OPEN ACCESS

CrossMark

A New Signal Estimator from the NIR Detectors of the Euclid Mission

Bogna Kubik[1], Remi Barbier[1], Eric Chabanat[1], Arnaud Chapon[2,3], Jean-Claude Clemens[2], Anne Ealet[2], Sylvain Ferriol[1],
William Gillard[2], Aurelia Secroun[2], Benoit Serra[2], Gerard Smadja[1], and André Tilquin[2]

[1] University of Lyon, UCB Lyon 1/CNRS/IN2P3, IPN Lyon, Lyon, France; bkubik@ipnl.in2p3.fr
[2] Centre de Physique des Particules de Marseille, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
[3] Now at Conseil et Etude en Radio-Protection, Cherbourg-en-Cotentin, France

Abstract

We describe how the Euclid detectors in the Near Infrared Spectrometer and Photometer (NISP) channel will be read out on board and present an analytic expression for the estimated fluence in each pixel with the associated quality factor of the fit per pixel. The method accounts for the Poisson like distribution of the data and includes the effects of noise correlations that arise after the coadding procedure of frames read non-destructively up the ramp during one exposure. The bias of the flux estimator presented in this paper is kept lower than 0.3% over a wide rang of scientifically interesting fluxes of Euclid. The associated error is by 6% lower than the commonly used formula derived in Rauscher et al. in the context of an equally weighted least squares fit. Moreover, the quality factor follows the very well known $\chi^2_{th}(x; n)$ distribution and thus provides a well behaved statistical tool to check the goodness of the ramp fit. The method is proposed in the context of a large amount of data per exposure, produced by the NISP detectors, that cannot be transferred to the ground for the subsequent processing. The method, which is validated using real and simulated test data, can be safely used by most near-infrared instruments which require very accurate measurements to be performed on board.

1. Introduction

In this work we describe how we plan to extract the signal from the near-infrared (NIR) detectors on board Euclid. A high accuracy of the flux estimate on board from a multiple sampling of a long (few hundreds of seconds) exposure in the NIR channel of Euclid, a rigorous control of the error and fit quality as well as the calibration with sub-percent precision are fundamental to ensure the scientific goals of the mission. For this purpose we introduce in this paper a new statistical estimator of the fluence and an associated quality factor of the fit to the data produced by the NIR detectors in the Euclid focal plane.

A common technique used in the low-noise science applications, and which will be applied to the Euclid NIR detectors, to reduce the effective readout noise of the pixels is to read out NIR arrays in the multiple accumulated sampling (MACC) mode. The accumulating signal is sampled up the ramp (UTR) as a function of time and the multiple reads are averaged within groups.

Euclid's telemetry limitations do not allow for a transfer to the ground of all the averaged groups for the subsequent processing. Only the image composed of the fitted slopes of the pixels for each exposure is sent to the ground. The flux integrated by the focal plane, sampled over more than 60×10^6 pixels, must be fitted in orbit in an analytic way with an algorithm subject to the CPU limitations. One should also be able, using the transferred data, to detect any abnormal behavior of the pixels. It is therefore necessary to send an image of quality factors for each exposure, in addition to the image of the fitted fluxes. The former should control the goodness of the fit and indicate the occurrence of any fit inconsistency caused by a cosmic ray hit, nonlinear response of the pixel or any electronic induced instability.

Usually an equally weighted least squares fit is applied to estimate the flux. Although this procedure has an analytic solution, the derivation of the corresponding quality factor requires a second pass over all the groups digitized UTR per pixel, which is impossible within the CPU limitations of Euclid. Moreover, the equally weighted least squares fit neglects the correlations between the digitized frames and treats the samples as Gaussian distributed with constant variance. It was noticed in Kubik et al. (2015a) that the
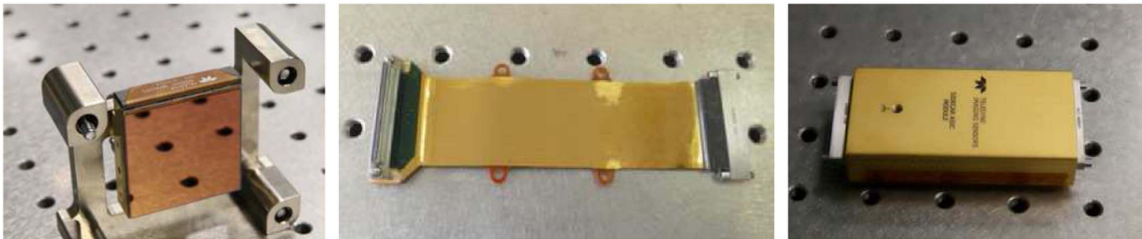
Figure 1. The detector system composed of a sensor chip assembly (left panel), cryo flex cable (middle panel) and sensor chip electronics (right panel). (A color version of this figure is available in the online journal.)

optimal calculation of the flux, in terms of signal to noise ratio, should use all the information contained in the covariance matrix. In this case, however, explicit analytic expressions for the parameters cannot be obtained and the parameters must be computed using numerical methods which are too demanding in terms of accessible memory and computing time.

In order to avoid the numerical computations the typically minimized statistics in the case of Poisson distributed data are the Cash statistics C, defined by Cash (1979). In practice, the absolute value of C statistics cannot be readily interpreted as an indicator of the fit quality (an accurate simulation is needed). It is therefore preferable to use the χ^2 statistics which measures the difference between the observed data and the model. Since for Poisson distributed random variables ΔG_i the expected value is equal to the variance, uncertainties in the data are never accurately known, nor are they usually the same for each data point. Thus, some approximation to the uncertainties in the data is needed. The χ^2 is usually approximated by the data-based summation which in our case takes the form

$$\chi^2 \approx \chi_D^2 = \sum_i \frac{(\Delta G_i - g)^2}{\Delta G_i} \tag{1}$$

or

$$\chi^2 \approx \chi_M^2 = \sum_i \frac{(\Delta G_i - g)^2}{g} \tag{2}$$

where D and M subscripts indicate whether the data or the model are used as weights and g stands for the best estimate of the parent population of ΔG_i. In the literature these two weighting choices are referred to as "Neyman's" and "Pearson's" respectively.

The deviations from Gaussianity invalidate both approximations when the counts per bin fall below ∼10–20. The obvious approximation, that of using each datum ΔG_i as the estimate of its variance is usually biased (Nousek & Shue 1989; Wheaton et al. 1995; Churazov et al. 1996). Wheaton et al. (1995) point that the failure of the χ_D^2 form is more precisely caused by the strong anticorrelations between the data and the weights when ΔG_i rather than $\overline{\Delta G_i}$ is used to approximate the uncertainties in the data ($\overline{\Delta G_i}$ stands for the expectation value of ΔG_i). In short, ΔG_i already contains the Poisson noise and thus wrongly

estimates the variance. A number of other approximations have been proposed to mitigate this effect (Kearns et al. 1995; Wheaton et al. 1995; Churazov et al. 1996; Mighell 1999) but none of them can be directly applied to the nondestructive readouts of the NIR integrating arrays.

As a very stringent error budget is required in the NIR channel of the Euclid mission we introduce in this paper an adapted statistical estimator of the fluence and the associated fit quality factor from Poisson distributed and correlated data. Both, the estimator and the quality factor have an analytic form and thus can be easily implemented in the on-board digital processing unit. The flux bias is kept under control in the range of representative science fluxes and the associated error is by 6% lower than the commonly used formula derived in Rauscher et al. (2007) in the context of an equally least squares fit.

The paper is structured as follows. Section 2 briefly describes the Euclid NIR detectors and the mission requirements. The principle of nondestructive reads in NIR integrating arrays and the origin of the correlations are also explained in this section. The new fluence estimator is introduced in Section 3. The derivation of the theoretical formulas is followed by a Monte Carlo based analysis where the estimator sensitivity to different sources of uncertainties is considered. Finally, in Section 4, we introduce the quality factor and show how it can be used to identify the anomalies occurring during the exposure.

Although our discussion is focused on the Euclid NIR array, we anticipate that much of what is discussed will be of interest to any mission using similar NIR sensors.

2. Detectors for Euclid Near Infrared Spectrometer and Photometer (NISP)

2.1. NISP Detectors Subsystem

The NISP Detector System (NI-DS) hosts the focal plane composed of 16 low-noise HAWAII-2RG (H2RG)[4] detectors

---

[4] HAWAII is an acronym for HgCdTe Astronomical Wide Area Infrared Imager, 2 denotes 2048 × 2048 pixels, R—reference pixels and G—guide window capability.

supplied by Teledyne (Beletic et al. 2008) and selected by NASA.

Each H2RG detector, presented in Figure 1, is equipped with a SIDECAR ASIC[5] sequenced by a dedicated firmware implementing the readout modes specified for the Euclid survey. The H2RG detectors are read using 32 parallel outputs. The acquisition firmware yields a minimum exposure time of 1.31 s with a pixel rate of 100 kHz.

Each H2RG detector will return an array of 2048×2048 pixels coded with 16 bits per pixel. The photosensitive 2040×2040 pixels array is surrounded by 4 rows and columns of reference pixels which are not connected to the detector photodiodes but contain a simple capacitor that electronically mimics the operating pixels. The reference pixels are important to track biases and temperature variations over long exposures (Moseley et al. 2010; Rauscher et al. 2012). The typical single frame read noise of the H2RG detectors is in a range from 10 to 20 e^-.
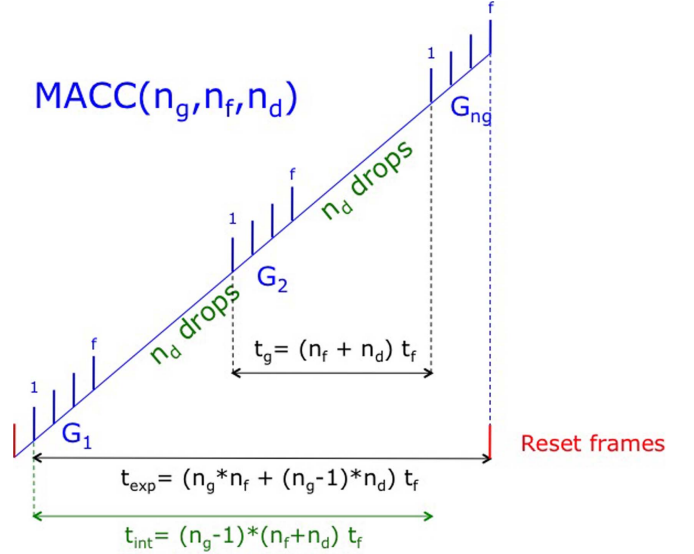
Similar detectors will be used by the future NASA mission *JWST* (NIRCAm, NIRSpec, NIRISS) Gardner et al. (2009). Previous generations of this type of detectors are in use in the *Hubble Space Telescope* Wide Field Camera 3 (HAWAII-1R) (Robberto et al. 2004) and in instruments on ground facilities such as the NIR integral field spectrograph at Gemini North at Manua Kea (HAWAII-2; McGregor et al. 1999) or X-shooter at the European Southern Observatory at Paranal (H2RG) (D'Odorico et al. 2004).

2.2. H2RG Multi-accumulated Readout Principle

For science observations NIR detectors, in general, acquire UTR sampled data at a constant frame cadence. A frame S_i is the unit of data that results from sequentially clocking through and reading out a rectangular area of pixels during a time t_f. As in the UTR mode individual frames suffer from a high readout noise, it is common to average the frames within groups G_k. This readout pattern is called a multiple accumulated sampling and frequently a common abbreviation MACC(n_g, n_f, n_d) is in use where n_g is the number of equally spaced groups sampled UTR, n_f is the number of frames per group and n_d is the number of dropped frames between two successive groups (see Figure 2). Before fitting the flux, the frames within groups are averaged. The signal in group G_k after averaging n_f frames is equal to

$$G_k = \frac{1}{n_f} \sum_{i=1}^{n_f} S_i^{(k)}. \tag{3}$$

The advantage of the coadding procedure is the reduction of the Gaussian distributed pixel readout noise σ_R, assumed here to be uncorrelated from frame to frame, down to the effective group

---

[5] SIDECAR—System for Image Digitization, Enhancement, Control And Retrieval, ASIC—Application Specific Integrated Circuit (™ Teledyne Imaging Sensors).

Figure 2. Multiple accumulated sampling MACC(n_g, n_f, n_d) with n_g—number of equally spaced groups sampled up the ramp, n_f—number of frames per group and n_d—number of dropped frames between two successive groups. (A color version of this figure is available in the online journal.)

read noise $\sigma_{\rm RG}$ defined by

$$\sigma_{\rm RG}^2 = \frac{\sigma_R^2}{n_f}. \tag{4}$$

Therefore, the larger the number of frames per group, the lower the readout noise associated to each group after coadding.

2.3. Importance of Working in the Differential Signal Mode

The correlations between data points are intrinsic to the non-destructive readouts of the integrating arrays. The covariance matrix C between points G_k and G_l was derived in Kubik et al. (2015a). The diagonal entries of C are not only sums of the readout noise and the shot noise, which would be the case for an UTR sampling, but they include correlations in the shot noise after the coadding procedure. The Poisson noise correlates all the points (G_k, G_l) with $l \neq k$ making all the entries of C different from zero and implying a use of numerical methods to inverse the matrix in the fitting procedure.

The merit of working in the group difference space $\Delta G_k = G_{k+1} - G_k$, instead of individual group values, is that a large part of Poisson noise contributions to the correlations between groups is removed. The covariance matrix D between the differences ΔG_k and ΔG_l was derived in Kubik et al. (2015a). The diagonal elements of D are defined in Equation (14) of Kubik et al. (2015a) which we rewrite below with the

notations adapted for this work

$$D_{kk}(\Delta G_k; g) = \frac{2\sigma_R^2}{n_f} + \frac{(n_d + 1)}{n_f + n_d}\frac{g}{f_e}$$
$$+ \frac{(n_f - 1)(2n_f - 1)}{3n_f(n_f + n_d)}\frac{g}{f_e}. \quad (5)$$

The first term of D_{kk} is the electronic readout noise of ΔG_k. Second term corresponds to the shot noise contribution between the last frame of the group G_k and the first frame of group G_{k+1} while the third term arises from correlations after the coadding procedure. In the above equation g is the signal accumulated between two successive groups. We explicit also the dependence on the conversion gain f_e [e^-/ADU] which converts the arbitrary digital units (ADU), recorded by the readout electronic, into electrons. In this way Equation (5) and the formulas in the following sections apply directly to the raw pixel data not converted into physical units, i.e., the inter-group flux g is measured in ADU per unit time $t_g = (n_f + n_d)t_f$, and the electronic readout noise σ_R is given in ADU. $D_{kk}(\Delta G_k; g)$ is the error on the measurement ΔG_k and depends on the inter-group flux g. We rewrite the D_{kk} in the simple form

$$D_{kk}(\Delta G_k; g) = (1 + \alpha)\frac{g}{f_e} + \frac{2\sigma_R^2}{n_f} \quad (6)$$

where the Poisson noise correlations are encoded in the coefficient α

$$\alpha = \frac{1 - n_f^2}{3n_f(n_f + n_d)}, \quad (7)$$

which depends only on the applied readout mode MACC (n_g, n_f, n_d) common to all the pixels.

The only non-null off-diagonal terms of the covariance matrix D are the terms $(k, k \pm 1)$. They are given by Equation (8) of Kubik et al. (2015a) which in our notations can be written as $D_{k,k\pm1}$. $D_{k,k+1}$ take into account the correlation of the Poisson and readout noise between two *consecutive* group differences (Kubik et al. 2015a).

It can easily be seen that at fluxes higher than $5\,e^-\,s^{-1}$ the off-diagonal terms can be neglected as they represent not more than 4% of the diagonal terms for the typical readout noise value $\sigma_R = 10\,e^-$ in the MACC(15, 16, 13) exposure mode.

The readout noise σ_R is assumed here to be uncorrelated from frame to frame, i.e., we neglect the $1/f$ noise component which correlates all the groups in the exposure. An estimate of the contribution of the $1/f$ noise to the signal variance was derived in Smadja et al. (2010) and Kubik et al. (2015b). It was shown that the $1/f$ noise component depends on the exact timing scheme adopted for the acquisition and can be significant for long spectroscopic exposures. The power spectrum of the noise

peaks at low frequencies, but it seems that the recent versions of the SIDECAR ASIC have a lower level of $1/f$ noise than measured in Rauscher et al. (2012). The estimation of its effects on the fitting procedure is out of the scope of this paper, nevertheless this issue will have to be reconsidered when the performance of the algorithms proposed in this work is evaluated with real data.

3. New Signal Estimator for the Multi-accumulated Sampling

The maximum likelihood is a well known method for estimating parameters in terms of the noisy data that they influence. For example a likelihood for a Poisson-Gaussian mixture has been proposed in Snyder et al. (1993) for image recovery from noisy data acquired with a charge-coupled-device (CCD) camera. As the CCD readout differs from the MACC scheme used for NIR arrays, the likelihood from Snyder et al. (1993) is unsuitable given the correlations arsing from the coadding procedure. Moreover its solution requires the execution of an iterative algorithm.

Here we propose a model that describes the measurements of nondestructive readouts form infrared integrated arrays which has an analytic solution and thus can be easily implemented in the on board electronics. We assume that the detector response is perfectly linear over its entire well depth, is thermally stable, and is not subject to any time dependent effects that vary during integrations or readouts. However, it is well know that this type of detectors have a non liner response mainly at high fluxes. The nonlinear behavior of the pixels can be corrected prior to the signal estimate described below with the scheme presented in Vacca et al. (2004). The nonlinearity correction method that will be applied to both, the flux estimate and to the quality factor, in the case of Euclid mission will be addressed in the subsequent paper (in preparation).

We define the likelihood function $\mathcal{L}$ for the infrared integrating arrays as:

$$\mathcal{L} = \prod_{i=1}^{n_g-1} \frac{1}{\sqrt{2\pi\sigma_{\text{eff}}^2(\Delta G_i, g)}} \exp\left(-\frac{(\Delta G_i - g)^2}{2\sigma_{\text{eff}}^2(\Delta G_i, g)}\right), \quad (8)$$

where the total error squared on the measured differences ΔG_k

$$\sigma_{\text{eff}}^2(\Delta G_k; g) = D_{kk}(\Delta G_k; g) \quad (9)$$

includes the effective electronic readout noise of a pixel σ_{RG}, the flux Poisson noise and the Poisson noise correlations arising during the coadding procedure.

The first and second derivatives of the logarithm of the likelihood

$$l = -2\log\mathcal{L} = \sum_{i=1}^{n_g-1} \frac{(\Delta G_i - g)^2}{(1+\alpha)\frac{g}{f_e} + 2\sigma_{RG}^2}$$
$$+ (n_g - 1)\log\left[(1+\alpha)\frac{g}{f_e} + 2\sigma_{RG}^2\right] + (n_g - 1)\log 2\pi$$
(10)

give the flux and the variance estimators, $\hat{g}$ and $\hat{\sigma}_{\hat{g}}^2$ respectively. The estimated flux value $\hat{g}$ is equal to

$$\hat{g} = \frac{1+\alpha}{2f_e}\left[\sqrt{1 + \frac{4f_e^2\sum_{i=1}^{n_g-1}(\Delta G_i + \beta)^2}{(n_g-1)(1+\alpha)^2}} - 1\right] - \beta \quad (11)$$

and the variance estimator $\hat{\sigma}_{\hat{g}}^2$ is

$$\hat{\sigma}_{\hat{g}}^2 = \frac{2(\hat{g}+\beta)^2}{(n_g-1)\left(1 + \frac{2f_e}{(1+\alpha)}(\hat{g}+\beta)\right)} \quad (12)$$

where we have introduced the coefficient β to simplify the notations:

$$\beta = \frac{2\sigma_R^2 f_e}{n_f(1+\alpha)}. \quad (13)$$

The introduced likelihood function is still an approximation since we assume the Gaussian probability distribution for ΔG_i which are actually Poisson distributed. Nevertheless it provides an unbiased fluence estimator in a large range of scientifically interesting signals for a space missions such as Euclid.

In the next sections we illustrate how accurately the flux and the variance are reconstructed from the estimators defined above using Monte Carlo simulation. To illustrate the impact of the improvements introduced in our method, namely the Poisson noise correlations encoded in the coefficient α and the flux dependent normalization factor in the likelihood function, we will show the systematic effects on the estimators $\hat{g}$ and $\hat{\sigma}_{\hat{g}}$ when these effects are not taken into account. So we will refer to the fully correct estimators as given in Equations (11) and (12) with a label "$\alpha \neq 0$." The label "$\alpha = 0$" will refer to the estimators where the Poisson noise correlations are neglected, namely we explicitly put $\alpha = 0$ in Equations (11) and (12). Finally, the "no likelihood correction" label will correspond to the estimators derived from the likelihood with a g-independent normalization factor. The systematic bias of the estimator derived from Equation (2) is partially due to the assumption that the likelihood normalization factor $1/\sqrt{2\pi\sigma^2}$ does not depend on the flux. This is obviously not true in the case of Poisson distributed random variables ΔG_i for which the uncertainty σ is proportional to the inter-group flux g. We account for this effect explicitly in Equation (8) as σ_{eff} is a function of g. The "no likelihood correction" corresponds to the
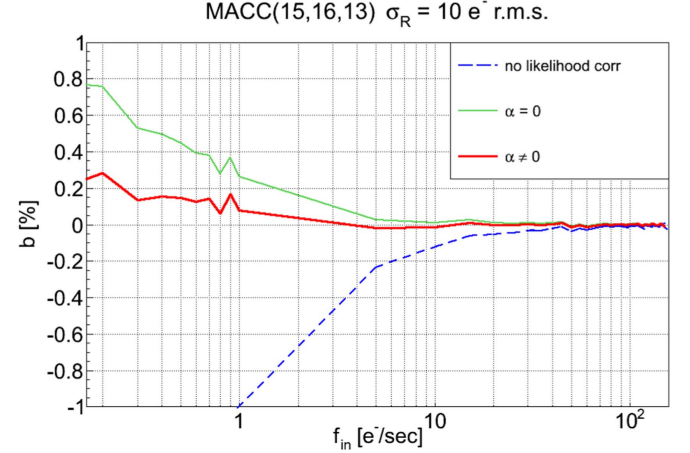
Figure 3. Bias of the flux estimator $\hat{g}$ in MACC(15, 16, 13) readout mode with $\sigma_R = 10\ e^-$ rms. The red bold solid line corresponds the bias of the estimator given in Equation (11). The green solid line corresponds to flux estimator with neglected Poisson noise correlations, namely with $\alpha = 0$ in Equation (11). Blue dashed line depicts the estimator derived from likelihood with constant normalization factor.
(A color version of this figure is available in the online journal.)

omission of the second term of the log-likelihood l before taking the derivatives, thus it is formally equal to the standard approximation of the minimized function given in Equation (2).

3.1. Accuracy of the Flux Estimation

As an example we take the readout mode MACC(15, 16, 13) proposed for the NISP instrument of the Euclid mission, but the results can be easily generalized to any other MACC readout schemes. The pixel readout noise is set to 10 electron rms—the typical value for the NIR detectors used in NISP. We simulate 10,000 nondestructive exposures with the input flux f_{in} ranging from $0.1\ e^-\ s^{-1}$ to $150\ e^-\ s^{-1}$. The flux range spans the expected signal range in Euclid from the dark current ($\sim 0.1\ e^-\ s^{-1}$) in the detector pixel, through the typical values of flat field exposures and faint objects (1–$20\ e^-\ s^{-1}$), up to bright objects (~ 100–$150\ e^-\ s^{-1}$). The signal accumulated between two successive groups g_{in} equals $f_{in}t_g = f_{in}(n_f + n_d)t_f$, where $t_f = 1.3$ s is the single frame read time and we set $f_e = 1\ e^-/\text{ADU}$.

The accuracy of the flux estimator is tested by computing the bias

$$b = \frac{f_{in} - \hat{g}/t_g}{f_{in}}. \quad (14)$$

The results are shown in Figure 3 as the function of the simulation input flux f_{in}. The bias of $\hat{g}$ (red bold line) is less than 0.3% for all the range of fluxes considered in this paper. For comparison, when we neglect the Poisson noise correlations the bias can be as high as 0.8% in the examined range of fluxes. The fluence estimate obtained without the likelihood
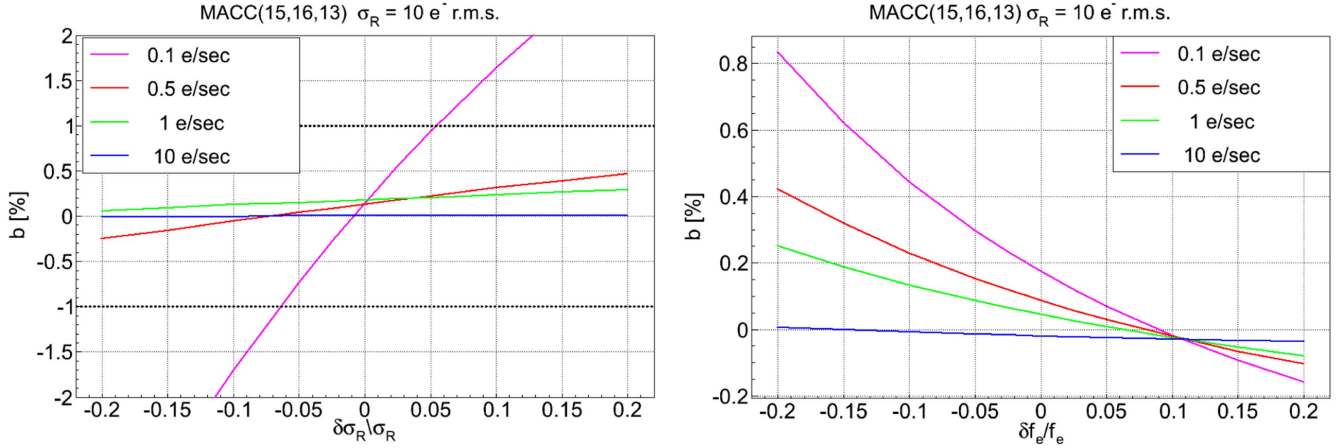
Figure 4. Left: flux bias introduced by a lack of knowledge of the exact readout noise value per pixel. Right: flux bias introduced by a lack of knowledge of the exact conversion gain f_e [e^-/ADU] value per pixel.
(A color version of this figure is available in the online journal.)

correction is overvalued by more than 1% for fluxes lower than about $1\,e^-\,s^{-1}$.

3.1.1. Sensitivity of the Signal Estimation to the Readout Noise and Conversion Gain

The likelihood function depends on two parameters, the readout noise and the conversion gain, which characterize the detectors. They will be measured with a given accuracy. It is important to control the sensitivity of the flux estimator to these parameters. In Figure 4 we show how the lack of knowledge of the exact value of the readout noise σ_R can introduce an offset of the fitted flux. For the ramps simulated at given flux level and with a constant readout noise value $\sigma_R = 10\,e^-$ rms we change the readout noise in Equation (11) by $\pm 20\%$ with respect to the simulation input σ_R. For fluxes of the order of $0.1\,e^-\,s^{-1}$ the bias is around 3% and it decreases with increasing flux down to only 0.1% for $1\,e^-\,s^{-1}$ when a 20% error on σ_R is used. If the average value of σ_R over the detector array is used on board to fit the signal this systematic bias can be removed on ground knowing the exact value of the readout noise per pixel and the bias variation as function of $\delta\sigma_R/\sigma_R$.

As the Poisson noise in the denominators of the likelihood Equation (8) is proportional to the flux, the error in the conversion gain f_e will also induce an error in the variance estimation. We illustrate this in the right panel of Figure 4 where the conversion gain is over- or underestimated by $\pm 20\%$ at most. The flux bias can reach 0.8% for the flux of $f_{in} = 0.1\,e^-\,s^{-1}$ and is reduced down to less than 0.3% for fluxes higher than $1\,e^-\,s^{-1}$.

As seen in Figure 3, for delta $\delta\sigma_R = 0$ and $\delta f_e = 0$, the flux biasis lower is lower than 0.2% for the fluxes between 10 and $0.1\,e^-\,s^{-1}$.

While in general the impact of the conversion gain uncertainty is lower than the bias created by the readout noise

error, both, the electronic readout noise σ_R and conversion gain f_e per pixel have to be measured with a precision better than 5% and 10% respectively if we want to avoid systematic effects higher than 0.5% in the flux estimation at low flux regime.

3.2. Variance of the Flux Estimator

The second derivative of the likelihood function, defined in Equation (12), provides an estimate of the flux variance. In Figure 5 we show the ratios $\mathrm{rms}[\hat{g}]/\hat{\sigma}_{\hat{g}}$ as the function of the simulation input flux f_{in} for three different values of the electronic readout noise $\sigma_R = 1, 5, 10\,e^-$. Solid (dashed) lines correspond to $\alpha \neq 0$ ($\alpha = 0$) in Equation (12). If the noise model is correct then the observed noise computed as the rms of the fitted unbiased slopes is equal to the analytic result and the ratio is equal to one.

According to Figure 5 the model is in a very good agreement with the observed rms of fitted slopes in all the range of the fitted fluxes for $\sigma_R = 1\,e^-$. At higher readout noise the variance is overestimated in the low flux regime. Namely, at low fluxes the readout noise correlations between the consecutive differences of groups become important. This correspond to the second term of the off-diagonal terms of the covariance matrix D, Equation (8), which we have neglected. As the electronic readout noise σ_R decreases, the flux threshold above which the observed rms of slopes and the analytic noise formula are in agreement also decreases. For example, at $\sigma_R = 5\,e^-$, the variance given by Equation (12) is compatible with the dispersion of the fitted slopes for fluxes higher than $1\,e^-\,s^{-1}$. Below that value, the off-diagonal entries $D_{k,\,k\pm1}$ are more than 3% of the diagonal terms D_{kk}. At $\sigma_R = 10\,e^-$, the corresponding flux threshold is $5\,e^-\,s^{-1}$. The likelihood corrective term does not introduce any improvement in the noise estimation in the readout noise dominated regime. Moreover when the Poisson noise correlations are neglected
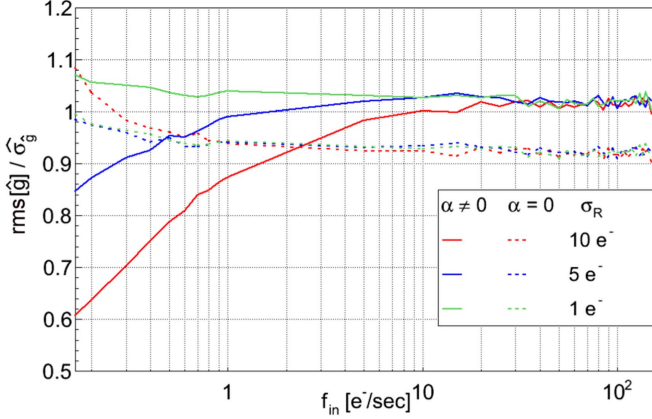
Figure 5. Ratios of the standard deviations of the fitted fluxes rms[$\hat{g}_i$] divided by the noise estimator $\hat{\sigma}_{\hat{g}}$ in MACC(15, 16, 13) readout mode for the simulated exposures with $\sigma_R = 1$ (green), 5 (blue), 10 (red) e^-. Solid lines correspond to $\hat{\sigma}_{\hat{g}}$ given in Equation (12). Dashed lines correspond to $\alpha = 0$ in Equation (12). (A color version of this figure is available in the online journal.)

Figure 6. Ratio of the standard deviation of the estimated flux rms[$\hat{g}$] (blue) and rms[$\hat{g}_{LSF}$] (black) normalized to the square root of the average inter-group flux $\langle g \rangle$ in MACC(15, 16, 13) readout mode with $\sigma_R = 10\ e^-$ rms. In red we plot the ratio $\hat{\sigma}_{RS}/\langle g \rangle$. (A color version of this figure is available in the online journal.)

the noise is systematically overestimated by 10% as compared to the observed rms of the data independently of the flux value.

A general expression for the signal variance $\hat{\sigma}_{RS}$ of an instrument using a multi-accumulated readout was proposed in Rauscher et al. (2007). The variance formula given by Equation (1) in Rauscher et al. (2007), normalized by a factor $(n_g - 1)$, applies to the inter-group signal $\hat{g}_{LSF}$ fitted with a non weighted least squares fit (LSF), a method that neglects any correlations of the sampled points and their Poisson distribution.

In Figure 6 we show the rms[$\hat{g}$] and the rms[$\hat{g}_{LSF}$], obtained from the simulated exposures with $\sigma_R = 10\ e^-$, normalized to the square root of the average inter-group flux
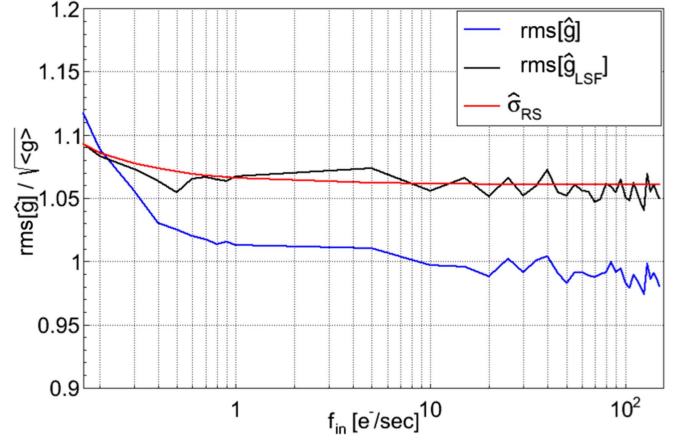
$$\langle g \rangle = \frac{f_{in}\,(n_f + n_d)\,t_f}{n_g - 1} \tag{15}$$

The theoretical prediction $\hat{\sigma}_{RS}$ is also shown in the plot. The difference between the two noise models is negligible for fluxes as low as $0.1\ e^-\ s^{-1}$. Yet, the error on the signal derived form the likelihood function is lower by 6% than $\hat{\sigma}_{RS}$ in the MACC (15, 16, 13) readout scheme for fluxes higher than $5\ e^-\ s^{-1}$. This translates directly into the quality of scientific data and in the increased figure of merit of the survey.

4. Quality Factor to Control the Flux Estimate

4.1. Quality Factor Definition

An important advantage of the proposed method is that it provides a statistical estimator, the "quality factor," which tests the compatibility of the data with the straight line fit. The quality factor verifies whether the data distribution follows the Poisson convoluted with Gauss distribution as it should be the case for the nondestructive readouts not subject to any

anomaly. We show here the properties of this estimator and how it can be used to control the data quality on ground.

The quality factor QF is defined as

$$QF = \sum_{i=1}^{n_g-1} \frac{(\Delta G_i - g)^2}{\sigma_{eff}^2(\Delta G_i, g)}\Big|_{g=\hat{g}_x}. \tag{16}$$

where the pseudo-flux estimator $\hat{g}_x$

$$\hat{g}_x = \sqrt{\sum_{i=1}^{n_g-1} \frac{(\Delta G_i + \beta)^2}{n_g - 1} - \beta} \tag{17}$$

is the value of g in the minimum of QF. Because Equation (16) is a pure product of Gaussian probability density functions in ΔG_i the quality factor should follow the well known $\chi_{th}^2(x; n)$ distribution with n degrees of freedom defined by

$$\chi_{th}^2(x; n) = \frac{(x^2)^{\frac{n}{2}-1}\exp^{-\frac{x^2}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}. \tag{18}$$

The number of degrees of freedom n in our case is equal to $n_g - 2$ and is independent of the flux value, the number of coadded frames n_f, and the number of drops n_d.

The quality factor is related to the pseudo-flux $\hat{g}_x$ through the relation

$$QF = \frac{2f_e}{(1 + \alpha)}[(n_g - 1)\hat{g}_x - (G_n - G_1)] \tag{19}$$

and therefore is related to the unbiased flux estimator $\hat{g}$ which can be expressed in terms of $\hat{g}_x$ as

$$\hat{g} = \frac{1 + \alpha}{2f_e}\left[\sqrt{1 + \frac{4f_e^2\,(\hat{g}_x + \beta)^2}{(1 + \alpha)^2}} - 1\right] - \beta. \tag{20}$$
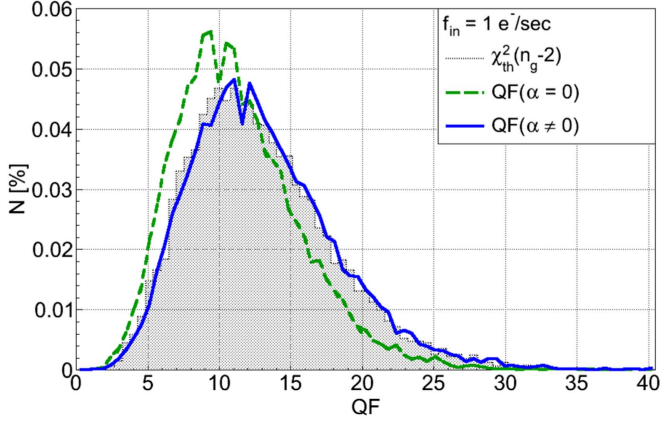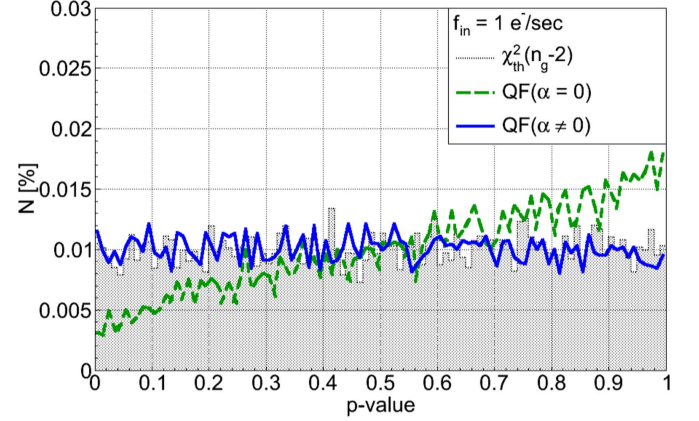
Figure 7. Quality factor QF distribution for the simulation input flux $f_{in} = 1\,e^{-}\,s^{-1}$ and the spectrometric readout scheme MACC(15, 16, 13). The theoretical $\chi^2_{th}(x; 13)$ defined by Equation (18) is plotted as gray shadowed zone.
(A color version of this figure is available in the online journal.)

Figure 8. p-value distribution for the quality factor QF with the simulation input flux $f_{in} = 1\,e^{-}\,s^{-1}$ and the spectrometric readout scheme MACC(15, 16, 13). The p-value of the theoretical $\chi^2_{th}(x; 13)$ defined by Equation (18) is plotted as gray shadowed zone.
(A color version of this figure is available in the online journal.)

In Figure 7 we show the distribution of the quality factors QF obtained in our simulations with the input flux $f_{in} = 1\,e^{-}\,s^{-1}$. Two cases are shown. In blue the quality factor is computed taking into account the Poisson correlations (i.e., $\alpha \neq 0$ in Equation (19)). In green, the Poisson correlations are neglected (i.e., $\alpha = 0$ in Equation (19)). The theoretical distribution $\chi^2_{th}(x; n)$ is plotted as the gray shadowed region. The quality factor with $\alpha \neq 0$ follows closely the expected distribution with the mean value of 12.99 ± 0.05. Furthermore, the flatness of the associated p-value distribution, shown in Figure 8, strongly indicates that the errors are well estimated in the present model, which is not the case if one neglects the Poisson noise correlations, i.e., with $\alpha = 0$ in Equation (19).

In order to explore the properties of the quality factor over the range of fluxes from $0.1\,e^{-}\,s^{-1}$ to $150\,e^{-}\,s^{-1}$ we plot in Figure 9, left panel, the mean value of the QF distribution divided by the mean of $\chi^2_{th}(x; 13)$ in the MACC(15, 16, 13) readout mode. In the right panel the ratio of the rms of the quality factor distribution divided by the rms of the distribution of $\chi^2_{th}(x; 13)$ is shown. Both ratios are expected to be one if the quality factor distribution is consistent with $\chi^2_{th}(x; 13)$.

We confirm thus, that the quality factor QF follows the $\chi^2_{th}(x; n)$ distribution for fluxes exceeding $0.5\,e^{-}\,s^{-1}$. The importance of the Poisson correlations encoded in the factor α is also illustrated in these figures, as for $\alpha = 0$ the properties of the quality factor deviate by 20% from the expected distribution. The same behavior is confirmed in our simulations for different electronic readout noise values σ_R ranging from 1 to $50\,e^{-}$ rms so that we can use the value of QF as a reliable indicator of the goodness of the fit over the large range of fluxes considered in this paper.

4.2. Validation of the QF Distribution with the H2RG Data

We have verified the distribution of the quality factor by using the engineering grade Euclid H2RG $\lambda_c = 2.3\,\mu$m detectors operated in our test facilities. The readout noise per pixel was computed on the dark exposures. The median value over the array, equal to $12\,e^{-}$ rms, was used to fit the flux on exposures with higher incident flux in the MACC(15, 16, 13) exposure. The average conversion gain of $1.36\,e^{-}/$ADU, computed using the Photon Transfer Curve method on flat field exposures, was set for each pixel. The distribution of the fluxes fitted over 2040×2040 pixels and the corresponding quality factors and p-values are shown in Figure 10. The distribution of the quality factors is centered on 13.31 which corresponds to the number of degrees of freedom in the MACC(15, 16, 13) exposures. The distribution of the p-values is flat indicating that the pixel readout errors are well estimated. An example of the exposure of one pixel with a p-value less than 0.001 is also shown. The pixel was subject to a cosmic ray hit during the exposure. This example confirms that pixels that behave badly can be flagged on the basis of the quality factor.

5. Conclusions and Discussion

The control of the systematic effects in the flux measurement, its bias and the error associated to the extracted signal are essential for the Euclid mission.

We have presented in this paper a new estimator of the flux, adapted to the multi-accumulated readout mode of the NIR detectors on board the NISP instrument. The estimator has an analytic form, so that the flux of the observed objects can be fitted in flight.
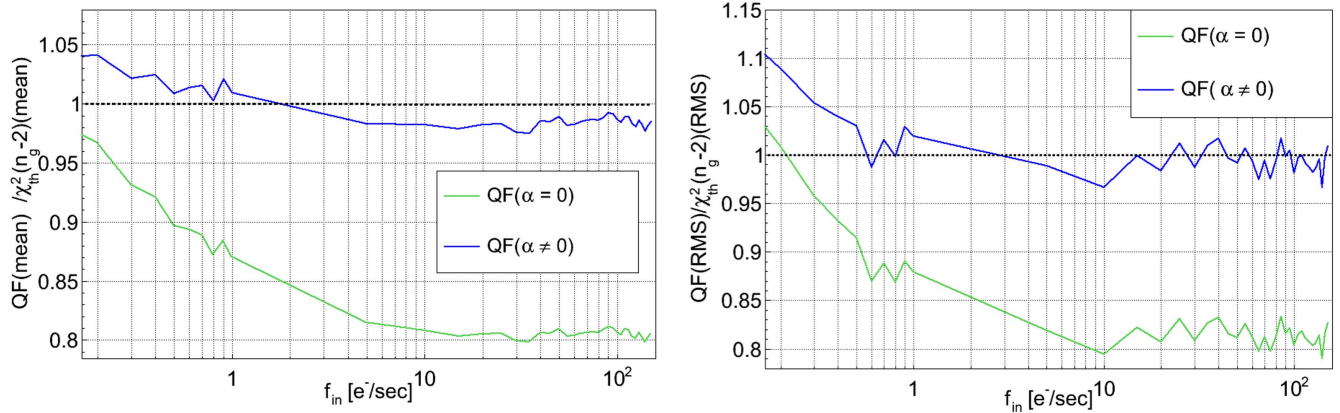
Figure 9. We plot in the left (right) panel the ratio of mean (rms) value of the quality factors QF distribution divided by the mean (rms) of $\chi^2_{th}(x; 13)$ as the function of flux.
(A color version of this figure is available in the online journal.)
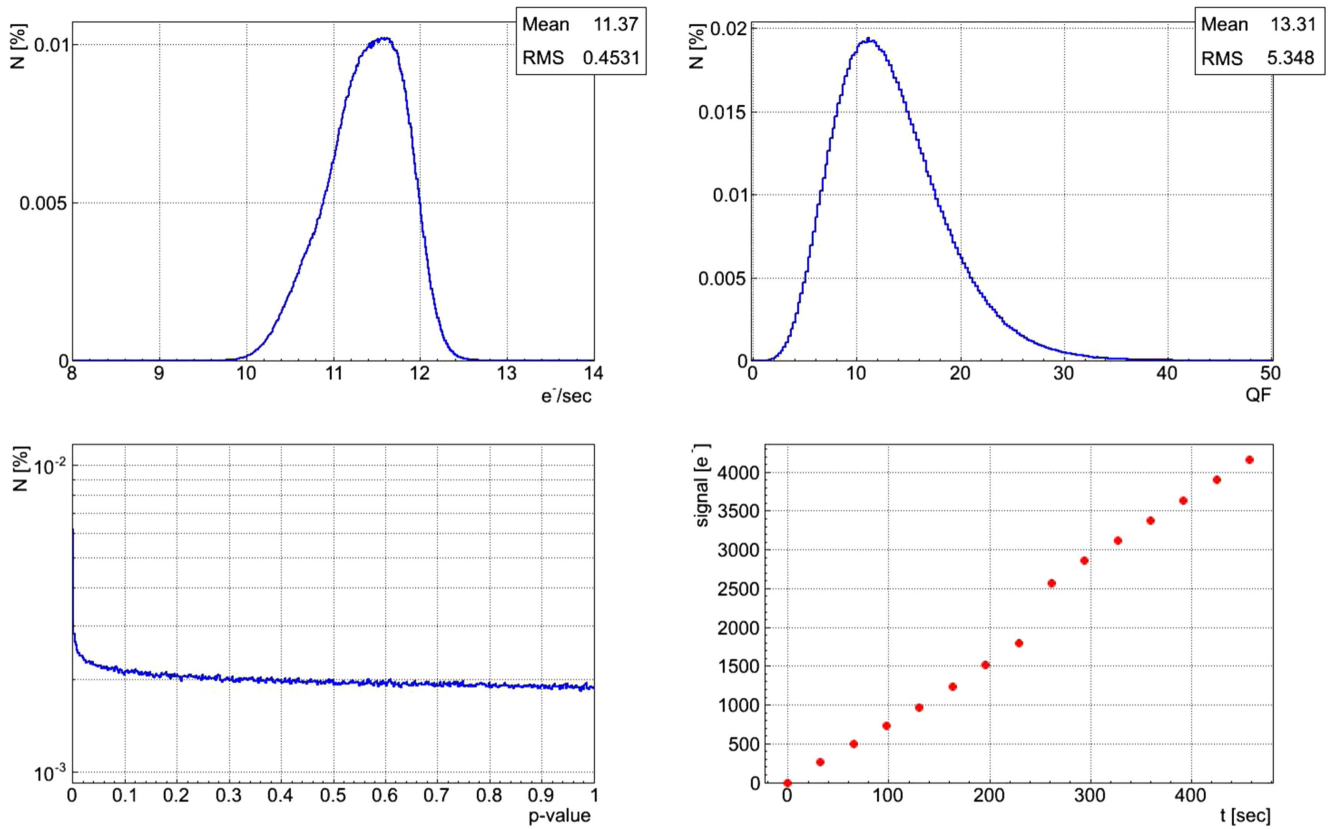
Figure 10. Distribution of the fitted fluxes (upper left), quality factors (upper right) and the corresponding p-values (lower left) on the data taken with a Euclid-like H2RG $\lambda_c = 2.3$ μm detector. An example of the ramp flagged as bad using the criteria of p-value <0.001 is also shown (lower right).
(A color version of this figure is available in the online journal.)

We have demonstrated by Monte Carlo simulations that the bias of the flux estimator can be controlled to less than 0.03% in the range of scientifically interesting fluxes of Euclid. That is assured by taking into account first, the flux dependent error in the likelihood formulation, and second, the correlations intrinsic to the nondestructive readouts of the infrared detectors.

We examined how the accuracy of the estimator is sensitive to the errors in the pixel readout noise and the conversion gain from digital units into electrons. If one wants to keep the

systematic bias lower than 0.5% then for the lowest fluxes the error on the readout noise and conversion gain must not exceed 5% and 10% respectively. If the average readout noise is used to fit the signal on board, then a correction could be applied on ground using the precise map of readout noise values per pixel. The error on the flux estimator introduced in this paper is lower by 6% than the noise derived in Rauscher et al. (2007) for fluxes higher than $5 \, e^- \, \mathrm{s}^{-1}$. This translates directly into a higher figure of merit of the survey.

Finally we have defined the quality factor which follows the $\chi^2_{\mathrm{th}}(x; n)$ distribution and can be therefore used as a reliable check of the consistency of the flux estimator.

In the context of the Euclid collaboration, we propose to compute in flight and to send to the ground the pseudo-flux $\hat{g}_x$, defined in Equation (17), and the quality factor defined in Equation (19). The unbiased flux $\hat{g}$ can be then computed on ground using formula (20).

The simulation used to validate our algorithms does not include non-Gaussian tails in the readout noise. The comparison with the data suggests that the performance should not be drastically lowered.

We have assumed that the detector is perfectly linear over its entire well depth, is thermally stable, and is not subject to any time dependent effects that vary during integrations or readouts. However, if the nonlinearity of the detector is not accounted for, the estimated source flux can be substantially lower than the true source flux. One can use the nonlinearity correction scheme described in Vacca et al. (2004) prior to the flux estimate in orbit. In the case of Euclid, the flux estimation will be applied on the raw exposures and the nonlinearity effects will be corrected on the ground with a method presented in the subsequent paper. A correction should also be applied on the quality factor to recover a correct QF for high flux where pixels have a non linear response but are not affected by any other anomaly. The exact method and results will be addressed in a further study.

We are convinced that the proposed flux estimate and the associated quality factor will help achieve the high precision scientific goals of the mission. We anticipate that the method can be applied by other missions that use similar detectors and readout schemes and that are subject to similar CPU and telemetry limitations.

References

Beletic, J. W., Blank, R., Gulbransen, D., et al. 2008, Proc. SPIE, 7021, 70210H
Cash, W. 1979, ApJ, 228, 939
Churazov, E., Gilfanov, M., Forman, W., & Jones, C. 1996, ApJ, 471, 673
D'Odorico, S., Andersen, M. I., Conconi, P., et al. 2004, Proc. SPIE, 5492, 220
Gardner, J. P., Mather, J. C., Clampin, M., et al. 2009, The James Webb Space Telescope, 1
Kearns, K., Primini, F., & Alexander, D. 1995, in ASP Conf. Ser. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco, CA: ASP), 331
Kubik, B., Barbier, R., Castera, A., et al. 2015a, JATIS, 1, 038001
Kubik, B., Barbier, R., Castera, A., et al. 2015b, NIMPA, 787, 315
McGregor, P. J., Conroy, P., Bloxham, G., & van Harmelen, J. 1999, PASA, 16, 273
Mighell, K. J. 1999, ApJ, 518, 380
Moseley, S. H., Arendt, R. G., Fixsen, D. J., et al. 2010, Proc. SPIE, 7742, 77421B
Nousek, J. A., & Shue, D. R. 1989, ApJ, 342, 1207
Rauscher, B. J., Arendt, R. G., Fixsen, D. J., et al. 2012, Proc. SPIE, 8453, 84531F
Rauscher, B. J., Fox, O., Ferruit, P., et al. 2007, PASP, 119, 768 (Erratum; 2010, PASP, 122, 1254)
Robberto, M., Stiavelli, M., Baggett, S. M., et al. 2004, Proc. SPIE, 5167, 166
Smadja, G., Cerna, C., Castera, A., & Ealet, A. 2010, NIMPA, 622, 288
Snyder, D. L., White, R. L., & Hammoud, A. M. 1993, JOSAA, 10, 1014
Vacca, W. D., Cushing, M. C., & Rayner, J. T. 2004, PASP, 116, 352
Wheaton, W. A., Dunklee, A. L., Jacobsen, A. S., et al. 1995, ApJ, 438, 322