THE ANALYSIS OF SOME BIVARIATE ASTRONOMICAL TIME SERIES

Marthinus Christoffel Koen

Submitted in fulfilment of the requirements for
the Master of Science degree at the
University of Cape Town

October 1993

ACKNOWLEDGEMENTS

I am grateful to my thesis supervisor, Professor Walter Zucchini, for telling me, amongst other things, about Cassandra. I appreciated his willingness to spend time discussing a wide range of topics in statistics, often at short or no notice.

I am also grateful to the former Director of the South African Astronomical Observatory, Professor Michael Feast, for allowing me to indulge my interest in data analysis; and to his successor, Doctor Robert Stobie, for continuing the tradition.

Mrs. Audrey Hultzer was responsible for the high quality of the diagrams.

ABSTRACT

In the first part of the thesis, a linear time domain transfer function is fitted to satellite observations of a variable galaxy, NGC5548. The transfer functions relate an input series (ultraviolet continuum flux) to an output series (emission line flux). The methodology for fitting transfer function is briefly described. The autocorrelation structure of the observations of NGC5548 in different electromagnetic spectral bands is investigated, and appropriate univariate autoregressive moving average models given. The results of extensive transfer function fitting using respectively the λ1337 and λ1350 continuum variations as input series, are presented. There is little evidence for a dead time in the response of the emission line variations which are presumed driven by the continuum.

Part 2 of the thesis is devoted to the estimation of the lag between two irregularly spaced astronomical time series. Lag estimation methods which have been used in the astronomy literature are reviewed. Some problems are pointed out, particularly the influence of autocorrelation and non-stationarity of the series. If the two series can be modelled as random walks, both these problems can be dealt with efficiently. Maximum likelihood estimation of the random walk and measurement error variances, as well as the lag between the two series, is discussed. Large-sample properties of the estimators are derived. An efficient computational procedure for the likelihood which exploits the sparseness of the covariance matrix, is briefly described. Results are derived for two example data sets: the variations in the two gravitationally lensed images of a quasar, and brightness changes of the active galaxy NGC3783 in two different wavelengths. The thesis is concluded with a brief consideration of other analysis methods which appear interesting.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

A typical example of the type of data with which the greater part of this thesis is concerned, is shown in Figs. 1(a) and (b). These are U (ultraviolet, short wavelength) and K (infrared, long wavelength) photometric observations of the active galaxy NGC 3783 (Glass 1992). It has been postulated that variations in the U band are due to events in the galactic nucleus, and that some of this high energy flux is absorbed and re-radiated by dust which is at some distance from the nucleus (Clavel, Wamsteker & Glass 1989). If this is correct, one expects to see changes in the K radiation lagging those in U, with the lag determined by the distance between nucleus and dust formations. Conversely, if one can determine the lag between the two sets of observations, the nucleus-dust distance follows. The implicit statistical model relating the observations z_K and z_U of the galaxy is

$$z_K(t) = Az_U(t+\tau) + \zeta(t)$$

where τ is the lag, A is constant, and $\zeta(t)$ is white noise; the quantity of primary interest is the lag τ.

Various techniques for estimating the time offset τ between two irregularly sampled light curves such as those in Fig. 1 have been used by astronomers. The simplest of these is linear interpolation to produce data at regularly spaced time points, followed by calculation of the cross-correlation function (ccf). The lag is then identified as that time shift corresponding to the maximum of the ccf. Edelson & Krolik (1988) have pointed out the dubiousness of interpolating the observations, and have proposed instead binning data pairs which are at more or less the same time separation, in order to obtain a better estimate of the ccf. These authors also remarked on the fact that autocorrelation in the individual series hampers interpretation of the ccf. This point is well known in the statistics literature (e.g. Box & Newbold 1971). The remedy which is usually applied, viz. removing the autocorrelation by prewhitening, cannot however be used for the data of Fig. 1, due to the irregular time spacing. This point, as well as the the effects of non-stationarity in the process means, will be discussed at some length in Part 2, which deals particularly with irregularly spaced observations.

Some progress may be made with the somewhat intricate problem of unevenly observed series by studying a simpler example; this may help to gain some insight into the nature of time series such as those in Fig. 1, and hence postulating a suitable statistical model. Fig. 2 shows a very rare type of astronomical time series: the observations were obtained at almost regular intervals of about four days. Since the object studied is also an AGN ("Active Galactic Nucleus", i.e. a galaxy showing brightness variations in its central regions), in this case the galaxy NGC 5548, one might hope that its behaviour will be similar to that of NGC 3783. In fact, in this case it is believed that variations in the continuum flux (Fig. 2a) drive those in the line or discrete emission features (Fig. 2b). Part 1 of the thesis deals with the analysis of data such as those in Fig. 2. The method used is standard: time domain transfer functions are fitted. In this case far more information than just the value of τ can easily be extracted from the observations.

By using a parametric model for the observations z_K and z_U, which is suggested by the results

Figure 1: Photometric observations of the variable galaxy NGC 3783 (Glass 1992). (a) K band (b) U band

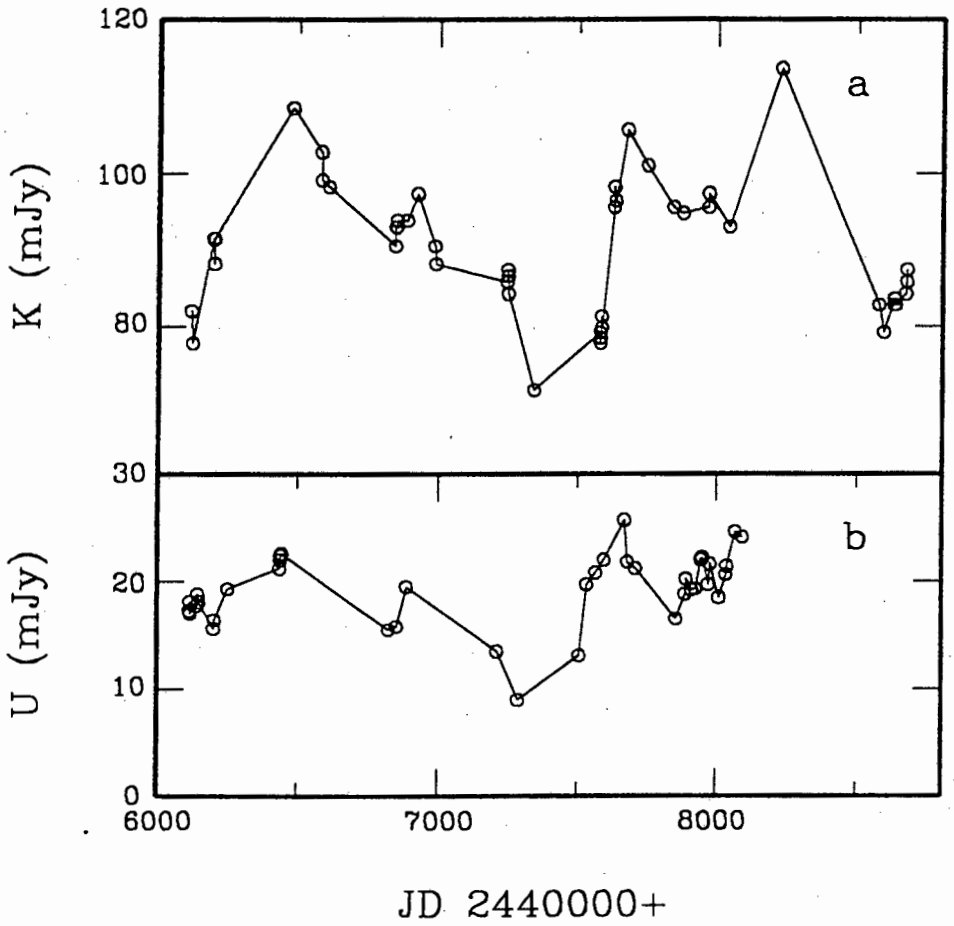
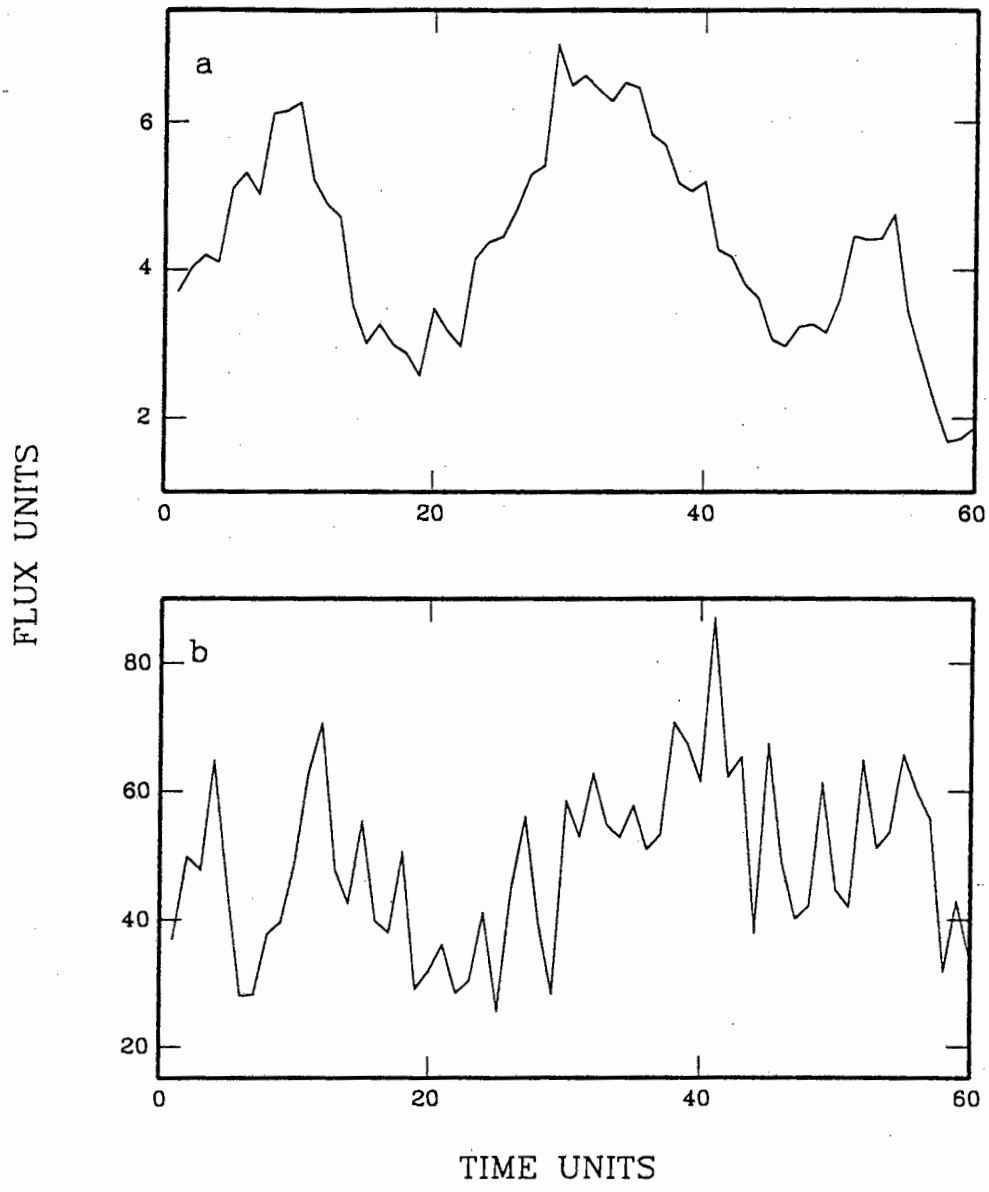


Figure 2: Two typical sets of IUE observations of NGC5548: (a) the λ1350 continuum (b) SiIV + OIV] λ1402 line emission (reduced by the Spectral Image Processing System - see CEA91). The unit of time is approximately 3.99 days.



of Part 1, the covariance matrix of the joint process can be written down for any assumed lag. On adopting a suitable bivariate distribution for the observations, the "true" lag can then be estimated by the maximum likelihood method. This programme is carried out in Part 2 for the data of Fig. 1, as well as for a second set of irregularly spaced observations of considerable current (early 1990s) astrophysical interest. The thesis is concluded by a brief consideration of a number of other approaches to studying the relationship between two irregularly observed time series.

PART 1. A TRANSFER FUNCTION ANALYSIS

1.1 INTRODUCTION

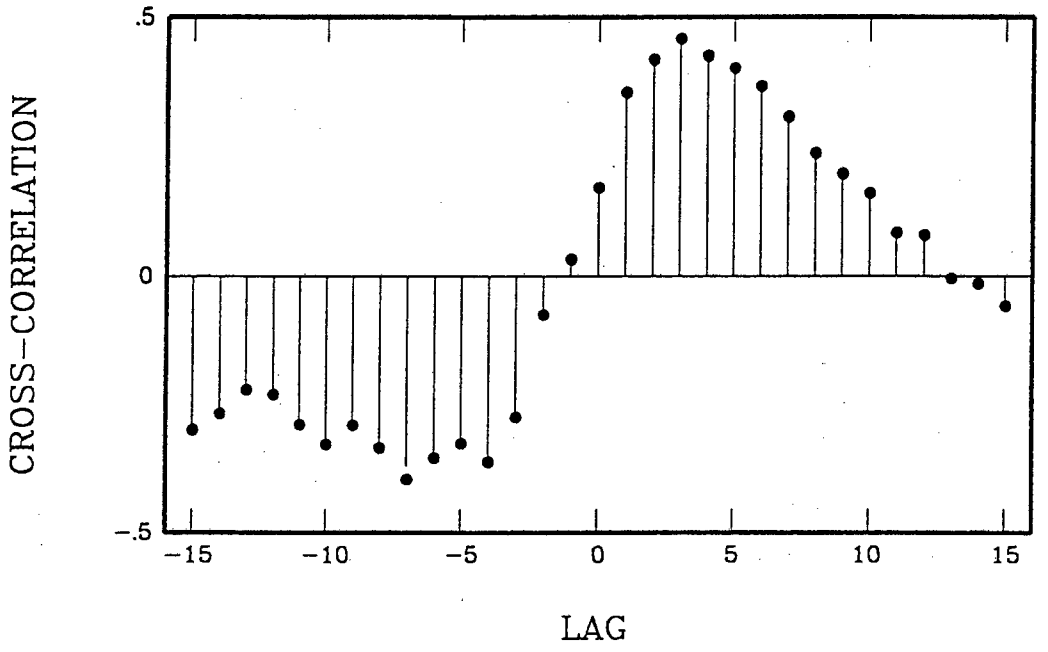
Clavel et al. (1991) (henceforth referred to as CEA91) describe and analyse the variability of the galaxy NGC5548 as observed with the International Ultraviolet Explorer (IUE) satellite over an eight month period. In particular they study the time relation between the variability in different electromagnetic spectral bands, and estimate the time delays between intensity fluctuations in the ultraviolet continuum and in emission lines which are presumed to be continuum-driven. The time delays are estimated by use of a cross correlation function (ccf) applied directly to the observations. Fig. 2 shows two typical series of observations obtained by CEA91, and Fig. 3 shows their ccf

$$r_{xy}(\tau) = \frac{1}{N S_x S_y} \sum_{t=1}^{N-\tau} [x(t) - \overline{x}]\,[y(t+\tau) - \overline{y}]$$

where $\overline{z}$ and S_z are the mean and the standard deviation of the N observations of z. It is concluded from the position of the maximum in the ccf plot that the SiIV + OIV] λ1402 emission fluctuations lag the λ1350 continuum fluctuations by three time units (i.e. about 12 days). Throughout, the continuum wavelength is referred to by the central wavelength of the band over which it was measured; the unit of measurement is Ångstroms. The roman numerals appended to the chemical species responsible for a given emission feature refer to the ionisation state, with "I" indicating neutral atoms. Thus "OIV" refers to three times ionised Oxygen. The square bracket indicates that the atomic transition is semi-forbidden, i.e. is unobservable in the laboratory under usual conditions. Note also that the IUE measurements were converted to electromagnetic flux units by two different techniques, the "Spectral Image Processing System" (SIPS) and "Gaussian Extraction" (GEX) methods. These are decribed in some detail by CEA91. The particular flux data set used will often be indicated in this thesis by use of the suffixes "G" or "S".

CEA91 mention a number of problems which hamper quantitative use of the ccf results, such as the absence of reliable error estimates for the time delays. They also point out that the large width of the ccf peaks (which causes some uncertainty in the value of the time delays) can be ascribed at least in part to the underlying autocorrelation of the continuum observations. This is an important point which has a bearing on the substantial ccf peaks found by CEA91; Jenkins (1979) gives an example of two series for which the highly significant cross correlation is entirely due to autocorrelation (see also Pierce 1977, and the very interesting paper by Box & Newbold 1971). It is not being suggested that all large cross correlations found by CEA91 are spurious: the point is that the possibility of an underlying mechanism which causes similar behaviour of all the series over time needs to be ruled out before any causal relationship can be definitely assumed. The question is thus whether (say) the continuum observations contain any information about the emission line variations which is not simply derivable from the autocorrelation structure of the latter series. In order to answer this, account needs to be taken of the role of autocorrelation

Figure 3: The cross-correlation function of the two data sets of Figure 2.



in the ccfs. A method for doing this will be described below.

This part of the thesis is concerned with the calculation of transfer functions, i.e. linear functions relating the continuum and emission line variations. The necessary statistical theory is described in the next section. As has been made clear in the preceding paragraph, the autocorrelations of the individual series are of importance. These are investigated in section 1.3. Section 1.4 presents the results of transfer function calculations.

1.2 TIME DOMAIN TRANSFER FUNCTIONS

A very condensed introduction to linear transfer function modelling is presented below; a standard text is Box & Jenkins (1976), which deals with the topic in great detail.

In its simplest form the transfer function relating the input series $\{x_t\}$ to the output series $\{y_t\}$ is defined by

$$y_t = U_0 x_{t-\tau} + U_1 x_{t-\tau-1} + U_2 x_{t-\tau-2} + \cdots + U_s x_{t-\tau-s}$$

where the U_j are constants and τ is the response lag or "dead time" between the input impulse x and the response y. The equation can be written somewhat more concisely in terms of a backshift operator B defined by

$$B\,x_t \equiv x_{t-1} \qquad B^2\,x_t \equiv x_{t-2} \qquad \cdots \qquad B^n\,x_t \equiv x_{t-n}$$

as

$$y_t = (U_0 + U_1 B + U_2 B^2 + \cdots + U_s B^s) x_{t-\tau} \tag{1}$$

In (1), and indeed in all the equations to follow, it is implicitly assumed that the statistical processes are of zero-mean form. This is easily arranged by subtracting the mean value of the series from each observation. Note though that where this adjustment is needed, the series mean constitutes an extra model parameter to be estimated.

In practice the relationship between the input and output series is stochastic, rather than deterministic, and (1) is suitably generalised to

$$y_t = (U_0 + U_1 B + U_2 B^2 + \cdots + U_s B^s) x_{t-\tau} + N_t \tag{2}$$

where N_t is a "noise" process. Information on the modelling of the univariate process N_t can be found in Box & Jenkins (1976) or any modern time series analysis textbook; see also Scargle (1981) for a description aimed at an astronomical audience. Here it is simply noted that N_t may in general be written as an autoregressive moving average (ARMA) process:

$$N_t = \sum_{i=1}^{p} \alpha_i N_{t-i} + \sum_{j=1}^{q} \beta_j \varepsilon_{t-j} + \varepsilon_t \tag{3}$$

where ε_t is a completely uncorrelated process. In many cases a single autoregressive (AR) or moving average (MA) term is sufficient to model "noise" processes.

It is often found that the impact of an impulse declines approximately geometrically over time, e.g.

$$y_t = U_0(1 + S_1B + S_1^2B^2 + \cdots)x_{t-\tau} + N_t$$

This may be written more concisely as

$$y_t = \frac{U_0}{1 - S_1B}x_{t-\tau} + N_t$$

A very general and flexible form of (2) is thus

$$y_t = \frac{\sum_{i=0}^{s} U_iB^i}{1 - \sum_{j=1}^{r} S_jB^j}x_{t-\tau} + N_t \tag{4}$$

The response lag τ may be conveniently absorbed into the numerator of (4):

$$y_t = \frac{\sum_{i=\tau}^{s+\tau} U_iB^i}{1 - \sum_{j=1}^{r} S_jB^j}x_t + N_t \tag{5}$$

which is the form of the transfer function to be used in this thesis.

The ccf of $\{x_t\}$ and $\{y_t\}$ is used to identify likely values of U_i, S_j and τ in (5). However, this is only possible if the distorting influence of the autocorrelation in the $\{x_t\}$-series has been removed. This is not as difficult as one might anticipate; both input and output series are prewhitened or filtered by the ARMA structure of the input series $\{x_t\}$. A general strategy for fitting transfer functions is outlined below:

(i) An ARMA model is fitted to the input series:

$$x_t = \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \xi_{t-j} + \xi_t \tag{6}$$

Inspection of the acf is an essential part of this procedure, which is described in detail by e.g. Box & Jenkins (1976), Pankratz (1983). A model is judged tenable if the $\{\xi_t\}$ in (6) are uncorrelated, and if the parameters α_i and β_j are statistically significant. If more than one model satisfies these requirements, that with the minimum value of the Akaike or Bayesian information criteria

$$AIC = \frac{2(p+q)}{N} + \ln \sigma^2 \qquad BIC = \frac{(p+q)\ln N}{N} + \ln \sigma^2$$

may be selected. In these expressions $p + q$ is the number of model parameters fitted, N the number of observations, and σ^2 the mean squared residual or residual variance (see e.g. Harvey 1989). (Note that when observations have been transformed to zero-mean form by subtraction of the series mean, the value of $p + q$ should reflect the extra parameter estimated). The information criteria supply an objective means for deciding between the opposing requirements of keeping the number of fitted parameters to a minimum and attaining a small residual sum of squares.

(ii) The ARMA model is used to filter both input and output series:

$$\xi_t = x_t - \sum_{i=1}^{p} \alpha_i x_{t-i} - \sum_{j=1}^{q} \beta_j \xi_{t-j}$$

$$\eta_t = y_t - \sum_{i=1}^{p} \alpha_i y_{t-i} - \sum_{j=1}^{q} \beta_j \eta_{t-j}$$

Note that the ξ_t are not subject to autocorrelation.

(iii) Calculate the ccf of the $\{\xi_t\}$ and the $\{\eta_t\}$.

(iv) Inspection of the ccf will suggest possible forms for the transfer function (5). Transfer function parameter calculation can be done with commercially available software such as GENSTAT, IMSL or BMDP.

(v) The residuals from the transfer function fit constitute the noise process $\{N_t\}$ in (5) and (3). An ARMA model is fitted to this series, using the procedure outlined in (i). The residuals $\{\varepsilon_t\}$ should of course be uncorrelated.

(vi) The transfer function model can be checked by calculating the ccf of the residuals $\{\varepsilon_t\}$ and the prewhitened input series $\{\xi_t\}$. It should be statistically zero.

1.3 ARMA MODELS OF INDIVIDUAL SERIES

It is evident from the discussion of prewhitening above that ARMA models of the input series are required. It is also useful to study the correlation structure of the output series as similar models can often be fitted to the noise process $\{N_t\}$ as to $\{y_t\}$.

Before proceeding, a word about the data. The ARMA and transfer function methods are designed for data equally spaced in time, whereas there is some slight irregularity in the times of the NGC5548 observations. The observations reported in CEA91 were therefore linearly interpolated to provide 60 data points at an equal spacing of 3.99 days. All four data tables in CEA91 were interpolated to yield observations at the same time points.

Figure 4 shows the acf of the data of Figure 2a. The acf is typical of a first order autoregressive [i.e. AR(1)] process with positive coefficient

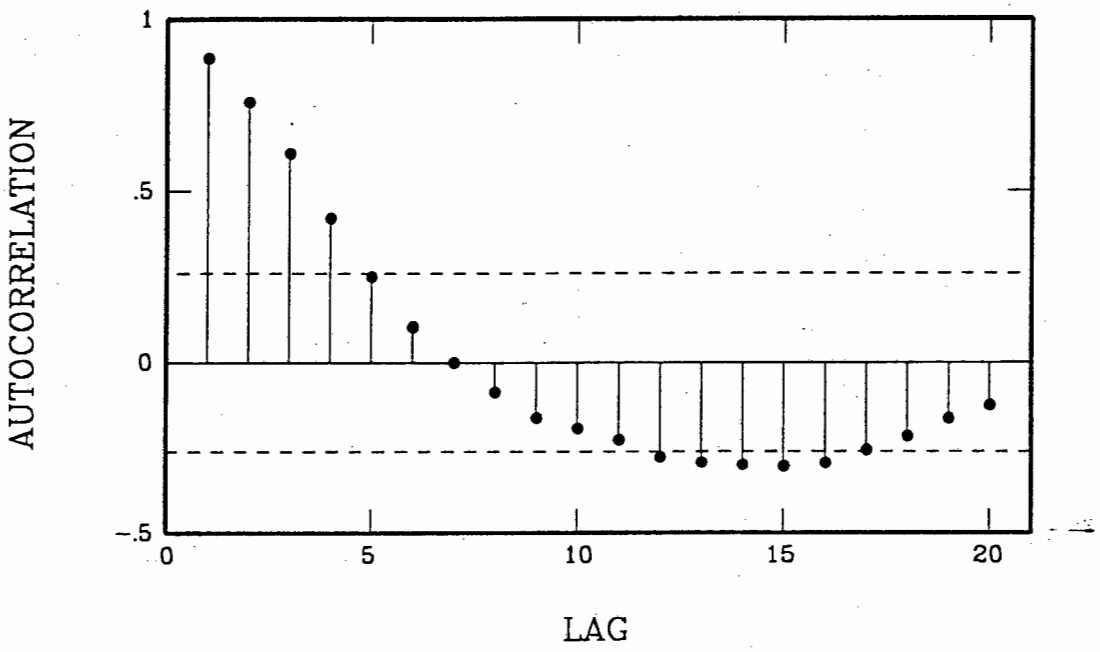
$$x_t = \alpha_1 x_{t-1} + \xi_t \qquad \alpha_1 > 0$$

(Box & Jenkins 1976). An AR(1) model in fact fits the λ1350 continuum observations quite well; $\alpha_1 = 0.94$ (s.e. 0.049) with residual variance $\sigma^2 = 0.302$.

The value of α_1 is not significantly different from unity. The AR(1) process with $\alpha_1 = 1$ is of course well known in the physical sciences as the random walk:

$$\Delta x_{t-1} = x_t - x_{t-1} = \xi_t \tag{7}$$

Figure 4: The autocorrelation function of the data in Figure 2a. The broken lines show approximate two standard deviation bounds for zero autocorrelation at a given lag. The standard deviation is given to good approximation by $N^{-1/2}$.



where the $\{\xi_t\}$ are uncorrelated random variables. Fixing $\alpha_1 = 1$ gives essentially the same residual variance as for the AR(1) model. Furthermore, the mean of the Δx_t may be assumed to be zero and hence does not need to be estimated, which again reduces the information criteria. Assuming $\alpha_1 = 1$ is therefore appealing from both the standpoints of the statistics and the physics of the problem and this value will be adopted in what follows.

The residual acf of (7) is shown in Figure 5. It is not immediately obvious whether the spike at lag 3 implies the need for further model components: when simultaneously studying autocorrelations at a range of lags (20 in this case), one might expect some apparently significant features to arise by chance. The portmanteau statistic Q which is proportional to the sum of the squared autocorrelations measures the overall significance of the acf (Box & Jenkins 1976). Calculated for the first 10 lags of the acf in Figure 5, Q is significant at only about the 18% level. Nonetheless, the situation is not straightforward: including a third order MA-term ($\beta_3 = 0.35$, s.e. 0.13) in the model for the observations decreases the residual variance from $\sigma^2 = 0.302$ to $\sigma^2 = 0.271$, which is a meaningful reduction according to both the AIC and BIC criteria. Furthermore, it is also possible to fit statistically significant lag 3 parameters to the continuum observations in other wavebands, so that a closer look is definitely called for. Figure 6 shows quite clearly that the origin of the large lag 3 correlation is the occurrence of large (in absolute value) brightness increments $|\Delta x_t|$ at $t = 4$, 10, 19, each followed by a similar increment exactly three time units later. This conclusion is confirmed by the scatterplot Figure 7, in which Δx_t is plotted against Δx_{t-3}; the three marked points are those which result from the observations noted above. A straight line fit to the scatterplot has a slope $\alpha_3 = 0.30$, which is significant at the 3% level.

Some judgement on the side of the modeller is needed to resolve the issue. Inspection of Figure 6 tells one that the features causing the large lag 3 correlation are restricted roughly to the first third of the series of observations. The large correlation is thus not a general feature of the data, as can be confirmed by studying the acf of series in which respectively the first or last 20 observations are deleted. Even more convincing, deletion of data points at $t = 4$ and $t = 19$ (the origins of the points marked A and B in Figure 7) gives a scatterplot for which the slope differs from zero at a mere 19% level; the Q statistic for this reduced data set is significant at only the 43% level (first 10 autocorrelations). It may be concluded that there is fairly strong evidence for lag 3 repetitions of large $|\Delta x_t|$ in the early part of the series, but that this is a transient feature of the λ1350 data. Similar considerations apply to the other continuum data.

Based on the above, the final ARMA model adopted for all the short wavelength continuum data sets is (7). Furthermore, in what follows the differenced form of all the observation series will be used, i.e. the flux increments over time units of 3.99 days rather than the actual brightnesses will be analysed. This is not only statistically expedient, but may be physically more meaningful than working with the actual flux levels.

Acceptable ARMA models for the observations reported in Tables 1 to 4 of CEA91 are given in Table 1 of this paper. Several series of flux increments are statistically equivalent to random walks. Only one data set, the λ2670 continuum measurements obtained by the Gaussian extraction reduction method, required more than one model parameter for adequate modelling.

Figure 5: The autocorrelation function of the differenced series $\Delta x_t = x_t - x_{t-1}$ where the $\{x_t\}$ are the data in Figure 2a.

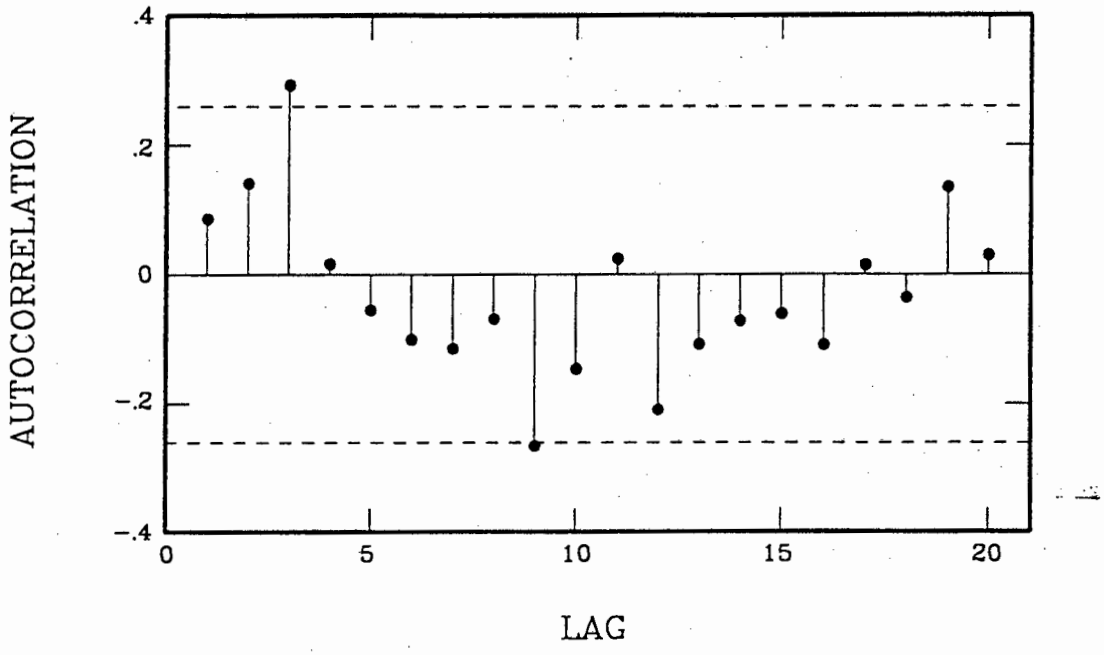


Figure 6: A plot of the differenced form of the data in Figure 2a.

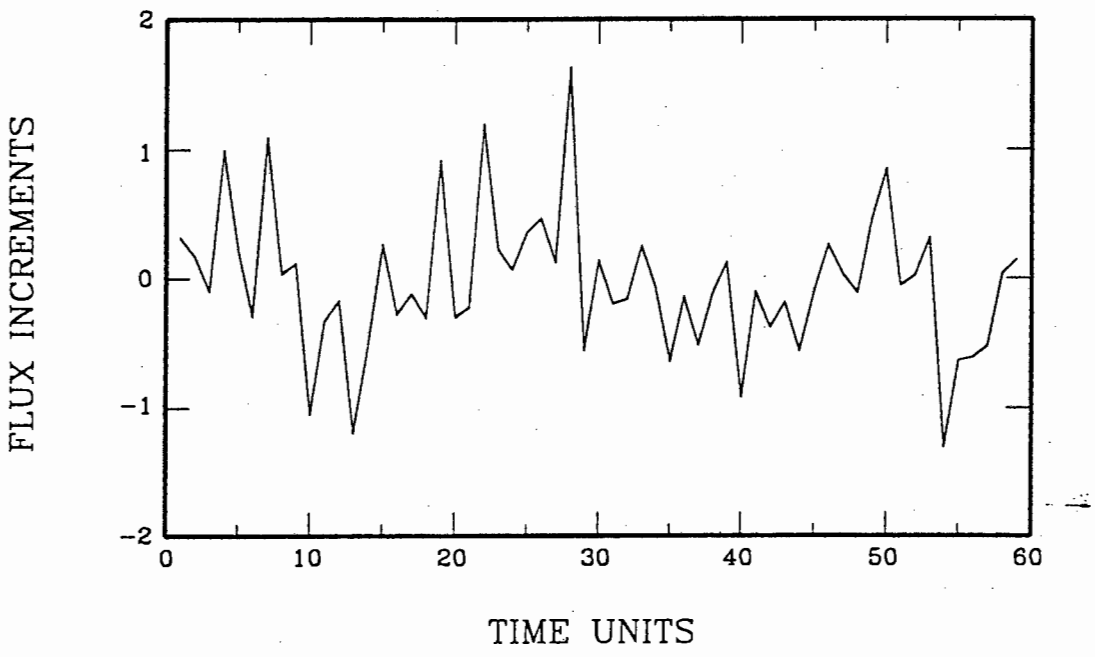


Figure 7: The data of Figure 6 plotted against itself lagged by three time units, i.e. Δx_t is plotted against Δx_{t-3}. Points of special interest are marked: A is (Δx_4, Δx_7); B is (Δx_{19}, Δx_{22}); C is (Δx_{10}, Δx_{13}).

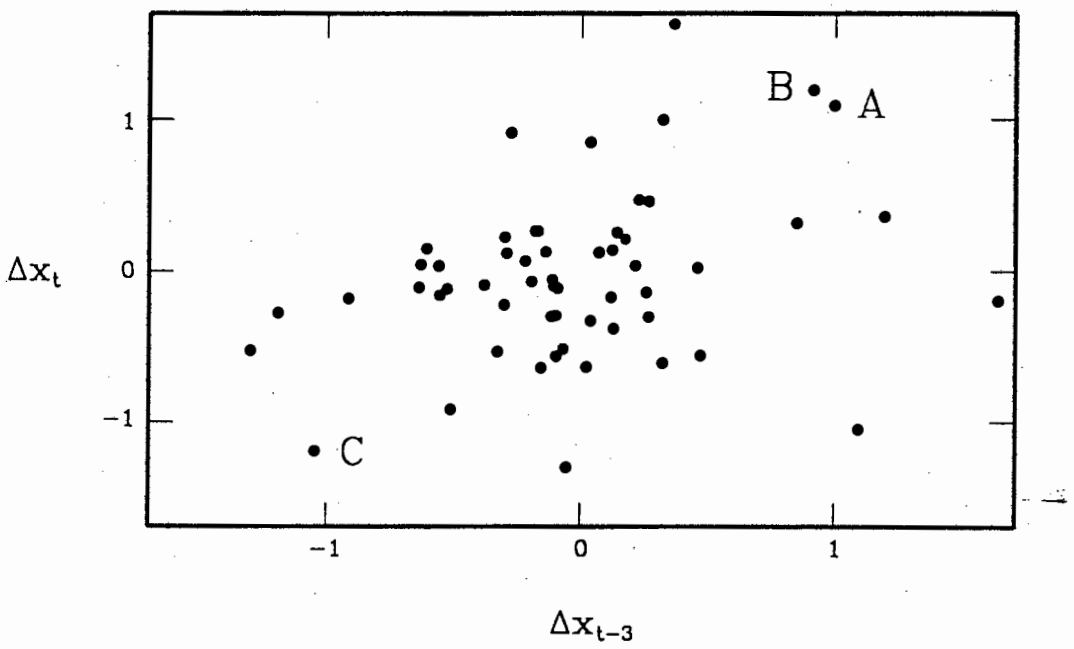


Table 1. Univariate models fitted to the data of CEA91: Suffix S denotes reductions by the IUE Spectral Image Processing System (SIPS); suffix G denotes reduction by the Gaussian Extraction technique (GEX; see CEA91 for details). FES is the Fine Error Sensor which measures wideband long wavelength flux. The Q-significance is the significance level of the portmanteau statistic of the first 10 autocorrelations.

Data Set	Parameter (s.e.)	Residual Variance	Q-Significance
λ1337 G		0.295	0.10
λ1350 S		0.302	0.18
λ1813 G		0.100	0.10
λ1840 S		0.112	0.37
λ2670 S	$\alpha_2 = 0.32$ (0.13)	0.022	0.54
λ2670 G	$\alpha_1 = 0.24$ (0.13)		
	$\alpha_2 = 0.31$ (0.13)	0.018	0.79
FES counts	$\alpha_1 = -0.35$ (0.12)	5.51	0.70
Lyα S	$\alpha_1 = 0.26$ (0.13)	1572.48	0.51
Lyα G		5797.85	0.10
CIV S		2867.59	0.58
CIV G		2549.82	0.46
SiIV S	$\beta_1 = -0.59$ (0.11)	158.47	0.69
SiIV G	$\beta_1 = -0.52$ (0.12)	144.75	0.58
HeII S		184.99	0.62
HeII G		315.05	0.88
CIII] S	$\beta_1 = -0.54$ (0.11)	319.23	0.56
CIII] G	$\beta_1 = -0.45$ (0.12)	225.19	0.81
NV G	$\beta_1 = -0.56$ (0.11)	1581.13	0.89
MgII S	$\beta_1 = -0.48$ (0.12)	51.59	0.62
MgII G	$\beta_1 = -0.73$ (0.09)	78.97	0.49

1.4 TRANSFER FUNCTIONS FITTED

Tables 2 and 3 contain salient features of the transfer functions fitted to the CEA91 data, with respectively the λ1350 and λ1337 data as input series. Before discussing these results, the fitting procedure is briefly described for the series of Figure 2.

Figure 8 shows the ccf between the differenced form of the two series displayed in Figure 1. The broken lines are approximate two standard error limits for the individual cross-correlations. The obvious starting point is to fit the model

$$y_t = U_1 x_{t-1}$$

where $\{x_t\}$ is the series of differenced λ1350 observations, and $\{y_t\}$ the differenced SiIV data. Inspection of the residual acf leads one to fit an MA(1) model to the noise $\{N_t\}$, giving

$$y_t = 5.03x_{t-1} - 0.74\varepsilon_{t-1} + \varepsilon_t$$

However, the ccf of the $\{\varepsilon_t\}$ and the $\{x_t\}$ (Figure 9) shows that the model is not adequate, requiring additional terms.

It is instructive to compare Figure 9 to Figure 10, the ccf obtained by prewhitening the input data by a β_3-term. The issue of a lag 3 term was discussed at some length in section 1.3 where it was concluded that it this feature of the continuum observations was not present throughout the data span. Nonetheless, taking account of it in calculating the ccf obviously serves the useful purpose of simplifying the latter. All ccfs reported below were therefore calculated after prewhitening the input series according to

$$\xi_t = x_t - \beta_3 \xi_{t-3}$$

The ccf of the residuals from the model

$$y_t = -8.68x_t + 12.35x_{t-1} - 0.84\varepsilon_{t-1} + \varepsilon_t$$

and the prewhitened input series is given in Figure 11; it appears satisfactory, and all coefficients are statistically significant. The acf of the $\{\varepsilon_t\}$ can be seen in Figure 12.

The following interesting points emerge from a study of the Tables:

(i) It is possible to model the longer wavelength continuum variations as being caused by λ1337 variations. There is some evidence in Table 3 that the relative contribution from lagged values of the λ1337 variations increases with increasing wavelength of the output series. It does not seem possible to adequately model longer wavelength continuum fluctuations in terms of λ1350 variations.

(ii) Models fitted to data obtained by the two different reduction techniques do not necessarily agree closely, and in some cases "good" models could be fitted to one but not the other data set. The implication is that any quantitative interpretation of the models should be done with caution.

Figure 8: The ccf of the differenced forms of the data in Figure 2. The broken lines show approximate two standard deviation bounds for zero cross-correlation at a given lag. The standard deviation at lag k is given by $(N-k)^{-1/2}$.

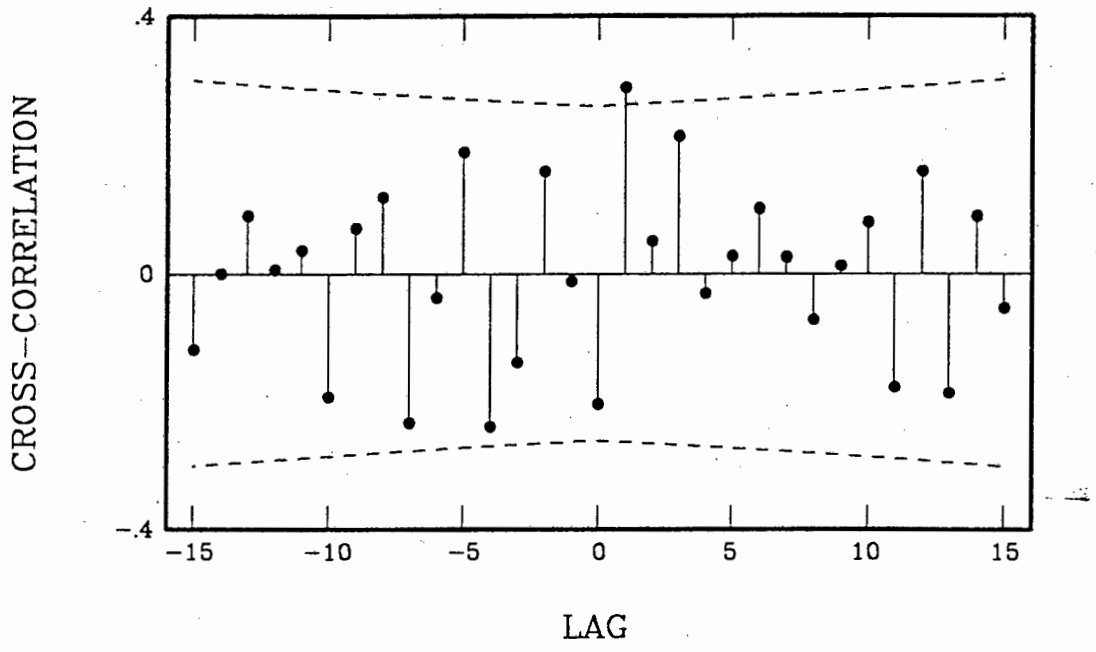


Figure 9: The ccf of the differenced input series of Figure 6 and the residuals of a two-parameter (U_1, β_1) transfer model fit to the differenced data of Figure 2b.

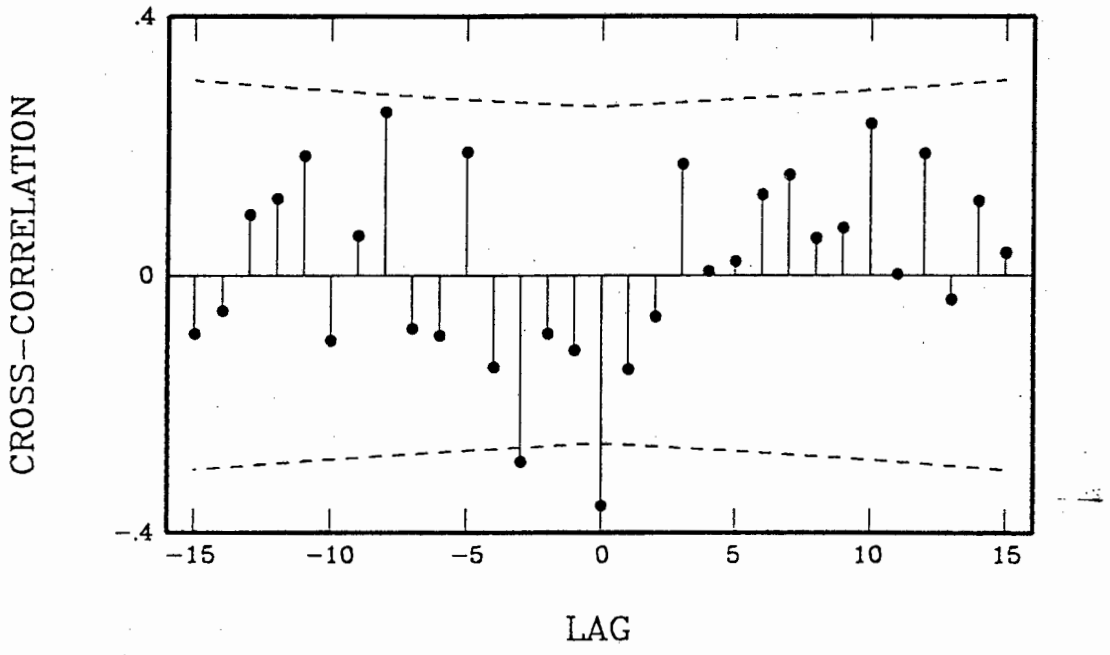


Figure 10: As in Figure 9, but with the input series prewhitened by a third order MA term.

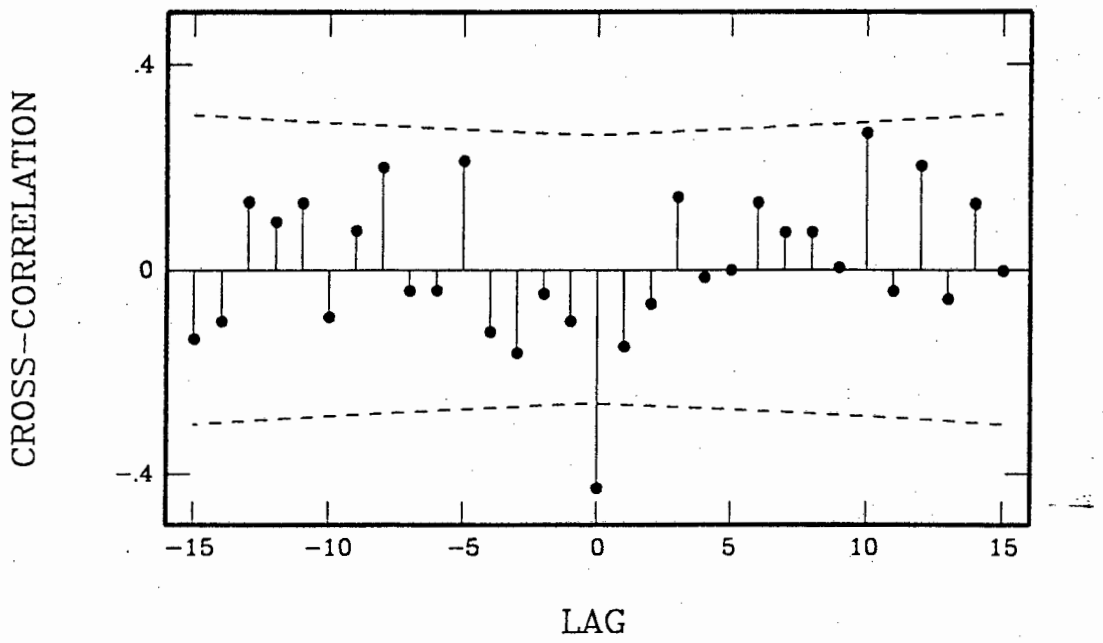


Figure 11: The ccf of the prewhitened data of Figure 6 and the residuals of a three-parameter (U_0, U_1, β_1) transfer model fit to the differenced data of Figure 2b.

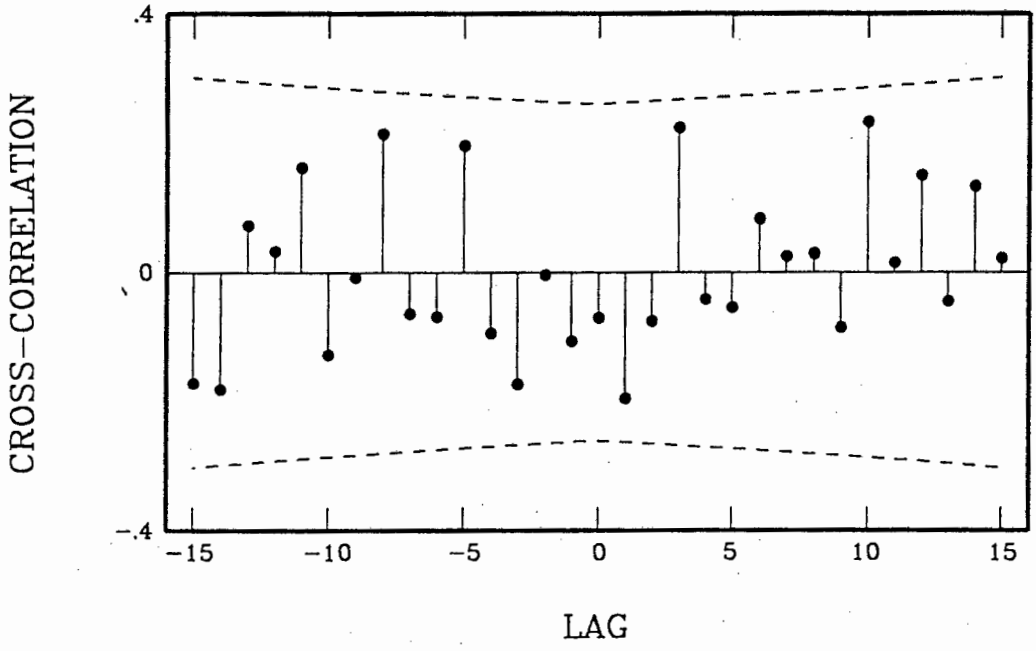


Figure 12: The acf of the residuals of a three-parameter (U_0, U_1, β_1) transfer model fit to the differenced data of Figure 2b.

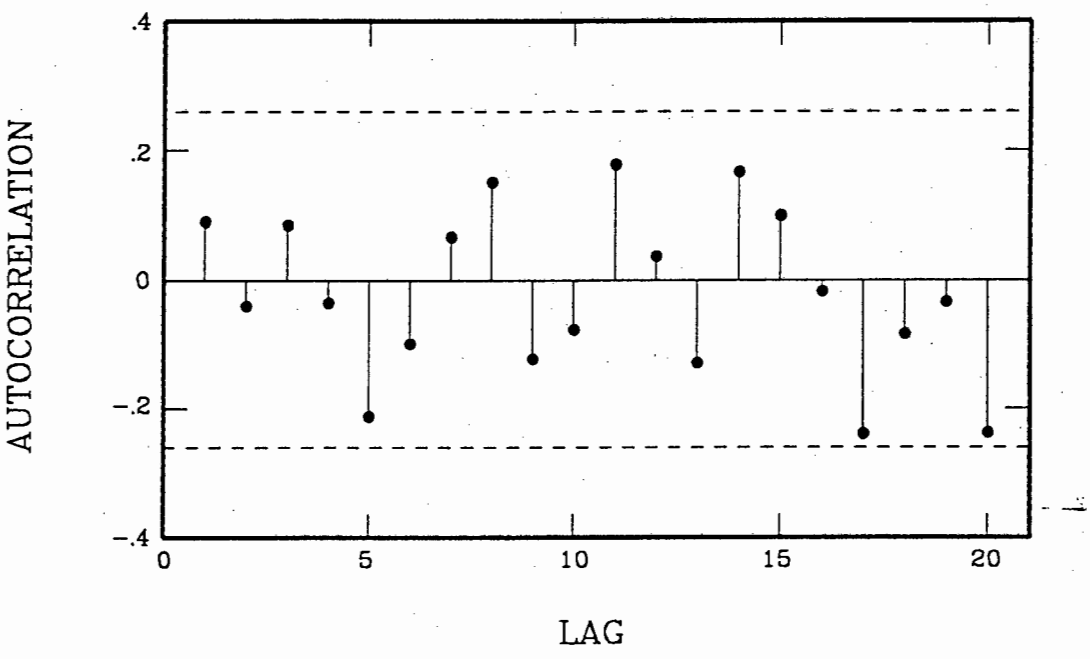


Table 2. Transfer functions fitted to the data of CEA91. The independent variable is the differenced λ1350 continuum observation series, i.e. the λ1350 flux increments. The models were fitted to the differenced dependent variable observations.

Data Set	Parameters (s.e.)	Residual Variance
λ1813 G	No physically realizable model found	
λ1840 S	No physically realizable model found	
λ2670 S	No physically realizable model found	
λ2670 G	No physically realizable model found	
FES counts	$U_0 = 1.29\ (0.22)$ $S_1 = 0.55\ (0.09)$ $\beta_1 = -0.77\ (0.08)$	3.45
Lyα S	$U_0 = 32.69\ (5.60)$ $S_1 = 0.66\ (0.08)$	811.64
Lyα G	$U_0 = 42.71\ (4.19)$ $S_1 = 0.73\ (0.03)$ $\beta_1 = -0.74\ (0.09)$	3019.21
CIV S	No physically realizable model found	
CIV G	$U_1 = 27.70\ (6.66)$ $S_1 = 0.65\ (0.11)$ $\beta_1 = -0.58\ (0.11)$	1571.76
SiIV S	$U_0 = -8.68\ (2.69)$ $U_1 = 12.35\ (2.64)$ $\beta_1 = -0.84\ (0.08)$	120.29
SiIV G	$U_0 = -5.66\ (2.53)$ $U_1 = 7.66\ (2.64)$ $U_3 = 4.60\ (1.96)$ $\beta_1 = -0.82\ (0.08)$	108.65
HeII S	$U_0 = 7.72\ (2.72)$ $\alpha_4 = -0.37\ (0.12)$	146.71
HeII G	$U_0 = 11.80\ (3.63)$ $U_2 = 7.10\ (3.61)$ $\beta_1 = -0.46\ (0.12)$	270.32
CIII] S	$U_0 = -8.44\ (2.88)$ $U_3 = 9.47\ (2.97)$ $\beta_1 = -0.69\ (0.10)$	271.46
CIII] G	$U_0 = -8.19\ (2.53)$ $U_3 = 9.22\ (2.60)$ $\beta_1 = -0.63\ (0.11)$	185.41
NV G	No physically realizable model found	
MgII S	No dependence	
MgII G	No dependence	

Table 3. As for Table 2, but with the λ1337 flux increments forming the independent variable.

Data Set	Parameters (s.e.)	Residual Variance
λ1350 S	$U_0 = 0.88\ (0.04)$ $U_1 = 0.073\ (0.036)$ $\beta_1 = -1.02\ (0.05)$	0.022
λ1813 G	$U_0 = 0.54\ (0.02)$ $U_2 = 0.094\ (0.02)$ $\beta_1 = -0.83\ (0.08)$	0.010
λ1840 S	$U_0 = 0.50\ (0.03)$ $U_2 = 0.13(0.03)$ $\beta_1 = -0.99\ (0.02)$	0.025
λ2670 S	$U_0 = 0.24\ (0.01)$ $S_2 = 0.26\ (0.05)$ $\beta_1 = -0.64\ (0.11)$	0.0070
λ2670 G	$U_0 = 0.24\ (0.01)$ $S_2 = 0.28\ (0.05)$ $\beta_1 = -0.62\ (0.11)$	0.0064
FES counts	$U_0 = 1.29\ (0.23)$ $S_1 = 0.53\ (0.09)$ $\beta_1 = -0.76\ (0.08)$	3.30
Lyα S	$U_0 = 33.54\ (5.50)$ $S_1 = 0.64\ (0.08)$	763.04
Lyα G	No physically realizable model found	
CIV S	No physically realizable model found	
CIV G	$U_1 = 23.69\ (6.28)$ $S_1 = 0.69\ (0.11)$ $\beta_1 = -0.52\ (0.12)$	1668.66
SiIV S	$U_0 = -6.83\ (2.73)$ $U_1 = 7.66\ (3.26)$ $U_3 = 4.22\ (1.93)$ $\beta_1 = -0.87\ (0.07)$	118.40
SiIV G	$U_0 = -6.38\ (2.53)$ $U_1 = 7.84\ (2.97)$ $U_3 = 4.89\ (1.83)$ $\beta_1 = -0.81\ (0.08)$	103.75
HeII S	$U_0 = 7.56\ (2.74)$ $\alpha_4 = -0.37\ (0.12)$	148.36
HeII G	$U_0 = 12.65\ (3.36)$ $U_2 = 7.17\ (3.32)$ $\beta_1 = -0.48\ (0.12)$	253.57
CIII] S	$U_2 = -6.27\ (3.18)$ $U_6 = 13.28\ (3.36)$ $S_2 = -0.54\ (0.23)$ $\beta_1 = -0.66\ (0.12)$	264.30
CIII] G	$U_0 = -6.92\ (2.61)$ $U_5 = 7.37\ (2.68)$ $S_2 = -0.64\ (0.15)$ $\beta_1 = -0.58\ (0.12)$	182.15
NV G	$U_0 = 11.26\ (3.36)$ $S_1 = 0.55\ (0.14)$ $\beta_1 = -0.91\ (0.06)$	1079.82
MgII S	No dependence	
MgII G	No dependence	

(iii) In a number of instances no entirely satisfactory transfer function could be found; such cases are denoted "No physically realizable model found" in the tables. This often involved the fact that a large cross-correlation at lag -3 was present in the residual ccf. The implication of this is that there are features in the output series *preceding* similar features in the input series, in addition to output feature following input features. This could be a chance phenomenom, or the simple underlying physical model of an input driving an output may be inappropriate. Note in particular that no satisfactory transfer functions for the longer wavelength continuum in term of λ1350 fluctuations, could be found. A similar problem can be seen in, for example, the MgII-λ1350 ccf of CEA91: the amplitude of the ccf is rather larger for negative lags than for positive, implying that line emission changes preferably precede continuum flux changes. This issue requires more extensive research.

Two example ccfs demonstrating the problem are presented in Figures 13 and 14.

(iv) The only case for which emission line variability appears to be independent of the continuum variations, is that of the Mg II line. An example ccf is shown in Figure 15. The only promising feature is the marginal cross-correlation at lag 3; however, the U_3-term in a transfer function fit was only 1.35 standard errors large.

(v) Almost all the models contain a U_0-term, i.e. the output series shows an instantaneous response to the input series. It is interesting that $U_0 < 0$ in a number of instances, i.e. the output decreases immediately if the input increases (SiIV, CIII]). The CIV series is the only one which shows unambiguous evidence for a dead time: U_0 is statistically zero. The response time τ is one unit (i.e. in the range 2 to 6 days).

(vi) Generally models based on respectively λ1337 and λ1350 variations as independent variable, are quite similar. The exception is the CIII] series, which shows evidence for responses at more than a six time units lag to λ1337 flux changes only. The CIII] response to λ1350 variations nonetheless also appears to be relatively stonger at long lags than that of other emission lines.

(vii) In all cases in which meaningful transfer functions could be fitted, the residual variance is substantially smaller than for the corresponding univariate model. This indicates that in most cases the series of short wavelength continuum flux variations contains information about the longer wavelength continuum and emission line fluxes. Residual variances found using λ1337 fluctuations as input are generally smaller than those obtained with λ1350 variations as independent variable.

Figure 13: The ccf of the differenced CIV S, and the differenced λ1337 continuum observations. Note the presence of large cross-correlations at both positive and negative lags. The residual ccf of all otherwise reasonable models also showed a large cross-correlation at lag -3.

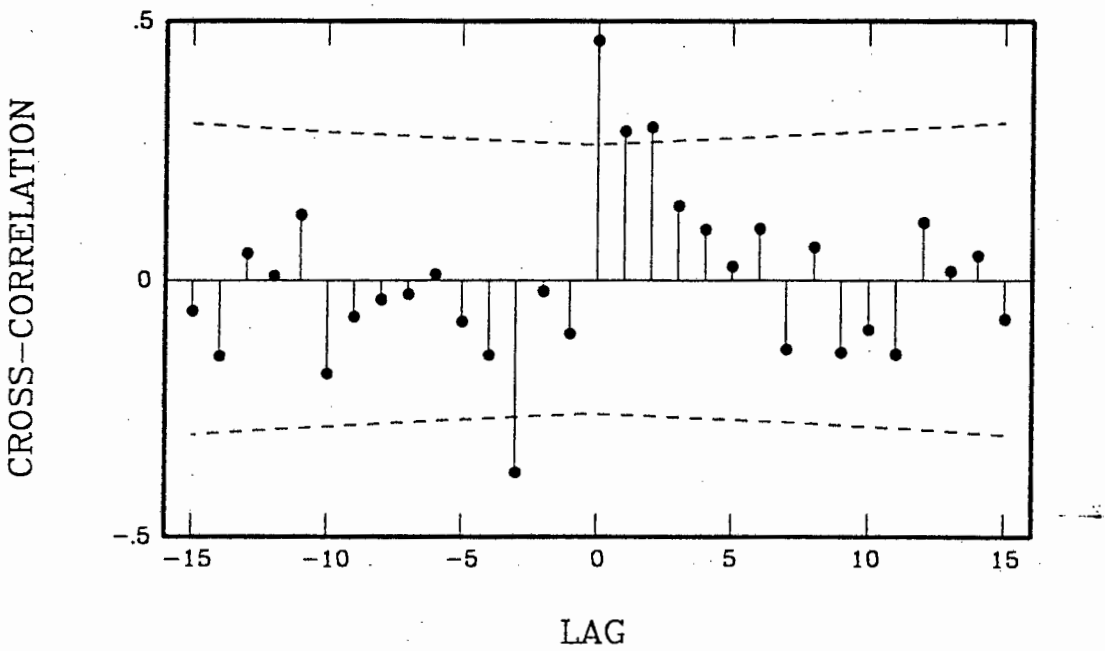


Figure 14: (a) The ccf of the differenced Lyα S, and the differenced λ1337 continuum observations. (b) The ccf of the same input series, and the residuals from a three parameter (U_1, S_1, β_1) transfer function. The residuals from other models gave similar results.

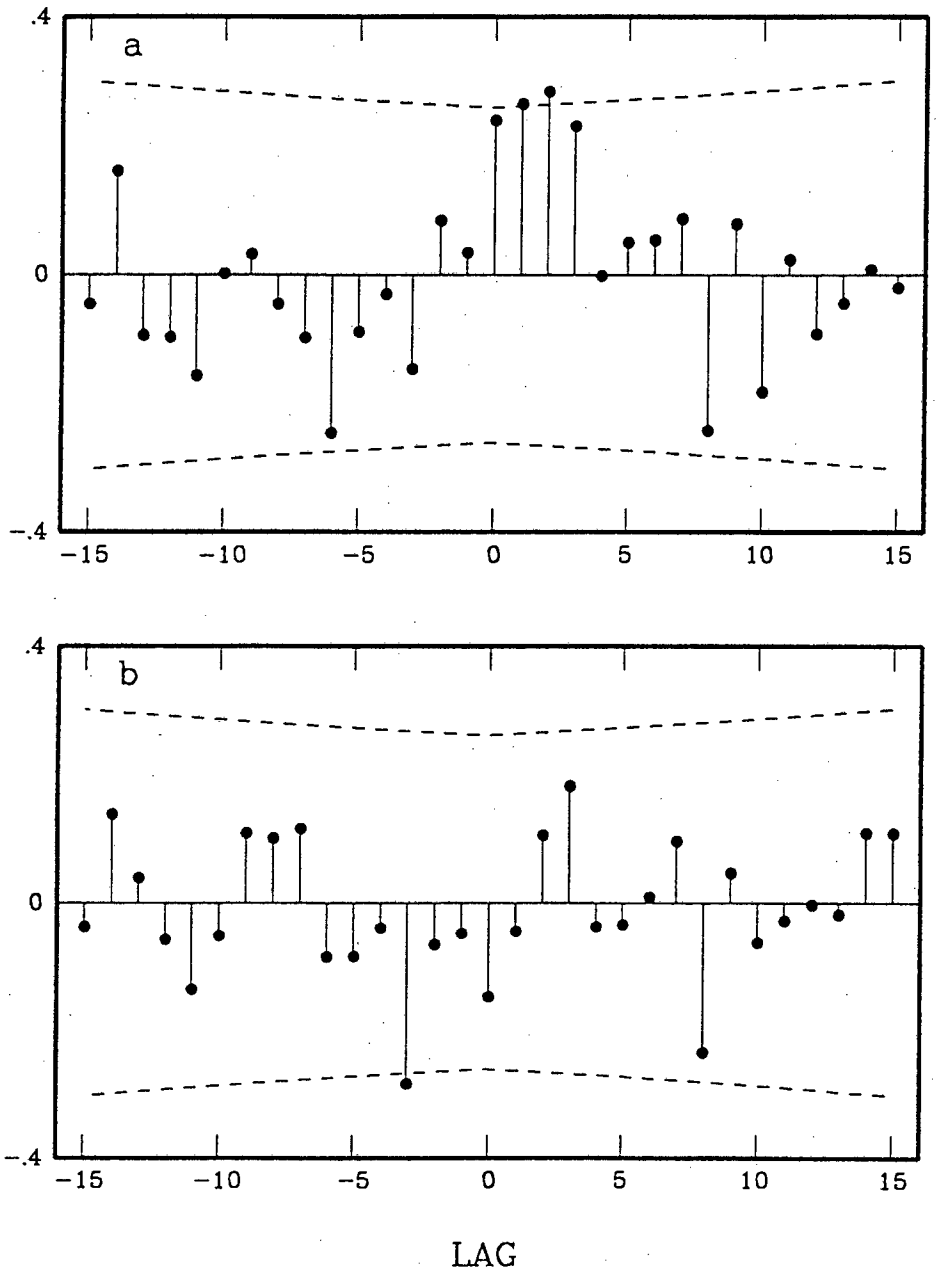
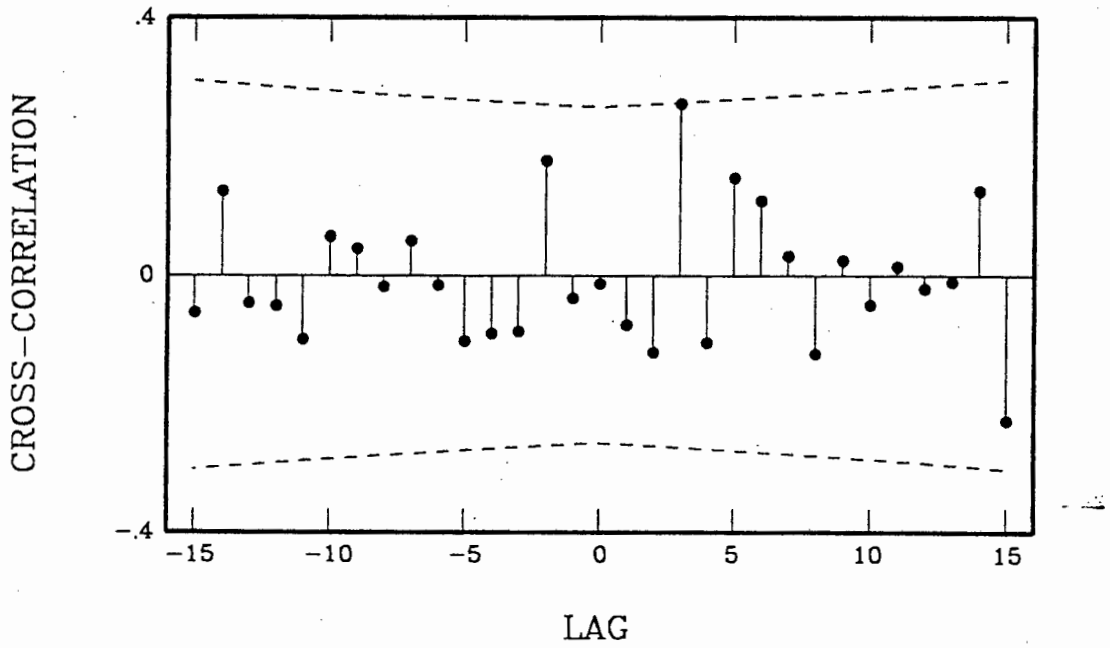


Figure 15: The ccf of the differenced MgII G, and the differenced λ1337 continuum observations.



PART 2. IRREGULARLY SPACED DATA

2.1 INTRODUCTION

There are several contexts in astronomy in which the phase lag between two time series of observations provide information about the geometry of the system under investigation. One of these, mentioned in the general introduction, is multi-wavelength observation of e.g. AGNs: large changes in blue light are expected to occur in the nuclei of the AGNs, with consequent reprocessing into red light at some large distance from the galactic centres (see e.g. Glass 1992). The time delay between flux changes in the blue and red is then a consequence of the travel time of the short wavelength radiation from nucleus to reprocessing site, and measurement of the lag allows the distance to be estimated. A second situation where the lag is of importance is in the brightness variations of the lensed components of e.g. quasars (see, for example, Press, Rybicki & Hewitt 1992). Quasars, or quasi-stellar objects, are thought to be compact, active, galactic nuclei. If there is a large mass, such as a galaxy, in the line of sight to a distant galaxy or quasar, two or more images of the distant object are sometimes seen. The effect is due to the curvature of space around the massive intervening galaxy, as predicted by the general theory of relativity. The phenomenom is referred to as gravitational lensing. Measurement of a time difference in the variations of the different quasar images allows conclusions to be drawn about the distance of the lensed object, something which is of cosmological importance.

In practice there are a number of difficulties in measuring lags between most bivariate astronomical time series. Some of these are discussed in section 2.2 below. Many of these problems are associated with the typically very irregular time spacing of the observations. The time delay estimation method described in section 2.3 is based on modelling the changing brightness of the object under study as a random walk phenomenom, and does not favour regular time spacing of observations. An example of each of the two applications discussed above is analysed below. The appropriateness of the assumed statistical model for the individual series is examined for these in the section 2.4. As is demonstrated in section 2.2 below, apparently highly significant lags may be identified between two series which are really unrelated. This obviously necessitates testing whether the two series are truly related. This point is addressed in section 2.5.

The methods given in section 2.3 are computationally intensive for moderately sized computers. A discussion of efficient algorithms for the necessary matrix calculations is presented in section 2.6. Next, lag estimation is carried out for each of the two examples. Concluding remarks to Part 2 are given in section 2.8.

2.2 TIME DELAY ESTIMATION METHODS USED IN ASTRONOMY

Broadly speaking, two different approaches to the determination of lags may be identified in the astronomy literature. The most direct of these is calculation of the cross correlation function (ccf) between the two sets of observations. For N equally spaced observations of two series

$\{x(t)\}$ and $\{y(t)\}$, the ccf at a trial lag τ is (e.g. Box & Jenkins 1976)

$$r(\tau) = \frac{1}{NS_xS_y} \sum_{t=1}^{N-\tau} [x(t) - \overline{x}][y(t+\tau) - \overline{y}],$$

where the standard notation for the mean and standard deviations of the two series has been followed. For irregularly spaced data the formula cannot be applied as it stands, and a modification which has been suggested is

$$r(\tau) = \frac{1}{NS_xS_y} \sum_{j=1}^{N(\tau)} [x(t_j) - \overline{x}][y(t_j+\tau) - \overline{y}] \tag{8}$$

where the summation is over t_j such that either or both series were observed at that time point, and $N(\tau)$ is the number of terms in the sum at lag τ. If only one of the series was observed at t_j, the value of the other is estimated by linear interpolation (Gaskell & Peterson 1987. The actual estimator suggested by these authors is slightly different, but essentially equivalent to the concise form above).

Edelson & Krolik (1988) have pointed out the uncertainty inherent in the data interpolation procedure required in (8), and suggested the estimator

$$r(\tau) = \frac{1}{n(\tau)S_xS_y} \sum_{i,j \in B} [x(t_i) - \overline{x}][y(t_j) - \overline{y}] \tag{9}$$

where the summation is over that set $B = B(\tau)$ of (i, j) such that $\mid t_j - t_i - \tau \mid < \Delta\tau/2$, for a suitable bin width $\Delta\tau$; $n(\tau)$ is the number of elements in B. The authors give a formula for the standard error of the $r(\tau)$, based on the scatter of the factors summed in (9). It is straightforward to add to this a simple expression which can be used as an approximate check on the significance of individual values of $r(\tau)$, provided the series are free of autocorrelation. Starting from the hypothesis that the two series are unrelated and hence uncorrelated, one has

$$\mathrm{var}[C(\tau)] = \mathrm{E}\,[C(\tau)^2] = \frac{1}{n(\tau)^2}\mathrm{E}\left\{\sum_B [x(t_i) - \overline{x}][y(t_j) - \overline{y}] \sum_B [x(t_k) - \overline{x}][y(t_m) - \overline{y}]\right\}$$

where E is the expectation (ensemble average) operator, and $C(\tau) = S_xS_yr(\tau)$ is the covariance of the two series at lag τ. Using

$$\begin{aligned} \mathrm{E}\,[x(t_i) - \overline{x}][y(t_m) - \overline{y}] &= 0 \\ \mathrm{E}\,[x(t_i) - \overline{x}][x(t_j) - \overline{x}] &\approx S_x^2\delta_{ij} \\ \mathrm{E}\,[y(t_i) - \overline{y}][y(t_j) - \overline{y}] &\approx S_y^2\delta_{ij} \end{aligned}$$

it follows that

$$\mathrm{var}[C(\tau)] \approx \frac{1}{n(\tau)} S_x^2 S_y^2$$

(The reason for the equalities above being approximate, rather than exact, is that the expectations are equal to the population variances, rather than the sample variances S_x^2 and S_y^2. It is assumed that the differences are small). Finally,

$$\mathrm{var}[r(\tau)] \approx n^{-1}(\tau)$$

If one is prepared to assume that the $r(\tau)$ are Gaussian in distribution, $\mid r(\tau) \mid$ may be compared to $2n^{-1/2}(\tau)$ to gauge whether it is significant at an approximately 5% level. Of course, if many correlations are simultaneously evaluated, the probability of finding some significant is considerably enhanced. In order to derive a portmanteau statistic Q which can be used to assess the complete ccf, it is first noted that, subject to the specifications above, the $r(\tau)$ are all independent for non-overlapping bins. By analogy with the case for regularly spaced observations (Box & Jenkins 1976)

$$Q = \sum_{\tau_1}^{\tau_2} n(\tau) r(\tau)^2$$

seems like a reasonable choice. Q will be distributed approximately chi squared with degrees of freedom equal to the number of cross-correlations evaluated.

The second basic method for lag determination is to identify that lag at which the two series fit together most "smoothly". Van Langevelde, Van der Heiden & Van Schooneveld (1990) have proposed minimising (with respect to the lag) the residual sum of squares when matching the two series. The function to be minimised is

$$R(\tau) = \frac{1}{n(\tau)} \sum_t \{[x(t) - \overline{x}] - [y(t+\tau) - \overline{y}]\}^2 \qquad (10)$$

where the summation is over $n(\tau)$ regularly spaced values obtained by interpolation. Various refinements were also suggested. Press, Rybicki & Hewitt (1992; referred to below as PRH) appear to have been the only authors who have produced a method which takes full account of the autocorrelation structure in the data. Their χ^2 criterion to be minimised is

$$\chi^2(\tau) = \mathbf{z}^t C^{-1} \mathbf{z} \qquad (11)$$

where $\mathbf{z} = \mathbf{z}(\tau)$ is a column vector containing the $\{x(t)\}$ and $\{y(t)\}$ in the time order corresponding to the lag τ; and C is the covariance matrix of the observations, with allowance for measurement errors. The series are transformed to the same mean level by adjusting according to their individual means calculated over the time span of overlap corresponding to τ. The authors estimate C by applying the Edelson & Krolik (1988) binning method mentioned above to the individual series, and noting that the correlation structures of the two series are similar.

It should be noted that in the applications dealt with by Van Langevelde et al. (1990) and PRH, it was assumed that the standard deviations, i.e. "amplitudes", of the x and y processes are the same. In general this will not be the case and it will be necessary to scale the series by their standard deviations to make them comparable.

Some of the difficulties involved in using the methods summarised above, have already been mentioned either explicitly, or by implication, elsewhere in the thesis, but will now be examined in more detail.

Edelson & Krolik (1988), amongst others, describe a problem in using the ccf of the raw data as a tool for determining time delays, namely the effects of autocorrelation in the individual series on the ccf. The origin of the problem is not difficult to see: similar autocorrelations in the two series can result in similar behaviour of the series over short timescales. The similarities in the time plots of the series will show up as high cross-correlations between the series. Fig. 16 shows three realisations of the series $y_t = 0.8y_{t-1} + \varepsilon_t$, where the ε_t are uncorrelated random variables with a Gaussian zero mean distribution. The three series are, by construction, unrelated. However, the three ccfs in Fig. 17 all show highly significant features. (The 2σ confidence limits are calculated from $\sigma \approx (N-k)^{-1/2}$, where N is the number of data and k is the unsigned lag). In the case of regularly observed series the effects of autocorrelation can be removed by a prewhitening procedure, as was done in Part 1; for irregularly spaced data it appears to be impossible. It is clear that autocorrelation in the individual series could similarly result in the spurious identification of lags by the Van Langevelde et al. (1990) procedure. The method of PRH, on the other hand, takes explicit account of the correlation structure of the series and in fact uses it in the time delay estimation.

A second type of problem arises if the series are non-stationary. Here only the most obvious type of non-stationarity, that of the mean (i.e. trends), will be discussed. Figure 18 shows realisations of two processes, each consisting of uncorrelated noise superposed on a small positive trend. The ccf of the the noisy parts of the series can be seen in Fig. 19a, while the ccf of the shown series appears in Fig. 19b. It is evident that the cross-correlation is amplified by the presence of the trend.

Application of PRH's method to data with trends deserves a careful evaluation. Their procedure is designed to work in situations in which the underlying process mean and standard deviation are ill-defined ("low frequency divergent" processes), as is evidently the case with the observations analysed in PRH (Fig. 20). (Evidence that non-stationarity in quasar time series may be rather common has been presented by Smith et al. 1993.) Nonetheless, the authors assume the data to be covariance stationary, i.e. $\mathrm{cov}[x(t), y(t+\tau)] = C(\tau)$, where $C(\tau)$ does not depend on t, but only on the lag τ. It is not difficult to show that this requires knowledge of the process mean as a function of time. Let, quite generally, $x(t) = f(t) + \varepsilon_t$ where ε_t is a zero-mean covariance stationary random process and $f(t)$ is a systematic trend. Then

$$C(\tau) = \mathrm{cov}[x(t+\tau), x(t)] = \mathrm{E}[x(t+\tau) - \mathrm{E}\,x(t+\tau)][x(t) - \mathrm{E}\,x(t)]$$

by definition. Thus

$$C(\tau) = \mathrm{E}[f(t+\tau) + \varepsilon_{t+\tau} - f(t+\tau)][f(t) + \varepsilon_t - f(t)] = c(\tau),$$

where $c(\tau)$ is the covariance function of the ε_t, independent of t by assumption. If, however, a *constant* mean value μ is used in the above formulae, one finds

$$C(\tau) = \mathrm{E}[x(t+\tau) - \mu][x(t) - \mu] = f(t+\tau)f(t) + \mu^2 - \mu[f(t+\tau) + f(t)] + c(\tau),$$

Figure 16: Three realisations of the process $y_t = 0.8y_{t-1} + \varepsilon_t$ for $t = 1, 2, \ldots, 200$. The ε_t are zero-mean random Gaussian variates.

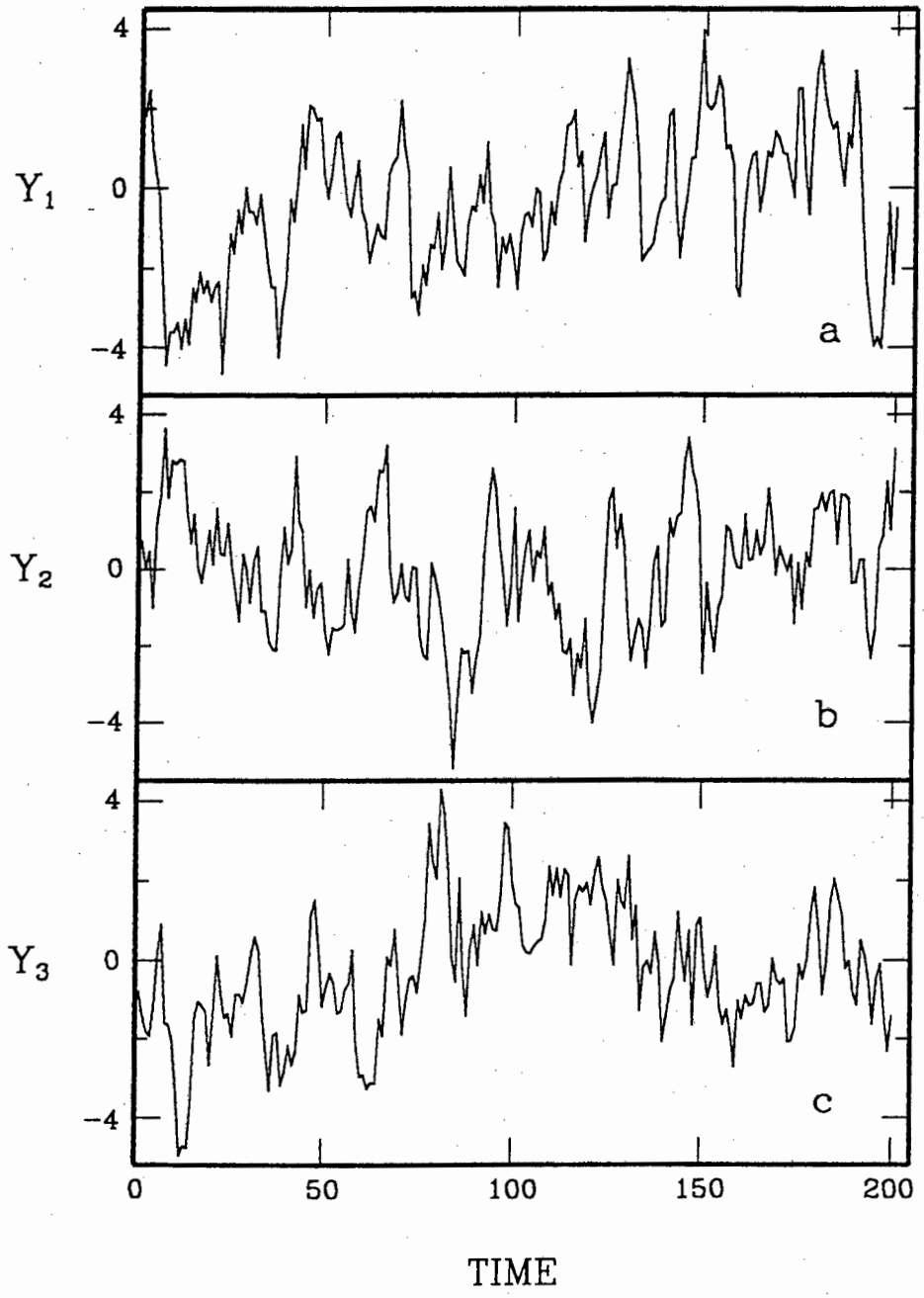


Figure 17: Cross-correlation functions for the series in Figure 16: (a) y_1 and y_2 (b) y_1 and y_3 (c) y_2 and y_3. The broken lines dileneate two sigma limits for zero cross-correlation.

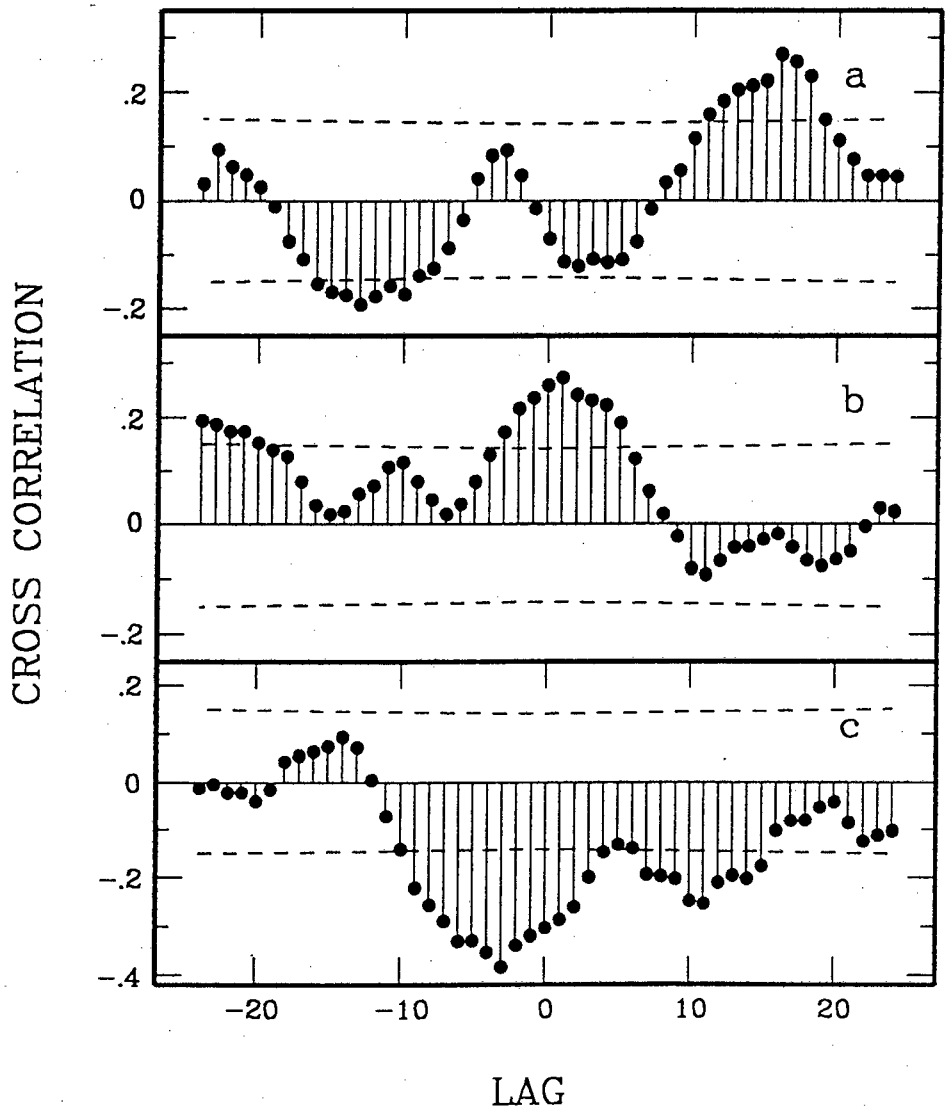


Figure 18: Two realisations of uncorrelated zero-mean Gaussian noise superposed on small trends: (a) $y_t = 0.005t + \varepsilon_t$ (b) $y_t = 0.0075t + \varepsilon_t$. In both cases $t = 1, 2, \ldots, 200$

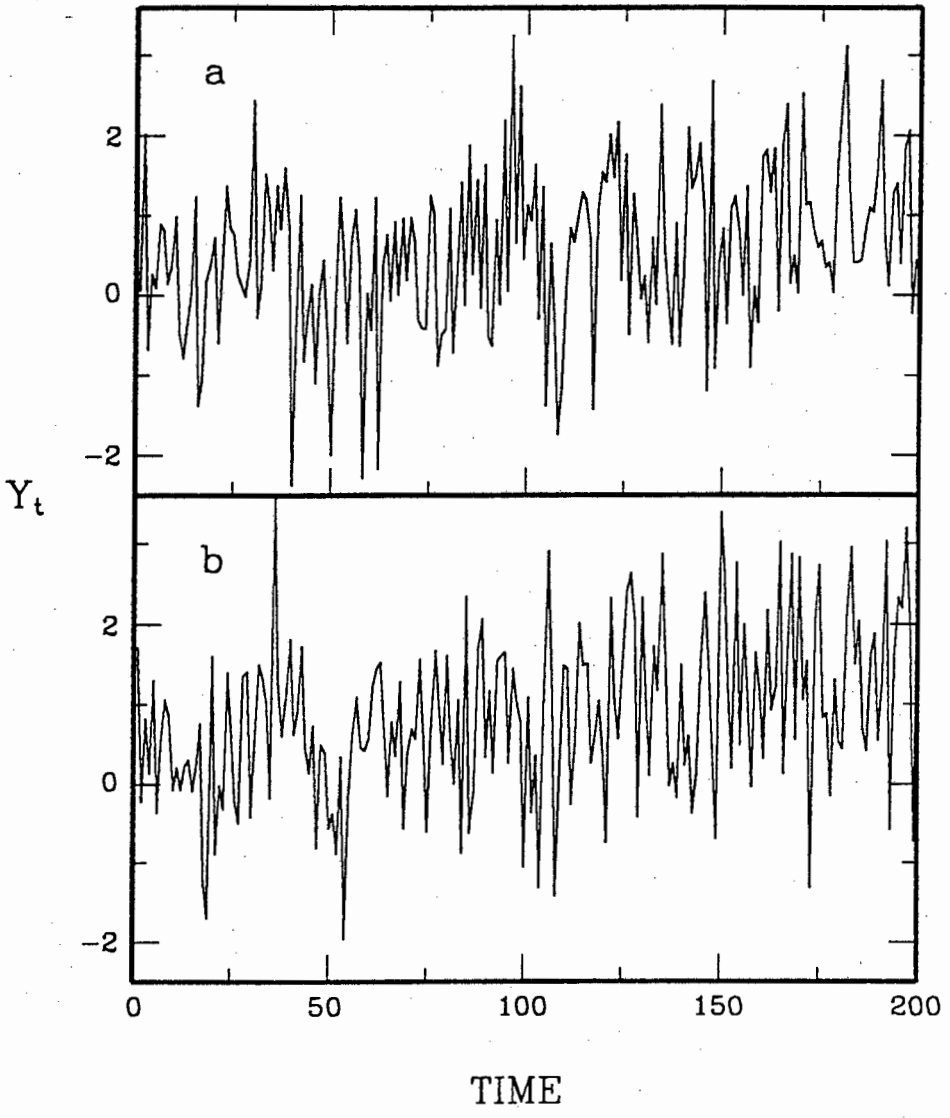


Figure 19: (a) Cross-correlation function for the noise series in Figure 18. Note that there is only one marginally significant value. (b) Cross-correlation function of the non-stationary series in Figure 18. Several significant values are evident.

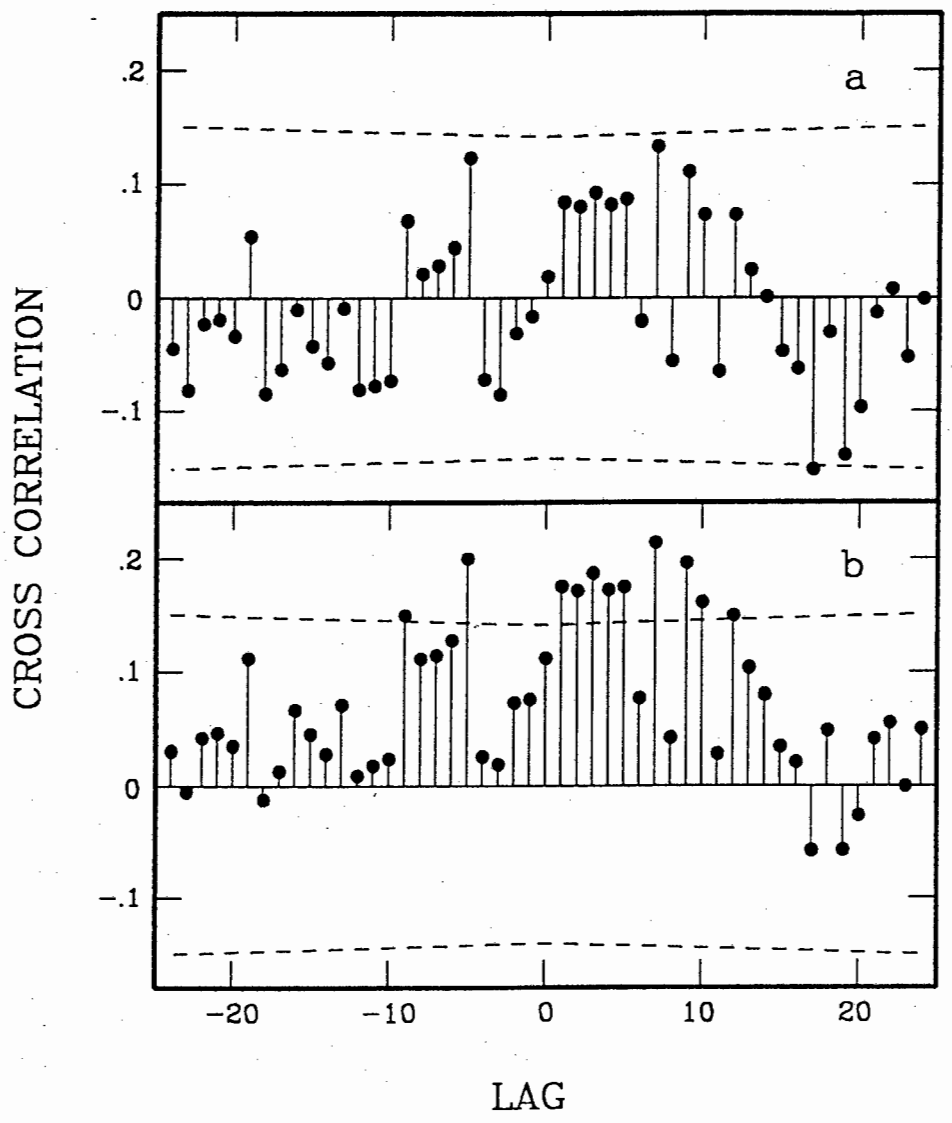
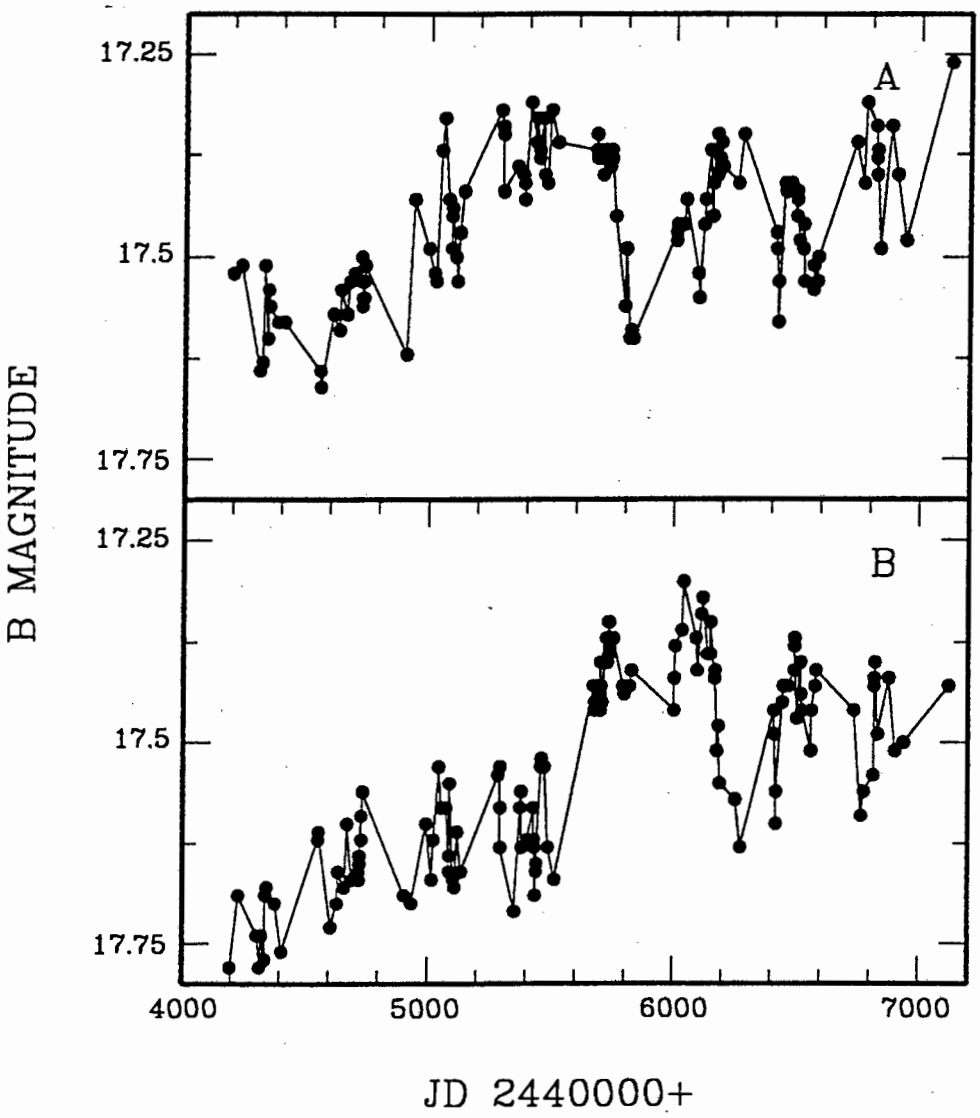


Figure 20: Observations of images A and B of the quasar 0957+561. The data are from Vanderriest et al. 1989.



which is not in general independent of t.

An interesting further example is provided by random walk processes. In terms of the increments $\Delta x_j = x(t_{j+1}) - x(t_j)$, the defining characteristics of the random walk are

$$\mathrm{E}\Delta x_j = 0 \qquad \mathrm{cov}(\Delta x_j, \Delta x_k) = \sigma^2 \parallel [t_j, t_{j+1}] \cap [t_k, t_{k+1}] \parallel \tag{12}$$

where σ is the random walk standard deviation and the factor between the set of parallel bars is the overlap between the two time intervals over which the increments Δx_j and Δx_k are observed. It follows that $\mathrm{cov}(\Delta x_j, \Delta x_j) = \sigma^2(t_{j+1} - t_j) \equiv \sigma^2 \Delta t_j$, while increments over disjoint time intervals are uncorrelated. Thus for $\tau > 0$,

$$C(\tau) = \mathrm{cov}[x(t_j + \tau), x(t_j)] = \mathrm{cov}[x(t_j), x(t_j)]$$

Now $x(t_j) = x(0) + \sum_{i=1}^{j} \Delta x_i$ (which is, incidentally, well known to be a non-stationary process), and hence

$$C(\tau) = \sum_{i=1}^{j} \mathrm{cov}(\Delta x_i, \Delta x_i) = \sigma^2 \sum_{i=1}^{j} \Delta t_i = \sigma^2 (t_{j+1} - t_1)$$

which is independent of τ but depends on t_j.

It is obvious that the series in Fig. 20 do not have well-defined mean values, yet these are bound to play a crucial role in correctly aligning the two series prior to determination of the time delay between them. In Part 1 above the problem of possibly non-stationary observations of NGC 5548 was circumvented by differencing the data and hence reducing both series to zero-mean stationary form (see also Box & Jenkins 1976). However, those data were rather unique in being regularly spaced, so that the differencing procedure is well-defined. In the next section a model for which differencing of irregularly spaced time series is meaningful, will be discussed.

2.3 LAG DETERMINATION WHEN THE SERIES ARE RANDOM WALKS

The basic model proposed is that the two sets of observations are random walks, i.e. the increments Δx_i $(i = 1, 2, \ldots, N-1)$ and Δy_j $(j = 1, 2, \ldots, M-1)$ satisfy the definition (12). A necessary extension is best introduced by consideration of the practicalities of obtaining photometric brightness measurements of visually extended objects such as galaxies: substantial measurement errors can be expected. In order to allow analysis of potentially noisy data, the random walk model needs to be generalised to include measurement error:

$$x(t_j) = x'(t_j) + \varepsilon_j$$

where x' is the unknown true value of the series at t_j, with increments satisfying (12), and ε_j is an uncorrelated zero-mean noise process. It follows that $\mathrm{E}\Delta x_j = 0$ and

$$\mathrm{cov}(\Delta x_j, \Delta x_k) = \begin{cases} \sigma_x^2 \Delta t_j + 2\sigma_\varepsilon^2 & k = j \\ -\sigma_\varepsilon^2 & k = j-1, j+1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

For later reference the equivalent equation for the y-process is

$$\mathrm{cov}(\Delta y_j, \Delta y_k) = [\sigma_y^2 \Delta T_j + 2\sigma_\eta^2]\delta_{kj} - \sigma_\eta^2[\delta_{k,j-1} + \delta_{k,j+1}] \tag{14}$$

where use has been made of the Kronecker delta function to write the relation in a concise form. The measurement error in $y(t)$ is denoted by η_t. Series y is observed at times T_j, which are in general different from the times t_k of observation of series x.

The above considerations apply in particular to data such as those shown in Fig. 1; all observations were obtained at the same site, with very similar equipment, so that the nature of the measurement errors is uniform over the timespan of the observations. For non-homogeneous data such as those in Figs. 20, individual measurement error estimates may be given. The appropriate forms of (13) and (14) are then

$$\begin{aligned} \mathrm{cov}(\Delta x_j, \Delta x_k) &= [\sigma_x^2 \Delta t_j + \sigma_{\epsilon,j+1}^2 + \sigma_{\epsilon,j}^2]\delta_{kj} - \sigma_{\epsilon,k}^2 \delta_{k,j-1} - \sigma_{\epsilon,k}^2 \delta_{k,j+1} \\ \mathrm{cov}(\Delta y_j, \Delta y_k) &= [\sigma_y^2 \Delta T_j + \sigma_{\eta,j+1}^2 + \sigma_{\eta,j}^2]\delta_{kj} - \sigma_{\eta,k}^2 \delta_{k,j-1} - \sigma_{\eta,k}^2 \delta_{k,j+1} \end{aligned} \tag{15}$$

For homogeneous data the estimator

$$\hat{\sigma}_\epsilon^2 = -\frac{1}{N-2} \sum_{j=1}^{N-2} \Delta x_j \Delta x_{j+1} \tag{16}$$

for the x error variance is suggested by Equation (13). Also,

$$\mathrm{E} \sum_{j=1}^{N-1} (\Delta x_j)^2 = 2(N-1)\sigma_\epsilon^2 + \sigma_x^2 \sum_{j=1}^{N-1} \Delta t_j$$

and thus

$$\hat{\sigma}_x^2 = \frac{1}{t_N - t_1} \left[\sum_{j=1}^{N-1} (\Delta x_j)^2 - 2(N-1)\hat{\sigma}_\epsilon^2 \right] \tag{17}$$

seems a reasonable estimator for the random walk variance. If the increments Δx_j and errors have Gaussian distributions, maximum likelihood estimators $\tilde{\sigma}_\epsilon^2$ and $\tilde{\sigma}_x^2$ may be obtained by maximising the log likelihood function

$$L = -\frac{1}{2}[(N-1)\ln 2\pi + \ln \mid \Sigma \mid + \Delta \mathbf{x}^t \Sigma^{-1} \Delta \mathbf{x}] \tag{18}$$

where $\Sigma = \Sigma(\sigma_\epsilon^2, \sigma_x^2)$ is the covariance matrix of the Δx_j, constructed according to (13), and $\Delta \mathbf{x}$ is a column vector with elements Δx_j.

Maximisation of L with respect to the two variances may be simplified by dividing each entry of the covariance matrix Σ by σ_ϵ^2; this leaves the tri-diagonal matrix Σ_*, with diagonals $2 + \sigma_x^2 \Delta t_j / \sigma_\epsilon^2$, and off-diagonals equal to -1. The explicit solution

$$\sigma_x^2 = \frac{1}{N-1} \Delta \mathbf{x}^t \Sigma_*^{-1} \Delta \mathbf{x} \tag{19}$$

for the random walk variance in terms of Σ_*, which contains the single unknown $q_x \equiv \sigma_\epsilon^2 / \sigma_x^2$, is obtained. The solution strategy is thus

(i) Select a trial value for q_x.

(ii) Calculate σ_x^2 from (19).

(iii) Substitute into (18) to find

$$L = L(q_x) = -\frac{1}{2}[(N-1)(1+\ln 2\pi) + (N-1)\ln\sigma_x^2 + \ln \mid \Sigma_* \mid]$$

(iv) Repeat steps (i) to (iii). New trial values for q_x are chosen by reference to previous results, in order that $L = L(q_x)$ eventually be maximised.

Note that negative estimates for variances or q_x should be replaced by the smallest realisable value of these quantities, namely zero. For negligible measurement error $\Sigma = \mathrm{diag}(\sigma_x^2 \Delta t_j)$, and it is easily shown that

$$\tilde{\sigma}_x^2 = \frac{1}{N-1}\sum_{j=1}^{N-1}(\Delta x_j)^2/\Delta t_j \tag{20}$$

which should be compared with (17); in (20) the process increments are weighted inversely with the length of the time interval over which they are measured. In view of (13), this certainly seems reasonable.

For the case of known measurement error variance the log likelihood function in (18) is a function of the single unknown σ_x^2.

Having dealt with the internal structures of the x and y series, we now turn to the relationship between them. The postulated model is

$$y(t) = Ax(t+\tau) + B \tag{21}$$

where A, B and τ are constant. Clearly $\tilde{\sigma}_y/\tilde{\sigma}_x$ is an estimator for A, while a rough estimate of B can easily be made after the lag τ has been determined. In order to find τ, note that

$$\mathrm{cov}(\Delta x_j, \Delta y_k) = A \parallel [t_j, t_{j+1}] \cap [T_k - \tau, T_{k+1} - \tau] \parallel \tag{22}$$

by the definition (12) and (21). The observations shown in Fig. 20 provide an example of data which are inhomogeneous in error magnitudes (Vanderriest et al. 1989), and in which the x and y measurements which are obtained contemporaneously may have correlated errors (see PRH). For such cases

$$\begin{aligned}\mathrm{cov}(\Delta x_j, \Delta y_k) &= A \parallel [t_j, t_{j+1}] \cap [T_k - \tau, T_{k+1} - \tau] \parallel + \rho\,[\sigma_{\epsilon,j+1}\sigma_{\eta,k+1}\delta(t_{j+1}, T_{k+1}) \\ &+ \sigma_{\epsilon,j}\sigma_{\eta,k}\delta(t_j, T_k) - \sigma_{\epsilon,j+1}\sigma_{\eta,k}\delta(t_{j+1}, T_k) - \sigma_{\epsilon,j}\sigma_{\eta,k+1}\delta(t_j, T_{k+1})]\end{aligned} \tag{23}$$

where δ is again the Kronecker function. It has been assumed that the correlation ρ between the measurement errors is constant, but (23) is easily generalised if contrary information is available.

For any assumed lag u, Equation (22), together with (13) and (14), (or (23) together with (15)) allow the covariance matrix $C = C(u)$ for the jointly time-ordered set of increments

$\{\Delta y_j, \Delta x_k\}$ to be constructed. The true lag τ may be estimated by that value of u which satisfies a suitable optimality criterion. A non-parametric possibility is the PRH criterion (11), where the vector z is made up of the $N-1$ increments Δx, and the $M-1$ increments Δy; C being $C(u)$ as described above. Note that whereas the PRH matrix had no zero elements, the matrix $C(u)$ is sparse. This is important, as will be made clear in section 2.6. Alternatively, if the process increments and measurement errors are Gaussian, the log likelihood is

$$L = -\frac{1}{2}\left[(N+M-2)\ln 2\pi + \ln \mid C \mid + \mathbf{z}^t C^{-1}\mathbf{z}\right] \tag{24}$$

and the lag is estimated by the value of u which maximises L. It is worth remarking that in essence the two criteria given above differ only in the presence of the factor $\ln \mid C \mid$ in (24).

This section is concluded by a consideration of the specification of confidence limits for the derived parameter values. Approximate confidence intervals for the random walk variance σ_x^2 and the ratio $q = \sigma_\epsilon/\sigma_x^2$ can be obtained by making use of the fact that asymptotically, the maximum likelihood estimators have a joint multivariate Gaussian distribution. It is useful to deal with the theory in some generality, as it has wider applications, e.g. in tests for systematic period changes in infrequently observed pulsating or eclipsing double stars (Lombard & Koen, in preparation). In general, one is interested in a vector θ of parameters, components being $\theta_1 = \sigma_x^2$ and $\theta_2 = q$ in the present instance. According to the asymptotic theory quoted above, the maximum likelihood estimate $\tilde{\theta}$ is multivariate Gaussian, with an expected value equal to the vector of true parameter values, and covariance matrix F^{-1}, where F is the Fisher information matrix (see e.g. Cox & Hinkley 1974. Of course, this is a large-sample result, and the level of approximation is unknown for finite samples). The elements of F are given by

$$F_{ij} = -\mathrm{E}\left[\frac{\partial^2 L}{\partial\theta_i \partial\theta_j}\right],$$

where E is the expectation (ensemble average) operator. Further calculations are facilitated by rewriting (11) in the form

$$L = -\frac{1}{2}\left[(N-1)\ln 2\pi + \ln \mid \Sigma \mid + \mathrm{trace}\,(\Sigma^{-1}G)\right], \qquad G \equiv \Delta\mathbf{x}\,\Delta\mathbf{x}^t$$

where the earlier form of the log likelihood function is used in the interest of greater generality. Using the rules of matrix differentiation (e.g. Graybill 1969, Bargmann 1984), it can be shown that

$$\begin{aligned}
\frac{\partial L}{\partial\theta_i} &= \frac{1}{2}\left\{\mathrm{trace}\left[\Sigma^{-1}(G-\Sigma)\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\right] - \mathrm{trace}\left[\Sigma^{-1}\frac{\partial G}{\partial\theta_i}\right]\right\} \\
\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} &= \frac{1}{2}\mathrm{trace}\left[\frac{\partial\Sigma^{-1}}{\partial\theta_i}(G-\Sigma)\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j} + \Sigma^{-1}(G-\Sigma)\frac{\partial}{\partial\theta_i}\left(\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right)\right] + \frac{1}{2}\left[\Sigma^{-1}\frac{\partial G}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right] \\
&\quad -\frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right] - \frac{1}{2}\mathrm{trace}\left[\frac{\partial\Sigma^{-1}}{\partial\theta_i}\frac{\partial G}{\partial\theta_j} + \Sigma^{-1}\frac{\partial^2 G}{\partial\theta_i\partial\theta_j}\right]
\end{aligned}$$

Now since $\mathrm{E}[G] = \Sigma$,

$$\begin{aligned} \mathrm{E}\frac{\partial L}{\partial \theta_i} &= -\frac{1}{2}\mathrm{E}\,\mathrm{trace}\left[\Sigma^{-1}\frac{\partial G}{\partial \theta_i}\right] \\ F_{ij} &= -\mathrm{E}\,\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} \\ &= \frac{1}{2}\mathrm{E}\left[-\Sigma^{-1}\frac{\partial G}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j} + \Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j} + \frac{\partial\Sigma^{-1}}{\partial\theta_i}\frac{\partial G}{\partial\theta_j} + \Sigma^{-1}\frac{\partial^2 G}{\partial\theta_i\partial\theta_j}\right] \end{aligned} \tag{25}$$

The general result (25) can now be specialised to the problem at hand. One obtains

$$F_{ij} = \frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_i}\Sigma^{-1}\frac{\partial\Sigma}{\partial\theta_j}\right]$$

In terms of the matrix Σ_*,

$$\begin{aligned} F_{11} &= \frac{1}{2}\sigma_x^{-4}\mathrm{trace}[I_{N-1}] = \frac{N-1}{2\sigma_x^4} \\ F_{12} = F_{21} &= \frac{1}{2\sigma_x^2}\mathrm{trace}\left[\Sigma_*^{-1}\frac{\partial\Sigma_*}{\partial q}\right] \\ F_{22} &= \frac{1}{2}\mathrm{trace}\left[\Sigma_*^{-1}\frac{\partial\Sigma_*}{\partial q}\Sigma_*^{-1}\frac{\partial\Sigma_*}{\partial q}\right] \end{aligned}$$

Introducing the more concise notation $S \equiv \Sigma_*^{-1}$,

$$\begin{aligned} F_{12} = F_{21} &= \sum_{j=1}^{N-1}\left[2S(j,j) - S(j,j-1) - S(j,j+1)\right] \\ F_{22} &= \sum_{k,j=1}^{N-1}\left[2S(k,j) - S(k,j-1) - S(k,j+1)\right]\left[2S(j,k) - S(j,k-1) - S(j,k+1)\right] \end{aligned} \tag{26}$$

where, by definition, $S(k,0) = S(k,N) = 0$ for all k. Similar results apply for the y-series.

For the case of known measurement error variances, $\partial\Sigma/\partial\sigma_x^2 = \mathrm{diag}(\Delta t_j)$ (cf. 8), hence

$$F_{11} = \frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}\mathrm{diag}(\Delta t_j)\right]^2 = \frac{1}{2}\sum_{k,j=1}^{N-1}\left[\Sigma^{-1}(k,j)\right]^2\Delta t_k\Delta t_j$$

Before discussing confidence intervals for the estimated lag τ, it is worth considering that at least two areas of uncertainty are not represented by the usual confidence specifications. First, criteria such as $\chi^2(\tau)$ in (11) or $L(\tau)$ in (18) may have several local extrema, from which the global value is selected. It is, however, conceivable that in some instances the true extremum will, due to random fluctuations, not be the global one. This effect has been pointed out in a different context by Hamon & Hannan (1974). Second, it has not yet been established that the

two series are truly related; the techniques described merely determine the "best" lag, *given* the relation $y(t) = Ax(t+\tau) + B$ applies. This point will be pursued in detail in section 2.4.

A confidence interval for the maximum likelihood estimate $\tilde{\tau}$ of the lag τ does not follow so readily as that for $\tilde{\sigma_x^2}$ (for example), as the analytic differentiation of the likelihood function with respect to τ is not possible. Instead, use may be made of the large-sample likelihood ratio result that

$$2[\ln L(\tilde{\tau}) - \ln L(\tau)] \sim \chi_1^2, \tag{27}$$

i.e. the quantity on the left is distributed as a chi squared variate with one degree of freedom (e.g. Mood, Graybill & Boes 1974). From (27), a 90% confidence interval consists of all τ such that the statistic is smaller than 2.71; the limits for 95% and 99% intervals are 3.84 and 6.64.

It is noted in passing that the likelihood ratio may be used to specify a multidimensional confidence region for a number of parameters (e.g. random walk variances, q_x, q_y and $\tilde{\tau}$) simultaneously as well: the appropriate degrees of freedom of the χ^2 distribution is the number of parameters under consideration.

2.4 VALIDATION OF THE RANDOM WALK MODEL

The process increments for the data of Fig. 20 are plotted in Fig. 21. The mean values are -0.002 and -0.003, and standard deviations are $s = 0.063, 0.064$ respectively. The graphs for the other example data set (Fig. 1) are similar; the means are 0.18 ($s = 2.88$) and 0.11 ($s = 7.55$). The assumption that the series of increments have zero mean values is thus entirely reasonable.

The assumption that the data are Normal may be verified without much trouble for error-free observations; the standardised increments $\xi_j = \Delta x_j / \sqrt{\delta t_j}$ will be Gaussian with mean zero and variance σ_x^2 (see (6)). However, for noisy data the ξ_j are not necessarily collectively Gaussian, even if the Δx_j and ε_j are Normally distributed. This is so because the standardisation is no longer appropriate; $\mathrm{var}(\xi_j) = \sigma_x^2 + 2\sigma_\varepsilon^2/\Delta t_j$, i.e. the ξ have different variances. Note though that those ξ_j corresponding to large Δt_j will be approximately Gaussian with variance σ_x^2. This was verified for the data of Fig. 20; the ξ_j calculated for series A deviates from Normality at the 6% level (chi squared test). However, selecting only those data with $\Delta t_j > 2$ ($N = 85$), the ξ_j were found to differ from Normality at only the 16% level, and increasing the restriction to $\Delta t_j > 4$ ($N = 48$), lead to a significance level of 44%. The ξ_j corresponding to the U and K data of Fig. 1 deviate from Normality at the 8% and 10% levels respectively, while choosing only values having $\Delta t_j > 2$ gives significances 28% and 35% ($N = 28$, 25 for U and K respectively).

Next, inspection of (13) or (15) shows that only increments with consecutive index numbers ought to be correlated. Calculation of the autocorrelation functions of the sets of ξ_j shows that the lag one value is negative in all four cases under consideration, and that no other small lag autocorrelation is significant in any of the series. Not surprisingly, the lag one correlation is largest (in absolute value) for those series for which large errors are quoted.

Finally, it follows from (13) that

$$\mathrm{E}(\Delta x_j)^2 = \sigma_x^2 \Delta t_j + \sigma_\varepsilon^2.$$

Figure 21: The magnitude increments of the data in Fig. 20. Note the well-defined mean values.

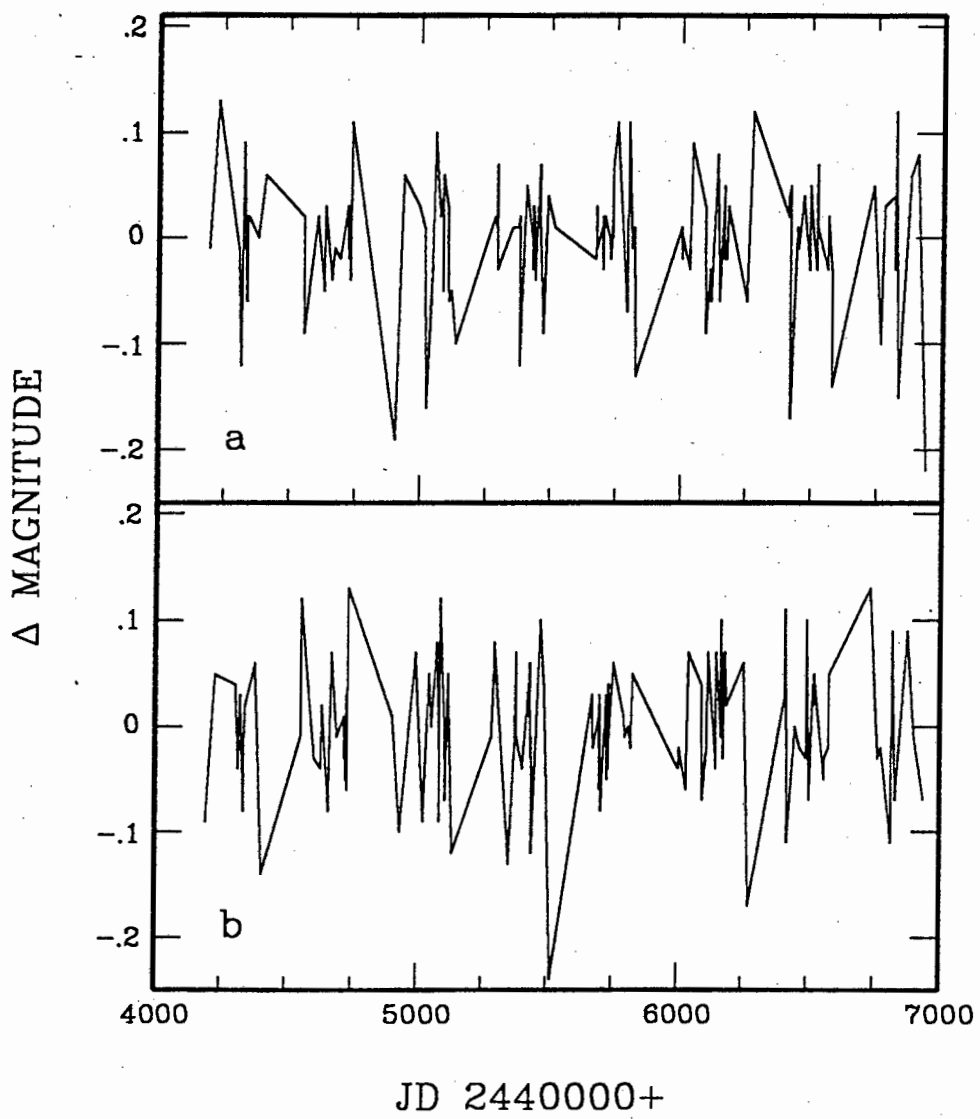
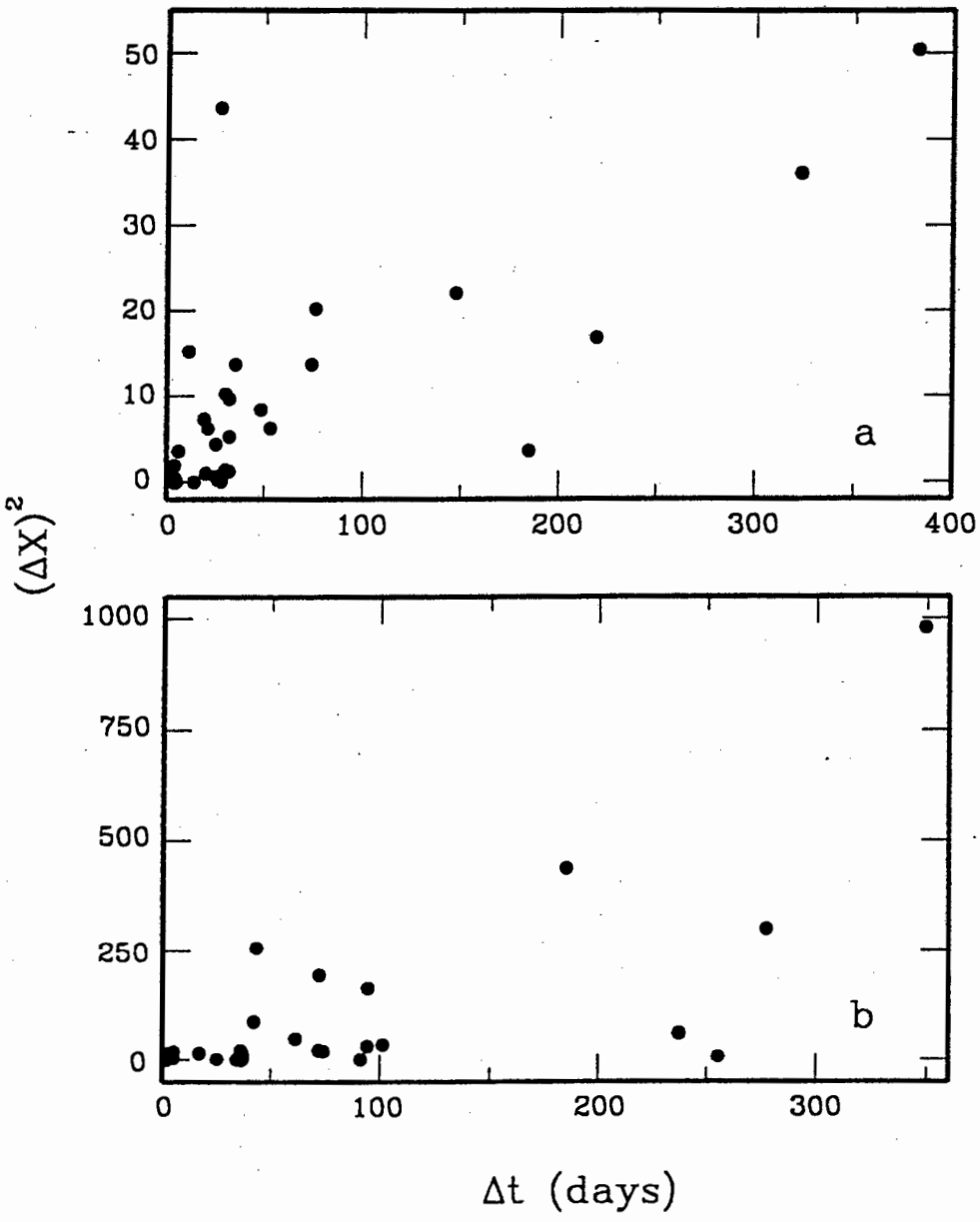


Figure 22: The squared increments $(\Delta K)^2$ (a) and $(\Delta U)^2$ (b) of the two sets of observations in Fig. 1 plotted against the time intervals between successive observations.



This suggests that a plot of Δx_j^2 against Δt_j should be linear, with positive slope and intercept. Fig. 22 shows that this is plausible for the homogeneous-error data of Fig. 1. (The reason why no attempt is made to apply regression techniques to the data of Fig. 22 is that aside from the obvious difficulties of outlying and high-leverage points, the data are not independent. This hinders the specification of confidence intervals for the slopes and intercepts; see the very interesting paper by Sterne 1934 where this problem is discussed in a different astronomical context).

The results of applying the theory developed in section 2.3 to the two example data sets, are shown in Table 4. The following comments can be made:

(i) The mean measurement error sizes quoted for the gravitational lens 0957+561 are 0.03 and 0.04 magnitudes for components A and B respectively (Vanderriest et al. 1989). The corresponding random walk standard deviations are 0.010, 0.008. It follows from (16) that observations should be taken several days apart in order to gain information on object variation, rather than on measurement accuracy.

(ii) As far as NGC 3783 is concerned, the solution for q given in column 5 of the Table shows that the measurement error and random walk standard deviations are comparable in size. The implication is that observations taken more than a day apart can, in principle, contribute new information about changes in the flux from the galaxy.

(iii) All the σ^2-values in column 2 are significant, as compared to the asymptotic standard deviations (ASDs) in column 3. The values of q in column 7 are not significant as compared to their ASDs in column 8. The latter point will be discussed in more detail below.

(iv) A number of computer simulation experiments were performed, in which artificial data sets with Gaussian errors and process increments resembling those of the observed data were generated and analysed. The results shown in column 5 of Table 1 indicate that the maximum likelihood estimator of σ^2 tends to underestimate the true random walk variance (i.e. column 2 value); in other words, the ML estimator is biased. Similarly, comparing columns 6 and 3 shows that the simulation standard deviation of the estimated variance is consistently lower than the ASD. The latter value is probably more reliable, particularly in view of the bias of the ML estimator.

(v) The naïve estimators (16) and (17) did not perform well. In almost 50% of the simulations of data set 2a, the σ_ϵ^2 was estimated as zero (equivalent to estimating $q = 0$). This is considerably worse than the 17% by the ML method. Comparison of columns 4 and 5 of the Table shows that the ML estimator of the random walk variance also fares rather better than (10).

(vi) The median, rather than the mean, of the q values found in the simulations, is quoted in column 9. The reason is that the true q is less than 1.5 ASDs from zero for both data sets 2a and 2b. One therefore expects to find (as indeed happens) a large number of zero solutions for q (i.e. the minimum possible value for this parameter). This clearly makes

Table 4. The results of analysing each component of the pairs of series separately. The series are: 0957+561, component A (1a) and B (1b); the first halves of series 1a and 1b (1c, 1d); the second halves of series 1a and 1b (1e, 1f); NGC 3783 U-band (2a) and K-band (2b). The meaning of the column headings are: σ^2 is the random walk variance, and ASD(σ^2) its asymptotic standard deviation; σ_m^2 is the random walk variance estimated by (17); σ_s^2 is the mean of the simulation random walk variances, and SD(σ_s^2) their standard deviation. q is the ratio of the measurement error variance to the random walk variance, and ASD(q) is its asymptotic standard deviation; q_s is the median of the simulation q-values, and SD(q) in the following column their standard deviation. N is the number of simulations the results in columns 5, 6, 9 and 10 are based on.

Series	σ^2	ASD(σ^2)	σ_m^2	σ_s^2	SD(σ_s^2)	q	ASD(q)	q_s	SD(q_s)	N
1a	9.64E-5	2.2E-5		6.69E-5	1.8E-5					500
1b	6.69E-5	1.8E-5		5.44E-5	1.5E-5					500
1c	6.63E-5	2.3E-5		4.78E-5	1.6E-5					200
1d	4.15E-5	1.8E-5		3.27E-5	1.4E-5					200
1e	1.19E-4	3.5E-5		9.95E-5	3.0E-5					200
1f	8.67E-5	3.1E-5		7.63E-5	2.6E-5					200
2a	0.217	0.064	0.16	0.211	0.058	0.85	1.04	0.82	1.44	500
2b	0.980	0.28	0.51	0.972	0.28	0.95	0.66	0.96	0.91	500

the mean value of q unsuitably large, and the median a better choice for representing a "typical" q. Note that q_s agrees quite well with the true value of q.

(vii) The standard deviations of the simulation q values, given in column 10, are substantially larger than the ASDs. The relatively large errors associated with estimating q are perhaps not too surprising, as it is the ratio of two unknown quantities. Much more accurate estimates of the measurement error variance ought to be attainable if it is estimated directly; the reader is reminded that computational ease, rather than accuracy, served as guideline in choosing the estimation method.

(viii) In most simulation experiments the distribution of the σ_s^2 was consistent with Normality.

2.5 TESTING RELATEDNESS OF THE TWO SERIES

The random walk model, and all the techniques described in section 2.2, aim to find that lag in Equation (21) which is "most consistent" with the data, in some sense. By implication, application of any of these methods will identify an optimum lag even for series which are completely unrelated. It is therefore crucial to verify that the x and y series are truly related.

A particularly simple, yet very powerful, technique can be used to test for relatedness if measurement error is negligible. It follows from (13) that if $\sigma_\epsilon \to 0$, the increments Δx_j are uncorrelated. The implication is that random permutations of the time order of the Δx_j are then statistically equivalent to the observed series (at least to second order in their joint moments). A non-parametric permutation test for relatedness may then be performed as follows:

(i) Estimate the lag as described in section 2.3, using either of the criteria (11) or (24).

(ii) Shuffle the Δx_j into a random order. Do the same with the Δy_k.

(iii) Determine the optimum value of the criterion in (i) for the two permutated series, which are, by construction, unrelated.

(iv) Repeat steps (ii) and (iii) a large number of times.

(v) Determine the percentile of the observed criterion value with respect to the permutation results; this gives a good indication of the "specialness" of the observed lagged relationship with respect to that obtained for unrelated $x - y$ sets.

The presence of measurement error introduces serial correlation in the sequence of process increments; hence, permutation tests cannot be used. It is necessary to resort to parametric tests, i.e. specific assumptions about the statistical distributions of the data need to be made. The general type of test procedure suggested here is that of "parametric bootstrapping" (e.g. Tsay 1992). This technique is used to test the overall adequacy of the final statistical model. The model introduced in section 2.3, for example, is fully specified by σ_x^2, σ_ϵ^2, σ_η^2, τ, A, B and the statement that the process increments and the errors are Gaussian. Statistically equivalent data sets, i.e. sets of "observations" $\{x(t_j)\}$ and $\{y(T_k)\}$ at the same time points as the actual data,

can then easily be computer-generated. The next step is choosing a functional which captures the essence of the proposed model. In theory the cross spectrum is probably the best functional for the problem in hand, but due to a number of technical difficulties may be rather tricky to use in practice. Instead, the criteria (11) and/or (18), evaluated for a number of trial lags, may be used. The chosen functional is calculated for each of the synthetic data sets, and the results used to construct suitable confidence bounds for the functional. (Good examples of this can be seen in Ripley 1988). The adherence of the actual observations to the model is then judged by checking its functional against these confidence bounds.

Finally, the reader is also referred to section 5 of PRH which describes interesting applications of similar techniques to the data of Figure 20.

2.6 DETAILS OF THE MATRIX COMPUTATIONS

The covariance matrix C which appears in Equations (11) and (24) is of size $J \times J$, where $J = N + M - 2$. For even moderate sample sizes (of the order of a few hundred observations) this makes the use of efficient numerical techniques imperative. There are two helpful circumstances: as an examination of e.g. (15) shows, C is sparse; and since it is a covariance matrix it is positive definite. The latter fact allows Choleski factorisation of the matrix, i.e. $C = U^t U$, so that

$$| C | = | U |^2 \qquad \mathbf{z}^t C^{-1} \mathbf{z} = \mathbf{z}^t U^{-1} (U^{-1})^t \mathbf{z} \tag{28}$$

The algorithm

$$\begin{aligned} u_{kk} &= \left(c_{kk} - \sum_{p=1}^{k-1} u_{pk}^2 \right)^{1/2} \\ u_{kj} &= \left(c_{kj} - \sum_{p=1}^{k-1} u_{pk} u_{pj} \right) / u_{kk} \qquad j > k \\ u_{kj} &= 0 \qquad j < k \end{aligned} \tag{29}$$

may be applied sequentially for $k = 1, 2, \ldots, J$ to find the elements of U in terms of those of C (Tewarson 1973). The utility of the factorisation lies in the upper triangular form of U, which allows easy evaluation of the functions $| U |$ and U^{-1} in (28). The elements a_{kj} of $A = U^{-1}$ are given by

$$\begin{aligned} a_{kk} &= u_{kk}^{-1} \\ a_{kj} &= -\sum_{p=k}^{j-1} a_{kp} u_{pj} / u_{jj}, \qquad j = k+1, k+2, \ldots, J \\ a_{kj} &= 0, \qquad j < k \end{aligned} \tag{30}$$

while

$$| U | = \prod_{k=1}^{J} u_{kk} \tag{31}$$

A brief description of the exploitation of the sparseness of C in the computer programming of the algorithms above is now given. Note first that it is unnecessary to store the elements of C: these are merely calculated as they are needed in the algorithm (29). Care needs to be taken with the ordering of the elements of $\mathbf{z}$, which again determines the arrangement of the entries in C, in order that U be as sparse as possible. The first ordering tried by the author, namely all $(N-1)$ values of Δx_j, followed by the $(M-1)$ values of Δy_k, proved very inefficient. The best arrangement appears to be according to the assumed time order of the increments (i.e. as determined by the current trial value of the lag τ. This of course implies that the order of the elements of $\mathbf{z}$, and hence C, changes with different trial lags).

Elements of U are stored in two vectors $V1$ and $V2$, containing respectively the diagonal and off-diagonal elements. A number of subsidiary vectors are used for storing the addresses of elements of U in the vector $V2$. Some of these are: $I1(j)$ indexes the U row number of entry j in $V2$; $I2(j)$ indexes the address of the next reference in $V2$ to the same U column as entry j; $J1(k)$ contains the address of the first reference in $V2$ to column k of U; and $J2(k)$ contains the address in $V2$ of the first reference to row k of U. Thus, $J1$ and $I2$ allow the elements of any given column of U to be rapidly located in $V2$, while the corresponding entry in $I1$ specifies the U row number. Since all elements of a given row in U are calculated sequentially (cf. (29)), $J2$ contains all the information necessary to locate U row elements.

The determinant of C is easily found as the squared product of all elements in $V1$ (see (28) and (31)). The second equation in (28) can be rewritten as

$$\mathbf{z}^t C^{-1}\mathbf{z} = \mathbf{z}^t A A^t \mathbf{z} = (A^t\mathbf{z})^t A^t\mathbf{z} = \left[\sum_{i=1}^{J}\sum_{j=1}^{J} a_{ji} z_j\right]^2 = \left[\sum_{j=1}^{J}\sum_{i=j}^{J} a_{ji} z_j\right]^2$$

where a_{ij} is the ij-th element of the upper triangular matrix A. The a_{ij} are calculated from the entries in $V2$ according to (30), with the necessary elements located by use of the subsidiary index vectors described above. Since the a_{ij} are calculated as needed, elements of the only non-sparse matrix of the problem, namely A, do not need to be stored. The major vectors in terms of storage requirements are $V2$, $I1$ and $I2$, which contain three times the number of non-zero off-diagonal elements in U, i.e. 3/2 times the number of non-zero off-diagonal elements in C. Thus, for example, if $N = M = 1000$, then the expected storage requirement is fewer than 10^4 elements. This should be contrasted with 10^6 storage allocations for the full matrix C alone.

2.6 EXAMPLE LAG DETERMINATIONS

2.6.1 0957+561

The likelihood function plotted in Fig. 23(a) is remarkably similar to the $\chi^2(\tau)$ plot of PRH. These authors found a best lag of -536 days, with a 90% confidence interval of ± 10 days. The corresponding results here is almost exactly the same, being a best lag of -538 days, with 95% confidence interval (-8,+10) days, as is easily established from the likelihood ratio statistic in Fig 23(b). Fig. 24 shows approximate 99%, 95% and 90% confidence envelopes for the likelihood function from 550 simulations based on the models 1a and 1b of Table 1, with an assumed lag

Figure 23: (a) The likelihood function for the data of Fig. 20 plotted against lag τ for the observations of the B component. The maximum is at a lag of -538.5 days. (b) The likelihood ratio statistic (27) plotted against component B lag.

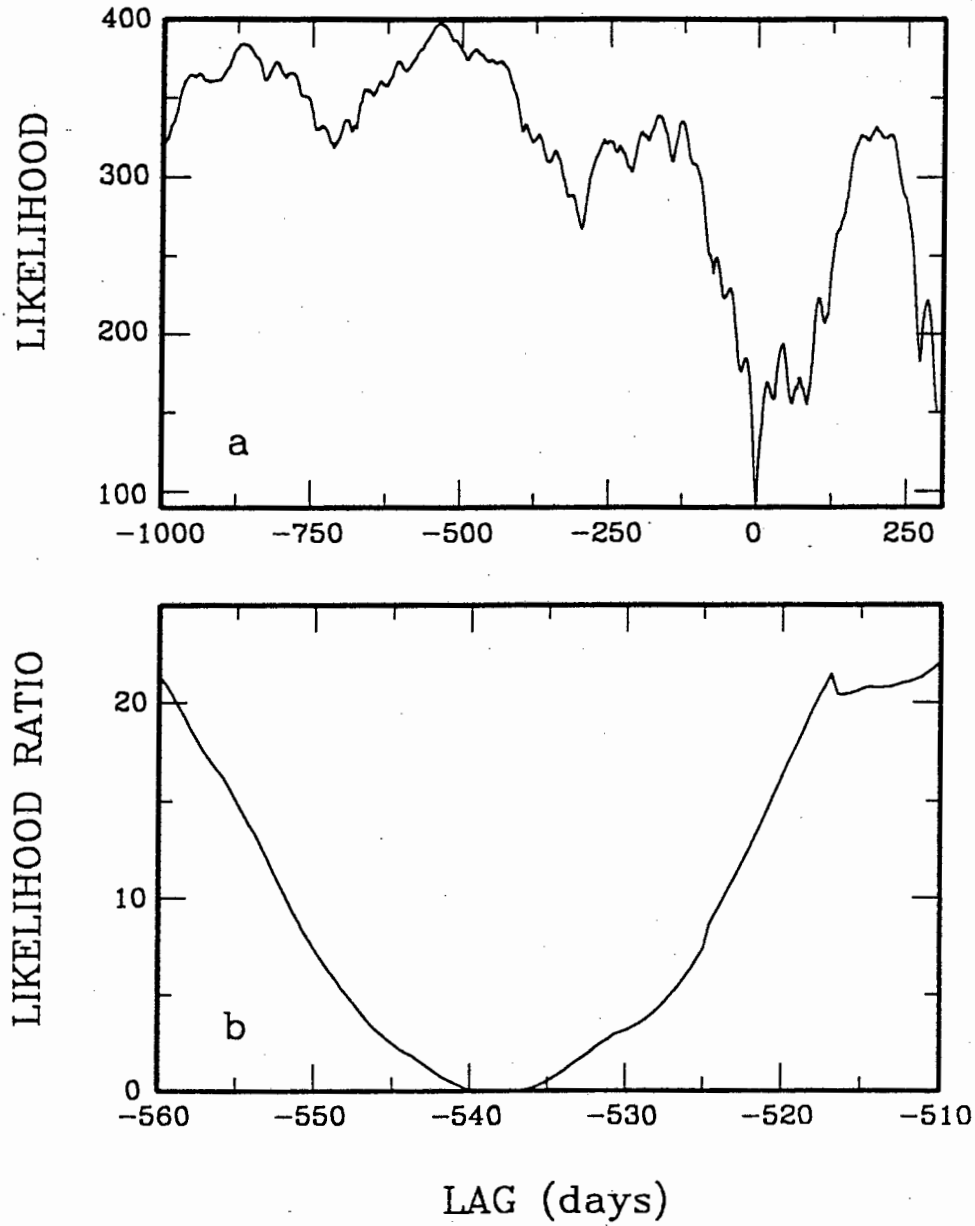
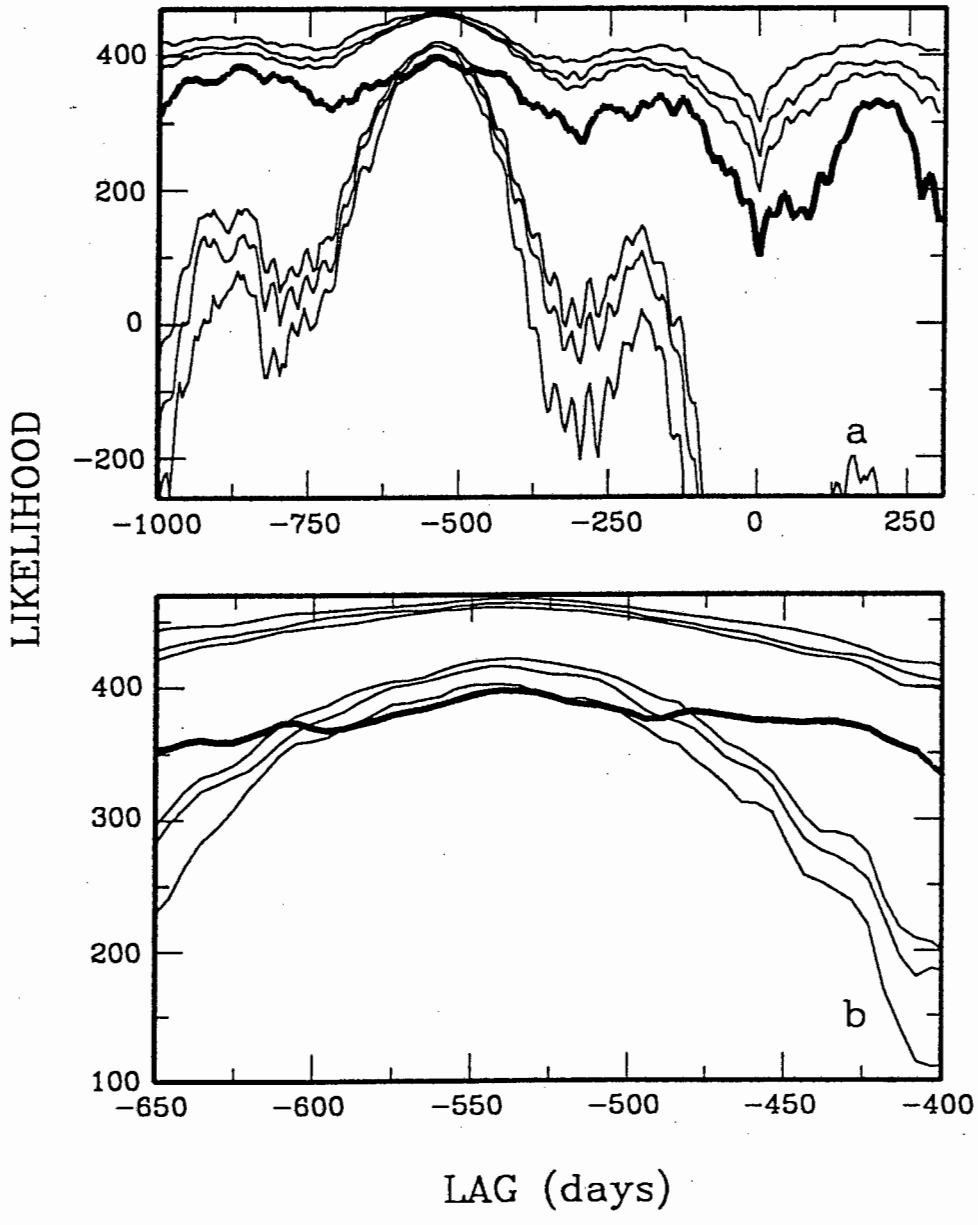


Figure 24: (a) Confidence envelopes for the likelihood function of Fig. 23(a). From outermost to innermost the probabilities are 99%; 95% and 90%. The observed likelihood is plotted as a heavy line. (b) Detail of (a), showing the deviation from the assumed model.



of -538 days. It is encouraging for the use of these envelopes that the most restrictive trial lag interval is indeed around the true lag. It is unfortunately also at these lags that the likelihood function of Fig. 23(a) lies outside the simulation bounds, suggesting that the assumed model is not entirely appropriate. A similar effect was found by PRH, who speculated that the slight discrepancy might be due to e.g. correlation between measurement errors of components A and B. Further investigation of the problem is warranted.

In order to test whether the estimated lag is well-defined, the first and second halves of the data may be analysed separately. It is necessary first to investigate whether the lag can be reliably estimated for such short data sets. Two hundred simulations of each of the first and second halves of the data were performed, with an assumed lag of -538 days of component B. The mean lag recovered form these simulations was -549 days (standard deviation 59 days) for the first, and -547 days (standard deviation 60) for the second half of the data. These results should be contrasted with the results of 550 simulations for the full data set: the mean lag was -539 days, with a standard deviation 19 days. There is clearly a large loss of accuracy when working with the shorter data spans. It should also be noted that the simulation confidence intervals for the lag are considerably wider than those from the likelihood ratio; the 95% interval, for example, is (-577, -502) days compared with (-546, -528) days from (27). This difference probably reflects the uncertainty in the random walk variances, which is not taken into account in the likelihood ratio statistic.

Figs. 25 and 26 contain plots of the likelihoods and likelihood ratios of the first and second halves of the data. The 95% confidence intervals from (20) consist of disjoint intervals in both cases: $(-635, -603) \cup (-541, -528)$ and $(-966, -929) \cup (-842, -835) \cup (-812, -806) \cup (-439, -420)$ for first and second halves respectively, with most probable values -621 and -951 days. For the first half of the data, the results appear to be consistent with simulation results. This is not so for the second half though: the most extreme solutions for the lag from 200 simulations were -872 and -191 days, so that the observed value -951 days appears incompatible with a true lag of -538 days. In Fig. 27 likelihood envelopes analogous to those in Fig. 24 are given for the two data sections.

It is perhaps a little surprising that the -538 day lag identified for the full data set is not recovered from either half. Its origin can nonetheless be seen in Figs. 25 and 26: the second most likely lag in Fig. 25 is at -536 days, while a lesser local extremum at around -533 days can be seen in Fig. 26.

The distribution of L_{max}, the maximum value of the likelihood function, is also of some interest. Although this quantity is ordinarily without intrinsic statistical meaning, it is useful in the context of parametric bootstrapping as an additional diagnostic for evaluating how well the postulated model represents the observations. For the full data set of Fig. 20, the mean and standard deviation of L_{max} was 441 and 12.0; the observed value was 397, 3.7σ below the mean. A similar result was obtained for the first half of the data (distribution mean 220, standard deviation 6.9, observed value 196), while the observed value $L_{max} = 203$ for the second half was more meaningful (distribution mean 211, standard deviation 9.7). In all these cases the distribution of L_{max} was consistent with Normality. A number of simulations (161) of two unrelated series were also performed. In this case the distribution of L_{max} was not Gaussian; the

Figure 25: As for Fig. 23, for the first half only of the data in Fig. 20.

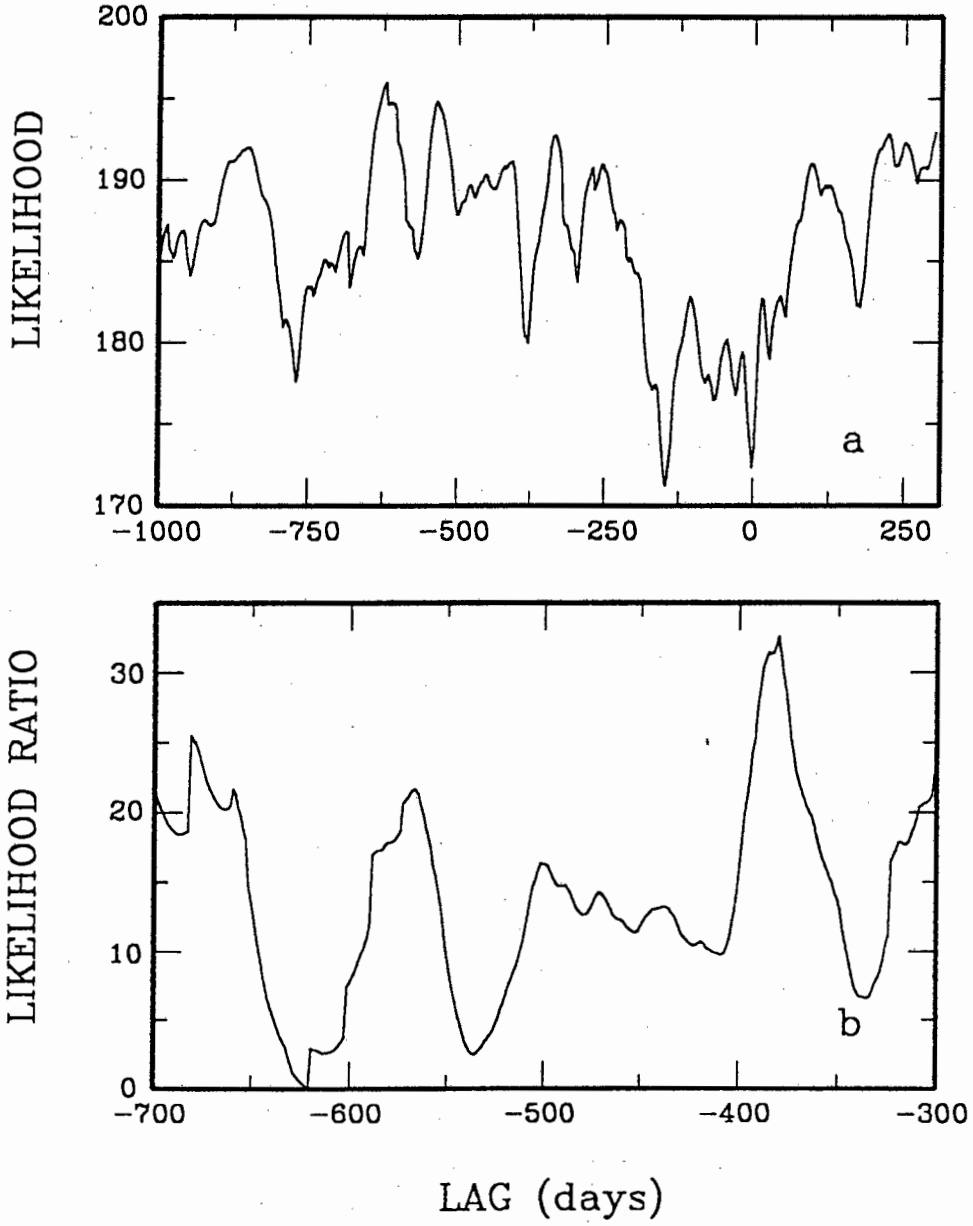


Figure 26: As for Fig. 23, for the second half only of the data in Fig. 20.

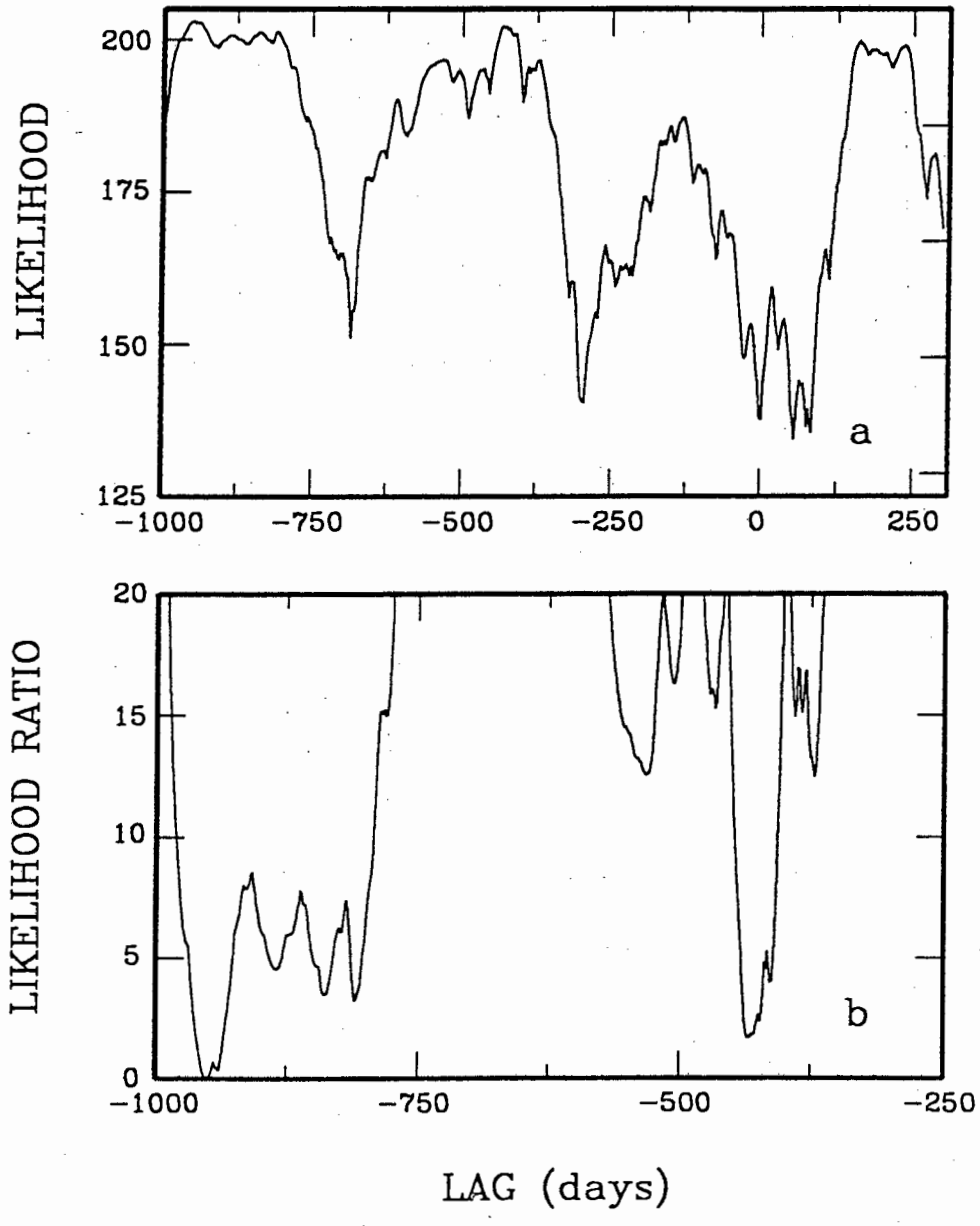


Figure 27: (a) Confidence envelopes for the likelihood function of Fig. 25(a). (b) Confidence envelopes for the likelihood function of Fig. 26(a). The outermost envelopes are for 95% confidence, innermost envelopes for 90% confidence. The observed likelihood is represented by the dotted line.

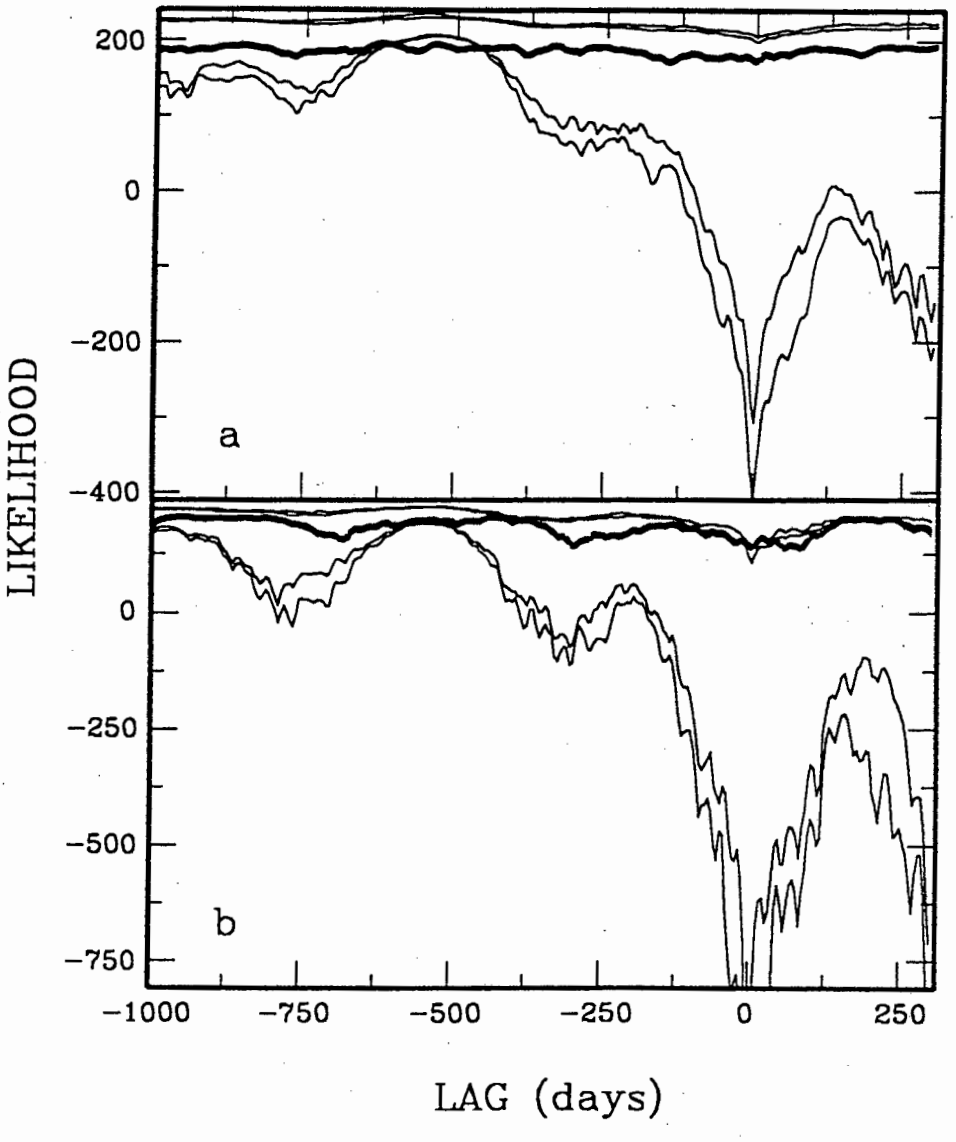
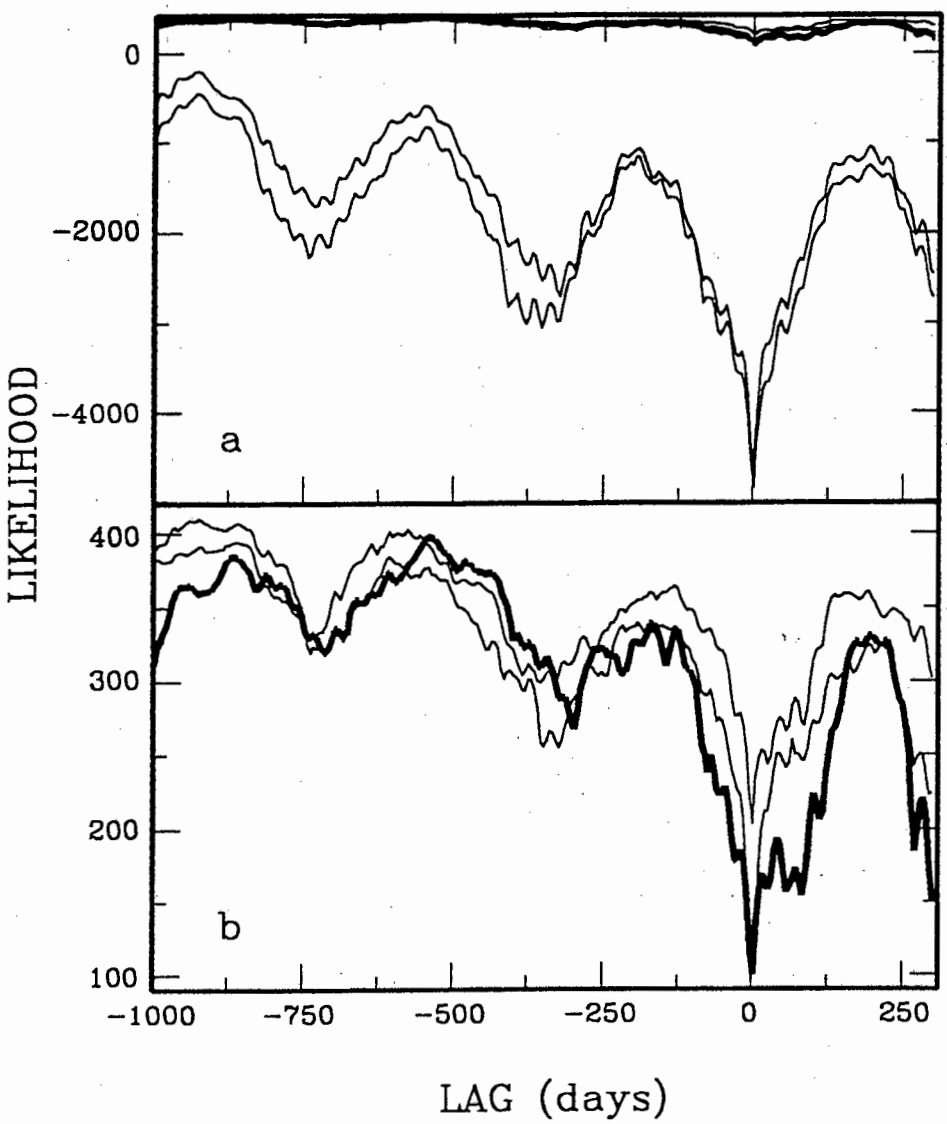


Figure 28: Confidence envelopes for the likelihood in the case where two series resembling those in Fig. 20 are unrelated. (a) Inner and outer envelopes are the 90% and 95% confidence bounds respectively. (b) The upper 90% and 95% bounds from (a), plotted on a larger scale. The heavy line shows the observed likelihood. The envelopes are based on the result of 161 simulations.



mean and standard deviation was 219 and 195, respectively. Of interest is that the probability of finding a value of $L_{max} \geq 397$ (the observed value for the data of Fig. 20) is about 18%. This may lead one to conclude that the 538 day lag identified, is spurious. Nonetheless, as the confidence envelopes based on these simulations show (Fig. 28), the likelihood function of the observations is certainly not typical of two unrelated series.

The histogram of Fig. 29, based on the simulations of two unrelated series, shows that some lags are much more prone to spurious identification: for example, there is a 48% chance of finding a lag in the range (-970, -904) days where none exists. The probability of finding a spurious lag in the range (-578, -513) days is 6%.

As a point of interest, 200 simulations of two series resembling those in Fig. 20, but with B lagging A by 200 days, were also generated. The mean and standard deviation of the estimated lags were -201 and 7.3 days, indicating that a 538 day lag may be relatively poorly constrained by the observation times.

2.6.2 NGC 3783

The likelihood function, and associated likelihood ratio, in Fig. 30 is interesting in that two disjunctive confidence intervals for the time shift of the K-band fluxes are identified. These are $(-88, -82)$ and $(-77, -74)$ days (95% confidence), with the most likely value $\tau = -84.4$ days. Comparison of the likelihood function with confidence envelopes derived from 500 simulations (Fig. 31) indicates that the statistical model based on a -84 day lag is acceptable. The mean and standard deviation of the lags estimated from these simulations were -87 and 12.3 days. The mean value of L_{max} from the simulations was -213, and the standard deviation 11.5; the observed maximum likelihood is -221.

The NGC 3783 data sets are unfortunately short ($N = 39$, 52 for the U and K observations respectively) so that experiments with subsets of the data are not feasible. Instead, Fig. 32 show the results of comparing the observed likelihood function to simulation envelopes based on three other lags. It is particularly interesting that the -75 day lag model (Fig. 32b) does not fare as well as the 84 day lag, despite the fact that it is within the 95% confidence interval based on the likelihood ratio.

As a final check, 400 sets of unrelated series with the same variances as the observed series were generated and the results analysed. The 90% and 95% confidence envelopes for the likelihood function are plotted in Fig. 33; the observed likelihood is consistently larger than the upper 95% curve over most of the domain of the plot. A histogram of the actual lags identified is shown in Fig. 34; there is a 6% chance of spuriously identifying a lag in the range (-90, -84) days. The probability of finding $L_{max} \geq -221$, the observed value, is about 14%.

2.7 CONCLUDING REMARKS

It is rather difficult to draw definite conclusions regarding the gravitationally lensed quasar 0957+561 on the basis of the evidence in section 2.6. Fig. 28 implies that the likelihood is large compared with that of series which are unrelated, while Fig. 24 shows that one should perhaps not take the identified lag of -538 days overly seriously. The latter point is also borne

Figure 29: The probability of identifying spurious lags between the two sets of observations of Fig. 20. The histogram is based on 161 simulations of unrelated series.

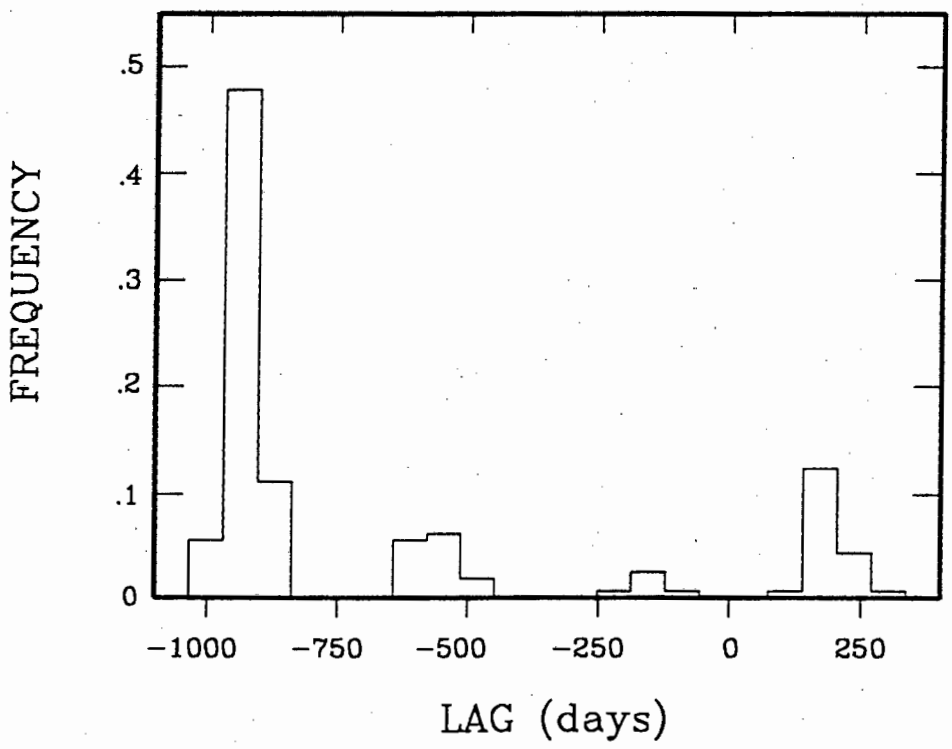


Figure 30: (a) The likelihood function of the data of Fig. 1. (b) The likelihood ratio. Note the two disconnected solutions.

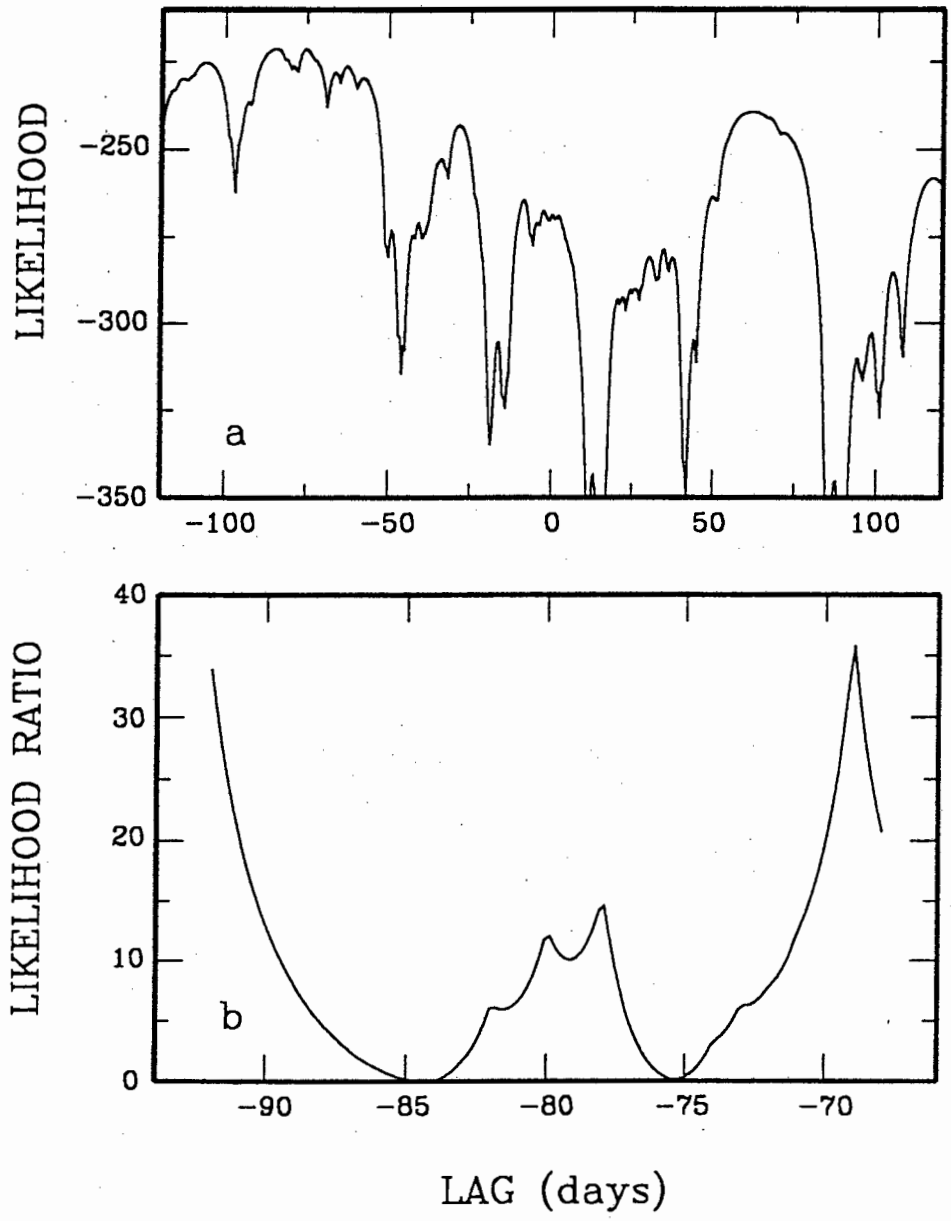


Figure 31: Ninety percent (inner bounds) and 95% simulation confidence envelopes for the likelihood function in Fig. 30 (heavy line).

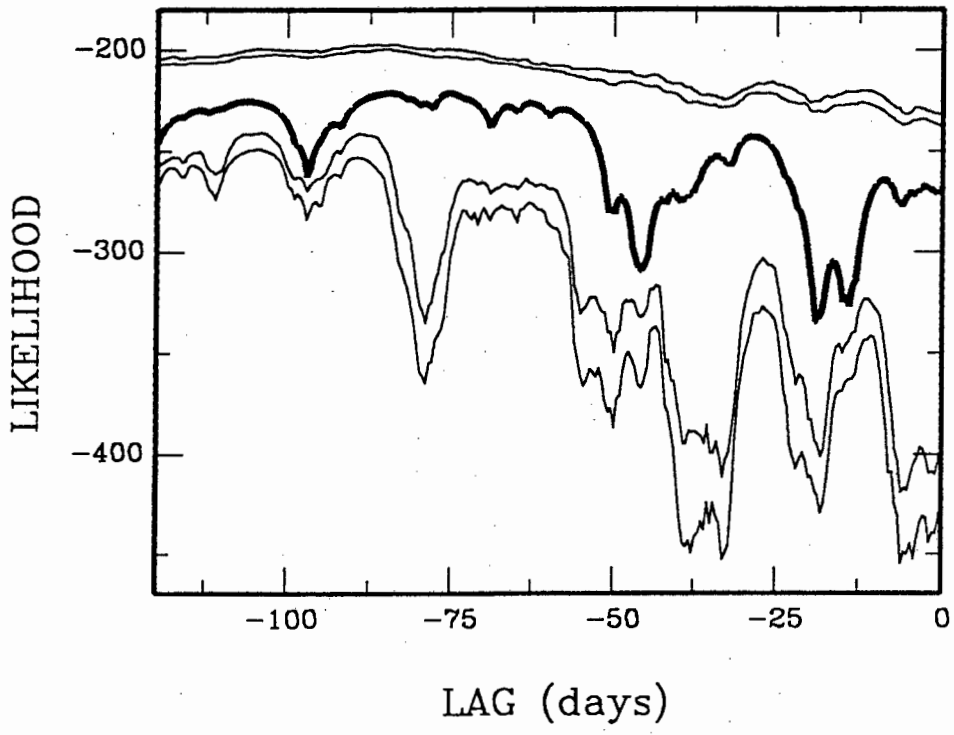


Figure 32: Confidence envelopes for the likelihood function of the data of Fig. 1, for assumed simulation lags of (a) -50; (b) -75; and (c) -100 days.

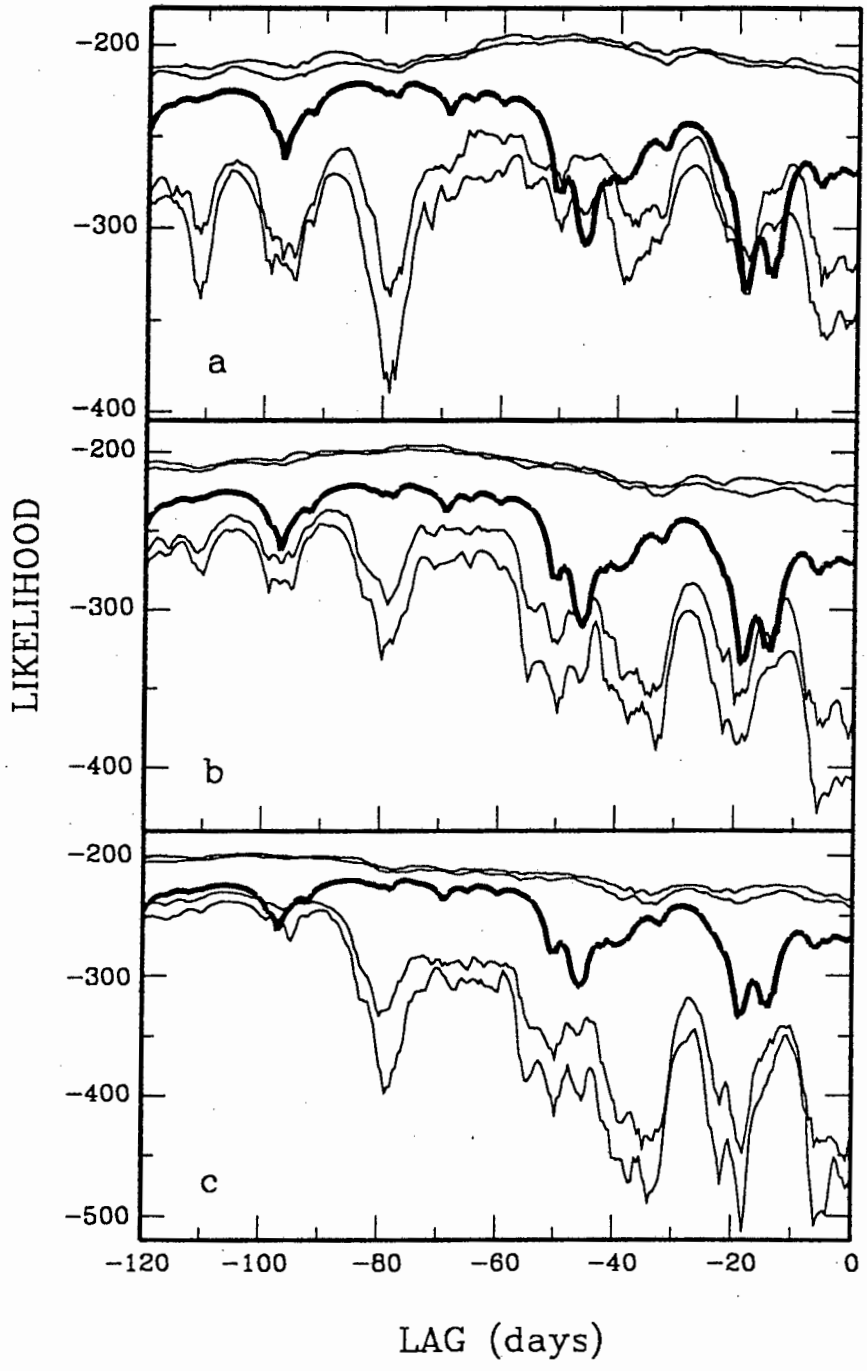


Figure 33: Confidence envelopes for the likelihood in the case where two series resembling those in Fig. 1 are unrelated. (a) Inner and outer envelopes are the 90% and 95% confidence bounds respectively. (b) The upper 90% and 95% bounds from (a), plotted on a larger scale. The heavy line shows the observed likelihood. The envelopes are based on the results of 400 simulations.

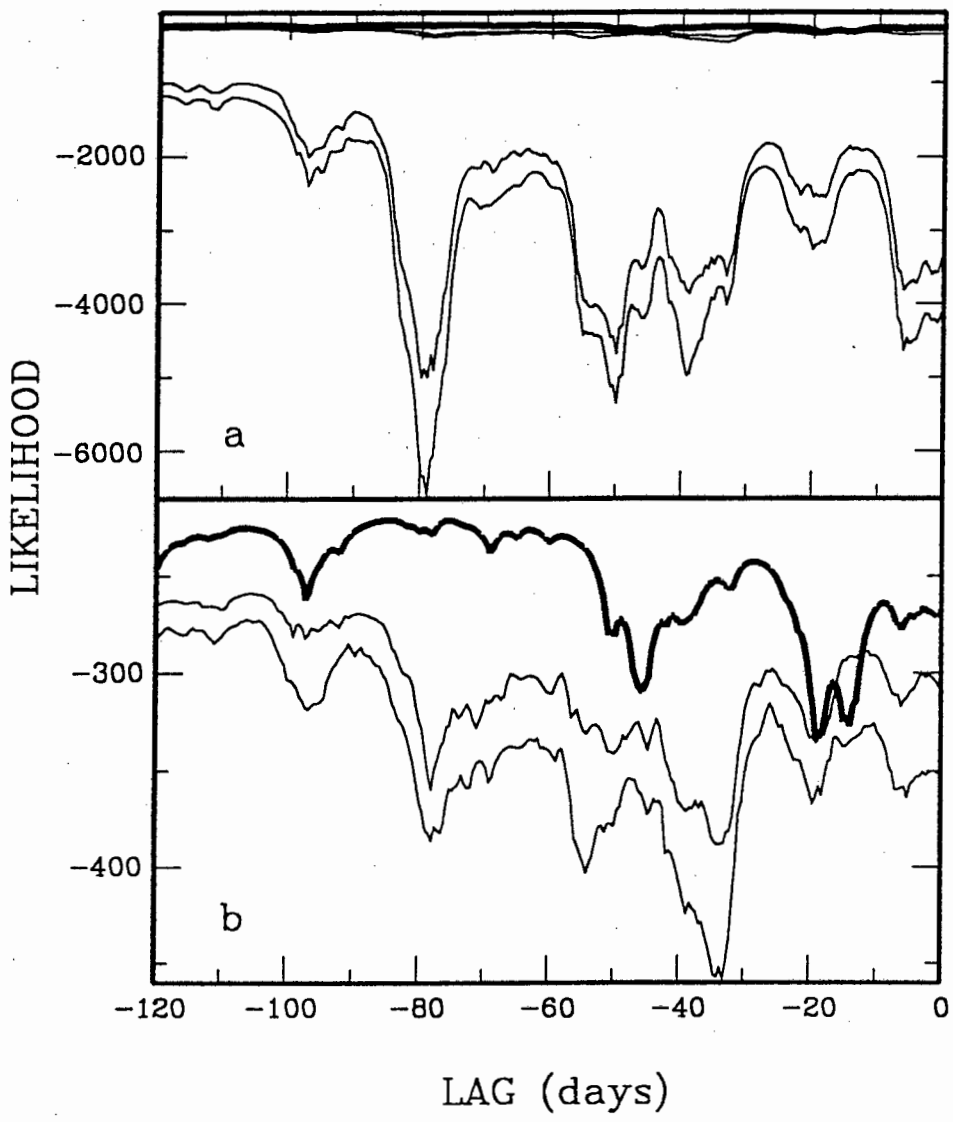
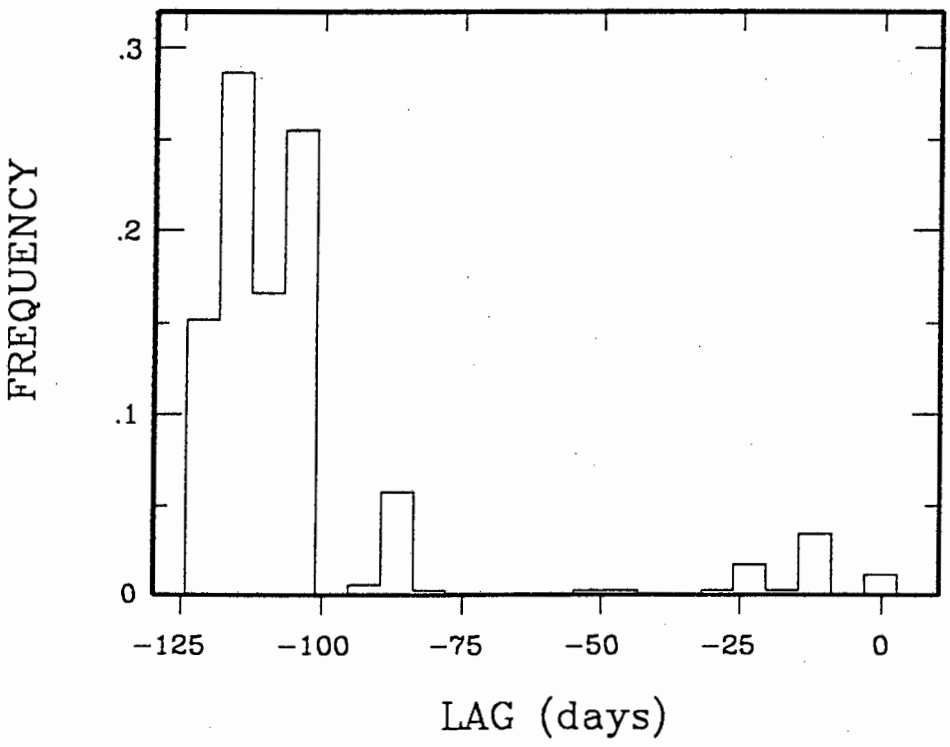


Figure 34: The probability of identifying spurious lags between the two sets of observations of Fig. 1. The histogram is based on 400 simulations of unrelated series.



out by Figs. 25 and 26. In the context of the model studied in this paper, the most appealing conclusion may be that the two series are in fact related, but not by any well-defined lag.

It must be pointed out that the model (15) used to calculate the covariance matrix of the 0957+561 data series pair ignores a further source of uncertainty: the σ_ϵ and σ_η are estimates, not perfectly known quantities. This is not accounted for, to give an example, in the simulation envelopes for these data sets. The effect of this uncertainty may be incorporated in the simulations in a crude way by making use of the well known distributional relationship between the estimated error variance s^2 and the ensemble mean value of the variance, σ^2:

$$N s^2/\sigma^2 \sim \chi^2_N\,, \tag{32}$$

where N is the number of degrees of freedom used for the estimation of s^2 (e.g. Mood, Graybill & Boes 1974). Of course, astronomers often do not estimate e.g. photometric errors by repeated measurements, hence values for the degrees of freedom should just be chosen to reflect approximately the accuracy of s^2. Simulation values for the measurement error variances may then be produced in accordance with (32) and used in the covariance matrix C instead of the quoted values.

FURTHER IDEAS

The simple random walk model (12) can be seen as a special case of the linear first order stochastic differential equation

$$\frac{d\,x(t)}{d\,t} = \alpha x(t) + \sigma_x \varepsilon(t) \tag{33}$$

(e.g. Vio et al. 1992). Of course, (33) can itself be generalised to higher order (Jones 1985) or non-linear forms (Vio et al. 1992). Clearly many different types of time series which are not random walks can be fitted by such equations (see Vio et al. 1992 for some examples). Some work has also been done on applying the theory of stochastic differential equations to multiple time series (Jones 1984, Harvey & Stock 1988). The author is not aware of any examples of lag determination by this method.

The results of Part 1 point to an oversimplification implicit in Part 2: the actual relation between the two series of observations may be more complicated than a single time lag. For example, in the case of multiwavelength observations of AGNs, the long wavelength radiation may show a response to changes in the blue radiation at more than one lag. This is not altogether surprising: it seems quite plausible, for example, that there might be simultaneous changes in the level of radiation over a broad spectral range, followed by different detailed behaviour over time at different wavelengths. There is, however, a potentially more serious problem for the analyst than the complexity of the true relation between $\{x(t)\}$ and $\{y(t)\}$: possible non-stationarity of any such relation. In the context of the approach of Part 2 above, such non-stationarity could be due to changes in the lag between the two series. This may be investigated for very long sets of observations by studying segments of the series and comparing the estimated lags. In a slightly more sophisticated treatment the lag could be calculated for many possibly overlapping data windows, and plotted as a function of the central window position (for more on this approach see the discussion by Harding & Sew-Hee following Bruce & Martin 1989, or Sew-Hee 1988). For shorter series the identification of non-stationarity will be much more difficult.

In the case of regularly spaced observations the relation between the series can be studied in much finer detail by the methods of time domain transfer function modelling (as in Part 1 of this thesis). It is desirable to be able to apply transfer function type analyses also to data of the form of Fig. 1. This will entail extension of the theory described in section 2.3 above. A possible approach to fitting models of the form

$$y(t) = A_1 x(t-\tau_1) + A_2 x(t-\tau_2) + \cdots + A_n x(t-\tau_n) + \zeta(t) \tag{34}$$

where $x(t)$ is a random walk, is now outlined.

A very important difference between (21) and (34) is that in the latter case $y(t)$ may not be a random walk, despite the fact that $x(t)$ may be one. The increments Δy_j and Δy_i are not necessarily uncorrelated for disjoint time intervals Δt_j and Δt_i:

$$\begin{aligned}\mathrm{cov}(\Delta y_j, \Delta y_i) &= (A_1^2 + A_2^2) \parallel [t_{y,j}, t_{y,j+1}] \cap [t_{y,i}, t_{y,i+1}] \parallel \\ &+ A_1 A_2 \parallel [t_{y,j} - \tau_1, t_{y,j+1} - \tau_1] \cap [t_{y,i} - \tau_2, t_{y,i+1} - \tau_2] \parallel \\ &+ A_1 A_2 \parallel [t_{y,j} - \tau_2, t_{y,j+1} - \tau_2] \cap [t_{y,i} - \tau_1, t_{y,i+1} - \tau_1] \parallel \end{aligned} \quad (35)$$

where, for the sake of simplicity $n = 2$, $\mathrm{var}(x) = 1$ and $\zeta(t) = 0$ have been taken in (34). For the purpose of studying the autocorrelation of the Δy, τ_1 plays no role, and (35) may be simplified to

$$\begin{aligned}\mathrm{cov}(\Delta y_j, \Delta y_i) &= (A_1^2 + A_2^2) \parallel [t_{y,j}, t_{y,j+1}] \cap [t_{y,i}, t_{y,i+1}] \parallel \\ &+ A_1 A_2 \parallel [t_{y,j}, t_{y,j+1}] \cap [t_{y,i} - \tau_*, t_{y,i+1} - \tau_*] \\ &+ [t_{y,j} + \tau_*, t_{y,j+1} + \tau_*] \cap [t_{y,i}, t_{y,i+1}] \parallel \end{aligned} \quad (36)$$

where $\tau_* = \tau_2 - \tau_1$. Equation (36) can be used as a basis for the maximum likelihood estimation of τ_*. Note that it will be necessary to simultaneously estimate A_1 and A_2. The likelihood function may be viewed as a continuous time analog of the autocorrelation function. Significance may be assessed by simulation. The value of τ_1 can then be determined by constructing the joint likelihood function of all Δx and Δy. The appropriate covariance matrix elements may be found from

$$\begin{aligned}\mathrm{cov}(\Delta y_j, \Delta x_i) &= A_1 \parallel [t_{y,j} - \tau_1, t_{y,j+1} - \tau_1] \cap [t_{x,i}, t_{x,i+1}] \parallel \\ &+ A_2 \parallel [t_{y,j} - \tau_* + \tau_1, t_{y,j+1} - \tau_* + \tau_1] \cap [t_{x,i}, t_{x,i+1}] \parallel \end{aligned}$$

Of course, (34) is still rather restrictive in allowing only a number of discrete lags between $y(t)$ and $x(t)$. In practice one might expect a continuous range of lags, as e.g. various parts of a dust complex respond to irradiation by high energy photons at different times. This could presumably be modelled by some continuous function of lag analogous to the linear transfer function (5), e.g.

$$y(t) = A \int_0^\infty f(q) x(t - \tau)\, d\tau + \zeta(t) \quad (37)$$

where A is a constant, and $f(\tau)$ is a suitable weighting function. A candidate for the function f is the exponential $\mathrm{e}^{-\alpha\tau}$ with $\alpha > 0$ a constant. The values of α and A could again be determined by maximum likelihood methods, and significance tests be performed by simulation.

There exits a substantial literature on frequency domain methods for dealing with models of precisely the form (37). Astronomers make frequent use of univariate spectral methods, but there are very few examples of the use of the cross spectrum of two series in the astronomy literature. Functions which can be derived from the cross spectrum are the coherency and the phase spectrum, which respectively measure the degree of dependence between the two series and the phase relationship between them (Bloomfield 1976). The phase spectrum may be used to estimate the lag between the two series, as discussed for example in Hannan & Thomson (1988). The application of these methods to the data of Fig. 1 is not straightforward though: first,

the data are unevenly spaced, so that special methods may be required for spectral estimation (Masry 1976, Marquard & Acuff 1982, Bronez 1988). Secondly, the series dealt with may be non-stationary in the mean. This would be of lesser importance if spectral leakage were minor, so that low frequencies could simply be excluded from the analysis. The results of preliminary studies indicate that the situation may be more complicated than this for some series of interest. Presumably the data will therefore have to be de-trended, as in the application to atmospheric carbon dioxide levels by Kuo et al. (1990).

REFERENCES

Bargmann, R. E. (1984). Matrices and Determinants, in Beyer W. H., ed., CRC Standard Mathematical Tables (27th edition). CRC Press, Boca Raton.

Bloomfield, P. (1976). Fourier Analysis of Time Series: An Introduction. John Wiley & Sons, New York.

Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis, Forecasting and Control. Holden-Day, Oakland.

Box, G. E. P. and Newbold, P. (1971). Some comments on a paper by Coen, Gomme and Kendall, J. Royal Statist. Soc. A, 134, 229-240.

Bronez, T. P. (1988). Spectral estimation of irregularly sampled multidimensional processes by generalized prolate spheroidal sequences, IEEE Transaction Acoustics Speech Signal Processing, 36, 1862-1873.

Bruce, A. G. and Martin, R. D. (1989). Leave-k-out diagnostics for time series, J. R. Statist. Soc. B, 51, 363-424.

Clavel, J., Wamsteker, W. and Glass, I.S. (1989). Hot dust on the outskirts of the broad-line region in Fairall 9, Astrophys. J., 337, 236-250.

Clavel, J. and 56 others (1991). Steps toward determination of the size and structure of the broad-line region in active galactic nuclei. I. An 8 month campaign of monitoring NGC5548 with IUE, Astrophys. J., 366, 64-81. (CEA91).

Cox, D. R. and Hinkley, D. V. (1974). Theoretical Statistics. Chapman & Hall, London.

Edelson, R. A. and Krolik, J. H. (1988). The discrete correlation function: a new method for analyzing unevenly sampled variability data, Astrophys. J., 333, 646-659.

Gaskell, C. M. and Peterson, B. M. (1987). The accuracy of cross-correlation estimates of quasar emission-line region sizes, Astrophys. J. Suppl., 65, 1-11.

Glass, I. S. (1992). Infrared variability of the Seyfert galaxy NGC 3783, Monthly Notices Roy. Astron. Soc., 256, 23P-27P.

Graybill, F. A. (1969). Introduction to Matrices with Applications in Statistics. Wadsworth Publishing Co., Belmont.

Hamon, B. V. and Hannan, E. J. (1974). Spectral estimation of time delay for dispersive and non-dispersive systems, Appl. Statist., 23, 134-142.

Hannan, E. J. and Thomson, P. J. (1988). Time delay estimation, J. Time Series Analysis, 9, 21-33.

Harvey, A. C. (1989). Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge

Harvey, A. C. and Stock, J. H. (1988). Continuous time autoregressive models with common stochastic trends, J. Econom. Dynamics & Control, 12, 365-384.

Jenkins, G. M. (1979). Practical Experiences with Modelling and Forecasting Time Series. Gwilym Jenkins & Partners, St. Helier.

Jones, R. H. (1984). Fitting multivariate models to unequally spaced data, in Parzen E., ed., Time Series Analysis of Irregularly Observed Data (Lecture Notes in Statistics No. 25). Springer Verlag, New York.

Jones, R. H. (1985). Time series analysis with unequally spaced data, in Hannan E. J., Krishnaiah P. R. and Rao M. M., eds., Handbook of Statistics Volume 5. Elsevier, Amsterdam.

Kuo, C., Lindberg, C. and Thomson, D. J. (1990). Coherence established between atmospheric carbon dioxide and global temperature, Nature, 343, 709-714.

Marquard, D. W. and Acuff, S. K. (1982). Direct quadratic spectrum estimation from unequally spaced data, Anderson O. D. and Perry M. R., eds., Applied Time Series Analysis. North Holland, Amsterdam.

Masry, E.(1976). Discrete-time spectral estimation of continuous-parameter processes - a new consistent estimate, IEEE Transactions Information Theory, IT-22, 298-312.

Mood, A. M., Graybill, F. A. and Boes, D. C. (1974). Introduction to the Theory of Statistics (Third Edition). McGraw-Hill, New York.

Pankratz, A. (1983). Forecasting with Univariate Box Jenkins Models. John Wiley & Sons, New York.

Pierce, D. A. (1977). Relationships - and the lack thereof - between economic time series, with special reference to money and interest rates, J. Amer. Stat. Ass., 72, 11-22.

Press, W. H., Rybicki, G. B. and Hewitt, J. N. (1992). The time delay of gravitational lens 0957+561. I. Methodology and analysis of optical photometric data, Astrophys. J., 385, 404-415. (PRH)

Ripley, B. D. (1988). Statistical Inference for Spatial Processes. Cambridge University Press, Cambridge.

Scargle, J. D. (1981). Studies in astronomical time series analysis. I. Modeling random processes in the time domain, Astrophys. J. Suppl., 45, 1-71.

See-Hew, K. K. K. Y. (1988). PhD Thesis, University of Cambridge.

Smith, A. G., Nair, A. D., Leacock, R. J. and Clements, S. D. (1993). The longer optical time scales of a large sample of quasars, Astron. J., 105, 437-455.

Sterne, T. E. (1934). The errors of period of variable stars.-Part I: The general theory, illustrated by RR Scorpii, Harvard Circular No. 386.

Tewarson, R. P. (1973). Sparse Matrices. Academic Press, New York.

Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis, Appl. Statist., 41, 1-15.

Van Langevelde, H. J., Van der Heiden, R. and Van Schooneveld, C. (1990). Phase lags from multiple flux curves of OH/IR stars, Astron. & Astrophys., 239, 193-204.

Vanderriest, C., Schneider, J., Herpe, G., Chevreton, M., Moles, M. and Wlérick, G. (1989). The value of the time delay $\Delta T(A,/,B)$ for the "double" quasar 0957+561 from optical photometic monitoring, Astron. & Astrophys., 215, 1-13.

Vio, R., Cristiani, S., Lessi, O. and Provenzale, A. (1992). Time series analysis in astronomy: an application to quasar variability studies, Astrophys. J., 391, 518-530.