AUTOMATING THE TRANSIT LEAST SQUARES WITH A PRESELECTION ALGORITHM

M. WINZEN

Summary

In 2019 René Heller and Michael Hippke published the so-called Transit Least Squares (TLS), which compared to the Box Fitting Least Squares also considers the stellar limb darkening and planetary ingress and egress. This improvement allowed finding another 18 exoplanet candidates by scanning the Kepler light curve data. Also in 2019 Hippke showed that detrending with Tukey's biweight filter and using a running window of three times the duration of a central transit yields the highest recovery rates for injected transits of simulated data.

So far, transit durations derived from manually chosen trial periods of circular orbits were used as input to the biweight filter (as for Kepler 160 TLS analysis [1]). The Preselection Algorithm directly identifies potential transits in the light curve in order to use three times the transit duration as running window for the detrended light curve, which is analyzed by the TLS afterwards. The advantage of firstly applying the biweight filter to enable the TLS to find also planets that usually would be hidden by noise is automized by using durations from transit candidates of Preselection Algorithm as input for the running window.

Transit candidates of the light curve adaptive Preselection Algorithm are also compared with each other by depth and duration in order to derive potential exoplanet periods. The result of the TLS can be compared to the potential exoplanet periods of the Preselection Algorithm. Other than the TLS the Preselection Algorithm is not sensitive to data gaps, which permits to automate the validation of TLS results regarding false positives due to data gaps.

Preselection Algorithm

The Preselection Algorithm consists of two steps:

1. Transit Scan
2. Transit Pairing (Creating Exoplanet Candidates) Transit Scan

The process of finding an exoplanet starts with identifying flux drops in the light curve that could potentially be a planet transit. The characteristics of a planet transit are shown in Fig. 1.

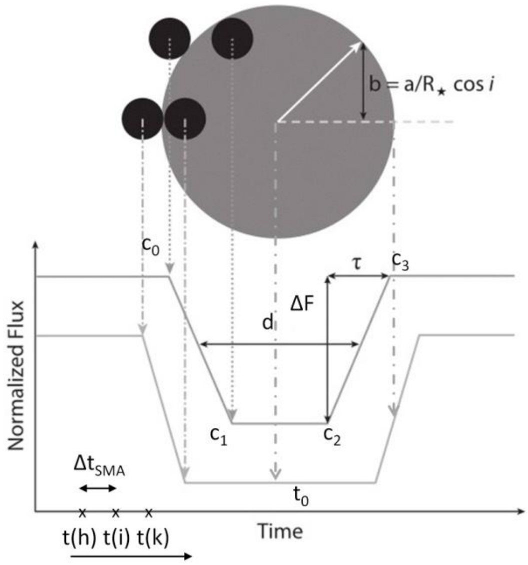

Fig. 1: Theoretical transiting exoplanet light curve [2]

The two main characteristics of a planet transit are the transit depth ΔF and the transit duration d. Furthermore, the transit can be modeled by the 4 basic points c_0, c_1, c_2, c_3, in which the light curve gradient is changing. In order to identify a transit and the 4 basic points the 3 running indices h, i, k are used separated by an approximate time interval of

$$\Delta t_{SMA} = \Delta t_{meas} N_{SMA}$$

Eq. 1

where Δt_{meas} is the regular interval between two measurements (for Kepler telescope $\Delta t_{meas} \cong 0.02$ [days]) and N_{SMA} is the number of measurements considered for a central simple moving average (SMA).

The following rules are applied for the basic points:

$$c_0 = i, \text{if} \;\; F_k < (1 - f_1 c_{SMA})F_i \quad \wedge \quad F_k - F_i < \; f_2(F_i - F_h)$$

Eq. 2

$$c_1 = i, \text{if} \;\; (1 - f_1 c_{SMA})F_k < F_i \quad \wedge \quad f_2(F_k - F_i) > \; F_i - F_h$$

Eq. 3

$$c_2 = i, \text{if} \ \ (1 - f_1 c_{SMA})F_k > F_i \quad \wedge \quad F_k - F_i > \ f_2(F_i - F_h)$$

Eq. 4

$$c_3 = i, \text{if} \ \ F_k > (1 - f_1 c_{SMA})F_i \quad \wedge \quad f_2(F_k - F_i) < \ F_i - F_h$$

Eq. 5

The multiplier f_2 has to be greater than 1 in order to describe a transit shown in Fig. 1. $f_2 = 2$ is a realistic multiplier and proofed to be practical in testing with transits from confirmed exoplanets. The multiplier f_1 is the minimal transit depth. In order to adapt to the variance of different stars, it is set to 0.5 % percentile of the normalized flux F. In case this setting does not find any transit candidates, the minimal transit depth is soften to 1 % percentile; then 2 %,4 %, 8 %, 16 % or even 32 % percentile. c_{SMA} is a multiplier taken into account the smoothing depending on the measurements N_{SMA} considered for the SMA:

$$c_{SMA} = \frac{1}{N_{SMA}}$$

Eq. 6

The following N_{SMA} are applied:

N_{SMA}	c_{SMA}	$\Delta t_{SMA,Kepler} = [\text{days}]$
3	0.333	0.06
4	0.25	0.08
5	0.2	0.1

Tab. 1: Number of measurements considered for the SMA and resulting parameters

After all 4 basic points of a transit candidate are defined further checks are applied. Firstly the limb darkening durations at the beginning and the end of a transit shall be of the same order:

$$\frac{1}{2}\big(t(c_3) - t(c_2)\big) < t(c_1) - t(c_0) < 2\big(t(c_3) - t(c_2)\big)$$

Eq. 7

The requirement for the transit depth is

$$\Delta F > f_1 c_{SMA}$$

Eq. 8

In order to avoid triggering transits caused by anomaly or by long term darkening due to varying star activity the transit duration is limited to

$$3\Delta t_{SMA} < d < 25[\text{hours}]$$

Eq. 9

Transit Pairing

After a transit candidate passed all requirements, it will be paired with other transit candidates forming exoplanet candidates. A transit candidate m will only be paired with posterior transit candidate n, if the following requirements are fulfilled:

$$\frac{8}{10}d_n < d_m < \frac{10}{8}d_n$$

Eq. 10

$$\frac{6}{10}\Delta F_n < \Delta F_m < \frac{10}{6}\Delta F_n$$

Eq. 11

The transit depth and transit duration of the exoplanet candidate are the averages of the two transits. The exoplanet candidate period is calculated by

$$T = \frac{t_{0,n} - t_{0,m}}{2}$$

Eq. 12

Area of application

The Preselction Algorithm described above was also executed with $N_{SMA} = 1{,}2$, but it was seen that it is too sensitive to anomalies and in consequence causing an infinite number of exoplanet candidates. The sensitivity to anomalies is also the reason for the requirement in Eq. 9. Therefore, the Preselection Algorithm is only sensitive to transit durations greater than 0.16 days.

Automated TLS Scan

The Preselection Algorithm prepares in between 1000 and 150000 exoplanet candidates with transit duration, period, epoch and transit depth. It would need a lot of time to run Tukey's biweight filter and afterwards the TLS for the transit duration of each candidate. This is not necessary. When the running window of Tukey's biweight filter is close to but not equal 3 times the actual transit duration of an exoplanet, the signal-to-noise ratio is already improved a lot. Therefore it is sufficient, if the exoplanets are sorted by transit duration and then split into 4 groups.

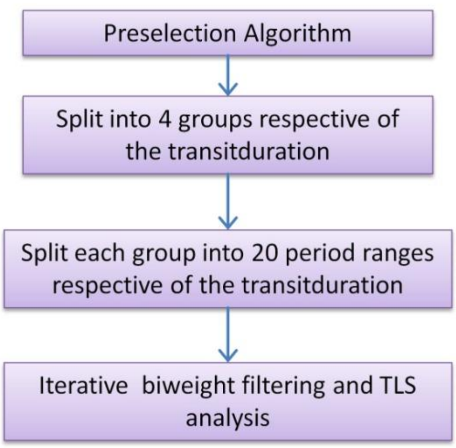

Fig. 2: Flowchart Automated TLS

Since the TLS is sometimes sensitive to data gaps, exoplanets might not be found. This risk can be reduced to run the TLS primarily with smaller period ranges instead of analyzing the whole period spectrum. Therefore the exoplanet candidates are sorted by period in each transit duration group and then split another time into 20 period ranges. In consequence Tukey's biweight filter and TLS have to be executed 80 times to scan a light curve for every group. The running window of the filter will be 3 times the mean of the transit durations in this group.

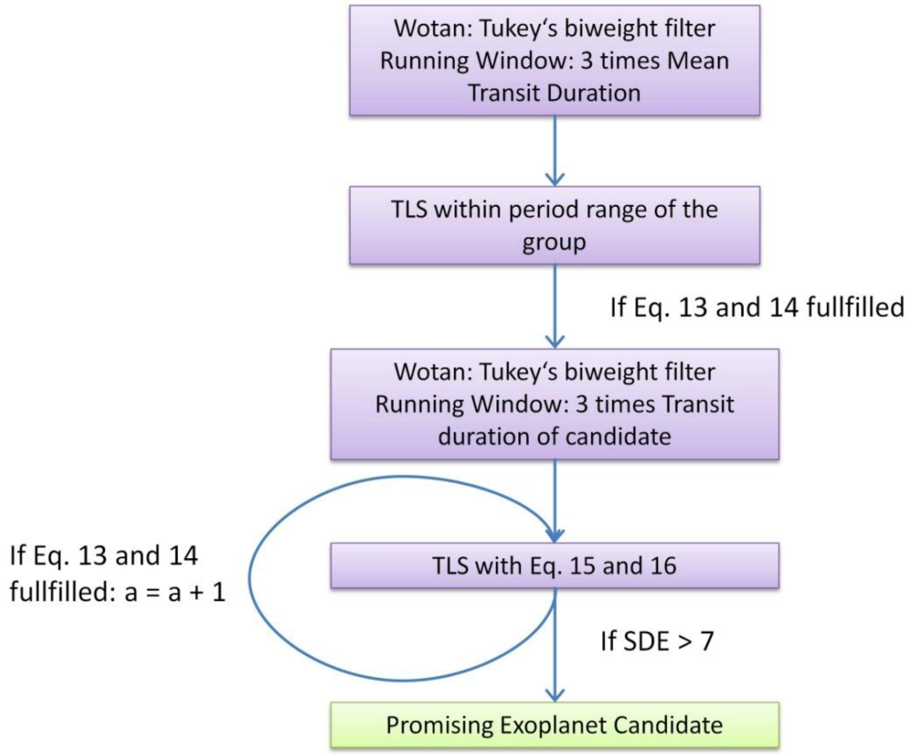

Fig. 3: Flow Chart Iterative Biweight Filtering and TLS analysis

The result of each TLS run is compared with the candidates of the Preselection Algorithm. A result is declared valid, if the following tolerance requirements are fulfilled:

$$\left|t_{0,TLS} - t_{0,Pre}\right| < 0.005\ T_{Pre}$$

Eq. 13

$$|T_{TLS} - T_{Pre}| < 0.001\ T_{Pre}$$

Eq. 14

If a valid result was found the results are refined by applying Tukey's biweight filter again with 3 times the transit duration of the result. Additionally, the period ranges of the TLS analysis are iteratively extended by

$$T_{min} = \left(1 - \frac{a^2}{100}\right) T_{Pre}\,; a = 3,4, \ldots 10$$

Eq. 15

$$T_{max} = \left(1 + \frac{a^2}{100}\right) T_{Pre}\,;\, a = 3,4, \ldots 10$$

Eq. 16

The iteration is aborted, if one of the requirements Eq. 13 and Eq. 14 is not fulfilled anymore. The last iteration of the TLS must show a Signal Detection Efficiency (SDE) greater than 7 to be a promising exoplanet candidate instead of a false positive.

Evaluation of the Automated TLS Scan

Firstly, the general functionality of the Automated TLS Scan shall be proven by testing with confirmed exoplanet lightcurve data. Afterwards the advantages of the period grouping and the primarily executed biweight filtering should be explained by affected examples.

Validation with confirmed Kepler planets

The Automated TLS Scan shall especially be used to search for Earthlike exoplanets. Therefore, the validation will only consider confirmed Kepler planets with a period greater than 50 days. Due to the sensitivity of the Preselection Algorithm explained before only confirmed Kepler planets with a transit duration greater than 0.16 days are tested. Other exoplanets and confirmed planets with only 2 transits will be removed from the lightcurve data. After data retrieval from NASA Exoplanet Archive [3] the Automated TLS Scan scans the lightcurve data and when a exoplanet candidate with an SDE greater than 7 has been found, the results will be compared. If the following tolerances are fulfilled, the confirmed exoplanet was successfully reproduced:

$$\left|t_{0,TLS} - t_{0,NASA}\right| < 0.001\ T_{TLS}$$

Eq. 17

$$\left|T_{TLS} - T_{NASA}\right| < 0.001\ T_{TLS}$$

Eq. 18

The results in Annex 4 show that 91.37 % of confirmed exoplanet results are reproduced by the Automated TLS. The Preselection Algorithm candidates included even 93.88 % of the confirmed exoplanets. There are 4 characteristics identified, which are potentially causing that some exoplanets are not found by the Automated TLS:

1. Star flux variation of 0.01 % and higher
2. Noise higher than 0.03 % of star flux
3. Transit durations longer than 11 hours
4. Data gaps in the order of the order of the exoplanet period

The first characteristic is regarding the TLS itself, although this was already improved by applying the period range grouping. The second and the third characteristic are features of the Preselection Algorithm. It could be tested, if this is improved by using $N_{SMA} = 6$. Since these characteristics apply summed up only to 2 % of the considered confirmed exoplanets, this is declined in favor of computing time. The 4 characteristic cannot be improved, since the Preselection Algorithm requires 2 neighbor transits and the probability to have them within the data is highly reduced in these cases. Anyway, this was only the case for 0.7 % of the considered exoplanets.

Evaluation of period grouping

The TLS algorithm is run on Kepler 264 lightcurve data. Since it is the target to reproduce the results of Kepler 264 c ($T_{NASA} = 140.101$ [days]; $t_{0,NASA} = 216.217$ [days]), the transit of Kepler 264 b ($T_{NASA} = 40.807$ [days]; $t_{0,NASA} = 155.716$ [days]) are filtered. The TLS algorithm is run for a period range between 120 days and 400 days. The best result is shown for $T_{TLS} = 166.121$ [days]. (Fig. 4)

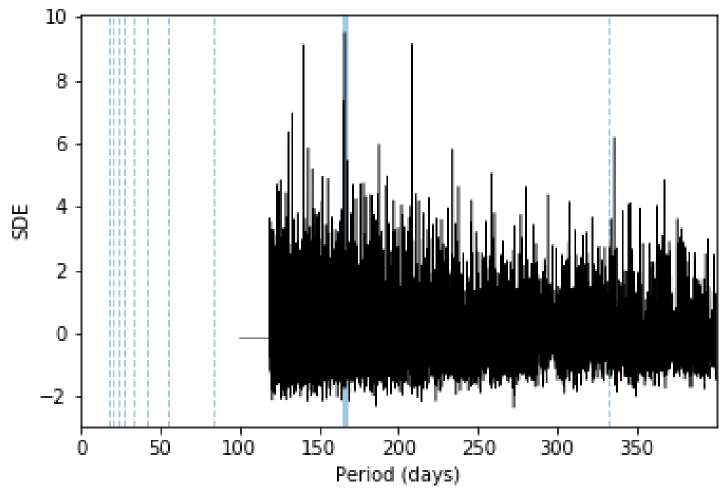

Fig. 4: SDE results for TLS on Kepler 264 with period range between 120 and 400 days (Kepler 264 b filtered)

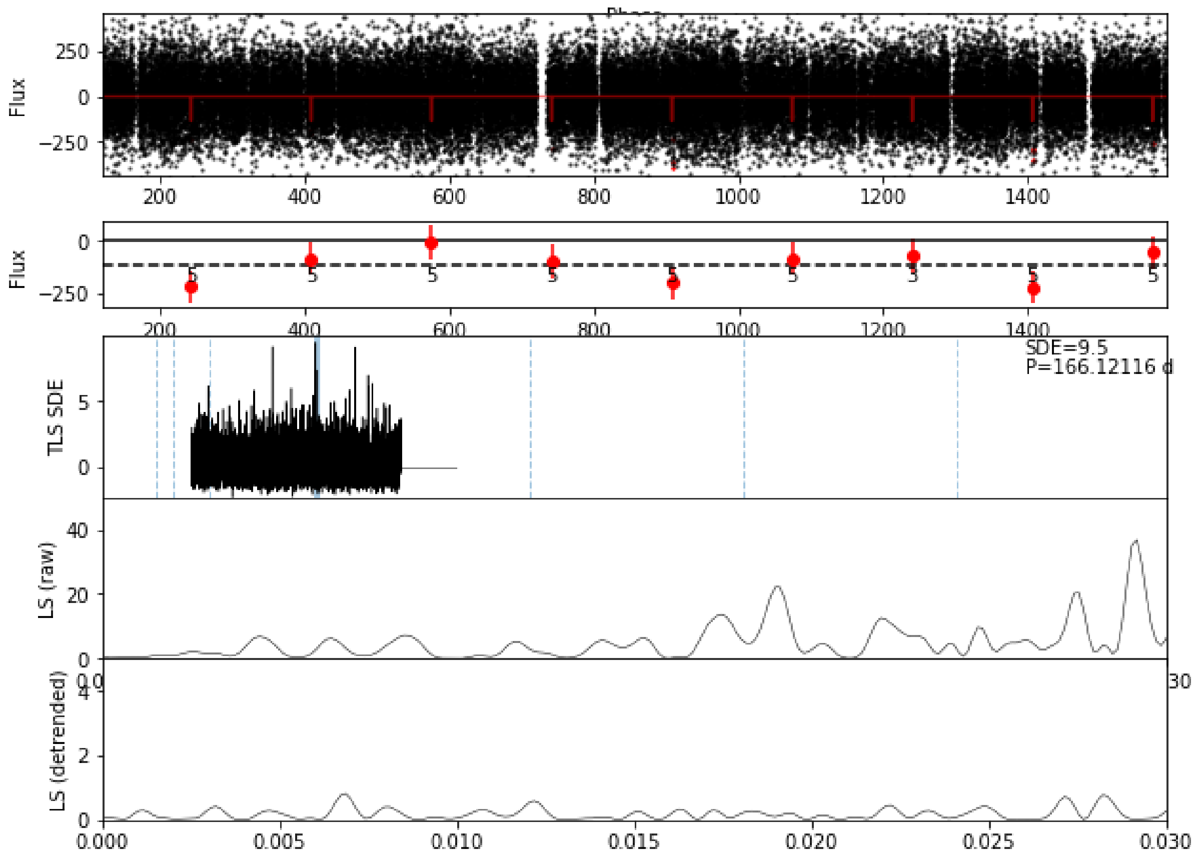

Fig. 5: Transit depth and number of measurements per transit for result from Fig. 4

It is very unlikely that the result is a new exoplanet. When looking at variety of transit depths shown in Fig. 5, the transit depth is has a high variation, while the star flux is almost not changing at all. Nevertheless, the TLS analysis chose this as best result instead of the confirmed planet Kepler 264 c.

When using the automated TLS including the Preselection Algorithm, the result period $T_{TLS} = 166.121$ [days] is grouped in another period range than Kepler 264 c. Therefore, Kepler 264 c is now found. Hence, the period grouping increases the probability that an exoplanet is not hidden by noise or variation in stellar flux.

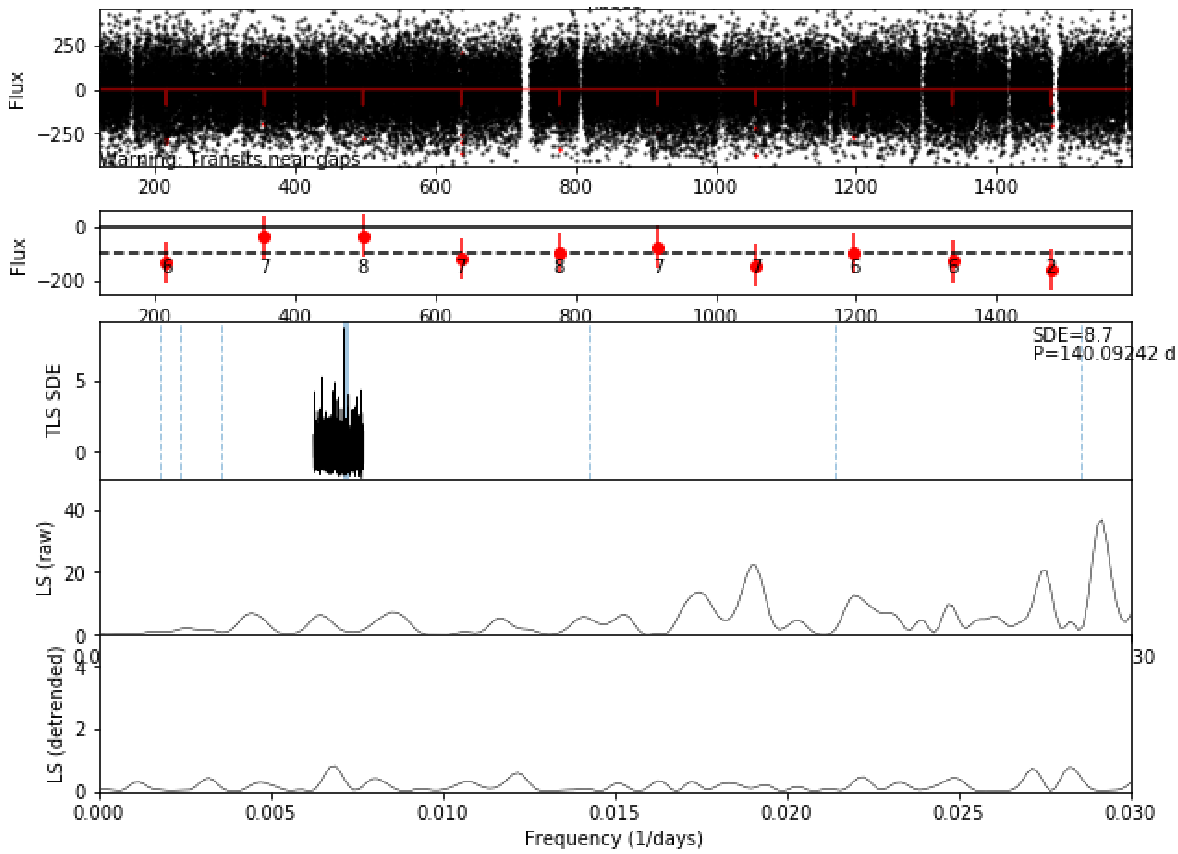

Fig. 6: Transit depth and number of measurements per transit for Kepler 264 c

A posterior comparison showed that Kepler 264 c transits from TLS are also within Preselection Algorithm candidates (see Annex 1 for Preselection Candidates). In conclusion using the Preselection Algorithm for confirmation of the TLS result makes the Automated TLS more robust.

Evaluation of primarily executed biweight filtering

The most important advantage of the Preselection Algorithm is the primarily executed biweight filtering in combination with ability to validate TLS results by Eq. 13 and Eq. 14. By applying the Automated TLS to KOI-5756 lightcurve data a promising exoplanet candidate is found (Annex 2: Preselection Candidates with SDE; Annex 3: Vetting Sheet).

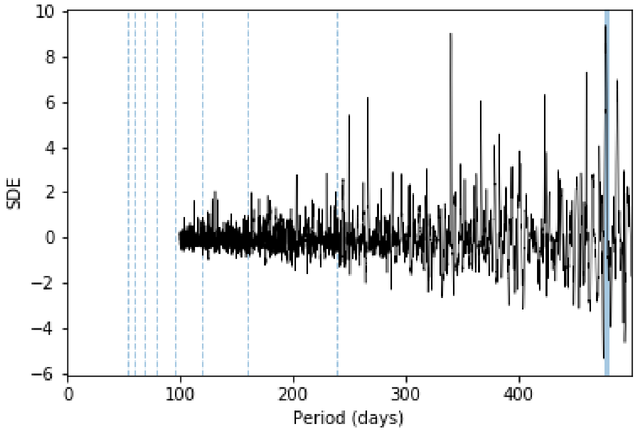

Fig. 7: SDE results for TLS on KOI-5756 lightcurve without biweight filtering

Fig. 7 shows several peaks in SDE. In consequence, there is no evidence for an exoplanet. The Preselection Algorithm delivers several candidates (Annex 2) and one of them is successful: When the running time window of Tukey's biweight filter is set to three times the transit duration $d_{Pre} = 0.311$ [days] (Running Time Window: 0.933 days) SDE greater than 7 is found for $T_{Pre} = 384.438$ [days]; $t_{0,Pre} = 585.896$ [days]. The signal could be even improved, when filtering with a running time window of 1.5 [days]. (Fig. 8)

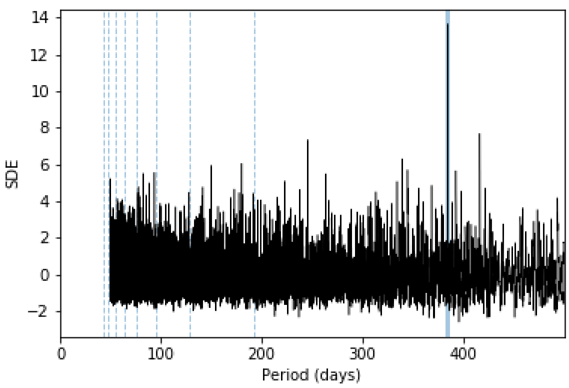

Fig. 8: SDE results for TLS on KOI-5756 lightcurve with biweight filtering; Running time window of 1.5 [days]

It is important to choose a good running time window for Tukey's biweight filter. When a running time window of 0.9 days is chosen instead of the 0.933 days defined by the Preselection Algorithm, the exoplanet candidate can already not be found anymore in the SDE diagram.

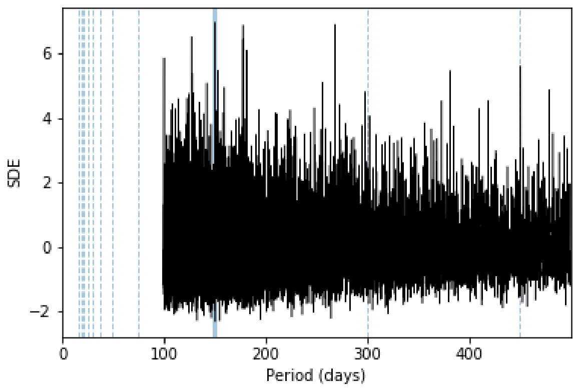

Fig. 9: SDE results for TLS on KOI-5756 lightcurve with biweight filtering; Running time window of 0.9 days

Conclusion

The Automated TLS including a Preselection Algorithm is fully automated including the possibility of primarily to TLS biweight filtering lightcurves. The results from the Preselection Algorithm are used to automatically validate TLS results and identify wrong results due to data gaps.

The TLS still struggles with lightcurves from stars with a high flux variation. This could not be solved by the automation.

Nevertheless, the check with confirmed planets from NASA Exoplanet Archive proofed that the automated TLS is working and is ideal to scan Kepler Data for further Earthlike exoplanet candidates.

Bibliography

[1] M. H. J. F. K. R. N. M. B. a. S. B. René Heller1, "Transit least-squares survey III," *Astronomy & Astrophysics,* p. 15, 04 06 2020.

[2] J. Johnson, "How Do You Find an Exoplanet?," Princeton, Princeton University Press, 2015, p. 65.

[3] NASA, "NASA Exoplanet Archive," 03 11 2021. [Online]. Available: https://exoplanetarchive.ipac.caltech.edu/docs/program_interfaces.html#kep_names_stellar_timeseries. [Accessed 29 11 2021].