Mem. S.A.It. Vol. 74, 984

Memorie della

Variable sources and data mining: an example

Marco Riello[1,2]

[1] Dipartimento di Astronomia, Università di Padova, Vicolo dell'Osservatorio 2, 35122 Padova, Italy, e-mail: riello@pd.astro.it
[2] INAF, Osservatorio di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy

Abstract. We present the reduction and archiving tools developed for our search for supernovae at intermediate redshifts at ESO. The data reduction recipes developed for the SN candidate selection are described. Since our program provided many byproducts (minor planets, active galactic nuclei, variable stars) we have prepared a MySQL database with a web interface to archive these variable sources as well as the tool which enables the identification of variables objects detected in previous observing runs. With this tool we are able to recover the variability history for many SN candidates on our search images which strongly reduce the contamination by other variable sources. Our database approach is general and easy to customize to match the needs of other variability projects, in particular it provides a natural framework to share data.

Key words. Astronomical data bases: miscellaneous – Supernovae – Surveys – Techniques: image processing –

1. Introduction

In the last couple of years we have carried out a supernova (SN) search with the main scientific aim of determining the SN rate at intermediate redshift ($z \leq 0.6$) (for details see the contribution by G. Altavilla, this Conference). For the detection of SN candidates we used the Wide Field Imager at ESO/MPE 2.2m telescope (La Silla, Chile). We surveyed 21 fields of ~ 0.25 deg^2 each (corresponding to a total useful area of ~ 5.1 deg^2) in the V and R bands. At each epoch we took for each field three exposures of 900s with a dithering of few arcsec to remove the cosmic rays, detector cosmetic defects and moving objects. For the spectroscopic confirmation of the SN candidates we used FORS1/2 at ESO VLT because the magnitude of our candidates is usually fainter than 22 mag. Due to the limitation of observing time, only a subsample (the brightest candidates) could be spectroscopically confirmed. Usually one VLT night for candidate confirmation was allocated about 10 days after two consecutive nights of search with the WFI.

Send offprint requests to: Marco Riello
Correspondence to: Vicolo dell'Osservatorio 5, 35122 Padova, Italy

2. Data Reduction Software

The pre-reduction, bias and flat fielding, are performed using IRAF[1] and MSCRED (Valdes 1998), a package specifically designed to handle mosaic images in multi-extension FITS format (MEF). Whenever it was possible, standard sky flats are improved by constructing a super flat using the scientific frames. The combination of dithered frames is a standard task which involves first the geometrical registration of the images, the subtraction of background, the determination of the photometric scaling factors and finally the stacking. We developed a robust script to perform these steps in automatic mode. The stacked images are finally provided with an astrometric calibration computed using the USNO2 reference catalog. The astrometry is used to identify each object in the field.

The search for SNe is based on the image subtraction of the new observation with a reference image obtained weeks to months before. First a proper reference image with a seeing close to the one to be searched is selected, then the two epochs are geometrically registered. If the seeing of the two epochs is very different it is necessary to degradate the best image to match the PSF of the worse one. Finally we use ISIS v2.1 software (Alard & Lupton 1998) to make the subtraction.

To search the difference image for residuals due to variable sources we use Sextractor (Bertin & Arnouts 1996). The catalog we get at this stage is usually heavily contaminated by spurious detections due to residuals left by bright/saturated stars and poorly removed cosmic rays. In a typical field, consisting of eight CCDs, over a thousand detections are usually found. To remove spurious detections from this catalog we developed an automatic procedure which sorts each detection based on several measured parameters: the FWHM of the object with respect to the mean, the distance of the residual from the center of the associated object if any (i.e. the host galaxy in the case of SN), the difference between the object magnitudes measured with different prescriptions and the Sextractor stellarity index. After some tuning of the threshold of each parameter with Monte Carlo simulations, we were able to reduce the number of detections of a factor 10 without significant loss of efficiency in the detection of good SN candidates. Our search technique detects SN candidates but also other variable sources like active galactic nuclei (AGN), quasars (QSO), variable stars and moving objects. After a visual inspection of the candidates with higher rank we attach a provisional classification (Supernova, AGN, variable star or moving object) to each reliable variable source after considering all the available informations. All these variable sources and the related informations are finally inserted into a database.

[1] IRAF is distributed by the National Optical Astronomy Observatories, which are operated by the Association of Universities for Research in Astronomy, Inc., under cooperative agreement with the National Science Foundation.

3. The database

A typical epoch usually provides from 5 to 15 variable sources depending on the seeing and the sky transparency. We have monitored 21 different fields over 25 observation runs which provided a total number of 185 epochs. With these numbers we need an efficient way to store and quickly access all the informations when needed. We have decided to implement a relational database based on MySQL. The main advantages for this choice are the following:

- It is an open source project which makes it hopefully longlife.
- It is completely free.
- It is available for many platforms, in particular for Linux.
- It benefits from a large community of experienced users, who can help to solve everyday problems, and developers who check and fix most bugs.

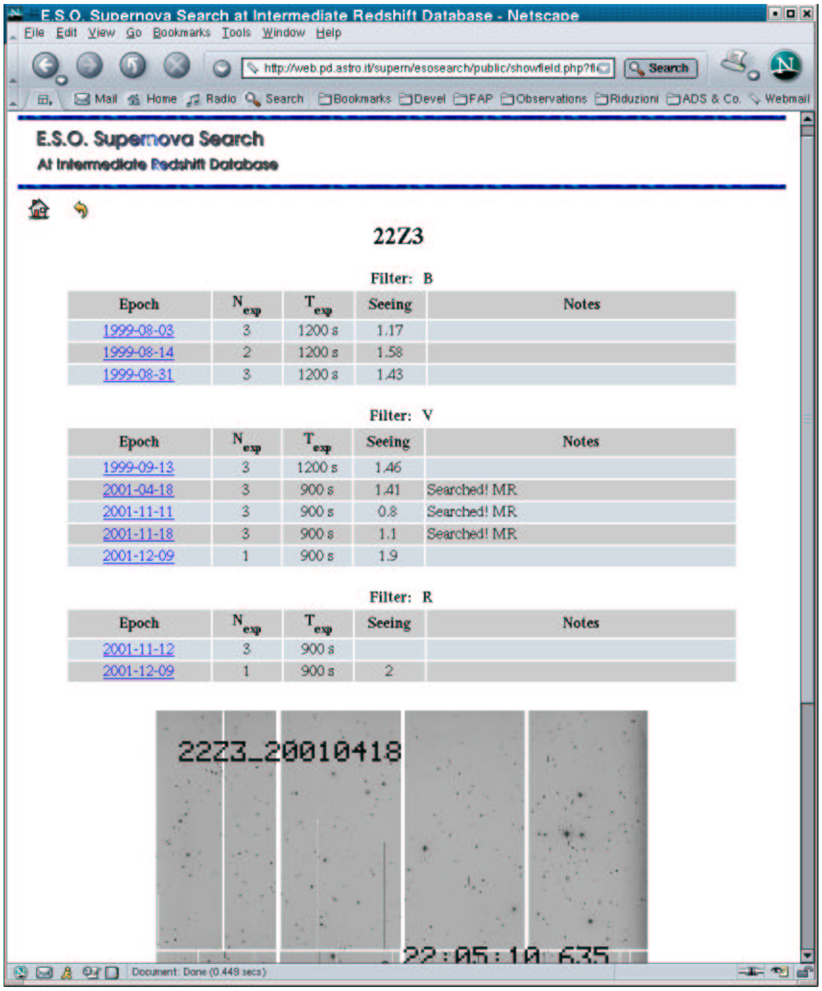
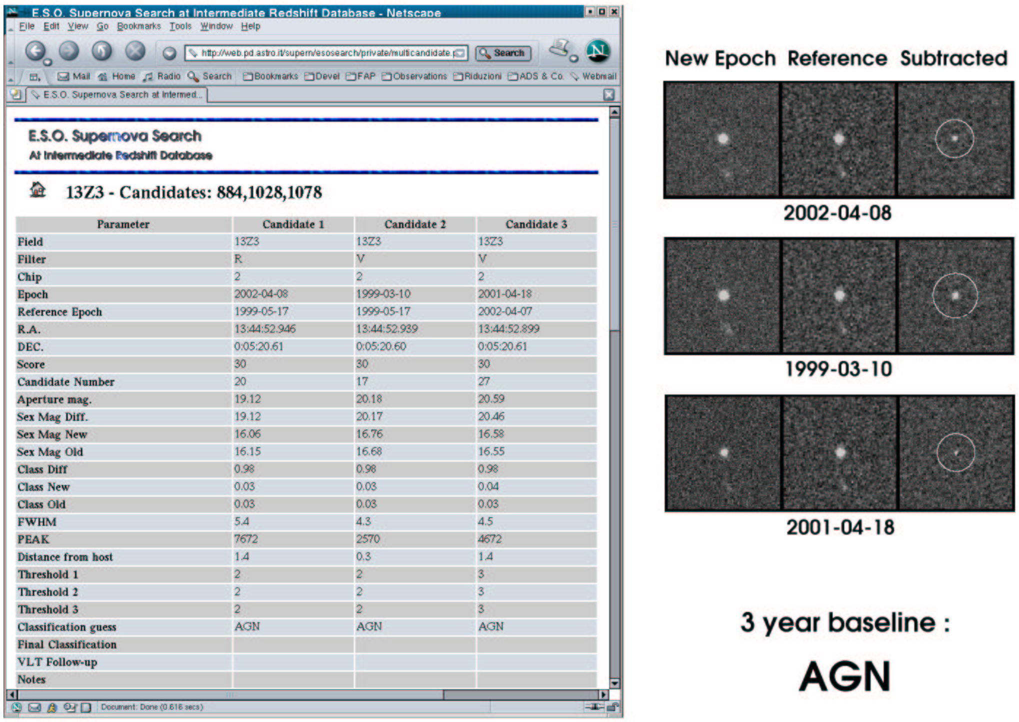


Fig. 1. Output web page generated by the query that selects all epochs available for a given field. For each epoch the relevant informations are reported: number of exposures, integration time of a single exposure, mean seeing and comments.

– It is easy to install and to maintain.
– It is easy to program a user interface and to provide external access via the web.

Our database contains all the informations about the monitored fields: the photometric zero points for each epoch, the list of epochs available and the related informations (seeing, exposure time etc.), and the list of variable sources detected. We developed also a web interface (`web.pd.astro.it/supern/esosearch/`) to access and operate the database. The user interface has been designed to be useful both during the observations, giving quick access to the informations on the single fields and epochs, and during the reduction and analysis phases. We stress again that, the web interface allows an easy access to the data because it does not require any special software but a web browser which nowadays is available on every type



Parameter	Candidate 1	Candidate 2	Candidate 3
Field	13Z3	13Z3	13Z3
Filter	R	V	V
Chip	2	2	2
Epoch	2002-04-08	1999-03-10	2001-04-18
Reference Epoch	1999-05-17	1999-05-17	2002-04-07
R.A.	13:44:52.946	13:44:52.939	13:44:52.899
DEC.	0:05:20.61	0:05:20.60	0:05:20.61
Score	30	30	30
Candidate Number	20	17	27
Aperture mag.	19.12	20.18	20.59
Sex Mag Diff.	19.12	20.17	20.46
Sex Mag New	16.06	16.76	16.58
Sex Mag Old	16.15	16.68	16.55
Class Diff	0.98	0.98	0.98
Class New	0.03	0.03	0.04
Class Old	0.03	0.03	0.03
FWHM	5.4	4.3	4.5
PEAK	7672	2570	4672
Distance from host	1.4	0.3	1.4
Threshold 1	2	2	3
Threshold 2	2	2	3
Threshold 3	2	2	3
Classification guess	AGN	AGN	AGN
Final Classification			
VLT Follow-up			
Notes			

Fig. 2. The left panel shows the output generated by the query which retrieves the informations for a variable object detected on different epochs. For each detection the page summarizes the parameters of interest (magnitudes, Sextractor stellarity index, celestial coordinates etc.) and then shows three stamps, one for the new epoch, one for the reference image and one for the subtracted image (right panel). In particular this figure shows the case of a stellar residual near the center of a diffuse object detected in three different epochs over a baseline of three year. Thanks to our tool this SN candidate can be classified with high confidence as an AGN.

of terminal. These are the main kind of operations implemented:

– **Basic Operations** : It is possible to find and retrieve the informations for a single field (finding charts, pointing coordinates), epoch (seeing, exposure time, number of exposures, photometric zero points), variable sources detected in each epoch (image, reduction parameters, magnitudes, stellarity indexes, etc.).
– **Database Management** : It is possible to add informations for the ongoing observing runs, modify the data already inserted as soon as the reduction process proceeds, delete records that are no longer relevant. These operations are performed through web forms which include the checks needed to guarantee the database integrity.
– **Variability History** : It is possible to check if any of the objects detected at given epoch was already found in previous ones thus obtaining the variability history of the field (see next section).

4. Variability History

The procedure described above is designed to detect all sources which have varied between the epoch of observation and the reference one. Although the ranking algorithm is rather effective in reducing spurious detections, we are still contaminated by other variable objects like AGNs and variable stars. We have found of great help in reducing this contamination the use of the variability history of the candidate: in fact if the candidate shows a record of previously observed variability not compatible with the light curve of a SN it must be rejected as SN candidate (but can be used for other science). We have thus developed a tool that, for any given object, scans the database for records with the same position within a customizable error box. This approach is especially effective for uncrowded fields for which the probability of having two overlapping variable objects is very low. The astrometric solution computed during the reduction phase is good enough to allow the use of a reasonable small error box (typically 3–4 arcsec). The results of the database search is formatted through a web page which presents, for each detection, all the relevant parameters available (magnitudes, score, celestial coordinates, etc.) and the picture for each detection showing the three dithered images, the combined image, the reference image and the subtracted image. The direct inspection avoids the few cases in which the parameters alone can be misleading. An example of this page is presented in Fig. 2.

5. Conclusions and Future Work

The approach to data mining for variable sources presented here have proved to be very effective for large data sets. We believe that the effort required to set up the database and the access interface is largely paid by the advantages described and it is general enough to be customized to match the requirements of different variability projects and to share data products between different research groups.
We are now working on the implementation of the follow–up database with all the spectra of observed objects and related data (redshift, classification, etc.). This work, started as an investigation of the capabilities of new software solutions, is now turning to a more mature phase addressing new issues like unsupervised data reduction.

Acknowledgements. I would like to thank Marco Rota for kindly helping me during the database design phase. This work has been financially supported by the Italian Ministry for University and Research through grant Cofin/2001.

References

Alard, C. & Lupton, R.H. 1998, ApJ 503, 325
Bertin, E. & Arnouts, S. 1996, A&AS 117, 393
Valdes, F. 1998, adass 7, 53