Development of a new digital version of "Cordoba Durchmusterung" stellar catalog.

Daniel Esteban Severin[1]
Diego Javier Sevilla[2]

Received: September 2015 – Accepted: October 2015

[1]Jefe de Trabajos Prácticos (dedicación exclusiva) en Depto. de Matemática de la EFB | FCEIA | Universidad Nacional de Rosario. Investigador Asistente de CONICET. http://www.fceia.unr.edu.ar/~daniel,

[2]Jefe de Trabajos Prácticos (dedicación exclusiva) en Depto. de Física y Química de la EFB | FCEIA | Universidad Nacional de Rosario

E-mail: daniel@fceia.unr.edu.ar

Introducción

A stellar catalog is basically a set of records where each one identifies a single star. Each record is composed of stellar coordinates plus other physical data associated with the corresponding star. Currently, there exist thousands of astronomical catalogues available to the public through Data Centers such as Strasbourg (http://cdsweb.u-strasbg.fr) which offers, among other services, free download access to complete catalogues in its repository *VizieR*, or online search access via its web site *SIMBAD*.

A single star has different designations according to the catalog being used. For example, *Sirius* is designated as α CMa (Bayer), 9 CMa (Flamsteed), BD -16° 1591 (Bonner Durchmusterung), GC 8833 (Boss General Cat.), HD 48915 (Henry Draper), HIP 32349 (Hipparcos), SAO 151881 (Smithsonian Astrophysical Obs.) and WDS J06451-1643 (Washington Double Star), among others.

Nowadays, catalogues are generated automatically from specific algorithms that reduce measurements and published in digital format through the Internet. But older catalogues, originally published in printed form, have been transcribed/digitized by hand. The one that we address in this work is *Cordoba Durchmusterung* (CD), which has more than 600000 stars between declination -22° and the south pole. This catalog was made at the Argentine National Observatory in the late nineteenth and early twentieth century [1]. Despite the age of this catalog, it is still relevant. In particular, there are many stars whose primary designation in *SIMBAD* belongs to this catalog, for instance: CD -23° 8634 or CD -23° 8645.

During the translation from old printed catalogues to digital format, *typo* errors are usually introduced. In the case of CD, transcription was carried out by a consortium of several institutions led by the *National Space Science Data Center* of *NASA*, and then was stored in *VizieR* as catalog I/114. An example of a typo error in this digital version is star CD -26° 16170, which has magnitude 6.5, but in the original printed version it is 9.5 [2]. Another disadvantage of catalog I/114 is the lack of information that the printed version has: a) stars that are double or multiple, b) stars that show coloration. In addition, the digital catalog does not provide cross-references, possibly because at the time of its preparation current catalogues were not complete enough to identify each star of CD. Some of the stars of CD have been identified as built in cross-references in other catalogues such as SAO or PPM.

An example of a CD star not identified is CD -22° 2. If the coordinate of star TYC 5844-275-1 is precessed to B1875.0 (the epoch of CD) and corrected by proper motion, and the angular separation between this star and CD -22° 2 is computed, the resulting distance is less than 4 arcsec. Since this value is within the accuracy of CD, and the next closer star of CD -22° 2 is

TYC 5844-246-1 with an angular separation of 593 arcsec, it is clear that the identification between the first two stars is right. At present, only some of the CD stars (about a quarter), have been identified.

Having a cross-identification between CD and other catalogues is relevant for several reasons. It helps detect errors in digital transcriptions. For instance, the error given above about CD -26° 16170 was detected since we knew that this star is identified with PPM 273886, a star with magnitude V 10. Note the remarkable discrepancy between the magnitude reported in the digital version of CD and the one of PPM star, which is closer to the given value in the printed version. Cross identifications also allow comparisons of observations obtained at different epochs, thus detecting changes in parameters such as the brightness of stars. On the other hand, objects that cannot be identified with counterparts in modern catalogues are also of interest, as they may be objects like asteroids, novae, supernovae, etc. that were then visible in that sector of the sky. Finally, the CD catalog was widely used throughout the twentieth century and there are publications identifying stars with CD designation. An example is the heavy use of CD and other *durchmusterung* in creating star charts to make visual estimates of variables stars [3].

There are precedents of digitization of printed catalogues and cross identifications with modern ones [4]. In particular *Uranometría Argentina*, another catalog compiled by the Argentine National Observatory, was transcribed independently by two different groups. On the one hand, the version of Paolantonio and Minniti which was made under a project of the Ministry of Science and Technology from the National University of Cordoba [5]. On the other, the version of Frederick Pilcher, that can be downloaded as catalog V/135A from *VizieR*. In the latter one, cross references to HD and SAO catalogues are provided.

With the help of a group of amateur astronomers, we intend to develop a new version of the Cordoba Durchmusterung catalog by removing as many typo errors as possible from digital catalog I/114, adding missing information that are present in the printed catalog (colored stars and double ones), and providing cross identifications with modern catalogues. Due to the magnitude of this work, and in order to assess the feasibility of it, in this article we address only a belt of the sky corresponding to CD stars with declination -23° (B1875.0). The work reported here was performed in 3 stages. In the first one, the list of color stars and double ones (from the area of sky previously mentioned) was transcribed to a digital format. In the second stage, a comparison between records of CD and other catalogues for which there is a known cross-identification was performed. In the third stage, the problem of cross-identification of CD records with a modern catalog was addressed through the implementation of an algorithm that solves such identification.

This paper is divided as follows. In Section 2, different aspects of CD catalog including conversion between its magnitude scale and modern ones are given. In Sections 3, 4 and 5, stages are presented and developed. In Section 6, format of the new digital catalog is explained. Finally, in Section 7, conclusions and future works are given.

The Cordoba Durchmusterung catalog

The *durchmusterungs* consists of 3 stellar catalogues that completely sweep the sky to account for stars up to about 10th magnitude [1]. The Bonner Durchmusterung (BD) includes stars from the North Pole to the declination -1°. The second, Südlicher Durchmusterung (SD), covers declinations between -2° and -23°. Finally, the CD covers from -22° to the South Pole and has 613778 stars and 175 non-stellar objects. Two overlapping degrees in the SD and CD aim to splice both catalogues properly.

In order to identify a star of CD, notation CD *-xx° yyyyy* is used, where *xx* is the declination degree of the star and *yyyyy* is given by an enumeration in ascending order according to the right ascension between stars with the same declination degree. The catalog offers positions and brightness up to 10th magnitude (in the scale used in such epoch). Positions are reported for the epoch B1875.0, where right ascension is given in hours, minutes and seconds with a precision of $0^s.1$, and declination is given in degrees and minutes of arc with a precision of 0'.1. Precision for the visual magnitude is 0.1. The printed catalog also has extra information: a list of colored stars (yellow – red) and a list of visually double (or multiple) stars, i.e. when the angular separation is small as they are observed with a low-powered eyepiece (15x). This information is given as footnotes. Moreover, the printed catalog also has cross-references with catalogues of that epoch [2].

Further corrections to the printed catalog were published. We use the *corrigenda* of the same volume [2], those reported by Thome [6] and by *NSSDC* [7].

For the sky area corresponding to declination -23°, the number of CD records is 18138: 18133 stars and 5 non-stellar objects which are distinguished with legend *neb.* in the field where magnitude should be reported: -23° 542, -23° 1186, -23° 13481, -23° 14459 and -23° 17047. The latter records should not be cross identified with other stellar catalogues.

For stars between -22° and -32°, the *probable* error is $\pm 0^s.42$ in right ascension and ±0'.23 in declination [2]. That is equivalent to $\sigma_{\alpha *}$ = 9.3 arcsec and σ_{δ} = 20.5 arcsec in terms of standard deviation respectively, where $\alpha* = \alpha.\cos(\delta)$. In further observations, for other sky areas, the error was reduced.

In order to estimate the error in brightness, it is necessary to convert between the old visual scale used in CD and Johnson V magnitude scale. In this work, we fit a quadratic polynomial by least squares to points representing magnitudes of CD with respect to the ones that are present in other catalogues, for cases where the cross-identification is known. We used *Hipparcos* and *PPM* catalogues. They are described more deeply in Section 4. For *Hipparcos*, we considered two cases: 1) only non-variable stars (VarFlag=1), and 2) all records, although HIP 67620 was discarded since the difference in magnitude with CD -23° 11328 is 3.5 units, too high to be a statistical error. The number of stars in each fitting was 763 (HIP without variables), 809 (HIP) and 3927 (PPM). On the one hand, we noticed that the difference between both HIP's fittings is negligible, i.e. discarding variable stars has little effects in the result. On the other, we observed that HIP's fittings are not suitable for high magnitude stars, since this catalog has not enough faint stars. We conclude that the best fitting is the one yielded by PPM. If $f(m) = Am^2 + Bm + C$ is the function that, for a given Johnson V PPM magnitude m, it returns the magnitude $f(m)$ in CD scale, then the coefficients obtained in the regression are: $A = -0.01335368$, $B = 1.076636$, $C = 0.2249828$, and the quadratic mean error is $\sigma_m = 0.2759$.

Addition of color and double stars to the digital catalog

In this stage, footnotes of pages 56 to 117 (stars with declination -23°) from printed catalog [2] was transcribed and stored in two text files: one for color stars with 90 records and the other for double stars with 194 records. Then, a simple algorithm saved this information to the digital catalog.

We noticed an error in the printed catalog: in page 65, CD -23° 3539 is marked as "color" although this star is not present in that page. On the other hand, CD -23° 2539 (a star from page 65) is bright and orange, so it is likely that there is a typo error in the printed catalog. A search in *SIMBAD* of CD -23° 2539 leads to an identification between this designation and HD 34087, a star with magnitude V of 7.4 and color index B – V = 1.03.

We also noticed that, in page 105, star CD -23° 14537 is marked as "cumulus". This makes sense because Messier 22 (NGC 6656) is located in the same position. We do not know why this star was not originally reported as a non-stellar object.

Besides the list of color and double stars, we also considered the possibility of transcribing cross references given in the printed version of CD, since they are absent in I/114. These references are given, for some CD stars, in an additional field where a combination of 2 letters is used for referencing 5 old catalogues where the star is displayed. However, the task proved to be very time-consuming and we discontinued it after the transcription of 266 records (around 8 pages of the printed catalog).

Comparison with other catalogues and correction of typo errors

There are stellar catalogues partially cross-referenced to CD. Thus, one can take advantage to compare parameters of stars (such as position or brightness) from these catalogues with those recorded in the digital version of CD and, in this way, find errors in the latter. Our hypothesis is based on the assumption that each parameter of CD is normally distributed with a standard deviation of σ whereas the same parameter is very accurate in the other catalog. Therefore, there is a high probability that a typo error in the digital version, i.e. I/114, is revealed when the parameter is compared against the other one and the difference is greater than 3σ, since only 0.3% of these differences would be due to the nature of the parameter.

Next, we describe the method we follow. We choose a catalog and scan all the records cross-referenced with a star CD -23°. For these records, we precess and correct by proper motion the position to B1875.0 and we get the magnitude on the CD scale. Then, an algorithm enumerates those cases where the difference in right ascension, declination or magnitude exceed a threshold, denoted $u_{\alpha*}$, u_{δ} and u_m respectively. Once the list is generated, dubious parameters are compared manually with the ones reported in the printed version. Thresholds were chosen in a way that the number of dubious cases are balanced for different parameters. The following catalogues were considered:

- *Hipparcos* (HIP). It is an astrometric catalog available at *VizieR* as I/239. The catalog is very accurate and almost all records are cross-identified to CD stars. Nevertheless, it has few records in contrast to CD: 811 records. After discarding HIP 62292 (CD -23° 10709) since there isn't enough information to precess its position to 1875, and those stars exceeding thresholds $u_{\alpha*}$ = 23.4 arcsec, u_{δ} = 51.9 arcsec, u_m = 0.84, we obtain 50 records. From these ones, only star CD -23° 9296 manifests a typo error.

- *Positions and Proper Motions* (PPM). It is an astrometric catalog available at *VizieR* as I/193 (South) and I/208 (Supplement). It is less accurate than HIP but it has more records cross-referenced to CD stars: 4583 records (roughly 25% of CD stars with declination -23°). However, 656 records do not report visual magnitude, so we only compare position for them. Thresholds are $u_{\alpha*}$ = 34 arcsec, u_δ = 62 arcsec, u_m = 0.75. We obtained 84 records for manual checking and we found typo errors in -23° 8777, -23° 9296 and -23° 12414. We also found a difference between printed catalog and I/114 for -23°5830, but this star is reported in the corrigenda of *NSSDC* [7], so we discarded it.

We also noticed that fields of -23° 12413 y -23° 12414 are swapped in I/114. We believe it was an attempt in the past to correct the order in right ascension. Since these stars are cross referenced in other catalogues according to the parameters given in the printed catalog, we turn back to the original parameters.

- *Cordoba A* (AGK). This catalog was made at the National Argentine Observatory in the early twentieth century as part of an international project named *Astronomische Gesellschaft Katalog*. It has 3389 records with cross references to CD -23°. The benefit of this catalog is that the coordinates of its stars was measured in almost the same epoch as CD, so error in position due to proper motion is low. Also, visual magnitudes given in AGK are similar to the ones given in CD, so they can be compared without making a transformation. Digital version can be downloaded from http://dc.zah.uni-heidelberg.de/arigfh/katkat/q/form by writing 986 in the field *Teleki*.

The catalog presents a drawback: each AGK record has the CD *yyyyy* number, but the declination *-xx*° is missing. Therefore, the latter value must be inferred from the CD catalog with a reasonable criterion. In our case, we choose the nearest CD star. For instance, if AGK 11856 is identified with CD 13169, and the distance between AGK 11856 and CD -23° 13169 (21 arcsec) is smaller than the one between AGK 11856 and CD -24° 13169 (461 arcsec), we can guarantee that AGK 11856 is really matched with CD -23° 13169. Thresholds are $u_{\alpha*}$ = 34 arcsec, u_δ = 62 arcsec, u_m = 0.96, and we obtained 107 records. Then, we found errors in -23° 8777, -23° 9296, -23° 9705 and -23° 12414.

- *Cape Photographic Durchmusterung* (CPD). This catalog, also at *Vizier* as I/108, was made in the same epoch as CD and we can use it for comparing against CD without the need to correct positions of stars by proper motion. Although cross references are missing in the CPD catalog, a cross identification performed by Rappaport and Warren between CD and CPD [8] is available at: ftp://dbc.nao.ac.jp/DBC/NASAADC/catalogs/4/4019

We have 8091 records with cross references to CD -23°. Since CPD does not report visual magnitude, we only compare positions. Thresholds are $u_{\alpha*}$ = 64 arcsec and u_δ = 94 arcsec, and we obtained 392 records. Then, we found errors in -23° 5485, -23° 6942, -23° 7042, -23° 8777, -23° 11817 and -23° 12414.

Cross-identification between CD and other catalogues

Currently, there are cross-identifications between CD and other catalogues such as HIP, SAO or PPM. Nevertheless, since CD has a larger number of stars than these catalogues, several stars of CD are not identified yet. As we have said in the introduction, CD -22° 2 is an example of a non-identified star. However, in the last decades new catalogues become available with a huge number of stars and other celestial objects (for instance, USNO-B1.0 has 1042618261 records), so now it is possible to perform a cross-identification between CD and one of these new catalogues, in order to assign a record to each CD star.

We prepared a cross-identification between CD and the catalog *Positions and Proper Motions eXtended* (PPMX, in *VizieR* I/312), since it has enough stars and, at the same time, it has enough accuracy to be able to obtain reliable positions of their stars for epoch B1875.0. Each record of this catalog has the following parameters: *coordinates* (RAJ2000, DEJ2000) and their *errors* (e_RAJ2000, e_DEJ2000) for a given *epoch* (epRA, epDE), *proper motion* (pmRA, pmDE) and its *error* (e_pmRA y e_pmDE), *magnitude* in Johnson V (Vmag) and its *error* (e_Vmag), among others [9]. We consider only those stars with magnitude up to 13.5 and with declination -23° for epoch B1875.0 (plus 10 arcmin in both sides). We get 119230 stars of PPMX.

A cross-identification is usually obtained by comparing parameters of two catalogues with a metric, e.g. the angular separation between positions of both records (reduced for the same epoch). Then, a correspondence between records is performed via a *matching* algorithm that chooses such assignments that minimize those angular separations. This technique is widely used nowadays, mainly in *Virtual Observatories* such as X-Match of Strasbourg Data Center. However, metrics used in these tools only consider the position in the sky whereas there are better metrics in the literature that include magnitude and other physical parameters [10], [11]. On the other hand, these tools do not return a one-to-one correspondence.

In order to achieve our goal, we designed and implemented a software that avoids the previous drawbacks and fulfills the following features:

- The generation of a cross-identification such that, for a given single CD star, a PPMX star is assigned (or none alternatively) and, for a given double CD star, two PPMX stars are assigned (or one or none alternatively).
- In the case of ambiguity, the algorithm chooses the most "probable" assignment.
- For single stars, the metric approximates the probability of the identification between a CD star and a PPMX star, taking into account position, brightness and the corresponding standard deviations.
- For double stars, the metric also takes into account the probability of the identification between a CD star and a pair of PPMX stars. In other words, for a given pair of PPMX stars to be considered a "candidate pair", additional statistics such as the angular separation and difference in magnitude between both components are deemed.

Regarding the last item, it was necessary to determine the mean angular separation and the difference in magnitude between the components of a candidate pair. We performed an inspection and manual identification of 7 double stars (-25° 8168, -30° 8494, -23° 9730, -25° 10131, -25° 10846, -25° 11101, -24° 11610) and we noticed that the angular separation between their components varies from 15 arcsec to 1 arcmin, with an average value of 34.9 arcsec and standard deviation of 13.65 arcsec. In the case of the difference between magnitudes, it is clear that if this difference is large, the fainter star is difficult to observe visually. Assuming that this difference behaves as a normal distribution centered at zero, we estimate a standard deviation of 0.915.

In order to compute metrics, we also need the error in position and brightness for each star of CD. Since this information is unavailable, we assigned a constant error for every star in CD. We assumed that the error in magnitude is the same as the one given by the quadratic fit in Section 2, i.e. $\sigma_m = 0.2759$. Regarding the error in position, two options could be chosen. The first one, to use the values suggested by Thome (which were computed by comparing positions between CD and SD catalog of Schönfeld which has low accuracy). The second one, to compare positions between CD and a modern catalog. We chose the latter option and we used PPM catalog. We first discarded 3 stars marked in PPM as "astrometrically problematic", and another 253 stars whose differences in right ascension/declination exceed 2 arcmin with respect to the same CD star. From the 4327 remaining records, we got $\sigma_{\alpha*} = 10.15$ arcsec y $\sigma_\delta = 22.74$ arcsec. Note that our values are similar to the ones obtained by Thome. PPM catalog was chosen since it has many stars cross-referenced with CD stars, and its accuracy is high. For instance, the star with highest uncertainty, i.e. the one that maximizes $\sigma^2_{\alpha*} + \sigma^2_\delta$, where $\sigma_{\alpha*}$ and σ_δ are standard deviations propagated to epoch B1875.0, is PPM 266698 with $\sigma_{\alpha*} = 0.57$ arcsec and $\sigma_\delta = 0.61$ arcsec (this fact is mainly due to the high uncertainty in the proper motion, which propagates at a rate of 6.3 and 6.2 mas/year over 98 years, since Ep = 1973). Note that these values are relatively small compared to the uncertainty of CD stars.

Now, we describe briefly our algorithm, which has been implemented in C++. Technical details will be the subject of an upcoming publication. Although the algorithm is designed to make a cross-identification between CD and PPMX, the ideas presented here can be applied without trouble to any pair of catalogues. These catalogues should comprise information about positions, brightness and their standard deviations, and possibly double stars in one of them.

In the first place, coordinates of PPMX stars are precessed and corrected by proper motion to besselian epoch of 1875 with subroutine *wcsconp* available in the library WCSTools [12]. The error is also propagated [13], i.e. standard deviations of right ascension and declination are computed from parameters given in PPMX catalog.

Then, a structure in Discrete Mathematics known as *bipartite graph* is created: each CD and PPMX star is represented by a vertex of this graph and, for a given pair of stars *(x,y)* where *x* is a CD star and *y* is a PPMX star, an edge between their corresponding vertices is added only if the angular separation between *x* and *y* is lower than a threshold (in our case, 6 arcmin). In order to avoid misidentifications, CD stars with declinations -22° and -24° also were considered. The resulting graph has 171904 vertices and 320542 edges.

In the next step, the graph is decomposed into connected components with a *Kruskal*-like algorithm. Our graph has 31092 components: 22862 isolated vertices, 3960 components consisting of one CD star and one or more PPMX stars (in this case the identification is trivial, since one assigns the most probable PPMX star to that CD star) and 4270 remaining components. The largest connected component corresponds to a densely populated area of stars of the Milky Way, with 10249 CD stars (of which 71 are double), 33071 PPMX stars and 133162 edges.

For a given component, cross-identification is performed as follows. If the component has double stars, candidate pairs (y_1, y_2) of PPMX stars are generated, where y_1 is the main star and y_2 the secondary one. We established a criterion based on angular separation and difference in magnitude. On the one hand, the distance between y_1 and y_2 must be lower than 80 arcsec. On the other, y_1 must not be fainter than y_2 in 1 unit of magnitude. The largest connected component of our graph has 1609 candidate pairs with this criterion.

Then, for each edge (x,y) with a single star x, the probability that x and y are the same star is computed. Similarly, for each edge (x,y_1) with a double star x, the probability that x and (y_1,y_2) are the same object is computed for all y_2. Also, probabilities that a star x is not matched with any star of PPMX, or a double star x is only matched with a single PPMX star are computed. The search of the *best* matching, i.e. that maximizes the overall probability, is performed through the optimization of an Integer Linear Programming Model and solved with GuRoBi 6.0.2 (http://www.gurobi.com). The objective function of the model sums the logarithm of the probabilities of the matched edges. In our graph, the largest connected component yields a model with 144167 binary variables and 43320 linear constraints, although it is solved in less than 4 seconds of time.

The overall processing time was 1114 seconds, over 1 CPU thread. The process was carried out on a computer equipped with an Intel i5 2.67Ghz.

Files and format of the new catalog

Two text files have been generated for storing the digital version of the new catalog. The first file, *cd.txt*, has the new version of CD catalog, including every star and non-stellar object given in the printed version for the declination -23° in *VizieR* format (similar to the one used in I/114), which is detailed as follows:

```
Byte-by-byte Description of file: cd.txt
--------------------------------------------------------------------------------
   Bytes Format  Units    Label       Explanations
--------------------------------------------------------------------------------
   1-  2  A2     ---      ---         [CD] The catalog prefix
   3-  5  I3     deg      zone        [-23] The declination zone
   6- 10  I5     ---      num         The number of the star within the zone
      11  A1     ---      suppl       [ ] Corrigenda flag: not needed here
  12- 15  F4.1   mag      mag        *Estimated visual magnitude
  16- 17  I2     h        RAh         Hours of right ascension, 1875
  18- 19  I2     min      RAm         Minutes of right ascension, 1875
  20- 23  F4.1   s        RAs         Seconds of right ascension, 1875
      24  A1     ---      DE-         [-] Sign of declination
  25- 26  I2     deg      DEd         Degree of declination, 1875
  27- 30  F4.1   arcmin   DEm         Minutes of declination, 1875
      31  A1     ---      dpl        *[D ] Double flag
      32  A1     ---      color      *[C ] Color flag
--------------------------------------------------------------------------------

Note on mag:
   20.0 = neb; 30.0 = var
Note on dpl:
   D = the star is double visually
Note on color:
   C = the star is yellow or red
```

The second file, *cross.txt*, stores the cross-identification between CD and PPMX, in the following format:

```
Byte-by-byte Description of file: cross.txt
--------------------------------------------------------------------------------
   Bytes Format  Units    Label       Explanations
--------------------------------------------------------------------------------
   1-  2  A2     ---      ---         [CD] The catalog prefix
   3-  5  I3     deg      zone        [-23] The declination zone
   6- 10  I5     ---      num         The number of the CD star within the zone
      11  A1     ---      suppl       [ ] Corrigenda flag: not needed here
      12  A1     ---      type       *[SUAB] Type of identification
  13- 18  F6.2   arcsec   dist        Angular separation between position of CD star
                                      and position of PPMX star
  19- 22  A4     ---      ---         [PPMX] The catalog prefix
  23- 37  A15    ---      PPMX        Name of PPMX star (HHMMSS.S+DDMMSS)
      38  A1     ---      m_PPMX      [pf] if name of PPMX star is identical
--------------------------------------------------------------------------------

Note on type:
   S = CD star is single
   U = CD star is double, but it was identified with a unique PPMX star
   A = PPMX star corresponds to the component A of CD star
   B = PPMX star corresponds to the component B of CD star
```

We also uploaded the following additional files: *color.txt* and *dpl.txt*, have lists of color and double stars prepared in the first stage; *ref.txt*, has a partial list of cross references prepared in the first stage; *comp_hip.txt*, *comp_ppm.txt*, *comp_agk.txt* and *comp_cpd.txt*, have lists of CD stars automatically generated in the second stage; *changes.txt*, has a list of modifications made to the digital catalog, in the format given in I/114/cdchg.

Conclusions

In this article, we described the initial phase of a greater project which consists of making a new version of Cordoba Durchmusterung, as faithful as possible to the original catalog, i.e. with the fewest transcription errors and with all the information the printed catalog provides. To estimate the scale of this project, we addressed only a small area of sky and concluded the following:

The transcription of color and double stars in the first stage was a straightforward task that required two or three hours. We conclude that it is feasible to address the remaining part of the CD catalog. Moreover, the availability of the list of double stars made it possible to properly identify them. Regarding the transcription of cross references given in CD, we noticed that the task was time-consuming and therefore we declined to continue it.

In the second stage, we noted that it isn't worth using several comparison catalogues, since we found the same typo errors in most of these catalogues whereas each catalog contributes few additional errors on its own. We conclude that PPM as the unique comparison catalog will be sufficient for the remaining part of CD.

The third stage is the most interesting one, since a new line of research is derived from it: how to identify double stars. Below, we propose a list of further works:

- Explore several metrics from the literature and propose new ones. Use these metrics on instances where a cross-identification is available and test them in order to check its success ratio. Also, try to integrate the information about star's color for improving the quality of the metric.
- In particular, for CD, use a larger sample of double stars in order to infer "angular separation" and "difference in magnitude" parameters. For other catalogues, propose new parameters for carrying out cross-identifications of stars with 3 or more components.
- Modify the current software tool in order to allow the computation of cross-identifications on any pair of catalogues.
- It is easy to see that the cross-identification problem is *polynomial* when there are no double stars. In that case, an instance can be solved efficiently through a matching algorithm such as Hungarian Algorithm [14]. However, a study of the *computational complexity* (and how to address it later) should be done when double stars are present in one of the catalogues.

Acknowledgements

We thank Ms. Andrea Jimenez, Mr. Catriel Caruso, Ms. Noelia Acosta Pedemonte, and our colleagues Guillermo Ibañez and Silvia Morales for their collaboration in this work. We are especially grateful to Prof. Santiago Paolantonio for answering our questions on the subject.

References

(1) MINNITI, E.; PAOLANTONIO, S. Córdoba Estelar. Desde los sueños a la Astrofísica. Historia del Observatorio Nacional Argentino. *Editorial Universidad Nacional de Córdoba*, 2013.

(2) THOME J. M. Zonas de Exploración (Córdoba Durchmusterung): declinación -22° a -32°. *Resultados del Observatorio Nacional Argentino*, 1892, Vol. 16.

(3) GLASBY, J. S. The Variable Star Observer's Handbook. 1st. Edition. *Sidgwick and Jackson*, 1971.

(4) LEQUEUX J. From Flamsteed to Piazzi and Lalande: new standards in 18th century astrometry. *Astronomy & Astrophysics*, 2014, Vol. 567, A26 pp. 1--9.

(5) MINNITI, E.; PAOLANTONIO, S. Uranometría Argentina 2001. Historia del Observatorio Nacional Argentino. Edición digital. http://historiadelaastronomia.wordpress.com/documentos/uranometria-argentina/uabicentenario

(6) THOME J.M. List of additional errors found in the Cordoba catalogues. *The Astronomical Journal*, 1896, Vol. 16, pp. 30.

(7) NASA Reference Publication 1299. Cordoba Durchmusterung, volume 1-4. *Technical Report, NASA Goddard Space Flight Center*, 1993.

(8) RAPPAPORT, B. N.; WARREN, W. H., Jr. Cross-identification of the Cordoba and Cape Photographic Durchmusterung. *Bulletin of the American Astronomical Society*, 1986, Vol. 18, pp. 897.

(9) ROESER, S.; SCHILBACH, E.; SCHWAN, H.; KHARCHENKO, N. V.; PISKUNOV, A. E.; SCHOLZ R.-D. PPM-Extended (PPMX), a catalogue of positions and proper motions. *Astronomy & Astrophysics*, 2008, Vol. 488, pp. 401--408.

(10) SUTHERLAND, W.; SAUNDERS, W. On the likelihood ratio for source identification. *Monthly Notices of the Royal Astronomical Society*, 1992, Vol. 259, pp. 413--420.

(11) BUDAVARI T.; SZALAY A. S. Probabilistic Cross-Identification of Astronomical Sources. *The Astrophysical Journal*, 2008, Vol. 679, pp. 301--309.

(12) *WCSTools*: Image World Coordinate System Utilities. Harvard-Smithsonian Center for Astrophysics. http://tdc-www.harvard.edu/wcstools

(13) KOVALEVSKY J.; SEIDELMANN P. K. Fundamentals of Astrometry. *Cambridge University Press*, 2004.

(14) COOK W. J.; CUNNINGHAM, W. H.; PULLEYBACK, W. R.; SCHRIJVER A. Combinatorial Optimization. *Wiley-Interscience*, 1998.