Exowarning mode: proposed detection algorithm

Céline QUENTIN, P. BARGE, R. CAUTAIN
Laboratoire d'Astrophysique de Marseille

CW9 – Estec, Noordwijk, 5-9 December 2005.

Requirements

- To detect signals with small transit counts (exclude harmonic or linear filters)
- To sort candidates following confidence level for oversampling
- To work within short delays
  - Fast enough to process 12000 LCs in 1-2 days

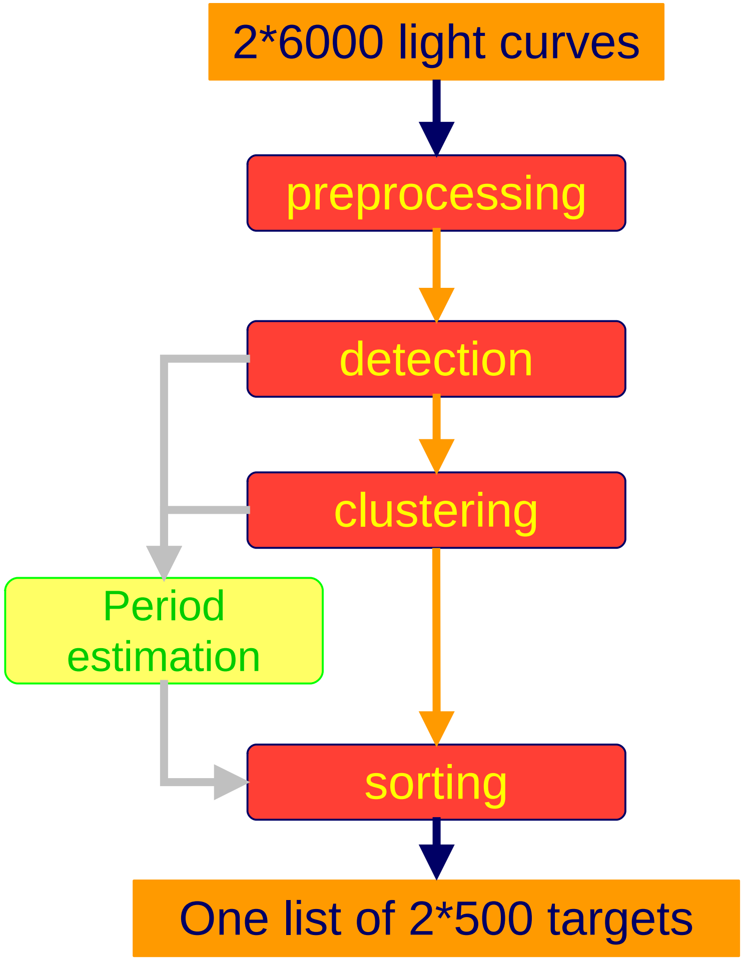
Proposed method

- Identification of transit-like shapes in the LCs
- Based on a morphological approach following the work of V. Guis (2005)

--> Morphological Individual Detector (MID)

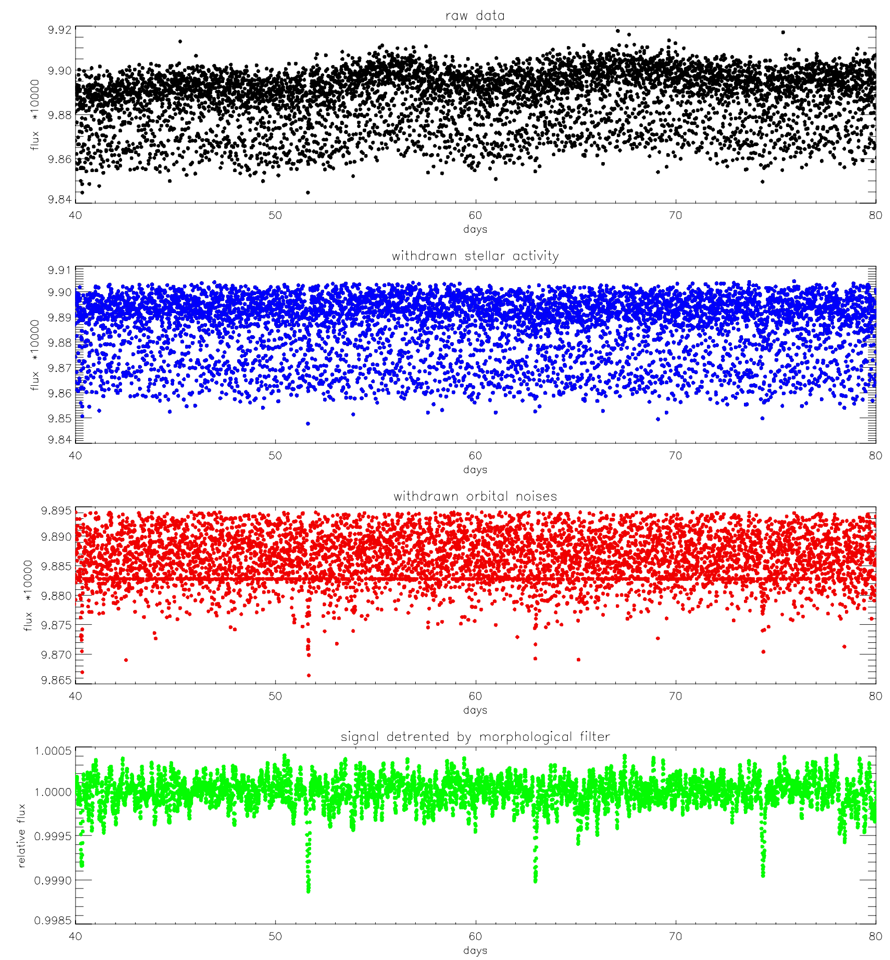
Algorithm overview

- Pre-processing
  - to filter stellar activity, orbital noises and scattered light
- Detection
  - In sliced data
  - Works on a "clean signal"
- Clustering
  - Discriminates noise from possible transit events
- Sorting of candidates
  - following confidence level
- Period estimation



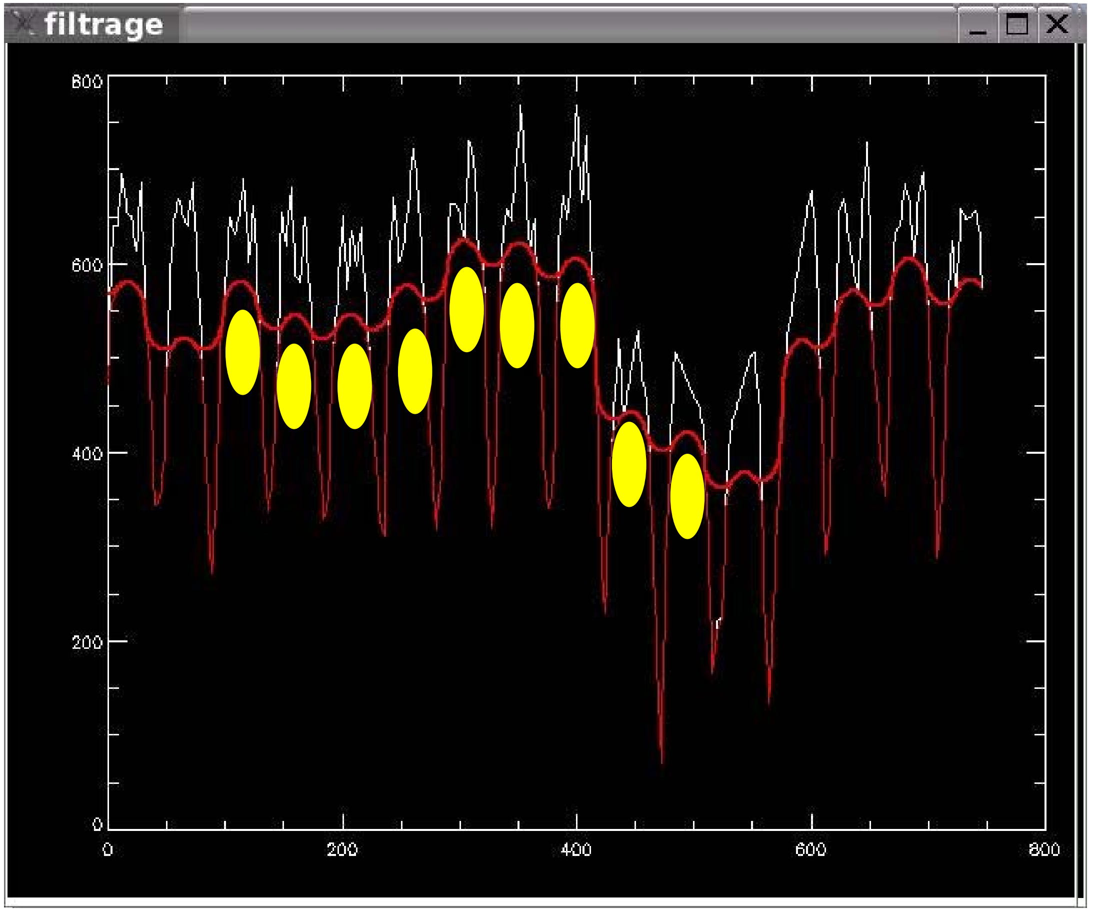
Pre-processing

- Three stages:
  - Non-linear filtering to remove slow variations of the signal (stellar activity)
  - Correction of orbital noises
  - Morphological filtering "gauging filter"



Gauging Filter

- **Use morphological operators:**
  - **erosion and dilatation** by a structuring element or gauge
  - **opening transformation** = dilatation(erosion)
  - **closing transformation** = erosion(dilatation)
- Idempotent transformation
- Remove
  - High frequencies
  - Residual SAA and scattered light

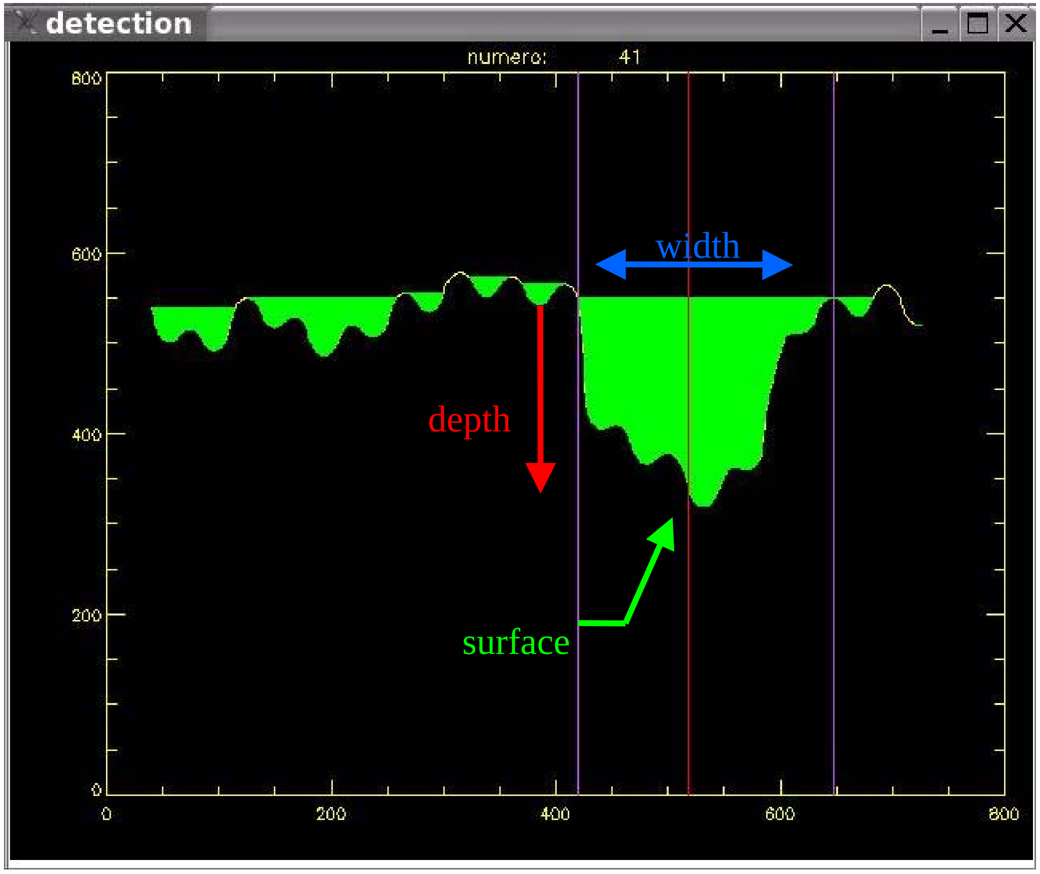


Detection

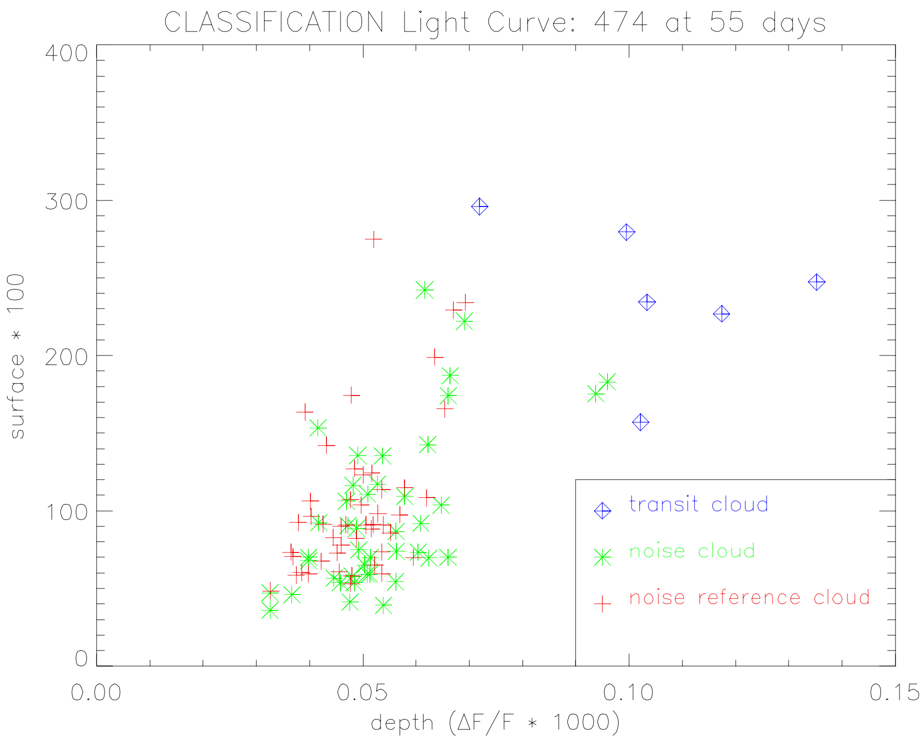
- Light curves are sliced in blocks of 36h each 24h
- Detection on each block
- Segmentation with the **waterline method**.
- Identification of the deepest feature
- *Determination of three parameters:*

  depth, width, surface,

  in the two parts of the signal.



Clustering and sorting



Projection in 2D-space (depth, surface) for one light curve in 55 days.

- *Three clusters on a map:*
  - *noise in opposite signal,*
  - *possible transit events or candidates,*
  - *noise features*
- Candidates are sorted following confidence level

  index defined with distances « noise cloud » / « transit cloud »

Tests and results

- This method was tested on simulated light curves (1000 « realistic » LCs produced during BT1)
- Comparison with the BLS algorithm

Days	BLS		MID		MIDCI	
	R	W	R	W	R	W
10	0	0	6	1	3	0
20	3	8	6	13	4	4
50	7	8	9	5	9	3
100	9	3	9	23	9	0
150	9	2	8	16	12	0

- BLS: Box Least Square
- MID: Morphological Individual Detector
- MIDCI: MID with statistical approach to remove « systematics »

--> Detection is better with MID than BLS during the first 20 days

Performances

Present results:

- MID satisfies the computing time limitation
- Long transits are better detected than short ones
- Good results when running on BT1 light-curves
- MID is better than BLS on 20-day long LCs
  (mainly for events with a small transit number)

Prepared Improvements:

- To estimate possible period of the transit candidates
- To better use collective information to remove systematics and reduce false alarm

Period determination

Specific algorithm for small transit number:

- Works on collections of transit candidates, even if some transits are missing :

T_1 (T_2) T_x T_3 (T_4) (T_5) T_6

- Estimates duration $D_{i,j}$ between events i & j
- Estimate uncertainty on the dates
- Search for a common divider:

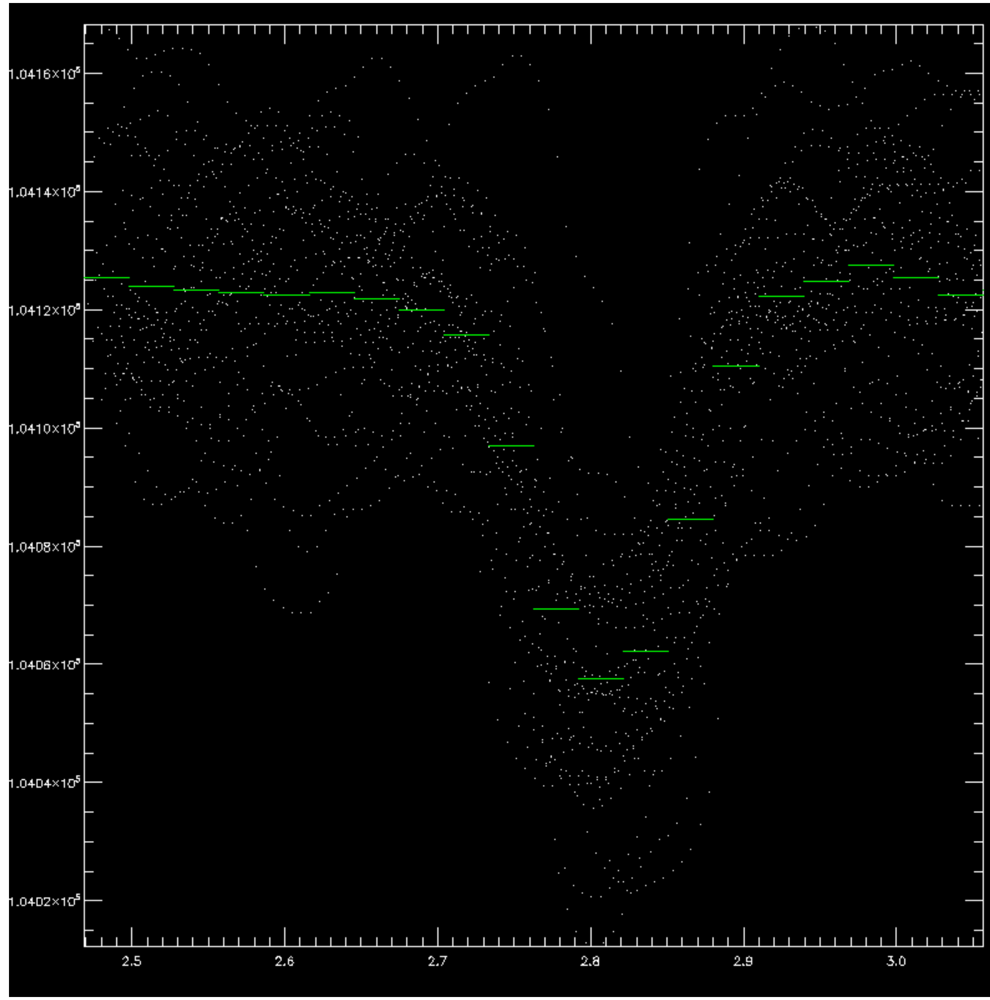
$P = (D_{6,1} - 2*D_{3,1})$ and $(D_{3,1} - 2*P) \sim$ zero

- A set with a "bad" event T_X ileads to unrealistic period and is rejected.

-->> The algorithm is fast and robust but still under development

Period likelihood estimation

- Once transits are detected (at least 2) a period can be estimated.
- A confidence level is obtained by folding the signal in a window(WPDM algorithm)



Current status and prospective

The proposed method:

- Is ready to be implemented in the operational chain
- It is well suited for weak transit numbers and is complementary to the BLS
- The clustering step will help to identify others classes of events than the planetary transits

However:

- The filtering step needs some improvements and will change when true data will be available
- The algorithm for the determination of the period needs to be complemented
- Systematics need to be removed, so collective informations will be treated by a PCA or sysrem