

---

# Parameter Estimation in Realistic Binary Microlensing Light Curves with Neural Controlled Differential Equation

---

Haimeng Zhao<sup>1,2</sup> Wei Zhu<sup>1</sup>

## Abstract

Machine learning method has been suggested and applied to the parameter estimation in binary microlensing events, as a replacement of the time-consuming, sampling-based approach. However, the equal-step time series that is required by existing attempts are rarely realized in ground-based surveys. In this work, we apply the neural controlled differential equation (neural CDE) to handle microlensing light curves of realistic data quality. Our method can infer binary parameters efficiently and accurately out of light curves with irregular time-steps (including large gaps). Our work also demonstrates the power of neural CDE and other advanced machine learning methods in identifying and characterizing transient events in ongoing and future ground-based time domain surveys, given that it is common for astronomical time series from the ground to have irregular sampling and data gaps. The extended journal paper can be found at [arXiv:2206.08199](https://arxiv.org/abs/2206.08199).

## 1. Introduction

Irregular time series are common in astronomical observations from the ground. Such irregularities complicate the application of machine learning methods. The standard method, recurrent neural network (RNN, Goodfellow et al., 2016), requires the time series to be regularly sampled. In order to apply RNNs, one then needs to transform the original, unevenly sampled time series into evenly sampled ones through imputation or interpolation (e.g., Charnock & Moss, 2017), but this may destroy or introduce false information into the original data (Che et al., 2018). Alternatively, one may incorporate the time steps between observations into the RNN (e.g., Naul et al., 2018), but this makes strong

assumptions about the temporal evolution of the underlying process that may not be applicable to more general situations (Rubanova et al., 2019).

The neural controlled differential equation, or neural CDE, has recently been proposed as a general method to deal with partially observed, irregularly sampled time series (Kidger et al., 2020; Kidger, 2022). Neural CDE originates from the combination of neural differential equations (Chen et al., 2018; Rubanova et al., 2019) and controlled differential equations (Lyons et al., 2007), and it handles irregular time steps and missing values by approximating the underlying process as a differential equation. In this work, we apply neural CDE to the parameter estimation in binary microlensing light curves with irregular time steps and long gaps, as a demonstration of its power in the identification and characterization of astronomical transient events in ground-based time domain surveys.

Gravitational microlensing has long been suggested as a powerful technique in detecting faint and even dark objects (Einstein, 1936; Paczynski, 1986). In recent years, it has been successfully used to the detection and characterization of  $> 100$  exoplanets and many more stellar binaries (e.g., Mao & Paczynski, 1991; Gaudi, 2012). The interpretation of a microlensing light curve originating from multiple lenses can be very challenging, due to the time-consuming computation of the light curve as well as the pathological likelihood landscape of the high-dimensional parameter space. As a result, the current analysis of microlensing events is still case-by-case, and each event requires hundreds of CPU hours as well as the supervision of domain experts (e.g., Zang et al., 2022).

Machine learning has been proposed as a promising solution to the efficient modeling of binary microlensing light curves (Vermaak, 2003; Zhang et al., 2021), in addition to its application to the identification of microlensing events (Wyrzykowski et al., 2015; Godines et al., 2019; Mróz, 2020). However, the existing methods for parameter inferences are not readily applicable to realistic microlensing data. In particular, Zhang et al. (2021) trained and evaluated their deep learning network on simulated light curves from the *Roman* microlensing survey (Penny et al., 2019), which contained  $\sim 10^4$  time stamps with equal steps. Ongoing

---

<sup>1</sup>Department of Astronomy, Tsinghua University, Beijing, China <sup>2</sup>Zhili College, Tsinghua University, Beijing, China. Correspondence to: Haimeng Zhao <haimengzhao@icloud.com>, Wei Zhu <weizhu@tsinghua.edu.cn>.

microlensing surveys from the ground, such as the Korea Microlensing Telescope Network (KMTNet Kim et al., 2016), see light curves with lower signal-to-noise (S/N) ratios and irregular samplings (including data gaps). These realistic light curves cannot be processed in the framework of Zhang et al. (2021).

In this paper, we develop the first pipeline that can efficiently and accurately perform parameter inference in binary microlensing events of realistic data quality. The proposed pipeline is illustrated in Fig. 1. A neural CDE is used to extract features from light curves. The features are then fed into a mixture density network (MDN, Bishop & Nasrabadi, 2006) to predict the posterior density of the microlensing parameters. We show with experiments that our approach is flexible and robust to irregular sampling and data gaps. Our work paves the way toward applying deep learning to realistic microlensing data and serves as a potentially universal framework for learning from irregular astronomical time series.

## 2. Data

A standard binary microlensing light curve can be fully described by the following parameters: time of closest approach ( $t_0$ ), impact parameter ( $u_0$ ), Einstein ring crossing time ( $t_E$ ), the source size scaled to the Einstein ring ( $\rho$ ), the binary mass ratio ( $q$ ), binary projected separation ( $s$ ), angle from the binary lens axis to the direction of the source trajectory ( $\alpha$ ), the source flux fraction ( $f_s$ ), and the total baseline magnitude ( $m_0$ ). Among these parameters,  $m_0$  will be subtracted in the preprocessing step, and thus the exact choice has no effect. Parameters  $t_0$  and  $t_E$  characterize the temporal evolution of the source trajectory and can be inferred separated by a semantic segmentation algorithm, the full description of which will be provided in a future work. We thus set  $t_0 = 0$  and  $t_E = 1$  in the present work. We also fix  $\rho = 10^{-3}$ . For the chosen (effective) cadence of observations, the finite-source effect is not expected to be detected, and even if it is, the localized finite-source feature is not expected to affect our inference of other parameters. The remaining parameters are sampled according to  $u_0 \sim \text{Uniform}(0, 1)$ ,  $q \sim \text{LogUniform}(10^{-3}, 10^0)$ ,  $s \sim \text{LogUniform}(0.3, 3)$ ,  $\alpha \sim \text{Uniform}(0, 360)$ , and  $f_s \sim \text{LogUniform}(10^{-1}, 10^0)$ .

We use the VBBL algorithm (Bozza et al., 2018) through the `MulensModel` (Poleski & Yee, 2019) package to generate binary microlensing light curves. Each light curve composes of 500 data points, randomly (i.e., irregularly) sampled in the time interval  $[-2t_E, 2t_E]$  and assigned with Gaussian noises of S/N=33 in flux. To mimic the missing observations due to bad weather or telescope failures, we dump data within a randomly selected window with 1/25 of the total duration. Simulated events without significant binary features

( $\chi^2/\text{d.o.f.} < 2$  based on the single-lens fit) are filtered. In the end, six batches of  $10^5$  light curves are generated. Five of them are used as training set, and the remaining one is split into validation and test sets. For comparison, we have also produced data sets with equal samplings and irregular samplings but no gaps.

## 3. Method

**Preprocessing** We first preprocess the data before applying neural CDE. Light curves are subtracted by the baseline magnitudes and then normalized to obtain a variance of unity across the training set. To improve the training speed and reduce the memory requirement, we follow the approach of Morrill et al. (2021) to down sample the time series. Specifically, the light curves are fed into a depth- $n_D$  log-signature transform, which outputs a shorter, steadier, and higher dimensional sequence summarizing the sub-step information. We set the log-signature depth  $n_D = 3$  and the shortened length  $l = 100$ . The output data have  $v = 5$  dimensions.

**Neural CDE** We then interpolate the processed signal  $\{(t_i, x_i)\}_{i=0}^n, x_i \in \mathbb{R}^{v-1}$  with natural cubic splines, yielding a continuous signal  $X_t \in \mathbb{R}^v, t \in [t_0, t_n]$  with  $X_{t_i} = (t_i, x_i)$ . Neural CDE then defines a latent state  $z_t \in \mathbb{R}^w, t \in [t_0, t_n]$  with  $w = 32$  dimensions, whose evolution is described by the differential equation (Kidger et al., 2020)

$$z_t = z_{t_0} + \int_{t_0}^t f_\theta(z_\tau) dX_\tau \quad \text{for } t \in (t_0, t_n]. \quad (1)$$

Here  $f_\theta : \mathbb{R}^w \rightarrow \mathbb{R}^{wv}$  is a neural network parameterizing the dynamics and is further controlled by the signal  $X_t$ . The initial value  $z_{t_0} = \xi_\theta(X_{t_0})$  is defined by another neural network  $\xi_\theta$ . The subscript  $\theta$  denotes the dependence on the learnable parameters. Intuitively, neural CDE gradually extracts features from  $X_t$  according to a learnt policy  $f_\theta$ , and stores its knowledge in its latent state  $z_t$ . The terminal value  $z_{t_n}$  is then expected to carry useful information that is passed to downstream tasks.

**MDN** Binary microlensing events frequently suffer from model degeneracies (e.g., Yee et al., 2021). Therefore, the posterior probability distributions are preferred over certain values of microlensing parameters in this inference problem (Zhang et al., 2021). Here we adopt MDN, a neural network that estimates probability density with a mixture of Gaussians. Compared to the masked autoregressive flow method of Zhang et al. (2021), MDN provides a straightforward and efficient alternative with explicit density estimation. We feed the terminal value  $z_{t_n}$  from the neural CDE into the MDN. It outputs the normalized weight  $\pi_i$ , mean  $\mu_i$ , and covariance matrix  $\Sigma_i$  of each of the  $n_G$  multivariate

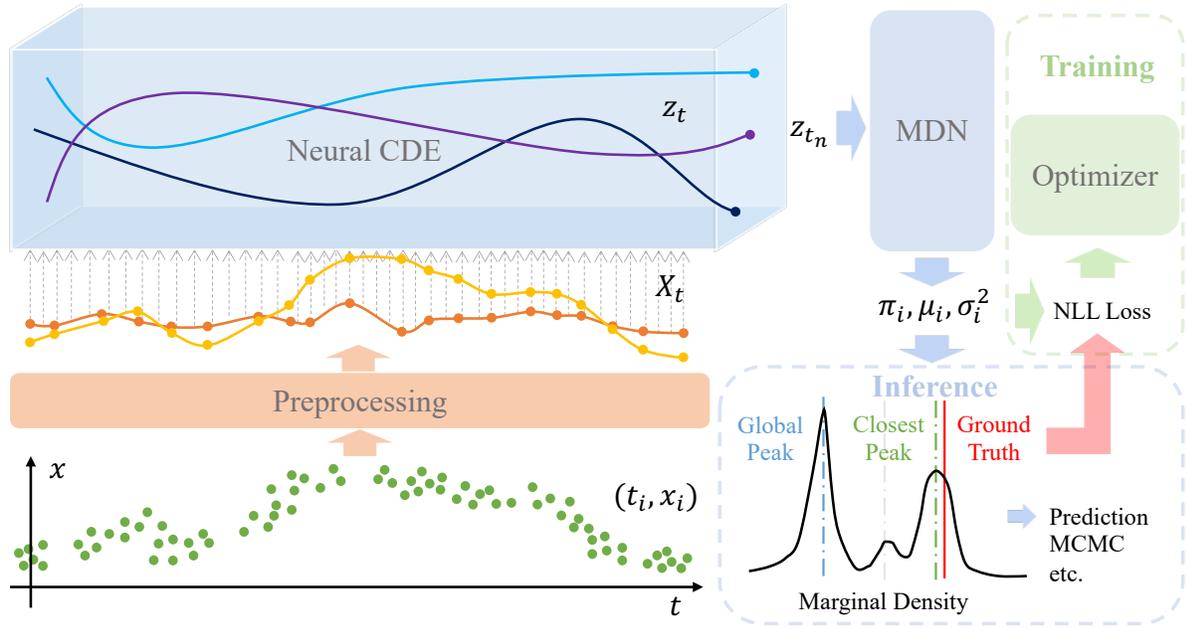


Figure 1. The schematic view of our proposed method. See Section 3 for the detailed description.

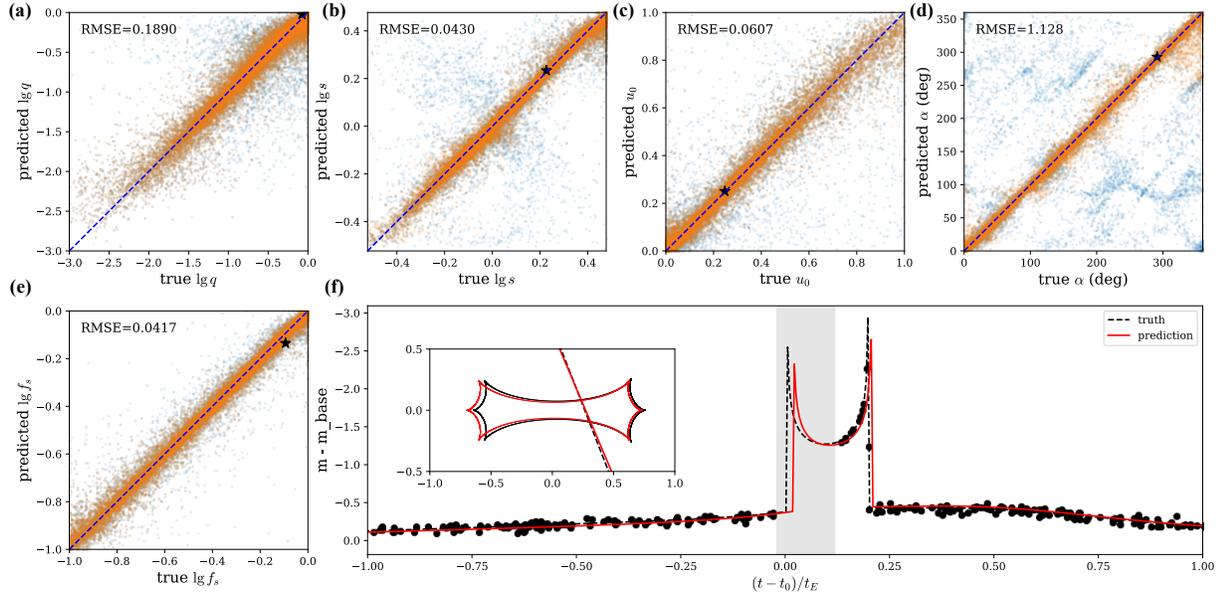


Figure 2. **Panels (a)–(e)**: Comparisons between ground truth and predicted values of microlensing parameters, namely  $\lg q$ ,  $\lg s$ ,  $u_0$ ,  $\alpha$ , and  $\lg f_s$ , respectively, based on 16384 binary light curves. Values corresponding to the global and closest peaks in the marginalized density distributions are shown in blue and orange, respectively. In each panel, the dashed line in each panel remarks the 1 : 1 line, and the accuracy of the prediction, measured by rooted mean squared errors (RMSEs), is indicated in the top left. **Panel (f)**: An example event whose binary anomaly is not covered with data points. Data points are shown as black dots, the input model is shown as the black dashed curve, and the data gap is marked as the gray shaded region. The red curve shows our model prediction, which matches the input reasonably well. The input and predicted values of the microlensing parameters are shown as black asterisks in panels (a)–(e). The inset shows the lensing geometry of both input and predicted models.

Gaussians. The probability density of the microlensing parameters,  $\omega \equiv (\lg q, \lg s, u_0, \alpha/180, \lg f_s)$ , is given by

$$p(\omega|z_{t_n}) = \sum_{i=1}^{n_G} \pi_i(z_{t_n}) \phi_{\mu_i(z_{t_n}), \Sigma_i(z_{t_n})}(\omega), \quad (2)$$

where  $\phi_{\mu, \Sigma}$  denotes the density of a multivariate Gaussian with mean  $\mu$  and covariance matrix  $\Sigma$ . For simplicity, we have assumed the multivariate Gaussians have diagonal covariance matrices. This preserves universality as long as  $n_G$  is large enough (Bishop & Nasrabadi, 2006). In the current network, we set  $n_G = 12$ .

**Network Details** The initial value network  $\xi_\theta$ , the evolution network  $f_\theta$  and MDN are constructed as residual fully connected networks with three residual blocks (He et al., 2016). Each block consists of two fully connected layers of width 1024. The evolution network has an additional activation of hyperbolic tangent to rectify the dynamic.

**Training** We use the ADAM optimizer (Kingma & Ba, 2014) and minimize the empirical negative log likelihood (NLL) to adjust the model parameters  $\theta$ . Specifically,

$$\theta = \operatorname{argmin} \{ -\mathbb{E}_{\text{training set}} [\ln p(\omega_{\text{true}}|z_{t_n})] \}. \quad (3)$$

The learning rate is set to  $10^{-4}$  initially and drops one percent after each epoch. Training samples are divided into batches of size 128 with label  $(\lg q, \lg s, u_0, \alpha/180, \lg f_s)$ . After three days of training on one NVIDIA Tesla V100 GPU, the average NLL loss drops to -8.76 on the validation set of size 1024. The optimized model is then used to infer the density of microlensing parameters (Equation 2).

## 4. Results & Discussion

We test our model on a set of 16,384 light curves generated in the same way as those in the training sample. The comparisons between the predicted and the ground truth parameter values are shown in panels (a)–(e) of Fig. 2. We follow Zhang et al. (2021) and adopt the peak in the marginalized density that is closest to the ground truth as the prediction of our model. This is justified on two bases. First, the closest peak is usually also the global peak of the density distribution.<sup>1</sup> Furthermore, the main purpose of our model is to identify all possible parameter sets that can closely fit the light curve. Once this purpose is achieved, one may employ the traditional optimization method to more efficiently locate the best solution(s). The accuracy of the prediction is measured by the root-mean-squared error (RMSE) between the ground truth and the prediction. For each microlensing parameter, the value of the RMSE is indicated in the upper left corner of the corresponding panel.

<sup>1</sup>To be quantitative, this is true in 90.6% and 86.5% of the test cases for parameters  $\lg q$  and  $\lg s$ , respectively.

Overall, our method can predict the microlensing parameters at high accuracy. Specifically, we can achieve typical fractional uncertainties of 43.5% and 9.9% on the mass-ratio  $q$  and the projected separation  $s$ , respectively. Compared to Zhang et al. (2021), which trained and evaluated on light curves consisting of  $10^4$  data points with S/N=200 and regularly sampled within the same time interval, our model achieves a comparable performance and less biased results based on light curves of more realistic quality (i.e., irregular sampling (including gaps) and orders of magnitude less data points). The achieved accuracy is high enough to ensure statistical analysis of the binary lens distribution. The prediction can also be used as input into the traditional sampling-based approach to further refine the parameters.

As an example, we show in the panel (f) of Fig. 2 light curves of a simulated event whose anomalous feature is affected by the injected data gap. Such a light curve cannot be properly analyzed via the standard RNN approach, whereas our model is able to predict the correct microlensing parameters and recover the “missing” anomaly.

Our model can analyze  $\sim 100$  light curves per second. For a comparison, the standard modeling based on Markov chain Monte Carlo (MCMC) approach typically requires the computation of  $\sim 10^4$  light curves, each taking  $\sim 1$  second, in order for a reasonable sampling of the posterior distribution. The initial localization of the rough solutions through a grid search requires even more light curve calculations. Our method thus remarks a speedup of  $\gtrsim 10^6$  over the existing approach.

## 5. Summary & Future Work

In this work, we apply the state-of-the-art machine learning method to the parameter inference in binary-lens microlensing events with realistic data quality. Unlike previous works, we train and evaluate on light curves that suffer from irregular sampling and missing data, both of which are frequently encountered in astronomical observations from the ground (and sometimes space as well). We show that our method can predict the key microlensing parameters efficiently and accurately. In the next step, we will incorporate the methods to infer the remaining microlensing parameters into the current pipeline and apply it to the analysis of real microlensing events.<sup>2</sup>

Irregular samplings and data gaps are common in time domain astronomy. Our work showcases the great potential of neural CDE and other advanced machine learning algorithms in the identification and characterization of astronomical transients, including but not limited to supernova, variable stars, exoplanet transit, and the coalescence of com-

<sup>2</sup>This has been implemented. See our journal paper arXiv:2206.08199 for details.

pact objects due to gravitational wave radiation.

## Acknowledgements

We thank Zerui Liu and Minghao Liu for discussions. This work is supported by the National Science Foundation of China (grant No. 12173021 and 12133005) and the science research grant from the China Manned Space Project with No. CMS-CSST-2021-B12. We also acknowledge the Tsinghua Astrophysics High-Performance Computing platform for providing computational and data storage resources.

## References

- Bishop, C. M. and Nasrabadi, N. M. *Pattern recognition and machine learning*. Springer, 2006.
- Bozza, V., Bachelet, E., Bartolić, F., Heintz, T. M., Hoag, A. R., and Hundertmark, M. VBBINARYLENSING: a public package for microlensing light-curve computation. *Monthly Notices of the Royal Astronomical Society*, 479(4):5157–5167, October 2018. doi: 10.1093/mnras/sty1791.
- Charnock, T. and Moss, A. Deep Recurrent Neural Networks for Supernovae Classification. *Astrophysical Journal Letters*, 837(2):L28, March 2017. doi: 10.3847/2041-8213/aa603d.
- Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. Recurrent neural networks for multivariate time series with missing values. *Scientific reports*, 8(1):1–12, 2018.
- Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.
- Einstein, A. Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field. *Science*, 84(2188):506–507, December 1936. doi: 10.1126/science.84.2188.506.
- Gaudi, B. S. Microlensing surveys for exoplanets. *Annual Review of Astronomy and Astrophysics*, 50:411–453, 2012.
- Godines, D., Bachelet, E., Narayan, G., and Street, R. A machine learning classifier for microlensing in wide-field surveys. *Astronomy and Computing*, 28:100298, 2019. ISSN 2213-1337. doi: <https://doi.org/10.1016/j.ascom.2019.100298>. URL <https://www.sciencedirect.com/science/article/pii/S2213133719300113>.
- Goodfellow, I., Bengio, Y., and Courville, A. *Deep learning*. MIT press, 2016.
- He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- Kidger, P. On neural differential equations. *arXiv preprint arXiv:2202.02435*, 2022.
- Kidger, P., Morrill, J., Foster, J., and Lyons, T. Neural controlled differential equations for irregular time series. *Advances in Neural Information Processing Systems*, 33: 6696–6707, 2020.
- Kim, S.-L., Lee, C.-U., Park, B.-G., Kim, D.-J., Cha, S.-M., Lee, Y., Han, C., Chun, M.-Y., and Yuk, I. KMTNET: A Network of 1.6 m Wide-Field Optical Telescopes Installed at Three Southern Observatories. *Journal of Korean Astronomical Society*, 49(1):37–44, February 2016. doi: 10.5303/JKAS.2016.49.1.037.
- Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Lyons, T. J., Caruana, M., and Lévy, T. *Differential equations driven by rough paths*. Springer, 2007.
- Mao, S. and Paczynski, B. Gravitational Microlensing by Double Stars and Planetary Systems. *Astrophysical Journal Letters*, 374:L37, June 1991. doi: 10.1086/186066.
- Morrill, J., Salvi, C., Kidger, P., and Foster, J. Neural rough differential equations for long time series. In *International Conference on Machine Learning*, pp. 7829–7838. PMLR, 2021.
- Mróz, P. Identifying Microlensing Events Using Neural Networks. *Acta Astronomica*, 70(3):169–180, September 2020. doi: 10.32023/0001-5237/70.3.1.
- Naul, B., Bloom, J. S., Pérez, F., and van der Walt, S. A recurrent neural network for classification of unevenly sampled variable stars. *Nature Astronomy*, 2:151–155, November 2018. doi: 10.1038/s41550-017-0321-z.
- Paczynski, B. Gravitational Microlensing by the Galactic Halo. *The Astrophysical Journal*, 304:1, May 1986. doi: 10.1086/164140.
- Penny, M. T., Gaudi, B. S., Kerins, E., Rattenbury, N. J., Mao, S., Robin, A. C., and Calchi Novati, S. Predictions of the WFIRST Microlensing Survey. I. Bound Planet Detection Rates. *The Astrophysical Journal Supplement Series*, 241(1):3, March 2019. doi: 10.3847/1538-4365/aafb69.
- Poleski, R. and Yee, J. C. Modeling microlensing events with MulensModel. *Astronomy and Computing*, 26:35, January 2019. doi: 10.1016/j.ascom.2018.11.001.

- Rubanova, Y., Chen, R. T., and Duvenaud, D. K. Latent ordinary differential equations for irregularly-sampled time series. *Advances in neural information processing systems*, 32, 2019.
- Vermaak, P. Rapid analysis of binary lens gravitational microlensing light curves. *Monthly Notices of the Royal Astronomical Society*, 344(2):651–656, 09 2003. ISSN 0035-8711. doi: 10.1046/j.1365-8711.2003.06851.x. URL <https://doi.org/10.1046/j.1365-8711.2003.06851.x>.
- Wyrzykowski, Ł., Rynkiewicz, A. E., Skowron, J., Kozłowski, S., Udalski, A., Szymański, M. K., Kubiak, M., Soszyński, I., Pietrzyński, G., Poleski, R., et al. OGLE-III microlensing events and the structure of the galactic bulge. *The Astrophysical Journal Supplement Series*, 216 (1):12, 2015.
- Yee, J. C., Zang, W., Udalski, A., Ryu, Y.-H., Green, J., Henderley, S., Marmont, A., Sumi, T., Mao, S., Gromadzki, M., Mróz, P., Skowron, J., Poleski, R., Szymański, M. K., Soszyński, I., Pietrukowicz, P., Kozłowski, S., Ulaczyk, K., Rybicki, K. A., Iwanek, P., Wrona, M., Albrow, M. D., Chung, S.-J., Gould, A., Han, C., Hwang, K.-H., Jung, Y. K., Kim, H.-W., Shin, I.-G., Shvartzvald, Y., Cha, S.-M., Kim, D.-J., Kim, S.-L., Lee, C.-U., Lee, D.-J., Lee, Y., Park, B.-G., Pogge, R. W., Bachelet, E., Christie, G., Hundertmark, M. P. G., Maoz, D., McCormick, J., Natusch, T., Penny, M. T., Street, R. A., Tsapras, Y., Beichman, C. A., Bryden, G., Novati, S. C., Carey, S., Gaudi, B. S., Henderson, C. B., Johnson, S., Zhu, W., Bond, I. A., Abe, F., Barry, R., Bennett, D. P., Bhattacharya, A., Donachie, M., Fujii, H., Fukui, A., Hirao, Y., Silva, S. I., Itow, Y., Kirikawa, R., Kondo, I., Koshimoto, N., Alex Li, M. C., Matsubara, Y., Muraki, Y., Miyazaki, S., Olmschenk, G., Ranc, C., Rattenbury, N. J., Satoh, Y., Shoji, H., Suzuki, D., Tanaka, Y., Tristram, P. J., Yamawaki, T., Yonehara, A., and MOA Collaboration. OGLE-2019-BLG-0960 Lb: the Smallest Microlensing Planet. *The Astronomical Journal*, 162(5):180, November 2021. doi: 10.3847/1538-3881/ac1582.
- Zang, W., Yang, H., Han, C., Lee, C.-U., Udalski, A., Gould, A., Mao, S., Zhang, X., Zhu, W., Albrow, M. D., Chung, S.-J., Hwang, K.-H., Jung, Y. K., Ryu, Y.-H., Shin, I.-G., Shvartzvald, Y., Yee, J. C., Cha, S.-M., Kim, D.-J., Kim, H.-W., Kim, S.-L., Lee, D.-J., Lee, Y., Park, B.-G., Mróz, P., Skowron, J., Poleski, R., Szymański, M. K., Soszyński, I., Pietrukowicz, P., Kozłowski, S., Ulaczyk, K., Rybicki, K. A., Iwanek, P., Wrona, M., and Gromadzki, M. Systematic KMTNet Planetary Anomaly Search. IV. Complete Statistical Sample of 2019 Prime-Field Microlensing Planets. *arXiv e-prints*, art. arXiv:2204.02017, April 2022.
- Zhang, K., Bloom, J. S., Gaudi, B. S., Lanusse, F., Lam, C., and Lu, J. R. Real-time likelihood-free inference of roman binary microlensing events with amortized neural posterior estimation. *The Astronomical Journal*, 161(6): 262, 2021.