Neurocomputing 72 (2008) 32–38

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Bayesian estimation of time delays between unevenly sampled signals

Markus Harva [a,*], Somak Raychaudhury [b]

[a] Laboratory of Computer and Information Science, Helsinki University of Technology, P.O. Box 5400, FI-02015 TKK, Espoo, Finland
[b] School of Physics and Astronomy, University of Birmingham, Birmingham B15 2TT, UK

ARTICLE INFO

Available online 26 September 2008

Keywords:
Delay estimation
Irregular sampling
Bayesian methods
MCMC
Kalman filtering

ABSTRACT

A method for estimating time delays between signals that are irregularly sampled is presented. The approach is based on postulating a latent variable model encoding the assumption of slow variability of the underlying source signal. The posterior distribution of the delay is obtained partly by exact marginalisation computable by a specific type of Kalman filter and partly by Markov chain Monte Carlo. Experiments with artificial data show the effectiveness of the proposed approach while results with real-world gravitational lens data provide the main motivation for this work.

1. Introduction

The estimation of a delay between two signals is concerned with the following question: given two time series $x_1(t)$ and $x_2(t)$, for which time lag τ does the relation $x_1(t) = ax_2(t-\tau) + b$ hold most accurately? The question is of importance in many fields of signal processing, from underwater acoustics to satellite positioning. Much of the work done in answering the question has dealt with evenly sampled signals, and the solution is most often based on the cross correlation function or a variation of it.

The problem drastically changes when the signals are unevenly sampled, which is a frequently encountered case in certain fields. In observational astronomy, for example, uneven sampling results from the sampling times being largely determined by external factors, such as observing conditions, instrumental availability, and scheduling. In the uneven sampling case, the cross correlation function cannot usually be exactly evaluated for any case other than $\tau = 0$. For other values, some form of interpolation needs to be performed, which means that one implicitly assumes that the signals are slowly varying. Even when that is a legitimate assumption, interpolation can pose a risk of inventing new data that have little to do with reality [5]. Consequently, several methods have been proposed as alternatives to the standard cross correlation function. Of these, perhaps the most widely used are the discrete correlation function [5] (especially the locally normalised version [12]) and the dispersion spectrum [13].

A recent study [4], involving one of the authors of this paper, shows that the delay estimation problem with irregular sampling, mainly in the context of the determination of time lags in gravitationally lensed multiple images of distant celestial objects, is far from completely solved.

In this paper, we propose a Bayesian approach to solving the delay estimation problem with unevenly sampled signals. We avoid any kind of interpolation on the noisy data by formulating a latent-variable model which exploits the assumption of the signals being slowly varying. We derive an inference algorithm for our model partly based on exact marginalisation and partly on MCMC methods [6]. Although the main motivation of this work is solving the delays in gravitational lensing, no domain specific assumptions are made, and the method is of general applicability as long as the assumption of slow variability can be made.

In the next section we delineate our modelling assumptions. In Section 3 the algorithm for computing the posterior probability distribution of the delay is derived. Section 4 reports both comparison experiments against other methods as well as results with real-world gravitational lensing data.

This article extends our earlier conference paper [7] on the same topic. Here, we give an alternative derivation for the computation of the marginal likelihood and report the experiments in more detail showing results against one additional comparison method.

2. Modelling assumptions

In the basic setting, we have two sets of observations, $x_1(t)$ and $x_2(t)$, measured at time instants $\{t'_i : i = 1, \dots, N\}$. Both of these observations are due to a common source $s(t)$, but the second observation is delayed by τ. With a fixed delay τ, we can think of having observations at $2N$ distinct time instants, $\{t'_i : i = 1, \dots, N\} \cup \{t'_i - \tau : i = 1, \dots, N\}$. By ordering these we obtain a signal $x(t_i)$, measured at time instants $\{t_i : i = 1, \dots, 2N\}$, which

* Corresponding author. Tel.: +358 9 451 3287; fax: +358 9 451 3277.
E-mail addresses: markus.harva@tkk.fi (M. Harva), somak@star.sr.bham.ac.uk (S. Raychaudhury).

0925-2312/$ - see front matter
doi:10.1016/j.neucom.2007.12.046

incorporates the samples from both x_1 and x_2. We denote the correspondence with $k(i) \in \{1,2\}$ such that $x(t_i) = x_{k(i)}(t'_j)$ for some j. The source can also be shifted and scaled individually for each of the two observations. Denoting the data as $\mathbf{X} = \{x(t_i) : i = 1, \ldots, 2N\}$, the source as $\mathbf{S} = \{s(t_i) : i = 1, \ldots, 2N\}$, the scaling coefficients and the shifts as a_k and b_k, respectively, and finally the log-variances of the noise as v_k, we can write the likelihood for the model parameters as

$$p(\mathbf{X}|\mathbf{S}, \{a_k\}, \{b_k\}, \{v_k\}, \tau) = \prod_{i=1}^{2N} \mathcal{N}(x(t_i)|a_{k(i)}s(t_i) + b_{k(i)}, e^{v_{k(i)}}). \quad (1)$$

Here, $\mathcal{N}(x|\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. As we have no a priori knowledge about the nature of the noise, the Gaussian assumption is justified by the maximum-entropy principle [9].

The mere likelihood above does not relate the individual observations to each other in any way, and thus using it alone it is not possible to estimate the delay. The additional assumption we make is that the source varies slowly in time, such that the next state $s(t_{i+1})$ is dependent on the previous state $s(t_i)$. We incorporate this assumption to the model by specifying a Wiener-process prior on $\mathbf{S}$:

$$\begin{aligned} p(\mathbf{S}|w, \gamma, \tau) &= \prod_{i=2}^{2N} p(s(t_i)|s(t_{i-1}), w, \gamma) \times p(s(t_1)) \\ &= \prod_{i=2}^{2N} \mathcal{N}(s(t_i)|s(t_{i-1}), (t_i - t_{i-1})^{\gamma} e^{w}) \mathcal{N}(s(t_1)|0, \sigma_s^2). \quad (2) \end{aligned}$$

The prior states that the fluctuations in the source are proportional to the temporal distance between consecutive observations. We see this formulation as least restrictive while still incorporating the prior knowledge. The constant σ_s^2 specifies the prior variance of the source at the first time instant. Since the data can be normalised as a preprocessing step, we can use a moderate value of $\sigma_s^2 = 10^2$.

On the surface, it might seem that the delay has no role in Eqs. (1) and (2). It of course does, but the effect is of structural nature as it affects the combined time instants t_i and consequently the order in which the data samples appear in the model. This is illustrated in Fig. 1.

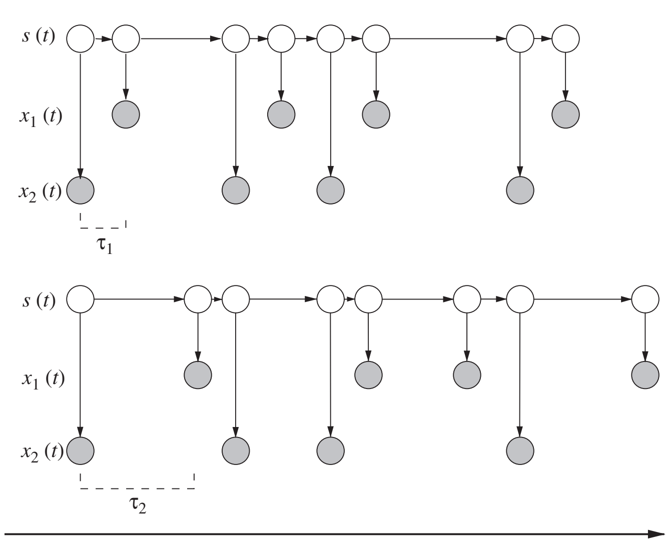

Fig. 1. A schematic illustration of how the time delay τ affects the structure of the model.

For the parameters we choose the following priors:

$$\begin{aligned} &p(\tau) = \mathcal{N}(\tau|\mu_\tau, \sigma_\tau^2), \\ &p(w) = \mathcal{N}(w|-5, 5^2), \\ &p(v_k) = \mathcal{N}(v_k|-5, 5^2). \end{aligned} \quad (3)$$

The constants μ_τ and σ_τ^2 are chosen to reflect the earlier knowledge of the delay (if such exists), but so that they do not constrain the delay too much. We have considered γ both as a variable and a constant, but have found that little difference is made in estimating it, and so we keep it fixed to $\gamma = 2$.

For the scale and shift parameters to be identifiable, we fix a_1 and b_1 to 1 and 0, respectively. Due to the normalisation, the values of the parameters a_2 and b_2 should not be too far from a_1 and b_1. Hence, for them we have the following priors:

$$\begin{aligned} &p(a_2) = \mathcal{N}(a_2|1, 1), \\ &p(b_2) = \mathcal{N}(b_2|0, 1). \end{aligned}$$

When we have knowledge of the uncertainties related to the measuring process, we want to incorporate that to the model as well. In that case, we actually postulate $\mathbf{X}$ as hidden variables from which we have the noisy observations $\mathbf{Y} = \{y(t_i) : i = 1, \ldots, 2N\}$ with *known* standard deviations $\sigma_y(t_i)$:

$$p(\mathbf{Y}|\mathbf{X}) = \prod_{i=1}^{2N} \mathcal{N}(y(t_i)|x(t_i), \sigma_y^2(t_i)). \quad (4)$$

This concludes the formulation of the model in the case of two observed signals. The same construction extends straightforwardly to the multiple (>2) signals case. Also, when several measurements from different wavelengths or distinct time intervals are available, we can use them jointly. If we have M datasets, we also have M set of parameters, excluding the delay (delays) which is (are) common to all datasets.

3. Learning the model

We are only interested in the posterior distribution of the delay $p(\tau|\mathbf{Y})$ and consider all the other variables in the model to be nuisance parameters which we would like to marginalise out. In the next subsection we show how to get rid of most of the nuisance parameters by exact marginalisation which results in a recursive formula for the marginal likelihood. This reduces the dimensionality of the parameter space radically making sampling from the rest of the variables effective. The sampling procedure is discussed in Section 3.2.

3.1. Marginalisation

We denote by $\boldsymbol{\theta}$ all the time-independent parameters in the model, excluding the delay. We are going to compute the marginal likelihood $p(\mathbf{Y}|\tau, \boldsymbol{\theta}) = \int d\mathbf{X}\, d\mathbf{S}\, p(\mathbf{Y}, \mathbf{X}, \mathbf{S}|\tau, \boldsymbol{\theta})$. This integral is analytically tractable and can be directly computed by recursive integration, and that is how it was evaluated in [7]. But it is perhaps more illuminating to derive the likelihood by noting that conditional on τ and $\boldsymbol{\theta}$, $p(\mathbf{Y}, \mathbf{X}, \mathbf{S}|\tau, \boldsymbol{\theta})$ is essentially a Kalman-filter model. By identifying the correspondence, the results concerning Kalman filtering apply. This derivation is straightforward but a bit lengthy, and hence it is postponed to Appendix A. Here, we directly proceed to give the expression for the marginal likelihood. First, we need the following notation:

$$\hat{y}(t_i) = \frac{y(t_i) - b_{k(i)}}{a_{k(i)}} \quad \text{and} \quad \tilde{y}(t_i) = \frac{e^{v_{k(i)}} + \sigma_y^2(t_i)}{a_{k(i)}^2}.$$

Now the ingredients of the marginal likelihood can be computed by the following recursive filter:

$$\tilde{s}(t_1) = (\sigma_s^{-2} + \tilde{y}^{-1}(t_1))^{-1},$$
$$\hat{s}(t_1) = \tilde{s}(t_1)\hat{y}(t_1)/\tilde{y}(t_1) \tag{5}$$

and for $i > 1$

$$\tilde{s}(t_i) = \{[(t_i - t_{i-1})^{\gamma} e^{w} + \tilde{s}(t_{i-1})]^{-1} + \tilde{y}^{-1}(t_i)\}^{-1},$$
$$\hat{s}(t_i) = \tilde{s}(t_i)\{\hat{s}(t_{i-1})/[(t_i - t_{i-1})^{\gamma} e^{w} + \tilde{s}(t_{i-1})] + \hat{y}(t_i)/\tilde{y}(t_i)\}. \tag{6}$$

Finally, equipped with the above quantities, we can express the marginal likelihood as

$$p(\mathbf{Y}|\tau, \boldsymbol{\theta}) = \prod_{i=2}^{2N} \frac{1}{|a_{k(i)}|} \mathcal{N}(\hat{y}(t_i)|\hat{s}(t_{i-1}), (t_i - t_{i-1})^{\gamma} e^{w} + \tilde{s}(t_{i-1}) + \tilde{y}(t_i))$$
$$\times \frac{1}{|a_{k(1)}|} \mathcal{N}(\hat{y}(t_1)|0, \sigma_s^2 + \tilde{y}(t_1)). \tag{7}$$

3.2. Sampling procedure

The final step in obtaining summaries from $p(\tau|\mathbf{Y})$ is to sample from $p(\tau, \boldsymbol{\theta}|\mathbf{Y})$. As an alternative, we considered numerically marginalising over $\boldsymbol{\theta}$, but the dimensionality is still high enough, especially in the case of more than one delay and more than two datasets, to render the task intractable.

The Metropolis–Hastings algorithm [8] is particularly suitable here, as it only requires us to be able to evaluate the unnormalised posterior density—something we can readily accomplish by multiplying the marginal likelihood of the previous section with the prior. In addition, a suitable jumping distribution needs to be formulated such that the resulting Markov Chain converges reasonably fast. In the experiments that follow, the form of the jumping distribution was either Gaussian or a mixture of Gaussians. The parameters were set up experimentally to obtain a rejection ratio that was not too small or too big.

A crucial part in using MCMC methods is the evaluation of the convergence of the chain. If the Markov chain has not converged to its equilibrium distribution, the samples do not come from the desired distribution making further inferences invalid. Although there is no fool proof method to assess the convergence, there exists several schemes that have proved to be useful in practice. One such method, and the one we have used here, is to compute the so-called potential scale reduction factor [3] from several parallel Markov chains started from random initial states. It should give an estimate of how much the scale of the empirical posterior could be reduced by obtaining more samples. Practically, it compares the first and second order moments of the parallel chains.

4. Experimental results

The main motivation of this work is the estimation of delays in gravitational lensing systems. But comparing methods using real datasets is difficult as the true delay remains unknown. Hence, we have made comparison experiments using artificial datasets, as then we can make well-justified claims about the accuracy of different methods. These are reported in the next subsection. Results with real-world gravitational lensing datasets are reported and discussed in Section 4.2.

4.1. Controlled comparisons

We compare the proposed approach to the following three widely used methods: locally normalised discrete correlation

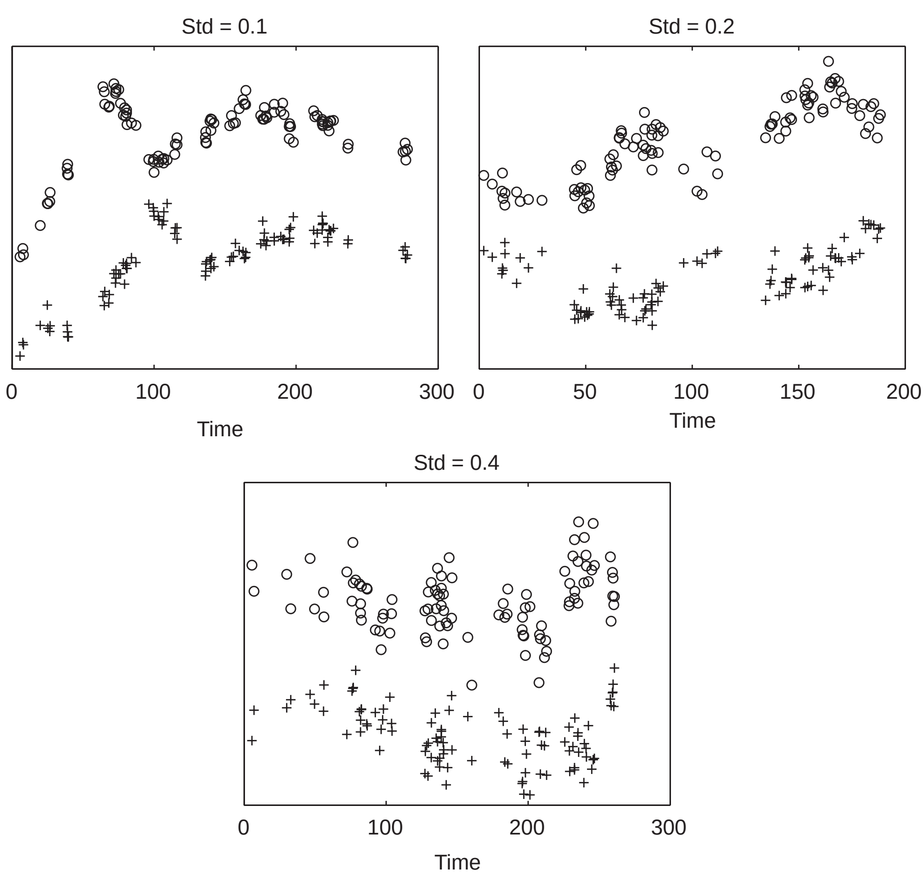

Fig. 2. An example dataset from each of the three classes.

function (LNDCF) [12], linear interpolation followed by standard cross correlation analysis (similar scheme used e.g. in [11]), and dispersion spectrum [13].

We did repeated experiments with artificial datasets generated from our model with a fixed delay of 35 units. The datasets were composed of two observed signals having 100 samples each. The sampling times t'_i were generated from the following mixture model:

$$p(\lambda) = \begin{cases} 0.2 & \text{when } \lambda = 0.1, \\ 0.8 & \text{when } \lambda = 1.0, \end{cases}$$

$$p(\Delta t|\lambda) = \mathscr{E}(\Delta t|\lambda),$$

$$t'_i = t'_{i-1} + \Delta t,$$

where $\mathscr{E}(x|\lambda)$ denotes the exponential distribution with inverse scale λ. This model was used to simulate the gaps that are typical in astronomical datasets. The source was generated using Eq. (2) with $\gamma = 2$ and $w = 2\ \ln 0.05$. The observations were generated from the distribution 1 with $a_1 = 1$, $b_1 = 0$, $a_2 = 0.8$ and $b_2 = 0.2$. The noise variances e^{v_k} were same for both observations and three different levels (resulting in three classes of datasets) of it were used, namely 0.1^2, 0.2^2 and 0.4^2. Since the source was normalised to unity variance, these noise variances directly translate to equivalent signal-to-noise ratios. No additional measurement noise was added to the data, so $\sigma^2_y(t_i) = 0$, $\forall i$. Each of the three classes contained 225 datasets, and an example from each of them is shown in Fig. 2.

The LNDCF method contains one tunable parameter: the binning size. To make the comparison as fair as possible, we optimised this by trying various bin sizes between 5 and 60. We found that the bin size of 10 produced the best results on average. With the linear interpolation approach one can vary the sampling frequency. Again we tried several possibilities between 0.01 and 1.0, but found that within this interval, it did not make much of a difference what the exact value was, so we set this parameter to 0.05. The dispersion spectra come in many varieties. We used the $D^2_{A,B}$ statistic [13] which is completely nonparametric. With all the above comparison methods, we computed the value of the appropriate statistic varying the delay between 0 and 70 with stepping of 0.2 and selected the delay giving the best value for the statistic.

The proposed approach contains several tunable parameters in the proposal distribution. But as opposed to the comparison methods, these can be selected without knowing the true delay, by monitoring the rejection ratio and adjusting the parameters appropriately. The prior on τ (see Eq. (3)) was chosen to be vague with $\mu_\tau = 0$ and $\sigma_\tau = 200$. We drew 10 000 samples from five independent chains, having sampled the initial values for the delay from the uniform distribution on [0, 70]. To reduce the possibility of including samples not coming from the posterior due to failure to converge, we tried to prune one to two of the chains by computing the potential scale reduction factor for each of the subset of chains and selecting the subset with the lowest

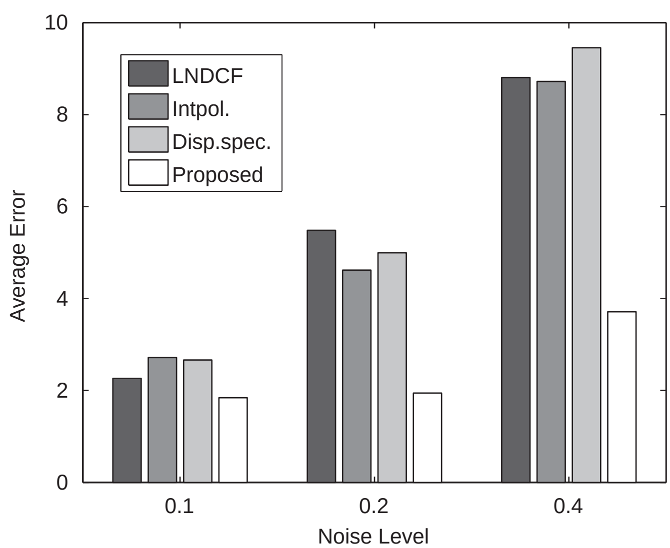

Fig. 3. Average errors of the methods.

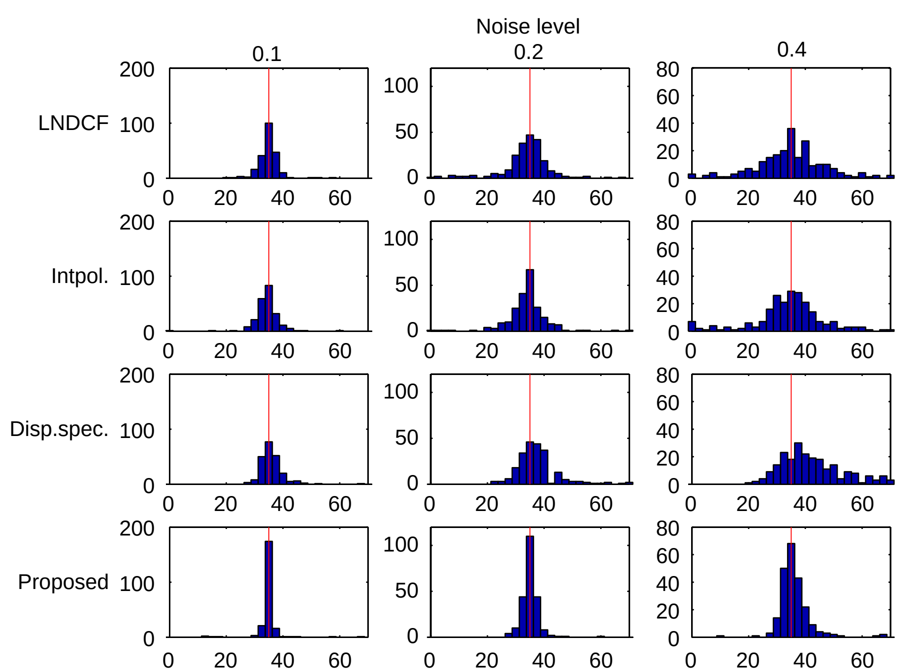

Fig. 4. The distributions of the estimates.

value. Finally, half of the samples from each of the chains were discarded as the burn-in period, and the rest of the samples were combined. To be able to compare against the other methods, we collapsed the posterior distribution of the delay to a point estimate by computing the sample mean.

The accuracy of the estimates was measured using the average error defined as

$$\mathrm{AE} = \frac{1}{225}\sum_{i=1}^{225} |\hat{\tau}_i - 35|,$$

where $\hat{\tau}_i$ is the estimate for dataset i. Ideally the AE would be zero, meaning an exactly correct estimate every single time. The scheme of randomly guessing a delay between 0 and 70 can be thought of as the baseline (in the experimental setting adopted, the comparison methods could produce an estimate only within this interval), in which case the average error would be (on average) 17.5.

Fig. 3 shows the accuracies of the methods. All the comparison methods seem to perform at a roughly similar level. The proposed approach is better on average with all noise levels and the difference becomes more prominent as the noise condition gets worse.

To give a fuller view to the results, we show the distribution of the estimates in Fig. 4. If one examines the histograms of the proposed method carefully, one can discover that there are a few small bars far from the bulk of the probability mass. These are to some extent comprised by estimates gone awry by the sampler failing to converge. When analysing any given dataset, the convergence can easily be assessed and the sampler rerun with possibly different parameters to obtain samples from the equilibrium (or sufficiently near it). Here, such fine tuning was not performed.

4.2. Estimating the delays in gravitational lensing systems

Gravitational lensing occurs when the light coming from a distant quasar is bent by the gravitational potential of an intervening galaxy, such that several images of the source are observed. Relativistic effects, as well as the different lengths of the light travel paths, affect the time it takes for the photons originating from the source to travel to the observer. If the source varies in brightness over time, a delay is perceived between the fluctuations in the intensity profiles of the images. The significance of estimating the delays in such systems stems from the early observation that they can be used in determining important cosmological quantities, since the delay provides a rare method of directly measuring distances over very large scales [14].

We have determined the delays of several lensing systems, and compared the results with earlier estimates, obtained with methods typically used by astrophysicists, on the same or similar data. This whole study is of general interest to the astronomical

Table 1
Our estimates of time delays (mean ± std) compared to previous results

System	Image	Our delay	Previous measures	Ref.
B0218 + 357		10.9 ± 0.7	10.5 ± 0.4	[2]
PG1115 + 080	AC	-11.7 ± 1.7	-13 ± 1	[1]
			-9.4 ± 3.4	[15]
	BC	-22.7 ± 1.8	-25 ± 1	[1]
			-23.7 ± 3.4	[15]

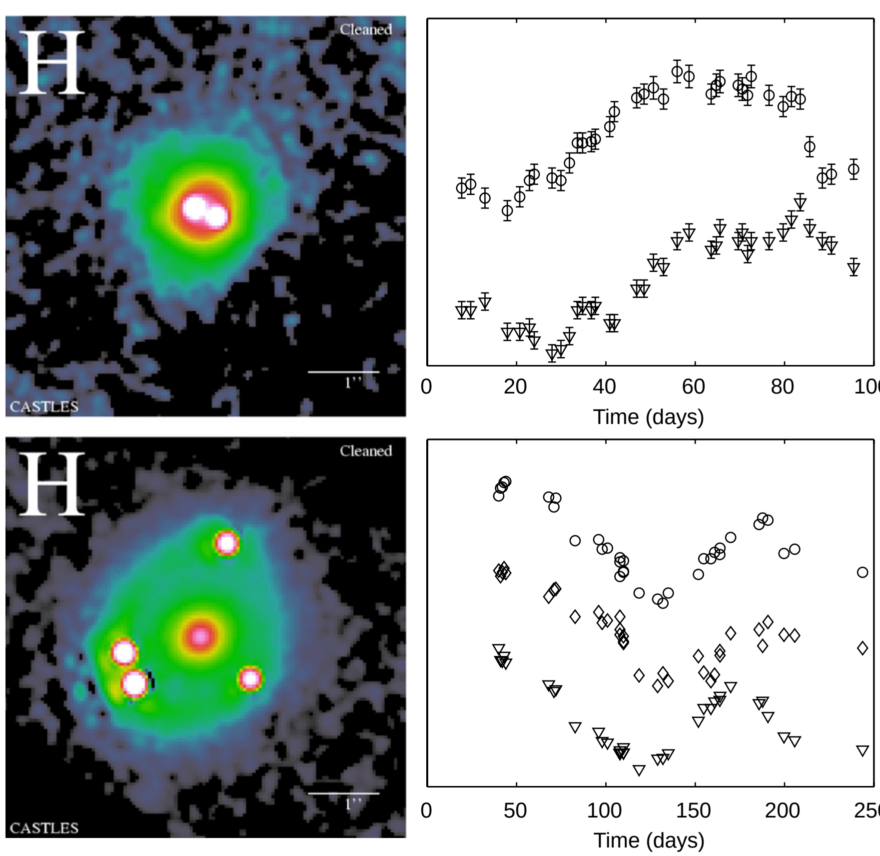

Fig. 5. The images and the corresponding intensity measurements for B0218 (top) and PG1115 (bottom). With PG1115, the two images closest to each other are merged so that there are only three distinct time series of intensity measurements. The images were obtained from CASTLES [10].

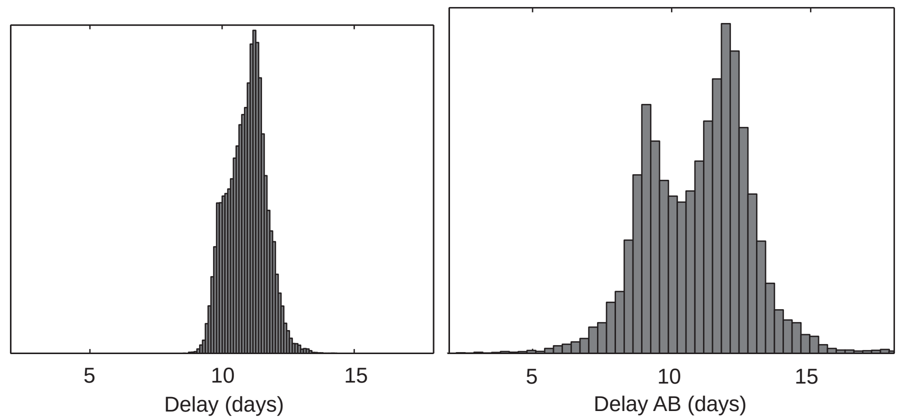

Fig. 6. Left: the posterior over the delay in B0218. Right: same for the delay AB in PG1115.

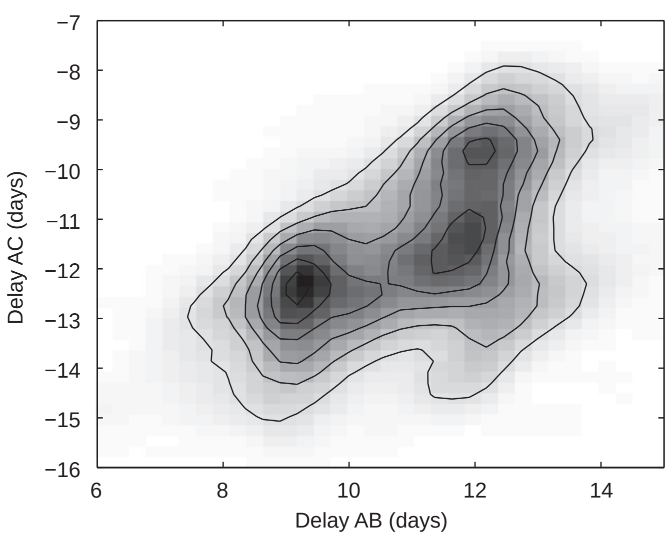

Fig. 7. The posterior distribution over the two delays in the lensing system PG1115. The time delays between the pairs of the three images ABC are related by $\tau_{AB} + \tau_{BC} = \tau_{AC}$.

community and will be detailed elsewhere. Since the true delays are and will remain unknown, we cannot make well-justified claims of having measured them better or worse than our predecessors. Instead, here we will point out the benefits of our kind of analysis, in the sense that, in addition to being more accurate, it equips us with full probability distributions over the quantities of interest. This is important knowledge since the uncertainties in the delays will translate to uncertainties in the derived quantities.

We shall concentrate on two systems which lie at both extremes on the well-determined/ill-determined axis. These are the systems B0218 + 357 (with two distinct images AB of the same source, hereafter B0218) and PG1115+080 (with at least three distinct images ABC, hereafter PG1115). Images of the two systems, along with their intensity profiles ("light curves") are shown in Fig. 5.

Table 1 summarises our results and previous measures of the delays. In the case of B0218, we find a delay that matches with the estimate in the previous work quite well. The situation is somewhat different with PG1115, which has arguably a far more complicated light curve. Earlier measures of the time delays between pairs of the three major images had produced contradictory results in the literature, as can be noted in Table 1. For the first time, our analysis, which produces a posterior distribution, shows why.

The posterior over the delay in B0218, shown in Fig. 6, is well behaved, having the probability mass tightly concentrated around the mean value. This is to be contrasted to the case of PG1115, where the posterior of one of the delays is badly multimodal spanning a wide range of values from 5 to 15. The uncertainty is even more visible in the joint distribution of the two delays shown in Fig. 7. It has at least two strong modes. In this light, the posterior average and standard deviation, that we customarily report in the table, are not that sensible statistics. But since the two strong modes have almost equal probability mass, we do not feel comfortable in computing the subsequent quantities based on just one of them either. Rather, we feel that the indeterminacy pointed out by this analysis should be taken as a hint that the system PG1115 (or rather the so far observed data from it) is not the best candidate for the computation of the derived quantities.

5. Conclusions

The estimation of a delay between unevenly sampled signals is a recurring problem in certain fields. It is also a drastically different problem compared to its evenly sampled counterpart, since the standard cross-correlation methods are not well justified and often produce questionable results. In this paper we proposed a Bayesian approach to solve the problem, derived the learning procedure and showed its effectiveness over several alternative approaches.

Appendix A. Derivation of the marginal likelihood

The derivation will rely on the following Kalman-filter model:

$$
\begin{aligned}
&p(h_1) = \mathcal{N}(h_1|0, \sigma_h^2),\\
&p(h_i|h_{i-1}) = \mathcal{N}(h_i|h_{i-1}, v_{hi}), \quad i = 2, \ldots, T,\\
&p(z_i|h_i) = \mathcal{N}(z_i|h_i, v_{zi}). \qquad \text{(A.1)}
\end{aligned}
$$

Above, h is the hidden variable and z the observed variable. The innovation- and noise variances, v_h and v_z, respectively, are time dependent but fixed. The marginal likelihood of this model, $p(\{z_i\}) = \int d\{h_i\}\, p(\{z_i\}, \{h_i\})$, is obtained by the following filtering equations:

$$\tilde{h}_i = \begin{cases} (\sigma_h^{-2} + v_{zi}^{-1})^{-1}, & i = 1, \\ [(v_{hi} + \tilde{h}_{i-1})^{-1} + v_{zi}^{-1}]^{-1}, & i > 1, \end{cases}$$

$$\hat{h}_i = \begin{cases} \tilde{h}_i z_i / v_{zi}, & i = 1, \\ \tilde{h}_i [\hat{h}_{i-1}/(v_{hi} + \tilde{h}_{i-1}) + z_i/v_{zi}], & i > 1. \end{cases} \tag{A.2}$$

From these, the marginal likelihood is computed as

$$p(\{z_i\}) = \prod_{i=2}^{T} \mathcal{N}(z_i | \hat{h}_{i-1}, v_{hi} + \tilde{h}_{i-1} + v_{zi}) \mathcal{N}(z_1 | 0, \sigma_h^2 + v_{z1}). \tag{A.3}$$

Now we need to identify the model of this paper with the above Kalman-filter model. The first step is to integrate out the error model:

$$\begin{aligned} p(y(t_i)|s(t_i), \tau, \boldsymbol{\theta}) &= \int \mathrm{d}x(t_i) p(y(t_i)|x(t_i)) p(x(t_i)|s(t_i), \tau, \boldsymbol{\theta}) \\ &= \int \mathrm{d}x(t_i) \mathcal{N}(y(t_i)|x(t_i), \sigma_y^2(t_i)) \mathcal{N}(x(t_i)|a_{k(i)} s(t_i) \\ &\quad + b_{k(i)}, e^{v_{k(i)}}) \\ &= \frac{1}{|a_{k(i)}|} \mathcal{N}(\hat{y}(t_i)|s(t_i), \tilde{y}(t_i)), \end{aligned}$$

where

$$\hat{y}(t_i) = \frac{y(t_i) - b_{k(i)}}{a_{k(i)}} \quad \text{and} \quad \tilde{y}(t_i) = \frac{e^{v_{k(i)}} + \sigma_y^2(t_i)}{a_{k(i)}^2}.$$

Now the model can be written as

$$\begin{aligned} & p(s(t_1)|\boldsymbol{\theta}, \tau) = \mathcal{N}(s(t_1)|0, \sigma_s^2), \\ & p(s(t_i)|s(t_{i-1}), \boldsymbol{\theta}, \tau) = \mathcal{N}(s(t_i)|s(t_{i-1}), (t_i - t_{i-1})^{\gamma} e^{w}), \\ & p(y(t_i)|s(t_i), \boldsymbol{\theta}, \tau) = \frac{1}{|a_{k(i)}|} \mathcal{N}(\hat{y}(t_i)|s(t_i), \tilde{y}(t_i)). \end{aligned} \tag{A.4}$$

Apart from the coefficients $1/|a_{k(i)}|$, the above equations are exactly as in the Kalman-filter model (A.1). Since the coefficients are independent of the variables to be integrated over ($s(t_i)$ that is), they pop out from the integral and end up as the factor $\prod_{i=1}^{2N} |a_{k(i)}|^{-1}$ in the final marginal likelihood. Now it is simply a task to find the correspondence between the variables in (A.1) and (A.4), then make the appropriate substitutions in (A.2) and (A.3), and finally multiply by the above mentioned factor. What follows, is exactly (5)–(7).

References

[1] R. Barkana, Analysis of time delays in the gravitational lens PG 1115 + 080, Astrophys. J. 489 (1997) 21–28.
[2] A.D. Biggs, I.W.A. Browne, et al., Time delay for the gravitational lens system B0218 + 357, Mon. Not. R. Astron. Soc. 304 (1999) 349–358.
[3] S.P. Brooks, A. Gelman, General methods for monitoring convergence of iterative simulations, J. Comput. Graphical Stat. 7 (4) (1998) 434–455.
[4] J.C. Cuevas-Tello, P. Tiňo, S. Raychaudhury, How accurate are the time delay estimates in gravitational lensing? Astron. Astrophys. 454 (3) (2006) 695–706.
[5] R.A. Edelson, J.H. Krolik, The discrete correlation function: a new method for analysing unevenly sampled variability data, Astrophys. J. 333 (1988) 646–659.
[6] A.E. Gelfand, A.F.M. Smith, Sampling-based approaches to calculating marginal densities, J. Am. Stat. Assoc. 85 (410) (1990) 398–409.
[7] M. Harva, S. Raychaudhury, Bayesian estimation of time delays between unevenly sampled signals, in: Proceedings of the International Workshop on Machine Learning for Signal Processing (MLSP'06), Maynooth, Ireland, 2006, pp. 111–116.
[8] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.
[9] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, Cambridge, UK, 2003.
[10] C.S. Kochanek, E.E. Falco, et al., Website of CfA-Arizona space telescope lens survey ⟨http://www.cfa.harvard.edu/castles/⟩.
[11] T. Kundic, W.N. Colley, et al., A robust determination of the time delay in 0957 + 561A, B and a measurement of the global value of Hubble's constant, Astrophys. J. 482 (1997) 75–82.
[12] J. Lehar, J.N. Hewitt, B.F. Burke, D.H. Roberts, The radio time delay in the double quasar 0957 + 561, Astrophys. J. 384 (1992) 453–466.
[13] J. Pelt, R. Kayser, S. Refsdal, T. Schramm, Time delay controversy on QSO 0957 + 561 not yet decided, Astron. Astrophys. 286 (1) (1994) 775–785.
[14] S. Refsdal, On the possibility of determining Hubble's parameter and the masses of galaxies from the gravitational lens effect, Mon. Not. R. Astron. Soc. 128 (1964) 307–310.
[15] P.L. Schechter, The quadruple gravitational lens PG 1115 + 080: time delays and models, Astrophys. J. Lett. 475 (1997) L85–L88.

Markus Harva received the M.Sc. degree in Computer Science and Engineering from the Helsinki University of Technology in 2004. He has been with the Adaptive Informatics Research Centre since 2002. Currently he is pursuing the Ph.D. degree working on approximate methods for Bayesian inference.

Somak Raychaudhury studied Physics at the University of Calcutta, India, and the University of Oxford, UK, and obtained his Ph.D. in Astrophysics from University of Cambridge, UK. Following positions at the Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA, and the Inter-University Centre for Astronomy and Astrophysics, Pune, India. He is currently a Senior Lecturer at the School of Physics and Astronomy, University of Birmingham, UK.