# A Bayesian method for the detection of planetary transits

C. Defaÿ[1], M. Deleuil[1,2], and P. Barge[1]

[1] LAM, Laboratoire d'Astrophysique de Marseille, BP 8, 13376 Marseille Cedex 12, France
[2] Université de Provence, CMI, 39 rue J. Curie, 13453 Marseille Cedex 13, France

**Abstract.** A new algorithm is proposed for the systematic analysis of light curves produced in high accuracy photometric experiments. This method allows us to identify and reconstruct in an automated fashion periodic signatures of extra-solar planetary transits. Our procedure is based on Bayesian methods used in statistical decision theory and offers the possibility to detect in the data periodic features of an unknown shape and period. Periodicities can be determined with a good accuracy, of the order of one hour. Shape reconstruction allows us to discriminate between transit events and possible artefacts. The method was validated on simulated light curves, assuming the data are produced by the instrument prepared for the future space mission, COROT. Moreover, stellar activity was accounted for using a sequence of the VIRGO-SOHO data. The probability to detect simulated transits is computed for various parameters: amplitude and number of the transits, apparent magnitude and variability level of the star. The algorithm can be fully automated, which is an essential attribute for efficient planetary search using the transit method.

**Key words.** stars: planetary systems – occultations – methods: data analysis

## 1. Introduction

The idea to detect planetary companions around stars by taking advantage of the small and periodic decrease of the stellar flux produced during successive eclipses was first proposed by Struve (1952).

From a pure geometrical point of view, the amplitude of the relative brightness variations are proportional to the square of the planet to star radius ratio: in the solar system it is of the order of 1% for a Jupiter-sized planet and of 0.01% for an Earth-sized planet.

Of course, such a possibility only arises in the eventuality that the observer's line of sight lies within the orbital plane of the planet. The probability for this to occur is, under the geometrical assumption, simply given by the star over planet radius ratio (0.0025 for the Sun-Jupiter system).

The actual probability of detection is in fact smaller than that since only a fraction of stars have planets. However, the present models of planet formation say nothing about the chances that a star will have planets. People preparing future space missions are faced to such an exercise and try to estimate some probability based on the T-Tauri disks observations and the percentage of exosolar planets discovered so far.

*Send offprint requests to*: C. Defaÿ
e-mail: `celine.defay@astrsp-mrs.fr`

On the other hand, the ability to detect in the flux of a star small drops due to the transit of one of its planet is determined by the properties of the emitter (the star), on that of the mask (the planet) and on the performances of the instrument. Thus, the capacity of transit detection depends on: (i) the spectral type, the apparent visual magnitude, the intrinsic variability of the star; (ii) the size and the orbital characteristics of the planet; (iii) the accuracy of the relative brightness measurements at various wavelengths. However, the complexity of the whole problem will be avoided in this paper which is only devoted to the search for an efficient detection method.

Three characteristics of a planetary transit can be used to identify possible events in the light curve of a target star: its specific shape, its periodicity and its intrinsic chromaticity. A number of people already addressed the question of transit detection using each of these criteria at once.

A method was first developed for the detection of planets in eclipsing binary systems using an a priori knowledge of the shape of the transits (Jenkins et al. 1995; Doyle et al. 1999); It is based on the standard matched filter approach and permits to detect planets larger than 3 $R_\oplus$ with a period of 60 days or less, with a probability of 90%. Such a method is efficient to detect and locate in the light curve a transit signal but does not permit its shape reconstruction.

Another method based on the differential limb-darkening in the red and the blue has been suggested to discriminate between a transit and possible artefacts (Rosenblatt et al. 1971; Borucki & Summers 1984). In fact, this is not an actual detection method but rather a discriminative test. Moreover it needs a larger photometric precision than currently available.

In this paper a new method is described for the detection of planetary transit building a simple algorithm, able to detect and to localize periodic transit-like features in the light curves of stars. It is based on the tutorial work of Bretthorst (1988, 1990) on Bayesian methods developed for the detection of simple periodic signals, sinusoidal or exponential, embedded in additive and Gaussian noises.

The algorithm is particularly well suited for white Gaussian noises. If the parent star is a solar one, its low variability does not really modify the noise characteristics and the algorithm remains efficient. Nevertheless the algorithm is not adapted for a larger stellar variability but a pre-processing of the data might allow to preserve results.

Our algorithm, developed in the specific case of the planetary transits, have some important advantages: (i) the period of the signal is determined easily, (ii) a reconstruction of the shape of the transit features is possible, (iii) detection can be fully automated.

Long term photometric data necessary to observe a series of transit are not available today. Ground based observations lack the accuracy and the continuity necessary for the detection of planets different from 51PegB like planets. Observation facilities are not presently available in space, except in the peculiar case of 47 Tucanae (2000) which was recently observed by HST, during only two weeks. In fact, appropriate data will be available in a near future from space observations: with the COROT satellite and with EDDIGTON or KEPLER, if accepted by ESA and NASA, respectively.

Simulated data have been used to build and finalize the detection method. So transits were modelized taking into account the limb darkening effect and light curves were modelized on the assumption that the signal of the stars is processed by an instrument similar to that developed for the COROT space mission, adding appropriate white noises to account for the instrumental noises and extrapolating the variability of the stars from Sun's one.

Section 2 describes our transit model and the way light curves are simulated. The detection method is presented in Sect. 3 and subsequently applied to the simulated data in Sect. 4. In this section the main results and performances are also discussed. Section 5 is probing the algorithm for various level of variability simulated by a simple scaling of Sun's activity. In Sect. 6 an automation of the algorithm is suggested.

## 2. Modelisation of the light curves

As long term photometric data are not presently available and do not have the appropriate accuracy, we test the proposed method on simulated data sets. Such data have
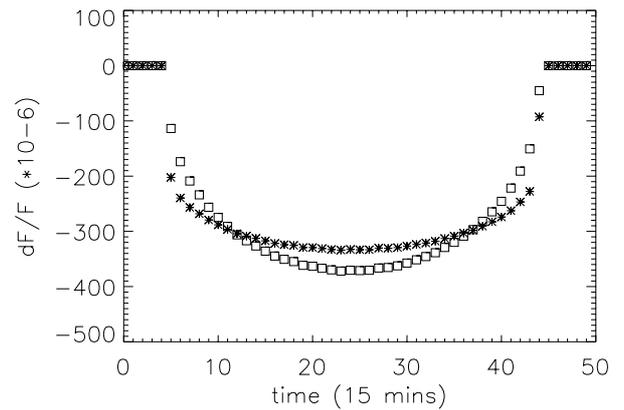


**Fig. 1.** The shape of a planetary transit at two different wavelengths. Units is in ppm: the rectangle and the star symbols refer to the shape at 400 nm and 820 nm, respectively. The characteristics of the planet are: $R_{\rm planet} = 2.8\ R_\oplus$, orbit$_{\rm planet}$ = 1 a.u., for a solar type star

been produced modelling the transit signal, the various instrumental noises and, also, the variability of the target stars.

### 2.1. The transit model

The transit of a planet in front of the disk of its parent star produces a periodic dimming of the star flux which can be characterized by three main parameters: the amplitude of the relative brightness variation, the duration of the event and the period.

Due to a planetary transit the light-curves of a target star eventually display a small drop whose shape depends on the observation wavelength, on the limb-darkening effect of the star and on the actual inclination of the orbital plane with respect to the line of sight. However, as we are only interested in exploring optimal detection methods, planetary transits will be modelized assuming zero orbital inclination.

Under this assumption the amplitude of the relative brightness variations only depends on the square of the planet to star radius ratio and the mean duration of the event is simply given by the standard expression:

$$\Delta t = 2R_*\sqrt{\frac{a}{\pi G M_*}},$$

where $a$ is the orbital radius.

Some example of typical transit parameters are given in Table 1. It must be noticed that the transit duration never exceeds a few hours.

Figure 1 illustrates the difference in the transit shape due to the chromatic effect and evidences the symmetric aspect of the transit signal. Chromatic information can improve planet detection however this question is out of the scope of the present paper and will be investigated in a further work.

**Table 1.** Transit parameters for a 51PegB like planet and a terrestrial planet for different spectral type of the parent star. $\delta F/F$ is the amplitude of the relative brightness variations, $\Delta t$ the duration of the transit, $T$ the orbital period, $\theta_{\mathrm{planet}}$ the temperature of the planet

| | 51PegB with $\theta_{\mathrm{planet}} = 1250$ K | | | 1.5 $T_\oplus$ planet with $\theta_{\mathrm{planet}} = 600$ K | | |
|---|---|---|---|---|---|---|
| Spectral type | $\delta F/F$ | $\Delta t$ (hours) | T (days) | $\delta F/F$ | $\Delta t$ (hours) | $T$ (days) |
| A8-F2 | 0.5% | 4.7 | 12.5 | 0.01% | 10.6 | 140 |
| G3-G7 | 1.1% | 2.6 | 3.6 | 0.022% | 5.9 | 40 |
| G8-K2 | 1.4% | 2.2 | 2.3 | 0.03% | 5 | 26 |
| K8-M2 | 2.7% | 1.3 | 0.8 | 0.048% | 2.9 | 8.4 |

## 2.2. The instrumental noises

Instrumental set-up and performances for the COROT space mission have been presented in a recent paper (Rouan et al. 1999). In this section, only the characteristics of the instrument of relevance for the subsequent simulations are summarized.

For a given amplitude of the signal, that is a planet of a given radius transiting in front of the disk of a star of a given spectral type, detection is possible only if the star visual magnitude is small enough. The corresponding limiting magnitude basically depends on the characteristics of the instrument (telescope, camera and CCDs) which set up the effective photons number, the size of the PSF (some 50 pixels in COROT's case) and some specific noises whose main contribution is the read out noise of the order of 15 $e^-/30$ s/pix. It also depends on the presence of other sources of photon such as stray light from the Earth and the Zodiacal light. This last one gives a dominant contribution which is of the order of 300 $e^-/30$ s/pix in our case. In consequence, the magnitude of the stars which can be monitored with sufficient photometric accuracy ranges from 12 to 16. The minimum sampling available for the data is imposed by the on-board exposure time and was set up to 15 min.

## 2.3. The stellar variability

Of course other sources of noise have an astrophysical origin as for example the one due to the star intrinsic variability, especially for spectral type later than the Sun.

The intrinsic variability of non-solar type stars is, however, still poorly known. Likely, it is in relation with the rotation and internal structure of the stars but also to the rotation of dark spots on the star surface.

The only long term data on the stars brightness variations are Sun's ones, recently obtained by the VIRGO experiment on-board the SOHO spacecraft. These data give the total solar flux (Fröhlich et al. 1997) and have no significant photons shot noise due to apparent magnitude of the Sun (Fig. 2 gives an example of the total Solar Variability).

They show that the relative brightness variation is lower than 0.01% (the amplitude corresponding to the transit of an Earth-sized planet) and would not bother appreciably the planetary transits detection.
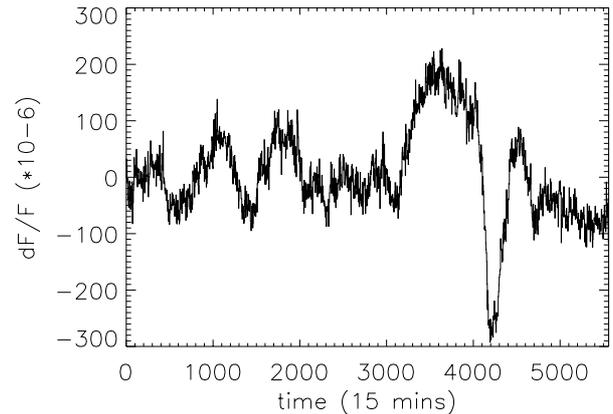


**Fig. 2.** Relative total flux variations of the Sun in ppm as observed by VIRGO (Fröhlich et al. 1997). Note the presence of a solar spot at $t = 4000$

Comparing the relative variations of the Sun to that of other solar-type stars, Uruh et al. (2000) typically found a scaling factor of the order of 0.8 and 0.9. On the other hand, in the case of active stars, Olàh et al. (1999) found that these variations can be ten times larger than Sun's ones. Of course such brightness variations can seriously perturb the detection. This case will be investigated in some details in Sect. 5 studying how this "noise" can be processed in order to improve planet detection.

Solarspots are frequently considered as the main source of artefacts because they are expected to mimic the effect of planetary eclipses. According to Vigouroux (1996) the crossing of a sunspot lasts about 14 days, which is fortunately much longer than planetary transits whose expected duration are several hours. On the other hand, the lifetimes of sunspot groups are quite sparse: approximately 50% of them lasts only two days, and 10% lasts more than 11 days (Richard 1997). The maximum lifetime of a sunspot group is about 2 or 3 rotations of the star. However, the point is not so clear for stars of a spectral type different from that of the Sun.

## 2.4. From noise characteristics to working assumptions

The three main types of noises in our simulated data are: photons, read-out and stellar noises.

(1) Photons noise describes a Poisson distribution which can be approximated by a Gaussian one when the
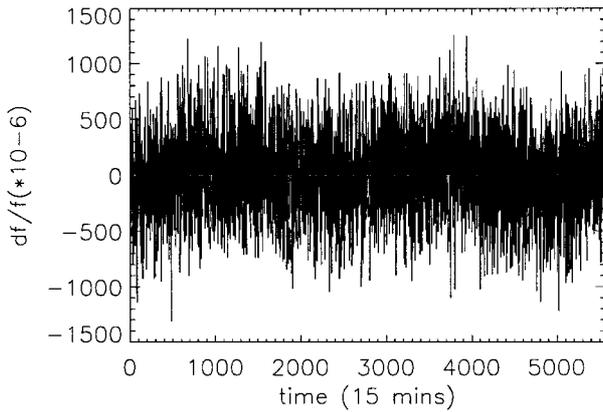
**Fig. 3.** Simulated transit against the total noise sequence for a Sun-like parent star with a magnitude $m_v = 13$. The radius of the planet is 2 $R_\oplus$; the transit has a 9.7 days period and a 5h duration

number of photon is large enough. It is basically a multiplicative process so that its variance is changing during a transit event. Nevertheless, the decrease in the number of photons is so small during an eclipse that the variance can be considered as a constant and the photons noise behaves as an additive one;

(2) The read out noise, whose characteristics are white and Gaussian (Adda 2000) is also an additive noise;

(3) Stellar variability is not well known and will be crudely assumed as an additive noise.

For instance, the relative amplitude of the various noises has been estimated explicitly for a given star ($m_v = 13$). It is approximately 0.03% for the total photons noise, 0.008% for the solar variability and 0.007% for the read-out noise. The photons noise clearly dominates all the other ones in the case of a solar-type star; this greatly simplifies the problem.

Figure 3 shows the two main components of typical simulated data: a planetary transit and a sequence of the total noise including the solar variability.

## 3. Description of the method

Bayesian methods are frequently used in statistical decision theory and offer also interesting possibilities for signal analysis. When looking for periodic features in time series, these methods allow to combine prior information on the expected signal with the information contained in the analysed data. So, detection is made easier because statistics is, in some way, guided.

Such methods were used by Gregory (1999) for the detection of a periodic signal of unknown shape and period in the general case where the noise is independent and Gaussian and the data are non-uniformly sampled. The model was developed by this author for the detection of periodic pulsar signals. It is based on a step-wise algorithm which is not really suited to the specific problem of the planetary transits and is time consuming.

### 3.1. Position of the problem

In the present work, the simulated data are assumed to be uniformly sampled with a time step short enough (15 min) as to fit correctly the expected profile of a planetary transit. Further, they are assumed to be a sum of two terms:

$$s(t_i) = f(t_i) + n(t_i) \qquad (1 \le i \le N), \qquad (1)$$

where the first one is the periodic signal and the second one is the total noise including the various components due to the instrument but also to the intrinsic variability of the target star.

The search for periodic transit features is made easier when the photon noise is assumed to dominate the "stellar" noise, since the second term $n(t_i)$ reduces to a single Gaussian white noise. Such an assumption holds, for example, in the case of quiet Sun-like stars. This assumption will be made in this section as the most convenient one to describe the new detection method we are proposing.

The first step is to expand the periodic signal in a standard Fourier series:

$$f(t) = \sum_{k=1}^{\infty} (a_k \cos(\omega k t) + b_k \sin(\omega k t)). \qquad (2)$$

For computational convenience we will only keep a finite number of terms in the series. After truncation to rank $m$, evaluated in Sect. 4.2., $f(t)$ is characterized by only $2m+1$ parameters: $(a_1, ..., a_m, b_1, ..., b_m, \omega)$, referred to as $\theta$ in the rest of the paper. Then, the problem is to determine $\theta$ from the data time series $s = (s(t_1), s(t_2), ..., s(t_N))$.

### 3.2. Estimate of the parameters

#### 3.2.1. The period of the expected signal

The last parameter in $\theta$ is the frequency $\omega$ of the signal to be detected. It is clearly independent of all the other ones which are Fourier coefficients and characterize the shape of the signal. In this respect they are usually called the "nuisance parameters" of the system, with no prior information about them.

Our goal is to determine the most probable value of $\omega$ for a discrete series of data. This posterior probability $P(\omega/s)$ can be derived thanks to Baye's theorem:

$$P(\omega/s) = \frac{P(s/\omega)P(\omega)}{P(s)}. \qquad (3)$$

In fact, as $P(s)$ is independent from $P(\omega)$, we get $P(\omega/s) \propto P(s/\omega)P(\omega)$.

Let $L(\omega) = P(s/\omega)$ denotes the likelihood function of the expected frequency $\omega$. In the absence of any prior information about this parameter a uniform density distribution has been assumed, so that the required determination of $P(\omega/s)$ reduces to an estimate of the likelihood function:

$$L(\omega) = \int_{-\infty}^{\infty} L(\theta) da_1 ... da_m db_1 ... db_m \qquad (4)$$

where $L(\theta) = P(s/\theta)$ is the global likelihood function.

Following standard methods in information theory and signal analysis, this expression must be maximized in order to get the most probable value of the likelihood function or "maximum likelihood" for $\omega$.

### 3.2.2. The shape of the signal

Once the period is found, the next step is to determine the Fourier coefficients $a_k$, $b_k, k = 1, ... m$ containing the information on the signal shape. They are simply derived maximizing the global likelihood function for any given value of $\omega$, satisfying the system of equations:

$$\frac{\partial L(\theta)}{\partial a_k} = 0, \frac{\partial L(\theta)}{\partial b_k} = 0 \qquad k = 1...m. \tag{5}$$

### 3.3. The global loglikelihood function $\log L(\theta)$

Under the assumption that the noise is white, additive and Gaussian with variance $\sigma^2$ and zero mean, the probability associated to each discrete time step $t_i$ writes:

$$P(n(t_i)) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-n(t_i)^2}{2\sigma^2}\right), \tag{6}$$

which, as the total noise $n$ is assumed to result from independent contributions, writes:

$$P(n) = \prod_{i=1}^{N} P(n(t_i)). \tag{7}$$

The probability of observing $s$ given the set of parameters $\theta$ and according to Eq. (1) is:

$$P(s/\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{-[s(t_i) - f(t_i)]^2}{2\sigma^2}\right\}, \tag{8}$$

which is also referred as the global likelihood $L(\theta)$ and writes in logarithmic form:

$$\log L(\theta) \propto -N\log(\sigma) + \sum_{i=1}^{N} \frac{-1}{2\sigma^2}[s(t_i) - f(t_i)]^2. \tag{9}$$

Using the expanded expression of $f(t_i)$ and simplifying the cross terms we can now rewrites Eq. (9) as:

$$\log L(\theta) = -N\log(\sigma)$$
$$+ \frac{-N}{2\sigma^2}\left[s^2 - \frac{2}{N}\sum_{k=1}^{m}(a_k\alpha_k + b_k\beta_k) + \frac{1}{2}\sum_{k=1}^{m}(a_k^2 + b_k^2)\right] \tag{10}$$

where, according to Bretthorst (1988) notation:

$$\alpha_k = \sum_{i=1}^{N} s(t_i)\cos(\omega k t_i), \qquad k = 1...m$$

$$\beta_k = \sum_{i=1}^{N} s(t_i)\sin(\omega k t_i), \qquad k = 1...m.$$

Integrating Eq. (10) over $a_k$ and $b_k$ we get:

$$\log L(\omega) \propto \frac{-N}{2\sigma^2}\sum_{k=1}^{m}\left(\frac{\alpha_k^2}{N^2} + \frac{\beta_k^2}{N^2}\right). \tag{11}$$

Note that the standard deviation of the total noise $\sigma$ is unknown but can be conveniently approximated by that of the data itself. Then, the most probable value of the frequency of the expected signal, $\omega_{lik}$, can be easily determined fitting at best the position of the maximum of $\log L(\omega)$.

Consequently, Eqs. (5) can be solved easily and lead to the Fourier coefficients:

$$a_k = \frac{2\alpha_k}{N}, \quad b_k = \frac{2\beta_k}{N}, \qquad k = 1, ..., m. \tag{12}$$

Therefore, the expression of the reconstructed signal writes:

$$f(t) = \sum_{k=1}^{m}\left(\frac{2\alpha_k}{N}\cos(\omega k t) + \frac{2\beta_k}{N}\sin(\omega k t)\right). \tag{13}$$

As a final comment, since no analytical model of the stellar variability is available, it is impossible to perform the same kind of computations when the simulated data include the stellar noise. This question will be specifically addressed in Sect. 5.

## 4. Implementation and validation of the algorithm

For computational convenience only a 60-day long time span has been simulated and will be used in the rest of the paper. In this section, the algorithm was validated on a simple case assuming that: (i) the visual magnitude of the star is 13, (ii) noise is white and Gaussian, (iii) stellar activity is neglected. The signal to noise ratio has been taken high enough so as to identify the transit signal. Validation is based on the capacity of the algorithm to identify and reconstruct a reference signal. It is subsequently performed with a 5-hours transit signal produced by a planet 2 times larger than the Earth.

### 4.1. Estimating the period of the reference signal

For a given data set the likelihood function is derived according to Eq. (13). This function displays a strong maximum (Fig. 4) which defines the Most Likely Period (MLP) but, in general, does not exactly coincide with the reference period. In fact, MLP depends on the chosen data sample and spreads around the reference period within an error band $\Delta\omega$. This error band can be estimated using a standard bootstrap technique. A MLP distribution is built in such a way (Fig. 5), generating a few hundred of data sets as to achieve statistically meaningful results. The standard deviation of this distribution sets up the incertitude on the period estimate and allows us to define a criterion for a "true" detection (i.e.) when the MLP is within $\Delta\omega$. In the case discussed in Fig. 5 it is of the order of 1.4 units (the unit is 15 mins) for a sample of 200 data.
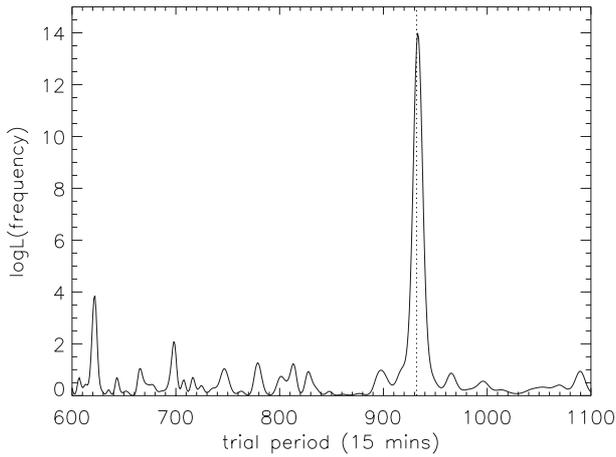
**Fig. 4.** Estimation of the most likely period in the case of a Gaussian white noise. The reference period is indicated by the dotted line. The magnitude of the star is 13; the transit has a period of 932 quarters of an hour with a duration of 5 hours; the planet radius is $R_{\mathrm{planet}}/R_{\oplus} = 2$. Note also the presence of harmonics: for example 2T/3 (620 units), 3T/4 (698 units)
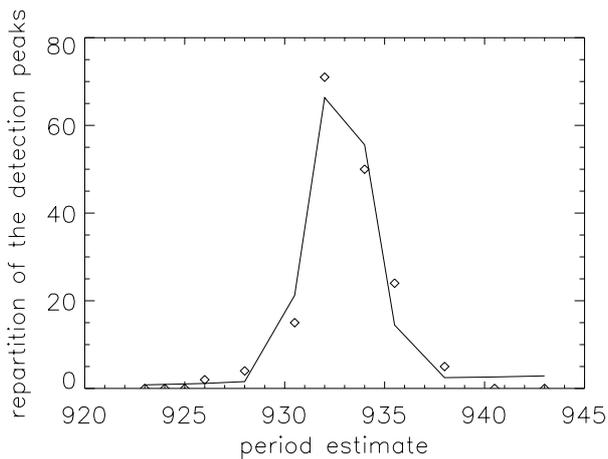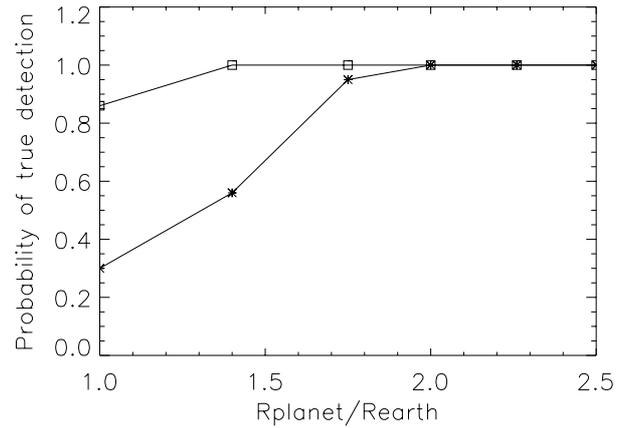


**Fig. 6.** Probability of true detection as a function of the planet radius, in earth units, assuming a pure white noise. The magnitude of the star is 13. Crosses correspond to 6 transits with a 5 hours duration, squares to 40 transits with a 3 hours duration. Results are obtained thanks to 200 bootstrap samples
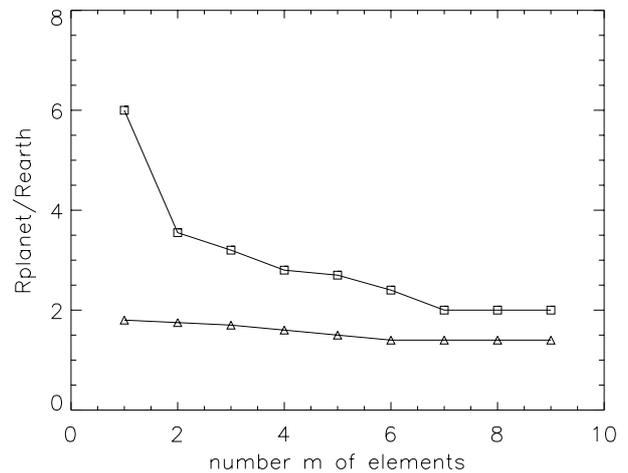


**Fig. 5.** Repartition of the detection peaks for 200 generated data (diamonds). The transit has a period of 932 quarters of an hour with a duration of 5 hours; the planet radius is $R_{\mathrm{planet}}/R_{\oplus} = 2$. The magnitude of the star is 13. The solid line is the best fit with a Gaussian function with $\mu = 932, \sigma = 1.4$



**Fig. 7.** Radius, in Earth-units, of the smallest detectable planet with a 99% confidence level, as function of the truncation rank $m$. The magnitude of the Sun-like parent star is 13. Squares correspond to a 6 transits detection with a 5h duration, triangles to a 40 one with a 3h duration

So, a detection is considered as a true one when the MLP is located within an error band of $\pm 2$ units. The accuracy on the period estimate has been tested for different transits numbers and was always found of the order of one hour.

The error band $\Delta\omega$ permits also to evaluate, using the same bootstrap technique, the probability of true detection as a function of the planet radius (i.e.) the detection efficiency of the algorithm. Therefore, for a given magnitude of the star and for each value of the planet radius, these computations provide a confidence level of the detection.

The results are plotted in Fig. 6. It is found that the minimal planetary radius detectable with a 99% confidence level is $2\,R_{\oplus}$ for 6 transits and $1.4\,R_{\oplus}$ for 40 transits.

The results will be subsequently presented with a 99% confidence level.

### 4.2. Truncation rank of the Fourier series

Of course, when computing $\log L(\omega)$, the accuracy depends on the value of the truncation rank $m$ of the Fourier series. The choice of this integer results from a compromise between computation time and detection efficiency. The dependence of the detection efficiency as a function of the truncation rank is illustrated in Fig. 7. The optimal value of $m$ is 7 and does not depend on the transit number.
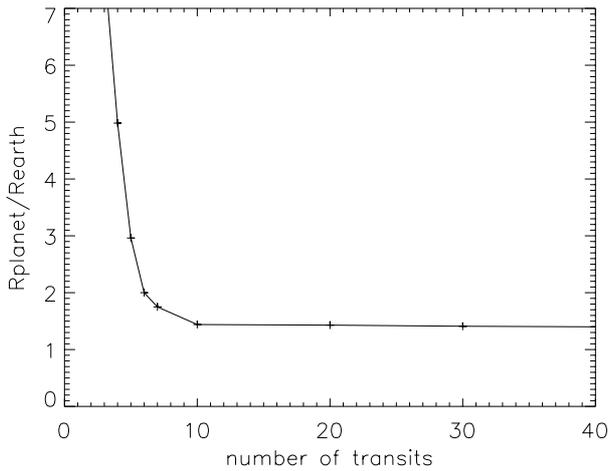
**Fig. 8.** Radius, in Earth-units, of the smallest detectable planet with a 99% confidence level, as function of the transit number $m$. The magnitude of the star is 13
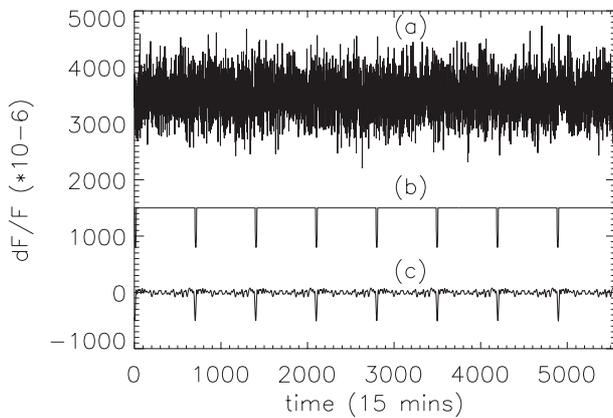


**Fig. 9.** Example of signal reconstruction: **a)** noise, **b)** reference signal, **c)** reconstructed signal. 8 transits with a 4h duration; the radius of the planet is $R_{\mathrm{planet}} = 2.5\,R_{\oplus}$. The magnitude of the star is 13

### 4.3. Dependence of the algorithm on the transit number

The detection efficiency depends on the periodicity of the signal, in obvious relation with the transit number. It can be estimated as the minimum detectable planet radius computed as a function of the transit number and at a 99% confidence level. For a number of transits greater than 10 this minimum radius tends towards a limiting value: 1.4 Earth-radius in the case of a star of $m_v = 13$ (see Fig. 8).

### 4.4. Reconstruction of the transit signal

As explained in Sect. 3, our method permits to evaluate the period of the signal but also to determine its shape through Eq. (13). Reconstructed in such a way, the signal is plotted in Fig. 9. In this example, the signal to noise ratio is low and an accurate reconstruction of the signal is difficult: the amplitude and duration of the reconstructed signal are found smaller than those of the reference. This
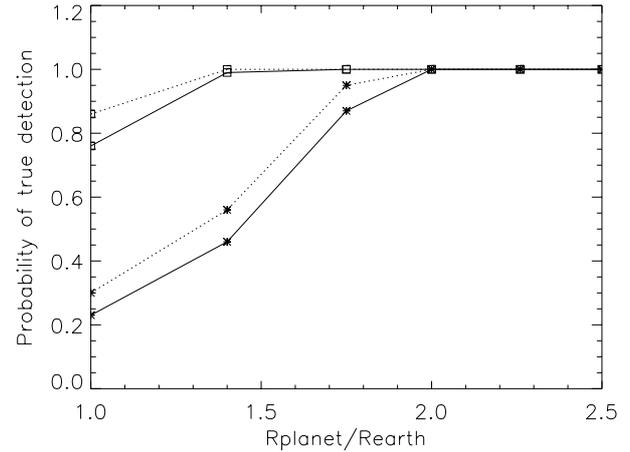


**Fig. 10.** Probability of true detection as a function of the planet radius, in earth units, for a pure white noise (dotted line), and for a Sun-like parent star (solid line). The magnitude of the star is 13. Crosses correspond to 6 transits with a 5h duration, squares to 40 transits with a 3h duration. The results are obtained with 200 bootstrap samples

comes from relatively strong noise and weak data sampling (16 points for a single transit).

## 5. Including a "coloured" stellar noise

In this section more realistic data are simulated including a sequence of the VIRGO-SOHO data for solar type target stars. In order to explore stronger level of stellar variability these data were also extrapolated using a scale factor whose values range from one to six and introducing a duplicated solar spot. In this case the noise is obviously different from a simple Gaussian white noise, anyway we use the same detection method.

### 5.1. Probability of correct detection for a Sun-like target star

The detection probability computed when a solar-like variability is taken into account is plotted as a function of the planet radius in Fig. 10. In this section, we used the VIRGO data without any extrapolation. Compared to the results obtained in the white noise case, detection is found to be only slightly decreased, by less than 7%. The change is so small, in the case of a solar-like variability, that the algorithm keeps nearly the same efficiency and has not to be changed.

Figure 11 gives, under the same assumption, the radius of the smallest detectable planet with a 99% confidence level as a function of the magnitude of the parent star. On one hand, planets larger than $5\,R_{\oplus}$ are detectable on target stars brighter than $m_v = 16$ (independently of the transit number), on the other hand, planets smaller than $2\,R_{\oplus}$ are detectable on target stars brighter than $m_v = 14$, depending on the transit number. 51PegB-like planets, for example, should be very easily detected, whereas
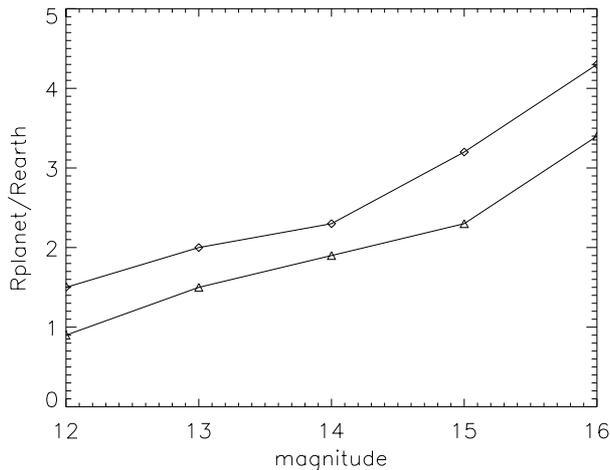
**Fig. 11.** Radius, in Earth-units, of the smallest detectable planet with a 99% confidence level, as function of the magnitude of the Sun-like parent star. Diamonds correspond to 6 transits with a 5h duration and triangles to 40 transits with a 3h duration
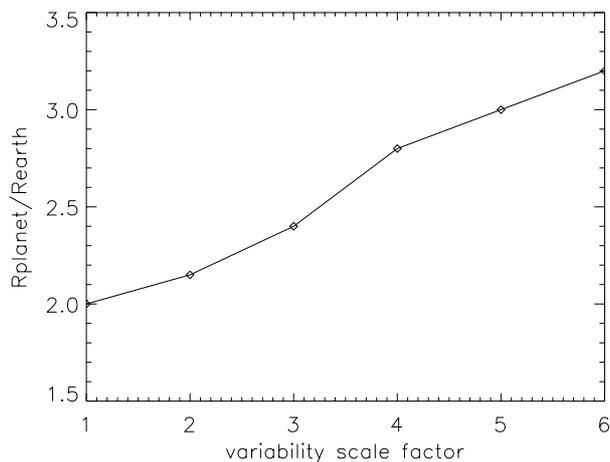


**Fig. 13.** Spectra of: **a)** the Sun's relative variations observed by VIRGO, **b)** a planetary transit signal. The transit duration is 10h. The planet has a radius of $2\,R_\oplus$ and a period of 15 days



**Fig. 12.** Radius, in Earth-units, of the smallest detectable planet with a 99% confidence level as a function of the variability scale factor. The magnitude of the star is 13. Data contains 6 transits with 5h duration

terrestrial ones need brighter target stars and long enough observation periods.

### 5.2. Detection when increasing the stellar variability level

As discussed above, solar variability does not really modify the detection efficiency. Nevertheless, it must be pointed out that Sun's variability is particularly weak.

Signatures of dark spots can be found in the SOHO data. One of them has been identified in the data sequence at our disposal and was used as a model to simulate more noisy light-curves. Our simulated stellar noise sequence contains three such dark-spot features in order to test how transit detection is affected.
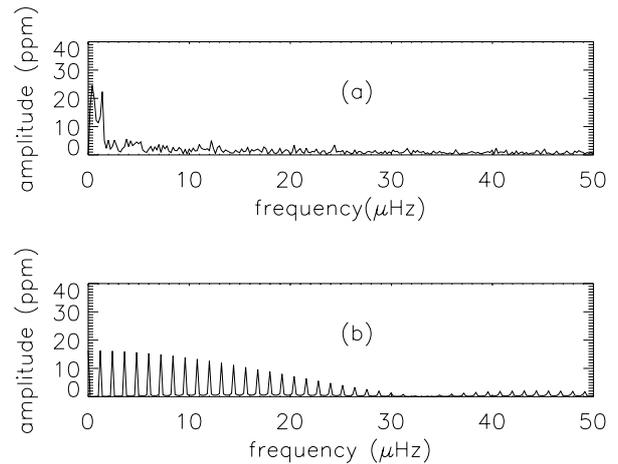
The consequences on the detection of a larger stellar variability can be explored evaluating the radius of the smallest detectable planet as a function of the scaling factor (Fig. 12).

Clearly our algorithm becomes inefficient for the large stellar variabilities. This was, indeed, foreseeable since our method assumes that the noise is white and Gaussian (see Sect. 3), whereas stellar noise is not. Since the stellar variability is too poorly known to be appropriately modelized the solution is to preprocess the data before applying the detection algorithm.

This pre-processing can be done by an additional filtering of the data in two ways.

(1) The first method consists in applying to the data a simple highpass filter. As appears in Fig. 13, the solar variability mainly contains low frequencies while the spectrum of the largest transit is more spread: the use of an highpass filter at 4 $\mu$Hz improves the results by removing the most disturbing frequencies.

(2) The second method consists in a whitening of the data. This operation can be performed dividing the Fourier transform of the signal by its modulus (Guérault et al. 1997). By construction this filter is well adapted to the noise itself.

Figure 14 shows the transit detectability after those preprocessing operations. Comparisons show clearly that for a scaling factor larger than four, the highpass filter leads to the detection of planet 0.8 times smaller than without pre-processing.

As regards the whitening operation it is completely inefficient for the low variability levels but can improve detection when the variability factor is larger than 4. So, the whitening filter is only necessary for the highest level of the stellar variability.
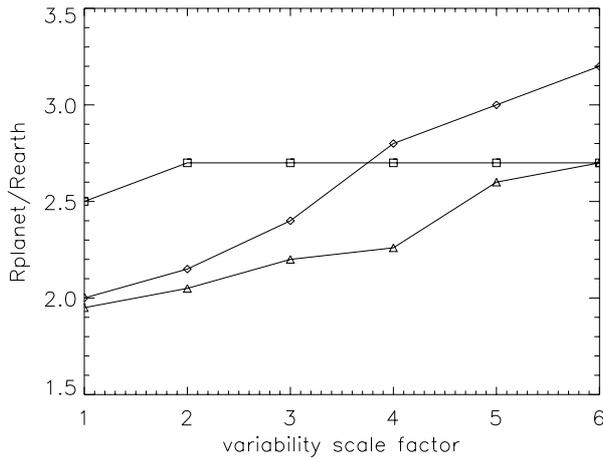
**Fig. 14.** Radius, in Earth-units, of the smallest detectable planet with a 99% confidence level, as function of the variability scale factor. The magnitude of the star is 13. Data contains 6 transits with 5h duration. Diamonds correspond to results without preprocessing, squares and triangles to results with a preprocessing, whitening or highpass filter, respectively
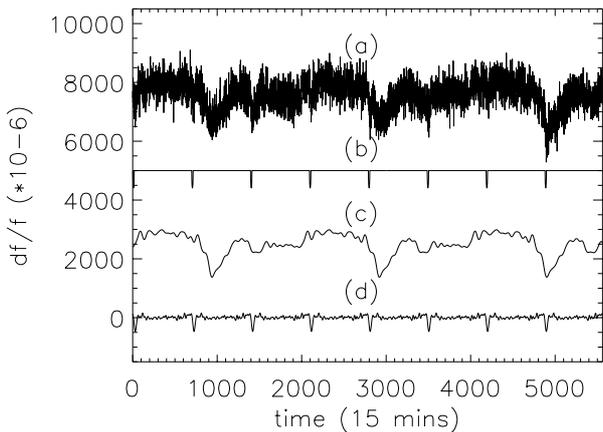


**Fig. 16.** Reconstructed signal for a 51PegB like planet around a Solar-type star ($m_v = 13$). The time sampling is 4 mins. The solid curve indicates the reference transit and diamonds the reconstructed signal

signal may be sufficient to characterize planetary transits discarding artefacts unambiguously.

Moreover, if the sampling of the light-curve and the signal to noise ratio are large enough, it is even possible to reconstruct the precise shape of each transit feature. In this case any eventual confusion with artefacts becomes impossible, as explicitly appears in Fig. 16.

## 6. Automation of the algorithm

To detect extra-solar planets with the transit method requires the study of thousands of light curves. Such a task cannot be planned by simply scrutinizing the data, even if pre-processed. Rather it is necessary to use automated data processing. The algorithm presented in this paper offers such a possibility, discussed in the present section.

### 6.1. Determination of a detection threshold

Our method is based on identifying a maximum in the log-likelihood function what was implicitly assumed in Sects. 3 and 4 devoted to the description and validation of the algorithm. Of course, when working on actual data, the first step is to identify a significant maximum in $\log L(\omega)$. A standard way to do is using a detection threshold. A likelihood peak will be considered as a potential detection if its amplitude is greater than a given threshold, which can be set up as a value proportional to the standard deviation of $\log(L(\omega))$. It is obviously an important parameter of the method since it directly controls the plausibility of the events, deciding whether a detection peak can be classified among possible planetary transits.

This optimal threshold is determined thanks to bootstrap sampling by computing the percentage of MLP higher than a given threshold in two different cases: (i) transits are included in the data-sets and detection is performed with a 99% confidence level for a given magnitude of the star (this implicitly sets up the faintest detectable



**Fig. 15.** Simulated and reconstructed signals (in the case of a solar type star): **a)** total signal (star with $m_v = 13$, planet with $R_{planet} = 2\,R_\oplus$ and transit duration of 5h), **b)** reference signal, **c)** reconstructed signal without highpass filter (signature of a solar spot), **d)** reconstructed signal with highpass filtering (signature of a transit)

### 5.3. Reconstruction of the transit signals

Examining a sequence of the VIRGO-SOHO data, signal reconstruction might be thought to be affected by artefacts coming from the stellar activity (Fig. 2).

Since the expected duration of a planetary transit always remains smaller than 20 hours, we use as a first step the previously defined high-pass filter. Figure 15 gives examples of some reconstructed signals, with or without the use of this filter. Without the use of an highpass filter the star spot is detected and reconstructed while the use of the filter permits to detect and reconstruct the transit. These results show that, in the limit of our present knowledge on the stellar activity, a simple filtering of the total
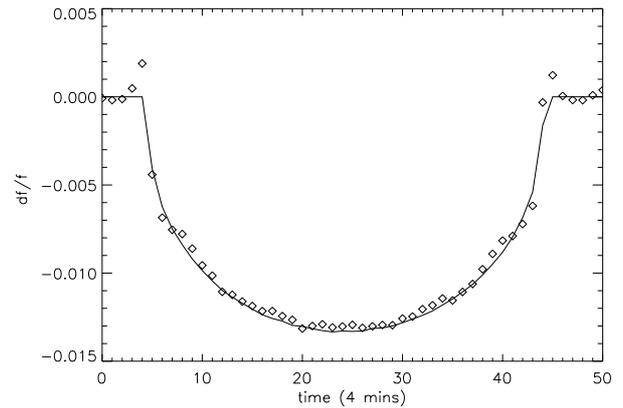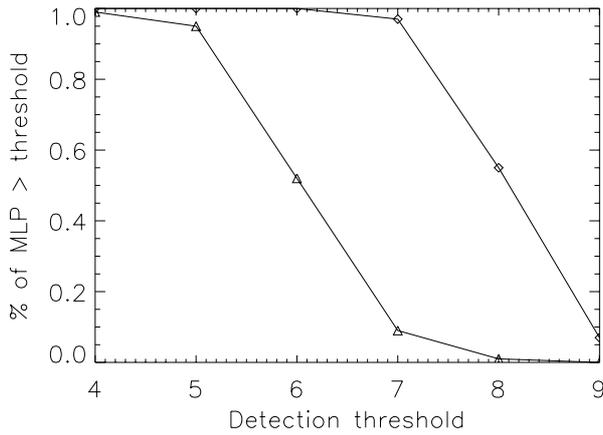
**Fig. 17.** Probability that the MLP for the true (diamonds) and false (triangles) detections are beyond the detection threshold for the radius 2 $R_\oplus$ of the smallest planet detectable with a 99% confidence level with 6 transits for $m_v = 13$. The results are obtained with a 200 bootstraping sample
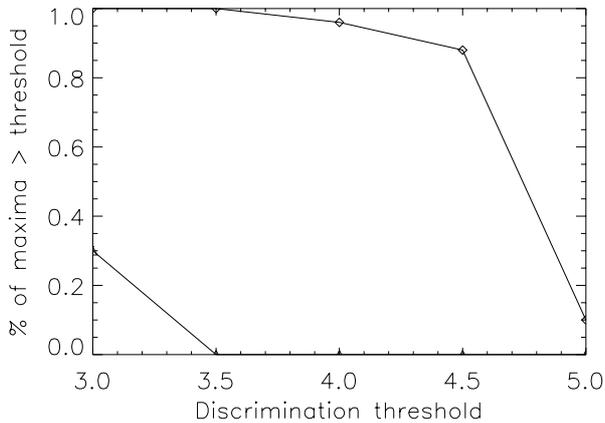


**Fig. 18.** Probability that the maximum peaks for the true (diamonds) and the false (triangles) discriminations are beyond the detection threshold. The radius of the smallest planet detectable with a 99% confidence level and a star with $m_v = 13$ is 2 $R_\oplus$ in the case of 6 transits. The results are obtained with a 200 bootstraping sample

signal i.e. the smallest planet); (ii) no transits are included and so, MLP only correspond to artefacts.

These results, plotted in Fig. 17, show that a threshold of $7\sigma$ is the optimal one to avoid at best artefacts while working at the detection limit.

### 6.2. A discriminating test

In what concerns the discrimination from the stellar variability, the method is nearly the same as previously defined in the case of the detection threshold, the only difference lies in the fact that we are working on the absolute value of the reconstructed signal. A discriminating threshold can be set up as a value proportional to the standard deviation of this signal. Computed in the same way as above the optimal threshold is found to be $3.5\sigma$ (see Fig. 18).

## 7. Discussion and conclusion

We propose a new method for the detection of periodic planetary transits using a standard tool in statistical decision theory, the so called Bayesian method. The algorithm is developed for uniformly sampled data and is based on a simple Fourier expansion of the signal. Bayes's theorem is used to determine the posterior probability of the transit frequency, the likelihood function $L(\omega)$. This function is the key point of the method since its knowledge allows to go back to two important parameters for planet detection: the period and the shape of the transit, in close relation with the size and the orbit of the planets.

The algorithm was tested on simulated data-sets produced according to the definition of the instrument of the COROT space mission. The expected accuracy on the period determination is found to be of the order of three time steps (45 min), when detection is performed at a 99% confidence level. In the case of a solar type star with $m_v = 13$ the radius of the smallest detectable planet is found to be 1.4 $R_\oplus$ for a 2-day periodic signal and 2 $R_\oplus$ for a 10-day periodic signal (with a 60-day long time span).

Our algorithm can be completely automated determining a significant maximum in the log-likelihood, $\log(L(\omega))$, thanks to a threshold to be evaluated. This is an essential property for an algorithm aimed at the search for planets with the transit method. Indeed, as the actual probability of planet detection is very small, a large number of target stars have to be monitored and a large number of lightcurves have to be processed in consequence.

Another important capacity of our method lies in the possible determination of the Fourier coefficients of the signal whose knowledge leads to an actual form reconstruction. Indeed, signal reconstruction enables to discriminate between the symmetric shape of a planetary transit, the asymmetric shape of a solar spot and the unspecific shape of an artefact. This discriminating test was easily implemented in the automated version of the algorithm.

A final remark is in relation with our starting assumption that the noise in the data-set is additive and Gaussian white noise. It is obviously well suited in the case of a dominant photon noise but is also adapted to the case of the coloured noise of stellar activity similar to the Sun's one. For large stellar variability a pre-processing of the data has been performed, using highpass filtering or whitening, as to cut the most disturbing frequencies and to come back to the efficient algorithm developed for the white noise case. This last point is an important one since it will help us seriously for the detection of planetary transits even at a high level of stellar variability.

## References

Adda, M. 2000, Photométrie CCD de très haute précision dans l'espace, Observatoire de Meudon, 128

Borucki, W. J., & Summers, A. L. 1984, Icarus, 58, 121

Bretthorst, G. L. 1988, Maximum-Entropy and Bayesian Methods, in Sci. Eng., 1, 75

Bretthorst, G. L. 1990, J. Mag. Res., 88, 552

Doyle, L. R., Deeg, H. J., Kozhevnikov, V. P., et al. 1999, ApJ, 535, 338

Fröhlich, C., Andersen, B. N., & Appourchaux, T. 1997, Solar Phys., 170, 1

Gregory, P. C. 1999, ApJ, 520, 361

Guérault, F., Signac, L., Goudail, F., & Réfrégier, P. 1997, Opt. Eng., 36, vol. 10, 2660

Jenkins, J. M., Doyle, L. R., & Cullers, D. K. 1996, Icarus, 119, 244

Oláh, K., Kolláth, Z., & Strassmeier, K. G. 2000, A&A, 356, 643

Richard 1997, The lifetime of a sunspot group, http://www.ips.gov.au/papers/richard/ spot-lifetime.html

Ronald, L., Gilliland, T. M., Brown, P., et al. 2000, ApJ, accepted

Rosenblatt, F. 1971, Icarus, 14, 71

Rouan, D., Baglin, A., Barge, P., et al. 1999, Phys. Chem. Earth, 24(5), 567

Struve, O. 1952, Observatory, 72, 199

Unruh, Y. C., Knaack, R., & Fligge, M. 2000, Are the Sun really tamer than those of other comparable solar-type stars?, PASPC, in press

Vigouroux, A. 1996, Étude de la variabilité solaire à long terme, Observatoire de la côte d'Azur, 60