

Analysing the database for stars in open clusters

I. General methods and description of the data

J.-C. Mermilliod¹ and E. Paunzen^{2,3}

¹ Institut d'Astronomie de l'Université de Lausanne, 1290 Chavannes-des-Bois, Switzerland

² Institut für Astronomie der Universität Wien, Türkenschanzstr. 17, 1180 Wien, Austria

³ Zentraler Informatikdienst der Universität Wien, Universitätsstr. 7, 1010 Wien, Austria

Received 7 May 2003 / Accepted 25 June 2003

Abstract. We present an overview and statistical analysis of the data included in WEBDA. This database includes valuable information such as coordinates, rectangular positions, proper motions, photometric as well as spectroscopic data, radial and rotational velocities for objects of open clusters in our Milky Way. It also contains miscellaneous types of data like membership probabilities, orbital elements of spectroscopic binaries and periods of variability for different kinds of variable stars. Our final goal is to derive astrophysical parameters (reddening, distance and age) of open clusters based on the major photometric system which will be presented in a follow-up paper. For this purpose we have chosen the Johnson *UBV*, Cousins *VRI* and Strömgren *uvby β* photometric systems for a statistical analysis of published data sets included in WEBDA. Our final list contains photographic, photoelectric and CCD data for 469 820 objects in 573 open clusters. We have checked the internal (data sets within one photometric system and the same detector technique) and external (different detector technique) accuracy and conclude that more than 97% of all investigated data exhibit a sufficient accuracy for our analysis. The way of weighting and averaging the data is described. In addition, we have compiled a list of deviating measurements which is available to the community through WEBDA.

Key words. open clusters and associations: general – catalogs

1. Introduction

The study of open clusters is very important in several respects. It allow one to estimate different important astrophysical parameters within individual clusters as well as the study of the wider solar neighbourhood concerning its structure.

For this purpose it is essential to have a homogeneous set of photometric and additional (e.g. membership probability, proper motion) data for a statistically significant number of open clusters.

One of the most compelling databases in this respect is the WEBDA interface which has been developed at the Institute for Astronomy at the University of Lausanne, Switzerland. It offers astrometric data in the form of coordinates, rectangular positions, and some proper motions, photometric data in the major systems in which star clusters have been observed (e.g. Johnson-Cousins *UBVRI*, Strömgren *uvby β* and Geneva 7-color), spectroscopic data, like spectral classification, radial velocities, rotational velocities. It also contains miscellaneous types of data like membership probabilities, orbital elements of spectroscopic binaries and periods of variability for

different kinds of variable stars. Finally a whole set of bibliographic references allows the community to locate the relevant publications for each individual cluster easily.

In this first paper we present the compilation of 573 open clusters for which photometric measurements are available within WEBDA. The statistical methods used to derived weighted means are described. Lists with objects showing deviating photometric measurements within one system and/or different observing techniques (photographic, photoelectric and CCD) were generated. We discuss the internal and external measurement accuracies based on a statistically significant sample of independent sources from the literature.

Our final goal of Paper II will be the determination of the ages, distances and reddening for the presented open clusters using the newest isochrones. This analysis will include the Johnson-Cousins *UBVRI* and Strömgren *uvby β* photometric systems and should supersede the work of Janes & Adler (1982) who presented a compilation of 434 open cluster of our Milky Way for which they summarized the reddening, ages and distances from 610 references in order to analyse the galactic structure. Their compilation was highly nonuniform since they made no attempt to redetermine the appropriate astrophysical parameters.

Send offprint requests to: E. Paunzen,
e-mail: ernst.paunzen@univie.ac.at

2. Description of the database

WEBDA (accessible via <http://obswww.unige.ch/webda/>) and its predecessor BDA has been developed since 1987 at the Institute for Astronomy, University of Lausanne, Switzerland by JCM. The progress of its development was described by Mermilliod (1988, 1992, 1995). We will give here a brief overview of its current status and content.

The database tries to collect all published data for stars in open clusters that may be useful either to determine the star membership, or to study the stellar content and properties of the clusters.

It is divided into three levels: 1) database; 2) cluster and 3) star level.

The database contents includes measurements in most photometric systems in which cluster stars have been observed, spectroscopic observations, astrometric data, various kinds of useful information, and extensive bibliography. It is possible to perform selection of clusters based on the amount of available data. The data are usually recorded in their original form, with an indication of the source, but also as averaged values or selected data when relevant.

The greatest effort has been spent in solving the identification problems raised by the definition of so many different numbering systems and a special interface has been developed to query the cross-reference tables.

Maps for more than 200 clusters have been scanned and included in the database. They are active maps and permit to retrieve basic data (e.g. positions, cross identifications and Johnson *UBV* values) simply by clicking on the star images.

The database structure uses the directory hierarchy supported by the Unix system. The main directory is the database itself. It contains several sub-directories: description of the database, help information, references, bibliography, programs and perl scripts. Each cluster defines an independent directory identified by its name and containing the available data in distinct files, one for each data type. This structure allows easy inclusion of any new cluster and any additional data type.

Whenever possible, the records of the various data files have the same structure: star identification, source, data. The files are organised sequentially and, within the files, the entries are sorted by star number and source reference.

The star identification is the main key to access the data, but it is also possible to use filters based on the bibliographic references or astrophysical parameters.

The database engine WEBDA is a relational database built upon the package “/rdb” developed by Manis et al. (1988) which is a high performance relational database management and application development system designed for Unix environments.

Samples of clusters can be obtained by performing a selection on the clusters parameters, i.e. coordinates, distances, ages, diameters. The form prepared permits to do the selection on all parameters simultaneously. Clusters may also be selected on active plots, drawing the clusters in right ascension versus declination plane, galactic longitude versus latitude, distance from the Sun or above the galactic plane.

Table 1. Excerpt of the content of WEBDA from the 8th of April 2003.

Data description	Clusters	Measurements	Stars
Fundamental			
Identifications	403	12 079	12055
Transit Tables	315		349171
Coordinates J2000	408	110 385	109256
Coordinates B1950	480	143 775	134028
Positions (round off)	482		142422
Positions (<i>x</i> , <i>y</i>)	514		461873
Double stars	198	2063	1631
Photometry			
<i>UBV</i> (photographic)	294	126 775	100221
<i>UBV</i> (photoelectric)	439	34 000	23038
<i>UBV</i> (CCD)	261	315 374	261135
<i>VRI</i> (Cousins)	43	1460	412
<i>VRI</i> (Cousins; CCD)	86	45 596	42788
<i>RI</i> (Cousins; CCD)	12	2803	2712
<i>VI</i> (Cousins; CCD)	192	286 357	257731
<i>VRI</i> (Johnson)	97	2598	2145
<i>u</i> by (photoelectric)	214	7260	4949
<i>u</i> by (CCD)	25	21 371	20277
β (photoelectric)	248	7277	4771
β (CCD)	16	2685	2414
Geneva 7-color	190	4618	4496
<i>RGU</i> (photographic)	79	10 369	10332
Spectroscopy			
MK types	300	10 397	6399
HD types	319	13 148	12625
$v \sin i$	107	4636	3199
V_r (mean)	92	3734	3492
V_r (individual)	214	44 606	5927
V_r (GPO)	10	702	699
V_r (RFS)	7	141	141
Orbits	59	419	275
Miscellaneous			
Proper motion (abs)	7		3653
Proper motion (rel)	12	6304	6302
Probability (μ)	81		39384
Probability (V_r)	8		655
Periods (Var)	50	2482	1905
X-ray flux	28	3910	3351
gK stars	260		5189
Am stars	34		110
Ap stars	84		218
Be stars	86		368
Blue stragglers	209		930
Spectroscopic binaries	49		934

The database is in a dynamic growing process as new data are published and included. Table 1 lists an excerpt of the content of WEBDA based on the status from the 8th April 2003. This date is also used as a “deadline” for our final analysis.

3. Analysis and compilation of the data

The final goal of this extensive statistical analysis is not only to investigate the consistency of the published data but more importantly to derive ages, reddening values and distances for a large number of open clusters. It is therefore necessary to select photometric systems for which enough data and appropriate isochrones are available. Table 1 lists the numbers for the most common photometric systems included in WEBDA. From a close inspection we have chosen the following photometric systems for our analysis:

- Johnson *UBV*; photographic, photoelectric and CCD measurements
- Cousins *VI*; CCD
- Cousins *VRI*; photoelectric and CCD
- Strömberg *wbyβ*; photoelectric and CCD
- *RGU*; photographic, for comparison.

We have not included the Johnson *VI* systems because there are usually only a few measurements for the brightest objects within one cluster making an isochrone fitting impossible. The Geneva 7-color system is outstanding compared to the other photometric systems. For most of the open clusters, only the brightest members are investigated whereas for a few ones also the faintest members were observed. Furthermore, WEBDA already includes the mean values for all these objects, so no improvement can be done within our analysis. We have used the data of the Geneva 7-color system for several open clusters in order to check the results from the other photometric systems.

The list with available photometric data in the mentioned above systems contains 573 open clusters (listed in Table 2) with a total of 469 820 objects.

The data analysis of the relevant photometric systems includes several different steps in order to perform a careful check of the homogeneity of the individual sources. Much effort was already spent to improve the homogeneity of WEBDA by investigating the published data, finding charts and listed coordinates (Mermilliod 1988, 1992, 1995). This process is very time consuming and not straightforward. During the first stages of our new analysis we have already found some wrongly identified objects and misprints in the literature. These errors have already been fixed in WEBDA. But we have to emphasize that these are only the “eye hitting” divergences, still there are many unsolved cases (see the lists mentioned above) which have to be investigated in the future.

As the first step of our analysis we have checked the intrinsic consistency of different sources for one photometric system (e.g. photographic Johnson *UBV* data) of all individual open clusters. In general, we have used the following (very conventional) limits for a measurement being “outstanding” if the difference of the data are larger than:

- 0.5 mag: *UBV* photographic; *RGU*
- 0.2 mag: *UBV*, *VRIc*, photoelectric and CCD; *wby*, photoelectric
- 0.1 mag: *wby*, CCD; *β* photoelectric and CCD.

The compiled list includes 4467 entries (2914 from photographic measurements). Excluding these objects, we have

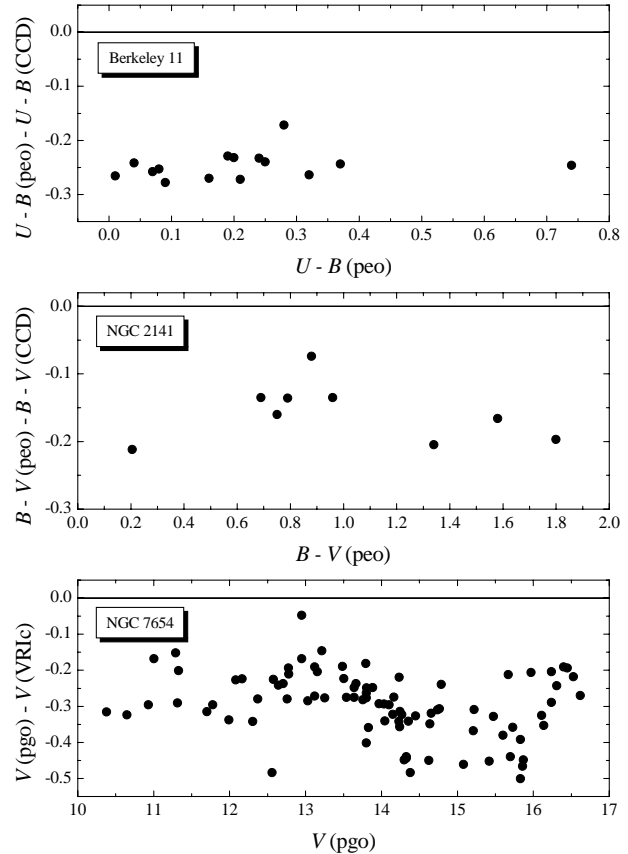


Fig. 1. Three examples for deviating measurements from different observing techniques as listed in Table 4.

searched for intrinsic correlations for data sets which have more than five objects in common using a simple linear correlation algorithm. We only find twelve statistically significant deviating cases for ten open clusters. These deviating cases are listed in Table 3. Paunzen & Maitzen (2002) reported one deviating case for NGC 6451 for which they were able to show that the photometry by Piatti et al. (1998) has an unidentified error and was therefore excluded from our analysis.

As a next step we have used the averaged mean values of different photometric systems (e.g. *UBV* photographic and photoelectric) to search for “external” discrepancies between measurements for the individual clusters. Again, a list of outstanding objects was created with the limits given as:

$$\text{limit}(\text{ext}) = \sqrt{\text{limit}(\text{system}_1)^2 + \text{limit}(\text{system}_2)^2}. \quad (1)$$

This list has 7061 entries. Table 4 shows the deviating data sets from this external check. Figure 1 shows three examples for the Johnson *UBV* system graphically.

Since it is well known that photographic measurements have in general larger errors we have used averages of photoelectric and CCD data only. If such data are not available, photographic ones were used. The final averaged weighted values were calculated following the approach described in Mermilliod & Mermilliod (1994) which is a two step iterative procedure. The first step consists of a weighted mean, the weight being the number of measurements to the $2/3$ power. The next step uses the differences between the weighted mean

Table 3. Deviating data sets within one photometric system. The errors in the final digits of the corresponding quantity are given in parenthesis.

Cluster	Photometric system	Set1	Set2	Mean	$N(\text{obj})$
Berkeley 64	$UBV: V$ (CCD)	Pandey et al. (1997)	Ann et al. (2002)	-0.161(44)	26
Markarian 50	$UBV: U - B$ (pgo)	Turner et al. (1983)	Grubissich (1965)	+0.174(27)	27
Melotte 71	$UBV: V$ (pgo)	Pound & Janes (1986)	Hassan (1976)	+0.388(102)	33
NGC 1348	$V Ic: V$	Ann et al. (2002)	Carraro (2002)	+0.197(55)	124
NGC 2244	$V Ic: V - I$	Park & Sung (2002)	Berghöfer & Christian (2002)	-0.237(68)	124
NGC 6611	$UBV: U - B$ (peo)	Hiltner & Morgan (1969)	Thé et al. (1989)	+0.083(22)	15
NGC 6791	$V Ic: V - I$	Garnavich et al. (1994)	von Braun et al. (1998)	+0.104(34)	12
NGC 6910	$UBV: V$ (peo)	Hoag et al. (1961)	Heiser (private communication)	+0.088(26)	13
NGC 7044	$UBV: V$ (CCD)	Aparicio et al. (1993)	Sagar & Griffiths (1998)	+0.135(36)	553
NGC 7654	$UBV: V$ (CCD)	Choi et al. (1999)	Stetson (2000)	-0.164(23)	49
	$UBV: B - V$ (CCD)	Choi et al. (1999)	Stetson (2000)	+0.153(22)	73
	$V Ic: V - I$	Choi et al. (1999)	Stetson (2000)	-0.231(32)	57

Table 4. Deviating data sets within different photometric systems.

Cluster	Photometric system	mean	$N(\text{obj})$
Berkeley 11	$U - B$ (peo, CCD)	-0.247(25)	15
	$B - V$ (peo, CCD)	-0.083(25)	17
Berkeley 58	V (pgo, CCD)	-0.346(84)	36
Bochum 10	V (pgo, CCD)	-0.346(84)	36
Haffner 8	$U - B$ (pgo, peo)	-0.303(86)	7
NGC 637	V (pgo, peo)	-0.268(60)	25
	V (pgo, CCD)	-0.303(76)	35
NGC 884	m_1 (peo, CCD)	+0.130(35)	17
NGC 1245	V (peo, CCD)	-0.241(40)	15
NGC 2112	V (peo, CCD)	+0.168(46)	16
	V (peo, $VRIc$)	-0.241(40)	15
	V (CCD, $ubvy$)	+0.221(52)	9
NGC 2141	$B - V$ (peo, CCD)	-0.158(41)	9
NGC 2158	V (peo, CCD)	-0.129(24)	8
	V (peo, $V Ic$)	-0.124(32)	8
NGC 3680	V (CCD, $V Ic$)	+0.077(12)	6
NGC 6005	V (CCD, $V Ic$)	-0.053(17)	529
NGC 7654	V (pgo, $VRIc$)	-0.296(89)	89
	V (peo, $VRIc$)	-0.157(28)	18
	V ($VRIc$, $ubvy$)	+0.147(41)	9

and the individual values to compute new weighted mean values. This procedure gives a lower weight to discrepant values.

We also find a few cases (e.g. for NGC 6705) for which data sets show a trend within different photometric systems, e.g. $(U - B)_{\text{Set1}} - (U - B)_{\text{Set2}}$ versus $(B - V)_{\text{Set1}}$. There is no straightforward solution for these data sets. However, we have excluded those data from our final analysis.

The complete tables with the available weighted mean values will be available at WEBDA only or upon request from the authors.

4. Discussion

For the analysis of the overall accuracies of the available data sets, we have used Gaussian distributions (Christensen 1996)

to fit the histograms of all mean values for the different data. The calculated histograms were normalized to the overall percentage of the sum. The results for the Johnson UBV and Strömgren $ubvy\beta$ photometric systems are summarized in Figs. 2 to 4. These figures include the most important fit parameters such as the mean value, width, R^2 , χ^2 as well as the number of data sets and objects. The corresponding histograms in the other photometric systems are not plotted since there are too few data points to compare which makes a statistically sound analysis impossible.

The way of calibrating observed magnitudes is either to simultaneously measure “well established” standard stars or to use already known standard transformations for the individual telescope and filter set. Both approaches are certainly not straightforward. Sung & Bessell (2000) summarize and discuss the problems concerning the variations of atmospheric extinction coefficients, transformation equations, different filter systems, CCDs as well as the difference between two sets of standard stars (SAAO and Landolt). They also include a compelling list of references concerning this special topic.

It is out of the scope of this statistical analysis to reproduce the used transformation technique of the individual references. We have to rely on the published data. The only possible check is to search for misidentified objects or typos. Otherwise undetected variability of any kind could also lead to several divergent observations, again a fact which we are not able to prove.

We will now discuss the internal and the external accuracy separately.

4.1. The internal accuracy

The most important check for the reliability of published data is the comparison with other independent measurements within the same photometric system and the same technique. Figure 2 shows the histograms (bin size is 0.02 mag) for the Johnson UBV system. We have separated the photographic, photoelectric and CCD measurements.

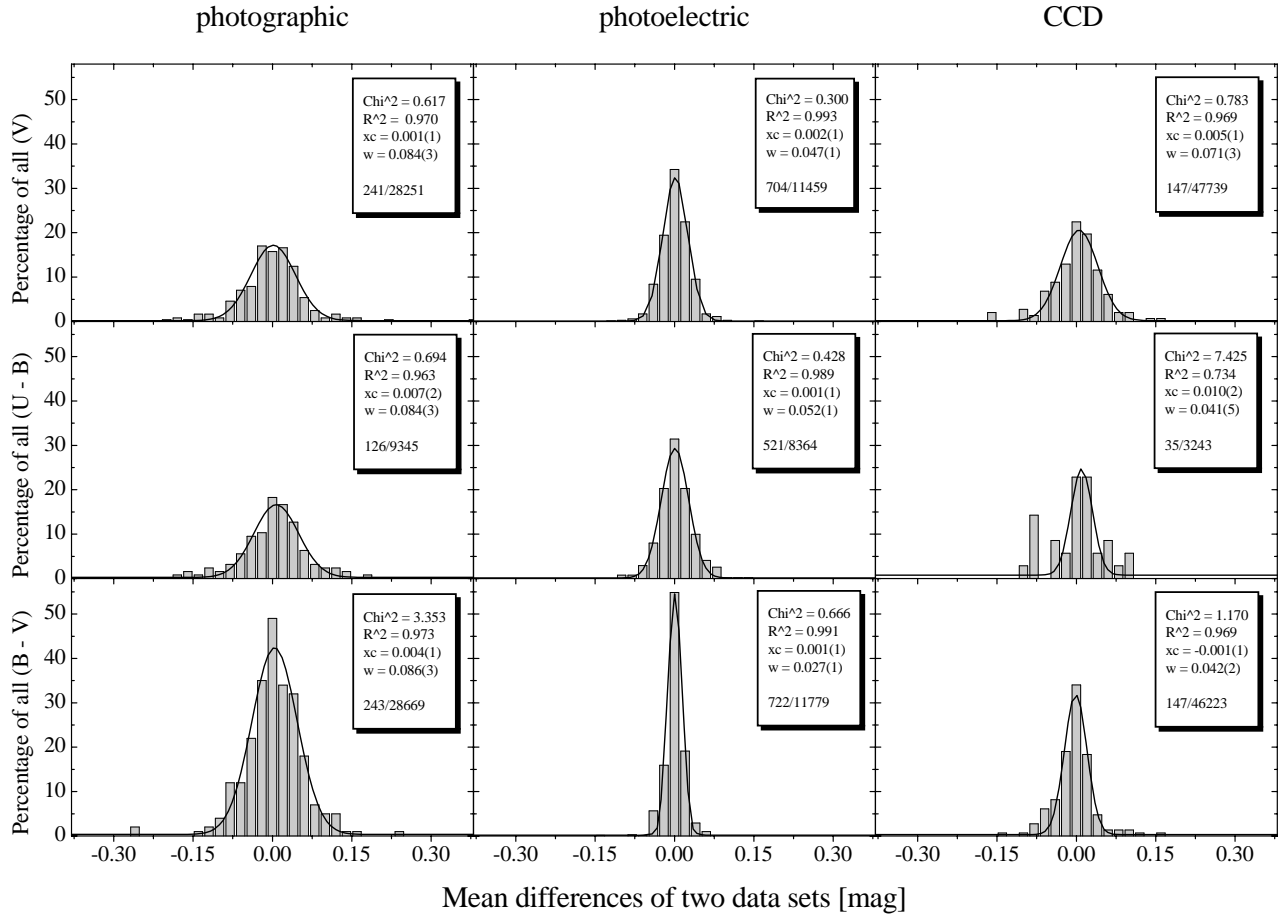


Fig. 2. Histograms of the internal accuracy of the Johnson UBV photometric system for photographic (left panel), photoelectric (middle panel) and CCD (left panel) measurements; listed are the most important parameters of the fitted Gaussian distributions together with the number of data sets and objects (N_1/N_2).

The histograms are based on a statistically significant number of individual data sets and objects. The only exception is CCD measurements for $U - B$. This is probably caused by the insensitivity of the modern CCD detectors in the ultraviolet region which makes the observations in the standard U band almost impossible. From Fig. 2 we are able to conclude:

- The bandwidth of the Gaussian distributions for the photographic measurements is twice (≈ 0.09 mag) as large as the corresponding ones from the other two sources.
- The only exception is the CCD V data which might be due to the relative faintness and thus the larger observational error reach with this technique.
- All mean values of the fitted distributions are close to zero.

The corresponding histograms for the Strömgren $wby\beta$ photometric system have a bin size of 0.01 mag and are shown in Fig. 3. The widths of the fitted distributions are all between 0.01 and 0.03 mag which shows the high quality of the published data.

In total, 4467 deviating measurements were found (Sect. 3) within 4056 different data sets and 266 779 objects. This corresponds to 1.7% which is an extremely low percentage. If we exclude the outliers from photographic measurements, this percentage even lowers to 0.6%.

4.2. The external accuracy

After having shown that the accuracies within the individual photometric systems are very good. We then investigated the errors for different measurement techniques and thus mainly different quantum efficiencies characteristics of detectors.

Figure 4 shows the result for the Johnson UBV photometric system (bin size is 0.02 mag). The results for the other systems are similar. We have summarized the data for V , $B - V$ and $U - B$, otherwise we would run into poor number statistics. However, it shows that the distributions for the three different indices with the same detector technique are essentially the same.

The bandwidth of the Gaussian distributions for comparison of the photoelectric data sets (upper and lower panel) is about 0.05 mag whereas the comparison of the CCD versus photographic measurements results in an almost three times higher value (0.12 mag). This reflects the most different quantum efficiency characteristics of these detectors whereas the photoelectric one is in between. In addition, the larger scatter may be due to the faintness of the photographically observed objects. Usually, the photoelectric data magnitude limit is brighter than that of the photographic one. We conclude that the photoelectric measurements are still the most valuable to connect the photographic to the CCD ones.

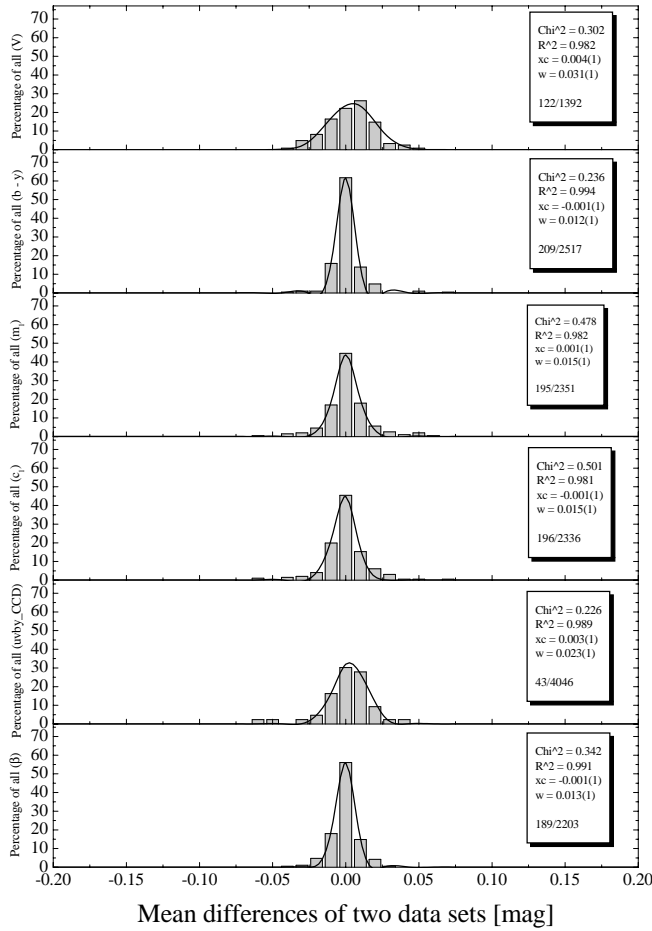


Fig. 3. Histograms of the internal accuracy of the Strömgren $uvby\beta$ photometric system. The upper four panels show the results for photoelectric $uvby$, the fifth CCD $uvby$ and the last panel all β measurements; listed are the most important parameters of the fitted Gaussian distributions together with the number of data sets and objects (N_1/N_2).

This analysis is based on 1960 data sets with 292 770 measurements. Again, the number of deviating data points is surprisingly low (7061 or 2.4%). This shows the high capability of WEBDA to analyse the astrophysical properties of open clusters in the Milky Way.

5. Conclusions and outlook

The enormous amount of photometric data within WEBDA was analysed in order to check for the internal and external accuracies of published data for open clusters. This analysis is based on photographic, photoelectric and CCD measurements for five different photometric systems which includes 573 open clusters and 469 820 objects. The way of weighting and averaging the data is described.

We have investigated 4056 data sets which have more than five objects in common and concluded that the internal accuracies are very good. The accuracy is best for the Strömgren $uvby\beta$ system and drops significantly towards photographic Johnson UBV data. Less than 2% of deviating measurements were found and tabulated.

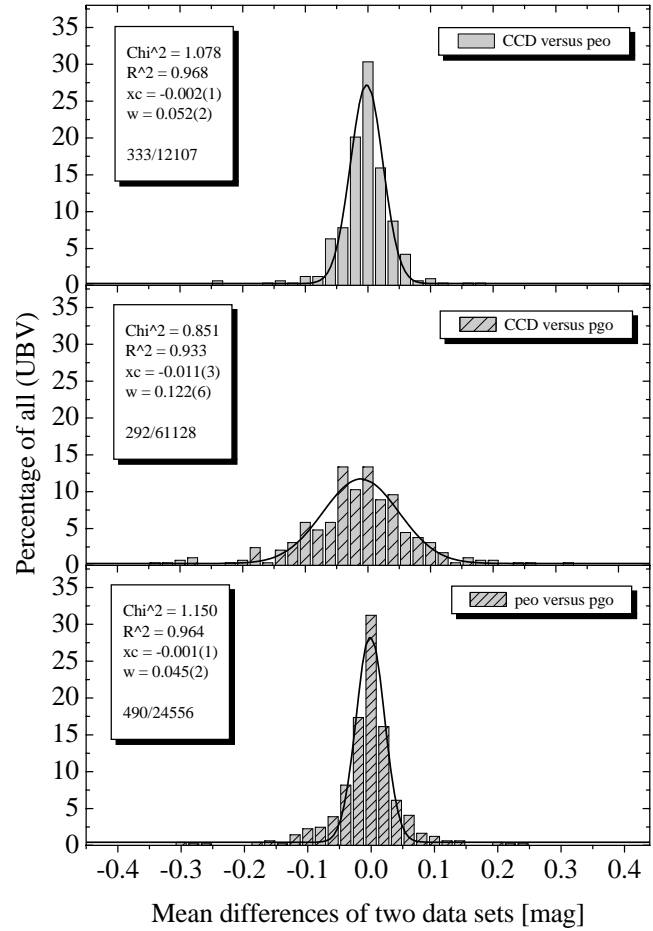


Fig. 4. Histograms of the external accuracy of the Johnson UBV photometric system for CCD versus photoelectric (upper panel) and photographic (middle panel) as well as photoelectric versus photographic (lower panel) measurements; listed are the most important parameters of the fitted Gaussian distributions together with the number of data sets and objects (N_1/N_2).

A surprisingly good agreement between photoelectric and photographic as well as CCD data was found. The higher error for CCD versus photographic data reflects the differences of the individual quantum efficiency curves of these systems. Nevertheless, the amount and homogeneity of data will allow us to derive astrophysical parameters such as the ages, distances and reddenings for the 573 open clusters investigated. This will be done in a second paper which includes isochrone fitting and the discussion of different statistical issues concerning the structure of our Milky Way.

Acknowledgements. EP acknowledges partial support by the Fonds zur Förderung der wissenschaftlichen Forschung, project P14984.

References

Ann, H. B., Lee, S. H., Sung, H., et al. 2002, *AJ*, 123, 905
 Aparicio, A., Alfaro, E. J., Delgado, A. J., Rodríguez-Ulloa, J. A., & Cabrera-Cano, J. 1993, *AJ*, 106, 1547
 Berghöfer, T. W., & Christian, D. J. 2002, *A&A*, 384, 890
 Carraro, G. 2002, *A&A*, 387, 479

- Choi, H. S., Kim, S.-L., Kang, Y. H., & Park, B.-G. 1999, *A&A*, 348, 789
- Christensen, R. 1996, *Analysis of Variance, Design and Regression* (London: Chapman & Hall)
- Garnavich, P. M., Vandenberg, D. A., Zurek, D. R., & Hesser, J. E. 1994, *AJ*, 107, 1097
- Grubissich, C. 1965, *Z. Astrophys.*, 60, 256
- Hassan, S. M. 1976, *A&AS*, 26, 9
- Hiltner, W. A., & Morgan, W. W. 1969, *AJ*, 74, 1152
- Hoag, A. A., Johnson, H. L., Iriarte, B., et al. 1961, *Pub. US Naval Obs. Second Ser.*, 17, 347
- Janes, K., & Adler, D. 1982, *ApJS*, 49, 425
- Manis, R., Schaffer, E., & Jorgensen, R. 1988, *Unix Relational Database Management* (Prentice Hall)
- Mermilliod, J.-C. 1988, *Bull. Inform. CDS*, 35, 77
- Mermilliod, J.-C. 1992, *Bull. Inform. CDS*, 40, 115
- Mermilliod, J.-C. 1995, in *Information & On-Line Data in Astronomy*, ed. D. Egret, & M. A. Albrecht, *Astrophys. and Space Sci. Library*, 203, 127
- Mermilliod, J.-C., & Mermilliod, M. 1994, *Catalogue of Mean UBV Data on Stars* (New York: Springer Verlag)
- Paunzen, E., & Maitzen, H. M. 2002, *A&A*, 385, 867
- Pandey, A. K., Durgapal, A. K., Bhatt, B. C., Mohan, V., & Mahra, H. S. 1997, *A&AS*, 122, 111
- Park, B.-G., & Sung, H. 2002, *AJ*, 123, 892
- Piatti, A. E., Clariá, J. J., & Bica, E. 1998, *ApJS*, 116, 263
- Pound, M. W., & Janes, K. A. 1986, *PASP*, 98, 210
- Sagar, R., & Griffiths, W. K. 1998, *MNRAS*, 299, 1
- Stetson, P. B. 2000, *PASP*, 112, 925
- Sung, H., & Bessell, M. S. 2000, *Publ. Astron. Soc. Austr.*, 17, 3, 244
- Thé, P. S., de Winter, D., Feinstein, A., & Westerlund, B. E. 1989, *A&AS*, 82, 319
- Turner, D. G., Moffat, A. F. J., Lamontagne, R., & Maitzen, H. M. 1983, *AJ*, 88, 1199
- von Braun, K., Chiboucas, K., Minske, J. K., & Salgado, J. F. 1998, *PASP*, 110, 810