

# Expected performances of the Characterising Exoplanet Satellite (CHEOPS)

## III. Data reduction pipeline: architecture and simulated performances

S. Hoyer<sup>1</sup>, P. Guterman<sup>1,2</sup>, O. Demangeon<sup>1,3</sup>, S. G. Sousa<sup>3</sup>, M. Deleuil<sup>1</sup>, J. C. Meunier<sup>1</sup>, and W. Benz<sup>4</sup>

<sup>1</sup> Aix Marseille Univ., CNRS, LAM, Laboratoire d'Astrophysique de Marseille, Marseille, France  
e-mail: [sergio.hoyer@lam.fr](mailto:sergio.hoyer@lam.fr)

<sup>2</sup> Division Technique INSU, BP 330, 83507 La Seyne Cedex, France

<sup>3</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal

<sup>4</sup> University of Bern, Physikalisches Institut, Gesellschaftstrasse 6, 3012 Bern, Switzerland

Received 16 July 2019 / Accepted 16 September 2019

### ABSTRACT

The CHaracterizing ExOPlanet Satellite (CHEOPS) is set to be launched in December 2019 and will detect and characterize small size exoplanets via ultra high precision photometry during transits. CHEOPS is designed as a follow-up telescope and therefore it will monitor a single target at a time. The scientific users will retrieve science-ready light curves of the target that will be automatically generated by the CHEOPS data reduction pipeline of the Science Operations Centre. This paper describes how the pipeline processes the series of raw images and, in particular, how it handles the specificities of CHEOPS data, such as the rotating field of view, the extended irregular point spread function, and the data temporal gaps in the context of the strict photometric requirements of the mission. The current status and performance of the main processing stages of the pipeline, that is the calibration, correction, and photometry, are presented to allow the users to understand how the science-ready data have been derived. Finally, the general performance of the pipeline is illustrated via the processing of representative scientific cases generated by the mission simulator.

**Key words.** methods: data analysis – techniques: image processing – techniques: photometric – space vehicles: instruments – methods: observational – planetary systems

## 1. Introduction

The CHaracterizing ExOPlanet Satellite (CHEOPS) is an ESA small mission to be launched in December of 2019. CHEOPS is designed as a follow-up instrument devoted to ultra high precision photometry in order to detect or precisely measure transits of small size exoplanets already known via radial velocity measurements or via transit searches (Broeg et al. 2014; Benz et al. 2018). The series of raw images acquired by the instrument will be automatically processed (with no external interaction nor interactive configuration) into a flux time series, ready for scientific analyses. As part of the Science Operations Centre, the data reduction pipeline (DRP) is in charge of producing these calibrated light curves, with associated intermediate products, which will be delivered to the scientific users. While the instrument performs ultra-high precision photometry like the CoRoT (Deru et al. 2015; Baglin et al. 2016) or *Kepler* (Jenkins et al. 2010a,b) missions, it presents different specificities that demand tailored approaches for the data reduction. In particular, the field of view of the instrument rotates around the line of sight, making background stars to roll around the target, and potentially periodically polluting its photometry. In addition, the point spread function (PSF) of the instrument measured in laboratory exhibits an extended irregular shape which, together with the temporal gaps in the data, challenges pipeline procedures, such as the detection and correction of cosmic rays hits, among others.

The present paper aims to provide to the community a complete description of the automated data reduction pipeline, as implemented in the pre-launch phase. The goal is to show how

the pipeline deals with CHEOPS specificities, allowing a better understanding of how the science-ready data have been derived. Also, this paper intends to serve as a reference for the possible use of additional pipeline products, which will complement further light curve analysis (e.g., with filtering or detrending algorithms). Finally, in the framework of the specific and strict mission photometric requirements, the expected performance of the DRP was estimated. For this, a series of simulated data for typical astrophysical configurations, as provided by the CHEOPS end-to-end simulator (Futyan et al. 2020, Paper II), were used. The reader can also find a description of the CHEOPS onground performance in Deline et al. (2020, Paper I).

The structure of the paper is the following: Sect. 2 recalls the mission profile and instrument specificities while Sect. 3 presents the pipeline architecture. Sections 4–6 detail the different processing steps operated by the main modules of the DRP and the implemented algorithms. Some processes that could be used indistinctly at any step of the pipeline are described in Sect. 7. The expected performance is reviewed in Sects. 8 and 9 summarizes and concludes this work.

## 2. Overview of the mission profile and instrument

A complete description of the instrument and the mission profile can be found in the CHEOPS Observers Manual<sup>1</sup> or in Broeg et al. (2018) and Deline et al. (2020). But to make reading

<sup>1</sup> <https://www.cosmos.esa.int/web/cheops-guest-observers-programme/ao-1>

easier, we provide a quick description of the key elements which direct the data reduction in the following paragraphs. CHEOPS will be settled on a 700 km altitude sun-synchronous orbit of about 101 min. The spacecraft is nadir locked and will continuously roll around the line of sight, ensuring a thermally stable environment for the payload radiators. Consequently, during one orbit, the background stars rotate around the optical axis of the telescope while the target star remains at the same location, modulo jitter perturbations. Because of its orbit at low altitude, it is expected that up to 40% of data could be lost due to the close passage of the Earth to the line of sight for targets far from the ecliptic in addition to the South Atlantic Anomaly (SAA) crossings (Pinheiro da Silva et al. 2008). Therefore, these losses translate into time gaps in the raw data products received by the DRP and consequently, in the final light curves delivered by the pipeline. As previously mentioned, the telescope will observe one single target at a time in a field of view of  $0.32^\circ$  in diameter. The telescope has an effective diameter of 30 cm, it has no shutter, and the focal plane is equipped with a  $1024 \times 1024$  pixels back-illuminated charge coupled device (CCD) with a pixel size of  $13 \mu\text{m}$  and a pixel scale of  $1''$ . It will operate at a nominal temperature of  $-40^\circ\text{C}$ . The focal plane is defocussed to deliver a large PSF with a 12 pixels radius encircling 90% of the flux. As a result of the combination of the Ritchey–Chrétien design and other specific features of the building of the telescope, the PSF exhibits sharp and peaky features at sub-pixel level. CHEOPS has no filter in the optical path and its bandpass covers the visible-near-infrared range of 330–1100 nm. The spectral transmission of CHEOPS is very similar to the *Gaia* G bandpass (Evans et al. 2018) as can be seen in Deline et al. (2020, Fig. 1). At launch, the telescope will have a cover for protection. The opening of the cover will occur during in-orbit commissioning after performing some tests and calibration observations of the instrument.

Full-array images will be downloaded for calibration or test purposes only. In nominal operation mode, only an image of  $200 \times 200$  pixels (default size), which is referred to as subarray, will be downlinked to ground with the associated housekeeping data. Each subarray image is usually formed by the stack of several shorter exposures, which allows one to avoid saturation during bright target observations, for example. In addition, the central region of the image is transmitted before the stacking as small imagerettes of typical size  $35 \times 35$  pixels, this thus provides a higher cadence sampling of the target. In fact, images and imagerettes are circular in order to downlink only the relevant region of the images and thus spare bandwidth.

The magnitude of targets are in the  $6 \leq V\text{-mag} \leq 12$  range but the instrument will also allow the observation of brighter or much fainter stars. To accommodate this large range of magnitudes the exposure time can be adjusted from 1 ms to 60 s. Depending on the selected exposure time, different detector read-out modes are set-up. These modes, called faint, faint-fast, bright and ultra-bright, consist of different read-out frequencies and different setting combinations of the detector read-out and the onboard processing of the image. Thus, each read-out mode has a specific duty cycle, as low as 8–50% for exposure times below one second, which leads finally to an image cadence between 1 s and 60 s (see Table 1 of CHEOPS Observers Manual for details). In addition, in order to reduce the amount of downlinked data, one image can be the stack of 1–60 short exposures.

The instrument is required to reach a photometric precision of 20 ppm for a star with a  $V$  magnitude in the range  $6 \leq V\text{-mag} \leq 9$  with 6 h of integration time, to allow the detection of an Earth-like planet around a G5V star with an orbital

**Table 1.** CDPP estimations from final light curves of case 1 and 2 (see text for description).

Science case	Int. Time (h)	Req. (ppm)	Modified CDPP		
			Theor.	Default	Optimal
1	6	20	5	6	5
2	3	85	62	117	83
2 w/o CR	3	85	62	69	67

**Notes.** The first column shows the integration time used for the metrics, the second column shows the photometric requirements for each case and the final columns shows the respective noise estimation of each analyzed light curve.

period of 50 days. At the faint end, the expected photometric precision is 85 ppm for a star of  $V\text{-mag} = 12$  in 3 h of integration time, which will allow the detection of Neptune-size planets transiting a K-type dwarf star with an orbital period of 13 days (Fortier et al. 2015; Benz et al. 2018). To achieve the high photometric stability, the instrument should operate in a thermally stable environment, minimize the various sources of straylight, and ensure a pointing stability of  $2''$  rms. This precision will be achieved by including the instrument in the attitude control loop.

### 3. Pipeline architecture

The DRP is run automatically once triggered by the processing framework. There is no interaction with external agents and there is no interactive configuration of the pipeline. The complete processing can be separated in three main steps: (1) the calibration module which corrects the instrumental response, (2) the correction module on charge of correcting environmental effects and (3) the photometry module which transforms the resulting calibrated and corrected images into a calibrated flux time series or light curve. Each of these modules consist in successive processing steps, as presented in Fig. 1, which are run sequentially as the output of one step is used as an input of the next one. The next sections detail the different processing steps and the adopted algorithms of each of these three main modules. Some additional modules which are used indistinctly at any point of the pipeline are described in Sect. 7.

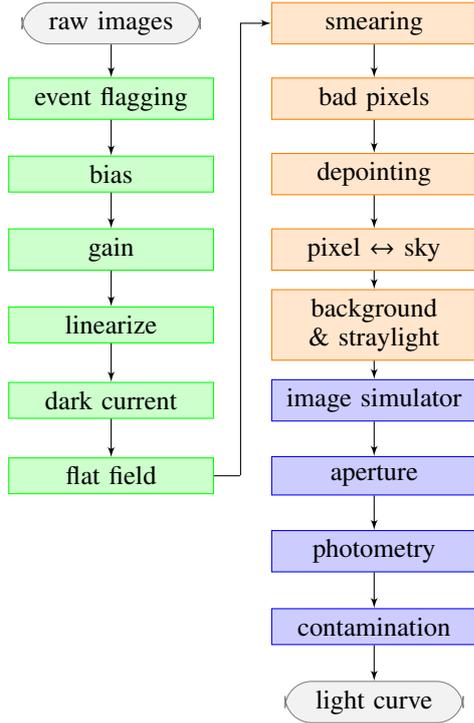
In addition to the reduced light curves, the pipeline generates a visit processing report. This report allows the user to get direct insight into the performance of each step of the data reduction.

### 4. Calibration

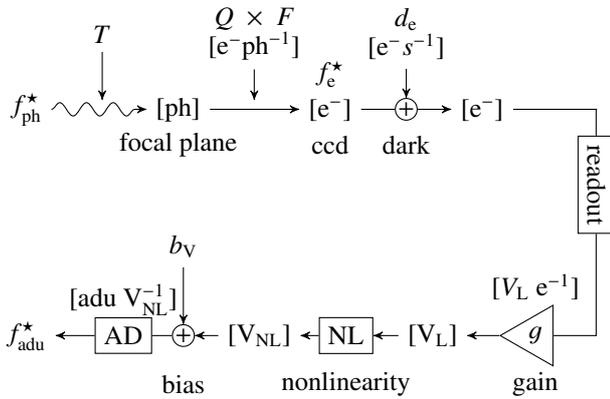
The calibration step transforms the raw images received from the instrument, into photo-electrons calibrated ones. It exploits the knowledge of the instrument derived from its characterization performed either in laboratory or in space during the commissioning phase, to invert the instrument response. Thus, the calibration module removes the bias introduced by the analog chain, restores the analog digital unit (ADU) image back to  $e^-$ , evaluates and corrects the dark current and the pixel response non-uniformity (PRNU or flat-field).

#### 4.1. Instrument model

The data reduction sequence results from the signal transforming steps from incident photons to raw images as illustrated in Fig. 2. The light flux  $f_{\text{ph}}^*$  entering the telescope is guided to the



**Fig. 1.** Data reduction flowchart. Green, orange and blue color are calibration, correction and photometry main modules, respectively.



**Fig. 2.** Signal chain. Following the main arrow,  $f_{\text{ph}}^*$  is the input photon flux. The units of successive transformations are given in brackets: [ph] photons,  $[e^-]$  electrons,  $[V_L]$  and  $[V_{NL}]$  linear and nonlinear volts, and [adu] the analog-to-digital units.  $T$  is the optical throughput,  $Q$  is the quantum efficiency,  $F$  the flat field,  $d_e$  is the dark current. The readout label is the frame transfer, the triangle represents the analog amplifier with its gain  $g$ , its nonlinearity NL and its bias voltage  $b_V$ . AD is the analog to digital converter. The output is the raw image  $f_{\text{adu}}^*$ .

focal plane through the optics with an optical throughput  $T$  that depends on the wavelength and incidence angle. The photons create opto-electrons in the CCD with a quantum efficiency rate  $Q$  depending on the wavelength. Different response from pixel to pixel creates a pattern that translates into PRNU, which is evaluated as a function of wavelength in laboratory. At the end of the exposure time, the frame is transferred in 25 ms to the CCD storage zone protected from light. A dark current leakage of  $\sim 0.05 e^- \text{pix}^{-1} \text{s}^{-1}$  adds to the photo-electrons during both exposure and readout process. The pixels are then serialized and their charge converted into voltage by the analog amplifier with

gain  $g$ , deviation from linearity NL, and an added polarization bias voltage  $b_V$  to prevent feeding the digital converter with possibly slightly negative voltage. The serialization lasts from  $\sim 1$  s to 4.63 s depending on the chosen reading mode. The result is the raw image in ADU received from the instrument.

The overall transformation from star's photons to ADU is:

$$f_{\text{adu}}^* = \text{AD}(\text{NL}[g \cdot (f_{\text{ph}}^* \cdot T \cdot Q \cdot F + d_e)] + b_V), \quad (1)$$

following the labeling of the different transformations presented in Fig. 2. The flux received on the focal plane can be retrieved by inverting Eq. (1):

$$f_e^* = \left[ L_e \left( \frac{f_{\text{adu}}^* - b_{\text{adu}}}{g} \right) - d_e \right] \cdot \frac{1}{F}, \quad (2)$$

where  $f_e^* = f_{\text{ph}}^* \cdot T \cdot Q_{e \text{ ph}^{-1}}$  is the input flux in units of electrons, the function  $L_e = [\text{AD}(\text{NL})]^{-1}$  is the inverse of the volts nonlinearity after digitization. This function is derived from laboratory measurements (Sect. 4.4). The digitized bias voltage is  $b_{\text{adu}}$  measured as explained in Sect. 4.2. The measurement of dark electrons  $d_e$  is described in Sect. 4.5.

The organization of the data reduction pipeline presented in Fig. 1 is derived directly from the signal restoration in Eq. (2). The first step in the calibration module is the event flagging, which is a general function of the pipeline responsible to flag images previous to any processing of data. This function is described in Sect. 7.1.

#### 4.2. Bias and readout noise

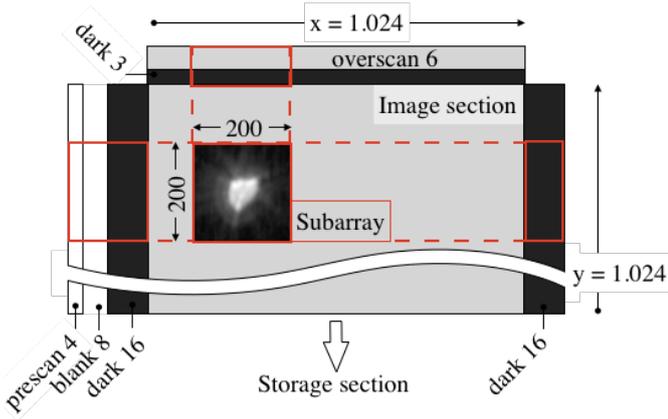
The bias voltage is a voltage added to avoid negative values due to readout noise in case of faint flux. For CHEOPS's CCD the bias voltage is regulated around  $b_{\text{adu}} \sim 609$  ADU per readout with a 10 ppm stability. The expected readout noise (ron) is  $\sim 3.5$  ADU per readout. Since the reference voltage used to generate the bias voltage can vary slightly with temperature, the bias is monitored and corrected using prescan pixels.

Prescan are virtual, empty pixels that contain neither photon nor dark current electrons and they are digitized before any real pixel. For prescan pixels, Eq. (1) simplifies to  $b_{\text{adu}} = \text{AD}(b_V)$ . The CCD pixels map is presented in Fig. 3, where the columns and rows correspond to the  $x$ - and  $y$ -axis, respectively. Prescans, take the form of 4 extra columns on this map.

To save bandwidth, only the median  $\bar{p}$  and standard deviation  $\sigma_p$  of the on board stacked prescan pixels are transmitted to ground. The pipeline then normalizes to a single readout value to estimate the bias  $\hat{b}_{\text{adu}} = \bar{p}/n$  and readout noise  $\hat{\sigma}_o = \sigma_p/\sqrt{n}$  of a single image, where  $n$  is the number of stacked exposures.

In practice, using such bias estimate would cause a significant increase of the white noise in the light curve on the order of  $(n_{\text{ap}}/\sqrt{n_{\text{presc}}}) \times \text{ron} \sim 100 \times \text{ron}$ , with  $n_{\text{ap}}$  being on the order of 3000 pixels in the aperture (assuming a 30 pixels radius) and  $n_{\text{presc}}$  the 800 prescan pixels.

To overcome this, the bias correction is separated into two steps: (1) a constant component is accurately estimated over the whole visit which can therefore be subtracted from images without creating noise. This constant component,  $\langle \hat{b}_{\text{adu}} \rangle$ , is the average of  $\hat{b}_{\text{adu}}$  over the visit; and (2) the time varying component is then corrected later by the general background correction (Sect. 5.3) that works on an image per image basis. This component is assumed to be small thanks to the high thermal stability of the instrument.



**Fig. 3.** Schematic view of the photo-sensitive area of the CCD. The  $200 \times 200$  square inside the  $1024 \times 1024$  full CCD is the region of interest, called subarray, transmitted to ground. Margins in the left: 4 prescan columns, 8 blank columns (unused), and  $16 + 16$  dark columns. Margins in the top: 6 overscan rows, 3 dark rows. The bottom storage section, not represented, mirrors the CCD, including margins. The arrows on the top and right of the diagram represent the  $x$ - and  $y$ -axis of the pixels of the full CCD, respectively.

Additionally, the bias difference between pixels is compensated by using a fixed bias frame recorded using null time exposures during ground calibration and updated regularly in flight. The pixel couplings that would result as image structures have actually been found to be negligible. The overall bias correction then resumes to:

$$I_b = I_r - n \cdot \langle \hat{b}_{\text{adu}} \rangle - n \cdot B_b, \quad (3)$$

with  $I_r$  the raw image,  $I_b$  the bias corrected image,  $B_b$  the zero-average bias frame and  $n$  the number of stacked images.

#### 4.3. Gain

The analog amplifier converts the individual charges into a low impedance voltage that feeds the AD converter. The amplifier response has three main characteristics: its offset or bias (Sect. 4.2), its slope and its nonlinearity (Sect. 4.4). The slope is the gain of the digital conversion process, given in units of  $\text{ADU } e^{-1}$ . The gain is influenced by several reference voltages and the temperature of the front end electronics. They are measured and provided to ground as numerical values in the housekeeping data, associated with each exposure.

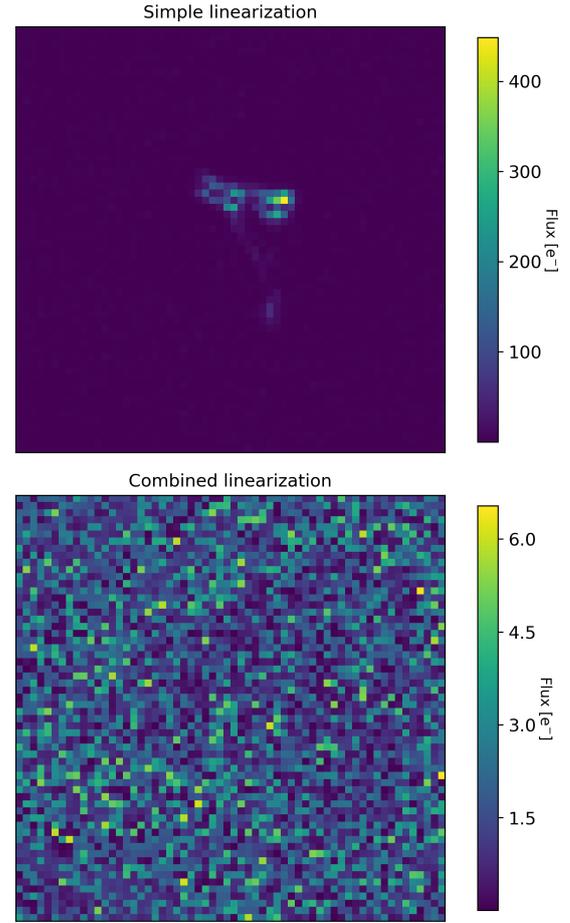
The laboratory characterization provides a model of gain  $g$  which depends of the input voltages and temperatures. This model is applied in the pipeline to correct each exposure:

$$I_g = I_b / g(T, V); \quad (4)$$

where  $I_b$  is the input bias-corrected image,  $I_g$  the output gain-corrected image,  $T$  and  $V$  the housekeeping temperatures and voltages. After correction, the resulting image is in units of photoelectrons, and can be used directly to determine the shot noise. Typical measured values of the gain in the nominal setup are around  $0.5 \text{ ADU } e^{-1}$ .

#### 4.4. Linearization

The classical linearization is the straightforward application of a correction law determined from laboratory measurements of a constant light beam through a series of increasing exposure



**Fig. 4.** Linearization residuals of a  $60 \times 60$  pix image of a  $V\text{-mag} = 9$  target built from 6 stacked readouts. *Top*: from direct application of the linearization correction to the stacked image. *Bottom*: obtained using the combined algorithm.

times. But such an approach does not work well on stacked images because the correction law itself is not linear. It should therefore be applied to individual readouts prior to stacking. Since the individual readouts are not always downloaded, the pipeline takes advantage of the imagerettes to mitigate this limitation. The linearization of imagerettes is combined with the linearization of stacked images when necessary. Because the position of the imagerette may change at each readout to follow the target's motion, some pixels are not present in all imagerettes of a given stacked image. Therefore, the algorithm completes the missing information by properly weighting pixels taken from the stacked image. Figure 4 shows the efficiency of this technique to restore linearity on the illustrative case of a  $V\text{-mag} = 9$  star whose images are built from 6 stacked readouts. The combined linearized stacked image in the bottom panel shows no imprint of the PSF compared to the classical linearization shown in the top panel where residuals of PSF are clearly visible. It is worth noticing the difference of intensity scale between the two panels. In conclusion, the linearization is applied to the gain-corrected image  $I_g$  to obtain the linearized image  $I_L$ . This step involves no change of units.

As an alternative to the onground processing, onboard correction of the nonlinearity could be performed. This will be investigated during the instrument commissioning, through a series of dedicated tests and, depending on the results, the pipeline could be updated accordingly.

#### 4.5. Dark current

The dark current is accumulated by a given pixel of an image from the beginning of the exposure to its readout. The dark current  $d_e$  is monitored using dedicated blind pixels on either side of the CCD: the 32 dark columns which are not exposed to light, see Fig. 3.

In the default configuration, the readout process starts by quickly transferring the full frame image, including margins, into the blind storage zone in  $t_z = 25$  ms. The image is then shifted down line by line during a total time of about 4.63 s into the serialization register where each pixel is in turn shifted into the digitizing electronics in a time  $t_x$  that depends on its  $x$ -position or column number.

Consequently, each pixel has a different dark current time accumulation, depending on its  $(x, y)$  position on the CCD. This is described by time map  $M(x, y)$  of the same dimensions as the CCD:

$$M(x, y) = n \cdot (t_{\text{exp}} + t_z + y \cdot t_y + x \cdot t_x), \quad (5)$$

where  $n$  is the number of stacked images,  $t_{\text{exp}}$  is the integration time of an individual image,  $t_y$  the line shift time and  $t_x$  the column shift time. The dark current estimation is a robust linear regression between dark pixels values and their lifetimes in the time map  $M$ . The typical dark current is  $\sim 0.05 \text{ e}^- \text{ s}^{-1}$ , resulting in  $\sim 1800 \text{ e}^-$  in a typical photometric aperture during one minute exposure.

To save telemetry, the dark margins are line by line averaged onboard into a single column of 200 mean dark pixels, that is the same  $y$ -axis size of the subarray. The time map  $M$  is averaged accordingly before the regression. Median and standard deviation are also provided as robust backup and controls in case a cosmic ray would hit the dark columns.

Similarly to the case of the bias correction (Sect. 4.2), the dark current correction is separated into a constant and a variable terms to optimize the signal to noise ratio (S/N) of the correction. The constant term is accurately determined by averaging the dark current over the full observation run. The correction of the variable component is left to the background correction step described in Sect. 5.3.

Finally, the dark current difference between two pixels is corrected by applying a fixed dark frame,  $D$ . The latter is derived from laboratory measurements. It has to be properly scaled in order to match the actual in-orbit conditions. The dark frame will be updated during the commissioning with exposures with the cover closed. The complete correction of the dark current constant term is then given by:

$$I_d = I_L - d \cdot M \cdot D \frac{d}{d_{\text{lab}}}, \quad (6)$$

where  $I_d$  is the dark-corrected image,  $I_L$  is the image after linearization in units of electrons,  $d$  is the constant in-orbit dark current and  $d_{\text{lab}}$  the constant dark determined in laboratory conditions.

#### 4.6. Flat field

CHEOPS uses a chromatic flat field correction to take into account the PRNU. The dependency of the flat field with the wavelength has been carefully assessed in the laboratory, resulting in a large set of monochromatic images (Deline et al. 2020). Depending on the wavelength, these measurements show noticeable structures: surface gradients for long wavelengths and strays for short wavelengths as can be seen in Fig. 5.

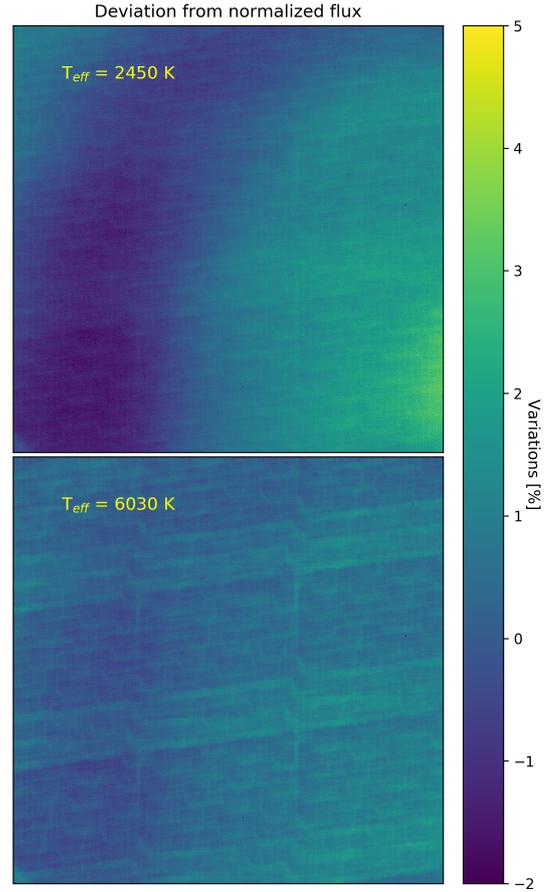


Fig. 5. Examples of flat field images. These FF images were derived from monochromatic images corresponding to the spectral energy distribution of a  $T_{\text{eff}} = 2450 \text{ K}$  (top) and a  $T_{\text{eff}} = 6030 \text{ K}$  star (bottom).

The flat field used for correction is a linear combination of several monochromatic measurements, weighted according to the effective temperature  $T_{\text{eff}}$  of the target. The determination of the input  $T_{\text{eff}}$  of the target is responsibility of the scientific user or observer of the visit. The DRP will use this value automatically as an input. A set of mean normalized  $T_{\text{eff}}$  indexed flat fields spaced out by  $\sim 150 \text{ K}$  is available for the correction. Thus, the pipeline uses the flat field image that better matches the target's temperature to perform the correction.

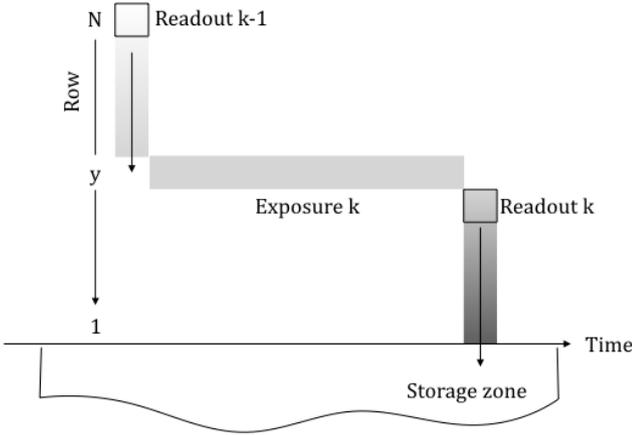
The flat field correction is the last stage of the calibration module of the DRP. The calibrated image is in units of photoelectron and passed to the correction stage which is described in Sect. 5.

## 5. Correction

The correction step aims at correcting individual calibrated frames from environmental effects such as smearing trails, bad pixels, and background and stray light pollution, as detailed in the following subsections. The pixel-to-sky step presented as the first box in the correction module in Fig. 1 is a general purpose function described in Sect. 7.

### 5.1. Smear correction

Because there is no shutter, the pixels remain exposed during the readout process. Therefore during the 25 ms of the frame transfer, each charge well collects light from each pixel crossed on its



**Fig. 6.** Illustration of the process of charge transfer.

way to the storage area. As a result, vertical trails do appear on the image (top panel of Fig. 7). The trails are generated by all stars on the CCD even when located outside the subarray image.

Figure 6 illustrates this effect in the case where the individual exposures are not stacked. At the end of exposure  $k - 1$  an empty charge well is created on the top of the CCD and will reach its integration position  $y$  after crossing the upper pixels  $N$  down to  $y + 1$  and collecting a fraction of their flux. When the exposure  $k$  begins, this pixel does thus already contain part of its future smear. At readout  $k$  it sweeps down through the rest of the CCD, but across a slightly different image because of the motion of the field.

As a result, the smear flux  $f_k(y)$  collected in pixel  $y$  of the image  $k$  is:

$$f_k(y) = \sum_{i=y+1}^N s_{k-1}(i) + \sum_{i=1}^{y-1} s_k(i), \quad (7)$$

where  $s_k(i)$  is the flux collected when the charge well passes under photo-site  $i$  during readout  $k$ . The first and second terms correspond to the contribution of the column above and below the pixel respectively. The smear problem is thus to estimate the contributions of the various photo-sites crossed by a given pixel of the image.

The basic approach would be to derive the contributions  $s(i)$  from the image itself as proposed by Powell et al. (1999) who subtracts the summed column of the image properly scaled by the transfer time, or Iglesias et al. (2015) which adapts that principle to varying illumination during the exposure. But, this approach does not apply to CHEOPS since only a part of the CCD is downloaded and because the image continuously varies over time: between two consecutive 1 min exposures the image is rotated by  $3.6^\circ$  and undergoes a different pointing jitter.

A set of overscan pixels is available to estimate the smear. Overscans take form of 6 rows of virtual pixels on the top of the image (see Fig. 3). An overscan is not a silicon pixel but an extra clocking at readout time that generates an empty well which immediately crosses the whole CCD following the image. Therefore, the overscans only contain the smear flux.

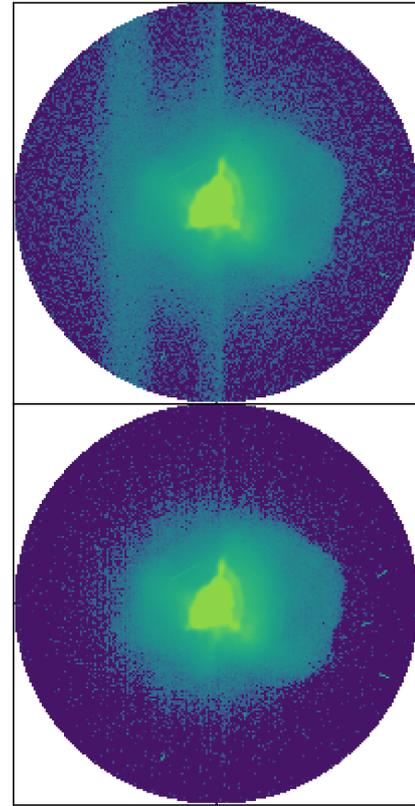
The contribution  $s_k(i)$  defined at Eq. (7) can be estimated from overscans by:

$$s_k(i) = \omega_k, \quad (8)$$

where  $\omega_k$  is  $1/N^{\text{th}}$  of the average overscan row at exposure  $k$ . The smear Eq. (7) then resumes to:

$$f_k(y) = (N - y)\omega_{k-1} + (y - 1)\omega_k. \quad (9)$$

Image of overscan smear correction



**Fig. 7.** Example of smear correction. *Top*: simulated  $200 \times 200$  exposure of a  $V\text{-mag} = 9$  target and one external contaminant. *Bottom*: same exposure after smear correction. The color scale has been adapted for better visualization.

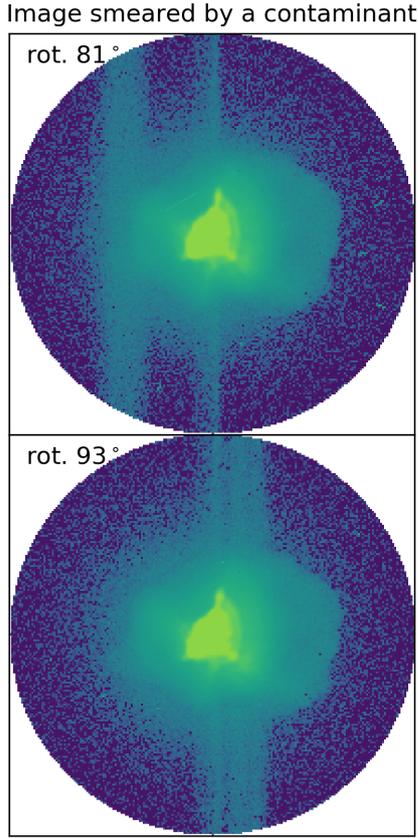
The correction consists on subtracting the estimated smear flux from the image. Figure 7 shows an isolated star of  $V\text{-mag} = 9$  before and after correction. The bright target generates large smear trails, visible in the top panel, due to an important number of stacked readouts. In the image in the bottom, the correction has been applied (e.g., Jenkins et al. 2010a; Rauer et al. 2014).

Although the correction looks fine in the images, it causes a significant increase of noise in the light curve. Due to briefness of the 25 ms transfer time, only few electrons are collected in the overscans and thus cause an important shot noise. That noise is amplified by the large area of the photometric aperture similarly to the bias correction (Sect. 4.2).

The smear estimate  $\omega_k$  from the overscans assumes an uniform column or trail. This assumption holds when the observed image is static or when the smear flux is dominated by the target as it was the case in others missions of photometric observations of transits such as CoRoT and Kepler.

Figure 8 shows that this is not necessarily true for CHEOPS. Indeed, the trails of the others stars rotating around the target can overlay the target, even when located far outside of the downloaded region of the CCD. Additionally a star which is present at exposure  $k - 1$  below a given pixel will leave a trace in the corresponding overscan  $\omega_{k-1}$  used for correction, but might have rotated away at exposure  $k$  and thus never been crossed by the pixel.

The solution comes from Fig. 9 which shows the light curve of the isolated smear pattern obtained by simulation. The peaks of the curve originate from the crossing trail of the an external star of Fig. 8 rotating outside the subarray.



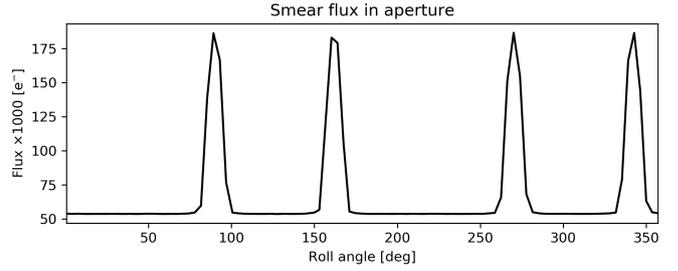
**Fig. 8.** Examples of smear trails in the subarray. The two images correspond to a different roll angles of one contaminant star of  $V\text{-mag} = 7$  rotating around a  $V\text{-mag} = 9$  target outside the subarray. Each exposure is labeled with its respective roll angle and a logarithmic scale was used for better visualization.

It is important to note that the flux outside the peaks, which originate only from the trail of the target itself (hereafter called self-smear) is nearly constant (e.g., the flat bottom of the curve in Fig. 9). That comes from the fact that the photometric aperture follows the motion of the target and consequently it sees a static pattern. Therefore it is not necessary to correct the self-smear which is left as is to avoid introducing noise.

Only the peaks above a certain threshold are corrected, using simulated images (Sect. 6.2) to determine the concerned exposures. The threshold is chosen to ensure that the noise introduced by the overscan based correction will be smaller than the disturbing smear signal.

## 5.2. Bad pixels

The bad pixels module detects and corrects for cosmic ray hits during the observation as well as for pixels with temporary or permanent abnormal response. Currently, three types of bad pixels are considered by the DRP. First type corresponds to cosmic rays: when high energy particles impacts the CCD they cause positive outliers in a pixel during a single exposure. These cosmic rays (CR) can affect one or several connected pixels, as well as dark and overscan CCD margins. The CR hits occur mainly during the SAA crossings that are not down-linked, but also spuriously outside the SAA. Second type of bad pixels are the hot and dead pixels that are permanent damaged pixels that suffer abnormally high or low flux response, respectively. Finally, the pipeline also detects the random telegraphic pixels, which are



**Fig. 9.** Example of the smear flux in the aperture from the target and external stars as a function of the roll angle.

unstable pixels whose state randomly flips between a normal behavior and an arbitrary high response, or just are affected by a high level of noise. Caused by irrecoverable radiation damages, the total number of telegraphic pixels is expected to increase during mission lifetime.

The CCD will be regularly monitored and an updated list of bad pixels will be issued and serve as an entry for DRP. The pipeline then notifies after each observation its own detections from the signal in the subarray window. The location of the subarray will be chosen to avoid hot and dead pixels in the aperture.

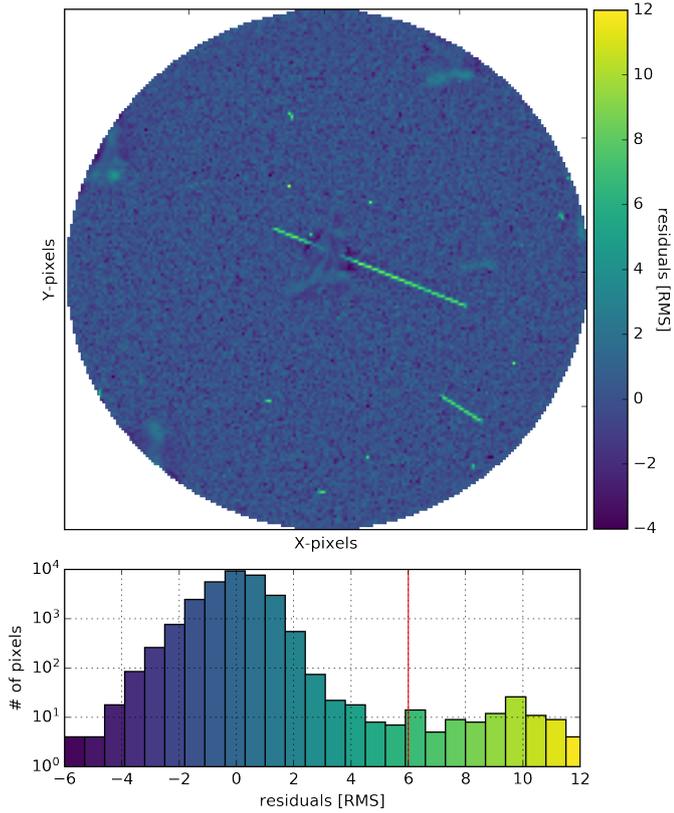
Simple approaches like sigma clipping to search for outliers in pixel flux time series is not relevant for CHEOPS because of the specific features of its data such as: (i) the noise is not stationary due the permanent rotation of the image and (ii) at pixel level the noise is dominated by the jitter noise, especially for the peaks of the PSF near the center of the target. To reduce the temporal variability of individual pixels due to the jitter, the bad pixels detection module begins by re-centering images and imagettes, shifting them in the opposite direction of the depointing. Then, to remove flux variations caused by the rotation of the images and the target's intrinsic variability affecting close pixels in the same way, the detection of bad pixels operates on the residuals. A residual  $r$  is the relative variation of a pixel compared to its neighbors:

$$r = \log \left( \frac{f}{f * k} \right), \quad (10)$$

where  $f$  is the image (resp. imagette) and  $k$  an unitary smoothing kernel of size  $10 \times 10$  (resp.  $6 \times 6$ ) pixels. The sign  $*$  accounts for convolution function. The advantage of using residuals is that each type of bad pixel has an specific footprint in these  $r$  images.

Special attention is brought to CR which are difficult to detect when embedded into the target main flux. For this reason, the detection is also performed in the imagettes where the same CR flux has a better contrast with respect to the reduced unstacked target flux. Both detections maps are then merged, taking into account the fact that the target's position follows the depointing.

As temporal outliers, CR are detected by sigma-clipping the residuals. The adopted threshold is adjusted to represent the best compromise between the number of detections and the number of false positives, avoiding the correction of false events. The thresholds are derived from a set of simulations over a wide range of target brightness,  $6 < V\text{-mag} < 12$ , and exposure times (see Sect. 8). Figure 10 shows an example of a residuals image used for the CR detection. One long trail of a cosmic ray crossing the target's PSF is detected in the upper image, except for the pixels inside the peaks of the PSF. The CR energy is indeed not large enough to stand out among the PSF pixels flux. The  $6\text{-}\sigma$  detection threshold is represented by the vertical line in the



**Fig. 10.** Example of cosmic rays detection. *Top*: example of a residuals image used for cosmic rays detection. *Bottom*: residuals distribution of the image. The vertical line represents the detection threshold.

lower histogram of residuals. All pixels above this threshold are flagged as CR. The evolution of the light curve of two pixels through the CR detection module is shown in Fig. 11.

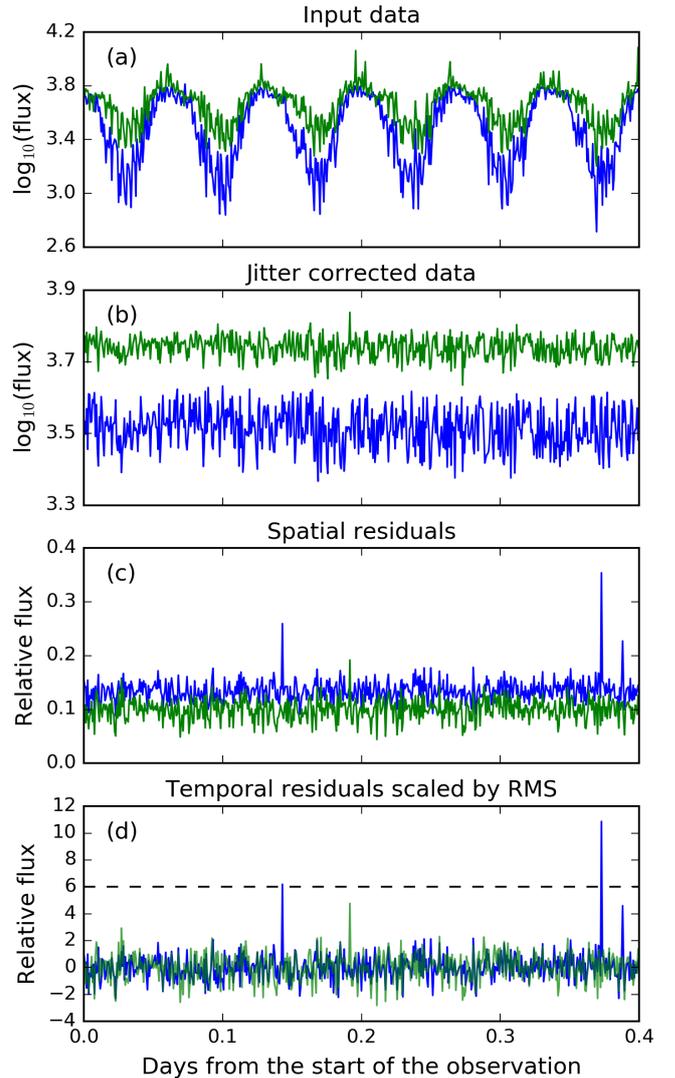
Once the bad pixel detection is performed, the DRP proceeds to the correction of the pixels hit by CR. This correction is done with a 2D cubic interpolation of neighbor pixels using the Python routine `interpolate.griddata` of the Scipy library.

Hot (dead) pixels are positive (negative) spatial outliers imprinted, in this case, on the temporal average of the residual images. They are detected via a spatial sigma clipping. A threshold as high as  $30\text{-}\sigma$  is necessary to avoid flagging pixels influenced by the peaks of the target’s PSF. No centering is applied before hot, dead and telegraphic pixels detection. An example of a temporally averaged residuals is shown in Fig. 12 for a target of  $V\text{-mag} = 6$ . The signatures of the PSF are clearly visible at the center of the image. Hot pixels appears as strong positive values on this map while dead pixels as negative ones.

Finally, the telegraphic pixels are detected as noisy pixels in the map of residuals variation over time. On the contrary, residuals of an ordinary pixel shows small variations over time. Nevertheless, a detection threshold of  $7\text{-}\sigma$  is used to avoid false detections in the PSF peaks. An example of a noise map is shown in Fig. 13. The effect of the jitter on the target is evident at the center. Finally, the Bad Pixel module outputs the corrected image cube and the 2D map of the bad pixel location.

### 5.3. Background

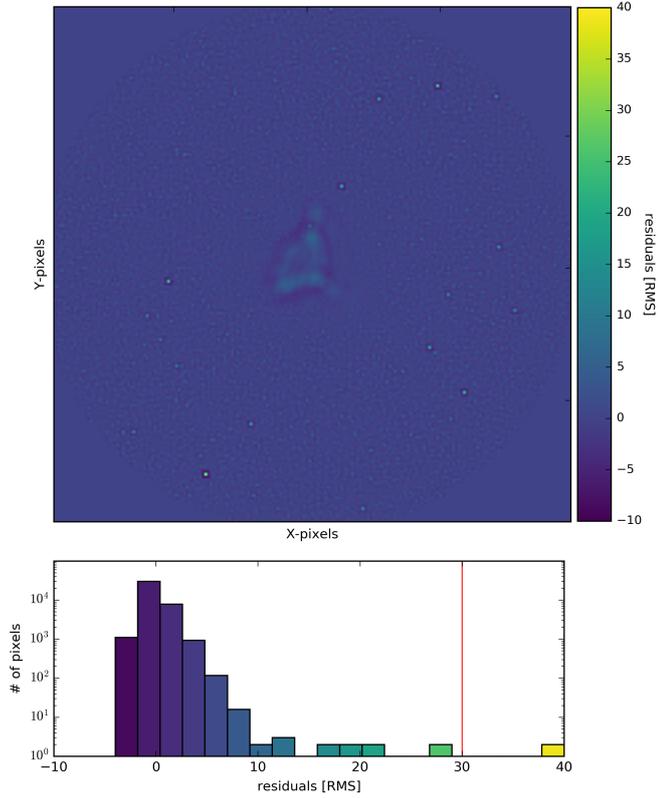
The zodiacal light, non resolved background objects, stray light from Earth and Moon inject a non constant flux offset over the CCD. This background flux depends primarily on the orbital



**Fig. 11.** Evolution of the light curves of two pixels close to target along the CR process. *Panel a*: initial flux. *Panel b*: flux after re-centering by the opposite de-pointing direction. *Panel c*: spatial residuals. *Panel d*: residuals normalized unitary RMS. The horizontal dashed line represents the current  $6\sigma$  threshold used to flag the cosmic ray hits.

phase and on the pointing of the telescope. In particular for CHEOPS, the background correction module plays an important role because of the satellite’s proximity to the Earth. The classical approach of background estimation based on selected background windows gives poor results for CHEOPS since the displacement of the stars due to the rotation of the image obliges to move them continuously, probing thus not all the time the same pixels and flux distribution. These changes translate into discontinuities in the estimated background time series which is ultimately imprinted in the light curve of the target by the correction.

Instead, DRP uses a histogram based method which is insensitive to the rotation of the field and maximizes the total background sampled flux. For this purpose, a large circular mask that excludes the central target is applied to each frame. This mask follows the de-pointing so that the probed region is always the same. An histogram is drawn from all pixels included in the background mask. The upper bound of the histogram is restricted to the admissible background level in order to exclude contaminating stars as well as the tails of the target’s PSF. Then, the mode of a fit skewed Gaussian is taken as the background value



**Fig. 12.** Example of hot pixels detection. *Top*: temporal mean of residuals normalized by the spatial median absolute variation (MAD) as an  $200 \times 200$  pix image. *Bottom*: histogram of the distribution of the residuals mean values where the vertical line marks the detection threshold of hot pixels.

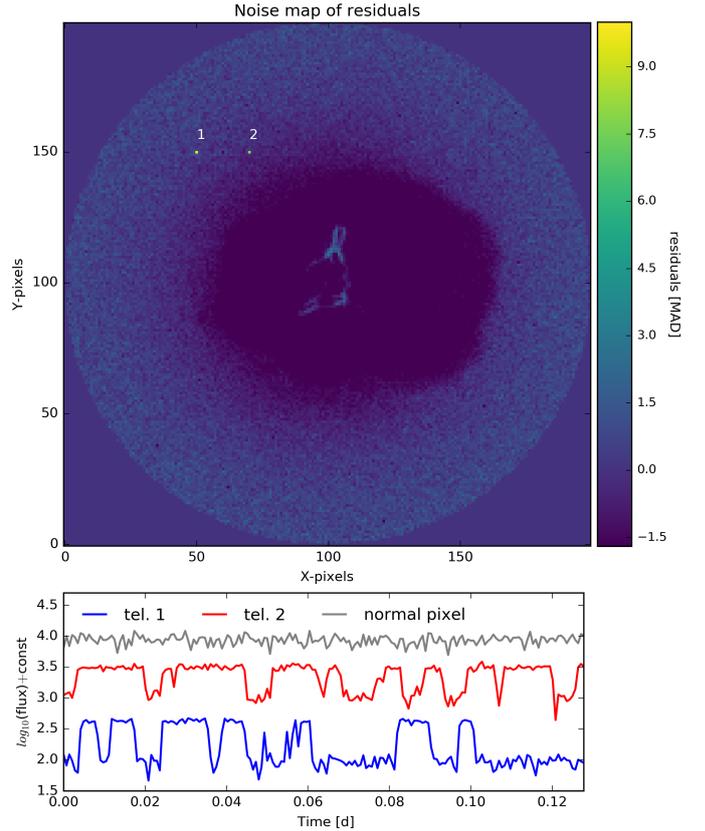
and subtracted from the image. Figure 14 shows an example of the masked background region, its respective histogram and the resulting background estimation. The background time series of a typical observation with a few faint background stars is shown in Fig. 15. The figure shows a clear correlation between the roll angle variation, and therefore the stray light level, and the background flux. For each visit, the background time series is delivered by the DRP and the corrected images are used as the starting point of the photometry extraction.

## 6. Photometry

After the data has been fully calibrated and corrected, the DRP performs an aperture photometry to deliver the final light curve. The aperture is a circular binary mask that follows the target’s displacements. The circular shape respects the intrinsic symmetry of the rotating experiment. To avoid sharp edges like a binary step-like contour, the border is weighted in relation with the pixels fraction covered by the mask.

To avoid area changes when the border shifts from a subpixel quantity, for a particular radius, only one disk template is computed using a null depointing and then applied to all depointings of the whole time series by using an antialiased shifting algorithm that strictly preserves the mask surface. Apertures of non-constant area would introduce artificial photometric noise in the light curves.

In fact, DRP provides four light curves each measured through a different aperture. The three first radii are pre-defined: 26, 33 (default aperture) and 39 pix, while the fourth radius is optimized for each visit (optimal aperture).



**Fig. 13.** Example of telegraphic pixels detection. *Top*: temporal noise of residuals normalized by the spatial MAD. *Bottom*: Light curves of the two flagged telegraphic pixels (green and blue lines) and one normal pixel (gray line).

The default aperture (33 pix) encompasses 97.5% of the PSF flux. The two other pre-defined apertures are lower (80%) and upper bounds (120%) of the default radius and used as controls. Figure 16 shows the flux of the PSF encompassed by each of the predetermined apertures.

The light curve  $f$  is simply the sum of the pixels inside the aperture, and weighted by the mask  $m$  depointed by  $(\delta x, \delta y)$ :

$$f = \sum_{\text{pix}} m(\delta x, \delta y) \cdot p, \quad (11)$$

with  $p$  being the concerned pixel.

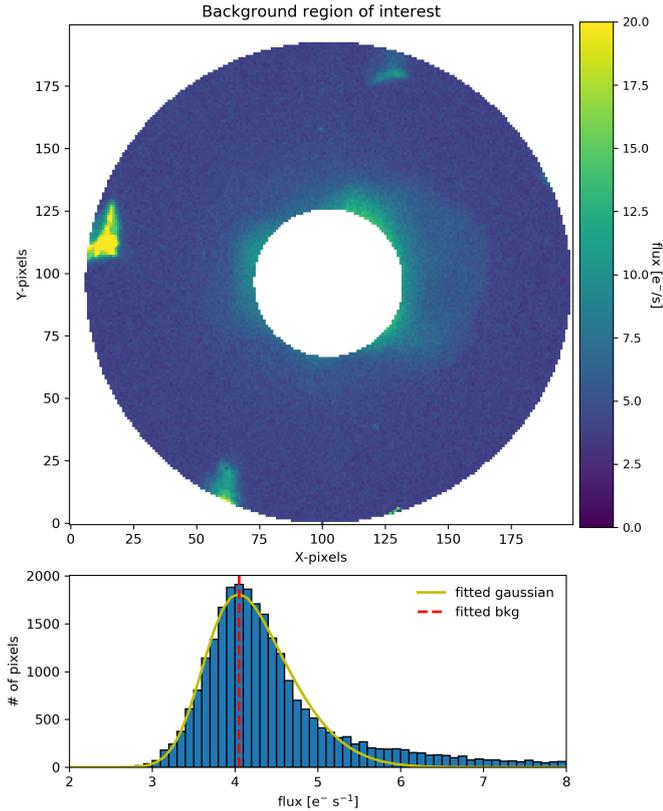
### 6.1. Optimal aperture

The optimal aperture is optimized for the visit. For instance, bright targets deserve a larger mask as their flux dominates further out from the center over the background and the readout noise. On the other hand dense star fields require a smaller aperture to better exclude contaminating stars.

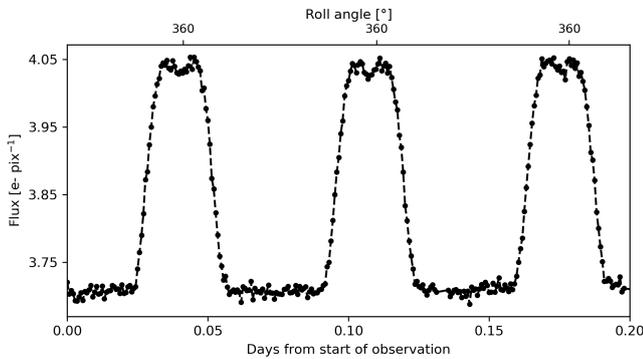
The optimal aperture corresponds to the radius that minimizes the noise to signal ratio:

$$\text{NSR} = \frac{\sqrt{f + c + \sigma_c^2 + \sigma_{\text{ron}}^2}}{f}, \quad (12)$$

where the numerator lists all the considered noise sources. The components  $f$  and  $c$  respectively accounts for the target and contamination shot noise inside the tested aperture. They are computed from image simulations, see Sect.6.2 for details. The noise

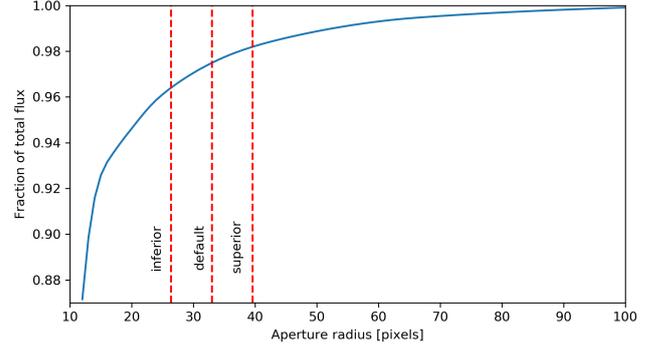


**Fig. 14.** Background estimation from one image. *Top*: masked region excluding the target. *Bottom*: histogram of the pixels in the background region after clipping extreme values shown together with the fit Gaussian function (orange) and the adopted background level (red dashed vertical line).

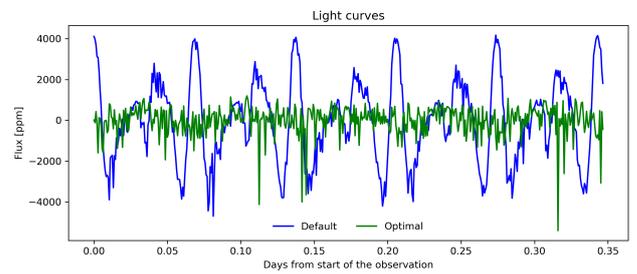


**Fig. 15.** Example of a background curve of a 5 h observation under typical observing conditions. The correlation with the roll angle (top axis) is evident.

$\sigma_c$  is the contamination variation caused by the ingress-egress of the contaminants, whose irregular PSF enters and exits the aperture mask along the rotation. The noise  $\sigma_{ron}^2 = n_{pix} \cdot n_{stack} \cdot ron^2$  is the readout noise estimated in Sect. 4.2, and transformed to electrons using the gain, of the  $n_{pix}$  pixels of the mask for an image composed of  $n_{stack}$  stacked readouts. Figure 17 shows the photometric improvement when using the optimal aperture if a  $V$ -mag = 9 contaminant is distant from the  $V$ -mag = 6 target by only 30 pixels. The default light curve is clearly degraded by the variable overlapping of the contaminant shown in Fig. 18. These flux variations are no longer present when applying the optimal circular mask which in this case was set by the optimization method with a radius of 15 pixels.



**Fig. 16.** Photometric growth curve of CHEOPS PSF. The vertical dashed lines represent the radii of the three pre-defined apertures used by the DRP.



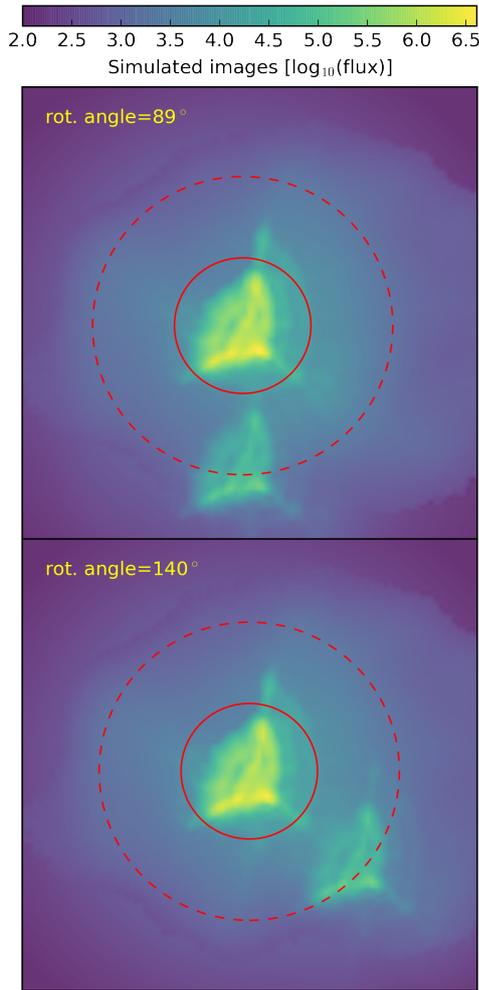
**Fig. 17.** Light curves of the default radius aperture (blue) and optimal aperture (green) for a  $V$ -mag = 6 target with a  $V$ -mag = 9 background star located at  $\sim 30$  pix distance.

Finally, besides the four light curves, the pipeline delivers as products complementary correction values that could help the user to perform a deeper analysis of the data. Among these products are the dark current, background and contamination light curves.

## 6.2. Image simulations

The pipeline builds up simulated images of the whole visit because it needs to estimate smear trails (Sect. 5.1) and contamination from the resolved nearby stars. The DRP internal simulator starts by making use of the World Coordinate System (WCS) of each exposure (see Sect. 7.3) jointly with the sky coordinates and the CHEOPS magnitudes of the stars extracted from an input catalog. This catalog is built for each observation by extracting from the *Gaia* DR2 catalog (Evans et al. 2018) the sky coordinates, the CHEOPS magnitude (obtained from the  $V$ - or  $G$ -band conversion) and the  $T_{eff}$  of each star in the field of view, and it is provided to the DRP as an input file associated to each observation. The internal simulator then uses this information to spread a reference PSF over the CCD coordinates of the stars with the flux scaled according to their CHEOPS magnitude, resulting in the expected simulated data set. The reference PSF comes from laboratory measurements during pre-launch instrument characterization. It will be later on replaced by the flight PSF derived from commissioning phase.

A double simulation is first built up: one with only the target in the field and the other with the resolved contaminants only. This pair is used to estimate the effect of the contaminant stars in the photometry (Fig. 19) and to compute the optimal aperture for the photometry (Sect. 6). In the figure the red circle represents the location of the photometric aperture that is used to compute the values  $f$ ,  $c$  and  $\sigma_c$  in Eq. (12).



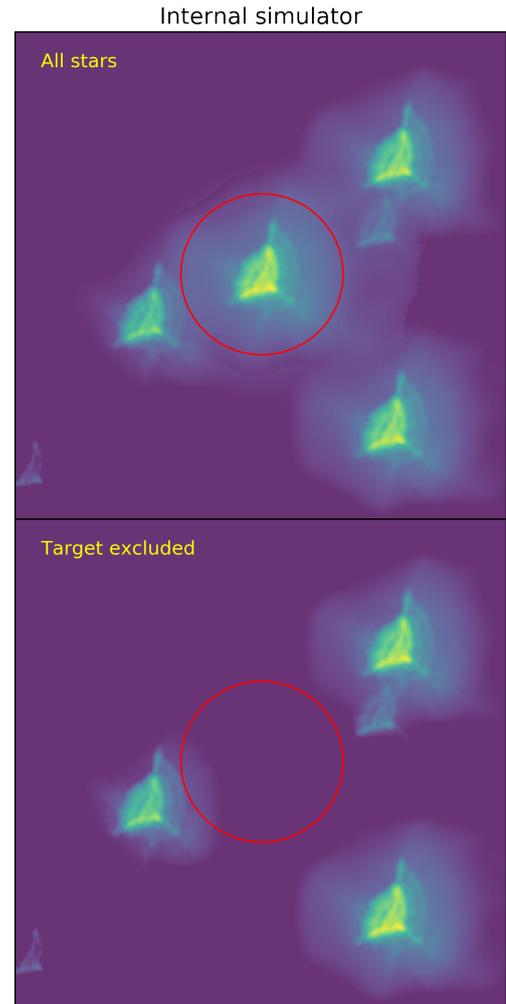
**Fig. 18.** Examples of two simulated images of an observed field composed of one  $V$ -mag = 6 target star and one  $V$ -mag = 9 background star located at  $\sim 30$  pix distance. The optimal and default aperture for photometry are represented by the solid and dashed red circles, respectively. Each image is labeled with their respective roll angle.

A second simulation over the whole CCD height which includes the non downloaded portion above the image is necessary to model the smear trails since any contaminant crossing that region let a trace in the smear trails. As the smear trails extend along the  $y$ -axis, the computation of their simulation is optimized by collapsing both PSF and star coordinates in the single spatial dimension of the  $y$ -axis. Here, the possible change of position during the exposure, which will produce a small dilution of the signal on the  $x$ -axis is not taken into account at the moment.

## 7. General purpose modules

### 7.1. Events flagging

Due to the low orbit, a significant fraction of measurements are lost due to the proximity of the Earth to the line of sight and the crossing of the SAA. The event-flagging module is in charge of identifying and flagging the exposures which are affected by a high stray light level or a high rate of cosmic ray impacts and when housekeeping temperatures are too high. The minimal angle values for a valid exposure are  $120^\circ$  for the Sun and  $5^\circ$  for



**Fig. 19.** Simulated images of the FoV for contamination estimation. Top: simulation including the target and all background stars. bottom: and all the stars but the target. The red circle represents the photometric aperture.

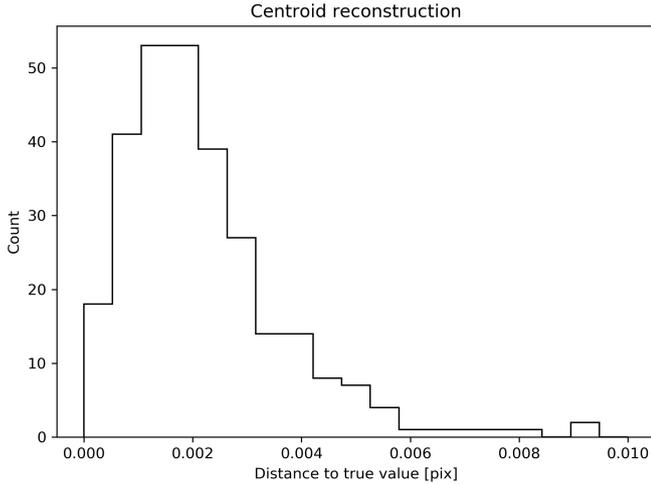
the Moon. There is also a provided stray light estimate used to flag high stray light levels in the images.

The ratio of bad exposures can be as high as 10 min per orbit on average for SAA and 40 min per orbit for the Earth occultation when the instrument line of sight is out of the ecliptic plane. Both types are not necessarily phased one over the other so they can overlap or happen at different time, lowering the duty cycle down to 50% in the worst case. There could also be situations where there is only one or two valid exposures between two consecutive gaps that must be dealt with.

Finally this module also verifies housekeeping temperature and checks for values that might lie outside predetermined bounds and could be responsible for bad measurements. The DRP takes this information into account and flags each exposure accordingly.

### 7.2. Centroids

An accurate reconstruction of the depointing is needed for some correction steps and for the photometry. The expected stability of the platform is about  $\pm 2$  pixels on the long term. For each image, the onboard estimate of the depointing is the starting point of centroid determination.



**Fig. 20.** Distribution of the distance between DRP computed target's centroids to the true values used in the simulations.

The centroid computation is applied to the image corrected from bad pixels and smearing. We use an iterative Gaussian apodization method (as in [Deline et al. 2020](#)). The algorithm starts by applying a Gaussian apodization on the target in order to reduce the influence of neighboring stars, image corners and pixels entering and exiting the image with jitter. The Gaussian mask parameters are  $\sigma = 10$  pix in relation with the PSF size, and the initial centering is the depointing provided by the onboard software. Then, the center of light is computed on the weighted image, resulting in a refined measurement. The mask is then re-centered accordingly and the process iterates until a convergence criterion is met. The convergence is usually obtained within 20 iterations.

The centroid estimation error is as low as  $2 \times 10^{-3}$  pix. Figure 20 presents the distance from the estimate to the true depointing introduced in the simulations. The centroid of an image may differ by a constant between the methods of DRP, on board software and reference PSF centering. To overcome this point, centroids around the center of the subarray are converted into depointings around (0, 0) by subtracting their own average. Thus, the result is consequently method independent.

### 7.3. Pixel and sky coordinates relationship

In order to pass from pixel coordinates to physical sky coordinates back and forth, the DRP uses the World Coordinate Systems (WCS) library ([Calabretta & Greisen 2002](#)). This library is commonly used in the astrophysics community as designed to easily store the sky coordinates in the data. Since the pipeline deals with several different images over the visit the WCS are stored in each individual image. The target has its coordinates RA and DEC defined. The centroid position in pixel coordinates and the rotation angle of each image define the reference point for WCS. The WCS rotation matrix is defined from the rotation angle coming from the raw data in order to be incorporated as WCS information. All necessary WCS keywords are stored in the metadata to be easily used with the WCS library to get the physical sky coordinates.

### 7.4. Report

After each run of the DRP an automatic report of the processed observation is generated in the form of a document provided to

the CHEOPS end user. This report is a digest of plots and metrics that walks-through the gradual evolution of the signal across the successive DRP steps. It is a fast way to identify possible noise sources in the final light curve, or any residual correlation that can exist between the target's flux and main observational parameters such as depointing, roll angle, etc. Usual metrics are point-to-point RMS, measurements of the scatter in some portions of the light curve, and a modified version of the combined differential photometric precision (CDPP; [Jenkins et al. 2010a](#); [Christiansen et al. 2012](#)) to account for the gaps in the data. It is worth noticing that no filtering nor any detrending algorithm is applied in the metrics themselves in order to preserve the full signal information and, in this way, to accurately see its evolution. Examples of the report can be found in CHEOPS guest observer website at ESA<sup>2</sup>.

## 8. Performance

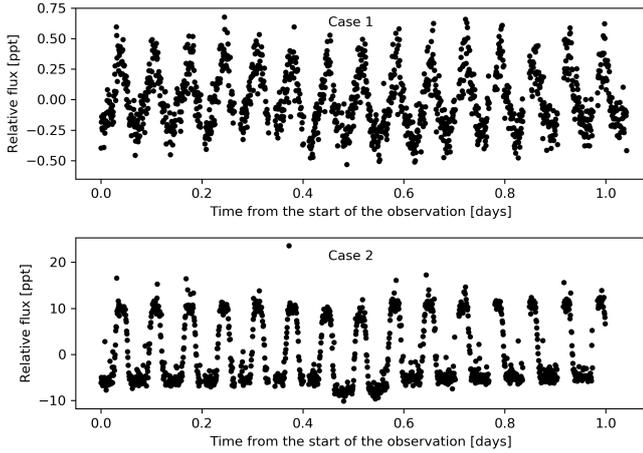
Two datasets were prepared using CheopSim simulator ([Futyán et al. 2020](#)) to illustrate the performance of the DRP and compare it with the CHEOPS science requirements in terms of photometric precision. The first simulation (case 1 hereafter) represents the observation of a transit of an Earth-size planet orbiting a  $V$ -mag = 6 G0V star with a period of 50 days. The second simulation (case 2 hereafter) corresponds to the observation of a faint star ( $V$ -mag = 12) with a transit of a Neptune-size planet in a 13 days orbit. CHEOPS photometric requirements for these science cases are 20 ppm in 6 h of integration time for case 1; and 85 ppm in 3 h of integration time for case 2.

Both simulations were built using an intermediate contamination environment by setting the appropriate options of CheopSim to MEDIUM (namely 1673 background stars in the field and  $\sim 2 e^{-} \text{pix}^{-1} \text{s}^{-1}$  of stray light) and taking into account the intrinsic stellar noise of its host star. Cosmic rays were randomly injected on both simulations together with 3 hot and 1 telegraphic pixels manually placed on the subarray window to avoid contaminating the target's PSF. The light curves of both simulations are shown in Fig. 21. The planet transits are barely visible in the simulator output light curves slightly processed by removing the simulated bias, dark and gain to convert the units (hereafter raw light curves). The cause is a strong correlation of flux with the position of the target on the CCD (case 1) and the background contamination in the aperture in case 2. The corresponding final DRP default light curves and their 10 min binned version are shown in Fig. 22. In addition, the theoretical light curves containing only photon and stellar noise, that is before the injection of any instrumental or environmental contamination, are also shown.

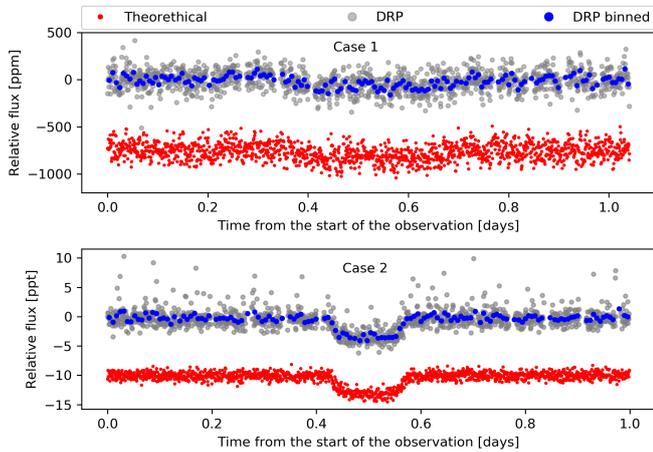
The modified CDPP at different time scales is used to assess the photometric precision of the light curves for each case. It accounts for the gaps in the light curves and is based on the mean of the unbiased variance estimates from a rolling window of a specific time length. Then, the reported CDPP value corresponds to the square root of the mean of the variances normalized by the maximum number of points in the rolling windows. This metric can be interpreted as the noise one would obtain by rebinning the light curve at the selected time scale, time correlated or red noise included. As explained before, no detrending nor filtering is applied within the metrics.

The obtained CDPP is shown at different time scales in Fig. 23. For case 1, the DRP light curve reaches a precision

<sup>2</sup> <https://www.cosmos.esa.int/web/cheops-guest-observers-programme>



**Fig. 21.** Light curves of raw data of case 1 (top) and case 2 (bottom). The raw data has been only corrected for bias, dark and gain for units adaptation.

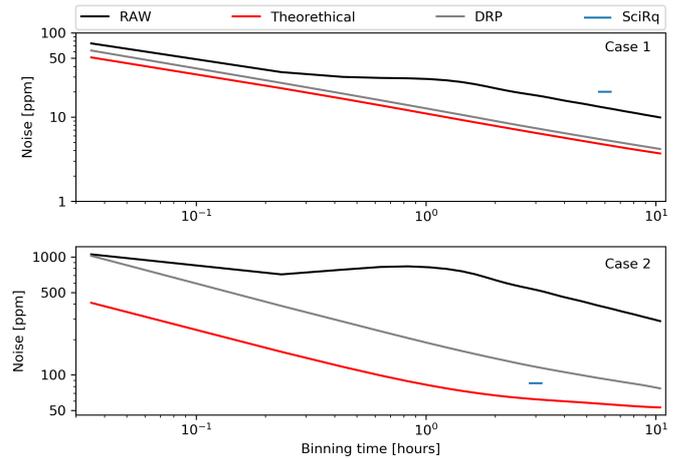


**Fig. 22.** DRP light curves with the default aperture (gray) in case 1 (top) and 2 (bottom). Blue points are the 10 min binned version. Red points are the unbinned theoretical light curve arbitrarily shifted for better visualization.

below 6 ppm in 6 h of integration time with a duty cycle of 90% in the 24 h of the visit.

The noise estimation for case 2 is 117 ppm in 3 h of integration for the same duty cycle. For its part, the optimal aperture, not shown in Fig. 23, delivers a dispersion of 83 ppm while the dispersion of the theoretical light curve is 62 ppm in the same conditions. The case 2 light curve is evidently affected by undetected cosmic rays (Fig. 22). This effect is not surprising since the long 60 s exposure time used in case 2 translates in a larger number of CR per image for an equivalent integrated flux (see for example Futyán et al. 2020). Furthermore, in this observation no imagettes are available to help the cosmic rays detection. Confirming this effect, a control simulation with no cosmic rays injected gives dispersion of only 70 ppm (resp. 67 ppm with the optimal aperture) comparable to the theoretical case of 62 ppm.

Regarding the detection of the hot and telegraphic pixels, the pipeline was able to recover both hot pixels out from a dark current 3 to 5 times larger than the usual one and the inserted telegraphic pixel was also correctly flagged. There were a few false detections, most of them close to the target, but they have no influence on the result since the DRP is not correcting but communicating them for posteriori long term analysis.



**Fig. 23.** Noise estimations for the case 1 and 2 light curves. The plots are the modified CDDP of the raw (black), default DRP (gray) and theoretical light curves. The photometric requirement for each case is represented by the blue dash at 6 and 3 h for case 1 and 2, respectively (see text).

## 9. Conclusions

The CHEOPS data reduction pipeline in its pre-launch version has been presented in this paper with a detailed description of the core processing stages of the calibration, correction and photometry modules. The particularities of CHEOPS data and their treatments by the pipeline have also been discussed. In addition, two representative examples of scientific cases for CHEOPS have been used to evaluate the expected performance of the pipeline. For each case, the achieved photometric precision is given at different time scales. These examples show that the results of the DRP are fully compliant with the scientific requirements of the mission.

Even for challenging observations, such as for faint targets (e.g., case 2 in Sect. 8) the light curve derived by the DRP is not far from ideal results. It was shown that for a  $V$ -mag = 12 target, the 3 h dispersion of the light curve derived with the optimal mask is very close to the noise level of the theoretical photometry: 83 ppm vs. 62 ppm, respectively. In fact, this could even be improved by performing clipping of photometric outliers or flux binning, for example. These treatments are left to the users since best results are usually reached by a case by case detailed analysis which strongly depends on the science goals of the observations. As shown in Sect. 8, the deviations from theoretical expected performance of CHEOPS, is not driven by instrumental effects but by the influence of external agents such as smear trails of background stars, cosmic rays or background contamination. The pipeline has proven that it is able to mitigate successfully these effects on the final photometry even though improvements, in particular, in cosmic ray detections are still under study and will be finally tested when the in-flight PSF is available.

DRP generates various output products the user will retrieve from the archive: four light curves calculated with different aperture sizes, each with its contamination curve and associated uncertainties. The user will also get the report automatically generated by the pipeline which intends to allow the user to follow what treatment has been done on the data, the quality of the processing and of the final result. In addition to these final products, the user will have the possibility to get additional by-products of the processing, such as, for example, bad pixel maps or the background light curve.

After the launch, the pipeline will be tuned and adapted to real in-flight data: algorithms and modules will be improved all along the mission lifetime with our increasing knowledge and understanding of the instrument to allow the best characterization of the transiting planets CHEOPS will observe. The pipeline and its associated reference files are under versioning control and therefore the data can be easily re-processed if it is required by CHEOPS project.

*Acknowledgements.* The DRP team thanks our referee, F. Claus, for his very careful reading of the paper and his valuable comments and suggestions which help us to improve the quality of the manuscript. We also thank D. Futyran for his valuable support on Cheopsim. We gratefully acknowledge the SOC team and the CHEOPS science team members who evaluated the pipeline results, for their constant support in the realization of the CHEOPS pipeline and their helpful suggestions which allowed significant improvements of key reduction steps. We thank K. Isaak for her careful reading and valuable feedback on the manuscript. We thank also A. Cameron and D. Queloz for providing comments that improved the paper. The team at LAM acknowledges CNES funding for the development of the CHEOPS DRP, including grants 124378 for O.D. and 837319 for S.H., and the support of the Direction Technique of INSU with P.G. assignment. S.G.S. also acknowledges support from FCT (FCT – Fundação para a Ciência e a Tecnologia) through Investigador FCT contracts nr.IF/00028/2014/CP1215/CT0002. O.D. is also supported by FCT contract DL 57/2016/CP1364/CT0004. This work was supported by FCT/MCTES through national funds (PIDDAC) by the grant UID/FIS/04434/2019, by FCT funds (PTDC/FIS-AST/28953/2017) and by FEDER – Fundo Europeu de Desenvolvimento Regional through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI-01-0145-FEDER-028953). *Software:* the DRP is developed in Python 3 (Python Software Foundation, <https://www.python.org/>), and makes use of Astropy (Astropy Collaboration 2013; Price-Whelan et al. 2018), Matplotlib (Hunter 2007), Numpy (<https://www.numpy.org/>), Scipy (Jones et al. 2001) among other Python open source libraries. Jupyter notebooks (Kluyver et al. 2016) were also used for the developing and testing of the code.

## References

- Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
- Baglin, A., Chaintreuil, S., Vandermarcq, O., & CoRoT Team 2016, *The CoRoT Legacy Book: The adventure of the ultra high precision photometry from space* (EDP Sciences), 29
- Benz, W., Ehrenreich, D., & Isaak, K. 2018, *CHEOPS: CHaracterizing ExOPlanets Satellite* (Springer), 84
- Broeg, C., Benz, W., Thomas, N., & Cheops Team. 2014, *Contrib. Astron. Observatory Skalnaté Pleso*, 43, 498
- Broeg, C., Benz, W., & Fortier, A. 2018, in *42nd COSPAR Scientific Assembly*, 42, E4.1-5-18
- Calabretta, M. R., & Greisen, E. W. 2002, *A&A*, 395, 1077
- Christiansen, J. L., Jenkins, J. M., Caldwell, D. A., et al. 2012, *PASP*, 124, 1279
- Deline, A., Queloz, D., Chazelas, B., et al. 2020, *A&A*, 635, A22 (Paper I)
- Deru, A., Chaintreuil, S., Baudin, F., Ferrigno, A., & Baglin, A. 2015, *Eur. Phys. J. Web Conf.*, 101, 06022
- Evans, D. W., Riello, M., De Angeli, F., et al. 2018, *A&A*, 616, A4
- Fortier, A., Beck, T., Benz, W., et al. 2015, in *Pathways Towards Habitable Planets*, 76
- Futyran, D., Fortier, A., Beck, M., et al. 2020, *A&A*, 635, A23 (Paper II)
- Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
- Iglesias, F. A., Feller, A., & Nagaraju, K. 2015, *Appl. Opt.*, 54, 5970
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010a, *ApJ*, 713, L120
- Jenkins, J. M., Caldwell, D. A., Chandrasekaran, H., et al. 2010b, *ApJ*, 713, L87
- Jones, E., Oliphant, T., Peterson, P., et al. 2001, *SciPy: Open Source Scientific Tools for Python*, [Online, <http://www.scipy.org/>]
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, eds. F. Loizides & B. Schmidt (IOS Press), 87
- Pinheiro da Silva, L., Rolland, G., Lapeyrere, V., & Auvergne, M. 2008, *MNRAS*, 384, 1337
- Powell, K., Chana, D., Fish, D., & Thompson, C. 1999, *Appl. Opt.*, 38, 1343
- Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, 156, 123
- Rauer, H., Catala, C., Aerts, C., et al. 2014, *Exp. Astron.*, 38, 249