Draft version September 28, 2024
Typeset using LaTeX **twocolumn** style in AASTeX631

A Deep, Archival Search for Nuclear Flares. SURF Project Final Report

Thuwaragesh Jayachandran[1]

[1] *California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA*

ABSTRACT

Tidal Disruption Events (TDEs) occur when a star is shredded by the tidal forces of a supermassive black hole (SMBH) and could potentially provide crucial insights into quiescent black holes. Recent observations of TDEs suggest that the traditional definitions applied to nuclear flares to be classified as TDEs could be restricting the domains of TDEs we observe. Further several TDE properties and their correlations with host galaxy properties are yet to be constrained. This project addresses these challenges using re-processed archival data from the Zwicky Transient Facility (ZTF). By developing software and utilizing a novel approach of exponentially smoothed moving average (ESMA) filters to detect flares from light curves, we build a comprehensive and inclusive catalog of flares. These flares will be analyzed to extract physical parameters and study their correlations with host galaxy properties, ultimately forming a broader and more precise definition of TDEs to better understand them.

1. INTRODUCTION

The occurance of TDEs produce multi-wavelength flares if the tidal radius is greater than the black hole event horizon, which in turn implies that there exists a maximum BH mass for an observable TDE (Hills 1975). Thus, studying TDEs could help understanding the lower mass portion of the black hole mass function. The accretion of roughly half of the stellar mass during a TDE also suggests that it could provide crucial insights into accretion disk physics. More importantly, by producing observable flares, TDEs equip us with a new way of observing the abundant quiescent black holes. Detecting TDEs still proves to be a challenging task. The traditional selection criteria used to do so could have missed several TDEs as suggested by recent observations of fainter and slower TDEs (Charalampopoulos et al. 2023; Yao et al. 2023).

In this project, we build a software that can process galaxy light curves by implementing Gaussian Process (GP) regression on them, and run the fitted curve through an exponentially smoothed moving average (ESMA) filter in search of significant rise and decline of flux values over a specified timescale, thus detecting any nuclear flares. The sensitivity of the detection software is then tested by flare simulation models comprising of a Gaussian rise followed by an exponential decay. Alongside we select and build a catalog of galaxies on which the detection software is to be run. With a complete picture of the range of parameter values that our detection software is sensitive to, we aim to run it on the selected list of galaxies, in search of any nuclear flares. These nuclear flares are then further analyzed by studying several physical properties extracted from those flares in order to identify any correlations with host galaxy properties. The data will also be used to constrain TDE rates.

2. METHODS

2.1. *Parsing Light Curves*

The galaxy light curves are data points plotted as flux value vs. time. The flux values were adjusted by deducting a weighted average of the data points before modified Julian date 58500, with the weights being chosen as $\frac{1}{error^2}$. This choice allows us to make zero point correction based on the left most data points and ensures that the color calculations do not have any negative values as inputs for the logarithms.

2.2. *Gaussian Process Regression*

A Gaussian Process (GP) is defined as a stochastic process (a collection of random variables indexed by time or space), such that every finite collection of those random variables has a multivariate normal distribution (Rasmussen & Williams 2006). With the specification of a covariance matrix as the 'kernel' of a GP, it is possible to define an expected numerical relationship between variables. GP regression provides a normal distribution as a prediction, rather than a single fitted value.

The irregularities in observations due to factors such as weather result in periodical gaps in light curves. Us-

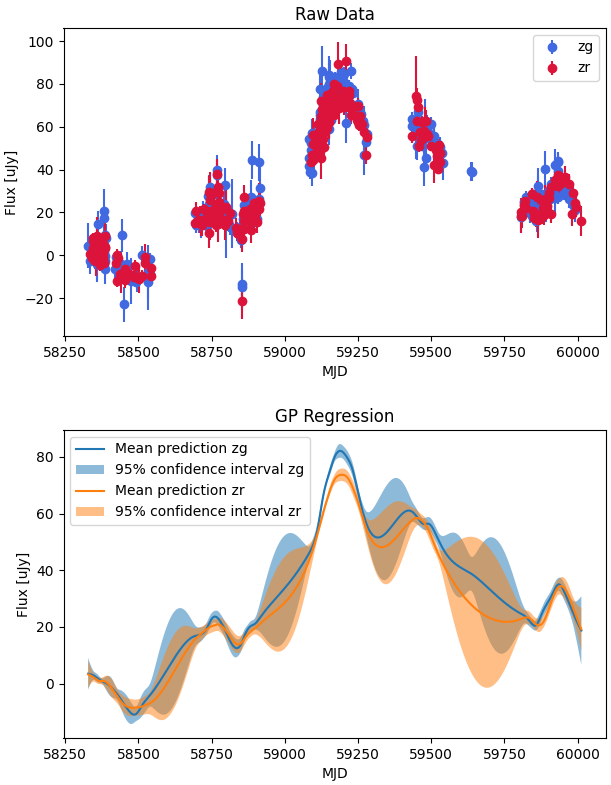

Figure 1. An example light curve data plotted as flux vs. time and the fit obtained for it using Gaussian Process regression

ing GP regression enables us to fit the light curve as a smooth graph and to fill between the observational gaps. The mean of the predicted normal distribution is used as the flux value and its standard deviation is taken as a measurement of the error band around it. This provides us with a more sensible fit for a light curve, accounting for errors.

We used GP regression with a Matern-3/2 kernel, as suggested in McLaughlin et al. (2024), and interpolated the light curves with a cadence of 2 days. `george` library in Python enabled implementing GP fit under a reasonable time limit.

2.3. *Exponentially Smoothed Moving Average Filters*

Exponential smoothing is a technique used for smoothing time series data with the exponential window function. It is calculated as a weighted average of the inputs with the weights exponentially decreasing over time. The simplest form takes the following formula,

$$s_i = \alpha x_i + (1-\alpha)s_{i-1} = s_{i-1} + \alpha(x_i - s_{i-1}) \quad \text{(1a)}$$

$$s_0 = x_0 \quad \text{(1b)}$$

Range of raw_j	Value chosen for x_j
$0.20 < raw_j$	+1.0
$0.06 < raw_j \leq 0.20$	+0.5
$0.03 < raw_j \leq 0.06$	+0.3
-0.03 $\leq$raw$_j \leq$ 0.03	0.0
-0.06 $\leq$raw$_j <$ -0.03	-0.3
-0.020 $\leq$raw$_j <$ -0.06	-0.5
$raw_j <$ -0.20	-1.0

Table 1. The step function used to map intermediate raw_j values to x_j values confined to [-1,1]

where x_i is the raw value at the ith timeseries point, s_i is the smoothed moving average value at the ith timeseries point, and α is the smoothing factor; $0 \leq \alpha \leq 1$.

By confining x_j to [-1,1], we confine s_j as well. A contiguous set of x_j's above some threshold will smooth out s_j above the same threshold. By expanding equation (1a) it can be seen that the weights of x_j is $\alpha(1-\alpha)^{i-j}$ for $1 \leq j \leq i$. This means that exponential smoothing places higher waits to the more recent x_j's as opposed to placing equal weights for all x_j's in a simple moving average function.

We model the change in flux values over adjacent timeseries points as our raw value, x_j. To do so, we utilize the mean values, y_j, and standard deviation values, σ_j, obtained from the GP fit. Before calculating x_j, we define an intermediate raw value as,

$$raw_j = \frac{y_j - y_{j-1}}{\sqrt{{\sigma_j}^2 + {\sigma_{j-1}}^2}} \quad (2)$$

The final value of x_j is calculated by confining raw_j to [-1,1] using the step function described in Table 1. Accordingly, a positive value of x_j indicates an increase in the flux and a negative value indicates a decrease in the flux. A higher absolute magnitude, $|x_j|$, is correlated with a steeper change in the flux and low uncertainty. The usage of a step function for mapping x_j to [-1, 1] is mainly because of its simplicity.

The x_j values are then smoothed using the EMA formulae, with $\alpha = \frac{\Delta t}{T}$ with $T = 9$ days and $\Delta t = 2$ days since we chose a cadence of 2 days. A rise of smoothed value s_j above 0.8 would be flagged as an onset of a flare. If an onset is then followed by a decline of the smoothed value below -0.8, it is considered and detected as a flare. Thus, a window of the smoothed value that begins with a +0.8 followed by a -0.8 is considered a flare. Figure 2 shows the GP fit and the real data points of a light curve along with the window detected as a flare by the software.

The specification of the step function and the choice of value for T were both based on experimentation on

a limited number of samples. Our light curves contain data in 2 bands, zg and zr. We treat each band separately and look for flares in either of them.

2.4. *Extraction of Physical Parameters of Detected flares*

Following the detection of flares using ESMA filters as described above, the following physical parameters were extracted from each of those flares, whenever possible.

1. Time taken for the rise from half value at peak to the peak.
2. Time taken for the decline from the peak to half the value at peak.
3. $g - r$ color at peak, along with error value.
4. Rate of change of $g-r$ colors for 5 timeseries points (10 days) following the peak, along with error values.

Parameter (2) could not be measured when the flare has not faded to half its value within the available data points. Parameters (3) and (4) could not be measured when the data contained only a single colour band. All these parameters were calculated from the GP fit of the light curve.

2.5. *Simulating Nuclear Flares to Test the Sensitivity of the Software.*

In order to test for the range of physical parameter values of flares to which the detection software is sensitive to, we created simulations of nuclear flares. To ensure the existence of observational gaps in the simulations, we used real ZTF curves and flattened them by converting each real data point into a simulated one drawn from a normal distribution with zero mean and standard deviation equal to the real error band. Then a model of Gaussian rise followed by an exponential decay with a randomly chosen flare peak epoch was added to the flat light curve to obtain the simulations. The light curve models used the following equations, as used in Yao et al. (2023):

$$L_\nu(t) = A_\nu \times \begin{cases} e^{(t-t_{\rm peak})^2/(2\sigma_{\rm rise}^2)} & t \le t_{\rm peak} \\ e^{(t-t_{\rm peak})/\tau_{\rm decay}} & t > t_{\rm peak} \end{cases} \quad (3a)$$

$$A_\nu = L_{\nu_0 {\rm peak}} \frac{B_\nu(T_0)}{B_{\nu_0}(T_0)} \quad (3b)$$

Here, $L_{\nu_0 {\rm peak}}$ is the peak luminosity in the reference band, which we chose to be the g-band ($\nu_0 = 6.3 \times 10^{14}$ Hz). Owing to spherical symmetry, the luminosity values in both sides of equation (3a) can be directly to peak flux values. The flare peak epoch $t_{\rm peak}$ was chosen from a uniform distribution of [0,1], indicating the fraction of the total time line of the light curve. Owing to spherical symmetry, the The timescale parameters, $\sigma_{\rm rise}$ and $\tau_{\rm decay}$, the peak temperature T_0, and the peak flux value in the reference band, $F_{\nu_0 {\rm peak}}$, were chosen from uniform distributions listed in Table 2. We generated 10,020 simulations of this model.

Parameter for flare simulation	Distribution
$F_{\nu_0 {\rm peak}}$	[0 μJy, 150 μJy]
$\sigma_{\rm rise}$, $\tau_{\rm decay}$	[2 days, 365 days]
$t_{\rm peak}$	[0, 1]
T_0	[10^3 K, 10^7 K]

Table 2. Uniform distributions from which parameter values were picked for simulating flares of Gaussian rise + exponential decay model. For $t_{\rm peak}$ the value refers to the fraction of the total timeline at which the flare peak is injected.

2.6. *Selection of Galaxies*

We selected galaxies with optical spectroscopy available in the Sloan Digital Sky Survey (SDSS), which might help in future studies. We queried for galaxy type objects with spectroscopical data and redshift < 0.1. This list was then cross matched with PANSTARRS to obtain the IDs of the retrieved galaxies in order to retrieve their light curves from ZTF. The final list contained 412,225 galaxies.

3. RESULTS

We tested our software on the 10,020 simulations created to understand the range of parameters in which it is able to detect flares. Figure 4 shows a summary of the software detection as scatter plots depicting the fraction of simulations detected in each bin, defined by the parameter interval.

With regards to $t_{\rm peak}$, the software was poor at detecting flares of very small (< 0.1) and of very large (> 0.95) flare peak epochs, as seen in the bottom right plot of Figure 4. This is a result of requiring an onset ($s_j > +0.8$) followed by a decline ($s_j < -0.8$) for a flare, which would elude flares with peaks at the very beginning or at the very end since they would have a less significant rise or decline (or even completely miss either the rise or decline if $t_{\rm peak} = 0$ or $t_{\rm peak} = 1$). If needed, this condition can be modified (by simply requiring only one of $s_j > +0.8$ or $s_j < -0.8$) in the software to enable the detection of flares with only a rising or a declining part as well.

In terms of sensitivity, there was much less distinction between $\sigma_{\rm rise}$ and $\tau_{\rm decay}$, since both of them are essen-

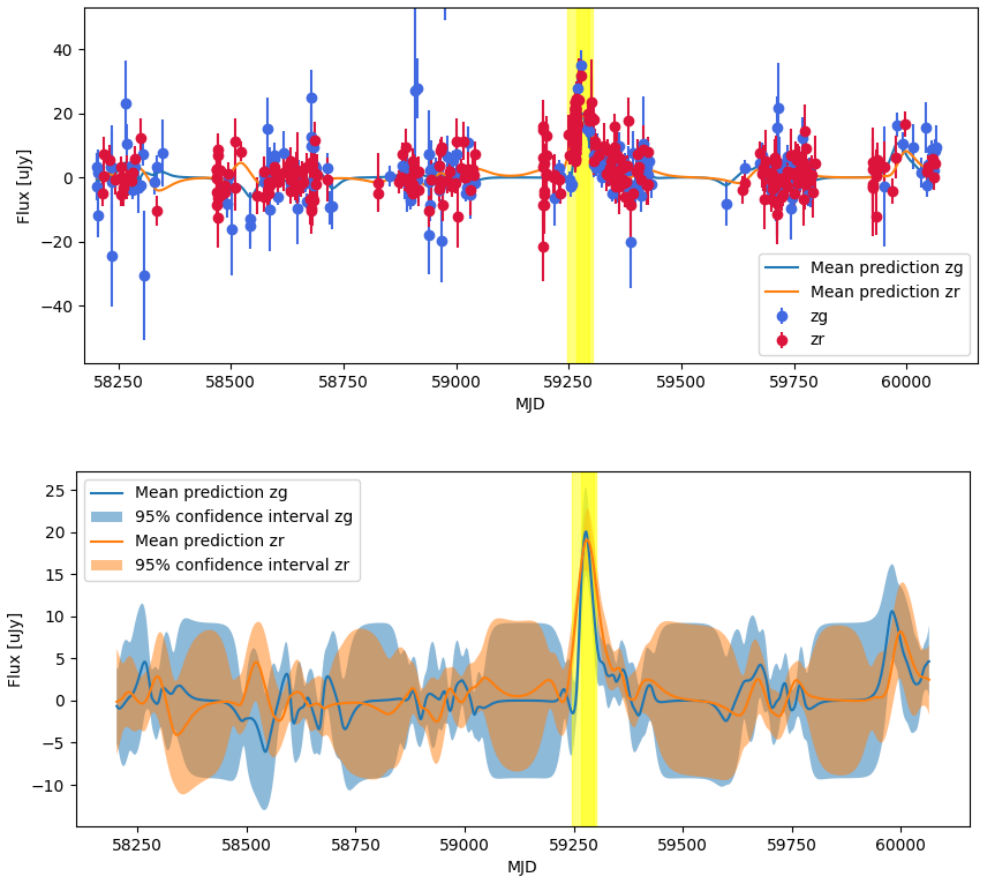

Figure 2. The datapoints (top) and the GP fit (bottom) of an example light curve that has been detected with a flare. The yellow coloured window depicts the time interval which begins with a smoothed value of +0.8 followed by -0.8.

tially timescale parameters defining the rise and the decline respectively. Increasing either of them results in a slower rise and decline respectively. The top right plot of Figure 4 shows the fraction of simulations detected in each bin of $F_{\nu_0\text{peak}}$ vs. τ_{decay}. It can be seen that τ_{decay} does not restrict the fraction of simulations detected by the software, for a given $F_{\nu_0\text{peak}}$. We obtained similar plots for T_0 vs. τ_{decay}, suggesting that our software is insensitive to the peak luminosity temperature, T_0, as well.

The parameter to which the software is most sensitive to is $F_{\nu_0\text{peak}}$. As can be seen in both the top and bottom right plots of Figure 4, there is a significant increase in the fraction of detections in each bin for simulations of higher peak flux in reference band (around $F_{\nu_0\text{peak}} > 50\mu\text{Jy}$). This is an expected result since a high $F_{\nu_0\text{peak}}$ value is very likely to yield see a contiguous set of positive values for x_j, followed by negative x_j's, thus satisfying the flare conditions.

Unfortunately, due to time constraints, we still have not run the software on the final list of galaxies to analyze them. We hope to execute this step in the upcoming weeks.

4. CONCLUSIONS

Fitting of light curves with GP regression followed by the usage of ESMA filters to detect for flares seem to be very responsive to flares with $F_{\nu_0\text{peak}}$ above a certain threshold and with the peak light epoch not around the very beginning or the end of the timeline. As can be seen from Equation (2) and Table 1, we basically need a contiguous increase in y_j's with small error bands (σ_j), for raw_j to be higher than the thresholds of the specified step function, and thus to smooth out s_j to be above +0.8. With the choice of $T = 10$ days, we can expect a minimum of around 13 days of contiguous $x_j = 1.0$ to smooth out to $s_j \geq 0.8$.

One thing to note is that most of the numerical values and equations in the detection software have been chosen through experimentation. This includes using a step function for mapping raw_j to [-1,1], the choice of threshold values in the step function, value of T, choice of raw_j, choice of thresholds for flare detection. With a

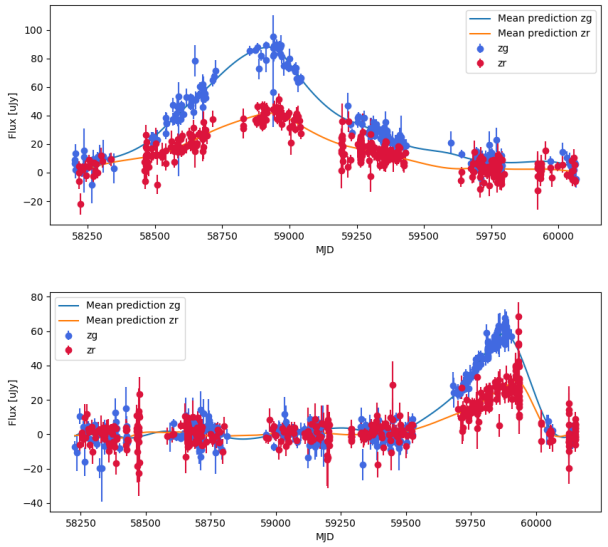

Figure 3. Two example simulations of nuclear flares. The parameter sets $(F_{\nu_0 \mathrm{peak}}, t_{\mathrm{peak}}, \sigma_{\mathrm{rise}}, \tau_{\mathrm{decay}}, T_0)$ for each model are as following. Top figure: (90 μJy, 0.398, 283 days, 319 days, 2.47×10^6 K), Bottom figure: (60 μJy, 0.876, 158 days, 36 days, 3.30×10^6 K). As given by the numbers and depicted in the figures, the top curve has a slower (longer time scale) rise and slower decline compared to the bottom curve. The flare peak epoch is earlier in the top figure and near the end in the bottom figure. Also, the top curve has a higher peak flux value in the reference band.

deeper mathematical analysis of ESMA filters or by experimenting with a larger data set, it is possible to build a more concrete algorithm which can be made sensitive for specified requirements.

Because our software was sensitive to detecting flares over a large timescale range, for peak flare flux above a certain threshold, it promises to able to detect both faster and slower TDEs. The threshold for peak flare flux can be adjusted by adjusting the specified step function, thus making it detect fainter flares as well. This marks an important progress towards building an algorithm that could detect TDEs that do not belong to the traditional definitions. We plan on running the code on the final list of galaxies in the upcoming weeks in anticipation of detecting interesting flares.

5. ACKNOWLEDGMENTS

Thank you to Professor Vikram Ravi, Assistant Professor of Astronomy at Caltech, and Jean Somalwar, graduate student in PMA at Caltech, for mentorship and guidance during my fellowship.

Thank you to the Flintridge Foundation for funding this project through the Caltech SURF Program.

REFERENCES

Charalampopoulos, P., Pursiainen, M., Leloudas, G., et al. 2023, A, 673, A95, doi: 10.1051/0004-6361/202245065

Hills, J. G. 1975, , 254, 295, doi: 10.1038/254295a0

McLaughlin, S. A. J., Mullaney, J. R., & Littlefair, S. P. 2024, , 529, 2877, doi: 10.1093/mnras/stae721

Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (The MIT Press)

Yao, Y., Ravi, V., Gezari, S., et al. 2023, ApJL, 955, L6, doi: 10.3847/2041-8213/acf216

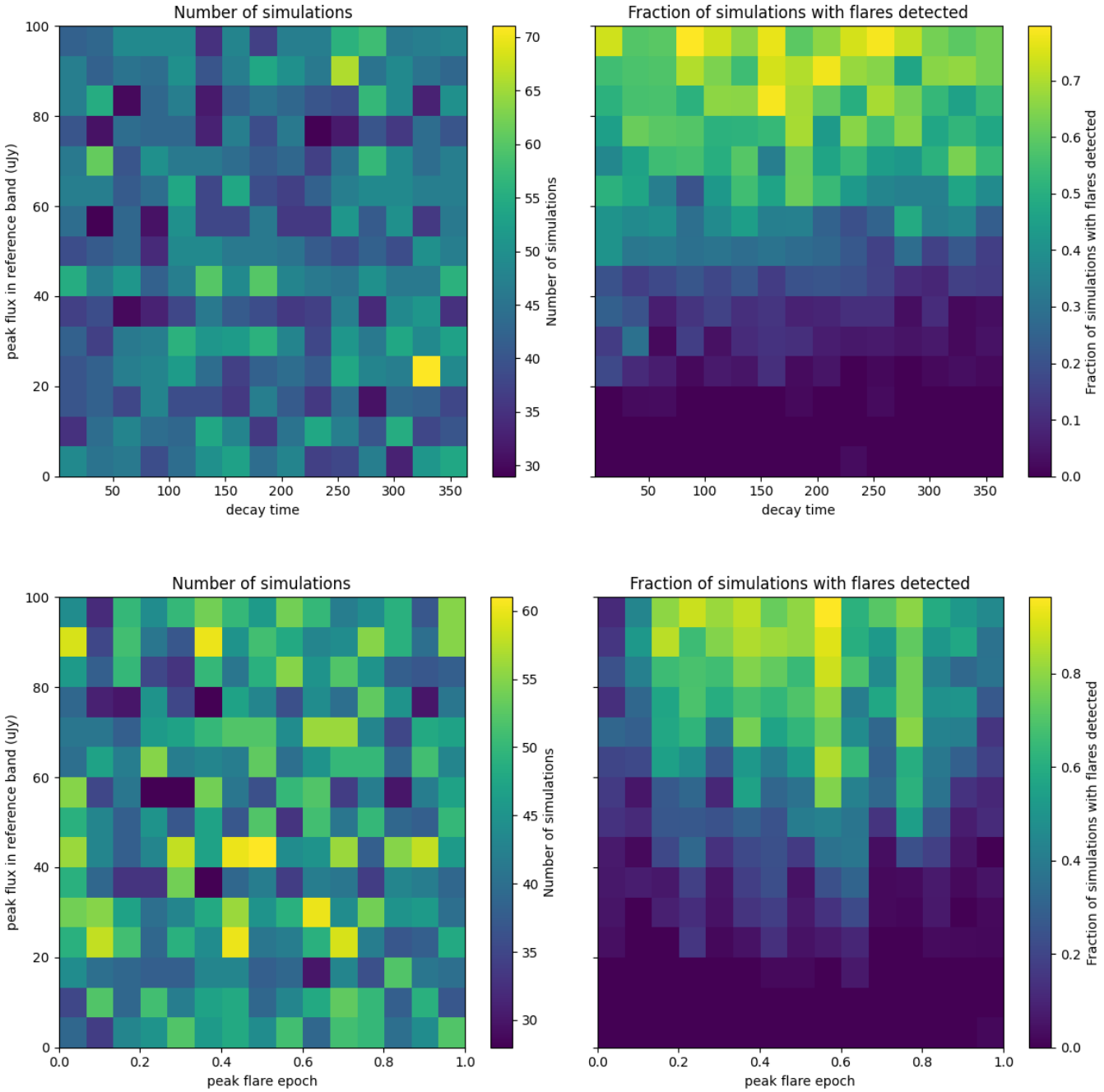

Figure 4. Left: Distribution of simulations created across different bins, plotted as $F_{\nu_0 \mathrm{peak}}$ vs. τ_{decay} (Top) and as $F_{\nu_0 \mathrm{peak}}$ vs. t_{peak} (Bottom). Right: The fraction of simulations flagged as consisting of a flare in the same bin of the corresponding left diagram.