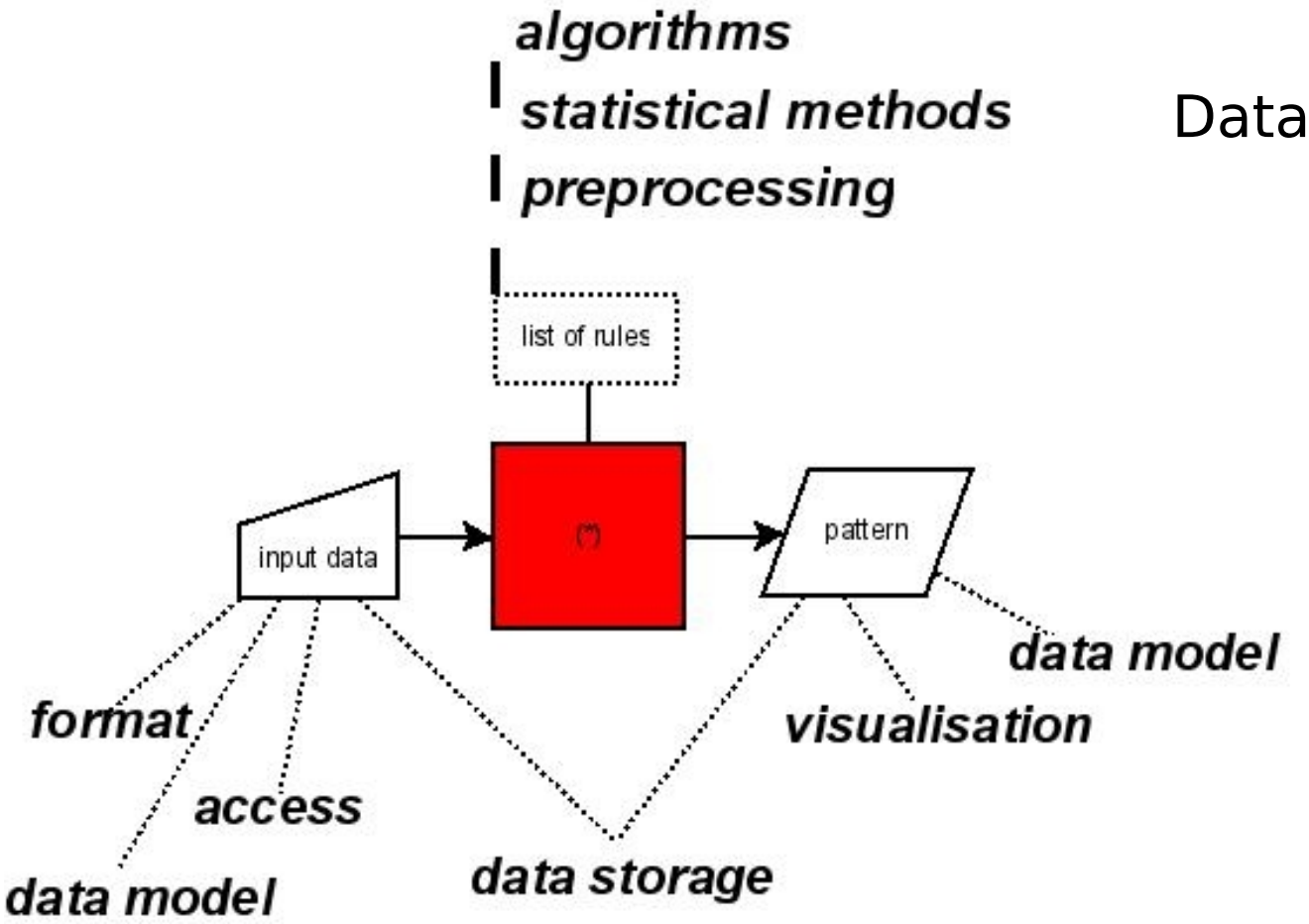
Virtual Observations 2016
Data Mining in Astronomy

Data Mining

Scientific Paradigm

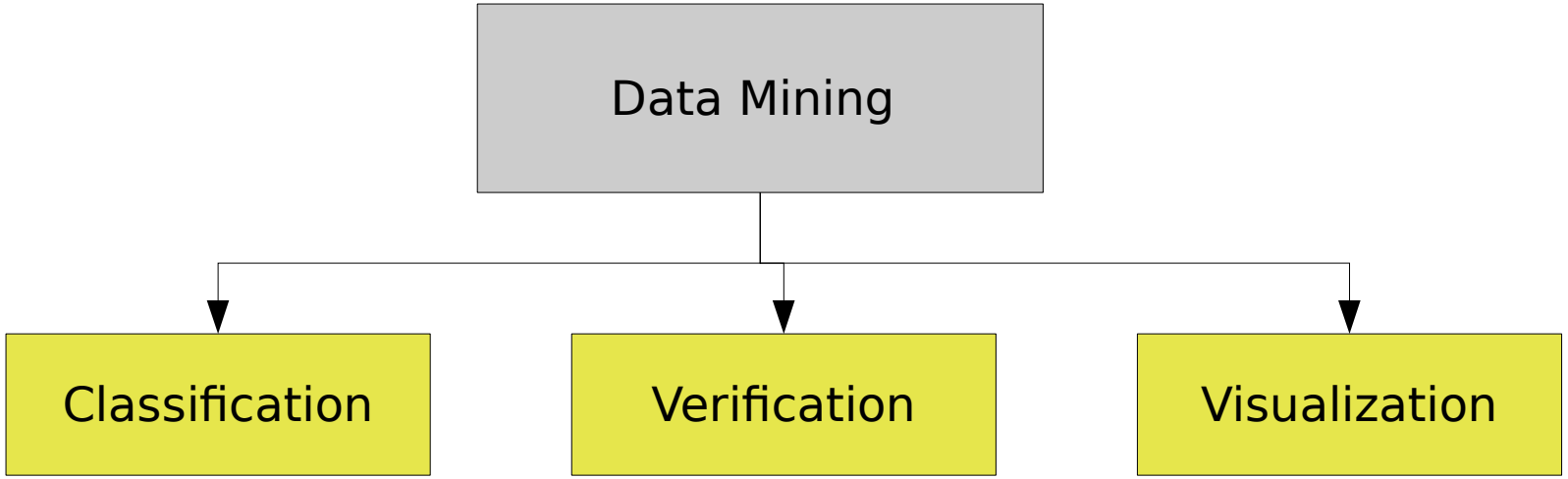
- First : Observations
- Second : Theory
- Third: Computer simulations
- Fourth: Data mining

Data Mining



Database: Data format
Access
Data model
Preprocessing
Processing

Data Mining: Connection



Machine Learning

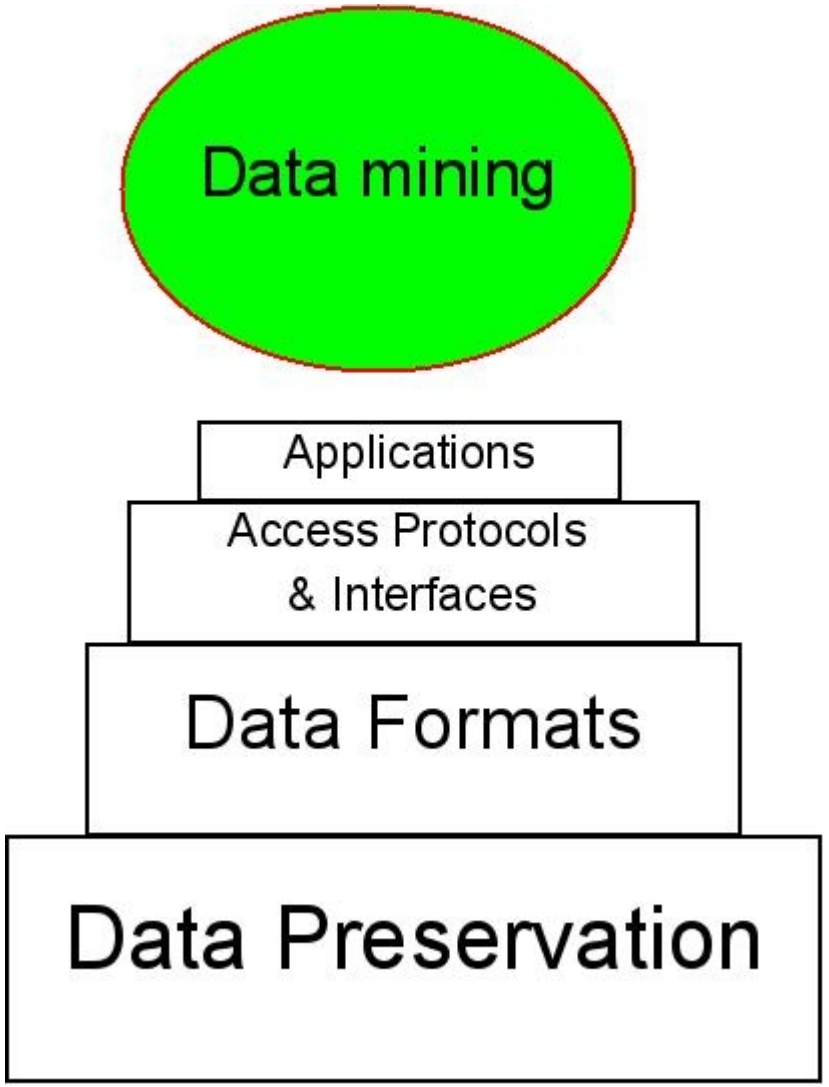
Statistics

Preprocessing

Pattern Recognition

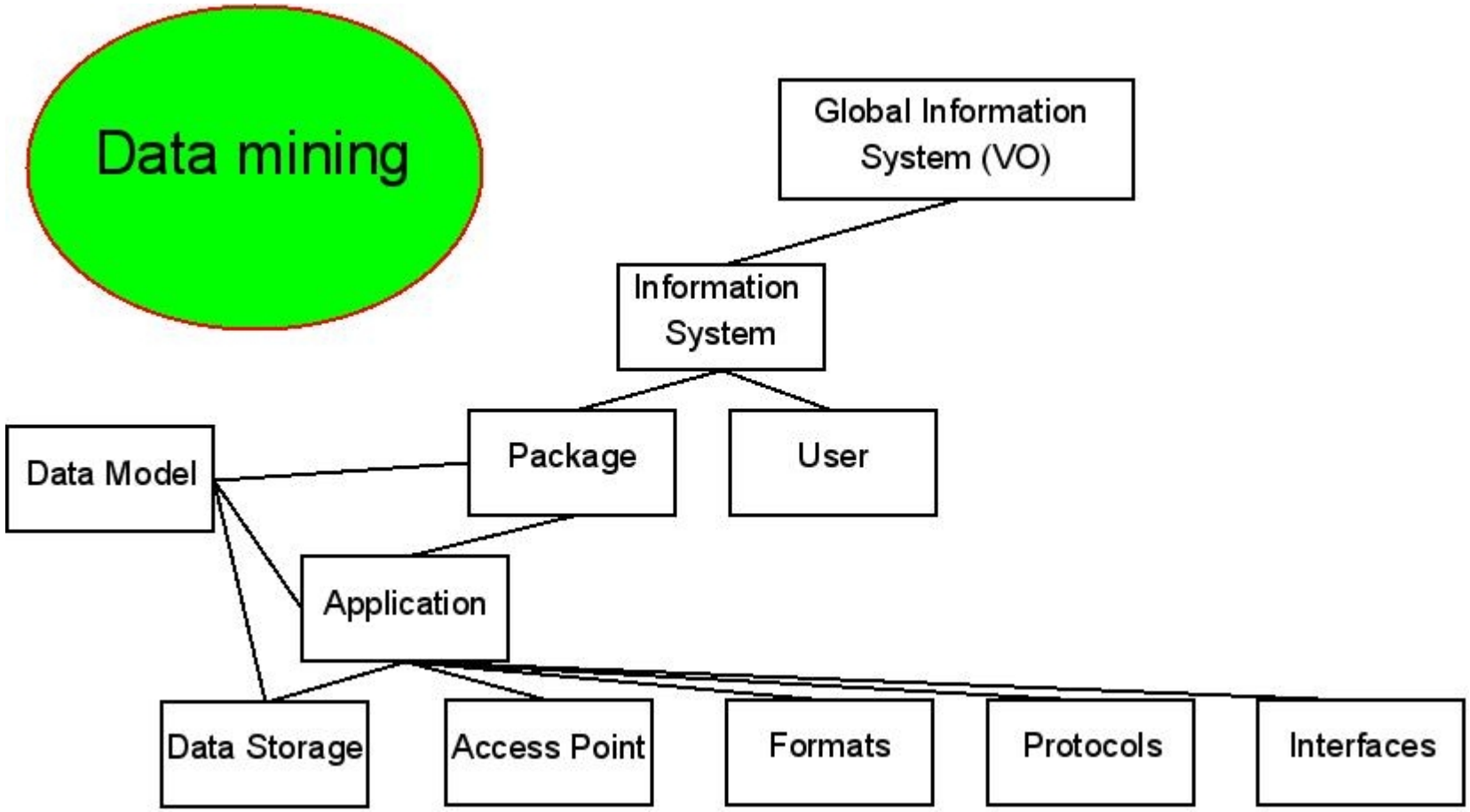
AI

Data Mining: Concept

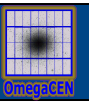


- Data Preservation: OAIS
- Data Formats: FITS +metadata
- Access & Interfaces: http, web interfaces, XML
- Applocations

Data Mining: Implementation



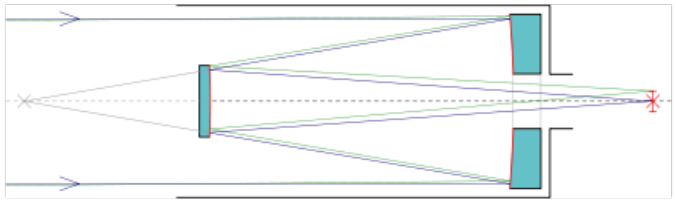
Data Mining in Astronomy: Examples of Use case

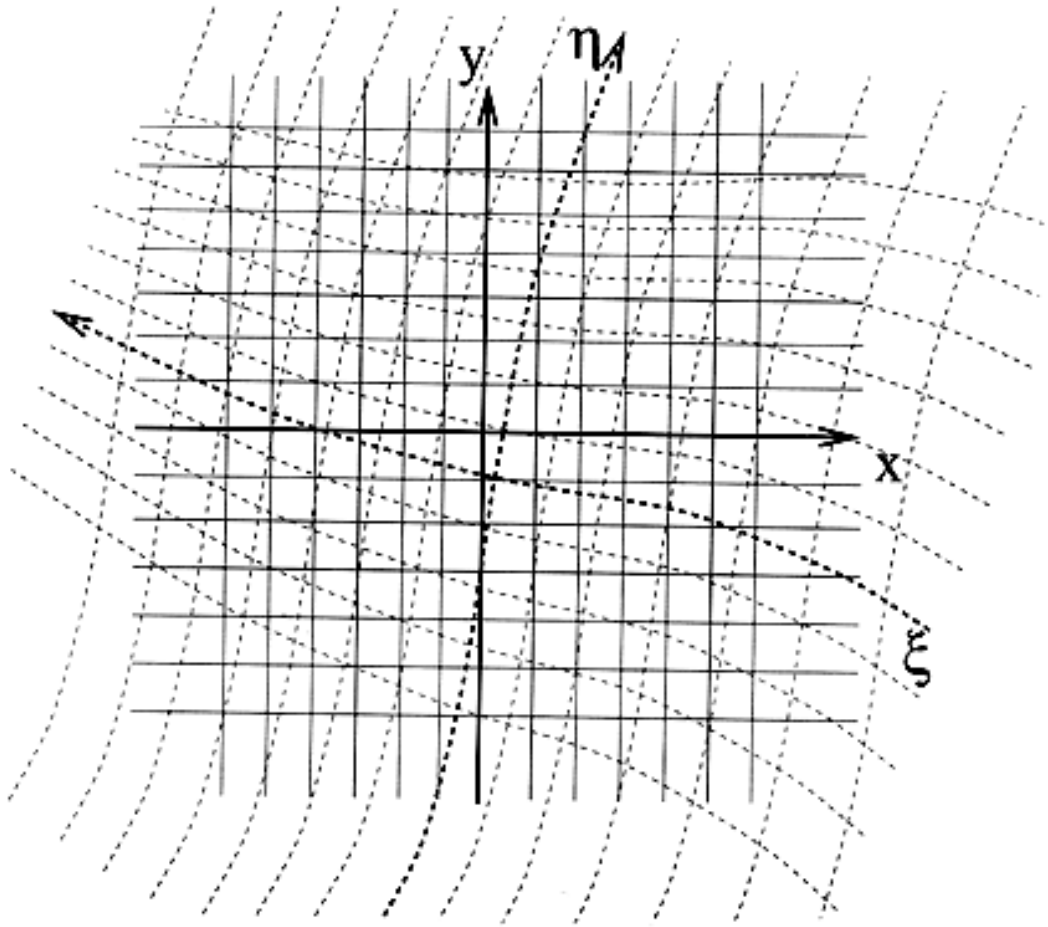


- Cross-identification
- Parametric search
- Pattern recognition
- Coordinate transformation
- Source extraction

Cross-identification

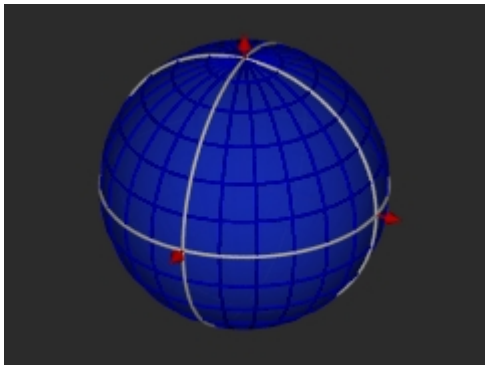
- Different coordinates
- Different epochs
- Reference system



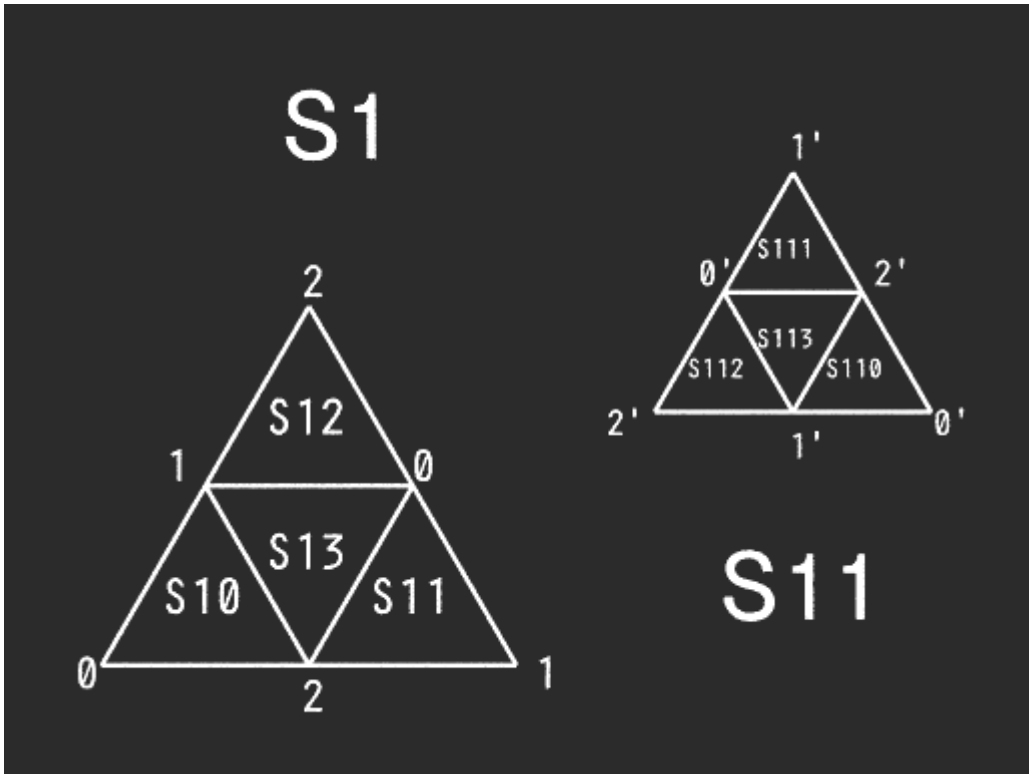


Cross-identification: HTM

http://skyserver.org/htm/



- Time consuming
- Distance



Parametric search

- Photometric and coordinate search
- Object parameters (shapes, classificators)
- SDSS: benchmark of 20 “typical” queries

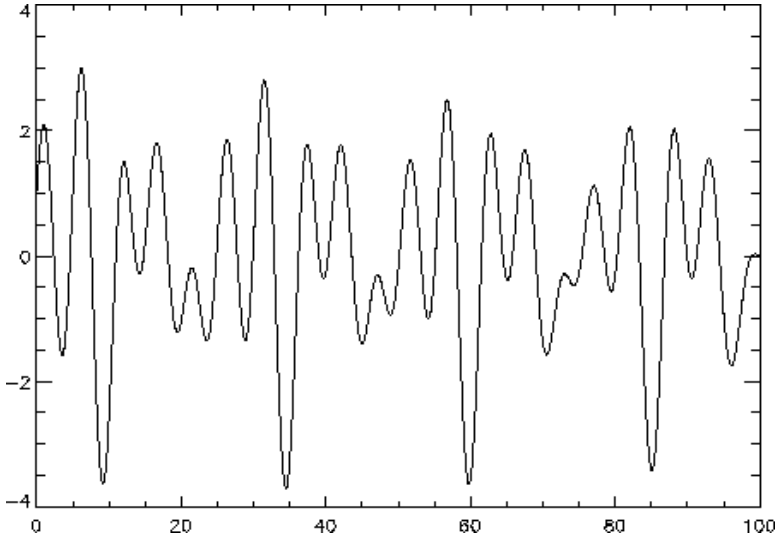
```
Search for Cataclysmic Variables and pre-CVs with White Dwarfs and
very late secondaries:
u-g < 0.4
g-r < 0.7
r-i > 0.4
i-z > 0.4
------------------------------------------------------------------
--------
SELECT run, camCol, rerun, field, objID,
u,g,r,i,z,
ra, dec
INTO ##results
FROM PhotoPrimary
WHERE (u - g) < 0.4 and
(g - r) < 0.7 and
(r - i) > 0.4 and
(i - z) > 0.4
```

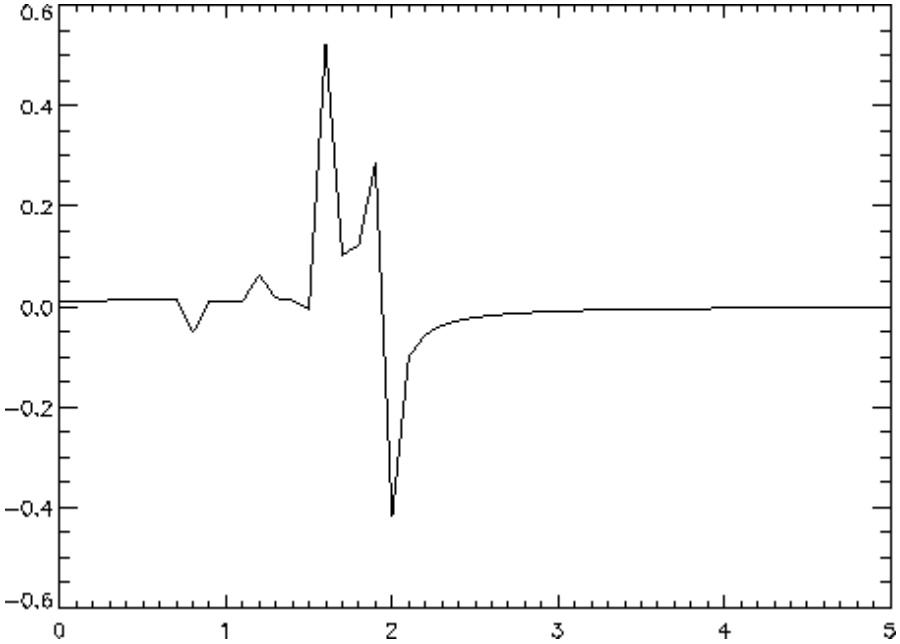
Pattern recognition

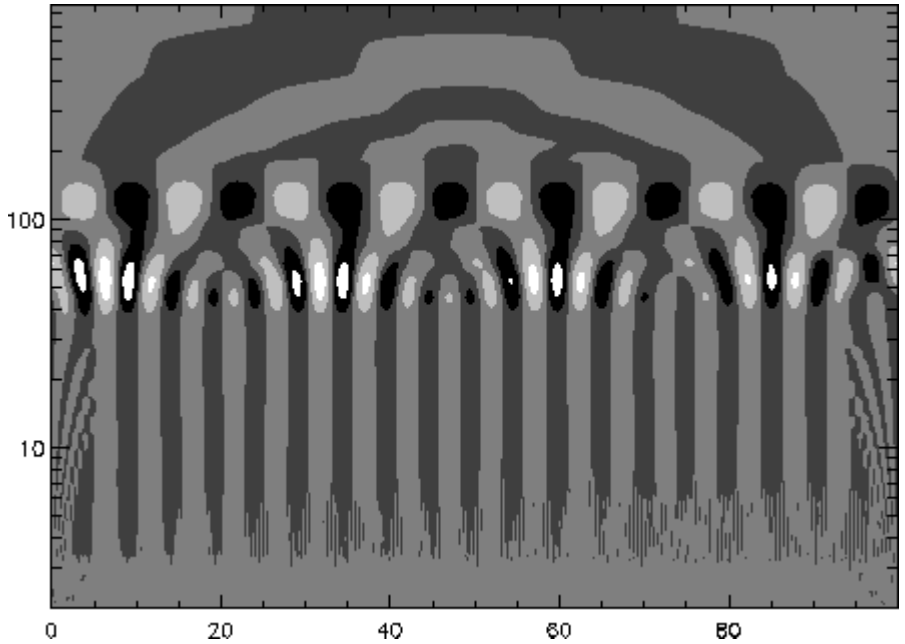
- Parametric
- Non-parametric

Coordinate transformation

- FFT
- Wavelets

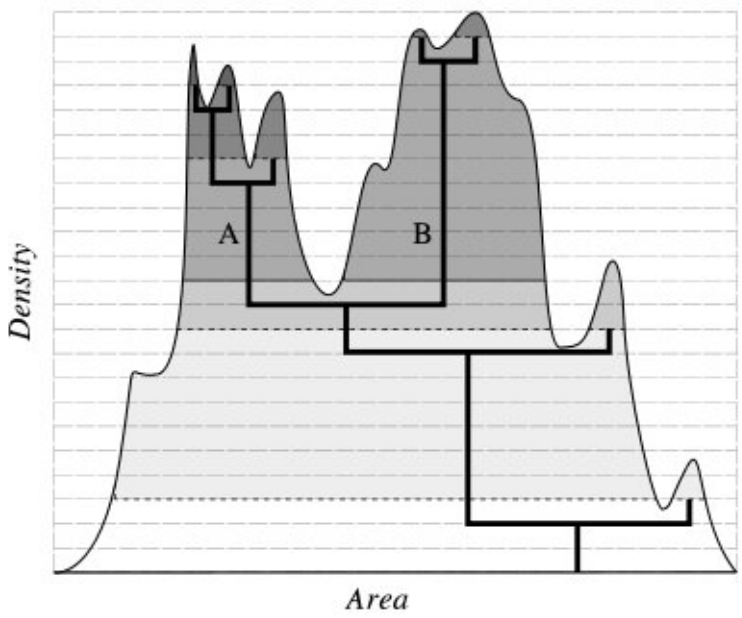


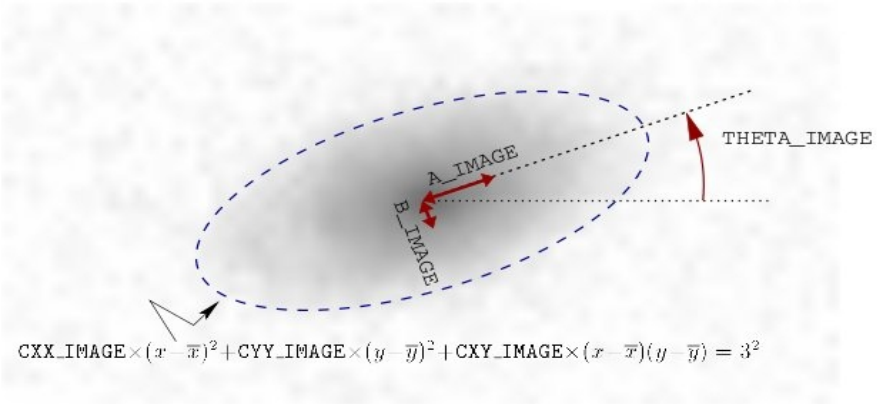




Source extraction

- Sextractor





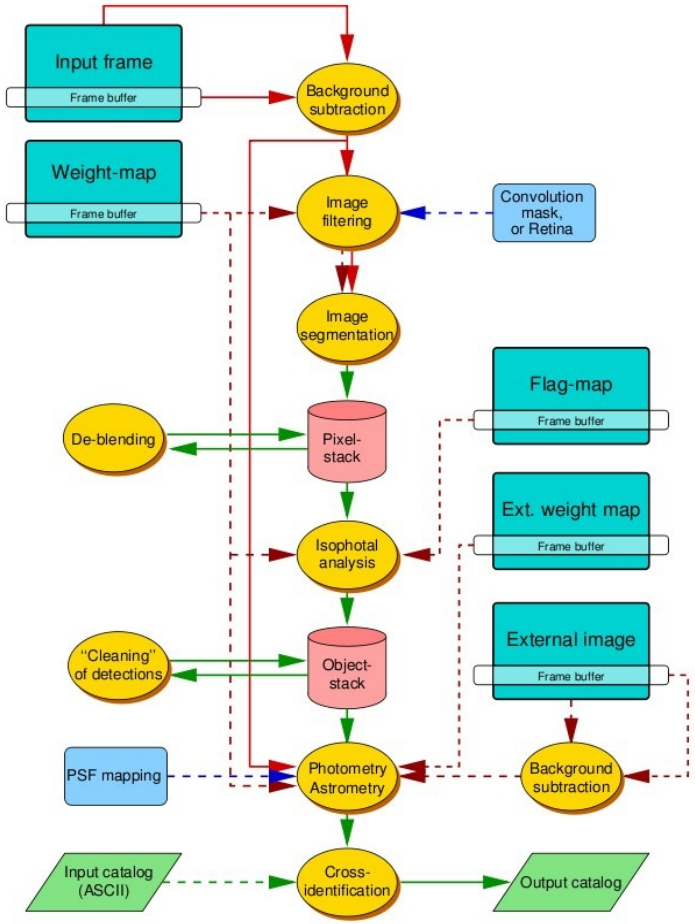


Figure 1: Layout of the main SExtractor procedures. Dashed arrows represent optional inputs.