



From the telescope to the computer: how to improve the archiving

L. Cuesta*, M. T. Eibe and A. Ullán

Centro de Astrobiología (INTA-CSIC), Carretera de Ajalvir, km 4, 28850 Torrejón de Ardoz, Madrid, Spain

Abstract. One of the most critical issues in operating robotic observatories is the data acquisition and storage. Most telescopes produce enormous amounts of information to be processed and stored. Finding an efficient way to archive the data, including quality control procedures, is essential. Virtual Observatory projects could be very helpful in these cases.

Keywords : robotic telescopes – data archiving – image quality

1. Introduction

It is well known that robotic telescopes play a very important role in many fields of astronomy (Baruch 1992; Helmer & Morrison 1985), in particular the study of exoplanets (Street et al. 2003; Konacki et al. 2003; Pollacco et al. 2008), Solar System bodies (Steel & Marsden 1996; Taff 1981; Viggh et al. 1997), GRBs (Castro-Tirado et al. 1999; Akerlof et al. 1994, 2003; Jelínek et al. 2005) and many others (Helmer & Morrison 1985). Over the last ten years the evolution of this type of telescopes has been impressive and most of the initial problems have been solved (Boyd et al. 1988; Downey & Mutel 1996). From the very beginning, robotic observatories have been mostly semiautomatic telescopes that can make some decisions for themselves, most importantly those related to the weather and the corresponding actions. Year after year the quality of the data obtained with robotic telescopes has been improved by using sophisticated algorithms (Kubánek et al. 2008) (see also www.audela.org, indilib.org and ascom-standards.org).

On the other hand, because of this better use of the telescopes, the production of

*email: cuestacl@cab.inta-csic.es



Figure 1. Archives are an important part of the telescopes.

data increases consequently. In addition, robotic telescopes are normally used to monitor some particular astronomical objects that require many images of short exposure times instead of few images of long exposure times. The result is that the amount of data obtained every night takes up a huge number of MegaBytes, normally more than 10 GigaBytes per night (for instance, SuperWASP produces about 100 GB/night, see www.superwasp.org) (see Fig. 1).

Although the capacity of storage is bigger and bigger every day, it is necessary to find a solution for this problem: at a rate of about 10 GB per night it is easy to accumulate around 4 TB per year.

2. Requirements of data archives

From an historical perspective, the requirements on processing and archiving has grown continuously. At the time of eye observation no processing was necessary, little analysis was possible and archiving was not necessary (see Fig. 2). When the photography came to astronomy, it began to be possible to perform some simple processing and analysis, but archiving remained very simple and it was not very expensive. Fi-

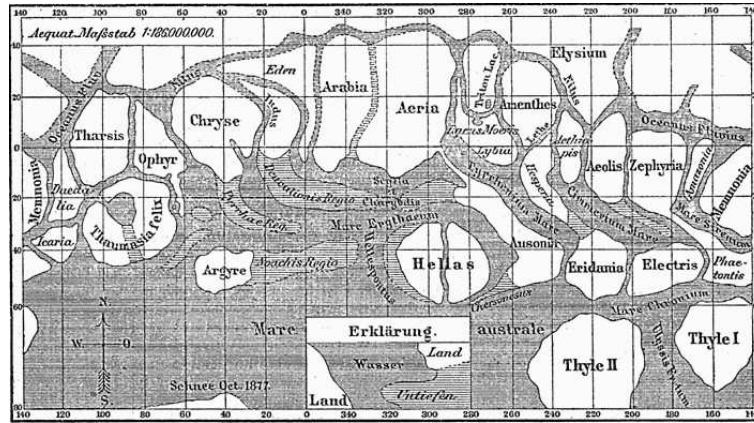


Figure 2. Map of Mars by Giovanni Schiaparelli.

nally, with the advent of the electronic devices (IPCS, photomultipliers, CCD) the requirements increased considerably. Now the processing of the data is mandatory (pre-reduction: map of bad columns, bias, dark, flat). In contrast, it is possible to perform a complete analysis of the data and to extract the maximum amount of information using different methods. However, archiving requirements increase very hard with time.

In some cases the solution is simply not to save repetitive data and only save the results. It depends on the use of the data and the type of archives to be created. In the case of restricted archives, the data is fully reduced and all the possible analysis is done (astrometry, photometry, spectroscopy, polarimetry). In these cases only the final data have to go to the archives (coordinates, photometry values, light curves). In most cases these archives belong to scientific projects with specific goals:

- Exoplanets
- GRBs
- NEOs
- Supernovae

The access to the data is provided only to the members of the project. Because not all the data has to be saved it only requires moderate disk space for the archive.

In the case of open archives the data is fully available to the whole scientific community and the entire series of data processing must be accessible. Then, the data is fully reduced and some analysis is done. Both raw and analyzed data are archived. In these cases, a large amount of disk space is required. General and specific scientific goals projects have use of this type of archives. Some good examples are data archives from most observatories (ING archive, see casu.ast.cam.ac.uk/casuadc/archives/ingarch,

CAHA archive, see caha.sdc.cab.inta-csic.es/calto/index.jsp SuperWASP archive, see www.wasp.le.ac.uk/public).

With this list of characteristics for both types of archives, the question is: What may be archived? In the case of open archives the answer is all the data, without restrictions. However, in the case of restricted archives, for which not all the data may be archived, the answer is: Only representative data and of good quality (see Fig. 3). The question of quality will be treated in next section. Here, we are going to give only some recommendations on how to reduce the size of archives with the mean criteria of saving as less images as possible. Firstly, for this kind of archives, redundant data are not relevant. Then, images taken on the same set of observations, with all the observation parameters equal (FOV, exposition time, filter, atmospheric conditions, etc) may be transformed in order to save only a representative image of each set. A good example is the observations taken with the aim of detect or characterize exoplanets for which a successive series of observations with the same parameters is required. The same happens with observations conducted to obtain the cycle of a variable star.

The best way to reach this objective is the transformation of every redundant image to any other information that takes up less space in the archive. Then, all the relevant information of each image is saved. In order to get the position information, the images may be transformed to astrometry information. Then, we get the individual WCS solution (Greisen & Calabretta 2002; Calabretta & Greisen 2002; Greisen *et al.* 2002) for every detected star in the field. The same may be applied to the flux received from every star. Then we will have a series of light curves for each set of observations.

Then, a good set of tools to select the data to be archive is needed. There are some automatic pipelines that perform these task (Ballester *et al.* 2008; Rit   *et al.* 2008). The real-time analysis of the data is mandatory if we want to fully exploit the data (Starr *et al.* 2008).



Figure 3. An example of data that may be and may not be saved: Only good data may be saved.

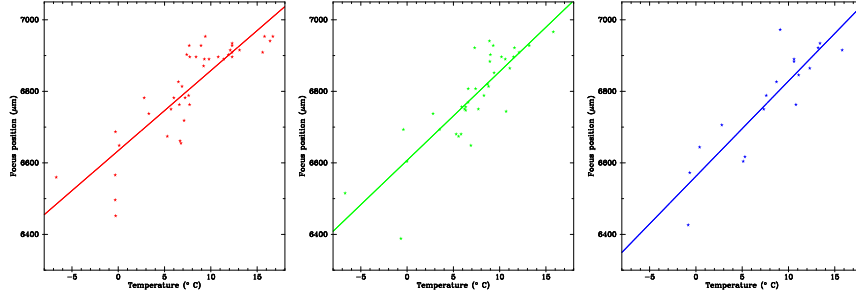


Figure 4. A typical example of variations on focus value with temperature for the filters R, V and B. For each case a linear fit is present.

3. Control quality

The quality of the data depends on many parameters related with the observation (telescope, instrumentation, weather conditions) and with the processes done after observation (data reduction, analysis). Then, it is necessary to determine which of these parameters affect the quality of the data and how it does so. This is the task of quality control. Quality control is a set of procedures to:

- Identify the degradation of data
- Study the ways to get the best accuracy
- Eliminate poor data

In the case of robotic telescopes, one of the typical causes of image degradation, and the consequent loss of quality of data, is deviations from the optimal focus. These deviations cause the observed stars profile to differ from a point-like source and increase the FWHM. There are several factors that cause this effect. Temperature changes cause expansion in the telescope structure leading to disruptions of focus (Eibe et al. 2011). Because the system has to work all the night on its own, a complete set of focus values for each value of temperature is the best way to reach a high quality on the data (see Fig. 4). In some cases some other weather conditions, such as humidity, or the position of the telescope, due to gravity flexures, can influence the focus (see Fig. 5 and Fig. 6).

The same expansions in the telescope structure could affect the pointing of the telescope. But, also, there could be a variety of systematic errors including unusual mount flexure, incorrect location and drive train peculiarities that produce the same effect. Then, the relative position of a particular star on the CCD varies (see Fig. 7). But not only the position varies, it will also present elongated shapes for the stars. To solve this problem a good fine-alignment procedure is required. It is not very

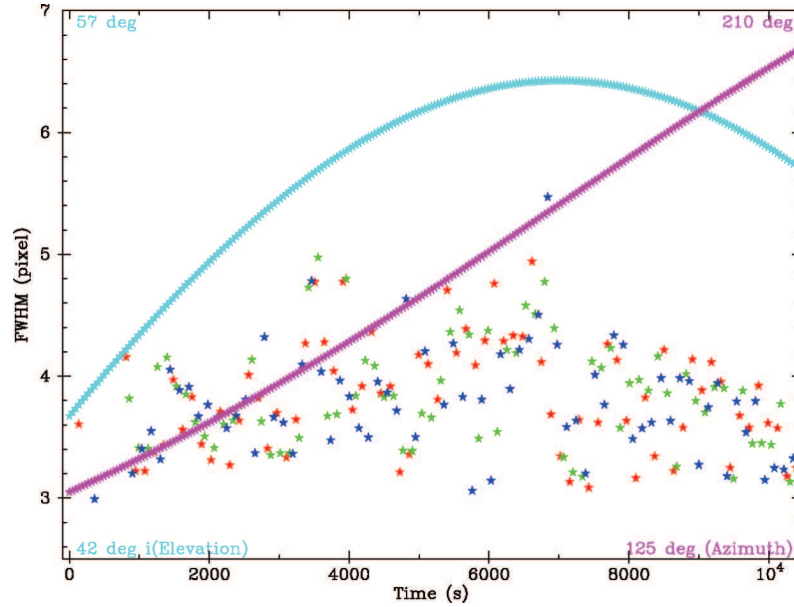


Figure 5. A typical example of variations on FWHM, for the three filters R (red dots), V (green dots) and B (blue dots), as a function of time, for a set of observations, as an indication of quality degradation. Variations on temperature (cyan dots, for which a box between 5 and 9 is used) and humidity (violet dots, for which a box between 25% and 65% is used) are also represented in order to find correlations.

difficult to obtain this. The idea of this procedure is to take a large collection of images which cover the entire sky and compare the real position with the target position. The differences are the errors. Then, we will have indications on the errors of the position of the telescope over all the movement space. With this procedure we will have a table of differences in Right Ascension and Declination for the entire sky (see Fig. 8). The system analyzes it and combines into a map of pointing errors. The resulting map will be automatically used by the telescope control system for all subsequent pointing operations by compensating the target position in real time (Eibe et al. 2011).

Another aspect for which data quality is very sensible is the correction for the flat frames (see Fig. 9). While some people prefer to use flat-fields from the dome (Marshall & DePoy 2005), because they are easier to take, some other people prefer flat-fields from the sky (Chromey & Hasselbacher 1995). Both have problems and advantages. The principal problem with flat-fields from the sky is that the period during the twilight available to take these is very short and it is very difficult to find a good field in the sky without any bright star that could be used every night. Also, at the time of twilight the sky is not completely dark and if the telescope is not sufficiently

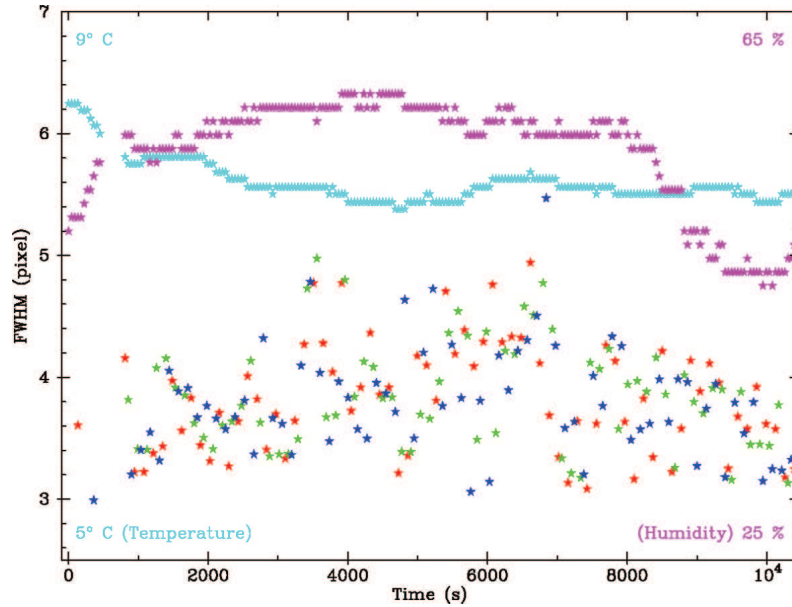


Figure 6. Equivalent to Fig. 5 but in this case the variations of elevation (cyan dots, for which a box between 42 deg and 57 deg is used) and azimuth (violet dots, for which a box between 125 deg and 210 deg is used) are also represented in order to find correlations.

well buffered some scattered light could appear on the edges of the CCD. However, the system is in normal conditions much more like a typical observation.

On the other hand, dome flat-fields are very easy to take. There is not any requirement on the time at which it may be taken. Also, it is possible to regulate dome illumination in order to simulate adequately the best conditions to reduce diffuse light. But it is not so easy to have a flat illumination on the CCD.

In any case, in order to create a master flat that does not introduce more noise, it is necessary to have a huge number of flat-fields and with a high level of counts each. It is also a good idea to study frequently that the flat structure has not changed. If that is the case, a new master flat has to be created. In any case, of course, the accuracy of this process depends a lot on the required precision of the data.

Although not as important as the flat frame problem, the question on the shutter map of the CCD is also a question to study carefully (Zissell 2000), mostly if the typical expositions are very short (see Fig. 10).

Finally, there are many other observation parameters that have a clear S/N dependence and that have to be taken into account in order to increase accuracy. Weather parameters are very important and, for instance, with high levels of humidity the

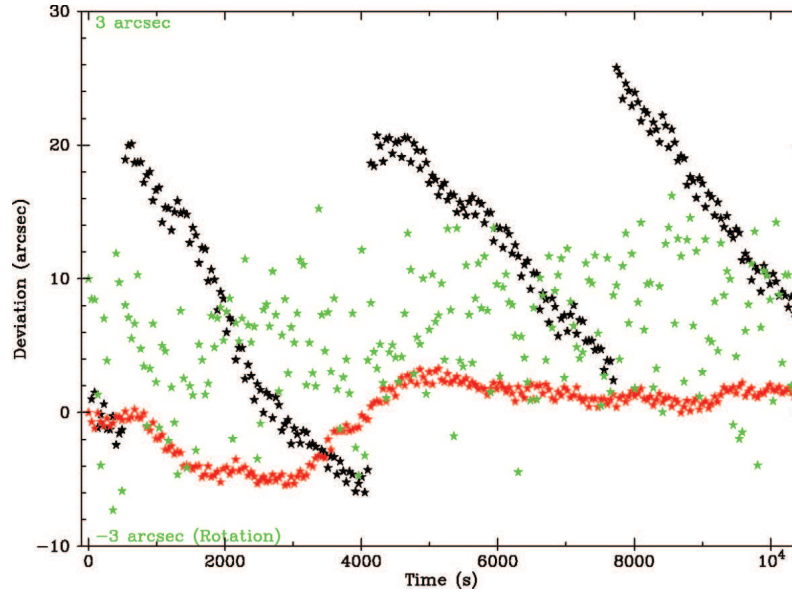


Figure 7. Typical example of variations on X (black dots) and Y (red dots) pixel deviations and field rotation (green dots) along time, for a set of observations, as an indication of tracking accuracy.

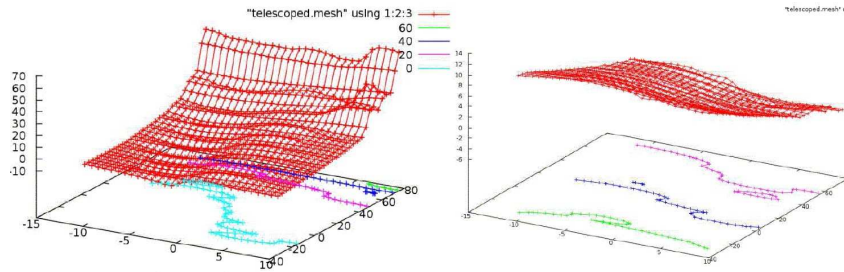


Figure 8. A typical example of a pointing model for both axes. Left panel represents deviations in Right Ascension (in arcmin) while the right panel represents deviations in Declination (in arcmin). Both deviations are calculated in the space of Right Ascension from -15 to 10 hours and Declinations from -40 deg to 80 deg.

degradation of the data is clear. Also, the temperature has a demonstrated influence on the data quality.

Two other parameters that are of key importance during the observations are the seeing and airmass. With very high seeing values the stellar profile width increases disastrously and it becomes impossible to get quality data. Experience shows that the seeing should not exceed values of $\sim 2''$. The airmass has a similar and combined ef-

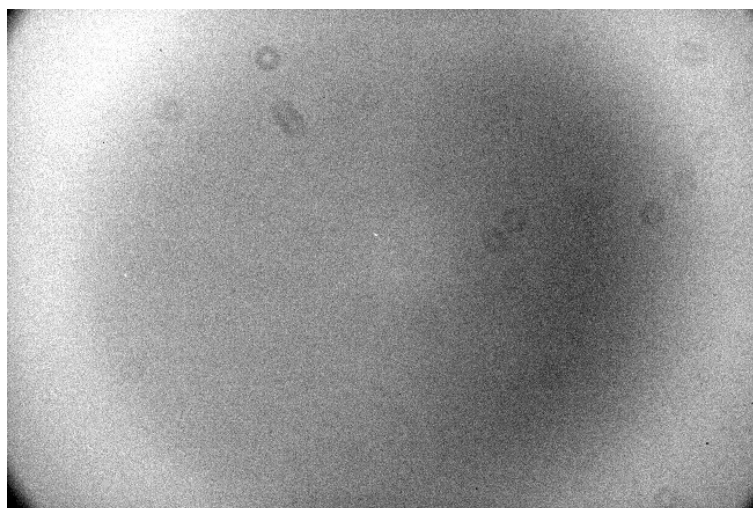


Figure 9. A typical flat frame taken from sky. Observe evidence for some dust and the effect of indirect light contamination on the edges.

fect. By increasing the amount of atmosphere traversed the effect of seeing grows and greatly increases the disturbances. Here, the experience is that it should not exceed values of ~ 2 in air mass, i.e., below 30 deg.

After the observation, it is necessary to choose the data that has sufficient quality to extract the information, either photometric, spectroscopic or astrometric. This makes multiple tools necessary to help in this task. First, it is very useful to analyze the FWHM of the observed stars and establish a quality criterion. Then you have to properly identify the stars in the field. WCS type tools provide a solution by comparison with astrometric star catalogues. Finally, it is important to choose a format for data that meets the standards of data files and that is widely used to allow sharing and exchange of data. The FITS format (Wells et al. 1981) (see <http://fits.gsfc.nasa.gov/>) is the most widely used and provides the best results, including a full header scheme.

When data are collected, they mostly become photometric data using different tools and processes. Proper use of these tools should provide photometric data of good quality. However, it is appropriate to take into account certain considerations. It is very important to choose properly the reference calibration stars. Of course, its magnitude must be constant and fixed. It is also very important to use stars with magnitudes similar to that of the star studied. Thus the calculations are more consistent and avoid the possible error caused by deviations from linearity in the CCD. To achieve this you can play with the values of exposure time and binning while making the observations. In essence, it is advisable to reach values of S/N as high as possible at each observation.

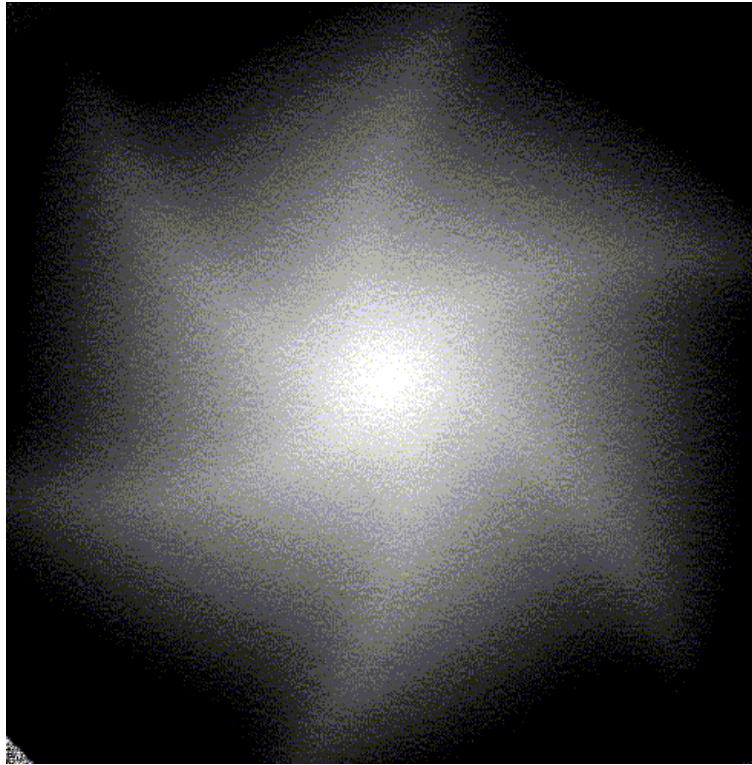


Figure 10. Typical shutter map of a CCD camera.

4. Virtual observatory

Modern Astronomy is experiencing an exponential growth of data volume driven by the new instruments and telescopes. However, our understanding of the Universe is increasing much more slowly mainly due to methodological bottlenecks. Virtual Observatory is a kind of accessible archive that is used by scientist to get access to data from any observatory. It is essentially a collection of data archives and software tools from many data centres to create a scientific research environment using the Internet to give the opportunity of a transparent and distributed access to data. Then scientists can discover, access, analyze, and combine nature and lab data.

Its purpose is to provide an efficient coordination among the different national initiatives in the framework of the Virtual Observatory and to achieve an effective integration of all the expertise in this research domain. The main goal of the Virtual Observatory is to ensure an efficient management (discovery, access, retrieval and analysis) of the information available in astronomical archives.

Behind the Virtual Observatory project idea there are many groups. At Inter-

national level: IVOA (International Virtual Observatory Alliance, see www.ivoa.net) was formed in June 2002 with a mission to facilitate the international coordination and collaboration in the field of VO. The IVOA now comprises 17 VO projects from Armenia, Australia, Brazil, Canada, China, Europe, France, Germany, Hungary, India, Italy, Japan, Korea, Russia, Spain, the United Kingdom, and the United States. The work of the IVOA focuses on the development of standards. At European level: Euro-VO has the top level objective of the coordination for the VO activities at European level. IVOA exists to develop inter-operability standards and provide with some standards (see www.ivoa.net/Documents/SIA and www.ivoa.net/Documents/latest/CharacterisationDM.html) to help robotic observatories to develop their archive access portal in accordance.

One of the most typical tools used in the VO framework is the data analysis and data mining. Both represent an excellent way to perform an efficient and systematic study of the vast amount of data available from the federation of VO archives. On the other hand, there are many fields in Astrophysics with a strong need of direct and rigorous comparisons between theoretical models or simulations and real data. However, the different architectures, programming codes, formats, etc, make it, most of the times, extremely difficult and inefficient. In order to define the requirements needed for a full interoperability between observations and models, an IVOA Theory Group was created to provide the access protocols to theoretical data and a theoretical models server.

In summary, Virtual Observatory projects have the following objectives:

- Construction of a community of science users
- Scientific Workshops
- Support to VO science programs
- Production of on-line tutorials
- Development of IVOA standards and services
- Support to data publishing in the VO
- Education and Outreach activities

Some examples of VO around the world are:

Euro-VO: the European VO; SVO: the Spanish Virtual Observatory; AstroGrid: the UK's Virtual Observatory Service; French-VO; ESO-VO; ESA-VO; National Virtual Observatory: the USA's VO; Virtual Observatory, India: the India's Virtual Observatory; Iran Virtual Observatory: the Iran's Virtual Observatory; IVOA (International Virtual Observatory Alliance).

Acknowledgments

The authors would like to acknowledge financial support from the Spanish programme AYA2009-14000-C03-02 starting in 2010.

References

- Akerlof C. W., Fatuzzo M., Lee B. et al., 1994, AIP Conference Proceedings, 307, 633
 Akerlof C. W., Kehoe R. L., McKay, T. A. et al., 2003, PASP, 115, 132
 Ballester P., Biereichel P., Kaufer A. et al., 2008, Astronomical Data Analysis Software and Systems XVII, 394, 577
 Baruch J. E. F., 1992, Vistas in Astronomy, 35, 399
 Boyd L. J., Genet R. M., Hayes D. S., 1988, in Hayes D. S., Genet R. M., eds, Automatic Small Telescopes, Fairborn Press
 Calabretta M. R., Greisen E. W., 2002, A&A, 395, 1077
 Castro-Tirado A. J., Soldán J., Bernas M., et al., 1999, AAPS, 138, 583
 Chromey F. R., Hasselbacher D., 1995, AAS, 186, 309
 Downey E. C., Mutel R. L., 1996, Astronomical Data Analysis Software and Systems V, A.S.P. Conference Series, 101, 380
 Eibe M. T., Cuesta L., Ullán, A. et al. 2011, MNRAS, 412, 1181
 Greisen E. W., Calabretta M. R., 2002, A&A, 395, 1061
 Greisen E. W., Calabretta M. R., Valdes F. G. et al., 2006, A&A, 446, 747
 Helmer L., Morrison L. V., 1985, Vistas in Astronomy, 28, 505
 Jelínek M., Kubánek P., Hudec R. et al., 2005, The Astrophysics of Cataclysmic Variables and Related Objects, 330, 481
 Konacki M., Torres G., Jha S. et al., 2003, Nature, 421, 507
 Kubánek P., Jelínek M., French J., et al., 2008, SPIE proceedings, 7019, 12
 Marshall J. L., DePoy D. L., 2005, arXiv:astro-ph/0510233
 Pollacco D., Skillen I., Collier Cameron A. et al., 2008, MNRAS, 385, 1576
 Rit   C., Slijkhuis R., Rosati P. et al., 2008, Astronomical Data Analysis Software and Systems XVII, 394, 605
 Starr D. L., Bloom J. S., Butler N. R., 2008, Astronomical Data Analysis Software and Systems XVII, 394, 609
 Steel D. I., Marsden B. G., 1996, Earth, Moon and Planets, 74, 85
 Street R. A., Pollaco D. L., Fitzsimmons A. et al., 2003, in Drake D., Seager S., eds, Scientific Frontiers in Research on Extrasolar Planets, ASP Conference Series, Vol. 294
 Taff L. G., 1981, PASP, 93, 658
 Viggh H. E. M., Stokes G. H., Shelly F. C., et al., 1997, AAS, DPS meeting #29, BAAS, 29, 959 1
 Wells D. C., Greisen E. W., Harten R. H., 1981, A&ASS, 44, 363
 Zissell R. E., 2000, JAAVSO, 28, 149