A Different Approach to Lunar Data Analysis:
The Application of Bayesian Statistics

Judit G. Ries, William H. Jefferys, and Peter J. Shelus
McDonald Observatory and Department of Astronomy
The University of Texas at Austin

In Lunar Laser Ranging, most of the photons detected are due to noise. Unfavorable meteorological conditions can further deteriorate the return signal, and some data cannot be extracted by the present method of analysis. Bayesian analysis can help us recover some of that data. Recovering the width of the returning laser pulse might provide information to improve the error estimate of the normal point. Earth orientation determination could be done simultaneously with data filtering, providing almost instantaneous Earth orientation parameters values.

In our investigation, the likelihood function of a generic lunar experiment has been derived. Instead of the marginalizing the distribution by numerically integrating for the possible range of parameters, the marginal distributions are obtained using a Markov chain Monte Carlo method. The method has been tested on poor quality simulated lunar data and demonstrated good recovery of the unknown parameters. Testing of the method on actual lunar data is underway.

Introduction

The only experiment of the NASA Apollo lunar missions still in operation is the Lunar Laser Ranging (LLR) experiment. The arrays of reflecting cornercubes that the Apollo astronauts left on the Moon, along with two other arrays delivered by Soviet spacecraft, do not require power, and their surfaces have not shown measurable degradation since they were deployed. Improvements in laser technology and timing devices have increased the accuracy of the range measurements since 1969, but LLR remains a technically and scientifically challenging measurement. The wide range of scientific results is summarized in [1].

LLR is the measurement of the round-trip travel time of a photon emitted from an Earth-based laser. Changes in travel time, that indicate changes in the separation between the transmitter and the reflector, contain a great deal of information about the Earth-Moon system which can be retrieved by estimating model parameters. An important signal is the error in predicted Earth Rotation Parameters (EOP's). One motivation for better identification of photons returning from the moon is to better determine these EOP's. However, the quality and quantity of lunar data depends strongly on atmospheric conditions and on the lunar phase. High humidity, atmospheric turbulence and high background noise can make the detection of lunar returns quite difficult. An additional incentive is to narrow the data gap in the around new and full Moon. By reducing the aliasing of observational effects we could also improve our estimate of relevant relativity parameters.

Data acquisition and the present filtering method

The basic elements of an LLR station are a laser, timing equipment, a telescope, and a computer. There is a substantial loss of signal due to transmission through the telescope and the atmosphere, laser beam divergence, the distance to the retroreflector, and array size. When the laser beam reaches the Moon, it is spread over an area of about 7 km in diameter. This results in a factor of 10^{21} decrease in the signal, making LLR a single photon detection experiment.

An essential requirement in collecting LLR data is an adequate model of lunar dynamics, atmospheric refraction, station coordinates and Earth orientation. Based on this model, the telescope can be accurately pointed to the retroreflector on the Moon. During observing, depending on the lunar phase (i.e., on the illumination of the lunar disk), and on the atmospheric conditions, ten to a thousand photons reach the detector. When the reflectors are in the dark, most of the detected photons are laser returns, but at other times, background photons can overpower the lunar returns because the retroreflector is illuminated by the Sun. The individual returns are later analyzed and compressed into normal points, presently of 1 cm precision. Data accuracy is now at about the 2-3 cm level.

In addition to the hardware spectral and temporal filters the data are processed through a software filter. The difference between the predicted and measured round-trip times, that is, the residuals, are calculated using the adopted model. For a perfect measurement and model the residuals should be zero. Different model errors introduce different signatures into the residuals, but the noise photons do not follow any pattern. We assume that the mathematical model we are using is correct except for errors in the predicted EOP's. During a run this error will cause the residuals of the returning photons to lie on a straight line of small but unknown slope. The background photon residuals are randomly distributed. Assuming that the background noise is uniformly distributed, to identify the lunar return the analyst looks for clumping, that is for significant deviation from a uniform distribution [2]. In practice, the residuals are binned (usually into 1 nanosecond bins) and the number of photons expected in a bin is calculated from the total number of detections. The software examines the bins in pairs, and looks for a significant deviation from this expected number (the slope and the width of the bins can be adjusted during this process). When it finds such pairs, all photons in the two bins are identified as lunar returns and compressed into normal points. The EOP are recovered through another step, using nightly corrections to the apriori EOP based on the normal points.

This approach breaks down if the number of the total returns is small, or if the noise level is high (Figure 1); in such cases the program cannot decide, based only on the maximum expected number of returns, whether the detected photons are from the retroreflector or not. In this case no photons are identified, and data can be lost. The returning laser pulse is wider than the outgoing pulse. If we can estimate ilaser pulse width, additional information could be obtained about the precision of the normal point. At present this is not taken into account, except that the bins are chosen to be wide enough to contain the smeared pulse. Since the EOP determination can be improved by increasing the amount of reliable data going into the calculations, and by recovering the parameters close to real time, the Bayesian approach can help in two ways: It could recover data deemed unusable before, and it eliminates the extra step of forming the normal points for EOP determination.

Bayesian analysis

Bayesian analysis requires the user to provide two things: A prior distribution, p(model), which is a function of all of the model parameters in the problem, and a likelihood function *L*(model; data) which describes, in a probabilistic sense, what sort of data we expect to see, given any particular choice of model parameters [3]. In our case we are interested in estimating the distance D to the Moon. The prior distribution represents our knowledge (and/or opinions) about the parameters, prior to taking some data set. It is reasonable to think that, from all the possible parameter values, some are more likely than others. The prior distribution is a measure of our own ignorance: if we are very sure of the value of the parameter, it will be sharply peaked, and if we are less sure, it will be spread out. The prior distribution can vary from user to user for various reasons, including the fact that different individuals have different prior information. Under many circumstances the result of a Bayesian analysis may not depend critically on the choice of prior, within a fairly wide class of priors. In other cases, care must be taken. If the prior is uncertain in such a way that the results do depend sensitively on the prior, one should carefully investigate the way the results depend on the prior.

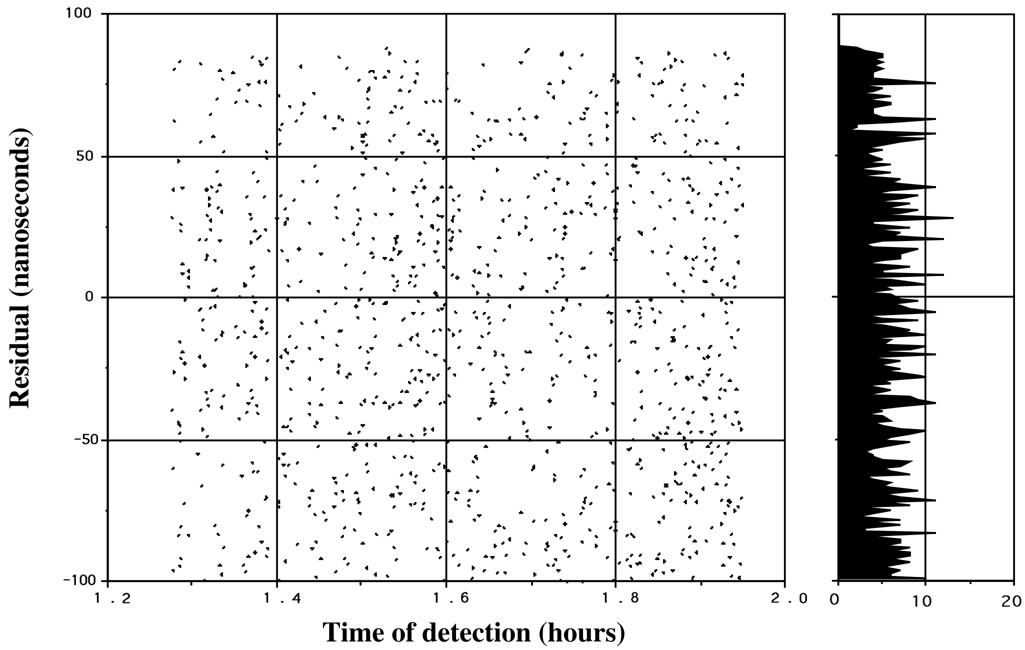

Figure 1. Poor data: The data are marginal, and the clumping is not obvious. The traditional filtering method breaks down. However, the ranging crew, with years of experience at lunar laser ranging, indicated that they think they obtained real laser returns in this run.

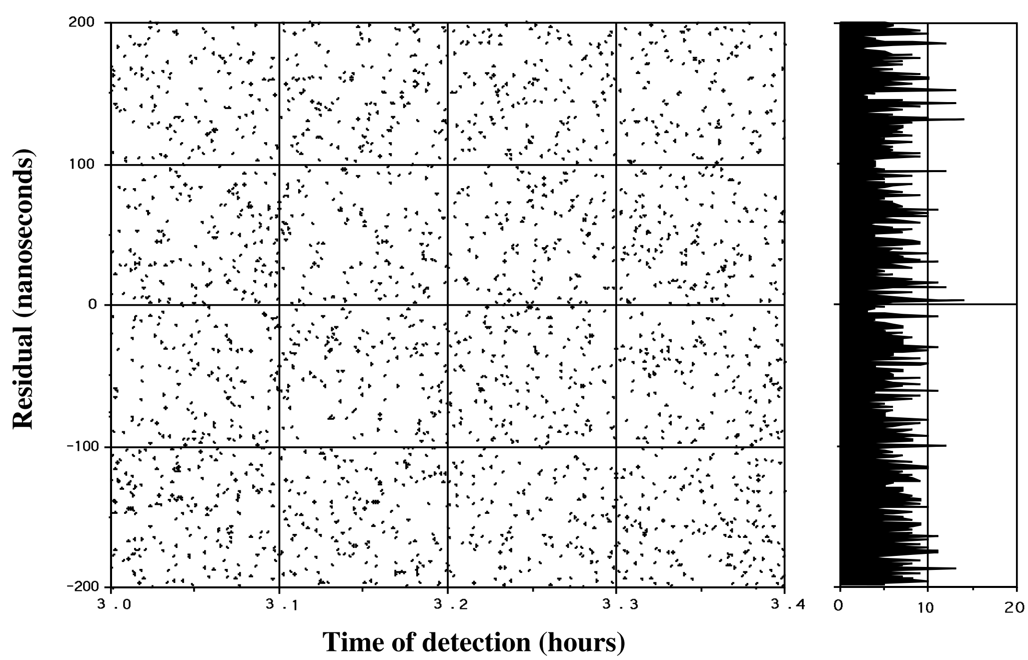

Figure 2. Simulated poor data: Simulated returns are plotted as a function of time. The 8 actual returns can barely be seen on the histogram and are difficult or impossible to see on the scatter diagram.

The likelihood function L(model; data) is any function that is proportional to the probability of obtaining certain data, given a model; but it is considered as a function of the model (i.e., the model parameters), since the data we obtain are fixed. Frequently the particular model is specified by a particular parameter set. For example, each possible value of D corresponds to a particular model, and the probability of obtaining a particular set of data (return timings, in the LLR case), depends upon the value of D. Data that are consistent with a high value of D are more likely to be obtained if the value of D is high than if it is low, and vice versa.

Expressed in the language of conditional probability,

$$L(\text{model; data}) \propto p(\text{data}|\text{model}) \tag{1}$$

The Bayesian prescription tells us that, by Bayes' theorem, the posterior distribution of the parameter given the data, p(model | data), is proportional to the prior distribution times the likelihood, with a normalization factor that is just the reciprocal of this product integrated over all models (i.e., sets of parameters). Thus

$$p(\text{model/data}) \propto L(\text{model; data}) \text{ x } p(\text{model}) \tag{2}$$

with proportionality factor C given by

$$C^{-1} = \int_{\text{all models}} L\ (\text{model;data}) \text{ x } p(\text{model}) \tag{3}$$

A key characteristic of the Bayesian approach is that the results are conditional on the data that have been actually observed. That is, the Bayesian analysis does not consider data that might have been observed but were not. This is displayed by the conditional nature of the posterior probability distribution. All results of interest can be derived from the posterior distribution.

By integrating out (marginalizing with respect to) parameters that are not of interest, we can obtain a posterior distribution that is a function of only those parameters that we are interested in. Such calculations are often the most difficult part of a Bayesian analysis. In many cases it is relatively staightforward to write down the likelihood function and even to construct the prior. But the integration of the posterior distribution may be difficult because it may be complex and not integrable in closed form. To handle this problem, a number of Monte Carlo methods have been developed in recent years that have proved rather effective. We will follow this strategy in our discussion after deriving the likehood function.

The Likelihood Function

Because we are counting discrete events that are, for all practical purposes, independent, Poisson statistics are applicable. Often it is possible, in large-signal situations, to approximate Poisson statistics by an appropriate normal approximation, but that route is not available here. We must deal with the Poisson nature of the data at the outset.

We approach the problem of writing down the likelihood by observing that, in a very short interval of time Δt, the probability that we detect no photons is
and the probability that we detect a single photon is

$$(r\Delta t)^1 \exp(-r\Delta t)\ \frac{1}{1!} = (r\Delta t)\exp(-r\Delta t) \tag{4}$$

where r is the expected rate per unit time of a detection. By making Δt very small, we can ignore the probability of two detections in the interval. In our case, the rate varies with time, $r = r(t)$, because the probability of the arrival of a photon is significantly enhanced during the very short interval that the range coincides with the actual light-time to the moon.

If the intervals Δt are disjoint, then the detected events are independent and the likelihood function is just the product of (4) and (5) over all intervals:

$$L(\text{model};\text{data}) \propto \exp\left(-\sum r(t)\Delta t\right)\left[\prod_{i=N} r(t_i)\right](\Delta t)^N \tag{6}$$

where the data t_i are the times when a photon was detected and N is the total number of detections. Since we only need the likelihood function up to a constant factor, we drop the last term $(\Delta t)^N$ and then take the limit as $\Delta t \to 0$ to obtain

$$L(\text{model };\text{data}) \propto \exp\left(- \int_{\text{all time gate open}} r(t)dt\right)\prod_{i=N} r(t_i) \tag{7}$$

where the integral is taken over the entire time that the range gate is open (or, equivalently, over the entire duration of the observation set, noting that the probability of detection $r = 0$ when the gate is closed). The same result is obtained when we analyze the problem using the approach suggested in [4, §2.1].

For our problem, we presume the following form for the rate r(t):

$$r(t) = \begin{cases} 0 \ \text{ if the gate is closed} \\ r_{bg} \text{ if the gate is open but the return is not expected} \\ r_{bg} + r_s \text{ if the pulse return is expected} \end{cases}$$

Here, r_{bg} is the background detection rate per unit time and r_s is the signal detection rate per unit time. We presume that the returning laser pulse is very narrow, i.e., a few hundred picoseconds. The pulse travels to the moon and is reflected back to the Earth. We do not know the time of arrival. We have a model for the return time that is dependent on certain parameters. We predict (on the basis of our model) that at some time t the center of the pulse will arrive; that is the expected time for the return of the pulse.

By using the "box" function $\Pi(x)$, which is 1 when $-0.5 \le x \le 0.5$ and zero otherwise, we can express a simple model for r(t) as follows:

$$r(t) = \sum_i \prod((t - t_{gi})/a_i)(r_{bg} + r_s \prod((t - t_{ri})/a_{pw})) \tag{8}$$

where in the ith shot, a_i is the length of time the range gate is open, t_{gi} is the mean of the opening and closing times of the range gate, a_{pw} is the pulse width, and t_{ri} is the predicted time of the pulse return. It is assumed that there is an unknown bias in the ephemeris of the Moon, so that there is an unknown offset in the Moon's distance; and that furthermore, the offset varies in time, linearly to first order, by an amount that is also unknown. Thus we can write

$$t_{ri} = b + c(t - t_m) \tag{9}$$

where b is the expected pulse return time at time $t = t_m$, t_m is the midpoint of the data run, and c is the slope of the unknown pulse return time. The unknown parameters b and c are to be determined. The expression for the likelihood function can be simplified by carrying out the integration and product. We see that

$$\int r(t)dt = Tr_{bg} + (m - k/2)r_s\Delta t \tag{10}$$

where T is the total time that the range gate was open, Δt is the assumed width of the returning pulse, k is the number of detections that occured within $\Delta t/2$ of the expected pulse arrival time t_{ri}, and m is the number of times that the range gate was open when the pulse return was expected at time t_{ri}. The factor 1/2 in the last term takes into account the fact that the pulse is detected, on average, halfway through the pulse width, an approximation that is convenient but inessential. We adopt it for the purpose of this calculation. The product can be written up to a constant factor as

$$r_{bg}^N\left(1+\frac{r_s}{r_{bg}}\right)^k$$

where as before N is the total number of detections. The formal unknown parameters in this problem are the detection rates r_{bg} and r_s, and the parameters b and c that describe the expected time of pulse arrival. Usually one would have a good idea of the errors of the ephemeris, and can bound b and c. For this investigation we adopted a simple prior that is uniform within a range typical of what might be expected for the parameter and zero outside that range. For some runs we took a worst-case stance by assuming that the pulse could return at any time that the range gate was open and setting the width of the prior accordingly. In actuality, that is far too pessimistic, but it allowed us to find out just how well we could pin down the actual return time from the data. We restricted the slope, c, so that over the entire run the expected time of arrival would not vary by more than 10 nanoseconds. The width of the priors on b and c were variable, i.e., we allowed ourselves to be very sure or quite ignorant, depending on the run.

Gibbs sampler

Once we have determined the prior and the likelihood, we write down the posterior distribution (up to a constant factor). We consider some of the parameters, in particular r_{bg} and r_s, to be “nuisance parameters”. That is, we are not much interested in their actual values. We are most interested in the marginal distributions of b and c, that provide the desired information about the Moon's orbital motion. The marginal posterior distributions $p(b|data)$ obtain and $p(c|data)$ are obtained by integrating over all of the other parameters to from the complete posterior $p(b,c,r_{bg},r_s|data)$. We might also be interested in the marginal distributions of r_s. During the initial runs, we assumed that we knew r_{bg} and r_s Under ideal conditions this information is actually well known since the characteristics of the laser pulse and the reflection process on the Moon are well understood, but can be influenced by weather and other conditions.

Our integration method was the Metropolis subchain Gibbs sampler, as suggested in [5] and described in [4, §.5.3]. The idea is to generate a Markov chain using the posterior distribution to generate each next step in the chain. The transition probabilities at each step are prescribed by the posterior distribution, in such a way that one is more likely to make a transition from a region of lower posterior probability to one of higher posterior probability than vice versa. Thus, the Markov chain tends to spend more time in regions of high posterior probability than in regions of low posterior probability. The Markov chain is defined so that a step is taken first in b, then in c, then in r_{bg}, then in r_s. This process is one iteration of the chain. The process is then repeated indefinitely, always starting the new step where the old one left off. After each iteration, the current values of the parameters obtained are tallied separately, to build marginal distributions for each parameter. It can be shown that under reasonable conditions, the resulting Markov chain yields marginal distributions that approach the actual marginals in the limit.

Analytical sampling schemes exist only for a narrow class of distributions. However, many probabilistic schemes are derived for such sampling. We applied Müller's ideas to generate trial steps from the one-dimensional distributions at each iteration, using a Metropolis-Hastings approach. The details are described in [4] but the basic procedure is as follows. Suppose we are ready to generate the next step in a parameter, say b. From a symmetric, but otherwise arbitrary distribution $q(\Delta b)$, generate a step Δb. The new trial value of the parameter is $b^* = b + \Delta b$. We accept or reject this trial value probabilistically based on a simple function of the posterior distribution at the two points b and b^*. In particular, with probability

$$\alpha(b^*, b) = \min\left[\frac{p(b^*, c, r_{bg} r_s \mid \text{data})}{\text{p}(\text{b}, c, r_{bg} r_s \mid \text{data})}, 1\right] \tag{11}$$

accept b^* as the new step; otherwise keep b. Then proceed in the same way with c, r_{bg} and r_s to complete the iteration.
A key advantage of this method is that it is unnecessary to work with the normalized posterior probability, since the normalization factor cancels out of the expression for α.

We chose $q(\Delta b)$ to be uniform over an interval that was typically 10% of the allowed variation in the particular parameter in question (i.e., the range of the prior for that parameter). This is somewhat crude, and other choices need to be investigated, but it is remarkable how effective this choice was. We repeated the sample/accept/reject procedure for the second parameter, then the third, and so on, until all parameters had been sampled once. At this point the current position of the point in parameter space is noted, and the whole process is repeated. Typically we would iterate on the order of 10,000 times to allow the calculation to "burn in", and then start recording the data for another 20,000 iterates. Finally, histograms of the later iterates' marginal distributions were tallied and plotted.

Results

The data used were a set of simulated returns [6]. The data consisted of 14,400 shots, of which 2313 generated detections. Most of those detections were noise; only 8 were actual returns, or 0.34%. This particular set of data was intended to represent data of poor quality. Figure 2 shows the simulated data.

The slope, c, generated by the simulator was +1.4; we used various starting values to see if the Gibbs sampler would find the actual slope. Also, we sometimes started the procedure well away from the actual bias to see whether we would converge on the actual bias (which for these data was b = +2.35 nanoseconds). We would typically start at positions like ± 10 nanoseconds from the true value; the range gate width was 400 nanoseconds for these data). Our biggest question was whether we would be able to determine b and c sufficiently well to tell when the returns came. The likelihood function can be expected to have a very narrow peak near the return. Would the Metropolis subchain Gibbs sampler find it? We tried a number of runs, first assuming that we knew the detection probabilities and, later, their marginals as well. The results were quite gratifying, as can be seen in the figures for one of the later runs.

In the run shown in Figures 1 and 2, we adopted a normal prior for b with mean 0 and standard deviation 15 nanosec, cut off at ± 30 nanosec and started at b = 0 nanosec. This prior represents the typical state of knowledge for actual runs. The prior on c was still relatively pessimistic: uniform for -10 £ c £ 10 and 0 outside that interval. The results were that the median of the smallest posterior distribution of b was +2.27 nanosec (true value +2.35 nanosec) with the smallest 80% Bayesian confidence interval (+2.25; 2.35) nanosec. The smallest 95% Bayesian confidence interval was (+1.89; 2:44) nanosec. Thus, the tails of the distribution are quite heavy relative to a normal distribution, and the center is very strongly peaked near the true value. This run is quite typical, and it shows that, despite the fact that the posterior probability is strongly peaked, the Metropolis subchain Gibbs sampler was able to handle it well and find the peak with no apparent difficulty.

The posterior distributions of the slope and of the rates are roughly normal, as expected, and not very interesting, so they will not be discussed here.

Figure 4 shows the 8 actual lunar returns, together with the median line from the Gibbs sampler calculation. The vertical scale has been blown up substantially; the error bars in the vertical scale are ± 200 picosec. The fit is very satisfactory. Indeed, we have been delighted with the way the Gibbs sampler homes in on the input answer even when the signal is so small that one can't pick it out by eye. This is very promising for practical application.

Conclusions

This paper describes a demonstration that a Bayesian approach can be used to analyze the results of LLR experiments. With poor data and somewhat pessimistic priors, one clearly and unambiguously picks out the return signal, even though over 99% of the photons detected were noise. So far we have merely scratched the surface of this project. Real data have not yet been considered and must be analyzed. It will be particularly interesting to see if and how well this method can recognize signal where the previous method cannot. We also need to investigate methods of improving the Gibbs sampler, e.g., convergence criteria and methods of deciding ideal step sizes for trial steps.

References

[1] J.O. Dickey, P.L. Bender, J.E. Faller, X.X. Newhall, R.L. Ricklefs, J.G. Ries, P.J. Shelus, A.L. Whipple, J.R. Wiant, J.G. Williams, and C.F. Yoder. Lunar laser ranging: A continuing legacy of the Apollo program. Science, 265:482-490, 1994.

[2] R. L. Ricklefs and P.J. Shelus. Poisson filtering of laser ranging data. In Proceeding of the Eighth International Workshop on Laser Ranging Instrumentation. Annnapolis: NASA conference Publication 3214, 1992.

[3] T.J. Loredo. From Laplace to Supernova 1987a: Bayesian inference in astrophysics. In P. Fogère, editor, Maximum Entropy and Bayesian Methods, pages 81-142. Kluwer Academic Publishers, Dordrecht, 1990.

[4] Martin A. Tanner. Tools for Statistical Inference. Springer- Verlag, New York, 1993.

[5] P. Müller. A generic approach to posterior integration and bayesian sampling. Technical report 91-09, statistics depart- ment, Purdue University, 1991.

[6] J.G. Ries and W.H. Jefferys. Application of bayesian statistics to lunar data analysis. Bulletin of the American Astronomical Society, 27(3):1200, 1995.

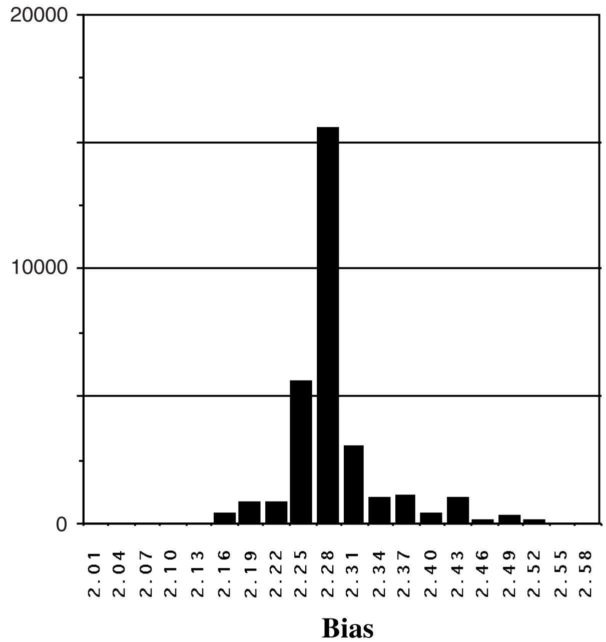

Figure 3. Posterior marginal distribution for the bias b that represents the expected return of the photons from the laser at time t_m. It is strongly peaked near the true value; 80% of the posterior probability is contained within an interval of 0.1 nanoseconds, and 95% within an interval of 0.55 nanoseconds.

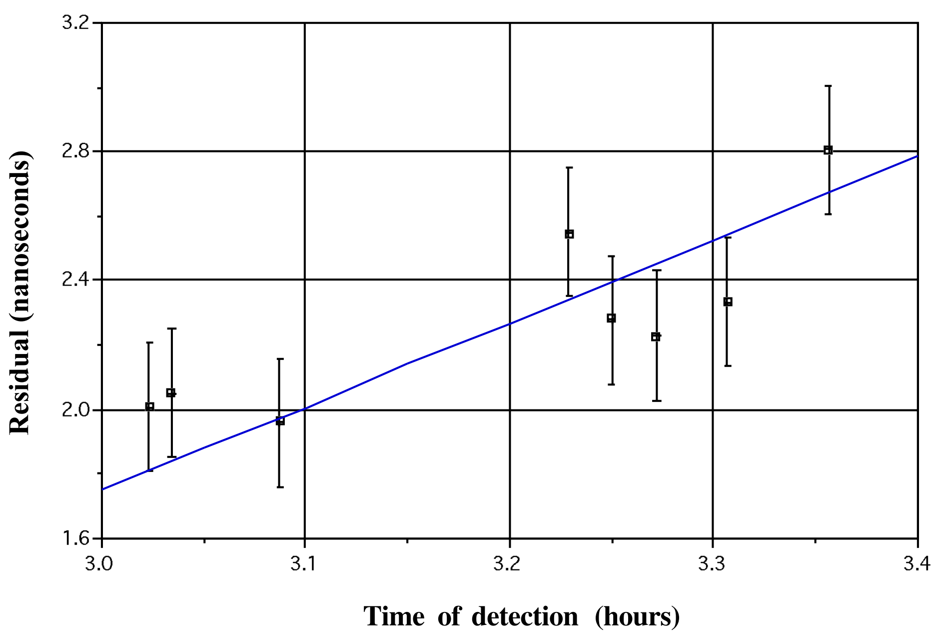

Figure 4. The eight actual returns are plotted together with the median line from the marginal distributions of b and c. Error bars indicate the uncertainty in return time.