Global Analysis of the TRAPPIST Ultra-Cool Dwarf Transit Survey

F. Lienhard,1⋆ D. Queloz,1 M. Gillon,2 A. Burdanov,3,4 L. Delrez,1,2,5 E. Ducrot,2 W. Handley,1 E. Jehin,5 C. A. Murray,1 A. H. M. J. Triaud,6 E. Gillen,1 A. Mortier,1 B. V. Rackham3,7,8

1 Cavendish Laboratory, JJ Thomson Avenue, Cambridge CB3 0HE, UK
2 Astrobiology Research Unit, Université de Liège, Allée du 6 Août 19, Sart Tilman, 4000, Liège 1, Belgium.
3 Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
4 Instituto de Astrofísica de Canarias, Vía Láctea s/n, 38205 La Laguna, Tenerife, Spain
5 Space Sciences, Technologies and Astrophysics Research (STAR) Institute, Université de Liège, 4000 Liège, Belgium.
6 School of Physics & Astronomy, University of Birmingham, Edgbaston, Birmingham B152TT, UK
7 Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
8 51 Pegasi b Fellow

Accepted 2020 July 9. Received 2020 July 9; in original form 2020 March 23

ABSTRACT
We conducted a global analysis of the TRAPPIST Ultra-Cool Dwarf Transit Survey – a prototype of the SPECULOOS transit search conducted with the TRAPPIST-South robotic telescope in Chile from 2011 to 2017 – to estimate the occurrence rate of close-in planets such as TRAPPIST-1b orbiting ultra-cool dwarfs. For this purpose, the photometric data of 40 nearby ultra-cool dwarfs were reanalysed in a self-consistent and fully automated manner starting from the raw images. The pipeline developed specifically for this task generates differential light curves, removes non-planetary photometric features and stellar variability, and searches for transits. It identifies the transits of TRAPPIST-1b and TRAPPIST-1c without any human intervention. To test the pipeline and the potential output of similar surveys, we injected planetary transits into the light curves on a star-by-star basis and tested whether the pipeline is able to detect them. The achieved photometric precision enables us to identify Earth-sized planets orbiting ultra-cool dwarfs as validated by the injection tests. Our planet-injection simulation further suggests a lower limit of 10 per cent on the occurrence rate of planets similar to TRAPPIST-1b with a radius between 1 and 1.3 $R_\oplus$ and the orbital period between 1.4 and 1.8 days.

Key words: atmospheric effects – techniques: photometric – planets and satellites: detection

1 INTRODUCTION
Ultra-cool dwarfs (UCDs: late M dwarfs and brown dwarfs with effective temperature below 2,700 K, Martín et al. 1999; Kirkpatrick 2005) make up a significant fraction of all stellar objects in the Galaxy. Counting dwarf stars of spectral type M7 and later including L, T, and Y dwarfs, the census of stars and brown dwarfs by Kirkpatrick et al. 2012 suggests that UCDs account for about 18 per cent of the stellar and substellar objects within 8 pc of the Sun. UCDs are excellent targets in the search for temperate transiting planets because of their low mass and size. The habitable zones of UCD stars are 30–100 times closer to their host star than that of the Sun due to their low temperature. Therefore, temperate planets exhibit a comparably short orbital period from one to a few days, which increases the likelihood of observing transits (Gillon et al. 2013). Since the radius of a mature UCD is about ten times smaller than the Sun (Chabrier & Baraffe 1997; Dieterich et al. 2014), the transit depth of an Earth-sized planet is of the order of 1 per cent, which is within the detection range of small ground-based telescopes. Additionally, planets orbiting UCDs are optimal targets for the characterisation of their atmospheres’ chemical composition via transmission spectroscopy.

⋆ E-mail: fl386@cam.ac.uk

© 2020 The Authors
(Kaltenegger & Traub 2009; de Wit & Seager 2013). However, despite their high frequency, the statistics of the planetary population of late M dwarfs and brown dwarfs are poorly understood (Delrez et al. 2018b). This is mainly due to the low intrinsic brightness of UCDs, which reduces the sample of stars within reach of radial velocity and transit searches. Additionally, UCDs can display a high degree of stellar activity complicating the planet detection via the radial velocity and the transit technique. Knowledge of planet occurrence rates comes mainly from the Kepler mission (e.g. Dressing & Charbonneau 2015). However, UCDs were not the primary targets of the Kepler mission. Furthermore, some ground-based surveys, such as MEarth (Berta et al. 2012) focusing on M dwarfs, were designed for planets with radii above $2R_{\oplus}$ at short orbital periods, which turned out to be rare (Berta et al. 2013).

Radial velocity and transit surveys indicate that tightly packed systems of low-mass planets are common around solar-type stars (Howard et al. 2010) and red dwarfs (Kirkpatrick et al. 2012; Dressing & Charbonneau 2013; Hardegree-Ullman et al. 2019). UCDs could host a comparable planetary population as simulations by Ormel et al. (2017) suggest. In the present analysis, we test whether this hypothesis is compatible with the results of the TRAPPIST Ultra-Cool Dwarf Transit Survey (TRAPPIST-UCDTS, Gillon et al. 2013), which discovered three temperate Earth-sized planets orbiting the late M dwarf TRAPPIST-1 located 12.1 pc from the Sun (Gillon et al. 2016). Further transit observations and analysis of transit timing variations have revealed that the system hosts 7 rocky, Earth-sized planets in almost perfectly edge-on and circular orbits (Gillon et al. 2017; Luger et al. 2017). No additional planets have been found around TRAPPIST-1 (Ducrot et al. 2020) or any of the other stars in this data set.

The data analysed in this paper was acquired with TRAPPIST-South (TRAn siting Planets and Planetesimals Small Telescope-South) (Gillon et al. 2011; Jehin et al. 2011), which is a robotic f/8 Ritchey-Chrétien 60 cm telescope located at the La Silla Observatory in Chile. TRAPPIST-UCDTS has been monitoring late M dwarfs and brown dwarfs since 2011 as a prototype survey for the more ambitious SPE CULOOS survey (Burdanov et al. 2018; Delrez et al. 2018b; Gillon 2018). Here, we reanalyse the data of 40 targets observed between 2011 and 2017 by TRAPPIST-South starting from the raw images. The pipeline created for this analysis was developed alongside the SPECULOOS Southern Observatory (SSO) pipeline (Murray et al. 2020).

The paper is structured as follows. In Section 2, we describe the sample used for this analysis. In Section 3, we describe the light curve detrending, removal of data points affected by non-planetary signals and the modelling of stellar variability and activity. In Section 4, the light curve of TRAPPIST-1, and thus the performance of the pipeline for the only planetary system found in the data set, is discussed. In Section 5, we estimate the energy of the flare events that we find in the light curves. The transit injection tests characterising the statistics of the survey are outlined in Section 6, while the results of the injection tests are described in Sections 7 and 8. Finally, we summarise our results, compare them to other surveys, and discuss the significance for multi-planetary systems around UCDs in Section 9.

2 TRAPPIST-UCDTS: INSTRUMENTATION AND METHODOLOGY

TRAPPIST-UCDTS was conducted mostly in the I+z′ filter due to the redness of the targets (cf. figure 6 in Delrez et al. 2018b). The response function of the TRAPPIST instrument is depicted in Fig. 2 (M. Gillon – priv. comm.). The field of view is equal to $22'' \times 22''$ with a resolution of 0.65 arcsec/pixel. More information can be found in Gillon et al. (2011). As TRAPPIST-UCDTS is a targeted survey, the exposure time is optimised for each target to maximise the signal-to-noise ratio and duty cycle and is typically between 30 and 60 seconds. Most parts of the light curves allow the detection of transits caused by Earth-sized planets. The UCDs have typically been observed for 50 hours, as visible in Fig. 1. The here quoted value is the total exposure time, not the total observation time dedicated to this target. A description of the survey and the first 20 light curves (5 M6, 6 M7, 4 M8, and 5 M9) is provided in Gillon et al. (2013). The first 40 light curves observed between 2011 and 2017 are characterised further in Delrez et al. (2018b) and Burdanov et al. (2018). The stars, and associated parameters based on Section 6.1, are listed in Table 1. Some of the target stars have been further observed within the TRAPPIST-UCDTS or SPECULOOS programs without identifying additional transiting planets.

3 METHODS

During the survey, the data were inspected for transits by eye on a day-by-day basis. We perform an automated global analysis for this study, which ensures consistency in how the data are treated and which enables us to estimate the probability of finding planets around UCDs using a clearly defined detection criterion. The individual steps in the automated global analysis are detailed in the following subsections.

Figure 1. Histogram of the added exposure time of all targets. The target with 175 hours total exposure time is TRAPPIST-1, which was observed more extensively due to the detection of a planetary system around it.
is then computed with `imcorelist`, while the mean exposure time and Span the number of days between the first and the last observation. In the column below RMS, we list the median nightly RMS of the differential light curves before division by the GP mean (i.e. as in Fig. 7 panel 3).

### Table 1. The coordinates are in J2015.5. The apparent $K_S$ magnitudes were retrieved from 2MASS (Cutri et al. 2003) and the $G-R_P$ colour index from Gaia DR2 (Gaia Collaboration et al. 2016, 2018). Radii and masses are estimated as described in Section 6.1. The dagger symbol indicates that the absolute $K_S$ magnitude was above 10, and hence we set the mass estimate to the approximate mass of an M9 star. The number of observation nights is shown below Nights, while DP indicates the number of measurements before bad-weather removal. Exp is the mean exposure time and Span the number of days of nights and between 7.07 pixels (4.6 arcsec) for all stars and images. Damasso et al. (2010) suggest an aperture radius between 2 and 3 times the mean FWHM of the field stars, which indicates that an aperture radius between 6 and 9 pixels is optimal for

### 3.1 Image reduction

To calibrate the science images, we employ the standard offset, bias, and dark subtraction followed by flat-field division. Thereafter, we perform aperture photometry using the CASUTools software (Irwin et al. 2004). For each target, 20 suitable images are stacked with CASUTools `imstack` to generate a high-quality image of the field and extract the coordinates (RA, DEC) of the stars using CASUTools `imcore`, CASUTools `wcsfit` and `astrometry.net` (Lang et al. 2010).

The flux of the stars in all images of the respective field is then computed with `imcorelist` by adding the pixel values within apertures at the coordinates RA, and DEC, and subtracting the background flux. To estimate the latter, `imcorelist` divides the image into a grid of scale size 64 pixels and computes the iteratively k-sigma clipped median on the grid sections. It then lightly filters these values and computes the background via bilinear interpolation. We account for proper motion by recomputing the coordinates of the stars in a target field for each night based on Gaia DR2 (Gaia Collaboration et al. 2016, 2018) proper motion data. This is necessary since the target stars are nearby and the time span between the first and the last measurement can be extensive, as visible in Table 1 or Fig. 8. The aperture radius is set to 7.07 pixels (4.6 arcsec) for all stars and images. Damasso et al. (2010) suggest an aperture radius between 2 and 3 times the mean FWHM of the field stars, which indicates that an aperture radius between 6 and 9 pixels is optimal for

---

**MNRS 000, 1–21 (2020)**
most of our science images. To further test this, we computed the RMS of the differential target light curves for different aperture radii. We did not remove bad-weather data, but we applied an iterative global 4-sigma clipping to remove outliers. Based on this analysis, we concluded that 7.07 pixels is the most suitable aperture radius yielding the lowest mean RMS. We chose a fixed aperture over a dynamic one to avoid adding structure to the light curves, which could result from a suboptimal background estimate that a dynamic aperture radii. We did not remove bad-weather data, but we applied an iterative global 4-sigma clipping to remove outliers.

3.2 Differential photometry

To correct for photometric effects of non-astrophysical origin, we divide the raw target light curve by the weighted average (hereafter called the Artificial Light Curve, ALC) of the median-normalised light curves of selected comparison stars in the same frame as the target star.

First, we define a set of potential reference stars in the same field as the target star. This set consists of all the stars with more than 100 detected electrons per second in their aperture. We also remove bright stars which lead to saturated pixels. We then arrange the median-normalised light curves in a matrix $M$ of dimension $(n,m)$ with $n$ the number of potential reference stars and $m$ the number of science images of the respective field. The matrix form enables us to compare the flux measurements in the same science image (the first axis) or in time (the second axis).

(i) The initial step consists of 3-sigma clipping along the first axis of $M$. The fluxes of the normalised light curves in the same science image are similarly affected by the airmass, and they are median-normalised along the second axis in $M$. Therefore, we can compare them to each other to remove outliers and to find stars that behave very differently compared to the median star in the field. The standard deviation of the normalised flux values in a science image is dominated by the fainter stars. Consequently, this step represents a coarse clipping on the brighter stars. These bright stars get a high weight in step (iii), which motivates us to sigma clip here since stars can display a high degree of variability which cannot be removed by sigma clipping along the second axis in matrix $M$.

(ii) A reference star’s light curve is discarded if more than 20 per cent of its values are removed in step (i).

(iii) We set the statistical weight $w_j$ of each remaining potential reference star $j$ to its median brightness before normalisation:

$$w_{\text{initial}} = \text{Md}_i(N_j),$$

where $N_j$ is the photon count for star $j$ in science image $i$ in a fixed aperture and $\text{Md}_i(N_j)$ is the median over all photon counts of star $j$.

By setting the statistical weights $w_j$ to the initial weights $w_{\text{initial}}$ as above, we compute the ALC as the weighted mean of the normalised light curves:

$$\text{ALC}_i = \sum_{j=1}^{n_{\text{ref}}} \frac{N_j w_j}{\text{Md}_i(N_j)} \left( \sum_{j=1}^{n_{\text{ref}}} w_j \right)^{-1},$$

where ALC$_i$ is the ALC value for the $i$th science image. We tested a more sophisticated noise model including read-out, dark, and scintillation noise as initial weights. This approach led to the same final weights.

(iv) By dividing each light curve in $M$ by the ALC, we get a first estimate of the reference stars’ differential light curves. We set the weight of each reference star to the inverse square of the RMS of its differential light curve:

$$w_j = \text{RMS}_i\left(\frac{N_j}{\text{Md}_i(N_j)} \frac{1}{\text{ALC}_i}\right)^{-2}$$

and recompute the ALC as in Eq. 2.

(v) We repeat step (iv) until the weights converge to within 0.00001, which is typically achieved after 20 iterations. Reusing the variability estimate to refine the weights in the ALC calculation is conceptually similar to the detrending approach in Broeg et al. (2005).

(vi) Some stars can have a higher RMS than similarly bright stars either due to continuous brightness variations caused by, for example, pulsation or binarity or because some parts of the light curve are noisy as a result of bad pixels, an imperfect flat-field correction, or flares. Other stars are perfectly quiet throughout the light curve. In the case of non-stationary noise which is not present in the other light curves, we overweight the noisy parts of the respective light curve and underweight the parts with low temporal RMS by assigning fixed weights in the ALC calculation (cf. Eq. 2). Thus, these variable stars can induce some spurious features in the ALC, but they can easily be identified and removed. For this purpose, we modelled the RMS of the normalised light curve of a calm star with median photon count $N'$ by fitting the following function with three fitting parameters ($a$, $b$, $c$) to the RMS of the stars with the lowest RMS in a given flux bin:

![Figure 2. Response function of the TRAPPIST instrument. The total response is computed by multiplying the different components (the mirror coating reflectance has to be taken into account twice). The limiting contributor on the lower end is the filter transmission while the upper end is determined by the mean CCD quantum efficiency.](image-url)
(vii) Using the estimate of the ALC from step (v), we
(i) The previous steps optimise the ALC with respect to
a typical star in the given target field, neglecting the

fact that noise can be locally correlated and that light
curves are colour-dependent. To improve the correction
of the target star, we increase the statistical weight of
reference stars near the target star as a function of an-
gular distance to the target star. More specifically, we
multiply the statistical weight of the reference stars by a
factor \( w_d \) which depends on the angular distance of
the respective star to the target star. The distance weight-
ing is computed for each target star field individually,
and it remains the same over the full light curve. Since
the distance weighting function should be flat near the
target star and decay further out, we use the following
functional form for \( w_d \):

\[
w_d = \frac{1}{1 + \left( \frac{d_{\text{target}}}{d_{\text{max}}} \right)^a}^5
\]

where \( a \) is a fitting parameter, \( d_{\text{target}} \) is the angular distance
between target and reference star, and \( d_{\text{max}} \) is the
angular distance to the star furthest from the target star.
Simultaneously with parameter \( a \) in equation (5), we set
a lower threshold for the colour since the target stars are
very red compared to the average field star. This cut-
off reduces the impact of colour-dependent atmospheric
effects on the differential light curves. For this purpose,
we calculate the RMS of the final light curves for a grid
of potential values of parameter \( a \) and colour thresholds
based on the Gaia DR2 \( G - G_{\text{RP}} \) colour index. We pro-
ceed with the combination of the two parameters which
minimises the RMS of the respective light curve. The
typical cutoff value for \( G - G_{\text{RP}} \) is around 0.5, while
the parameter \( a \) in the distance weighting scheme is more
variable but typically around 2.4.

3.3 Removal of bad-weather data

Clouds or hazes can lead to spatial inhomogeneity in the
transparency of the atmosphere, which hampers the effec-
tiveness of differential photometry. Additionally, these and
similar effects decrease the number of photons that reach the
detector, thus increasing photon noise. We tackle this prob-
lem by removing all flux measurements where the running
mean of the ALC is below 0.5 or where the running RMS
exceeds 10 per cent. For the running RMS, we compute the
RMS of typically 15 flux values closest to the point in time
where we intend to estimate the scatter in the light curve.
The same procedure is applied for the running mean.
Fig. 4 exemplifies that the running RMS of the differential
light curve is increased if the running RMS of the
ALC is high. It also shows that the minimal RMS of the
target light curve increases as a function of the ALC RMS,
which is consistent with the findings of Berta et al. (2012).

3.4 Removal of flares

Flares are characterised by an abrupt rise followed by a
slower decrease in brightness. They can pose significant
problems for all of the following light curve optimisation
steps, and they tend to confound the transit search. There-
fore, it is beneficial to find and remove data points affected
by flares. Various approaches were chosen in other surveys to

\[
\text{RMS}(N^j) \propto \frac{\sqrt{\sigma_{\text{photon}}^2 + \sigma_{\text{const}}^2 + \sigma_{\text{scintillation}}^2}}{N^j}
\]

where \( \sigma_{\text{photon}} \) is the photon noise, \( \sigma_{\text{const}} \) is the constant noise,
and \( \sigma_{\text{scintillation}} \) is the scintillation noise.

\[
\text{RMS}(N^j) \propto a \sqrt{\frac{1}{N^j} + \frac{b}{(N^j)^2} + c}.
\]

Poison noise is proportional to \( \sqrt{N^j} \), while scintillation
noise is proportional to \( N^j \) (Young 1967). In addition,
there is a constant error term to account for read-out
and dark current noise.

Stars with an RMS 20 per cent above the minimal RMS
of a calm star, calculated in Eq. (4), are identified as vari-
able stars. We exclude these stars from the reference star
set, as shown in Fig. 3. The theoretical noise estimate,
computed as in Damasso et al. (2010) eq. 3, is included
in Fig. 3 to compare to our fit of the RMS of the entire
light curves including unfavourable nights.

\[
\text{RMS}(N^j) \propto \frac{\sqrt{\sigma_{\text{photon}}^2 + \sigma_{\text{const}}^2 + \sigma_{\text{scintillation}}^2}}{N^j}
\]

\[
\text{RMS}(N^j) \propto a \sqrt{\frac{1}{N^j} + \frac{b}{(N^j)^2} + c}. \tag{4}
\]
address this problem. Berta et al. (2012), for instance, performed a grid search on their data modelling a flare as a fast rise in flux followed by an exponential decay. They removed a night’s light curve if they found a flare with an amplitude above 4 sigma. Since stars were often observed during one full night in TRAPPIST-UCDTS, we cannot discard a whole night if there was a flare because it would mean losing a significant amount of data. The MEarth telescopes observe a star roughly every 20 minutes, while TRAPPIST-UCDTS focuses on one target at a time with a cadence of the order of a minute. We therefore had to find a criterion to find flares and remove these from the light curves.

Using *Kepler* data, Davenport et al. (2014) generated a median flare template from 885 flares observed on the M4 star GJ 1243 to analyse the different phases of a flare. Their template suggests a flux rise, characterised by a fourth order polynomial, followed by an initial impulsive exponential decay which smoothly transitions into a more gradual exponential decay phase. As other authors (e.g. Walkowicz et al. 2011; Loyd & France 2014), we approximated a flare’s shape as a single exponential decay. This simple model is sufficient for our purposes since we iteratively fit and remove flares and because a gradual flux decay within the typical flux range is not expected to confound the transit search code.

Flare candidates in the differential target light curves are identified by evaluating criterion (6) with condition (7) for all flux values:

\[
\frac{2f_j - f_{j-2} - f_{j+3}}{\sigma_j} > 12
\]  

(6)

\[
2f_j - f_{j-2} - f_{j+3} > 0,
\]  

(7)

where \( f_j \) is the \( j \)th flux measurement of the respective star and \( \sigma_j \) is the RMS of the 60 data points closest to \( j \).

Criterion (6) can be understood as follows. For any data point, we compare the potential flux peak \( f_j \) to adjacent flux measurements to find flares by exploiting the fact that there is a peak in the light curve and that the flux rises more rapidly than it decays. If there is a flare in the light curve and it peaks at \( f_j \), then \( f_j - f_{j-2} \) captures the flux rise and is positive. In this case, \( f_j - f_{j+3} \) is a measure for how quickly the flux decays after the flare peak and is positive as well. The first component of criterion (6) favours strong signals consisting of a very sharp rise in flux followed by a fast decay since in this case both \( f_j - f_{j-2} \) and \( f_j - f_{j+3} \) are positive and add up.

In the second part of criterion (6) we exploit the asymmetry of a flare’s shape since \( f_j - f_{j-2} \) is expected to be much greater than \( f_j - f_{j+3} \). In this case, \( f_j - f_{j+3} \) was chosen over \( f_j - f_{j+1} \) since \( f_j - f_{j+1} \) captures the asymmetry better. To avoid false signals from noisy data, we divide both parts of criterion (6) by the sigma clipped standard deviation of the flux differences\(^2\). The cut-off value was set to 12 by inspecting the value of criterion (6) for the smaller flares that we still intend to detect and remove. Assuming the shape of the flare to consist of an abrupt rise followed by an exponential decay, setting the cadence to one minute, the characteristic decay time to 5 minutes, and the standard deviation to 0.5 per cent, we find the value of criterion (6) exceeds 12 if the ratio of the flare amplitude to the standard deviation is about 3. Preliminary tests with SPECULOOS data suggest that a lower cut-off might be adequate for other data sets (C.A. Murray – priv. comm.)

Condition (7) ensures that signals which resemble inverted flares are not removed. We always use \( f_j - f_{j-2} \) instead of the more intuitive \( f_j - f_{j-1} \) because the typical exposure time (40 seconds) is relatively long, and thus a flare can begin just shortly before the shutter is closed. This results in a less steep rise in flux in the light curve that does not accurately represent the actual flux evolution of the star. Thus, we avoid the potential distortion of the true flare signal by comparing the potential peak to the penultimate data point before it. Also, it is possible that the flare indeed rises over

\(^2\) For each data point \( j \), we compute the difference between two adjacent flux measurements for the 60 data points closest to \( j \). Towards the beginning or end of a night’s observation period, the number of included data points gradually shrinks to 30. A flare in the data leads to a high standard deviation of the flux differences, which decreases the strength of the detection signal. Therefore, we clip the flux differences iteratively with \( \sigma_{upper} \) equal to 6 and \( \sigma_{lower} \) equal to 10. A lower sigma leads to a higher number of detections but also to more false positives, which are expected to be filtered out by requiring the flare amplitude to exceed a certain threshold.

Figure 4. Comparison between the running RMS of the differential target light curve (not GP-detrended but corrected for flares and cosmic rays) and the ALC. The RMS of the target light curve is low if the conditions are favourable. Adverse conditions such as clouds or hazes, identified by the high RMS in the Artificial Light Curve (ALC), can cause some scatter in the target light curve. We remove any flux value where the RMS of the ALC is above 10 per cent. CUBERHELIX colour scheme (Green 2011).
Global Analysis of the TRAPPIST Ultra-Cool Dwarf Transit Survey  

3.5 Removal of measurements affected by cosmic rays

The electron count within an aperture is significantly increased if a cosmic ray hits a pixel. To circumvent this problem, we discard all images for which the flux of the target star is four times the running RMS above the running mean of its light curve.

3.6 Light curve detrending from pixel position

The flat-field calibration corrects for pixel sensitivity variations. We refine this calibration and correct residual differential pixel sensitivity variations and potential gaps between pixels, which can affect the light curve if the target moves on the CCD. This effect can be modelled by fitting a 2D second order polynomial as a function of pixel position of the target to the differential target light curve before and separately after the meridian flip of the telescope due to the German equatorial mount. Dividing the target light curve by the fitted polynomial for each observation night before and after the meridian flip removes most of the flux dependence on the pixel position. The standard deviation of the star’s CCD position during an observation night, before or after the telescope flip, is typically about 1 pixel. A very similar approach was taken by Berta et al. (2012) who corrected for correlations between light curve trends and external parameters such as the pixel position in their instrumental systems model. They also allowed different baselines before and after the meridian flip, which is conceptually the same as our approach. Similar to Berta et al. (2012), we do not include a correction for a potential correlation between the FWHM of the target star and its light curve because we find the Pearson correlation coefficient to be usually between -0.1 and 0.1 indicating no photometrically relevant correlation.

3.7 Light curve detrending using Gaussian Process model

We employ Gaussian Process (GP; Rasmussen & Williams 2006) regression to model the intrinsic stellar brightness variations, as well as residual instrumental and atmospheric variations not fully accounted for in our initial detrending. The model is intended to capture the general trend over the typical variation timescale, but not short abrupt signals.

We model long-term variations due to effects such as precipitable water vapour (Bailer-Jones & Lamm 2003; Berta et al. 2012; Murray et al. 2020), varying target FWHM, and intrinsic flux variations due to starspots or the inhomogeneous cloud coverage of UCDs. Short-term fluctuations in brightness, which can also result from precipitable water vapour changes (Murray et al. 2020), should remain in the corrected light curves, however.

GP modelling requires an informed choice of kernel covariance function $k$, the mean function $\mu$, and an estimate of the uncertainty of each observation $\sigma_j$. The kernel is used to compute the covariance between every two measurements (Eq. 8 in the present analysis), which results in an $n \times n$ covariance matrix for $n$ data points. To account for noise in the data, the squared uncertainty of the data points is added on the diagonal. Combined with the mean function, the covariance matrices then provide a Gaussian distribution at each point of interest.

We use the GP package Celerite (Foreman-Mackey et al. 2017), which is computationally fast but restricts us in the choice of the kernel. In general, the computational cost of GPs scales cubically with the number of data points. Celerite, however, exhibits linear scaling, which is advantageous for the analysis of a large data set such as ours.

Several kernels were tested, of which the stochastically-driven damped simple harmonic oscillator (SHO) kernel with
the quality factor set to $1/\sqrt{2}$ and the mean function $\mu$ set to the median flux yielded the best results regarding flexibility and applicability to all light curves. This treats an intermediary path between smooth periodic and rough stochastic variations. The kernel of the SHO with the mentioned quality factor is equal to:

$$k(\tau) = S_0 \omega_0 \exp\left(-\frac{|\omega_0 \tau|}{2}\right) \cos\left(\frac{\omega_0 \tau}{\sqrt{2}} \frac{\pi}{4}\right).$$

where $\tau$ is the time difference between two measurements. The free hyperparameters $S_0$ and $\omega_0$ are computed by maximising the likelihood of the data given the model using the L-BFGS-B non-linear optimisation routine (Byrd et al. 1995; Zhu et al. 1997). L-BFGS-B is based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Fletcher 1987) and allows solving large non-linear optimization problems with bounds using the gradient projection method. We include the entire light curve of each target star for this procedure to capture the global behaviour of the light curve and lower the impact of transits on the GP hyperparameters. The mean of the Gaussian distribution predicted by the GP for each time point then serves as a model for the flux originating from the star.

The uncertainty $\sigma_y$ that we need to assign to each data point influences how closely the GP fits the data. We aim at modelling the stellar light curves, but there could be additional planetary transits affecting the measured brightness of their host star. Thus, we set the uncertainty of a flux measurement to the running RMS and add an additional component to the running RMS to reflect this source of uncertainty regarding the true brightness of the star alone as further explained in Section 6.4. In this analysis, we added 7 ppt to the running RMS ($\sigma_y^2 = \text{RMS}_{\text{running}}^2 + 0.007^2$) in quadrature, which works well in the case of TRAPPIST-1, for example, as shown in Fig. 6.

The running RMS increases at a transit, which additionally lowers the risk of overfitting the light curve and thus hiding transits. To maximise the increase of the uncertainty estimate near a transit, we must make sure that a typical transit fits into the range of data points that we include for the running RMS. For computational efficiency, we set the running RMS of each data point again to the RMS of a fixed number of data points $n_{dp}$ nearest to the respective data point. Since a transit of a habitable planet orbiting a UCD is expected to take about one hour, we choose $n_{dp}$ for each star and day such that the median of the time difference between the last and the first included data point is equal to one hour. Transits are not expected to be ubiquitous in the light curve, and therefore we do not expect them to significantly impact the final GP hyperparameters and consequently the estimate of the typical variation time scale.

The FWHM of the kernel function is generally between 1 and 9 hours, while the logarithm of the amplitude hyperparameter varies between -18 and -12.

The most important products of the different steps described in this section are displayed for two stars in Figs. 7 and 8.

![Figure 6. Gaussian Process fit of the normalised differential light curve with a transit of TRAPPIST-1b vs. Heliocentric Julian Date (HJD) applying different measurement uncertainty estimates (blue, violet, red, and orange solid lines). The black dots show the light curve values binned in 7-minute bins. Using the running RMS as the measurement uncertainty leads to a GP mean which closely follows the light curve trends. Quadratically adding a constant additional error term to the running RMS leads to a GP fit that still captures the overall variability of the star itself but less of the planetary transits.](image)

### 3.8 Transit search

To find transits in the GP-detrended differential light curves, we use the Box-fitting Least Squares algorithm (BLS) by Kovács et al. (2002). BLS folds the light curve on a given number of periods, bins the folded light curves, and fits a box-shaped transit to them. It then provides us with the transit parameters which best fit the analysed light curve. Namely, the period, transit depth, signal residue, and the bin number in the folded light curve where the transit starts and ends respectively. From this, we can reconstruct the position of the BLS-transits in the light curves, and visually compare these potential transits to the neighbouring light curve regions and the ALC to assess the BLS output.

We tested 10,000 orbital periods ranging from 0.8 to 10 days, but no additional planets except those of TRAPPIST-1 were found.

### 4 PIPELINE OUTPUT FOR TRAPPIST-1

We assessed the performance of our pipeline using the data collected on TRAPPIST-1 since it is the only star in the data set known to host transiting planets. Fig. 9 shows the TRAPPIST-1 light curve folded with the BLS peak period and the respective box-shaped transit. It shows that we correctly identify all transits of TRAPPIST-1b (in red) using the completely automated pipeline. We then removed the transits of TRAPPIST-1b from the respective light curve and reran the GP-detrending and BLS. This revealed three out of four transits of TRAPPIST-1c present in the light curve since BLS found an alias of the true period to be the most likely orbital period. Fig. 10 displays the TRAPPIST-1 light curve folded with the new BLS peak period. All but one of the TRAPPIST-1c transits (in red) are located within the region where BLS predicts a transit.

TRAPPIST-1c was transiting on the second observation night while TRAPPIST-1g did so on the third. The pipeline identifies these two transits as originating from the same
planet if the raw data of the first three observation nights is fed in. Thus, it would have predicted the planetary system very early on.

To get an idea of the accuracy of the BLS output parameters, we compare our results to the TRAPPIST system parameters as derived from Spitzer data by Gillon et al. (2017). The transit depth for planets b and c (0.576 per cent, 0.586 per cent respectively) are lower than the Spitzer estimates (0.7266±0.0088 per cent, 0.687±0.010 per cent) while the transit durations for planets b and c (27±4 min, 44±12 min) are near the Spitzer values (36.40±0.17 min, 42.37±0.22 min). This is mainly due to the
box-shaped transit, which is not an adequate representation of a transit for a transit depth or duration analysis but is sufficient for an efficient transit search. Given this effect, the impact of the GP fitting on the transit depth is expected to be minor for the TRAPPIST-1 light curve. The retrieved periods are both accurate up to less than 2 minutes if we account for the aliasing factor in the period of planet c.

We removed the transits of planets b and c in the TRAPPIST light curve, leaving gaps at the respective positions, and included the resultant light curve in the following analysis.
5 FLARE ENERGIES

The population of early and mid-type M dwarfs hosts a comparably high fraction of flaring stars. In a study by Günther et al. (2019) based on two months of TESS mission data (Ricker et al. 2014), about 10 per cent of early M dwarfs and about 30 per cent of mid M dwarfs showed observable flares. The flare statistics for late M dwarfs and brown dwarfs are less well understood. We detected a considerable number of flares in our UCD data set and estimated the flare energies as part of this global analysis. For this, we followed the approach outlined in Shibayama et al. (2013), which is based on the assumption that the spectrum of white-light flares is consistent with blackbody emission at a constant effective temperature. We set the effective temperature of the flare $T_{\text{flare}}$ to 9,000 K (as in Kretzschmar 2011; Günther et al. 2019). Based on the normalised light curve evolution, we know the excess flux $C'$ originating from the flare relative to the stellar flux, i.e. the ratio between the observed luminosity of the flare divided by the observed luminosity of the star. Assuming the effective temperature $T_{\text{eff}}$ of the UCDs to be 2,700 K and that the effective temperature of the flare is constant, we computed the area $A_{\text{flare}}$ of the flare taking account of the response function of the TRAPPIST instrument $R_{\text{f}}$ and using the stellar radii $R_{\ast}$ in Table 1:

$$A_{\text{flare}}(t) = C'(t)\pi R_{\ast}^2 \int R_{\text{f}} B_{\text{flare}}(T_{\text{eff}}) \lambda d\lambda \int R_{\text{f}} B_{\text{flare}}(T_{\text{flare}}) \lambda d\lambda. \tag{9}$$

The response function $R_{\text{f}}$ is equal to the CCD window transmission multiplied by the quantum efficiency, the reflectance of the mirrors (twice) and the transmission function of the I+z' filter. The estimate of the flare area then serves to compute the bolometric flare luminosity for each flux measurement by the Stefan-Boltzmann law, whereas the total flare energy is computed by integrating over the duration of the flare:

$$E_{\text{flare}} = \int_{\text{flare}} \sigma_{SB} T_{\text{flare}}^4 A_{\text{flare}}(t) dt, \tag{10}$$

where $\sigma_{SB}$ is the Stefan-Boltzmann constant. Shibayama et al. (2013) estimated the uncertainty in the flare energy to be ±60 per cent. Since we used ground-based measurements while Shibayama et al. (2013) used Kepler data, we expect the uncertainty of our estimates to be higher due to the uncertainty in our ALC solution. We neglected the transmission of the atmosphere in Eq. 9 because it affects the result at a level below 0.1 per cent of the flare energy as computed using ESO’s SkyCalc Sky Model Calculator based on Noll et al. (2012) and Jones et al. (2013).

The results are displayed in Table 2. The range of the flare energy estimates is similar to but lower than the values for M dwarfs found by Günther et al. (2019). For most targets there is an effective temperature estimate derived by the SPECULOOS team as described in the appendix of Gillon et al. (2020). The majority of these effective temperatures is near 2,700 K, the value we used to compute the flare energy estimates. Computing the flare energies with the effective temperature of the stars set to 2,000 K leads to results which are one order of magnitude lower than those in Table 2.
6 TRANSIT INJECTION TESTS

To assess the sensitivity of the survey, we calculated the probability of detecting planets orbiting their stars with a given orbital period and radius. This task involves estimating the probability that the planetary orbit allows transits (P(geometry)) as well as the probability that we identify a planet in such an orbit (P(identify)). The former can be computed based on the stellar and tested planetary parameters alone. The latter is estimated by simulating the effect of such planets on the light curves and testing whether we can recover the transits. The planet occurrence rate for UCDs are not well-known and could vary within the group of UCDs. Due to the size of our sample, we treat the UCDs as a homogeneous group in this analysis.

6.1 Planetary and stellar parameters

We tested the ability of the pipeline to detect planets for different orbital parameters and planetary radii by generating 10,000 sets of orbital period (P), phase, inclination (i) and planetary radius (Rp) for each star. Since there are 40 targets in UCDTS, we evaluated the light curves resulting from 400,000 different configurations in total. In this analysis, we restricted the orbits to circular ones. The period, radius, and phase were randomly drawn from a uniform distribution, whereas for the inclinations i we can exploit the fact that \( \cos(i) \) is uniformly distributed for randomly chosen orbits. We did not test orbits with inclinations which lead to impact parameters greater than 1 (i.e. inclination angles smaller than \( \frac{\pi}{2} - \arctan\left(\frac{R_p}{a}\right) \) with \( R_p \) being the stellar radius and \( a \) the semi-major axis) since these transits are unlikely to be recovered by BLS and for an observer they are difficult to distinguish from stellar or atmospheric effects. We account for the orbits with impact parameter above 1 in Section 6.3 with \( P_{\text{geometry}} \).

The Mandel & Agol (2002) transit model in PyTransit (Parviainen 2015) was used to compute model transit light curves.

In addition to the randomly drawn orbital parameters, PyTransit needs the stellar masses, radii, and limb darkening coefficients. The limb-darkening coefficients were set to \([0.65, 0.28]\), which are those of TRAPPIST-1 for the I+z filter in Gibbons et al. (2016) inferred from Claret & Bloemen (2011). The extent of limb-darkening varies by stellar host, but all stars are similar to TRAPPIST-1 in spectral type. We found the results to depend only marginally on the limb-darkening coefficients.

The stellar mass was derived as in Benedict et al. (2016) using the apparent 2MASS Ks-band magnitude and the Gaia DR2 parallax. The absolute Ks magnitudes of seven targets are above 10, whereas for the remaining stars the mass estimation polynomial leads to a very low mass estimate. To avoid using non-physical stellar masses, we set a minimum of 0.075 M⊙ to the stellar mass estimate. This mass was chosen because it is the approximate mass of an M9 star (Reid & Hawley 2005). Some of the targets could be brown dwarfs for which we overestimate the mass by setting a minimum of 0.075 M⊙. Consequently, this leads to an underestimated probability of finding the planet in a transiting orbit. However, setting a lower limit does not change the results significantly and thus, the results overall do not depend critically on the exact value of the lower limit. We neglect reddening since all targets are within 30 pc and therefore within the Local Bubble where reddening is negligible (Holmberg et al. 2007).

Next, we needed to estimate the stellar radius to infer the transit depth of a planet with a given size. The relationship between the Ks-band magnitude and the radius for M dwarfs described in Mann et al. (2015) provides the desired estimate for Ks below 9.8. Since the radius-magnitude relationship still provides estimates in the expected radius range even if the Ks-band magnitude is slightly above 9.8, we used the Mann radius estimate for all target stars. The uncertainties of the Gaia parallax and the Ks magnitude measurements are low and Mann et al. (2015) report a fractional residual of less than 5 per cent for their fitting polynomial. However, we applied their polynomial for stars at the
upper end of the absolute magnitudes considered. Therefore, the estimates suffer from low number statistics.

As the last input parameter of PyTransit, the semi-major axis scaled by the stellar radius was derived using Kepler’s third law.

Light curves with simulated planets were then computed by multiplying the real final light curves by the model light curves from PyTransit. Running the BLS algorithm on the product then provided us with a guess for the location, depth, and duration of the injected transits. Next, we need to define when BLS result is to be considered as a correct recovery.

6.2 Identification criterion

We know at which time point in the light curve we injected transits and where BLS detected a transit-like signal. Thus, we compared the time of the injected transits in the light curve to the BLS-transit region to assess whether we correctly identify the injected planet. Whenever BLS recovered at least half of the transit duration at half of the transit depth (the area shaded in orange in Fig. 11), the respective transit is labelled as detected. We count an injected planet as recovered if at least two injected transits were detected by the mentioned criterion.

Other authors such as Giacobbe et al. (2012) or Petigura et al. (2013) used a criterion based on the recovered period. We found this period-based criterion to work well for calm light curves. In all other cases it is, however, too restrictive since BLS seeks to include any light curve feature resembling a transit. A transit-like signal in the original light curve can therefore cause the BLS period to differ significantly from the injected period and its aliases even if BLS found all or most injected transits. We assume that the BLS algorithm is used as an initial tool to find potential planetary transits. As done in TRAPPIST-UCDTS, the potential transits can subsequently be vetted by analysing, for instance, the stellar FWHM, the ALC, or the light curves of stars of the same or a similar spectral type. In other studies (e.g. Berta et al. 2013; He et al. 2017), a planet was counted as recovered if a single transit with detection significance above a certain threshold was found. Berta et al. (2012, 2013) additionally propagated the uncertainties associated with their detrending pipeline into the detection significance, which is favourable for MEarth, a multi-telescope survey targeting a high number of stars. TRAPPIST-UCDTS consists of a significantly lower number of stars but operates at a higher cadence than MEarth, which enables rigorous vetting of the transit candidates. Furthermore, the low number of stars allows us to keep track of the impact of the detrending procedure. If we can correctly identify two transits originating from the same planet, we can schedule follow-up observations to observe a third transit. For these reasons, the two-transit detection criterion was preferred.

Applying our identification criterion to each BLS result, we counted the number of correctly recovered planets in a given radius and period range to assess the probability ($P_{\text{identify}}$) to identify such planets. Additionally, we calculated the results for $P_{\text{geometry}}$ in the same radius and period range and combined the two probability estimates to get the detection probability.

\[ P_{\text{detect}} = P_{\text{planet}} \cdot P_{\text{geometry}} \cdot P_{\text{identify}} \]  

with

- $P_{\text{planet}}$: Probability that the tested planet orbits the respective star. This is set to 1 except where stated otherwise.
- $P_{\text{geometry}}(R_p, M_p, P)$: Probability that the inclination of the planetary orbit lies within the accepted inclination angle range (impact parameter smaller than 1). Using Kepler’s third law, we get the semi-major axis $a$ from the orbital period and the stellar mass. The geometric probability for an orbit with impact parameter below 1 is equal to $\frac{a}{R_p}$.
- $P_{\text{identify}}(R_p, M_p, R_p, P)$: Probability of identifying the planetary transits (cf. Section 6.2) if the planet orbits within the accepted inclination angle range.

Figure 11. Illustration of the detection criterion. A synthetic light curve was multiplied by a transit light curve generated with PyTransit to display a potential light curve with an injected transit. The area shaded in orange in the dashed line box indicates where the injected transit dips below half of its maximal transit depth. The transits suggested by the Box Least Squares algorithm can be shifted due to noise or other transits in the light curve. We regard a transit as detected if the BLS transit (shaded in violet in the solid line box) overlaps at least half of the transit at half depth (shaded in orange in the dashed line box).
6.4 Effect of GP-detrending on identification probability

The identification probability mainly depends on how we detrend the light curves before or while searching for transits. In the following, we discuss why we detrend the light curves using a GP before injecting transits and looking for them with BLS (detrend-inject-recovery test). This approach is intended to be as close as possible to the actual planet detection probability in a survey within which we can analyse light curves in detail and we can apply more sophisticated techniques. The approach taken might appear optimistic. However, a more optimistic detrending approach will lead to a more conservative lower limit on the occurrence rate of planets like TRAPPIST-1b later on. Furthermore, we investigated the effect of the detrending on our occurrence rate estimates to assess the impact on our conclusions.

Injecting transits first, GP-detrending after that and subsequently looking for transits (inject-detrend-recovery test) is not optimal since it necessarily decreases the transit depth. Inspecting a light curve by eye as done in TRAPPIST-UCDTS would likely yield a higher detection probability, and hence inject-detrend-recovery testing underestimates the actual probability of spotting transits. Additionally, detrending after every transit injection is computationally expensive. Detrending and searching for transits simultaneously would be optimal but is even more computationally expensive and therefore not suited for our statistical analysis.

One necessary component for computing the GP hyperparameters is the uncertainty estimate of the flux measurements. By setting this to the running RMS of the light curve and adding an additional term $\sigma_{\text{add}}$ to it in quadrature, we can adjust how closely we fit the data using the GP, as visible in Figs. 6 and 12.

The GP mean for flat light curves is also flat. This means that the detrend-inject-recovery tests yield a very similar detection probability as a simultaneous approach irrespective of $\sigma_{\text{add}}$. Therefore, we do not need to optimise $\sigma_{\text{add}}$ for these light curves. Instead, we perform inject-detrend-recovery tests on a set of 16 variable light curves with planets similar to TRAPPIST-1b and determine the $\sigma_{\text{add}}$ that maximises the detection probability. This gives us an optimal $\sigma_{\text{add}}$ for each star. Using a GP with $\sigma_{\text{add}}$ set to the mean of these 16 optimised $\sigma_{\text{add}}$, which is equal to 7 ppt, ensures that we mainly remove the general trend and not small transit-like features. We then apply this GP trained with inject-detrend-recovery tests for our detrend-inject-recovery tests. By this procedure, we seek to emulate a simultaneous search.

The impact of the GP-detrending depends on the depth of the transits since the identification probability decays very quickly with transit depth. Therefore, the detection probability of small planets is more prone to being affected by GP-detrending. For a typical variable star, the identification probability for the two test cases is compared in Fig. 12 if we inject planets similar to TRAPPIST-1b. The transit detection efficiency over a larger parameter space is compared in Table 3 for two stars, which shows that the detection probabilities of the two cases overall converge as we add an additional error term to the running RMS. In Table 3, we can see again that for the detrend-inject-recovery test case, the identification probability decreases with increasing $\sigma_{\text{add}}$ since a lower $\sigma_{\text{add}}$ leads to a flatter light curve. Transits injected into these flatter light curves are easier to detect. In the inject-detrend-recovery test case, the identification probability increases with increasing $\sigma_{\text{add}}$ because we remove less of the transit depth (cf. Fig. 6).

In Section 7, we derive a lower limit on the occurrence rate of planets similar to TRAPPIST-1b. As outlined above, this value depends on the detrending. Therefore, we compute it for different detrending scenarios.

6.5 Expected number of planets in the data set

The probability of finding $n$ planets in the entire sample is described by a Poisson binomial distribution. The mean of this distribution is equal to the number of planets which we expect to find in the data set, defined as

$$E_{\text{detected planets}}(R_p, P) = \sum_{j=1}^{\text{nr of target stars}} P_{\text{detect}_j}(R_p, P),$$

in which $R_p$ is the radius of the planet, $P$ is its orbital period, and $P_{\text{detect}_j}$ is computed as described in Eq. 11. In Fig. 13 we illustrate $E_{\text{detected planets}}(R_p, P)$ with the occurrence rate of the tested planets set to 1.

Additionally, we can calculate the probability of not finding any planet for a given star and the probability of
R

Radius \[\oplus\] 1.5

As visible in Fig. 14, an Earth-sized planet with an orbital period above 3 days is unlikely to be detected even if these planets are assumed to be frequent. Furthermore, we can see that the detection of at least one close-in planet in the data set is likely if there is one around every single star with randomly drawn orbital parameters (such as the inclination).


\[
P_{\text{no-detection}}(R_p, P) = \frac{\text{nr of target stars}}{\prod_{j=1} \left(1 - P_{\text{detect}}(R_p, P)\right)}.
\]

(13)

As visible in Fig. 14, an Earth-sized planet with an orbital period above 3 days is unlikely to be detected even if these planets are assumed to be frequent. Furthermore, we can see that the detection of at least one close-in planet in the data set is likely if there is one around every single star with randomly drawn orbital parameters (such as the inclination).
7 INJECTION RESULTS

In Fig. 15 the probability of not detecting a single planet, assuming that 10 per cent of all stars host a planet of the tested type in a randomly chosen orbit, is displayed. This analysis indicates that the probability of finding at least one planet similar to TRAPPIST-1b (in radius and orbital period) in the entire data set is equal to 5 per cent if the occurrence rate for such planets is equal to 10 per cent. In other words, in this case the probability of not detecting a planet like TRAPPIST-1b is equal to 95 per cent and the TRAPPIST team would have been very lucky to find a planet as they did. Since TRAPPIST-1b was found in this survey, we can conclude that the occurrence rate is likely to be above 10 per cent.

In Fig. 16, we display the probability of finding at least one planet in the data set as a function of the occurrence rate for three different detrending scenarios. The results quoted here refer to the case where we detrend before injecting planets and quadratically add 7 ppt to the initial uncertainty estimate except where stated otherwise.

Additionally to the result that the occurrence rate is likely to be above 10 per cent if the likelihood of finding at least one planet is expected to be greater than 5 per cent, we see that an occurrence rate of only 2 per cent would lead to a likelihood of 99 per cent of not detecting a planet like TRAPPIST-1b in such a survey.

7.1 Dependence on detrending

As a first estimate of $P_{\text{planet}_{\text{sim}}}$, the lower limit on the occurrence rate of planets like TRAPPIST-1b, we derive the occurrence rate for which the probability of no planet detection in the data set is equal to 95 per cent:

$$0.95 = \prod_{j=1}^{\text{nr of target stars}} \left(1 - P_{\text{planet}_{\text{sim}}} \cdot P_{\text{geometry}_j} \cdot P_{\text{identify}_j}\right).$$

We estimate that the geometric probability is accurate up to 10 per cent, in which case the identification probability dominates the result.

If we overfit the light curve in the GP-fitting or any of the previous steps, we produce very flat light curves, within which injected transits are easy to identify compared to transits injected in more variable light curves. Thus, overfitting before the planet injection leads to a higher identification probability $P_{\text{identify}}$.

As visible in equation (14), the lower limit on the occurrence rate $P_{\text{planet}_{\text{sim}}}$ decreases as $P_{\text{identify}}$ increases. Consequently, choosing a more optimistic detrending leads to a more conservative estimate of $P_{\text{planet}_{\text{sim}}}$.

Ideally, we would search for transits while simultaneously fitting stellar variability. Finding an efficient algorithm that performs light curve detrending simultaneously with the transit search, and is suited for statistical analyses, is out of scope of this paper. To test how different non-simultaneous detrending scenarios affect the lower limit, we computed it for three different cases.

Using a GP to detrend without simultaneously fitting for transits decreases the detection probability since the GP will try to fit the injected transits. Thus, dividing by the GP mean prediction reduces the depth of the injected transits. We might have found transiting planets by visual in-

Figure 15. Probability of no planet detection assuming that 10 per cent of all stars have a planet of the tested type. The black diamond shows the period and radius of TRAPPIST-1b.

Figure 16. Probability of finding at least one planet similar to TRAPPIST-1b as a function of the occurrence rate resulting from Eq. (14) for three detrending scenarios. The optimistic scenario (orange) describes the detection probabilities if we detrend the light curves without removing any part of the transits. In the intermediate scenario (violet-red), we still correct for variability without removing transits but we compromise on how well the stellar variability is removed. In the pessimistic scenario (blue), we use the probabilities to identify planets in light curves which have been GP-detrended after the transit injection. In the intermediate scenario, if the occurrence rate is below 10 per cent (14 per cent in the pessimistic scenario, 7 per cent in the optimistic scenario), the probability of not finding any planet similar to TRAPPIST-1b in period and radius orbiting a ultra-cool dwarf, exceeds 95 per cent, which means that the probability of finding at least one planet is below 5 per cent.
we drop the dependence on stellar radii and masses and denote the marginalised detection probability for star \( j \) as \( F_j \). The probability of finding a planet around star \( j \) results as \( r \cdot F_j \), where \( r \) is the number of planets we expect to find within the mentioned radius and period range.

Adding over all possible combinations of single-planet systems and weighting them with their corresponding probability, we finally get:

\[
P(D \mid r) = r f(\Theta_{1b}) \cdot \prod_{j \in S} (1-rF_j).
\]

where \( f(\Theta_{1b}) \) is the probability of finding TRAPPIST-1b around its host star and \( S \) is the set of all target stars except TRAPPIST-1. In fact, the detection probabilities for the injected planets also depend on the probability that a human confirms a correct BLS planet detection. This varies depending on the stellar and planetary parameters and the individual observer, but a characterisation of this function is out of scope of this paper. As explained in Section 6.2, we assume that the flexible design of the survey and the definition of the identification criterion allow us to set the confirmation probability of a correct BLS detection to 1.

In Fig. 17 and 18, we show \( P(D \mid r) \) of TRAPPIST-UCDTS (dashed red line). Furthermore, we averaged over the detection probabilities to further estimate the shape of \( P(D \mid r) \) for a higher number of surveyed stars. Since the Kepler-informed prior puts a low weight on some regions of the parameter space where the detection probability is high, the marginalised detection probabilities are lower compared to the results derived using a uniform prior. Therefore, the peak of \( P(D \mid r) \) is at a higher \( r \) in Fig. 18 than in Fig. 17 for the same number of stars.

The low number of target stars restricts us from drawing strong conclusions on the occurrence rate of short-period planets hosted by UCDs. A larger sample would be more indicative, as illustrated in Fig. 17.

From a Bayesian perspective (e.g. Hall et al. 2018), if one chooses a prior distribution on \( r \), then the likelihood \( P(D \mid r) \) may be inverted to a posterior distribution \( P(r \mid D) \). If the prior is uniform \( P(r) = 1 \) then \( P(r \mid D) \propto P(D \mid r) \), and the curves in Fig. 17 may be interpreted as Bayesian degrees of belief in the unknown parameter \( r \).

8 LIKELIHOOD ANALYSIS

Alternatively, we can calculate the likelihood of the occurrence rate \( r \). \( P(D \mid r) \) is the probability of finding a data set \( D \) with exactly one planet similar to TRAPPIST-1b around TRAPPIST-1 and no planets around the other surveyed stars, given an occurrence rate \( r \) within the parameter range 0.9 to 3 days and 1 to 4 \( R_\oplus \). This accompanies the probability of finding a planet around at least one of the surveyed stars as described in section 7. We outline the derivation below, while a detailed explanation can be found in Appendix A.

First, we marginalised the detection probabilities over the analysed radius and period range (0.9 to 3 days, 1 to 4 \( R_\oplus \)) because the planetary and orbital parameters of a potential undiscovered planet are unknown. For this, it is necessary to choose a prior reflecting our knowledge about where in the parameter space planets are likely to be found. Since this is unknown for UCDs, we tested two different priors. As our first prior, we chose a uniform prior. By using this prior for the marginalisation of the detection probabilities, equal weight is given to all potential planets within the tested parameter range. This prior may not be optimal given that for low-mass stars e.g. hot Neptunes are more rare than their warm counterparts (e.g. Dressing & Charbonneau 2015; Hirano et al. 2018). As the second prior we chose the normalised occurrence rates of M dwarfs as computed by Dressing & Charbonneau (2015) using Kepler data. The following steps were carried out separately for the two sets of marginalised detection probabilities resulting from the two priors. The different results are shown in Figs. 17 and 18.

9 CONCLUSIONS

9.1 Survey results

We performed aperture photometry, generated differential light curves and cleaned these from non-planetary signals (cf. Section 3) and also from slow stellar and residual instrumental variability (cf. Sections 3.7, 6.4). Through injection-recovery tests we could then infer, inter alia, the number of planets we expect to find in such a survey (cf. Section 7). The results rely on our estimates of the stellar parameters in Table 1 and the applied identification criterion in Section 6.2. We expect to find 0.52 planets similar to TRAPPIST-1b in radius and period and 0.2 planets similar to TRAPPIST-1c if all stars analysed in this study are orbited by a planet of the tested type. Therefore, it is not surprising that only one system was detected.

Using the pipeline, we would have found the planets b
He et al. (2017) performed a planet injection analysis on brown dwarf light curves acquired by the Spitzer Space Telescope. Their study lead to an expectation value for the number of discovered planets of (0.6, 0.81, 0.83, 0.85, 0.89) within the radius bins (1±0.25, 1.5±0.25, 2±0.25, 2.5±0.25, 3±0.25) $R_\oplus$ for an orbital period between 1.28 and 3 days. We expect to find (0.24, 0.48, 0.59, 0.62, 0.64) planets in the same parameter range. They observed 44 brown dwarfs with a median observation time of 20.9 hours. Thus, they have observed their targets for a shorter amount of time. However, they count one well-observed planetary transit as a detection while we require the detection of at least two transits. Furthermore, their observations are almost continuous while ours can be quite spread out in time, which can adversely affect the detection probability. The expectation value in He et al. (2017) flattens out for radii above 1.5 $R_\oplus$ while this is the case for radii above 2 $R_\oplus$ in our analysis. This is likely the case because the Spitzer light curves do not contain short-term variations due to, for example, precipitable water vapour.

Injection-recovery tests by Demory et al. (2016) on 189 late M dwarfs observed in K2’s Campaigns 1–6 yielded a recovery efficiency of 10 per cent for planets similar to TRAPPIST-1b. K2 has been monitoring each campaign field quasi-continuously for approximately 80 days, which is much longer than our typical summed exposure time of 50 hours. Nevertheless, we got a detection probability of 30 per cent in the period and radius bin of TRAPPIST-1b. Possible reasons for this difference are the fact that late M dwarfs are faint in the Kepler passband and that the Kepler cadence of 30 minutes can smear out the transits. Furthermore, Demory et al. (2016) used a criterion based on the BLS period, which is likely to yield a lower detection probability for active stars compared to our detection criterion defined in Section 6.2.

### 9.2 Multiplanetary systems

Planetary systems like TRAPPIST-1, consisting of multiple, similarly sized planets in edge-on, coplanar orbits, produce comparably deep transits with the characteristic, sharp transit shape. A human analysing a light curve of such a planetary system would connect the transits produced by different planets and further observe this target star. The same is true for the BLS algorithm, as in the case of TRAPPIST-1, it first connects the transits of TRAPPIST-1c and TRAPPIST-1g before finding the transits of TRAPPIST-1b, which are very clear compared to the first two transits. Thus, multiplanetary systems similar to TRAPPIST-1 are ideal targets for surveys such as TRAPPIST-UCDTS. Generally, multiplanetary coplanar systems lead to a higher identification probability. If all surveyed stars are orbited by two planets with planetary radii between 1 and 1.3 $R_\oplus$ at randomly drawn inclinations, one of them with a period between 1 and 1.4 days and the other with a period between 1.8 and 2.3 days, we estimate the probability of detecting at least one of the planets to 69 per cent (cf. Fig. 14). If coplanarity is assumed, however, the innermost planet is the dominant contributor to the detection probability which is thus above 57 per cent for the mentioned case. We must conclude in both cases that a close-in planetary system, consisting of one or more planets, is relatively likely to be discovered in a survey like

---

**Figure 17.** Probability of the survey result as a function of the occurrence rate $r$ for different numbers of surveyed stars divided by the maximum of the respective probability curve. A uniform prior was used to marginalise over the unknown planetary parameters. The most likely occurrence rate for TRAPPIST-UCDTS is currently equal to 1 but shrinks to 0.3 as we observe 200 target stars without detecting another planetary system.

**Figure 18.** Probability of the survey result as a function of the occurrence rate $r$ except that the normalised Kepler occurrence rates for M dwarfs were used as the prior. This prior essentially excludes planets with orbits below 2 days and radii above 2 $R_\oplus$. and $c$ automatically, which validates the pipeline. An indication for the existence of planet $c$ would have been noticed first because it transits during the second observation night and its transit is clearly visible and easily identified by the pipeline.

In accordance with previous results (e.g. Dressing & Charbonneau 2015; Demory et al. 2016), we find that hot mini-Neptunes are likely to be rare around cool stars since the probability of missing these planets is low (cf. Fig. 14) and we found none in the data set. Figs. 13 and 14 further indicate that we are sensitive to Earth-sized planets with short periods of the order of up to three days only.
TRAPPIST-UCDTS, if these systems are common around UCDs. Therefore, we expect to find some multi-planet systems in surveys such as SPECULOOS (Burdanov et al. 2018; Delrez et al. 2018b; Gillon 2018; Jehin et al. 2018), consisting of multiple telescopes fully dedicated to the search of exoplanets orbiting UCDs and achieving a higher photometric precision than TRAPPIST-South, if compact coplanar planetary systems are frequent. The latter is supported by evidence that compact multiple systems are relatively common around mid-type M dwarfs. Considering planets with periods of less than 10 days, Muirhead et al. (2015) found that 21\% of photometric signals with simultaneous variability correction this indicates that a rigorous computational search for periodic transit signals with simultaneous variability correction is necessary.

9.3 Transit detection challenges

With few exceptions (Dalal et al. 2019; Huber et al. 2013), measurements of the obliquities of stars hosting multiple planets (e.g. Hirano et al. 2020; Sanchis-Ojeda et al. 2012; Hirano et al. 2012; Albrecht et al. 2013; Chaplin et al. 2013) suggest that the axis of rotation of the host star is likely to be close to parallel to the orbital axis of its planets. Stars with close-in transiting planets, are therefore unlikely to exhibit a quiet light curve since the starspot and the inhomogeneous cloud coverage expected on the stellar surface of late-type M dwarfs and brown dwarfs (Metchev et al. 2015; Goldman 2005) will constantly change as seen from Earth due to stellar rotation. This is critical given the fast rotation of many brown dwarfs and very-low-mass stars of the order of one day (e.g. Irwin et al. 2011; Scholz 2016). Additionally, precipitable water vapour can cause variability in the light curves (Bailer-Jones & Lam 2003; Murray et al. 2020) further complicating the confirmation of a transit candidate.

Therefore, it is likely that some transits, especially grazing transits, have not been identified by eye if the time scale and amplitude of the brightness variations during the respective night are comparable to those of planetary transits. This indicates that a rigorous computational search for periodic transit signals with simultaneous variability correction is necessary.

ACKNOWLEDGEMENTS

We thank the anonymous referee for their thorough review which helped elevate the quality of this work. FL gratefully acknowledges a scholarship from the Fondation Zdenek et Michaela Bakala. The research leading to these results has received funding from the European Research Council under the FP/2007-2013 ERC Grant Agreement no 336480 and from the ARC grant for Concerted Research Actions, financed by the Wallonia-Brussels Federation. TRAPPIST is funded by the Belgian Fund for Scientific Research (Fonds National de la Recherche Scientifique, FNRS) under the grant FRFC 2.5.594.09.F, with the participation of the Swiss National Science Foundation (SNF). MG and EJ are F.R.S.-FNRS Senior Research Associates. AM acknowledges support from the senior Kavli Institute Fellowships. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France.

DATA AVAILABILITY

Data will be available via VizieR at CDS.

REFERENCES

Albrecht S., Winn J. N., Marcy G. W., Howard A. W., Isaacson H., Johnson J. A., 2013, ApJ, 771, 11
Bailer-Jones C. A. L., Lam 2003, MNRAS, 339, 477
Benedict G. F., et al., 2016, AJ, 152, 141
Berta Z. K., Irwin J., Charbonneau D., Burke C. J., Falco E. E., 2012, AJ, 144, 145
Berta Z. K., Irwin J., Charbonneau D., 2013, ApJ, 775, 91
Broeg C., Fernández M., Neuhäuser R., 2005, Astronomische Nachrichten, 326, 134
Burdanov A., Delrez L., Gillon M., Jehin E., 2018, SPECULOOS Exoplanet Search and Its Prototype on TRAPPIST. p. 130, doi:10.1007/978-3-319-55337-7_130
Byrd R. H., Lu P., Nocedal J., Zhu C., 1995, SIAM Journal on Scientific Computing, 16, 1190
Chabrier G., Baraffe I., 1997, A&A, 327, 1039
Chaplin W. J., et al., 2013, ApJ, 766, 101
Claret A., Bloemen S., 2011, A&A, 529, A75
Cutri R. M., et al., 2003, VizieR Online Data Catalog, p. II/246
Dalal S., Hébrard G., Lecavelier des Étangs A., Petit A. C., Bouvier V., Laskar J., König P. C., Correia A. C. M., 2019, A&A, 631, A28
Damasso M., Giacobbe P., Caledides P., Sossieti A., Lattanzi M. G., Bernagazzhi, Bertolini E., Smart R. L., 2010, PASP, 122, 1077
Davenport J. R. A., et al., 2014, ApJ, 797, 122
Delrez L., et al., 2018a, MNRAS, 475, 3577
Delrez L., et al., 2018b, in Proc. SPIE. p. 1070011, doi:10.1117/12.2321475
Demory B.-O., Queloz D., Alibert Y., Gillen E., Gillon M., 2016, ApJ, 825, L25
Dieiterich S. B., Henry T. J., Jao W.-C., Winters J. G., Hosey A. D., Riedel A. R., Subasavage J. P., 2014, AJ, 147, 94
Dressing C. D., Charbonneau D., 2013, ApJ, 767, 95
Dressing C. D., Charbonneau D., 2015, ApJ, 807, 45
Ducrot E., et al., 2020, arXiv e-prints, arXiv:2006.13826
Fletcher R., 1987, Practical methods of optimization. Wiley, Chichester New York
Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, AJ, 154, 220
Gaia Collaboration et al., 2016, A&A, 595, A1
Gaia Collaboration et al., 2018, A&A, 616, A1
Giacobbe P., et al., 2012, MNRAS, 424, 3101
Gillon M., 2018, Nature Astronomy, 2, 344
Gillon M., Jehin E., Magain P., Chantry V., Hutsemékers D., Manfroid J., Queloz D., Udry S., 2011, in European Physical Journal Web of Conferences. p. 06002, doi:10.1051/epjconf/2010106002
Gillon M., Jehin E., Funel A., Magain P., Queloz D., 2013, in European Physical Journal Web of Conferences. p. 03001, doi:10.1051/epjconf/20134700001
Gillon M., et al., 2016, Nature, 533, 221
Gillon M., et al., 2017, Nature, 542, 256
Gillon M., et al., 2020, arXiv e-prints, arXiv:2002.04798
Goldman B., 2005, Astronomische Nachrichten, 326, 1075
Green D. A., 2011, Bulletin of the Astronomical Society of India, 39, 289
Appendix A: Occurrence Rate

In the following, we summarise the stellar and planetary parameters \( (R, P, M_*, R_*) \) of the potential star-planet system \( j \) in \( \Theta_j \). These parameters for the whole survey are thus encoded in \( \Theta \). Additionally, we abbreviate TRAPPIST-1 as T1. \( \Theta_{T1} \) thus stands for the planetary radius and period of TRAPPIST-1b and the stellar radius and mass of its host star TRAPPIST-1.

The probability of detecting a planet around star \( j \), if it hosts a planet, can be derived using

\[
P(\text{detect planet} \mid \Theta_j, N_j = 1) = P_{\text{identify}} \cdot P_{\text{geometry}} = f(\Theta_j).
\]

(A1)

We use \( \bar{N} \) to describe the potentially hidden planetary population of the target stars. \( N_j \) is set to 1 if there is a planet around star \( j \) and it is set to 0 otherwise. We assume that the false positive probability, i.e. the probability of detecting a planet around star \( j \) if it does not host a planet, is equal to 0:

\[
P(\text{detect planet} \mid \Theta_j, N_j = 0) = 0.
\]

(A2)

Additionally, we assume that we correctly infer the existence of the planet if the detection criterion in Section 6.2 is met. We set the probability that there is a planet orbiting planet \( j \) to:

\[
P(N_j = 1) = r.
\]

(A3)

Analogously,

\[
P(N_j = 0) = 1 - r.
\]

(A4)

Next, we have to marginalise over detection probabilities since we do not know the planetary parameters of the potential unknown planets. For this, we have to make an assumption about the occurrence rates of planets within the chosen parameter range. In case 1) we marginalise using a uniform prior. In this case, we assume that we have no indication about the occurrence rate of planets orbiting UCDs. The parameter space covering short orbits and large radii, where we have a high detection probability but planets seem to be rare, is thus treated the same as short orbit Earth-sized planets which might be more common. The occurrence rates for UCDs are unknown, but there are occurrence rate estimates for M dwarfs in general. In case 2), we use the occurrence rates from M dwarfs derived from the Kepler mission data by Dressing & Charbonneau (2015), normalised over the parameter space, as our new prior. We thus essentially exclude hot Neptunes. In both cases, we set the minimal planetary radius \( R_{\text{min}} \) to 1 \( R_\oplus \), the maximal radius \( R_{\text{max}} \) to 4 \( R_\oplus \), the shortest orbital period \( P_{\text{min}} \) to 0.9 days, and the longest period \( P_{\text{max}} \) to 3 days to match the bin choice in Dressing & Charbonneau (2015) figure 12.

Case 1) Uniform prior: To marginalise over the unknown planetary radius and period, we integrate over a uniform prior \( P(R, P) \):

\[
F_j = \int_{R_{\text{min}}}^{R_{\text{max}}} \int_{P_{\text{min}}}^{P_{\text{max}}} f(\Theta_j) dR dP \left( \frac{R_{\text{max}} - R_{\text{min}}}{P_{\text{max}} - P_{\text{min}}} \right).
\]

(A5)

Case 2) Normalised Kepler occurrence rates for M dwarfs as prior: To marginalise over the unknown planetary radius and period we integrate over a uniform prior \( P(R, P) \):

\[
F_j = \int_{R_{\text{min}}}^{R_{\text{max}}} \int_{P_{\text{min}}}^{P_{\text{max}}} f(\Theta_j) dR dP.
\]

(A6)
where \( p(\Theta_j) \) are the occurrence rates as in Dressing & Charbonneau (2015) figure 12 but normalised over our integration range since we want to use the parameter \( r \) as the number of planets per star within this range. In both cases we assume our stellar masses and radii to be correct.

For both cases we continue as follows. Our data \( D \) is the survey outcome that we initially find T1b but no other planets. The assumption here is, that T1b is the true trigger for the detection of the other TRAPPIST-1 planets.

• A: detect T1b in the T1 light curve.
• B: detect no other planets in the other light curves.

which is equal to

\[
P(D|\Theta) = r f(\Theta_{T1b}) \cdot \prod_{j \in S} (1 - r F_j).
\]

This paper has been typeset from a TeX/LaTeX file prepared by the author.