Progress Report for Adapting APASS Data Releases for the Calibration of Harvard Plates

Edward J. Los
Harvard College Observatory, 60 Garden Street, Cambridge, MA 02138

Presented at the 100th Annual Meeting of the AAVSO, October 8, 2011; received November 21, 2011; revised January 31, 2012; accepted January 31, 2012

Abstract The Digital Access to a Sky Century @ Harvard (DASCH) has scanned over 19,000 plates and developed a pipeline to calibrate these plates using existing photometric catalogues. This paper presents preliminary results from the use of the AAVSO Photometric All-Sky Survey (APASS) catalogue releases DR1, DR2, and DR3 for DASCH plate calibration. In the optimum magnitude 10–12 range of the DASCH patrol plates, the median light curve RMS with APASS calibration is 0.10–0.12 magnitude, an improvement from the 0.1–0.15 magnitude median light curve RMS with GSC 2.2.3 calibration.

1. Introduction

The Harvard College Observatory plate collection consists of approximately 525,000 photographs produced by over eighty telescopes spanning over 100 years from about 1885 to 1992. The goal of the Digital Access to a Sky Century @ Harvard project (DASCH; Grindlay 2009; see http://hea-www.harvard.edu/DASCH/) 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 18,000 plates and and used the SExtractor program (Bertin and Arnouts 1996) to extract an average of 90,000 objects per plate.

One of the primary goals of this project is to extract photometric data from these plates. The DASCH pipeline presented in Laycock *et al.* (2010) creates a calibration curve for each plate by using robust locally weighted scatterplot smoothing to match SExtractor isophotal magnitudes to magnitudes in a standard catalogue. The choice of a standard catalogue has been a continuing project issue because the catalogue must be all-sky and must match the magnitude 4–17 range of the plate collection. We are currently using the *Guide Star Catalog* version 2.3.2 (GSC 2.3.2; Lasker *et al.* 2008). The relatively poor photometric accuracy of the *Guide Star Catalog* is balanced by a good overlap in magnitude range with the Harvard plate collection and all-sky coverage. However, DASCH team member Sumin Tang reports that light curve searches using the *Kepler Input Catalog* (KIC; Brown 2011) for calibration showed less error contamination. This paper introduces a new metric, the normalized outlier count, in an attempt to quantify the relative value of calibration catalogues with respect to such light curve searches.

Fortunately, the American Association of Variable Star Observers (AAVSO) has similar requirements for a bright star photometric catalogue and we welcome the initiative of AAVSO to produce the AAVSO Photometric All-Sky Survey (APASS; AAVSO 2009; see http://www.aavso.org/apass). We appreciate the willingness of the AAVSO to provide preliminary releases of their data and recognize that all results reported in this paper are preliminary until the final release of the APASS catalogue. Catalogue calibration of the APASS data for 18,308 plates takes 1.7 days on our current cluster configuration. Database entry is single threaded and takes an additional 3.3 days for the 16,640 plates that produce valid data with the APASS DR3 release. Because of these long processing times, the continued scanning of new plates, and the evolution of the pipeline algorithms, the design of strictly controlled experiments is not easily achieved.

The Harvard plate collection has full sky coverage, but we have been concentrating on the regions of the sky shown in Table 1. As a result, the current DASCH sky coverage from scanned plates is shown in Figure 1.

The first APASS data release (DR1) covered primarily regions North of +40 degrees declination and matched only the Northern part of our Kepler field for which we already had calibration magnitudes available from the *Kepler Input Catalog*. The second APASS data release (DR2) (Figure 2) provided a good match with our LMC field for which we had only GSC 2.3.2 calibration magnitudes. We began our tests with the APASS DR1 and DR2 releases and then migrated to the APASS Data Release 3 (DR3) as soon as it became available in August 2011. This paper discusses our experience with the DR1 and DR2 releases and the changes implemented with the DR3 release.

2. APASS DR1 and DR2 releases

For maximum performance, the calibration catalogues are reformatted into compact binary files sorted by "gsc_bin_index" which is a simple spatial hash index consisting of 1/64 degree declination bands beginning at the celestial South pole. Within each declination band, stars are sorted by right ascension into an integral number of bins approximately 1/64 degree wide. Although there is no closed form mathematical solution for the bin layout, a lookup table provides quick conversion of RA and Declination to one of the 168,966,386 gsc_bin_index values.

Figure 3 shows the steps for reformatting the calibration catalogue. Since APASS objects currently have no catalogue identification number, these objects are given a designation similar to the SDSS designations: "APASS Jhhmmss.s+ddmmss" where the J2000 Right Ascension is hh:mm:ss.ss and and the declination is +dd:mm:ss in sexagesimal format. The DASCH pipeline requires two magnitudes, a primary magnitude close to the blue spectral range where the

Harvard plates are most sensitive, and a secondary magnitude used to calculate plate-specific color corrections. With the GSC 2.3.2 catalogue, DASCH uses photographic blue and red (IIIaJ and IIIaF) magnitudes and for the KIC, SDSS g and r magnitudes. For DR1 and DR2 releases, the current study selected the APASS g and r magnitudes.

It is necessary to filter the entries for accuracy. Previous experience with KIC calibration of the DASCH plates show that the best accuracy produced by the DASCH photometric pipeline is about 0.1 magnitude RMS. Therefore, the combined APASS catalogue uses only observations for which more than one measurement is available for the SDSS g and r magnitudes and the magnitude error of each measurement is less than 0.1 magnitude. (Because of a software bug, the preferred filtering on color errors added in quadrature was not implemented in the DR1/DR2 phase of this study.) We also use a color check to filter out g-r colors which are not in the range of –1.0 and 2.5 mag. Of the 8,702,597 entries on the DR1 and DR2 releases, there are 7,173,325 entries between magnitudes 7 and 19 which meet the above criteria.

Because APASS DR1 and DR2 have overlapping fields on the celestial equator, the question arises as to whether the combined catalogue contains any duplicate objects. Matching the combined DR1 and DR2 catalogue against itself produced 79,173 matches, the majority of which were within the stated accuracy of 0.25 magnitude and 2.5 arcsec pixel size of the APASS telescopes. Before writing out the contents of a gsc_bin_index, a search for matches within a 1.0 arcsec radius accepts only the matching entry that has the lowest RMS of the magnitude accuracies added in quadrature. For the DR1 and DR2 combined catalogue, a total of 68,510 entries were rejected.

Because the earliest DASCH scanned plate to date was exposed on January 24, 1886, there is a need to correct all catalogue positions for proper motion. The final step is to match the combined the APASS catalogue with the UCAC3 catalogue (Zacharias 2010) and add the UCAC3 proper motions to the APASS data. Using the same search algorithm described above, there were 6,231,923 UCAC3 matches for the 7,104,815 APASS entries.

3. APASS DR1 and DR2 results

Figure 4 shows that the coverage of the APASS-calibrated stars is most complete in the LMC region and the Northern half of the Kepler region. There was also some coverage near the equator for the M44 and 3c273 regions. Table 2 shows that 68% of the plates which yielded data for the GSC 2.3.2 catalogue also yielded data for the APASS catalogue, and about 25% of the stars that provided good data with GSC 2.3.2 also provided good data with APASS.

Three important metrics of DASCH photometry quality are derived from light curves which have at least ten points and pass all of the DASCH quality checks. Figure 5 shows the DASCH pipeline yield as a function of magnitude for all of the calibration catalogues. A major deficiency with the DR1 and DR2 combined release is its ninth magnitude limit for bright stars. Figure 6 shows the median RMS of DASCH light curves as a function of magnitude. By plotting the median RMS, this figure excludes variable stars. The best accuracy of DASCH light curves comes in two ranges, magnitude 9 to 12 and magnitude 15 to 17, because of the different types of telescopes used to generate the Harvard plate collection. Most of the plates that we scan are patrol plates with small aperture objectives which have a limiting magnitude of 13 to 14. Magnitudes dimmer than 15 are covered by a smaller collection of plates produced by telescopes with larger objectives. For objects dimmer than 11th magnitude, there is little differentiation among the various catalogues, suggesting that errors in the plates dominate. Brighter than 11th magnitude, the combined DR1 and DR2 catalogue is comparable to the KIC and noticeably better than GSC 2.3.2. The third metric shown in Figure 7 reflects the use of of the DASCH database to search for new variable stars. For this search, Sumin Tang (2012) defined a series of light curve parameters which show promise in identifying flares and other unusual variability. One such parameter is "nburst3" which is the number of points in a light curve that are at least 0.4 magnitude brighter than the median magnitude of the light curve. Figure 7 shows a "normalized nburst3" which is the sum of nburst3 for all light curves that have valid values of nburst3 divided by the number of light curves that have valid values of nburst3. For stars dimmer than magnitude 12 the three catalogues produce comparable results, although the APASS catalogue produces the highest outlier rate around magnitude 14. For stars brighter than magnitude 12, the APASS calibration is significantly better. These results should be taken with some caution: A problem with the "normalized nburst3" metric is that it does not reproduce Sumin Tang's qualitative assessment that the KIC produces fewer false outliers than the GSC 2.3.2 catalogue.

4. APASS DR3 release

Figure 8 shows that release of the APASS DR3 dataset in August 2011 provides expanded coverage in the Southern celestial hemisphere to include more of the DASCH M44, 3C273, and LMC fields. Since this release incorporates the DR1 and DR2 releases, there was no need to combine DR3 with the previous releases. This new release provided an opportunity to make four changes in the reformating of the APASS catalogue shown in Figure 9. First, the search radii for duplicates was expanded from 1 to 2 arcsec. Second, objects rejected for photometry use are now retained in the calibration catalogue to avoid confusion in searching for novae and other uncatalogued objects. These rejected objects are flagged as "variable" so that the pipeline code does not use them for plate calibration, but does assign a DASCH magnitude from the calibration data. Third, the DR3 release was merged with the *Tycho-2* catalogue (Høg 2000) as described below to improve coverage in the magnitude 4 to 9 range. Fourth, a bug was corrected in which not only individual magnitude uncertainties greater than 0.1 magnitude were rejected, but the sum of these uncertainties added in quadrature were rejected. Correction of this bug, however, did not occur until the final processing run involving B and V Johnson magnitudes.

The *Tycho-2* catalogue was first filtered to reject B and V RMS values greater than 0.1 magnitude. While this step rejected nearly 60% of the stars, Figure 10 shows that most of the stars rejected were dimmer than 10th magnitude and would have been replaced by APASS stars.

The next step was to transform the *Tycho-2* magnitudes into the SDSS color system. Because the KIC also incorporates Tycho-2 stars, we used equations 6a and 6b in Brown *et al.* (2011) which are repeated below:

$$g = 0.54B + 0.46V - 0.07 \quad (1)$$

$$r = -0.44B + 1.44V + 0.12 \quad (2)$$

Since there was time for another photometry processing iteration, this iteration used the Johnson B and V data available with the APASS release. Høg (2000) says that the *Tycho-2* magnitudes are "close to Johnson B and V" and recommends that transformations not be used because these transformations are dependent on the luminosity class and reddening of each star. Since there are 55,718 direct matches between *Tycho-2* and APASS positions, these matches can be used to derive a transformation. However, the resulting transformation shows a standard deviation of 0.45 magnitude and is valid over only a 1- to 2-magnitude range. Consequently we followed the recommendation of Høg (2000) and used no transformation of the *Tycho-2* B–V data.

In combining *Tycho-2* with APASS DR3, there were 1,103,581 duplicates in 21,418,221 records. Preference was given to the object with the lowest magnitude error and to APASS stars as shown in Table 3.

The final step of matching the combined catalogue with the UCAC3 catalogue to obtain proper motions proceeded as described above. For the 20,434,140 objects in the combined SDSS g-r catalogue, we found UCAC3 proper motions for 16,671,675 stars. For the 20,314,720 objects in the combined Johnson B–V catalogue, we found UCAC3 proper motions for 16,580,857 stars.

Because refinements in object positions between DR2 and DR3 may cause the APASS designation to change, it is important to flush the photometry database of all references to the earlier APASS catalogue. This flushing is accomplished by a version ID which is incremented whenever new results for the same plate are inserted into the photometry database.

5. APASS DR3 results

When compared with Figure 4, Figure 11 shows expanded coverage of APASS-calibrated stars South of 30 degrees declination. Table 4 must be interpreted with caution because it reflects *Tycho-2* calibrated stars as well as APASS calibrated stars. The new results are presented in the previously discussed Figures 5, 6, and 7. These figures contain data for both the APASS Johnson B and V calibration and the APASS SDSS g and r calibration.

Interpretation of these figures must recognize several underlying variables which may skew the results. First, there are the variations in coverage of the three catalogues which may produce systematic errors dependent on sky conditions. The GSC 2.3.2 catalogue may also have systematic variations between the different plate series used for the Northern and Southern hemispheres and for the Magellanic clouds. Experiments with subsets of Figure 6 show that there is indeed a large effect of sky position on the median RMS values between magnitudes 5 and 9. RMS values for Large Magellanic Cloud stars are higher than for Kepler satellite field stars. Second, there is the use of the *Tycho-2* catalogue to fill in the three calibration catalogues. Where APASS and *Tycho-2* overlap, most stars brighter than magnitude 9 have *Tycho-2* calibrations. Where there is no overlap, the *Tycho-2* calibrations can extend to magnitude 12. Our version of the GSC 2.3.2 catalogue uses independently derived transformations from *Tycho-2* to SDSS magnitudes. We learned that the KIC does not use a similar tranform. (According to Monet (2011), equations 6a and 6b in Brown *et al.* (2011) were used only for the calculation of the effective KIC magnitude.) Finally, there is the division of the DASCH data into patrol and deep-field telescopes. Each telescope is good for an optimum range of magnitudes, but this range is also a function of the exposure time for each individual plate and emulsion type.

Given the above uncertainties and the fact that both DASCH and APASS are continuing works in progress, the author can draw only limited conclusions at this stage. Except for the magnitude 13 to 15 range, the APASS calibration catalogue provides better results than the GSC 2.3.2 and KIC catalogues. The APASS B and V data produce comparable results to the APASS g and r data, though the former produce slightly fewer outliers over a large magnitude range. Finally, the attempt to merge in the *Tycho-2* catalogue should be abandoned in favor of a smaller but better-behaved dataset between magnitudes 8 and 11. Since we currently process every newly scanned plate with both the GSC 2.3.2 and the APASS catalogues, researchers seeking coverage for brighter stars should use the GSC 2.3.2 calibrated data. Subsequent to the processing of the APASS DR3 data, the DASCH team implemented a proposal by Sumin Tang to improve saturated star calibration, but the results from this change are not yet available. We are also interested in testing the new UCAC4 catalogue as a photometry calibration catalogue when it becomes available.

6. Acknowledgements

Thanks to the other members of the DASCH photometry team, principal investigator Prof. Jonathan Grindlay, Sumin Tang, and Dr. Mathieu Servillat for their advice and comments. Additional contributors to the DASCH project are the Harvard College Plate Stacks Curator, Alison Doane; hardware engineer, Robert Simcoe; and our transcribers, plate cleaners, and scanners. Arne A. Henden of AAVSO provided us with early releases of the APASS data. Both Arne Henden and the anonymous reviewer provided valuable comments on the drafts of this paper. This work was supported in part by NSF grants AST0407380 and AST0909073 and now also the Cornel and Cynthia K. Sarosdy Fund for DASCH. The AAVSO APASS project was made possible with funding by the Robert Martin Ayers Sciences Fund.

References

Bertin, E., and Arnouts, S. 1996, *Astron. Astrophys., Suppl. Ser.*, **117**, 393.

Brown, T. M., Latham, D. W., Everett, M. E., and Esquerdo, G. A. 2011, *Astron. J.*, **142**, 112.

Grindlay, J., Tang, S., Simcoe, R., Laycock, S., Los, E., Mink, D., Doane, A., and Champine, G. 2009, in *Preserving Astronomy's Photographic Legacy: Current State and the Future of North American Astronomical Plates,* ed. W. Osborn and L. Robbins, ASP Conf, Ser., 410, Astron. Soc. Pacific, San Francisco, 101.

Henden, A. A., Welch, D. L., Terrell, D., and Levine, S. E. 2009, *Bull. Amer. Astron. Soc.*, **41**, 669.

Høg, E., *et al.* 2000, *The Tycho-2 Catalogue of the 2.5 Million Brighest Stars*, *Astron. Astrophys.*, **355**, L27.

Lasker, B., *et al.* 2008, *Astron. J.*, **136**, 735.

Laycock, S., Tang, S., Grindlay, J., Los, E., Simcoe, R., and Mink, D. 2010, *Astron. J.*, **140**, 1062.

Monet, D. 2011, private communication (September 10).

Simcoe, R. J., Grindlay, J. E., Los, E. J., Doane, A., Laycock, S. G., Mink, D. J., Champine, G., and Sliski, A. 2006, in *Applications of Digital Image Processing XXIX*, ed. A. G. Tescher, Proc. SPIE, 6312, 35.

Tang, S. 2012, in preparation.

Zacharias, N., *et al.* 2010, *Astron. J.*, **139**, 2184.

Table 1. DASCH Primary Coverage Regions.

Region	*R.A. (2000.0)* h m	*Dec. (2000.0)* ° ′
M44	08 40	+19 41
3C273	12 30	+02 03
Baade's Window	18 03	–29 00
Kepler Satellite	19 20	+45 00
Large Magellanic Cloud	05 23	–69 45

Table 2. Photometry Pipeline Output for APASS DR1 and DR2.

Catalogue	*Plates*	*Magnitudes*	*Stars Matched*	*Light curves**
GSC 2.3.2	15,942	1,566,000,000	72,411,000	4,307,000
KIC	3,673	322,733,487	3,978,000	533,000
APASS DR1 and DR2	10,853	812,740,385	2,834,509	1,115,651

**The light curve contains at least two points that match all of the DASCH acceptance criteria.*

Table 3. Truth Table for Star Selection Between *Tycho-2* and APASS Stars Within 2 Arcsec. There are sixteen possible combinations of catalogue source and RMS level. The note marks show the preferred entry.

Matches (g–r)	*Matches (B–V)*	*First Duplicate Source*		*Second Duplicate Source*	
3956	3137	Tycho-2	RMS < 0.1[b]	Tycho-2	RMS < 0.1[b]
1505	1706	Tycho-2	RMS < 0.1[a]	Tycho-2	RMS > 0.1
31	25	Tycho-2	RMS > 0.1	Tycho-2	RMS < 0.1[a]
484	956	Tycho-2	RMS > 0.1[b]	Tycho-2	RMS > 0.1[b]
74787	55740	Tycho-2	RMS < 0.1	APASS	RMS < 0.1[a]
53171	37827	Tycho-2	RMS < 0.1[a]	APASS	RMS > 0.1
222573	264450	Tycho-2	RMS > 0.1	APASS	RMS < 0.1[a]
38373	30812	Tycho-2	RMS > 0.1	APASS	RMS > 0.1[a]
105587	81575	APASS	RMS < 0.1[a]	Tycho-2	RMS < 0.1
293755	346859	APASS	RMS < 0.1[a]	Tycho-2	RMS > 0.1
85118	63032	APASS	RMS > 0.1	Tycho-2	RMS < 0.1[a]
51744	44900	APASS	RMS > 0.1[a]	Tycho-2	RMS > 0.1
25800	87652	APASS	RMS < 0.1[b]	APASS	RMS < 0.1[b]
29950	34671	APASS	RMS < 0.1[a]	APASS	RMS > 0.1
25337	26382	APASS	RMS > 0.1	APASS	RMS < 0.1[a]
91410	23777	APASS	RMS > 0.1[b]	APASS	RMS > 0.1[b]

Notes: a: Selected candidate. b: Selected candidate has the lowest RMS.

Table 4. Photometry Pipeline Output for APASS DR3.

Catalogue	*Plates[a]*	*Magnitudes*	*Stars Matched*	*Light curves[b]*
GSC 2.3.2	16989	1,666,000,000	76,209,000	4,325,000
APASS DR3 g–r	16642[c]	1,282,000,000	8,109,840[c]	3,052,942[c]
APASS DR3 B–V	16640	1,291,720,897	8,323,460	2,618,816

Notes: a: These totals reflect the additional 1,188 plates of the LMC scanned during August, 2011.
b: The light curve contains at least two points that match all of the DASCH acceptance criteria.
c: These numbers may include stale entries from previous processing iterations.

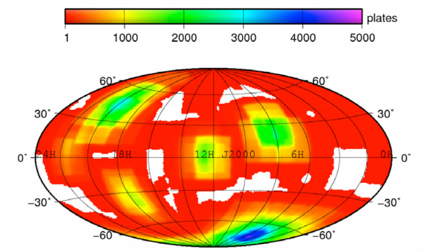

Figure 1. Current DASCH sky coverage from scanned plates.

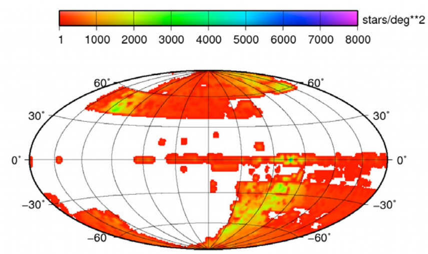

Figure 2. Coverage of APASS Data Releases 1 and 2.

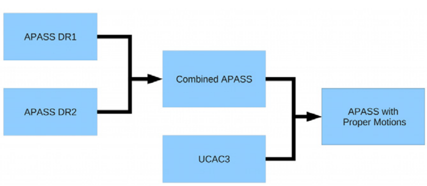

Figure 3. Preparation of DASCH/APASS DR1 and DR2 calibration catalog.

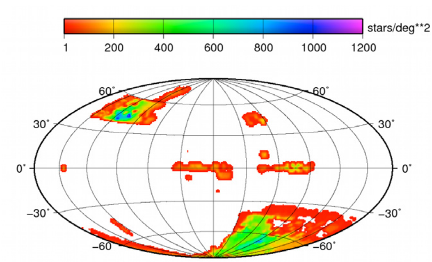

Figure 4. Coverage of APASS DR1 and DR2 calibrated plates.

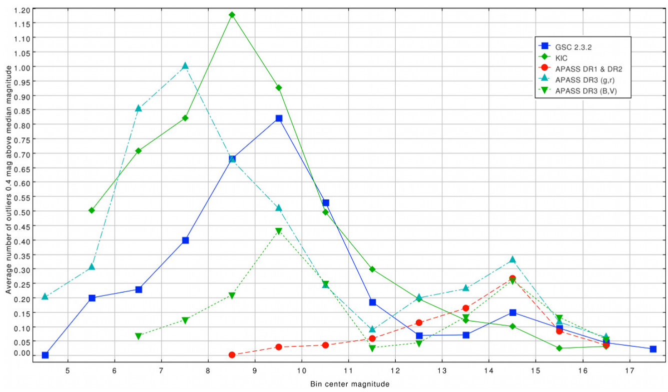

Figure 5. Number of DASCH light curves with at least ten good points after photometric processing.

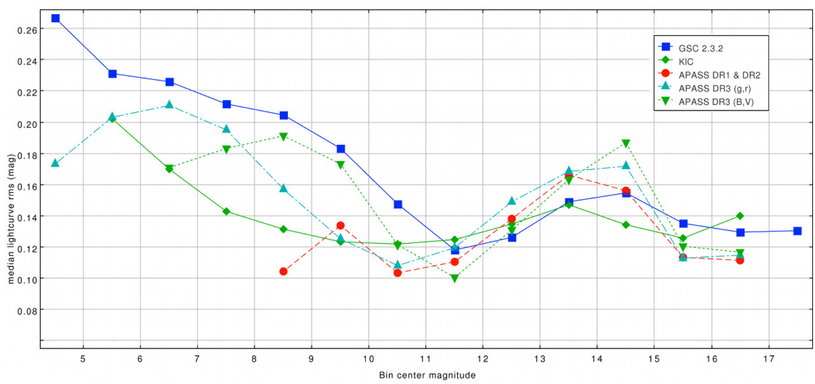

Figure 6. Median RMS of DASCH light curves with at least ten good points after photometric processing.

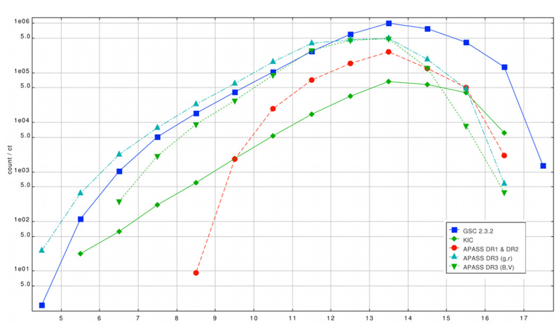

Figure 7. Average count light curve outliers: points 0.4 magnitude above the median light curve magntiude.

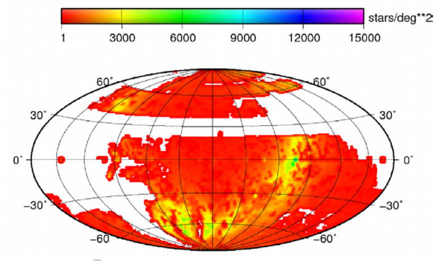

Figure 8. Coverage of APASS Data Release 3.

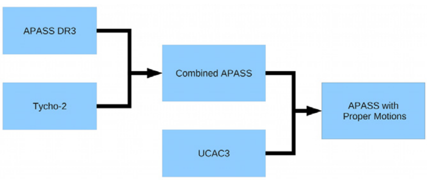

Figure 9. Preparation of DASCH/APASS DR3 calibration catalog.

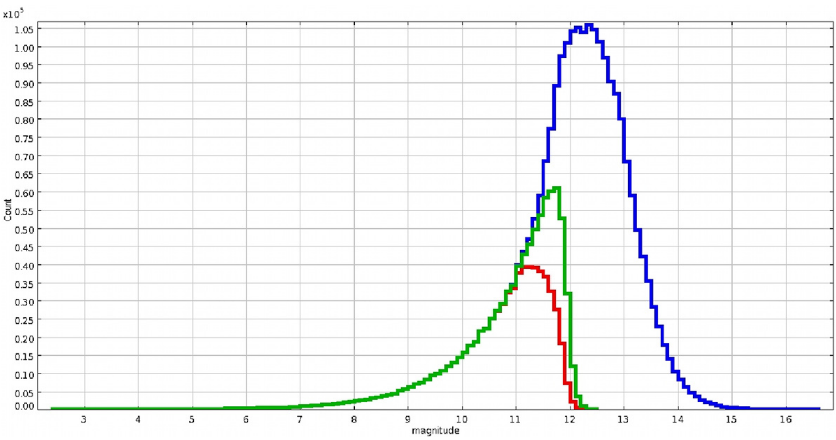

Figure 10. Magnitude Distribution of the Tycho2 Catalog. The full dataset (blue) contains all objects with valid magnitudes. The next most inclusive dataset (green) contains all objects with individual magnitude RMS values less than 0.1 mag. The most restricted dataset (red) contains objects with magnitude uncertainties added in quadrature less than 0.1 mag.

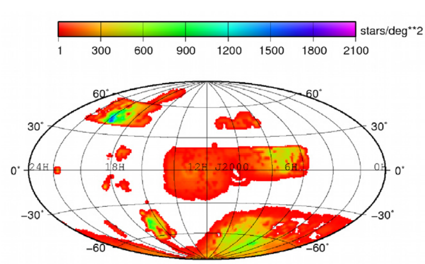

Figure 11. Coverage of APASS DR3 calibrated plates.