Volume Title
*ASP Conference Series, Vol. **Volume Number***
Author

The DASCH Data Processing Pipeline and Multiple Exposure Plate Processing.

Edward Los[1], Jonathan Grindlay[1], Sumin Tang[1], Mathieu Servillat[1], and Silas Laycock[2]

[1]*Harvard College Observatory*

[2]*Department of Physics, University of Massachusetts at Lowell*

Abstract. Digital Access to a Sky Century @ Harvard (DASCH) is a project to digitize the collection of approximately 525,000 astronomical plates held at the Harvard College Observatory. This paper presents an overview of the DASCH data processing pipeline, with special emphasis on the processing of multiple-exposure plates. Such plates extended the dynamic range of photograph emulsions and improved photometric accuracy by minimizing variations in plate development procedures. Two approaches are explored in this paper: The repetitive use of astrometry.net (Lang et al. 2010) and local correlation searches. Both procedures have yielded additional quality control checks useful to the pipeline.

1. Introduction

The Harvard College Observatory plate collection consists of approximately 525,000 photographs produced by over 80 telescopes spanning over 100 years from about 1885 to 1992. Until recently, the primary technique for analyzing this data was for a researcher to visit the stacks and visually compare objects of interest with a nearby sequence of comparison stars. The goal of the Digital Access to a Sky Century @ Harvard[1] (DASCH) project (Grindlay et al. 2009) is to digitize this entire collection and provide photometry measurements for all objects. With the successful completion of a high speed plate digitizer (Simcoe et al. 2006) we have digitized over 10,000 plates and used E. Bertin's SExtractor program (Bertin & Arnouts 1996) to extract an average of 80,000 objects per plate. The analysis of this digitized data (Laycock et al. 2010), (Tang et al. 2010), and (Tang, S. in preparation) presents a number of challenges which are no longer encountered with modern CCD photographic techniques. One of these challenges is the presence of many plates which have multiple exposures of the same objects.

1.1. Types of Multiple Exposure Plates

Plates with multiple exposures were produced in five ways:

- Multiple exposures were taken on the same plate before development. This procedure guaranteed that all stars on the plate received the same processing on the

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

same emulsion. Figure 1 shows a triple exposure made of the M44 field. Often, the North Polar region was included as show in in Figure 2 so that the researcher could compare the Harvard North Polar Sequence with the region of interest. By varying the exposure time as shown in Figure 3 the researcher could extend the limited dynamic range of photographic emulsions.

- The dynamic range of photographic emulsions could also be extended by the use of a Pickering or Racine Wedge (Leavitt 1917) which produced the ghost objects of bright stars shown in Figure 4. The wedges were glass disks placed near the telescope objective which deflected a known percentage of the light into secondary objects. The wedge angle was usually small enough to avoid significant spectral spreading of the star. If the primary object is too bright for accurate photometric measurement, a researcher can estimate the brightness of the secondary object and then apply a fixed correction which is a function of the wedge area divided by the objective area.

- Another technique for extending the dynamic range makes use of coarse gratings (King 1931) to produce secondary objects as shown in Figure 5. The geometry of the grating placed near the telescope objective determines the percentage of light falling into each diffraction order.

- Some plates in the Damon South Yellow series have ghost objects at a 276 arcsec separation from primary objects. Since there is no record of Pickering Wedges being used for this telescope, there is a possibility that either the filter or one of the lens elements may have been misaligned.

- Some multiple exposure plates were accidental, caused either by operator error or by drive clock failures.

Often methods were combined as shown in Figure 6 which shows a multiple exposure grating plate.

It is essential for the DASCH pipeline to identify which plates have multiple exposures in order to avoid false detections of transient events such as flares or asteroids.

A secondary goal is for the DASCH pipeline to match each object on a multiple-exposure plate with the correct logbook entry so that the plate data can contribute to the lightcurve of each star covered by the plate.

1.2. Number of Multiple Exposure Plates

Now that 217,000 of the 525,000 logbook entries have been transcribed and entered into a MySQL database, a simple query shows that 5,842 or 2.6% of the transcribed plates have multiple exposures. Figure 7 shows that the bulk of the transcribed plates are from 1895 to 1950 meaning that the Damon series, the only active telescopes during the 1970's and 1980's, are currently under-represented. Figure 8 show that multiple exposure plates were popular from 1905 to 1915; from 1925 to 1935; and in 1944.

Figure 9 shows the percentage of multiple exposures for series that have at least 1000 plates transcribed. Neither the widest field, smallest arcsec/pixel scales nor the smallest field highest arcsec/pixel scales were favored for multiple exposures. Instead, the 7.5 cm Ross-Fecker (rb and rh series) telescopes and the 20 cm Ross-Lundin (ir series) telescopes were favored.

The use of coarse gratings is normally marked with the letter “G” in the logbook plate class entry, but plate class entries have been transcribed for only 97,986 plates so far. Figure 10 shows that only 18 of these entries are grating plates used in the 3-inch Ross Fecker (rh series) telescope between 1928 and 1945. While these transcriptions suggest that less than 100 plates used coarse gratings, the total may be as high as 8000 because of issues with the logbook plate classification system.

Unfortunately, no mention of the use of Pickering Wedge plates has been found in the logbooks. To supplement this logbook information, the plate scanner operators have been asked to note any plates which show multiple exposures, spectra, or gratings. These notations become a permanent part of the scanner database and flag all photometry measurements generated from these plates.

2. DASCH pipeline

An overview of the DASCH pipeline appears in Figure 11. The two key steps involved in the processing of multiple exposure plates are the “Pickering Wedge Filter” and the “Multiple Exposure Loop”.

2.1. Plate Preparation

Before any scanning can occur, the relevant entries in the logbook must be transcribed and entered into the MySQL scanner database. For single exposure plates, the exposure date is the most important logbook entry, but for multiple exposure plates the logbook sky positions are important for assigning exposure dates to all of the objects. Both plate jackets and plates with ink annotations, such as the ones in Figure 6, are photographed with a Nikon D200 camera. All ink annotations on the reverse side of the plate from the emulsion must then be cleaned to avoid confusion with astronomical objects.

2.2. Mosaic Generation, WCS fitting, and Source Extraction

The digitizer generates sixty 45 mm square tiles in a 10 x 6 tile pattern to cover at typical 20 x 25 cm plate at least twice for every object, because of the half-steps along the plate width. The scanning process also generates flat field exposures of blank plates. Instead of a single dark frame, however, the procedure takes frames of a blank plate at 16 exposure levels to characterize nonlinearities in each of the four CCD readout amplifiers. The mosaicing process registers and combines these images to produce a single mosaic of approximately 780 megapixels. In addition, the process generates a 1/16 size thumbnail mosaic which is also used in preliminary stages of WCS fitting.

The SExtractor program (Bertin & Arnouts 1996) next generates object lists from both the mosaic and the thumbnail. The WCS fitting procedure first begins with astrometry.net (Lang et al. 2010) on the thumbnail and moves to successively accurate fits using WCStools (Mink 1999). Finally, a polynomial fit removes lens distortions from the original telescope. A companion paper (Servillat 2010) describes this procedure in greater detail.

2.3. Pickering Wedge Filter

The Pickering Wedge Filter is used to flag ghost objects by performing a spatial correlation within a limited region around bright stars. The procedure requires that these objects be identified on at least one plate in a telescope series. Such identification is usually made when a plate has an excessive number of objects not matched to the calibration catalog. To date, table 1 lists the known series which used Pickering Wedges, or in the case of the "dsy" telescope series, had ghost objects possibly caused by misaligned optics. The table also contains a representative astrometric error for the series normally obtained from statistical studies of the WCS fitting procedures mentioned above.

The procedural steps are as follows:

1. Select the 300 brightest stars on the plate that are not too close to the plate edges and do not have a full-width half maximum that is greater than the expected Pickering Wedge object separation. An object too close to the sides of the plate is within 2.5% of the total plate width from the edge. An object too close to the top and bottom of the plate is within 2.5% of the total plate height from the edge.

2. For each of these stars, find all of the stars in a ring of the radius of table 1 and a width of twice the astrometry error of table 1. Figure 12 is a map of all of these rings superimposed on each other. Note the clump of objects on the left side of the annulus.

3. This annulus is next divided into annular bins with a width along the circumference approximately equal to three times the astrometry error. The bin angle is defined as 0 degrees along the Sextractor X axis, and increases counter clockwise to 90 degrees along the Sextractor Y axis. Figure 13 shows that there is a clearly defined peak near 180 degrees and this peak corresponds with a dense spot in Figure 12. If there is more than one peak, the plate is likely a grating plate. These grating plates are currently rejected by the Pickering Wedge procedure.

4. Find the average and standard deviation of the star count in all of these bins. For each bin, calculate an excess value which is $Excess = (star\_count\_in\_bin) - (bin\_average) - 3 * (bin\_rms)$ If this excess value is positive for any bin, then the plate is likely to be a Pickering Wedge plate.

5. If a Pickering Wedge plate identification has been made, there is a need to make a more precise determination of the location of the wedge object within the annular bins. Start be defining a rectangular area that encompasses the bin with the peak star count and the two adjacent bins. For all stars within this area, plot the magnitude difference between the primary and secondary objects as a function of primary magnitude. The result, which should look like Figure 14 has two populations of objects, an upward sloping population of Pickering Wedge objects and a downward sloping population of field stars and grain noise.

   To get an accurate astrometric position, take the sum of the two magnitudes and sort the table of candidates based on this sum. (For the mc00380 case, the candidates with the highest value of the sum will be in the lower left at approximately MAG_ISO(pri) = -17). For the top 10-1000 candidates, calculate the average positional deviation as a function of the number of candidates, and take the new position of the wedge object to be the average position for the number of candidates at which the positional deviation is a minimum. Repeat the process casting out any outliers greater than one times the average positional deviation. The final collection of stars define a point ((aveMAG_ISO,aveMagDiff) which is (-17,3.5) on the plot of figure 14) at the average values of MAG_ISO for the bright star and the difference in MAG_ISO for both stars.

Assuming a positive slope, define the population of Pickering Wedge objects by the following linear equation:

magdiff = K + MAG_ISO

where K is an arbitrary constant which is determined by including the above average point on the line:

aveMagDiff = K + aveMAG_ISO

and assuming a negative slope, the population of field stars by the following equation:

magdiff = MAG_ISO_med - MAG_ISO

where, because the number of field stars dominates the distribution, MAG_ISO_med can be found as the one-sigma clipped median of the entire population.

Define a critical magnitude, MAG_ISO_crit as the intersection of these two lines. (Both straight lines are shown in figure 14). Because of the nonlinearity of the Pickering Wedge Population, add a margin of 1.0 magnitude to this critical magnitude.

6. Go through the entire Sextractor population and select as Pickering Wedge objects all objects which meet the following criteria:

    - The positional deviation of the object from the average position calculated above is less then three times the clipped average positional deviation calculated above.
    - The magnitude of the bright object is brighter than the critical magnitude calculated above including the 1.0 magnitude margin.
    - The magnitude of the dim object is brighter than MAG_ISO_med calculated above plus the rms magnitude of the entire population after one-sigma clipping.

    In Figure 14 these selected wedge objects are shown as crosses and ordinary star objects are shown as dots. The result for mc00380 shown in in Figure 15 shows that the Pickering Wedge objects do not coincide, most probably because of differences in optical path through the original telescope. The algorithm is pessimistic because it is better to falsely identify a non-Pickering Wedge object than to have an unidentified Pickering Wedge object.

2.4. Star Matching, Defect Filter, Photometry Calibration, and Magnitude Calculation

These steps are described in more detail in Laycock et al. (2010), Tang et al. (2010), Tang, S. (in preparation), and on the DASCH website[2]. After wedge filtering, objects are matched either to the GSC2.3.2 catalog (Lasker et al. 2009) or the Kepler Input

[2]see http://hea-www.harvard.edu/DASCH/photometry.php

Catalog[3] (for calibration of the Kepler satellite field). A defect filter removes emulsion defects, dust, and development defects by comparing PSF characteristics of matched objects with unmatched objects. Next, the plate is divided into nine annular bins and a colorterm algorithm estimates the spectral response of the plate emulsion together with any filters that may have been used. For each annular bin, a lowess curvefitting algorithm is used to generate a calibration between the Sextractor instrumental magnitudes and the blue catalog magnitudes. Finally, a local correction algorithm is used on a 50 x 50 grid to account for variations in emulsion and/or sky conditions. All of the objects on the plate are assigned magnitudes using the lowess calibration curve and the local bin corrections.

2.5. Multiple Exposure Loop

The concept behind multiple exposure processing is straightforward: remove the objects successfully matched to the calibration catalog from the SExtractor source list and submit the remainder source list to astrometry.net. At the completion of the WCS fitting procedures, the new WCS parameters are applied to all objects in the original SExtractor dataset and the pipeline proceeds normally from Pickering Wedge filtering to magnitude calculation. In keeping track of the multiple datasets, it is important to distinguish the “Exposure Number” from the “Solution Number”. The “Exposure Number” is the order of the time-stamped exposures as they appear in the observatory logbooks. The “Solution Number” is the order in which the astrometry.net finds individual WCS solutions and this order is not necessarily related to the order of exposures in the logbooks. In order to reconcile these two orders, it is necessary to consider every possible combination of Exposure Number and Solution Number and choose the combination which minimizes the sum of the scalar distances between the logbook plate centers and the astrometry.wcs plate centers. As more solutions are processed, the assigned Exposure Numbers for earlier Solution Numbers may change. In such a case, the photometry results of the earlier solutions must be recalculated because the newly assigned exposure date changes the zenith distance of the object and the estimated extinction correction applied to the catalog magnitude.

Two ambiguities arise: If a catalog object from one solution coincides with a different catalog object from another solution, then the result is considered a “Multiple Exposure Blend”. Since GSC2.3.2 galaxies do not have magnitudes, any multiple exposure blend with a galaxy is always flagged. However, if the catalog magnitudes show that all of the stars within a multiple exposure blend do not change the brightness of the brightest star by 0.1 magnitude, then the blend is ignored.

If an object is not matched to any catalog object, then it could be an asteroid or a stellar flare. However, there is a fundamental uncertainty with the date of this object because the actual solution number to which the object belongs is unknown. These objects receive a special flag indicating this date uncertainty.

Two sanity checks are necessary to prevent infinite loops. These sanity checks also detect unrealistic answers occasionally produced by the sixth-order polynomial fitting used by IRAF/ccmap and SCAMP; and the multivariate amoeba fitting used by wcstools. First, if the plate center of newly found WCS solution is within 40 pixels of a previous of a previously found WCS solution, then the solution is rejected. Second, if

[3]see http://www.cfa.harvard.edu/kepler/kic/format/format.html

a Multiple Exposure Blend is detected with the same catalog object, then the solution is also rejected because all or part of it overlaps with a previous solution.

2.6. Photometry Database

The output of the pipeline is a set of Starbase (Roll 1996) tables which are not suitable for rapid access of individual lightcurves for display and statistical analysis. Since traditional relational database models may not meet the needs of scientific studies (Gray 2005) (Ailamaki et al. 2010), magnitude measurements are stored in a set of binary files optimized for performance. Supporting star-specific and plate-specific data appear in MySQL tables. A detailed database design document will appear in a future publication.

3. Results and Discussion

The techniques described in this paper are probabilistic in nature: they can not guarantee detection of every WCS solution and might provide false WCS solutions as well. It is necessary to make some estimate of failure and false positive rates.

The combination of astrometry.net and Doug Mink's wcstools has been remarkably reliable in finding the first WCS solution of a photograph plate. Currently there are only 26 unsolved plates out of a sample of 10572 plates, a 99.75% success rate. An estimate of the false positive rate may be made by studying the distribution of calculated plate scales for the X and Y axes and the distribution of angle between the X and Y axes. For 10529 plates, 352 or 3.3% show at least one of these three parameters exceeding 3 sigma. Since the astrometry.net authors have demonstrated that their false positive rate should be neglegible, we believe that most of these 352 outliers show issues with multivariate-amoeba fitting used by wcstools and the sixth-order polynomial fitting used by IRAF/ccmap. For telescopes where sufficient plates have been scanned to determine the plate scale, an additional check using these three criteria is possible for detection of poor WCS fits.

The multiple exposure loop correctly detected all three exposures in figure 1. The very best performance occurred with detection of all 6 exposures of plate ma04979. The algorithm also detected 3 of the 9 exposures in figure 3 and 6 of the 9 exposures in a similar plate, mc12688, before the objects became too dim for further progress. However, in figure 2, the algorithm correctly identified only the pole exposure and one of the two Cygnus exposures.

The multiple exposure loop did not work well with the grating plate in figure 5. While astrometry.net correctly identified one set of first order objects, the wcstools procedure returned to the original zero-order solution. There is a need to remove not only the zero-order objects but also half of the first order objects.

The multiple exposure loop provided additional quality control checks by detecting at least 244 plates in which the second solution was within 40 pixels of the first solution, and 48 plates where the second solution overlapped the first solution. These plates cannot currently provide good photometry data because they are either grating plates or plates with bad astrometrical solutions.

A more quantitative assessment is difficult because transcription of full logbook entries did not occur until approximately 800 plates had been scanned and because the selection of multiple exposure plates for scanning had been avoided prior to the completion of the multiple exposure loop code. As a result, there is only a small sample 87 multiple exposure plates scanned with full logbook transcriptions. Results from this sample show successful detection of 31% of the double exposures and 18% of the triple exposures. On the other hand, there were 481 non-Pickering-Wedge plates which showed more solutions than described in the transcribed logbook entries. There is a need for further examination of the plates with missed exposures and more complete transcriptions of multiple-exposure logbook entries.

The Pickering Wedge filter flagged 857 of 11461 scanned plates. It is not possible to assess the completeness of this sample, because use of the Pickering Wedge filter has not been documented well in the logbooks. Use of this filter has been extended to detect a possible optics misalignment in the "dsy" telescope. These results suggest that a general near-neighborhood search algorithm would be useful for better detection of grating objects and double-peaked star PSF's.

In summary, the methods described in this paper represent useful approaches to the analysis of multiple-exposure plates.

Acknowledgments. This work is funded by National Science Foundation grants AST-0407380 and AST-0909073. Additional contributers to the DASCH project are the Harvard College Plate Stacks Curator, Alison Doane; hardware engineer, Robert Simcoe; and our transcribers, plate cleaners, and scanners.

Table 1. Known Pickering Wedge Telescope Series

Series	Separation arcsec	Error arcsec
mc	143.0	2.2
b	270.0	8.9
dsy	116.4	3.2
i	287.0	10.8
i	276.0	10.8

References

Ailamaki, A., Kanetre, V., & Dash, D. 2010, Communications of the ACM, 53, 68
Bertin, E., & Arnouts, S. 1996, Astronomy and Astrophysics Supplement, 317, 393
Gray, J. 2005, QUEUE, 3, 31
Grindlay, J., Tang, S., Simcoe, R., Laycock, S., Los, E., Mink, D., Doane, A., & Champine, G. 2009, in Astronomical Society of the Pacific Conference Series, edited by W. Osborn & L. Robbins, vol. 410 of Astronomical Society of the Pacific Conference Series, 101
King, E. S. 1931, A Manual of Celestial Photography (Boston: Eastern Science Supply Co.), 1st ed.
Lang, D., Hogg, D. W., Mierle, K., Blanton, M., & Roweis, S. 2010, AJ, 139, 1782. 0910.2233
Lasker, B. M., Lattanzi, M. G., McLean, B. J., Bucciarelli, B., Drimmel, R., Garcia, J., Greene, G., Guglielmetti, F., Hanley, C., Hawkins, G., Laidler, V. G., Loomis, C., Meakes, M., Mignani, R., Morbidelli, R., Morrison, J., Pannunzio, R., Rosenberg, A., Sarasso, M., Smart, R. L., Spagna, A., Sturch, C. R., Volpicelli, A., White, R. L., Wolfe, D., & Zacchei, A. 2009, AJ, 136, 735
Laycock, S., Tang, S., Grindlay, J., Los, E., Simcoe, R., & Mink, D. 2010, AJ, 140
Leavitt, H. S. 1917, Annals of Harvard College Observatory, 71, 106

Figure 1. Plate i52750 showing a triple exposure of M44. North is to the left.

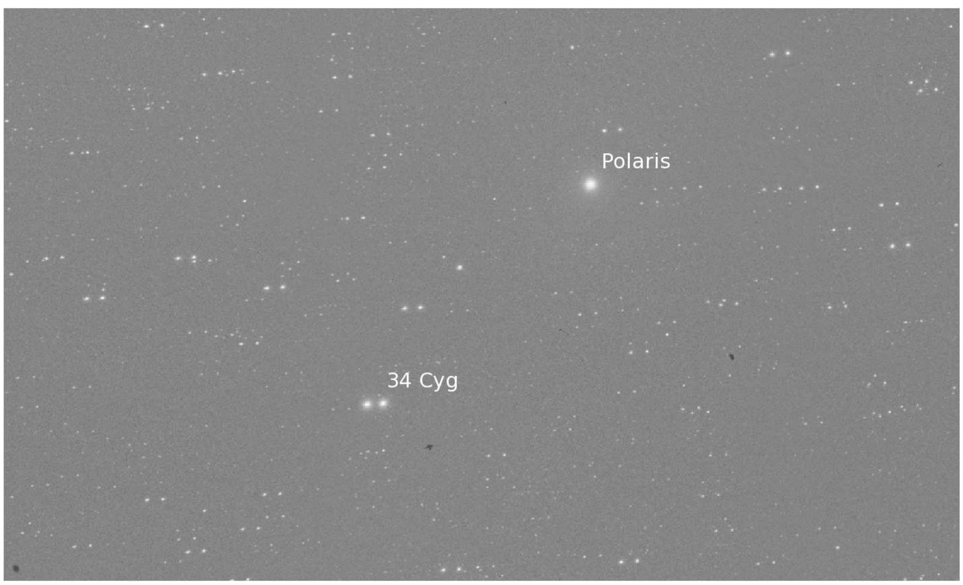

Figure 2. Plate i48332 showing a triple exposure: A single of Polaris and a double of a Cygnus starfield.

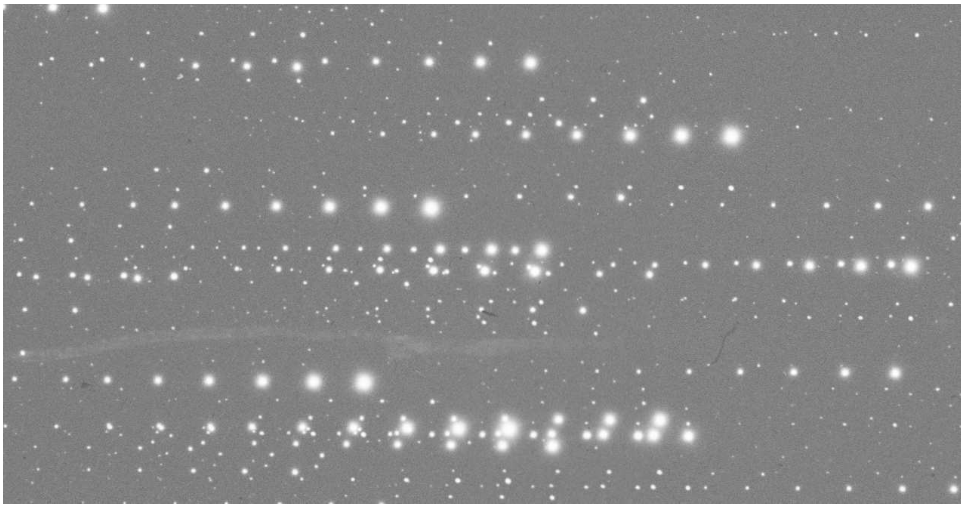

Figure 3. Plate mc05077 showing multiple exposures of the M44 field. Nine exposures were taken with exposure times decreasing by 50% for each successive exposure.

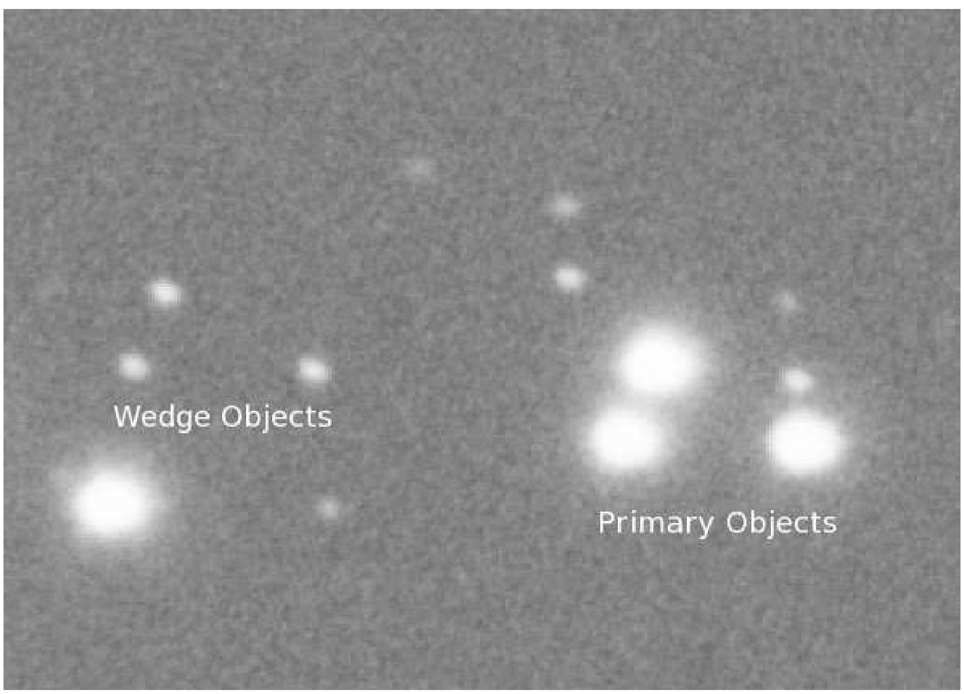

Figure 4. Plate i31090 showing examples of Pickering Wedge objects

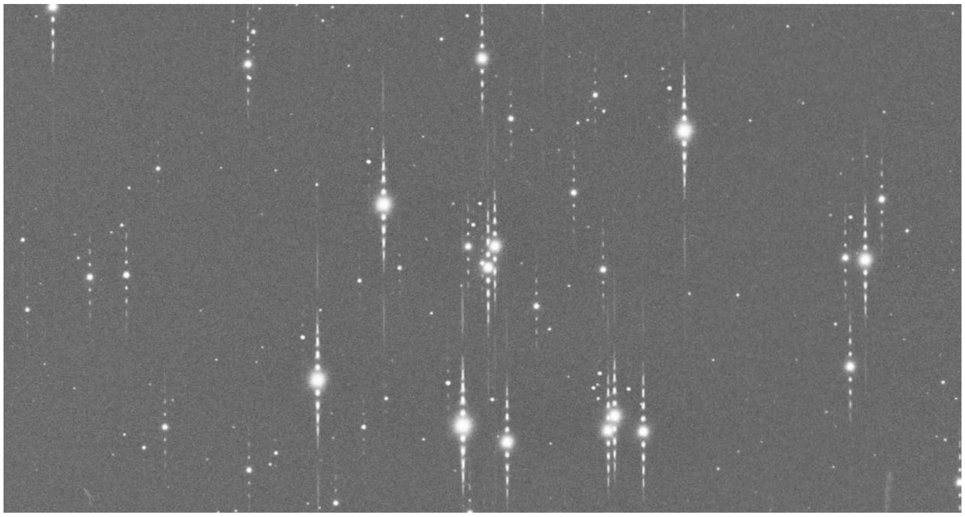

Figure 5. Plate mc08175 showing the use of a coarse objective grating.

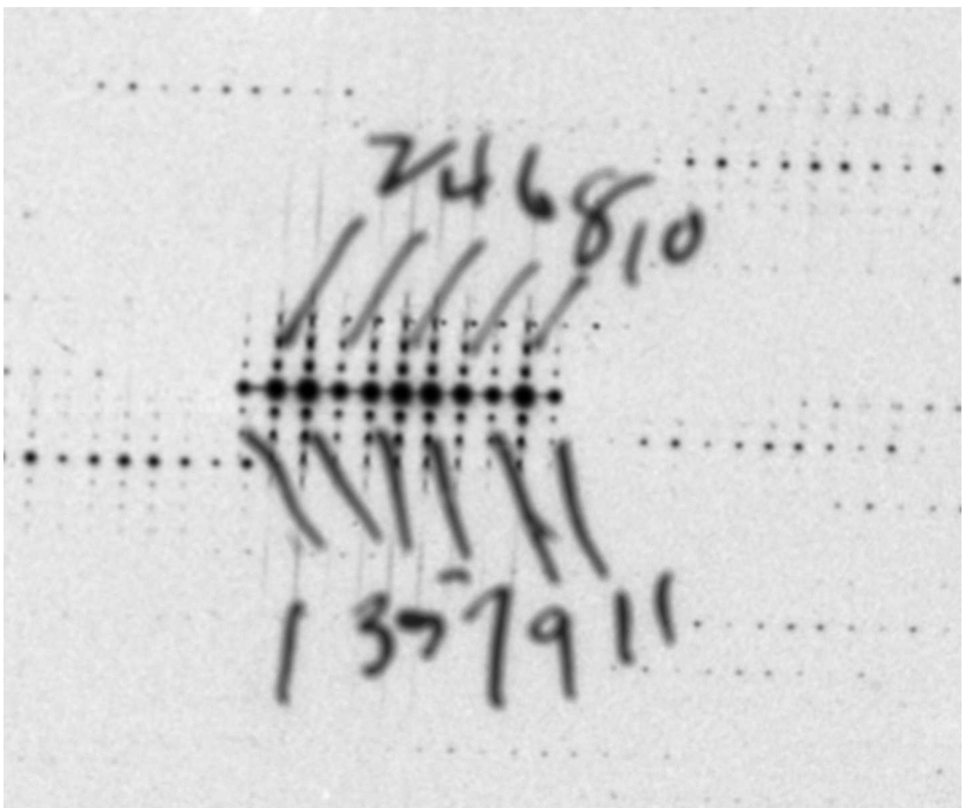

Figure 6. Plate mf00274 showing the combination use of mutiple exposures and a coarse objective grating

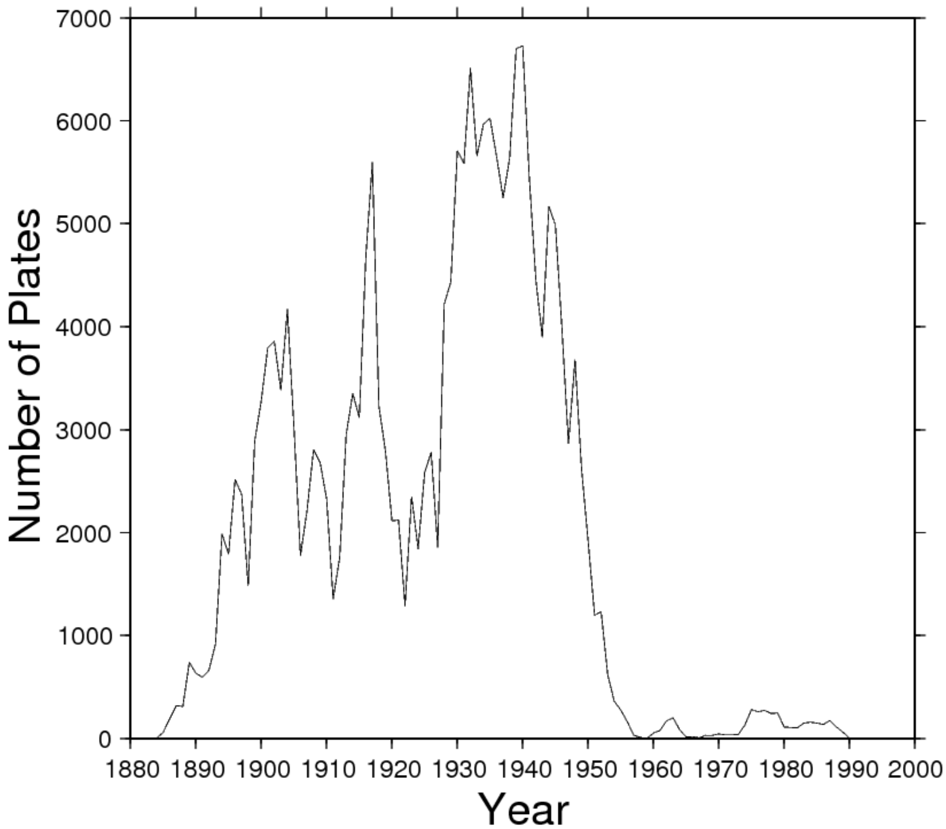

Figure 7. The number of transcribed plates per year.

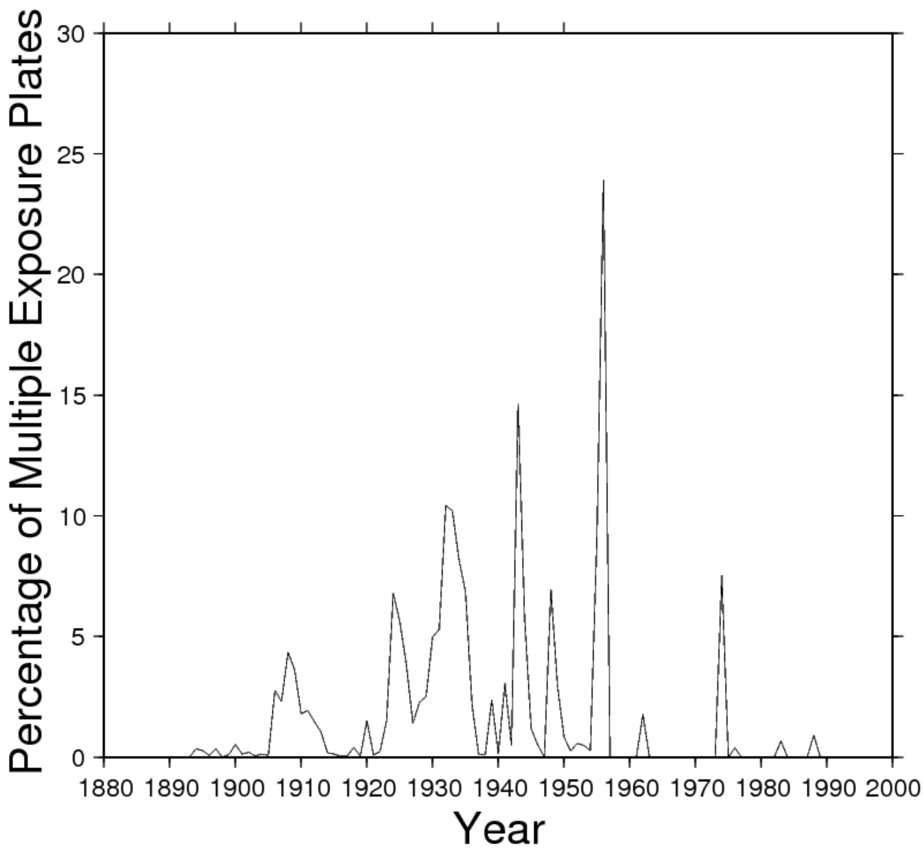

Figure 8. Percentage of transcribed multiple exposure plates per year.

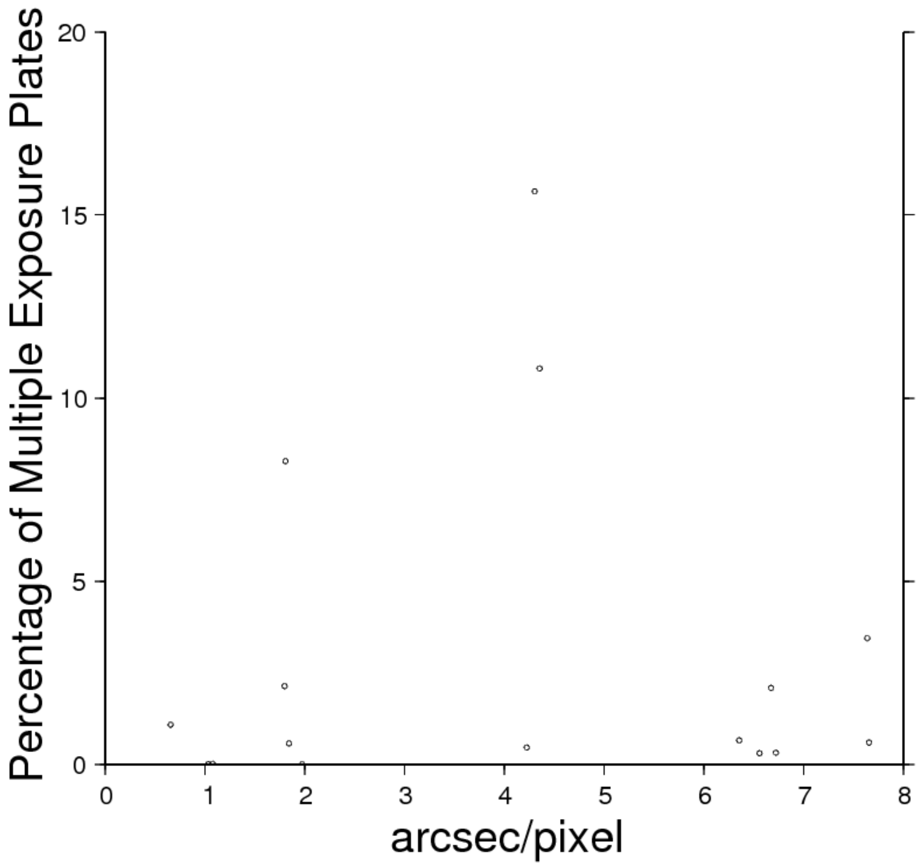

Figure 9. Percentage of multiple exposure plates as a function of telescope plate scale for transcribed series having at least 1000 plates

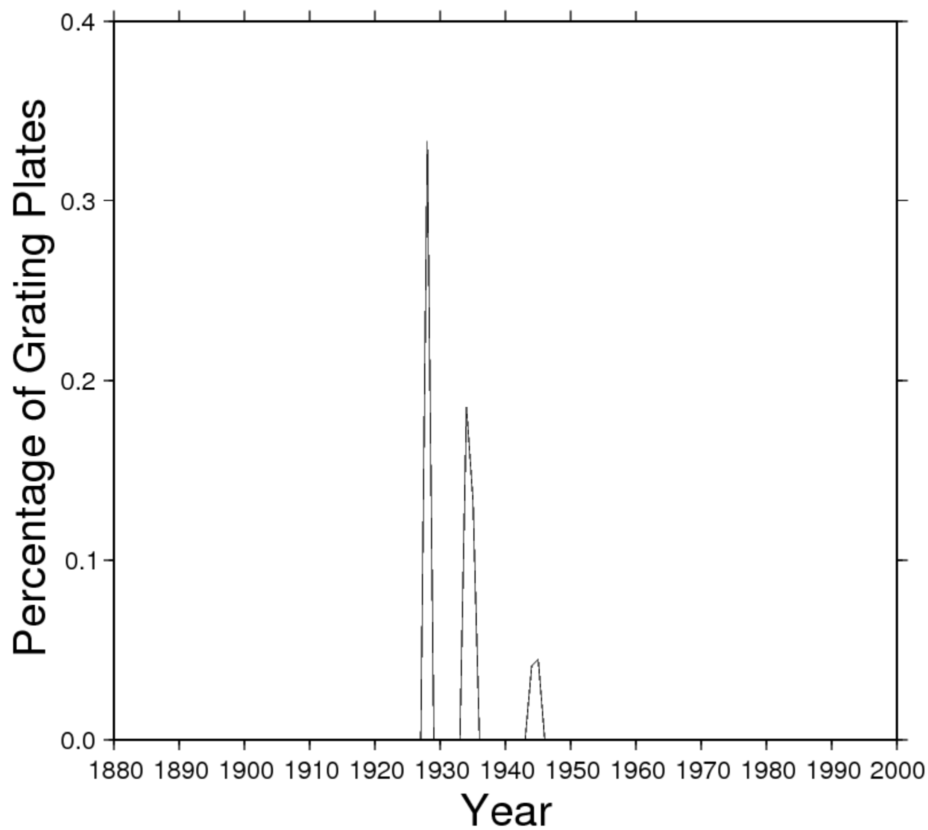

Figure 10. Percentage of grating plates per year

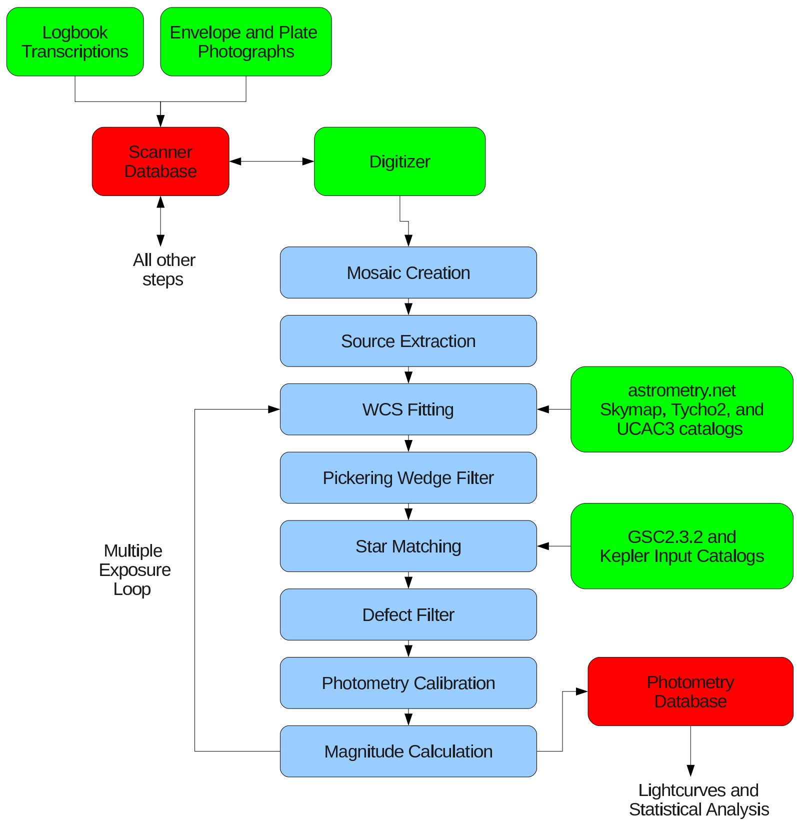

Figure 11. The DASCH object processing pipeline.

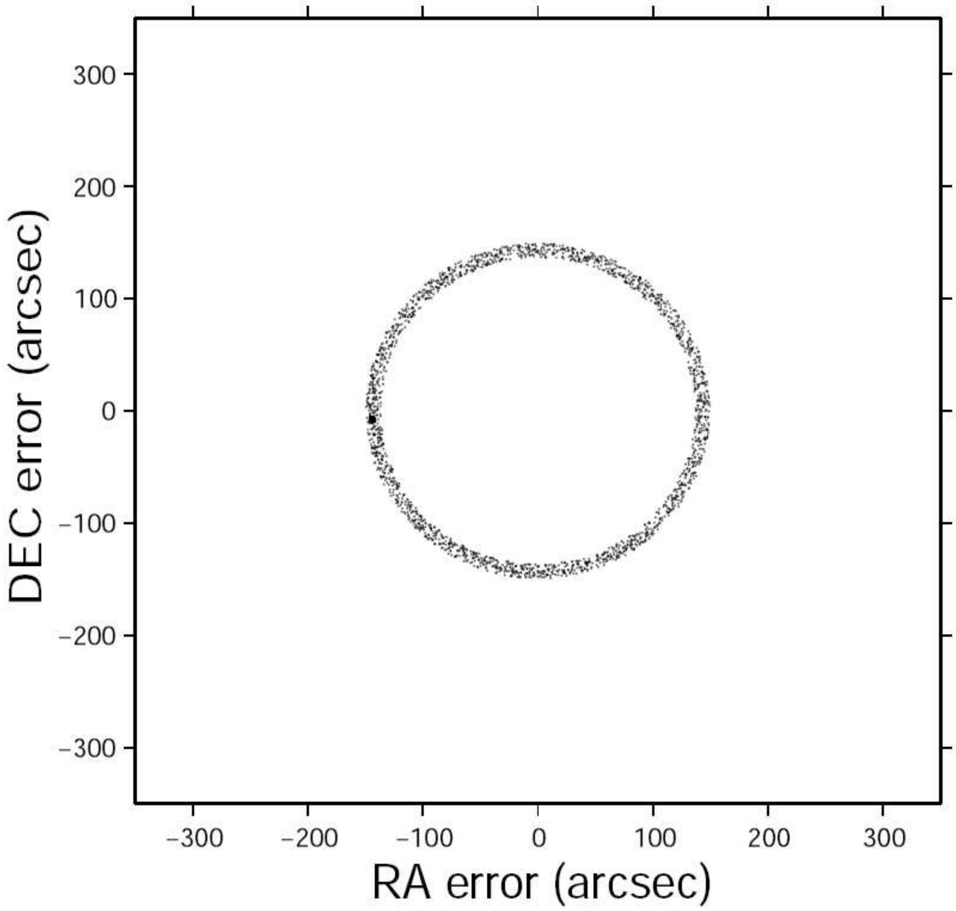

Figure 12. Pickering Wedge initial search annulus for mc00380

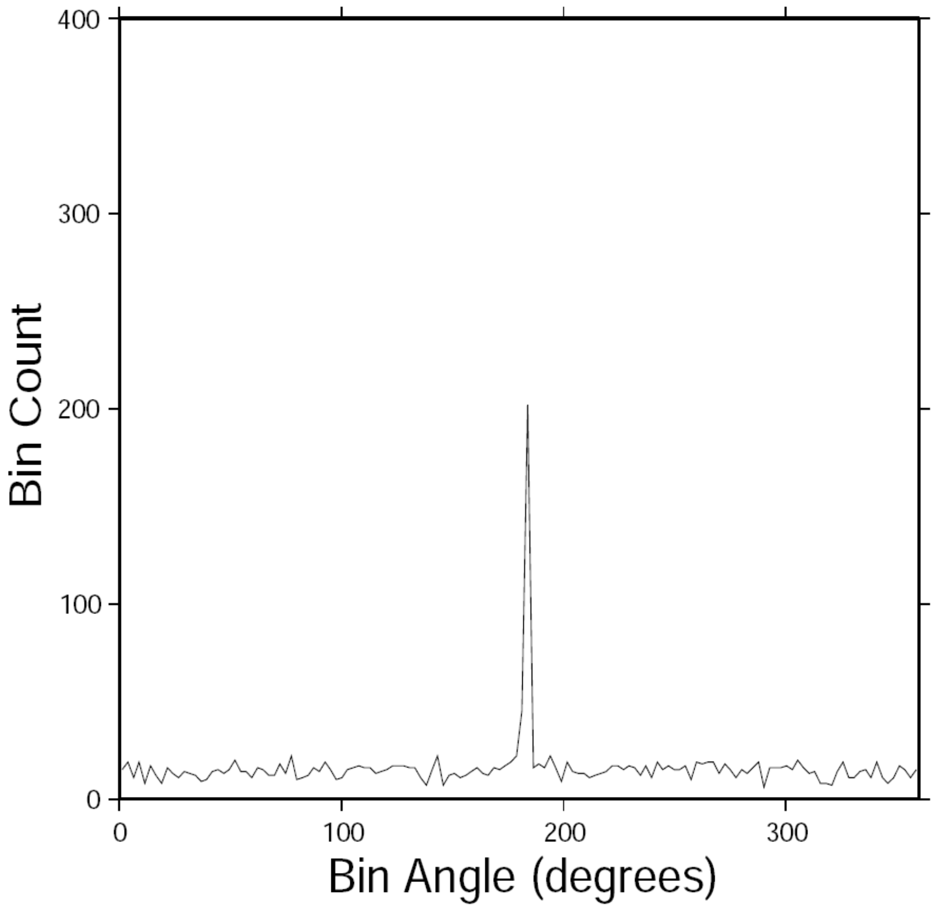

Figure 13. Pickering Wedge object counts vs angle for mc00380.

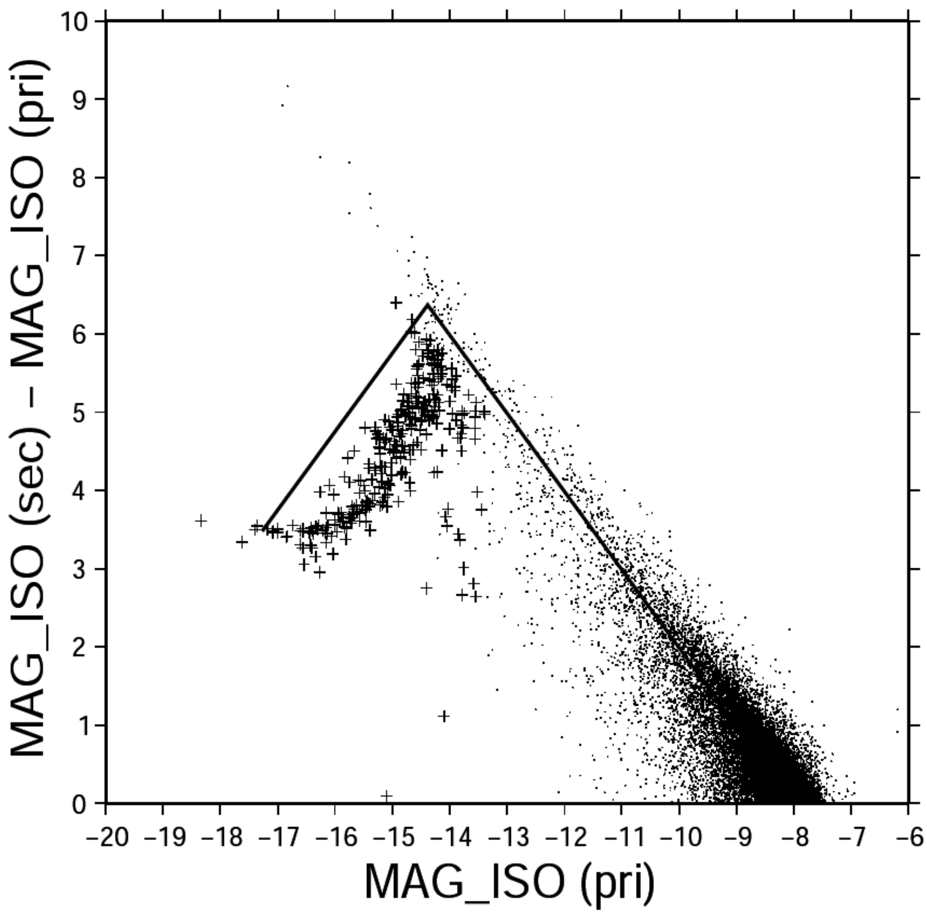

Figure 14. Pickering Wedge objects in instrumental brightness (MAG_ISO) space for mc00380

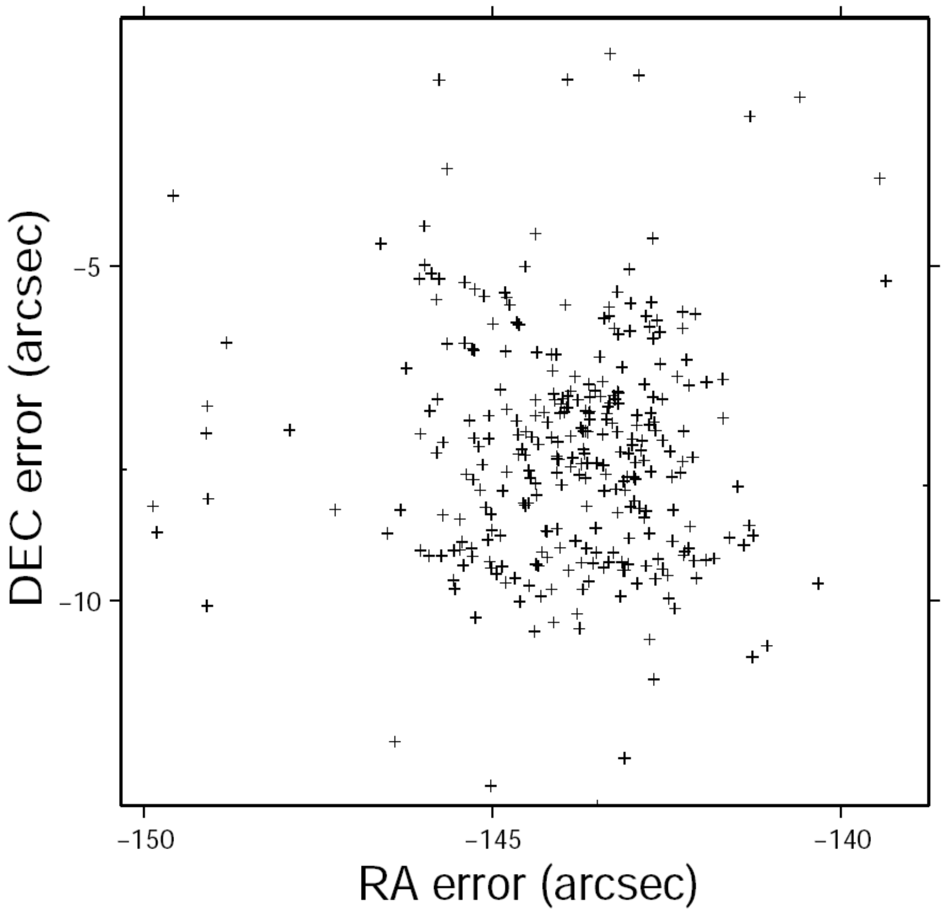

Figure 15. Map of flagged Pickering Wedge Objects for mc00380.

Mink, D. 1999, in Astronomical Data Analysis Software and Systems VIII, edited by D. Mehringer, R. Plante, & D. Roberts (Seattle: Astronomical Society of the Pacific), vol. 8 of A.S.P. Conference Series, 498

Roll, J. 1996, in Astronomical Data Analysis Software and Systems V, edited by G. H. Jacoby, & J. Barnes (Seattle: Astronomical Society of the Pacific), vol. 5 of A.S.P. Conference Series, 536

Servillat, M. 2010, in Astronomical Data Analysis Software and Systems XX, edited by A. Evans, I.and Accomazzi, D. Mink, & A. Rots (Seattle: Astronomical Society of the Pacific), vol. 20 of A.S.P. Conference Series

Simcoe, R. J., Grindlay, J., Los, E. J., Doane, A., Laycock, S. G., Mink, D. J., Champine, G., & A., S. 2006, in Applications of Digital Image Processing XXIX, edited by A. Tescher, vol. 6312 of Proceedings of the SPIE

Tang, S., Grindlay, J., Los, E., & Laycock, S. 2010, ApJ Letters, 710, 77