

Microlensing on different timescales in the lightcurves of QSO 0957+561 A,B

J. Pelt^{1,5}, R. Schild^{2,5}, S. Refsdal^{3,5}, and R. Stabell^{4,5}

¹ Tartu Observatory, EE-2444 Tõravere, Estonia

² Harvard-Smithsonian Center for Astrophysics, MS-19 60 Garden Street, Cambridge, MA 02138, USA

³ Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg-Bergedorf, Germany

⁴ Institute of Theoretical Astrophysics, P.O. Box 1029, Blindern, N-0315 Oslo, Norway

⁵ Centre for Advanced Study, Drammensveien 78, N-0271, Oslo, Norway

Received 20 February 1998 / Accepted 26 May 1998

Abstract. A statistical analysis of an extended data set on a time basis of 17 years of optical observations of the gravitationally lensed quasar 0957+561 is presented. We use all available photometry to estimate the time delay between *A* and *B* images as 416.3 ± 1.7 days. Using this refined value we investigate basic statistical properties of the difference curve between the *A* and the time shifted *B* image. We find a significant variability of 0.25 mag on a time scale of 5 years, which we interpret as microlensing by compact objects in the lensing galaxy.

We show that masses down to about $10^{-5} M_{\odot}$ may explain this microlensing event. Solar mass objects alone can more or less be excluded. However, a distribution of compact objects with a small fraction ($\lesssim 5\%$) in solar mass objects, seems to be possible. The best constrained parameter is the radius of the source which roughly should be $3 \times 10^{15} \text{ cm}$ (within a factor of 3).

With the present noise level an additional variability due to the microlensing on shorter time scales (days to months) is difficult to prove with a reasonable significance. All steps of the analysis are carried out by using simple statistical tests and with as few *a priori* assumptions as possible.

Key words: methods: statistical – methods: data analysis – quasars: individual: 0957+561A,B – gravitational lensing

1. Introduction

The problem of time delay estimation for the double quasar 0957+561A,B is generally considered as having a consistent solution 417 ± 3 days (Kundić et al. 1997, Haarsma et al. 1997 and references therein). In the context of the Hubble constant determination this is sufficiently accurate. But if we are interested in a more detailed study of the microlensing on all relevant time scales we must have a secure value for the time delay with at least 1-2 day precision. This is why we revisit the problem of the time delay estimation. Using all data available to us and methodology developed by Pelt et al. (1994, Paper I) and Pelt et

al. (1996, Paper II) we confirm the value of 416 – 417 days and use it as a basis for refined variability analysis. We also try to explain why the determination of the correct time delay value was so complicated and controversial.

Using the established time delay value we are now in a position to divide the total observed variability of the quasar images into four separate statistically defined components: source variability, long term variability (years) due to the microlensing, a possible rapid microlensing (days to months) and observational noise. We do our analysis with simple statistical tests and procedures. The reason for that is twofold: first - we can be sure that applied procedures are well known and consequently all our computations can easily be repeated, and secondly - we can avoid any artifacts of the more modern but less understood methods which often tend to distort the results in unexpected ways, especially when the data sets involved have a complex nature (irregular timing, long gaps, non-uniform precision of different measurements). There are parallel developments which use certain complex interpolation schemes (Thomson & Schild 1997) or wavelet analysis (Schild 1998) but in the current report we want to lay a groundwork for more advanced studies.

In Sect. 2 we describe the data sets and the process of merging them into one master data set. Sect. 3 is devoted to the time delay determination. As a byproduct of the method used we also get a numerical estimate of the difference curve between *A* and time shifted *B* (microlensing). In Sect. 4 we use simple statistical tests to analyse residuals after the main fitting procedure. In the discussion we try to put the obtained estimates into the context of current understanding of microlensing physics.

1.1. Notation

In the following presentation we often use combined statistical weights

$$W_{i,j} = \frac{W_i W_j}{W_i + W_j}, \quad (1)$$

for the pair of observer estimated weights $W_i = \frac{1}{\sigma_i^2}$, $W_j = \frac{1}{\sigma_j^2}$. The additional weights $S_{i,j}$ are used to include in summations

Send offprint requests to: R. Stabell⁴

only pairs of observations whose observing times, t , do not differ by more than δt :

$$S_{i,j} = \begin{cases} 1, & \text{if } |t_i - t_j| \leq \delta t, \\ 0, & \text{if } |t_i - t_j| > \delta t, \end{cases} \quad (2)$$

or downweight the pairs linearly according to the distance of time points:

$$S_{i,j}^* = \begin{cases} 1 - \frac{|t_i - t_j|}{\Delta t}, & \text{if } |t_i - t_j| \leq \Delta t, \\ 0, & \text{if } |t_i - t_j| > \Delta t. \end{cases} \quad (3)$$

The downweighting parameter Δt can often be considered as a *decorrelation length* (see Paper II).

Below all the brightness estimates are given in magnitudes and all time arguments are given in days.

2. Data sets

Our new analysis is based on five data sets. The largest of them is the significantly extended (when compared to the data set used in Paper II) data set compiled by R. Schild and D. Thomson. It contains 1074 pairs of *A* and *B* image measurements in the R photometric band from 16 Nov. 1979 to 20 Nov. 1996; we will denote it as CR (combined) data set. The second set is a subset of CR which contains only 958 observations made at Mt Hopkins (SR data set). The two data sets measured by Kundić et al. (1995,1997, ftp astro.princeton.edu/elt/:0957) contain 100 *A* and 101 *B* image measurements in the R band and 97 *A* and 98 *B* image measurements in the G band (data sets KR and KG). Finally, we also used the data set obtained by Oscoz et al. 1996 (data by private communication) which contained originally 44 A,B pairs in the R photometric band. Because we were interested in comparison of different photometries we skipped the two first observations (from an earlier year) from this data set. The resulting data set of 42 pairs will be denoted OR.

2.1. Stated and estimated precision

All five data sets contain photometric measurements and estimates of standard errors. To check the internal consistency of the data and also to get some idea about possible high frequency variations (generally from one day to the next) we calculated also values for the following simple statistic:

$$\sigma_{est}^2 = \frac{\sum_{i=1}^{N-1} W_{i,i+1} (y_i - y_{i+1})^2}{2 \sum_{i=1}^{N-1} W_{i,i+1}}, \quad (4)$$

where $y_i, i = 1, \dots, N$ is the sequence of measurements and $W_{i,j}$ are the combined weights. Generally we included into the average only pairs which were observed in adjacent nights ($\delta t \leq 1.5$); however for sparse data sets KR,KG and OR we used also $\delta t = 3.5$. In Table 1 the computed estimates σ_{est} are listed together with the mean standard errors given by the observers

$$\sigma_{std} = \sqrt{\frac{\sum_{i=1}^N \sigma_i^2}{N}}. \quad (5)$$

Table 1. Stated and estimated standard errors

Data set	Image	N_{pairs}	σ_{std} mag	σ_{est} mag	δt days
CR	A	681	0.0137	0.0116	1.5
CR	B	681	0.0162	0.0115	1.5
SR	A	632	0.0108	0.0114	1.5
SR	B	632	0.0113	0.0113	1.5
KR	A	10	0.0107	0.0084	1.5
KR	B	10	0.0107	0.0080	1.5
KR	A	55	0.0107	0.0116	3.5
KR	B	55	0.0107	0.0117	3.5
KG	A	9	0.0116	0.0081	1.5
KG	B	9	0.0118	0.0109	1.5
KG	A	55	0.0116	0.0123	3.5
KG	B	56	0.0118	0.0123	3.5
OR	A	20	0.0229	0.0261	1.5
OR	B	30	0.0200	0.0176	1.5
OR	A	20	0.0229	0.0229	3.5
OR	B	30	0.0200	0.0167	3.5

It is encouraging to see that both numbers are generally in quite good agreement for all the involved data sets ($\sigma_{est} > \sigma_{std}$ would reveal fast fluctuations, $\sigma_{est} \approx \sigma_{std}$ absence of significant day to day variability). The *B* image precision for the CR data is probably even underestimated by the observers. On the basis of this observation alone we can conclude that high level daily variability is generally ruled out. Any daily microlensing and/or source variability must be less than the mean errors of the reported observations, or about 0.01 mag expressed as an rms deviation. Periods of quiescence in the quasar's brightness have been noted by Schild 1990. However, some infrequent sharp rises or drops are not ruled out.

2.2. Merging of data sets

To get as precise and stable estimates as possible for the time delay we merged the CR, KR and OR data sets into one master data set (MR). The KR set was chosen to be merged into the master data set because of smaller color differences between it and the CR data sets. We also considered appropriate color corrections (see Kent 1985), but found that they do not fully describe the discrepancies between the photometries. It is important to note that the observers rely more on the G band photometry. However, there is a certain divergence between the KG and the KR photometry, to be discussed below. We assume that at the current level of precision these differences do not play a significant role for our conclusions.

Before merging, the KR and the OR data sets were formally shifted in magnitudes to match the SR data. The magnitude offsets were calculated from observed pairs which were obtained on the same night in different observatories ($\delta t < 0.5$). The corresponding magnitude corrections are given in Table 2. The corrections for the OR photometry are large because the original magnitudes were given relative to a nearby star. The master data set MR contains 1216 measurements for the *A* image and 1217 measurements for the *B* image. The stated and estimated mean

Table 2. Magnitude shifts relative to CR for merged data sets

Data set	Image	N_{pairs}	δm
KR	A	43	-0.635
KR	B	44	-0.430
OR	A	15	14.512
OR	B	15	14.709

standard errors are $\sigma_{std} = 0.0140$ and $\sigma_{est} = 0.0126$ for the A image, $\sigma_{std} = 0.0159$ and $\sigma_{est} = 0.0124$ for the B image.

3. Time delay

The seemingly simple procedure of time delay estimation for the two lightcurves turned out to be much more complicated than was first anticipated. References to previous work can be found in Papers I and II. Recent events are described in Kundić et al. 1997, Haarsma et al. 1997, Oscoz et al. 1997 and Goicochea et al. 1998. Now it can be safely said that there are two types of nuisance factors which strongly influence time delay calculations. The first one is heterogeneity of observational data in terms of its sparseness and unequal quality. The second one is significant additional variability (besides source variability) in one or both observed signals from the source quasar. We assume here that this additional variability is due to microlensing by stars or other compact objects, probably in the lensing galaxy (see Chang & Refsdal 1979). In the following analysis we divide the overall change in both components into distinct statistical dispersion components - source variability, long timescale (years) microlensing, possible rapid microlensing (days to months) and observational noise.

3.1. Dispersion spectra

To estimate the time delay τ between images A and B we use basically the same methods as described in Paper I and II.

Our data model is:

$$A_i = q(t_i) + \epsilon_A(t_i), i = 1, \dots, N_A, \quad (6)$$

and

$$B_j = q(t_j - \tau) + l(t_j) + \epsilon_B(t_j), j = 1, \dots, N_B, \quad (7)$$

where $q(t)$ is the source variability of the quasar, $l(t)$ is a constant difference in magnitudes plus additional variability in time due to microlensing (for simplicity we assign it fully to the B curve), $\epsilon_A(t)$ and $\epsilon_B(t)$ are observational errors. We generate a combined light curve C by taking all values of A as they are and “correcting” the B data by $l(t)$ and shifting them by τ :

$$C_k(t_k) = \begin{cases} A_i, & \text{if } t_k = t_i, \\ B_j - l(t_j), & \text{if } t_k = t_j + \tau, \end{cases} \quad (8)$$

where $k = 1, \dots, N$ and $N = N_A + N_B$. The dispersion spectra

$$D^2(\tau) = \min_{l(t)} D^2(\tau, l(t)). \quad (9)$$

can now be computed and searched for significant minima with respect to τ . In this paper we use two methods to estimate dispersions. The simplest dispersion estimate (Paper I) is:

$$D_1^2 = \min_{l(t)} \frac{\sum_{k=1}^{N-1} W_{k,k+1} G_{k,k+1} (C_{k+1} - C_k)^2}{2 \sum_{k=1}^{N-1} W_{k,k+1} G_{k,k+1}}, \quad (10)$$

where the $W_{k,k+1}$ are combined statistical weights (Eq. 1) and $G_{k,k+1}$ are weights which switch off pairs of C_k, C_{k+1} which happen to be lying on one and the same curve (A or B).

The slightly more complicated dispersion estimate (see Paper II):

$$D_2^2 = \min_{l(t)} \frac{\sum_{n=1}^{N-1} \sum_{m=n+1}^N W_{n,m} S_{n,m}^* G_{n,m} (C_n - C_m)^2}{2 \sum_{n=1}^{N-1} \sum_{m=n+1}^N W_{n,m} S_{n,m}^* G_{n,m}}, \quad (11)$$

uses linear downweighting by $S_{n,m}^*$ and includes generally more pairs into the summation.

The reason we use not only D_2^2 but also D_1^2 in this paper is the following. First, the simplest scheme does not contain arbitrary parameters and this gives to it a certain edge over the more complicated methods. Second, it was found that smoothing of dispersion spectra by additional weighting can introduce extra bias due to the asymmetry of the details in the curves compared. In Paper II we found that there can be systematic errors up to a level of 2 days (in the context of the data sets treated then). As we can see below, we slightly underestimated the bias. Recent photometric events which were observed (0.13 mag in 22 days during the so called “Kundić” event, Kundić et al. 1995, 1997) show that intrinsic variability can contain quite sharp episodes. The underlying assumption (implicit) in Paper II, that high frequency intrinsic fluctuations (on a time scale less than ten days) are mostly due to observational errors, was not fully justified. In this paper we want to utilize the information in the data sets on all the relevant timescales.

Before going further with application of the simple scheme to the master data set MR we try (once again) to explain why we carefully avoid gap filling approaches in our analysis.

3.2. Filling of the gaps can be misleading!

In Figs. 1 and 2 the A and B lightcurves for the master data set are depicted with smoothed versions of them. The smoothing procedures we used are described in detail in Appendix A. For this example we used a local parabolic fit with downweighting parameter $\Delta t = 300$ (the results are not strongly dependent on the chosen value). The continuous curves are obtained by calculating local parabolic fits (polynomial degree $d = 2$) for every fifth day and by connecting the obtained estimates with straight lines.

In Fig. 3 two spectra are depicted, the lower one is computed from the data set which is obtained by sampling the continuous estimate of A and B curves exactly at the moments where actual observations were made. The upper spectrum is computed using continuous curves (sampled with a time step 5 days). It is amazing to see, how the minimum of the dispersion spectrum moves from around 420-430 days to the 530-540 day region. This illustrates well how combining values in the gaps with the

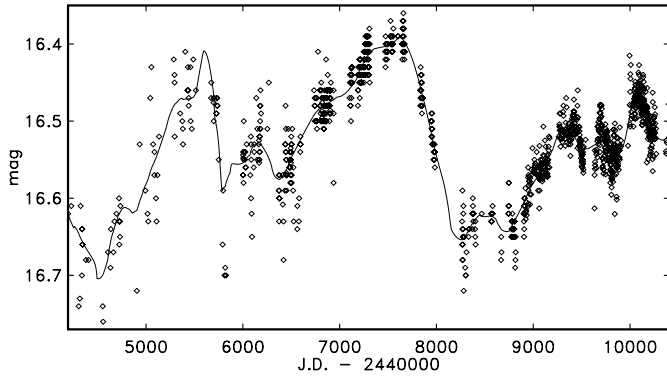


Fig. 1. Master data set, image A. The continuous line is the smoothed version of it (downweighting parameter $\Delta t = 300$ days, smoothing polynomials degree $d = 2$)

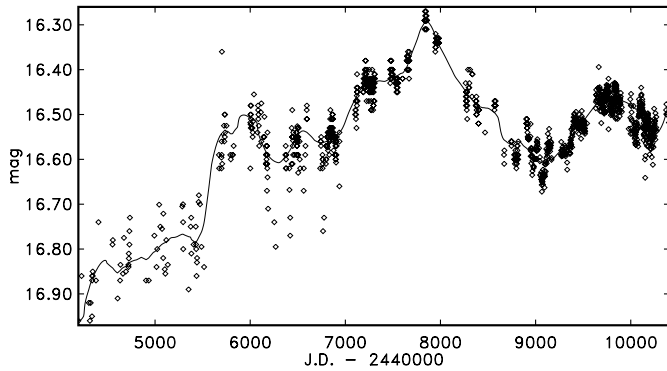


Fig. 2. Master data set, image B. The continuous line is the smoothed version of it ($\Delta t = 300$ days, smoothing polynomials degree $d = 2$)

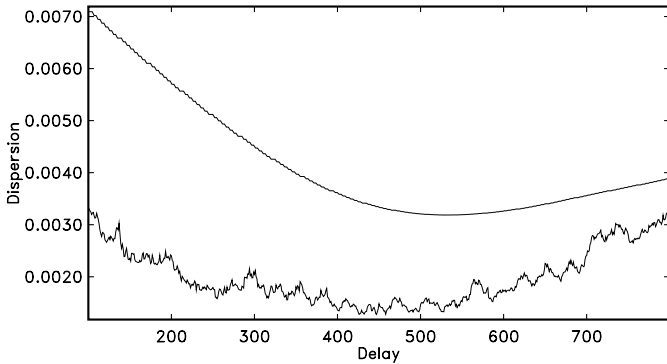


Fig. 3. Dispersion spectra D_1^2 for master data set with gaps filled (upper curve) and with original gaps (lower curve)

actual measurement data can shift the minimum. Changing the smoothing and the gap filling method or parameters of the procedure can change the way the spectrum is distorted. An overall drastic shift of the minimum can often be observed.

We still believe that the optimal interpolation method of Press et al. 1992a, 1992b failed not because of nongaussianity of the underlying process (see Press & Rybicki 1997), but because of unfortunate sampling and distortions caused by microlensing.

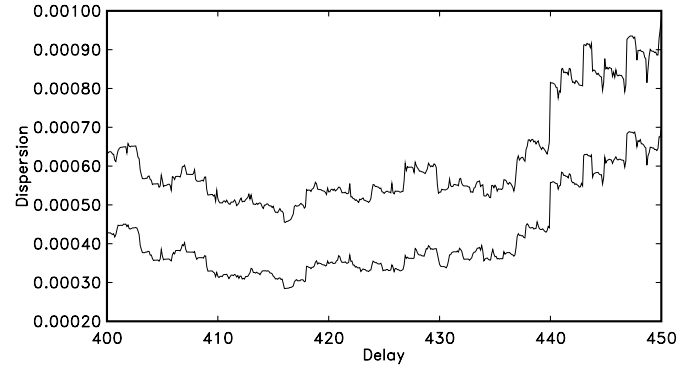


Fig. 4. Dispersion spectra $D_1^2(\tau)$ for two iterations, downweighting parameter $\Delta t = 720$ days for smoothing with local parabolic fit $d = 2$

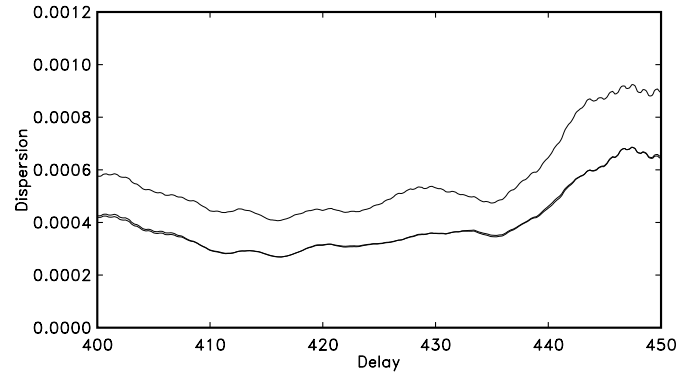


Fig. 5. Dispersion spectra $D_2^2(\tau)$ for three iterations, downweighting parameter $\Delta t = 720$ days for smoothing with local parabolic fit and $\Delta t = 2.5$ days for dispersion estimate

3.3. Basic iteration

The algorithm described above for time delay estimation contains a free function $l(t)$ which can be determined in two ways. We can postulate a parametric model for $l(t)$ (for instance a polynomial) and then minimize over the shift τ and the parameters in the model of $l(t)$ all in one procedure. Or we can proceed iteratively. For the current analysis we chose the second method. First we estimated the shift $\tau \approx 416$ days from the raw data. Then we estimated the $l(t)$ curve (see below). This estimate was then used to correct the B curve. After that the modified B curve was used together with the original A curve to get a new estimate for τ etc.

The $D_1^2(\tau)$ dispersion measure does not depend on any arbitrarily chosen parameter. For $D_2^2(\tau)$ however, we must fix the value for the downweighting parameter Δt . In Paper II we used the value $\Delta t = 20$ days because it gave statistically stable and robust spectra. Here we are interested in smaller details of the spectrum and consequently choose $\Delta t = 2.5$ days for $D_2^2(\tau)$. For both dispersion estimation methods we needed only one or two iterations to obtain convergence (see Table 3 and Figs. 4-5). The difference curve $l(t)$ was estimated in the following way. For every time point t in the shifted B curve ($B^{(\tau)}$) we looked for the nearest point (in time) in the A curve. The A and $B^{(\tau)}$ pair was included into the fit if the distance between A and

Table 3. Iterations to obtain correct value for τ

Iteration	Delay	D_1^2
0	416.1	0.0004547
1	416.1	0.0002845
		D_2^2
0	416.0	0.0004067
1	416.3	0.0002691
2	416.3	0.0002685

$B^{(\tau)}$ did not exceed $\delta t = 5.5$ days. The time points for differences were calculated as a simple mean between A timepoint and $B^{(\tau)}$ timepoint. More than half of the total differences were effectively exact hits ($\delta t < 0.5$ days). Weights for differences were calculated according to Eq. 1. After computing the differences we used local quadratic least squares fits with linear downweighting with $\Delta t = 720$ (see Appendix A. and Eq. 3) to get a continuous estimate for $l(t)$.

The entire iteration procedure is robust against moderate changes in the setup. Use of smoothing with different downweighting parameters did not change delay values by more than one day.

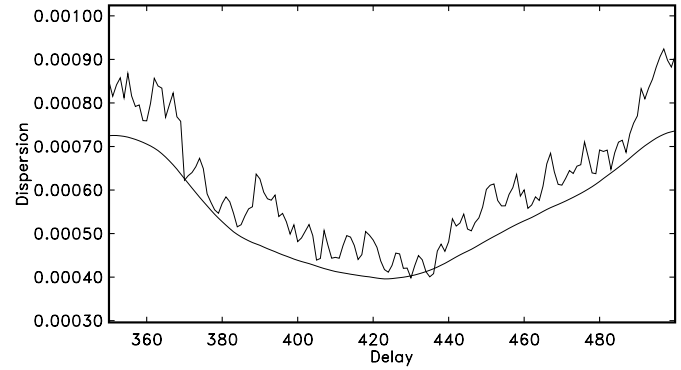
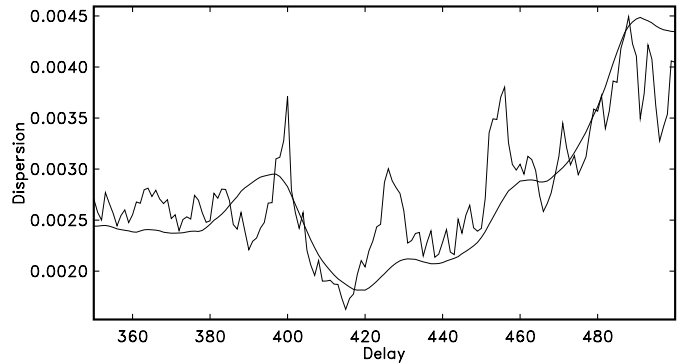
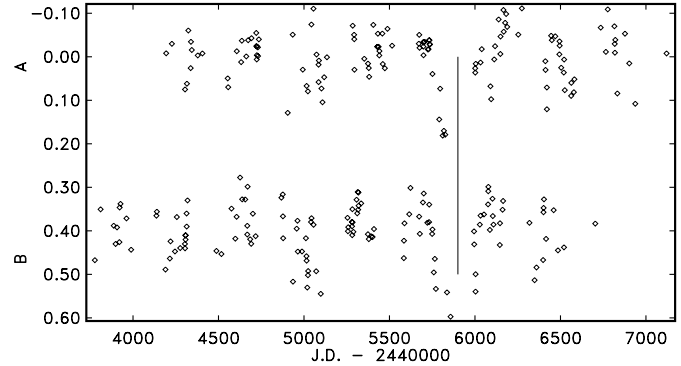
We estimated the rms error for the D_2^2 minimum by using bootstrapping as in Papers I and II. The final result 416.3 ± 1.7 agrees well with the value obtained from KG data set by Kundić et al. 1997.

3.4. $423 \neq 416$?

We were slightly surprised by getting the $\tau = 416$ day value from the master data set because the statistically stable value from Paper II (423 ± 6) was somewhat larger (see also Oscoz et al. 1997). To fully understand the reason why we earlier got a slightly different value (although nearly inside 1σ error limits) we revisited our previous analyses. In Fig. 6 the D_1^2 and D_2^2 dispersion spectra for the first 831 data point pairs compiled by R. Schild (as used in Paper II) are depicted. The D_1^2 curve is so erratic that it is not reasonable to assume that one or the other local minimum is a correct time delay estimate. However, the D_2^2 curve shows a nice, unique and stable minimum in the region.

In Fig. 7 the spectra of D_1^2 and D_2^2 are plotted for the Vanderriest 1989 data (as used in Paper I). There is a sharp minimum at 415 days in the D_1^2 spectrum and a less sharp minimum at 418 days in the D_2^2 spectrum (after removal of the low frequency component). Both of them (taking into account the photometric precision) are in extremely good agreement with recent estimates (Kundić et al. 1997). But still, from a statistical analysis it is quite easy to see that this minimum results mostly from the relatively fast decrease and following increase in the A lightcurve, later repeated in the B lightcurve (“Vanderriest episode”) depicted on Fig. 8.

In conclusion, it seems now that the quasar source variability indeed contains some decreases or increases like the “Vanderriest episode” and the “Kundić episode” on relatively short time scales. Often these events are mixed with observational

**Fig. 6.** Dispersion spectra for Paper II optical data, D_1^2 – erratic, D_2^2 – smooth**Fig. 7.** Dispersion spectra for Paper I optical data, D_1^2 – erratic, D_2^2 – smooth**Fig. 8.** “Vanderriest episode” in the early data (marked by a vertical line). The B data are shifted by 417 days and by 0.4 mag.

uncertainties, possible microlensing events or by the results of excursions due to the unknown systematic errors on different time scales. They can also be masked statistically when we use heterogeneous data sets with widely different error levels. Due to the general asymmetry of the large scale features in the source curve it can then happen that smoothed estimates of the type D_2^2 (with $\Delta t > 10.0$) can be slightly biased and sharp minima in the D_1^2 curves can be misleading (see Fig. 6). A similar conclusion was reached by Schild & Thomson 1997 who found that structure in the brightness curves causes multiple local minima and different calculations gave significantly different results;

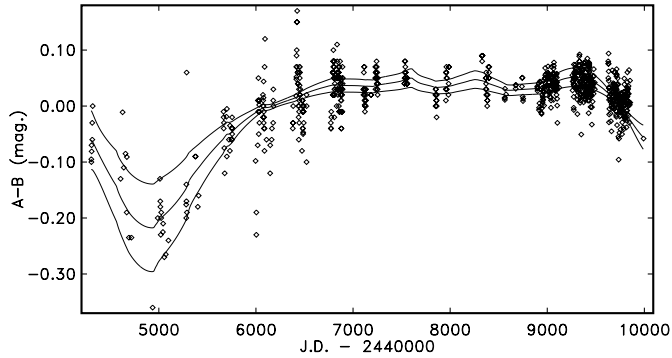


Fig. 9. Estimate for $A - B^{(\tau)}$ and its smoothed version. Upper and lower continuous curves depict 1σ spread corridors obtained from bootstrap runs

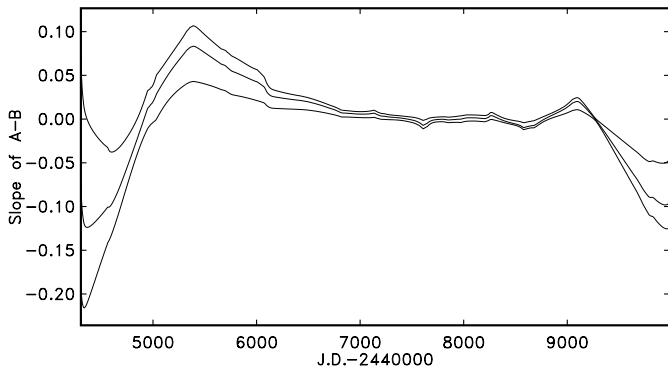


Fig. 10. Estimate for a slope of $A - B^{(\tau)}$ with 1σ error corridors

their final 404-day time delay estimate, computed by using an evolved gap filling scheme, has a large error bar consistent with the present 416-417 day value.

3.5. Microlensing curve

By removing a smooth long timescale component $l(t)$ from the B curve we largely improved the fit between the A and the time shifted B (below $B^{(\tau)}$) curve and decreased the overall dispersion estimate by 37% for D_1^2 and 34% for D_2^2 (see Table 3). As a byproduct of this iteration scheme we got an estimate for $l(t) = A - B^{(\tau)}$ curve (see Fig. 9) and also an estimate for its slope (see Fig. 10). The estimates of slopes are computed from the same locally fitted polynomials as the curve itself. This approach allows us to suppress wild oscillations in the slope estimates by the methods which use numerical differentiation of already computed curve estimates.

The error corridors depicted in Figs. 9 and 10 can not be interpreted as point estimates for every single point on the curves. They simply show 1σ limits for the distribution of 1000 bootstrap runs. Every single run is a continuous smooth curve (as is the basic estimate itself) and consequently its points are strongly correlated. Without any doubt the microlensing curve significantly differs from constant but it is very hard to quantitatively estimate the overall significance for such a complicated data set.

By changing smoothing methods or procedure parameters (say downweighting parameter Δt) we can slightly change the appearance of the $l(t)$ curve, but the total minimum to maximum variability level of 0.27 mag. and the maximum steepness of rising and downfalling parts of 0.083 mag per year and -0.124 mag per year, respectively, remain practically the same for a wide range of setups. This allows us to bracket some physically important parameters (see Discussion).

3.6. Analysis of residuals

Unfortunately the smooth component, which we interpret as the microlensing, is not enough to describe the full $A - B^{(\tau)}$ variability. The residual standard deviation after removal of this smooth component from the $A - B^{(\tau)}$ curve is 0.0242. If we compare this value with the expected rms error value 0.0197 we are still left with unexplained variability at the level $\sigma = 0.0145$. For so heterogeneous data as our master data set this can be considered as insignificant. But when we make analogous computations with more exact and homogeneous data obtained at Mt Hopkins (SR) during the last four years we again find a significant residual variability. The expected error in this case is just $\sqrt{2} \times 0.0100 = 0.0141$, the standard deviation for $A - B^{(\tau)}$ is 0.0210 and consequently the unexplained variability is at the level of 0.0155.

In the remaining parts of the paper we will carefully examine characteristics and the possible sources for this excess variability.

4. Short time scale variability

From this point onward we fix the time delay value $\tau = 416.3$ and start to analyse the available data sets assuming that this is the correct value. Because the observational sequences contain relatively sharp increases and decreases, on time scales of days and weeks, this decision is quite important (see additional discussion in Sect. 4.3.1 and Table 7). For other values of possible delays the pattern of residuals for the very short time scale features can be quite different (at least on the level of quantitative estimates).

4.1. Cross dispersion between data sets

Because of the relatively low amplitude of the unexplained short timescale variability we cannot use the master data set as a basis for analysis.

The differences of the photometries (see Appendix B) in different observatories are well demonstrated in Table 4 and in Fig. 11 (worst fit case) and Fig. 12 (best fit case). The cross-diagrams show quite large scatter. However it is hard to find any hints about systematic differences in the qualities of the photometries. Even when we consider the closeness of OR estimates to SR (when compared with OR distance to KR or KG photometry) we can be misled because of the small number of points in the OR data set. The observers of OR data also give

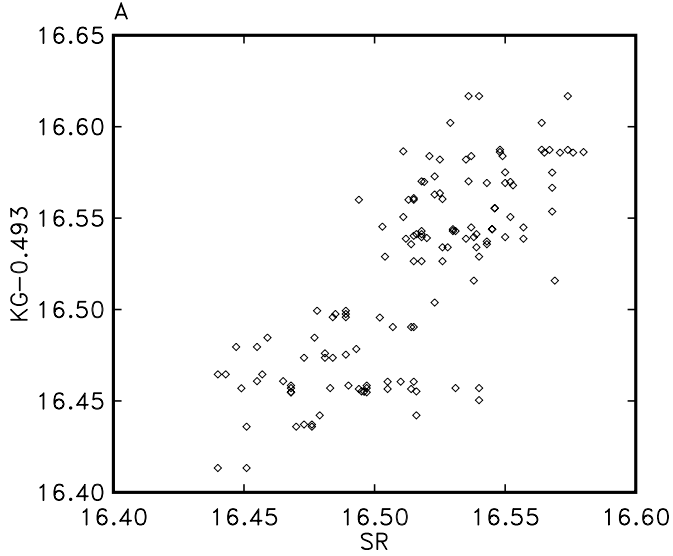


Fig. 11. Scatter between SR and KG, A image, “worst case”

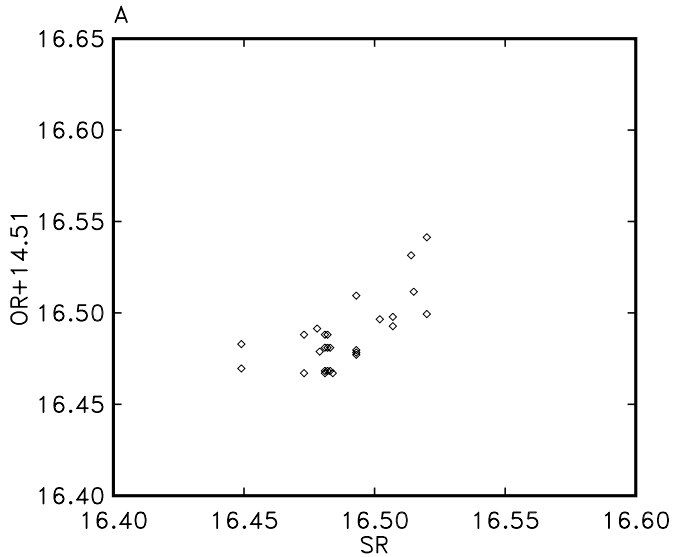


Fig. 12. Scatter between SR and OR, A image, “best case”

somewhat higher rms errors for their data, when compared with other groups.

We comment specifically on some points from Table 4. In the rows SR,OR for images A and B we note the small dispersion estimates 0.0103 and 0.0104 respectively. This suggests that the Mt Hopkins (SR) photometries and Oscoz (OR) photometries agree quite well. This agreement is partly due to the fact that observation nights that were common for SR and OR data occurred at times with good seeing (compare Figs. 11-12). In the rows KR,OR comparing Kundić and Oscoz photometries for images A and B, the dispersion estimates are 0.0207 and 0.0108 respectively, suggesting that there can be some systematic error in the KR photometry A curve. The Kundić KG photometry agrees quite well with Oscoz photometry, but the largest differences

Table 4. Cross-dispersions $D_{x,y}$ (see Appendix B) between data sets

Data sets	Image	N_{pairs}	$D_{x,y}$
SR,KR	A	134	0.0181
SR,KR	B	136	0.0174
SR,KG	A	133	0.0235
SR,KG	B	135	0.0196
SR,OR	A	27	0.0103
SR,OR	B	27	0.0104
KR,KG	A	114	0.0141
KR,KG	B	114	0.0237
KR,OR	A	33	0.0207
KR,OR	B	33	0.0108
KG,OR	A	26	0.0158
KG,OR	B	26	0.0150

are between SR and KG measurements and between KR and KG photometries in B curve.

It is also interesting to note that undoing the background galaxy correction (see Sect. 4.3) for SR B image data significantly improves the fit between SR and OR (from 0.0104 to 0.0088) and between SR and KR (from 0.0174 to 0.0147), but the fit with KG data becomes slightly worse (from 0.0196 to 0.0215). As the cross-dispersion analysis does not rule out conclusively any of the data sets, in the following we will consider them all to be equally informative for the task of establishing the nature of excess variability on the shorter time scales.

4.2. Dispersion of $A - B^{(\tau)}$

The simplest test to analyse excess variability is to look at the distribution of differences $A_i - B_j^{(\tau)}$ when $A_i, B_j^{(\tau)}$ occur nearby after shifting the B data by the assumed time delay.

Put formally, we will construct from the original data sets $A_i, i = 1, \dots, N_A$ and $B_j^{(\tau)}, j = 1, \dots, N_B$ a new sequence:

$$C_k = C\left(\frac{t_i + t_j}{2}\right) = A_i - B_j^{(\tau)}, \quad (12)$$

with statistical weights

$$W_k = W_{i,j} S_{i,j}^*, \quad (13)$$

where $S_{i,j}^*$ downweights linearly the pairs according to the time difference between A_i and $B_j^{(\tau)}$. Using the C_k sequence we can now form a dispersion estimate which will depend on the downweighting parameter Δt :

$$D^2(\Delta t) = \frac{\sum_{k=1}^K W_k C_k^2}{2 \sum_{k=1}^K W_k}, \quad (14)$$

where K is the total number of pairs $A_i, B_j^{(\tau)}$ with nonzero weight. When Δt increases the dispersion $D^2(\Delta t)$ will also generally increase and asymptotically it must approach the value

$$D^2(\infty) = \frac{1}{2N_A N_B} \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} W_{i,j} (A_i - B_j^{(\tau)})^2. \quad (15)$$

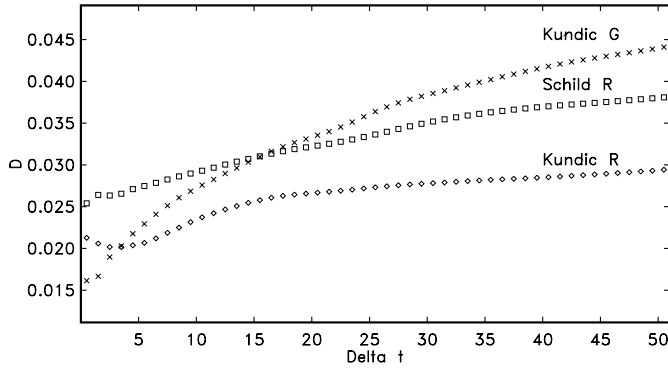


Fig. 13. Dependence of D on Δt

If there is no excess source of variability, then the lowest values of $D^2(\Delta t)$ must be about $2\sigma^2$ where σ is the mean observational error. To have a fair base for comparisons we cut from the SR data set a subset (in time) which has the same time span as the KR and KG data. In Fig. 13 the three curves for $D(\Delta t)$ are depicted. Although all three data sets are quite similar when plotted together, the curves on Fig. 13 show strong differences. There are three important differences to be noticed.

First, and most prominent, is the fact that the KG photometry attains the expected noise level for small values of Δt . Correspondingly - for the short time episode covering the common span of A and $B^{(\tau)}$ photometry for the KG data there are no excess variabilities at all. However, observing this, we must take into account that the total number of $A_i, B_j^{(\tau)}$ pairs for short Δt is rather small: 15 for $\Delta t = 0.5$, 25 for $\Delta t = 1.5$, 57 for $\Delta t = 2.5$ etc. The SR subset is more abundant, the number of pairs is 28 for $\Delta t = 0.5$, 79 for $\Delta t = 1.5$, 131 for $\Delta t = 2.5$.

The second observation is that the KR and KG data sets behave differently. Whereas the general dispersion of differences for KR is lower than for KG, in the first part of the graph its dispersion is higher. We interpret this as follows: the overall amplitude of KR variability is smaller than in KG and consequently the differences (after removing the mean) tend to be less. However, because both KR data are noisier than the KG data, we see the excess dispersion for low Δt values.

Finally, once again we clearly see that the SR data contain an excess variability component well above the expected error level.

In summary we have now conflicting evidence. The KG photometry can be well explained in the context of a simple lensing model, without any need to introduce concepts of short time scale excess variability. The KR photometry seems to be of poorer quality (this is also claimed by the observers). And finally the SR photometry which is more abundant shows consistent excess short time variability at least on the level twice that expected from a simple model. None of the data sets can be disregarded or ignored on the ground of poor quality (we checked this by computing cross-dispersions and also by introducing third party data as a calibration device, see Table 4).

Table 5. Stated and estimated errors and total variability for SR(1992-1996) subset

Image	N_{pairs}	σ_{std}	σ_{est}	σ_{tot}
A	471	0.0099	0.0103	0.0371
B	471	0.0094	0.0100	0.0576

Table 6. Minima in the D_2^2 spectra for SR(1992-1996) subset

Data	Statistic	τ	D_2^2
Original	D_1^2	416.5	0.000348
Original	D_2^2	419.5	0.000417
Uncorrected	D_1^2	416.5	0.000275
Uncorrected	D_2^2	419.0	0.000333
Detrended	D_1^2	416.0	0.000169
Detrended	D_2^2	419.0	0.000223

4.3. Fast variability in SR data set

To analyse more closely the variability of $A - B^{(\tau)}$ differences on shorter timescales we analysed the most densely sampled subset of the SR data - the last four years. There are 471 pairs of A and B measurements in this subset. Principal data for both images are given in Table 5, where σ_{std} is the mean rms deviation estimated by the observers, and σ_{est} is defined in Eq. 4. The total variability σ_{tot} (weighted standard deviation for the all data) for the B data is significantly higher than for the A data. One of the reasons for this is that the values of the B image for SR data are corrected for an underlying galaxy contribution using a fixed value 18.3 mag. as an estimate for the galaxy brightness (see Schild & Cholfin 1986). For uncorrected B values the total variability is less ($\sigma_{tot} = 0.0482$). It was interesting to see that the fit between A and $B^{(\tau)}$ is significantly better for uncorrected data (see Figs. 14-15 and Table 6). In aperture photometry the total amount of light inside the aperture depends strongly on seeing conditions. On the nights with bad seeing a significant part of the light is dispersed out of the aperture. This effect was expected to be slight because the galaxy image is more diffuse than the quasar image, and images with $FWHM > 2.5''$ were rejected. However some effect may be still present and consequently it is not totally correct to use a constant galaxy brightness value for all the nights independent of seeing conditions. Since the seeing data are not available for a realistic correction we chose to use uncorrected B image photometry, which gives better fit between A and $B^{(\tau)}$.

4.3.1. Statistical characterization of the excess variability

In Fig. 16 the estimated smooth microlensing curve is depicted together with the original values of $A - B^{(\tau)}$ (only exact hits - pairs with $\delta t \leq 0.5$; downweighting parameter $\Delta t = 720$ was chosen the same as for the master data set). The change from year to year is statistically significant (see Appendix C.) at least at the level 99.999% (100 000 bootstrap runs, 0 hits, 53 adjacent observations). However, after removal of the low frequency component the residuals can be considered as nonrandom only with significance 84.0% ($\sigma = 0.0178$, 100 000 bootstrap runs,

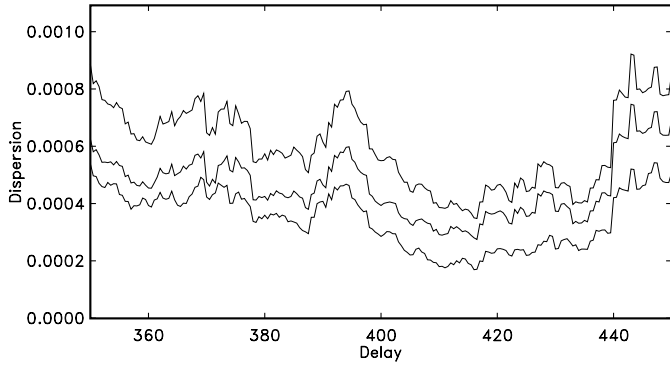


Fig. 14. Dispersion spectra $D_1^2(\tau)$ corrected for galaxy contribution data (upper curve), uncorrected data (middle curve) and detrended uncorrected data (lower curve)

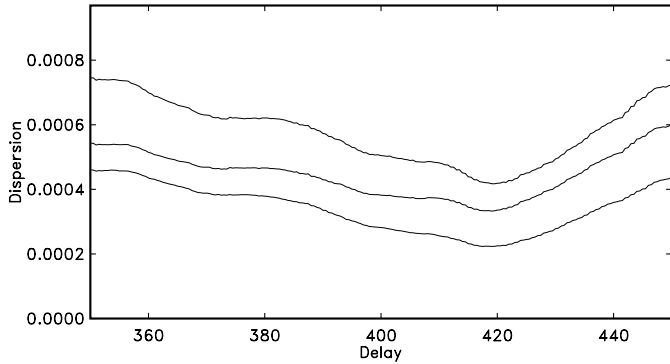


Fig. 15. Dispersion spectra $D_2^2(\tau)$ for original data (upper curve), uncorrected data (middle curve) and detrended uncorrected data (lower curve)

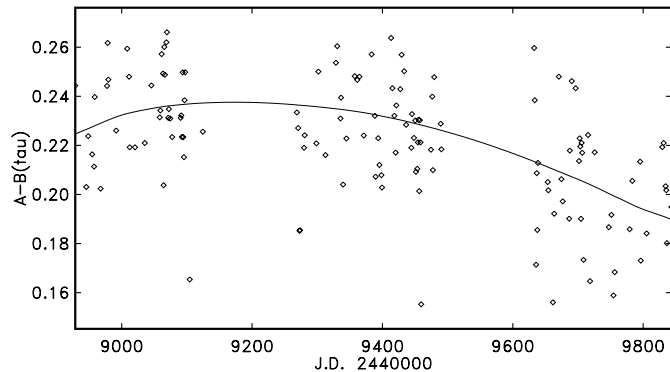


Fig. 16. Microlensing curve in SR(1992-1996)

15972 hits, 53 adjacent observations). In Table 7 the analogous results are depicted for delays 415 – 420 days. The absolute minimum at 417 days is additional evidence that this delay value can be considered as the most probable. But the most important result comes from bootstrap calculations. We see that the pattern of residuals very strongly depends on the delay value used. Apparently this is a result of two important properties of our data sets: very uneven distribution of observations in time and quite fast variability of the source and possible small amplitude

Table 7. Statistics of residuals for different time delays after removal of the smooth component

Delay	σ	Significance
415	0.0194	33.7%
416	0.0182	39.0%
417	0.0178	15.7%
418	0.0192	6.1%
419	0.0198	17.9%
420	0.0193	4.4%

fast microlensing. Slight changes in the applied time delay can cause large changes in the pattern of $A - B^{(\tau)}$ residuals.

5. Discussion

Using the simple methods of Papers I and II we re-estimated the time delay value for the extended data set of 1216 A and 1217 B image measurements. The obtained delay $\tau = 416.3 \pm 1.7$ was used to estimate the $A - B^{(\tau)}$ curve. Using a simple smoothing scheme we divided the obtained differences into two components. The low frequency component with full amplitude 0.27 ± 0.08 mag is certainly a significant feature. This is proved not only by the current analysis but also by the fact that its existence had a strong impact on the time delay controversy around QSO 0957+561 (see Schild & Cholfin 1986, Vandierriest et al. 1989, Press et al. 1992a, 1992b, Papers I,II and Kundić et al. 1995, 1997). The physical reality of the second, high frequency variability component (on the level $\sigma = 0.013 - 0.015$ mag) can not be established firmly using the available data. It is important also to note that there is a certain gray area between the low frequency variations of the $A - B^{(\tau)}$ curve and the high frequency variations for which observational sequences are available during separate years. The changes with typical time scales around one year are difficult to analyse because of the long gaps between different seasons.

The most important information in the $A - B^{(\tau)}$ lightcurve is the largest variation, Δm , and the corresponding timescale Δt . During the 17 years of monitoring the most prominent event is a variation with $\Delta m = 0.25$ mag in $\Delta t = 5$ yrs, from the beginning of 1982 to the end of 1986, see Fig. 9. From an analysis similar to that in Refsdal & Stabell 1993 (hereafter RS) for QSO 2237+0305, we find that a microlens variation in $A - B^{(\tau)}$ of 0.25 mag or more is highly improbable in QSO 0957+561 if the source radius is larger than $20R_0$. We therefore take $R = 20R_0$ as an upper limit for the relative source size. Here R_0 is the Einstein ring radius of a typical microlens projected into the source plane, and we have used lens parameters $\kappa_A = 0.22$, $\gamma_A = 0.17$, $\kappa_B = 1.24$ and $\gamma_B = 0.9$ (J. Lehar, 1997, private communication). Furthermore, by a simple scaling of Eq. 6 in RS we get (with $F \equiv R/R_0$):

$$\frac{F}{5} 30t_0 \gtrsim \Delta t = 5 \text{ yrs} \gtrsim \frac{F}{5} 8t_0, \quad (16)$$

since the case discussed in RS was $R = 5R_0$. From numerical tests we have found that the scaling argument is roughly correct

for $F \geq 1$. Note that we have here included an upper limit of Δt which follows by inspecting Figs. 3 and 5 in RS (the longest event lasted $30t_0$).

With $R_0 = 5 \times 10^{16} \sqrt{M/M_\odot}$ and $t_0 = R_0/V \approx 25\text{yrs} \sqrt{M/M_\odot} \times 600\text{kmks}^{-1}/V$ it is now possible to constrain M :

$$\begin{aligned} \frac{M}{M_\odot} &\lesssim 1.4 \times 10^{-2} F^{-2} \left(\frac{V}{600\text{kmks}^{-1}} \right)^2 \left(\frac{\Delta t}{5\text{yrs}} \right)^2, \\ \frac{M}{M_\odot} &\gtrsim 10^{-3} F^{-2} \left(\frac{V}{600\text{kmks}^{-1}} \right)^2 \left(\frac{\Delta t}{5\text{yrs}} \right)^2. \end{aligned} \quad (17)$$

Note that V is the effective transversal velocity in the source plane with the time measured by the observer (see Kayser et al. 1986). For $F = 20$ (and $V = 600\text{kmks}^{-1}$) we get a lower and upper limit for M/M_\odot of about 2.5×10^{-6} and 3.5×10^{-5} respectively, and for $F = 1$ we get 10^{-3} and 1.4×10^{-2} . We thus see, as also for the Einstein Cross (see RS), that microlens masses down to about $10^{-5} M_\odot$ can not be ruled out.

For $F < 1$ our conclusions are less certain since the simple scaling argument on which the inequalities of Eq. 17 are based, may not be applied anymore. By inspecting simulated light curves for $F < 1$ we find that lens masses as high as $1M_\odot$ are possible but rather improbable. One difficulty with so high masses and small values of F is that caustic crossing events usually have too large amplitudes. However, with a small fraction ($\lesssim 5\%$) of the lens mass in solar mass objects and the rest in objects down to about $0.05M_\odot$ or less, the observed microlens event has a higher probability to occur. Extensive numerical simulations must be carried out before more precise statements can be made.

The best constrained parameter is the absolute radius of the source which roughly should be $R \approx V\Delta t/2$ (within a factor of 3), see RS. Note that this is also valid for much smaller values of R/R_0 when the events are caused by the crossing of a caustic. For $V = 600\text{kmks}^{-1}$ we then get $R \approx 3 \times 10^{15}\text{cm}$. It is interesting to note that the two lens systems QSO 2237+0305 (see Wambsganss et al. 1990 and RS) and QSO 0957+561 indicate about the same value of R .

A more detailed analysis of the ‘‘allowed’’ regions in the (M, R) plane will follow in a subsequent paper (see also the interesting parallel development by Schmidt & Wambsganss 1998).

6. Conclusions

We demonstrated that the 416 – 417 day value for the time delay between A and B images of the double quasar can be firmly determined from the available data. This gives additional weight to the estimate obtained by Kundić et al. 1997 using only a single episode in the lightcurves.

The obtained time delay value was then used for the estimation of the microlensing curve $A - B^{(\tau)}$. We divided this curve into two parts. The long timescale component is an unquestionably significant feature, but this can not be said about the variability on shorter timescales.

The idea of observing microlensing on the shorter timescales (6-120) days is of course quite intriguing and it is very hard to ignore the evidence of it in the original SR data. We had hoped to find strong evidence for it from a careful comparison of the photometries of several observatories, but in spite of dedicated observing the intercomparisons were inconclusive because of inadequate overlap of the data sets, and apparent systematic errors in some data sets. We were able to significantly reduce the dispersion between the A and $B^{(\tau)}$ component by using data uncorrected for the galaxy brightness. It is not ruled out that the use of more detailed information about seeing conditions during the SR photometry allows one to apply these corrections more precisely. The physical implications of the possible short timescale variability will be discussed in a subsequent paper.

Acknowledgements. We would like to thank the Research Council of Norway who through the Nordic Scholarship Scheme for the Baltic Countries and Northwest Russia supported financially one of the authors (J.P.). We also thank the J. Wambsganss and the referee for several helpful comments.

Appendix A: smoothing techniques

To get estimates for the microlensing curve we explored three types of smoothing methods:

- Fitting polynomials to the full $A - B^{(\tau)}$ data set.
- Spline fitting with unequally spaced nodes.
- Local fit of low degree polynomials with linear downweighting.

The first approach is good in terms of its simplicity. But actual trial calculations demonstrated that the global nature of the polynomial fit can introduce unwanted artifacts. For instance the 9-th degree polynomials can have up to the 8 alternating extrema, and they tend to show themselves in our case, where the difference curve is nearly flat most of the time.

The second method is in principle more flexible. However, there is no automatic way to assign positions for fitting knots for such heterogeneous data that we have here. Manual positioning is ruled out if we want to follow the general methodology of objective assessment.

The third method has its own pitfalls. There are some points in time where the local fit is sensitive to minor changes in time or placement of observations. However, because it is implementable as a fully automatic procedure with only two free parameters (polynomial degree and downweighting parameter Δt) we chose it as the basic smoothing procedure. There is an additional important bonus in the third method; we can estimate the slope of the approximation curve together with its value. For the first two methods the slope estimate will be effectively the slope of the approximation curve and all artifacts of the estimation procedure will manifest themselves strongly in the slope. For the local smoothing the slope is estimated for every single point separately and it is basically as stable as is the local approximation itself.

For an arbitrary time point t we estimate a probable value for the microlensing curve $\hat{y}(t)$ by minimizing:

$$D^2(t) = \min_{p_d} \sum_{i=1}^N S^*(t - t_i) W_i (y_i - p_d(t_i - t))^2, \quad (\text{A1})$$

where $p_d(t)$ is a low degree polynomial and the $S^*(t)$ -s are additional weights

$$S^*(t) = \begin{cases} 1 - \frac{|t|}{\Delta t}, & \text{if } |t| \leq \Delta t, \\ 0, & \text{if } |t| > \Delta t. \end{cases} \quad (\text{A2})$$

The estimate $\hat{y}(t)$ is then the value of the best fitting polynomial ($\hat{y}(t) = p_d(t)$) and the slope can be estimated as the derivative $p'_d(0)$ of the local polynomial.

The error bars for the estimates are computed using the bootstrap method (see Efron & Tibshirany 1986) with reweighting. For every bootstrap run we reshuffled the normalized residuals $\epsilon_i = \frac{\hat{y}(t_i) - y(t_i)}{\sigma_i}$, $i = 1, \dots, N$, added them to the estimate

$$y^*(t_i) = \hat{y}(t_i) + \sigma_i \epsilon^*, \quad (\text{A3})$$

and repeated the estimation with a new data set. The row of artificial errors ϵ^* is just a random sample taken (without replacement) from the set of original errors $\epsilon(t_i)$. Doing so we obtained 1000 different solutions whose distribution corridors are depicted in Fig. 9. An important point here is that all bootstrap runs had effectively the same weighting structure as the main estimation run. This is very important, if we take into account the unevenly distributed precision of our data set.

Appendix B: cross-dispersion estimation

In order to estimate the cross-dispersion between two data sequences $x_i, i = 1, \dots, N1$ and $y_j, j = 1, \dots, N2$ we used the minimum

$$D_{x,y}^2 = \min_a \frac{\sum_{i=1}^N 1 \sum_{j=1}^N 2S_{i,j} W_{i,j} (x_i - y_j + a)^2}{2 \sum_{i=1}^N 1 \sum_{j=1}^N 2S_{i,j} W_{i,j}}, \quad (\text{B1})$$

where $S_{i,j}$ and $W_{i,j}$ are defined in Sect. 1.1. Inclusion of the correction a allows us to take into account differences in magnitude levels. Because $\delta t = 0.5$ (exact hits) usually gave very low numbers of pairs to be compared, we used instead $\delta t = 1.5$.

Appendix C: significance estimation for the trends

In order to search for significant departures from randomness in the residuals we modified the well known method of Durbin&Watson (see Draper & Smith 1981). For the row of observations $y_i(t_i), i = 1, \dots, N$ and for statistical weights $W_i(t_i)$ we computed a dispersion estimate

$$D^2(\delta t) = \frac{1}{2N} \sum_{i=1}^{N-1} S_{i,i+1} W_{i,i+1} (y_i - y_{i+1})^2, \quad (\text{C1})$$

where the $S_{i,j}$ -s exclude well separated pairs (see Eq. 2) and $W_{i,j}$ are combined weights (see Eq. 1). We then compared it

with the same kind of sums which were computed with reshuffled data. The significance of departure from randomness was then estimated as the ratio of cases when the original dispersion was less than the dispersion for the random row of the same data points. We used typically 100 000 runs of reshuffling to compute the levels of significance.

References

- Chang K., Refsdal S., 1979, Nat. 282, 561
 Draper N.R., Smith H., 1981, *Applied Regression Analysis*, John Wiley & Sons
 Efron B., Tibshirany R., 1986, Statistical Science 1, 54
 Goicoechea L.J., Oscoz A., Mediavilla E., Buitrago J., Sierra-Ricart M., 1998, ApJ 492, 74
 Haarsma D.B., Hewitt J.N., Lehár J., Burke B.F., 1997, ApJ 479, 102
 Kayser R., Refsdal S., Stabell R., 1986, A&A 166, 36
 Kent S., 1985, PASP 97, 165
 Kundić T., Colley W.N., Gott III R., et al., 1995, ApJ 455, L5
 Kundić T., Turner E.L., Colley W.N., et al., 1997, ApJ 482, 75
 Oscoz A., Serra-Ricart M., Goicoechea L.J., Buitrago J., Mediavilla E., 1996, ApJ 470, L19
 Oscoz A., Mediavilla E., Goicoechea L.J., Sierra-Ricart M., Buitrago J., 1997, ApJ 479, L89
 Pelt J., Hoff W., Kayser R., Refsdal S., Schramm T., 1994, A&A 256, 775 (Paper I)
 Pelt J., Kayser R., Refsdal S., Schramm T., 1996, A&A 305, 97 (Paper II)
 Press W.H., Rybicki G.B., 1997, in: *Astronomical Time Series*, eds. D. Maoz et al., Kluwer Academic Publishers, p.61
 Press W.H., Rybicki G.B., Hewitt J.N., 1992a, ApJ 385, 404
 Press W.H., Rybicki G.B., Hewitt J.N., 1992b, ApJ 385, 416
 Refsdal S., Stabell R., 1993, A&A 278, L5 (RS)
 Schild R., 1990, AJ 100, 1771
 Schild R., 1998, A wavelet exploration of the Q0957+561 A,B brightness record, ApJ, submitted
 Schild R., Cholfin B., 1986, ApJ 300, 209
 Schild R., Thomson D.J., 1997, in: *Astronomical Time Series*, eds. D. Maoz et al., Kluwer Academic Publishers, p.73
 Schmidt R., Wambsganss J., 1998, Limits on MACHOs from microlensing in the double quasar Q0967+561, A&A, submitted
 Thomson D.J., Schild R., 1997, in: *Time Series Analysis in Astronomy and Meteorology*, eds. T. Subba Rao and M.B. Priestley, Chapman & Hall, p.187
 Vanderriest C., Schneider J., Herpe G., et al., 1989, A&A 215, 1
 Wambsganss J., Paczyński B., Schneider P., 1990, ApJ 358, L33