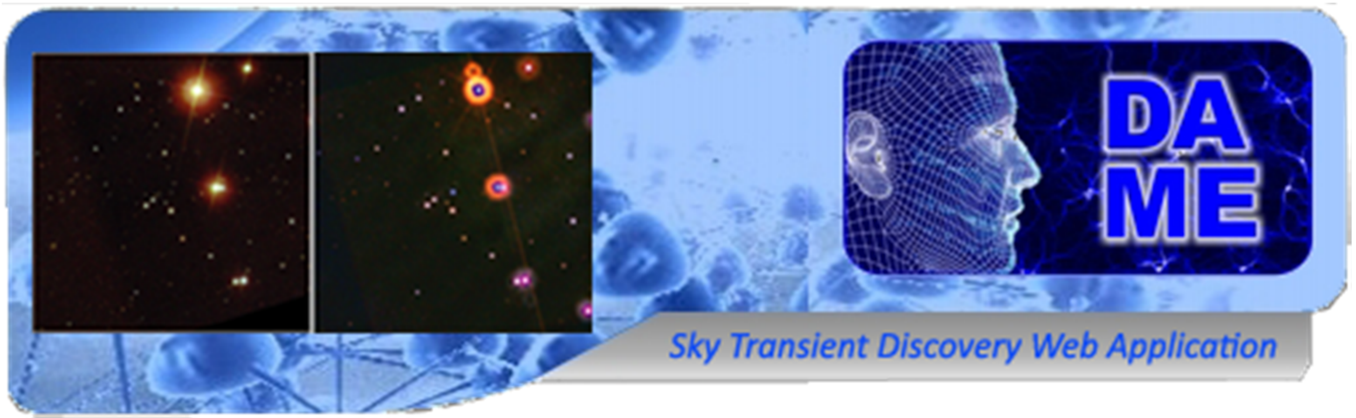


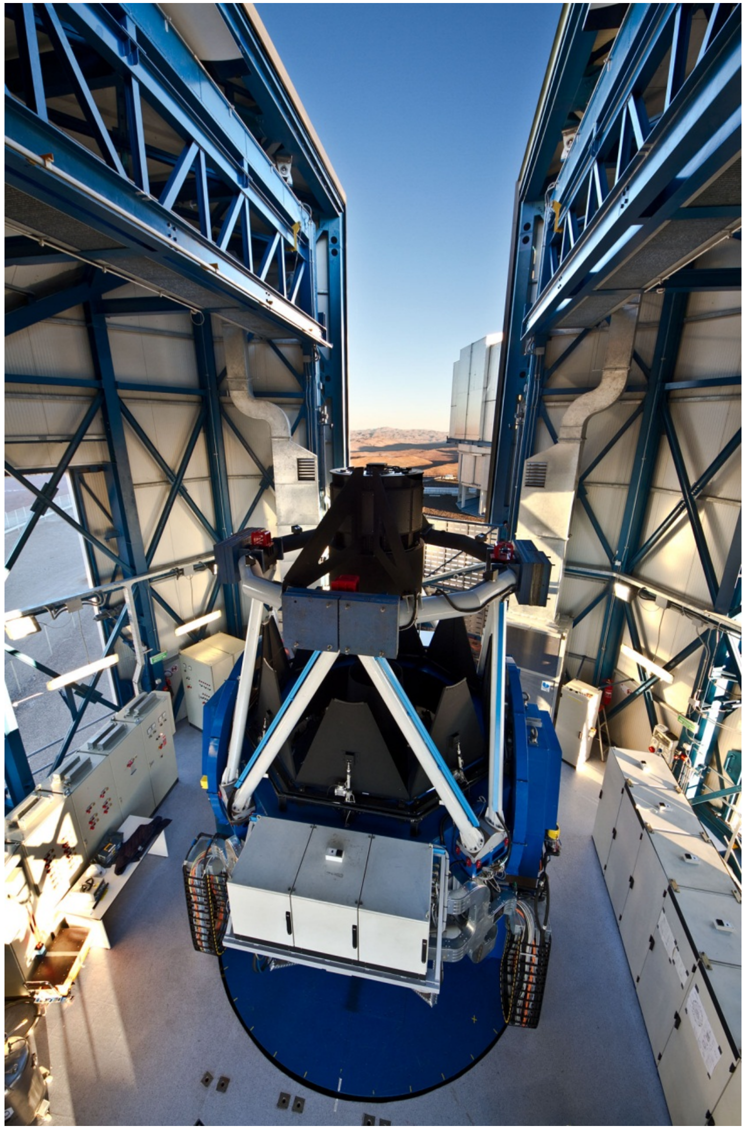
STraDiWA

Sky Transient Discovery Web Application:
A simulation environment for transient discovery

M. Annunziatella,
M. Brescia, S. Cavuoti, G. Longo, A. Mercurio

astromeeting

Time Domain Astronomy



VLT Survey telescope.

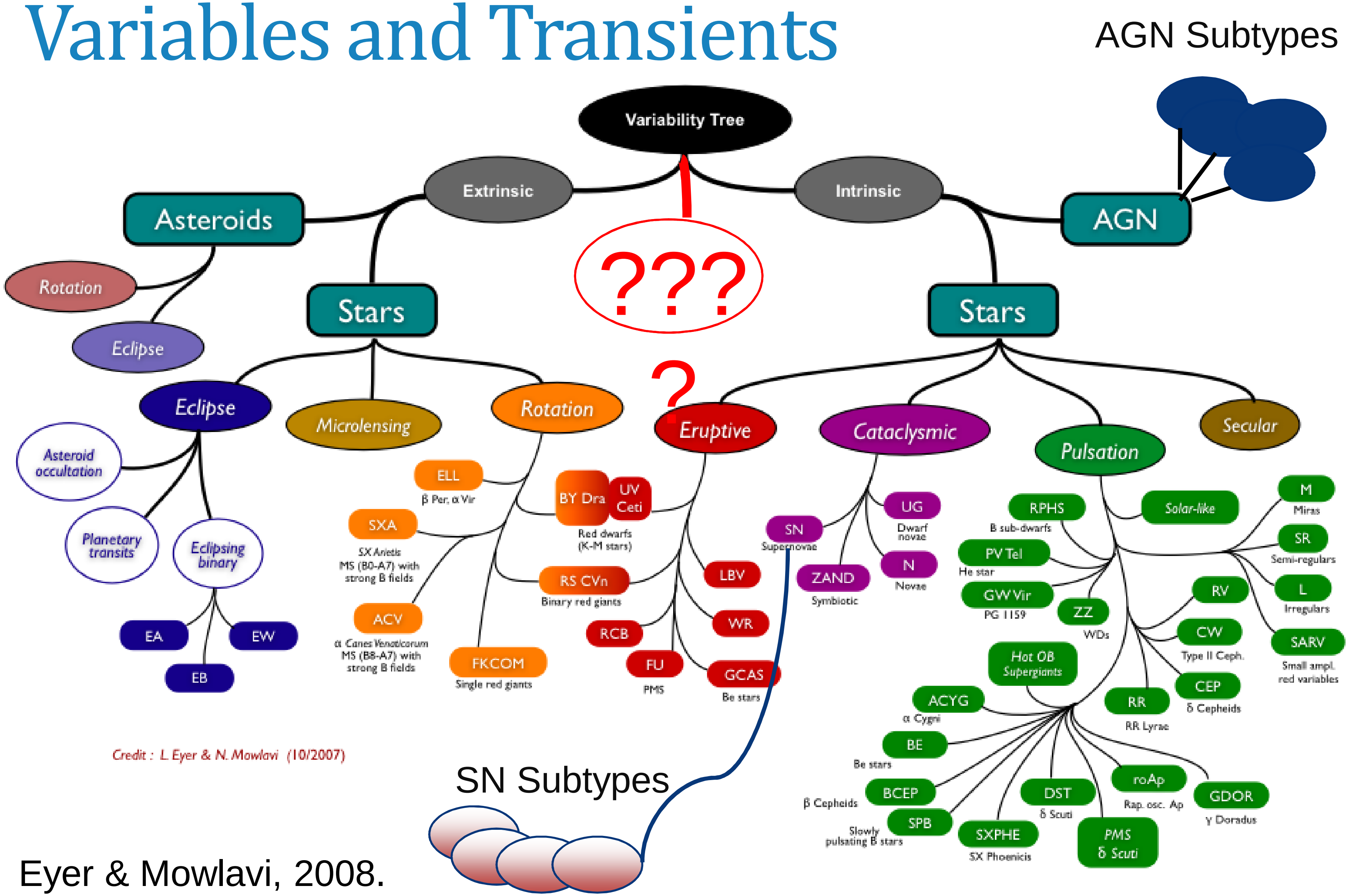
Synoptic sky surveys:

•Past: OGLE, MACHO, have already collected millions of light curves.

•Currently operating: CRTS, PTF, PanStarrs 1, VST, collect data streams of $\sim$ 0.1 TB /night and detect $\sim 10-10^2$ transients per night.

•Forthcoming: GAIA, LSST, EUCLID, will move us to the Petabyte regime and will detect $\sim 10^5 - 10^6$ transients event per night.

Semantic Tree of Astronomical Variables and Transients



Eyer & Mowlavi, 2008.

astromeeting

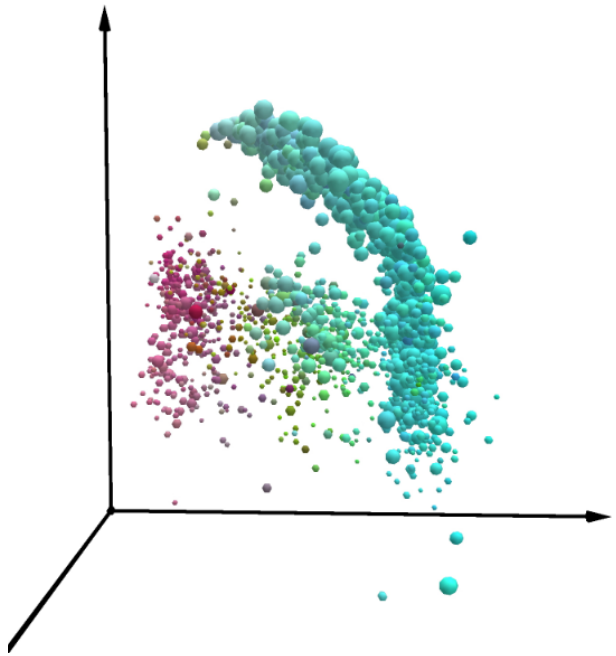
Problems

- How to characterize variable objects (light curves, other statistical indicators)?
- Knowledge base built on the data themselves (CRTS) or rather on simulated ones (STRADIWA)?
- How to solve the computational challenge (DAMEWARE)?
- How to find the unknown? Throwing away all the known (classification) or searching for intrinsic partitions of the Parameter Space (clustering)?

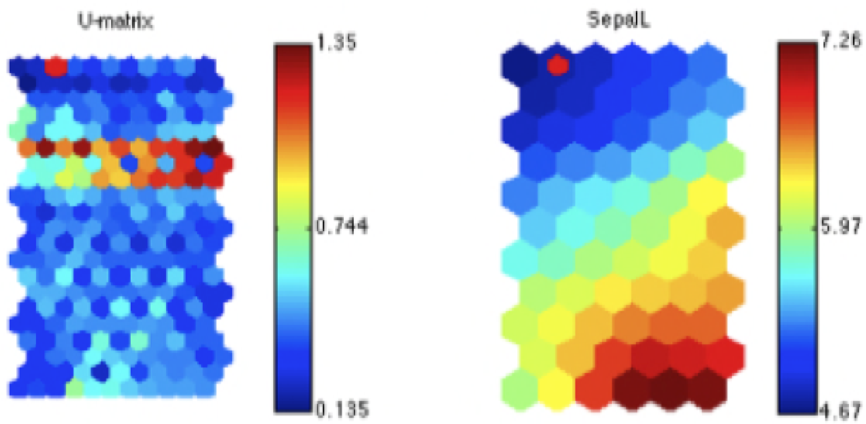
Unsupervised Clustering in Feature Space

(in coll. with Caltech)

- Unsupervised Machine Learning
- Can be used to determine the number of classes and cluster the input data in classes on the basis of their statistical properties only
- Search for Outliers, Trajectories, etc.
- Methods: SOM, K-means, Hierarchical Clustering, etc.



- Given a set of features, which ones are the most discriminating between different classes?

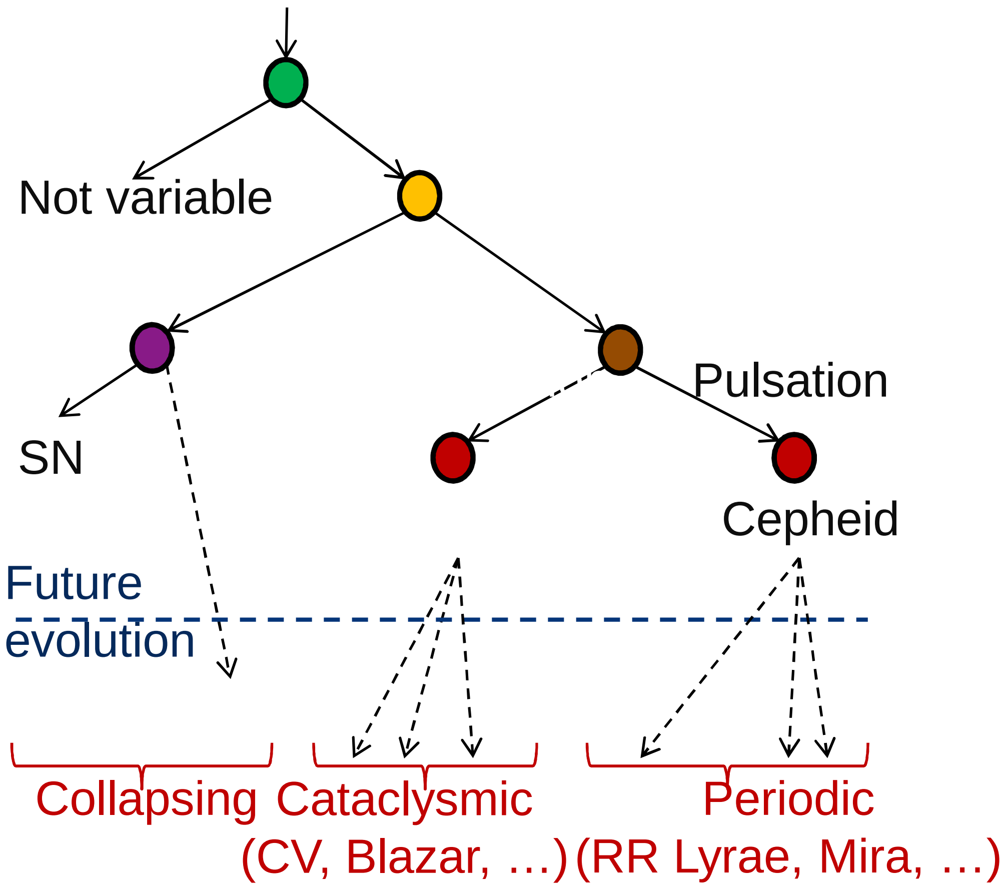


A Hierarchical Approach to Classification

Different types of classifiers perform better for some event classes than for the others

First approach is to use light curves to separate different groups of classes

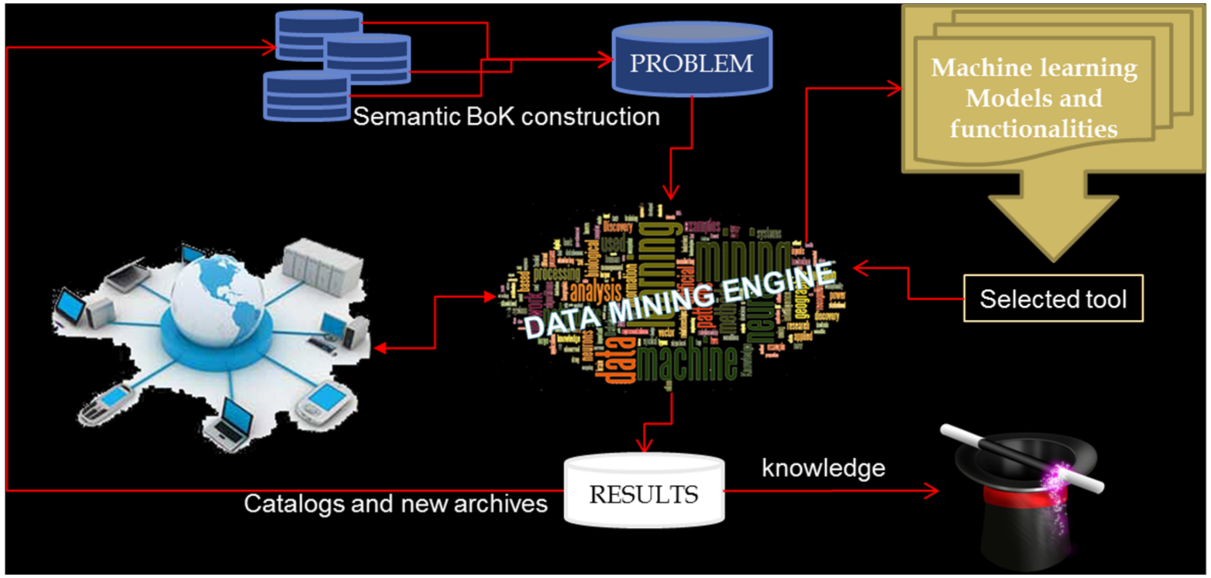
We propose to find which is the best classifier for a particular type of variable object. The evaluation of classifiers requires benchmarks, hence reference template images (or catalogues) on which to test the various models.



Data mining & Exploration Tool

Inspired by human brain features: high-parallel data flow, generalization, robustness, self-organization, pruning, associative memory, incremental learning, genetic evolution.

It is a web application for data mining experiments, based on WEB 2.0 technology



Multi Layer Perceptron trained by:
- Back Propagation
- Quasi Newton
- Genetic Algorithm

Support Vector Machines

Genetic Algorithms

Self Organizing Feature Maps

K-Means

Multi-layer Clustering

Principal Probabilistic Surfaces

Classification

Regression

Clustering

Feature Extraction

next …

Bayesian Networks

Random Decision Forest

MLP with Levenberg-Marquardt

The data mining WEB Application

DAMEWARE - DAta Mining Web Application REsource

Web-based app for massive data mining based on a suite of machine learning methods on top of a virtualized hybrid computing infrastructure.

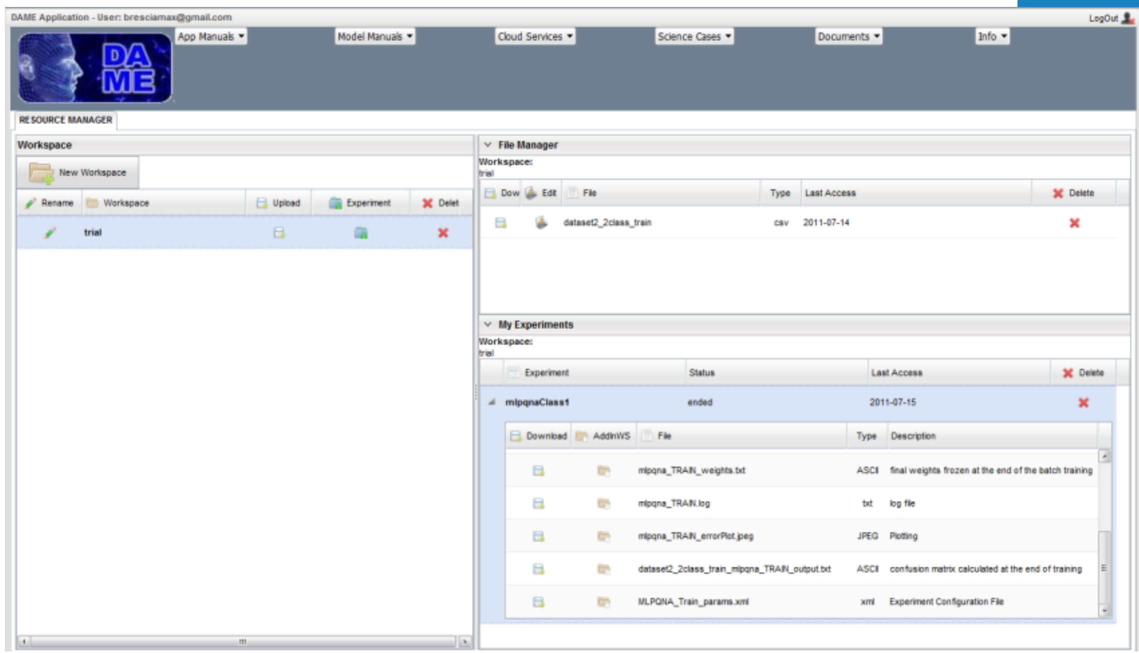
Intro page + manuals http://dame.dsf.unina.it/beta_info.html

Beta Release available http://dame.na.astro.it:8080/MyDameFE/

demo videos http://www.youtube.com/user/DAMEmedia

The release 1.0 will be deployed on a CLOUD, including GRID farm of S.Co.P.E

- Private user account after registration;
- Data files (CSV, ASCII, FITS-image, FITS-table, VOTable);
- Classification models;
- Regression models;
- Clustering models;
- Feature Extraction models;
- Editing files for experiment setup (join, split, sort, shuffle, scale etc.);
-



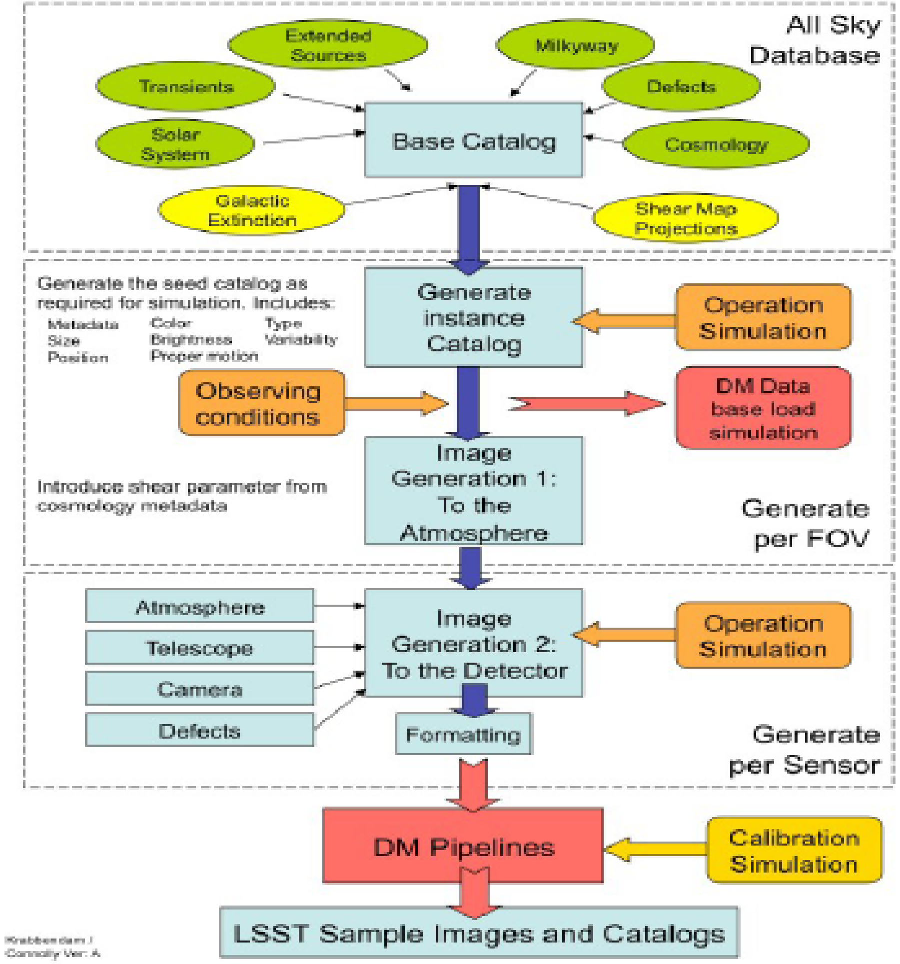
astromeeting

The Strategy

In order to be as realistic as possible, this simulation has to take into account as many as possible relevant factors:

- **Instrument setup.**
- **Observing site conditions.**
- **Survey strategy (how many bands, how deep).**
- **Sampling mode.**
- **Realistic distribution of pre-modeled variable objects.**

LSST Simulation



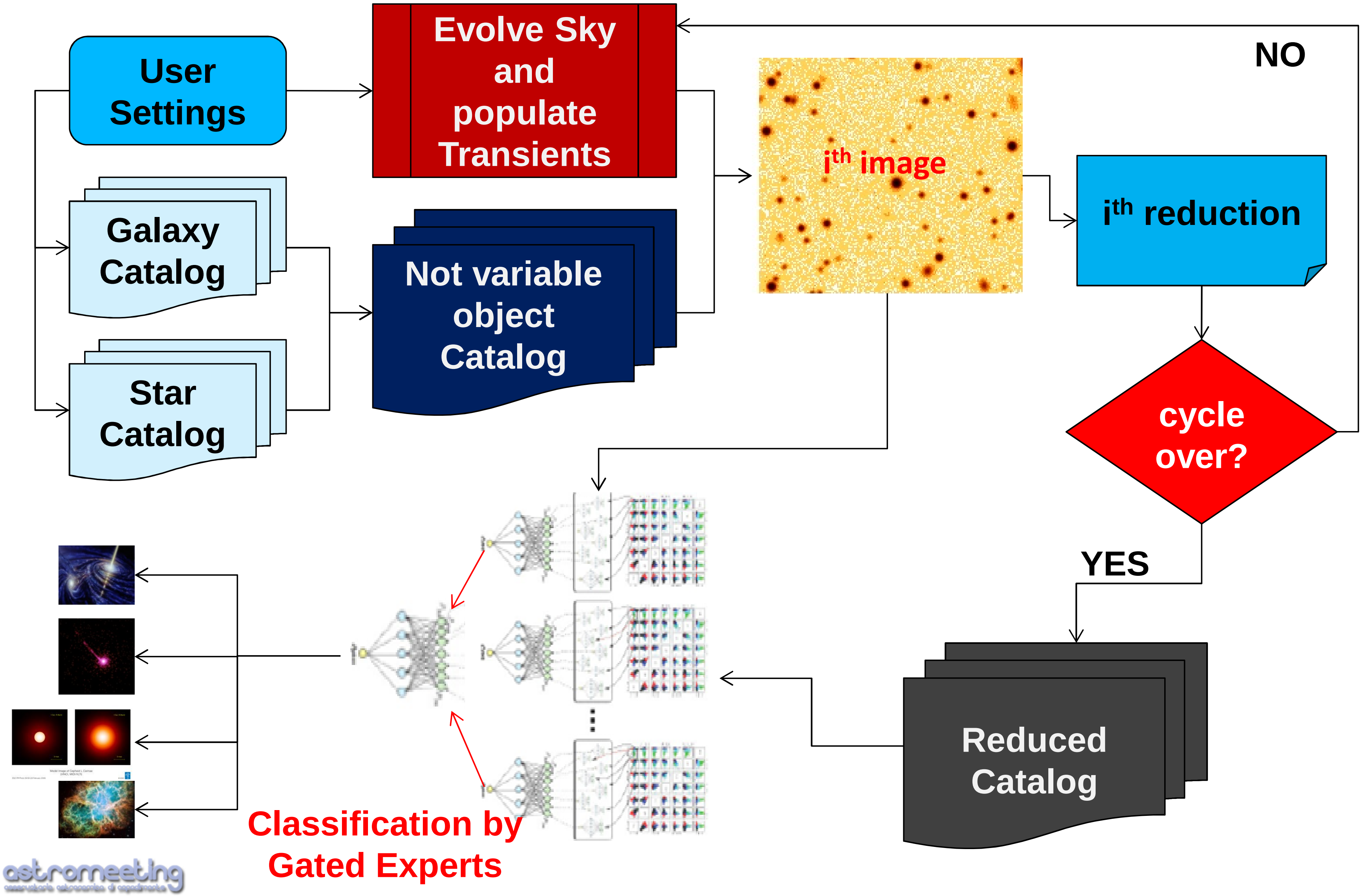
The LSST Image Simulation Group has implemented a simulation pipeline whose main scope is to test detection algorithms.

Transients are generated randomly loosing comprehensive theoretical treatment.

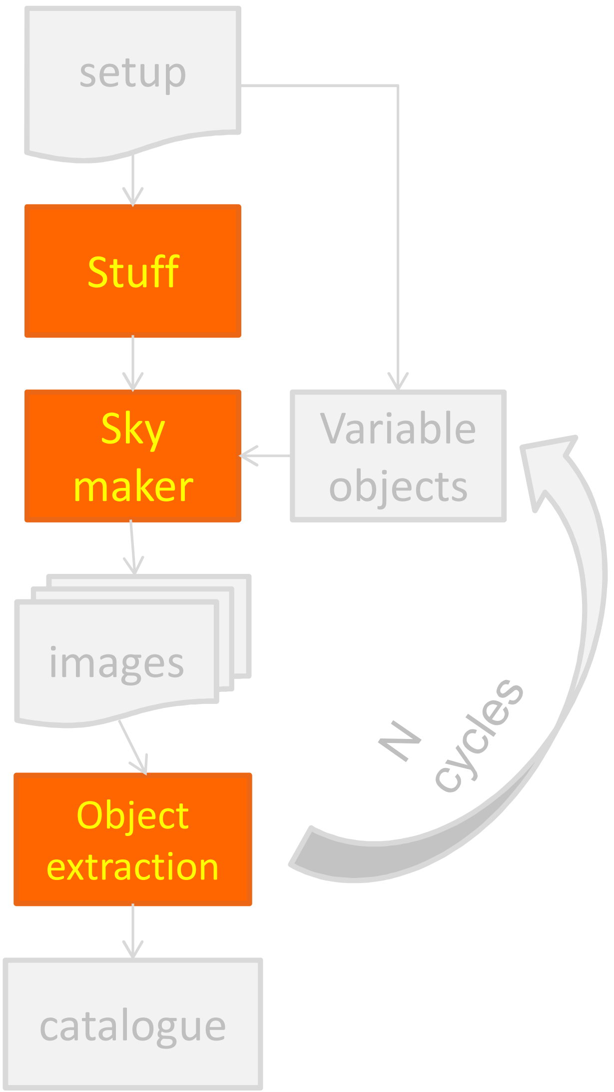
Our proposal is to include specific theoretical models for each variable object in the simulations of real images.

http://www.noao.edu/lsst/opsim/

STraDiWA Workflow



Starting Tools

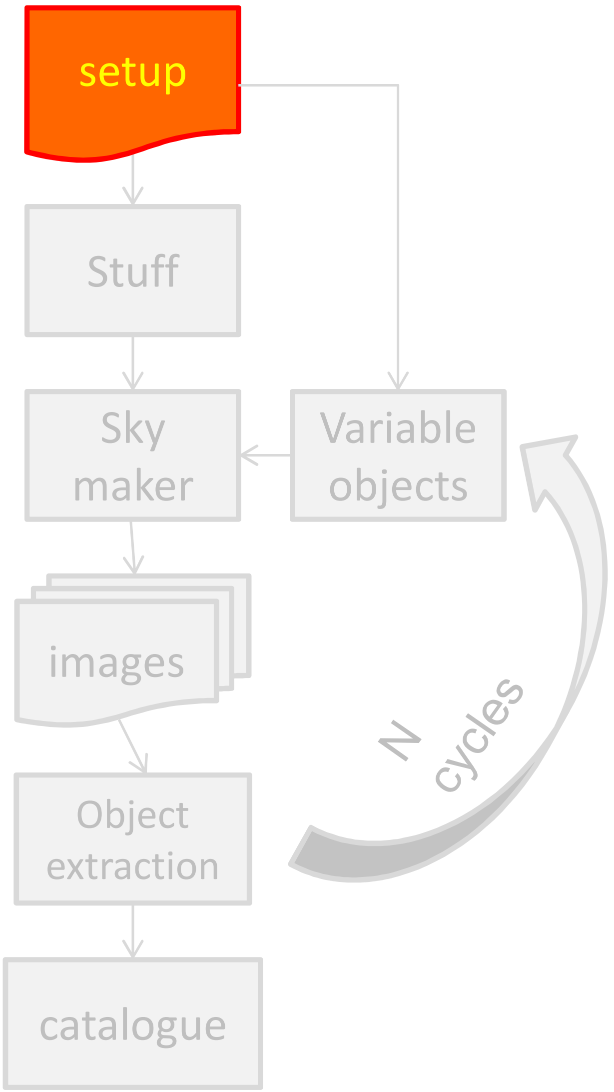


STUFF: Simulation of Galaxy background.

SkyMaker: Produces an image starting from a catalog of galaxies, produced by STUFF, adding a RANDOM field of stars.

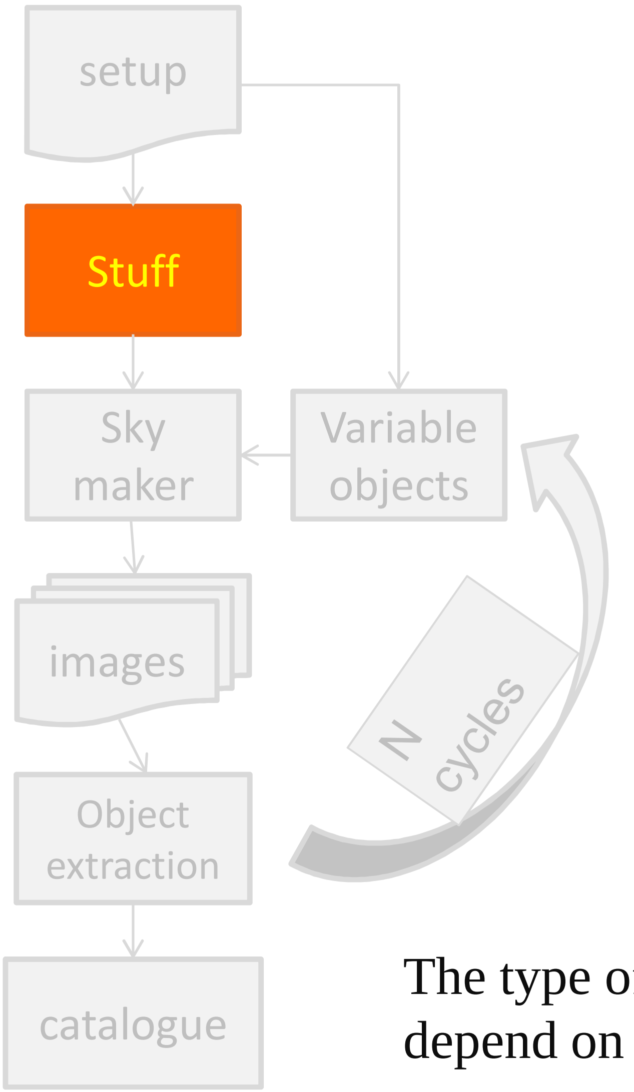
Software for catalog extraction.

Setup parameters



- Number of bands.
- Sampling: can be even or uneven.
- Observing Condition: the seeing can be assigned for each epoch or can be extracted randomly within a given range.
- Type and distribution of variable objects.

Stuff: Creation of the static sky I

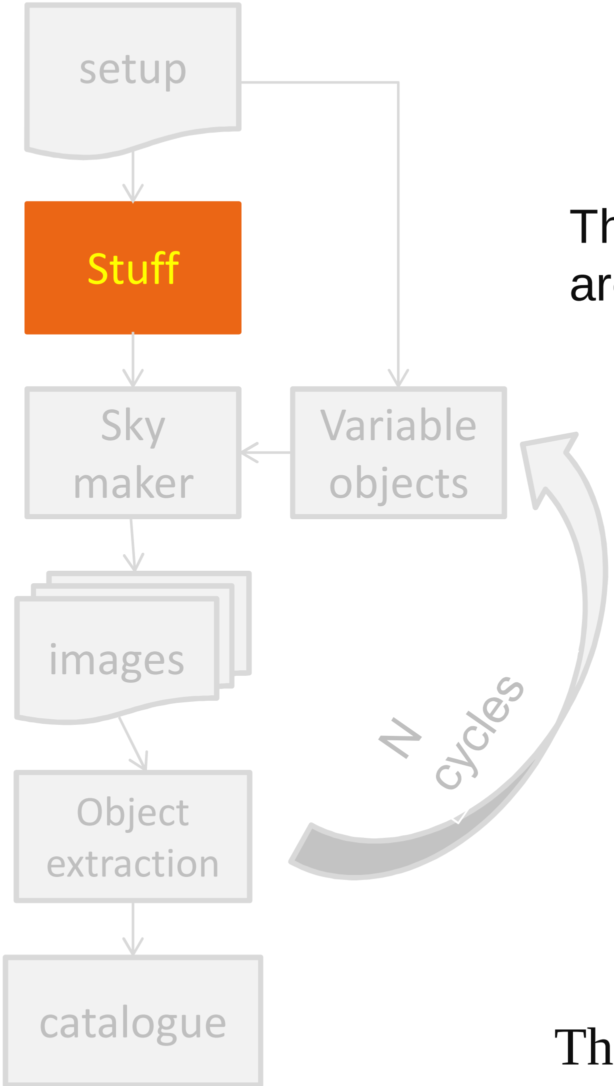


Stuff generates a catalog of galaxies which take in to account:

- Cosmological models.
- Luminosity distribution of the sources.
- Redshift distribution of the sources.
- Color distribution of the sources.

The type of these distributions and the parameters which they depend on can be all modified in the configuration file.

Stuff: creation of the static sky II

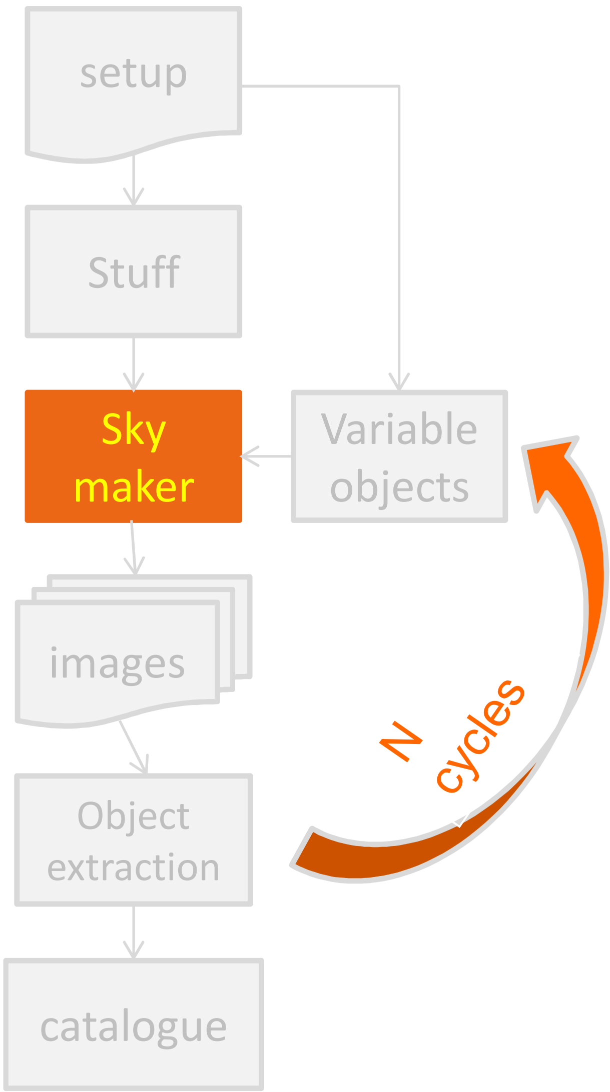


The main parameters to be opportunely set in Stuff are:

IMAGE_WIDTH	(pixel)
IMAGE_HEIGHT	(pixel)
PIXEL_SIZE	(arcsec/pixel)
MAG_LIMITS	(mag)
GAIN	(e-/ADU)
PASSBAND_OBS	

There are 120 filters available in Stuff, covering a wavelength range from 0.29 to 87.74831 μm.

SkyMaker: Instrumental simulation and image production I

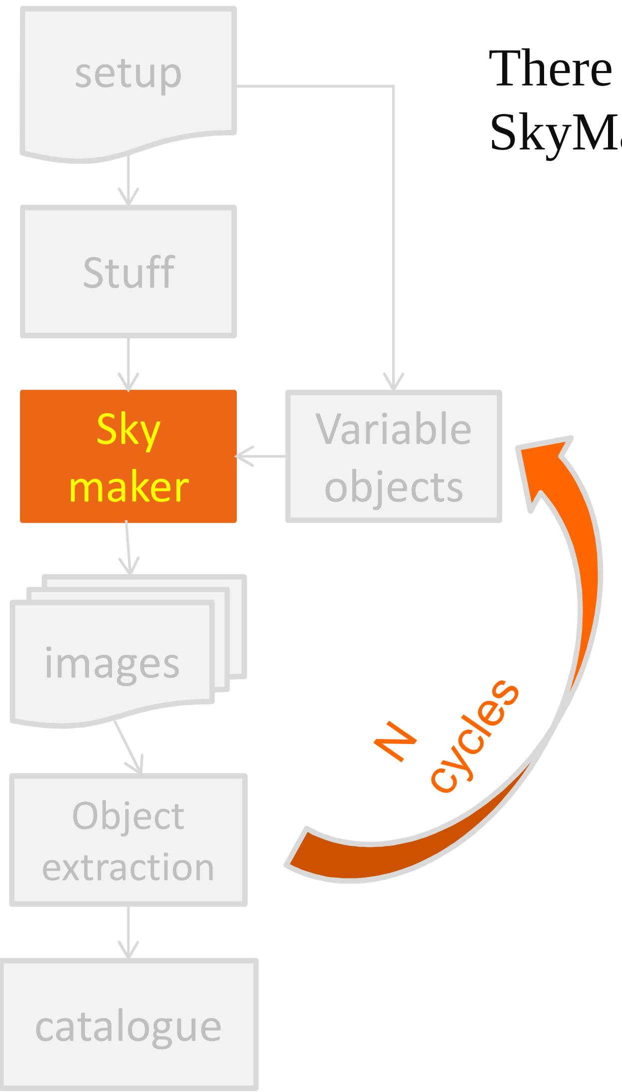


The simulation of optics and detector, as well as the image rendering are carried out by SkyMaker.

SkyMaker works through several steps:

- PSF modeling.
- Source modeling.
- Adding a uniform sky background.
- Applying Poissonian photon white noise and Gaussian read-out noise of the detector to the image .

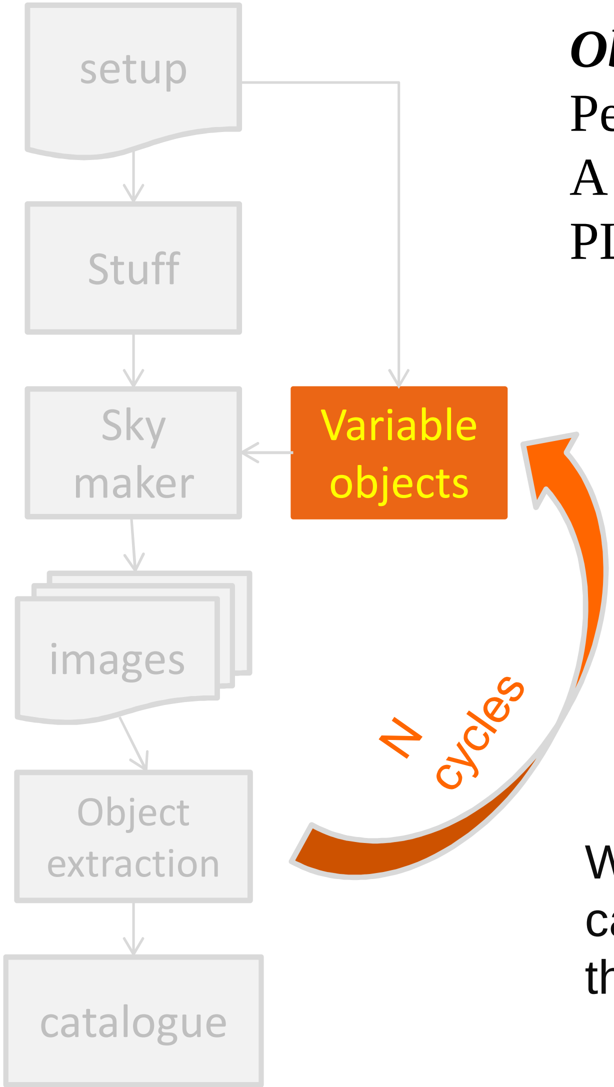
SkyMaker: Instrumental simulation and image production II



There are several parameters which have to be set in SkyMaker configuration file.

Detector	GAIN (e-/ADU) WELL_CAPACITY (e-) SATUR_LEVEL (ADU) READOUT_NOISE (e-) EXPOSURE_TIME (s) MAG_ZEROPOINT (ADU per second) PIXEL_SIZE (pixel size in arcsec)
Signal	WAVELENGTH (microns) BACK_MAG (mag/arcsec2)
Pupil feautures	M1_DIAMETER (m) M2_DIAMETER (m) ARM_COUNT ARM_THICKNESS (mm) ARM_POSANGLE DEFOC_D80 (arcsec) SPHER_D80 (arcsec) COMAX_D80 (arcsec) COMAY_D80 (arcsec) AST00_D80 (arcsec) AST45_D80 (arcsec) TRI00_D80 (arcsec) TRI30_D80 (arcsec) QUA00_D80 (arcsec) QUA22_D80 (arcsec)

Rules for variable objects: Classical Cepheids



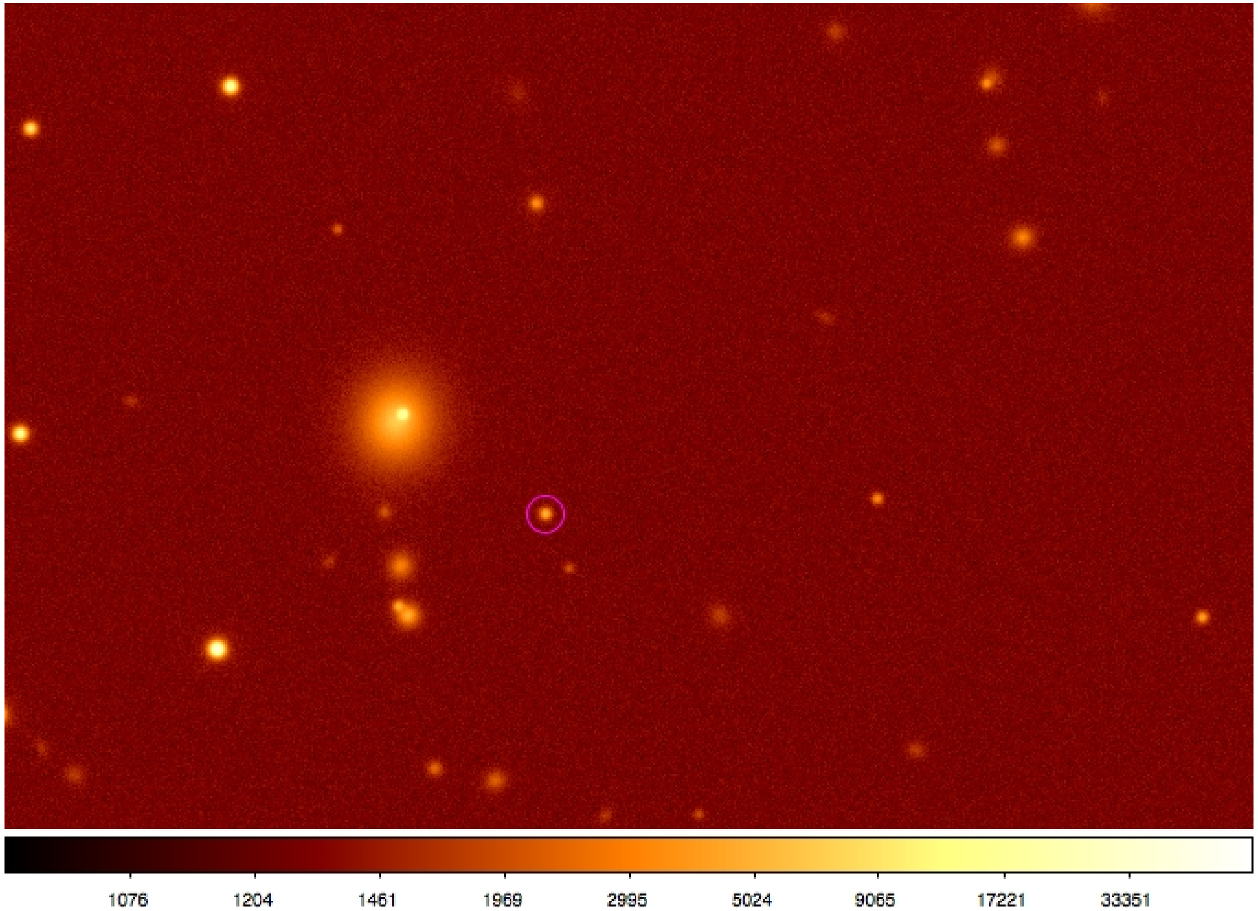
Observational properties:
Period Range: 1 – 80 days.
A = 0.1 – 2 mag.
PL relationship.

A classical Cepheid is modeled:

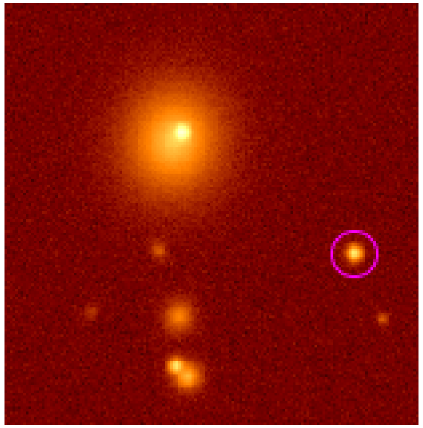
- Assuming a sinusoidal law.
- Imposing a PL luminosity relation.

We used the coefficients for the mean PL relation calibrated in Bono et al. 2010 and references therein.

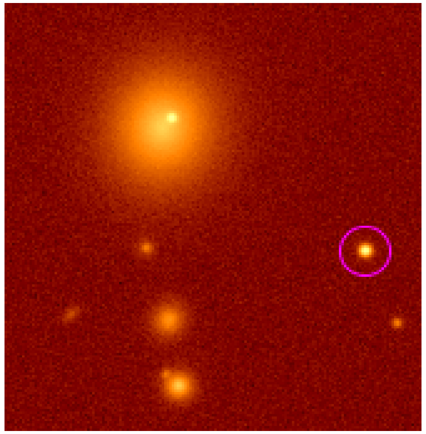
Image of a Classical Cepheid



Stamp of the image of a Classical Cepheid of Period of about 20 days at t = 0 days.

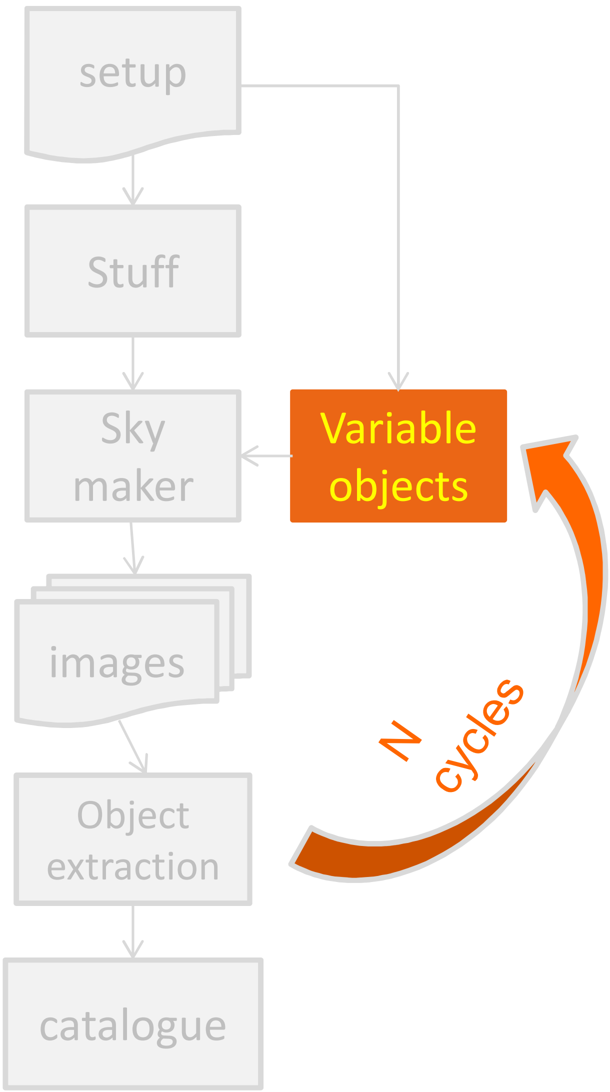


Cepheid at t = 13 d.



Cepheid at t = 35 d.

Rules for variable objects: Type Ia Supernovae I



As model for type Ia Supernovae we used an analytical function, used in Contardo, Leibundgut, and Vacca 2000 for the fit of a sample of type Ia Supenovae.

Linear decay

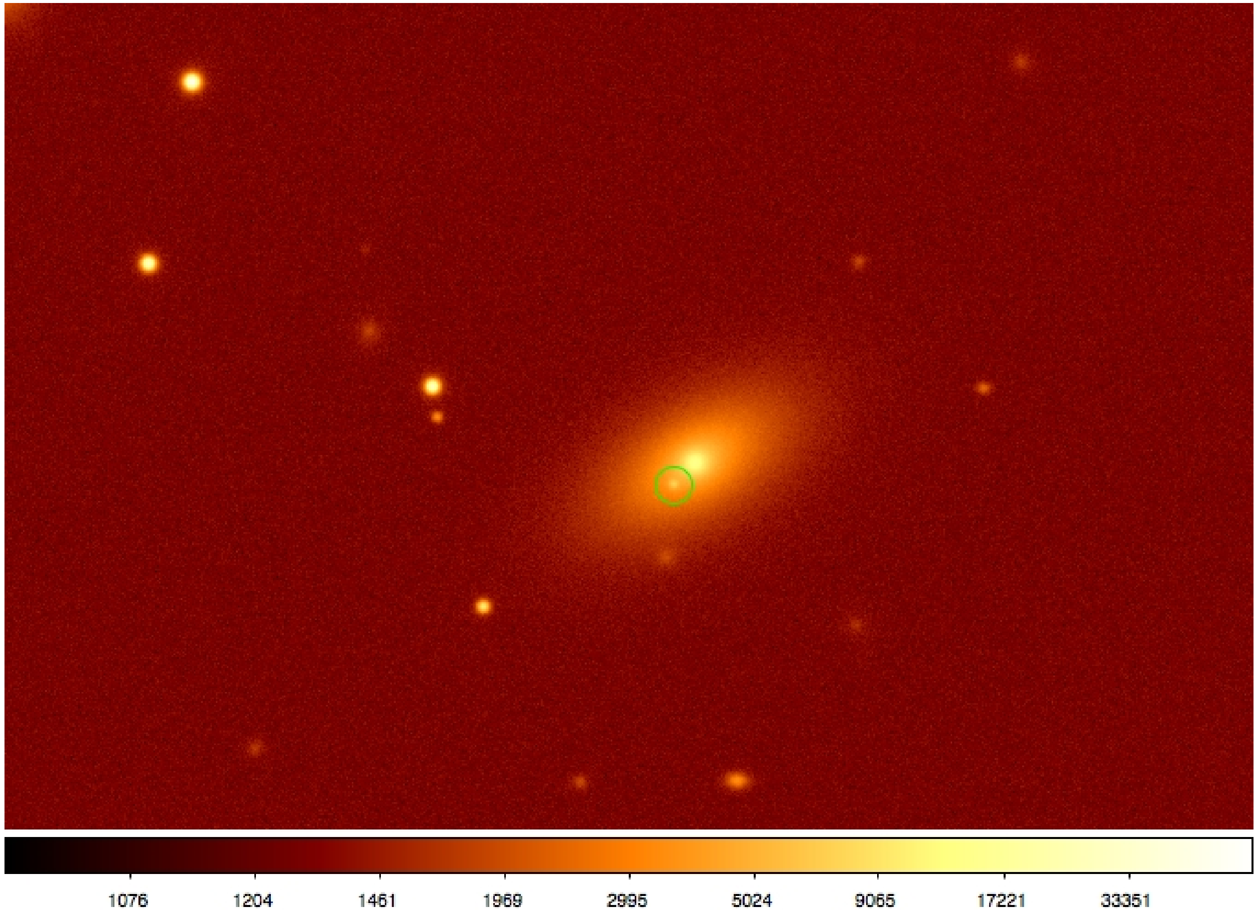
Second Maximum phase

$$m(t) = \frac{f_0 + \gamma(t - t_0) + g_0 e^{\frac{(t-t_0)^2}{2\sigma_0^2}} + g_1 e^{\frac{(t-t_1)^2}{2\sigma_1^2}}}{\left(1 - e^{\frac{\tau - t}{\theta}}\right)} \qquad (1)$$

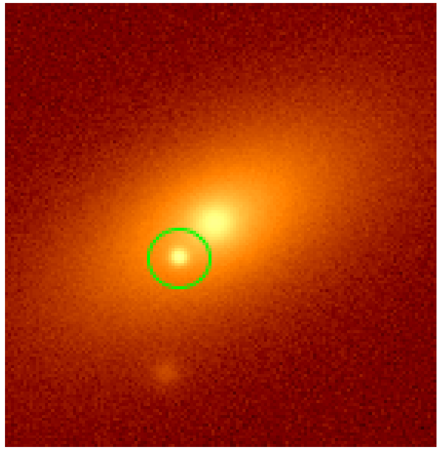
First Maximum

Exponential rising of the light curve

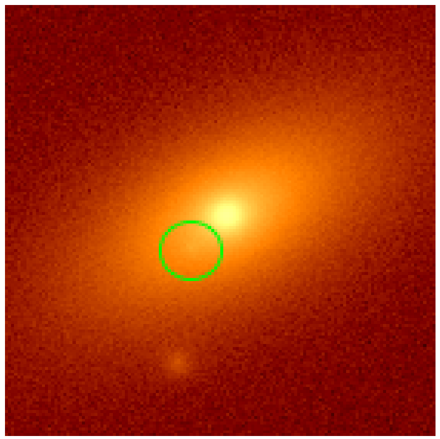
Image of a type Ia Supernova



Stamp of the image of a type Ia at t = 0 d. and B band

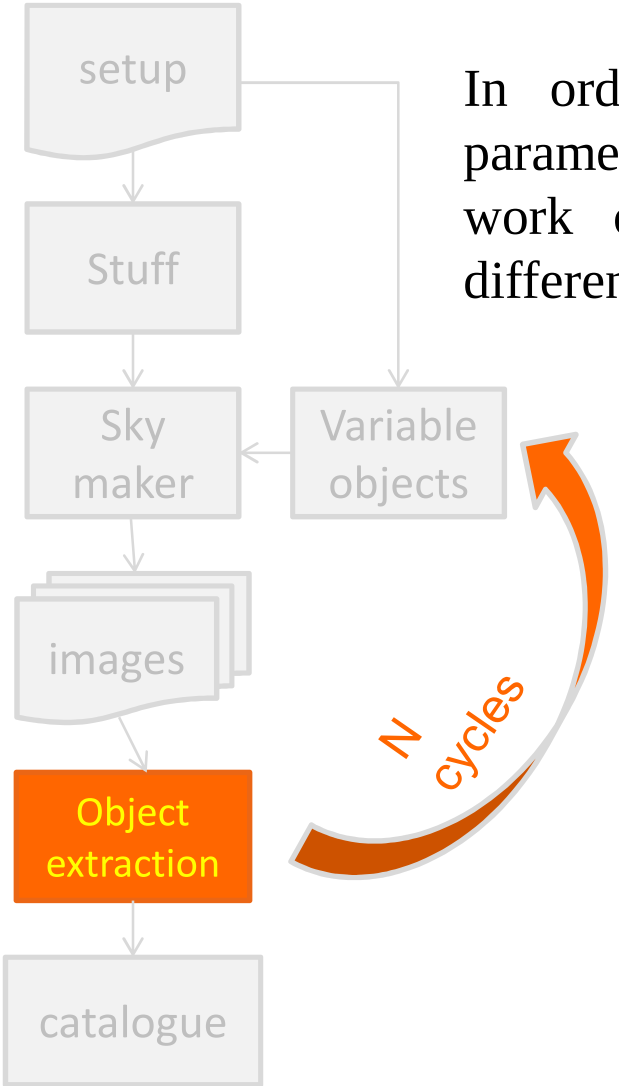


Stamp of a type Ia at t = 13 d.



Stamp of a type Ia at t = 63 d.

Comparison of object extraction packages I



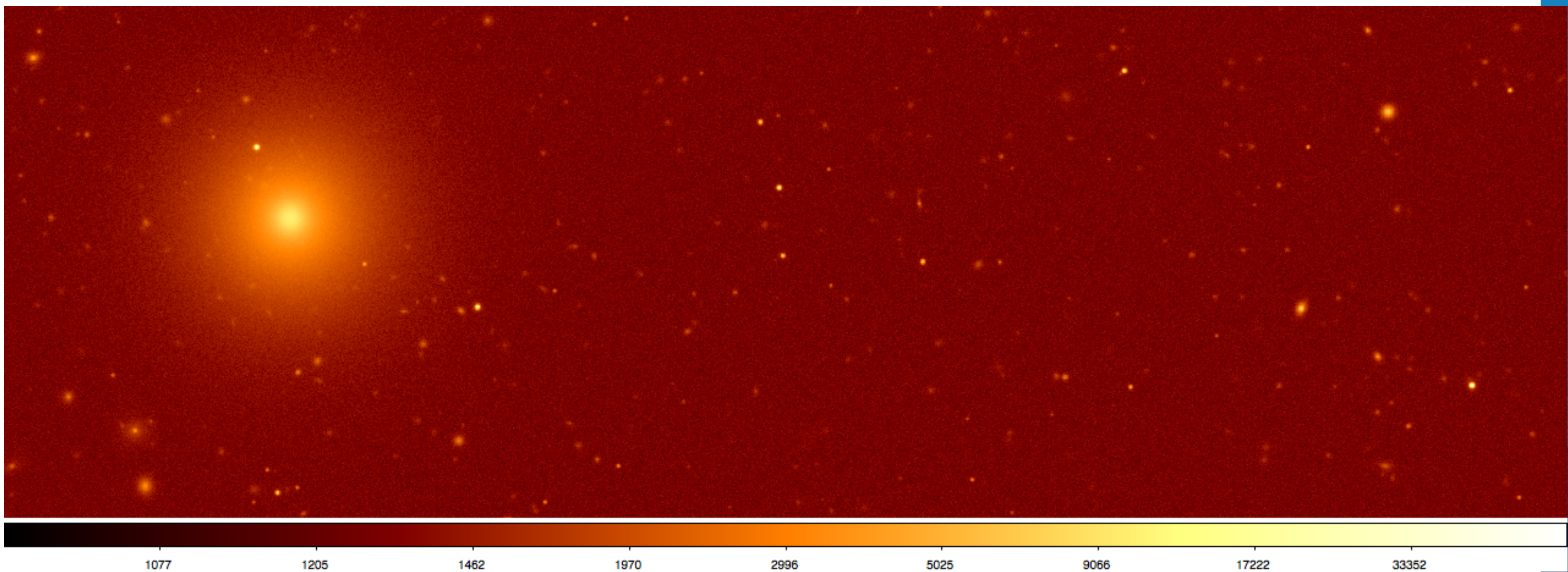
In order to verify the consistence of the imposed parameters and taking into account that the classifiers work on catalogs, instead of images, we compared different algorithms for catalog extraction.

We chose to make a compared between Sextractor + PSFEx(Bertin & Arnouts 1996, Bertin 2011) and DAOPHOT + ALLSTAR (Stetson 1987, Stetson 1994).

Similar work: Becker et al., 2007. They use as **‘true’** the results produced by *photo* (Lupton et al.).

Comparison of object extraction packages II

The images used for the comparison was simulated using the characteristic of the VST optics, and the using ¼ of the size of the camera. We used an exposure time of 1500s and set the magnitude limits between 14 and 26 magnitude and the seeing at 0.7, an average value in Cerro Paranal. All the following comparisons are referred to a johnson/B image.



Stamp of a region of the simulated image in B.

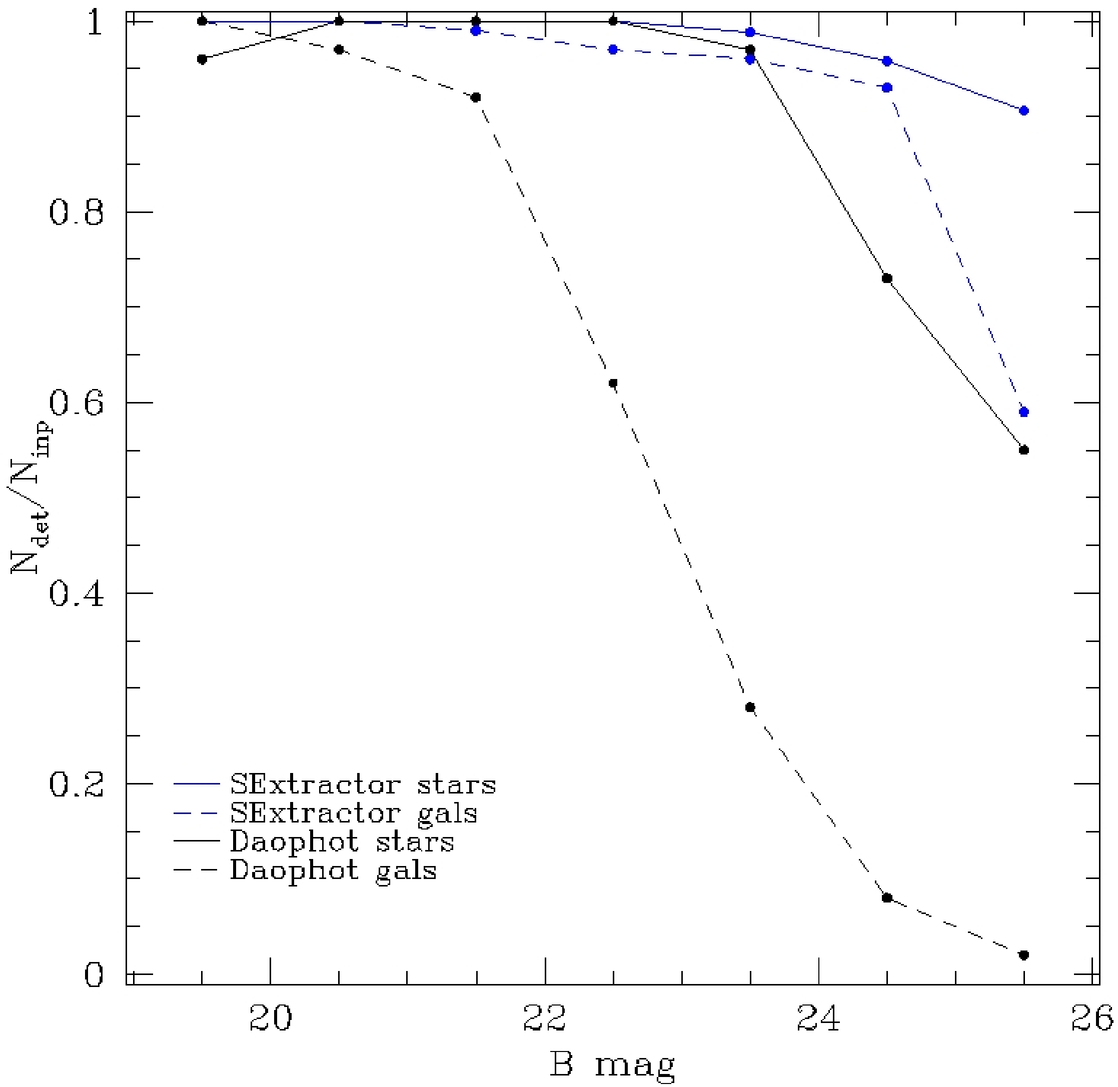
Comparison of object extraction packages II

The comparison between the source extraction algorithms was made evaluating the following aspects:

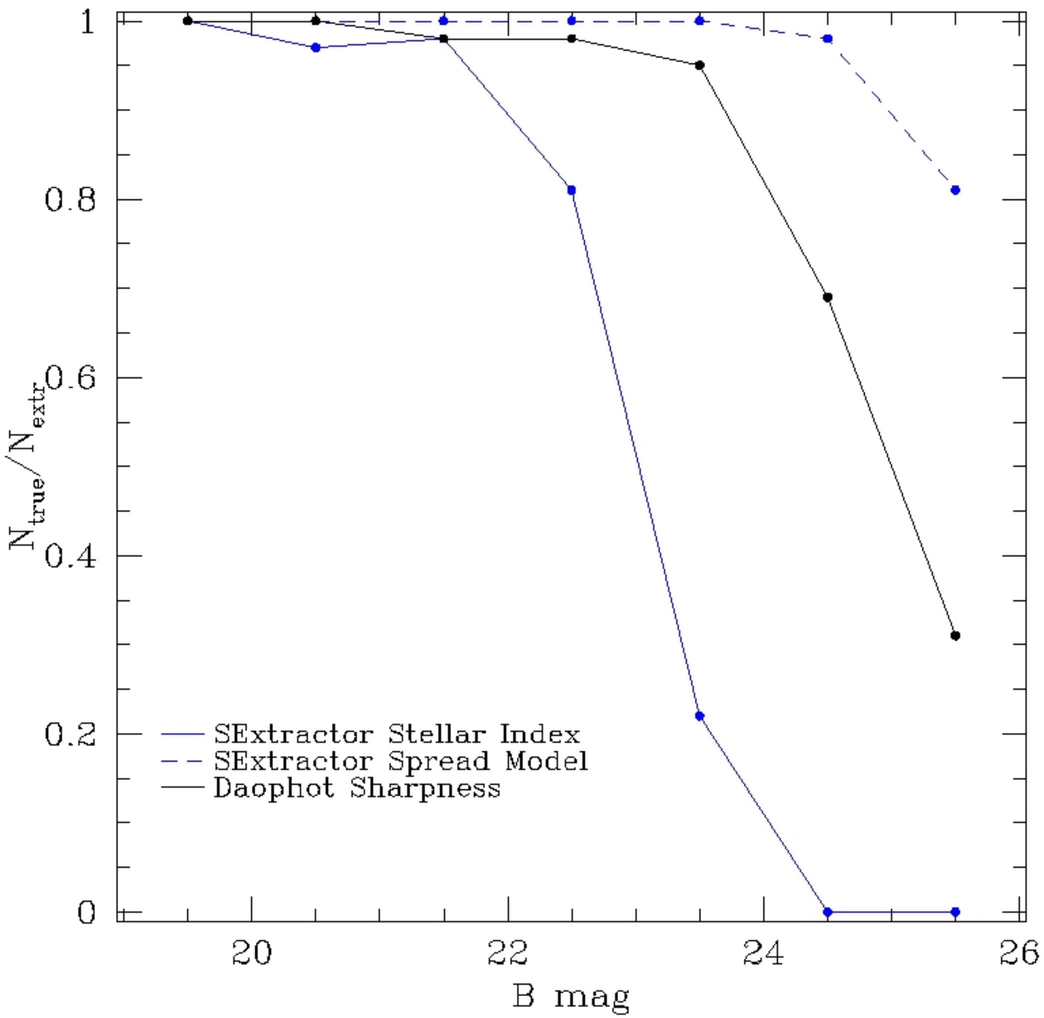
- Photometric depth of the extracted catalog.
- Purity of the extracted catalog.
- Photometry.
- Centroids.

Annunziatella M., Mercurio A. et al. , *Inside catalogs: a comparison of source extraction algorithms*, PASP (in preparation).

Completeness of the catalog



Purity of the catalog



Source classification

Sextractor + PSFEx:

•Stellarity Index.

•Half-light radius.

•μ_{max}.

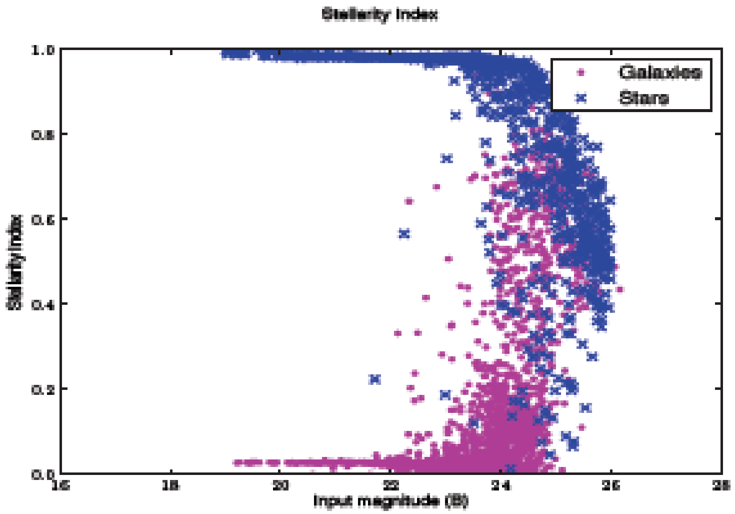
•Spread Model.

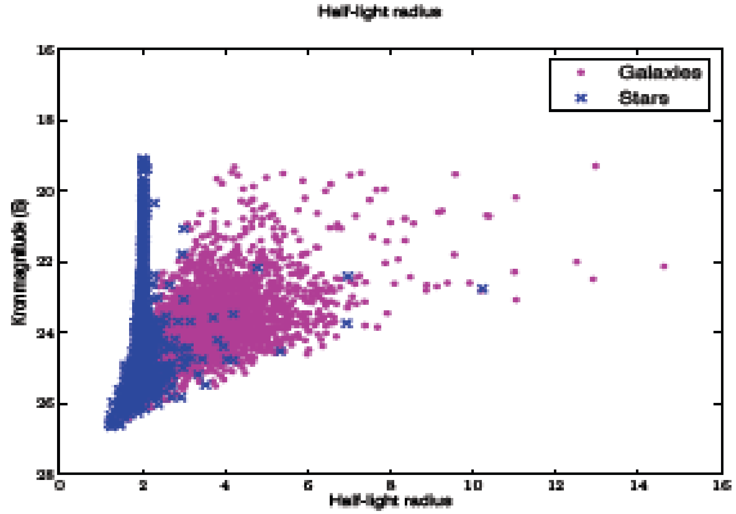
DAOPHOT + ALLSTAR:

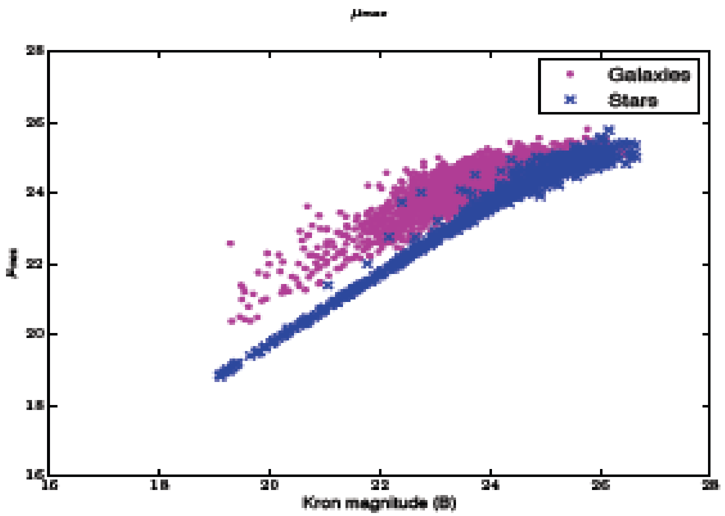
•Sharpness.

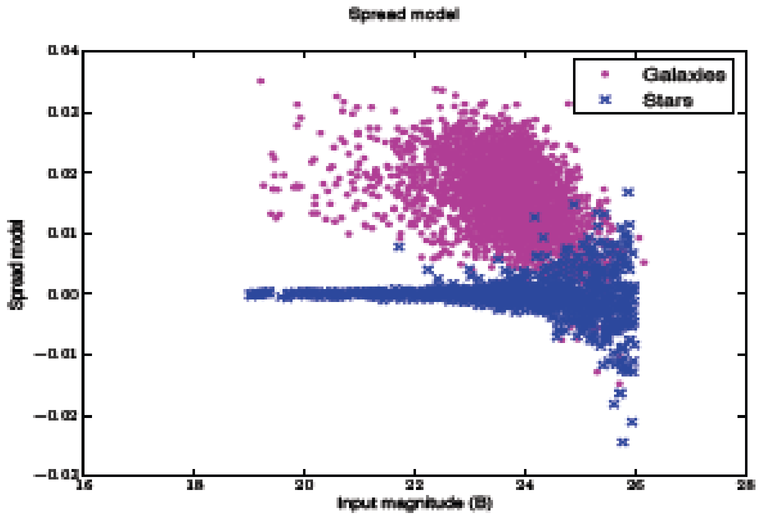
•χ.

Source classification: SExtractor + PSFEx

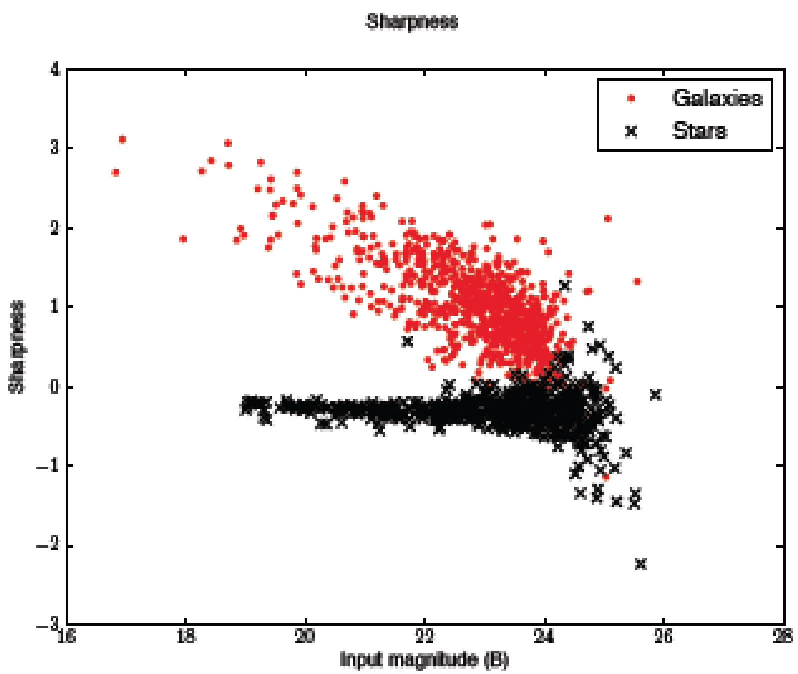


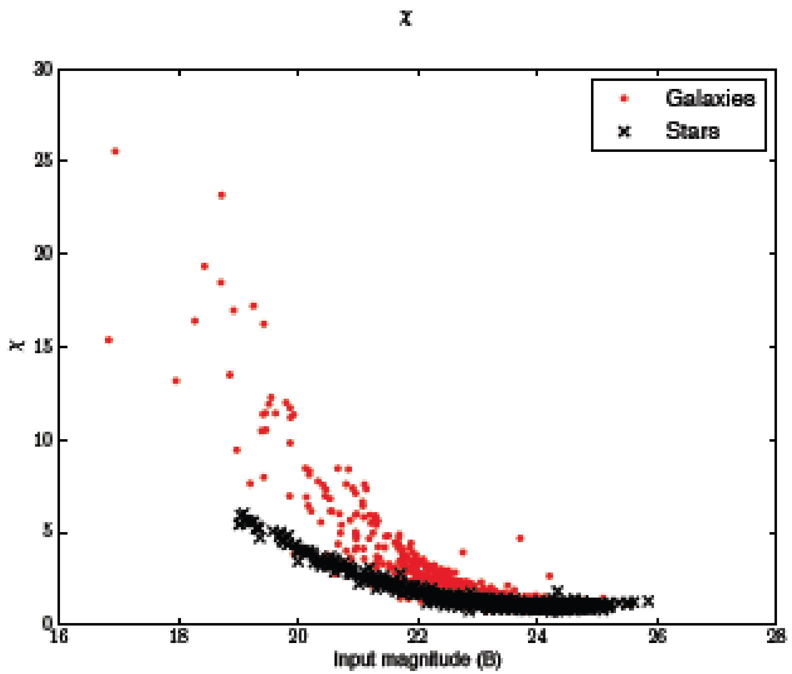




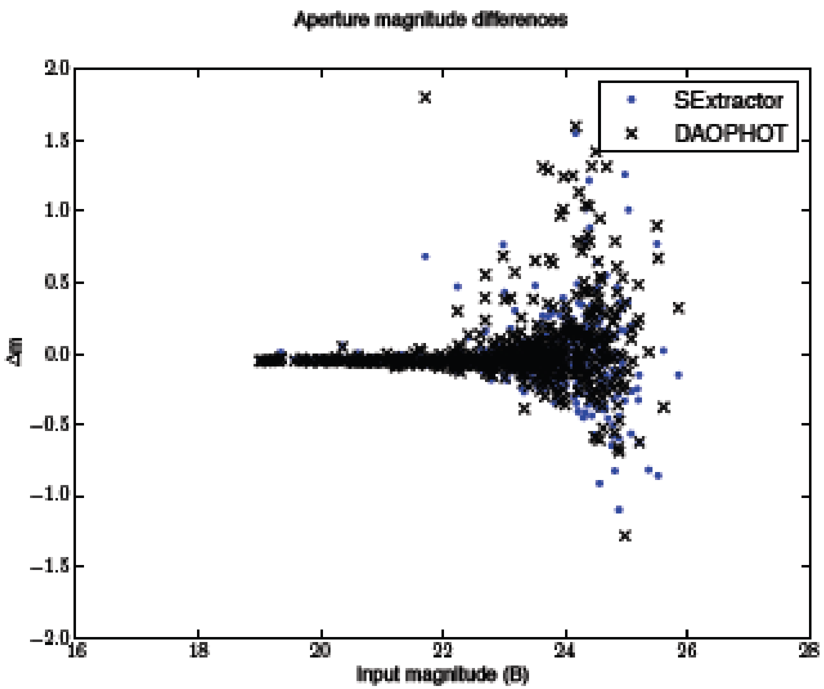


Source classification: DAOPHOT+ALLSTAR

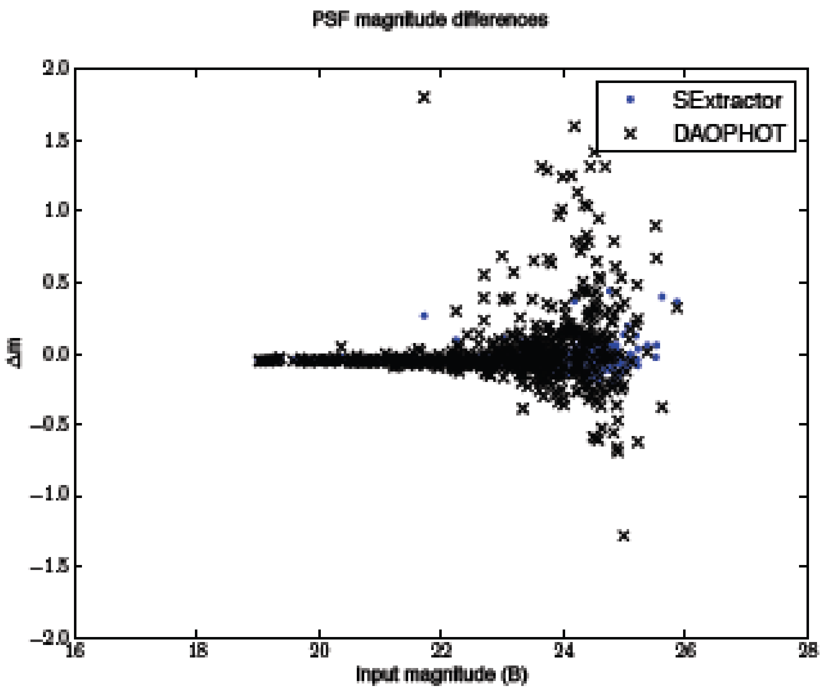




Photometry

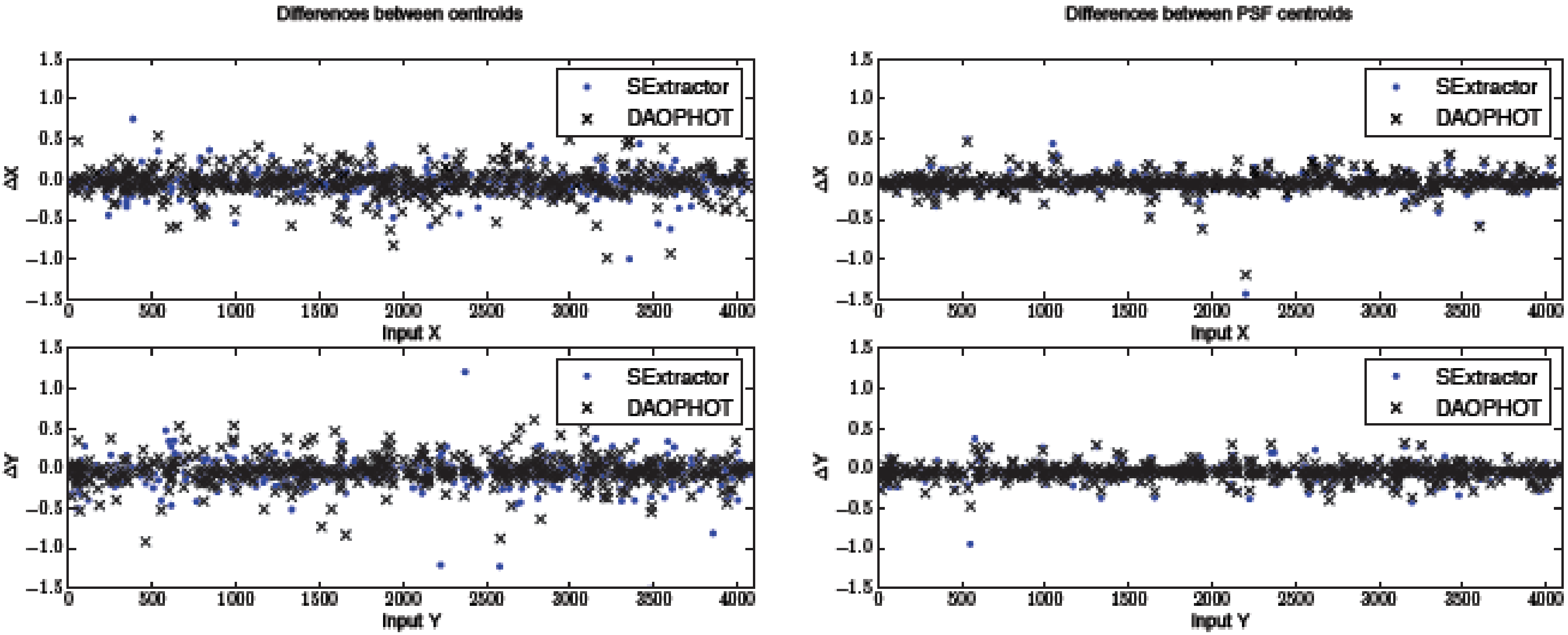


Aperture magnitude differences
SExtractor
DAOPHOT
Δm
Input magnitude (B)



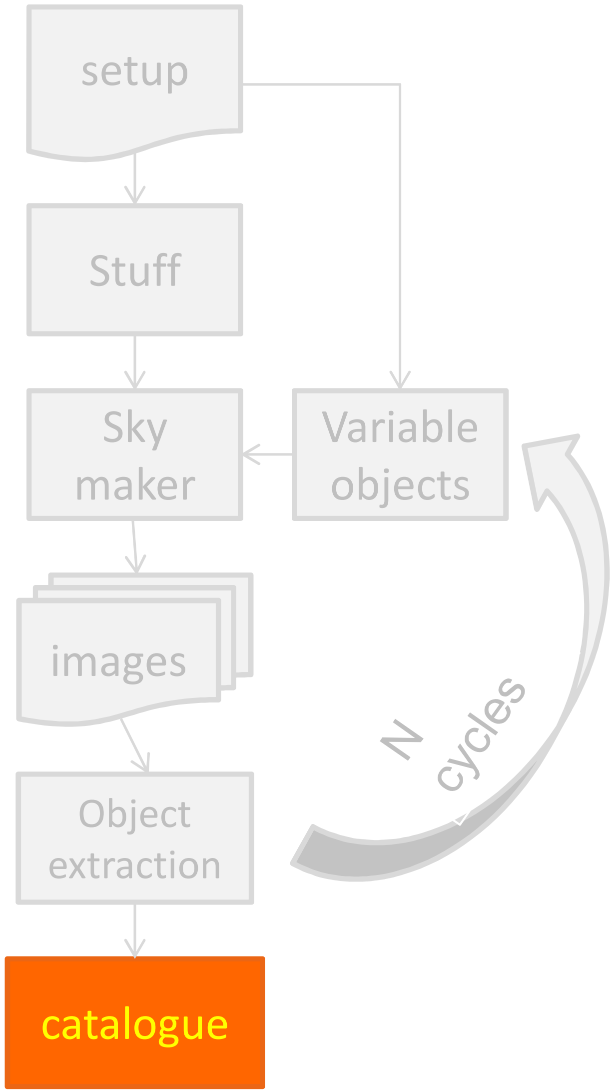
PSF magnitude differences
SExtractor
DAOPHOT
Δm
Input magnitude (B)

Centroids



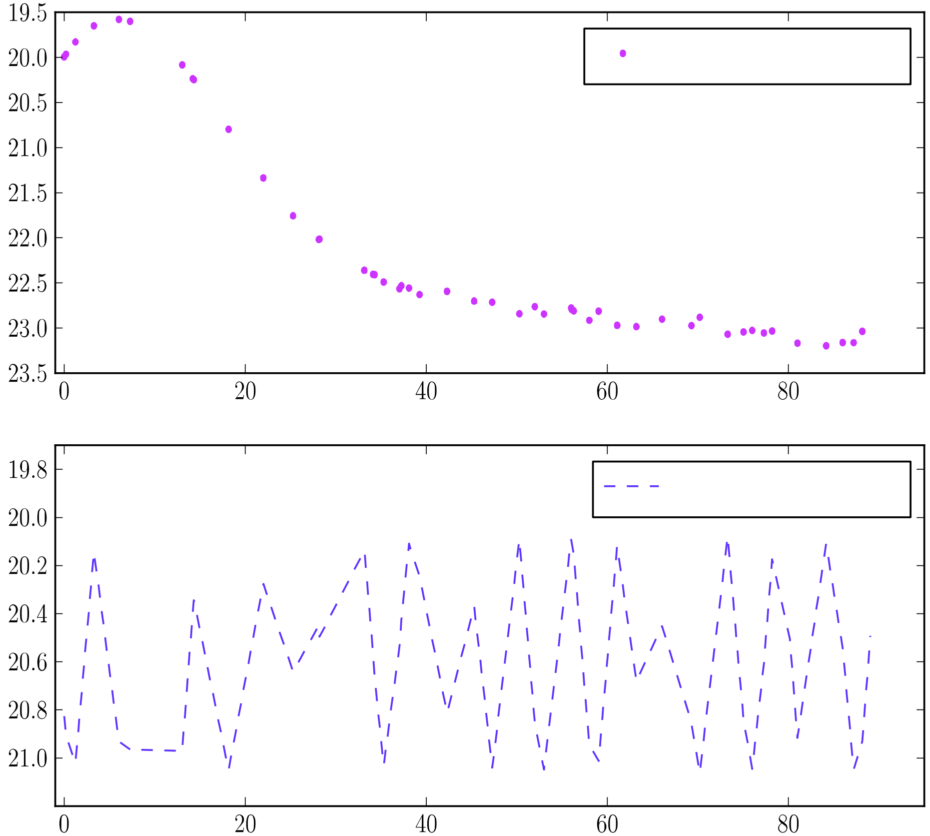
Differences between centroids
Differences between PSF centroids
SExtractor
DAOPHOT
ΔX
ΔY
Input X
Input Y

Single Catalog



- Obtain a single catalog of all the epochs.
- Establish train and test rates.
- Specify the number of the epochs and how they must be chosen:
  - N random epoch different for each object.
  - N random epoch equal for each object.
  - N epoch equally spaced.
- Decide whether or not we want Host Galaxies without SNe.

First classification test: the Simulations



Magnitude limits 18, 25.

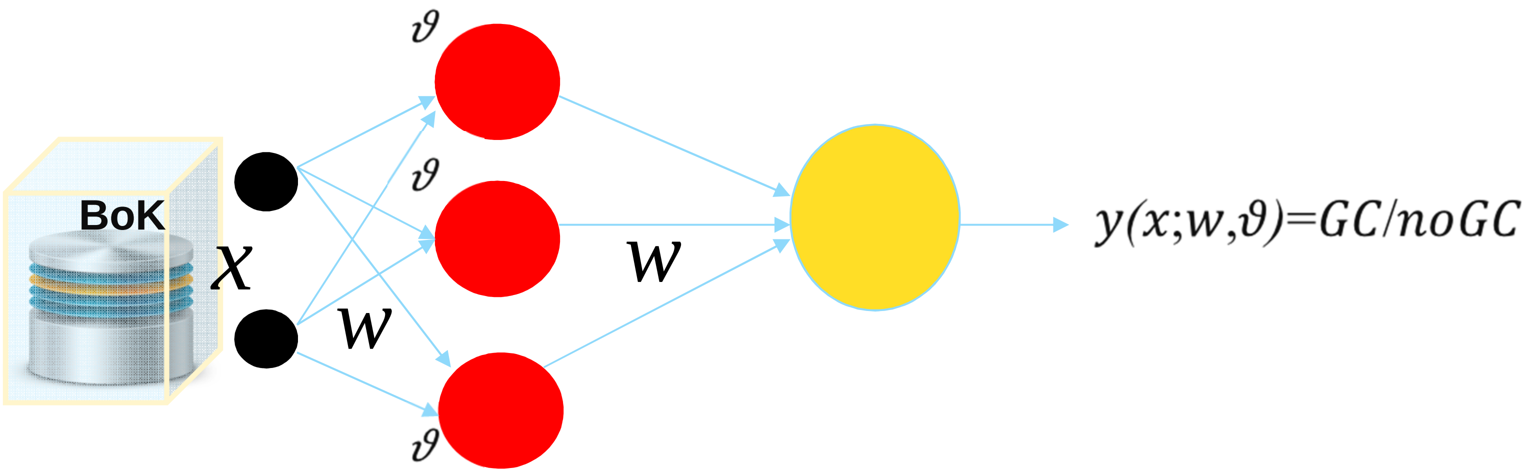
B band.

Size of detector 5120 pixel.

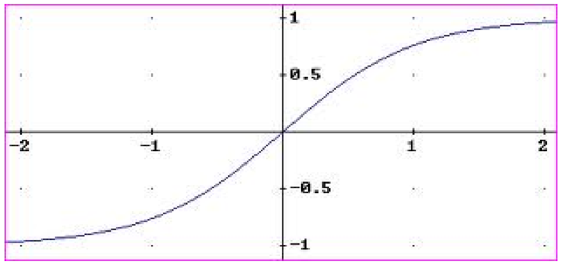
Types of variable objects:

- Type Ia Supernovae with their host galaxies;
- Classical Cepheids;
- Random Objects;
- Host galaxies without their Supernovae.

First classification test: the Simulations- Multi Layer Perceptron

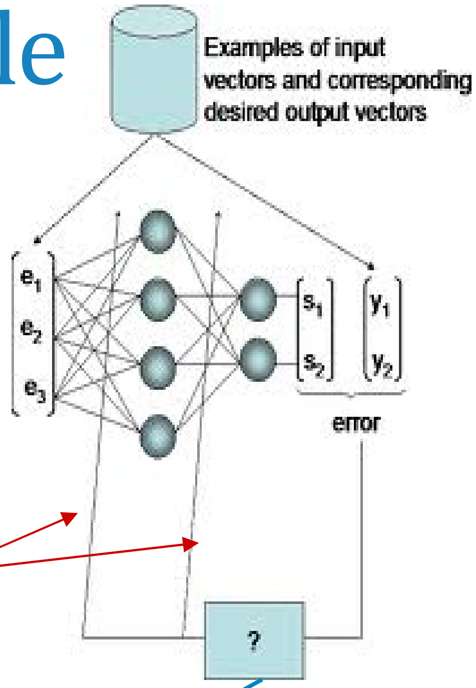


$$y(x;w,\vartheta) = \sum_{i=1}^{M} activ\_func(W_i{}^T x - \vartheta_i)$$



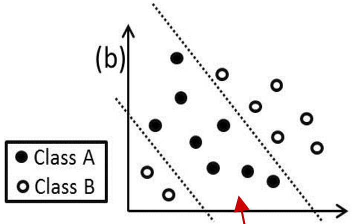
Hyperbolic tangent

Quasi Newton learning rule



$$\min_{w} E(w) = \frac{1}{2P}\sum_{p=1}^{P} E_p(w) = \frac{1}{2P}\sum_{p=1}^{P}\left(y(x^p; w) - d^p\right)^2$$

E_p error related to the *p*-th pattern



$$w^{k+1} = w^k + \alpha^k d^k$$

$d^k \in R^N$ **DIRECTION OF SEARCH**

$\alpha^k \in R$ **STEP**

$$\nabla^2 E(w^k) d^k = -\nabla E(w^k)$$

Hessian approx. (QNA)

More Info on MLP-QNA model for classification and regression in astrophysical contexts

(Brescia et al. 2011- MNRAS, accepted, arxiv:1110.2144)
(Cavuoti et al. 2012 – A&A, submitted, arxiv:1206.0876)

Simulation nr. 1 – The BoK

OBJECTS	TYPE	FULL	TRAIN	TEST
Variable	SN-Ia	80	64	16
	Cepheid	80	64	16
	Random	80	64	16
	Host Galaxies with SN	80	64	16
Not Variable	Stars	216	172	44
	Galaxies	1259	1007	252
	Host Galaxies without SN	11	8	3
TOTAL		1806	1443	363

Simulation nr. 2 – The BoK

OBJECTS	TYPE	FULL	TRAIN	TEST
Variable	SN-Ia	206	160	46
	Cepheid	200	160	40
	Random	200	160	40
	Host Galaxies with SN	206	160	46
Not Variable	Stars	617	493	124
	Galaxies	3510	2808	702
	Host Galaxies without SN	17	18	5
TOTAL		4956	3959	1003

Statistical Results

3 quality evaluation criteria

- **CA - Classification Accuracy**: fraction of objects correctly classified (either variable or not-variable), with respect to the total number of objects in the sample.
- **CO - Purity**: fraction of variable objects correctly classified as variable.
- **CN - Contamination**: fraction of not variable objects erroneously classified as variable.

	Object Number		TRAINING			Object number		TEST		
	Not Variable	Variable	CA %	CO %	CN %	Not variable	Variable	CA %	CO %	CN %
SIM1	1187	256	98,96	96,59	0,51	299	64	87,60	56,72	5,41
SIM2	3319	640	98,92	97,61	0,82	831	172	88,28	67,96	7,26

astromeeting

Forthcoming developments

In the next future we plan to:

- Add other types of variable objects to the simulation, in particular expanding the category of SNe and pulsating variables.

- investigate the statistical indicators according to the collaboration with Djorgovski et al.

- test other classifiers: MLP, SVM, SOM, PPS, WGE.

We are open to collaborations.