Astronomical Data Analysis Software and Systems IX
ASP Conference Series, Vol. 216, 2000
N. Manset, C. Veillet, and D. Crabtree, eds.

An Assessment of Data Analysis Software at ST ScI

R. A. Shaw, R. L. White, P. Greenfield

Space Telescope Science Institute

Abstract. We present an assessment of software utilities that are used within the *HST* community for data analysis. The assessment also provides insight into current hardware, OS, and analysis system use patterns. Future ST ScI software development will emphasize software components that are inter-operable with new or existing data analysis systems.

1. The HST User Survey

An in-depth assessment of support for users of *HST* data was conducted during the fall of 1998. This effort was the most extensive survey of the *HST* user community conducted to date, and consisted of a detailed survey of all categories of users, followed by focussed discussions with selected scientists and staff within and outside ST ScI. The focus groups served to help interpret and validate the survey results, and to probe a number of questions in greater depth. The aims of the assessment were to provide guidance to ST ScI management on the perceived quality of *HST* services, and to identify those areas that could be reduced or streamlined so that resources can be re-directed to support for NGST. We present here that part of the assessment concerning support for science calibration and analysis software.

2. Hardware and Operating Systems

The survey provided an opportunity to inquire of the hardware platforms and operating systems that *HST* users choose for their data analysis (Figure 1). While the popularity of Sun workstations and Solaris was perhaps predictable, the variety of other choices for platform/OS combinations in use must shape the planning for future system and application development. The popularity of IBM/PC compatibles and the Linux operating system has grown from near zero a few years ago to second most cited platform, and it is a very common choice of platform/OS among respondents who use more than one combination. We expect the popularity of PC/Linux to continue to grow, particularly as more users acquire machines for home use. Respondents were also asked to characterize the level of professional support they receive for OS and application upgrades on their systems. Although most (77%) reported full or at least partial support, a significant minority do not. We suspect the fraction without full support may track the expansion of computers purchased for home use, and if so, it will be important to assure that software systems are easy to configure.

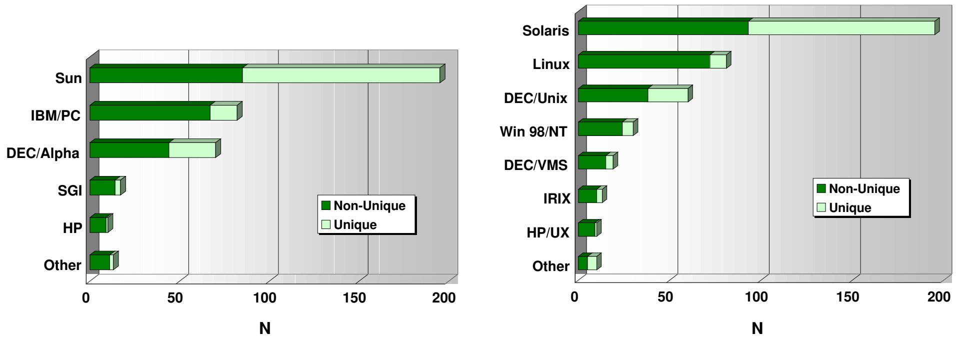

Figure 1. The hardware platforms (*left*) and OSs (*right*) used for *HST* data analysis. Respondents who indicated more than one platform/OS (*dark bars*) are distinguished from those who did not (*light bars*).

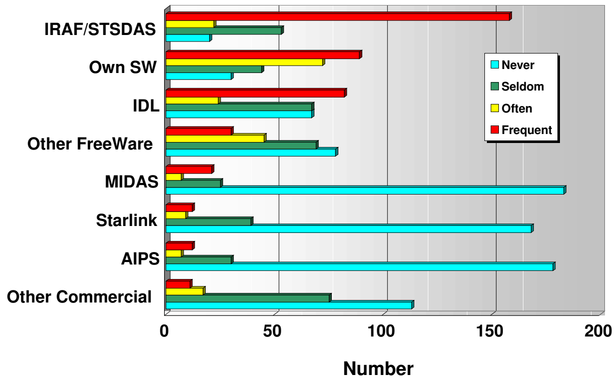

Figure 2. Data analysis software used to analyze *HST* data, ordered by frequency of use. The popularity of users' own software suggests a striking degree of self-sufficiency in the community for the development of scientific data analysis applications. Examples of freeware called out by users: SAOimage, SuperMongo, Tiny Tim, and xv.

3. Choice of Data Analysis Software

Respondents were asked to indicate what software they use for *HST* data analysis, and with what frequency. The results, in Figure 2, show that IRAF/STSDAS and IDL are very popular for analyzing *HST* data, while some other major systems are not, in part because the *HST* community is dominated by US astronomers who observe primarily in the optical and UV bands.

What was perhaps most surprising to us was the high popularity of data analysis applications written by the respondents themselves, which among other things suggests a fair degree of self-sufficiency in the user community for software development. While the survey did not ask respondents to specify the nature of their home-grown software (i.e., whether compiled programs, scripts, etc.), a close examination of the data suggests that IRAF CL scripts and IDL procedures

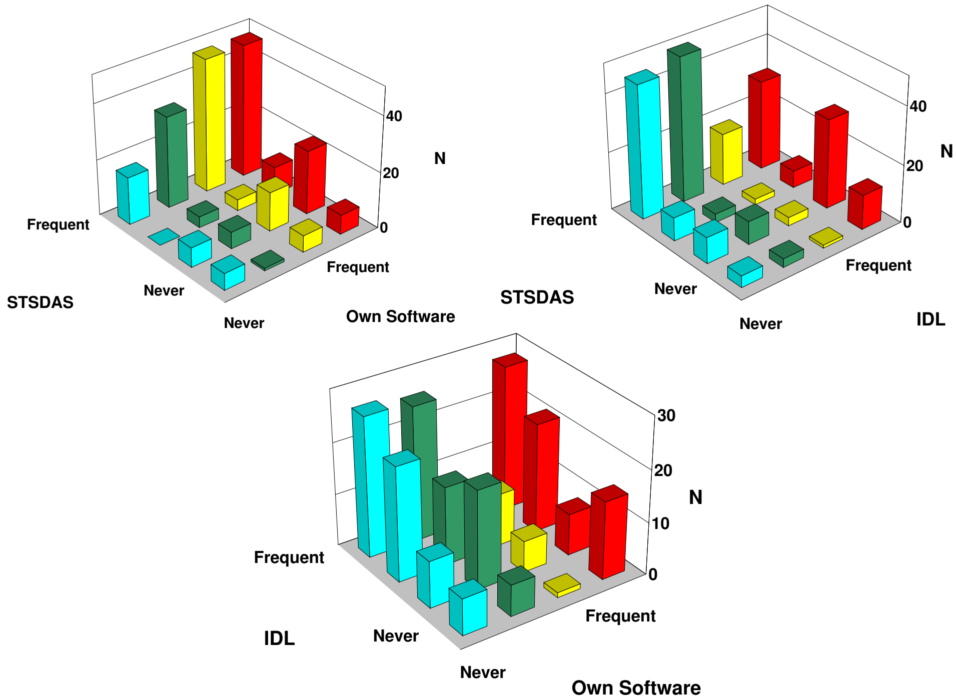

Figure 3. The frequency with which *HST* data analysis is performed using the three most popular sources of software: IRAF/STSDAS, user-written software, and IDL. The authors of home-grown software form a population that is distinguishable from STSDAS and IDL users, though there is also substantial overlap among these groups.

cannot wholly account for the results. Figure 3 shows the relative frequency of use among the most popular choices of analysis software noted in Figure 2.

The popularity of user-written software confirms our impression that users are keen to develop their own techniques for analyzing their data, and are willing and sufficiently able to express their ideas in software. The distinction in the popularity between user-written software and that offered in the major systems suggests that neither IDL nor the IRAF CL fully addresses their needs.

4. Evaluation of STSDAS

One of the challenges for ST ScI management in the coming years will be to find a way to scale back the resources devoted to *HST*, to that they may be redirected to NGST. The results of this assessment have provided a relatively clear picture of the relative value that users place on the software and services that ST ScI provides to the *HST* community. We found that, while users value the STSDAS software and related services, high quality calibration pipelines are the most critical for their *HST* data analysis. A more detailed examination of the respondents' use and evaluation of major STSDAS packages shows that the generic data manipulation and format conversion applications see the most use, followed by general data analysis packages (fitting, isophotal analysis, etc.); the use of many other applications (e.g., plotting, synthetic photometry) is spotty. While the quality of the packages was generally rated highly, there were quite a number of complaints about the difficult learning curve for STSDAS both as a system and for many of the applications, and about the difficulty in using the most complex tasks. These problems have been a consistent weakness of the IRAF system for many years. The survey respondents, by a wide margin, also reinforced the need to continue to export STSDAS to the external community. Such a practice is also valued internally within our development group, as it helps ensure that the software remains relatively portable and of high quality.

5. Implications for Future Development

The results of this assessment are providing useful guidance in revising the long-range plan for science software development at ST ScI, and to some degree it has shaped near-term projects as well. Future software projects will place more emphasis on technologies that enable even greater self-sufficiency for users of *HST* data, and that lessen the dependence on the IRAF system. Such a strategy will allow ST ScI to focus on developing critical software that the community is least well equipped to provide, while leveraging their experience in data analysis where possible. Current efforts along these line include:

- C-language bindings to IRAF libraries (De La Peña & Farris 2000).
- A Python-based replacement for the IRAF CL (Greenfield & White 2000; De La Peña 2000).
- Development of calibration pipelines in a standard language (ANSI C) that can be run stand-alone or from other environments, such as the IRAF CL. These pipelines also use FITS as their run-time disk data format, and are re-entrant to a high degree, to enable inter-operability with other software.
- New system libraries, such as *hstio*, that isolate the dependency of STSDAS applications on external data structures, and upon IRAF data I/O.

The STSDAS system itself will actually shrink in size before it grows: the next release will not include a number of little-used tasks, such as those for generating most calibration reference files. In the longer term, STSDAS development for *HST* will probably become more focussed on applications that address areas specific to *HST* data. More general applications, such as that for analysis of spectra, and system development are more likely to be undertaken in the context of support for NGST. Future system development will emphasize inter-operability with other analysis systems (and with host-level applications in general), and will seek to build on modern, open technologies for enhancing major system components, such as the Help system, facilities for modern graphical interfaces, and new, Web-based services.

References

De La Peña, M. D. 2000, this volume, 63

De La Peña, M. D. & Farris, A. 2000, this volume, 677

Greenfield, P. & White, R. L. 2000, this volume, 59