

# The largest ground-based catalogue of M-dwarf flares

A. D. Lavrukhina,<sup>1,2\*</sup> B. Demkov,<sup>3</sup> K. Malanchev,<sup>4</sup> M. V. Pruzhinskaya<sup>1</sup> and E. E. O. Ishida<sup>5</sup>

<sup>1</sup>*Lomonosov Moscow State University, Sternberg astronomical institute, Universitetsky pr. 13, Moscow, 119234, Russia*

<sup>2</sup>*Lomonosov Moscow State University, Faculty of Physics, Leninskie Gory 1-2, Moscow, 119991, Russia*

<sup>3</sup>*Independent Researcher*

<sup>4</sup>*McWilliams Center for Cosmology & Astrophysics, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA*

<sup>5</sup>*Université Clermont Auvergne, CNRS, LPCA, Clermont-Ferrand, F-63000, France*

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

We present the largest ground-based catalogue of M-dwarf flares to date, comprising 1,229 time-resolved events identified from the Zwicky Transient Facility (ZTF) data release 17. Using high-cadence ZTF observations collected between April 2018 and September 2020, we analyzed over 93 million variable light curves containing 4.1 billion photometric measurements. Flare candidates were initially identified through a machine-learning pipeline trained on simulated flare light curves, generated by injecting TESS-based flare templates into ZTF observational data. The candidates were then processed through an extensive post-filtering pipeline combining a machine-learning classifier, additional metadata gathering, and human inspection. For 655 flares with reliable Gaia-based distance estimates and well-sampled light-curve profiles, we derived bolometric energies ranging from  $10^{31}$  to  $10^{35}$  erg. A clear correlation is observed between flare frequency and spectral subtype, with a sharp increase toward later M dwarfs, particularly near M4-M5, coinciding with the transition to full convection. Using the subset of 680 flaring stars with known vertical distances from the Galactic plane  $z$ , we estimated how the fraction of flaring stars varies with Galactic height and found a negative trend, indicating that the likelihood of flare activity decreases with increasing  $|z|$ . The resulting catalogue represents the most comprehensive ground-based sample of M-dwarf flares available and establishes a framework for flare detection and classification in upcoming wide-field surveys such as the Vera C. Rubin Observatory Legacy Survey of Space and Time.

**Key words:** stars: flare – stars: late-type – stars: activity – surveys – methods: data analysis

## 1 INTRODUCTION

M-dwarf flares are highly energetic transient phenomena characterized by rapid energy release, with typical bolometric energies ranging from  $\sim 10^{29}$  to  $10^{36}$  erg and durations spanning from minutes to several hours (Günther et al. 2020; Hawley et al. 2014). These events manifest across the entire electromagnetic spectrum, from radio to X-ray wavelengths, with the most dramatic enhancements typically observed in the UV. The physical origin of these flares is intimately linked to magnetic reconnection processes and dynamo mechanisms (Hawley & Pettersen 1991; Kowalski et al. 2013). The enhanced magnetic activity observed in M-dwarfs stems from their fully convective internal structure, which facilitates efficient magnetic field generation and amplification (Browning 2008; Kitchatinov et al. 2014). The study of M-dwarf flares consequently represents a critical intersection of stellar and planetary science, as frequent energetic outbursts may significantly impact the habitability of nearby exoplanets.

There is a widely supported hypothesis that young stars exhibit flares more frequently than the older ones (West et al. 2015; Newton et al. 2017). This can be attributed to the stronger magnetic fields generated by the rapid rotation of stars, which is characteristic of

young stellar objects. The idea that younger stars produce flares more often has strong backing from both observational data and theoretical understanding. However, in order to confirm this trend with higher statistical confidence, larger and more homogeneous flare samples across a broad range of stellar ages are still required.

The abilities of modern time-domain surveys has enabled systematic searches for stellar flares across a wide range of spectral types, with M dwarfs being of particular interest due to their strong magnetic activity. Space-based missions such as *Kepler* (Borucki et al. 2010) and the Transiting Exoplanet Survey Satellite (TESS; Ricker et al. 2014) provide continuous monitoring with high photometric precision, allowing the detection of both frequent small-amplitude flares and rare energetic events. For example, Yang & Liu (2019) reported  $\sim 1.6 \times 10^5$  flares on about 3400 stars from *Kepler* data, while Yang & Liu (2019) identified 8695 flares in the first TESS data release and Pietras et al. (2022) extended this number to  $\sim 1.4 \times 10^5$  flares over the first three years of the mission. CHEOPS high-cadence monitoring combined with wavelet-based denoising enabled the detection of 291 microflares with energies as low as  $10^{26}$  erg (Poyatos et al. 2025). More recently, Galletta, G. et al. (2025) presented a TESS-based catalog of 17,229 flares from nearby M dwarfs ( $d \leq 10$  pc), probing energies down to  $10^{29}$  erg and demonstrating the large diversity of flare occurrence rates among stars of similar luminosity.

Ground-based surveys complement the space missions by cover-

\* E-mail: lavrukhina.ad@gmail.com

ing larger sky areas and probing brighter, rarer flares. The Zwicky Transient Facility (ZTF; [Bellm et al. 2019](#)) has been employed for systematic flare searches, including machine-learning approaches based on active anomaly detection, which identified more than one hundred flares ([Voloshina et al. 2024](#)). Other surveys have also contributed valuable flare samples: hundreds of flares in Sloan Digital Sky Survey (SDSS; [York et al. 2000](#)) and Two Micron All Sky Survey (2MASS; [Skrutskie et al. 2006](#)) data with maximum  $u$ -band amplitudes up to  $\sim 4.5$  mag ([Davenport et al. 2012](#)), 62 flares up to  $\sim 2$  mag in  $V$ -band in the ASAS-SN M-dwarf catalog ([Rodríguez Martínez et al. 2020](#)), 96 flares from DECam data ([Webb et al. 2021](#)), 22 fast flares from the Tomo-e Gozen project with on e-second cadence ([Aizawa et al. 2022](#)), and 132 flares in the TMTS survey ([Liu et al. 2023](#)).

The remaining of this paper is organized as follows. Section 2 describes the ZTF data release (DR) 17 dataset, the high-cadence subset, and all preprocessing steps. Section 3 details the flare-search pipeline, including cadence and variability-based pre-filtering (Section 3.1), 3.2 simulating process of flares built from TESS templates for training, 3.3 feature extraction and classification with Random Forest and CatBoost combined via a blending ensemble, and 3.4 post-filtering using Minor Planet Checker<sup>1</sup> (MPChecker), image-quality metrics, and Pan-STARRS1 (PS1) colors. Section 4 presents the resulting catalog of M-dwarf flares, statistics across spectral subtypes and Galactic height, energy – duration estimates and their correlations. Finally, the Appendix provide the full list of expert-crafted features.

## 2 DATA

The Zwicky Transient Facility (ZTF) is a wide-field sky survey conducted with the 48-inch Schmidt-type Samuel Oschin Telescope at Palomar Observatory ([Bellm et al. 2019](#)). It has a field of view of 47 square degrees observed in the  $g$ ,  $r$ , and  $i$  passbands, with a typical 30-second exposure reaching a median  $r$ -magnitude limit of approximately 20.6 mag. In phase I (March 2018 – Sept 2020), ZTF observed the northern sky with a three-day cadence and the Galactic plane with a one-day cadence. In phase II (Oct 2020 – Sept 2023), half of the ZTF telescope time is dedicated to a uniform two-day cadence public survey in the  $g$ - and  $r$ - bands. Data from this public survey are released every two months as data releases.

The ZTF DR photometric pipeline performs source extraction for individual frames and then cross-matches these sources across all frames within a single pair of observation field/CCD quadrant and passband. This process leads to a single sky source potentially being represented by multiple ZTF DR objects due to overlapping observation fields and the three passbands used.

In this work we use public and private survey data from ZTF DR17 as target data sets for searching for M-dwarf flares. This release covers the time span from March 2018 to March 2023. The DR 17 source catalog files contain 756 billion source detections, and 4.66 billion light curves derived by cross-matching the single-exposure data to the ZTF reference catalog.

M-dwarf flares are rapid events, requiring a high observational cadence to detect and accurately capture their flare profiles. The specialized high-cadence Galactic plane survey, conducted by [Kupfer et al. 2021](#) with a 40-second cadence, is particularly well-suited for this purpose. All observations of the high-cadence Galactic plane survey were obtained in  $r$ -band. Unfortunately, the bulk-downloadable

data releases do not include any metadata indicating the specific survey to which individual observations belong. As a result, we implemented a specialized filtering process based on the number of observations, the time delay between consecutive observations, and the total duration of the corresponding segment of the light curve (for further details, see Section 3.1). The outcome of this process is a dataset of light curve segments that satisfies the filtering criteria. Since multiple segments from the same light curve may meet these criteria, it is possible for more than one segment to share the same ZTF object identifier (OID). To differentiate between them, each light curve segment is characterized by both its ZTF OID and the first observation's heliocentric modified Julian date (HMJD).

## 3 SEARCH PIPELINE

Usually, M-dwarf flare light curves have single- or multi-modal fast-rise exponential-decay profiles. These profiles can be detected with numerical fitting, especially in the cases of high signal-to-noise ratio and good observational cadence.

Even in these cases some false positive detections are still possible. For example, if time series are not long enough to also include some observations before and after the flare. In such cases, other types of variability may produce profiles which would have a satisfactory model fit, for example, a single period of a fast-pulsating star (see Section 3.5). This type of false positives may be reduced by using a non-parametric light-curve shape model; in this paper, we use machine learning algorithms. Machine learning models may also help to detect flares with unusual shapes, reducing false-negative rate if examples of such shapes are given in the training set.

Despite these improvements, in the output of our machine learning models, we found many types of light curves causing false positive detections: image defocusing issues, asteroid occultations, etc. Such problems can be solved best by fetching additional data from external services, which is not feasible for the whole input data set. Even after all machine learning and metadata filters, expert analysis of candidate flares is required. This leads us to a "funnel" pipeline where each step reduces the number of candidates to an amount feasible for the next step.

Here is the outline of this pipeline:

- Section 3.1: ZTF light-curve data are prefiltered based on cadence and variability.
- Section 3.2 and Section 3.3: classification model selects candidates based on their light curves.
- Section 3.4: filtering of selected candidates based on the image quality, external data sources, possible asteroid occultations, and stellar colours.
- Section 3.5: expert analysis of the candidates.

We made use of computational cloud resources for data pre-processing, training, prediction, and post-filtering. A computational node with 32 CPU cores and 256 GB of RAM was used for the following tasks: data pre-processing took 19 hours, model training took 4 hours, and prediction took 26 hours.

### 3.1 Pre-filtering

The first step in the search pipeline is data selection, which is analogous to what we have done in [Voloshina et al. \(2024\)](#). Since we are specifically looking for stellar flares, our focus is limited to variable light curves. To form a dataset of variable light curves, we used a

<sup>1</sup> <https://www.minorplanetcenter.net/cgi-bin/checkmp.cgi>

reduced  $\chi^2$  statistics for the constant magnitude model (number of parameters  $d = 1$ ):

$$\text{reduced } \chi^2 \equiv \frac{1}{N-d} \sum_i \left( \frac{m_i - \bar{m}}{\delta_i} \right)^2, \quad (1)$$

where  $m_i$  and  $\delta_i$  are magnitude and its observational error,  $\bar{m}$  is the weighted mean magnitude, and  $N$  is the number of observations. The pre-filtered dataset consists of light curves with a value of the reduced  $\chi^2$  statistics greater than 3.

The second goal of pre-filtering was to extract observations made during the high-cadence campaign. In the absence of image-processing pipeline meta-information in the bulk-downloadable ZTF DRs, we selected the light curves for our final sample based on the following criteria:  $r$ -band observations only; at least 10 observations per light curve; at most 30 minute delay between two subsequent observations; total duration is at least 30 minutes.

The final number of light curves after application of all described criteria is 93,660,131, the total number of detections is 4,145,625,951, the mean number of observations per light curve is  $\sim 44$ . The data spans from April 2018 to September 2020, which indicates a probable ZTF observation strategy change in ZTF Phase II. It is important to mark that in the final sample we can have several light curves which relates to the same object, since during our procedure we segmentate all such light curves to the pieces which satisfy described criteria. The total number of unique ZTF DR object identifiers is 54,800,806, but the actual number of sources may be a little lower due to observational field overlaps.

### 3.2 Flare Simulation

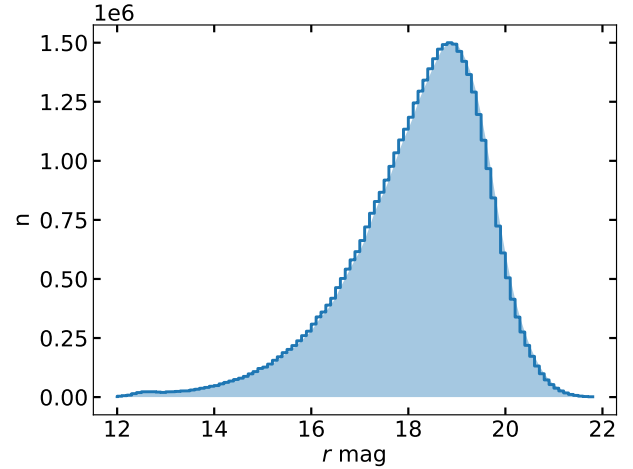
The absence of a comprehensive sample of M-dwarf flares, detected by ZTF, and sufficient for training a ML model, leads to the necessity of flare simulations. In this section, we describe a process of flares simulation based on M-dwarf flares detected in the TESS survey data (Davenport et al. 2014). The proposed method of flares simulation is based on utilizing TESS survey data as a starting point and combining it with time grids and observational magnitudes and errors from ZTF data. We prefiltered TESS light curves based on relative amplitude and intrinsic noise. A total of 436 unique TESS light curves were utilized as templates. Multiple flares were detected in some of these curves, resulting in 769 extracted flare segments (templates) that were subsequently used for further simulation. The full pipeline of flares simulation consists of further steps:

(i) Original light curves from the TESS survey are in fluxes, so as a first step for the initial template light curve all observations and observational errors were transformed from fluxes to magnitude. The magnitude light curve is linearly interpolated to produce a magnitude template function of time,  $f(t)$ .

(ii) For each simulated flare, we randomly select the ZTF observational field from the distribution based on all high-cadence light curves.

(iii) We randomly select a time-series from all ZTF light curves belonging to high cadence observations. Here and below, the high-cadence observations are selected by the same criteria as listed in Sect. 3.1, but without  $\chi^2$ -based filtering.

(iv) The peak time  $t_0$  of the simulated flare is selected randomly between the beginning and the end of the time series. This allows us to simulate different situations of flare detection (e.g. a full light curve, only descending or only ascending part) and is used partially as a data augmentation method together with selecting different time grids.



**Figure 1.** Distribution of magnitudes in  $r$ -passband from ZTF DR17 high-cadence data.

(v) In order to make the stellar magnitude distribution realistic, we sampled the quiescent magnitude  $m_0$  from the empirical distribution presented in ZTF high-cadence  $r$ -band observations, see Fig. 1. The result “perfect” (noise-free) magnitude is a sum of the quiescent magnitude and the time-dependent template magnitude  $f(t - t_0)$ .

(vi) The next step involved generating the observational magnitude uncertainty. We assume that the noise amplitude is distributed conditionally over the source magnitude and observational conditions, which both vary over time. For this reason we sample noise amplitude  $\sigma_i$  from an empirical distribution of ZTF errors, measured on the same exposure (ZTF expid) and CCD quadrant (fieldid & rcid) as the simulated observation. This empirical distribution is also conditional on magnitude: we use linear grid with 50 bins between maximal and minimal value of observed mag to bin observations.

(vii) As a final step, we simulate observational magnitude as a Gaussian random variable based on the “perfect” magnitude and the noise amplitude:

$$m_i \sim N(f(t_i - t_0) + m_0, \sigma_i^2),$$

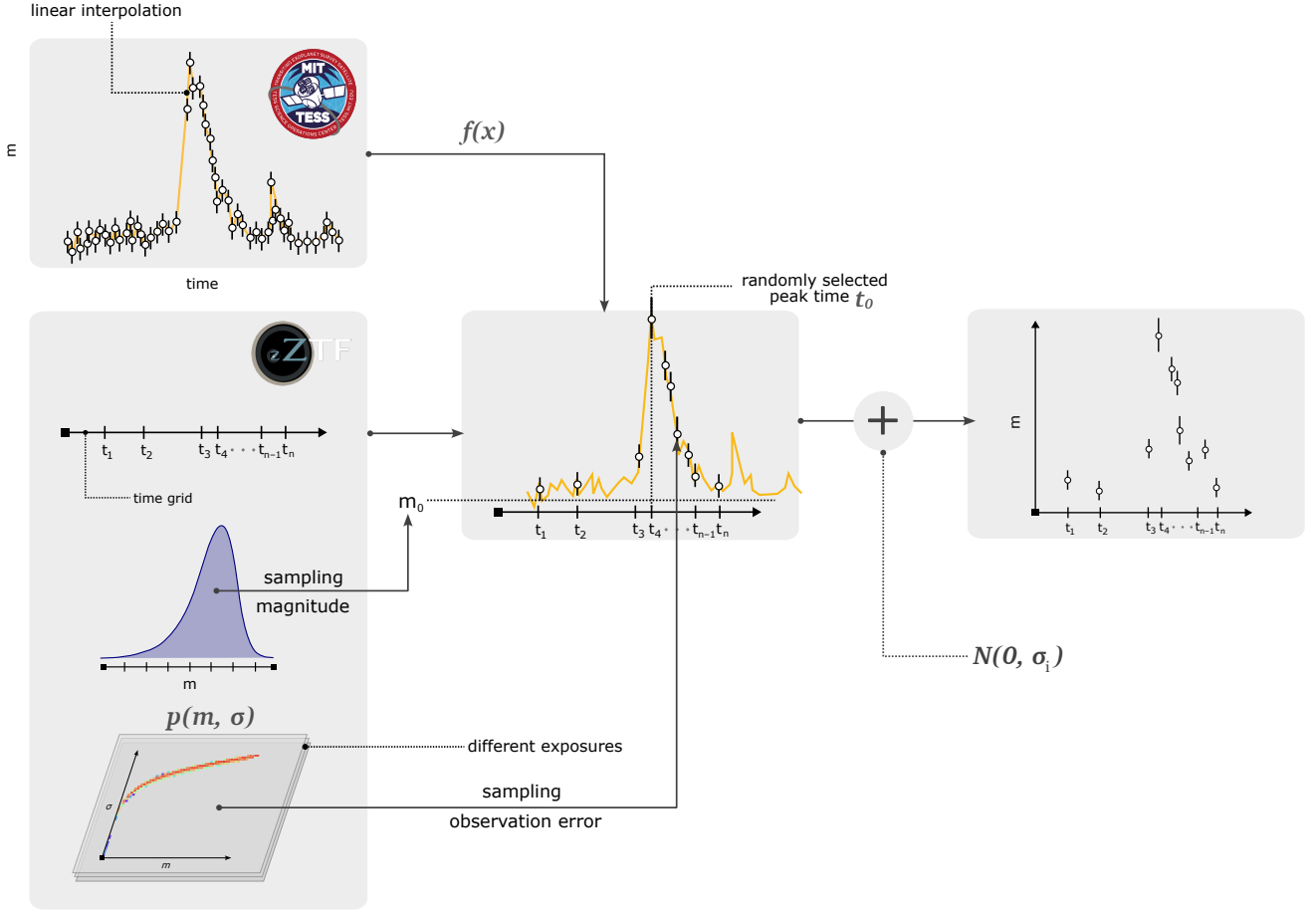
where  $m_0$  is the magnitude of the quiescent star,  $m_i$  is the observed magnitude,  $f(x)$  is the linear interpolation of the TESS template,  $t_i$  is the time of the observation,  $t_0$  is the shift due to the selection of peak time, and  $\sigma_i$  is the observation error.<sup>2</sup>

The simulation pipeline is sketched as a flow chart in Fig. 2.

### 3.3 Light-curve classification

We developed a machine learning model for selecting flaring candidates from the pre-filtered dataset. For model training, we generated

<sup>2</sup> Our exploration of the ZTF data showed that Gaussian distribution with provided observational error describes instrumental noise pretty well for  $m \gtrsim 15$  mag. We tested it with the empirical distribution of the relative residual from the mean object magnitude  $(m_i - \bar{m})/\sigma_i$ . The distribution was found to be sufficiently close to the standard normal distribution.



**Figure 2.** Flowchart of the flare light-curve simulation process. Original TESS flare templates are converted from flux to magnitude, interpolated, and injected into randomly selected ZTF high-cadence time grids with realistic noise and brightness distributions. Further details of the simulation procedure are provided in Section 3.2.

620 755 flaring curves by the procedure described in the previous section; these light curves have “positive” labels. For the “negative” class we randomly selected 620 755 from  $\sim 94 \times 10^6$  objects of the target dataset described in Section 3.1, assuming that the chance of accidentally including a flare light curve is negligible. The result labeled dataset was split into training and test datasets in 3:1 ratio.

Since our simulation pipeline may produce some light-curve features unique to artificial light curves, the machine-learning models we train may become biased toward using these features for binary classification instead of, or in addition to, those describing stellar flares. If this occurs, it could lead to less effective classification, as the models would not encounter these artificial features in real data and would assign lower scores to real flares. To address this, we used an additional test set consisting of 99 ZTF light curves identified by Voloshina et al. (2024). After training and validation of the models using the datasets described above, we perform a final quality test on this small dataset to make the final decision on model selection.

Next, in this section, we provide a detailed examination of the feature extraction process, light-curve classification models used, and metrics applied.

### 3.3.1 Feature extraction

Since number of detections, duration, and cadence of light curves are very diverse over the sample, we pre-process them before applying machine learning analysis. We extract two different classes of features: 33 expert-selected features, and 9,996 MiniRocket features. Expert-selected features are applied to the original light curves using `light-curve`<sup>3</sup> package, the inputs are arrays of time, magnitude and magnitude errors (Malanchev 2021). The full list of expert-selected features is given by Appendix A.

The second feature set is obtained using MiniRocket convolution algorithm (Dempster et al. 2021). First, light curves are interpolated with a third-order Akima spline to a uniform time grid of 100 points, spanning from the minimum to the maximum observational time, using the `scipy` package (Virtanen et al. 2020). The interpolated light curves are normalized with the standard scaler: magnitude values are scaled and shifted so the mean is zero and the variance is unity. MiniRocket employs a set of 84 predefined convolution kernels of size nine, with kernel values of either 2 or  $-1$ . The interpolated data is convolved with these kernels, and the resulting features are quartiles of the convolution outputs.

To reduce the number of highly correlated MiniRocket features

<sup>3</sup> <https://github.com/light-curve/light-curve-python>



and balance their contribution with expert-selected features, we apply principal component analysis (PCA) to reduce the dimensionality from 9,996 to 47. The number of PCA components is chosen such that 90% of the total data variance is explained by the PCA model.

As the result, the total of 80 features are extracted for each light curve and machine learning models are applied to them.

### 3.3.2 Classification models

The input feature vectors are high-dimensional and may have highly correlated components due to the residual correlation of PCA-transformed MiniRocket features and the correlation between some of the expert-selected features (see Figure A3 of Malanchev et al. 2021). This motivated us to use decision-tree ensemble models for the machine-learning classification pipeline, as they are typically more robust to feature correlation. Another advantage of decision-tree models is that their hyperparameters can be more easily configured to prevent overfitting, which was a significant risk in this research due to the use of both simulated and real data in the training dataset.

We trained two models using the same training dataset: Random Forest (Ho 1995) and CatBoost (Dorogush et al. 2018). We used the `scikit-learn` (Pedregosa et al. 2011) implementation of Random Forest, setting the following parameters to non-default values: the number of trees to 100 and the tree depth to 36 (the last was selected by the grid-based hyperparameter tuning). For CatBoost, we set the learning rate to 0.01, the tree depth to 10, and the maximum number of iterations to 5000. Additionally, we monitored the learning curve progress every 10 iterations for an early termination. We set 99:1 true-to-false class weights for both algorithms. Both models output scores are calibrated to much probability of flare class in the validation set.

We composed the models to create a third ensemble model, Blender. Blender aggregates outputs from both Random Forest and CatBoost and yields the minimum calibrated score. This approach makes the model more strict to false positives: for a given threshold score the object is classified as a flare, only if both models classified it so.

Table 1 shows the performance of the models according to the following metrics:

- precision: ratio of correctly labeled flares to all objects labeled as flares for the default classification score threshold value;
- recall: ratio of the correctly labeled flares to all flares for the default threshold value;
- F1 score is a harmonic mean of the precision and the recall;
- average precision score (AP): integral under the precision–recall curve, see Fig. 3.

The default threshold for test set was 0.98, so the final classifier data output would have feasible size to look through.

Table 1 shows that the Blender model never shows the best performance by any metrics. However, we found that the Blender model shows better performance for our particular use-case. Figure 3 shows that the Blender model shows better dynamics for small recall / large precision region, which is what we are targeting.

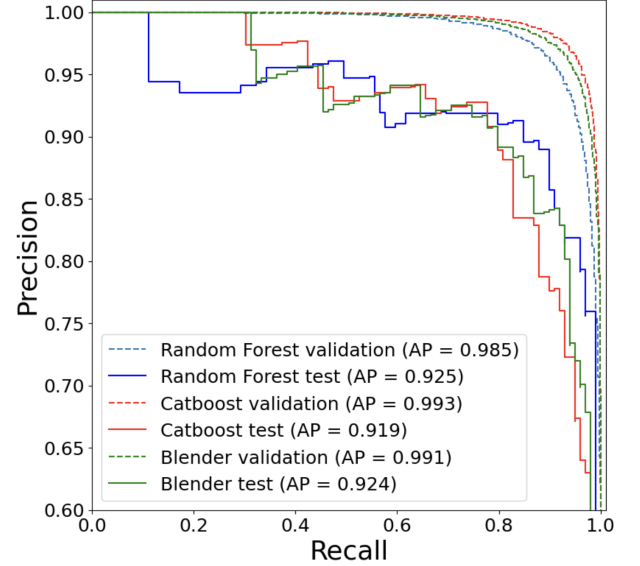
### 3.3.3 Early experiments and model selection

After we trained the initial version of Random Forest and CatBoost, we made an experiment of applying them to the target dataset. To get an estimation of efficiency of the models on the target dataset, we evaluated them and, for each model, randomly selected a hundred

**Table 1.** Evaluation metrics on validation and test sets (classification threshold 0.98).

Metric	Random Forest		CatBoost		Blender	
	Validation	Test	Validation	Test	Validation	Test
Precision	0.997	<b>0.961</b>	0.997	0.929	0.999	0.952
Recall	0.622	0.495	0.732	<b>0.525</b>	0.613	0.404
F1 score	0.766	0.653	0.844	<b>0.671</b>	0.759	0.567
Average precision	0.985	<b>0.925</b>	0.993	0.919	0.991	0.924

Bold indicates the best *test* value per metric.



**Figure 3.** Precision Recall curve for validation and test datasets.

objects classified as flares. We visually classified these small datasets and made two observations: precision of the models are much smaller than what we see in Table 1, and that a lot of mis-classification are from the asteroid occultation. For the final estimate of precision for this experiment we drop all asteroid occultations, since they are easy to filter out on the post-filtering step (see Section 3.4.1). The precision for the CatBoost was found to be 0.030, for Random Forest 0.054. Those precision values would not allow us to efficiently classify the objects, so we decided to make an ensemble model. We tried both mean-score value (precision 0.052) and minimum-score value (precision 0.077). We chose to use the latest model, Blender, which was described by the previous section.

We re-run this experiment multiple times when changing input dataset (we initially used ZTF DR14) and collected a dataset of both flares and non-flares. Since we also found that a lot of cases of mis-classification are caused by the various photometric pipeline problems, we used this small labeled dataset to built an ML classification model for the post-filtering step.

### 3.4 Post-filtering

Early experiments showed that among the candidates the Blender classifier mis-classifies there are different astronomical phenomena, observational and pipeline artifacts: periodic variables, asteroid occultations, defocusing in crowded fields, image artifacts, etc (see Section 3.5 for more details). To improve the efficiency of the manual classification, we apply three post-filtering stages to the candidates selected by the light-curve classification model.

### 3.4.1 Asteroid occultations

First, we check for nearby minor planets using the MPCChecker service. We set search radius to be 15 arcseconds and query for time of the maximum brightness. The number of detected signals from known asteroids is 7582, which  $\sim 15\%$  of the total number of candidates given by the light-curve classifier.

Since this step required access to the MPCChecker service, it may not be done in the beginning on the whole target dataset of dozen millions of objects.

### 3.4.2 Classification model based on image quality metrics

A lot of the candidates come from crowded Galactic fields, where photometry of stars can be contaminated by nearby objects if the image seeing changes between observations. We have already detected such kind of bogus variability in previous researches, [Malanchev et al. \(2021\)](#) gave an example of a seeing change causing stochastically-looking behavior of a light curve, and [Voloshina et al. \(2024\)](#) demonstrated the contamination of flare-like light-curves with these kind of artifacts. That made us believe that if brightening is correlated with seeing change then the detection is likely to be bogus ([Krushinsky et al. 2025](#)). Since this correlation can be tricky to simulate, we built a machine-learning algorithm on a few observation metadata features, and trained the model on data labeled by [Voloshina et al. \(2024\)](#) and Section 3.3.3.

We chose to use point spread function full width on the half maximum<sup>4</sup> and sharpness of the image source. We used a linear regression binary classifier based on six features: seeing and sharpness at the first observation and the maximum light, and the difference between these two seeing and sharpness values. All features were scaled to have zero mean and unity variance.

These exposure metadata are not available in the bulk-downloadable ZTF DRs, so we query the IRSA light-curve API<sup>5</sup> for each specific object. For training we used 607 non-flare light-curves and 120 flare light-curves. 2617 candidates for which the metadata was not available were dropped from the target sample.

Using a threshold of positive class probability equal to 0.9998, 4763 candidates were classifier as flare candidates and proceed to the next post-filtering step.

### 3.4.3 Color-based M-dwarf selection

During the visual analysis of the candidates we found that objects which are bluer than M dwarfs in quiescence never show real flare activity. Since we had more than 4 thousand candidates to go, we decided to make focus of this paper on M-dwarf flares only. Using Pan-STARRS1 DR2 colors we kept all objects with  $r - i$  color larger than 0.42.

## 3.5 Visual inspection

During the process of candidates analysis obtained after application of the classification model and the postfiltering we visually scrutinized light curves to sort out such misclassifications:

<sup>4</sup> We call this value “seeing” in this section to avoid a possible confusion with full the width on the half maximum of the light-curve shape we introduce in Section 4.3.

<sup>5</sup> [https://irsa.ipac.caltech.edu/docs/program\\_interface/ztf\\_lightcurve\\_api.html](https://irsa.ipac.caltech.edu/docs/program_interface/ztf_lightcurve_api.html)

- **Periodic variable stars.** Figure 4 presents a representative example of a misclassified light curve. Such misclassifications are primarily caused by limitations in the observing cadence in combination with the intrinsic periodicity of the source. Due to the sparse and irregular sampling of the light curves, only short segments of the periodic variability are observed. These segments can mimic the shape of a flare, particularly when they capture a rapid rise and decay phase. Consequently, in the absence of a full-phase light curve, it becomes difficult to distinguish periodic variables from genuine flaring events based solely on partial observational data.

- **Photometric artefacts.** An example of misclassified light curves caused by photometric artifacts is shown in Fig. 6. Instrumental effects can produce short-lived apparent brightenings with the same asymmetric morphology – fast rise followed by a slower decay – typical of stellar flares. For a broader overview and a publicly available dataset of such artefacts in ZTF DR3, see [Sreejith et al. \(2026\)](#). In our data the main contributors are:

- **Defocusing** (see Fig. 6a): overlapping PSFs of nearby stars (blending) broaden and reshape the point-spread function. Small changes in seeing or focus can temporarily increase the contribution of neighbouring sources to the target aperture, producing an apparent impulsive brightening followed by a gradual return, thus mimicking a flare-like rise-decay profile.

- **Ghosts** (see Fig. 6b): ghosts, caused by internal reflections within lenses, filters, or their barrels, often appear as a diffuse smudge around the target; their motion into and out of the aperture creates transient brightenings with asymmetric morphology.

- **Frame edge** (see Fig. 6c): frame edge effects occur when the target lies near the detector boundary, where small pointing shifts can move part of the PSF in or out of the frame or aperture, producing step-like or spike-like flux changes that can mimic flares.

- **Symmetric light curves.** We also found a light curve with a nearly symmetric shape ZTF OID 805205100020986, whose morphology is more consistent with alternative phenomena e.g. self-lensing binaries ([Crossland et al. 2024](#)) than with a genuine flare.

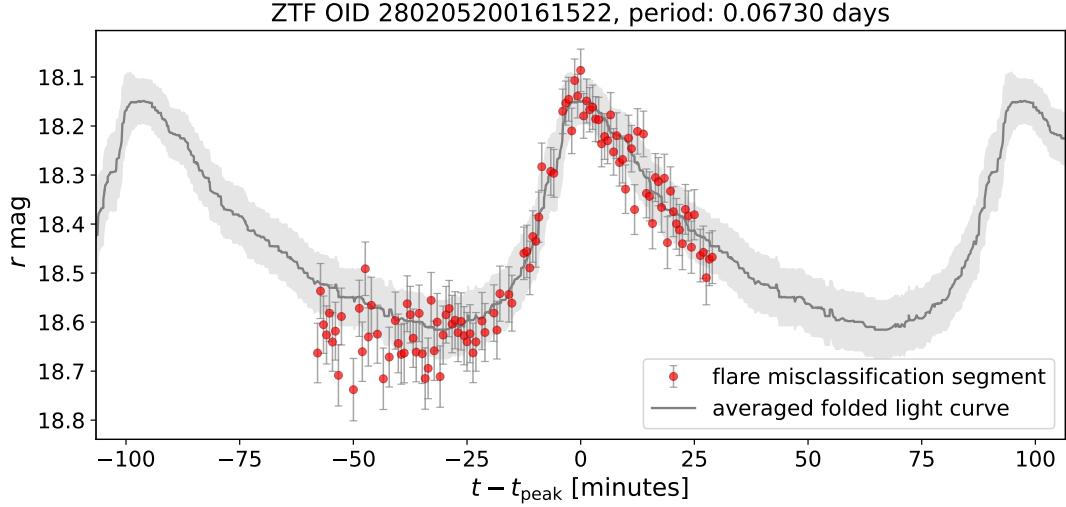
All described misclassifications arise because, in the absence of complementary information, the light-curve morphology alone is often insufficient to reliably distinguish genuine stellar flares from unrelated periodic variability, instrumental artefacts, or other transient phenomena.

---

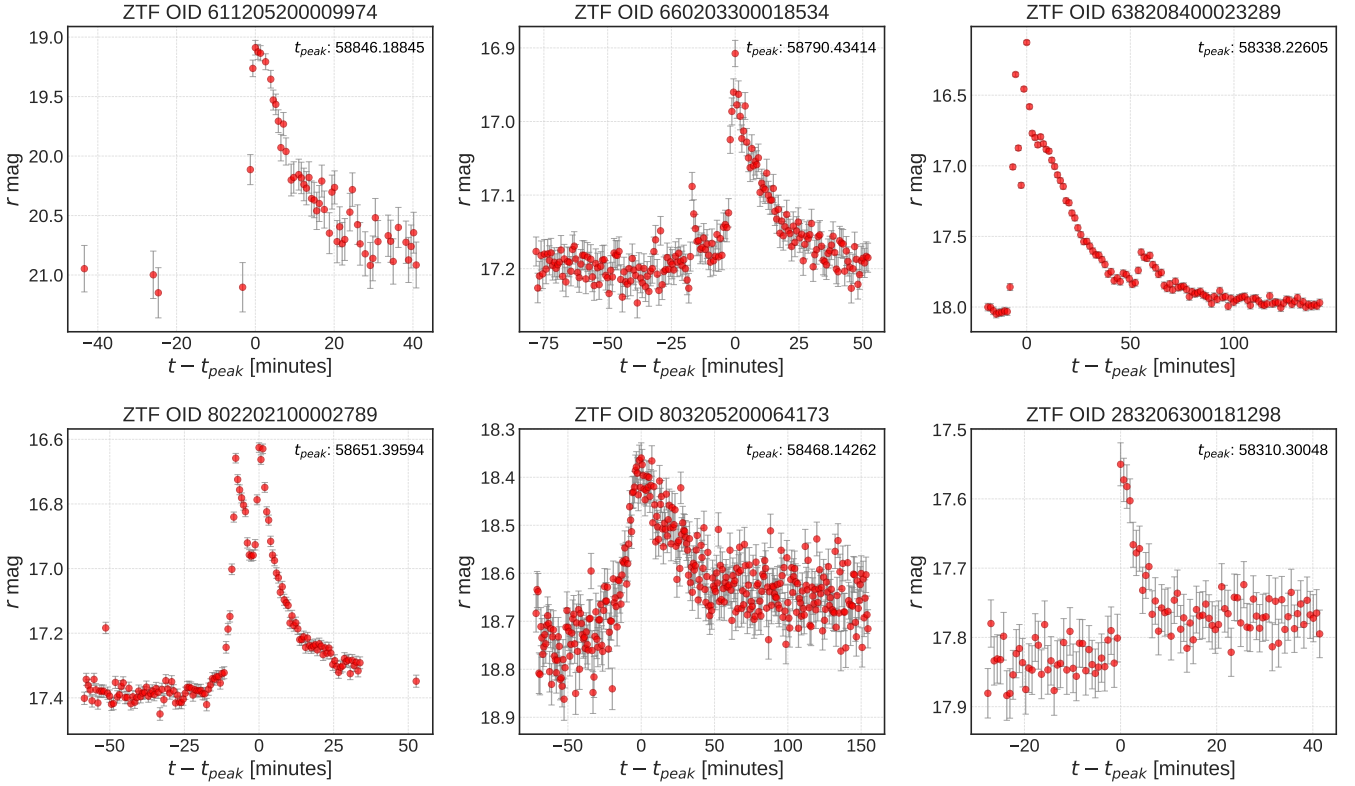
After applying the described pipeline, we obtained 2 898 candidates, and classified 1235 of them as M-dwarf flares. Six of these flares were observed in two ZTF fields and appeared as two different light curves in our dataset. This changes the total number of unique flares to 1229.

## 4 RESULTS

In this section we analyze the population of flaring stars and physical properties of flares. We build an M-dwarf flare catalog (Section 4.1). We use this catalog to estimate a conditional distribution of flare frequency over spectral sub-classes (Section 4.2) and vertical distance from the Galactic plane (Section 4.4). Finally, we provide various physical parameters based on flare light curves, such as flare duration and its total energy (Section 4.3).



**Figure 4.** Light curve of a  $\delta$ -Scuti star misclassified by the classification pipeline. The grey line shows the period-folded light curve of all observations, smoothed with a median filter. The light grey shaded area indicates the average deviation. Red points highlight high-cadence observations only, which were incorrectly classified as a stellar flare by our classification algorithm.



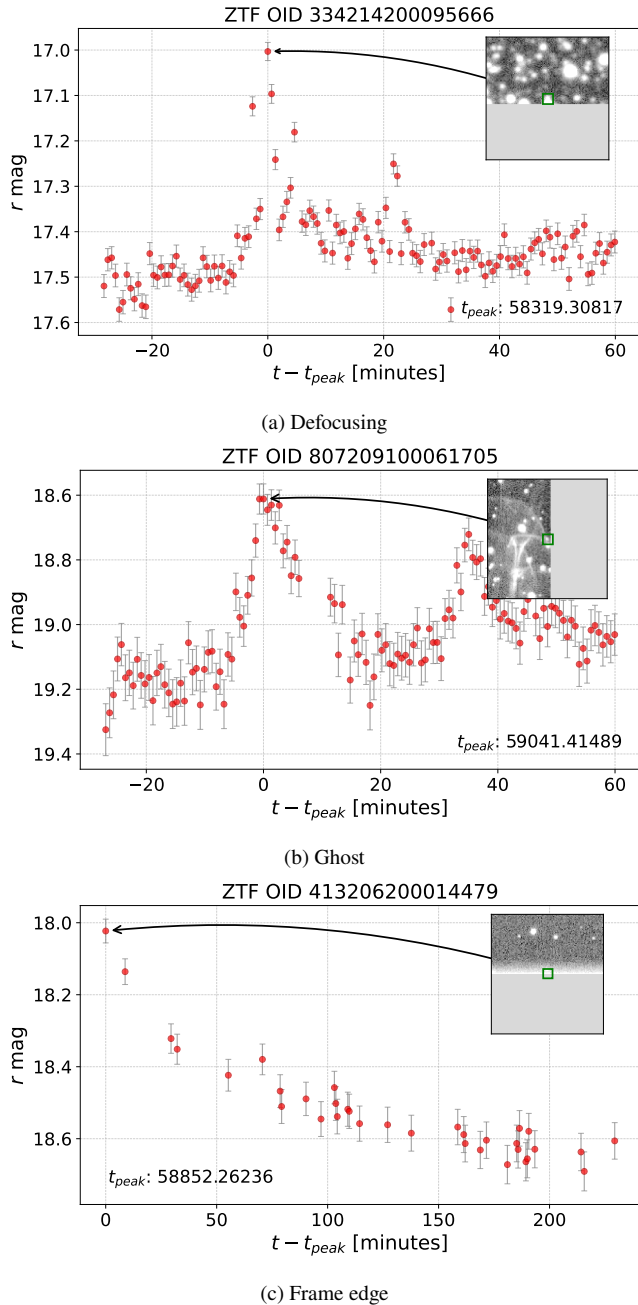
**Figure 5.** Representative examples of found stellar flare activity in M dwarf stars. The light curves demonstrate the diverse morphology of flare events, ranging from simple single-peak flares to complex multi-peak structures, with amplitudes varying from subtle brightness increases to dramatic outbursts exceeding several magnitudes.

#### 4.1 M-dwarf population in ZTF

We selected a dataset of M-dwarf candidates from the entire ZTF DR17 data set in the r-band using Gaia and Pan-STARRS data using

LSDB<sup>6</sup> package (Caplar et al. 2025). First, we cross-matched all ZTF DR17 r-band objects ( $\sim 2.2$  billion) with Pan-STARRS1 DR2 object catalog, selecting the nearest neighbor within 1 arcsecond.

<sup>6</sup> <https://lsdb.io>



**Figure 6.** Photometric artefacts that mimic stellar flares. The panels show short-duration, asymmetric brightenings – fast rise followed by slower decay – produced by instrumental/systematic effects: defocusing and blending of neighbouring PSFs (6a); ghosts from internal reflections (6b); and proximity to the frame edge (6c). Such events are often misclassified by our classification pipeline as stellar flares; for a broader overview and publicly available ZTF DR3 dataset of such artefacts, see Sreejith et al. (2026).

Following Kowalski et al. 2009 we filtered by  $(r - i) > 0.42$  and  $(i - z) > 0.23$  Pan-STARRS colors and magnitude errors in these bands to be smaller than 0.1. We use this sample for color-based spectral classification in Section 4.2. We further refer to this catalog as PS1-matched, it contains 880,640,576 objects.

Next, we cross-matched PS1-matched with Gaia DR3 catalog, again selecting the nearest neighbor within 1 arcsecond. We filter Gaia objects to have an upper limit of effective temperature (based

on RP/BP spectra modelling) to be below 3800 kelvins, keeping M and later classes only (Rajpurohit et al. 2013). For the further analysis we also need access to distances, so we join the result catalog with Gaia Early DR3 distances catalog (Bailer-Jones et al. 2021). Since we rely on distances in further analysis we filter the catalog to have parallax to parallax error ratio greater than three. This catalog is much smaller than the previous one and contains 27,420,797 objects. We further refer to this catalog as Gaia-matched.

## 4.2 Spectral classification

The primary analysis is conducted using the effective temperature ( $T_{\text{eff}}$ ) from Gaia. In the final Table 2, we provide the spectral classification based on Gaia  $T_{\text{eff}}$  when available. If Gaia data is not available, we present the spectral subclass derived from PS1 photometry, without applying any dust correction.

We used two methods to evaluate a spectral subclass for two different subsets of flares. The first subset consists of the stars from Gaia-matched with available effective temperature, thus we are using ratios from Malkov et al. (2020) to define a star’s spectral subclass. The total amount of flaring stars in this sample – 680 (we drop recurrent flares from the same source and take into account only total number of flaring stars). The second one consists of the stars from PS1-matched sample with available photometric data (the total amount of flaring stars – 1212).

To estimate the spectral subclass of each star, we use the probabilistic classification approach introduced by Kowalski et al. (2009), which models the distribution of M-dwarfs in the  $(r - i, i - z)$  colour-colour diagram as a set of two-dimensional Gaussians. Each spectral subtype  $\mathcal{M}$  (from M0 to M7) is characterized by a mean vector  $\mu_{\mathcal{M}}$  and a covariance matrix  $\Sigma_{\mathcal{M}}$ .

For a given star with colour vector  $\mathbf{x} = (r - i, i - z)$ , the likelihood of belonging to subclass  $\mathcal{M}$  is computed as:

$$p_{\mathcal{M}}(\mathbf{x}) = \frac{1}{2\pi\sqrt{|\Sigma_{\mathcal{M}}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu_{\mathcal{M}})^T \Sigma_{\mathcal{M}}^{-1}(\mathbf{x} - \mu_{\mathcal{M}})\right) \quad (2)$$

The spectral subclass is then assigned based on the maximum probability  $p_{\mathcal{M}}$  across all  $\mathcal{M}$ .

The result of such classification is shown in Fig. 7. We present comparison between the number of stars of the certain spectral subclass versus corresponding number of flares for both subsets (see Fig. 8a and Fig. 9a) as well as a comparison of flares frequencies per spectral subclass (Fig. 8b and Fig. 9b). Based on the analysis of Gaia-matched there is a clear increasing trend between a flare frequency and a spectral subtype with a steep rise near M4-M5 where stars, tentatively, become fully convective and therefore generate more flares which is consistent with previous researches (Günther et al. 2020; Webb et al. 2021). Oppositely there is no such evidence from PS1-matched sample analysis apparently due to the absence of color correction for the galactic reddening. The distribution of spectral sub-classes is shifted towards the older spectral sub-classes, due to interstellar extinction.

## 4.3 Physical parameter estimation

### 4.3.1 Flare energy

To conduct the energy estimation, we use the Gaia-matched sample with available distance estimations. For the recurrent events, we estimate energy for each flare separately. We exclude flares with unusual shapes or poorly sampled light curves (e.g., flares showing only a



**Table 2.** Preview of the final table of flares parameters. The full table will be available online upon publication.

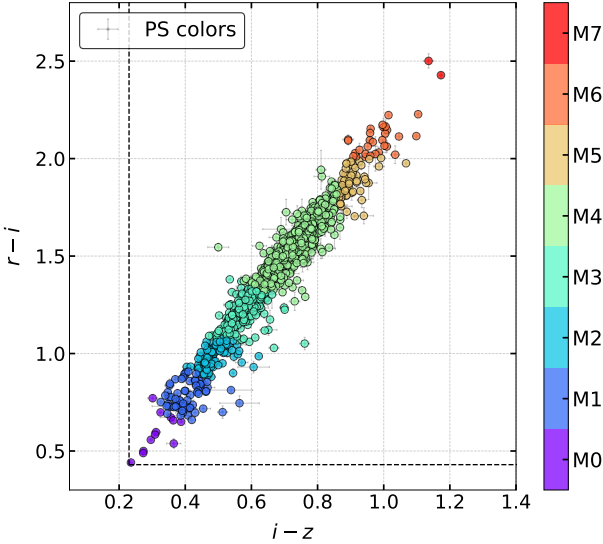
ZTF DR OID	$\alpha$ , deg	$\delta$ , deg	distance, pc	$A_r$ <sup>1</sup> , (mag)	$t_{\text{peak}}$ <sup>2</sup> , MJD–58000	FWHM <sup>2</sup> , hours	amplitude <sup>2</sup> , $\Delta$ mag	n points <sup>2</sup>	spectral class
736211200022871	0.5619	49.3449	361.96 <sup>+25.37</sup> <sub>–28.37</sub>	0.1570 <sup>–0.0262</sup>	2042.4119	6.57	0.632	56	M2
806208200015545	0.2700	61.6103	132.18 <sup>+0.77</sup> <sub>–0.97</sub>	0.0000	1473.1295	2.02	0.621	54	M3
806212300069862	0.5732	62.2869	<sup>†</sup> 1949.39 <sup>+1871.13</sup> <sub>–774.08</sub>	0.9577 <sup>+0.6199</sup> <sub>–0.2250</sub>	2032.4054	0.93	1.626	26	M4
806216100027505	1.0642	65.6713	645.17 <sup>+102.19</sup> <sub>–82.13</sub>	1.4394 <sup>+0.5219</sup> <sub>–0.0518</sub>	2032.4363	7.99	0.525	41	M3
833208300019023	0.9214	67.6663	<sup>†</sup> 1407.72 <sup>+492.90</sup> <sub>–305.74</sub>	3.5330 <sup>+0.0262</sup> <sub>–0.1047</sub>	2038.4438	7.34	0.503	30	M4
...									

<sup>1</sup> For sources with reliable geometric distances, the three-dimensional Milky Way dust map Bayestar19 (Green et al. 2019) was used to estimate extinction. If such distance information was unavailable, extinction values were taken from the Galactic Dust Reddening and Extinction map by Schlafly & Finkbeiner 2011.

<sup>2</sup> The peak time, FWHM, amplitude, and number of points were obtained from parametric flare-profile fits. For sparsely sampled light curves, the peak time corresponds to the epoch of minimum observed magnitude, and the amplitude was computed as the difference between this minimum and the quiescent magnitude inferred from the model. FWHM values were measured only for events with sufficient data to construct a reliable flare profile.

<sup>3</sup> Spectral classification is based on *Gaia* DR3 effective temperature ( $T_{\text{eff}}$ ) when available. If no *Gaia* information was present, the spectral subclass was derived from *PSI* photometry without applying any dust correction.

<sup>†</sup> Objects with  $\text{Plx}/e_{\text{Plx}} < 5$  according to *Gaia* DR3 parallax estimates.



**Figure 7.** Colour-colour diagram for PS1-matched sample. The distribution of stars in the ( $r-i$ ,  $i-z$ ) plane is shown, with colored regions indicating the loci corresponding to spectral subclasses M0–M7 used for photometric classification in this work.

decay phase) and apply the quality cut on the parallax-to-error ratio:

$$\frac{\varpi}{\sigma_{\varpi}} > 3.$$

The final sample used for energy calculation contains 655 flares. The energy estimation method follows the approach developed in Voloshina et al. (2024) and is summarized here for completeness. We assume that flare emission can be approximated by black-body radiation with  $T_{\text{flare}} = 9000$  K (Hawley & Fisher 1992), yielding the bolometric luminosity:

$$L_{\text{flare}}(t) = \sigma_{\text{SB}} T_{\text{flare}}^4 d^2 \frac{F_r(t)}{B_r}, \quad (3)$$

where  $F_r(t)$  is the extinction-corrected flux in the ZTF  $r$ -band, and  $B_r$  is the band-averaged black-body intensity. Following Shibayama et al.

OID	FWHM, minutes	E, $10^{33}$ erg
257207200014674	2.82	$0.58 \pm_{0.06}^{0.07}$
257207200014907	3.13	$0.88 \pm_{0.03}^{0.03}$
259205400044242	7.53	$4.25 \pm_{1.01}^{1.75}$
259207300021409	2.99	$1.16 \pm_{0.04}^{0.05}$
280211400109529	2.18	$0.68 \pm_{0.03}^{0.02}$
...		

**Table 3.** Preview of bolometric energy estimations for a sample of flare candidates. The full table will be available online upon publication. Energy errors reflect uncertainties in geometric distance estimates (see Sec. 4.3.1).

(2013) and Yang et al. (2017), we assume the geometric correction factor is unity.

Observed magnitudes are converted to fluxes using 3D extinction maps (Green et al. 2019), and the light curves are fit with a modified semi-empirical flare model from Mendoza et al. (2022), as described in Voloshina et al. (2024). The total flare energy is then obtained by integrating the luminosity:

$$E_{\text{flare}} = \int L_{\text{flare}}(t) dt. \quad (4)$$

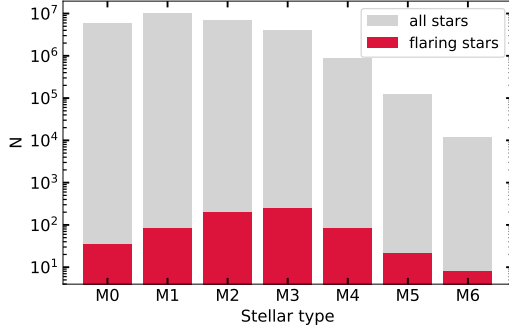
A power-law relation is found between flare duration (see details about calculation of flare duration  $t_{\text{FWHM}}$  in Section 4.3.2) and bolometric energy estimation:

$$E_{\text{flare}} = 10^{32.045} \times t_{\text{FWHM}}^{1.306}, \quad (5)$$

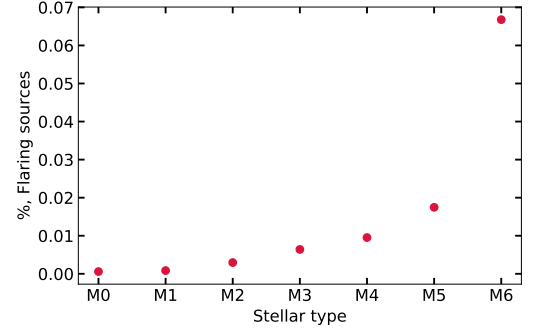
where  $t_{\text{FWHM}}$  is the full width at half maximum (in minutes) derived from the best-fit model light curve.

#### 4.3.2 Flare fit parameters

For each flare profile, we evaluated a set of parameters based on the semi-empirical model from Mendoza et al. (2022), including the amplitude, full width at half maximum ( $t_{\text{FWHM}}$ ), and the number of observations during the flare.

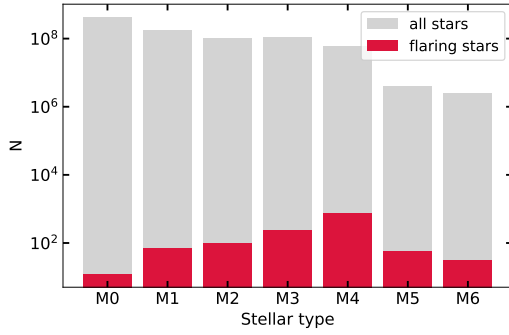


(a) Number of flares per spectral subclass.

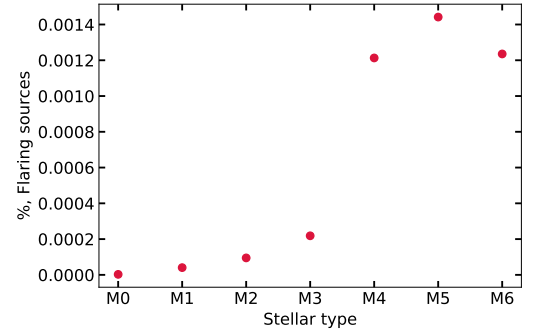


(b) Fraction of flaring sources per subclass.

**Figure 8.** Statistics of flaring sources by spectral type for the Gaia-matched sample. (8a) Total number of stars of each M-dwarf subclass (grey) and those exhibiting flares (red). (8b) Fraction of flaring sources relative to the total number of stars per spectral subclass.

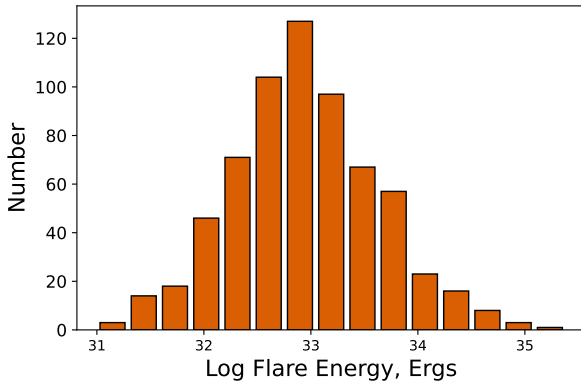


(a)



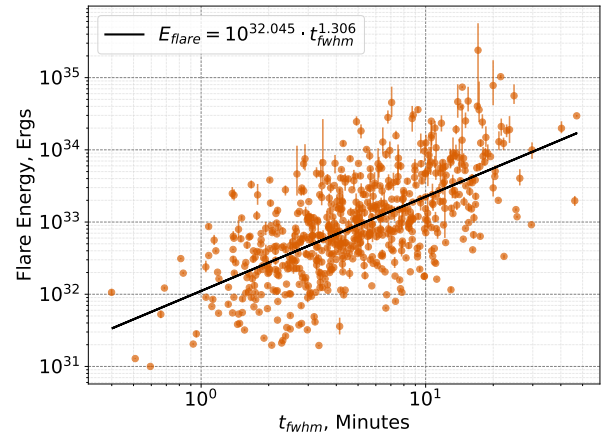
(b)

**Figure 9.** Statistics of flaring sources by spectral type for the PS1-matched sample. (9a) Total number of stars of each M-dwarf subclass (grey) and the number of stars exhibiting flares (red). (9b) Fraction of flaring sources relative to the total number of stars of the corresponding subclass.



**Figure 10.** Histogram showing the number of flare events per logarithmic energy bin. The sample spans approximately  $10^{31} - 10^{35}$  erg, derived for 655 Gaia-matched flares with reliable distance estimates.

In cases where a light curve contains multiple flares, we treated each flare separately by isolating and analyzing them independently. We performed a parametric fit to each flare profile to obtain a continuous representation of the flare. The amplitude was calculated as the difference between the maximum and minimum of the modeled



**Figure 11.** Bolometric flare energy as a function of the full width at half maximum ( $t_{fwhm}$ ) of the modeled light curves. The best-fit power-law relation  $E_{\text{flare}} = 10^{32.045} t_{\text{FWHM}}^{1.306}$ .

flux. To estimate the FWHM, we measured the time interval between points where the modeled flux equals half of the amplitude.

To determine the number of data points associated with a flare, we used the following criterion: any point in the light curve with

observed flux exceeding the quiescent stellar flux ( $f^*$ , as defined by the model) by more than  $3\sigma$  – where  $\sigma$  is the average observational uncertainty over the selected segment – was classified as part of the flare.

The resulting parameters for the representative subset of flares are listed in Table 2.

#### 4.4 Spatial distribution

Magnetic activity in low-mass stars provides a valuable tracer of stellar age and Galactic structure. Because vertical distance from the Galactic plane  $|z|$  correlates with stellar age through kinematic heating, the fraction of magnetically active or flaring M dwarfs can serve as a probe of the temporal evolution of magnetic activity across the Galactic disk. Our goal is therefore to quantify how the proportion of flaring stars – rather than the rate of individual flare events – varies with Galactic height.

For each observing field and vertical bin, we measured the total number of stars,  $N_{\text{star}}$ , and the number of stars exhibiting at least one detected flare,  $N_{\text{flare}}$ . The data are naturally described by a binomial process:

$$N_{\text{flare}, i, f} \sim \text{Binomial}(N_{\text{star}, i, f}, p_{i, f}), \quad (6)$$

where  $p_{i, f}$  is the probability that a randomly selected star in bin  $i$  and field  $f$  is flare-active. We emphasize that  $p_{i, f}$  represents the likelihood that a star belongs to the flaring (magnetically active) population, not the instantaneous probability of an individual flare event.

We modeled the dependence of this probability on Galactic height using a binomial generalized linear model (GLM) with a logit link function. The logit function,

$$\text{logit}(p_{i, f}) = \log\left(\frac{p_{i, f}}{1 - p_{i, f}}\right), \quad (7)$$

maps probabilities in the range  $0 < p_{i, f} < 1$  to the full real axis, allowing us to express the expected log-odds of flaring activity as a linear combination of predictors:

$$\text{logit } p_{i, f} = \alpha + \beta \log |z_{i, f}| + \gamma_f. \quad (8)$$

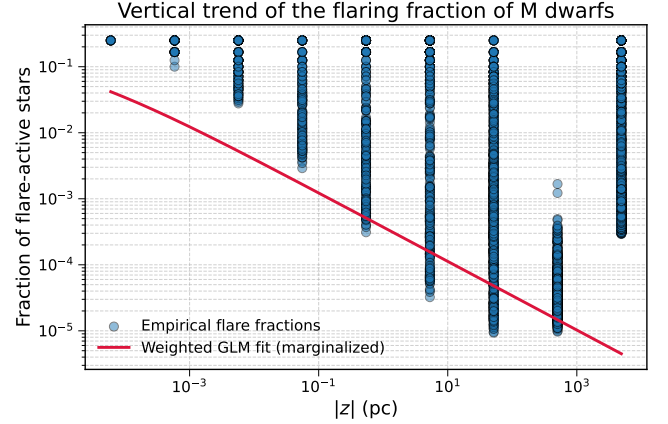
Fitting this model to the observed data yields a highly significant negative slope of

$$\beta = -0.52 \pm 0.03 \quad (p < 10^{-15}), \quad (9)$$

indicating that the odds of a star being flare-active decrease by roughly a factor of three for every order-of-magnitude increase in  $|z|$ . This trend remains consistent when tested with Poisson and negative binomial GLMs, confirming that it is robust to the assumed statistical family. The observed decline of  $p(|z|)$  with Galactic height reflects the progressive aging of stellar populations away from the midplane and the corresponding decay of magnetic activity with time. Figure 12 shows the vertical distribution of flare-active M dwarfs and the fitted generalized linear model (GLM) for the flaring fraction as a function of Galactic height. The blue points represent empirical flare fractions measured in individual fields, computed as the ratio  $N_{\text{flare}}/N_{\text{star}}$  in each  $|z|$  bin. The solid red curve shows the weighted binomial GLM prediction marginalized over fields and scaled by the total number of stars per field.

## 5 CONCLUSIONS

In this work, we have constructed and presented the largest catalogue to date of M-dwarf flares from a ground-based survey, con-



**Figure 12.** Vertical trend of the flaring fraction of M dwarfs. Blue points show empirical flare fractions per field, while the solid red curve represents the weighted binomial GLM prediction marginalized over fields and scaled by stellar counts. Both axes are logarithmic.

sisting of 1,229 unique flaring events identified in data from the Zwicky Transient Facility. This catalogue provides a substantial, well-characterized sample for studying the energetic phenomena on low-mass stars.

Our results were achieved through a multi-stage pipeline trained on simulated examples to efficiently scan through billions of light curves and isolate flare events. Training examples were simulated by injecting flare templates from the TESS space telescope into actual ZTF quiescent light curves, ensuring the model learned to identify flares within the specific noise properties and cadence of the ZTF survey. To maximize the purity of the final catalogue, we used a post-filtering process that successfully removed a significant number of false positives by: (1) cross-match against minor planet ephemerides; (2) deploying a second ML model to filter out instrumental artifacts based on image quality metadata; and (3) using Pan-STARRS colors to select for M-dwarf candidates. The developed pipeline is adaptable for detecting not only flares but also various other transient phenomena, such as short transients (for example, Fast Radio Bursts).

The resulting catalogue offers detailed characterizations of both the flares and their host stars. By leveraging photometric data from Pan-STARRS and astrometric data from Gaia, we analyzed the distribution of flare hosts across spectral subtypes. Our analysis confirms previous findings, revealing a clear increase in flare frequency with later spectral types, with a significant peak around the M4–M5 subtypes. This observation is consistent with the theoretical transition to fully convective interiors in M-dwarfs, which is thought to enhance dynamo action and magnetic activity. Furthermore, for a high-quality subset of 655 flares with reliable Gaia distance measurements, we estimated bolometric energies and durations.

Establishing robust methodologies for flare detection and characterization is crucial for the era of upcoming large-scale transient surveys. Our developed pipeline and post-filtering strategy provide a blueprint for efficiently processing the unprecedented volumes of data expected from these next-generation facilities. Missions like the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST), the Nancy Grace Roman Space Telescope, *ULTRASAT*, and the Argus Array will each contribute unique capabilities to our understanding of the dynamic sky. By demonstrating the successful extraction of a large and pure sample of M-dwarf flares from ZTF data, our research paves the way for even more comprehensive studies of stellar activity across the vast parameter space that these future sur-

veys will unlock, ultimately enhancing our understanding of stellar evolution, exoplanet habitability, and the transient universe.

## DATA AVAILABILITY

The ZTF light-curve data underlying this article are available from the NASA/IPAC Infrared Science Archive<sup>7</sup>. The final catalogue of selected flares and their derived parameters will be made publicly available via GitHub<sup>8</sup> and the Vizier database upon publication.

## ACKNOWLEDGEMENTS

A. Lavrukhina and M. Pruzhinskaya acknowledge support from a Russian Science Foundation grant 24-22-00233, <https://rscf.ru/en/project/24-22-00233/> for the analysis of M-dwarf population and their physical properties. Support was provided by Schmidt Sciences, LLC. for K. Malanchev. The computational resources were provided by Yandex Cloud<sup>9</sup> as part of a grant for scientific research. K. Malanchev and E.E.O. Ishida acknowledge support from 2024 CNRS International Emerging Actions A100011.

## REFERENCES

- Aizawa M., et al., 2022, *PASJ*, **74**, 1069  
 Bailer-Jones C. A. L., Rybizki J., Foesneau M., Demleitner M., Andrae R., 2021, *AJ*, **161**, 147  
 Bellm E. C., et al., 2019, *PASP*, **131**, 018002  
 Borucki W. J., et al., 2010, *Science*, **327**, 977  
 Browning M. K., 2008, *ApJ*, **676**, 1262  
 Caplar N., et al., 2025, *arXiv e-prints*, p. arXiv:2501.02103  
 Crossland A., Bellm E. C., Klein C., Davenport J. R. A., Kupfer T., Groom S. L., Laher R. R., Riddle R., 2024, *The Open Journal of Astrophysics*, **7**  
 D’Isanto A., Cavuoti S., Brescia M., Donalek C., Longo G., Riccio G., Djorgovski S. G., 2016, *MNRAS*, **457**, 3119  
 Davenport J. R. A., Becker A. C., Kowalski A. F., Hawley S. L., Schmidt S. J., Hilton E. J., Sesar B., Cutri R., 2012, *ApJ*, **748**, 58  
 Davenport J. R. A., et al., 2014, *ApJ*, **797**, 122  
 Dempster A., Schmidt D. F., Webb G. I., 2021, in Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. KDD ’21. ACM, p. 248–257, doi:10.1145/3447548.3467231, <http://dx.doi.org/10.1145/3447548.3467231>  
 Dorogush A. V., Ershov V., Gulina A., 2018, CatBoost: gradient boosting with categorical features support (*arXiv:1810.11363*), <https://arxiv.org/abs/1810.11363>  
 Galletta, G. Colombo, S. Prisinzano, L. Micela, G. 2025, *A&A*, **698**, A180  
 Green G. M., Schlafly E., Zucker C., Speagle J. S., Finkbeiner D., 2019, *The Astrophysical Journal*, **887**, 93  
 Günther M. N., et al., 2020, *AJ*, **159**, 60  
 Hawley S. L., Fisher G. H., 1992, *ApJS*, **78**, 565  
 Hawley S. L., Pettersen B. R., 1991, *ApJ*, **378**, 725  
 Hawley S. L., Davenport J. R. A., Kowalski A. F., Wisniewski J. P., Hebb L., Deitrick R., Hilton E. J., 2014, *ApJ*, **797**, 121  
 Ho T. K., 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition. pp 278–282 vol.1, doi:10.1109/ICDAR.1995.598994  
 Kim D.-W., Protopapas P., Bailer-Jones C. A. L., Byun Y.-I., Chang S.-W., Marquette J.-B., Shin M.-S., 2014, *A&A*, **566**, A43  
 Kitchatinov L. L., Moss D., Sokoloff D., 2014, *MNRAS*, **442**, L1

- Kowalski A. F., Hawley S. L., Hilton E. J., Becker A. C., West A. A., Bochanski J. J., Sesar B., 2009, *The Astronomical Journal*, **138**, 633  
 Kowalski A. F., Hawley S. L., Wisniewski J. P., Osten R. A., Hilton E. J., Holtzman J. A., Schmidt S. J., Davenport J. R. A., 2013, *ApJS*, **207**, 15  
 Krushinsky V. V., Pruzhinskaya M. V., Semenikhin T. A., The SNAD team 2025, *Research Notes of the American Astronomical Society*, **9**, 156  
 Kupfer T., et al., 2021, *Monthly Notices of the Royal Astronomical Society*, **505**, 1254–1267  
 Lavrukhina A., Malanchev K., Kornilov M. V., 2023, *Research Notes of the American Astronomical Society*, **7**, 199  
 Liu Q., et al., 2023, *MNRAS*, **523**, 2193  
 Malanchev K., 2021, light-curve: Light curve analysis toolbox, Astrophysics Source Code Library, record ascl:2107.001 (ascl:2107.001)  
 Malanchev K. L., et al., 2021, *Monthly Notices of the Royal Astronomical Society*, **502**, 5147–5175  
 Malkov O., Kovaleva D., Sichevsky S., Zhao G., 2020, *Research in Astronomy and Astrophysics*, **20**, 139  
 Mendoza G. T., Davenport J. R. A., Agol E., Jackman J. A. G., Hawley S. L., 2022, *AJ*, **164**, 17  
 Newton E. R., Irwin J., Charbonneau D., Berlind P., Calkins M. L., Mink J., 2017, *ApJ*, **834**, 85  
 Otsu N., 1979, *IEEE Transactions on Systems, Man, and Cybernetics*, **9**, 62  
 Pedregosa F., et al., 2011, *Journal of Machine Learning Research*, **12**, 2825  
 Pietras M., Falewicz R., Siarkowski M., Bicz K., Preš P., 2022, *ApJ*, **935**, 143  
 Poyatos J., Fors O., Gómez Cama J. M., Ribas I., 2025, *Astronomy & Astrophysics*, **699**, A242  
 Rajpurohit A. S., Reylé C., Allard F., Homeier D., Schultheis M., Bessell M. S., Robin A. C., 2013, *A&A*, **556**, A15  
 Ricker G. R., et al., 2014, in Space Telescopes and Instrumentation 2014: Optical, Infrared, and Millimeter Wave. p. 914320 (*arXiv:1406.0151*), doi:10.1117/12.2063489  
 Rodríguez Martínez R., López L. A., Shappee B. J., Schmidt S. J., Jayasinghe T., Kochanek C. S., Auhettl K., Holoién T. W. S., 2020, *ApJ*, **892**, 144  
 Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, **737**, 103  
 Shibayama T., et al., 2013, *The Astrophysical Journal Supplement Series*, **209**, 5  
 Skrutskie M. F., et al., 2006, *AJ*, **131**, 1163  
 Sreejith S., et al., 2026, *New Astron.*, **122**, 102466  
 Stetson P. B., 1996, *PASP*, **108**, 851  
 Virtanen P., et al., 2020, *Nature Methods*, **17**, 261  
 Voloshina A. S., et al., 2024, *Monthly Notices of the Royal Astronomical Society*, **533**, 4309–4323  
 Webb S., et al., 2021, *MNRAS*, **506**, 2089  
 West A. A., Weisenburger K. L., Irwin J., Berta-Thompson Z. K., Charbonneau D., Dittmann J., Pineda J. S., 2015, *ApJ*, **812**, 3  
 Yang H., Liu J., 2019, *ApJS*, **241**, 29  
 Yang H., et al., 2017, *ApJ*, **849**, 36  
 York D. G., et al., 2000, *AJ*, **120**, 1579

## APPENDIX A: LIST OF EXPERTS FEATURES FOR TRAINING

We define light curve as a list of triples of magnitude  $m_i$ , its error  $\delta_i$  and observation time  $t_i$ , where  $i = 0..N - 1$  and  $N$  is the number of observations.

### A0.0.1 Mag min

**A0.0.2 Bazin fit** Five fit parameters and goodness of fit (reduced  $\chi^2$  of the Bazin function developed for core-collapsed supernovae:

$$f(t) = A \frac{e^{-(t-t_0)/\tau_{fall}}}{1 + e^{-(t-t_0)/\tau_{rise}}} + B, \text{ where } f(t) - \text{flux observation}$$

<sup>7</sup> <https://irsa.ipac.caltech.edu/>

<sup>8</sup> <https://github.com/snad-space/flare-classifier>

<sup>9</sup> <https://yandex.cloud>



**A0.0.3 Median Buffer Range Percentage** Fraction of points within  $\text{Median}(m) \pm q \times (\max(m) - \min(m))/2$ . We used  $n = 0.01$  in our analysis.

**A0.0.4 Amplitude** The half amplitude of the light curve:

$$\frac{\max(m) - \min(m)}{2}. \quad (\text{A1})$$

**A0.0.5 Percent Difference Magnitude Percentile** Ratio of  $p$ -th inter-percentile range to the median:  $\frac{Q(1-p) - Q(p)}{\text{Median}(m)}$ . We used  $p = 0.05$  in our analysis.

**A0.0.6  $\eta^e$**  Kim et al. (2014) Generalisation of  $\eta$  for unevenly time series.

$$\eta^e \equiv (t_{N-1} - t_0)^2 \frac{\sum_{i=0}^{N-2} \left( \frac{m_{i+1} - m_i}{t_{i+1} - t_i} \right)^2}{(N-1) \sigma_m^2}. \quad (\text{A2})$$

Our version of this feature differs from Kim et al. (2014).

**A0.0.7 Anderson Darling Normal** Unbiased Anderson-Darling normality test statistic:  $\left(1 + \frac{4}{N} - \frac{25}{N^2}\right) \times \left(-N - \frac{1}{N} \sum_{i=0}^{N-1} (2i+1) \ln \Phi_i + (2(N-i)-1) \ln(1 - \Phi_i)\right)$ , where  $\Phi_i \equiv \Phi((m_i - \langle m \rangle)/\sigma_m)$  is the cumulative distribution function of the standard normal distribution,  $N$  – the number of observations,  $\langle m \rangle$  – mean magnitude and  $\sigma_m = \sqrt{\sum_{i=0}^{N-1} (m_i - \langle m \rangle)^2 / (N-1)}$  is the magnitude standard deviation.

**A0.0.8 Standard Deviation** Standard deviation of magnitude:

$$\sigma_m \equiv \sqrt{\sum_i (m_i - \langle m \rangle)^2 / (N-1)}$$

**A0.0.9 Median** We define  $\text{Median}(x) \equiv Q_x(0.5)$ .

**A0.0.10 Mean** We define mean of the sample  $\{x_i\}$  as

$$\langle x \rangle \equiv \frac{1}{N} \sum_i x_i. \quad (\text{A3})$$

**A0.0.11 Von Neumann  $\eta$**  Kim et al. (2014)

$$\eta \equiv \frac{1}{(N-1) \sigma_m^2} \sum_{i=0}^{N-2} (m_{i+1} - m_i)^2. \quad (\text{A4})$$

**A0.0.12 Median Absolute Deviation** D’Isanto et al. (2016) Median of the absolute value of the difference between magnitudes and their median:

$$\text{Median}(|m_i - \text{Median}(m)|). \quad (\text{A5})$$

**A0.0.13 Beyond  $n$  Std** D’Isanto et al. (2016) The fraction of observations beyond  $\langle m \rangle \pm n \sigma_m$ :

$$\frac{\sum_i I_{|m - \langle m \rangle| > n \sigma_m}(m_i)}{N}. \quad (\text{A6})$$

We use  $n = 1$  and  $n = 2$  in our analysis.

**A0.0.14 Skew** The skewness of magnitude distribution:

$$G_1 \equiv \frac{N}{(N-1)(N-2)} \frac{\sum_i (m_i - \langle m \rangle)^3}{\sigma_m^3}. \quad (\text{A7})$$

**A0.0.15 Weighted Mean** Weighted mean magnitude:  $\bar{m} \equiv \frac{\sum_i m_i / \delta_i^2}{\sum_i 1 / \delta_i^2}$

**A0.0.16 Excess Variance** Measure of the variability amplitude:  $\frac{\sigma_m^2 - \langle \delta^2 \rangle}{\langle m \rangle^2}$ , where  $\langle \delta^2 \rangle$  – mean squared error.

**A0.0.17 Stetson  $K$**  Stetson (1996) Stetson K coefficient described light curve shape:

$$K \equiv \frac{\sum_i \left| \frac{m_i - \bar{m}}{\delta_i} \right|}{\sqrt{N \sum_i \left( \frac{m_i - \bar{m}}{\delta_i} \right)^2}}. \quad (\text{A8})$$

**A0.0.18 Cusum** Kim et al. (2014) The range of magnitude cumulative sums:

$$\max_j (S_j) - \min_j (S_j), \quad (\text{A9})$$

where

$$S_j \equiv \frac{1}{N \sigma_m} \sum_{i=0}^j (m_i - \langle m \rangle). \quad (\text{A10})$$

**A0.0.19 Inter Percentile Range**  $Q(1-p) - Q(p)$ , where  $Q(n)$  and  $Q(d)$  –  $n$ -th and  $d$ -th quantile of magnitude sample. We use  $p = 0.25$  in our analysis.

**A0.0.20 Percent Amplitude** Maximum deviation of magnitude from its median:  $\max_i |m_i - \text{Median}(m)|$

**A0.0.21 Maximum Slope** Maximum slope between two sub-sequential observations:  $\max_{i=0 \dots N-2} \left| \frac{m_{i+1} - m_i}{t_{i+1} - t_i} \right|$

**A0.0.22 Kurtosis** The kurtosis of the magnitude distribution

$$G_2 \equiv \frac{N(N+1)}{(N-1)(N-2)(N-3)} \frac{\sum_i (m_i - \langle m \rangle)^4}{\sigma_m^4} - \frac{3(N+1)^2}{(N-2)(N-3)}. \quad (\text{A11})$$

**A0.0.23 Mean Variance** Standard deviation to mean ratio:  $\frac{\sigma_m}{\langle m \rangle}$

**A0.0.24 Otsu Split** Otsu (1979); Lavrukhina et al. (2023) Difference of subset means, standard deviation of the lower subset, standard deviation of the upper subset and lower-to-all observation count ratio for two subsets of magnitudes obtained by Otsu’s method split. Otsu’s method is used to perform automatic thresholding. The algorithm returns a single threshold that separate values into two classes. This threshold is determined by minimizing intra-class intensity variance  $\sigma_W^2 = w_0 \sigma_0^2 + w_1 \sigma_1^2$ , or equivalently, by maximizing inter-class variance  $\sigma_B^2 = w_0 w_1 (\mu_1 - \mu_0)^2$ . There can be more than one extremum. In this case, the algorithm returns the minimum threshold.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.