

Data Analysis for OJ 287 from LCOGT

ASTR518: Astronomical Instrumentation & Observation, Fall 2024, Project Report

NINGYUAN XU¹

¹*Steward Observatory and Department of Astronomy, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA*

ABSTRACT

OJ 287 is known to have a binary black hole system at its center and their gravitational interactions lead to periodic flares in observation. The study of the variability of OJ 287 combined with multi-wavelength data could improve our understanding of the binary black hole system and their accretion disks. Thus in this paper, I present a basic data reduction pipeline for observations of OJ 287 from the Las Cumbres Observatory global telescope network (LCOGT). The pipeline involves bias and dark frame subtraction, flat field normalization, background modeling and subtraction, aperture and PSF photometry, and SNR calculation. I then apply this pipeline to 10 observations of OJ 287 to obtain a simple lightcurve over one month, which shows large fluctuations.

1. INTRODUCTION

Since the first discovery of OJ 287 over a century ago, OJ 287 has been the main interest of many astrophysicists due to its periodic flares. Located about 4 billion light years away, OJ 287's flare is still noticeable long before the modern large telescope, with luminosity exceeding the entire Milky Way. Due to these properties, OJ 287 is classified as the BL Lac object, a subtype of active galactic nuclei (AGN). The first theoretical model is proposed in the 1980s, which uses a binary black hole system and their gravitational interactions to explain the periodic outburst ([Sillanpaa et al. 1988](#)). Later in the 1990s, a more detailed model involving a secondary black hole crossing the accretion disk of the primary black hole is proposed ([Lehto & Valtonen 1996](#)). Since then astrophysicists use numerical simulations to test these models and compare to the observation. Recently, more studies on OJ 287 lead to various modifications on the original binary black hole model: [Tanaka \(2013\)](#) argues the distinguishability between tidal disruption event and the cavity created within the binary system, [Britzen et al. \(2023\)](#) proposes that the precession of jet is responsible for long-term variability. Furthermore, the study of OJ 287 also improve our understanding of jet mechanism, physical processes within the accretion disk, and the black hole properties such as mass and spin ([Valtonen et al. \(2024\)](#), EHT Collabration et al. in prep).

After the proposal of the binary black hole model, people combine this theoretical model with the observational pattern to predict the appearance of flaring event of OJ 287. Such approach is successful in predicting the 2015 and 2019 outburst ([Valtonen et al. 2016](#); [Laine et al. 2020](#)). This method also predicts a flaring event in October 2022 but there is an absence of outburst in the observational data, which implies that the previously-estimated black hole mass

is either too high or the precession is weaker (Komossa et al. 2023). The inconsistency between model prediction and observation demonstrates the imperfections of the current model. In order to resolve this issue, we should proceed from two directions. On the observational side, a more detailed observation of the lightcurve and variability of OJ 287 is needed, especially using multi-wavelength data to capture all physical processes. On the theoretical side, numerical simulations based on new model of accretion disk and binary black hole system should be generated and we can use ray-tracing code to obtain the predicted observational signature. By comparing the observed multi-band data to the predicted observational signature, we can better quantify the discrepancy and gradually resolve it.

In this report, I will focus on the observational side and provide a basic data reduction pipeline to process data from LCOGT (Brown et al. 2013). In Section 2 I discuss the condition of observation, telescope, and instrument used. Then in Section 3 I outline each sub-step of the data reduction pipeline and the lightcurve generated from this pipeline. Finally in Section 4 I discuss uncertainties associated with the current procedures and the outlook of the project.

2. METHODS

In this project, I use data from 10 observations from the LCOGT. A summary of two observations can be found in Table 1. It is noticeable that the seeing limit on 11-16 is pretty bad while the sky is really bright on 11-17. The reason I choose these two days is that these are the originally designated observation date for Bok 90". And the following data reduction will be focused on these two observations. The 11-16 data is obtained on a 0.4m telescope at McDonald Observatory using a QHY600 CMOS camera. The 11-17 data is obtained on a 0.4m telescope at Teide Observatory using a QHY600 CMOS camera. Both of them have exposure time of 100s and use SDSS-R band filter. The CCD they use has a Gain of $0.77e^-/\text{ADU}$ and dark current of $0.01e^-/\text{s/pixel}$ at 200K. The read noise is slightly different, with $3.1e^-/\text{pixel}$ for McDonald and $3.18e^-/\text{pixel}$ for Teide. The CCD has 9600×6422 pixels with plate scale of $0.74''/\text{pixel}$ and physical size $3.8\mu\text{m}$. However, the data use binned pixel with shape 2400×2400 , so the effective plate scale is about $2.74''/\text{pixel}$.

Table 1. Observation Summary

Time (UTC)	Site	Instrument	Read Noise (e^-/pixel)	Seeing (")	Mean Airmass	Temp (°C)	Sky Brightness (mag/arcsec 2)
2024-11-16 09:46:28.046	McDonald Observatory	sq31 ^a	3.1	41.38	1.20	14	17.17
2024-11-17 02:33:59.536	Teide Observatory	sq32	3.18	3.99	1.57	4.8	14.38

^a sq refers to the QHY600 CMOS camera.

3. RESULTS

After using DS9 to inspect the raw data, I find that there is an offset in the coordinate. To correctly calibrate the WCS coordinate, I apply [Astrometry.net](#) to the raw data. Astrometry.net matches the field of view of the input data with its database and returns the corresponding WCS coordinate. All the following analysis are performed on the calibrated data.

3.1. Bias & Dark Frame Subtraction

A bias frame (or zero-exposure frame) measures the dark current and any offset in the readout electronics. It captures the noise inherent in the sensor when no light is entering the camera. A dark frame captures the thermal noise that accumulates in the detector during an exposure. It is used to correct for hot pixels on the sensor and is important for a longer exposure time, where thermal effects become more prominent.

To perform the bias subtraction, I use 5 bias frames for each telescope and instrument collected on the same day as the raw data. The master bias frame is the average of these 5 frames. A similar procedure is used to perform dark subtraction, but it needs normalization due to the difference in exposure time. For the data I use, the exposure time for the dark frame is 300s. The master dark frame is the average of these normalized 5 frames. Then the subtraction is performed by subtracting the master bias and master dark frames from the raw data.

3.2. Flat Field

A flat field measures the spatial variations of the detector, which can be caused by pixel-to-pixel sensitivity variation (instrumental imperfection), dust spots, vignetting, cosmic rays, and etc. Normalizing the data by the flat field can remove some spatial variations. To apply the flat field, I use one frame for each telescope and instrument collected on the same day as the raw data. However, the flat field data (9600×6422) is not in binned pixels as the raw data (2400×2400). To address this, I use a 2nd-order polynomial to interpolate the flat field data and re-bin it to the shape of raw data. Since the flat field should have a mean value of count per pixel close to 1, I then normalize the flat field data. Finally, the flat field correction is performed by dividing the bias-and-dark-subtracted data by the normalized flat field.

The 11-17 raw data, bias-and-dark-subtracted data, flat-field-corrected data, and flat field used can be found in Fig. 1. From Fig. 1 we can see that the dark and bias frames contribute about 1/3 of the count per pixel and thus removing them is important for my observation. Besides, the flat field shows that the edge of the detector is less sensitive to the central part and the flat-field-corrected data also displays a similar pattern (edge seems contaminated). Thus, the following analysis will only be performed on the central region of the raw data, where the spatial variations are small.

3.3. Background Modeling

In order to estimate the background rate, I use two approaches: the first one applies the DAOPHOT MMM algorithm to the flat-field-corrected data; the second one uses DAOStarFinder and PSF photometry to extract all sources and model the background. The first approach is very straightforward and gives a rate of $8.01 \pm 0.07e^-/\text{s/pixel}$ for 11-16 data and $4.81 \pm 0.06e^-/\text{s/pixel}$ for 11-17 data.

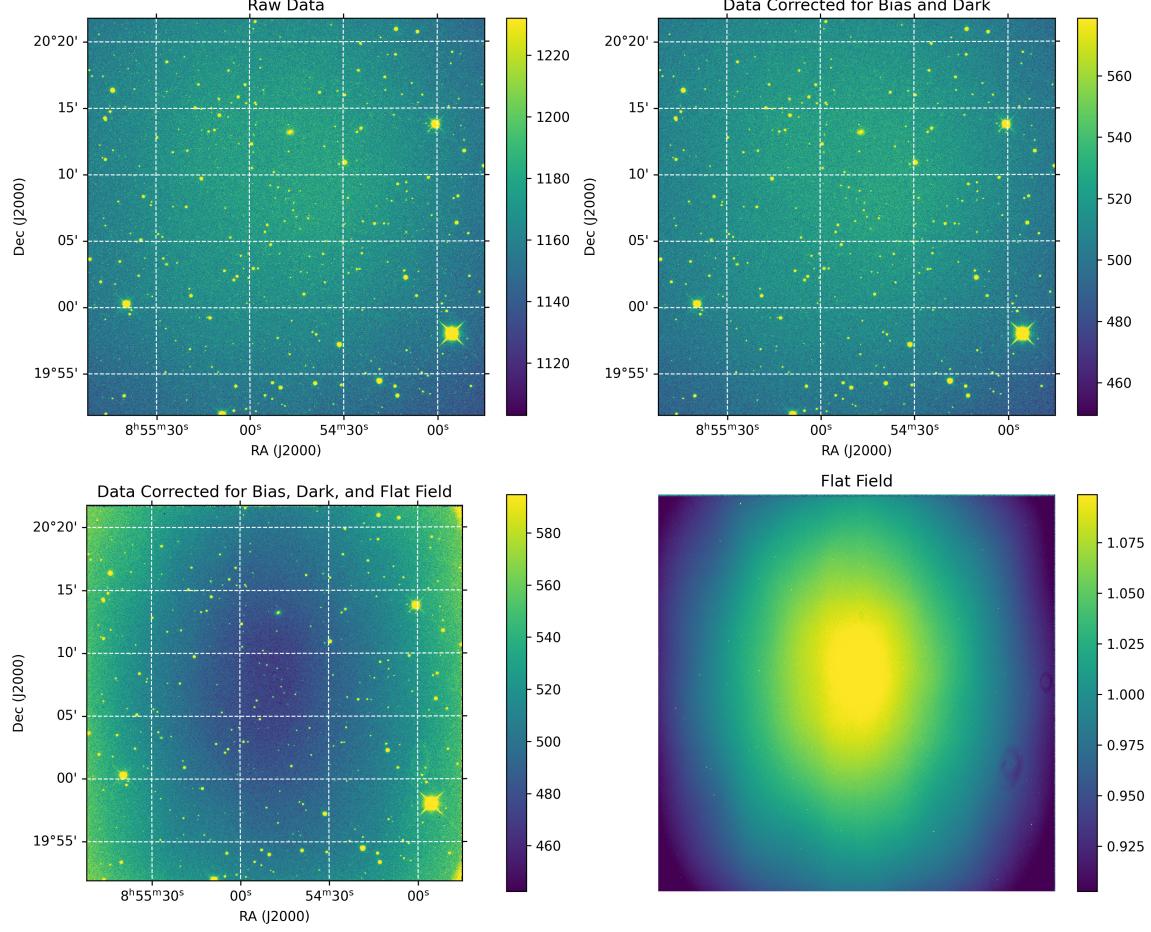


Figure 1. Upper Left: The raw data obtained on 11-17 with calibrated WCS coordinate. Upper Right: Data after bias and dark frame subtraction. Lower Left: Data normalized to the flat field. Lower Right: Normalized flat field data. The colorbar represents the count per pixel.

For the second approach, I assume the PSF has a Gaussian shape and the detection threshold is $5 - \sigma$ for the DAOStarFinder. The algorithm identifies 61 stars for 11-16 data and 47 stars for 11-17 data. Then the source model is created by using the identified sources from DAOStarFinder and the aperture radius $r = \frac{2 \times \text{Seeing}}{\text{PlateScale}} [\text{pixel}]$, assuming a Gaussian PSF. Thus, the modeled background is given by subtracting the source model from the flat-field-corrected data. Finally, the background rate can be calculated by taking the mean and standard deviation from the $3 - \sigma$ clipped, modeled background. Using this approach, I obtain $7.98 \pm 0.37 \text{e}^-/\text{s}/\text{pixel}$ for 11-16 data and $4.80 \pm 0.29 \text{e}^-/\text{s}/\text{pixel}$ for 11-17 data, which is consistent with the first approach.

The modeled background and identified sources for the central region of 11-16 data is shown in Fig. 2. From Fig. 2 we can see that the star finding algorithm identifies most sources, including OJ 287, and the background contributes a large portion of the detected counts per pixel.

3.4. PSF & Aperture Photometry

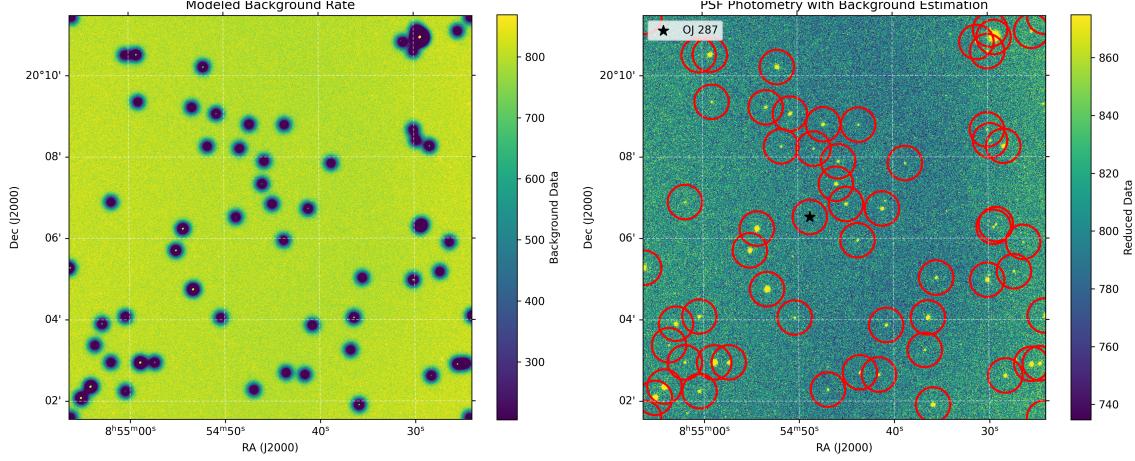


Figure 2. Left: the modeled background rate by the DAOStarFinder and PSF photometry. Right: the central region of 11-16 data after flat field correction; red circles represent the identified stars by DAOStarFinder and the black star is our source, OJ 287.

In this section I will outline the two methods I use to calculate the counts rate from the source: aperture and PSF photometry. For aperture photometry, I first calculate the sum of all counts within the circular aperture with radius r , C_t ; then the background rate C_b is calculated as the sum of all counts within the annulus between r and $2r$, and normalized by area. Here the background rate is obtained using the second approach as discussed in Section 3.3. The count rate from the source is thus given by $C_s = C_t - C_b$. Using this method, I get $1583\text{e}^-/\text{s}$ for 11-16 data and $74\text{e}^-/\text{s}$ for 11-17 data. The number of pixels within the aperture is $n_{\text{pix}} = 3681$ and 37 for 11-16 and 11-17 data, respectively.

For PSF photometry, I fit the source with a general 2D Gaussian Distribution with the following formula:

$$I(x, y) = A \cdot \exp \left(-\frac{1}{2} \left(\frac{(x - x_0)^2}{\sigma_x^2} + \frac{(y - y_0)^2}{\sigma_y^2} - 2 \cdot \frac{(x - x_0)(y - y_0)}{\sigma_x \sigma_y} \cdot \cos(2\theta) \right) \right) + \text{offset} \quad (1)$$

where A is the amplitude (peak intensity) of the Gaussian, (x_0, y_0) is the coordinate center, σ_x, σ_y are standard deviations of the Gaussian, θ is the angle of rotation of the Gaussian ellipse and offset represents the background value. The count rate from the source is then given by:

$$C_s = 2\pi \cdot A \sigma_x \sigma_y / t_{\text{exp}} \quad (2)$$

Using this method, I obtain $4520\text{e}^-/\text{s}$ for 11-16 data and $225\text{e}^-/\text{s}$ for 11-17 data. These values are about 3 times larger compared to the values obtained from aperture photometry, which might be caused by the fitted size of PSF being larger than the actual PSF or the overestimation of background rate in the previous step.

The source from 11-17 observation in the flat-field-corrected data and background-subtracted data, along with aperture and fitted PSF, is shown in Fig. 3. From Fig. 3 we can see that our modeled background works well in reducing the background noise and the fitted Gaussian PSF is consistent with the actual source. Additionally, the distribution of counts within the aperture is very similar to a Gaussian profile with central brightness and surroundings with descent brightness. I don't see an Airy disk in this case and this might be caused by following reasons: 1. The system is under-sampled: the central wavelength of SDSS-R filter is 623nm, diameter of the telescope is 0.4m. So the

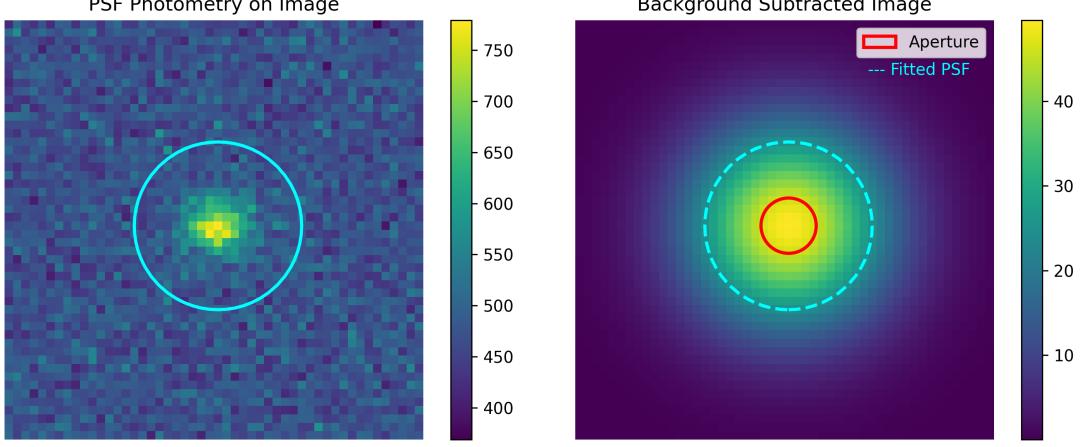


Figure 3. Left: region around OJ 287 in the flat-field-corrected data with cyan line showing the fitted PSF. Right: background subtracted data around OJ 287; red circle represents the calculated aperture while the cyan dashed circle represents the fitted Gaussian PSF.

diffraction limit is about $0.392''$, but the plate scale is given as $2.42''/\text{binned pixel}$ ($0.74''/\text{pixel}$). Thus the detector cannot fully resolve the diffraction pattern. 2. The seeing condition is not good for this observation, about $4''$. 3. The background subtraction can be refined using more advanced technique and more robust selection criterion.

3.5. SNR Calculation

In order to calculate the SNR, I use the following formula:

$$SNR = \frac{S_0 Q T G}{\sqrt{S_0 Q T G + S_s n_{pix} Q T G + S_d n_{pix} T G + (RG)^2}} \quad (3)$$

where S_0 is the count rate from the source, S_s is the sky background rate, S_d is the dark current, R is the readout noise, Q is the quantum efficiency of the instrument, T is the exposure time, and G is the gain of the detector. There is no source specifying the quantum efficiency of these instruments so I assume $Q = 0.7$ in the calculation. n_{pix} is given by the number of pixels within the aperture, as shown in Section 3.4. S_0 is obtained using the PSF photometry discussed in Section 3.4 and S_d is obtained using the modeled background discussed in Section 3.3. Thus, for 11-16 data the calculated SNR is 27.49 while for 11-17 data is 70.52. These values make sense because the seeing limit of 11-16 data ($41.38''$) is much worse compared to the seeing limit of 11-17 data ($3.99''$). I also use an [online calculator](#) to estimate the SNR on these telescopes and instruments. The estimated SNR for exposure time of 100s is about 20, which is generally consistent with the calculation.

3.6. Lightcurve

For the other 8 observations, I apply the same data reduction procedures and a lightcurve is obtained using the count rate estimated by the second approach in Section 3.4, $\text{SourceRate} = C_s / n_{pix} [\text{e}^-/\text{s}/\text{pixel}]$. The lightcurve is shown in Fig. 4, from which we can see that there are large fluctuations within this time range. However, the significant drop on 11-16 should be caused by the bad seeing limit ($41.38''$). Other fluctuations need further data and a more robust

reduction pipeline for inspection. Past study shows that flare events of OJ 287 occur with a period of several years (Laine et al. 2020), so the fluctuations observed in my data are likely systematics.

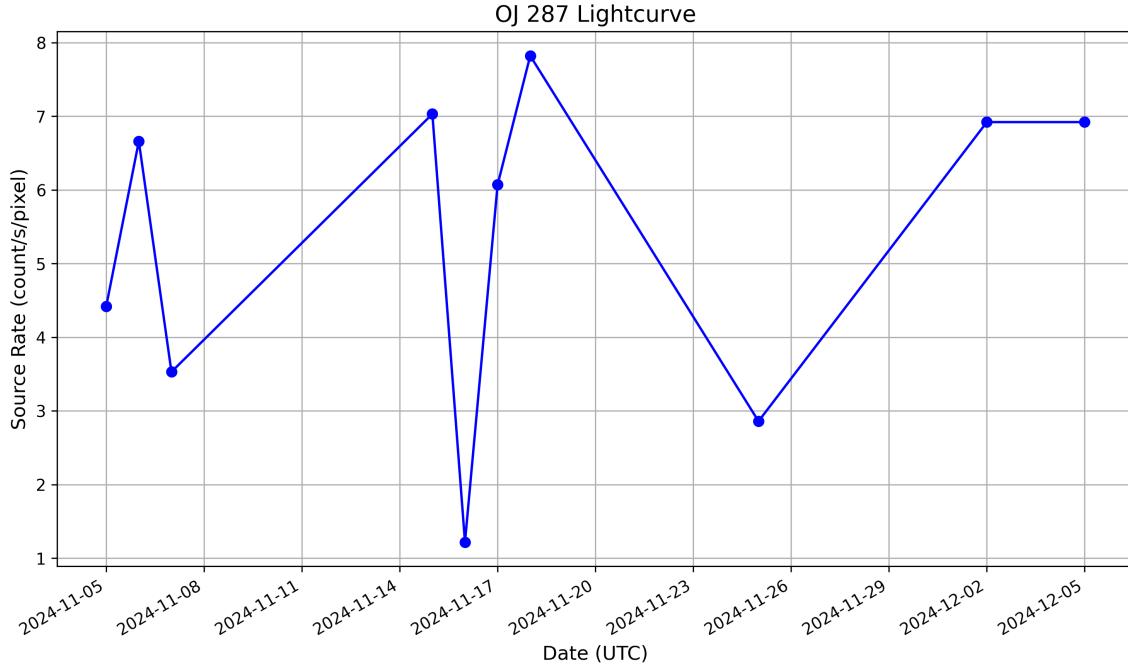


Figure 4. Lightcurve of OJ 287 using the data reduction pipeline discussed in Section 3 over one month.

4. DISCUSSION & CONCLUSION

This small project demonstrates the basic pipeline of data reduction and the reduced data is ready for photometric analysis given its moderately high SNR. However, there are still some uncertainties associated with the reduction pipeline. The bias frame contains uncertainty coming from variations in the bias levels, which can differ slightly across frames. Dark frame uncertainty comes from variations in dark current across the frame and there might be residual structure after subtracting the dark frame. Uncertainty in the flat field arises from the precision of the flat field calibration, and its quality depends on the number of flat frames taken and the stability of the instrument. In background modeling, the accuracy of the psf fitting and DAO star finder can introduce uncertainty, especially if the PSF is not well-matched to the observed object. This can also be impacted by nearby sources or variations in the sky background. Furthermore, for variable objects like OJ 287, the uncertainty in the measurement can also depend on the time between observations and any short-term variability in the object that was not accounted for. Finally, any issues with the instrument, such as misalignment or imperfect optical components, can introduce systematic errors in the measurements. Thus, a more robust and advanced reduction pipeline that addressed these issues is needed for research-level use. The implication of this project is that the observation of OJ 287 in the visible light band can achieve high enough SNR on a relatively small telescope with moderate exposure time. Thus it would be much easier

to obtain high-quality data with short exposure time on large and modern telescope to perform further analysis on OJ 287's properties.

My main research interest of OJ 287 lies on the interaction between binary black holes and their implication on the accretion disk. Thus, a series of observation is needed in order to obtain high-quality lightcurve and analyze variability. In addition, as pointed out by (Prince et al. 2021), multi-wavelength observation is also important for the study of OJ 287 system: x-ray and radio band can be used to study jet while LISA can capture the gravitational wave generated by the binary black holes. On the other side, numeric simulations like GRMHD (General Relativistic Magneto-Hydrodynamics) simulations are also needed to model the system in detail and compare to the observation.

All data and codes can be found in this [Github repository](#).

5. ACKNOWLEDGMENT

This work makes use of observations from the Las Cumbres Observatory global telescope network. The data are based on observations made with the QHY600 CMOS camera on the 0.4m telescope at McDonald Observatory in Texas and Teide Observatory in Spain.

REFERENCES

- Britzen, S., Zajaček, M., Gopal-Krishna, et al. 2023, ApJ, 951, 106, doi: [10.3847/1538-4357/acbbcc](https://doi.org/10.3847/1538-4357/acbbcc)
- Brown, T. M., Baliber, N., Bianco, F. B., et al. 2013, Publications of the Astronomical Society of the Pacific, 125, 1031, doi: [10.1086/673168](https://doi.org/10.1086/673168)
- Komossa, S., Grupe, D., Kraus, A., et al. 2023, MNRAS, 522, L84, doi: [10.1093/mnrasl/slad016](https://doi.org/10.1093/mnrasl/slad016)
- Laine, S., Dey, L., Valtonen, M., et al. 2020, The Astrophysical Journal Letters, 894, L1, doi: [10.3847/2041-8213/ab79a4](https://doi.org/10.3847/2041-8213/ab79a4)
- Lehto, H. J., & Valtonen, M. J. 1996, ApJ, 460, 207, doi: [10.1086/176962](https://doi.org/10.1086/176962)
- Prince, R., Agarwal, A., Gupta, N., et al. 2021, A&A, 654, A38, doi: [10.1051/0004-6361/202140708](https://doi.org/10.1051/0004-6361/202140708)
- Sillanpää, A., Haarala, S., Valtonen, M. J., Sundelius, B., & Byrd, G. G. 1988, ApJ, 325, 628, doi: [10.1086/166033](https://doi.org/10.1086/166033)
- Tanaka, T. L. 2013, MNRAS, 434, 2275, doi: [10.1093/mnras/stt1164](https://doi.org/10.1093/mnras/stt1164)
- Valtonen, M. J., Zola, S., Ciprini, S., et al. 2016, ApJL, 819, L37, doi: [10.3847/2041-8205/819/2/L37](https://doi.org/10.3847/2041-8205/819/2/L37)
- Valtonen, M. J., Zola, S., Gupta, A. C., et al. 2024, ApJL, 968, L17, doi: [10.3847/2041-8213/ad4d9b](https://doi.org/10.3847/2041-8213/ad4d9b)