An anomaly detector with immediate feedback to hunt for planets of Earth mass and below by microlensing

M. Dominik,[1][*][†] N. J. Rattenbury,[2] A. Allan,[3] S. Mao,[2] D. M. Bramich,[4] M. J. Burgdorf,[5] E. Kerins,[2] Y. Tsapras[5] and Ł. Wyrzykowski[6,7]

[1]*SUPA, University of St Andrews, School of Physics & Astronomy, North Haugh, St Andrews KY16 9SS*
[2]*Jodrell Bank Observatory, Macclesfield, Cheshire SK11 9DL*
[3]*School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL*
[4]*Isaac Newton Group of Telescopes, Apartado de Correos 321, 38700 Santa Cruz de La Palma, Canary Islands, Spain*
[5]*Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf, Birkenhead CH41 1LD*
[6]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*
[7]*Warsaw University Astronomical Observatory, Al. Ujazdowskie 4, 00-478 Warszawa, Poland*

ABSTRACT

The discovery of OGLE 2005-BLG-390Lb, the first cool rocky/icy exoplanet, impressively demonstrated the sensitivity of the microlensing technique to extrasolar planets below $10\,M_\oplus$. A planet of $1\,M_\oplus$ instead of the expected $5\,M_\oplus$ for OGLE 2005-BLG-390Lb (with an uncertainty factor of 2) in the same spot would have provided a detectable deviation with an amplitude of ∼3 per cent and a duration of ∼12 h. While a standard sampling interval of 1.5–2.5 h for microlensing follow-up observations appears to be insufficient for characterizing such light curve anomalies and thereby claiming the discovery of the planets that caused these, an early detection of a deviation could trigger higher-cadence sampling which would have allowed the discovery of an Earth-mass planet in this case. Here, we describe the implementation of an automated anomaly detector, embedded into the eSTAR system, that profits from immediate feedback provided by the robotic telescopes that form the RoboNet-1.0 network. It went into operation for the 2007 microlensing observing season. As part of our discussion about an optimal strategy for planet detection, we shed some new light on whether concentrating on highly magnified events is promising and planets in the 'resonant' angular separation equal to the angular Einstein radius are revealed most easily. Given that sub-Neptune mass planets can be considered being common around the host stars probed by microlensing (preferentially M and K dwarfs), the higher number of events that can be monitored with a network of 2-m telescopes and the increased detection efficiency for planets below $5\,M_\oplus$ arising from an optimized strategy gives a common effort of current microlensing campaigns a fair chance to detect an Earth-mass planet (from the ground) ahead of the *COROT* or *Kepler* missions. The detection limit of gravitational microlensing extends even below $0.1\,M_\oplus$, but such planets are not very likely to be detected from current campaigns. However, these will be within the reach of high-cadence monitoring with a network of wide-field telescopes or a space-based telescope.

Key words: gravitational lensing – methods: observational – planetary systems.

1 INTRODUCTION

After Mao & Paczynski (1991) first pointed out that microlensing events can be used to infer the presence of extrasolar planets or

place limits on their abundance, this technique has now become established with several claimed detections (Bond et al. 2004; Udalski et al. 2005; Beaulieu et al. 2006; Gould et al. 2006). The discovery of OGLE 2005-BLG-390Lb (Beaulieu et al. 2006; Dominik, Horne & Bode 2006), estimated to be five times more massive than the Earth, with an uncertainty factor of 2, under the lead of the Probing Lensing Anomalies NETwork (PLANET)/RoboNet campaign demonstrated that microlensing not only can detect massive gas

[*]Royal Society University Research Fellow.
[†]E-mail: md35@st-andrews.ac.uk

giants, but can also detect planets that harbour a rocky/icy surface under a thin atmosphere. Moreover, it provided the first observational hint that cool rocky/icy planets are actually quite common, as previously predicted by simulations based on core-accretion models of planet formation (Ida & Lin 2005).

It was already estimated by Bennett & Rhie (1996) that there is a non-negligible chance of 1–2 per cent for detecting an Earth-mass planet located at about 2 au from its host star by means of observing a few per cent deviation in a microlensing light curve. However, such a discovery requires photometric measurements on a few hundred microlensing events, assuming that a fair fraction of the host stars are orbited by such planets.

A sufficient number of events can only arise from monitoring dense fields of stars. With a probability of $\sim 10^{-6}$ for a star in the Galactic bulge being magnified by more than 34 per cent at any given time due to the bending of light caused by the gravitational field of an intervening foreground star (Kiraga & Paczynski 1994), and such a microlensing event lasting of the order of a month, one namely needs to monitor 10^7–10^8 stars. This was achieved by microlensing surveys like the Optical Gravitational Lensing Experiment (OGLE) (Udalski et al. 1992), MAssive Compact Halo Objects (MACHO) (Alcock et al. 1993), Expérience de la Recherche d'Objets Sombres (EROS) (Aubourg et al. 1993) and Microlensing Observations in Astrophysics (MOA) (Muraki et al. 1999) with a roughly daily sampling. Moreover, all these surveys have been equipped with real-time alert systems (Udalski et al. 1994; Alcock et al. 1996; Bond et al. 2001; Glicenstein 2001; Udalski 2003) that notify the scientific community about ongoing microlensing events. This allows to schedule follow-up observations that provide an increased photometric accuracy, a denser event sampling, and/or coverage during epochs outside the target visibility from the telescope site used by the respective survey campaign.

The PLANET collaboration[1] established the first telescope network capable of round-the-clock nearly-continuous high-precision monitoring of microlensing events (Albrow et al. 1998) with the goal to detect gas giant planets and to determine their abundance. For being able to detect deviations of 5 per cent, PLANET aims at a 1–2 per cent photometric accuracy. With a typical sampling interval of 1.5–2.5 h allowing a characterization of planetary anomalies on the basis of at least 10–15 data points taken while these last, the required exposure time then limits the number of events that can be monitored. For bright (giant) stars, exposure times of a few minutes are sufficient, so that PLANET can monitor about 20 events each night or 75 events per observing season, but this reduces to about six events each night or 20 events per season for fainter stars, for which exposure times reach 20 min (Dominik et al. 2002). In 1999, MACHO and OGLE-II together provided about 100 microlensing alerts, out of which only seven were on giant source stars. This severely limited PLANET in its planet-detection capabilities: rather than 75 events, only about 25 could be monitored per season. The OGLE-III upgrade, in effect from 2002, had a major impact on the potential of microlensing planet searches, paving the way towards the now nearly 1000 microlensing events per year provided by the alert systems of the OGLE[2] and MOA[3] surveys. The much larger number of events arising from this upgrade allowed OGLE itself to obtain meaningful constraints on planets of Jupiter mass (Tsapras et al. 2003; Snodgrass, Horne & Tsapras 2004), while OGLE and

MOA have even demonstrated that such planets can in fact be detected by their surveys (Bond et al. 2004). However, for studying less-massive planets, their sampling is insufficient. At the same time, the OGLE-III upgrade enabled PLANET to exploit its full theoretical capability and, moreover, it gave PLANET a reliable chance to detect planets of a few Earth masses, provided that these are not rare around the stars that cause the microlensing events. The discovery of OGLE 2005-BLG-390Lb (Beaulieu et al. 2006; Dominik et al. 2006) explicitly proved the sensitivity of the PLANET observations to planets in that mass range.

Microlensing events are also regularly monitored by the Microlensing Follow-Up Network (MicroFUN) team.[4] However, rather than exploiting a permanent network, MicroFUN concentrates on particularly promising events and activates target-of-opportunity observations should such an event be in progress. Besides 1-m class telescopes, their stand-by network includes a larger number of small (down to 0.3 m diameter) telescopes operated by amateur astronomers, which are well suited to observe the peaks of events over which the source star makes a bright target.

Since the PLANET network is restricted in its capabilities of monitoring ~ 25 per cent of the currently alerted events with the observational requirements, the planet-detection rate could be boosted by using larger (2-m) telescopes or clusters of 1-m class telescopes. In fact, such an upgrade is required in order to obtain a sample that allows a reliable test of models of the formation and evolution of planets around K and M dwarfs. RoboNet-1.0[5] (Burgdorf et al. 2007) marks the prototype of a network of 2-m robotic telescopes, not only allowing for a fast response time, but also allowing for a flexible scheduling by means of the multi-agent contract model provided by the eSTAR project[6] (Allan, Naylor & Saunders 2006; Allan et al. 2006). eSTAR is a key player in the Heterogeneous Telescope Networks (HTN) consortium and involved in the International Virtual Observatory Alliance (IVOA) standards process.

If one aims at the discovery of Earth-mass planets, the standard follow-up sampling of 1.5 h usually does not produce the amount of data required to characterize the corresponding signals, and with less-frequent sampling one even faces a significant risk of missing any hint for a deviation from an ordinary microlensing light curve. However, planets of Earth mass and even below can be discovered by shortening the sampling interval to ~ 10 min once a regularly sampled point is suspected to depart from a model light curve that represents a system without planet. In order to properly trigger such anomaly alerts, all incoming data need to be checked immediately, and prompt action needs to be taken within less than ~ 15 min. The amount of data and the required response time for achieving a good detection efficiency for Earth-mass planets are, however, prohibitive for relying on human inspection. Therefore, we here describe the implementation of an automated anomaly detector that exploits the opportunities of immediate response and flexible scheduling of a network of robotic telescopes. A first similar warning system, dubbed EEWS, had been installed by OGLE in 2003 (Udalski 2003) which, however, involves further human inspection and operates with a single telescope. In contrast, our design needs to succeed without any human intervention and take care of a heterogeneous telescope network. The underlying algorithm follows previous experience on the assessment of anomalies. We explicitly aim at reaching a significant detection efficiency to Earth-mass

---

[1] http://planet.iap.fr.
[2] http://ogle.astrouw.edu.pl/ogle3/ews/ews.html.
[3] http://www.massey.ac.nz/∼iabond/alert/alert.html.

[4] http://www.astronomy.ohio-state.edu/∼microfun/.
[5] http://www.astro.livjm.ac.uk/RoboNet/.
[6] http://www.estar.org.uk.

planets with the current survey/follow-up strategy of microlensing planet searches.

This paper is organized as follows. In Section 2, we describe the modelling of ordinary microlensing events with particular emphasis on the importance of robust parameter estimates, not confused by outliers, in order to properly identify real deviations. While Section 3 deals with the general strategy for detecting low-mass planets by microlensing, we derive a suitable concept for an anomaly detector in Section 4. The embedding of the SIGNALMEN anomaly detector, that went into operation for the 2007 microlensing campaign, into the eSTAR project is discussed in Section 5, before its algorithm is described in Section 6. Section 7 then discusses the prospects of the SIGNALMEN anomaly detector for discovering planets of Earth mass and below. In Section 8, we provide a short summary and final conclusions.

Appendix A makes a point on the inability to detect planets at the resonant separation in some of the observed events.

2 ORDINARY LIGHT CURVES AND ANOMALIES

The bending of light due to the gravitational field of a foreground '*lens*' star with mass M at distance D_L causes an observed background '*source*' star at distance D_S to be magnified by (Einstein 1936):

$$A(u) = \frac{u^2 + 2}{u\sqrt{u^2 + 4}}, \tag{1}$$

if both objects are separated on the sky by the angle $u\theta_E$ with θ_E denoting the *angular Einstein radius*

$$\theta_E = \sqrt{\frac{4GM}{c^2}\left(D_L^{-1} - D_S^{-1}\right)}. \tag{2}$$

With the assumption that lens and source star move uniformly, where μ is the absolute value of their relative proper motion, the separation angle can be parametrized as

$$u(t) = \sqrt{u_0^2 + \left(\frac{t - t_0}{t_E}\right)^2}, \tag{3}$$

where u_0 denotes the closest approach at epoch t_0, and $t_E = \theta_E/\mu$ is a characteristic event time-scale.

Each set of observations with a specific telescope and filter comprises a data archive s of observed fluxes $F_i^{[s]}$ and their error bars $\sigma_{F_i^{[s]}}$ at epochs $t_i^{[s]}$. With the source flux $F_S^{[s]}$ and background flux $F_B^{[s]}$ depending on the data archive s, one observes symmetric light curves

$$F^{[s]}(t) = F_S^{[s]} A[u(t)] + F_B^{[s]} \tag{4}$$

peaking at t_0.

Estimates for $(t_0, t_E, u_0, F_S^{[s]}, F_B^{[s]})$ can then be obtained by minimizing

$$\chi^2 = \sum_{k=1}^{m} \sum_{i=1}^{n^{[k]}} \left(\frac{F^{[k]}(t) - F_i^{[k]}}{\sigma_{F_i^{[k]}}}\right)^2. \tag{5}$$

While we use the CERN library routine MINUIT for determining (t_0, t_E, u_0), the source and background fluxes $F_S^{[s]}$ and $F_B^{[s]}$ for any choice

of (t_0, t_E, u_0) simply follow from linear regression as

$$F_S = \frac{\sum \frac{A(t_i)F_i}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \sum \frac{A(t_i)}{\sigma_i^2} \sum \frac{F_i}{\sigma_i^2}}{\sum \frac{[A(t_i)]^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \left[\sum \frac{A(t_i)}{\sigma_i^2}\right]^2},$$

$$F_B = \frac{\sum \frac{[A(t_i)]^2}{\sigma_i^2} \sum \frac{F_i}{\sigma_i^2} - \sum \frac{A(t_i)}{\sigma_i^2} \sum \frac{A(t_i)F_i}{\sigma_i^2}}{\sum \frac{[A(t_i)]^2}{\sigma_i^2} \sum \frac{1}{\sigma_i^2} - \left[\sum \frac{A(t_i)}{\sigma_i^2}\right]^2}, \tag{6}$$

where the summations run from 1 to $n^{[k]}$, $\sigma_i \equiv \sigma_{F_i}$, and the index $[k]$ has been dropped. Any archive s can only be included if it contains at least 3 data points.

The characteristic form of the light curve described by equation (4) is based on the assumption that both source and lens star are single point-like objects that are moving uniformly with respect to each other as seen from the Earth. Apart from planets orbiting the lens star, significant deviations, so-called *anomalies* can, however, also be caused by binarity or multiplicity of lens or source, the finite angular size of the stars, or the revolution of the Earth (parallax effect).

Since it is our primary goal to detect light curve anomalies, it is essential to ensure that our adopted model is reasonably correct. However, frequently the data do not allow strong constraints to be placed on the model, in particular during early phases of the event. It is a well-known fact that OGLE announce a fair fraction of their events with the prediction of quite high peak magnification, whereas it turned out later that most of these peak at much lower magnifications. As studied in some detail by Albrow (2004), this is related to the fact that χ^2-minimization is equivalent to obtaining a maximum-likelihood estimate of the model parameters if the data are assumed to follow a Gaussian distribution, which is biased, that is, its expectation value does not coincide with the true expectation value of the considered quantity. Using the statistics of previously observed OGLE events, a Bayesian estimate that can be obtained by adding an effective penalty function to χ^2 comes closer to the expectation value (Albrow 2004). While the estimated value can be tuned by this, one does not fully get around the problem of large indeterminacy of the model parameters.

A further problem arises from the necessity to avoid that our model is driven towards data outliers. Otherwise, real anomalies would be missed while points matching an ordinary light curve would seem deviant. As a consequence, we would face the problem of not being able to distinguish between ongoing anomalies and further data requiring an adjustment of model parameters. Therefore, we apply a more sophisticated algorithm for estimating the model parameters that is rather invulnerable to outliers.

The model can be made to follow the bulk of the data by downweighting points according to their respective residual (e.g. Hoaglin, Mosteller & Tukey 1983) as follows. With the residuals

$$r_i^{[k]} = \frac{F^{[k]}(t) - F_i^{[k]}}{\sigma_{F_i^{[k]}}} \tag{7}$$

and the median of their absolute values $\tilde{r}^{[k]}$ for each data archive, we give further (bisquare) weight

$$w_i^{[k]} = \begin{cases} \left[1 - \left(\frac{r_i^{[k]}}{K\tilde{r}^{[k]}}\right)^2\right]^2 & \text{for } \left|r_i^{[k]}\right| < K\tilde{r}^{[k]} \\ 0 & \text{for } \left|r_i^{[k]}\right| \geqslant K\tilde{r}^{[k]} \end{cases} \tag{8}$$

to each data point, where we adopt $K = 6$ for the tuning constant. The choice of the weights, equation (8), means that data points

whose absolute residuals exceeds K times their median are ignored. This procedure is repeated until the formal χ^2 converges. However, we need to deal with non-linear models which are prone to several possible χ^2 minima. In contrast to linear models, it can therefore happen that this procedure leads to periodic switching between different minima, where nevertheless a subsequence converges to each of these. In this case, we have to live with the absence of a unique minimum and choose that one with the lowest χ^2. With the formal χ^2 not being dominated by outliers, we can also reliably adjust the relative weight between different data archives k after each iteration step, so that all $(\chi^2)^{[k]}/n^{[k]}$ coincide, preventing the estimation of model parameters being influenced by the collective over- or under-estimation of error bars.

3 DETECTION OF LOW-MASS PLANETS

It was pointed out by Mao & Paczynski (1991) that planets orbiting the lens star can reveal their existence by causing significant deviations to microlensing light curves. They also found that the probability to detect a planet becomes resonant if the angular separation from its host star is comparable to the angular Einstein radius θ_E, which reflects the fact that the detection of planets is aided by the tidal field of their host star. However, as pointed out in Appendix A, for a given event, in particular for larger impact parameters, the detection probability of smaller planets can actually drop to zero for angular separations close to θ_E rather than reaching a maximum. In such case, only slightly wider or closer separations can be probed. It is a lucky coincidence that the gravitational radius of stars and distances within the Milky Way combine in such a way that the angular Einstein radius converts to a projected separation $D_L\theta_E \sim 2$ au for $M = 0.3\,M_\odot$, the typical mass of the lens stars, assuming $D_S \sim 8.5$ kpc and $D_L \sim 6.5$ kpc. Gould & Loeb (1992) quantified the prospects for detecting planets from microlensing signatures by finding that Jupiter-mass planets distributed uniformly within angular separations $0.6\,\theta_E \leqslant d\theta_E \leqslant 1.6\,\theta_E$, comprising the so-called *lensing zone*, have a probability of 15 per cent of being detected among microlensing events with peak magnifications $A_0 \geqslant 1.34$, corresponding to the source entering the Einstein ring (of angular radius θ_E) of the lens star, that is, $u_0 \leqslant 1$. As shown by Griest & Safizadeh (1998), this probability increases significantly if one restricts the attention to events with larger peak magnifications, where about 80 per cent is reached for $A_0 \geqslant 10$. Since the area subtended on the sky by angular source positions that correspond to a significant deviation decreases towards smaller planet masses, both a shorter duration of the planetary signal and a smaller probability to observe it result. In contrast, the signal amplitude is only limited by the finite angular size of the source, where significant signal reductions start arising once it becomes comparable or larger than the size of the region for which a point source provides a significant deviation. However, Bennett & Rhie (1996) estimated that Earth-mass planets still have a 1–2 per cent chance of providing a signal in excess of a few per cent.

Planets around the lens star affect the light curve only by means of two dimensionless parameters, namely the planet-to-star mass ratio q and the separation parameter d, where $d\theta_E$ is the instantaneous angular separation of the planets from its host star (i.e. the lens star). With typical relative proper motions between lens and source stars of $\mu \sim 15\,\mu\text{as}\,\text{d}^{-1}$, microlensing events on Galactic bulge stars are usually observable for about a month or two, whereas planetary deviations last between a few hours and a few days, depending on the mass of the planet. In contrast to other indirect techniques, microlensing therefore obtains a snapshot measurement of the planet

rather than having to wait for it to complete its orbit. This gives microlensing the unique capability of probing planets in wide orbits whose periods otherwise easily exceed the lifetime of a project or its investigator.

With many events on offer from the OGLE and MOA surveys and only limited resources available for follow-up observations, one needs to make a choice which of these to monitor and how frequently to sample each event. With the goal to maximize the number of detections of planetary deviations, a prioritization algorithm that spreads the available observing time over the potential targets has been devised by Horne (in preparation), which forms a central engine of the RoboNet observing strategy. Any such strategy must be based on observables, model parameters arising from the collected data, or any other data statistics. As Horne (in preparation) pointed out, each data point carries a detection zone with it, composed of the angular positions for which a planet would have caused a detectable deviation. Unless finite-source effects begin diminishing the detectability of planets (Han 2007), detection zones grow with the current magnification. Moreover, the same photometric accuracy can be achieved with smaller exposure times for brighter targets. An efficient prioritization algorithm therefore needs to be based on both the current magnification and the brightness along with the time when the last observation was carried out, where taking into account the latter avoids obtaining redundant information. Such a prioritization of events, however, does not consider how well an observed deviation allows to constrain its nature of origin and it also assumes that the model parameters of the ordinary light curve are known exactly.

If the effect on the microlensing light curve is dominated by a single planet, the lens system can be fairly approximated as a binary system consisting of the star and this planet. Gravitational lensing by a binary point-mass lens has been studied in great detail for equal masses by Schneider & Weiß (1986) and later generalized for arbitrary mass ratios by Erdl & Schneider (1993). On the other hand, Chang & Refsdal (1979) have discussed lensing by bodies of different mass scales. While their target of interest was the brightness variation of individual images of quasi-stellar objects that are gravitationally lensed by an intervening galaxy, a very similar situation arises for planets orbiting a lens star. Similarly to individual stars in the galaxy splitting an image due to lensing by the galaxy as a whole into 'microlensing', a planet can further split one of the two images due to lensing by its host star if it roughly coincides in angular position with that image. Dominik (1999) has further investigated the transition towards extreme mass ratios and shown how the case described by Chang & Refsdal (1979), the so-called *Chang–Refsdal lens*, is approached. The derived expansions into series have later been used by Bozza (1999) for discussing the case of multiple planets. Binary lenses in general and planetary systems in particular create a system of extended caustics, consisting of the angular positions for which a point-like source star would be infinitely magnified. While sufficiently small sources passing the caustics can provide quite spectacular signals, planets are more likely to already reveal their existence on entering a much larger region surrounding these.

For less-massive planets, there are usually two separate regions for positions of the source star that lead to detectable planetary signals, which are related to two types of caustics. Only if the angular separation of the planet from its host star is in a close vicinity to the angular Einstein radius θ_E, where the corresponding range is broader for more massive planets, a single caustic results and these regions merge. Otherwise, there are one or two *planetary caustics* which are located around positions for which bending of its light due

to the gravitational field of the lens star causes the source to have an image at the position of the planet, and a *central caustic* which can be found near the lens star (Griest & Safizadeh 1998; Dominik 1999). As Bozza (1999) demonstrated, the planetary caustics associated with different planets are almost always separated and any kind of interference between these is quite unlikely. In contrast, Gaudi, Naber & Sackett (1998) pointed out that the central caustic is always affected by the combined action of all planets. However, it is likely, although not guaranteed, that there is a hierarchical order among the effects of different planets, so that a linear superposition is a fair approximation (Rattenbury et al. 2002; Han 2005).

While the absence of any deviations near the peak of extreme highly magnified ordinary events that are related to the source potentially approaching the central caustic poses strict limits on the abundance of low-mass planets (Abe et al. 2004; Dong et al. 2006), their actual discovery from this kind of deviations suffers from several complications. While the linear size of the detection region around planetary caustics scales with the square root of the planet mass, it is proportional to the planet mass itself for the central caustic (Chang & Refsdal 1979; Griest & Safizadeh 1998; Dominik 1999; Chung et al. 2005; Han 2006). Therefore, the finite angular size of the source star is more likely to cause a significant reduction of the signal amplitude. Moreover, the characterization of the nature of origin for such deviations is significantly more difficult than for deviations related to planetary caustics. The latter provide further information by means of the time elapsed between the peak of the background ordinary light curve and the deviation, whereas central-caustic deviations involve a higher degree of model degeneracies with more prominent finite-source and parallax effects. In any case, a promising sensitivity to Earth-mass planets is only reached for lens–source impact parameters $u_0 \lesssim 5 \times 10^{-4}$, which occur at a rate of less than one per year.

On the other hand, the non-negligible probability to detect planetary signals if the source passes in the vicinity of planetary caustics offers a fair chance of detecting a planet of Earth mass by also making use of the large number of events that exhibit lower magnifications at a given time. Given these facts, it is not a surprise that the first sub-Neptune mass planet whose existence could be reported on the basis of microlensing observations, OGLE 2005-BLG-390Lb (Beaulieu et al. 2006), produced a 15 to 20 per cent signal at a magnification $A \sim 1.3$ about 10 d after an observed peak at magnification $A_0 \sim 3$ (see Fig. 1) rather than a deviation within a highly magnified peak.

While the mass of OGLE 2005-BLG-390Lb is about $5\,M_\oplus$, uncertain to about a factor of 2 (Dominik 2006), a planet of $1\,M_\oplus$ in the same spot would still have produced a signal with an amplitude of $\sim$3 per cent, lasting $\sim$12 h rather than about twice that long. The actual sampling would have been insufficient for discovering such a planet in this configuration, but the situation would have been different had our sampling interval been decreased to 10–15 min on the suspicion of a first deviation. This case explicitly shows how an anomaly detector can help us in not missing short-lasting small deviations (related to low-mass planets). By requiring an initial sampling that is just dense enough for an ongoing anomaly being alerted before most of it has passed, it moreover allows to monitor a sufficient number of events for providing a reasonable number of planet discoveries. The main gain of the anomaly detector will indeed be achieved for detecting planets from perturbations related to planetary caustics at lower and moderate magnification, whereas a high-cadence sampling can already be scheduled a priori for (predictable) high magnifications without the need for any further alert.
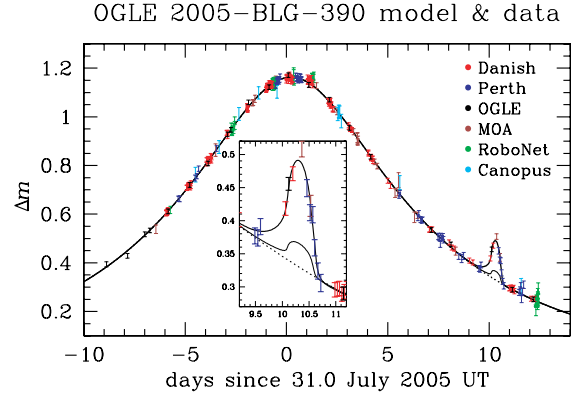

Figure 1. Model light curve of microlensing event OGLE 2005-BLG-390 along with data taken with the Danish 1.54 m at ESO LaSilla (Chile), red, the Perth 0.6 m (Western Australia), blue, and the Canopus 1.0 m (Tasmania), cyan, by PLANET, the Faulkes North 2.0 m (Hawaii), green, by RoboNet-1.0, the OGLE 1.3 m (Chile), black, and the MOA 0.6 m (New Zealand), brown, where $\Delta m = 2.5 \log A(t)$ has been plotted along with $m_i = 2.5 \log A_i$. The $\sim$15 per cent deviation lasting about a day revealed the existence of a planet with $m \sim 5.5\,M_\oplus$ (uncertain to a factor of 2), while an Earth-mass planet in the same spot would have caused a 3 per cent deviation lasting about 12 h (thin line). The time-scale of this event is $t_E = 11.0$ d, while $d = 1.610$ and $q = 7.6 \times 10^{-5}$. Moreover, $u_0 = 0.359, t_0 = 31.231$ 2005 July UT, and the angle between the vector from the planet to its host star and the source trajectory is $\alpha = 157°.9$, where the less centre of mass is to the right-hand side. Finally, the source star moves by its own radius relative to the lens within $t_\star = 0.282$ d. The dotted line refers to a model light curve in the absence of a planet.

The ability of detecting an anomaly depends on how well earlier data constrain the model describing an ordinary light curve. For large model parameter uncertainties, it becomes hard to distinguish a real deviation from a necessary model revision due to a previous misestimate, for which χ^2 adjustments are not a reliable indicator due to the intricate parameter space and poor knowledge about the measurement uncertainties. Therefore, the anomaly detection is more efficient after the peak of a microlensing has passed rather than prior to it (cf. Udalski 2003), where the ability is particularly vulnerable to data gaps. Thus, if the increased detection efficiency for low-mass planets that is achieved by means of the anomaly detector is a relevant goal for a monitoring strategy, it is sensible to give preference to events past peak over those pre peak for comparable magnifications. Although it is more difficult to decide whether a deviation from a previous model is real or due to a model misestimate if constraints on its parameters are weaker, it is more likely that a suspected deviation occurs and is reported. This has the by-effect that more data will be collected in this case, which in turn strengthens the model parameter constraints. Despite the fact that the higher magnification around the peak allows for accurate data being taken with shorter exposure times, the weak constraints on the position of the peak make it rather difficult to detect an ongoing anomaly there, unless the peak region is monitored quite densely and no data gaps occur.

4 CONCEPT FOR AN ANOMALY DETECTOR

If reported data deviate from the expected light curve, this could either mean that there is a real effect, the deviation could be of statistical nature, or the data could simply be erratic by any means. It is therefore impossible to arrive at an appropriate judgement about the presence of anomalies on the basis of a single deviating data point.

However, such a point should raise suspicion that an anomaly is indeed ongoing. Our anomaly detector, dubbed SIGNALMEN, profits from the fact that real-time photometry and robotic telescope operation allow immediate feedback. Rather than having to rely on a fixed sampling rate for a given event, we can request prompt further observations once the modelling of incoming data indicates a deviation from an ordinary light curve.

Based on the collected data, the anomaly detector can arrive at one out of the following three possible conclusions and assign a corresponding status to the event:

(i) there is no ongoing anomaly (ordinary);
(ii) there is an ongoing anomaly (anomaly); or
(iii) not sure what is going on (check).

While the last option, corresponding to a suspected, unconfirmed anomaly, does not look appealing at first sight, it actually marks the strength of the feedback concept. In this case, we urgently request further observations on the same target, thereby providing the anomaly detector with further data on which it can base the decision in subsequent runs. In a 'recheck and repeat' strategy, data whose absolute model residual is among the largest trigger further observations, and this process is repeated until a decision about whether there is an anomaly can be taken with the desired significance.

The art of optimizing an anomaly detector is in finding the appropriate balance between not missing planetary anomalies and avoiding false alerts. The availability of immediate feedback opens the possibility of using a rather low initial trigger level on the first suspicion of an anomaly, which gives us a fair chance of detecting low-amplitude anomalies at an early stage. The early detection is a vital feature for being able to discover Earth-mass planets. In contrast, we do not care that much about the detection of anomalies that have already been missed or are mostly over. A low initial trigger, however, means that we will need to spend a significant amount of time on collecting evidence against the presence of an anomaly if the point that triggered the 'check' observations does not constitute a real deviation. As pointed out in more detail in the following section, we aim at rechecking 5 per cent of the incoming data for anomalous behaviour, while about four to five further points are expected to be required for providing sufficient evidence against. This means that we spend about 20 per cent of our observing time on checking potential anomalies. By basing the criterion for a significant deviation on a comparison of the model residual of the tested data point with those of earlier data, we pay respect to the fact that the true scatter of data is not properly reflected by the size of the reported error bars and can be non-Gaussian.

We also account for the fact that data collected with different telescopes may arrive in blocks rather than point by point and not necessarily in time-sequence. Moreover, all data are subject to change, which not only means that reported (F_i, σ_{F_i}) might alter between two runs of the anomaly detector, but also means that data at certain epochs might disappear, whereas additional data at other epochs prior to the most recent data point might be released. By not making any distinction between whether 'new' data are released in a block or arise from recent point-by-point observations, we also take care of the possibility that an anomaly is already apparent in the latest data update.

Our robust fitting scheme is rather powerful in identifying outliers and therefore gives us some protection against failures of the real-time photometry and weird results that might be the consequence. We have implemented a further test for distinguishing between havoc photometry and ongoing anomalies which produces an alert urging to check the data reduction. However, there is no way getting around the point that the capabilities of the anomaly detector will rise or fall with the quality of the real-time data analysis. In principle, one can also investigate correlations with observing conditions such as the reported seeing or sky brightness. However, such information may not be provided for all considered sites, so that we try to avoid relying on it as long as possible.

5 ANOMALY DETECTOR EMBEDDING AND EXTERNAL INTERFACES

The intelligent-agent architecture of the eSTAR project constitutes the harness inside which the SIGNALMEN anomaly detector operates. Thereby, it provides autonomous decision-making by means of software, which allows to build systems that learn and adapt. The eSTAR system provides the feedback loop by feeding the SIGNALMEN anomaly detector with real-time data, which then replies with an expert opinion that allows the eSTAR system to solve the distributed-scheduling problem of how to distribute follow-up requests over the network in order to maximize the chances of detecting and characterizing an extrasolar planet.

The eSTAR project serves as a meta-network between existing proprietary robotic telescope networks built upon a peer-to-peer agent based architecture (Wooldridge 2002), which cuts across traditional notions that running such a network requires a 'master scheduler'. Instead, eSTAR can be viewed as a collaborative multi-agent system using a contract model. The crucial architectural distinction of such a system is that both the software controlling the science programme and those embedded at the telescope acting as a high-level interface to the native telescope control software are equally seen as 'agents'. A negotiation takes place between these agents in which each of the telescopes bids to carry out the work, with the user's agent scheduling the work with the agent embedded at the telescope that promises to return the best result. This preserves the autonomy of individual telescope operators to implement scheduling of observations at their facility as they see fit, and offers adaptability in the face of asynchronously arriving data. For instance, an agent working autonomously of the user can change, reschedule, or cancel queries, workflows or follow-up observations based on new information received. The eSTAR architecture represents a 'turn-key' system for autonomous observations of transient events, and therefore is ideal for microlensing follow-up.

The agents are also capable of responding in real time to external alerts (White et al. 2006; Williams & Seaman 2006), so-called Virtual Observatory Events (VOEvents).[7] While OGLE and MOA alerts are being translated into this format, the detection of an anomaly by SIGNALMEN will also be reported by means of a VO-Event.

Besides the communication by means of software agents, the design of the SIGNALMEN anomaly detector also contains interfaces for output to human observers and upload of data provided by any other observing campaign. Currently, data from PLANET, OGLE, MOA and MicroFUN are fed in. Moreover, we will keep two separate mailing lists for notification on the decision in favour of an ongoing anomaly ('anomaly' status) and on the detection of deviant points ('check' status), which everyone is free to subscribe to. While dense follow-up by other teams is much encouraged in this case, the 'check' status will be invoked frequently (several times each night) and mainly serves to steer the internal feedback with the

---

[7] http://www.voevent.org/.

robotic telescopes of the RoboNet network and in second instance with the other telescopes involved in the PLANET/RoboNet campaign. In addition to providing real-time notification of suspected or ongoing anomalies, we will publish up-to-the-minute plots showing collected data along with a model light curve, whose parameters have been determined by the SIGNALMEN anomaly detector.

On the suspicion of an anomaly, a fast response with further observations is crucial for either confirming or rejecting this hypothesis. While robotic telescopes can react almost instantaneously, human observers need to be informed by e-mail or other means of communication, which adds some delay. Only if an observatory is staffed and the observer frequently monitors incoming e-mail, the feedback loop can be closed. This works reasonably well with the current PLANET network, where observers are present at the telescope on each night with suitable weather during the observing season. However, telescopes that are only activated on a target-of-opportunity basis, such as several of those used by MicroFUN, might miss the short-notice call. In any case, the success of the strategy is limited by the need to find out whether a suspected anomaly is present or not with the use of telescopes that have already monitored the microlensing event of interest. The value of data from other sites is limited to providing early useful data if it turns out that an anomaly is ongoing, but these contain rather little information about whether the light curve deviates.

While so far, we have implemented an algorithm that alerts us on suspected or ongoing anomalies, it neither gives us a recommendation of the best anomaly sampling interval, for which we simply assume an initial choice of 10 min, nor does it inform us when the anomaly is over and we can return to the standard follow-up sampling rate. Both of these issues currently need to be dealt with by human interaction through an internal webpage automatically listing events that are considered to deviate from ordinary light curves.

6 THE ANOMALY DETECTOR ALGORITHM

6.1 Basics, data statistics, and deviations

The implementation of the SIGNALMEN anomaly detector described in the following is a first sketch, matching the primary requirements. It involves some basic statistical tests, building upon prior experience. More sophisticated tests can be designed and added, should it turn out that these yield significant improvements on the decision process. During the 2007 season, SIGNALMEN will log all incoming data, the anomaly indicators, current model parameters, and its decisions, which will provide a valuable basis for further tuning. Our algorithm involves several constants that can be adjusted. Their values can be changed by editing a configuration file rather than requiring alteration of the source code itself. In the following, we list our default setting in brackets.

With the source and background fluxes, $F_S^{[s]}$ and $F_B^{[s]}$, depending on the data archive s, residuals need to be compared by means of the magnifications

$$A_i = \frac{F_i - F_B^{[s(i)]}}{F_S^{[s(i)]}} \tag{9}$$

rather than the measured fluxes F_i, where the uncertainties of A_i are given by

$$\sigma_{A_i} = \frac{\sigma_{F_i}}{|F_S^{[s(i)]}|}. \tag{10}$$

In general, the reported error bars σ_{F_i} are not a proper reflection of the true scatter, which moreover frequently deviates from a Gaussian distribution. In particular, data provided by OGLE come with severely underestimated photometric uncertainties for $I \leqslant 15$, whereas these are about the right size for $15 \leqslant I \leqslant 18$ and overestimates for faint targets $I \geqslant 18$. One of the sources of this behaviour is that the photometric reduction packages usually do not take into account further systematic uncertainties. We therefore correct for this fact by adding a systematic error SYST_ERR (0.003) in quadrature to the uncertainty of the reported magnitude. Moreover, rather than relying on σ_{F_i}, we assess the scatter by means of two statistics, namely the median scatter $\hat{\delta}^{[s]}$ and the critical scatter $\delta_{\rm crit}^{[s]}$. By calculating the residuals

$$\delta_k = \frac{A(t) - A_k}{\sigma_{A_k}} \tag{11}$$

for each archive s and sorting the $n^{[s]}$ values $(\delta_k^{[s]})^2$ in ascending order, we find

$$\hat{\delta}^{[s]} = \begin{cases} \left[\left(\delta_{(n^{[s]}+1)/2}^{[s]} \right)^2 \right]^{1/2} & \text{for} \quad n^{[s]} \text{ odd} \\ \left\{ \frac{1}{2} \left[\left(\delta_{n^{[s]}/2}^{[s]} \right)^2 + \left(\delta_{n^{[s]}/2+1}^{[s]} \right)^2 \right] \right\}^{1/2}, \\ \qquad\qquad\qquad \text{for} \quad n^{[s]} \text{ even} \end{cases} \tag{12}$$

and with the critical real index

$$\nu^{[s]} = n^{[s]} (1 - \text{DEV\_PERC}) + 1, \tag{13}$$

we determine

$$\delta_{\rm crit}^{[s]} = \begin{cases} \left[(\delta_{n^{[s]}})^2 \right]^{1/2} & \text{for} \quad \nu^{[s]} \geqslant n^{[s]} \\ \left[\beta^{[s]} \left(\delta_{\lceil \nu^{[s]} \rceil} \right)^2 + \left(1 - \beta^{[s]} \right) \left(\delta_{\lfloor \nu^{[s]} \rfloor} \right)^2 \right]^{1/2}, \\ \qquad\qquad\qquad \text{for} \quad \nu^{[s]} < n^{[s]} \end{cases} \tag{14}$$

where $\beta^{[s]} = \nu^{[s]} - \lfloor \nu^{[s]} \rfloor$, DEV_PERC (0.05) denotes the probability for an absolute deviation in excess of $\delta_{\rm crit}^{[s]}$, and $n^{[s]}$ is the number of data points for archive s. With a deviation threshold DEV_SIG (2), we require for a significant deviation both

$$|A_j - A(t_j)| > \text{DEV\_SIG} \, \sigma_{A_j} \max \left\{ 1, \hat{\delta}^{[s(j)]} \right\} \tag{15}$$

and

$$|A_j - A(t_j)| > \sigma_{A_j} \delta_{\rm crit}^{[s(i)]} \tag{16}$$

to hold. For Gaussian errors bars, both conditions can be made to coincide. In order to allow for a proper evaluation of the scatter, we require that at least MIN_DATA_TEST (6) data points and data from at least MIN_NIGHTS (2) previous nights have been collected. Otherwise, the statistical scatter is likely to be underestimated and therefore false alerts are almost certain.

With our robust-fitting algorithm that downweighs or even ignores outliers and the fact that we rely on the median scatter and trigger on the absolute residual exceeding that of a fixed percentage of data, we are well able to distinguish between low-quality data and real deviations. In particular, this allows us to achieve a low false alert rate. The requirement of obtaining significant data statistics before assessing deviations comes at the price of some inability to identify deviations in fast-rising events with high-magnification peak. However, this does not significantly affect the planet-detection prospects, since a high-cadence sampling will be carried out for these events irrespective of suspected anomalies in the data.

6.2 Data sequence and modelling

SIGNALMEN assumes that events do not exhibit anomalies at the time these are first announced by the OGLE or MOA microlensing

surveys. For each data archive we keep track of the latest collected data point and restart our assessment for anomalies at the epoch t_{new} corresponding to the earliest 'new' point among all archives. In order to assess the data point by point, we sort these in time-sequence and step through points $k \leqslant n$ with $t_k \geqslant t_{new}$, where n is the index of the most recently observed data point. For each event, we store the time-ranges for which anomalies were considered to be ongoing, and the parts of these ranges prior to t_{new} are then excluded from fits for models of an ordinary light curve. Moreover, on each run of SIGNALMEN on a specific event, we note the epoch $t_c \geqslant t_{new}$ for which an ongoing anomaly was first suspected, and administrate a list of all points l with $t_l \geqslant t_c$ that were found to deviate, which form the current anomaly sequence. When considering all data with $t \leqslant t_k$, the deviation of a point with index j ($t_c \leqslant t_j \leqslant t_k$) can be determined with respect to the following models that include all data with indices i that fulfill the following.

  (i) *'Previous'*. $t_i < t_k$, exclude data within an anomaly time-range prior to t_{new} or in the current anomaly sequence;
  (ii) *'Current'*. $t_i \leqslant t_k$, exclude data within an anomaly time-range prior to t_{new} or in the current anomaly sequence;
  (iii) *'All non-deviant'*. $t_i \leqslant t_k$, exclude data within an anomaly time-range prior to t_{new} or in the current anomaly sequence, but include $i = j$; and
  (iv) *'All-data'*: $t_i \leqslant t_k$, exclude data within an anomaly time-range prior to t_{new}.

If there is no current anomaly sequence, that is, none of the points $k \leqslant n$ has been found to deviate so far, the *'all-data'* and *'all non-deviant'* models coincide with the *'current'* model. Since model predictions can be expected to fail frequently, our initial assessment of a deviation is solely based on the *'current'* model, which includes the latest considered point k. Should this point fail to deviate significantly by means of the conditions given by equations (15) and (16), the *'current'* model becomes the *'previous'* model and k is increased. Otherwise, $t_c \equiv t_k$ and data point k is added to the current anomaly sequence. While the *'previous'* model is retained, it also becomes the *'all non-deviant'* model, whereas the *'current model'* also becomes the *'all-data'* model. For increased k, further tests will be performed for data j ($t_c \leqslant t_j \leqslant t_k$).

6.3 Anomalies: accept or reject?

If a current anomaly sequence has been found, SIGNALMEN will try to figure out whether further data points provide evidence in favour of an ongoing anomaly or against it, leading to finishing up with 'anomaly' or 'ordinary' status. If the current data do not allow to arrive at either of these conclusions, the 'check' status is invoked. In this case, the markers for the latest data points for each of the archives are set so that the current anomaly sequence is reassessed on the next run of SIGNALMEN. This avoids the necessity to store further information about this sequence and also easily allows for a potential revision of these critical data in the meantime.

    Data taken after t_c that are found not to deviate significantly from the *'current'* model can provide evidence against the presence of an ongoing anomaly. However, simply counting the number of non-deviant points is not a viable option since these might have larger uncertainties than the deviant points. This happens in particular if later data originate from different sites, while even for the same site it cannot be guaranteed that the same photometric uncertainty can be retained. Since data with large scatter and therefore no indication of an anomaly must not be used as evidence against, it is unavoidable that the photometric uncertainties are taken into account. Moreover,

we also need some characteristic for the amplitude of the anomaly which we would like to decide about whether it is real or not. Let us consider the fractional deviation

$$\varepsilon_i = \frac{A_i - A(t_i)}{A(t_i)}, \tag{17}$$

and for a deviant point l define ε_l as the anomaly level. With $\sigma_{\varepsilon_j} = (\sigma_{A_j} \max\{1, \hat{\delta}^{[s(j)]}\})/A(t_j)$, we then obtain the weighted average over all non-deviating points j after the deviant point (i.e. $t_j > t_l$)

$$\overline{\varepsilon} = \frac{\sum \frac{\varepsilon_j}{\sigma_{\varepsilon_j}^2}}{\sum \frac{1}{\sigma_{\varepsilon_j}^2}}. \tag{18}$$

and its standard deviation

$$\sigma_{\overline{\varepsilon}} = \left(\sum \frac{1}{\sigma_{\varepsilon_j}^2} \right)^{-1/2}. \tag{19}$$

The anomaly is then called off if

$$\overline{\varepsilon} < \varepsilon_l/2 - \text{REJECT\_SIG } \sigma_{\overline{\varepsilon}} \qquad (\text{for} \quad \varepsilon_l > 0)$$
$$\overline{\varepsilon} > \varepsilon_l/2 + \text{REJECT\_SIG } \sigma_{\overline{\varepsilon}} \qquad (\text{for} \quad \varepsilon_l < 0) \tag{20}$$

with a default setting REJECT_SIG $= 1$ and the additional requirement that at least MINPTS_REJECT (4) non-deviant points have been collected. For Gaussian data with constant error bars, we find the anomaly call-off typically not requiring more than five measurements. However, this can take significantly longer if only data with large effective error bars (corrected for actual scatter) can be acquired.

    If the data point k has been found not to deviate, we also reassess the current anomaly sequence with respect to the *'all non-deviant'* model. If an anomaly cannot be confirmed or discarded, just testing points in sequence against the *'current'* model can either lead to missed anomalies or lead to false alerts if the model is not well constrained. We therefore determine the residuals with respect to a model that includes all points found not deviating (and their scatter). This also allows taking into account an increased scatter present in more recent data. Previously deviant data that do not fulfill the new criterion are removed from the current anomaly sequence, which might lead to a revision of t_c and leave SIGNALMEN with an empty current anomaly sequence. In the latter case, SIGNALMEN will continue as if no deviant points were found in the current run. We also require that all data points in the current anomaly sequence deviate to the same side. Therefore, it will be shortened if necessary to meet this condition.

    Similarly, if the most recently considered data point k is found to deviate to the opposite site as the previous data, a new current anomaly sequence is started at $t_c \equiv t_k$ and the previous sequence is abandoned.

    A stronger hint for an anomaly being ongoing is obtained if the data point k deviates to the same side as the previous points in the current anomaly sequence. Once the current anomaly sequence contains at least two data points, we start testing the collected data against an *'all-data' model*, which also contains the points in the current anomaly sequence. With the earlier tests we avoided that the model of an ordinary event is driven towards points that deviate from it, which allows us to call off an anomaly if further points follow an ordinary light curve without getting confused by outliers. However, we also need to take care of the fact that more weight than just that of a single point might be needed to correct for a bad earlier estimate of model parameters. As a compromise, we adopt less-strict criteria, namely that the residuals of the last MINPTS_ANOMALY

(5) points are all of the same sign and at least MINPTS_ALL_ANOM (3) points deviate significantly. If earlier data in the current anomaly sequence cannot match these criteria, the sequence is shortened and t_c is revised.

A further test is based on the overlap between the points in the current anomaly sequence and non-deviant points falling in between. With the *'all-data'* model, we determine

$$\delta A_i = A_i - A(t_i). \tag{21}$$

If for a non-deviant point j following a deviant point l for which $\delta A_l > 0$, one finds

$$\delta A_j + 2\sigma_{A_j} \max \left\{ 1, \hat{\delta}^{[s(j)]} \right\} < \delta A_l - 2\sigma_{A_l} \max \left\{ 1, \hat{\delta}^{[s(l)]} \right\} \tag{22}$$

or the equivalent relation to hold for the subsequent deviant point, the non-deviant point is considered to contradict point l deviating, which is therefore removed from the current anomaly sequence. For $\delta A_l < 0$,[8] the corresponding condition reads

$$\delta A_j - 2\sigma_{A_j} \max \left\{ 1, \hat{\delta}^{[s(j)]} \right\} > \delta A_l + 2\sigma_{A_l} \max \left\{ 1, \hat{\delta}^{[s(l)]} \right\}. \tag{23}$$

Finally, we realize that the photometric reduction might fail occasionally and produce weird results. A common characteristic that can be distinguished from real anomalous behaviour are sudden changes between a rise and fall. We therefore determine the pattern of significant increase or decrease of the magnification amongst the data in the current anomaly sequence. Should there be more than one change in direction, SIGNALMEN abstains from the claim that an anomaly is ongoing. This 'zig-zag test' is only used as the final criterion once all other conditions for an ongoing anomaly are fulfilled. For two deviant points l and $m > l$, a significant increase is characterized by

$$\delta A_m - 2\sigma_{A_m} \max \left\{ 1, \hat{\delta}^{[s(m)]} \right\} > \delta A_l + 2\sigma_{A_l} \max \left\{ 1, \hat{\delta}^{[s(l)]} \right\}, \tag{24}$$

whereas a significant decrease arises by exchanging l and m. If there is no significant change between neighbouring points, a significant increase is assessed with respect to the lowest of these points while a significant decrease refers to the highest of these.

To summarize, SIGNALMEN concludes that there is an ongoing anomaly if all of the following criteria are satisfied.

(i) The anomaly is not called off by means of a series of at least MINPTS_REJECT (4) non-deviant points with a weighted-average fractional deviation significantly [measured by REJECT_SIG (1.0)] closer to zero than half of the fractional deviation of the previous deviant point;

(ii) the most recent deviant points form a sequence of at least MINPTS_ANOMALY (5) points that were found to deviate to the same side from the *'current'* model and the *'all non-deviant'* model;

(iii) the residuals with respect to the *'all-data'* model of at least the last MINPTS_ANOMALY (5) points in the current anomaly are all of the same sign;

(iv) at least MINPTS_ALL_ANOM (3) points in the current anomaly sequence deviate from the *'all-data'* model;

(v) there are no non-deviant data points in between those in the current anomaly sequence that significantly fail to overlap with them; and

(vi) data in the current anomaly sequence do not change more than once between a significant increase and decrease.

If these criteria are fulfilled for $k = n$, that is, at the end of the collected data, SIGNALMEN activates the 'anomaly' mode. Should these

be fulfilled earlier ($k < n$) only, SIGNALMEN finishes with 'ordinary' status, but a file notifying about a missed anomaly is written. If just the zig-zag test fails, SIGNALMEN notifies about problems with the photometric reduction and suspends evaluation of data archives for which recent data showed more than one change in direction in the suspected anomaly sequence. Such a case needs human intervention and should be dealt with at high priority.

7 PROSPECTS WITH THE ANOMALY DETECTOR

In order to demonstrate what can be achieved with the anomaly detector, let us use the event OGLE 2005-BLG-390, which already allowed us to detect a planet of $5\,M_\oplus$ (with a factor of 2 uncertainty), as an illustrative example and starting point of the
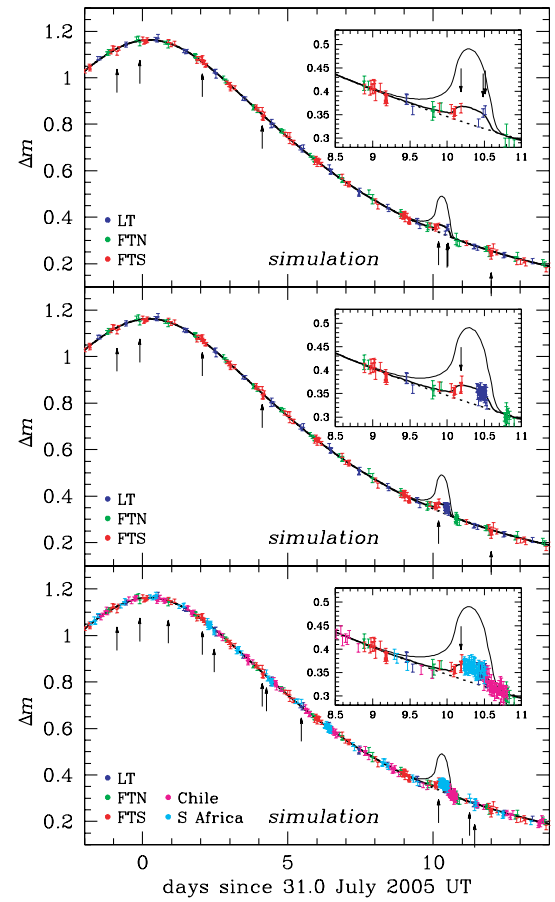

Figure 2. Detection of a hypothetical Earth-mass planet in microlensing event OGLE 2005-BLG-390 located in the same spot as OGLE 2005-BLG-390Lb with the RoboNet telescope network. Simulated data for the different telescopes are shown in different colours along with theoretical light curves, where a thin solid line corresponds to the actually derived model for that event (with OGLE 2005-BLG-390Lb), a bold solid line to a model with an Earth-mass planet, and a bold dotted line to a model without planet. The arrows mark data points that have been found to deviate from the best-fitting model available at that epoch. While the top panel shows only data collected with the standard sampling, the middle and bottom panel include further data with high-cadence (10 min) sampling after having triggered on the anomaly. For the top and middle panels, only the current RoboNet-1.0 telescopes have been considered, whereas for the bottom panel, the availability of two additional similar telescopes in Chile and South Africa has been assumed.

---

[8] Obviously, there is no $\delta A_l = 0$ case.

discussion. Fig. 2 shows the model light curve for the corresponding configuration again, where the planet OGLE 2005-BLG-390Lb has been replaced by a 1-$M_\oplus$ version in the same spot, but rather than the collected data, we now show simulated data related to the three robotic 2-m telescopes that currently comprise the RoboNet-1.0 network: the Liverpool Telescope (LT), the Faulkes Telescope North (FTN), and the Faulkes Telescope South (FTS). According to the target observability from the different sites at the event coordinates [RA = $17^\mathrm{h}54^\mathrm{m}19\overset{\mathrm{s}}{.}19$, Dec. = $-30\overset{\circ}{.}22\overset{\prime}{.}38\overset{\prime\prime}{.}3$ (J2000)], requiring that the target is at least $20°$ above the horizon, synthetic data have been generated where the average sampling interval is $\Delta t = (2\,\mathrm{h})/\sqrt{A}$ and the photometric accuracy is 1 per cent at baseline and smaller as the event brightens, following photon noise statistics, where Gaussian errors have been assumed. A systematic error of 0.5 per cent has been added in quadrature. For the time of the next observation, a Gaussian fluctuation of 20 per cent of the sampling interval has been adopted, and while its photometric uncertainty itself fluctuates by 12.5 per cent. Moreover, a drop-out probability of 5 per cent on each data point has been assumed.

While it would have been rather easy to detect and characterize a planet like OGLE 2005-BLG-390Lb, the standard sampling would have given rather little evidence for an Earth-mass planet in the same spot, and a characterization would not have been possible. The arrows in Fig. 2 indicate data points that were found to deviate by the anomaly detector, given the best-fitting model that could have been obtained at that time, based on all previous data. Further data after the first four trigger points would have indicated that there is no ongoing anomaly, but the deviation is rather of statistical nature. In contrast, subsequent data points after the first trigger point that fall on to the real anomaly would have confirmed the deviation and finally led to the activation of high-cadence anomaly monitoring. This example, however, also shows a critical weakness of the current RoboNet-1.0 network, namely its lack of round-the-clock coverage. In particular, one sees that the southern telescope offers a much longer time-window for our purpose than either of the northern telescopes. Just after the opportunity of taking a further point after activation of the 'check' status, the target could not have been followed anymore with the FTN and it would have been necessary to wait for the LT for acquiring subsequent measurements. This demonstrates that provision of a fast response also implies the availability of a telescope at the requested time. Further telescopes available in South Africa and Chile (see Fig. 2) would have allowed a coverage of the anomaly sufficient to detect an Earth-mass planet.

The detection of OGLE 2005-BLG-390Lb was eased by the source being a giant star with a radius $R_\star \sim 9.6\,\mathrm{R}_\odot$, which not only allowed to obtain an accurate photometry with rather short exposures but also increased the probability of detecting a significant deviation. While the large angular size led to a reduction of the amplitude, which did not matter because it remained at the 15–20 per cent level, a larger range of orientations or impact parameters than for a smaller source would have created on observable signal. Moreover, the duration of the planetary deviation is dominated by the source star moving by its angular size, giving a rather comfortable time-span, which would still have been $\sim 12\,\mathrm{h}$ for an Earth-mass planet. While a main-sequence star could have provided a signal with larger amplitude, the probability to observe it would have been smaller and it would not have lasted for that long.

If one replaces the source star of OGLE 2005-BLG-390 with an eight times smaller version, an Earth-mass planet in the same spot as OGLE 2005-BLG-390Lb would become undetectable since the smaller source would not enter a region for which significant deviations result. For the configuration shown in Fig. 3, the angle of the
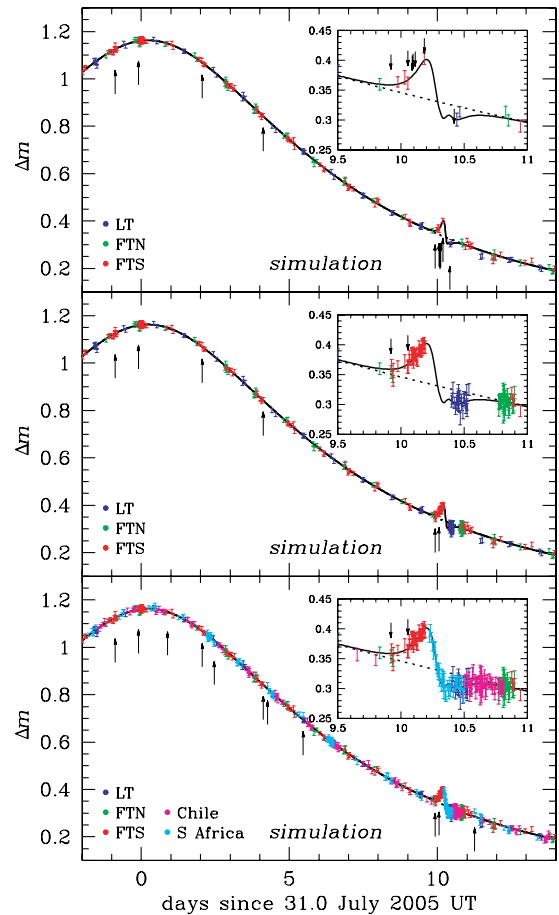

OGLE 2005-BLG-390 → MS star, $m = 1\,M_\oplus$

Figure 3. Detection of a hypothetical Earth-mass planet in an event resembling OGLE 2005-BLG-390 at the same separation as OGLE 2005-BLG-390Lb, but for a main-sequence source star with $R_\star \sim 1.2\,\mathrm{R}_\odot$. In fact, the source size has been assumed to be 8 times smaller than for the giant observed in OGLE 2005-BLG-390. Since with the smaller source, the planet is missed for the original angle $\alpha = 157\overset{\circ}{.}9$ (Beaulieu et al. 2006) between the line from the planet to its host star and the source trajectory, where the lens centre of mass is to the right, a slightly different angle $\alpha = 158\overset{\circ}{.}2$ has been adopted, for which a 5 per cent deviation results. Otherwise, this figure is analogous to Fig. 2.

source trajectory relative to the planet–star axis has therefore been slightly adjusted, resulting in a 5 per cent deviation. Achieving good photometry on the fainter target is more difficult and requires longer exposure times. Nevertheless, PLANET has demonstrated a photometric accuracy of even less than 0.5 per cent on a main-sequence star is possible provided that it is fairly isolated rather than in a crowded area. While for the previously discussed case involving a giant source star, signal amplitudes significantly exceeding 3 per cent cannot result, the shown 5 per cent deviation is not even near the limit for main-sequence stars, for which very strong signatures become possible should the source happen to cross a caustic. One also sees that the duration of the planetary deviation has not decreased by a factor of 8 as the source size did. Contrary to the giant source star case, the signal duration is now roughly given by the time in which a point source passes the region of angular positions that lead to significant deviations, and remains $\sim 8\,\mathrm{h}$ for the prominent peak. The angular size of the source star itself is reflected in the small peak within the time-span over which the brightness in

presence of the planet is smaller than without. As before with a giant source star, the proper characterization of the planetary anomaly is not possible with the standard 2-h sampling, while high-cadence sampling after having suspected or detected an anomaly will solve the problem, provided that telescopes are available to observe the target. Interestingly, in a very early stage of the anomaly, one of the data points appears to be higher just by chance. Further data taken at the sampling interval of 10 min, however, do not confirm a significant deviation, so that only after the next data point taken with the standard sampling rate the high-cadence anomaly monitoring remains active.

After having found that the discovery of Earth-mass planets does not constitute the limit of what can be achieved with microlensing survey/follow-up campaigns equipped with an automated anomaly detector, let us look into how far one can go. In fact, the rather large separation $d \sim 1.6$ of OGLE 2005-BLG-390Lb from its host star did not offer a very fortunate configuration. Let us therefore also consider $d \sim 1.25$ and see how the signal amplitude and duration are affected. As Fig. 4 shows, even for a planet with mass $m = 0.1$ $M_\oplus$ located at 1.25 θ_E from its host in an OGLE 2005-BLG-390-like event, a signal of 10 per cent lasting about 2.5 h can result on a main-sequence source star. The early detection of such a short signal with a standard survey sampling interval of 1–2 h and the anomaly monitoring each become challenging. In fact, the rather short time gap of about 40 min between the FTN in Australia and a telescope in South Africa is sufficient for missing the crucial falling part of the planetary anomaly in our simulation. This also demonstrates the extraordinary value of a telescope at the western edge of Australia and/or in southern central Asia. Nevertheless, the discovery of planets with masses of even $0.1\,M_\oplus$ or below by ground-based microlensing campaigns remains a matter of probability rather than possibility.

8 SUMMARY AND CONCLUSIONS

For probing models of planet formation and evolution and thereby taking an important step towards our understanding of the origin of living organisms as we know them, microlensing will remain a competitive and complementary technique with respect to other methods for the foreseeable future with its unique sensitivity to low-mass planets in wider orbits. It is unlikely (although not impossible) that microlensing will provide a timely discovery of a planet on which conditions similar to those on the Earth can exist that are known to support the formation of life forms, with less than 3 per cent of planets in any mass range expected to orbit suitable host stars at suitable radii (Park et al. 2006). While both transit and radial-velocity surveys approach the required orbital range for such habitable planets from closer orbits, essentially all planets that can be expected to be detected by microlensing reside in wider orbits. The discovery of the first extrasolar planets already demonstrated impressively how little one can infer about the origin of the Solar system if the study remains restricted to itself. Similarly, one should not expect that a study just of habitable planets will allow to arrive at a well-understood picture of their formation. Instead, a reliable test of theories should involve data spanning over a wider and surrounding region. Moreover, microlensing allows to obtain a planet sample not only around stars in the Galactic disc but also around those in the Galactic bulge, thereby probing two distinct populations.

We have shown that our SIGNALMEN anomaly detector, which went into operation for the 2007 microlensing observing season, allows to adopt a three-step strategy of survey, follow-up, and anomaly monitoring. The basis of this strategy is formed by the microlens-
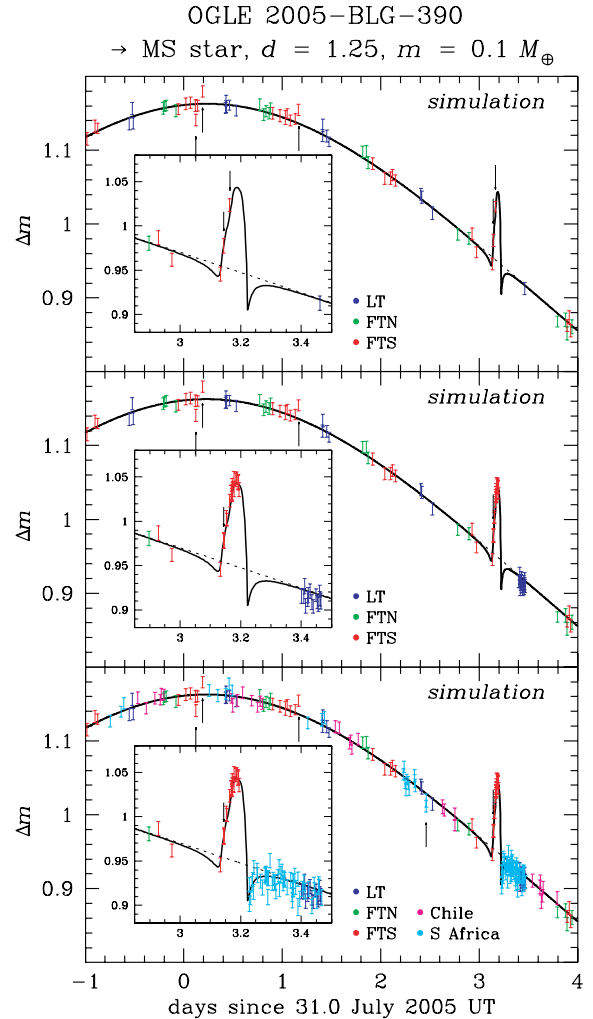

Figure 4. Coverage of a microlensing event resembling OGLE 2005-BLG-390 with a main-sequence source star ($R_\star \sim 1.2\,R_\odot$) instead of the 8 times larger giant, and a planet of $0.1\,M_\oplus$ at an angular separation of $d = 1.25$ times the angular Einstein radius θ_E. The angle between planet-to-star axis and source trajectory, with the lens star to the right, has been chosen as $\alpha = 126°6$, so that the source crosses the planetary caustic. As for Fig. 2, the trigger points of the anomaly detector are indicated by the arrows, and the middle and bottom panels include the high-cadence follow-up after anomaly suspicion or detection. While the top and middle panels only show data corresponding to the current RoboNet-1.0 network, the availability of further similar telescopes in Chile and South Africa has been assumed for the bottom panel.

ing events provided in real time by the OGLE and MOA surveys out of which a sufficient number are then monitored with a network of 2-m telescopes so that deviations due to Earth-mass planets are unlikely to be missed and a reasonable number of low-mass planets is expected be detected over the next few years. In particular, it is only required that the follow-up observations provide an early enough trigger of an ongoing anomaly, whereas a proper characterization need not be ensured by these, because this will be achieved by the high-cadence anomaly monitoring after an anomaly has been suspected or detected. In fact, the use of an anomaly detector becomes quite efficient if many points are required for proper characterization of a signal rather than being able to claim a detection from a single deviant point. The expected detections will

provide a powerful test of models of planet formation and evolution around K and M dwarfs. While planets of Earth mass appear to have some particular appeal, they do not provide the hard limit for ground-based microlensing searches. As shown by one of the examples discussed in Section 7, the anomaly detector allows us to go even below $0.1\,M_\oplus$, although such detections are not very likely to occur. However, these are reasonable goals for a network of ground-based wide-field telescopes or a space-based telescope (Bennett et al. 2004).

By only checking for significant deviations from an ordinary microlensing light curve, our current anomaly detector is blind to the nature of origin of the observed deviation. While it is ensured that planetary deviations due to planets of Earth mass and even below can be detected, more than 90 per cent of all deviations are due to other causes, such as finite-source effects, an Earth–Sun parallax, or stellar lens or source binaries. In order to distinguish these from planetary deviations and at the same time to obtain appropriate estimates for the urgency and frequency of second-level follow-up observations, a full real-time modelling taking into account all these effects would be required. Optimally, the prioritization of events would follow the expected constraints on the planet characteristics rather than just maximizing their detectability while ignoring the chances of properly characterizing a potential planet. We plan to implement such a system in the near future.

ACKNOWLEDGMENTS

REFERENCES

Abe F. et al., 2004, Sci, 305, 1264
Albrow M. D., 2004, ApJ, 607, 821
Albrow M. D. et al., 1998, ApJ, 509, 687
Alcock C. et al., 1993, Nat, 365, 621
Alcock C. et al., 1996, ApJ, 463, L67
Allan A. et al., 2006, in Lewis H., Bridger A., eds. Proc. SPIE Vol. 6274, Advanced Software and Control for Astronomy SPIE, Bellingham WA, p. 8
Allan A., Naylor T., Saunders E. S., 2006, Astronomische Nachrichten, 327, 767
Aubourg E. et al., 1993, Nat, 365, 623
Beaulieu J.-P. et al., 2006, Nat, 439, 437
Bennett D. P., Rhie S. H., 1996, ApJ, 472, 660
Bennett D. P. et al., 2004, in Mather J. C., ed., Proc. SPIE Vol. 5487, Optical, Infrared, and Millimeter Space Telescopes. SPIE, Bellingham WA, p. 1453
Bond I. A. et al., 2001, MNRAS, 327, 868
Bond I. A. et al., 2004, ApJ, 606, L155
Bozza V., 1999, A&A, 348, 311

Burgdorf M. J., Bramich D. M., Dominik M., Bode M. F., Horne K. D., Steele I. A., Tsapras Y., 2007, P&SS, 55, 582
Chang K., Refsdal S., 1979, Nat, 282, 561
Chung S.-J. et al., 2005, ApJ, 630, 535
Dominik M. et al., 2002, P&SS, 50, 299
Dominik M., Horne K., Bode M. F., 2006, Astron. Geophys., 47, 25
Dominik M., 1999, A&A, 349, 108
Dominik M., 2006, MNRAS, 367, 669
Dong S. et al., 2006, ApJ, 642, 842
Einstein A., 1936, Sci, 84, 506
Erdl H., Schneider P., 1993, A&A, 268, 453
Gaudi B. S., Sackett P. D., 2000, ApJ, 528, 56
Gaudi B. S., Naber R. M., Sackett P. D., 1998, ApJ, 502, L33
Gaudi B. S. et al., 2002, ApJ, 566, 463
Glicenstein J.-F., 2001, in Menzies J. W., Sackett P. D., eds, ASP Conf. Ser. Vol. 239, Microlensing 2000: A New Era of Microlensing Astrophysics The EROS Microlensing Alert System. Astron. Soc. Pac., San Francisco, p. 28
Gould A., Loeb A., 1992, ApJ, 396, 104
Gould A. et al., 2006, ApJ, 644, L37
Griest K., Safizadeh N., 1998, ApJ, 500, 37
Han C., 2005, ApJ, 629, 1102
Han C., 2006, ApJ, 638, 1080
Han C., 2007, ApJ, 661, 1202
Hoaglin D. C., Mosteller F., Tukey J. W., eds, 1983, Understanding Robust and Exploratory Data Analysis. Wiley, New York
Ida S., Lin D. N. C., 2005, ApJ, 626, 1045
Kiraga M., Paczynski B., 1994, ApJ, 430, L101
Mao S., Paczynski B., 1991, ApJ, 374, L37
Muraki Y. et al., 1999, Prog. Theor. Phys. Suppl., 133, 233
Park B.-G., Jeon Y.-B., Lee C.-U., Han C., 2006, ApJ, 643, 1233
Rattenbury N. J., Bond I. A., Skuljan J., Yock P. C. M., 2002, MNRAS, 335, 159
Schneider P., Weiß A., 1986, A&A, 164, 237
Snodgrass C., Horne K., Tsapras Y., 2004, MNRAS, 351, 967
Tsapras Y., Horne K., Kane S., Carson R., 2003, MNRAS, 343, 1131
Udalski A., 2003, Acta Astron., 53, 291
Udalski A. et al., 1992, Acta Astron., 42, 253
Udalski A. et al., 1994, Acta Astron., 44, 227
Udalski A. et al., 2005, ApJ, 628, L109
White R. R. et al., 2006, Astronomische Nachrichten, 327, 775
Williams R. D., Seaman R., 2006, in Gabriel C., Arviset C., Ponz D., Enrique S., eds, ASP Conf. Ser. Vol. 351, VOEvent: Information Infrastructure for Real-Time Astronomy. Astron. Soc. Pac., San Francisco, p. 637
Wooldridge M., 2002, An Introduction to Multi-agent Systems. John Wiley & Sons, Chichester

APPENDIX A: THE INABILITY TO DETECT PLANETS AT THE RESONANT SEPARATION IN SOME EVENTS

In general, angular positions for the source star relative to the lens composed of a star and its planet, for which a significant deviation in the observed light curve as compared to a lens star without planet results, form regions around the caustics of the lens system. For point sources, these regions grow as the angular separation parameter d, where $d\theta_E$ gives the angular separation of the planet from its host star, approaches unity. This indeed implies, as Mao & Paczynski (1991) pointed out, that planets around $d \sim 1$ are most easily detected among all events that occur. However, this does not imply that this also holds for each of the observed events.

Finite-source effects may cause an increase of the detection efficiency by means of the larger source catching more easily a region of significant deviation without bringing the signal amplitude below the detection threshold, or these can lead to a reduction, in particular if the finite source subtends regions corresponding to deviations of

opposite signs. In fact, Gaudi & Sackett (2000) found that a reduction of the detection efficiency most prominently affects separations around the resonant $d \sim 1$, where the wide-separation side suffers more than the close-separation side.

In any case, there is already another effect that prohibits the detection of less-massive planets around $d \sim 1$. Given that the central caustic is found at the position of the lens star and the centre of the planetary caustics is separated by $|d - 1/d|$ from it, the regions for which a significant deviation results might fall inside a circle whose radius is given by the impact parameter u_0 of a given event. In this case, these cannot be traced by the source trajectory. For small planet-to-star mass ratios q, the critical separations approach (cf. Gaudi et al. 2002)

$$d_\pm = \frac{1}{2}\left(\sqrt{u_0^2 + 4} \pm u_0\right) \tag{A1}$$

$$\simeq 1 \pm \frac{u_0}{2} \quad (u_0 \ll 1), \tag{A2}$$

while for larger q, an increasingly larger size of the caustics and the associated regions of significant deviations allows to enter the range $d \in (d_-, d_+)$ further and further. The fact that for larger lens–source separations, less-massive planets cannot be detected in increasingly broader regions around the angular Einstein radius θ_E is apparent in the examples showing the detection efficiency as a function of the separation parameter d for a few choices of the planet-to-star mass ratio in figs 5 and 8 provided by Gaudi & Sackett (2000), but unfortunately not discussed there.

Table A1 shows the critical separations for selected values of u_0. In particular, (low-mass) planets within the lensing zone can only be detected in events with $u_0 \leqslant 1$ (corresponding to $A_0 \geqslant 1.34$).

Table A1. Forbidden regions for the planet separation as a function of the event impact parameter.

u_0	d_-	d_+
1.5	0.5	2
1.0	0.62	1.62
0.7	0.71	1.41
0.5	0.78	1.28
0.4	0.82	1.22
0.3	0.86	1.16
0.2	0.90	1.10
0.1	0.95	1.05

Notes. $d_\pm$ are the critical separations for which the centre of the planetary caustic(s) falls inside a circle of radius u_0 around the host star. For sufficiently small mass ratios, this will prevent planets with separation parameters in the range (d_-, d_+) from being detected in events with impact parameter u_0.

Given these findings and the fact that a planet actually did reveal its existence in the event OGLE 2005-BLG-390, it is less surprising that with $u_0 = 0.359$, its angular separation from its host star in units of θ_E is the rather large $d = 1.61$, whereas a detection at, for example, $d = 1.1$ would have been impossible.

This paper has been typeset from a TEX/LATEX file prepared by the author.