FULL TITLE
ASP Conference Series, Vol. **VOLUME**, ©**YEAR OF PUBLICATION**
NAMES OF EDITORS

DASCH to Measure (and preserve) the Harvard Plates: Opening the ∼100-year Time Domain Astronomy Window

J. Grindlay, S. Tang, R. Simcoe, S. Laycock,[1] E. Los, D. Mink, A. Doane & G. Champine

Harvard Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

[1]Presently at Gemini Observatory, Hilo, HI

Abstract. The temporal Universe is now possible to study on previously inaccessible timescales of days to decades, over a full century, with the planned full-digitization of the Harvard plate collection. The Digital Access to a Sky Century @ Harvard (DASCH) project has developed the worlds highest-speed precision plate scanner and the required software to digitize the ∼500,000 glass photographic plates (mostly 20 x 25 cm) that record images of the full sky taken by some 20 telescopes in both hemispheres over the period 1880 – 1985. These provide ∼500-1000 measures of any object brighter than the plate limit (typically B ∼14-17) with photometric accuracy from the digital image typically Δm ∼0.1-0.15 mag, with the presently developed photometry pipeline and spatially-dependent calibration (using the Hubble Guide Star Catalog) for each plate. We provide an overview of DASCH, the processing, and example lightcurves that illustrate the power of this unique dataset and resource. Production scanning and serving on line the entire ∼1 Pb database (both images and derived light curves) on spinning disk could be completed within ∼3-5 y after funding (for scanner operations and database construction) is obtained.

1. Introduction

The era of astronomical surveys in the time domain has begun, with a variety of telescopes for which the primary science is time variability studies of astrophysical objects. Spurred in part by high-energy astrophysical processes and sources for which variability is the rule, not the exception, entire space observatories, most notably the Rossi X-ray Timing Explorer (RXTE), have been devoted (in large part) to time variability studies of primarily accreting X-ray binaries, but including M dwarfs (flare stars) to bright active galactic nuclei. Variability surveys have launched Time Domain Astronomy (TDA) over the past decade in the optical domain: the QUEST survey for RR Lyraes (Vivas et al. 2001) and the Palomar-QUEST survey (Djorgovski et al. 2007) are but two which incorporate newly established or refurbished (respectively) facilities for temporal studies. TDA is of course also central to the major new rapid-cadence and deep-sky survey projects now underway: PanSTARRS (Kaiser et al. 2002) has had first light with its telescope (2.4 m) on Mauna Kea and routine observing will begin shortly; and LSST (Tyson et al. 2003; Walker 2003) has its ambitious 8.4 m wide-field telescope under active development with first light expected by ∼

2016 from its Cerro Pachon site in Chile. While these (and more) are impressive beginnings for the emerging growth of temporal astronomy, they are necessarily so far constrained to timescales $\leq$20y, the approximate beginning of the modern digital era for systematic time variability studies of astrophysical phenomena.

Obviously much longer temporal records exist for astronomical data. However, these are usually sparse and not continuous or if so, for only relatively short timescales. Visual records of the most extreme variability have been crucial in certain astrophysical contexts, such as the Chinese Court astrologers' recording of the supernova now identified with the Crab nebula, thereby fixing its birth year (1054 AD) and even month/day (July 4, on the modern calendar). But these records of visual phenomena are necessarily limited to the brightest objects (visual magnitudes $\leq$6.5) and provide relatively sparse coverage with relatively few (sometimes only one) measurements of a given object. They also suffer, of course, from accuracy, usually no better than $\sim$0.5 mag. Although much of stellar variability was established with a century of astronomical imaging on glass plates, these data (in plate archives) have not been widely available because they have not been digitized and possible to analyze with automated pipelines. There is now the capability, and the project, to bring the historical era of astronomical plate imaging into the modern digital fold. The Digital Access to a Sky Century @ Harvard (DASCH) (Grindlay et al., in preparation) is now underway to develop the tools and software to digitize and calibrate the $\sim$100y collection of $\sim$500,000 glass plates of astronomical images of the full sky (northern and southern hemispheres) that presently occupy 3 floors of a building at the Harvard College Observatory (HCO). When each plate (most are 20 x 25 cm) is digitized into 11μm pixels with 12 bit depth, the $\sim$750 Mb (with Metadata) per plate becomes a total data volume of $\sim$400 Tb of mosaiced images that should be available on spinning disk for rapid download over the web. Since each mosaic is tiled from 60 scanner/CCD images (each 4 x 4 cm) that are half-step overlapping in the "X" direction of the scanner (and only slightly overlapping in the "Y") direction), the total archival data volume of the HCO plate collection is $\sim$1100 Tb. Allowing for the fact that the HCO plates include $\sim$28,000 large (35 x 43 cm) A-series plates and for various software tools and catalogs needed for the DASCH analysis, the total disk storage required is $\sim$1200 Tb or 1.2 Pb.

In this paper, we first describe the broad outline of how DASCH works, from the scanner to an abbreviated summary of the processing software. We then provide a few example light curves to show the power (and promise of what's to come) of making a digital record of the sky. We conclude with a projected timeline for carrying out the full digitization and posting the full dataset for public access, as well as the long-term plans for disposition of the Harvard plates.

2. Overview of DASCH

The astronomical plate collection of the Harvard College Observatory (HCO) is the world's largest with (mostly) 8 x 10 inch (20 x 25 cm) glass plate negative emulsions each recording a separate image of the sky (taken with $\sim$20 separate telescopes, most with apertures in the 0.1- 0.3 m range, in both northern and southern hemispheres). The unique feature of this collection of wide-field (typically 10° on a side) images of the astronomical sky is it's ∼100 y span, from 1881 - 1985 (with a partial gap from ∼1952 - 1963) that provides the longest and best sampled astronomical record of the sky available. Any given region (or object) typically appears on ∼500 - 1000 plates over this century-long period. This allows the most complete study of long-term (days, months to decades) variability study of astronomical objects on timescales ∼100 y, which will take nearly a half century to replace with modern digital images.

Accordingly, in 2002 the lead author resurrected his earlier (∼1988) concept to digitize the Harvard plates to make them generally available on line as a digital archive and initially reduced astrometric/photometric catalog. Whereas the earlier considerations to digitize the Harvard plates were not feasible due to the early state of both CCDs and low-cost digital processing and data storage in the early 1990s, by 2002 both were becoming feasible. A proposal was submitted to NSF in November 2003 to design and build a high-speed precision x-y table scanner with to digitally step a photographic plate across a fixed CCD, which would image the plate (1:1 magnification through a telecentric lens) illuminated from below by a bright LED array. The full scanner system was designed by R. Simcoe to use a high-end precision X-Y scanner from Aerotech, Inc. (Simcoe et al. 2006).

The digitized plates, with resolution and dynamic range fully as good as the original plates, are first processed with SExtractor (Bertin and Arnouts 1996) for object detection and initial isophotal photometry. The resulting list of objects is then solved for their (spatially dependent) World Coordinate System (WCS) using an initial solution obtained with Astrometry.net (Hogg et al. 2008) to derive accurate plate centers and then with a more precise and spatially dependent solution using WCS tools (Mink 2006) followed by higher order corrections such as ccmap (in IRAF) or SWARP (Bertin and Tissier 2007). The astrometric solution then enables matching with photometric wide-field catalogs for a spatially-dependent photometric calibration of the digital scan of a given plate. We now use the Hubble Guide Star Catalog, GSC (Lasker et al. 1985), with photometric accuracy about 0.2 mag (Lasker et al. 2008) for initial photometric calibrations and light curves of every stellar object detected and resolved (typically >1 x 10^5) on each plate. The initial astrometry and photometry pipeline processing system is summarized by Laycock et al. (2009), with now higher order corrections and refinements to the pipeline as described by Tang et al. (in preparation) and is briefly outlined below.

We summarize the key components of DASCH as follows:

Scanner: The initial fabrication of the scanner was done at Aerotech, Inc. (Pittsburgh, PA) and completed in 2004. This included the custom scanner base (designed to be possible to fit into the HCO scanner lab set up for DASCH). A custom plate loading system was designed and fabricated (Simcoe et al. 2006) to allow semi-automated loading of either two "standard" (8 x 10 in) plates side by side, or a single A-plate (14 x 17 inch). The scanner and plate loading system is all under control of a fast PC running Windows/XP (for compatibility with the frame grabber driver) with software developed by E. Los of the DASCH team. All DASCH analysis is done with two high-speed quad-core Linux systems. Additional improvements continue to be made to the system as more experience is gained – such as full-plate pre-scan exposures to measure mean plate density (to set the optimal exposure times gated by the LED array) over the whole plate rather than the initial test measure at the plate center done with the scanner itself.

Scanning totals: As of May, 2009, we have scanned ∼5100 plates covering 6 fields: M44 (∼600 plates), 3C273 (∼1400 plates), Baade's Window (∼900 plates), and 3 PG quasars: PG084+349, PG1211+143 and PG2130+099 – each with 700 plates. These fields were chosen to include a well-calibrated field (M44) with which to develop our initial astrometry and photometry software (Laycock et al. 2009), a high latitude field (3C273) with, of course, a particularly interesting bright quasar to measure its long-term variability and compare with its historical visually derived lightcurve from the Harvard plates (Smith and Hoffleit 1963). Initial science is being done with these first fields scanned (Grindlay et al., in preparation; Tang et al., in preparation), but the primary goal has been to develop the optimum astrometry and photometry tools and overall scanner processing pipeline. A brief summary of the project status, including plate totals, is given on the DASCH website.[1]

Plate handling/cleaning: While the scanning time for a single plate is ¡1min, including loading and unloading from the loading platen, the time required to clean the plate of accumulated dust or (for ∼20% of plates) annotation marks (on the glass side of the plate) to denote objects of interest, is ∼5 min/plate. A semi-automated plate cleaning machine has been designed but not yet built in prototype form to optimize its operation. This will be required for the eventual production scanning (when funding permits) of the ∼400 plates/day needed to scan and process the full collection in ∼3 y. The pre-scanning operations now also include high-resolution digital photography of the plate before cleaning (to preserve historical records of annotations) and the plate jackets.

Preparation of Plate Metadata: Production scanning will proceed by celestial region (not object field) to continuously tile the sky. This will require completion of the acquisition of plate metadata (plate coordinates, as originally recorded; exposure times; etc.) which is now available (on line) for only about 170,000 plates or ∼1/3 the total. Volunteer G. Champline completed a heroic effort to obtain digital photographs of all pages of all ∼1200 original telescope logbooks, as these contain the original record(s) of what was done for each exposure and each plate (approximately 20% of the plates are multiple exposures, with the telescope pointing shifted slightly between exposures, to obtain shorter time-scale variability information and/or calibration data – by stepping to shorter exposure times, for "precise" relative calibration). The ∼80,000 JPG images from these digital photos are being used by volunteers at the American Museum of Natural History (AMNH) in NYC, under the direction of M. Shara, to do keyboard entry of the required data (1-2 lines of characters) given that cursive 19th century script is not amenable to OCR. We have proposed to speed up the process by hiring data entry services (as done successfully in a test region) to complete the data entry within 1-2 y. This would also for the first time allow users to see approximate exposure depth and coverage of a given field, since the

[1]http://hea-www.harvard.edu/DASCH/status.php

limiting magnitude on digitized scans can now be derived quite precisely (Laycock et al. 2009; Tang et al., in preparation) and from this a reasonably accurate estimate of limiting magnitude can be derived from just the exposure time for any given plate series. Sky coverage for plates scanned thus far and approximate numbers of plates expected to have limiting magnitude ∼16 (plotted only for the ∼1/3 of plates with available metadata) are shown in Figure 1. Several deep fields, with ∼700 plates (or more) reaching B ∼16 are apparent, and perhaps a third of the sky is covered to B ∼16 on ∼300 (or more) plates.

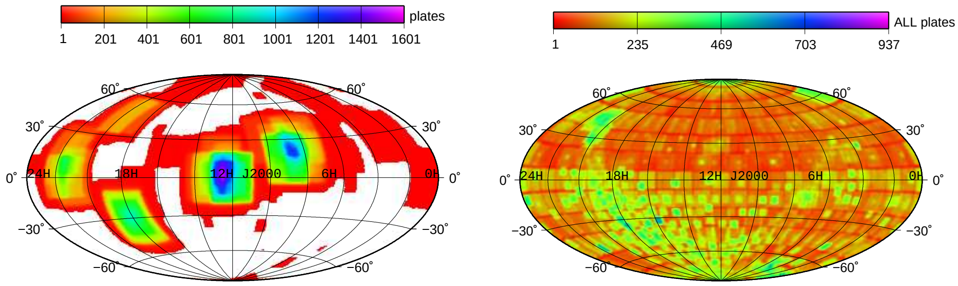

Figure 1. a) RA/Dec cords for ∼5100 plates scanned for DASCH; b) Approximate number of plates reaching limiting mag B ∼16 vs. RA/Dec for plate centers as derived from exposure times vs. limiting mag using the ∼1/3 of the plate collection with on-line Metadata.

Pipeline analysis of the scan data: The principal steps for processing of individual scans are described in Laycock et al. (2009). Further enhancements to the processing system are described in Tang et al. (in preparation), and the complete pipeline will be summarized in Los et al. (in preparation). In brief, the key elements are:

1. Mosaic the 60 scanner tile sub-images (4 x 4 cm on the plate) recorded from the 5(*2) x 6 array that a nominal 20 x 25 cm plate is scanned (with 5(*2) denoting the 10 half-steps to achieve 5 steps in the 20 cm or X-direction on the plate) into a single image. This is presently done by simply relying on the ∼0.1 μm absolute scanner motion setting accuracy(!) for alignment of the 11 μm image pixels from one sub-image tile to the next into the mosaic. Slight improvements to the registration of tiles could include accurate matching of stars in the overlap region (both X and Y directions) but has not yet been incorporated.

2. Run SExtractor (hereafter SE; Bertin and Arnouts 1996) for object detection at low S/N threshold above background grain noise. Experiments have shown that thresholds as low as ∼1.5σ above noise level are optimum for achieving maximum sensitivity, when various "filters" are applied to reject artifacts. SE produces a catalog of isophotal magnitudes for all objects detected, along with extensive image parameters used in subsequent analysis.

3. Run astrometry.net for *initial* plate center coordinates by fitting bright stars only (above a SE isophotal magnitude limit corresponding approximately to B ∼10. The initial astrometry is done against the Tycho catalog

(Høg et al. 2000). This initial solution then enables the accurate fit with WCSTools and IRAF/ccmap (at present, possibly SWARP in future) for higher order corrections for the full plate. For the ~20% of plates that have multiple exposures, the astrometry is done again for all *unmatched stars* (see below for matching) to derive the accurate WCS for exposures 2, 3, ... which are then processed separately through the complete pipeline (iterative WCS solutions have been demonstrated, but not yet fully implemented in the pipeline).

4. Match the complete catalog of object positions (for exposure 1; hereafter only this is considered) to the GSC2.3 catalog (McLean et al 2000; Lasker et al. 2008). Initial matches are made on position only but are subsequently checked for approximate magnitude matching of the GSC vs. instrumental (isophotal) magnitudes.

5. Filter out defects (plate scratches; incomplete cleaning smudges) by rejecting objects (or flagging them) with image properties not consistent with matched GSC stars in that *local bin* (see below) on the plate. This allows for and preserves spatial variations in the stellar psf.

6. Set blending flags for objects either already flagged by SE as blended or objects found by SE which from GSC matching should be blended with neighbors. These are likely to produce poorer photometry and so are not used in the stars selected to form the photometric calibration of the plate.

7. Search for "Pickering Wedge" plates (a fraction of the plates were exposed with a non-dispersive calibration thin wedge prism that produced a secondary image of each primary image at a fixed displaced position and fixed magnitude offset) and flag those objects either detected by SE or regions that should be detected and may produce blended images as PW secondary images (this step is not yet fully automated).

8. Fit a color term, C, for the full plate to derive a magnitude offset (usually $<$0.1mag) to be applied given the derived color sensitivity of that plate as determined by fitting Δm = m_{SE} - B = C * (B - R), where m_{SE} is the Iso mag from SE, and B and R are the GSC magnitudes in (approximately) Johnson bands B and R. The color term C is derived by minimizing the rms for Δm for every matched star (with B, R magnitudes in GSC) in the central region of the plate (to avoid chromatic aberration effects near the edges). Most of the Harvard plates are blue-sensitive emulsions with response similar to the GSC blue plates, in which case C = 0, whereas a red-sensitive emulsion would be fit with C = -1. To deal with air mass variations and thus colors over the wide-field of the plates, an expected atmospheric reddening correction is also derived for every star (given its elevation during the exposure) for a (small) magnitude correction.

9. Construct the calibration sequence between m_{DASCH}, the isophotal mag from SE, vs. GSC mag, B, for given plate by doing an *rlowess fit* (Cleveland 1981) to the scatter plot that forms the sequence. Due to the effects of vignetting and psf variation from center to edge of (any) plate, we calibrate in 8 equal area radial bins from plate center to within ~5 mm of the

plate edge. The remaining outermost region of the plate is designated "bin 9" and is generally useful only for gross magnitude variations in (bright) objects.

10. While the annular-bin calibration takes out the major effects of the telescope optics, it does not remove spatially dependent sensitivity variation effects on the emulsions (e.g. non-uniform developing, large scale defects, etc.). The digital image of each plate is therefore divided into a 50 x 50 grid (*local bins*) and for the stars in a given bin (typically ∼200) the median of the magnitude offset, $\Delta m_{bin} = m_{DASCH}$ - B (with m_{DASCH} the calibrated magnitude, including the effects of C) is computed for each bin. This array of Δm_{bin} values is locally smoothed (to remove effects of sparse samples in some bins) and the magnitudes of each star then corrected by this locally determined, and empirical, offset.

11. From this final determination of stellar magnitudes, a catalog of the full digitized plate is constructed and its values (positions, magnitudes, flags) for each object are entered into a MYSQL database that contains the full derived data for all plates scanned. Separate routines then permit rapid interrogation of this database for any object (or cone search field) to allow lightcurves with given properties to be extracted (∼1 sec per lightcurve). The present (May 2009) construction of this database has not yet entered the Baade's Window field (very crowded) and so contains results from ∼3900 plates scanned, which yield a total of ∼19.1M stars and a total of ∼432.5M measurements of those stars.

3. Representative Light Curves from DASCH

Example lightcurves from early DASCH processing (on the M44 field) are included in Laycock et al. (2009) and some are shown on the DASCH website. These include "random" field stars (to show photometric accuracy, typically with rms = 0.10 mag for B, R ∼ 11-13. Light curves for a number of known variables (CVs, Mira variables, etc.) are also shown. The exciting new science from DASCH will be the many new variables discovered; many have been found in analysis of scans of just one field (M44), as described by Tang et al. (in preparation). Here we include two "representative" examples, together with followup spectra obtained at the Whipple Observatory (using the FAST spectrograph). First, a long term variable (possible K giant and RSCVn binary) with ∼0.6 mag variations over ∼10 y and a spectrum that shows Hα in emission is shown in Figure 2. A second example is a probable eclipsing B star, shown in Figure 3.

These are just two of some ∼400 new variables in the M44 field with Δm $\geq$+1 mag variations and $\geq$3 consecutive points at least 2σ away from the mean or with significant negative fluctuations (eclipses or dips) as in Figure 3. Some of the more remarkable variables include stars with monotonic slow (∼100 y) brightening or dimming which may be extreme examples of RCrB star dust shells (they typically have M star spectra) or ∼5 y total duration rapid-rise, exponential decay, ∼1 mag flares which may be examples of new symbiotic stars. A much larger group of isolated "novae" (single measurement brightenings) is also found but requires further development of psf and plate defect filters to

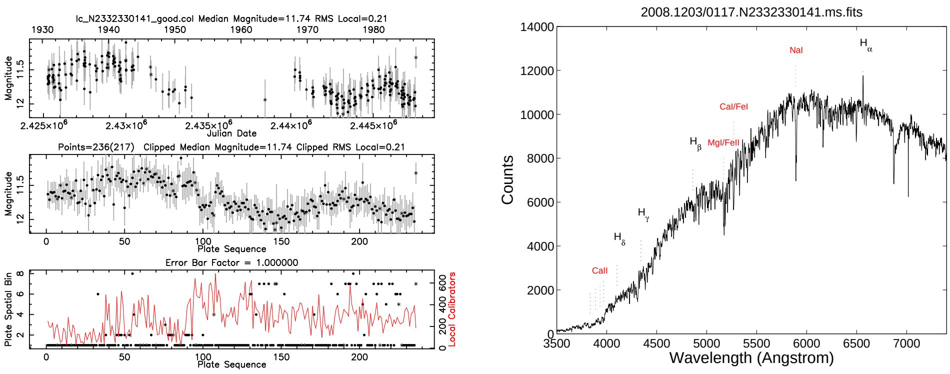

Figure 2. Left) Light curve of GSC-N233233041 showing 0.5 mag variations over ∼5 y timescales. Right) Spectrum showing it is approximately a K2 star with Hα in emission, possibly an RS CVn system.

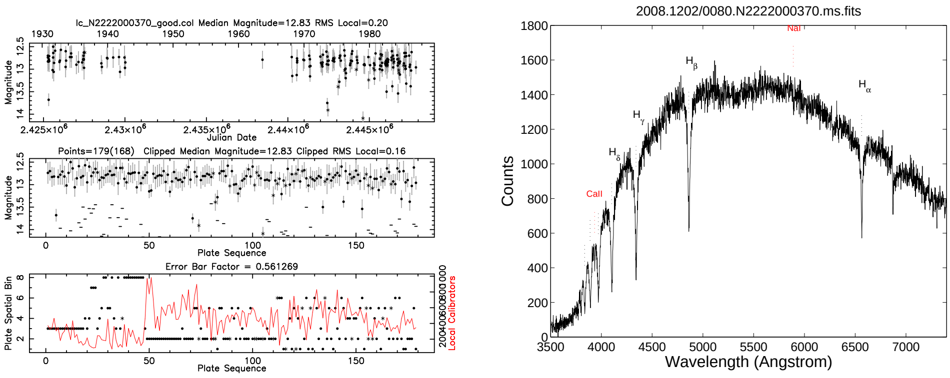

Figure 3. Left) Light curve of GSC-N2222000370 showing probable eclipses. Right) Probable B star counterpart as identified with spectrum from FAST.

eliminate more than the present ∼90% rejection rate of false events that could otherwise be random matches (≤2″) to background much fainter (B,R ∼18) GSC stars.

4. Plans and Prospects for DASCH

The ∼5 y history of DASCH has produced what may be the world's fastest and most precise astronomical plate scanner as well as an extensive software package that is close to pipeline status and can convert ∼400 scanned plates per day into efficiently-stored digital images and object catalogs readily accessible from a MYSQL database for analysis. The significant science that this will enable is wide-ranging ((Grindlay et al., in preparation)) and the synergies with the much more sensitive and rapid cadence surveys imminent (PanSTARRS) and planned (LSST) are immense.

Over the next 1-2 y, the final major upgrades to the scanner system should be made, a semi-automated plate cleaning system will be brought on line together with a plate bar-coding system, and the remaining key modules for the pipeline (improved blending and local photometric calibration corrections) will be largely completed. The goal is to begin production scanning of the HCO plates by 2011. With sufficient support (for personnel to operate the cleaning system, the scanner, plate movers, and – primarily – system software maintenance and improvements), the ∼500,000 plates can be scanned within 3 y. As plates are cleaned and scanned, they will be inserted in new protective envelopes, and re-packed in shipping boxes for permanent storage off-site. PARI is a logical storage repository, but a final decision will involve other considerations (from Harvard) as well. The plates must be preserved for "ground truth" (and history!), though the visual quality of the scans is fully equal to the original.

Whereas plate preservation is fundamental given the unique nature of this collection, DASCH is motivated first and foremost by the desire to make full access to the data public and immediately available. The first 6 fields now scanned will be made available (as digital images) along with our derived photometry and database as soon as resources (disk storage; servers; etc.) permit – likely within the next year. As the production scanning and (overnight) processing continues, the digital archive and database will grow and be immediately available. The time domain images could be linked for "movies" of limited regions (any region ...) of sky for public and professional viewing. This unique temporal record of the sky will likely be migrated into both World Wide Telescope (WWT) and Google Sky, which (we hope) will provide long-term storage and mirror site access to the eventual ∼1.2 Pb database which now seems large but will be "modest" within a decade. It is time for a DASCH to TDA.

Acknowledgments. We are grateful to the NSF for support under grant AST-0407380. We thank volunteers both at Harvard (A. Sliski, W. Toomey) and AMNH (M. Shara and Museum volunteers) for their generous help.

References

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Bertin, E. & Tissier, G. 2007, in ASP Conf. Ser. Vol. 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill & D. J. Bell (San Francisco: ASP), 507

Cleveland, W. S. 1981, The American Statistician, 35, 54

Djorgovski, G., Baltay, C., Mahabal, A., et al. 2007, BAAS, 38, 170

Høg, E., Fabricius, C., Makarov, V. V., et al. 2000, A&A, 355, L27

Hogg, D. W., Blanton, M., Lang, D., Mierle, K. & Roweis, S. 2008, in ASP Conf. Ser. Vol. 394, Astronomical Data Analysis Software and Systems, ed. R. W. Argyle, P. S. Bunclark & J. R. Lewis (San Francisco: ASP), 27

Kaiser, N., Aussel, H., Burke, B., et al. 2002, Proc. SPIE, 4836, 154

Lasker, B. M., Sturch, C. R., McLean, B. J., et al. 1990, AJ, 99, 2019

Laycock, S., Tang, S., Grindlay, J., et al. 2009, AJ, in press (and arXiv:0811.2005)

McLean, B. J., Greene, G. R., Lattanzi, M. G. & Pirenne, B. 2000, in ASP Conf. Ser., Vol. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 145

Mink, D. 2006, in ASP Conf. Ser. Vol. 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel, C. Arviset, D. Ponz & E. Solano (San Francisco: ASP), 204

Simcoe, R. J., Grindlay, J. E., Los, E. J., et al. 2006, Proc. SPIE, 6312, 17

Smith, H. & Hoffleit, D. 1963, Nat, 198, 650

Tyson, J. A., Wittmana, D. M., Hennawi J. F. & Spergelb, D. N. 2003, Nucl. Phys. B Proceedings Supplements, 124, 21
Vivas, A. K., Zinn, R., Andrews, P., et al. 2001, ApJ, 554, L33
Walker, A. R. 2003, MmSAI, 74, 999