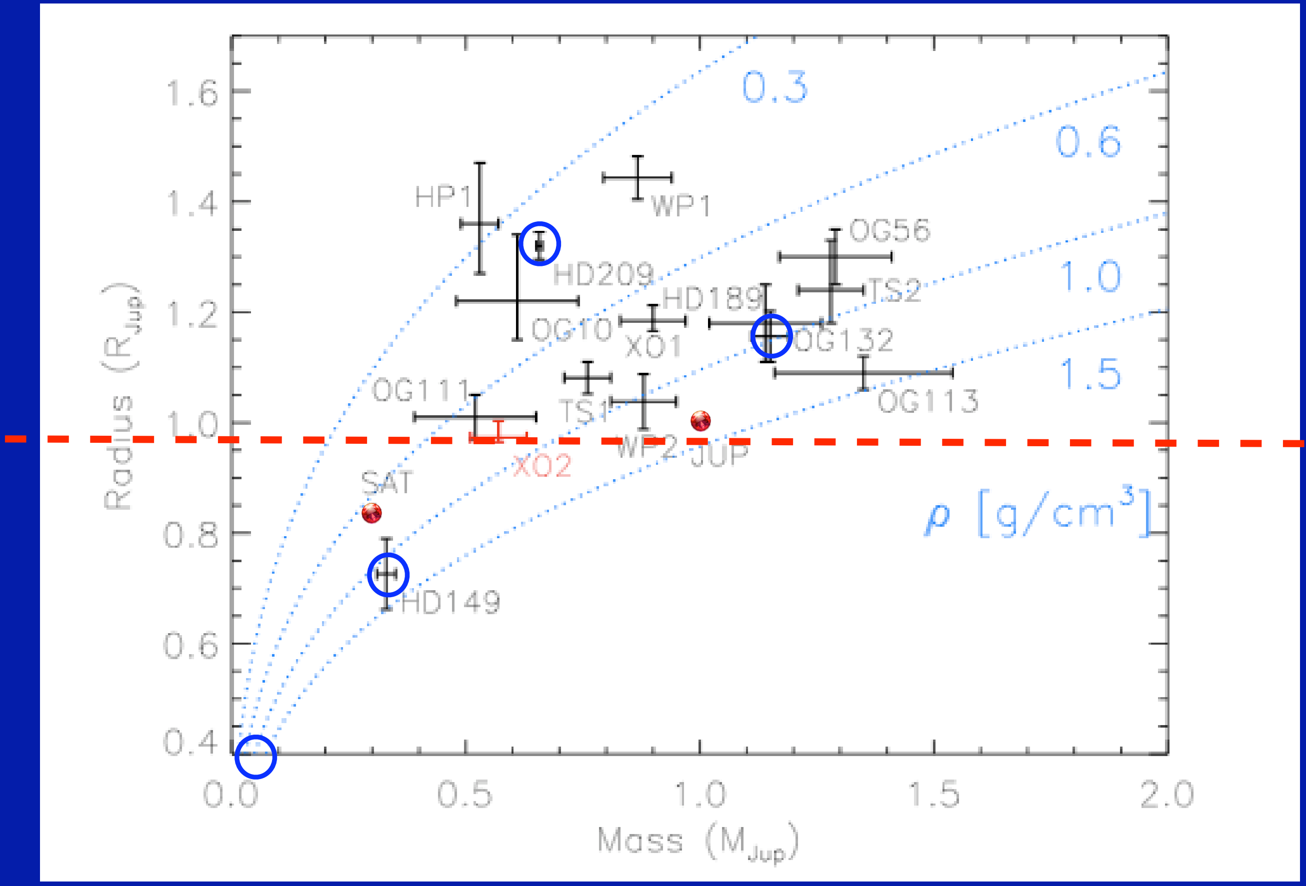
Data Reduction Strategies and Challenges:

White, Pink and Red Noise; Data Trend Filtering & Systematics

Frédéric Pont Geneva Observatory

The mass-radius relation for transiting EGP



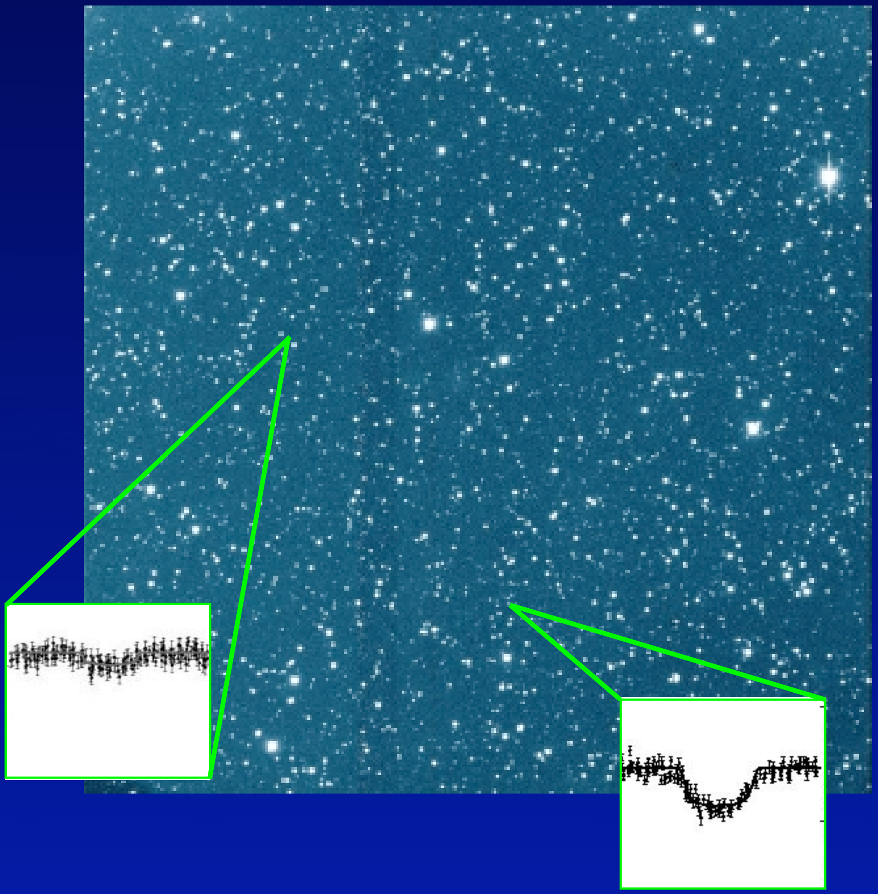
Field transit searches with CCD cameras

The challenge :

Monitor 10^5-10^6 stars

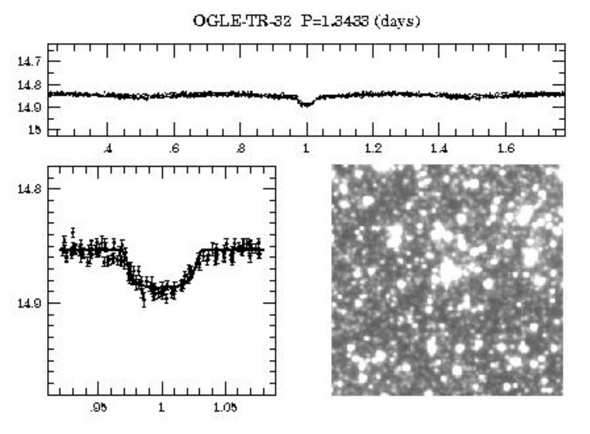
for 10^1 - 10^2 nights

at 10^{-2} - 10^{-3} accuracy

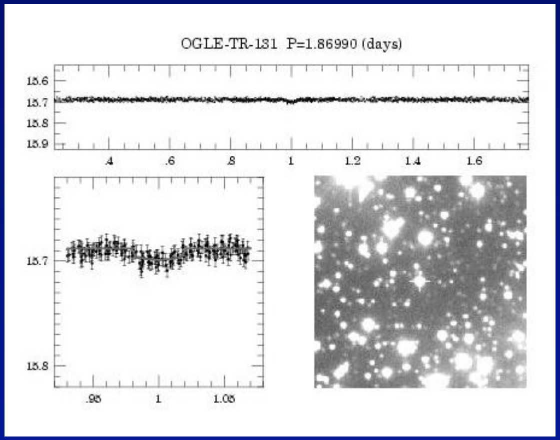


Most transiting planets are near the detection threshold

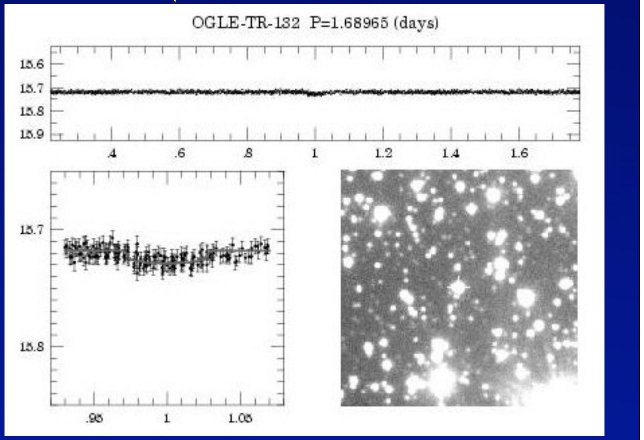
an eclipsing binary



a transiting planet

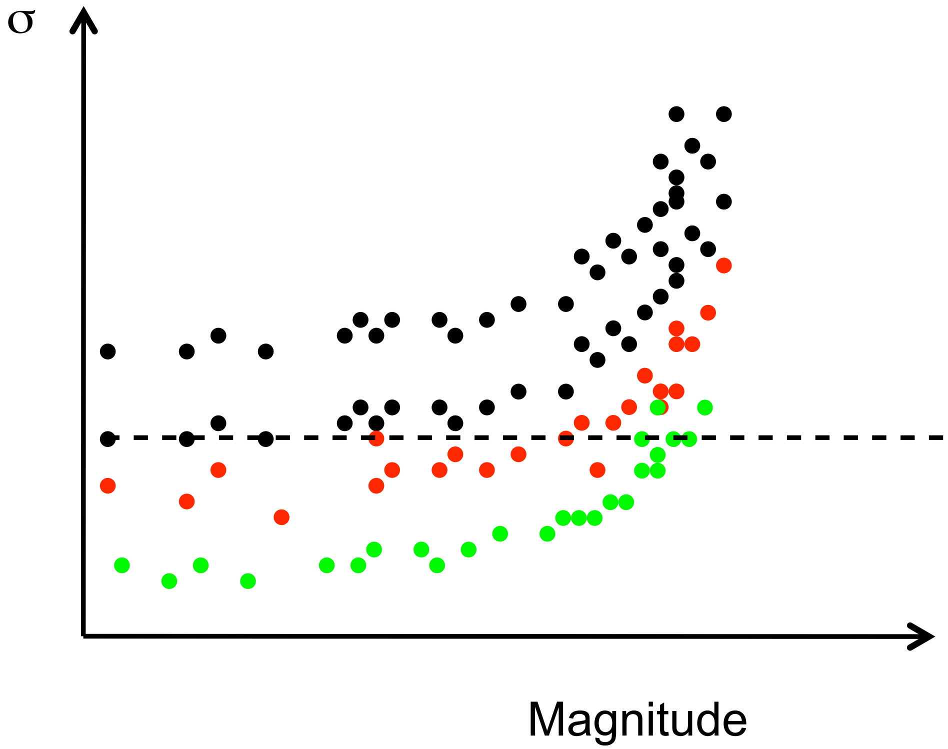


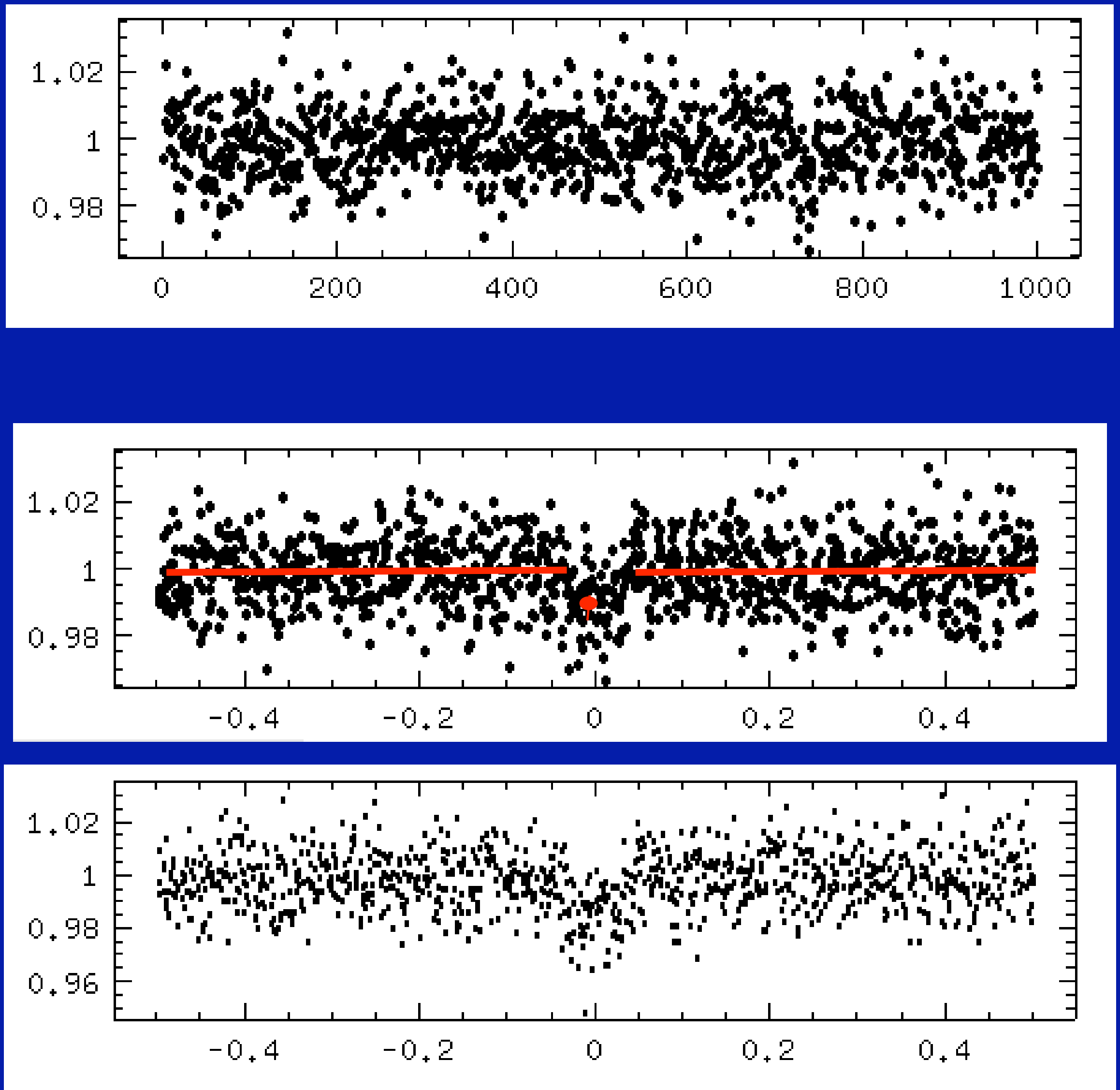
nothing at all!



Systematics / red noise in surveys

EXPLORE/OC	Sara Seager
Monitor	Suzanne Aigrain
SuperWASP	Andrew Cameron, Keith Horne
HAT	Gaspar Bakos
TrES	Tim Brown Hans Deeg
BEST	Heike Rauer, Anders Erikson





$$\mathrm{SNR} = \frac{\text{depth}}{\sigma / \sqrt{n}}$$

Detection threshold:

SNR > 7 - 10

(10^5 - 10^8 stat. tests)

Field transit searchs with CCD cameras

Dream and reality

Orders of magnitude :

Number of pixels in camera 10^7 pixels

Mean distance between targets $>10^1$ pixels

⇒ Number of targets 10^5

Accuracy : better than 10^{-2}

Accuracy over transit duration (2-4 hours): 10^{-2}

⇒ hot Jupiter transits detected at 10 sigmas

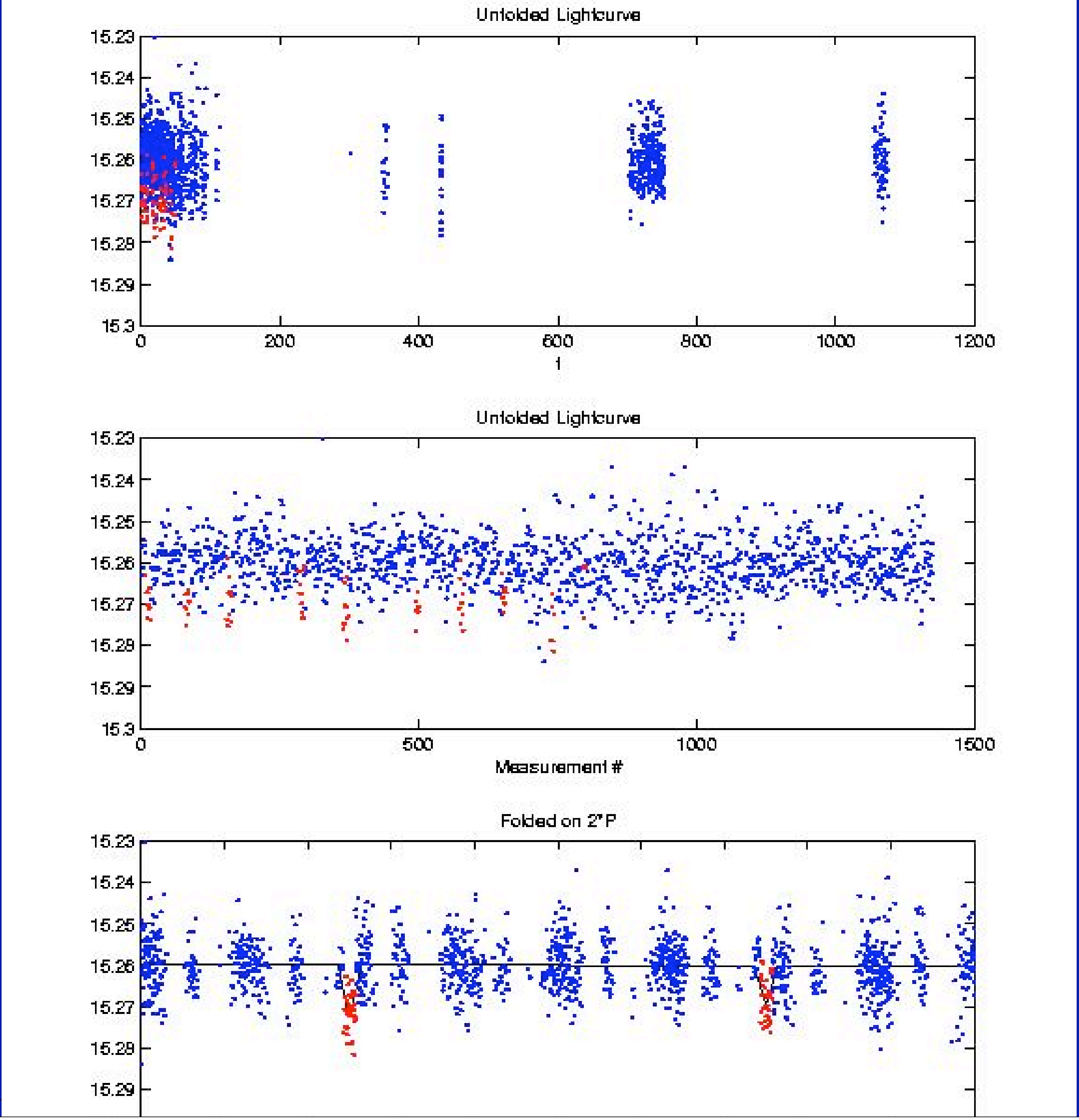
Number of transiting hot Jupiters : $0.1 \times 0.01 \times 10^5 = 10^2$

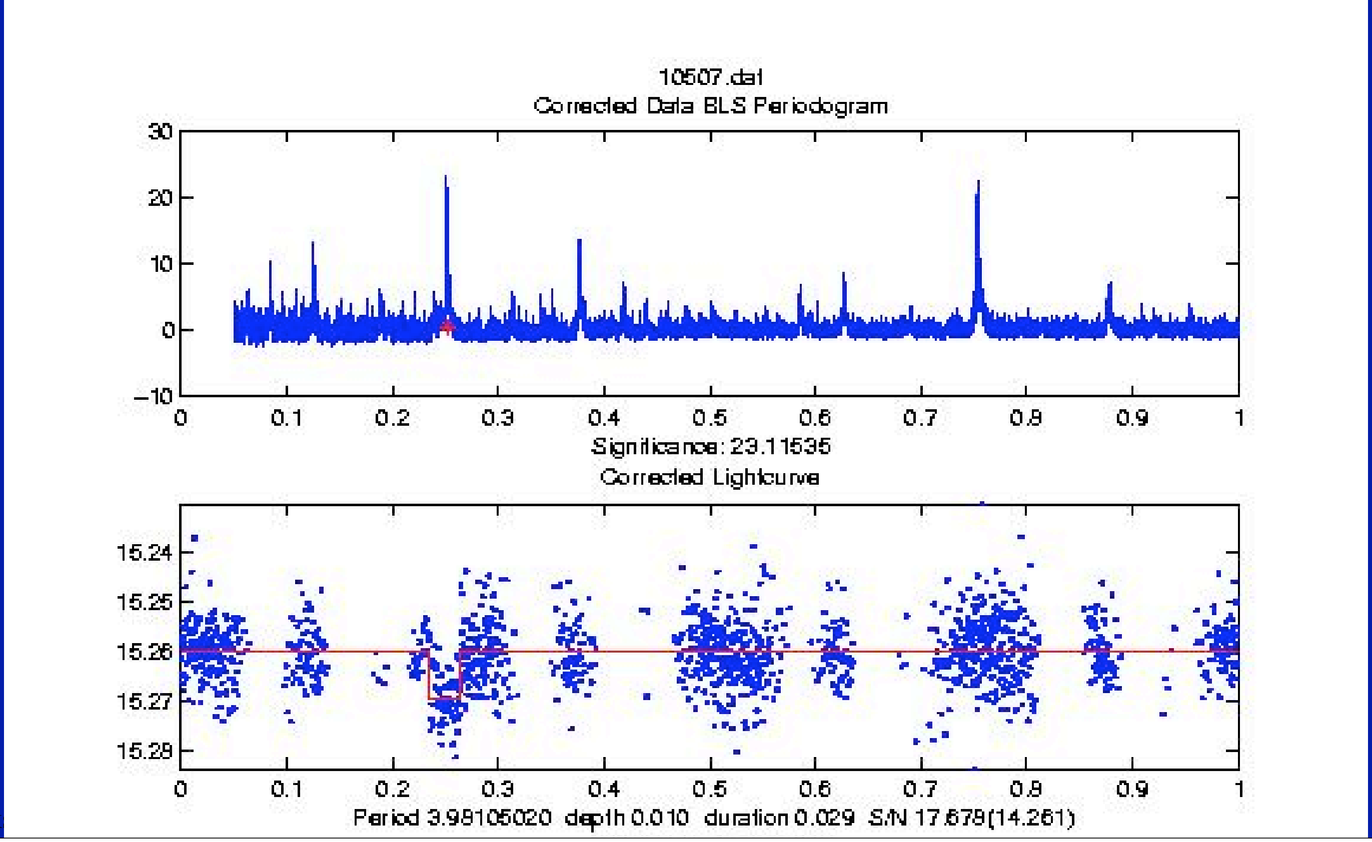
Number of detected transiting planets : dozens (e.g. Horne 2003)

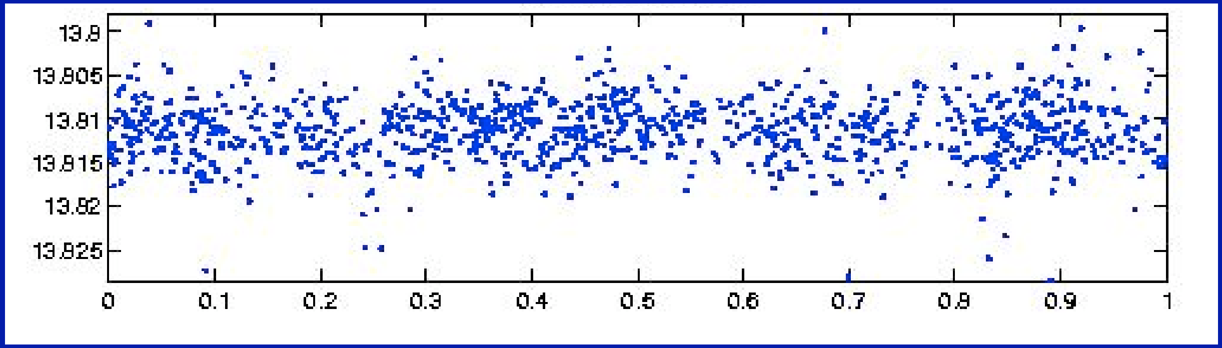
Actual detections :

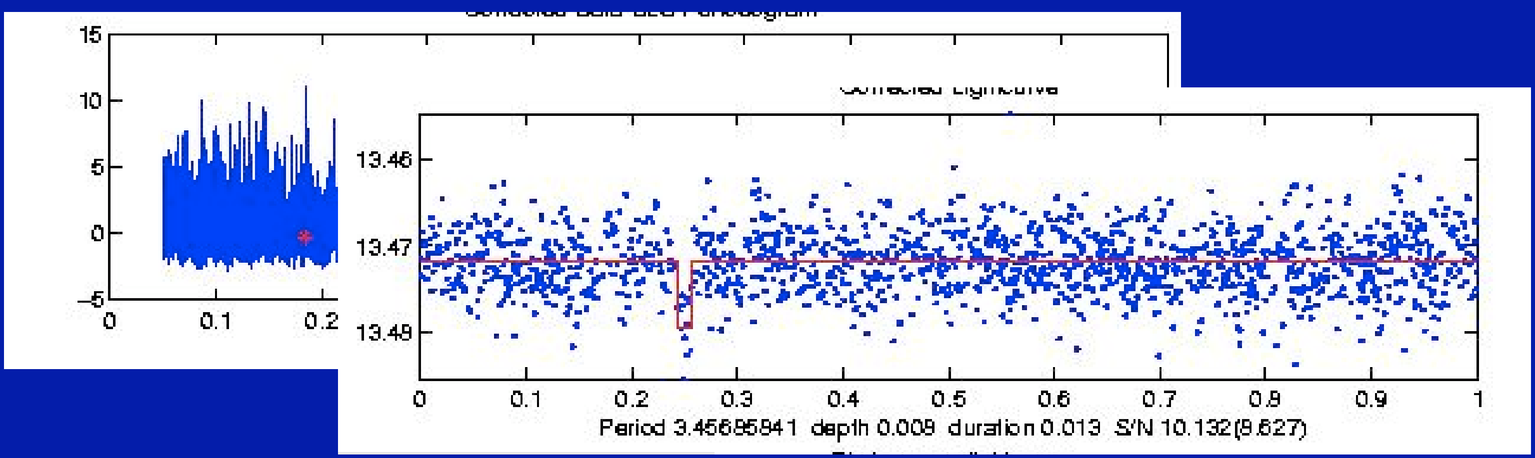
1 -2 transiting hot Jupiter per year per season for major surveys

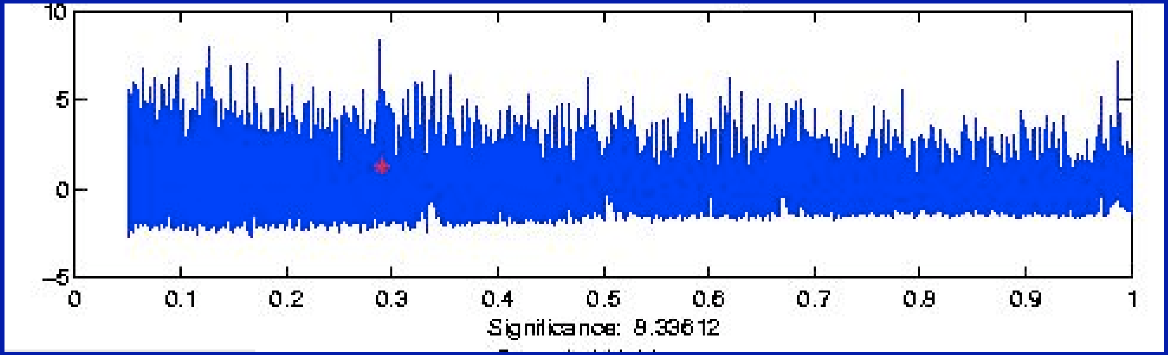
0 for minor surveys

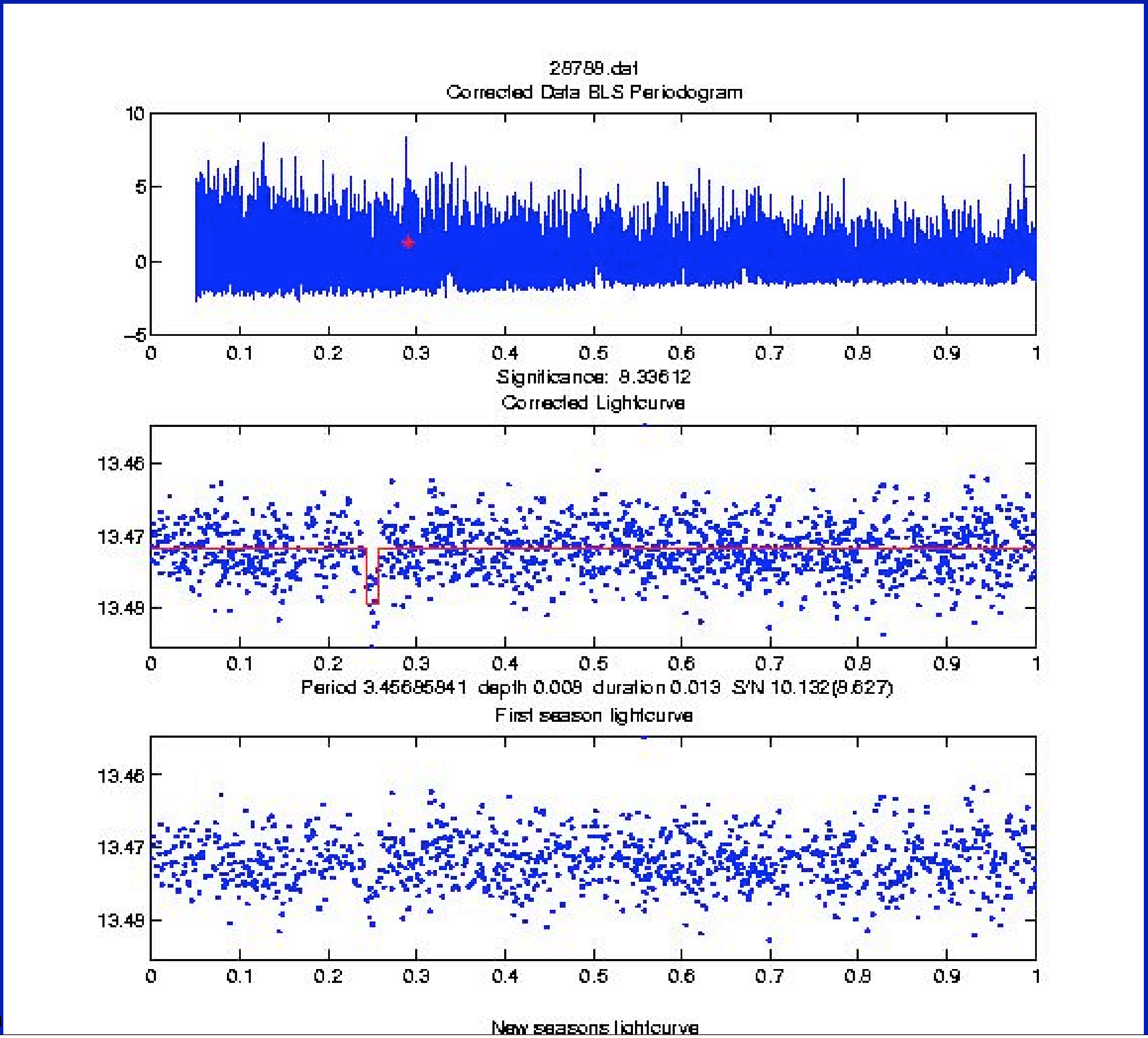

Unfolded Lightcurve
15.23
15.24
15.25
15.26
15.27
15.29
15.29
15.3
0
200
400
600
800
1000
1200
t
Unfolded Lightcurve
15.23
15.24
15.25
15.26
15.27
15.29
15.29
15.3
0
500
1000
1500
Measurement #
Folded on 2*P
15.23
15.24
15.25
15.26
15.27
15.28
15.29


10507.dat
Corrected Data BLS Periodogram
30
20
10
0
−10
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Significance: 23.11535
Corrected Lightcurve
15.24
15.25
15.26
15.27
15.28
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Period 3.98105020 depth 0.010 duration 0.029 S/N 17.678(14.261)


13.8
13.805
13.81
13.815
13.82
13.825
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


15
10
5
0
−5
0
0.1
0.2
13.46
13.47
13.48
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Period 3.45685841 depth 0.009 duration 0.013 S/N 10.132(8.627)


10
5
0
−5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Significance: 8.39612


28789.dat
Corrected Data BLS Periodogram
10
5
0
−5
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Significance: 8.33612
Corrected Lightcurve
13.46
13.47
13.48
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Period 3.45685841 depth 0.009 duration 0.013 S/N 10.132(8.627)
First season lightcurve
13.46
13.47
13.48
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
New seasons lightcurve

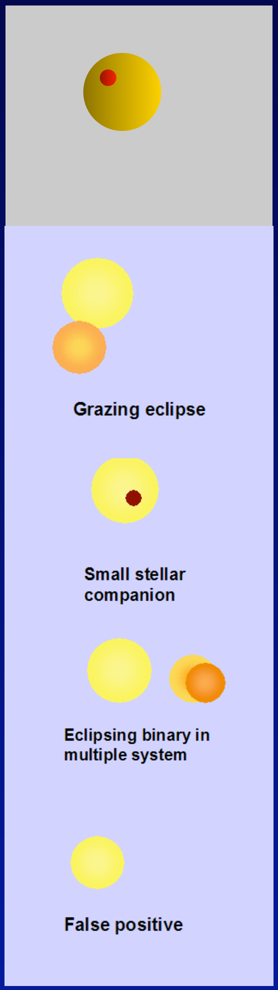
Nature of OGLE planetary transit candidates

Transiting gas giant planets

Eclipsing binaries

- grazing
- low-mass companion
- multiple systems and blends

False positives of the transit detection



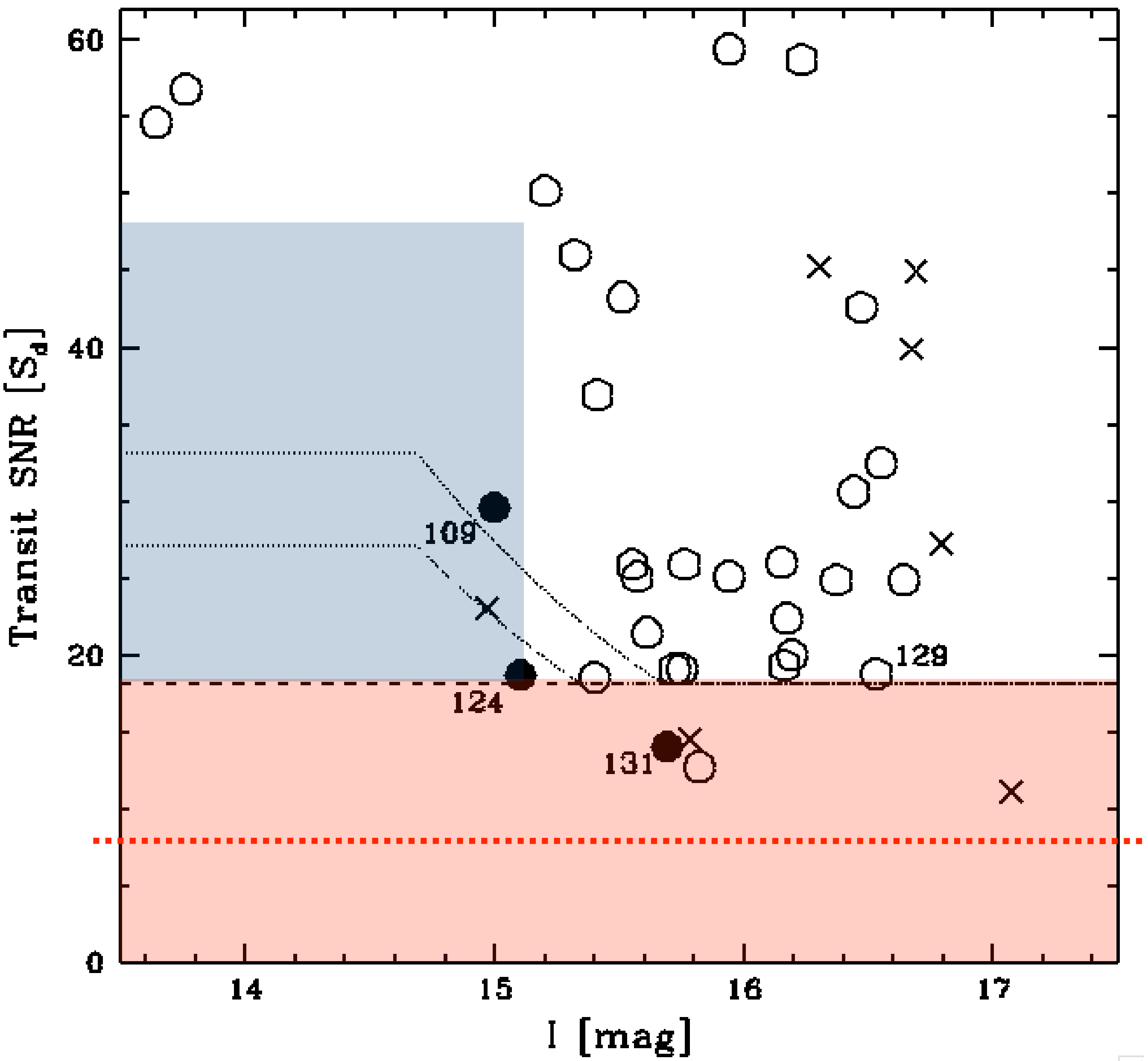
[5]

[>20]

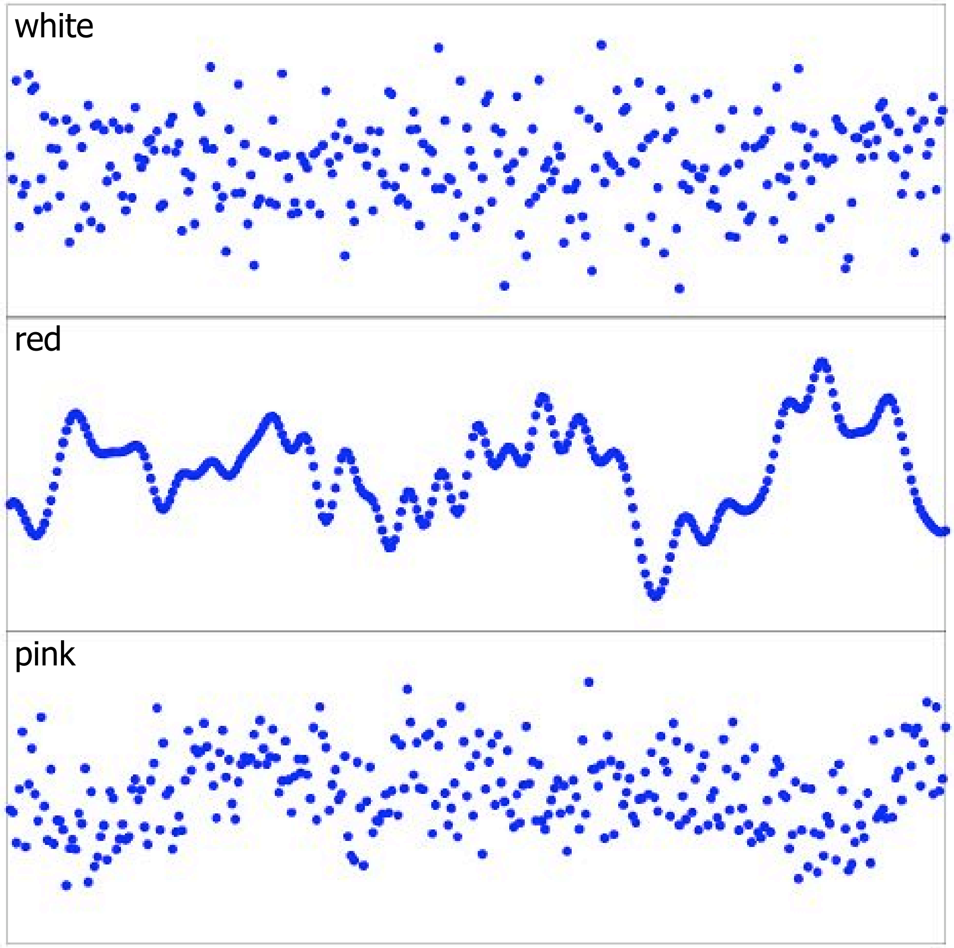
[>100]

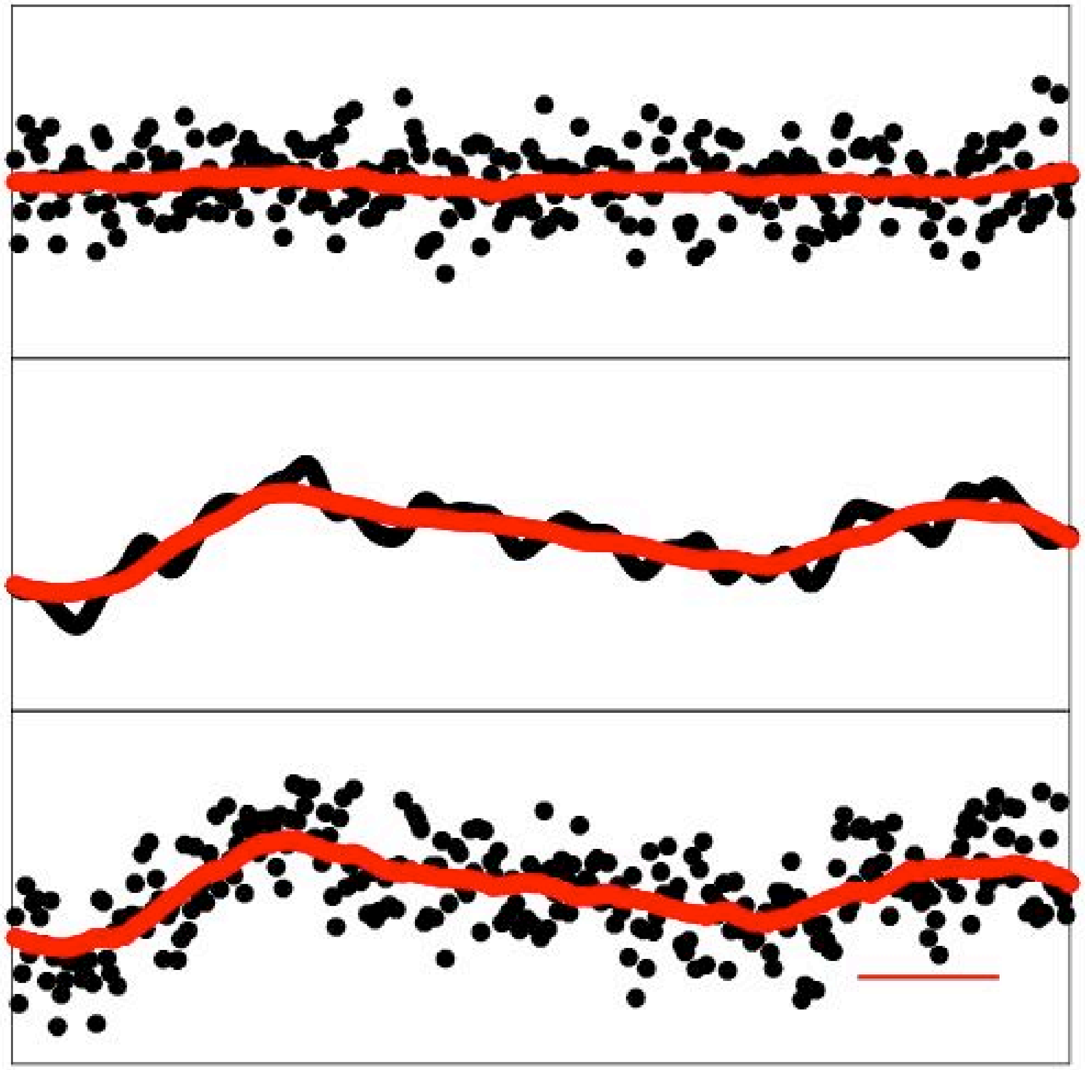
[>50]

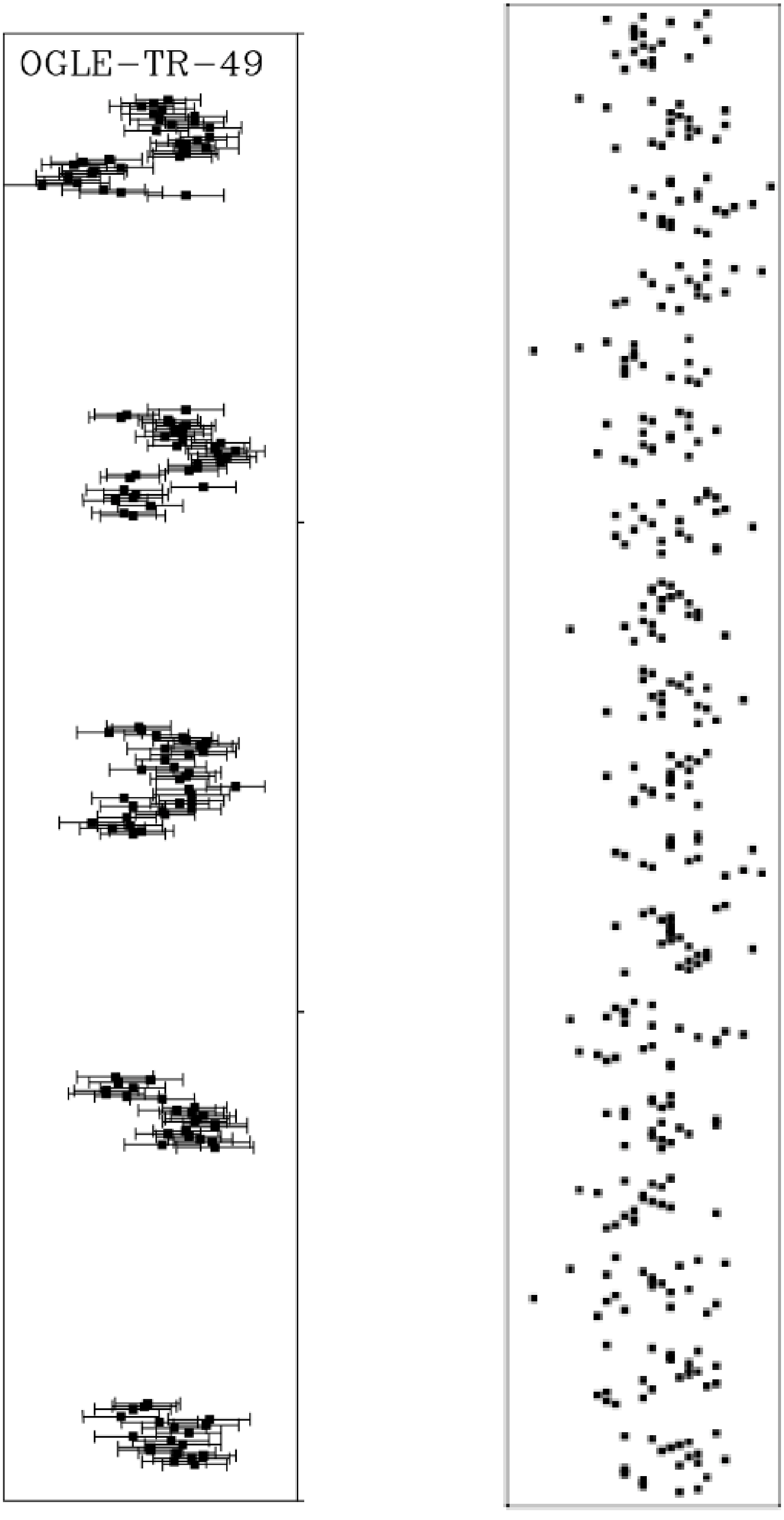
[10-20]


Transit SNR [S_d]
I [mag]
60
40
20
0
14
15
16
17
109
124
131
129

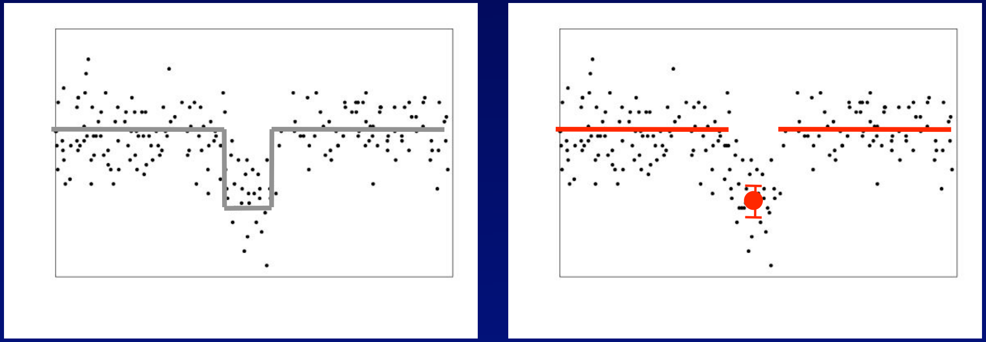
Colours of noise






OGLE−TR−49

Significance of transit detections



Transit detection signal-to-noise:

$$\mathrm{SNR} = \frac{\mathrm{depth}}{\sigma / \sqrt{n}}$$

Detection threshold:

SNR > 7 - 10

(10^5 - 10^8 stat. tests)

For a normal Hot Jupiter :
σ ~ 5 mmag, depth ~ 1%, n ~ 3 transits x 10 points

⇒ ***SNR ~ 10*** ***most HJ should be detectable***

Transit detection significance with **real** photometric noise

$$\mathrm{SNR}^2 = \frac{\mathrm{depth}^2}{\sigma^2 / n \quad + 1/ n^2 \Sigma\, \mathit{cov}(x_i,x_i)}$$

In the regime relevant to transit surveys, the red term dominates!

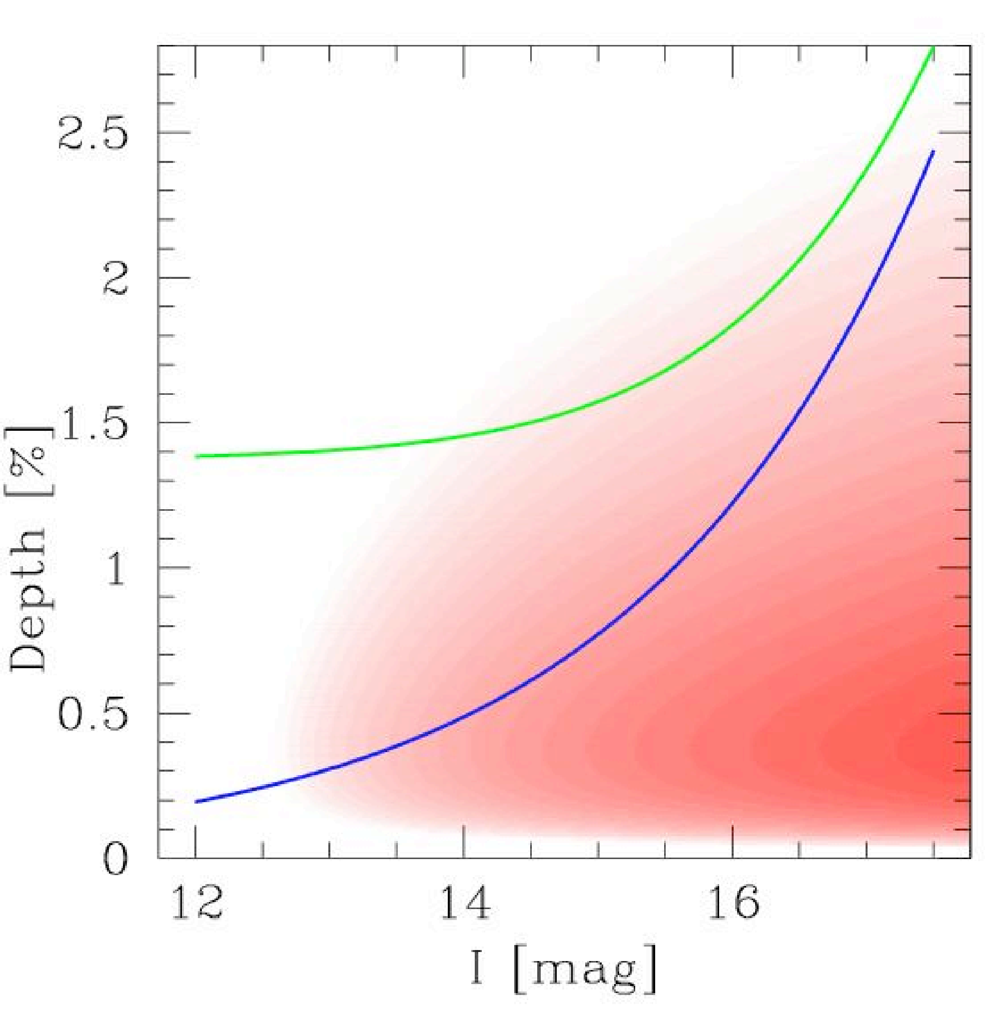
white noise (photon+sky+scintillation)

σ = 3-10 mmag, n=20-50, σ^2 / n = 0.1 - 0.5 mmag ^{2}

red noise (systematics from seeing, weather, tracking)

$1/ n^2 \Sigma\, \mathit{cov}(x_i,x_i) = (2\text{-}5\ \mathrm{mmag})^2 = 4 - 25\ \mathrm{mmag}^2$

Detection threshold with red noise (systematics)



- Factor 3-5 in threshold!
- Weaker dependence on magnitude
- Steeper dependence on period

Detection statistic with red noise

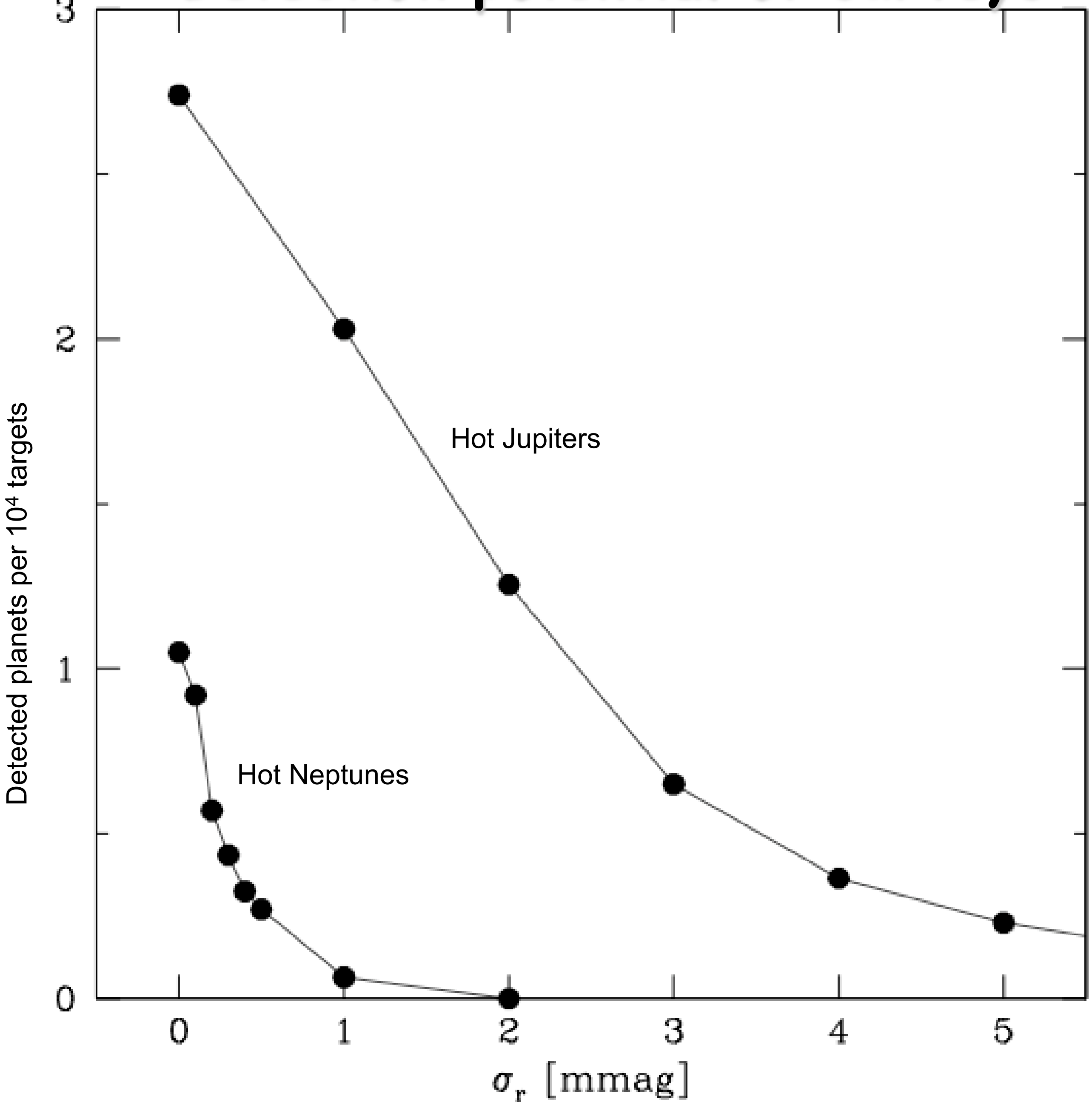
$$S_r \equiv \frac{d}{\sigma_d} = \frac{d}{\sqrt{\frac{\sigma_0^2}{n} + \frac{1}{n^2}\sum_{i\neq j} C_{ij}}} \; .$$

$$S_r^2 = d^2 \frac{n^2}{\sum_{k=1}^{N_{tr}} n_k^2 \mathcal{V}(n_k)}$$

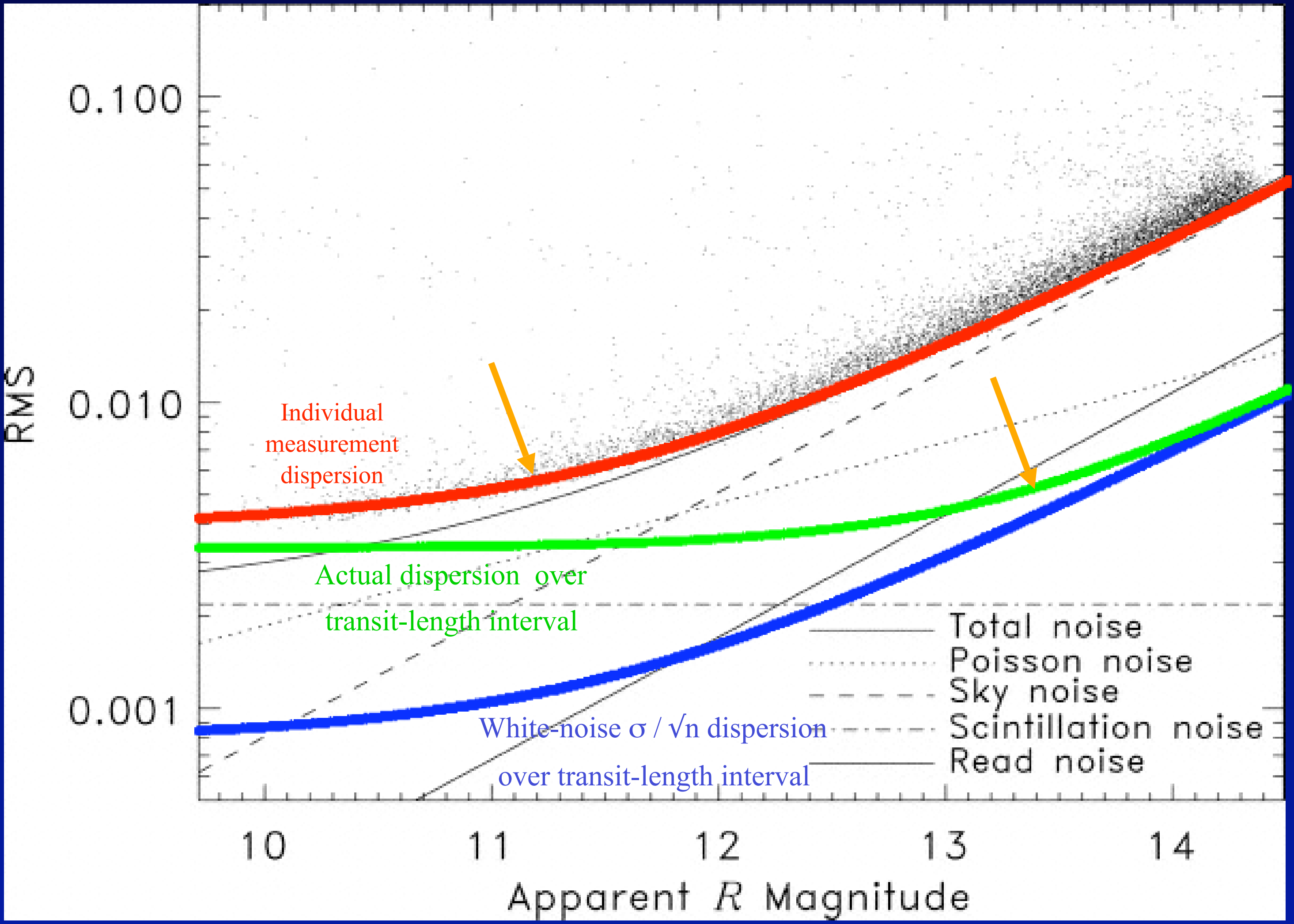
$$S_r^2 = \frac{d^2 n^2}{\sum_{k=1}^{N_{tr}} n_k^2 \left(\frac{\sigma_w^2}{n_k} + \sigma_r^2\right)} \; .$$

(Pont, Zucker & Queloz 2006)

Detection potential of surveys



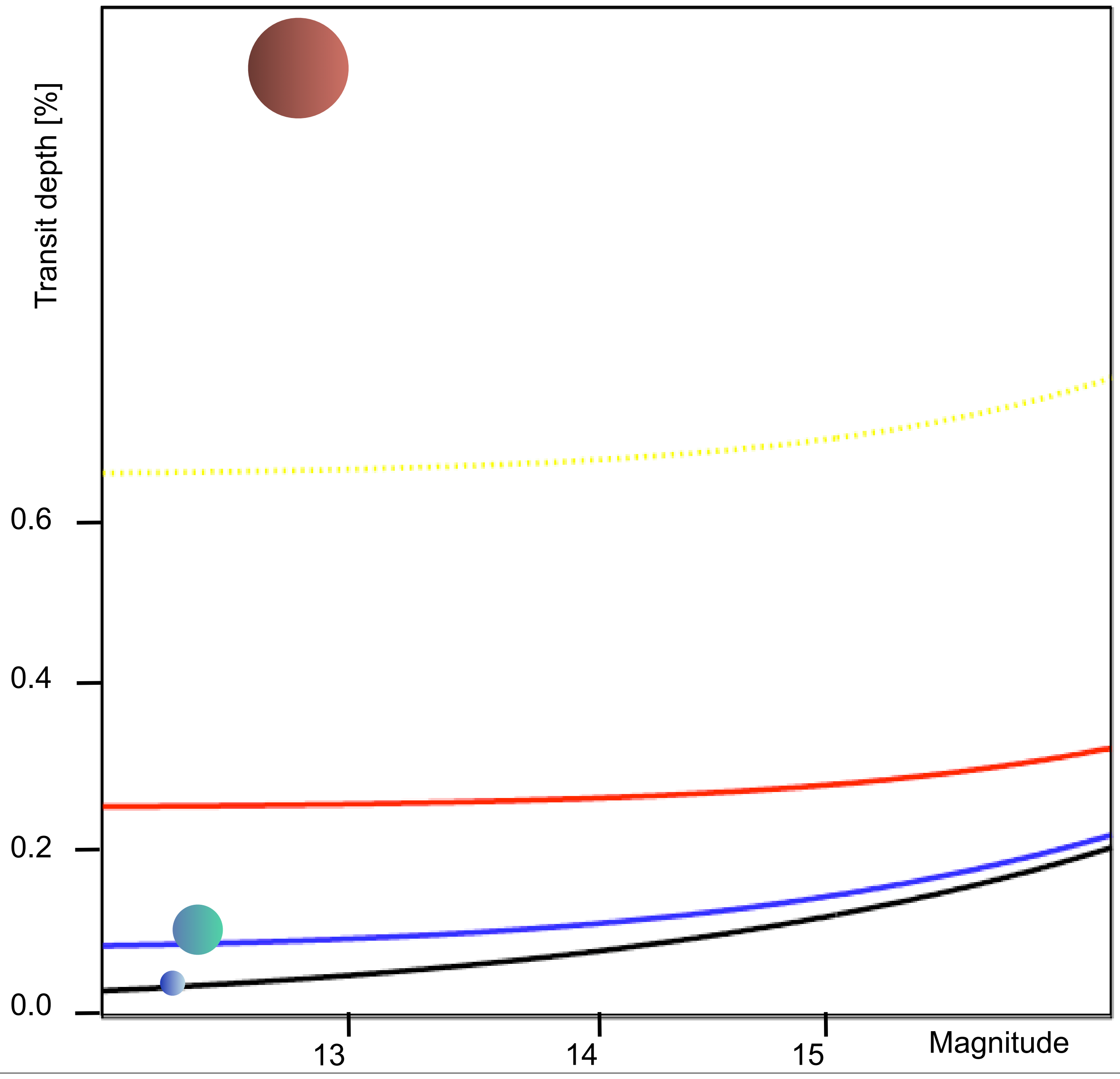
Noise and detection in ground-based transit surveys



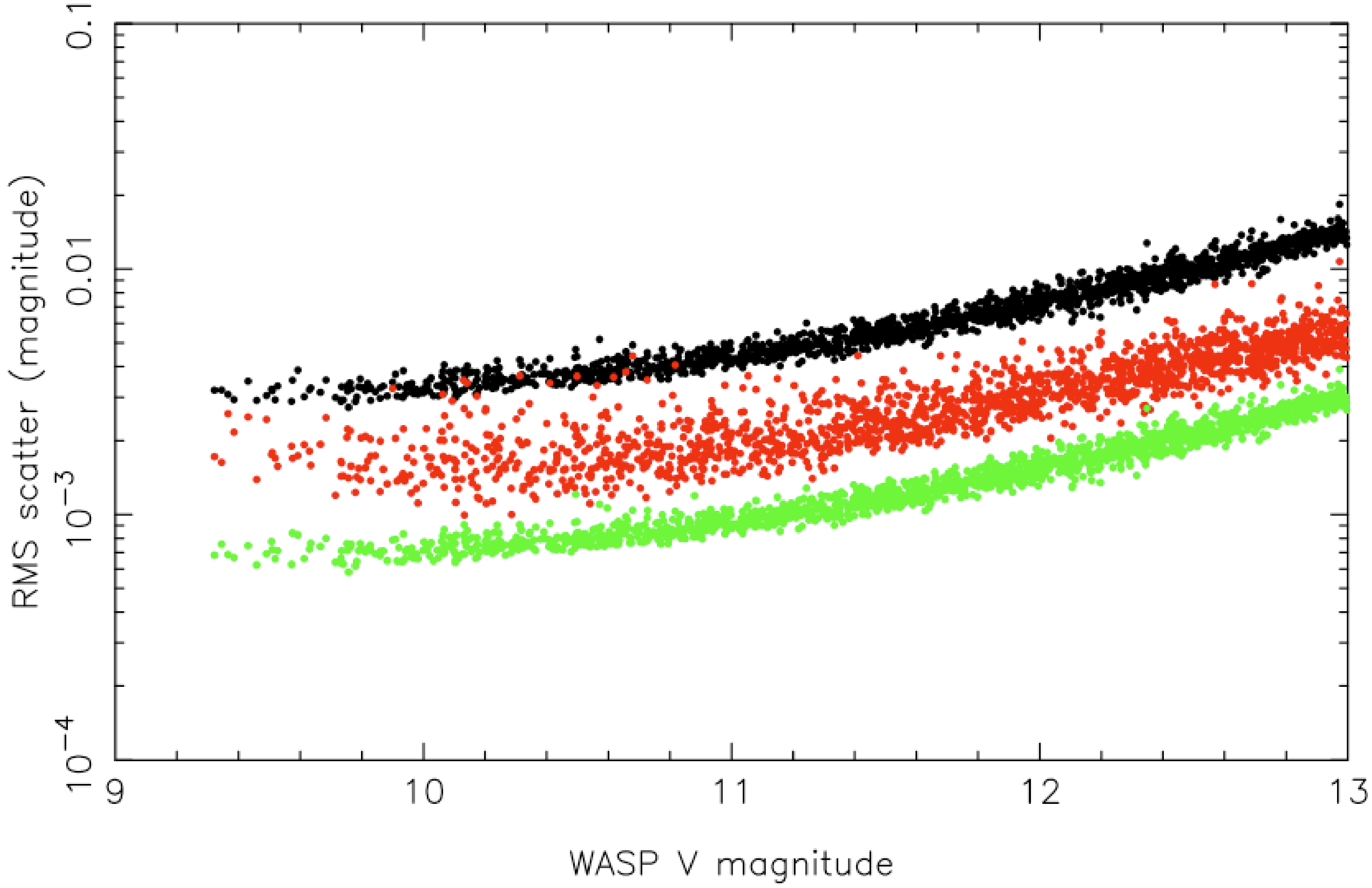
Individual measurement dispersion

Actual dispersion over transit-length interval

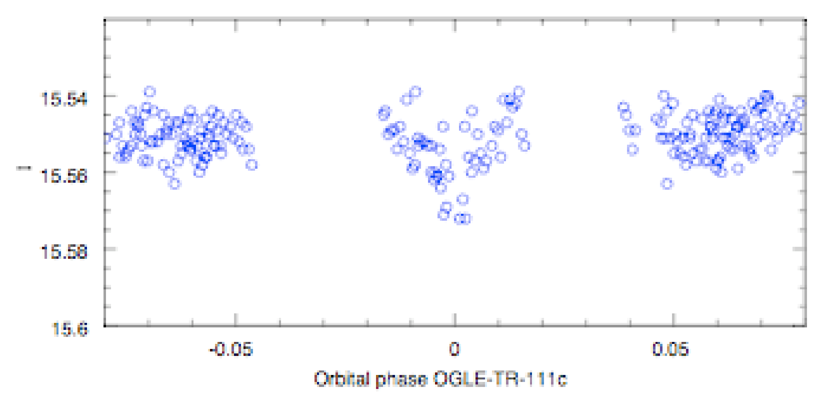
White-noise σ / √n dispersion over transit-length interval

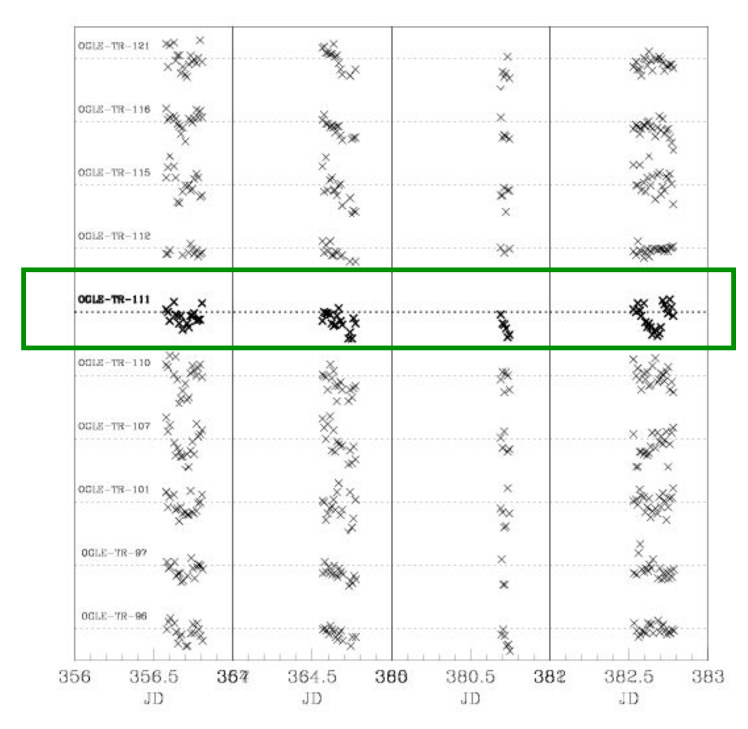

Transit depth [%]
0.6
0.4
0.2
0.0
13
14
15
Magnitude

Example: Red noise in SWASP data after detrending algorithm

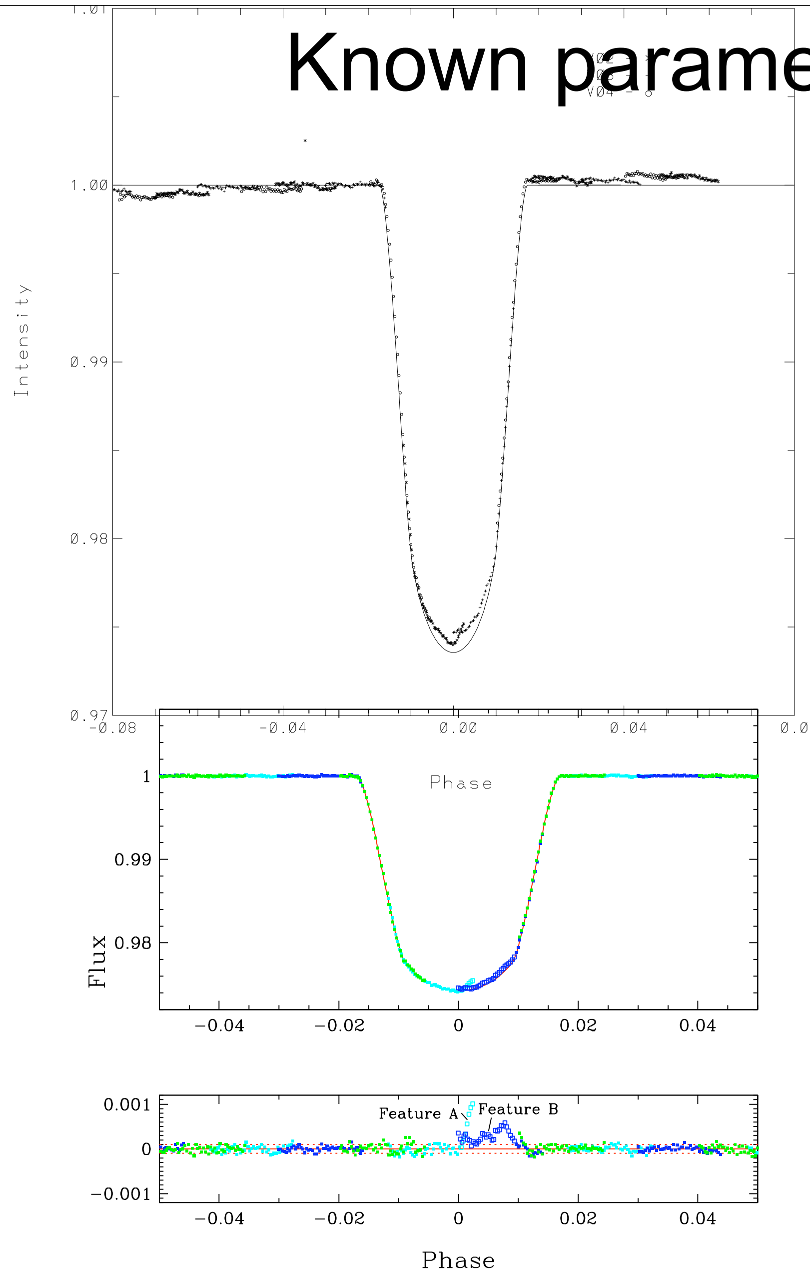


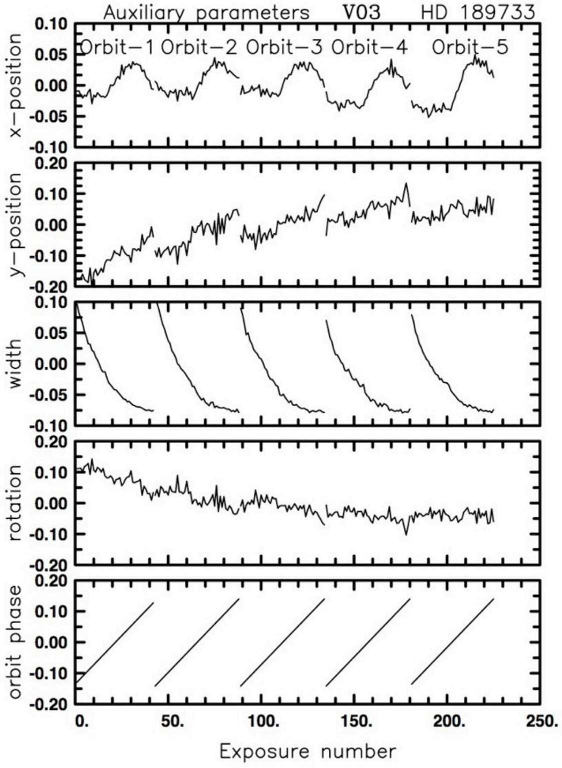
Second transiting planet around OGLE-TR-111 ?





Known parameters decorrelation





Example from HST data on transiting HD189733b

Systematics Removal
in large sets of light-curves

SysRem

(Tamuz, Mazeh & Zucker 2005)

Trend Filtering Algorithm

(Kovacs, Bakos & Noyes 2005)

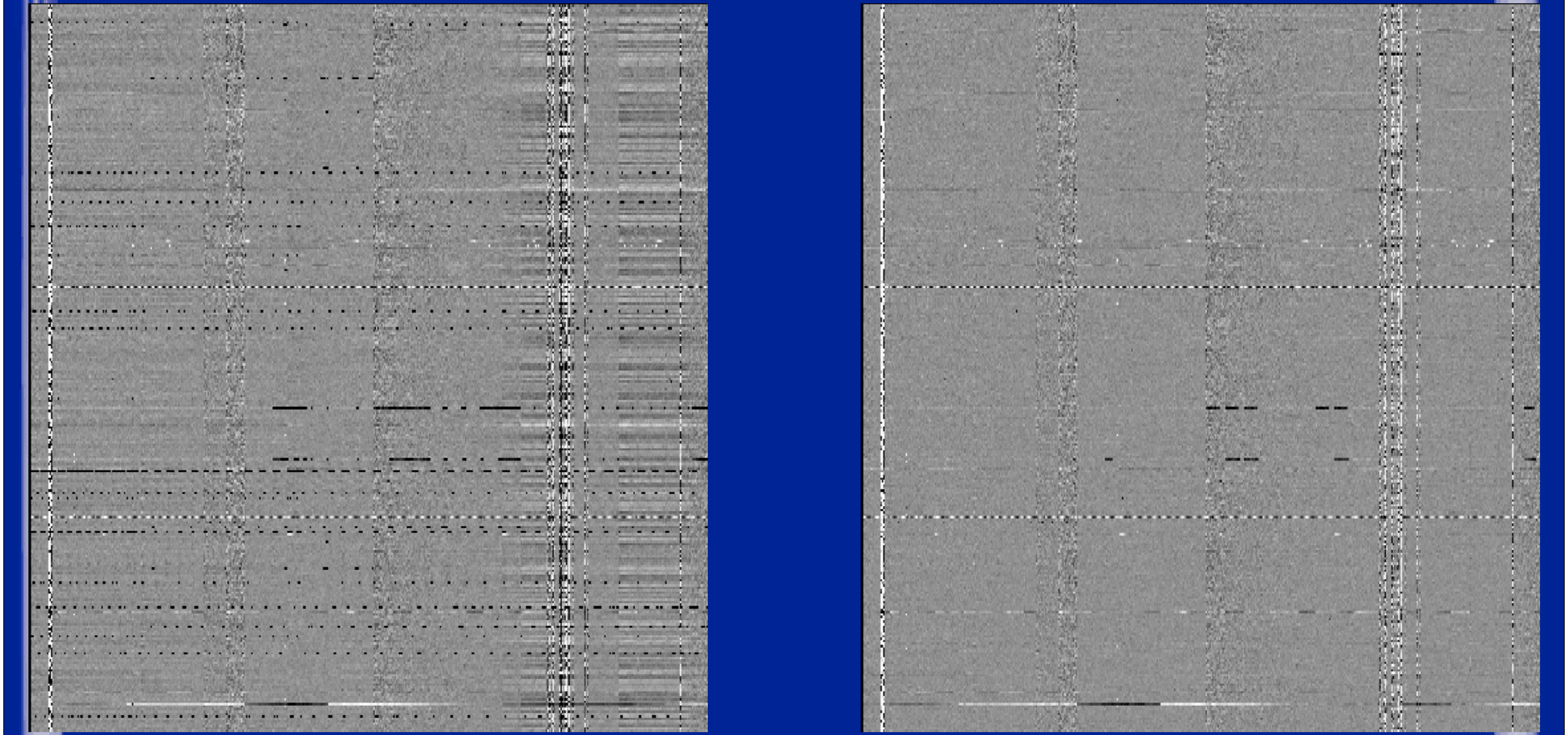
What kind of 'systematics'?

- Colour-dependent atmospheric extinction
  - c_i – colour, a_j – airmass
- Contaminating light (moon, earth)
- Position-dependent CCD response
- etc…

Huge photometric datasets:

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & \cdots & \cdots \\ r_{21} & r_{22} & \cdots & \cdots & \cdots \\ r_{31} & & & & \\ r_{41} & & & & \\ \cdots & & & & \\ \cdots & & & & \\ \cdots & & & & \\ \cdots & & & & \end{pmatrix}$$

Example: SuperWASP 16h30+28 field
300 stars, 2549 observations spanning 100 days



Before

After

r_i = transit (?)

+ white noise

+systematics

ri(j) = transit (?)
+ white noise
+ c_i^1 airmass (j)
+ c_i^2 temperature(j)
+ c_i^3 ???

Finding the $\mathbf{c}_i$ and $\mathbf{a}_i$?

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & \cdots & \cdots \\ r_{21} & r_{22} & \cdots & \cdots & \cdots \\ r_{31} & & & & \\ r_{41} & & & & \\ \cdots & & & & \\ \cdots & & & & \\ \cdots & & & & \\ \cdots & & & & \end{pmatrix}$$

$$\boxed{S^2 = \sum_{i,j} \left(\frac{r_{ij}}{\sigma_{ij}} \right)^2}$$

Find c_i and a_j
that minimize:

$$\boxed{S^2 = \sum_{i,j} \frac{(r_{ij} - c_i a_j)^2}{\sigma_{ij}^2}}$$

Star no. i:

$$S_i^2 = \sum_j \frac{(r_{ij} - c_i a_j)^2}{\sigma_{ij}^2}$$

Assume a_j are known, solve for c_i

$$c_i = \frac{\sum_j r_{ij} a_j / \sigma_{ij}^2}{\sum_j a_j^2 / \sigma_{ij}^2}$$

Image no. j:

$$S_j^2 = \sum_i \frac{(r_{ij} - c_i a_j)^2}{\sigma_{ij}^2}$$

Now c_i are known, solve for a_j

$$a_j = \frac{\sum_i r_{ij} c_i / \sigma_{ij}^2}{\sum_i c_i^2 / \sigma_{ij}^2}$$

Applying Sys-Rem to OGLE data

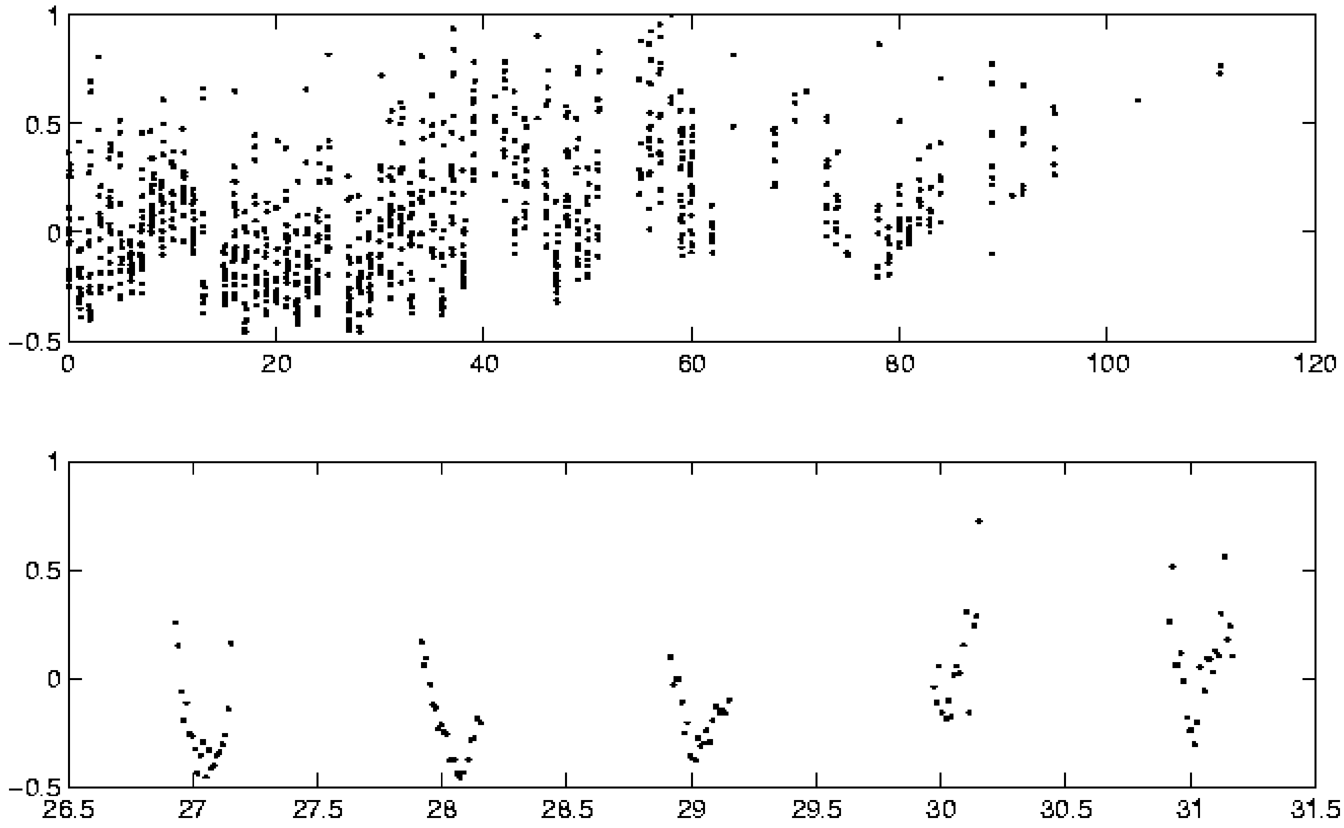


Figure 1. First pattern showing "parabolic" change every observing night.

Applying Sys-Rem to OGLE data

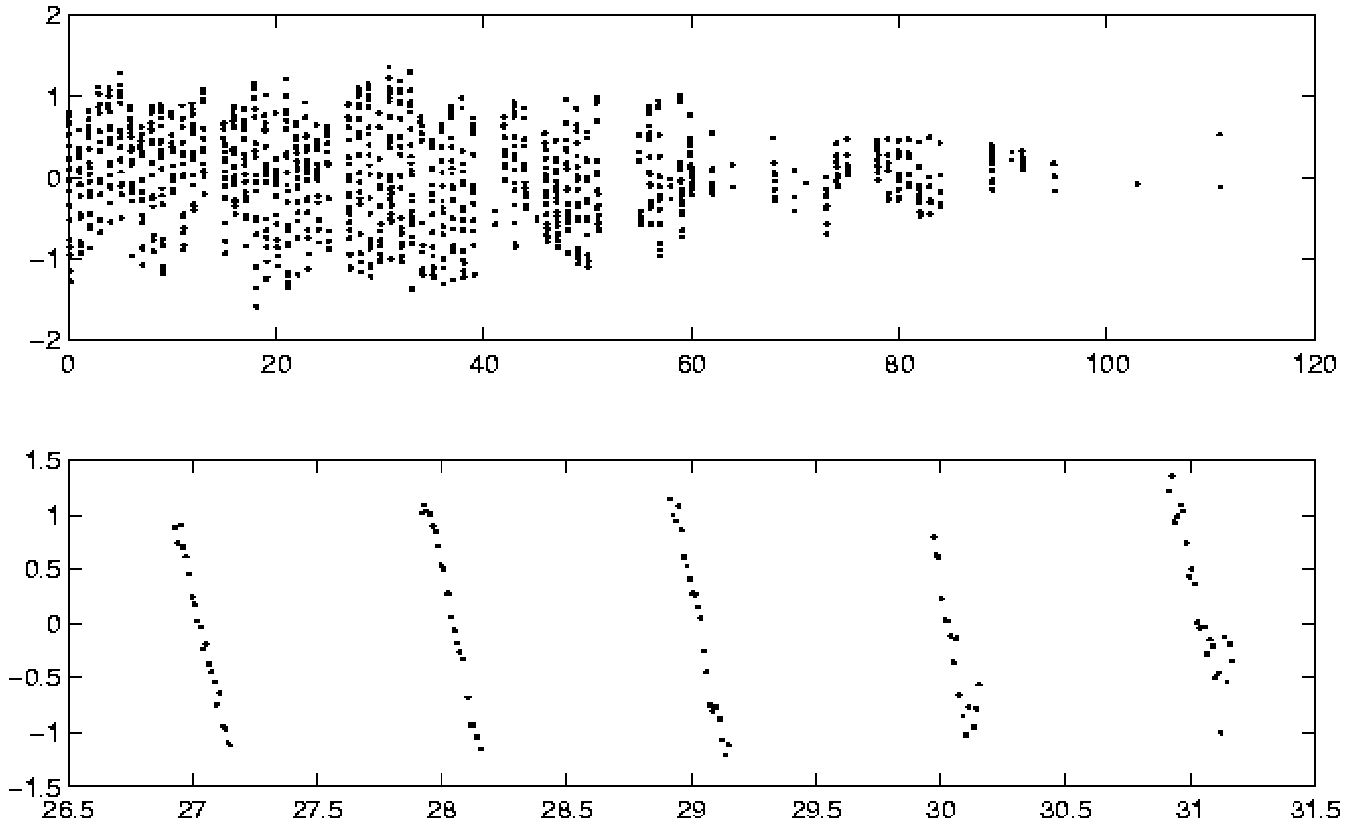


Figure 2. Second component showing linear monotonous change every observation night.

Applying Sys-Rem to OGLE data

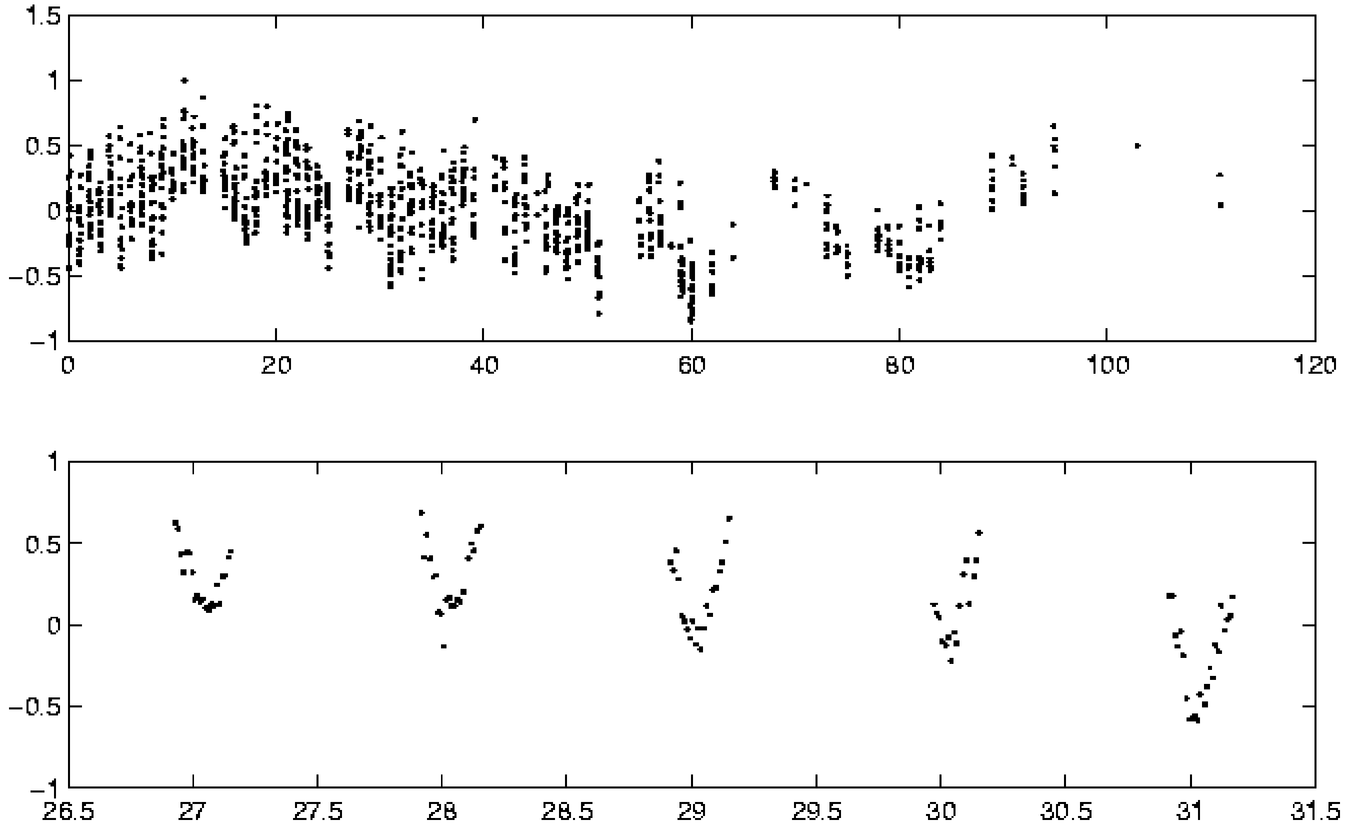


Figure 3. Third component showing similar behaviour to the first.

Applying Sys-Rem to OGLE data

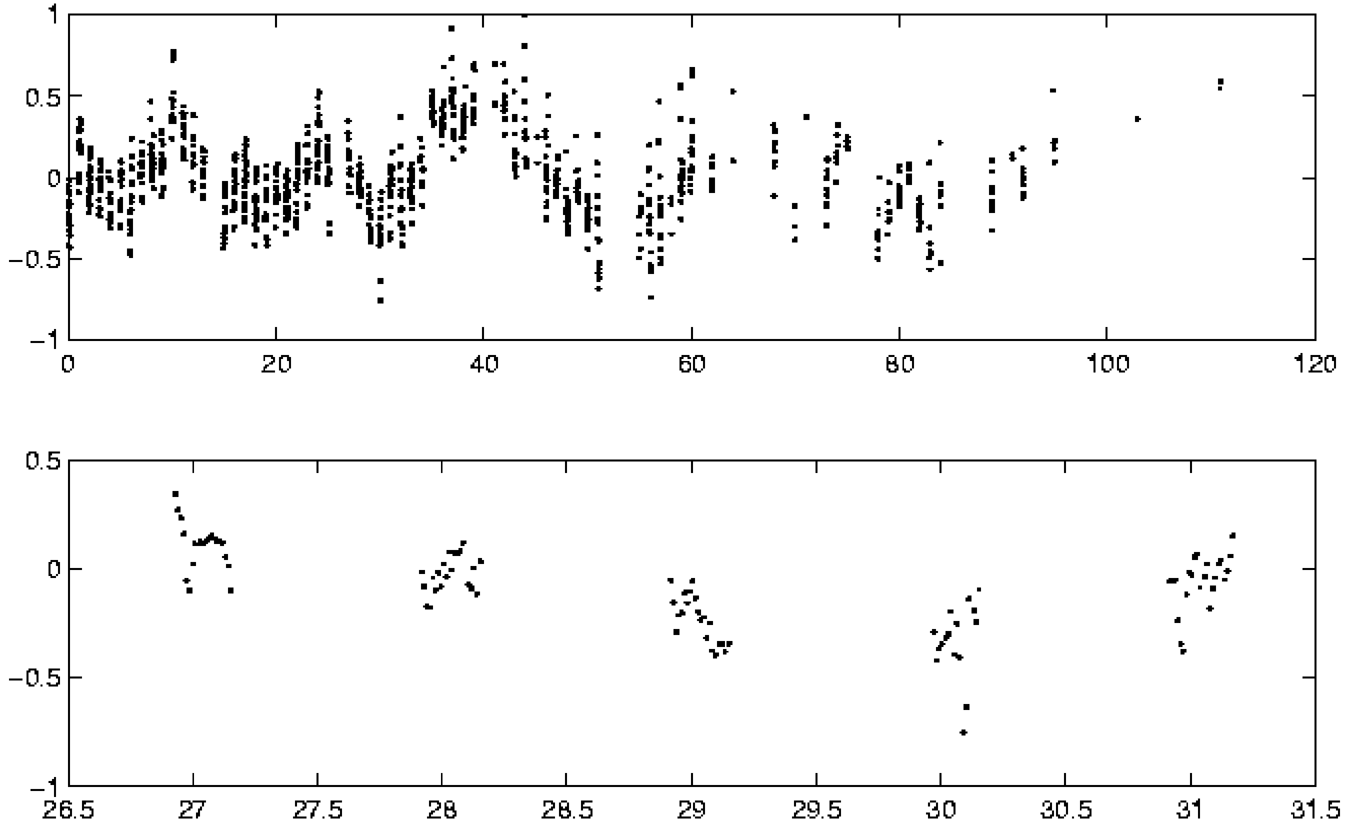
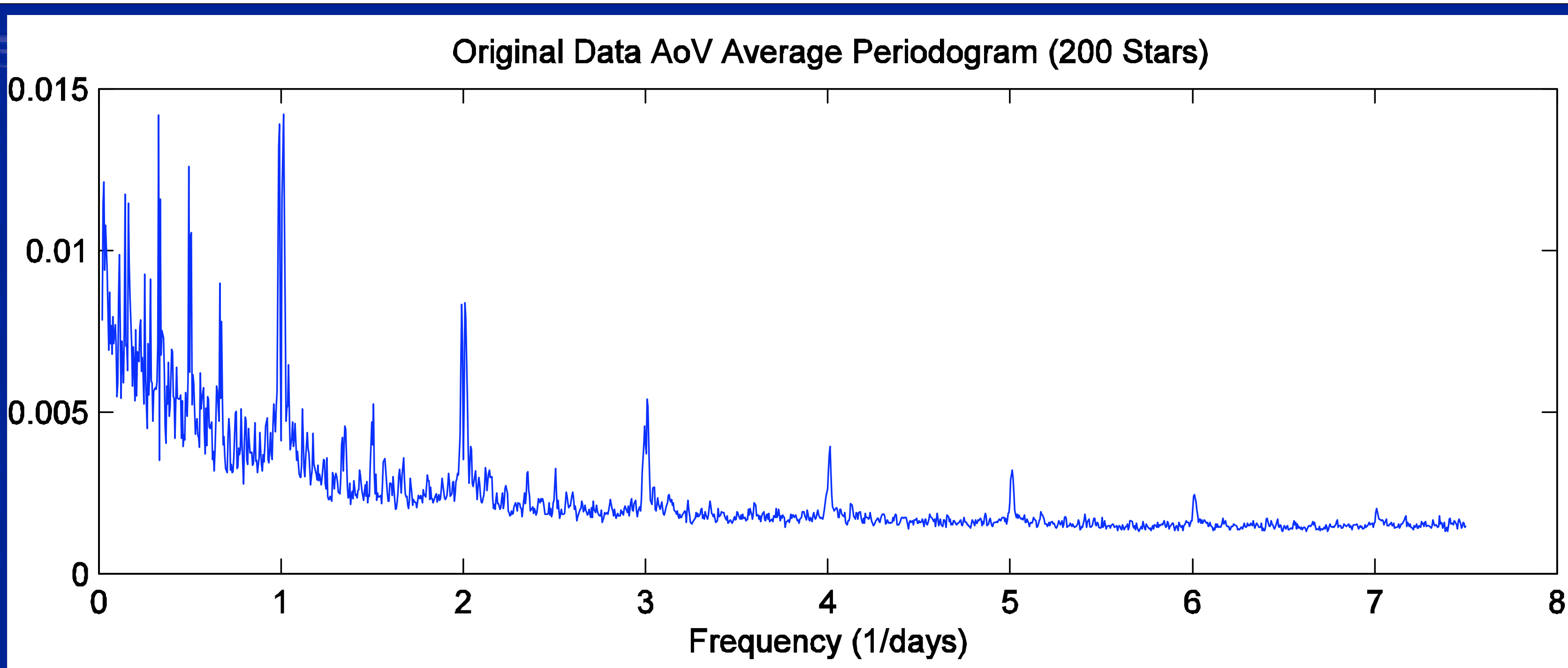
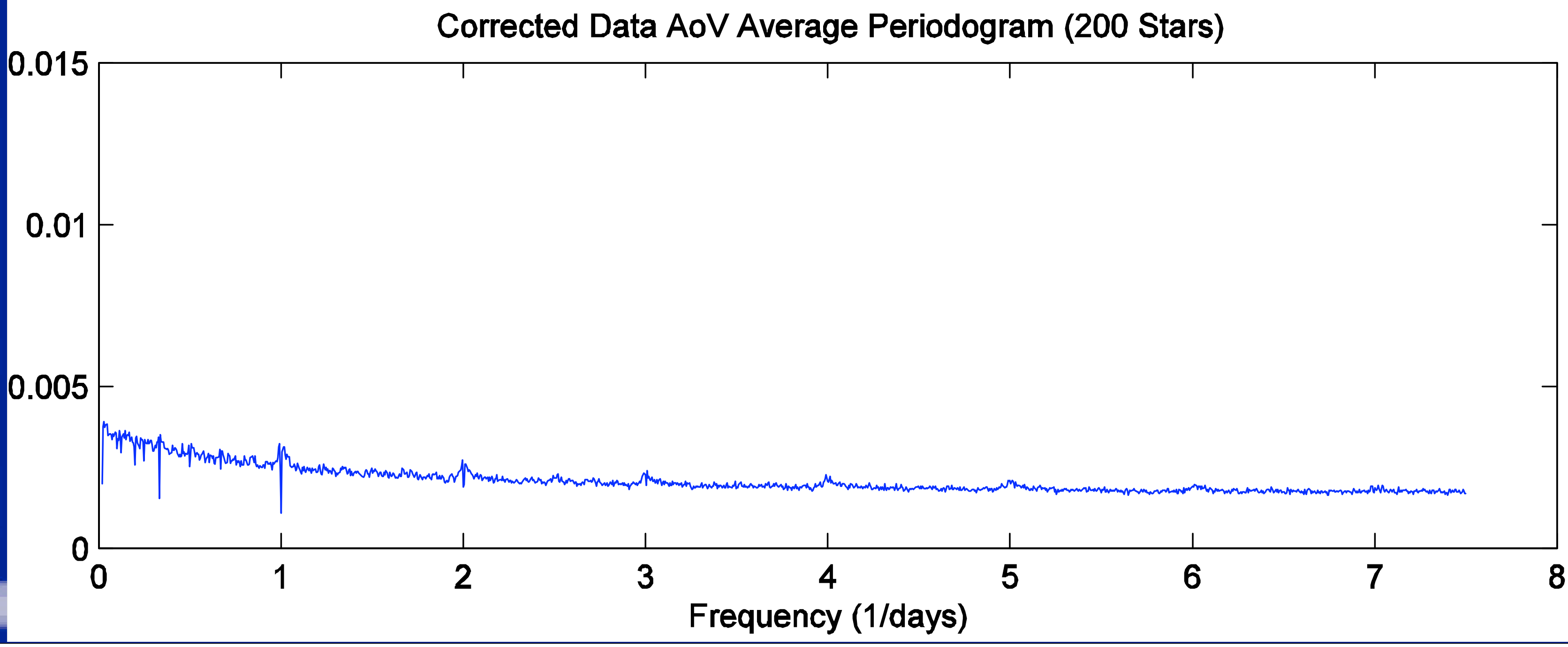


Figure 4. Sixth component showing considerable long term variation.

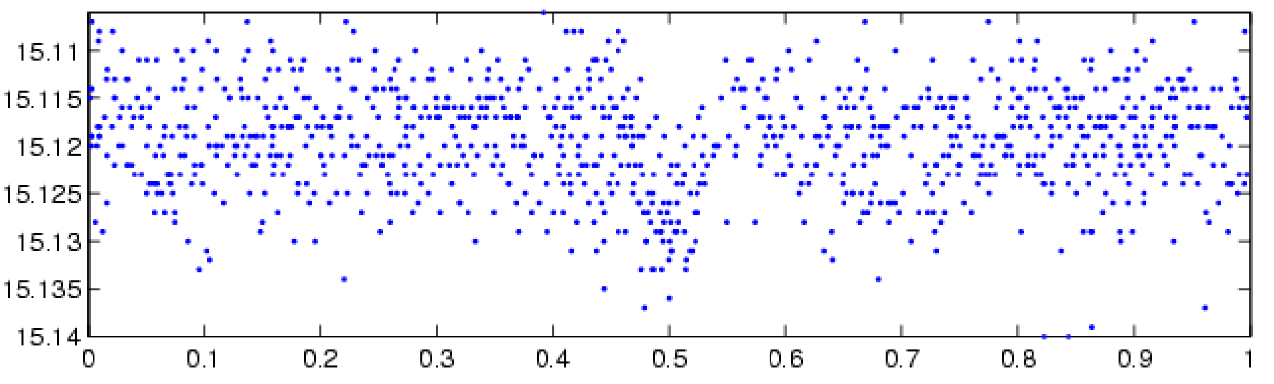


Original Data AoV Average Periodogram (200 Stars)

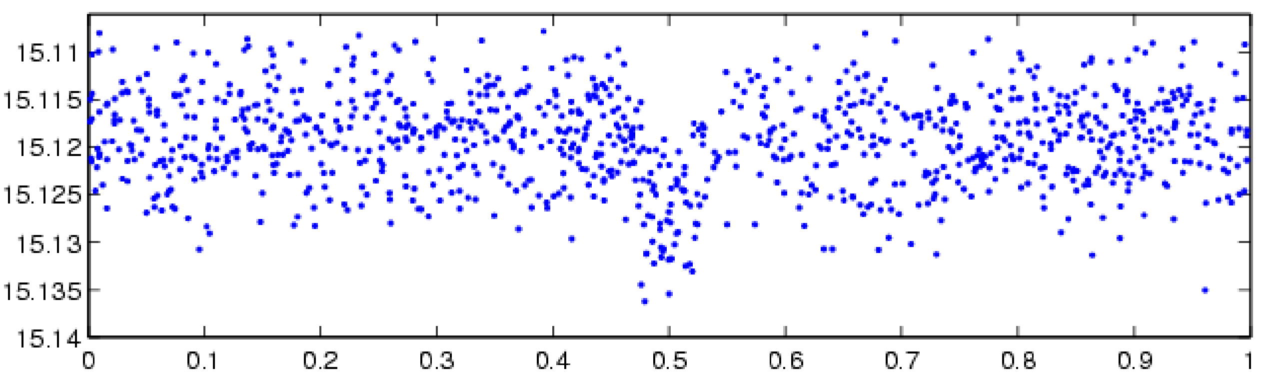


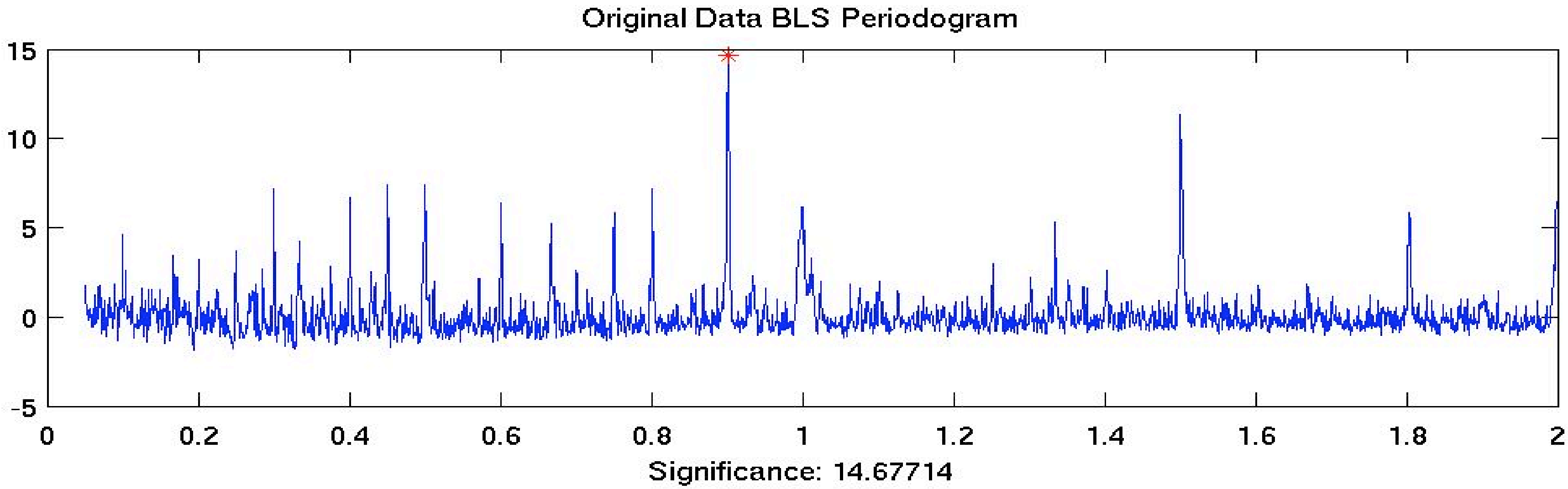
Corrected Data AoV Average Periodogram (200 Stars)

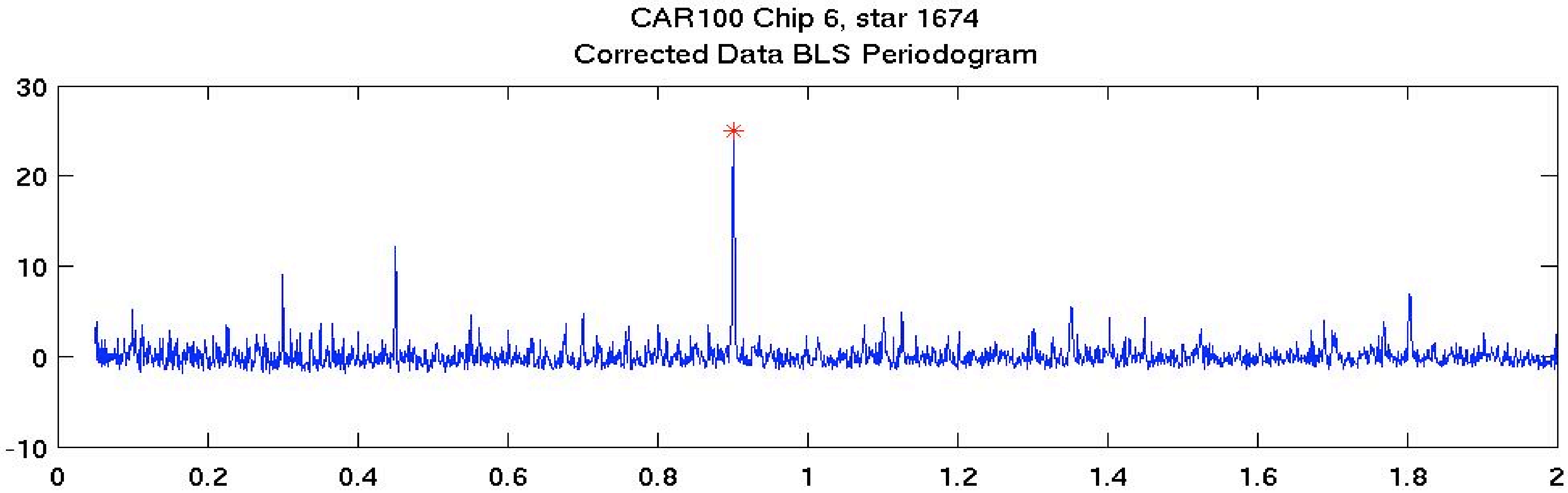
OGLE-TR-132, Before…



… and after




Original Data BLS Periodogram
15
10
5
0
-5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Significance: 14.67714


CAR100 Chip 6, star 1674
Corrected Data BLS Periodogram
30
20
10
0
-10
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

Super**WASP**

Wide Angle Search for Planets

D.J. Christian (Belfast)
W.I. Clarkson (Open University)
A. Collier Cameron (St Andrews)
N.A. Evans (Keele)
A. Fitzsimmons (Belfast)
C.A. Haswell (Open University)
C. Hellier (Keele)
S.T. Hodgkin (Cambridge)
K. Horne (St Andrews)
S.R. Kane (St Andrews)
F.P. Keenan (Belfast)
T.A. Lister (St Andrews)
A.J. Norton (Open University)
D. Pollacco (Belfast)
R. Ryans (Belfast)
I. Skillen (ING)
R.A. Street (Belfast)
R.G. West (Leicester)
P.J. Wheatley (Leicester)

Queen's University Belfast

University of St Andrews
St Andrews, Fife, KY16 9AJ, Scotland

UNIVERSITY OF CAMBRIDGE

ISAAC NEWTON GROUP
LA PALMA

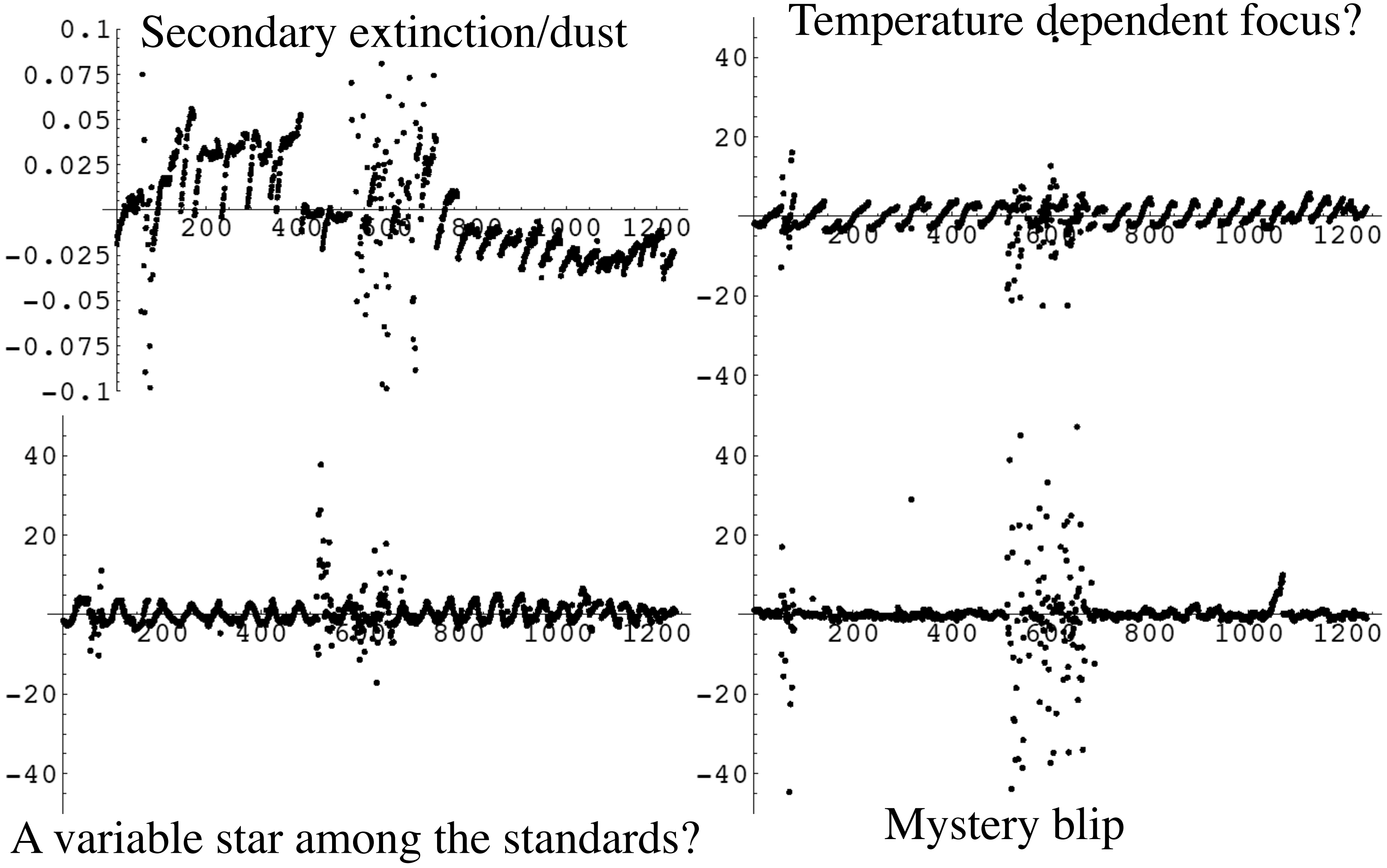
INSTITUTO DE ASTROFISICA DE CANARIAS · IAC ·

KEELE UNIVERSITY

University of Leicester

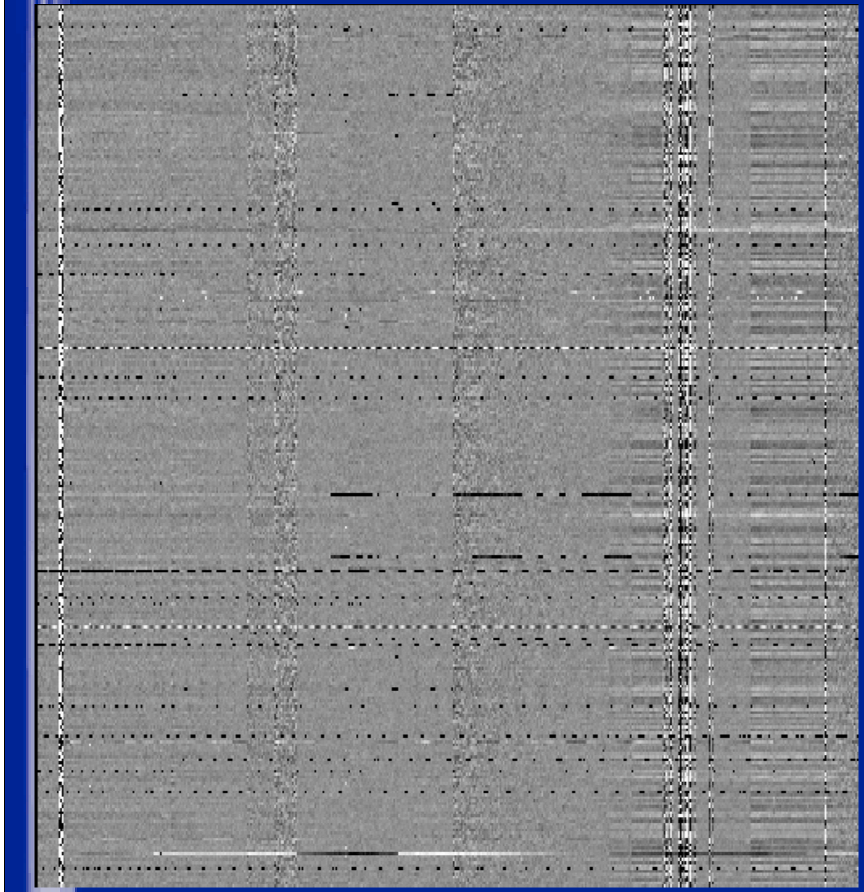
The Open University

12h43+28 field: First four SysRem components (arbitrary units)



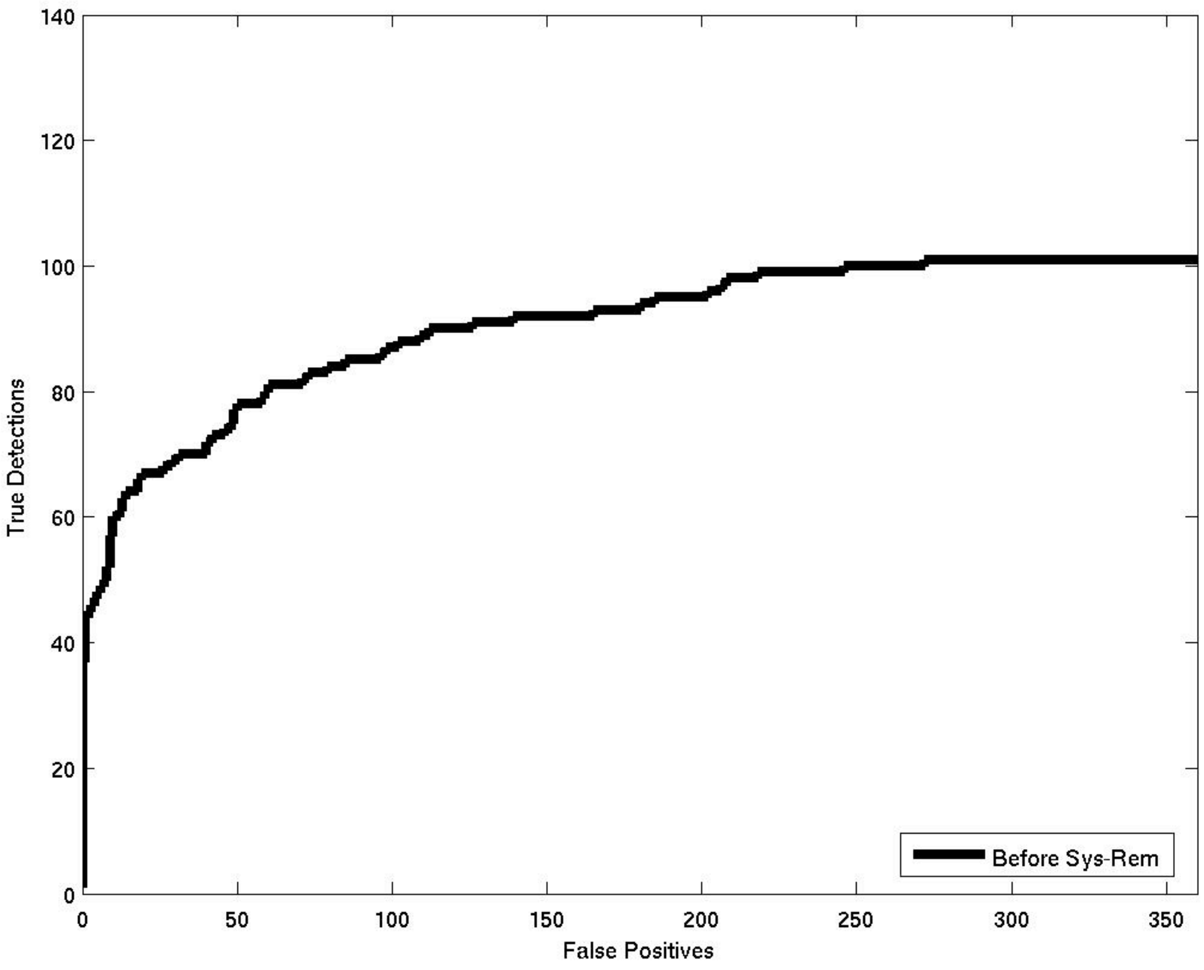
Example: SuperWASP 16h30+28 field

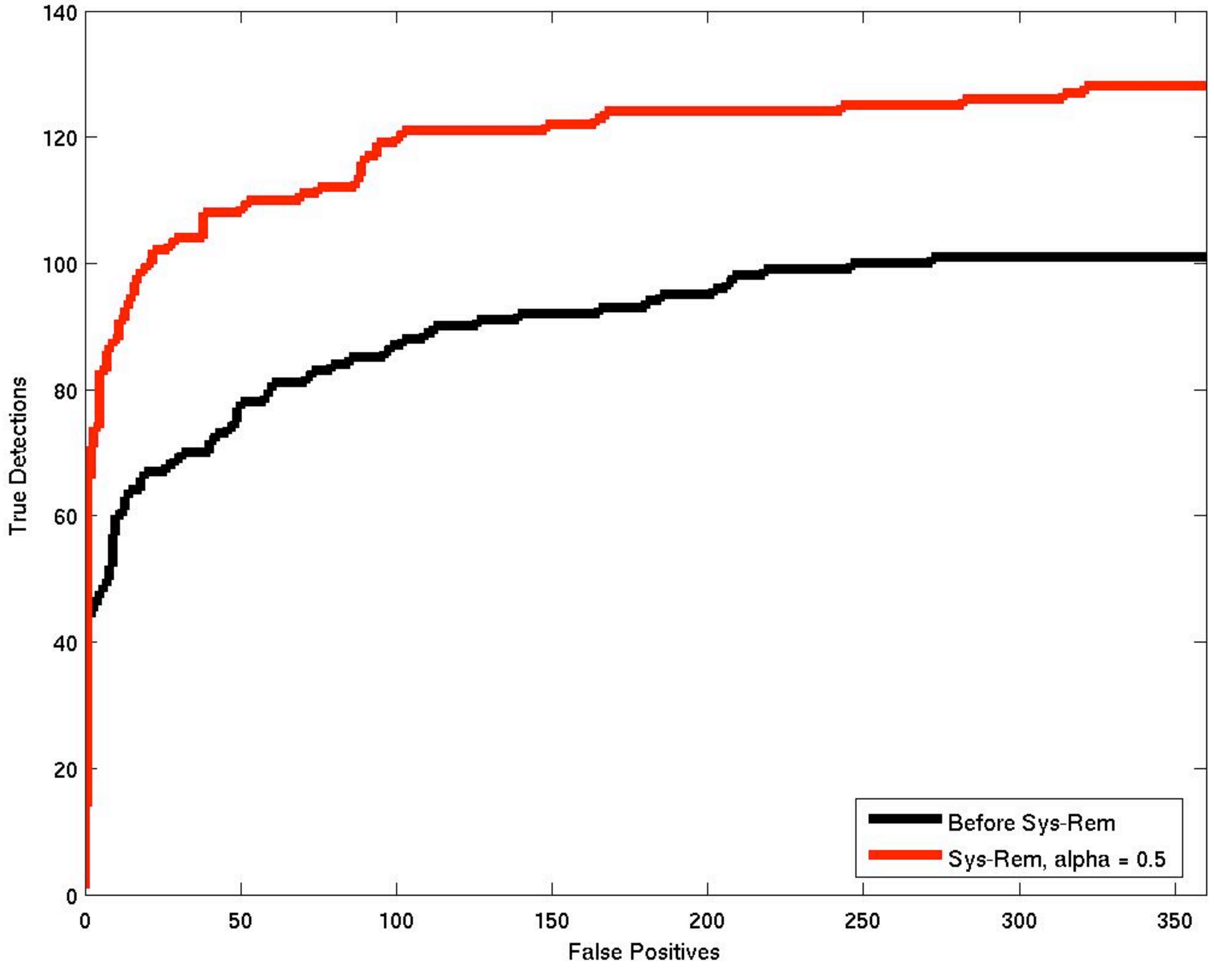
300 stars, 2549 observations spanning 100 days

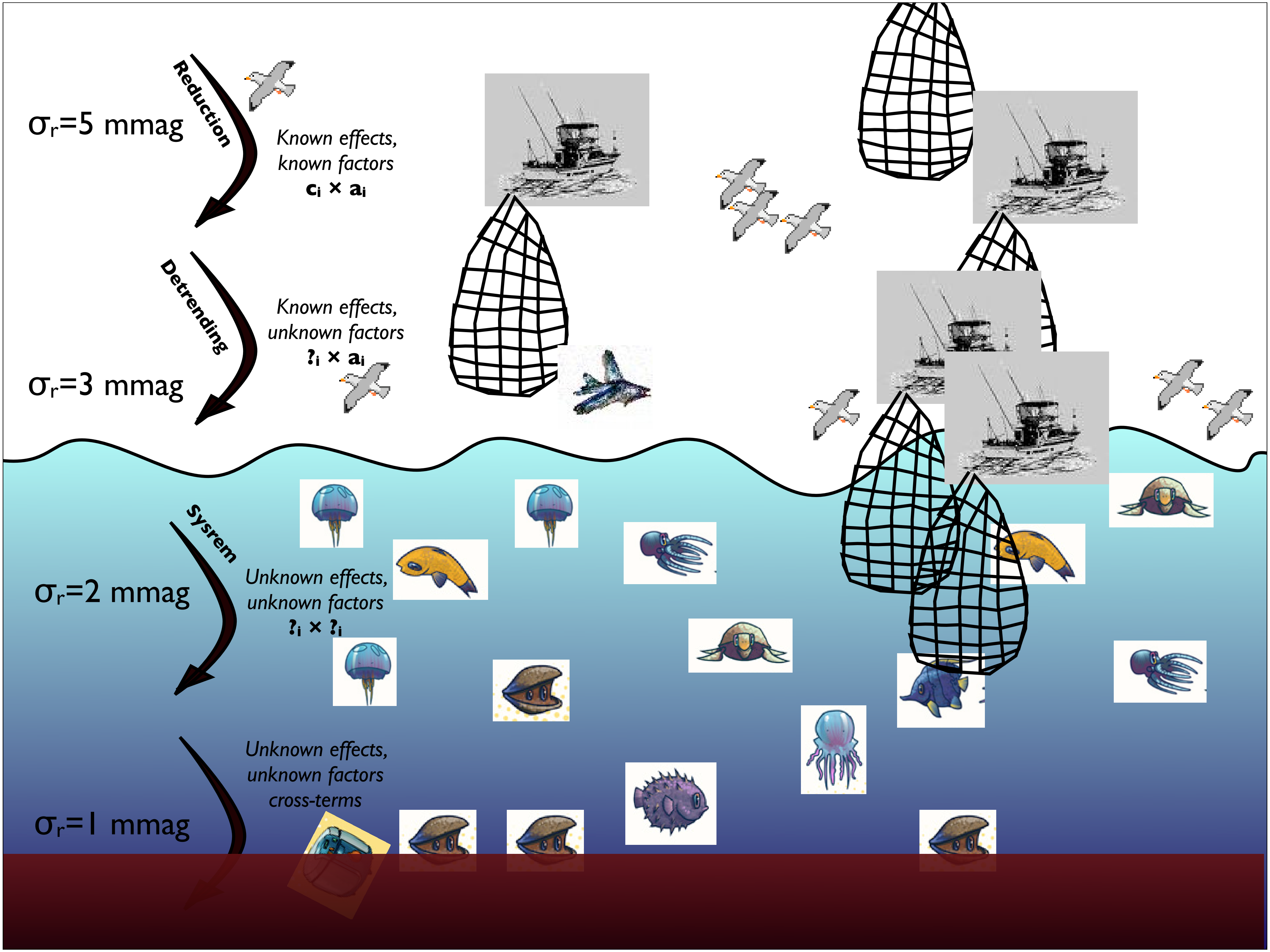


Before

After


140
120
100
80
60
40
20
0
True Detections
0
50
100
150
200
250
300
350
False Positives
Before Sys-Rem


140
120
100
80
60
40
20
0
True Detections
0
50
100
150
200
250
300
350
False Positives
Before Sys-Rem
Sys-Rem, alpha = 0.5



σ_r=5 mmag

Reduction

Known effects,
known factors
$c_i \times a_i$

Detrending

Known effects,
unknown factors
$?_i \times a_i$

σ_r=3 mmag

Sysrem

σ_r=2 mmag

Unknown effects,
unknown factors
$?_i \times ?_i$

Unknown effects,
unknown factors
cross-terms

σ_r=1 mmag