

Crowded Field Photometry and Difference Imaging

Przemysław Woźniak*

Space Science and Applications, ISR-1

Los Alamos National Laboratory

MS-D466

Los Alamos, New Mexico, 87545 USA

E-mail: wozniak@lanl.gov

The purpose of this lecture is to introduce the basic methodology that enables massive high precision photometric monitoring of crowded fields in microlensing surveys. Some background is given on the development of techniques and tools that ultimately led to the adoption of the image differencing approach in a variety of projects in optical astronomy requiring time-series photometry. The effects of field crowding on photometry of individual sources and statistical properties of surveys are summarized. The discussion of photometric techniques starts from Point Spread Function (PSF) fitting in conventional crowded field codes, but the main focus is on difference imaging. The Alard & Lupton algorithm for image subtraction and PSF matching based on convolution is the corner stone of photometric data pipelines in the current generation microlensing surveys. The presentation covers a detailed description of the basic algorithm and its extension to kernels with strong gradients. To the extent possible, practical advice and performance comparisons are given to help implementers and users of the method.

The Manchester Microlensing Conference: The 12th International Conference and ANGLES Microlensing Workshop

January 21-25 2008

Manchester, UK

*Speaker.

1. Introduction

Much of the motivation to develop numerical techniques and software that can handle object crowding in astronomical images comes from the study of gravitational microlensing [23], typically confined to extremely dense stellar fields in the Local Group. Within that context one is particularly interested in detecting variability (mostly flux changes, but also motions). This naturally drove the methodology of microlensing surveys toward relative measurements and ultimately pixel-by-pixel image differencing. The relative emphasis has shifted away from treating the photometry of a crowded image as a collection of individual object measurements toward a global decomposition of the image flux into variable and constant parts, sometimes referred to as AC/DC. This approach—to a large extent pioneered and developed into a useful tool by the microlensing community—is highly relevant and increasingly popular in other astronomical applications. The obvious examples are the spin-off projects of the microlensing surveys themselves such as investigations of stellar variability in partially resolved populations, [29, 6], structure studies of the Milky Way and nearby galaxies [30, 5], planet detection using microlensing anomalies [9, 17] and the transit method [16, 21], and long term monitoring of microlensing in strongly lensed quasars [40, 39, 35]. Crowding is also an intrinsic problem in photometry of explosive transients on cosmological scales (Supernovae, Gamma-Ray Bursts) detected against a complicated (but essentially static) extended profile of the host galaxy [10, 22].

The exercise set accompanying this lecture introduces a simple software package for Difference Image Analysis. It shows a complete example of how to build high S/N light curves of crowded fields starting from a sequence of flat-fielded images. The input data utilized in these exercises are actual OGLE-II scans of the BUL_SC3 field¹, one of the densest regions of the Galactic bulge covered by the OGLE-II microlensing survey. The software presented here is my own implementation of the Alard & Lupton algorithm for optimal image subtraction [1, 2] with several supporting programs required for a real life project. A modified version of this package is currently running in standard data pipelines of the OGLE project. Another modified version, the DIAPL package² by Wojtek Pych provides many usability and algorithmic improvements on my original code. It also comes with a fairly complete manual.

2. Astronomical image formation and pixel sampling

The main points that justify the numerical treatment are summarized without going into unnecessary details. Starting from the “true” image $O(x, y)$ that would be observed in the absence of Earth’s atmosphere, the combined effect of the air turbulence, the imaging optics, and the telescope tracking errors can be modeled as a convolution with seeing $S(x, y)$:

$$O * S(x, y) \equiv \int O(u, v) S(x - u, y - v) du dv \quad (2.1)$$

The signal is then convolved with the pixel response function $P(x, y)$ and sampled at regular intervals, i.e. multiplied by a forest of equally spaced 2-D delta functions III_2 :

¹All imaging data used in this lecture are part of the OGLE-II survey: courtesy of Prof. Andrzej Udalski from Warsaw University Observatory.

²<http://users.camk.edu.pl/pych/DIAPL>

$$\text{III}_2(x, y) \equiv \sum_{m, n} \delta_2(x - m, y - n) = \sum_{m, n} \delta(x - m) \delta(y - n). \quad (2.2)$$

For a point source $O(x, y) = \delta_2(x, y) = \delta(x) \delta(y)$ the result is the Point Spread Function (PSF):

$$S * P(x, y) \quad (2.3)$$

or

$$\text{PSF}(x, y) \equiv \sum_{m, n} S \otimes P(m, n) \delta_2(x - m, y - n) \quad (2.4)$$

after sampling. The properties of $\text{III}_2(x, y)$ allow switching back and forth between the integral sign and summation over arrays of pixels, and similarly, between convolving continuous functions $*$ and pixel meshes \otimes .

A band-limited function is defined to have a Fourier transform that is zero everywhere except inside a finite range of frequencies. In other words, very high frequencies are not present in the signal. This is equivalent to the existence of a cutoff frequency f_c such that the Fourier spectrum is zero for $|f| > f_c$. Band-limited functions have a remarkable property with respect to sampling. The sampling theorem states that a function with the cutoff frequency f_c is fully specified by its values taken at equal intervals (samples), provided that the sampling rate is higher than $2f_c$, or the spacing between the samples is less than $1/(2f_c)$. For a strictly band-limited function $F(x)$ it is possible to find an interpolation formula that reproduces the exact value of F at any x , as long as under-sampling is avoided.

Astronomical telescopes have a finite resolution and deliver approximately band-limited images. A useful rule of thumb for the marginal sampling rate is 2.5 pix/FWHM, where FWHM is the Full Width at Half Maximum of the PSF. Under-sampling causes aliasing in the frequency spectrum and complicates interpolation. With properly sampled images the exact shape of the pixel response $P(x, y)$ is not important (normally assumed to be a top hat), and there is little difference between integrating over a pixel and taking the value at the center. The OGLE-II camera is a good example of an efficient design with 0.4'' pixels and 1.3'' FWHM median seeing, rarely reaching below 1.0''.

3. Effects of object crowding in microlensing surveys

The stellar fields monitored by microlensing surveys are extremely crowded (Figure 1). When the surface number density reaches $\sim 0.1/\text{FWHM}^2$, overlaps between the PSF profiles of individual stars and flux blending are common. It is virtually impossible to describe all systematic effects associated with PSF estimation, source detection, and photometry/astrometry in crowded fields. The details depend on specific data sets and image processing tools. A comprehensive study of blending related systematics can be found in [28]. A detailed discussion of parameter estimation in automated analysis of crowded fields was published in [19]. Here, we only highlight the most important factors that routinely occur in microlensing work.

1. Magnitude bias near detection threshold

The detection threshold in severely crowded images is set by the confusion limit and changes from one location to another in response to fluctuations in density. The background level is set by merging PSF wings and faint cores. Even for relatively bright stars that still produce a local peak of intensity, the information about the wings of the profile is scrambled and parameter estimation is effectively based on the inner PSF core. A PSF model that systematically underpredicts the flux in stellar wings will result in overestimated background levels. The same discrepancy in background subtraction that hardly affects bright stars can produce a large magnitude offset in faint stars and introduce a strong apparent non-linearity in the photometry near the detection threshold.

2. Imperfect deblending

The level of crowding typical for microlensing surveys will eventually overwhelm the ability of the deblending algorithm to separate correctly very close and/or faint objects (cf. section 4). A fraction of fainter stars effectively disappears within a combined PSF footprint of brighter stars, making them appear even brighter and their profiles slightly wider. This happens across the magnitude range, but the details depend on the local slope of the Luminosity Function (LF), the size of the seeing disk, and the number density of stars. The PSF fitting codes also have a tendency to produce halos of spurious faint sources around bright stars due to residuals left behind an imperfect modeling of stellar wings.

3. PSF modeling and Luminosity Function

The failure to deblend a fraction of stars causes an effective broadening of the PSF model estimated from bright stars and changes the observed shape of the LF (Figure 2). The general trend is a magnitude dependent shift toward the bright end of the distribution and a rapid turn-over near the confusion limit.

4. Parameter degeneracy

Commonly used estimators of background, object centroid, and flux for individual sources become very noisy and may be biased. This significantly worsens the impact of the well known parameter degeneracy in microlensing light curve models [37, 18] that ultimately propagates to uncertainties in event time-scales and optical depth measurements.

5. Centroid bias and centroid motion for variable objects

The mean centroid of light for a source at \mathbf{r}_s and a blend at \mathbf{r}_b representing the effective influence of all contributing blends is:

$$\mathbf{r}(t) = \frac{\mathbf{r}_s F_s A(t) + \mathbf{r}_b F_b}{F_s A(t) + F_b}, \quad (3.1)$$

where F_s is the baseline flux of the source, F_b is the flux of the blend, and $A(t)$ is the magnification. The total baseline flux is $F_0 = F_s + F_b$, and setting $f_s = F_s/F_0$, $1 - f_s = F_b/F_0$ we have:

$$\mathbf{r}(t) = \frac{\mathbf{r}_s f_s A(t) + \mathbf{r}_b (1 - f_s)}{A(t) f_s + (1 - f_s)}. \quad (3.2)$$

The effective centroid at baseline is then:

$$\mathbf{r}_0 = \mathbf{r}_s f_s + \mathbf{r}_b (1 - f_s) \quad (3.3)$$

and the centroid offset from the baseline value is:

$$\Delta \mathbf{r} = \frac{\mathbf{r}_s f_s A(t) f_s + \mathbf{r}_b (1 - f_s)}{f_s A(t) + (1 - f_s)} - \mathbf{r}_0 = (\mathbf{r}_s - \mathbf{r}_b) \frac{f_s (1 - f_s) (A(t) - 1)}{f_s A(t) + (1 - f_s)} \quad (3.4)$$

Brightness changes in blended objects produce correlated centroid motions that may significantly affect the PSF photometry in fixed position mode. In a typical OGLE-II Galactic bulge field (~ 150 stars brighter than $I = 17$ mag per square arcmin and $1.3''$ FWHM seeing) there will be about 30% of all detectable events and 10% of bright events with $\Delta r > 0.2''$, corresponding to a centroid motion up to $1/2$ pix. Out of 15 OGLE-I microlensing events, 7 showed a noticeable correlation between the light centroid and magnification [15].

6. Time scale bias in microlensing events

For a blended event with $f_s < 1$, fitting a standard point mass, point source microlensing light curve model that assumes no blending ($f_s \equiv 1$) overestimates the impact parameter u_0 and, more importantly, underestimates the time-scale t_E (Figure 3). The time-scale distribution derived that way is systematically shifted toward lower t_E and develops a tail of events with short apparent durations.

7. Systematics of microlensing optical depth

The most frequently used optical depth estimator is $\tau \sim N^{-1} \sum_i t_{E,i} / \varepsilon(t_{E,i})$, where N is the total number of monitored stars and t_E is the Einstein ring crossing time. If blending is neglected, one might expect a large systematic bias in the optical depth from underestimated time-scales. In practice, it turns out that for a magnitude limited event sample, the bias in t_E toward lower values is closely compensated by a corresponding underestimate in the number of stars N (Figure 4).

8. Blending in bright stars

Strongly blended events are harder to detect because they are “diluted” by contributions to the total light that are not magnified. It may seem intuitive that blending should not be a significant factor for events that occur in apparently bright sources standing out in the LF. In reality, blending in bright stars is common and for most purposes must be taken into account. In the OGLE-II galactic bulge sample of events with baselines brighter than $I_{RC} \sim 15.6$ mag designed to select Red Clump Giants, the fraction of sources actually below the magnitude threshold is $\sim 38\%$ [32].

9. Negative blending

Occasionally, light curve models that account for blending turn out the best fit source fraction $f_s > 1$, i.e. the microlensed flux exceeds the total flux of the source and all contributing blends. Negative blend contributions are clearly an artifact. Simulations of fully synthetic images that approach the complexity of the actual CCD frames collected by the OGLE-II survey [28] link the main cause of these anomalies to the performance limit of the deblending routine in the presence of noise. Another explanation is over-subtracted background for a star that coincides with a negative fluctuation in the number density of blends, and sits in a “hole” in the background [24].

4. Conventional PSF fitting

Photometric codes that reduce images into catalogs of sources must take care of a considerable number of details. There is a trade-off between the level of sophistication and speed that is particularly felt in massive surveys such as microlensing searches. Similarly, the expected accuracy must be balanced against the range of applicability. In the generic case of arbitrary object profiles (e.g. spiral galaxies) field crowding severely complicates image segmentation, i.e. estimating how many objects are present and where. For the purpose of modeling the images of partially resolved stellar populations, however, it is safe to assume that the signal consists of point sources (PSF profiles) superimposed on a smooth background. Table 1 illustrates major issues involved in PSF photometry of crowded fields and two possible ways of addressing them. There are many excellent photometric tools available to the astronomical community, each with their specific strengths [31, 27, 11]. Historically, the DOPHOT software [27] played a particularly important role in microlensing as the basic number crunching tool adopted early on by the MACHO and OGLE surveys.

The number density of stars typical for microlensing work is so high that the actual number of sources detected by the deblending algorithm becomes very sensitive to seeing. The standard technique to control this and stabilize the photometry on a time series of images with variable seeing is to reduce the number of free parameters by fixing the number of objects and their centroids. A hint to the algorithm about the approximate object magnitudes further accelerates and improves convergence. In practice, the best prior information available for each field is the object catalog obtained from a regular DOPHOT run on a single image with exceptionally good seeing, the so called “template”. Using bright and relatively isolated stars the template positions and magnitudes are transformed to the reference frame of each individual exposure and serve to prime the computation. The “warm start” and the fixed position mode turned out to be truly enabling features and their introduction in DOPHOT was stimulated by microlensing searches.

The complexities of PSF fitting in crowded fields are greatly reduced if the perturbing influence of the neighboring stars can be removed. Direct image differencing is a very productive approach to time-series photometry of variable objects (Figure 5). Table 2 is intended to give a rough idea of the development of concepts and implementations leading from classic PSF fitting codes to a wide adoption of the image differencing method in microlensing searches and other sky monitoring surveys.

	DAOPHOT	DOPHOT
PSF model	Empirical PSF (analytical model fit + sub-sampled table of residuals)	Analytical PSF (pseudo-Gaussian)
PSF gradient	Originally fixed PSF shape, then 1st order variation with a weighted sum of 3 fixed PSFs, then ...	Originally fixed PSF shape, then 2-D polynomial fit for each shape parameter using an ensemble of stars
Background estimator	Local background estimates based on a large pixel annulus (mode)	Local sky level fitted for each object, then a global model for the ensemble
Detection	Convolve with a lowered Gaussian filter and identifies local intensity peaks	Finds local intensity peaks between a pair of progressively fainter flux thresholds
Pixel value	Integrates PSF over square pixels	Evaluates PSF at each pixel
Deblending	Examines significance and flux contributions of stars in PSF group	Classifies extendedness, goodness of fit test with two-PSF model
Algorithm	Simultaneous fitting of relatively isolated and self-contained groups of stars	Iterative fitting and subtraction of progressively fainter stars with parameter refinement
Optimization	Linearized least squares fit with non-linear model	Non-linear least squares
Warm starts	Sufficiently modular to enable warm starts	Optional warm starts and fixed position mode

Table 1: Comparison of two approaches to profile fitting photometry of point sources in crowded fields.

5. Image differencing based on PSF deconvolution with Fourier division

In principle, a PSF-matching kernel for a given pair of images can be calculated by deconvolving the PSF of one image from the other. Taking a Fourier transform maps the convolution in data space into multiplication of transforms. If $P_i(x, y)$ is the PSF of a program image, and $P_r(x, y)$ is the PSF of the reference image to be subtracted, we can write for the kernel $K(x, y)$:

$$K = FT^{-1} \left(\frac{FT(P_i)}{FT(P_r)} \right), \quad (5.1)$$

1987	DAOPHOT (Stetson et al.) [31]
1992	OGLE and MACHO surveys, modified DOPHOT [34, 3]
1993	DOPHOT (Schechter, Mateo & Saha) [27]
1996	Pixel lensing with Fourier Division (Tomaney & Crotts) [33] PEIDA software for EROS (Ansari) [7]
1998	Robust global subtraction algorithm (Alard & Lupton) [1]
1999	MACHO DIA analysis (Alcock et al.) [4]
2000	Extension of AL algorithm to variable kernels (Alard) [2] ISIS package (Alard) cdophot (Reid, Sullivan, & Dodd) [26]
2001	OGLE DIA package (Woźniak) [38]
2002	DIA based standard OGLE and MOA pipelines [36, 12]
2005	DIAPL extensions/modifications (Pych)
2008-?	DIA pipelines in LSST, Pan-STARRS [20, 14]

Table 2: Approximate development timeline of techniques and tools for crowded field photometry.

where FT and FT^{-1} denote a Fourier transform and its inverse. Unfortunately, this simple formula has several shortcomings: 1) in crowded fields PSF is ill defined 2) results rely on availability of isolated stars and estimating the PSF of each image 3) the method requires very high S/N input data 4) in the presence of noise, the algorithm is unstable and there is no good way to enforce that the end result makes sense 5) noise dominates the PSF wings, where the game is 6) results are very sensitive to under-sampling and aliasing 7) handling spatially variable solutions and finding enough clean information in the image is hard 8) sky backgrounds have to be matched separately. The sensitivity to noise can be controlled using a smoothness constraint. One way to accomplish this is to use the fact that the PSF of a typical astronomical image can be approximated with a Gaussian and the Fourier transform of a Gaussian is also a Gaussian. The main idea is to replace noisy wings of the Fourier transform with an analytical approximation [25]. After additional modifications, this approach has been successfully applied in the regime of extreme blending (pixel lensing) to search for microlensing events in the direction of M31 [33], and to analyze the MACHO Galactic bulge microlensing data [4].

6. PSF matching with constant kernels

In typical astronomical crowded field applications we can safely assume that only a small fraction of sources in the image change their brightness. Then, a vast majority of image pixels would only vary due to accidental factors such as different sky background, transparency, and most importantly due to PSF variations. These factors need to be “matched” and taken out of consideration before a meaningful pixel-by-pixel difference image can be obtained, usually requiring some form of interpolation applied to at least one of the images (see section 10). The main ideas behind the Alard & Lupton algorithm [1] are: 1) avoid Fourier transforms and perform the calculation in data space 2) insist on linear decomposition of the PSF-matching kernel 3) propose a particular basis for the kernel that works with a wide range of images.

Assuming that a program image $I(x,y)$ is given on the same pixel grid as the reference image $R(x,y)$ to be subtracted, we can model the former as a convolution of the latter with a PSF-matching kernel $K(u,v)$. We have:

$$I(\mathbf{r}) \equiv (K \otimes R)(\mathbf{r}) \equiv \int K(\mathbf{r} - \mathbf{r}')R(\mathbf{r}') d\mathbf{r}', \quad (6.1)$$

where $\mathbf{r} \equiv (x,y)$ and $\mathbf{r} - \mathbf{r}' \equiv (u,v)$. The goal is to find the PSF-matching kernel that minimizes the residuals between the quantities on both sides of equation 6.1, so the equality holds in the “least squares sense”. For the moment we assume that the kernel K we are trying to find is constant across the image, i.e. independent of x,y . A more fundamental assumption is that the kernel has the form $K(u,v) = \sum_i a_i K_i(u,v)$, where K_i are fixed preselected shapes. In other words we only consider solutions that are linear in coefficients a_i . This greatly simplifies and accelerates the computation. Furthermore, the impact of such constraints on the final result is insignificant as long as the adopted set of K_i can absorb all detectable principal components of K .

The next step toward the solution is changing the order of integration over all image pixels and summation over kernel components

$$K \otimes R = \left(\sum_i a_i K_i \right) \otimes R = \sum_i a_i (K_i \otimes R), \quad (6.2)$$

that allows one to model a given program image as a linear combination of some other images C_i , namely the convolutions of the reference image with individual components of the kernel. We have $I(x,y) = \sum_i a_i C_i(x,y)$, where $C_i \equiv K_i \otimes R$. The simplest choice of the cost function to minimize that leads to a closed form solution for a_i is the familiar $\chi^2 = \int (I - \sum_i a_i C_i)^2 / \sigma^2 dx dy$. In some applications skipping the pixel variance weights $1/\sigma^2$ may improve the quality of the final result by shifting the relative emphasis toward handling the systematics of bright stars, since $\sigma(x,y) \propto \sqrt{I(x,y)}$. With the above assumptions the problem of finding coefficients a_i is reduced to solving a linear equation:

$$\mathbf{M}\mathbf{a} = \mathbf{V}, \quad (6.3)$$

where

$$M_{ij} = \int \frac{C_i(x,y)C_j(x,y)}{\sigma(x,y)^2} dx dy \quad (6.4)$$

$$V_i = \int \frac{I(x,y)C_i(x,y)}{\sigma(x,y)^2} dx dy \quad (6.5)$$

The integrals are over all useful pixels of the image. Note that only the right hand side vector depends on the program image $I(x,y)$. The least squares matrix only depends on the reference image $R(x,y)$ and therefore can be evaluated once for a large set of images differenced with the same template.

So far we have not addressed the issue of how to select the “kernelets” K_i , except for stating that they have a fixed shape. The original formulation of the Alard & Lupton (1998) algorithm advocates the use of a few (~ 3) bivariate Gaussians, each warped with a low order 2-D polynomial:

$$K(u, v) = \sum_n e^{-(u^2+v^2)/2\sigma_n^2} \sum_{0 \leq p+q \leq D_n} a_{pq} u^p v^q \quad (6.6)$$

$$K(u, v) = \sum_n \sum_{p,q} a_{pq} u^p v^q e^{-(u^2+v^2)/2\sigma_n^2} \quad (6.7)$$

$$0 \leq p+q \leq D_n, \quad \text{and} \quad 0 \leq p, q \leq D_n \quad (6.8)$$

“Unrolling” the double loop over Gaussian components n and monomial terms pq we have

$$K_i(u, v) = u^{p_i} v^{q_i} e^{-(u^2+v^2)/2\sigma_{n_i}^2}, \quad (6.9)$$

where the flattened index is $\{n, pq\} \rightarrow i$, and $i = 1, \dots, \sum_n (D_n + 1)(D_n + 2)/2$ for a polynomial of degree D_n . Indeed, after accounting for spatial variability of the kernel (to be discussed shortly), this simple choice of basis functions is all that is needed to handle most PSF-matching problems likely to occur in optical astronomy, including microlensing surveys. However, the Alard & Lupton algorithm does not require any particular form of basis functions K_i . Such details can be completely factored out of implementations and provide considerable flexibility to configure the model at run time (section 10).

7. Handling differential background

The minimum level of complication that any realistic image subtraction algorithm must handle is the presence of a smooth background that varies from one location to another within a single image and changes from one exposure to another. This is particularly important in microlensing surveys, where background estimates for individual images are highly uncertain due to severe crowding. The model now includes the background difference $B(x, y)$ between the images being subtracted:

$$I \doteq B + K \otimes R, \quad (7.1)$$

Provided that the background component is a linear combination of fixed functions that can be evaluated at every pixel, the modification amounts to adding more vectors (images) to the global linear fit in equations 6.1 and 6.2. The only practical constraint is that one expects $B(x, y)$ to be smooth, and a 2-D polynomial dependence in image coordinates is a reasonable default:

$$B(x, y) = \sum_i a_i B_i(x, y) = \sum_{p,q} a_{pq} x^p y^q \quad (7.2)$$

$$0 \leq p+q \leq D_B, \quad \text{and} \quad 0 \leq p, q \leq D_B \quad (7.3)$$

For the background polynomial of order D_B , the set of vectors $C_i(x, y)$ in the global fit to $I(x, y)$ now contains $N_B = (D_B + 1)(D_B + 2)/2$ smooth images $B_i(x, y)$ that do not depend on the reference image $R(x, y)$. Assuming the canonical Alard & Lupton basis from section 6, the remaining $N_K = \sum_n (D_n + 1)(D_n + 2)/2$ vectors belong to Gaussians enumerated by n , each having a fixed width and warped with a polynomial of order D_n . This information is summarized as:

$$C_i(x, y) = \begin{cases} B_i \equiv x^{p_i} y^{q_i} & i = 1 \dots N_B \\ R \otimes K_i & i = 1 + N_B \dots N_K + N_B \end{cases} \quad (7.4)$$

8. Spatially variable kernels

In most real world astronomical images the PSF shape changes depending on the location within the image, and the pattern of this variability will also change with time. As a result, for a generic pair of images the PSF-matching kernel will have a non-negligible gradient across the full image format. This is certainly the case for medium- to wide-field images characteristic of microlensing surveys. To some degree the problem can be addressed using frame subdivision, but the results are likely to be sub-optimal. The size of the subframe that can be subtracted assuming a constant kernel tends to become prohibitively small, especially for data sets collected in drift-scan mode. This section describes an extension of the original Alard & Lupton algorithm that explicitly treats variable kernels [2].

The basic idea is to allow the kernel coefficients to vary across the image format. As before, in situations of practical importance the change in $a_i(x, y)$ can be parameterized as a linear combination of constant functions such as 2-D polynomials:

$$a_i \rightarrow a_i(x, y) = \sum_{r,s} a_i^{rs} x^r y^s \quad (8.1)$$

The PSF-matching kernel becomes

$$K(u, v, x, y) = \sum_i a_i(x, y) K_i(u, v), \quad (8.2)$$

and the new least square vectors are

$$\tilde{C}_i^{rs} = x^r y^s (K_i \otimes R) = x^r y^s C_i \quad (8.3)$$

One can think of this new model in terms of the original vectors $C_i \equiv K_i \otimes R$ from section 6 taken with weights that are themselves (smooth) images constructed using a recipe analogous to the one used in section 7 for the background:

$$\{i^{rs}\} \rightarrow n \quad (8.4)$$

$$C_n = K_{i_n} \otimes R = C_{i_n} \quad (8.5)$$

$$P_n = x^{r_n} y^{s_n} \quad (8.6)$$

Index n covers a triple loop over r, s, i with r_n, s_n, i_n repeating for more than one value of n . It is straightforward to write down the matrix elements and the right hand side vector for the new extended problem:

$$\tilde{M}_{nm} = \int \frac{\tilde{C}_n \tilde{C}_m}{\sigma^2} dx dy \quad (8.7)$$

$$\tilde{M}_{nm} = \int \frac{C_n P_n C_m P_m}{\sigma^2} dx dy = \int \frac{C_n C_m P_n P_m}{\sigma^2} dx dy \quad (8.8)$$

However, the new least square matrix \tilde{M} has N^2 times the number of elements in M , the old least square matrix for a constant kernel (Equation 6.4), where N is the number of spatial basis vectors. The default model of spatial variability is a 2-D polynomial of order D_s with $N = (D_s + 1)(D_s + 2)/2$ vectors. The number of operations required to directly evaluate all elements as integrals over all useful pixels in the image is starting to grow substantially and the computational cost is becoming impractical. The solution is suggested by the observation that, unless something is seriously wrong with the optics, the functions $a_i(x, y)$ should not have high frequency components. Rather, one expects the spatial dependence of the kernel shape to be slow, certainly much slower than the spatial scale of the kernel itself (a few FWHMs). The quantity under the integral sign is a product of high and low frequency components. The trick is to consider the integral over a pixel region (domain) small enough to ignore the change in $P_n(x, y)$ over its extent. Then, the full image area Ω is broken into domains (usually squares or rectangles), and so is the corresponding 2-D integral.

$$\int_{\Omega} dx dy \rightarrow \sum_k \int_{\Omega_k} dx dy \quad (8.9)$$

$$\Omega_k \subset \Omega \quad k = 1, \dots, N_{\Omega} \quad (8.10)$$

The value of the low frequency component for the entire domain is very close to the value at the central pixel of the domain $P_n(x, y) \simeq P_n(x_k, y_k)$ and can be pushed in front of the integral sign. What is left of the integral, is exactly one of the matrix elements for the constant kernel problem in section 6, so the new matrix elements for a small pixel domain ($\sim 10 \times 10$ FWHM²) are very well approximated as products of the old matrix elements with the value of the position dependent weight at the domain center. An equivalent description is that, within a single domain, each old matrix element generates many elements of the new matrix through products with spatially variable weights evaluated only once for the middle pixel. The final step is to sum contributions from all domains for every element of the new expanded matrix.

$$\tilde{M}_{nm} = \sum_k P_n(x_k, y_k) P_m(x_k, y_k) \int_{\Omega_k} \frac{C_n C_m}{\sigma^2} dx dy \quad (8.11)$$

$$\tilde{M}_{nm} = \sum_k P_n(x_k, y_k) P_m(x_k, y_k) Q_{nm}^k \quad (8.12)$$

$$Q_{nm}^k = \int_{\Omega_k} \frac{C_{i_n}(x, y) C_{j_m}(x, y)}{\sigma^2} dx dy = M_{i_n j_m}^k \quad (8.13)$$

$$\tilde{V}_n = \sum_k P_n(x_k, y_k) Q_n^k \quad (8.14)$$

$$Q_n^k = \int_{\Omega_k} \frac{I(x,y)C_{i_n}(x,y)}{\sigma^2} dx dy = V_{i_n}^k \quad (8.15)$$

To summarize, the final prescription for evaluating the extended matrix and the right hand side vector to solve the least squares problem for the spatially variable kernel is:

- Divide the image into a number of pixel domains.
- For each domain calculate the least squares matrix and right hand side vector for the constant kernel problem.
- Assemble the least squares matrix and right hand side vector for the full variable kernel problem by combining the contributions from all domains with spatial weights evaluated at the center of each domain.

The same prescription captured in a corresponding equation reads:

$$\tilde{M}_{nm} = \sum_k P_n(x_k, y_k) P_m(x_k, y_k) M_{i_n j_m}^k \quad (8.16)$$

$$\tilde{V}_n = \sum_k P_n(x_k, y_k) V_{i_n}^k \quad (8.17)$$

Note that it is acceptable for domains to overlap slightly or cover only a fraction of the total area. It is important, however, to secure a good sampling of the area over which we are seeking the solution. In very crowded fields both uniformly spacing domains or centering them on bright stars works well. The latter method is usually better in less crowded regions because it naturally avoids fitting large areas of flat background that increase the noise but do not contribute any signal. It becomes a necessity in cases of severe but very localized crowding in the middle of a sparse field that tends to happen in strongly lensed quasars with multiple images [40, 39, 35].

9. Flux conservation

The PSF-matching kernel given by equation 8.3 has a norm that is generally free to change with the global image coordinates (x, y) . When applied to an image, a kernel with a drifting norm can effectively redistribute the flux over large distances and introduce photometric gradients. This could be potentially of use for correcting a relative transparency gradient or a large scale flat field problem. Apart from such unusual cases, however, it is safest to enforce a constant norm over the entire image, i.e. $\int K(u, v, x, y) du dv = \text{const}$. Fixing the norm of the kernel can be accomplished by rearranging the basis vectors. Many terms of the type given in equation 6.6 do not contribute to the kernel norm at all, because they are anti-symmetric and have vanishing integrals over the (infinite) 2-D surface, in practice over a suitably centered pixel mesh. The symmetric terms can always be normalized to 1, and then all but one “neutralized” by subtracting a single term that will carry the full norm of the entire linear combination. We have:

$$\int \tilde{K}_i(u, v) du dv = \begin{cases} 1, & i = 0 \\ 0, & i > 0 \end{cases} \quad (9.1)$$

and by simply removing the (x, y) dependence of the coefficient in front of K_0 (here reducing the a_0 polynomial to a single number), we obtain a kernel model

$$K(u, v, x, y) = a_0 K_0(u, v) + \sum_{i>0} a_i(x, y) [K_i(u, v) - K_0(u, v)] \quad (9.2)$$

with a constant norm

$$\int K(u, v, x, y) du dv = \int \sum_i a_i(x, y) \tilde{K}_i(u, v) du dv = a_0 \quad (9.3)$$

10. From images to light curves

1. Preprocessing: Taking a difference of two images implies that one of them must be interpolated to the pixel grid of the other. In this process one obtains a sequence of registered and resampled images in which there are no systematic offsets between centroids of the same stars. The input images are normally assumed to have no significant rotation, but crude pointing shifts up to ~ 100 pix are expected with additional higher order distortions induced by the optics, correlated atmospheric turbulence and imperfect tracking (especially in drift scan mode). A single image with good seeing, well behaved PSF and low background serves as the astrometric template that defines the pixel grid for all subsequent calculations. Optimally, the pointing of the template should fall near the middle of the pointing distribution for the entire data set. Typical steps performed at this stage are: 1) find crude pixel offsets between all program images and the astrometric template 2) pan program images to cover approximately the same area of the sky 3) obtain a list of object positions to be cross-identified between each program image and the template 4) cross-identify the same objects in each program image and the template 5) fit a 2-D coordinate transform (e.g. a pair of 2-D polynomials) to register each image on the template reference coordinate system 6) resample images using transformed pixel locations.
2. Constructing templates: The reference image (a.k.a. the template) to be subtracted pixel-by-pixel from each program image is typically a co-added stack of selected images with the best overall quality. This is motivated primarily by S/N considerations (Figure 6). In case of uncorrelated noise the variance of the difference of two images is the sum of the variances of individual images (a $\sqrt{2}$ loss of S/N for two identical exposures of the same field). Subtracting a mean of ~ 20 independent images from a single program image will only decrease the S/N of the result by a few percent. Automated selection of images suitable for template construction is generally tricky. The problems to avoid include: 1) bad tracking 2) bad seeing 3) high background 4) low transparency e.g. due to high airmass 5) under-sampling 6) bimodal PSF and other mishaps.
3. Computing difference images: At this stage the Alard & Lupton algorithm is used to subtract the reference image from all program images. The result is a sequence of difference images suitable for extraction of variable objects and PSF and/or aperture photometry. One typically begins by tuning the width and shape of the kernel without any spatial variability (Figure

- 7). The first objective is to account for PSF changes from image to image (seeing changes, focus etc.). The solution is further optimized by tuning spatial variability of the kernel and removing correlated residuals (Figure 8). To find an acceptable fit, one typically needs to:
- 1) adjust the size of the kernel mesh
 - 2) adjust the size and arrangement of kernel fitting domains
 - 3) provide enough fixed width components to the kernel so that all characteristic frequencies are represented
 - 4) experiment with Gaussian sigmas and polynomial orders
 - 5) experiment with the order of the polynomial modeling the (x, y) dependence of the kernel
 - 6) iteratively adjust remaining parameters. Note that in drif-scan data taking mode (e.g. OGLE-II) temporal variability contributes to spatial variability.
4. Assembling light curves: Performing actual photometric measurements on difference images and assembling them in light curves of individual objects can be accomplished in a number of ways. Here we will describe a particular approach to emphasize several important differences between working with images interpolated to the same pixel grid, and a more traditional case of photometry on unregistered images. Before we can compute DIA light curves, we need to be able to evaluate the PSF profile at any location within the reference image. This implies fitting some sort of global model of the image $\text{PSF} = \text{PSF}(x, y)$. Then, knowing the local shape of the PSF matching kernel (from the DIA solution), we can evaluate the PSF model for any object in any image by directly convolving two small pixel meshes. Once the difference images have been obtained, we are free to place circular apertures and filter pixels through PSF models anywhere in the frame. But of course it is most efficient to perform measurements only at locations covered by PSFs of variable objects. In many cases it is productive to obtain the DIA measurements of constant objects as well. Blending additionally complicates AC/DC flux assignment between objects. Note that for any variable object, regardless of its amplitude, there may be difference images with pixel counts buried in noise. This happens when the single image flux is very close to the reference image flux. One possible approach to selecting objects of interest and measuring their centroids for subsequent photometry is to combine a sequence of difference images into a single "variability image" and locate clusters of variable pixels. This can be done using single pixel light curves on which some measure of variability is computed and stored as the pixel value of the variability image. Statistics that have been utilized include: cumulative absolute deviation (optionally weighted by noise), cumulative variance, or special filters sensitive to transient variability.

Aperture photometry is of some use in selected DIA applications. However, in very crowded fields such as the Galactic bulge one normally works with profile photometry, since the variable signal can itself be moderately crowded. Using each input object centroid prepares a local PSF model (for a given x and y within the entire frame), shifts the model to a proper fractional pixel and dumps it onto a pixel mesh. A statistical estimator of the object flux is evaluated using a circular area of pixels around the centroid and the model mesh. In order to minimize the variance of the final measurement, all model parameters are fixed at known values except the PSF amplitude. This reduces the problem to a linear least squares fit with a single free parameter that can be written down in closed form (and takes little time to compute). Figure 9 demonstrates the ability of the DIA algorithm to improve the S/N ratio of light curves over conventional PSF fitting.

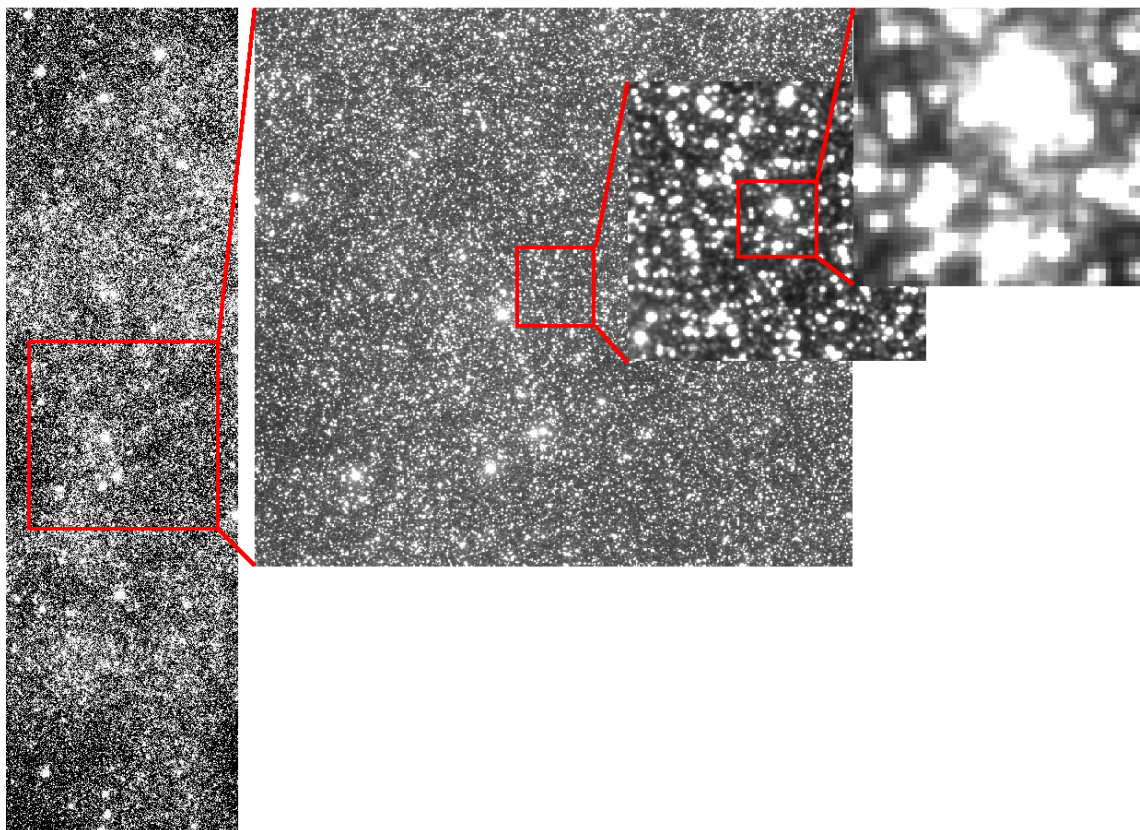
5. Estimating reference fluxes: Placing the light curve on a magnitude scale requires the knowledge of the object flux on the reference image (Figure 10 and 11). There are several possible approaches to this problem, each with their specific pros and cons depending on the application. We will only enumerate the possibilities and show the effects of underestimating and overestimating the reference flux. Note that in crowded fields source confusion and flux blending conceal true baseline levels of variable and constant objects independently of the applied photometric technique. We have: $f(t_i) = f_{\text{ref}} + f_{\text{DIA}}(t_i)$, and $m = \text{const} - 2.5 \times \log_{10} f$. Options to consider are: 1) measuring f_{ref} on the reference image using the same photometric code that was applied to difference images (if the mag zero point for the field was obtained with a conventional PSF code such as DOPHOT, we need to convert the flux units, usually by comparing relatively isolated stars) 2) comparing each DIA light curve to its calibrated conventional counterpart i.e. fit f_{ref} in $m_i = \text{const} - 2.5 \times \log_{10}(\Delta f_i + f_{\text{ref}})$ 3) using external information such as blending fractions and reference fluxes from Hubble Space Telescope (HST) observations 4) in case of transient objects, ensuring the reference flux $f_{\text{ref}} = 0$ with a proper choice of the template.
6. Other details:
- (a) Note that it is not necessary to know the PSF of either image to find the PSF-matching kernel. The method actually turns crowding into an advantage by using all information in the image. In a crowded image of a stellar field, every pixel contains information about the PSF.
 - (b) The method is reasonably robust against slight under-sampling
 - (c) The light centroid of variable objects in correctly subtracted images is unbiased.
 - (d) Kernel components with “dipole” shapes similar to Gaussian derivatives have a flux shifting property. They can be used to remove residual image mis-registration or to detect subtle motions [13].
 - (e) If the approximate light curve of a particular variable object is known (e.g. from a preliminary reduction of the same data), it is possible to find a linear combination of images in a given sequence that cancels the contribution of non-variable objects [16]. This way one can construct an image of the lensed light free of blends.
 - (f) When preparing reference images, it is important to keep in mind that optimal image co-addition in the most general case is an unsolved problem.
 - (g) The effective PSF of variable stars is slightly different than that of constant stars in case of a large spread of the PSF shape between the individual images stacked during template preparation.
 - (h) Good interpolation techniques matter. The plain old linear interpolation is biased (it cannot produce a value larger than either interpolated sample).
 - (i) Pixels inside the PSF footprint of variable stars must be clipped out of the fit (e.g. using iterative sigma clipping), and image regions dominated by defects or flat background must be masked.

- (j) To get the variance of the convolved image, convolve the variance image with the square of the kernel.
- (k) Caching the matrix for the least squares problem, and reusing it for entire sequence of images to be subtracted, is what makes the Alard & Lupton method sufficiently fast to use in microlensing surveys. This means that the weights of individual pixel contributions to the matrix can only depend on the reference image, implying that the “variance weighting” can only be approximate.
- (l) Separable kernels of the form $K_i(x,y) = f(x)g(y)$ are much faster to compute ($2 \times N$ ops versus N^2 , where N is the size of the kernel mesh).
- (m) There is still much room for experimenting with new choices of the basis functions: shapelets, orthonormal polynomials, axially symmetric functions, etc. For some problems fitting each pixel of the kernel mesh as a separate parameter may be the way to go. In the latter case the function K_i takes the form of a pixel mesh with one of the pixels set to 1 and the rest to 0.

References

- [1] Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325
- [2] Alard, C. 2000, A&AS, 144, 363
- [3] Alcock, C., et al. 1999, PASP, 111, 1539
- [4] Alcock, C., et al. 1999, ApJ, 521, 602
- [5] Alcock, C., et al. 2000, AJ, 119, 2194
- [6] Alcock, C., et al. 2004, AJ, 127, 334
- [7] Ansari, R. 1996, Vistas in Astronomy, 40, 519
- [8] Axelrod, T. S., et al. 1998, Proc. SPIE, 3349, 152
- [9] Beaulieu, J. P., et al. 2008, ArXiv e-prints, 808, arXiv:0808.0005
- [10] Becker, A. C., et al. 2004, ApJ, 611, 418
- [11] Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
- [12] Bond, I. A., et al. 2001, MNRAS, 327, 868
- [13] Eyer, L., & Woźniak, P. R. 2001, MNRAS, 327, 601
- [14] Fang, J., & Aspin, C. 2007, Bulletin of the American Astronomical Society, 38, 806
- [15] Goldberg, D. M., & Woźniak, P. R. 1998, Acta Astronomica, 48, 19
- [16] Gould, A., & An, J. H. 2002, ApJ, 565, 1381
- [17] Gould, A., et al. 2006, ApJ, 644, L37
- [18] Han, C. 1999, MNRAS, 309, 373
- [19] Irwin, M. J. 1985, MNRAS, 214, 575
- [20] Ivezić, Z., Tyson, J. A., Allsman, R., Andrew, J., Angel, R., & for the LSST Collaboration 2008, ArXiv e-prints, 805, arXiv:0805.2366

- [21] Konacki, M., Torres, G., Jha, S., & Sasselov, D. D. 2003, *Nature*, 421, 507
- [22] Miknaitis, G., et al. 2007, *ApJ*, 666, 674
- [23] Paczynski, B. 1996, *ARA&A*, 34, 419
- [24] Park, B.-G., et al. 2004, *ApJ*, 609, 166
- [25] Phillips, A. C., & Davis, L. E. 1995, *Astronomical Data Analysis Software and Systems IV*, 77, 297
- [26] Reid, M. L., Sullivan, D. J., & Dodd, R. J. 2001, *Astronomical Data Analysis Software and Systems X*, 238, 306
- [27] Schechter, P. L., Mateo, M., & Saha, A. 1993, *PASP*, 105, 1342
- [28] Smith, M. C., Woźniak, P., Mao, S., & Sumi, T. 2007, *MNRAS*, 380, 805
- [29] Soszynski, I., et al. 2007, *Acta Astronomica*, 57, 201
- [30] Stanek, K. Z., Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., Mateo, M., & Krzeminski, W. 1997, *ApJ*, 477, 163
- [31] Stetson, P. B. 1987, *PASP*, 99, 191
- [32] Sumi, T., et al. 2006, *ApJ*, 636, 240
- [33] Tomaney, A. B., & Crotts, A. P. S. 1996, *AJ*, 112, 2872
- [34] Udalski, A., Szymanski, M., Kaluzny, J., Kubiak, M., & Mateo, M. 1992, *Acta Astronomica*, 42, 253
- [35] Udalski, A., et al. 2006, *Acta Astronomica*, 56, 293
- [36] Udalski, A. 2003, *Acta Astronomica*, 53, 291
- [37] Woźniak, P., & Paczynski, B. 1997, *ApJ*, 487, 55
- [38] Woźniak, P. R. 2000, *Acta Astronomica*, 50, 421
- [39] Woźniak, P. R., Alard, C., Udalski, A., Szymański, M., Kubiak, M., Pietrzyński, G., & Żebruń, K. 2000, *ApJ*, 529, 88
- [40] Woźniak, P. R., Udalski, A., Szymański, M., Kubiak, M., Pietrzyński, G., Soszyński, I., & Żebruń, K. 2000, *ApJ*, 540, L65



POS (GMC8) 003

Figure 1: Field crowding in images collected by microlensing surveys poses significant challenges for photometric algorithms. The first image in this series of progressively smaller chunks of the BUL_SC3 field monitored by the OGLE-II survey covers the full $2K \times 8K$ pix scan, and the size of the last one is a mere 50×50 pix.

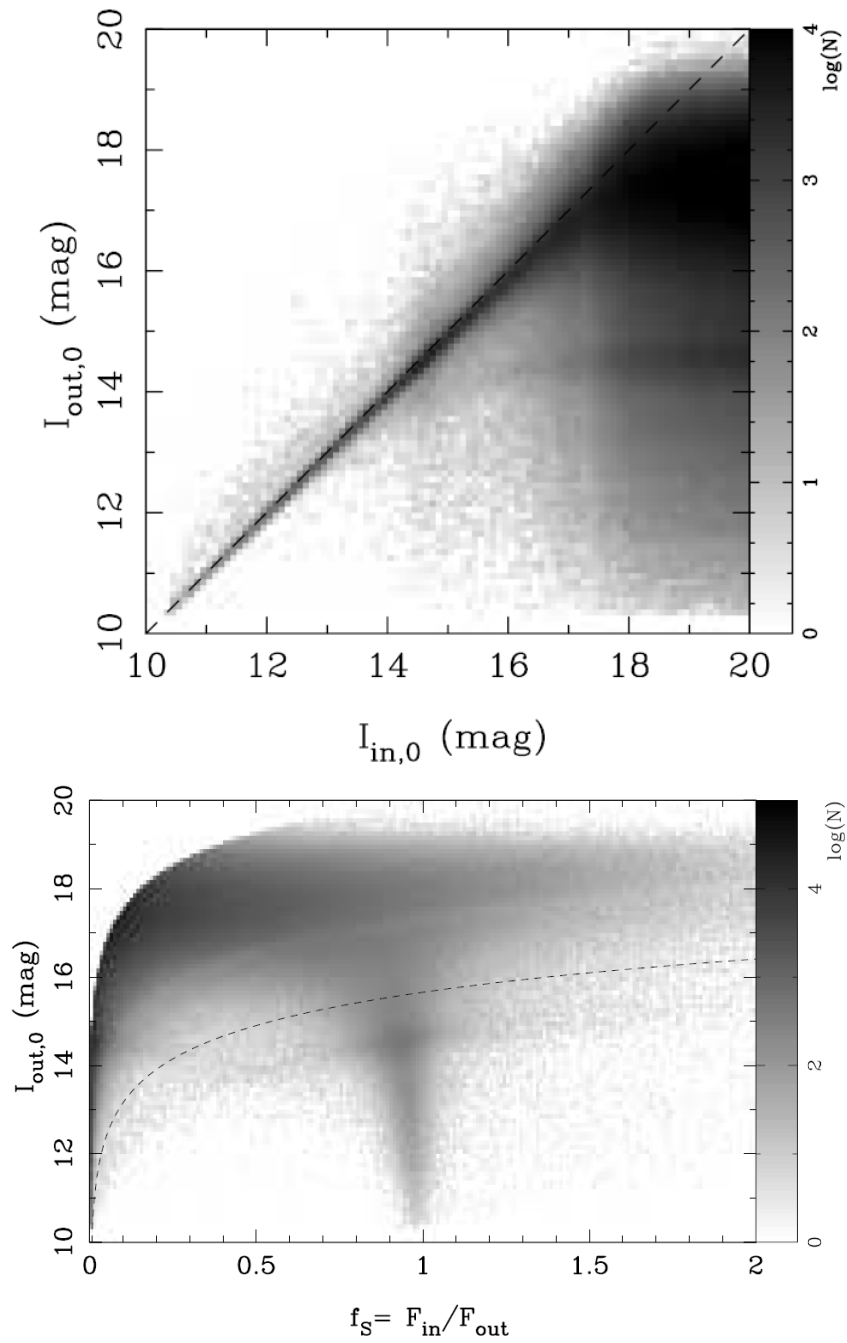


Figure 2: Deblending performance of DOPHOT based on full image simulations of one of the OGLE-II Galactic bulge fields using a realistic deep LF from HST observations (from [32]). The plot shows the recovery of magnitudes (top), and the corresponding distribution of source fractions in random microlensing events (bottom).

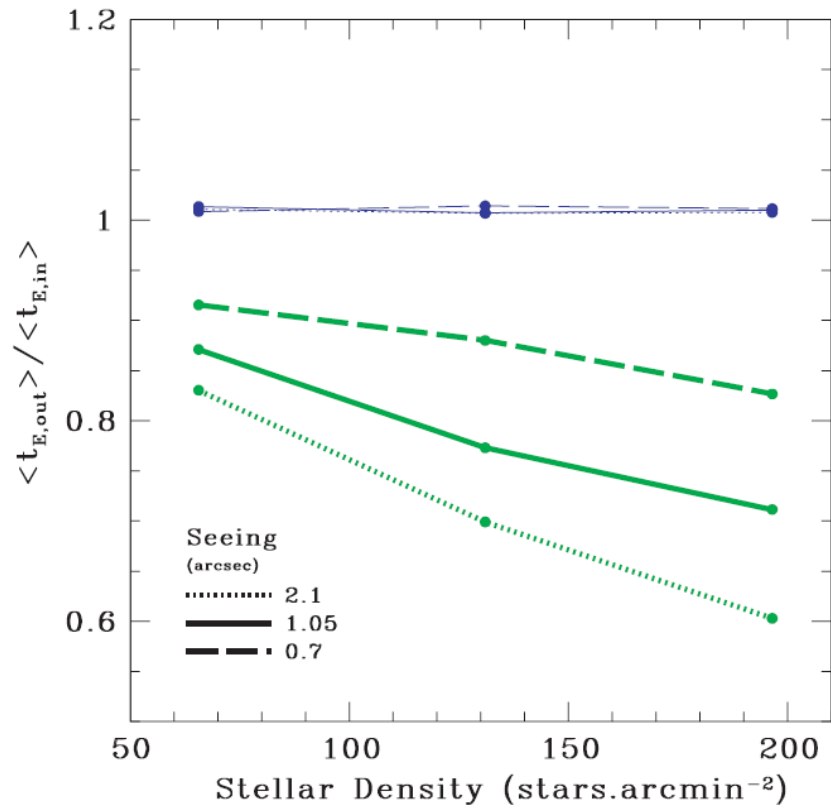


Figure 3: Time-scale bias in magnitude limited samples of microlensing events introduced by the assumption of negligible blending (from [28]). The mean efficiency corrected time-scale resulting from four-parameter fits (green symbols) is compared to that obtained from five-parameter fits (blue symbols).

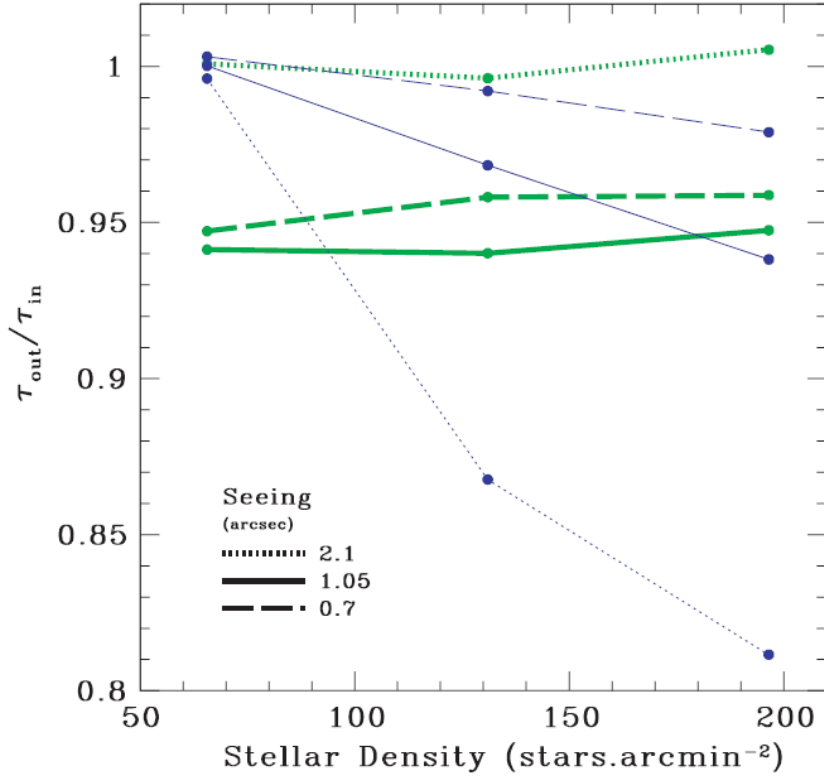


Figure 4: Systematics of microlensing optical depth τ in full image level simulations of several survey scenarios (from [28]). The estimates using an incorrect assumption of negligible blending (green symbols) are compared to estimates based on five-parameter model fits and unbiased time-scales (blue symbols). Assuming $f_s \equiv 1$ in a magnitude limited sample attributes all observed events to a much smaller population of sources than effectively monitored. The near cancellation of bias in the microlensing time-scale with the overestimated number of events holds over a large range of field densities and seeing. The alternative approach is using the fitted f_s values to reject events with source magnitudes below the sample threshold and utilize unbiased time-scales. However, for extremely crowded fields observed under poor seeing the “correct” estimator may actually be more biased, because of a net increase in the number of stars above the magnitude threshold due to blending.

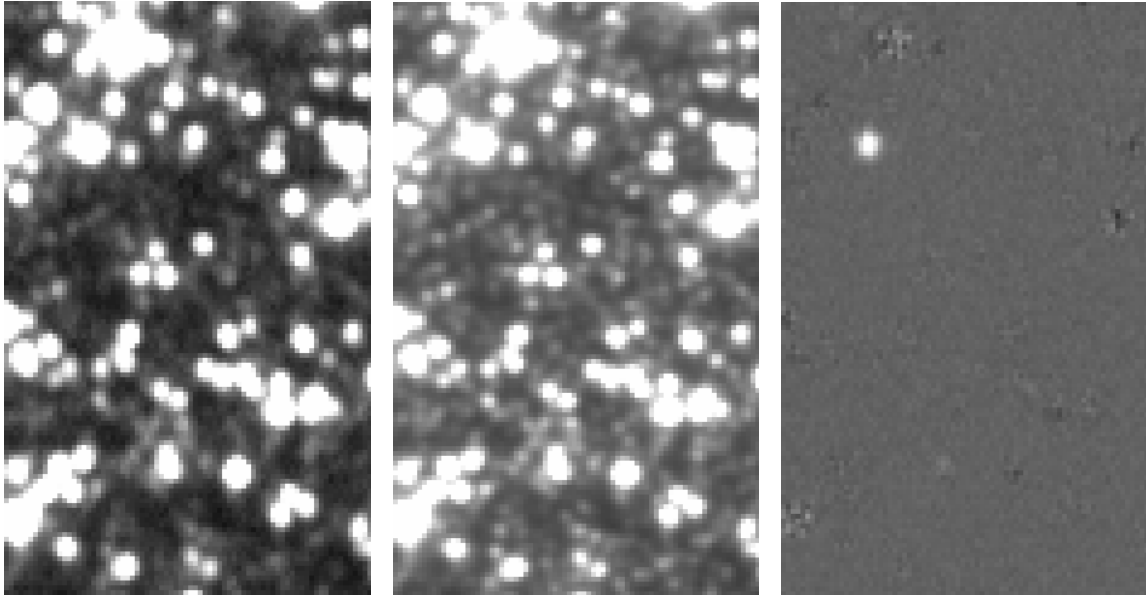


Figure 5: The power of difference imaging. By subtracting from a single program image (left) a high S/N reference image convolved with the PSF-matching kernel (middle), we obtain a difference image with the constant part of the signal removed (right). The PSF photometry of variable objects based on the difference image is largely free of the complexities associated with severe crowding. The errors in the difference flux are typically within 20% of the photon noise limit.

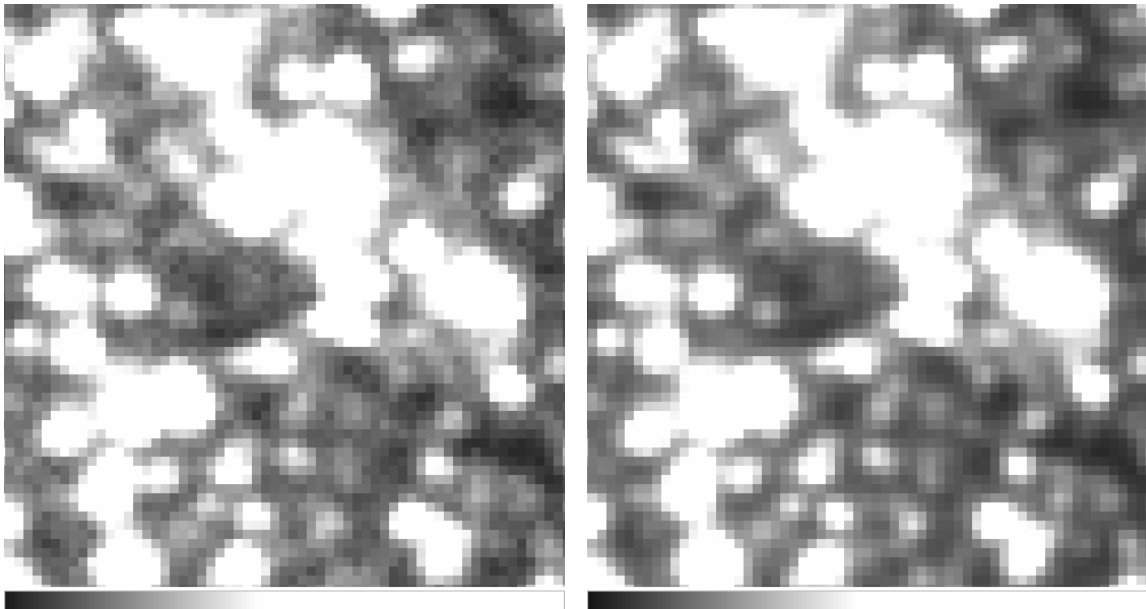


Figure 6: Typical S/N improvement between a single OGLE-II image (left) and a reference image combining 20 individual frames (right). High quality templates are critical to obtaining good difference images and optimizing the final photometric accuracy of DIA light curves.

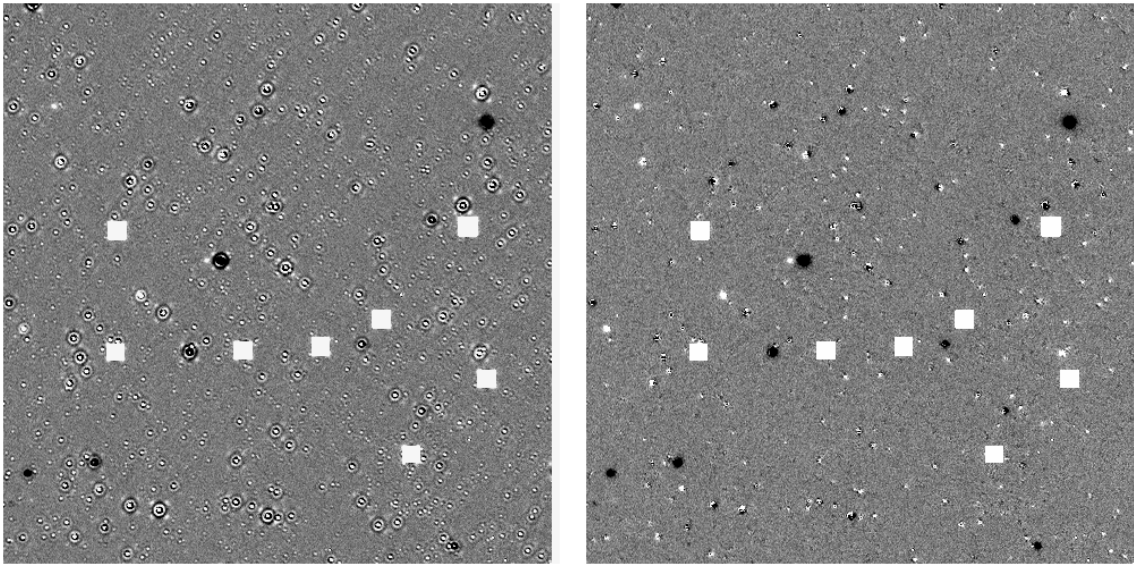


Figure 7: Tuning the local shape parameters of the PSF-matching kernel. In this example showing a pair of 512×512 pix difference images, the model with two relatively broad Gaussian components ($\sigma = 1.5$ and 4.5 pix) and 2-nd order polynomials fails to represent high spatial frequencies dominating the inner core. The tension between the core and the wings produces characteristic ringing effect (left). The model with three Gaussians ($\sigma = 0.75, 1.35,$ and 2.43 pix) and 3-rd order polynomials provides a much better fit with residuals close to the photon noise limit (right). Saturated stars are masked by white squares.

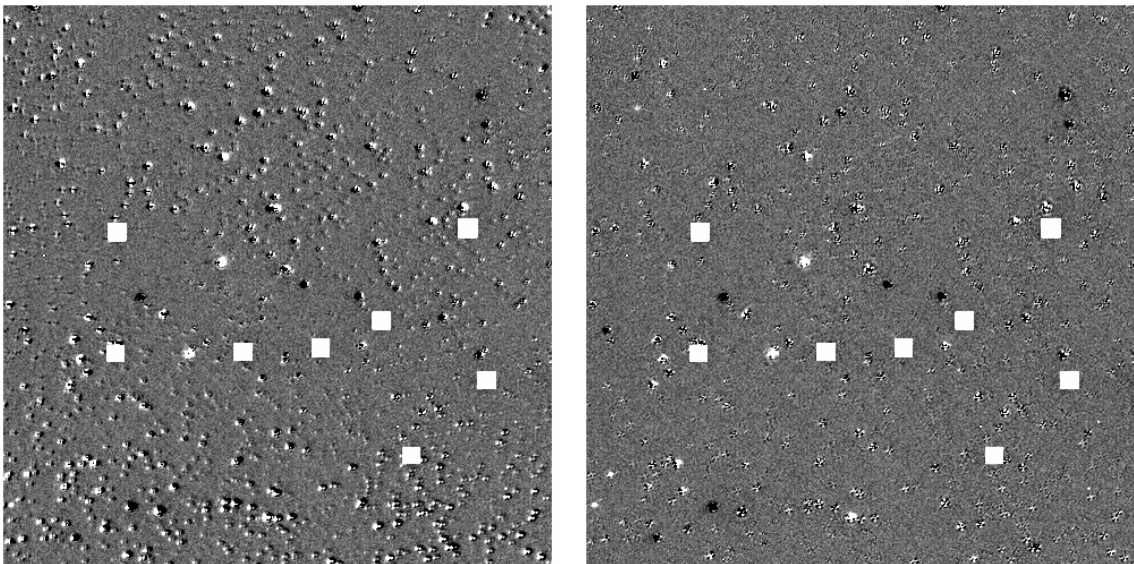


Figure 8: Fitting PSF-matching kernels with strong gradients. The frame used in this example shows a particularly strong PSF variation from top to bottom caused by a fluctuation in tracking rate during a drift-scan. A constant PSF-matching kernel provides a very poor fit to this data and produces characteristic dipole-like residuals with correlated orientations (left). Allowing the shape of the kernel to change across the image dramatically improves the final result (right). Here, a 3-rd order polynomial was used to model the gradient in kernel coefficients. Saturated stars are masked by white squares.

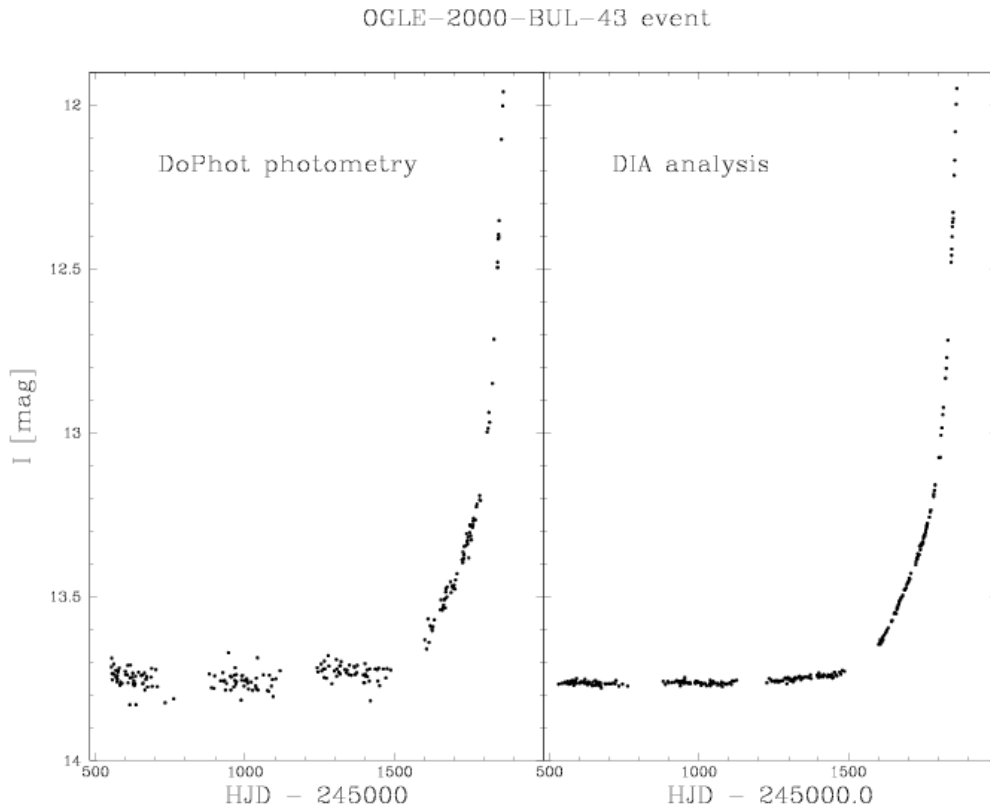


Figure 9: Comparison of photometry obtained using a conventional PSF fitting code (left) and difference imaging (right). The plotted microlensing light curve is for OGLE-BLG-2000-43, a parallax event discovered by the OGLE-II survey. The measurements based on the DIA technique have 4–5 times better S/N ratio. In this particular case the improvement is somewhat extreme due to the presence of a tight blend. A factor 1.5–2.0 decrease in the error bar is more typical for the OGLE-II data.

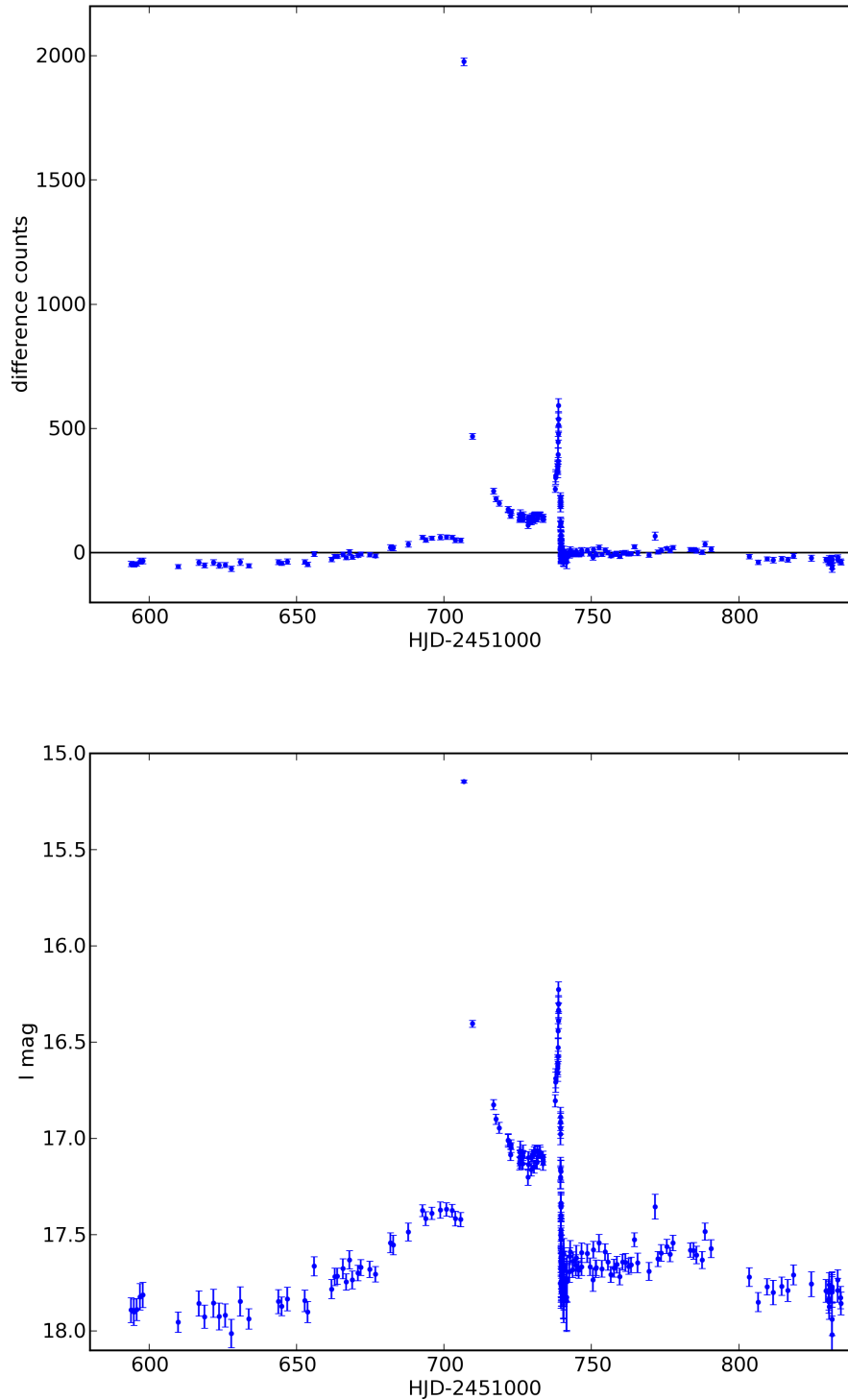


Figure 10: DIA light curve of OGLE-BLG-2000-38, a binary microlensing event discovered by the OGLE-II survey. In order to convert the linear DIA units to magnitudes, the light curve must be shifted by the total flux of the object in the reference image (the reference flux). Note that the reference flux is above the baseline level for microlensing sources that are effectively magnified in the reference image.

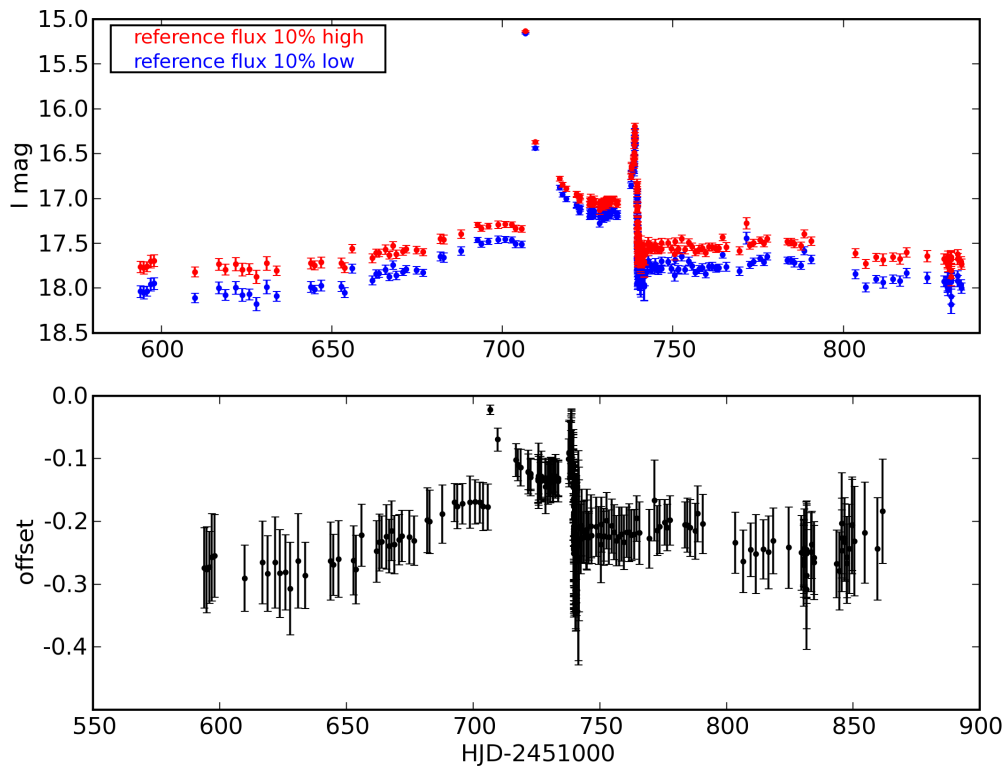


Figure 11: Influence of the reference flux errors on the DIA light curves. Using an incorrect value of the object flux to convert between the linear DIA flux units and magnitudes results in magnitude dependent offsets and affects the shape of the light curve.

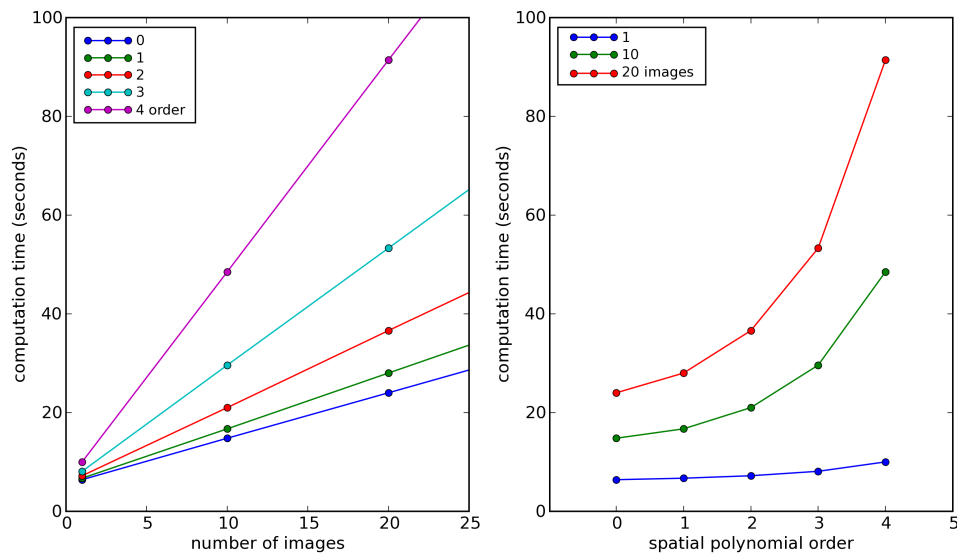


Figure 12: Processing time required to subtract the same template image (512×512 pix) from a sequence of program images. The first difference image is by far the most expensive to compute, because the least squares matrix only depends on the reference image and can be reused in the remaining calculations (left). The processing time increases rapidly with the number of basis vectors in the least squares fit (right). The tests were performed on a commodity dual processor 32-bit Intel Xeon (3.6 GHz) machine.