

Search for exocomet transits in Kepler light curves

Ten new transits identified

P. Dumond^{1,2,*}, A. Lecavelier des Etangs¹, F. Kiefer^{1,3}, G. Hébrard¹, and V. Caillé^{1,4}

¹ Institut d'astrophysique de Paris, CNRS, Sorbonne Université, 98^{bis} boulevard Arago, 75014 Paris, France

² CRAL, Ecole normale supérieure de Lyon, Université de Lyon, UMR, CNRS 5574, 69364 Lyon Cedex 07, France

³ LIRA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Université de Paris, 5 place Jules Janssen, 92195 Meudon, France

⁴ Institut Pierre-Simon Laplace, Sorbonne-Université, CNRS, Paris, France

Received 19 June 2025 / Accepted 15 October 2025

ABSTRACT

Although the Kepler telescope was retired over a decade ago, it continues to offer a rich dataset for uncovering new astrophysical objects and phenomena. In this study, we conducted a comprehensive search for exocometary transit signatures within the Kepler light curves, using a machine learning approach based on a neural network trained on a library of theoretical exocomet transit light curves. By analyzing the light curves of 201 820 stars, we identified candidate events through the neural network and subjected the output to filtering and visual inspection to mitigate false positives. Our results are presented in three catalogs of increasing ambiguity. The first tier catalog includes 17 high-confidence exocometary transit events, comprising seven previously reported events and ten newly identified ones, each associated with a different host star. The second tier catalog lists 30 lower-confidence events that remain consistent with possible exocometary transits. The third tier catalog consists of 49 more symmetric photometric events that could be exocometary transits, exoplanet mono-transits, or false positives due to eclipsing binaries mimicking transits. Contrary to previous studies, which suggested that the cometary activity was favored by stellar youth, we find a broad age distribution among candidate host stars, including several red giants. This challenges the general idea of a decline in cometary activity with stellar age and underlines the need for further investigation into the temporal evolution of exocometary activity in planetary systems.

Key words. methods: data analysis – techniques: photometric – surveys – comets: general

1. Introduction

Exocomets are rocky or icy minor bodies that become active when they approach their parent star on a highly eccentric orbit (Strøm et al. 2020). Because they produce a huge tail of gas and dust close to periastron, when they transit in front of the star, they become detectable using spectroscopy (for the gaseous tail) and photometry (for the dust tail). The first exocomets were thus identified in the 1980s in the β Pic system, through the detection of variable absorption features in the Ca II lines of the stellar spectrum (Ferlet et al. 1987; Kiefer et al. 2014a). Later on, a few other exocometary systems were found using spectroscopy, for example HD 172555 (Kiefer et al. 2014b; Grady et al. 2018) and 49 Cet (Montgomery & Welsh 2012; Miles et al. 2016). A list of spectroscopically identified exocometary systems can be found in Strøm et al. (2020).

For the detection of exocomets using photometry, one needs to obtain photometric measurements at high accuracy, typically at the 10^{-3} – 10^{-4} level over the full transit duration, i.e., for one to three days (Lecavelier des Etangs et al. 1999, hereafter LdE99a). This capability was reached only during the last decade with space missions such as Corot, Kepler, TESS, and Cheops. Detections have thus been obtained with Kepler toward KIC 3542116 and KIC 11084727 (Rappaport et al. 2018), and KIC 8027456 (Kennedy et al. 2019), with TESS toward β Pic (Zieba et al. 2019; Pavlenko et al. 2022; Lecavelier des Etangs et al. 2022), and

with Cheops toward HD 172555 (Kiefer et al. 2023). Here we also mention the detection with Kepler of a photometric event, possibly periodic, which has been interpreted as the transit of a string of five to seven exocomets in front of KIC 8462852 (Kiefer et al. 2017).

All these events were identified thanks to the particular shape of the light curve of an exocomet transit, as theoretically predicted at the end of the 1990s in LdE99a: the characteristic of an exocomet transit light curve is the asymmetry caused by the cometary tail passing in front of the star after the nucleus. This causes a sharp decrease in the star light followed by a slow return to the normal brightness (see also Lecavelier des Etangs 1999, hereafter LdE99b). However, exocomet transit light curves can also be more symmetric when the longitude of the periastron of the orbit is about 90° from the line of sight, because here the cometary tail is aligned with the line of sight. In this case, the shape of the transit light curve is similar to that of an exoplanet transit (LdE99a).

The detection of exocometary transits based on photometry is complementary to the detection through spectroscopy. It allows the physical characteristics of dust particles to be explored, such as the dust particle size distribution, the distance to the host star at the time of transit, and the dust production rate (Luk'yanyk et al. 2024). Because transits give direct access to the geometrical extent and the optical thickness of the transiting dust cloud, coupled with models of dust production, this allows the estimate of the dust production rate and the size of the comets nucleus. In the case of β Pic, a statistical analysis

* Corresponding author: pierre.dumond@ens-lyon.fr

of the photometric events allowed the estimate of the nucleus size distribution that was found to be similar to that of asteroids and comets in the Solar System as set by collisional equilibrium (Lecavelier des Etangs et al. 2022).

Although the characterization of exocomets paves the way to understanding the dynamical and chemical processes occurring in young planetary systems, only a very limited number of exocometary systems have been identified to date. As a result, our current understanding of the diversity, frequency, and evolutionary significance of exocometary activity remains incomplete. Systematic searches have been performed using spectroscopy in the Ca II line (Montgomery & Welsh 2012; Welsh & Montgomery 2013, 2018; Rebollido et al. 2020; Bendahan-West et al. 2025), resulting in a few detections (see review in Strøm et al. 2020). Spectroscopy is efficient in detecting tiny amounts of gas, and thus allows a sensitive search; the result is that detectable spectroscopic transits seem to be more frequent than detectable photometric events. For instance, in the case of β Pic, several exocomet transits are routinely detected every day in spectroscopy (Kiefer et al. 2014a), while only 30 photometric transits were detected in 156 days of TESS observations (Lecavelier des Etangs et al. 2022). However, spectroscopic searches for exocomets are inherently limited by their need to observe each target star individually and the need for long observation times with high spectral resolution. In contrast, photometric surveys can monitor several thousand stars simultaneously, making them significantly more efficient for large-scale searches for exocometary transits.

In that spirit, Rappaport et al. (2018) and Kennedy et al. (2019) began a deep search for exocometary transits in the Kepler light curves. To identify exocometary transits, they used a modified model of planetary transit or a combination of exponential and Gaussian tails, as also done later on by Zieba et al. (2019), Pavlenko et al. (2022), and Lecavelier des Etangs et al. (2022). Although this empirical approach does not carry physical information about the cometary bodies and tails, it is easy to implement and thus to identify candidates in the large number ($\sim 200\,000$) of light curves provided by Kepler.

Here we propose to use an automated search without the use of analytical functions to describe the exocometary transit light curves. For that we combined a search algorithm based on a neural network together with a library of photometric transits to train the network. The library is an updated version of the library of LdE99b, which has been recalculated using faster CPUs than those used 20 years ago, covering a wide range of parameters, and therefore a wide range of transit shapes that need to be identified. We applied this automated search to the Kepler data (Borucki et al. 2010) with the aim of producing a new catalog of exocometary system candidates. The methods are described in Sects. 2–4. The search in the Kepler data is presented in Sect. 5. The final catalog of candidates, divided into three tiers, is given in Sect. 6. The results are discussed in Sect. 7.

2. Training of the neural network

2.1. Training principle

The method used for the detection of exocomets in the Kepler data is based on a neural network using supervised learning. For the network learning, we used a library of theoretical light curves divided into three disjoint sets of labeled light curves: the training set, the validation set, and the test set. Each set contains the same number of light curves with and without an exocomet transit (see Sect. 2.2). Each light curve is labeled to indicate if it includes a transit or not.

Our objective is to make the algorithm identify the curves containing an exocomet transit among all the light curves collected by the Kepler satellite and to recover the transit time when a transit is present. Using an iterative procedure, the algorithm optimizes its parameters thanks to the training set. At the end of each iteration step called an epoch, the network performance is tested using the validation set. This allows us to control the overfitting of the algorithm (e.g., it becomes specific to the training set and loses in generality) and to adjust the hyperparameters. The learning phase is stopped when the network shows the best performance in the validation set. The final performance of the algorithm is given by its evaluation on the test set.

2.2. Light curve production

The light curves used to build the training, validation, and test set are the Presearch Data Conditioning (PDC) data from Kepler Data Release 25 (Thompson et al. 2016). We used the Quarters 1 to 17 and divided them into subquarters in case of interruption of the acquisition within a given quarter: two subquarters are separated by at least four consecutive photometric measurements missing in the light curve.

For each subquarter, a training, validation, and test set were constructed to overcome their possible specificity. For each subquarter, the training set consists of 36 000 curves, while the validation and test sets contain 12 000 each. A curve consists of ten days (480 photometric measurements at a cadence of one measurement every ~ 30 minutes) of acquisition of the luminosity flux from a star randomly chosen without repetition.

To ensure efficient learning and prevent the algorithm from falling into obvious traps, several modifications were applied to the Kepler PDC data. First, all data points deviating by more than 5σ from the mean of the light curve were removed, provided that their neighboring points remained within the 5σ range. In addition, each curve was normalized to have a zero mean and a standard deviation of 1, as needed for optimal learning by the algorithm. Finally, the possible missing measurements, i.e., at most three consecutive missing measurements within the same subquarter, were filled using a linear interpolation.

For the learning process, half of the ten-day light curves were labeled “0”, indicating the absence of an exocomet transit. Given that the probability of a randomly selected Kepler light curve containing an exocomet transit is extremely low, we assume that all light curves labeled “0” are free of exocometary transits. This subset was used to train the algorithm to learn the intrinsic noise patterns in the Kepler data. For the remaining half of the ten-day light curves, we added one synthetic exocomet transit by multiplying the light curve by a transit from the theoretical library (Sect. 2.3), and labeled these curves “1”, indicating the presence of an exocomet transit.

2.3. Simulation of exocomet transits

In order for the algorithm to learn to identify the shape of an exocomet transit in the light curves, we added theoretical exocomet transits to half of the curves that make up the training data. These theoretical exocomet transit light curves were calculated using the method described in LdE99a. We produced a library of 2200 exocomet transit light curves, which is an updated version of the library published 20 years ago by LdE99b. Among all the available simulations, we selected only those with a gas production rate $\log_{10}(P/(1\text{kg s}^{-1}))$ between 6 and 7.5 and an impact parameter $b < R_*$, where R_* is the radius of the star. This choice guarantees that the used transit light curves are sufficiently deep.

An impact parameter greater than one stellar radius or a production rate that is too low leads to a noisy and very weak extinction and would therefore have no interest for the training. Before adding the simulated transits to the Kepler light curves, we multiplied the theoretical transit light curve by a random constant such that the depth of the transits is in the interval $[2\sigma, K + 2\sigma]$ where K is the depth of the simulated transits and σ the standard deviation of the Kepler light curves. We thus trained our model for a signal-to-noise ratio higher than 2 to exclude the cases where the transit would not be detectable. The transit times of the theoretical transits added to the ten-day light curves are chosen randomly, with a uniform distribution between the beginning and the duration of the simulated transit before the end of the light curve.

Finally, in order to ensure that the training data are representative of the diversity of possible transits, we used a uniform random distribution of the distance of the periastron q in the set $\{0.2; 0.3; 0.5; 0.7; 1.0\}$ astronomical units and of the longitude of the periastron ω from -157.5° to $+157.5^\circ$ with a step of 22.5° , as in LdE99b.

The theoretical transits that have been used are totally different from the training, validation and the test sets. To guarantee that the evaluation of the algorithm's performance is unbiased, the choice of the transits for each set was made randomly from the whole set of theoretical transits.

3. The neural network model

3.1. Architecture

The deep learning algorithm developed in this study is mainly based on the convolution layer sequence adapted for pattern recognition. We relied on the architecture of existing algorithms for the detection of exoplanet transits (Shallue & Vanderburg 2018; Dattilo et al. 2019; Zucker & Giryes 2018; Chintarungruangchai & Jiang 2019). These algorithms, which all have a similar architecture, are optimized for periodic transit detection. They have been modified to be efficient for the detection of single transits as the orbital period of an exocomet is expected to be much longer than the Kepler mission duration. Two major modifications have been made. Squeeze-excited blocks (Hu et al. 2018) have been added to the convolution blocks. The authors indeed show that such blocks can improve the performance of a pattern-matching algorithm by improving the interdependence relations between the convolution layers without adding computational cost. The addition of the long short-term memory (LSTM) (Hochreiter & Schmidhuber 1997) and the gated recurrent unit (GRU) (Cho et al. 2014) layers also resulted in significant performance improvements. The algorithm was implemented in TensorFlow, an open-source machine learning framework (Abadi et al. 2016). We optimized the neural network parameters thanks to KerasTuner (O'Malley et al. 2019).

Figure 1 describes our final model, which is mainly a one-dimensional convolutional network. The activation function of all the convolutional layers is the rectified linear unit (ReLU) function (Jarrett et al. 2009). The output is a list of two real numbers. The first uses a sigmoid function whose range is (0,1). It is the probability returned by the algorithm that the ten-day input light curve contains an exocomet transit. An output value close to 1 indicates high confidence that the considered light curve has one transit, while an output value close to 0 means that there is no transit. The second output layer is activated by a linear function in the range (0,1). It gives the position of the identified transit in

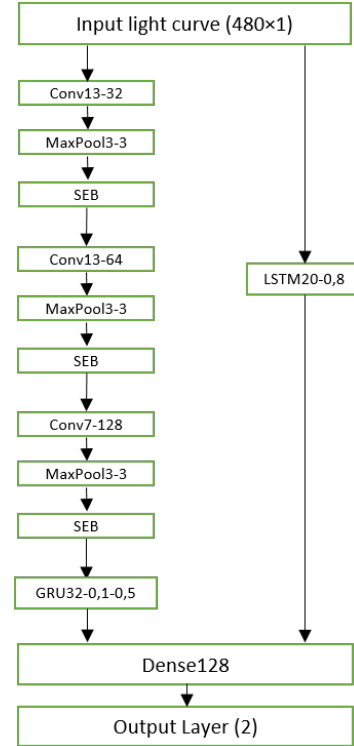


Fig. 1. Architecture of our best performing neural network model. The input data is a light curve of 480 timesteps, the output is a list of two real numbers between 0 and 1: one for the probability of detection of an exocometary transit and one for the position of the transit in the light curve. Convolutional layers are denoted Conv <kernel size> <number of feature maps>, max pooling layers are denoted MaxPool <window length> <stride length>, fully connected layers are denoted Dense <number of units>, LSTM layers are denoted LSTM <units> <dropout>, GRU layers are denoted GRU <unit> <droupout> <recurrent dropout>, and the squeezed-excitation blocks are denoted SEB.

the light curve. If no transit is detected, the layer should return a value close to zero.

3.2. Training

For each subquarter, we trained the model using 36 000 light curves for 150 training epochs, which means that we performed 150 complete passes through the entire training dataset to update the model weight parameters. We used the Adam optimization algorithm (Kingma & Ba 2015) to minimize the cross-entropy error function for the classification of the light curves and the mean absolute error function for the position of the transit. We used a learning rate of 10^{-4} and a batch size of 5000. The saved model in the one for which the validation loss reaches its minimum; it is usually obtained after 50–80 iterations (epochs). Therefore, training the model over 150 epochs is considered sufficient.

4. Evaluation of the neural network performance

We evaluated the performance of our neural network with respect to several metrics in the same way as Dattilo et al. (2019) did for their model. We computed all the metrics over the test set (rather than the training or validation sets) to avoid using any data that

were used to optimize the parameters or hyperparameters of the model.

4.1. The metrics

The metrics used to evaluate the performance of a classifier are the following:

1. Accuracy, the fraction of curves correctly classified by the model;
2. Precision (reliability), the fraction of correctly classified exocomets, i.e., the fraction of curves classified as exocomet candidates (False positives + True positives $N_{Fp} + N_{Tp}$) that are indeed true positives (1-labeled):

$$P = \frac{N_{Tp}}{N_{Fp} + N_{Tp}}; \quad (1)$$

3. Recall (completeness or true-positive rate), the fraction of the curves 1-labeled, which therefore includes an exocomet transit (true positives + false negatives, $N_{Tp} + N_{Fn}$), that the model classifies as exocomet candidates (N_{Tp}):

$$R = \frac{N_{Tp}}{N_{Tp} + N_{Fn}}; \quad (2)$$

4. False-positive rate (FPR), the fraction of total 0-labeled curves, which therefore do not include any exocomet transit (false positives + true negatives, $N_{Fp} + N_{Tn}$) that the model classifies as exocomet candidates:

$$FPR = \frac{N_{Fp}}{N_{Fp} + N_{Tn}}; \quad (3)$$

5. Area under the receiver-operator characteristic curve (AUC; see Figure 2), the probability that the model, if it receives two light curves, one 1-labeled curve including an exocomet transit and one 0-labeled curve without transit, would rank the 1-labeled curve higher than the 0-labeled curve;
6. Mean absolute error, the absolute error for all 1-labeled curves between the true position of the transit and the predicted position by the model.

The first five metrics are related to the detection of the presence of an exocomet in the light curves, while the last metric is related to the position in time of the transit in the light curve. Except for the AUC, the values of these metrics depend on the classification threshold chosen for the model. This threshold is the value of the probability of detection yield by the algorithm above which we consider that there is a detection. In Fig. 2 we show the evolution of the recall as a function of the FPR for various thresholds, usually called the receiver-operator characteristic (ROC) curve. It can be seen that the algorithm differentiates well between true transits and noise, as it achieves very low FPR recall, while keeping a reasonably large recall; our model reaches an AUC of 98.88% for the quarter Q1. We also plot the precision as a function of the recall. This curve shows the trade-off between having no false positives (high precision) and identifying all exocomet transits (high recall).

Considering the huge number of light curves in the Kepler data, the goal was to drastically reduce the number of false positives. We thus chose a classification threshold of 0.99. With this threshold, our model reaches an accuracy of 89.5% for the quarter Q1. The precision of the model is 99.8%, which means that we reach a high true positive rate, and the recall is 79.1%, which means that about 20% of the true exocomet transits are missing

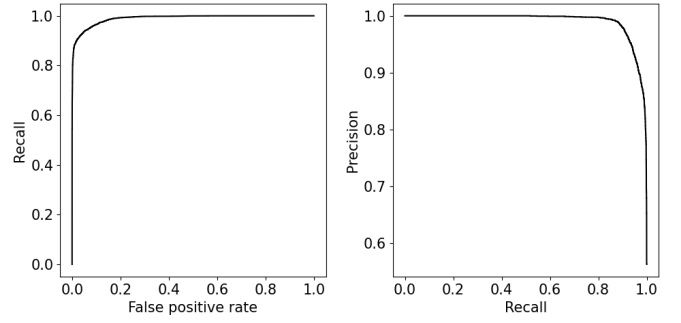


Fig. 2. Receiver-operator characteristic curve (left panel) and the precision vs. recall curve (right panel). The ROC curve shows the recall (true-positive rate) of the model against the ability to recognize false positives (false-positive rate) for different classification thresholds. Our model is highly successful at identifying false positives, as shown by the high AUC value (see Table A.1). The plot of the fraction of exocomets that the model classified as exocomets (recall) vs. the fraction of correctly classified planets (precision) shows the trade-off between having no false positives (high precision) and identifying all exocomet transits (high recall).

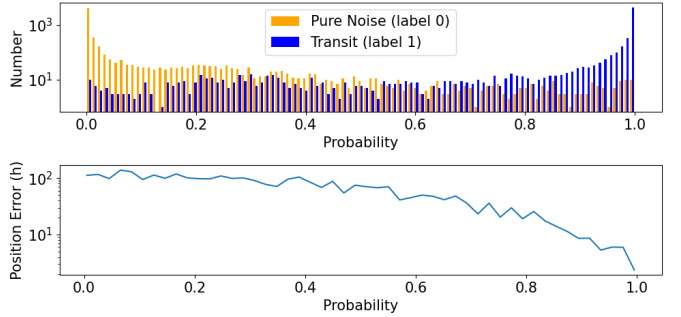


Fig. 3. Histogram of the results on the test set and error on the transit position found by the network. Top panel: histogram of the result of the neural network applied to the test data. Most of the 1-labeled light curves with an exocomet transit yield a probability close to 1, while there are only a few false positives (0-labeled light curves yielding a high probability). Above the chosen threshold, there is more than two order of magnitude between the number of true positives and false positive. Bottom panel: mean error of the position of the transit as a function of the probability of the presence of a transit assigned by the network to the light curves. This error is a few hours at most for the transits identified by the algorithm.

in the final classification. This trade-off aims to accept some loss in the finding of exocomet transits of about 20% to avoid a large number of false positives. Given the large number of light curves in the Kepler data and the small number of exocomet transits in these data, even a tiny fraction of false positives can significantly pollute the results. The metrics values are found to be similar for all quarters (see Table A.1).

4.2. Quality of the output

The quality of the model classification can also be visualized by the histogram of the results on the test set (top panel of Fig. 3). The closer the output from the neural network is to 1 (resp. 0), the higher the probability that the ten-day light curve contains (resp. does not contain) an exocomet transit. The histogram shows that the neural network works as expected: on the left-hand side, most of the light curves characterized by a low probability are mostly noise (orange histogram), while on the right-hand side, the large

majority of the curves characterized by a probability close to 1 include a simulated exocomet transit (label 1; blue histogram) and with very few false positives. The bottom graph of Fig. 3 shows the mean error of the position of the transit as a function of the probability assigned by the algorithm when it is applied to the curves containing a simulated transit (label 1, with known position). One can see that when the transit is not found (probability close to 0), the error on the predicted position is large, about 3 days, whereas it is on the order of a few hours only when the probability is above 0.99. Thus, one can be confident in the time of the transit given by the algorithm for curves with a probability higher than this threshold.

5. Search for exocomets in Kepler data

5.1. Application of the model to Kepler data

To apply the algorithm presented in the previous sections, all the light curves of the Kepler data were cut into ten-day windows obtained in the same way as the samples used to construct the training, validation, and test sets described in Sect. 2.2. In order not to lose any transits located at the edge of the windows, an overlap of 2 days was applied between two consecutive ten-day time intervals.

However, despite the level of precision achieved, the algorithm provided a large number of false positives due to the large number of ten-day windows: approximately 10^6 windows per quarter lead to about 10^3 positive candidate detections, among which less than a dozen real transits should be identified (see LdE99a and discussion in Sect. 7). To reduce the number of false positives, we decided to filter the candidates using constraints on the shape of the light curves of the detected photometric events.

5.2. Additional criteria to filter false positives

After all the Kepler light curves have passed through the neural network, we end up with a set of 56 525 exocomet candidates. To filter these candidates, we decided to add constraints on the shape and on the context on the transit (noise, quality flags).

5.2.1. Filtering through the shape of the transit candidates

The shapes of the candidate transits were determined by fitting them via the model proposed by Lecavelier des Etangs et al. (2022) together with a second-order polynomial baseline:

$$f_{\text{exo}}(t) = \begin{cases} Ct^2 + Dt + E & t < t_0, \\ (Ct^2 + Dt + E)(1 + K(e^{-\beta(t-t_0)} - 1)) & t_1 > t > t_0, \\ (Ct^2 + Dt + E)(1 + K(e^{-\beta(t-t_0)} - e^{-\beta(t-t_1)})) & t > t_1 \end{cases} \quad (4)$$

The parameters t_0 and t_1 represent respectively the time at which the transit starts and the time at which it reaches its minimum; K represents the characteristic depth of the transit, while $1/\beta$ corresponds to the characteristic crossing time of the transit; and C , D , and E are the coefficients of the polynomial baseline. The fit of the light curve by the transit model can be easily performed using the position of the detected candidate in the light curve as provided by the network output.

To accept or reject a proposed detection given by the neural network, we fit the light curve of the corresponding photometric event to obtain the values of the parameters K , β , and $\Delta T = t_1 - t_0$ that characterize the shape of the event. We then check that these

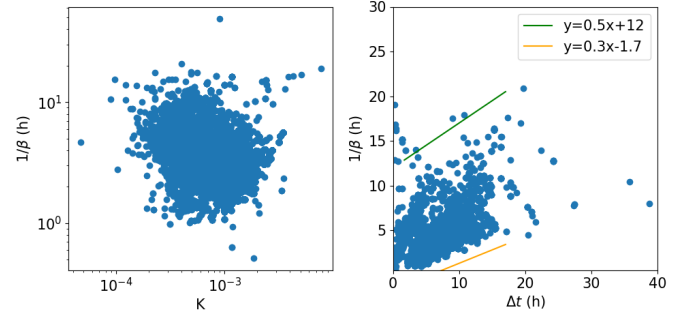


Fig. 4. Distribution of the value of the parameters obtained by fitting the 2163 light curves in the library of the simulated exocomet transits. The parameters K , $\Delta T = t_1 - t_0$, and $1/\beta$ correspond to the depth, ingress duration, and crossing time of the transit, respectively.

values are in the domain of validity for an exocometary transit. The domain of validity is obtained by considering the parameters space occupied by the fits of the theoretical light curves in the library of numerical simulations of exocometary transits, as shown in Fig. 4 and summarized in the first three lines of Table 1.

It is found that in the theoretical light curves K is mostly between 10^{-4} and 5×10^{-3} , ΔT is between 1.7 and 17 hours (corresponding to a periastron between ~ 0.02 to 2 au, Lecavelier des Etangs et al. 2022), and β^{-1} is in the domain above 0.5 hours, and between $0.3\Delta T - 1.7$ and $0.5\Delta T + 12$. Moreover, the parameters C and D have to verify the condition $|C|, |D| < 10^{-3}$. This ensures that the background is stable enough in the neighborhoods of the transit. When this is not the case, the detected transit appears to be only noise.

5.2.2. Quality flags

We observed that the presence of certain quality flags indicates an alteration of the light curve leading to a systematic detection by the network. Thus, candidates whose transit position is less than 1 day away from a quality flag 3 (spacecraft is in coarse point) or 12 (impulsive outlier removed before cotrending) were discarded from the candidate list.

5.2.3. Presence in the list of KOIs

To avoid periodic transits, which constitute roughly half of the candidate transits output by the algorithm, stars belonging to the list of Kepler objects of interest (KOIs) were also removed.

5.2.4. Other cases of false positives

We found three other circumstances that can mislead the algorithm and then result in obvious false positives: (1) significant background noise on a timescale close to that of a cometary transit; (2) a ramp-like shape in the light curve, i.e., a very sudden decrease in flux followed by a more or less rapid return to the normal level; (3) the presence of a mono-transit of a possible exoplanet that results in a very symmetrical brightness dip. In the following, we suggest the criteria to address each of these three cases.

To characterize the presence of noise anomaly, the Δ_{RMS} was calculated. It is defined as

$$\Delta_{\text{RMS}} = \sum_i (f_{\text{exo}}(t_i) - \bar{F}_3) / \sigma, \quad (5)$$

Table 1. Parameter values for filtering the candidate transits.

Parameter	Lower bound	Upper bound	Definition
K	10^{-4}	5×10^{-3}	Transit depth
β^{-1}	$\begin{cases} 0.5\text{h} \\ 0.3\Delta T - 1.7 \end{cases}$	$0.5\Delta T + 12$	Characteristic crossing time
ΔT	1.7h	17h	Ingress duration
$ C , D $		10^{-3}	Baseline coefficients
Δ_{RMS}	36		
$\Delta\chi_{\text{pl}}^2$	5		
$\Delta\chi_{\text{rp}}^2$	10		

where $f_{\text{exo}}(t_i)$ is the exocomet model fit to the flux where the candidate transit is found, \bar{F}_3 is the mean of the flux without the three lowest points, and σ is the standard deviation of the flux. Only candidate transits with $\Delta_{\text{RMS}} > 36$ were conserved. This criterion was very useful to remove candidates that were detected due to a few points significantly lower than the average.

To detect the presence of red noise that mimics an exocomet transit, we calculated the correlation product between the exocomet model fit to the candidate transit and the light curve over all the considered subquarters. If the maximum of the correlation \max_1 is reached more than one day away from the transit time or if its second maximum \max_2 is significantly close to the first maximum ($(\max_1 - \max_2)/\max_1 < 0.15$), the candidate is removed, considering that in this case the noise in the light curve resembles the identified transit, which is therefore probably not a real exocomet transit, but background red noise.

The cases of the ramp-like light curves mimicking an exocometary transit have been discarded by calculating the difference between the fit with a comet model (Eq. (4)) and a fit with a simple ramp model defined as

$$f_{\text{ramp}}(t) = \begin{cases} Ct^2 + Dt + E & t < t_0 \\ (Ct^2 + Dt + E)(1 - Ke^{-\beta(t-t_0)}) & t > t_0 \end{cases} \quad (6)$$

We calculate $\Delta\chi_{\text{rp}}^2$ defined as

$$\Delta\chi_{\text{rp}}^2 = N \left(\frac{\sum_i (f_{\text{ramp}}(t_i) - F(t_i))^2}{\sum_i (f_{\text{exo}}(t_i) - F(t_i))^2} - 1 \right), \quad (7)$$

where $N = 120$ is the number of points in the fitted window.

The photometric events that are almost as well fit with the ramp model as with the exocomet model ($\Delta\chi_{\text{rp}}^2 < 10$) were discarded from the candidate list.

Finally, to identify the cases of a symmetrical transit that could be due to the passage of an exoplanet with a long period (a mono-transit), we calculated the difference $\Delta\chi_{\text{pl}}^2$ between the χ_{pl}^2 of a model of a planetary transit calculated using the equations of Mandel & Agol (2002) and the χ_{exo}^2 of an exocomet transit model (Eq. (4)). It can be written as

$$\Delta\chi_{\text{pl}}^2 = N \left(\frac{\sum_i (f_{\text{pl}}(t_i) - F(t_i))^2}{\sum_i (f_{\text{exo}}(t_i) - F(t_i))^2} - 1 \right). \quad (8)$$

The photometric events having $\Delta\chi_{\text{pl}}^2 < 5$ have a symmetrical shape and were therefore not taken into account in our final selection of exocomets. Such transits were taken into account to establish the list of symmetrical mono-transits (see Sect. 6.3).

However, this criterion is not used as a sharp cut to distinguish the exocomet transits from the symmetrical mono-transits but only as a help. The final distinction is made through visual inspection.

After applying these criteria, the number of candidate transits per quarter varies between 50 and 100 per quarter, leading to a total of 1349 transits. It then becomes possible to proceed for each of them to a visual inspection to know if the candidate is or is not an exocomet transit. This final inspection is necessary because, although the criteria put in place make it possible to eliminate the vast majority of false positives, there still remain obvious false positives. We note here that the criteria were deliberately chosen to be conservative so as not to accidentally eliminate a true exocomet transit. The number of criteria was chosen so that the final number of transits to be visually inspected would be no more than a few thousand. Otherwise, visual inspection would be impractical. It would be more inefficient and inaccurate to add extra criteria than to visually check the current list of 1349 transits.

To verify the robustness of the detected transits, we performed an additional test. We applied our detection procedure to the data after reversing the time. If the proportion of detected exocomets is the same in the reversed time data as in the original data, this may question the relevance of the detected transits. However, as this analysis is time-consuming, we performed this analysis on one-half of the quarters. Of the 17 quarters, we analyzed eight randomly chosen quarters: Q1, Q2, Q6, Q9, Q10, Q12, Q15, and Q16. We found only one transit that clearly mimics an exocomet transit: KIC_3129239 in Q9. Given that approximately 17 robust transits were detected in the Kepler data (see Sect. 6), we can infer from this study that about 10–15% of these transits can be false positives. Although the same number of detections in the time-reversed data would have disproven the exocomet hypothesis, the result of this test supports the idea that there is an astrophysical signal in the data that is consistent with exocomet transits.

6. A new catalog of transiting exocomets

In this section we present the final list of exocomet transits resulting from the selection procedure. We divided the list into three different tiers based on the likelihood of a genuine exocometary detection. In the first tier, the photometric events are the ones that we consider as the most likely due to the transit of an exocomet. The second tier is for the events possibly due to comets, while the third tier gathers the most symmetric dip in the light curve, for which the transit of an exoplanet remains a possibility.

Table 2. First tier catalog of exocometary transits.

Stars (KIC)	Time (BKJD)	Quarter	K($\times 10^{-4}$)	$\beta^{-1}(h)$	$\Delta T(h)$	AD($\times 10^{-4}$)	$\Delta\chi_{\text{pl}}^2$	$\Delta\chi_{\text{polynomial}}^2$
†3542116	161.5	1	4.74	6.5	9.92	3.71	18.32	279.74
†*3542116	742.6	8	4.17	1.36	6.05	4.12	0.6	348.83
†*3542116	792.9	8	8.62	4.16	4.77	5.88	15.57	285.27
†3542116	992.1	10	10.12	3.62	14.68	9.94	28.67	2993.05
†3542116	1175.8	12	12.98	3.40	9.35	12.15	8.58	2670.38
3662483	1132.1	12	6.81	6.38	6.87	4.49	12.98	204.52
4078638	355.5	4	83.21	6.30	3.22	33.30	47.89	358.90
4481029	172.5	2	9.08	8.57	4.64	3.80	14.31	235.74
5206257	925.0	10	10.98	4.75	4.47	6.69	5.04	322.26
5514200	670.1	7	34.48	12.13	7.27	15.53	19.15	155.20
*6263848	775.9	8	10.22	3.29	4.90	7.92	7.96	151.93
6927963	416.8	4	33.60	9.13	2.83	8.96	6.54	115.89
7660548	966.2	10	15.74	5.95	5.09	9.05	26.10	219.47
†8027456	1449.0	15	8.0	9.90	14.12	6.08	16.78	401.18
8738545	238.1	2	17.0	4.92	3.67	8.94	7.64	92.53
10484683	309.5	3	29.85	14.05	8.01	12.97	87.66	853.82
†11084727	1076.2	11	15.29	3.92	8.20	13.40	88.14	2368.80

Notes. Column 1 gives the name of the star by its KIC identifier. Columns 2 and 3 indicate the position of the transit. Columns 4 to 6 give the fit parameters associated with the exocomet model (Eq. (4)). Column 7 gives the absorption depth. Column 8 presents the difference in χ^2 between the cometary model and the planetary model (Mandel & Agol 2002) and column 9 between the cometary model and a second-order polynomial fit. The transits that were first eliminated by the criterion based on the maximum of the correlation product (see Sect. 5.2.4) are flagged with an asterisk (*). The five transits already known are flagged with a † symbol.

6.1. The first tier catalog of exocometary transits

The first tier catalog of detected exocometary transits is given in Table 2 and the plots of the corresponding light curves are given in Appendix B. It is particularly noteworthy that our neural network is able to find all the transits already identified by previous works: KIC 3542116 in Quarters 1, 10, and 12; KIC 11084727 as found by Rappaport et al. (2018); and KIC 8027456 as found by Kennedy et al. (2019). This independent retrieval of the same exocometary transits provides confidence in our procedure.

Nonetheless, after application of all the search procedure described above, it appeared that two transits of KIC 3542116 identified by Rappaport et al. (2018) in Quarters 8 were not in our list of detections. These transits were eliminated by the criterion based on the maximum of the correlation product that aims to remove the detections whose pattern occurs several times in the same subquarter (Sect. 5.2.4). This criterion also has the disadvantage of eliminating active cometary systems where at least two comet transits occurred in the same quarter, as is the case for KIC 3542166 in Quarter 8. Therefore, we made a visual inspection of all photometric events eliminated because of this criterion in order to recover possible obvious exocometary transits that were wrongly rejected. We thus found the two transits of KIC 3542116 in Quarter 8, and in addition identified a new transit in front of KIC 6263848. This last case does not correspond to a star with frequent cometary transits, but one that had been deleted due to an edge effect on the calculation of the correlation function. These two cases are flagged by an asterisk in Table 2.

For all of these detection, we calculate the $\Delta\chi_{\text{polynomial}}^2$ defined as

$$\Delta\chi_{\text{polynomial}}^2 = N \left(\frac{\sum_i (f_{\text{polynomial}}(t_i) - F(t_i))^2}{\sum_i (f_{\text{exo}}(t_i) - F(t_i))^2} - 1 \right), \quad (9)$$

where $f_{\text{polynomial}}$ is the second-order polynomial that best fits the detection. For all detections, we obtained large $\Delta\chi_{\text{polynomial}}^2$

(larger than about 100). This highlights that the detection cannot be interpreted just as noise patterns.

The properties of the parent stars of the identified exocomets as tabulated by Kepler are given in Table 3 (Brown et al. 2011). For KIC 3542116 and KIC 3662483, the values for T_{eff} , $\log_{10} g$, metallicity, and radius are from Zhang et al. (2025). From the T_{eff} and $\log_{10} g$ we can obtain an estimate of the stellar type (Gray 2008), which is given in the last column of Table 3. It appears that most of the stars are main-sequence stars, with the exception of KIC 4078638, which is likely a red giant. For these stars, the upper limits on the age estimate given by Zhang et al. (2025) are all above 10^9 years, with the exception of KIC 4481029 with an upper limit of 9×10^8 years and KIC 8027456 with an age estimate of 5×10^8 years.

6.2. The second tier catalog of possible exocometary

After visual inspection of the list of candidates produced by our algorithm, we came up with a list of interesting photometric events that can be due to transiting exocomets. This includes transits that pass through all the criteria described in Sect. 5, but whose shape or signal-to-noise ratio makes it difficult to decide whether they are actually transits. The resulting list of possible exocometary transits in this second tier catalog is given in Table 4, and the plots of the corresponding light curves are given in Appendix C.

The properties of the corresponding parent stars in the second tier catalog are given in Appendix D (Brown et al. 2011). Here again, estimates of the stellar types given in the last column of the table show that most of the stars are on or close to the main sequence, with the exception of KIC 2984102 and KIC 11153134, which have a $\log_{10} g$ below 3 and are therefore likely red giants. For all stars in this second tier catalog, the upper limit of the age estimate given by Zhang et al. (2025) is above 10^9 years, with the exception of KIC 2984102, KIC 5294231, and KIC 7183123 with

Table 3. Information on the stars in the first tier catalog.

Kepler ID	RA (J2000)	Dec (J2000)	Kepler Mag	Teff (K)	Log g (cm/s ²)	Metallicity	Radius (R _⊙)	Stellar type
3542116	19 22 52.939	+38 41 41.51	9.979	6766	4.208	-0.190	1.454	F2 V
3662483	19 42 09.199	+38 43 01.38	10.541	5848	3.945	0.107	1.786	G1 IV
4078638	19 46 16.611	+39 06 15.23	13.156	4783	2.758	0.150	8.331	K3 III
4481029	19 43 13.210	+39 32 32.03	11.303	8865	3.968	0.010	2.310	A0 IV
5206257	19 46 33.826	+40 22 57.97	12.051	6017	4.435	-0.105	1.052	F6 V
5514200	18 58 08.306	+40 46 29.32	14.377	5651	4.500	-0.243	0.955	G4 V
6263848	18 52 30.566	+41 36 33.34	13.172	5850	4.365	-0.224	1.137	G2 V
6927963	18 56 15.946	+42 27 52.20	13.807	5711	4.441	-0.012	1.030	G4 V
7660548	18 47 22.807	+43 23 05.39	13.094	6158	4.369	-0.050	1.148	F5 V
8027456	19 25 15.838	+43 51 33.55	9.697	8732	3.766	0.065	2.974	A0 IV
8738545	18 58 37.262	+44 55 15.24	13.227	5091	3.581	-0.162	3.101	K1 IV
10484683	19 47 54.377	+47 40 14.16	12.588	6326	4.543	-0.009	0.934	F5 V
11084727	19 28 41.191	+48 41 15.14	9.987	6762	4.067	-0.152	1.726	F3 V

Table 4. Second tier catalog of possible exocometary transits (same legend as in Table 2).

Stars (KIC)	Time (BKJD)	Quarter	K($\times 10^{-4}$)	β^{-1} (h)	ΔT (h)	AD($\times 10^{-4}$)	$\Delta\chi^2_{\text{pl}}$	$\Delta\chi^2_{\text{polynomial}}$
2984102	345.7	3	36.42	15.9	10.12	17.14	61.90	430.26
4826941	385.8	4	21.58	15.22	7.85	8.69	3.76	126.17
5195476	1508.8	16	16.80	5.09	3.67	8.31	3.42	45.92
5294231	411.4	4	4.92	3.26	2.13	2.36	31.78	285.46
5511746	1105.8	12	13.31	12.33	5.06	4.48	2.41	44.54
5514277	670.4	7	63.11	13.57	9.88	32.64	13.96	160.95
5534083	1167.7	12	9.34	10.45	3.33	2.55	7.22	43.16
6119914	747.2	8	31.41	9.36	6.63	15.95	3.54	50.65
7183123	336.4	3	8.43	8.36	5.98	4.31	28.32	321.42
7201827	409.9	4	43.64	4.83	2.83	19.34	5.77	48.98
7220968	1325.7	14	18.60	5.50	6.23	12.61	2.61	46.42
7611858	1275.8	14	6.46	9.77	5.19	2.67	7.63	109.47
7669613	315.9	3	13.24	13.66	6.31	4.89	12.77	121.72
8111649	934.0	10	83.04	23.69	10.28	29.24	16.36	123.61
8144412	639.1	7	86.39	12.00	2.95	18.84	7.06	140.49
8374877	769.1	8	7.78	7.04	8.46	5.44	13.67	300.35
8489495	1351.8	14	16.43	8.44	15.04	13.67	10.70	47.92
8625941	1505.0	16	5.16	8.44	9.80	3.54	8.10	66.25
8683719	1340.2	14	19.21	6.43	14.25	17.12	7.59	195.79
8738569	1462.3	15	13.18	8.30	4.30	5.33	8.84	56.11
8843356	702.2	7	8.84	10.54	2.99	2.18	4.66	46.51
9025688	1078.2	11	13.29	7.51	3.57	5.03	58.03	58.03
9388975	680.3	7	18.41	3.36	2.47	9.59	6.76	60.28
9531080	748.0	8	10.23	9.31	6.56	5.18	16.13	53.73
10329957	1385.1	15	18.47	12.90	4.10	5.03	9.62	64.04
11124157	704.2	7	15.01	10.93	2.00	2.51	11.48	45.20
11153134	739.7	8	18.93	3.15	7.99	17.44	11.42	345.83
11246607	696.4	7	18.06	15.10	8.84	8.00	7.04	165.29
11515196	644.3	7	16.40	7.98	3.84	6.26	34.15	138.30
11701976	931.7	10	23.79	11.98	2.70	4.80	13.24	40.42

age estimates of 5×10^8 , 1×10^8 (with an upper limit of 3×10^8), and 9×10^8 years, respectively.

6.3. The third tier catalog of symmetric transits

After visual inspection of the list of candidates produced by our algorithm, we also came up with a list of interesting

photometric events that appears to be symmetrical. This is not surprising as some of the exocomet transits in the library used for training the network are indeed symmetrical (see Figs. 2 and 3 in LdE99a). However, we cannot exclude that these are due to mono-transits of exoplanets or to false positives mimicking planetary or cometary transits. Several scenarios might cause such false positives, including diluted or undiluted eclipsing

binaries. To identify these cases it would be necessary to carry out additional analyses that are beyond the scope of the present paper (see, e.g., [Crossfield et al. 2016](#)). Finally, although an exocometary transit classification cannot be made, these detections deserve to be mentioned. They are given in the third tier catalog of symmetric transits listed in Appendix E.

We compared these detected photometric events with the lists of single transits published by [Huang et al. \(2013\)](#), [Wang et al. \(2015\)](#), [Foreman-Mackey et al. \(2016\)](#), and [Herman et al. \(2019\)](#). It appears that four events in our list were already identified by [Foreman-Mackey et al. \(2016\)](#) and [Herman et al. \(2019\)](#). The transit in front of KIC 8410697 was found in both studies. It is interpreted as being due to an exoplanet with a radius that is 0.7 times that of Jupiter. The photometric event of KIC 10321319 was found by [Foreman-Mackey et al. \(2016\)](#) and is interpreted as being due to the transit of an exoplanet of 0.16 Jupiter radius. The photometric event of KIC 6196417 was found by [Herman et al. \(2019\)](#) and interpreted as being due to the transit of an exoplanet of about 0.7 Jupiter radius. The transit in front of KIC 10668646 at the time of BKJD=1449.3 was already identified by [Foreman-Mackey et al. \(2016\)](#), but the exoplanet transit scenario was rejected because of a centroid shift in the data. However, we also identified another photometric event in the same target at BKJD=196.3 with a different shape and a lower absorption depth. These few examples show that our list of newly identified photometric events that could be due to transits of exocomet with symmetrical shape of the light curve deserves further investigations, which are beyond the scope of the present paper.

7. Conclusion

Although the Kepler mission ended more than 10 years ago, the available data is still worth deep data mining in search of new discoveries. Here we presented a new search for exocometary transits in the Kepler light curves. We used a machine learning technique with a neural network trained using a library of theoretical exocomet transit light curves inherited from the work of LdE99b. After parsing the light curves of close to 200 000 stars through the neural network, despite the application of several filters to eliminate most of the false positive, a visual inspection of the outcome of the network was still needed. We ended up with three catalogs of interesting objects. The first tier catalog is composed of a total of 17 exocometary transits, including seven previously identified transits and ten new transits in front of ten different stars. The second tier catalog is a list of 30 photometric events that appear to be of second quality and that we qualify as possible exocometary transits. Finally, the third tier catalog presents the list of interesting photometric events that are symmetrical and may be due to transits of either an exoplanet or an exocomet with a periastron at 90° from the line of sight.

The complete Kepler data represent a four-year photometry survey of more than 170 000 distinct stars. However, as noted in [Kennedy et al. \(2019\)](#), not all stars were observed for the full duration of the mission. As a result, we can consider 150 000 as the approximate number of equivalent stars that were observed for the full duration of the mission. In other words, our analysis covers the equivalent of about 600 000 star-years. LdE99a showed that, for a survey covering 30 000 star-years of targets with Solar System–like cometary activity and achieving a photometric precision of a few 10^{-4} , the expected number of exocomet transit detections is on the order of ten. Therefore, assuming that each star harbors a cometary system similar to our own, we could expect roughly 200 detections with a perfect search in the Kepler

data. Finding a few dozen exocometary transits can therefore be considered a satisfactory result in agreement with what could be expected using reasonable estimates of the algorithm efficiency and cometary activity of the surveyed stars.

LdE99a showed that, for a survey covering 30 000 star-years of targets with Solar System–like cometary activity and achieving a photometric precision of a few 10^{-4} , the expected number of exocomet transit detections is on the order of ten.

Unfortunately, the Kepler stars are rather faint, and our catalogs contain stars with magnitude between 9.7 and 15.9. It is therefore not possible to undertake a spectroscopic follow up with the hope of confirming the exocometary nature of the photometric events. All the more because with only one transit per star in about four years, exocometary transits are rare and not predictable in time. However, we can anticipate that the same techniques could be applied to search for exocometary transits in the current TESS data and in upcoming PLATO data. For these two instruments, the surveyed star in the input catalog are significantly brighter than the Kepler stars. We can thus expect that the exocometary systems to be discovered in the near future can be studied in detail, particularly through spectroscopy.

Previous searches ([Rappaport et al. 2018](#); [Kennedy et al. 2019](#)) provided a list of three stars that seem to be rather young. This supported the anticipated conclusion that cometary activity is likely correlated with stellar age, with an expected decline over time. However, our catalogs do not support this idea; we found a wide spread of age for both the first and second tier catalogs, with possible exocometary transits in front of stars classified as red giants. The issue of cometary activity with age thus remains open, and further studies will be needed to clarify this important question of the evolution of planetary systems.

Note added to proofs. After this manuscript was submitted we became aware of a paper by [Norazman et al. \(2025\)](#) that used machine learning techniques to search for exocometary transits in sectors 1–22 of the TESS light curves. They found three additional transits compared to those already known, which highlights the utility of this technique for the search of such transits.

Acknowledgements. We thank Dr. Neda Heidari for her help in identifying already known mono-transits. We acknowledge support from the CNES (Centre national d'études spatiales, France). This work has made use of the Infinity Cluster hosted by Institut d'Astrophysique de Paris. We thank Dr. Guilhem Lavaux for his invaluable help in making efficient use of the GPUs in this cluster. This paper includes data collected by the Kepler mission and obtained from the MAST data archive at the Space Telescope Science Institute (STScI). Funding for the Kepler mission is provided by the NASA Science Mission Directorate. STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5–26555. This research has made use of the NASA Exoplanet Archive, which is operated by the California Institute of Technology, under contract with the National Aeronautics and Space Administration under the Exoplanet Exploration Program.

References

- Abadi, M., Barham, P., Chen, J., et al. 2016, *TensorFlow: A System for Large-Scale Machine Learning*
- Bendahan-West, R., Kennedy, G. M., Brown, D. J. A., & Strøm, P. A. 2025, *MNRAS*, **537**, 229
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, *Science*, **327**, 977
- Brown, T. M., Latham, D. W., Everett, M. E., & Esquerdo, G. A. 2011, *AJ*, **142**, 112
- Chintarunguangchai, P., & Jiang, I.-G. 2019, *PASP*, **131**, 064502
- Cho, K., van Merriënboer, B., Gulcehre, C., et al. 2014, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, eds. A. Moschitti, B. Pang, & W. Daelemans (Doha, Qatar: Association for Computational Linguistics), 1724

- Crossfield, I. J. M., Ciardi, D. R., Petigura, E. A., et al. 2016, *ApJS*, **226**, 7
- Dattilo, A., Vanderburg, A., Shallue, C. J., et al. 2019, *AJ*, **157**, 169
- Ferlet, R., Hobbs, L. M., & Vidal-Madjar, A. 1987, *A&A*, **185**, 267
- Foreman-Mackey, D., Morton, T. D., Hogg, D. W., Agol, E., & Schölkopf, B. 2016, *AJ*, **152**, 206
- Grady, C. A., Brown, A., Welsh, B., et al. 2018, *AJ*, **155**, 242
- Gray, D. F. 2008, *The Observation and Analysis of Stellar Photospheres* (Cambridge, UK: Cambridge University Press)
- Herman, M. K., Zhu, W., & Wu, Y. 2019, *AJ*, **157**, 248
- Hochreiter, S., & Schmidhuber, J. 1997, *Neural Comput.*, **9**, 1735
- Hu, J., Shen, L., & Sun, G. 2018, in 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT: IEEE), 7132
- Huang, X., Bakos, G. Á., & Hartman, J. D. 2013, *MNRAS*, **429**, 2001
- Jarrett, K., Kavukcuoglu, K., Ranzato, M., & LeCun, Y. 2009, in 2009 *IEEE 12th International Conference on Computer Vision*, 2146
- Kennedy, G. M., Hope, G., Hodgkin, S. T., & Wyatt, M. C. 2019, *MNRAS*, **482**, 5587
- Kiefer, F., Lecavelier des Etangs, A., Boissier, J., et al. 2014a, *Nature*, **514**, 462
- Kiefer, F., Lecavelier des Etangs, A., Augereau, J.-C., et al. 2014b, *A&A*, **561**, L10
- Kiefer, F., Lecavelier des Étangs, A., Vidal-Madjar, A., et al. 2017, *A&A*, **608**, A132
- Kiefer, F., Van Grootel, V., Lecavelier des Etangs, A., et al. 2023, *A&A*, **671**, A25
- Kingma, D. P., & Ba, J. 2015, Adam: A Method for Stochastic Optimization
- Lecavelier des Etangs, A. 1999, *A&AS*, **140**, 15
- Lecavelier des Etangs, A., Vidal-Madjar, A., & Ferlet, R. 1999, *A&A*, **343**, 916
- Lecavelier des Etangs, A., Cros, L., Hébrard, G., et al. 2022, *Nature*, **12**, 5855
- Luk'yanyk, I., Kulyk, I., Shubina, O., et al. 2024, *A&A*, **688**, A65
- Mandel, K., & Agol, E. 2002, *ApJ*, **580**, L171
- Miles, B. E., Roberge, A., & Welsh, B. 2016, *ApJ*, **824**, 126
- Montgomery, S. L., & Welsh, B. Y. 2012, *PASP*, **124**, 1042
- Norazman, A., Kennedy, G. M., Cody, A. M., et al. 2025, *MNRAS*, **542**, 1486
- O'Malley, T., Bursztein, E., Long, J., et al. 2019, KerasTuner
- Pavlenko, Y., Kulyk, I., Shubina, O., et al. 2022, *A&A*, **660**, A49
- Rappaport, S., Vanderburg, A., Jacobs, T., et al. 2018, *MNRAS*, **474**, 1453
- Rebollido, I., Eiroa, C., Montesinos, B., et al. 2020, *A&A*, **639**, A11
- Shallue, C. J., & Vanderburg, A. 2018, *AJ*, **155**, 94
- Strøm, P. A., Bodewits, D., Knight, M. M., et al. 2020, *PASP*, **132**, 101001
- Thompson, S. E., Caldwell, D. A., Jenkins, J. M., et al. 2016, *Kepler Data Release 25 Notes*, *Kepler Science Document KSCI-19065-002*, eds. J. Jenkins, & M.R. Haas, 3
- Wang, J., Fischer, D. A., Barclay, T., et al. 2015, *ApJ*, **815**, 127
- Welsh, B. Y., & Montgomery, S. 2013, *PASP*, **125**, 759
- Welsh, B. Y., & Montgomery, S. L. 2018, *MNRAS*, **474**, 1515
- Zhang, B., Huang, Y., Beers, T. C., et al. 2025, *ApJS*, **277**, 6
- Zieba, S., Zwintz, K., Kenworthy, M. A., & Kennedy, G. M. 2019, *A&A*, **625**, L13
- Zucker, S., & Giryes, R. 2018, *AJ*, **155**, 147

Appendix A: Performance of the algorithm over all the quarters

Table A.1: Performance of the neural network.

Quarter	Subquarter	Accuracy	Recall	Precision	FPR	AUC	MAE
1	0	0.8945	0.7910	0.9975	0.0020	0.9888	2.3404
2	0	0.8665	0.7347	0.9977	0.0017	0.9907	2.2237
2	1	0.8616	0.7245	0.9982	0.0013	0.9900	2.0792
2	2	0.8743	0.7498	0.9984	0.0012	0.9893	2.0455
3	0	0.9108	0.8240	0.9972	0.0023	0.9896	2.0329
3	1	0.9068	0.8162	0.9969	0.0025	0.9905	2.1186
3	2	0.8903	0.7835	0.9964	0.0028	0.9895	2.1638
4	0	0.8862	0.7745	0.9972	0.0022	0.9909	2.3073
4	1	0.9017	0.8065	0.9961	0.0032	0.9924	2.4661
4	2	0.8854	0.7728	0.9974	0.0020	0.9920	1.9370
5	0	0.8994	0.8013	0.9969	0.0025	0.9919	1.9726
5	1	0.8898	0.7818	0.9972	0.0022	0.9916	2.1857
5	2	0.8951	0.7923	0.9973	0.0022	0.9919	2.5873
6	0	0.8979	0.7972	0.9983	0.0013	0.9906	2.2020
6	1	0.8908	0.7833	0.9979	0.0017	0.9911	1.9720
6	2	0.8860	0.7738	0.9976	0.0018	0.9910	2.4570
7	0	0.8792	0.7605	0.9972	0.0022	0.9899	2.1862
7	1	0.8752	0.7522	0.9978	0.0017	0.9909	1.7915
7	2	0.9128	0.8267	0.9986	0.0012	0.9935	2.1101
8	0	0.9004	0.8018	0.9988	0.0010	0.9911	2.0729
8	1	0.9027	0.8065	0.9986	0.0012	0.9926	2.1139
9	0	0.9068	0.8157	0.9976	0.0020	0.9930	2.1293
9	1	0.9017	0.8062	0.9965	0.0028	0.9916	2.3157
9	2	0.8983	0.7992	0.9969	0.0025	0.9918	2.3348
9	3	0.8798	0.7610	0.9980	0.0015	0.9916	2.2389
10	0	0.8755	0.7532	0.9971	0.0022	0.9896	2.7444
10	1	0.8758	0.7538	0.9971	0.0022	0.9889	2.0156
10	2	0.8829	0.7672	0.9983	0.0013	0.9890	2.7258
11	0	0.8830	0.7685	0.9968	0.0025	0.9908	2.0722
11	1	0.8941	0.7900	0.9977	0.0018	0.9913	1.9576
11	2	0.8907	0.7828	0.9981	0.0015	0.9921	2.4971
12	0	0.8876	0.7767	0.9981	0.0015	0.9921	2.3743
12	1	0.9028	0.8075	0.9977	0.0018	0.9930	2.3531
12	2	0.9038	0.8103	0.9967	0.0027	0.9913	2.4029
13	0	0.8981	0.8000	0.9952	0.0038	0.9913	2.8353
13	1	0.8431	0.6875	0.9981	0.0013	0.9872	2.4796
13	2	0.9037	0.8117	0.9947	0.0043	0.9915	2.5127
14	0	0.8654	0.7342	0.9955	0.0033	0.9877	2.4772
14	1	0.8801	0.7632	0.9961	0.0030	0.9897	2.4096
14	2	0.8788	0.7592	0.9978	0.0017	0.9915	3.1279
15	0	0.8363	0.6743	0.9973	0.0018	0.9869	2.5724
15	1	0.8378	0.6780	0.9966	0.0023	0.9826	3.3750
15	2	0.8490	0.7000	0.9972	0.0020	0.9862	3.1479
16	0	0.8238	0.6483	0.9987	0.0008	0.9870	2.4968
16	1	0.8665	0.7355	0.9966	0.0025	0.9868	2.1849
16	2	0.8445	0.6907	0.9976	0.0017	0.9858	2.4707
17	0	0.8744	0.7510	0.9971	0.0022	0.9899	2.3075

Appendix B: Cometary transit light curves

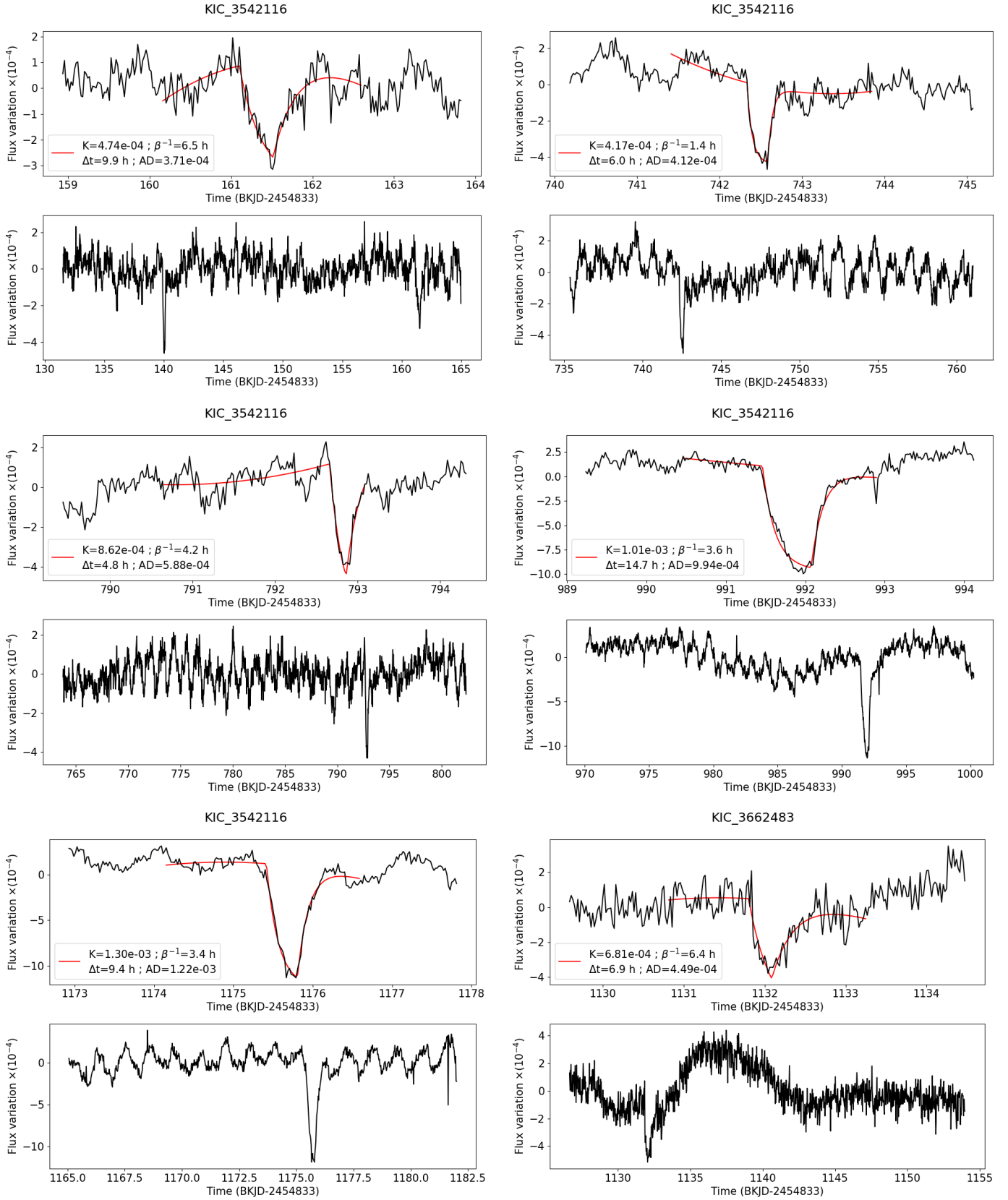


Fig. B.1: Plots of the light curves of the detected cometary transits in the first tier catalog. For each transit, the light curve is shown in the top panels over 5 days centered on the transit time and with an exocomet transit fit superimposed (red line). The bottom panels show the light curves over a larger time frame.

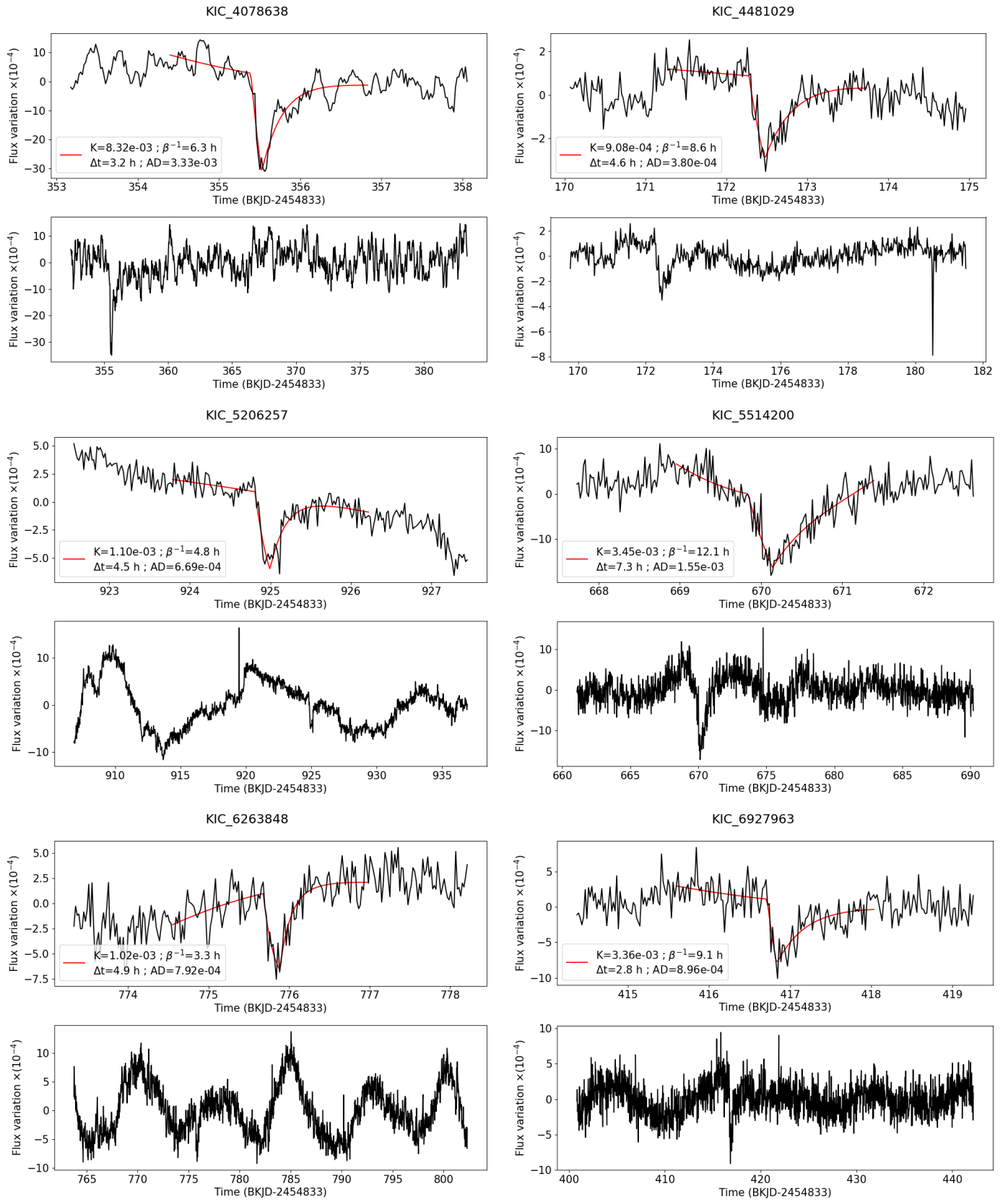


Fig. B.1: Continued

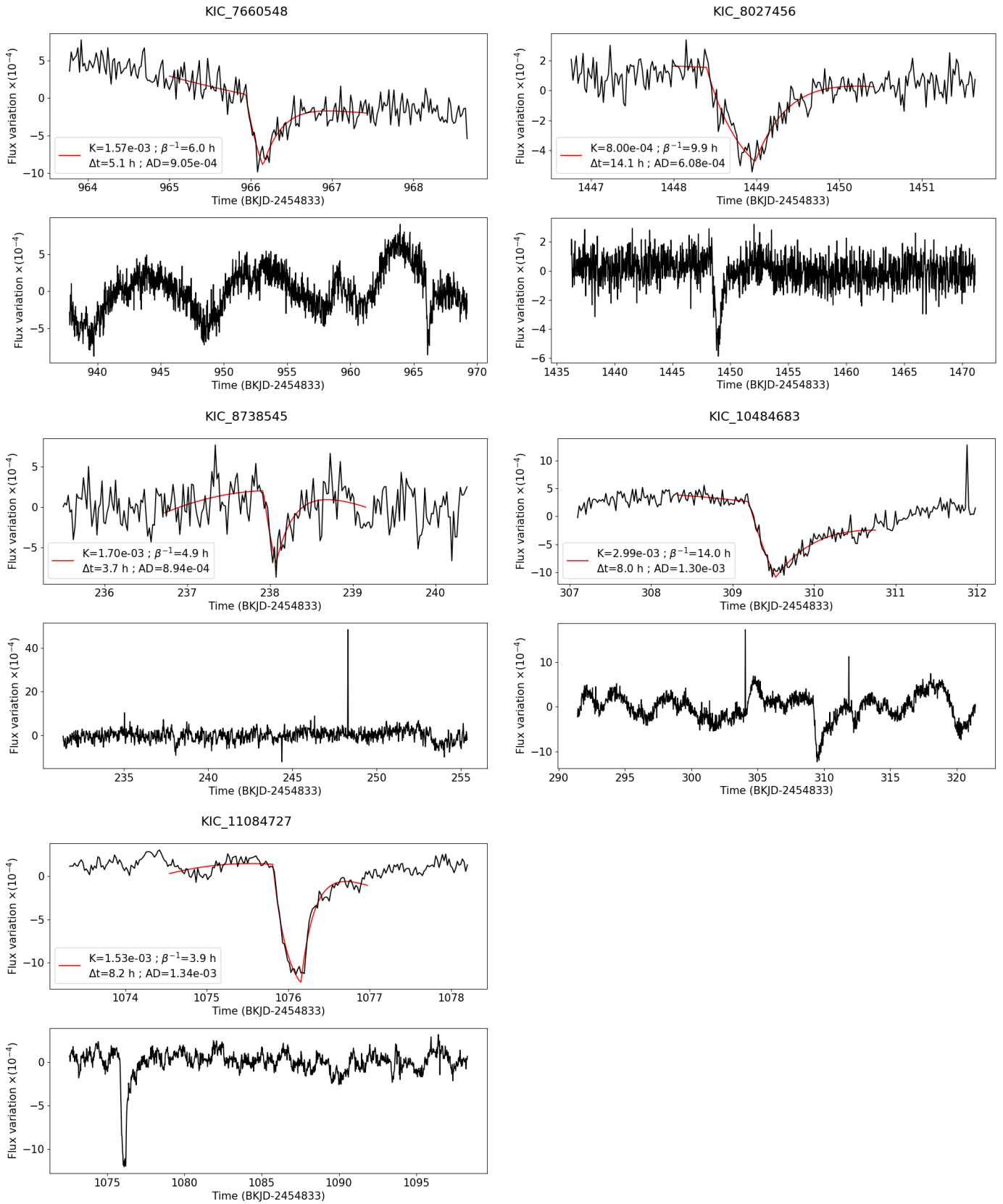


Fig. B.1: Continued

Appendix C: Possible exocometary transit light curves

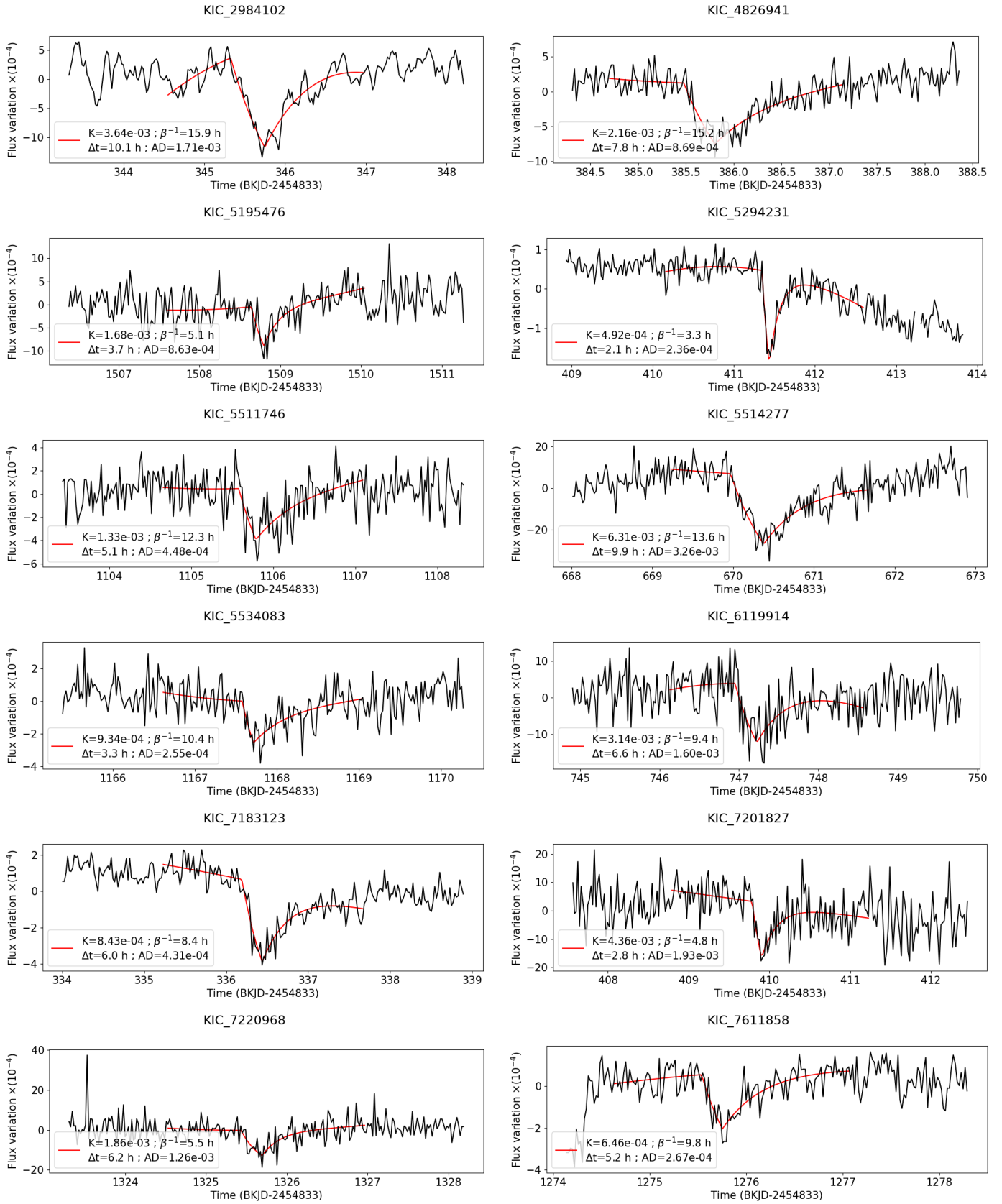


Fig. C.1: Plots of the light curves of the possibly detected cometary transits in the second tier catalog. The plots cover 5 days and are centered on the transit time and with an exocomet transit fit superimposed (red line).

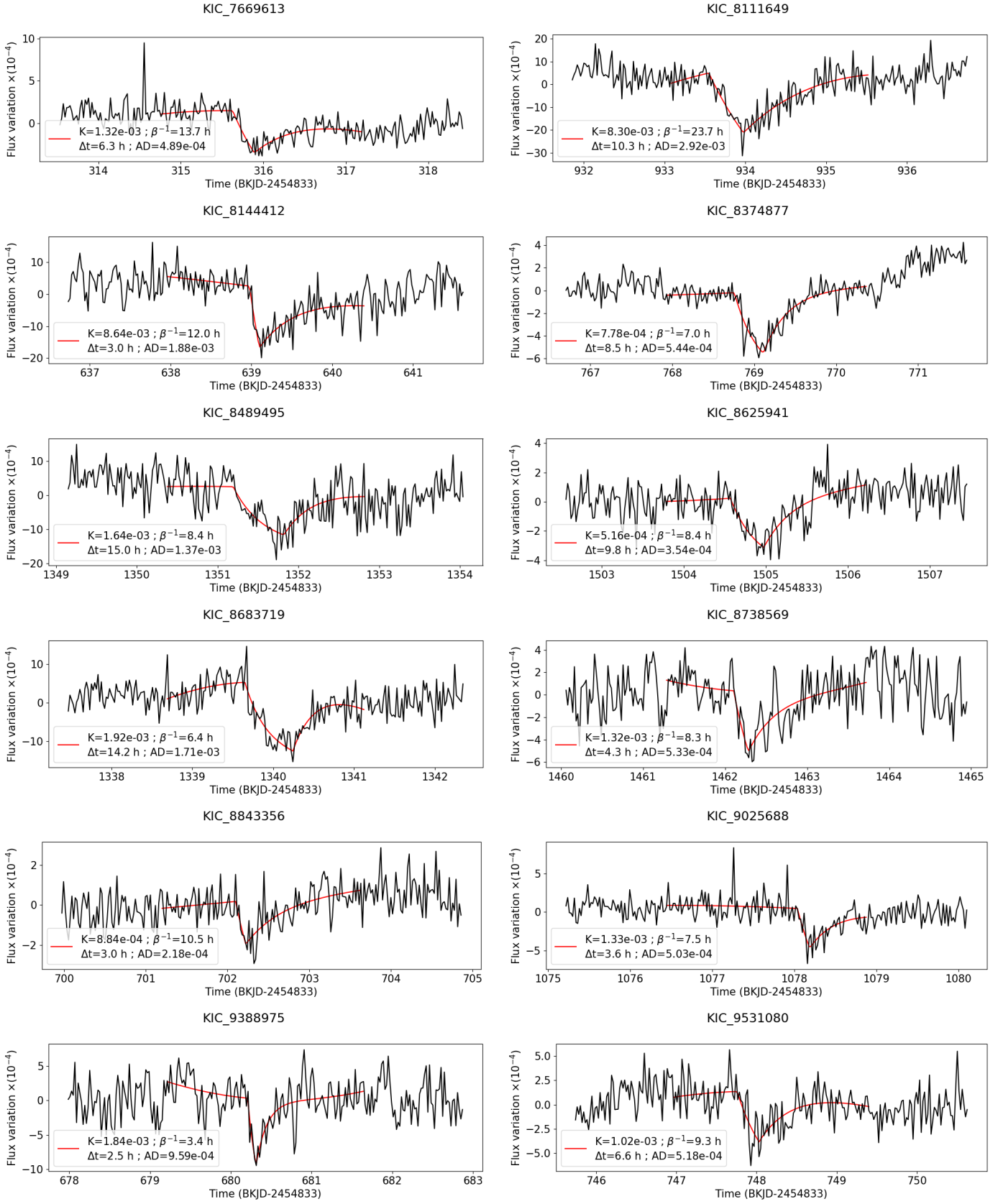


Fig. C.1: Continued

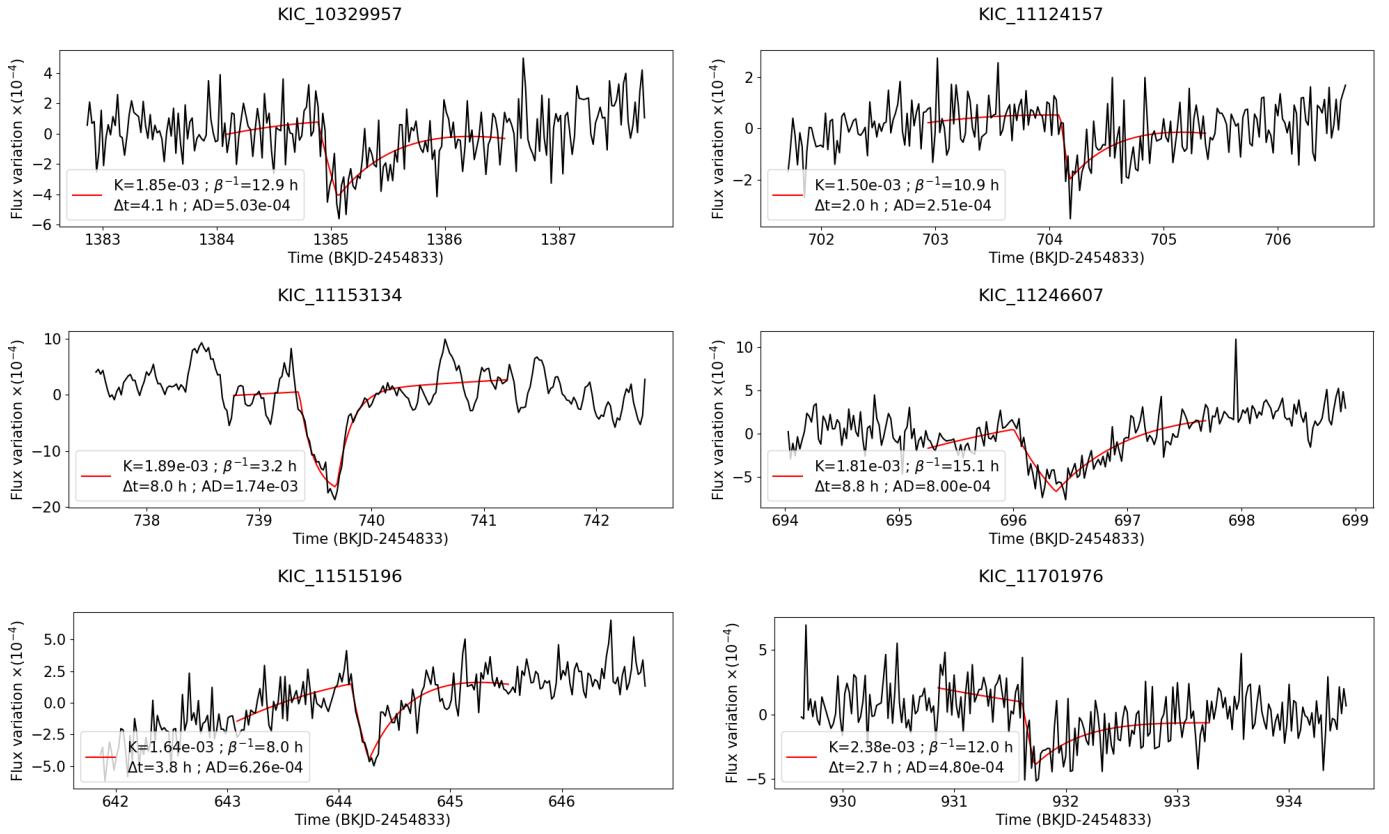


Fig. C.1: Continued

Appendix D: Information on the stars of the second tier catalog

Table D.1: Information on the stars in the second tier catalog.

Kepler ID	RA (J2000)	Dec (J2000)	Kepler Mag	Teff (K)	Log g (cm/s²)	Metallicity	Radius (R_⊙)	Stellar type
2984102	19 20 24.302	+38 09 41.26	11.271	5064	2.543	-0.167	10.715	K1III
4826941	19 16 42.927	+39 58 17.69	13.917	5748	4.425	-0.257	1.051	G3V
5195476	19 36 47.573	+40 19 41.48	13.925	4848	3.504	-0.214	3.406	K2IV
5294231	19 46 54.941	+40 26 00.17	9.842	9552	4.082	0.032	2.107	A0V
5511746	18 53 09.113	+40 46 26.87	13.216	5928	4.266	-0.277	1.290	G1V
5514277	18 58 16.380	+40 47 18.02	15.366	5401	4.601	-0.242	0.824	G7V
5534083	19 27 57.571	+40 45 16.38	12.602	6416	4.186	-0.261	1.458	F4V
6119914	19 22 29.443	+41 24 30.31	15.882	6167	4.819	-0.036	0.660	F5V
7183123	19 02 20.712	+42 47 30.55	11.207	8005	3.703	-0.145	3.020	A0IV
7201827	19 29 03.487	+42 44 03.41	15.409	5562	4.815	-0.098	0.638	G5V
7220968	19 48 59.026	+42 44 39.52	15.463	5015	4.483	0.35	0.921	K1V
7611858	19 32 36.857	+43 17 11.54	11.187	5684	4.477	-0.315	0.984	G4V
7669613	19 06 56.856	+43 18 44.17	13.044	5754	4.246	-0.382	1.313	G3V
8111649	19 45 47.110	+43 58 36.44	15.776	6080	4.576	-0.044	0.888	F6V
8144412	18 47 41.122	+44 04 04.40	15.213	5721	4.428	0.043	1.047	G3V
8374877	19 40 27.934	+44 19 09.62	12.083	6510	4.021	-0.505	1.809	F4V
8489495	19 18 40.234	+44 33 44.75	15.214	4912	4.328	0.068	1.112	K2V
8625941	19 28 55.855	+44 44 53.77	12.783	5689	4.602	-0.318	0.844	G4V
8683719	19 18 58.464	+44 49 28.02	14.717	5989	4.398	-0.035	1.100	G0V
8738569	18 58 41.198	+44 57 51.12	12.771	4907	3.548	0.173	3.227	K2IV
8843356	19 58 40.882	+45 01 02.64	12.039	5419	3.694	-0.387	2.687	G7IV
9025688	19 32 42.634	+45 23 41.57	13.382	6445	4.165	-1.598	1.500	F4V
9388975	18 56 00.802	+45 57 05.90	14.292	5163	4.493	-0.938	0.923	K0V
9531080	19 35 07.706	+46 09 35.64	13.861	6280	4.268	-0.346	1.306	F5V
10329957	19 10 18.223	+47 27 04.43	13.58	5002	4.567	-0.024	0.824	K1V
11124157	18 56 39.605	+48 45 44.21	12.244	4775	3.954	0.071	1.869	K3IV
11153134	19 53 11.854	+48 46 20.82	12.007	4941	2.719	-0.037	8.746	K2III
11246607	19 32 21.149	+48 58 47.78	13.169	6251	4.387	-0.253	1.127	F5V
11515196	19 42 06.725	+49 26 57.88	13.568	4754	4.15	0.345	1.403	K3V
11701976	19 02 02.592	+49 49 17.65	13.403	6210	4.398	-0.014	1.111	F5V

Appendix E: Information on the stars of the third tier catalog

Table E.1: Third tier catalog of symmetric transits that could be either exocometary transits or exoplanet mono-transits.

Stars (KIC)	Time (BKJD)	Quarter	$K(\times 10^{-4})$	$\beta^{-1}(\text{h})$	$\Delta T(\text{h})$	$AD(\times 10^{-4})$	$\Delta\chi_{\text{pl}}^2$	$\Delta\chi_{\text{polynomial}}^2$
2983000	759.1	8	18.38	0.84	16.33	18.38	-1.94	483.82
2993038	550.9	6	15.22	2.46	4.72	12.99	-1.57	1165.8
3222471	673.1	7	97.7	5.49	14.59	90.83	-70.29	2436.25
3346436	270.4	3	8.48	3.03	3.05	5.38	3.42	182.43
3755854	1492.3	16	3.53	1.07	6.11	3.51	1.03	45.73
5184479	534.6	5	8.87	1.04	19.61	8.87	-11.23	316.53
5184479	1203.0	13	12.17	1.97	17.91	12.17	-3.69	409.71
5305217	673.1	7	33.0	5.46	20.87	32.28	-7.41	170.59
5456365	270.8	3	35.42	3.85	10.37	33.03	-2.28	323.73
5905878	1343.3	14	2.25	1.46	7.55	2.24	-7.35	353.12
5905878	1441.3	15	0.89	1.49	8.14	0.89	-2.01	60.16
5967153	1394.5	15	5.06	4.05	7.9	4.34	-22.09	141.91
5975275	213.3	2	9.78	2.57	10.57	9.62	2.7	179.4
^b 6186417	959.1	10	33.19	1.09	18.42	33.19	-2.36	645.89
6387193	651.7	7	78.79	2.73	12.43	77.96	-48.64	3817.28
6515488	1420.4	15	28.85	2.98	9.54	27.67	-7.02	359.44
6804821	1009.1	11	12.97	1.6	22.41	12.97	-6.09	317.96
7047396	298.8	3	11.35	2.0	5.57	10.65	-4.56	305.5
7105703	1509.0	16	20.54	2.86	6.53	18.45	-4.32	186.11
7213651	1239.2	13	21.53	1.66	6.78	21.17	-9.23	214.24
7465971	342.7	3	11.14	1.14	9.75	11.14	-1.07	200.89
7983622	471.9	5	65.06	4.63	13.76	61.73	-15.71	732.08
8005892	608.5	6	96.11	1.2	9.78	96.09	-75.32	9905.87
8007462	1458.6	15	2.44	2.56	5.4	2.14	-7.15	93.52
8159297	440.7	4	18.1	1.87	4.62	16.57	-1.87	132.24
8313257	1149.1	12	25.05	2.17	17.4	25.05	-8.72	320.72
^{a,b} 8410697	542.5	6	54.16	1.0	17.87	54.16	-12.07	5694.83
8496108	277.7	3	19.41	9.57	3.83	6.4	-6.55	62.5
8617888	171.8	2	55.2	1.83	5.44	52.4	-4.99	149.3
9016734	666.0	7	61.61	5.57	13.7	56.35	-48.25	1624.94
9016734	1230.1	13	27.52	5.5	14.31	25.48	-7.51	522.79
9413755	1564.2	17	19.5	1.63	4.4	18.18	-16.75	217.06
9775416	1398.1	15	5.04	4.99	10.63	4.44	-2.36	141.64
10024862	1494.0	16	20.95	1.68	13.29	20.94	-2.57	121.87
10157075	369.5	4	26.75	1.93	5.57	25.27	-2.17	114.74
10192453	757.5	8	18.64	5.63	1.73	4.94	-29.2	76.45
^a 10321319	554.5	6	4.76	4.77	13.56	4.48	-29.18	174.87
10334763	549.7	6	7.23	0.74	10.88	7.23	-3.39	412.26
10397849	506.5	5	19.93	2.35	7.95	19.25	0.15	289.02
10450889	239.6	2	60.03	1.58	2.35	46.44	1.45	192.69
10556420	756.3	8	3.59	0.64	6.03	3.59	-1.17	364.07
10643786	1421.2	15	12.62	3.23	5.12	10.04	3.16	81.51
10668646	196.3	2	42.81	0.93	8.74	42.8	-20.57	1376.58
^a 10668646	1449.3	15	57.08	1.36	8.64	56.98	-38.4	3632.55
10909733	898.7	9	3.41	1.37	6.17	3.37	4.64	67.65
11561379	137.7	1	13.51	2.11	7.45	13.12	-1.64	103.03
11921843	1234.8	13	3.85	0.62	14.83	3.85	-2.49	287.33
12066509	632.3	7	38.71	0.78	12.41	38.71	-0.31	1883.85
12644038	584.7	6	36.26	1.49	6.37	35.77	3.56	199.23

Notes. Four transits were already identified by [Foreman-Mackey et al. \(2016\)](#) (flagged with “a” in the first column) and [Herman et al. \(2019\)](#) (flagged with “b”).