The APO Data Analysis Working Group Report

Thursday 3rd September, 2020

Adam Kowalski[1], James R. A. Davenport[2], Jon Holtzman[3], Xinyu Dai[4], Sarah Tuttle[2], Shane Thomas[5], and Nancy Chanover[3,5,6]

[1]University of Colorado
[2]University of Washington
[3]New Mexico State University
[4]University of Oklahoma
[5]Apache Point Observatory
[6]3.5-m Director

Overview

The APO Data Analysis Working Group (DAWG) was formed as an immediate outgrowth of the 3.5-m strategic planning meeting that was held in Sunspot, NM, in June 2019. At that meeting, the participants agreed that an important goal for the observatory is to have easily accessible, reliable, quick-look software for all facility instruments and a recommended starting path for reducing data from those instruments to yield scientific quality data. Thus, the DAWG was established with the following charge:

1. Collect and evaluate the currently existing pipelines/reduction directions etc.
2. Identify a common set of requirements for these tools (common keywords/metadata, etc.).
3. Determine what is missing, i.e. what software we need but do not currently have.
4. Identify a path forward for filling those gaps.
5. Develop a long term data management plan for the observatory, including archiving functions (cost-benefit analysis of building our own vs. using existing archives), and exploring the role that APO can play in championing ground-based astronomical data reduction software (e.g. by partnering with astropy).

The DAWG is composed of Jon Holtzman, James Davenport, Xinyu Dai, Sarah Tuttle, Adam Kowalski (chair), Nancy Chanover, and Shane Thomas. We met five times over Zoom from August 2019 - July 2020, and all meetings were recorded.

ARC Users' Survey on Data Reduction Facilities

To address items #1 and #3 above, we sent a Google survey to the ARC users in late 2019. In this survey we asked users to tell us how they reduced their 3.5-m data, and to rate their satisfaction with their method. The survey is contained in the Appendix. For each instrument that they use, we requested them to briefly tell us which tools/pipelines/cookbooks that they employ. We asked them to rate their satisfaction, ranging from 1 ("It's horribly difficult") to 5 ("It's very smooth"). Finally, we asked the users about any tools and resources that we might not be aware of and about any general tools that they wished were provided by the APO.

Survey Results and Findings

In summary, there were ~ 40 replies, and the results have been compiled and synthesized. Several users made us aware of promising avenues for IRAF/PyRAF reduction tool replacements (and which are not linked on the APO wiki):

- PyPeit: a reduction tool that can be used for ARCES and possibly for our longslit instruments; several pipelines for other telescopes (mostly 8-10m class) exist.

  https://pypeit.readthedocs.io/en/latest/

  https://pypeit.readthedocs.io/en/latest/new_spectrograph.html

- CERES: a reduction tool used for ARCES, requires Python 2.7.

  https://github.com/rabrahm/ceres

- Mosasaurus: A reduction tool that is used for DIS/transit spectroscopy (Zach Berta-Thompson).

  https://github.com/zkbt/mosasaurus

- blaze.py: a PyRAF longslit reduction pipeline used to auto-reduce DIS spectra for a PhD thesis (Kowalski).

- The ARCES ETC:

  https://joss.theoj.org/papers/10.21105/joss.01130

- Ds9match: match science frames to guider frames - Kowalski /Ketzeback authors

- scat.

- Time series / frequency analysis tools written by Anjum Murkadam and Scott Klineman for agile data.

- calendar ephemeris tools at APO.

- photutils (astropy).

- Astropy's replacement of imexam:

  https://imexam.readthedocs.io/en/0.9.1/

- Adam Kowalski's Python replacement of DS9+imexam (for on the fly longslit spectroscopy quick look):

  https://www.dropbox.com/s/g5ygs99a6w2a304/imexam_python_v2.1.tar.gz?dl=0

  https://www.dropbox.com/s/smbs3ylka1kr7l7/DST_HSG_sample.tar.gz?dl=0

  (These two links may be deprecated at some point; for future versions see Adam Kowalski's webpage: https://afkowalski.bitbucket.io)

Note, the DAWG does not make specific recommendations or assessments about these various tools and pipelines. We recommend that this list be provided on the APO webpage/wiki (see recommendations section below).

Other tools mentioned by users (that are linked to the APO wiki) are AESOP (ARCES) and PyDIS (DIS). There were many individually constructed pipelines that were mentioned (some that seem to use IRAF, some for which it was not clear). The authors of these (unnamed) pipelines include Jessica Roberts (for ARCTIC/NICFPS), Kolby Weisenburger, Jon Holtzman, Jean, Pey-Lian Lim, Hannah Lewis, and John Barentine.

A more detailed breakdown per instrument is as follows:

- Of the 19 responses for **ARCES**, 13 used IRAF (typically a custom .cl script or Karen Kinemuchi's guide) and 6 used either CERES or Aesop. The average satisfaction was 3.3/5.
- Of the 24 responses for **DIS**, 17 used IRAF and the average satisfaction was 3.7. Four used PyDIS.
- Of the 12 responses for **ARCTIC**, 7 used IRAF and 3 used AstroImageJ with an average satisfaction of 4.3.
- Of the 10 responses for **TSpec**, 9 used TspecTool (IDL) with an average satisfaction of 4.1.
- Of the 5 responses for **NICFPS**, 4 used IRAF with an average satifaction of 3.0.
- Of the 5 responses for **AGILE**, 2 used IRAF and 2 used AstroImageJ with an average satisfaction of 4.

The number of replies that use their "own script" without specifying one (IRAF, etc) are 4.

Overall, we found that there is still a significant base of IRAF/PyRAF users. The average satisfaction for reducing ARCES and DIS data was below 4, whereas for the other instruments (besides NICFPS) the satisfaction of the reduction pipelines was above 4. There is a general desire for each instrument page to provide at least a basic data reduction guide / recommendation, especially for ARCES. While experienced users generally find Karen Kinemuchi's IRAF reduction guides sufficient, there is a steep learning curve for new users. Most of the responses from the Google Survey were likely from experienced users and thus it was biased toward their needs.

It was discussed whether the easiest and cheapest solution for providing quick-look reduced data for the users was to host a standalone, "legacy" desktop (not connected to the internet and not accessible to users) that operates PyRAF and runs an auto-reduce script after every observing run. It would be straightforward to configure K. Kinemuchi's ARCTIC and ARCES guides and A. Kowalski's PyRAF/DIS script to run automatically. However, there were concerns about the security of such a system as well as concerns that this path would actually entail significant work (one always needs to prep, prune, and examine the raw data especially if there are instrument issues, bad weather, or non-standard setups) This idea does not adequately prepare for future instruments that will certainly not have PyRAF reduction suites, and so it was disfavored by the majority of the DAWG as a sustainable solution.

Some requests in the Users' Survey for APO are:

1. to provide a repository for standard calibration data for each instrument, routinely updated and archived and accessible to the users.

2. to provide quick look ("approximately correct") reductions, especially useful for new users.

3. to pursue improved ARCES and TSpec reduction tools: TSpecTool is written in IDL, and all ARCES tools rely on PyRAF (e.g., apall) to some extent, so having formal replacements of these two reduction pipelines was requested. There is no pipeline for NICFPS, which is a barrier for new users.

4. to provide validated exposure time calculators for each instrument.

5. to provide a quick look tool while observing, for spectroscopy it'd be nice to have a light-weight interactive tool that's not IRAF to help assess S/N, wavelengths, resolution, spectral type, etc on the fly

6. John Bally had a detailed request: APO should provide de-distorted ARCTIC and NICFPS images. This will enable assembly of mosaics in a few easy steps with functions such as IRAF/ imcombine. Although distortion coefficients are expected to be relatively stable over long-time intervals, it would be good to devote engineering time for prevision re-determination of distortion coefficients once a year, perhaps after summer shut-down.

The DAWG does *not* formally recommend all of these (and others that are not listed above) requests at this time. We include these for the record of user input that might be addressed partially or fully if the formal recommendations of the DAWG (below) are completed.

Additionally, there were some issues that users expressed concern about, such as providing a recommended way to flat field ARCTIC data, which were outside the scope of our recommendations. We believe that an organized system of communication among users, such as a more active Wiki, could alleviate some of these problems.

DAWG Recommendations on Data Reduction Facilities

Many of the widely used tools that observational astronomers have employed for decades (IDL, IRAF, Fortran) are currently in the extended process of being deprecated or replaced by software written in more popular computer languages, such as Python or Java [1]. For aperture photometry, AstroImageJ is becoming a standard tool [2], but one-to-one replacements of the spectroscopic tools of IRAF are lacking. Several efforts at ground-based observatories across the US are underway to standardize raw data and provide new data reduction facilities: at Keck, Luca Rizzi and Josh Walawender are leading efforts to develop a Data Reduction Pipeline (DRP) following the Joint Astronomy Centre's ORAC-DR software guidelines [3]. Gemini has built software called DRAGONS along similar principles to what Keck is pursuing. STScI is planning STPIPE. Astropy has photutils.

The DAWG recommends that APO provide a comprehensive and updated list of the possible data reduction tools and guides on each instrument webpage. In particular, the lists on the wiki should be supplemented with the tools mentioned by the user survey, as outlined above (Pypeit, Mosasauraus, etc..).

The DAWG recommends that funding and support be provided to develop robust exposure time calculators and quick-reduction tools for every instrument, perhaps using parts of the existing pipelines, or components of PyRAF. Though we recognize several large-scale projects for centralized tool(s) are unfolding (e.g. astropy), both exposure time and quick-look tools would provide valuable and specific improvements to the 3.5-m user experience. If funding and support is provided by APO, these tools could be developed by a combination of observatory staff and researchers or students from member institutions. We note however they are unlikely to be robust for general use without some Observatory testing and oversight. Although having these tools sooner would be better, there is not a hard deadline for their need. These two tools would likely improve the efficient use of the 3.5-m for new users especially.

DAWG Recommendations on Data Storage and Archiving

This issue is more straightforward, and the agreement among committee members was easier to attain. **We recommend that APO purchase disk space to store data for longer periods of time.** Local disk storage is relatively cheap, and we believe APO should commit to regularly (perhaps annually) increasing its data storage capacity in order to keep user data for $\sim$10 year timescales, rather than the currently enforced 1 year limit. Other possible solutions include cloud storage architechtures, which can be investigated further by Shane or Tracy. We do not recommend creating a queryable data archive such as MAST or IRSA (which was requested by some users), and we do not recommend anything within the data storage that requires changes to TUI. Overall we would like to see improvements to the capacity and longevity of raw data available to users.

An ancillary recommendation is to establish common header keywords across all instruments.

An ancillary recommendation is that weather information, cloud cam data, weather records, and related observatory metadata be stored for longer periods of time.

The DAWG also recommends APO establish and maintain a calibration and example data repository for all instruments on the 3.5-m. This will help users plan observations and track instrument changes (e.g., sensitivity degradation, contamination, chip illumination, etc..) on their own. Spectrophotometric standard star spectra, arc lines, flat fields can be periodically provided for each instrument. Some of the nightly checks that the observing specialists already perform would be useful for this repository. This may also help address the general user need for exposure time calculators, train new users on basic data reduction, and provide a basis for many quick-look data reduction scenarios.

TUI

Though not specifically within the charge of the DAWG, we identified and discussed TUI as a critical tool for the continued success of the Observatory.

DAWG Recommendation for TUI

The DAWG committee recommends identifying a pathway to continue support, maintenance, and improvements for TUI, such as porting to python 3.X, integrating new instruments, adding new functionality or observing modes.

Appendix: Users' Google Survey

APO 3.5m Data Analysis Working Group – User Software Survey

To help develop the observatory's software strategy, we need to know what tools users are currently using to reduce and analyze their data.

For each instrument that you're using, briefly tell us which tools/pipelines/cookbooks you're using. Feel free to write as much description as you like about your workflow, thoughts on difficulties, or improvements!

ARCES (Echelle)

ARCES (Echelle) Reduction & Analysis Tools

What do you use to reduce ARCES data? e.g. IRAF, PyRAF, AESOP, K. Kinemuchi's cookbook, other, nothing?

Your answer

ARCES (Echelle) Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

Dual Imaging Spectrograph (DIS)

DIS Reduction & Analysis Tools

What do you use to reduce DIS data? e.g. IRAF, PyRAF, PyDIS, blazeR300, other, nothing?

Your answer

DIS Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

NICFPS

NICFPS Reduction & Analysis Tools

What do you use to reduce NICFIPS data? e.g. IRAF, PyRAF, other, nothing?

Your answer

NICFPS Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

AGILE

AGILE Reduction & Analysis Tools

What do you use to reduce AGILE data? e.g. IRAF, PyRAF, Astropy photutils, AstroImageJ, other, nothing?

Your answer

AGILE Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

TripleSpec

TripleSpec Reduction & Analysis Tools

What do you use to reduce TripleSpec data? e.g. IRAF, PyRAF, Tspectool (IDL), other, nothing?

Your answer

TripleSpec Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

ARCTIC

ARCTIC Tools

What do you use to reduce ARCTIC data? e.g. IRAF, PyRAF, Astropy photutils, AstroImageJ, K. Weisenburger's pipeline, J. Huehnerhoff's cookbook, other, nothing?

Your answer

ARCTIC Tools Satisfaction

	1	2	3	4	5	
It's horribly difficult	○	○	○	○	○	It's very smooth!

Anything Else?

Any other tools/resources/ideas we should be aware of?

e.g. other pipelines or cookbooks you've found useful, efforts to develop tools we might not be aware of, etc

Your answer

Any general tools you wish existed for APO data?

What new or improved resources might help reduce the burden for reducing, analyzing, and publishing your data from APO?

Your answer

Email for follow-up questions or feedback?

Your answer

Page 1 of 1

Submit

Never submit passwords through Google Forms.

References

1. Bobra, M. G. *et al.* A Survey of Computational Tools in Solar Physics. **295**, 57 (2020). DOI 10.1007/s11207-020-01622-2. 2003.14186.
2. Collins, K. A., Kielkopf, J. F., Stassun, K. G. & Hessman, F. V. ASTROIMAGEJ: IMAGE PROCESSING AND PHOTOMETRIC EXTRACTION FOR ULTRA-PRECISE ASTRONOMICAL LIGHT CURVES. *The Astronomical Journal* **153**, 77 (2017). DOI 10.3847/1538-3881/153/2/77.
3. Jenness, T. & Economou, F. ORAC-DR: A generic data reduction pipeline infrastructure. *Astronomy Computing* **9**, 40–48 (2015). DOI 10.1016/j.ascom.2014.10.005. 1410.7509.