N.N. Samus[123]

[1]*Institute of Astronomy, Moscow*

[2]*Sternberg Astronomical Institute, Moscow University*

[3]*Eurasian Astronomical Society*

Big Data and Variable Stars

BRICS Astronomy Workshop

“Astronomical Data and Computation”

Ekaterinburg, Russia, September 6, 2016

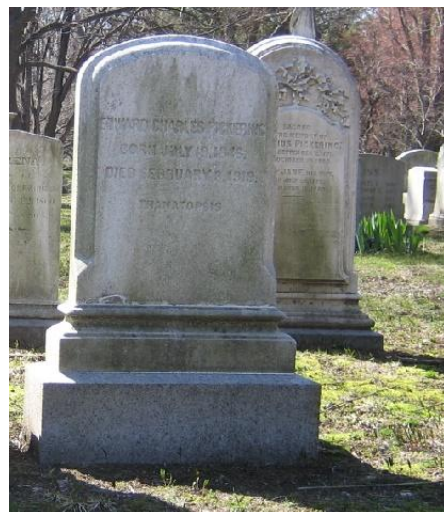
The first catalog of variable stars (E. Pigott, 1786) contained 12 objects. Almost a century later (1877), F.W.A. Argelander's catalog consisted of 18 stars (variability of one of them remains unconfirmed till now). Typically, there were no more than 100 photometric observations (usually much less) per star.

F.W.A. Argelander
1799–1875

I'll try to demonstrate that the science of variable stars has become a science of big data in many aspects by now and that the volume of data creates a number of problems remaining unsolved.

Since Argelander’s times, especially since the advent of astronomical photography in variable-star science (first introduced as a general tool of variable-star research in Harvard Observatory by its Director, Edward Pickering), the number of known variable stars shows a quick growth. By the end of the 20th century, CCD studies resulted in a new acceleration.

E. Pickering (1846–1919)



Pickering’s grave at Mt.Auburn cemetery in Cambridge, MA, USA. Photo of 2009.

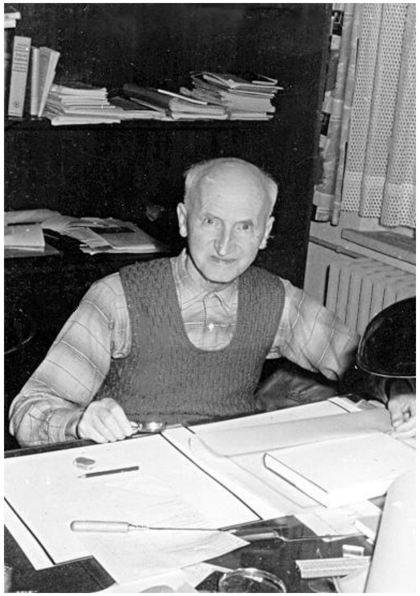
The era of analog astronomical photography almost exactly covers the calendar 20th century. The number of analog astronomical photographs in the world is estimated to exceed 2 millions. Collecting information on stacks of astronomical photographs kept at observatories of the world is coordinated by M.K. Tsvetkov (Bulgaria) in the WFPDB database,
http://www.wfpdb.org

Dr. M. Tsvetkov

Ekaterinburg, September 5–7, 2016

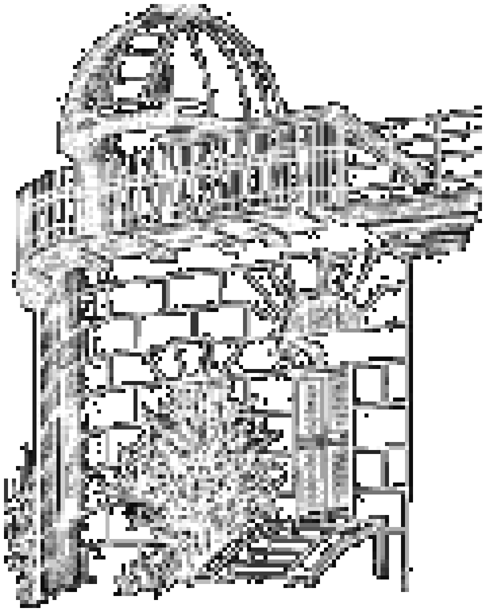
Harvard Observatory keeps the world-largest plate archive (about 500000 plates). Most of them were taken for variable-star research. The Europe's largest plate stacks are those of Sonneberg Observatory (about 300000). These plates were also taken mainly for variable-star research.

Construction of the new building of Sonneberg Observatory (Thuringia, Germany, 1957 – 1958)

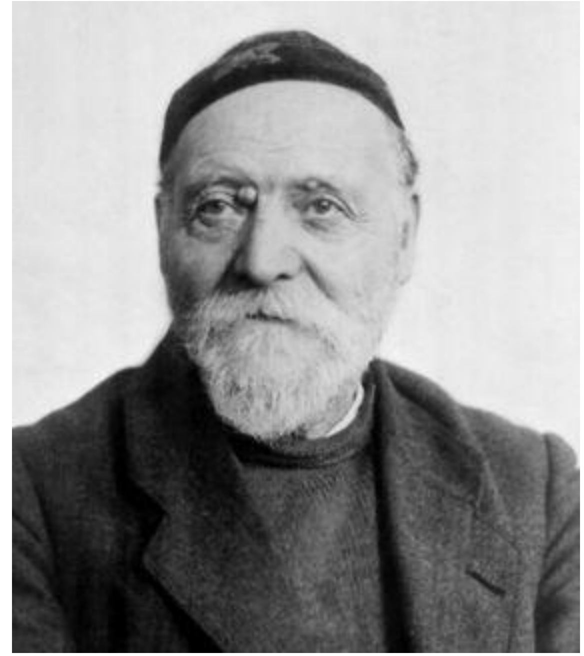


Cuno Hoffmeister (1892–1968) founded Sonneberg Observatory in 1925. He personally discovered about 10000 variable stars.

Asiago Observatory in Northern Italy (belongs to Padua University) keeps a collection of some 50000 photographs taken with two Schmidt telescopes (67/92 and 40/50 cm).

Odessa Observatory (Ukraine) keeps 104000 plates in its plate stacks, apparently the largest collection in the former USSR. About 84000 of these plates were taken in Mayaki village using the 7-telescope system shown above. Some 10000 plates in this archive are survivors of the old Simeiz stacks.

First attempts to take sky photos at Moscow's Presnya Observatory: Prof. Aristarch Belopolsky (1854–1934) in 1883, before his move to Pulkovo (1888), using wet colloid emulsions.

Presnya Observatory still exists, now belonging to Sternberg Institute. 1954: the new SAI building at Lenin Hills. 1950s: the new observing station in Crimea. 2014: the first light of the 2.5-m telescope at a new SAI's observatory near Kislovodsk.

Crimean laboratory of the Moscow University with its 125-cm reflector

The dome of the new 2.5-m reflector near Kislovodsk

SAI's 2.5-m telescope

The present Moscow plate stacks were founded by S.N. Blazhko (1870–1956) in 1895.

Cases of Moscow plate stacks

MOSCOW PLATE STACKS (SAI)

D, cm	*F*, cm	Field, degrees	m_{lim}	Years	*N*	Site(s)
10	64	20×28	13–14	1895–1933	1100	Moscow
16	82	16×22	14	1933–1956	2700	Moscow
23	230	6×6		1955–1991	10000	Moscow etc.
38	640	1.4×1.4	14	1902–1972	6400	Moscow
40	160	10×10	17–18	1948–1996	22300	Kuchino, Crimea
50	200	3.5×3.5	18–19	1958–2004	10000	Moscow, Uzbekistan
50	200	Spectra		1959–2004	2300	Crimea
70	1050	0.6×0.6	13–18	1961–1995	9500	Moscow

Plus other plates and films of inferior importance (from Shugarov, Antipin, Samus, and Danilkina 1999, with corrections and additions).

**Modern use of plate stacks requires digitization.
One of those who developed the machine used to digitize Harvard plate stacks, Bob Simcoe, near his machine. April 2009.**

Digitizing plates will keep them from eventual disasters, like that in Harvard Observatory on January 18, 2016. The Cambridge City water main under the observatory ruptured. 61000 plates were flooded. None of these plates were digitized by that time, some of computers and scanners also suffered. The rescue operation is under way; it is expected that serious losses of information will be avoided.

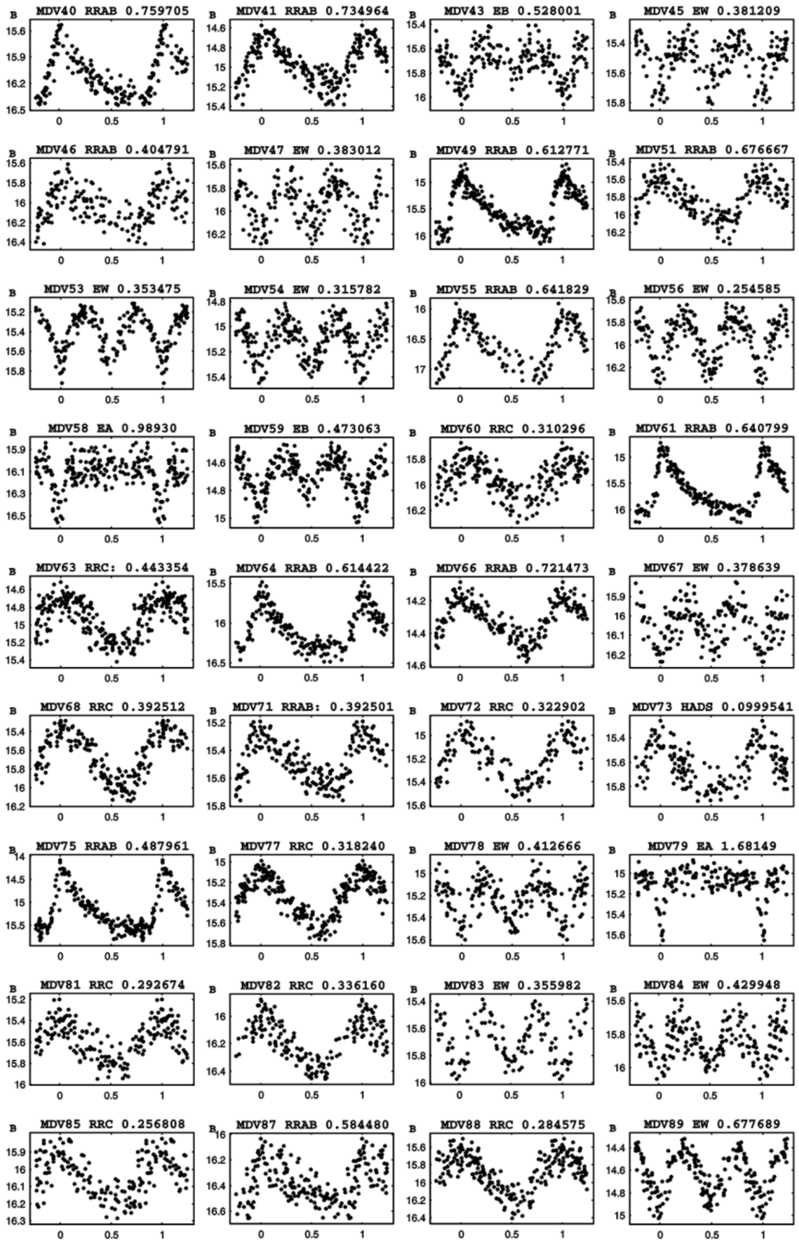
The basement level of the Harvard Observatory plate stacks (January, 2016)

CREO EverSmart Supreme scanners in SAI (were used in 2006–2011)

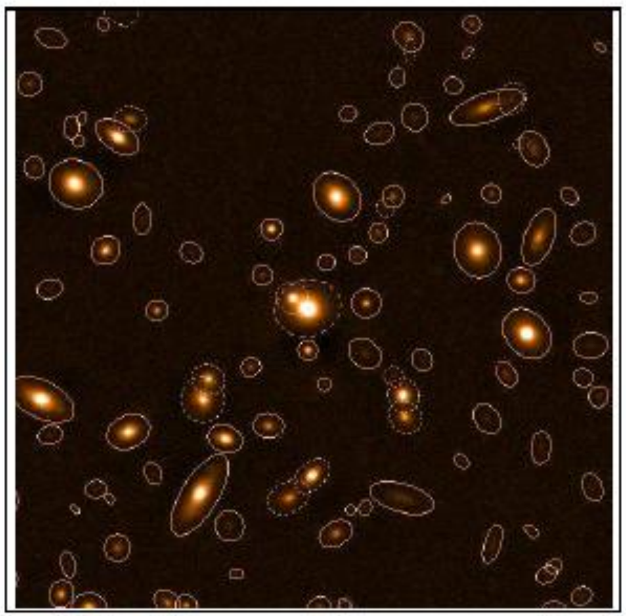
Ekaterinburg, September 5–7, 2016

The new Epson Expression 11000XL scanner, with a unit for scanning transparent large-size material (resolution 2400 dpi), was ordered upon advice from M.K. Tsvetkov and obtained on October 28, 2013. After that, we resumed scanning Moscow plate stacks. Each scan provides gigabytes of data.

Scans of Moscow plates are being searched for new variable stars using software developed and continuously improved by Dr. Kirill Sokolovsky. Though the plates were earlier searched for variables using traditional techniques, our automated search gives thousands of new discoveries. The results of new search are called “Moscow Digital Variables” (MDV).



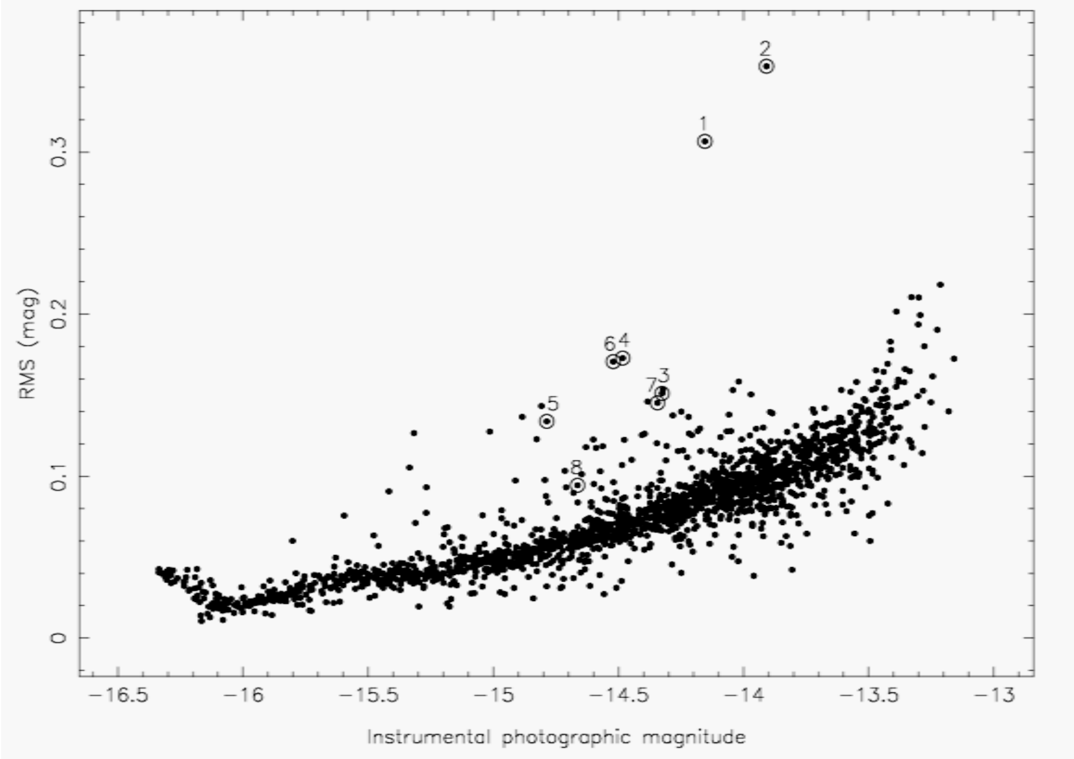
Reductions of scans



The SExtractor code;

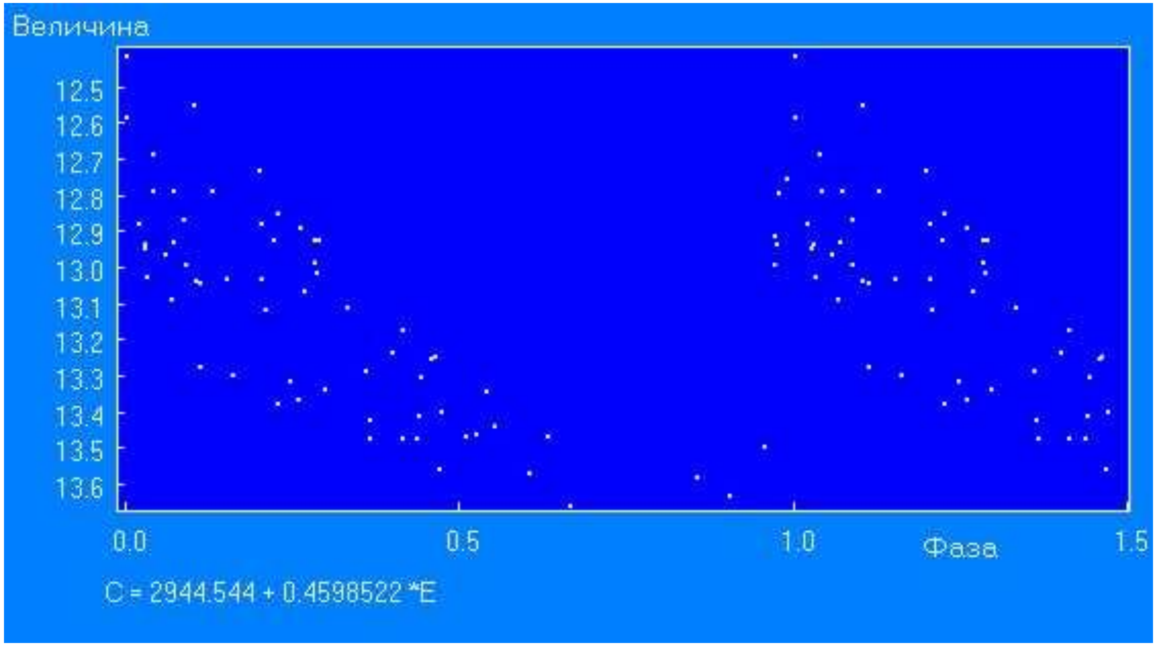
- **detects objects;**
- **determines their rectangular coordinates and estimates magnitudes (with an arbitrary zero point).**

For plates of the 40-cm astrograph, the full magnitude range requires two SExtractor runs, for bright and faint stars.



Identification of stars on different plates, zero-point correction, preliminary selection of variable-star candidates: the VaST code developed in our team by K.Sokolovsky and A. Lebedev. After the candidate selection, they are checked star by star and the period search is performed, usually with the incorporated WINEFK code (see below).

The WINEFK code by V.P. Goranskij (SAI) is an effective time-series analysis tool (e.g. for analyzing photometry of variable stars). Incorporates several period-search algorithms, provides possibilities of plotting light curves, rejecting bad observations, analyzing parts of the time series separately.



The ASAS-3 light curve of TX Pav generated with the WINEFK code

New versions of VaST already select periodic candidates as more promising candidates for variable-star discoveries.

VARIABLE STAR CATALOGS

E. Pigott 1786 12 stars (2 SNe, 1 Nova, 4 Miras, 2 Cepheids, 2 eclipsing stars, and the unique star P Cyg)

F. Argelander 1844 18 stars (1 still not confirmed)

E. Schoenfeld 1875 143

"Astronomische Gesellschaft" (issued yearly):

R. Prager 1926 2906

H. Schneller 1942 9476

Moscow compilers:

GCVS I 1948 Single volume, 10930

GCVS II 1958 Two volumes, 14708, cross-id tables

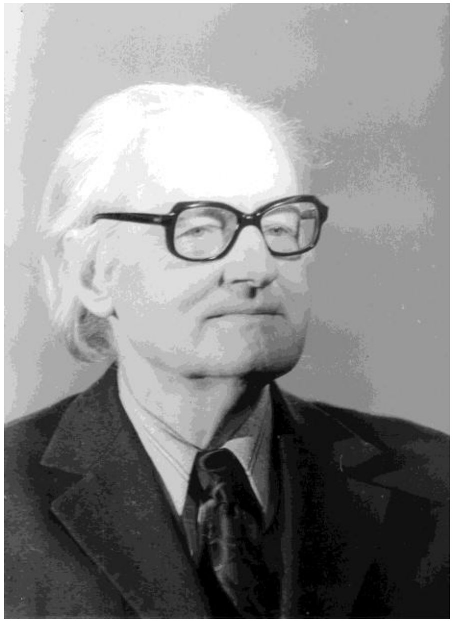
GCVS III 1969–1971 Three volumes, 20437

GCVS IV 1985–1995 Five volumes, 28435 plus about 12000 extragalactic variable stars, cross-id tables

By August, 2016 51853

GENERAL CATALOG OF VARIABLE STARS (GCVS): SINCE 1946, ON BEHALF OF THE IAU

P.P. Parenago
(1906–1960)



B.V. Kukarkin
(1909–1977)

P.N. Kholopov
(1922–1988)

Ekaterinburg, September 5–7, 2016

The GCVS team (photo of 2009)

Variability types for catalogs: hundreds of them (with subtypes) known. General users prefer to have few types (for better understanding). Advanced users want to have more and more types for categories of variables of their personal interest.

Decades of attempts to develop a usable system of automatic classification based on photometry. The catalog compilers would be happy with a system giving 2 – 3 percents of erroneous classifications; most systems gave 10 or more percents of mistakes.

Only very recently, usable systems did appear (for example, the machine learning package by D.-W. Kim and C.A.L. Bailer-Jones, 2016, based on 16 light-curve parameters, gives reasonably good results for seven main variability classes, with subtypes).

The GCVS is traditionally restricted to sufficiently well-studied variable stars (with at least variability type known) of our Galaxy (with the exception of those in globular star clusters). Special catalogs of variable stars in globular clusters are being compiled in Canada. The number of their objects can be roughly estimated as 6000.

Helene Sawyer Hogg (1905 – 1993), founder of the catalog of variable stars in globular clusters. Her work is continued by Christine Coutts Clement.

Extragalactic variable stars are normally outside the scope of the GCVS. The only GCVS attempt to cover them was made in the 4th edition (Volume V, 1995): 10979 variable stars (with the exception of supernovae) in 35 galaxies plus 984 supernovae. After that, the number of known extragalactic variables began to increase exponentially, and the GCVS team stopped to follow their discoveries.
SAI catalog of supernovae: D.Yu. Tsvetkov, N.N. Pavlyuk, O.S.Bartunov. More than 5000 entries. Its authors also write that the number of objects becomes too large to be managed on a star-by-star basis.

If catalogs like GCVS include only confirmed and sufficiently well-studied stars, we need so-called catalogs of suspected variable stars.

Astronomische Gesellschaft:

A catalog by E. Zinner and two catalogs by R. Prager (1929 – 1937), not replacing but appending each other (8020 stars)

GCVS team:

Replacement

CSV I (1951) 5835 + 2299 (1–5835, 100001–102999)

Continuation

CSV II (1965) 3079 + 838 (5836–8904, 103000–103137)

Replacement

NSV (New Suspected Variables) catalog (1981), 14811

Continuation (1998)

Numbers from 15001, 11206 stars

Principles of GCVS compiling, with human star-by-star analysis of information, are time-consuming, and thus GCVS is lagging behind the exponential growth of discoveries.
The American Association of Variable Star Observers (AAVSO) established the AAVSO Variable Star Index with a lower level of authorization. On August 15, 2016, it contained 398161 stars. The compilers also begin to feel overloaded with work.

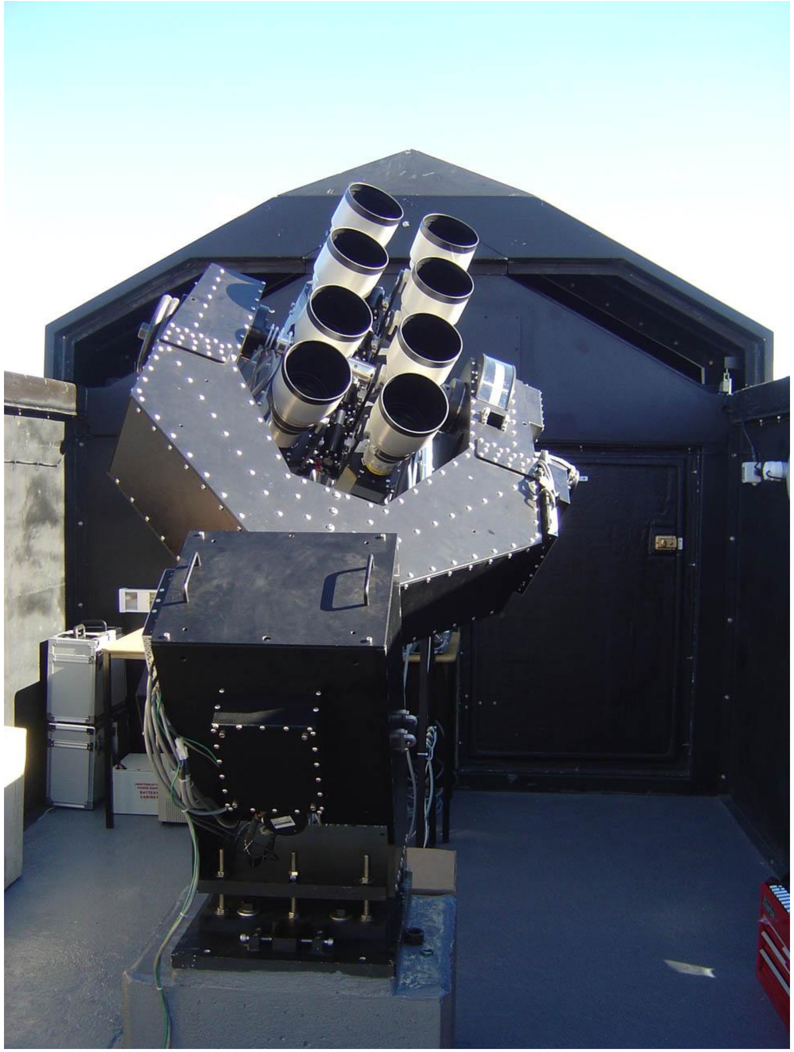
Sebastian Otero (AAVSO), Argentina, VSX coordinator

ASAS-3 (The All Sky Automated Survey)
Headed by Dr. G. Pojmanski (Warsaw)

- **All the southern sky, to +30^{o} in the north**
- **Small cameras, 2000 – 2009**
- ***V* photometric band, 2048 × 2048 CCDs, 8^{o}.8 × 8^{o}.8 field, 15.5″/pixel**
- **Millions of stars measured, with *V* down to 15^{m}.0 (more realistically, to 14^{m}.5)**
- **About 30000 new variable stars announced by the survey authors**
- **All data in open access, with possibilities to discover variables open to everyone**

NSVS (The Northern Sky Variability Survey)

- **A part of the ROTSE-I project**
- **The whole northern sky, to –38^{o} in the south, small cameras**
- **Only data for April1999 – March 2000 in open access**
- **Red instrumental system, 2035×2069 CCDs, 8^{o}.2 × 8^{o}.2 field, 14.4″/pixel**
- **14 million stars, *r* = 8^{m}–15^{m}.5**
- **Thousands of new variables discovered; possibilities of new discoveries almost exhausted by now**

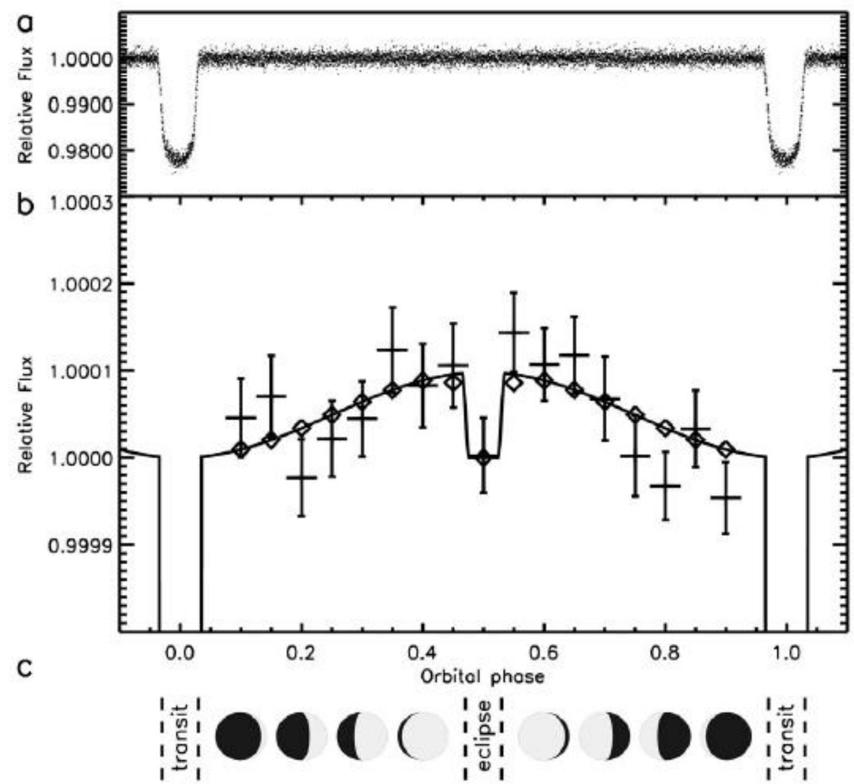


The SuperWASP project is the British exoplanet program. In each hemisphere, one instrument with 8 small cameras. Thousands of observations per star; angular resolution and limiting magnitude are like for ASAS-3 or ROTSE. Despite user-unfriendly data format, advanced users widely use SuperWASP for data mining. Instrumental effects were not sufficiently removed from the observations.

Catalina Sky Survey

Photometry from the Catalina Sky Survey (initially aimed at detecting transients) is in open access since 2012. 50–70 cm Schmidt telescopes are being used. Northern and southern sky, with the Milky Way band and poles excluded. Working magnitudes: 13 to 19. Good angular resolution.

Telescope: 27 cm
Corot main tasks:
Asteroseismology, search for exoplanets
Operation: 2006 – 2013



The launch of the Corot satellite with a Russian rocket

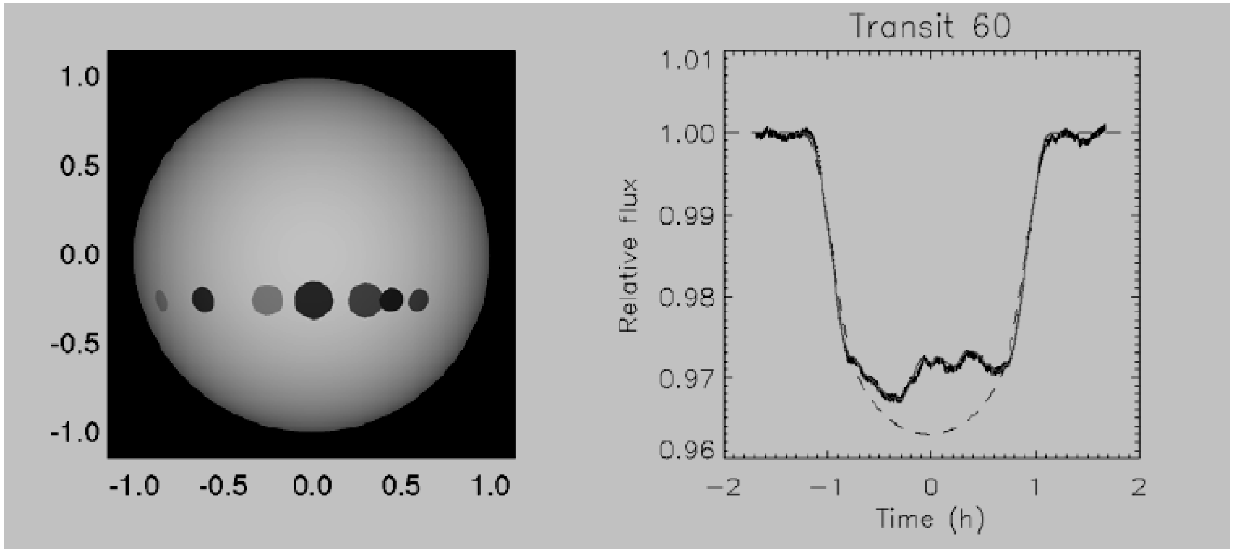
The first observation of phases and an eclipse of the exoplanet HD 189733b (I. Snellen et al., 2009)

There is no generally accepted definition of a variable star!

My provisional definition:
We call a star variable if the flux from it at the outer edge of the Earth's atmosphere (in visual, UV, or near-IR range) varies by an amount detectable given the existing photometric accuracy during the time covered with observations with such accuracy.

Neither the Astronomische Gesellshaft nor the IAU ever introduced the lowest amplitude of brightness variations needed to call the star a variable.

The number of known variable stars increases exponentially with improved techniques of ground-based photometry and beginning space-borne observations.



CoRoT data permit detection and mapping of star-spot groups covered by an exoplanet (A. Valio, 2011).

Launched in March,2009. The Kepler mission telescope is 95 cm in diameter. For 4 years, it continuously watched the field at the border of Cygnus, Lyra, and Cepheus. By August, 2016, discovery of 2327 exoplanets has been confirmed (some 10000 discoveries in total), several of them Earth-size objects in habitable zones. After failure of the original orientation system, the new K2 project was started, with stabilization by sunlight pressure.

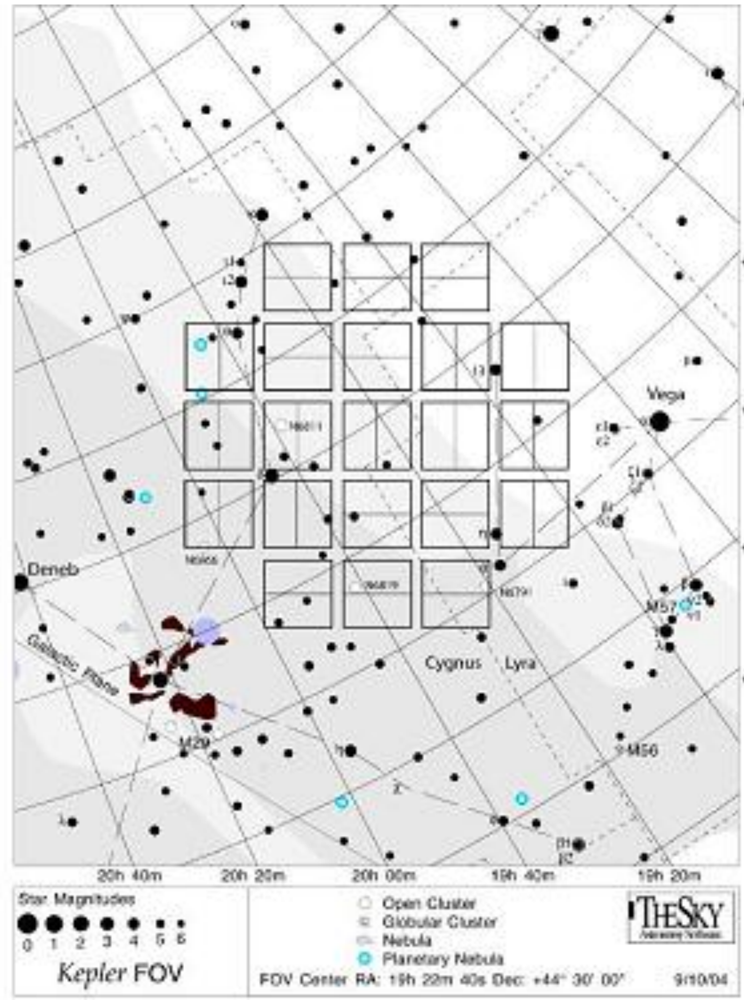
KEPLER mission

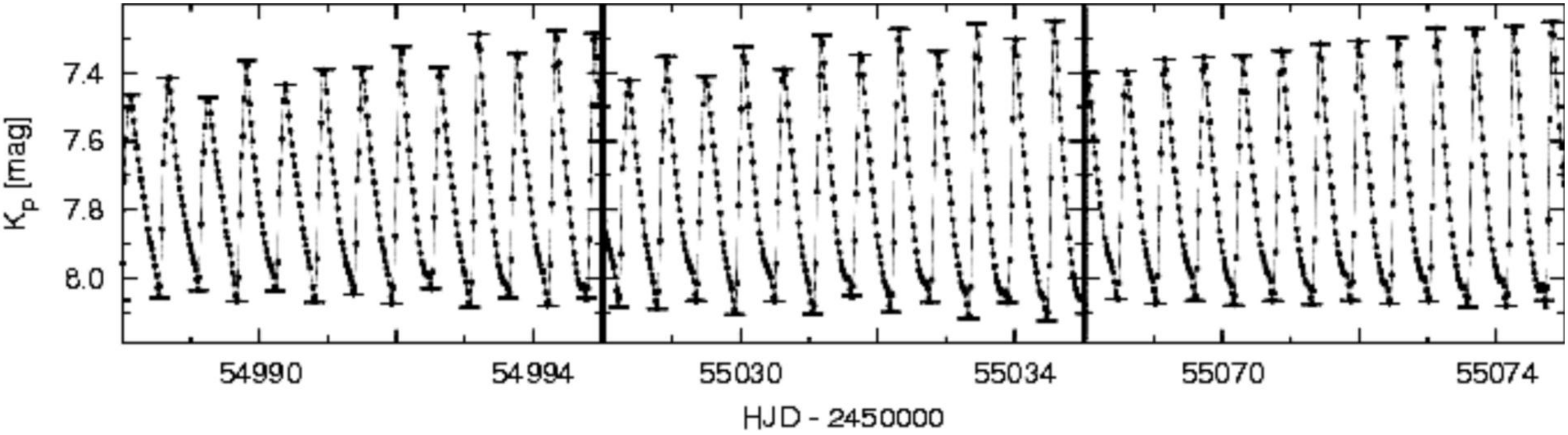
42 CCDs, 2200×1024

More than 150 000 stars in the program

Discovered thousands of eclipsing variable stars.

Cycle alteration for the variable RR Lyrae: the pulsation period being about half a day, the alteration phoenomenon cannot be detected in ground-based observations.





Different heights of odd and even maxima of RR Lyr in Kepler’s data

Ekaterinburg, September 5–7, 2016

KEPLER mission

Preliminary results:

Among ~ 150 000 program stars:

~ 60 000 are periodic variables;

~ 34 000 stars vary with poorly pronounced periodicity or aperiodically (G. Basri et al., 2011)

Given the KEPLER photometric accuracy, two thirds of all stars are variable.

USNO-B1.0 catalog (2003):

1 042 618 261 objects (stars and galaxies), ~ 1 billion stars to 20–21^{m} , about 700 million variables detectable with Kepler accuracy

Specialized variable-star catalogs (like GCVS) and variable-star databases (like VSX) based on star-by-star analysis will become impossible very soon. However, if we do not keep reliable knowledge on variable stars, we will discover the same things again and again.

My view of the possible way out is to include a section on variability characteristics into major catalogs, like USNO-B1.0. So far, this was done only in the HIPPARCOS catalog (about 100000 stars). HIPPARCOS variable stars are subdivided into periodic and unsolved variables.

СПАСИБО!
THANK YOU!