A BAYESIAN APPROACH TO THE ANALYSIS OF TIME SYMMETRY IN LIGHT CURVES: RECONSIDERING SCORPIUS X-1 OCCULTATIONS

ALEXANDER W. BLOCKER¹, PAVLOS PROTOPAPAS^{2,3}, AND CHARLES R. ALCOCK³ ¹ Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA ² Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA ³ Initiative in Innovative Computing, Harvard University, 60 Oxford Street, Cambridge, MA 02138, USA

Received 2009 April 6; accepted 2009 June 26; published 2009 August 4

ABSTRACT

We present a new approach to the analysis of time symmetry in light curves, such as those in the X-ray at the center of the Scorpius X-1 occultation debate. Our method uses a new parameterization for such events (the bilogistic event profile) and provides a clear, physically relevant characterization of each event's key features. We also demonstrate a Markov chain Monte Carlo algorithm to carry out this analysis, including a novel independence chain configuration for the estimation of each event's location in the light curve. These tools are applied to the Scorpius X-1 light curves presented in Chang et al., providing additional evidence based on the time series that the events detected thus far are most likely not occultations by trans-Neptunian objects.

Key words: Kuiper Belt – methods: data analysis – methods: numerical – methods: statistical – stars: individual (Scorpius X-1)

1. INTRODUCTION

As new surveys accumulate vast databases of astronomical light curves (Hodapp et al. 2004; Starr et al. 2002), the identification and analysis of events in these time series are becoming the increasingly important tasks (e.g., Preston et al. 2009; Stewart 2009). One of the topics of such analyses is examining the symmetry of such events. Traditionally, symmetry has been analyzed by examining the skewness of each event, sometimes after stacking the light curves (for examples, see Jones et al. 2006, 2008). However, this approach is not satisfactory for X-ray light curves due to the small number of counts involved, particularly during suspected occultation events. Additionally, the nonparametric approach of examining skewness has significantly lower power than a parametric analysis with a well-parameterized model, especially considering the large mean-squared errors of common skewness estimators (see Joanes & Gill 1998). This means that, for a given false-positive rate, a well-designed parametric analysis will have a lower false-negative rate than a nonparametric approach.

The debate over the nature of the observed Scorpius X-1 events has raged in the literature since 2006. It began with the publication of Chang et al. (2006), which claimed the detection of at least 58 occultations of Scorpius X-1 by small (diameter < 100 m) trans-Neptunian objects (TNOs). The size distribution of these small TNOs bears important clues of the dynamical evolution of the early solar system (e.g., Pan & Sari 2005; Kenyon & Bromley 2004), yet the vast majority of the TNO population is still way beyond the limit of direct observation by even the largest ground-based telescopes. Occultation of background stars by TNOs is the only present observational method able to reach the smaller objects (Bailey 1976; Lehner et al. 2009; Roques et al. 2006; Bickerton et al. 2008; Bianco et al. 2009). This finding was challenged later that year by Jones et al. (2006). The latter group claimed that the observed events were unlikely to be occultations by TNOs because the stacked profile for these events was asymmetric. However, this analysis was not entirely satisfactory from a statistical perspective. It

was performed on shifted and stretched versions of the original data and did not take into account the discrete nature of the observations, thereby failing to use all available information from the data. The stretching we refer to is the normalization of event widths using the full widths at half-maxima (FWHMs) from Gaussian fits, which can distort results by combining observations with fundamentally different timescales. However, the most important statistical issue is that the analysis of Jones et al. (2006) provided no measures of uncertainty or significance with regard to asymmetry.

Chang et al. responded to this analysis in Chang et al. (2007), wherein they claimed that they had identified 12 events in the Scorpius X-1 data that appeared to be actual occultation events, while acknowledging that most dip events found thus far were most likely artifacts. However, this claim was challenged again by Jones et al. (2008). Jones et al. presented evidence that the observed events were due to charged particle events in the detectors. Their argument relied again, in part, on the claim that the observed events were not temporally symmetric. Unfortunately, their analysis had not proceeded beyond a simple stacking-based approach (analogous to Jones et al. 2006), so the validity of these conclusions remains uncertain.

Most recently, Liu et al. (2008) presented a new analysis of 72 ks of data of Sco-X1 taken in the year 2007. Liu et al. concluded that no significant dips, which might be real occultation by 60–100 m TNOs, were observed.

In this paper, we introduce a new parameterization for unimodal events in light curves, called the bilogistic event profile. This parameterization can describe a wide range of possible events and naturally yield estimates of physically interesting quantities such as the FWHM of the event. Additionally, our parameterization can be applied in both the Poisson and Gaussian regimes.

Additionally, we demonstrate a computational approach to inference with the bilogistic event profile using Markov chain Monte Carlo (MCMC) methods. This includes the introduction of a novel approach to the construction of a proposal distribution



Figure 1. Bilogistic event profile for c = 50, $\alpha = 40$, $k_1 = k_2 = 1$, and $\tau = 50$. Varying h_1 and h_2 from 1 to 16.



Figure 2. Bilogistic event profile for c = 50, $\alpha = 40$, $h_1 = h_2 = 8$, and $\tau = 50$. Varying k_1 and k_2 from 1 to 8.

for the center of the event, improving the efficiency of our simulations.⁴

Finally, we apply our parameterization and computational approach to the occultation events identified in the Scorpius X-1 *Rossi X-Ray Timing Explorer (RXTE)* data in Chang et al. (2006) and Chang et al. (2007). Our analysis lends further support to the argument that the events identified thus far are instrumental in origin, as suggested by Jones et al. (2006) and Jones et al. (2008).

In Section 2, we describe the details of our data model for the Scorpius X-1. We present the details of our MCMC approach in Section 3. The results of applying our method to simulated data are discussed in Section 4. Our primary scientific results on the Scorpius X-1 light curves comprise Section 5. We conclude the paper in Section 6.



Figure 3. Simulated light curves for symmetric events. For the top event, $h_1 = h_2 = 5$ and $k_1 = k_2 = 2$. For the bottom event, $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$.

2. MODEL

2.1. Context and Specification

We have a time series of counts $\{y_t\}$, $t \in \{0, ..., T-1\}$. We assume that each count is distributed as an independent Poisson random variable, conditional on the source intensity at time *t*:

$$y_t \sim \text{Poisson}(\lambda_t)$$

We also believe that an event of some kind has occurred in this time series. For the sake of this discussion, we assume that this event is characterized by a dimming of the source (e.g., a transit). However, this framework can be extended trivially to events character by increased intensity (e.g., supernovae) or the high-count regime (with a Gaussian distribution replacing the Poisson). The key constraint is that we know a priori the sign of the expected deviation from the baseline intensity. We characterize the intensity λ_t as

$$\lambda_t = c - \alpha g(t; \tau, \theta),$$

where $\lim_{t\to\infty} g(t; \tau, \theta) = \lim_{t\to-\infty} g(t; \tau, \theta) = 0$ and $\sup_{\mathbb{R}} g(t; \tau, \theta) = g(\tau; \tau, \theta) = 1$. We will call $g(t; \tau, \theta)$ the event profile, as it characterizes the pattern of intensity changes throughout the event. The parameter τ is the time index about

⁴ In Metropolis–Hastings-type MCMC methods, each step of the Markov chain consists of two stages. In the first, a new value for the variable is generated from a proposal distribution with a known density. Then the proposed value is accepted or rejected as the next value of the chain with a certain probability (calculated using the target and proposal densities). By constructing a proposal distribution that is similar to the target distribution we are attempting to draw from with our MCMC method, we are able to improve the efficiency of our algorithm.



Figure 4. Histogram of posterior distributions of *c* and α for simulated events. For the top event, $h_1 = h_2 = 5$ and $k_1 = k_2 = 2$. For the bottom event, $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$.

which the event is hypothesized to be time-symmetric, and θ is a vector of parameters characterizing the event. *c* sets the baseline intensity of the source, and α sets the peak size of the deviation from the baseline intensity during the event. All of these parameters must be estimated from the data.

Putting all of these points together with the log-likelihood for the model, we obtain

$$\ell(c, \alpha, \tau, \boldsymbol{\theta}; \mathbf{y}) = \sum_{t=0}^{T-1} [y_t \ln(\lambda_t) - \lambda_t] + \text{const.}$$
(1)

Following the Bayesian approach, we can express the posterior probability as a function of the likelihood and the prior probabilities as

$$p(c, \alpha, \tau, \theta \mid \mathbf{y}) \propto \ell(c, \alpha, \tau, \theta; \mathbf{y}) \times p(c, \alpha, \tau, \theta).$$
 (2)

Having obtained the posterior of our parameters (up to a normalizing constant), we have all the information needed to make inferences about the parameters of interest. We can characterize our inferences about the locations of parameters via their posterior means and medians. Our uncertainty about these parameters can be characterized by 68% posterior intervals or posterior standard deviations. However, it still remains to specify a functional form for the event profile $g(t; \tau, \theta)$.



Figure 5. Histogram of key symmetry-related parameters for simulated events. For the top event, $h_1 = h_2 = 5$ and $k_1 = k_2 = 2$. For the bottom event, $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$.

2.2. Bilogistic Event Profile

The choice of the event profile $g(t; \tau, \theta)$ is quite important to this analysis. One must achieve the proper balance between parsimony in parameterization and giving the model enough flexibility to fit a wide range of event shapes. To that end, we introduce the bilogistic event profile:

$$g(t; \tau, h_1, h_2, k_1, k_2) = \frac{1 + e^{\frac{-h_t}{k_t}}}{1 + e^{\frac{|t-\tau| - h_t}{k_t}}}$$
(3)

$$h_t = \begin{cases} h_1 & t < \tau \\ h_2 & t \geqslant \tau \end{cases}$$
(4)

$$k_t = \begin{cases} k_1 & t < \tau \\ k_2 & t \geqslant \tau. \end{cases}$$
(5)

This event profile is characterized by four parameters (in addition to the location parameter τ). The first two parameters $(h_1 \text{ and } h_2)$ characterize the (approximate) half-maxima of the event's amplitude represented as deviations from the event's center; we will refer to them as the "half-life" parameters. This holds only approximately because of the correction for continuity in the numerator (the $e^{\frac{-h_1}{k_1}}$ term), which ensures



Figure 6. $(k_1 - k_2)$ vs. $(h_1 - h_2)$ for simulated events. For the top event, $h_1 = h_2 = 5$ and $k_1 = k_2 = 2$. For the bottom event, $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$.

that the event profile is continuous at time τ . Thus, $(h_1 - h_2)$ characterizes the departure of our event from symmetry in terms of the half-maxima, and (h_1+h_2) characterizes the (approximate) FWHM for our event.

The second set of parameters $(k_1 \text{ and } k_2)$ characterizes the rate at which the intensity changes during our event; we will refer to them as the "curvature" parameters. The first parameter (k_1) characterizes how rapidly our source's intensity diminishes at the beginning of our event, and the second parameter (k_2) characterizes how rapidly the intensity returns to the baseline at the end of our event. If $k_1 \ll h_1$ and $k_2 \ll h_2$, as is often the case, the curvature parameters can be viewed as pivoting our event profile through the half-intensity points set by h_1 and h_2 , with higher values corresponding to sharper changes in intensity. These features of the bilogistic event profile can be seen in Figures 1 and 2, where we fix one set of parameters $(h_1$ and h_2 or k_1 and k_2) and vary the other.

To perform inference using this model, we first consider it in a Bayesian light, assuming flat priors on all parameters except τ . We place one restriction on τ a priori. We assume that the event is centered within the interior of our light curve. In this context, we take interior to mean far enough from the boundaries that the event is contained within our light curve. This assumption is reasonable for our analysis because the light curves we are working with have been extracted and preprocessed to contain one complete event each. Thus, we impose a prior probability of zero on values of τ within *m* milliseconds of the edge of the light curve for our analysis; for the analysis presented here,



Figure 7. Distributions of posterior medians of $(k_1 - k_2)$ and $(h_1 - h_2)$ for 1000 simulated events (per plot). For the top plot, $h_1 = h_2 = 5$ and $k_1 = k_2 = 2$. For the bottom plot, $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$.

we set m = 10 milliseconds. This assumption also improves the computational properties of our posterior simulations by preventing them from becoming "stuck" at values of the τ parameter that are near the ends of the light curve. The end result of these restrictions is that our prior on τ is uniform on the set [m, T - m].⁵

We can then use a Markov chain Monte Carlo (MCMC) technique (the Metropolis–Hastings algorithm) to draw from the posterior distribution of our parameters. From these draws, we can calculate quantities such as the posterior median and 68% posterior intervals for $(h_1 - h_2)$ and $(k_1 - k_2)$ to characterize the deviation of our events from symmetry and our uncertainty about these properties. For an excellent introduction to the use of the Metropolis–Hastings algorithm for statistical inference, please consult Chib & Greenberg (1995).

3. COMPUTATIONAL DETAILS OF POSTERIOR SIMULATIONS

Although the model presented above is quite appealing, the resulting posterior distribution for our parameters of interest is

⁵ However, for other analyses of time symmetry in events (such as the analysis of longer duration events that are not fully contained in a light curve), these assumptions could be relaxed.



Figure 8. Distribution of normalized sample autocorrelations for split light curves from Chang et al. (2007) with standard normal density.

not analytically tractable. Thus, to make inferences about these parameters, we simulate a series of samples from their joint posterior distribution. We use a blocked Metropolis-Hastings algorithm for these simulations. This means that, instead of drawing new values for all seven of our parameters at once and accepting or rejecting them based on the Metropolis-Hastings rule, we vary only a subset of the parameters at time, holding all others fixed. The blocks used are: (1) τ , (2) c and α , and (3) h_1 , h_2 , k_1 , and k_2 . So, for each iteration of our algorithm, we first draw a new value of τ from our proposal distribution and choose whether to accept it based on the Metropolis-Hastings acceptance probability. This acceptance probability is calculated holding the values of all other parameters (c, α, h_1 , h_2 , etc.) fixed. Next, we draw new values of c and α from our proposal distribution for these parameters and calculate an analogous acceptance probability. Finally, we perform a similar procedure on the shape parameters (h_1, h_2, k_1) , and k_2). Blocking in this way allows us to take advantage of the structure in the model in constructing our simulation algorithm, improving its efficiency. For more information on blocked Metropolis-Hastings methods, please consult Section 6.2 of Chib & Greenberg (1995).

3.1. Initialization

We look to the data to obtain starting values for our MCMC algorithm. The first step is to obtain a smoothed version of our signal y_t , which we will denote as \tilde{y}_t . This is generated by taking a centered moving average of the original signal y_t . For the analyses presented here, a window of 5 ms wide was used (for light curves sampled at 500 μ s increments), although this can be modified. The important point on the selection of the window is that it should be narrow enough to localize the event while still performing some noise reduction. If the window is too wide, our smoothed light curve will not capture the structure of the event; as a result, it would not be useful in determining initial values for the magnitude (α) or in determining the location of the event in the light curve. However, if the window is too narrow, it will not well smooth the light curve. As a result, our initializations would be far noisier than necessary.

Using \tilde{y}_t , we can construct our initialization values and a proposal distribution for τ . Focussing on initialization first, we set $c_0 = \text{median}(\tilde{y}_t)$ and $\alpha_0 = c_0 - \min_t(y_t)$. We use a robust



Figure 9. 68% posterior intervals and medians for FWHM $(h_1 + h_2)$ for Chang et al. (2007) occultation events. Second plot is zoomed.

measure of the typical baseline rate (the median). The motivation for this initialization of c is that, if the event is brief relative to the duration of the light curve and our smoothing window is not too wide, our smoothed light curve should be near the baseline count rate c for most of the times. In contrast, if we simply used the overall mean, it would be contaminated by the event of interest. Our initialization for α is relatively straightforward; we take our estimate of the baseline rate (c_0) and subtract the minimum observed count rate to obtain an estimate of the magnitude of our event.

To initialize h_1 and h_2 , we take a similar approach. First, we assume that a symmetric initialization will be an acceptable starting point for our analysis. We then calculate $h_0 = \frac{1}{2}\#\left\{t: y_t < \frac{\text{median}(\tilde{y}_t) + \min_t(y_t)}{2}\right\}$ and set h_1 and h_2 to h_0 .⁶ Based on the above discussion of c_0 and α_0 , this approach to initializing h_1 and h_2 appears quite natural. We are using our initial estimates of c and α to estimate the FWHM of our event and attributing equal parts of this time to h_1 and h_2 . Finally, we initialize k_1 and k_2 manually. Our posterior simulations were quite insensitive to the initializations of these parameters; values between 1 and 20 yielded similar results.

 $^{^{6}~}$ # is the cardinality operator, which gives the number of items in the given set.



Figure 10. 68% posterior intervals and medians for $(h_1 - h_2)$ for Chang et al. (2007) occultation events. Second plot is zoomed.

3.2. Metropolis-Hastings Steps

Keeping in mind the prior discussed in Section 2.2, we can obtain a proposal distribution for τ using the procedure outlined in Algorithm 1.

Algorithm 1 Construction of Proposal Distribution for τ Require: \tilde{y}_t for $t \in \{1, ..., T\}, m \in \mathbb{N}, p > 0, \epsilon > 0$ $\tilde{w}_t \leftarrow (\max_t (\tilde{y}_t) - \tilde{y}_t)^p$ $\tilde{w}_t \leftarrow \tilde{w}_t + \epsilon$ if t < m or (T - t) < m then $\tilde{w}_t \leftarrow 0$ end if $w_t \leftarrow \frac{\tilde{w}_t}{\sum_t \tilde{w}_t}$

Calculate points along the cumulative distribution function (CDF) as $p_t = \sum_{s \leq t} w_s$ Calculate the CDF F(t) via the linear interpolation of $\{p_t\}$ Calculate the quantile function q(p) via linear interpolation Calculate the density function f(t) as the derivative of F(t) for $t \notin \mathbb{N}$

At the conclusion of this procedure, we have the density, cumulative density, and quantile functions for our proposal



Figure 11. 68% posterior intervals and medians for $(k_1 - k_2)$ for Chang et al. (2007) occultation events. Second plot is zoomed.

distribution of τ . Through a series of numerical experiments, we found that the values of p between 2 and 5 typically produced the best proposal distributions; p was set to 4 in the subsequent analyses. The ϵ term serves to ensure that all w_t are nonzero, avoiding an issue with the interpolation of our quantile function (where a w_t of zero leads to a vertical segment). We use linear interpolation because it is the only spline method that produces identical functions when the axes are exchanged. This is vital in our application because we are conducting separate interpolations of $\{p_t\}$ to obtain our CDF and its inverse (the quantile function). If we had used any other type of polynomial interpolation, the inverse of the calculated spline for the CDF would not be another polynomial spline. However, with a linear spline, the inverse of the interpolated function is simply another linear spline. Thus, we can avoid complex calculations and use standard interpolation routines to calculate our quantile function. Our method of generating a proposal for our location parameter is quite versatile, adapting naturally to the presence of multiple potential events in the light curve. We will discuss some potential applications of it beyond this analysis in Section 6.

The proposal generated by the above procedure allows us to use an independence chain Metropolis–Hastings method for τ . With this class of methods, the proposal distribution for the



Figure 13. Light curve for event 107 from Chang et al. (2007).

parameter being simulated is independent of that parameter's current value in the MCMC simulation. If the proposal is well constructed, this can greatly improve the convergence properties of the simulations; in the ideal case, we would make our proposal the same as our posterior to maximize efficiency. However, for such a procedure to be effective, one must have some prior knowledge of the parameter's posterior distribution. This is what the above procedure extracts from the light curve.

Using the initial values and proposal distribution obtained above, we construct proposal distributions for the parameters of our model (excepting τ). All proposals are Gaussian and are thus characterized by their means and standard deviations:

$$(c, \alpha)$$
 : $\mu = (c_0, \alpha_0), \ \sigma = \left(\sqrt{\frac{c_0}{T}}, \sqrt{\alpha_0}\right)$ (6)

$$(h_1, h_2)$$
 : $\mu = (h_1^{i-1}, h_2^{i-1}), \quad \sigma = \left(\frac{\sqrt{h_0}}{2}, \frac{\sqrt{h_0}}{2}\right)$ (7)

$$(k_1, k_2)$$
 : $\mu = (k_1^{i-1}, k_2^{i-1}), \quad \sigma = \left(\frac{\sqrt{k_0}}{2}, \frac{\sqrt{k_0}}{2}\right).$ (8)

With these proposals and the blocking scheme described previously, we have fully specified our MCMC simulation approach. We now demonstrate its effectiveness with simulated data.



Figure 14. Histogram of posterior medians for $(h_1 - h_2)$ for Chang et al. (2007) occultation events.



Figure 15. Histogram of posterior medians for $(k_1 - k_2)$ for Chang et al. (2007) occultation events.

4. RESULTS ON SIMULATED SYMMETRIC EVENTS

To test our approach, we simulated two 500 timestep light curves with symmetric events resembling those observed in the Scorpius X-1 data. For the first simulation, we set c = 50, $\alpha = 40$, $h_1 = h_2 = 5$, $k_1 = k_2 = 2$, and $\tau = 300$. For the second, we narrowed the event by setting $h_1 = h_2 = 1.5$ and $k_1 = k_2 = 1$. All other parameters remained as in the first simulation. The resulting light curves can be seen in Figure 3.

The results of the posterior simulations are summarized in Figures 4–6. From the first of these (Figure 4), we observe that the posterior distributions for both the baseline count rate (*c*) and the magnitude of the event (α) are centered around their true values of 50 and 40, respectively, for the wider event. For the narrower event, the posterior distribution for α is much wider and exhibits significant skew; however, it still covers the true value. The posterior on *c* is significantly more concentrated than the posterior on α in both the cases, as we would expect; we can use nearly every point in the light curve to infer the value of *c*, but we can only use the points during the event to infer the value of α . The posterior distributions of $(h_1 - h_2)$ and $(k_1 - k_2)$ are considerably more diffuse, as Figure 5 demonstrates. These posteriors are, however, centered near the true value of 0. The



Figure 16. Events 4, 19, 41, and 66 from Chang et al. (2007).

relationship between the posterior distributions of these two quantities can be seen from Figure 6. There appears to be a significant negative dependence between the two deviations.

The properties of our estimates for symmetry-related parameters $((h_1 - h_2)$ and $(k_1 - k_2))$ are further supported by a second set of simulations. Using each of the parameter values given above, we simulated 1000 light curves and ran our algorithm on each one. This allows us to quantify some frequentist properties of our Bayesian estimators, such as the coverage of our posterior intervals.⁷ The results of these simulations are summarized in Figure 7. Both distributions appear centered around zero. Coverage properties are not perfectly calibrated (as this is a Bayesian method), but they appear reasonable. In our 1000 simulated light curves with wider events, 938 of our 68% posterior intervals for $(h_1 - h_2)$ included the true parameter value of zero, as did 816 of our 68% posterior intervals for $(k_1 - k_2)$. Similarly, in our 1000 simulated light curves with narrower events, 989 of our 68% posterior intervals for $(h_1 - h_2)$ included the true parameter value of zero, as did 808 of our 68% posterior intervals for $(k_1 - k_2)$. Thus, in both simulated cases, our method provided greater than nominal coverage.

Overall, the results of this exercise with simulated events serve to increase our confidence in the validity of our method. We now apply our model and algorithm to the analysis of the Scorpius X-1 dip events.

5. RESULTS ON SCORPIUS X-1 EVENTS

Using the above model, we analyzed all 107 events from Chang et al. (2007). Each light curve consisted of 205

⁷ The coverage of an interval estimator is defined as the percentage of intervals that will cover the true value of the parameter of interest across repeated sampling. By checking the coverage of our key estimators, we can quantify their classical error rates in addition to their Bayesian properties. For further details, Wasserman (2004) provides an excellent reference in Sections 6.3.2 and 11.9.



Figure 17. $(k_1 - k_2)$ vs. $(h_1 - h_2)$ for event 90 from Chang et al. (2007).

equally spaced observations spanning approximately 100 ms (each observation covers approximately 500 μ s). All data were assembled by Chang et al. from Rossi X-Ray Timing Explorer/Proportional Counter Array (RXTE/PCA) archival data spanning 1996 through 2002 (564 ks in all). According to Jahoda et al. (2006), measurement errors should not be significant given the data spacing and number of Proportional Counter Units (PCUs) used by Chang et al. (2007). As both of these groups also note, variation in the light curve should follow a Poisson distribution-conditional on the source intensity. For occultation events, this assumption should hold, as the count rate at time t would be reduced from $\lambda(t)$ to $(1 - p)\lambda(t)$, where $\lambda(t)$ is the unobstructed count rate, and p is the proportion of incident photons obstructed by the occulting body. However, dead-time and coincidence-event effects, which could induce autocorrelation in our data (thereby violating the assumptions of our model), remain a concern.

To check for autocorrelation in our light curves, we split each light curve into two series of 92 observations each, removing the central 21 observations. This removes the dip in each light curve, allowing us to obtain estimates of the intrinsic autocorrelation of the observations with systematic (event-driven effects). For each of the resulting 214 series, we estimate the first 10 autocorrelations. As noted in Box & Pierce (1970), if our data are actually white noise, the kth autocorrelation is distributed approximately as a normal random variable with a mean of zero and variance $\frac{n-k}{n(n+2)}$ for moderately large sample sizes. Thus, to check for autocorrelation in our data, we first normalize each estimated autocorrelation by dividing by its theoretical standard deviation. This yields 2140 observations (10 for each series), each of which should follow a standard normal distribution if our light curves exhibit no autocorrelation. Their distribution is plotted against the density function of a standard normal in Figure 8. We also conducted a Kolmogorov-Smirnov test⁸ to determine if we could reject the standard normal distribution for our normalized autocorrelations. We obtained a p-value of 0.798, indicating that there is not sufficient evidence to reject the standard normal distribution. Therefore, we conclude that our data do not exhibit significant autocorrelation.

For each event, we ran our MCMC algorithm for 400,000 iterations, discarding the first 200,000 as burn-in.⁹ To run in a



Figure 18. $|h_1 - h_2|$ vs. estimated diameter for 57 of the 58 events from Chang et al. (2006) (58th event omitted due to anomalous nature).

parallel Harvard's Odyssey cluster, these runs required approximately 400 s in total (approximately 43,000 CPU seconds). Typical acceptance rates for the individual Metropolis–Hastings steps were between 10% and 40%. The results of these simulations are summarized in Figures 9–11.

From Figure 9, we see that the FWHM of a typical event from Chang et al. (2007) is between 1 and 2 ms (with 68% posterior probability). Events 2 and 107 appear to be the only significant deviation from this pattern with FWHMs of approximately 3.5 and 6.5 ms, respectively. Visual inspection of these events (in Figures 12 and 13) reveals that both are unusual in appearance. Event 2 is characterized by a more gradual change in intensity than most and appears to have two minima. Event 107 appears to have two clear, significant minima.

The posterior intervals for $(h_1 - h_2)$ in Figure 10 appear to be relatively evenly scattered around zero. However, although most of these intervals (90 of 107) cover zero, there is notable tendency toward the negative values of $(h_1 - h_2)$. This can be seen in the posterior median, with 81 of the 107 posterior medians below zero. Using a classical sign test of the hypothesis that each posterior median is equally likely to be positive or negative (Dixon & Mood 1946), we obtain a p-value of 9.4×10^{-8} , indicating that there is a significant deviation from symmetry in the relative half-maxima for this collection of events. The sign test is conducted by first counting the number of positive posterior medians. We then calculated the probability of observing a result at least as far from an even split of positive and negative medians as was observed, assuming that each median is equally likely to fall above or below zero. In this case, the *p*-value was calculated as $2 \times P(X \leq 26|p = \frac{1}{2})$, where X is the number of observed positive medians. Under the null hypothesis that positives and negatives are equally likely, $X \sim \text{Bin}(107, \frac{1}{2})$, so the preceding probability can easily be calculated using standard techniques for the binomial distribution. The pattern of negative deviations in $(k_1 - k_2)$ can also be seen in Figure 14; the deviation of this distribution from the symmetric case (seen in Figure 7) is clear. Events 2 and

⁸ For details on the Kolmogorov–Smirnov test, see Massey (1951).

⁹ In MCMC simulations, it is common to discard the earliest portion of the results as burn-in to ensure that we are only basing our conclusions on draws from the Markov chain's stationary distribution (which is, by design, our posterior).



Figure 19. $|k_1 - k_2|$ vs. estimated diameter for 57 of the 58 events from Chang et al. (2006) (58th event omitted due to anomalous nature).

107 appear somewhat anomalous in Figure 10; this is consistent with the unusual structure of these events discussed above.

Finally, the posterior medians for $(k_1 - k_2)$ are concentrated below zero (90 of 107 are negative). Inspection of the posterior intervals in Figure 11 indicates that very little posterior support is typically found above zero (only 55 of the 107 intervals included zero). A classical sign test on the posterior medians (as above) yields a *p*-value of 3.5×10^{-13} . This indicates that the observed pattern of posterior medians is quite unlikely to occur if the true distribution of posterior medians is evenly distributed above and below zero. This is consistent with the findings of Liu et al. (2008), as a negative value of $(k_1 - k_2)$ indicates that the count rate decreased more quickly than it increased. The pattern of negative deviations in $(k_1 - k_2)$ can also be seen in Figure 15; the deviation of this distribution from the symmetric case (seen in Figure 7) is again apparent.

The concentration of $(h_1 - h_2)$ and $(k_1 - k_2)$ across the events indicates that the observed events are characterized by a relatively rapid decrease in the count rate followed by a more gradual return to the baseline level; that is, the typical event profile is right-skewed. This pattern can be seen in Figure 16, which shows four representative events from Chang et al. (2007).

It is interesting to note that $(h_1 - h_2)$ and $(k_1 - k_2)$ are typically negatively correlated (when examining their joint posterior distribution) for approximately symmetric events. An example of this appears from our analysis of event 90 from Chang et al. (2007). The joint posterior distribution of $(h_1 - h_2)$ and $(k_1 - k_2)$ can be seen (as a scatterplot) in Figure 17. Two points are immediately apparent from this figure. First, the deviation of the distribution from the origin indicates that this event is most likely asymmetric. Second, $(h_1 - h_2)$ and $(k_1 - k_2)$ are strongly negatively correlated (r = -0.618). The intuition behind this outcome is clear. We are allowing for two possible deviations from symmetry in our model: deviations based on the halfmaxima $(h_1 - h_2)$ and deviations based on the rate of change $(k_1 - k_2)$ in our source intensity. For very short events, we can not be certain to which of these deviations we should attribute the asymmetry of our event, as we can, to some extent, tradeoff between them. Our posterior distribution simply reflects this trade-off.

It is also important to note that, when analyzed using our method, the events identified as potentially noninstrumental in origin by Chang et al. do not appear to deviate significantly from the remaining events in the FWHM, half-maxima deviation, or curvature deviation. This supports the conclusions of Chang et al. (2007) and Liu et al. (2008) that the events detected thus far (including those identified as possibly noninstrumental) are most likely the result of dead-time effects and other instrumental contamination, not TNOs occulting Scorpius X-1.

At the distance of the Kuiper Belt (~ 40 AU) the size of the objects given the width of the events is ~ 50 m, and that is close to the *Fresnel scale*, which is $\sqrt{\frac{\lambda D}{2}}$, where λ is the wavelength and D is the distance. We use $\lambda = 0.3$ nm (4 keV) because most of the *RXTE*/PCA-detected photons from Sco X-1 are at this energy. As a result, these occultation events are diffraction-dominated phenomena (see Roques et al. 1987 and Nihei et al. 2007), and we expect that diffraction effects would produce small count rate increases on either side of the dips. According to Jones et al. (2006, 2008), these increases should be detectable when combining information from all events, if not for each individual event.

Also TNOs of this size are not expected to be spherically symmetric (Roques et al. 1987) and therefore the resulting occultation events in the light curves are not expected to be time symmetric. However, for TNOs of the size of the Fresnel scale or smaller the shape of the object does not translate into asymmetry of the time series. Roques et al. (1987) has studied light curves from different shaped TNOs and demonstrated that when the effective size of the TNOs is smaller than the Fresnel scale the resulting events in the light curve do not differ from the light curves signatures of regularly spherically shaped TNOs. The sizes of the 107 objects studied here are within the range of the Fresnel scale, so some asymmetry on the light curves should be manifested. Figures 18 and 19 show the measures of asymmetry as a function of the size of the object for 57 of the 58 events from Chang et al. (2006). As can be seen from the figures, there is no clear relation between the estimated size of an object and the level of asymmetry for its corresponding event. This provides further evidence that the observed dips in intensity do not correspond to TNO transit events, as Jones et al. (2008) suggest.

6. CONCLUSIONS

Based on our analysis of the suspected Scorpius X-1 occultation events, we conclude that there is significant evidence for temporal asymmetry in these events. This finding is consistent with the suggestion of Jones et al. (2008) that the observed dips in brightness are due to dead-time effects. We also conclude that the events identified in Chang et al. (2007) and Liu et al. (2008) as possibly noninstrumental in origin are most likely not occultations, as they previously found for the majority of their proposed occultation events.

We also note that the approach presented herein can be applied (with minor modifications) to the study of events in nearly any light curve. Such studies need not be limited to examining the symmetry of events, as the parameterization of our bilogistic event profile and flexible Bayesian approach allow inference to be made on quantities such as the FWHM. For example, our approach could be extended to the analysis of exoplanetary transits, supernovae, quasars or any other transient event. Furthermore, the method we have presented for constructing a proposal distribution for the center of each event can be applied quite generally. Even when using a physically motivated model for occultations with parameters such as the body size, angle of incidence, etc. (as opposed to our reduced-form approach), one must account for uncertainties about the location of the event in the light curve when making inferences about the parameters of interest. This can be quite challenging using classical methods, and even MCMC methods do not offer an panacea. Our approach to constructing an independence chain proposal distribution for the location parameter in such models makes them quite a bit more tractable, especially in the case of multiple events.

We gratefully acknowledge the support from NSF AST-0501681. We also thank the participants in the Harvard Astrostatistics Seminar for their feedback, particularly Xiao-Li Meng. The simulations in this paper were run on the Odyssey cluster supported by the Harvard FAS Research Computing Group.

REFERENCES

- Bailey, M. E. 1976, Nature, 259, 290
- Bianco, F. B., Protopapas, P., McLeod, B. A., Alcock, C. R., Holman, M. J., & Lehner, M. J. 2009, AJ, 138, 568
- Bickerton, S. J., Kavelaars, J. J., & Welch, D. L. 2008, AJ, 135, 1039

- Box, G. E. P., & Pierce, D. A. 1970, J. Am. Stat. Assoc., 65, 1509
- Chang, H., King, S., Liang, J., Wu, P., Lin, L., & Chiu, J. 2006, Nature, 442, 660
- Chang, H.-K., Liang, J.-S., Liu, C.-Y., & King, S.-K. 2007, MNRAS, 378, 1287
- Chib, S., & Greenberg, E. 1995, Am. Stat., 49, 327
- Dixon, W. J., & Mood, A. M. 1946, J. Am. Stat. Assoc., 41, 557
- Hodapp, K. W., et al. 2004, Astron. Nachr., 325, 636
- Jahoda, K., Markwardt, C. B., Radeva, Y., Rots, A. H., Stark, M. J., Swank, J. H., Strohmayer, T. E., & Zhang, W. 2006, ApJS, 163, 401
- Joanes, D., & Gill, C. 1998, J. R. Stat. Soc. D: Stat., 47, 183
- Jones, T. A., Levine, A. M., Morgan, E. H., & Rappaport, S. 2006, Astron. Telegr., 949, 1
- Jones, T. A., Levine, A. M., Morgan, E. H., & Rappaport, S. 2008, ApJ, 677, 1241
- Kenyon, S. J., & Bromley, B. C. 2004, AJ, 128, 1916
- Lehner, M. J., et al. 2009, PASP, 121, 138
- Liu, C.-Y., Chang, H.-K., Liang, J.-S., & King, S.-K. 2008, MNRAS, 388, L44
- Massey, F. J. J. 1951, J. Am. Stat. Assoc., 46, 68
- Nihei, T. C., Lehner, M. J., Bianco, F. B., King, S.-K., Giammarco, J. M., & Alcock, C. 2007, AJ, 134, 1596
- Pan, M., & Sari, R. 2005, Icarus, 173, 342
- Preston, D., Protopapas, P., & Brodley, C. 2009, Proc. 9th SIAM International Conf. on Data Mining, 61
- Roques, F., et al. 2006, AJ, 132, 819
- Roques, F., Moncuquet, M., & Sicardy, B. 1987, AJ, 93, 1549
- Starr, B. M., et al. 2002, SPIE Conf. Ser., 4836, 228
- Stewart, I. M. 2009, arXiv:0901.3276v1
- Wasserman, L. 2004, All of Statistics: A Concise Course in Statistical Inference (Berlin: Springer)