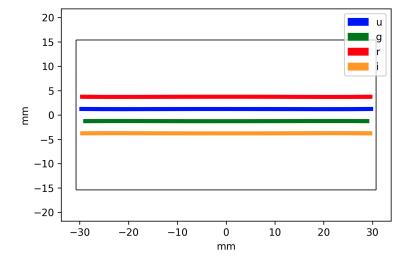


Figure 2: The four quasi-orders of the SOXS UV-VIS spectral format aligned linearly along the direction of the detector columns.

3. DATA REDUCTION PIPELINE REMIT

The main purpose of the SOXS Data Reduction pipeline is to use SOXS calibration data (typically, but not necessarily, collected close in time to the science data), to remove all instrument signatures from the SOXS scientific data frames, convert this data into physical units and deliver them with their associated error bars to the ESO SAF as Phase 3 compliant science data products, all within a timescale shorter than a typical SOXS science exposure. The pipeline must also support the reduction of data taken in each of the available SOXS observation modes. The primary reduced pipeline product will be a detrended, wavelength and flux calibrated, telluric corrected 1D spectrum with UV-VIS + NIR arms stitched together (see Section 5).

Although the underlying data reduction processes to convert the raw detector data to fully-reduced, fluxand wavelength-calibrated science-ready data are complex and multi-stepped, soxspipe has been designed with

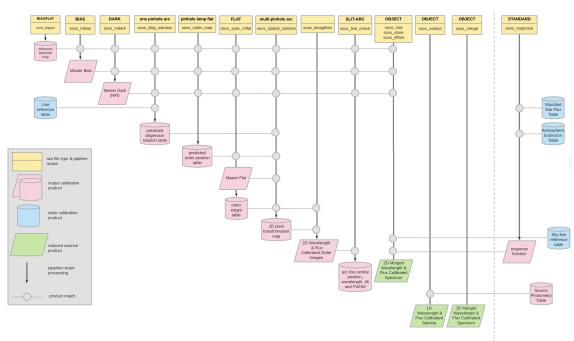


Figure 3: The SOXS Spectroscopic Data Reduction Cascade. Each of the vertical lines in the map depicts a raw data frame, the specific recipe to be applied to that frame and the data product(s) output by that recipe. Horizontal lines show how those output data products are used by subsequent pipeline recipes. Time loosely proceeds from left to right (recipe order) and from top-to-bottom (recipe processing steps) on the map.

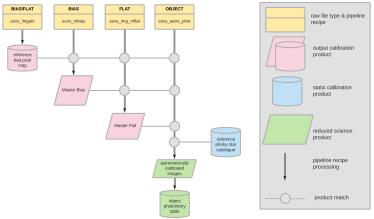


Figure 4: SOXS Imaging Data Reduction Cascade. See caption of Figure 3 for details.

a core aim of providing end-users with an easy-to-install, simple-to-use, clear, well-documented command-line interface while also allowing the pipeline to be run in a fully automated state; streaming reduced SOXS data into the ESO SAF without need for human intervention. Once users have miniconda* or anaconda† installed on their local machine, the pipeline can be installed via a single command and typically takes < 1 min to install.

conda create -n soxspipe python=3.8 soxspipe -c conda-forge

The static calibration files required by the pipeline are shipped alongside the code, removing the burden often

 $[*]miniconda\ https://docs.conda.io/en/latest/miniconda.html\\$

[†]anaconda https://www.anaconda.com



Figure 5: The pipeline development environment. The code is hosted and version controlled on Github, collaborative programming and prototyping is performed within Jupyter notebooks, Jenkins is used for Continuous Integration, code is distributed via PyPi and Conda and documentation is written host on readthedocs.

required of pipeline users to separately download and manage these files. This has the added benefit of these files being version controlled alongside the code so the end-user will always have access to the suite of calibration files associated with the specific version of the pipeline they have installed on their machine.

The pipeline will also generate Quality Control (QC) metrics to monitor telescope, instrument and detector health. These metrics are to be read and presented by the SOXS health-monitoring system.²⁷

4. PIPELINE ARCHITECTURE AND DEVELOPMENT ENVIRONMENT

Presently, the astronomical community have overwhelmingly adopted Python as their scripting language of choice and there are a plethora of well-maintained, mature python packages to help with basic data-reduction routines, visualisation, user-interaction and data manipulation. It was a natural choice therefore to develop the SOXS pipeline in Python 3. We have implemented an object-orientated composition and the pipeline is designed to be primarily driven from the command line. The concept of 'recipes', originally employed by ESO's Common Pipeline Library (CPL), has been adopted to define the modular components of the data reduction workflow. These recipes can be connected together to create an end-to-end data-reduction cascade, taking as input raw and calibration frames from the SOXS instrument and processing them all the way through to fully reduced, calibrated, ESO Phase 3 compliant science products (see Figures 3 and 4). Recipes are named with the prefix 'soxs' followed by a succinct description of the recipe (e.g. soxs_mbais for the master bias creation recipe). There are also many reusable functions designed to be called from multiple recipes; these are referred to as 'utilities' in soxspipe.

The pipeline has been built with an agile development philosophy that includes adaptive planning and evolutionary development. As with any software project, one of the greatest risks is knowledge loss due to a team member leaving before project completion. To mitigate this risk we have employed pair-programming techniques to share knowledge, both explicit and tacit, between two developers. In times of travel bans and remote working a JupterHub server with Python-based notebooks, shared screens and video conferencing tools have been essential to executing these techniques.

The SOXS End-to-End (E2E) simulator¹⁴ is capable of producing simulated 2D images in the SOXS format that take into account the main optical behaviour of the system (grating dispersion, sampling, PSF, noises and position of various resolution elements coming from full ray-tracing). By using test-driven development throughout the development process, combined with 'extreme' mock data generated from the E2E simulator, we can verify the pipeline is not only able to reduce a typical data set but also data that is far from ideal. This extreme data helps us push the pipeline to the limits of its capabilities and allows us to defensively develop against the edge-case scenarios the pipeline will most certainly experience at some point in production mode. Figure 6 gives an indication of the quality of the reductions achieved by soxspipe when reducing E2E calibration frames.

Table 1: Final data-products generated by the SOXS pipeline.

Product	Description
1D Source Spectra	1D spectra in FITS binary table format, one for each arm.
	Each FITS spectrum file will contain 4 extensions:
	1. Wavelength- and flux-calibrated spectra with absolute flux correction
	via scaling to acquisition image source photometry,
	2. an additional spectrum with correction for telluric absorption via
	MOLECFIT,
	3. the variance array and
	4. the sky-background spectra.
1D Merged Source Spectrum	1D UV-VIS & NIR merged spectrum in FITS binary table
	format with PDF visualisation. This spectrum
	will be rebinned to a common pixel scale for each arm.
	This spectrum file will also have the same 4
	extensions described above.
2D Source Spectra	A 2D FITS image for each spectral arm containing wavelength
	and flux calibrated spectra (no other corrections applied) allowing
	users to perform source extraction with their tool of choice. Note that
	rectification of the curved orders in the NIR introduces a source
	of correlated noise not present in extractions performed on the
	unstraightened orders as done by the pipeline.
Acquisition Camera Images	ugrizy astrometrically and photometrically (griz only) calibrated to
	Refcat2 ²⁹

The pipeline code is open-source, hosted on Github[‡] and connected to a Jenkins Continuous Integration/Continuous Deployment server via Github's webhooks. Any new push of code to a branch on the GitHub repository triggers a new 'build' of the code on the CI server where all unit tests are run. If all tests pass the branch can be merged into main development branch. If it is the main/production branch being tested, and all tests pass, then a new dot release version of the code is automatically shipped to PyPI[§] and conda-forge ready for deployment.

5. DATA PRODUCTS AND DATA FLOW

soxspipe will reduce data into a set of final data products (see Table 1 for details) which shall meet ESO Phase 3 standards 'out-of-the-box'. This has the benefit of allowing us to build an automated workflow (see Figure 7) to reduce data directly on the La Silla summit immediately after the data is acquired by the NTT and SOXS and then stream the reduced data directly into the ESO SAF³⁰ in Garching, Germany. Owners of the data will then be able to access the fully-reduced data alongside the raw data within minutes of the shutter closing on their observation. This low-latency, automatic reduction is possible thanks to the fixed format of SOXS (apart from exchangeable slit) allowing calibration frames to be prepared ahead of time before science data reductions. The SAF then acts as both a data distribution solution and also fulfils the SOXS consortium's legacy archive requirements.

Access to the 'open stream' method of shipping reduced data directly to the ESO SAF will initially require the ESO Archive Science Group to review and verify a moderately sized collection of soxspipe reduced data. Once the quality and content of the data produced by the pipeline have met ESO Phase 3 standards we will then be allowed to ship data products to the archive without further need of passing through a gatekeeper. The pipeline will automatically reduce on all point-source targets above an AB magnitude of r = 19 (with the stretch goal of r = 20). For sources below this magnitude, the pipeline will attempt to automatically reduce the data but may require some user interaction to optimise object extraction.

[‡]https://github.com/thespacedoctor/soxspipe

[§]https://pypi.org/project/soxspipe/

6. CONCLUSIONS

The SOXS Pipeline soxspipe has been designed and written in object-orientated Python 3 using an agile framework of development. Built with core aims of allowing for fast, automatic reduction of raw data, streaming reduced data into the ESO SAF without need for human intervention, while also providing end-users with a simple-to-use, well-documented command-line interface, it is our hope that the pipeline will help facilitate the success of SOXS in the years to come.

REFERENCES

- [1] Schipani, P. et al., "Development status of the SOXS spectrograph for the ESO-NTT telescope," *Proc. SPIE* **11447**, 1144709 (2020).
- [2] Aliverti, M. et al., "The mechanical design of SOXS for the NTT," Proc. SPIE 10702, 1070231 (2018).
- [3] Aliverti, M. et al., "Manufacturing, integration and mechanical verification of SOXS," *Proc. SPIE* **11447**, 114476O (2020).
- [4] Biondi, F. et al., "The assembly integration and test activities for the new SOXS instrument at NTT," *Proc. SPIE* **10702**, 107023D (2018).
- [5] Biondi, F. et al., "The AIV strategy of the common path of Son of X-Shooter," Proc. SPIE 11447, 114476P (2020).
- [6] Brucalassi, A. et al., "The acquisition camera system for SOXS at NTT," Proc. SPIE 10702, 107022M (2018).
- [7] Brucalassi, A. et al., "Final design and development status of the acquisition and guiding system for SOXS," Proc. SPIE 11447, 114475V (2020).
- [8] Capasso, G. et al., "SOXS control electronics design," Proc. SPIE 10707, 107072H (2018).
- [9] Claudi, R. et al., "The common path of SOXS (Son of X-Shooter)," Proc. SPIE 10702, 107023T (2018).
- [10] Claudi, R. et al., "Operational modes and efficiency of SOXS," Proc. SPIE 11447, 114477C (2020).
- [11] Colapietro, M. et al., "Progress and tests on the Instrument Control Electronics for SOXS," *Proc. SPIE* 11452, 1145225 (2020).
- [12] Cosentino, R. et al., "The vis detector system of SOXS," Proc. SPIE 10702, 107022J (2018).
- [13] Cosentino, R. et al., "Development status of the UV-VIS detector system of SOXS for the ESO-NTT telescope," *Proc. SPIE* **11447**, 114476C (2020).
- [14] Genoni, M. et al., "SOXS End-to-End simulator: development and applications for pipeline design," *Proc. SPIE* **11450**, 114501B (2020).
- [15] Kuncarayakti, H. et al., "Design and development of the SOXS calibration unit," *Proc. SPIE* **11447**, 1144766 (2020).
- [16] Ricci, D. et al., "Architecture of the SOXS instrument control software," Proc. SPIE 10707, 107071G (2018).
- [17] Ricci, D. et al., "Development status of the SOXS instrument control software," *Proc. SPIE* 11452, 114522Q (2020).
- [18] Rubin, A. et al., "MITS: the Multi-Imaging Transient Spectrograph for SOXS," *Proc. SPIE* **10702**, 107022Z (2018).
- [19] Rubin, A. et al., "Progress on the UV-VIS arm of SOXS," Proc. SPIE 11447, 114475L (2020).
- [20] Sanchez, R. Z. et al., "Optical design of the SOXS spectrograph for ESO NTT," Proc. SPIE 10702, 1070227 (2018).
- [21] Sanchez, R. Z. et al., "SOXS: effects on optical performances due to gravity flexures, temperature variations, and subsystems alignment," *Proc. SPIE* 11447, 114475F (2020).
- [22] Schipani, P. et al., "The new SOXS instrument for the ESO NTT," Proc. SPIE 9908, 990841 (2016).
- [23] Schipani, P. et al., "SOXS: a wide band spectrograph to follow up transients," *Proc. SPIE* **10702**, 107020F (2018).
- [24] Vitali, F. et al., "The NIR spectrograph for the new SOXS instrument at the NTT," *Proc. SPIE* **10702**, 1070228 (2018).

- [25] Vitali, F. et al., "The development status of the NIR arm of the new SOXS instrument at the ESO/NTT telescope," *Proc. SPIE* **11447**, 114475N (2020).
- [26] Young, D. et al., "The SOXS data reduction pipeline," Proc. SPIE 11452, 114522D (2020).
- [27] Marty, L. et al., "The quality check system architecture for soxs," Proc. SPIE (in press 2022).
- [28] Landoni, M. et al., "The soxs scheduling system," Proc. SPIE (in press 2022).
- [29] Tonry, J. L., Denneau, L., Flewelling, H., Heinze, A. N., Onken, C. A., Smartt, S. J., Stalder, B., Weiland, H. J., and Wolf, C., "The ATLAS All-Sky Stellar Reference Catalog," The Astrophysical Journal 867, 105 (Nov. 2018).
- [30] Romaniello, M., "ESO's Science Archive Facility," in [American Astronomical Society Meeting Abstracts #218], American Astronomical Society Meeting Abstracts 218, 305.03 (May 2011).

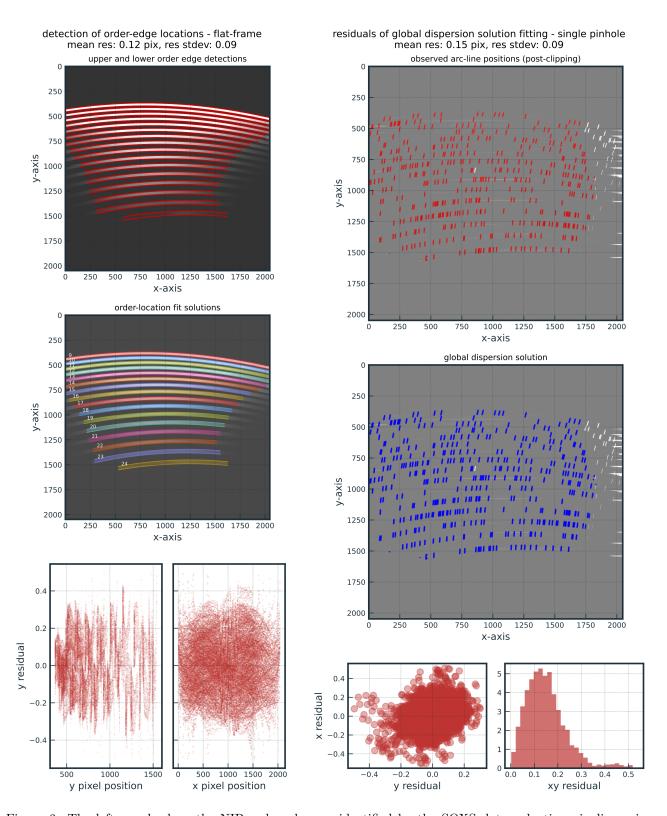


Figure 6: The left panels show the NIR order-edges as identified by the SOXS data-reduction pipeline using a master-flat frame created from a set of full slit flat-lamp frames generated by the E2E simulator. On the right, the resulting final dispersion solution and residuals as fitted by the SOXS data-reduction pipeline using a simulated arc-lamp frame obscured by a multi-pinhole mask. The arc lines detected in the frame (top right image panel) are used to fit a global dispersion solution (middle right image panel). The residuals of the fits as compared to measured order-edge and arc-line locations can be found in the bottom panels.

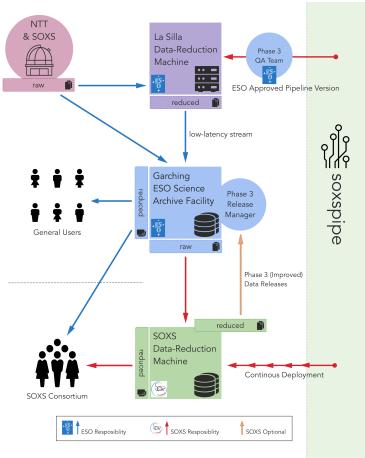


Figure 7: The proposed SOXS data flow. Data is acquired by the NTT and SOXS on the summit of La Silla, Chile (top right). Raw data is reduced on the summit (top centre) and transferred within minutes to the ESO SAF (Garching, Germany) where data-right owners can access it (central in blue). In parallel, the SOXS consortium will also reduce their data on a remote machine (probably cloud-based) with a leading-edge version of the pipeline (bottom in green). If at any point it is decided that new development of the pipeline has led to significantly improved data products compared to those hosted on the SAF the consortium may opt for a complete reprocessing and replacement of the data on the SAF via a dedicated Phase 3 Data Release (orange arrow).