# Exploring the Eclipsing Binary Yield of the Large Synoptic Survey Telescope

Ava E. Polzin

Honors Thesis
*Northwestern University*
*Department of Physics & Astronomy*

Spring 2018

Advisor: Dr. Aaron M. Geller

DRAFT VERSION MAY 2, 2018
Typeset using LaTeX **twocolumn** style in AASTeX61

# EXPLORING THE ECLIPSING BINARY YIELD OF THE LARGE SYNOPTIC SURVEY TELESCOPE

A. POLZIN,[1] A. GELLER,[1,2] A. MILLER,[1,2] AND K. BREIVIK[1]

[1]*Northwestern University, Department of Physics & Astronomy and the Center for Interdisciplinary Exploration and Research in Astrophysics*
[2]*Adler Planetarium, Department of Astronomy*

## ABSTRACT

When the Large Synoptic Survey Telescope (LSST) goes online in the early 2020s, it will provide astrometric and photometric data on $10^{10}$ stars, generating approximately 20 terabytes of data per night for the duration of its 10-year project run. LSST will be uniquely adept at observing and identifying a variety of eclipsing binaries due to the breadth of the survey (covering twice the area of Sloan Digital Sky Survey in six different photometric bands), its short, 3-day cadence, and its long operational duration. Searching through the data for eclipsing binaries will be a feat unto itself for the sheer (and staggering) volume. Herein, we investigate the expected number of binaries LSST will observe, detect, and properly characterize, as well as constrain the observing parameters that will maximize the number of eclipsing binaries detected. Using LSST-similar conditions and a full galactic model, we complete various simulations of the eclipsing binaries in LSST's field of view and determine that, given the telescope's location and technical specifications, the survey will be able to detect 0.6% of all binaries in the southern sky and properly characterize 9% of those detected. This observed and identified portion of all southern-sky binary star systems is primarily characterized by short periods (<10 days), low eccentricities (<0.1), and radius ratios near unity.

*Keywords:* binaries: eclipsing — keyword2 — keyword3

## 1. INTRODUCTION

The Large Synoptic Survey Telescope (LSST) will allow for the most comprehensive survey of eclipsing binaries (EBs) in the southern sky to date when it begins operating in the early 2020s (LSST Science Collaboration et al. 2009). The telescope will collect approximately 20 terabytes of data per night over the ten years of its projected run; that vast amount of information will be singularly useful for the study of eclipsing binary stars. (Binaries are systems in which two stars are gravitationally bound and orbit around each other. EBs are those binaries with an inclination relative to observation of almost 90 degrees, allowing one star to be observed passing in front of its companion.) This particular utility is due primarily to LSST's short ($\sim$3 days) cadence, large amount of continuous sky coverage, and improved sensitivity (up to magnitude 24.5 in certain photometric bands).

Despite the strict condition on their inclination, EBs are relatively abundant - LSST should detect approximately 24 million EBs alone (Prša et al. 2011), making them a rich area of study. Often more important than eclipsing binaries themselves, however, is the stellar information they provide: primarily orbital period (as we will further demonstrate herein) and other orbital characteristics such as relative radii and inclination. When combined with radial velocity measurements, EB observations can also provide the masses of the binary's constituent stars (Torres et al. 2010).

In exploring a more refined measure of just how effective LSST's eclipsing observations will be, it is important to acknowledge the existing body of research in this area. In 2011, Prša et al. found that there was a period recovery rate of approximately 28% among LSST-observed eclipsing binaries. Their work employed a sample of 10,000 binaries with periods between 0.5 and 1,000 days and utilized a neural network to analyze output results. In 2017, Wells et al. recovered 71% of their sample, including only already known and characterized eclipsing binaries from the Kepler survey, which primarily have periods less than 10 days (and do not have any periods greater than 1,000 days).

Unlike Prša et al. (2011), we will generate observations for individual binaries using a randomly sampled time series with an approximately three day cadence across all six filters. While prior studies have used `OpSim` (Delgado et al. 2014) - LSST's own observation simulator - to generate observations with native irregularity (due to weather, maintenance, etc.), our randomized times (providing around 200 observations per filter, spread over 3,650 days) offer a similar set of chosen observation dates and retain the fluctuations that make `OpSim` so agreeable. Our sample will also be much larger than the 10,000 binaries utilized in the 2011 study, which will allow us to more accurately assess the factors that contribute most to correct binary identification. In addition, we keep a running tally of all binaries in the southern sky, those that may be detectable by LSST, and those that we have properly characterized. We also retain information about binaries excluded for being unobservable in LSST's survey (see Section 3.1).

By contrast, Wells et al. (2017) interpolated over existing light curves, taken from the Kepler Eclipsing Binary Catalog (KEBC) – containing just under 3,000 binaries, and used `OpSim` to reapply new observation times and conditions. Their use of already successfully detected binaries might also contribute to their higher yield, for the sample being inherently biased toward those eclipsing binaries that are most conducive to ready observation (i.e., those with shorter periods, lower eccentricities, and radius ratios near unity). This may be particularly true for Kepler being capable of detecting only up to magnitude 19 (Kirk et al. 2016).

Understanding LSST's eclipsing binary yield before the telescope goes online will allow astronomers to better plan for use of the extensive data. It is therefore important that theoretical investigations such as ours appropriately approximate the experimental behavior and results we will see beginning in the early 2020s. We describe our methods below. In Section 2, we will lay out the details of our simulation, explaining in more detail the eclipsing binary sample we will pull from as well as the means by which we return periods from generated observations. In Section 3, we examine our returned data, looking specifically at output distributions of our identified populations. In Section 4, we analyze our results, exploring their ramifications.

## 2. METHODOLOGY

### 2.1. *EB sample distribution*

We return a Milky Way-like distribution with `COSMIC` (Breivik & Larson 2018). `COSMIC` is a modified version of the standard binary star evolution model `BSE` (Hurley et al. 2002), which generates data for tens of thousands of binary systems at a time.

From this sample distribution, we draw stellar radii, masses, luminosities, and positions as well as orbital properties such as eccentricity and period. This population will be referred to as *All binaries* for the remainder of the paper.

Further, while we choose not to filter stars with positive declinations (assuming a uniform distribution, and that only distance will have a substantive bearing on observability), we exclude binaries with the following characteristics:

1. Inclinations less than the minimum necessary for observation of an eclipse ($Incl_{min}$). That minimum is calculated[1]:

$$Incl_{min} = 90 - 2 \times \arctan\frac{(r_1 + r_2)}{a} \times \frac{180}{\pi} \quad (1)$$

The radii of the binary's constituent stars (in astronomical units – AU) are indicated by $r_1$ and $r_2$. The semi-major axis (also in AU) is given by a and calculated:

$$a = \sqrt[3]{\frac{G(m_1 + m_2)P_{in}^2}{4\pi^2}} \quad (2)$$

$P_{in}$ indicates the input period (see Section 2.1) and is given in days. G is the gravitational constant. The masses of the binary's constituent stars (in solar masses) are indicated $m_1$ and $m_2$.

2. Radii exceeding the maximum necessary (accounting for eccentricity) for the binary to remain detached. Binaries are excluded when the following conditions are not met (Eggleton 1983):

$$0 < \frac{r(0.6(\frac{m_1}{m_2})^{2/3} + \ln(1 + \sqrt[3]{\frac{m_1}{m_2}}))}{0.49a(\frac{m_1}{m_2})^{2/3} \times (1 - ecc)} < 1 \quad (3)$$

The radius of either constituent star is indicated r. The orbital eccentricity is indicated ecc.

3. Apparent magnitudes that exceed the maximum LSST is capable of detecting (24) or that are less than the minimum LSST is capable of detecting (11). These magnitude limits are meant to be approximate across all bands (rather than taking into account the individual bands' sensitivities). Apparent magnitude (mag) is calculated:

$$mag = mag_{\odot} - 2.5\log_{10}((L_1 + L_2) \times (\frac{dist_{\odot}}{dist_{kpc}})^2) \quad (4)$$

This specific calculation is chosen to go between a binary's constituent luminosities and its magnitude using distances of kpc. The distance to the binary is given as $dist_{kpc}$. The apparent magnitude of the sun is given $mag_{\odot}$; the distance to the sun (1 AU) is given $dist_{\odot}$ and is in kpc. $L_1$ and $L_2$ give the constituent star's luminosities in units of solar luminosity.

4. Periods greater than twice the total observing time of 3,650 days, such that even the half period will not be fully sampled.

We keep careful track of the total number of binaries as well as the number potentially detectable with LSST to allow us to study which factors most frequently impair observation and which factors have only nominal impact on overall EB observation in the southern sky. This post-exclusion population will be considered *LSST-detectable* from here on.

### 2.2. *Light curve generation – `ellc`*

We use the python package `ellc` (Maxted 2016) to generate our binary light curves. (Prša et al. 2011 used `PHOEBE` (Prša & Zwitter 2005) instead.) While `ellc` can only handle detached binaries, it has a number of advantages over similar packages. The primary benefit of `ellc`, however, is its speed. Free parameters we provide based on the EB sample distribution (see Section 2.1) include: orbital period (which we will refer to as the "input period" or $P_{in}$ for the rest of the paper), stellar masses, stellar radii, stellar luminosities, orbital eccentricity, distance/position, and inclination. Other parameters, such as effective temperature, semi-major axis radius, and system error are calculated using the parameters listed above.

Using five of the six SDSS/LSST filters (u, g, r, i, and z are all available within `ellc` – y is not), we generate five light curves per binary source. Each source is sampled approximately 200 hundred times over the projected LSST run using a random distribution spread over the time of observation. (Prša et al. 2011 use all six filters and so have up to 50 more observations per source.) Like `OpSim`, this allows us to account for irregularities in LSST's cadence by producing a record of observations at uneven times within the scope of a generally three day cadence.

### 2.3. *Returing a period – `gatspy`*

We then use `gatspy` (VanderPlas & Ivezić 2015) to attempt to recover the input periods, allowing us to determine whether we could recover the true binary parameters with LSST observations.

We want to translate our now `ellc`-generated light curves into a returned period that matches the observations. One popular method to accomplish this is use of a Lomb-Scargle periodogram. (A periodogram plots possible periods vs. powers – likelihoods that the given period is correct.) Generally Lomb-Scargle accounts for one band; `gatspy` allows one to search for a period across all photometric bands simultaneously, returning a "multiband" periodogram in addition to the regular single band periodograms. We will rely on the results of the multiband periodogram – from here on, we will define this multiband result as the "output period" (or $P_{out}$).

[1] We recently discovered a bug in our minimum inclination calculation. We have changed this in our code and will use the corrected version going forward, though we do not anticipate that our results will be drastically affected.

The Lomb-Scargle method itself is a uniquely useful tool in astronomy for its allowance of non-uniform sampling (due to observing times being inherently uneven in their distribution – for only being at night, in good weather, ...). Other, similar periodogram packages use underlying calculations that disallow inhomogeneous (multiband) observations.

By including all five bands in a single period analysis, we increase the overall number of observations, and (presumably) increase our ability to recover the input period, as compared to analyses of any single band alone.

Defining a successful identification in the same manner as Wells et al. (2017), output period to input period ratios of 0.5, 1, and 2 will be permissible within a tolerance of 0.1 as each corresponds to a particular physical case that is easily observable in the light curves of the binary in question:

$$\frac{P_{out}-\frac{1}{2}P_{in}}{\frac{1}{2}P_{in}} \text{ or } \frac{P_{out}-2P_{in}}{2P_{in}} \text{ or } \frac{P_{out}-P_{in}}{P_{in}} \leq 0.1$$

We will refer to EBs that meet this criterion as *Identified EBs*.

## 3. ANALYSIS

### 3.1. *Exclusion statistics*

For this investigation, we generated 5,775,865 binaries via the methods of Section 2.1. In the following we will refer to this total sample of binaries as *All binaries*. Of those, we deem 34,849 to be *LSST-detectable* and 3,090 to be *Identified EBs*. This suggests that 0.6% of all southern sky Milky Way EBs will be detectable by the Large Synoptic Survey Telescope.

Examining the exclusions that eliminate 99.4% of the sample population, we find that – in descending order of frequency:

- 99.2% are excluded for giving non-eclipsing inclination (point 1 in Section 2.1).
- 24.7% are excluded for being either too faint or too bright as assessed by the apparent magnitude (point 3 in Section 2.1).
- 23.9% are excluded for having a period outside our limits (point 4 in Section 2.1).
- 17.6% are excluded for not being detached systems (point 2 in Section 2.1).

These samples are not mutually exclusive, and therefore the above percentages sum to >100%.

### 3.2. *Returned distribution statistics*

To analyze our data, we use cumulative distributions in conjunction with a two-sample `SciPy` Kolmogorov-Smirnov (K-S) statistic (Jones, Oliphant, Peterson et al. 2001) to compare our three samples, namely *All binaries*, *LSST-detectable EBs*, and *Identified EBs* (terms defined in Sections 2.1 and 2.3). We use the K-S test to quantitatively compare our *Identified* population to our entire generated one of *All binaries*, and we use histograms for a qualitative comparison of our *Identified* and *LSST-detectable* populations.

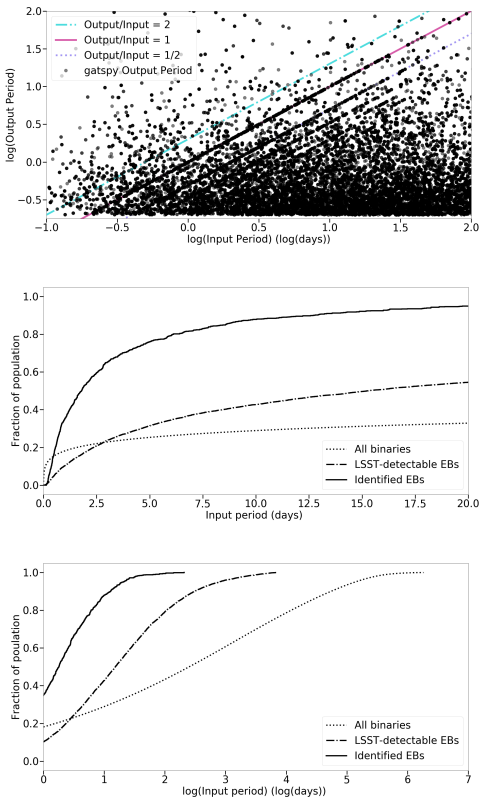


**Figure 1.** At top, we return a log-log plot of our input period to our output period for only those binaries considered *LSST-detectable*. Cyan, magenta, and purple lines highlight the *Identified EBs* in the plot and provide specific ratio values. The cumulative distributions (the latter in log) show *All binaries*, *LSST-detectable EBs*, and *Identified EBs*.

From the log-log period plot and the cumulative distributions, in Figure 1, it is clear that period has some bearing on *detectability* and *identifiability*. In comparing the *All binaries* and *Identified binaries*, we return a K-S p-value of 0.0, suggesting the population differences were so significant, the distributions were not even comparable and the test failed. We take this to mean that the distributions are distinct.

From Figure 2, the *Identified EB* preference for mass ratios near unity is clear. The cumulative distribution and the returned K-S stat of $1.787 \times 10^{-3}$ suggest that *All binaries*, *LSST-detectable EBs*, and *Identified EBs* are generally distinguishable.

Assessing the various populations' radius ratios, it becomes apparent that the *LSST-detectable* and *Identified* dis-

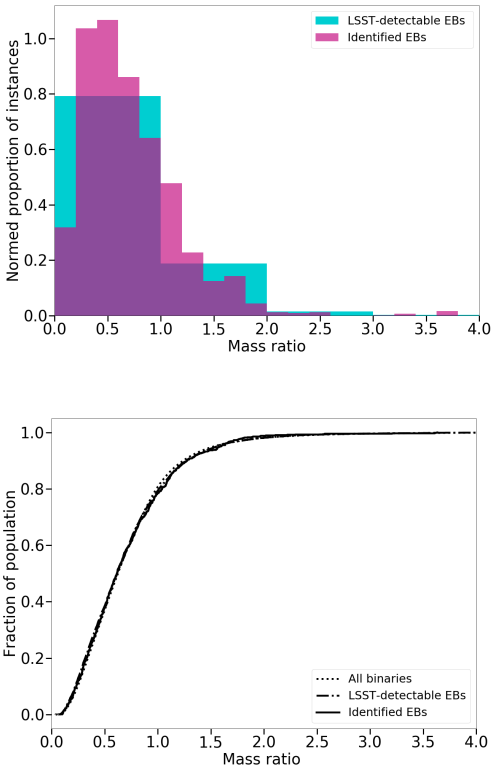


**Figure 2.** At top, we show the histogram distribution of the mass ratios *Identified EBs* (in magenta) relative to *LSST-detectable EBs* (in cyan). At bottom, we show a cumulative distribution mass ratios of *All binaries*, *LSST-detectable binaries*, and *dentified binaries*.

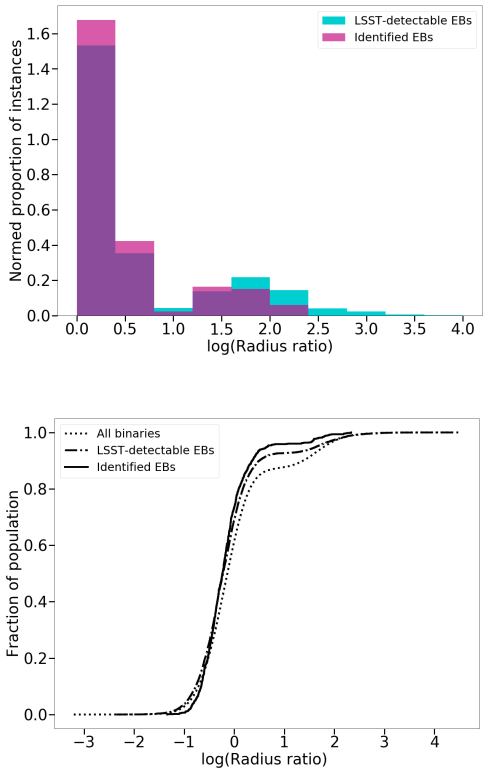


**Figure 3.** The above follow the format of Figure 2 for radius ratio.

tributions are relatively similar with a small bias toward radius ratios of order unity in both samples. Direct comparison between the *Identified* and *All binary* populations, however, yields a K-S p-value of $1.648 \times 10^{-47}$.

*Identified EBs* show a distinct low eccentricity preference. Though *LSST-detectable binaries* apparently more closely follow the *All binary* distribution, the K-S stat for the *Identified-All* comparison returns a p-value of $3.949 \times 10^{-295}$. This suggests that the various populations are extremely different from one another.

Because the inclination parent distribution is uniform, excluding the low end significantly alters the *LSST-detectable* and *Identified* samples relative to the *All binary* population. A comparison of *Identified* and *All binaries* returns a K-S p-value of 0.0. Once the EB has been detected, it shows no apparent inclination preference as is evident from the flat *Identified EB* distribution in Figure 5.

While Figure 7 demonstrates little preference on the part of *Identified* versus *LSST-detectable EBs* for distance, the *All binary* total distribution compared to the *Identified EB* distribution returns a p-value of $6.020 \times 10^{-23}$. There is also a bias towards apparent magnitudes just in the middle of LSST's observing range. The K-S stat p-value for the *Identified* and *All binary* comparison in apparent magnitude is 0.0.

As anticipated, we find that *Identified EBs* have the following general characteristics:

- short periods (<10 days)
- intermediate mass ratios (0.5-1)
- similar radii (radius ratio $\sim$1)
- low eccentricity (<0.1)
- intermediate apparent magnitudes (16-18; LSST's observing range is $\sim$11-24.5)

## 4. DISCUSSION

### 4.1. *Confirming consistent distributions*

We use the K-S test to determine if we can statistically distinguish the *Identified* and *All binary* distributions. By definition, a larger K-S distance statistic indicates more divergent distributions. A larger K-S distance statistic corresponds with a smaller K-S p-value (for a given sample size). We will consider a K-S p-value <0.01 to reject a hypothesis that the two distributions were drawn from the same parent distribution. We will lay out details of this analysis in the following subsections.

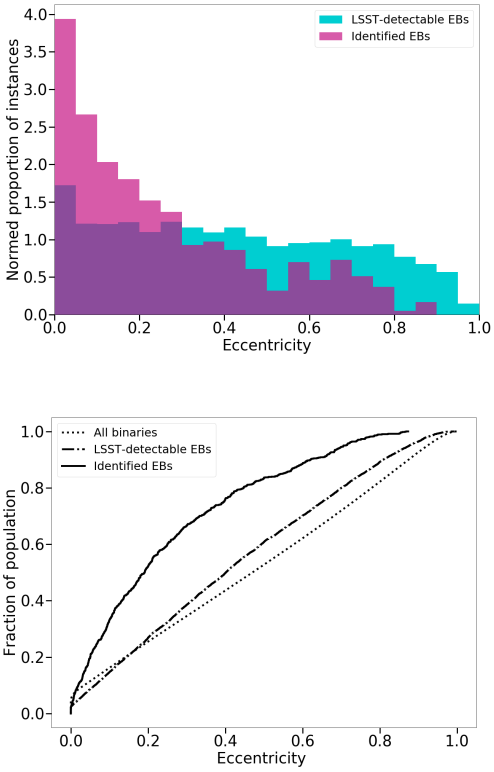


**Figure 4.** The above follow the format of Figure 2 for eccentricity.

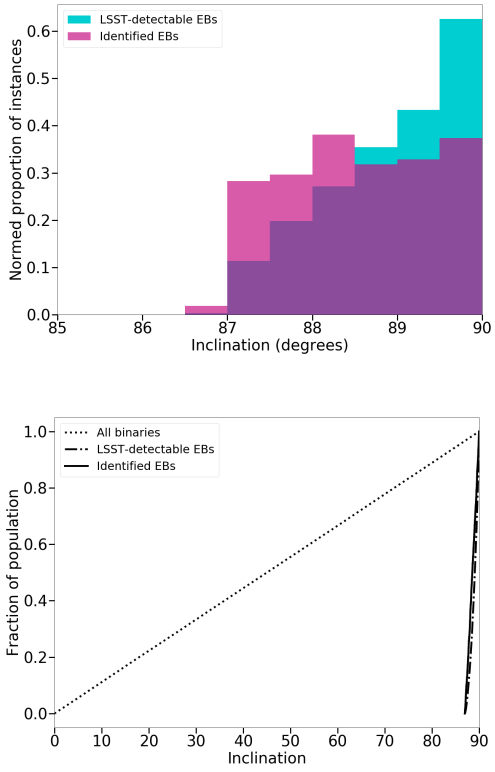


**Figure 5.** The above follow the format of Figure 2 for inclination.

### 4.2. *Returned period*

We find that *LSST-detectable* and *Identified binaries* favor lower periods. This is expected as lower period EBs will be more heavily sampled over LSST's ten year survey duration, while those with longer periods may never be observed in eclipse.

As evident from Figure 1, many of the unidentified binaries have longer periods that experience little to no magnitude change and return exceptionally low (fractional day) periods. For lack of necessary observations, `gatspy` returns the minimum period in its designated search range.

The K-S test also indicates that period distributions defined by *All binaries* and *Identified EBs* are significantly distinct.

### 4.3. *Returned mass*

Quantitatively, the K-S stat tells us that we can statistically distinguish the mass ratio distribution of *Identified EBs* and *All binaries* with respect to mass ratio. Qualitatively, Figure 2 suggests that mass ratios of 0.5 - 1.0 are marginally favored. Within this range, the dips in a light curve would be distinct and phase-folding packages will not mistake two peaks for a single peak sampled twice. This is presumably the sweet spot for identification, and likely accounts for the predominance of identified binaries that return the whole input period.

### 4.4. *Returned radius and eccentricity*

*LSST-detectable EBs* and *Identified EBs* both had radius ratios ranging on the order of 0.01 to 100. Most binaries in both of these samples, however, had much smaller ratios ($\sim$1-3). Larger radius ratios were excluded from even the *LSST-detectable* distribution which was likely due to these being part of attached or semi-attached systems (both of which are excluded for not being supported by `ellc`). Regardless, too large a radius ratio would be difficult to detect for the light curve being much less differentiable than in cases of a radius ratio near unity.

Eccentricity plays a role in the radius exclusion as well. As expected the *LSST-detectable* population is generally less eccentric than *All binaries*. EBs with extremely high eccentricities were already excluded from the detectable distribution for making the system attached or semi-attached at various parts of the orbit.

That the *Identified EBs* are still less eccentric than the *LSST-detectable* population indicates that eclipse time is a factor. Depending upon our orientation to the binary, its eccentricity could make an eclipse almost imperceptible for it being the most rapid part of a star's transit. As we see with period, eccentricity plays a significant role in whether a detected binary will be correctly characterized/identified.

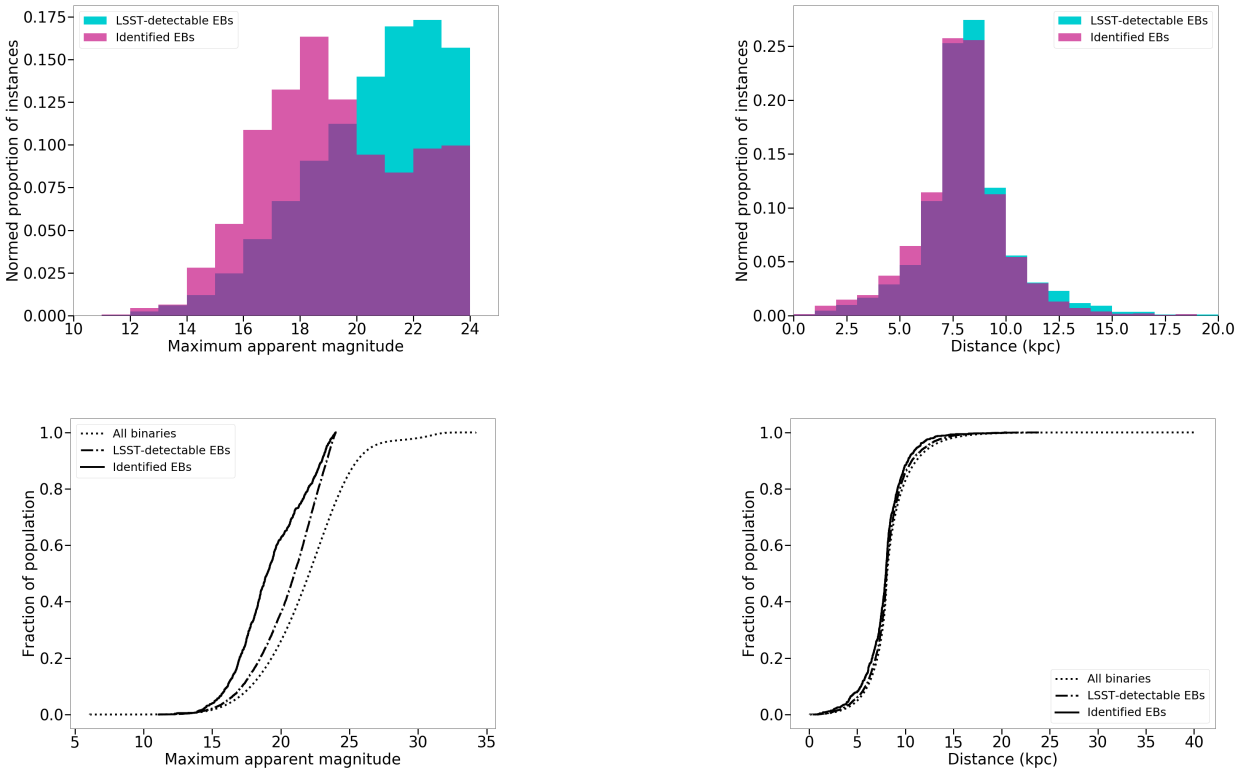


**Figure 6.** The above follow the format of Figure 2 for apparent magnitude.




**Figure 7.** The above follow the format of Figure 2 for distance.

### 4.5. *Returned inclination*

Despite inclination's considerable role in narrowing which binaries are *LSST-detectable*, we find that once the exclusions for non-eclipsing inclinations have been made, *Identified binaries* demonstrate no preference for higher or lower inclinations, instead returning a flat distribution.

The cumulative distributions raise an interesting point in this case – the K-S test indicates that the population of *All binaries* is significantly different from the population of *LSST-detectable EBs* and *Identified EBs* as the low (<85 degrees) end has been cut off.

### 4.6. *Returned apparent magnitude and distance*

Distance and apparent magnitude are intrinsically linked in our exclusions – more distant binaries will have to be much brighter intrinsically to still be detectable within the same range of magnitudes. Despite less distant stars generally returning lower apparent magnitudes, we find that *LSST-detectable* and *Identified EBs* are readily distinguishable from the population of *All binaries* in the southern sky. There is a small bias within the *Identified EBs* compared to *LSST-detectable binaries* toward closer stars – this is likely to do with the more obvious apparent magnitude bias we also see.

Like distance, the apparent magnitude distributions for *All binaries* and *LSST-detectable EBs* are similar, but distinguishible. The steep drop-off on either side of magnitudes 11 and 24 in both *Identified* and *LSST-detectable* populations is due to exclusions made for LSST's observing capabilities. There is also a bias toward binaries with magnitudes in the middle of that range (16-18) within the *Identified EB* population. This is likely due to light curves having lower uncertainties well-within the photometric bounds on LSST.

### 4.7. *EB yield*

Prša et al. (2011) and Wells et al. (2017) quote 28% and 71% identification respectively compared to our 8.9%. It is likely that this is simply a factor of differences in our methods and samples. Both Prša et al. and Wells et al. limit their EBs to only those with periods under 1,000 days – given the high performance of shorter period binaries and the increasing intrinsic frequency of binaries toward longer periods, this is a significant difference from our exclusion at 3,650 days. Additionally, Wells et al. utilize a catalog of already-identified/characterized binaries, interpolating over existing light curves. These binaries are already conducive to characterization with data that automatically excludes fainter binaries (see Introduction), so are likely especially conducive to characterization under the improved conditions of LSST.

When we cut our populations to match the restrictions set by Prša et al. (2011) – including only those binaries with periods less than 1,000 days and apparent magnitudes between 16 and 22 – we find that our recovered percentage of *Identified EBs* of those that are *LSST-detectable* increases from 8.9% to 10.5%. While it does not make up the difference between Prša's yield and ours, the increase is significant enough to confirm that apparent magnitude and period cuts alter the yield. It follows that other differences in our respective distributions will have as much impact. We plan to further investigate this difference to better understand why our results appear to be in tension.

### 4.8. *Future scientific promise*

We recognize that the K-S p-value relies on the sample size, which is very large for our theoretical *All binaries* sample, and therefore may exaggerate the difference between, e.g., *All binaries* and *Identified EBs*. With this in mind, we find that each *Identified EB* population is statistically distinct from its parent distribution of *All binaries*. We plan to investigate other methods to quantify the difference between these samples (which may be particularly important for cases where our qualitative comparisons show little difference, but the K-S p-value shows a significant distinction).

Still, clearly some of the *Identified EBs* are formally distinct from the *All binaries* distributions. This raises the question of whether the *Identified* distribution can be used to infer the characteristics of the *All binary* population (that makes up all binaries in the southern sky).

Investigating this relationship may be the subject of future work – specifically, examining whether the observed characteristics of the *Identified EB* population can be used to infer intrinsic characteristics of *All binaries*.

## 5. CONCLUSION

During its survey, LSST will be able to detect 0.6% of all southern sky binaries. Of those detected, there will be sufficient data to identify and characterize 9%.

Our work agrees with past studies in noting that shorter period binaries are more likely to be detected and are more likely to have the binary's characteristics (such as orbital period) correctly identified. This is, perhaps, an intuitive result as LSST's increased sampling will still favor systems that eclipse more frequently.

The *Identified EBs* will predominantly have shorter periods (<10 days), lower eccentricities (<0.1), and radius ratios near unity when compared to the original All binary population.

Going forward, we plan to implement `OpSim` – LSST's own observation simulator – in order to ensure that our simulation is as LSST-like as possible. Additionally, we plan to utilize other period recovery means to discern how much impact (if any) our use of `gatspy` had on our result.

This research was supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. This work was funded by a 2017 LSSTC Enabling Science Grant Award # 2017-UG05, to P.I. Aaron Geller, and also by a donation to CIERA.

*Software:* ellc, gatspy, COSMIC, Numpy, SciPy

## REFERENCES

Breivik, K., & Larson, S. 2018, in preparation

Delgado, F., Saha, A., Chandrasekharan, S., et al. 2014, in Proc. SPIE, Vol. 9150, Modeling, Systems Engineering, and Project Management for Astronomy VI, 915015

Eggleton, P. P. 1983, ApJ, 268, 368

Hurley, J. R., Tout, C. A., & Pols, O. R. 2002, MNRAS, 329, 897

Kirk, B., Conroy, K., Prša, A., et al. 2016, The Astronomical Journal, 151, 68

LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, ArXiv e-prints, arXiv:0912.0201 [astro-ph.IM]

Prša, A., & Zwitter, T. 2005, ApJ, 628, 426

Prša, A., Pepper, J., & Stassun, K. G. 2011, AJ, 142, 52

Torres, G., Andersen, J., & Giménez, A. 2010, A&A Rv, 18, 67

VanderPlas, J. T., & Ivezić, Ž. 2015, ApJ, 812, 18

Wells, M., Prša, A., Jones, L., & Yoachim, P. 2017, PASP, 129, 065003