WAVELET ANALYSIS OF HUGE STELLAR CATALOGUES

A.S. TSVETKOV, V.V. VITYAZEV, I.I. KUMKOVA
St. Petersburg State University
University pr., Peterodvorets, St. Petersburg, 198504, Russia
e-mail: a.s.tsvetkov@inbox.ru; vityazev@list.ru

ABSTRACT. We present the wavelet technique for searching the heterogeneities of stellar density in the data of NOMAD (Naval Observatory Merged Astrometric Dataset) catalogue which contains more than a billion stars. The known and unknown globular and open clusters have been detected in various photometric bands up to V=18. A lot of artifacts in NOMAD data were found in addition. This technique can be used for catalogues of the nearest future including the products of GAIA mission.

1. NOMAD DATABASE

The huge stellar catalogues which became available in the last years require new processing procedures. At present, the world largest stellar catalogue is the Naval Observatory Merged Astrometric Dataset (NOMAD). It contains more than 1 billion stars.

The NOMAD catalogue is a compilation of six astrometric catalogues – Hipparcos, UCAC2, YB6, Tycho-2 and USNO-B. In addition, it includes the photometric data from the 2MASS project. Finally, photometry in six bands (B, V, R, J, H, K) is available for about the half of stars. The distribution of NOMAD stars via their magnitude V is shown in Fig. 1

The average stellar density of the catalogue is tens of thousand stars per square degree, in some directions it is about 1 million stars per square degree. It is interesting that the catalogue contains even the stars belonging to bright galaxies. This abundance of data poses the problem of searching the unknown stellar clusters in our Galaxy. This paper demonstrates the application of the wavelet technique for detecting the irregularities in stellar density.

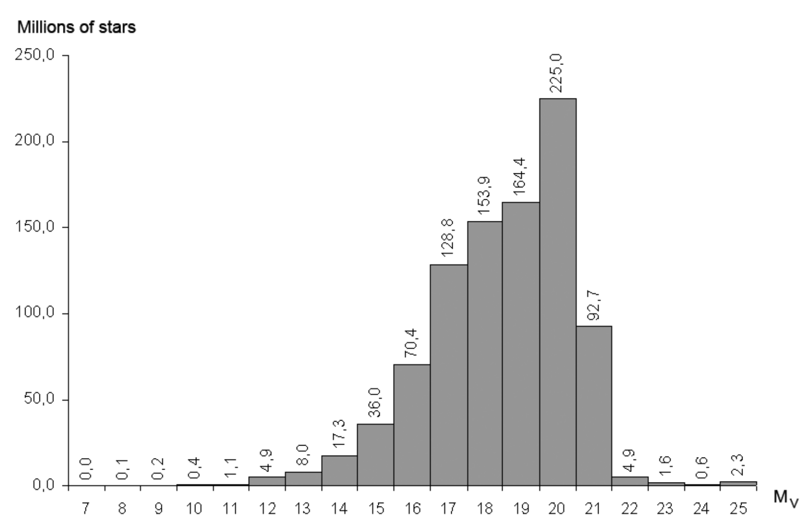

Figure 1: Distribution of NOMAD stars via magnitude V

2. ONE DIMENSIONAL WAVELET TRANSFORM

Processing the giant catalogues is severe time consuming work. To make the search of clusters as fast as possible we need to use the simplest wavelet technology. Let us consider the stellar density as a one-dimensional function of longitude along a circle of some galactic latitude. In practice, the average densities in narrow belts may be used. Since we are hunting the small scale features the 1-D wavelets may be used.

Let x_k, $k = 1...N$ are the discrete values of stellar density, Δt is the interval of sampling:

$$x_k = x(t_k), \qquad t_k = \Delta t \cdot k, \qquad k = 1, 2, \ldots, N. \tag{1}$$

We use the MHAT (Mexican hat) wavelet function

$$\psi(t) = (1 - t^2)e^{-t^2/2}. \tag{2}$$

It is important to note that the zero and first-order moments of the function (2) vanish.

The discritized wavelet transform can be written as follows:

$$W(a_i, b_j) = \frac{1}{n(a_i, b_j)} \sum_{k=1}^{N} x_k^{\circ} \psi^* \left(\frac{t_k - b_j}{a_i} \right), \qquad n(a_i, b_j) = \sum_{k=1}^{N} \exp \left(-\frac{1}{2} \left(\frac{t_k - b_j}{a_i} \right)^2 \right). \tag{3}$$

Here

$a_i = i\Delta a$ is a wavelet scale, $i = 1, \ldots, N_a$; $N_a < N$ defines the value of index i, corresponding to maximal scale;

$b_j = j\Delta b$ is a coordinate shift, $j = 1, \ldots, N$.

It is reasonable to adopt equal values for the scale and coordinate shift intervals:

$$\Delta a = \Delta b = \Delta t. \tag{4}$$

Further, the function (2) has strong localization in spatial domain. It is possible to show that

$$\int_{-L}^{+L} (1 - t^2)e^{-t^2/2}\, dt < 0.01 \tag{5}$$

for $L \approx 3.6$. It allows to decrease noticeably the number of terms in equations (3). Defining the "index-radius" of wavelet as

$$j^* = \left[\frac{La_i}{\Delta b} \right], \tag{6}$$

and taking into account (4) and (6) we finally obtain

$$W_{i,j} = \frac{1}{n_{i,j}} \sum_{k=j-j^*}^{k=j+j^*} x_k^{\circ} \psi^* \left(\frac{k-j}{i} \right), \qquad n_{i,j} = \sum_{k=j-j^*}^{k=j+j^*} \exp \left(-\frac{1}{2} \left(\frac{k-j}{i} \right)^2 \right). \tag{7}$$

The problem of edge effect (the restriction of admissible value of index j in equations (7)) is not important in our case since our functions are periodical: ($x_{k+N} = x_k$ and $x_{k-N} = x_k$). Equations (7) make the computational procedure very fast.

It is a common practice to use the squares of $W_{i,j}$. In our task we prefer the values

$$S_{i,j} = W_{i,j} \cdot |W_{i,j}, \tag{8}$$

since the positive and negative values $S_{i,j}$ help distinguish between congregation and rarefication of stars. The array of $S_{i,j}$ (or $|S_{i,j}|$) is often called the *scalogram*.

The left image in Figure 2 demonstrates the wavelet analysis of the simulated distribution of x_k. One can see the different representation of two artificial "clusters" of different size. The "diffraction" effects with negative density are noticeable near sharp boundaries of the heterogeneities.

The right image represents the results of wavelet analysis of a strip along the small circle of celestial sphere bounded by galactic latitudes $+31.5° \leq b \leq +32.5°$. The set of data x_k is the average values

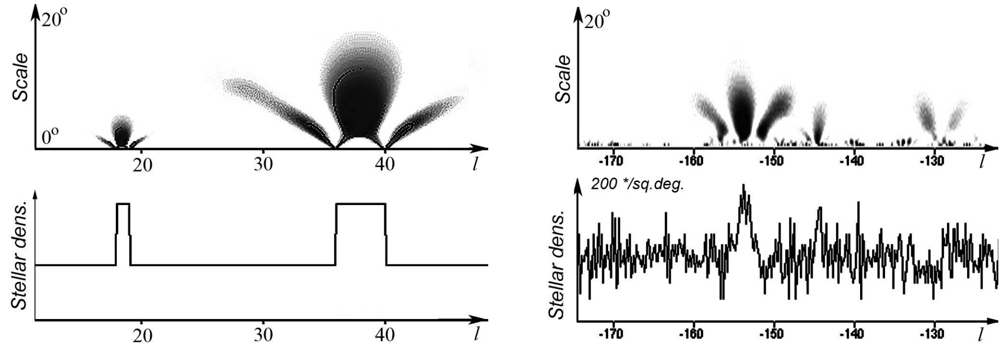

Figure 2: Wavelet transform of artificial (left) and real (right) data

of stellar density for stars $10 \leq J \leq 12$ within the interval $\Delta t = 10'$. Near $l = -153°$ one can see the distinct concentration produced by open cluster M 44 (Praecepe – "Hive"). We will see later that wavelet analysis can detect not only such obvious concentrations but substantially smaller ones.

3. WAVELET SURVEY OF THE CELESTIAL SPHERE

The fast algorithm for searching local heterogeneities in stellar density was proposed. It consists of the following steps:

- The sampling of stars from NOMAD according to some selection criteria (for example, the magnitude range);
- The forming of narrow strips (1° or 0.5° wide) along circles of equal galactic latitude. These strips are divided into small areas with constant longitude intervals ($10'$ or $5'$);
- Calculating the values x_k of average stellar density for each area in all strips;
- Application of the one dimensional wavelet technique to all strips.

The transformation of a scalogram (2-D array $S_{i,j}$) into *a scalegram* (1-D array G_j)

$$G_j = \frac{1}{i_{max} - i_{min} + 1} \sum_{i=i_{min}}^{i_{max}} S_{i,j} \tag{9}$$

allows to visualize all the celestial sphere using for example the Aitoff-projection (Fig. 3). The variation of values i_{min} and i_{max} results in the different visible details in the scalogram.

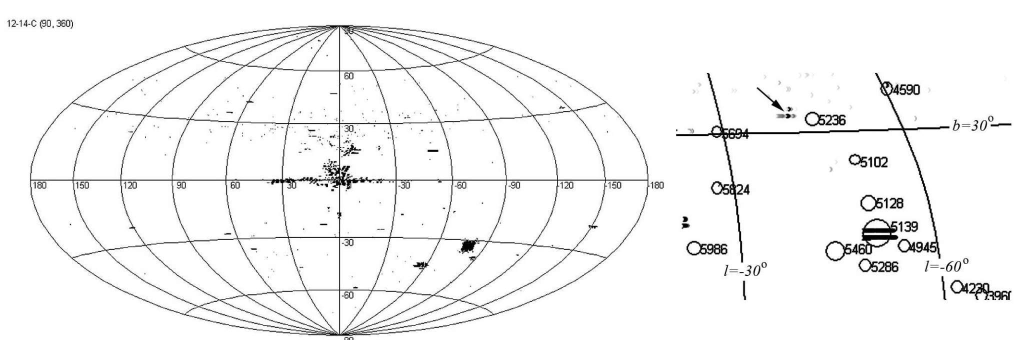

Figure 3: The general view of wavelet survey for stars of $12^m - 14^m$ (left) and its detail (right)

The analysis of an image like Fig. 3 gives evidence of three kinds of objects:

- The known clusters detected by wavelet technique;
- The known clusters which are not detected by wavelet technique;
- The unknown objects detected by wavelet technique.

The first class of objects is obvious. The explanation of the second situation is that some clusters are not visible in all magnitude ranges. For instance, the Pleiades (M 45) is imperceptible in the sample of faint stars ($12^m - 14^m$) but is easily detected in the sample of bright stars ($8^m - 10^m$).

The most interesting class of objects is the third one. Probably, we have found some unknown stellar clusters, for example, in the region near $l = -45.5°$, $b = +32°$ (Fig. 4, left). Still, we could not determine the origin of some strange objects like (Fig. 4, center).

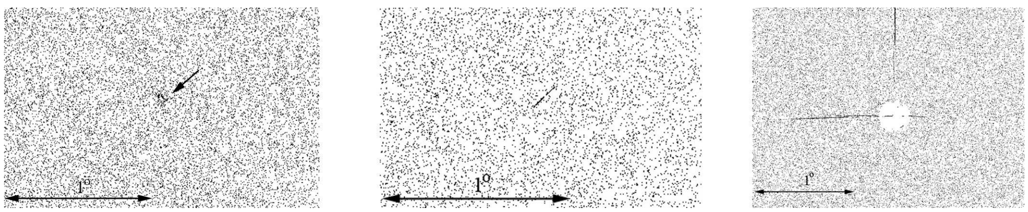

Figure 4: The unknown cluster (left), the object of unknown origin (center) and the artifact (right)

The most prevalent unknown objects are found to be the observational artifacts of the catalogue. The Fig. 4 (right) shows distinctly that the catalogue data contains even the diffraction rays near bright stars as usual stars.

4. CONCLUSION

Earlier, nebulous spots in the sky frequently fooled comet discoverers. To help them Charles Messier compiled his famous catalogue of stellar clusters, galaxies and nebulae visible in telescopes of XVIII century. Now, diffraction spikes and other artifacts in NOMAD put obstacles in the way of stellar cluster hunting. To help detecting new stellar clusters one need to have a catalogue of all known artifacts in the biggest data set of the XXI century. Our further goals are: the compilation of a catalogue for all detected aggregates in NOMAD; census of the NOMAD's artifacts and, finally, constructing the catalogue of new stellar clusters.

5. REFERENCES

NOMAD, Naval Observatory Merged Astrometric Dataset: `http://www.navy.mil/nomad` (2002).

The Hipparcos and Tycho Catalogues 1997: ESA SP-1200 (1997).

UCAC1: Zacharias, N. et al., 2000, ASP Conference 216, 427.

UCAC2: Zacharias, N. et al., 2004, AJ 127, 3043.

YB6: Monet, D.G., 2004: complete scan of NPM and SPM plates.

Tycho-2: Hog, E. et al. 2000, A&AS 355, L27

USNO-B: Monet, D. G. et al., AJ 125, 984 (2003)

2MASS: `http://www.ipac.caltech.edu/2mass/releases/allsky/`

J. D. McEwen, M. P. Hobson, D. J. Mortlock, A. N. Lasenby, 2005, IEEE Transactions on Signal Processing.

Sulentic, J.W., Tifft, W.G., 1977, Revised New General Catalogue of Nonstellar Astronomical Objects, Tuscon.

Harris W.E., 1996, A catalog of parameters for globular clusters on the Milky Way, AJ 112, 1487.

Dias W.S., Alessi B.S., Moitinho A., Lepine J.R.D., 2002, New catalog of optically visible open clusters and candidates (V2.8), A&AS 389, 871.

Vityazev, 2001, V.V. Wavelet analysis of time series, SPb (in Russian)