

# Progress Report

5th November 2010

## 1 Introduction

This is the fifth progress report on the Austria In-kind Contribution in the context of the *Data Reduction Software Project*, subprojects A and C, covering the period from August to October 2010.

After the midterm review meeting was held at ESO headquarters on 1st October 2010, work is currently concentrating on the action items that were developed in response to the midterm report and the code prototype (`esosoft-0.0.3`, where the name “esosoft” is still provisional and will be changed soon). The midterm review was mostly concerned with subproject A, deliverables DR01, DR02 and DR03.

The main events since the previous status report were:

- **Oliver Czoske** visited ESO from 16 to 28 August 2010, for discussions on the prototype release of DR01 and DR02 and the workflow for DR03.
- An updated version of the prototype for DR01 and DR02 was submitted to ESO on 17 August (`esosoft-0.0.2`). The midterm report for project A was submitted on 3rd September, along with a new version of the prototype (`esosoft-0.0.3`). Responses to the midterm report were received from ESO on 17 September and answered on 25 September.
- The midterm review meeting for the Austrian in-kind was held on 1<sup>st</sup> October 2010 in Garching. The entire team from Vienna attended.
- Two students are now helping with the project: **Harald Leibinger** (Institute for Astronomy) is conducting photometric tests on the performance of the sky subtraction recipes (DR01), in collaboration with **Oliver Czoske** and **Werner Zeilinger**. **Christoph Wiesmeyr** (NuHAG) is investigating novel ways of dealing with fringing in spectroscopic data, together with **Hans Feichtinger** and **Darian Onchis**.
- **Saptarshi Das** spent most of October on a long-planned vacation at home in India. Since there are only two full-time coders on our team, implementing the coding changes mentioned at the midterm review meeting was slowed down significantly.

- An Austrian In-kind status seminar was held at the Oberösterreichisches Landesbildungszentrum in Kefermarkt from 2 to 4 November with presentations from the three groups involved in ESO In-kind projects (Vienna, Innsbruck, Linz). The Vienna team contributed with presentations by **Oliver Czoske**, **Werner Zeilinger**, **Christoph Wiesmeyr** and **Hans Feichtinger**.

## 2 DR03: Stacking of one-dimensional spectra

The subject of DR03, stacking of one-dimensional spectra is handled by a new recipe called `esofsoft_stack_1D_spectra`. The set of frames that is input to the recipe contains a list of one-dimensional spectra of the same object, in principle (though not necessarily) taken with the same instrument, possibly from different observing blocks and different nights. The spectra should be tagged with `1D_SPECTRUM`. If noise spectra are available, there should be one for each spectrum. Noise spectra are tagged `1D_ERROR` and contain standard errors for the fluxes, i.e. the square root of the variance.

The recipe takes a very general approach and provides for spectra sampled on different wavelength grids and with no or only imperfect flux calibration.

The wavelength grid of the output spectrum can be specified by recipe parameters giving the type of wavelength grid: linear (the default) or logarithmic and three of the four parameters: starting wavelength  $\lambda_0$ , final wavelength  $\lambda_1$ , dispersion  $\Delta$ , length of spectrum  $N$ .

In the **linear** case,  $\Delta = \Delta\lambda = \lambda_i - \lambda_{i-1}$ , and the four parameters are connected by

$$\lambda_1 - \lambda_0 = N \cdot \Delta\lambda. \quad (1)$$

The wavelength of pixel  $i$  is given by

$$\lambda_i = \lambda_0 + i\Delta. \quad (2)$$

In the **logarithmic** case,  $\Delta = \lambda_i/\lambda_{i-1}$ , and the four parameters are connected by

$$\lambda_1/\lambda_0 = \Delta^N. \quad (3)$$

The wavelength of pixel  $i$  is given by

$$\lambda_i = \Delta^i \lambda_0 \quad (4)$$

or

$$\log \lambda_i = \log \lambda_0 + i \log \Delta \quad (5)$$

The correspondence between our recipe parameters and the respective fits keywords is given in Table 1. In terms of fits keywords, Eqs. (2) and (5) read

$$\lambda_i = \text{CRVAL1} + (i - \text{CRPIX1}) \cdot \text{CDELTA1} \quad (6)$$

$$\log \lambda_i = \text{CRVAL1} + (i - \text{CRPIX1}) \cdot \text{CDELTA1} \quad (7)$$

If no WCS parameters are specified, all spectra are resampled to the wavelength grid of the first spectrum in the SOF.

Table 1: Correspondence between recipe parameters and FITS WCS keywords for linear and logarithmic wavelength scales.

FITS keyword	linear	logarithmic
DC-FLAG	0	1
CRPIX1	1	1
CRVAL1	$\lambda_0$	$\log \lambda_0$
CDEL1	$\Delta = (\lambda_1 - \lambda_0)/N$	$\log \Delta = (\log \lambda_1 - \log \lambda_0)/N$
NAXIS1	$N = (\lambda_1 - \lambda_0)/\Delta$	$N = (\log \lambda_1 - \log \lambda_0)/\log \Delta$
CTYPE1	LINEAR	LINEAR
CUNIT1	Angstrom	Angstrom

The appropriate algorithm to transform a spectrum from one wavelength grid to another depends on the units of the data values in the spectrum. A spectrum containing counts, integrated fluxes or fluxes, in units of  $\text{erg cm}^{-2}$  or  $\text{erg cm}^{-2} \text{s}^{-1}$  can be viewed as a histogram, where each bin is assigned a value  $\int f_\lambda d\lambda$ , integrated over the bin size. A transformation from one wavelength grid to another is equivalent to a transformation from one set of bins to another. For this type of spectra, we have implemented a simple *drizzle* algorithm which computes the overlap between an input pixel and an output pixel and assigns a corresponding fraction of the input flux to the output pixel.

The other case is spectra which contain flux *densities*, in units of  $\text{erg cm}^{-2} \text{s}^{-1} \text{\AA}^{-1}$ . Flux-calibrated spectra are typically of this kind. The spectra sample a continuous function  $f_\lambda$  at the set of wavelength values. Their transformation requires a resampling algorithm, which interpolates between the wavelength values of the original grid.

In practice, the difference between resampling and rebinning will be slight (by the mean value theorem) except at places where the spectrum is very steep.

An automatic distinction between the two cases is difficult. An obvious place to check would be the FITS keyword BUNIT. We would need to know whether this keyword is set consistently by all the instrument pipelines.

The regridded spectra are next arranged in row-stacked form in a matrix, where each row corresponds to one spectrum. It is assumed that all the spectra are of the same object and taken with the same instrument, but possibly in different OBs and different nights. Unless the spectra have undergone perfect spectro-photometric calibration, there may be differences in the continuum, which can be a constant scaling (e.g. due to varying atmospheric transparency) or a difference in shape (e.g. if spectra were taken at significantly different air mass). We correct for such differences by fitting a low-order polynomial  $P(\lambda)$  by minimizing

$$\|f_{\text{ref}}(\lambda) - P(\lambda)f_i(\lambda)\| = \sum_{k=1}^{N_{\text{spec}}} (f_{\text{ref},k} - P(\lambda_k)f_{i,k})^2 \quad (8)$$

for spectrum  $i$ . The first spectrum in the SOF is taken as the reference spectrum.

Finally, the spectra are averaged using one of the methods described in the report *DR01-Sky subtraction*, i.e. min-max rejection,  $\kappa\sigma$ -clipping or median. A (weighted) arithmetic mean can

be used if the spectra are deemed free of cosmic rays and other artefacts.

### 3 Photometric tests of sky subtraction

We test the photometric fidelity of the sky subtraction routines by inserting fake sources into raw (basic calibrated) images, processing the latter and comparing the magnitudes of the fake sources as measured in the sky-subtracted images to the input magnitudes.

Fake sources are created using the programmes `skymaker` and `stuff`<sup>1</sup> to create images with fake sources. `skymaker` simulates images with randomly distributed “stars” and “galaxies”, as well as realistic background noise. “Stars” are created by `skymaker` with a shape determined by a given PSF. They are randomly distributed across the image, and their magnitudes are distributed according to a power law distribution within specified magnitude limits. Along with the image, a list of positions and magnitudes of the stars is output. `stuff` creates a list of galaxies modelled as the sum of a de Vaucouleurs bulge and an exponential disk with realistic distribution of parameters, and magnitudes in different filters determined from a set of galaxy SEDs. The galaxy list can be used as input to `skymaker` to include galaxies in the simulated image.

We create two images of the same “star” field with `skymaker`. The first image is created with no background (left panel of Fig. 1. This noise-free image is added to the real images (middle panel of Fig. 1 from one OB which then undergo the sky subtraction as described in the report *DR01-Sky subtraction* (right panel in Fig. 1). Output magnitudes are measured off the sky-subtracted images with `SExtractor`. If these magnitudes are compared to the input magnitudes we find a systematic offset of 0.1 mag, with the input magnitudes being brighter than the output magnitudes. This is due to the fact that in the sky-subtracted images the wings of the stars vanish in the background noise, whereas the input magnitudes are total magnitudes including the wings of the PSF. We therefore create a second fake image with a (constant) background with a surface brightness corresponding to the sky in the filter band under consideration. This image then has the same background noise as the sky-subtracted images but no structure in the background. The sky-subtracted images may have residual structure due to imperfect sky subtraction. The goal of the sky subtraction is to get the sky-subtracted images to be as close as possible to the noisy simulated images, hence these are the correct comparison images. Comparing the output magnitudes to the magnitudes measured off the noisy simulated image largely removes the offset.

We show examples in Figs. 2, 3 and 4. In Fig. 2, ISAAC  $K_s$ -band images of the GOODS field F01 were used, taken from a single OB. 2500 fake stars were created with magnitudes between 16 and 26 and added to the ISAAC images. The sky background in the  $K_s$  band was taken as  $13.1 \text{ mag arcsec}^{-2}$  (Cuby et al. 2000) and the photometric zeropoint of ISAAC as  $K_{s,o} = 24.08$  (Vega system, taken from the ISAAC QC web pages). The panels show the results for three different algorithms for the determination of the sky correction: median,  $\kappa\sigma$ -clipping and the mean-median ratio constraint.

The figure shows that for bright objects ( $K_s \lesssim 19$ ), the output magnitudes agree well with the input magnitudes. Outliers are fake stars which are close to real objects and could not be measured reliably. For fainter objects, the magnitude differences are biased towards negative

<sup>1</sup><http://www.astromatic.net/software/stuff>, <http://www.astromatic.net/software/skymaker>

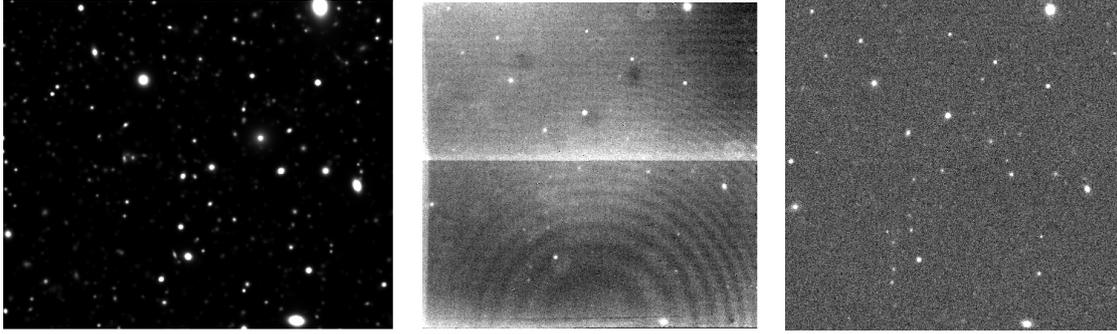


Figure 1: Left: noise-free image with stars and galaxies created with stuff and skymaker. Middle: ISAAC  $K_s$ -band image with added fake sources. Right: Sky subtracted image. The field of view of the left image equal to the total field covered by all dithered exposures in the OB.

values, i.e. the output fluxes are underestimated compared to the input fluxes. It is at present not clear, why that is so.

Fig. 3 shows the results from applying the 2-pass method to VIMOS  $I$ -band images. These images are strongly affected by fringing (see the previous progress report from 2010-08-01), and it is this feature that is removed here. Again the correspondence is very good for magnitudes brighter than around 20 mag. Interestingly, the tilt at the faint end is opposite for the mean-median ratio compared to the other two methods.

Finally, Fig. 4 shows the results of a test of the smooth background correction to a set of VIMOS  $R$ -band images of the CDFS.

A more detailed and systematic description of the sky subtraction tests will be provided in a separate report.

## 4 DR02: Justification for basing astrometry recipes on Terapix software

Deliverables DR02 concerns the astrometric calibration of imaging data and their assembly into contiguous mosaic images. An example of a data set to which this task applies is the ISAAC imaging of the GOODS field (e.g. Retzlaff et al. 2010). Further imaging data of the same field exist from the Wide Field Imager (WFI) and VIMOS.

Czoske (2002) developed tools for a similar purpose and a specific data set from the CFH12k camera on CFHT. He later adapted and applied these tools (based on a set of Perl scripts and Iraf tasks) to WFI data. Lessons learned from this work were (i) the accuracy of absolute astrometry is limited by the positional accuracy of the reference catalogue used to define the astrometric system and to correct for (relative) distortions in the imaging data; (ii) the image quality of the mosaic can be optimised by determining transformations (from input to output pixel grid) that minimise the rms spread of target positions of objects in the imaging data (i.e. relative astrometry). The accuracy achieved was around one tenth of a pixel (rms of target positions).

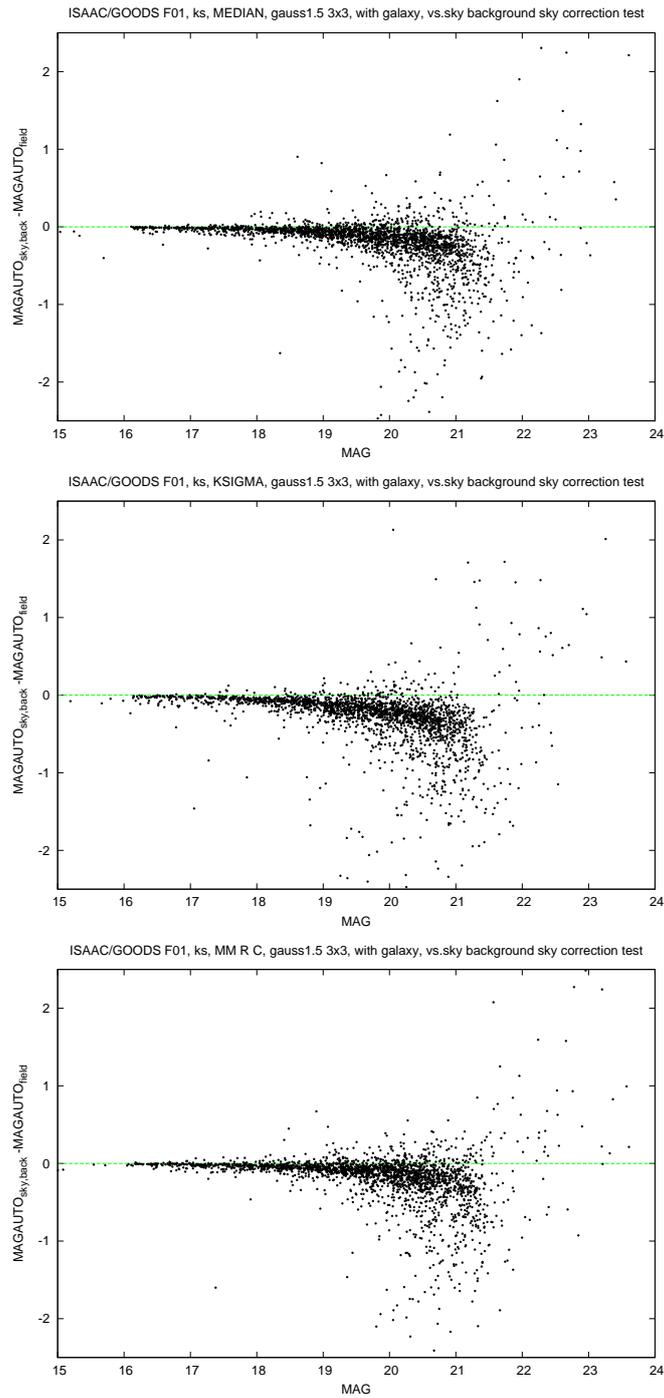


Figure 2: Comparison of input and output magnitudes of fake sources added to ISAAC  $K_s$ -band images of one exposure of GOODS field F01. The plots show sky subtraction with different estimators for the background correction image. Top: median. Middle:  $\kappa\sigma$ -clipping. Bottom: mean-median ratio.

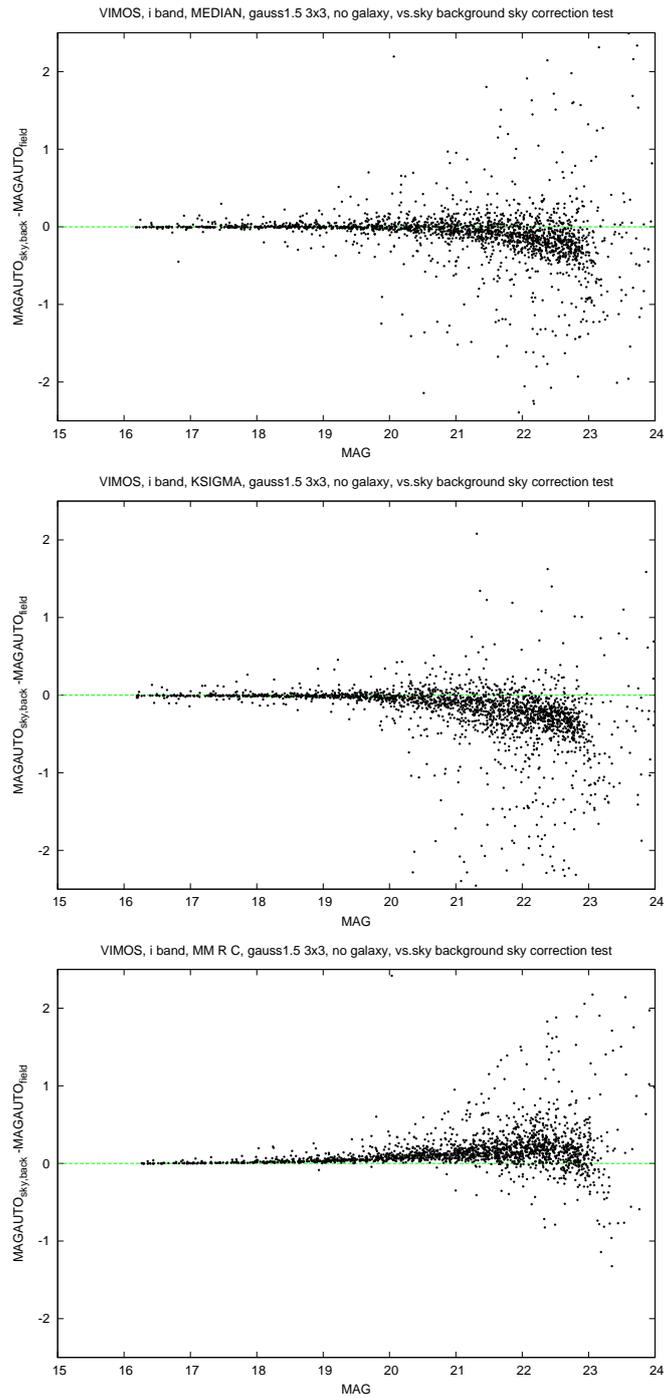


Figure 3: Comparison of input and output magnitudes of fake sources added to VIMOS *I*-band images of the GOODS/Chandra Deep Field South. The feature that is to be subtracted from these images is strong fringing. The plots show background/fringe subtraction with different estimators for the background correction image. Top: median. Middle:  $\kappa\sigma$ -clipping. Bottom: mean-median ratio. The magnitude scale has an arbitrary zero point in this case.

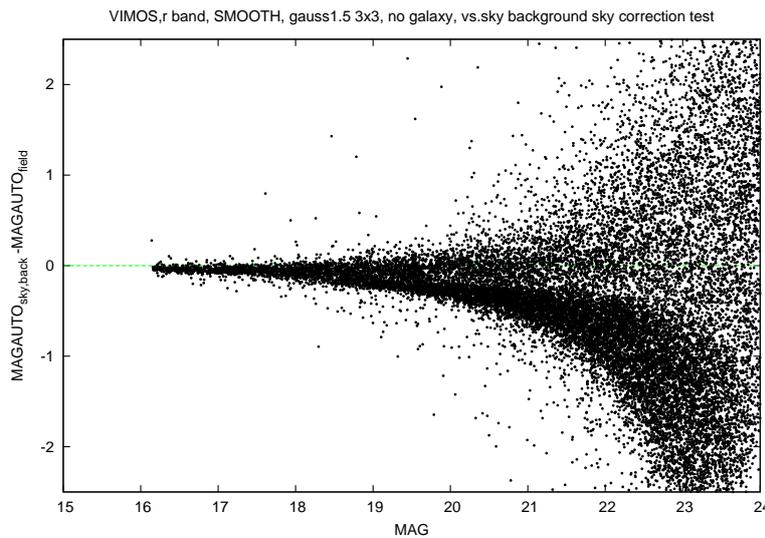


Figure 4: Comparison of input and output magnitudes of fake sources added to VIMOS *R*-band images of the GOODS/Chandra Deep Field South. Here, a smooth model of the sky background is subtracted.

Similar accuracies were reported from other calibration approaches (e.g. Kaiser et al. 1999).

Criteria for the development of astrometric recipes for DR02 were:

- The main quality criterion for a good astrometric calibration is preservation of the point spread function of the input images (in the case of varying seeing, this would be the size of the PSF in the worst-seeing exposure). PSF degradation results if the same object in different exposures is mapped to slightly different places in the output mosaic. In the worst case this might lead to double or multiple images of the same object in the mosaic.
- Mosaicing of images should be possible even in cases where no external reference catalogue is available. The astrometric solution should be based on internal data, i.e. objects in overlapping regions of exposures covering adjacent fields.
- Recipes should be instrument independent and applicable to single-extension as well as multi-extension fits files (from mosaic cameras such as HAWK-I or WFI).

As mentioned, optimal PSF quality is achieved by constraining the astrometric transformations for all exposures at the same time by minimising the spread of target positions of objects in the data. This solution is stabilized and put on an absolute reference frame by including an astrometric reference catalogue for the field.

The determination of astrometric solutions should be based on object catalogues. An alternative that one might consider is based directly on the imaging data using cross-correlation to match images. Such an approach is occasionally used to determine shifts between images (e.g. in the Iraf task `xregister`) but is hardly feasible if image distortions have to be taken into account.

Rather than implement such a procedure from scratch, it was decided early on to base the instrument-independent recipes for DR02 on the suite of programmes `SExtractor`, `Scamp` and `SWarp` developed by Emmanuel Bertin at the Terapix centre of the Institut d’Astrophysique de Paris. This is the only publicly available software suite that fulfils all of the requirements listed above. The programmes have been described in detail in the report *DR01 – Astrometry*. The algorithms used in `Scamp` as well as their implementation have been checked by Tomasz Hrycak. It was found that some details could be improved from a mathematical point of view, e.g. the transformation model could be described as a sum of orthogonal polynomials (Legendre or Chebyshev polynomials) rather than a sum of numerically less stable monomials. However, the cost of implementing such a change and the need to specify such a model in the form of FITS header keywords were judged prohibitive. Empirically, the model in terms of monomials works fine, and we have never encountered numerical problems in practice.

Quantitative testing of the astrometric recipes based on the Terapix programmes is in progress and will be described in a separate report.

## 5 Default value for window size in NIR background subtraction

The recipe `esosoft_compute_bkg` estimates the background in NIR images as a (robust) running mean over the sequence of exposures within an observing block. So far, the default value for the half window size (a recipe parameter) was set to `nhalf_window=7`. It would be desirable to make this default value adaptive to the number of exposures in the OB and their exposure time, taking into account the typical variability timescale of the NIR sky.

The R function `runmed` implements a general running median filter. The man page<sup>2</sup> mentions a default value for the (full) window size  $k$  used in an implementation of the algorithm of Haerdle & Steiger (1995) by Berwin Turlach:

$$k = 1 + 2 * \min( (n - 1) \% \% 2, \text{ceiling}(0.1 * n) ) \quad (9)$$

The R function does not actually use this default value, presumably because it does not take into account the characteristics of the data sequence other than its length  $n$ .

In the case of NIR images taken with the VLT in service mode, we can argue as follows to derive a variant of Eq. (9) which provides a reasonable adaptive default for the window size in our data. A standard OB at the VLT has a length of 60 minutes, hence the length of an exposure in a sequence of  $n$  exposures is  $60/n$ . Considering the typical variability timescale of the NIR sky background, we would like to have a half window covering about 10 minutes, hence the half window should have a length of about  $n/6$ . This suggests the following modification of Eq. (9):

$$k = 1 + 2 * \min( (n - 1) \% \% 2, \text{ceiling}(n/6.) ) \quad (10)$$

In this formulation,  $k$  is actually a step function of  $n$  which is always above  $n/6$  (see Fig.). Using this for the background computation is conservative in terms of signal-to-noise because it includes more exposures than necessary to cover a 10 min half window.

<sup>2</sup><http://stat.ethz.ch/R-manual/R-devel/library/stats/html/runmed.html>

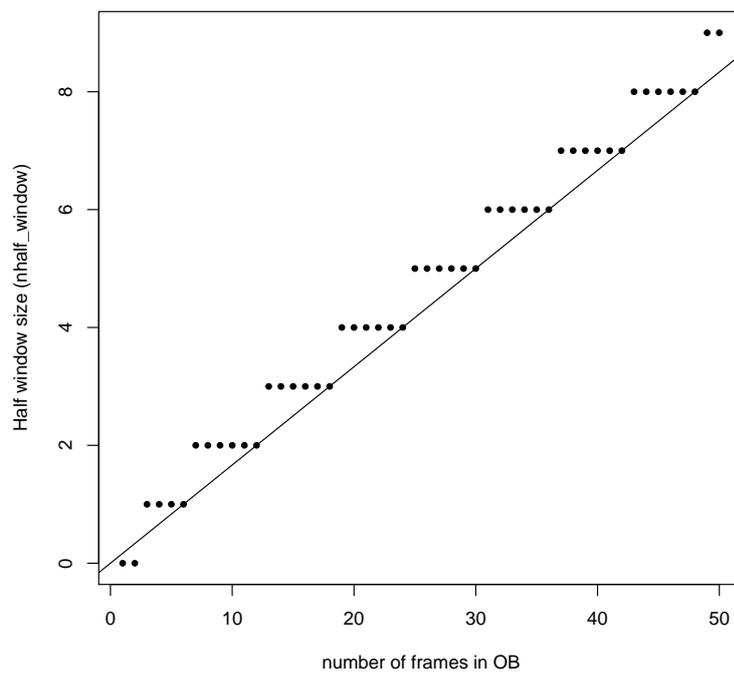


Figure 5: Points give the default value for the half window size (Eq. 10) used in `esosoft_compute_bkg`. The line gives the lower bound for the half window size of  $n/6$ .

## 6 Status of midterm action items

The following table gives the current status of the action items that were defined at the midterm review meeting. The new release of the code will be prepared by 15 November.

RIX-#	Status
WFR-01	HAWK-I background subtraction has been implemented in the new release, cf. AGA-05
WFR-03	Photometric tests so far have focused on point sources (see report); tests on parameters relevant for extended objects such as growing radius and threshold are in progress
WFR-07	The kit will be available in the new release, testing of the data sets is in progress
WFR-10	Performance tests of the algorithms are included in the report
WFR-19	see WFR-10
WFR-21	Tests are still in progress
AGA-02	implemented at the package configuration level
AGA-05	see WFR-01
AGA-10	implemented in the new release
AGA-17	see progress report
AGA-25	implemented in the new release
AGA-34	implemented in the new release
LLU-01	implemented in the new release
LLU-02	implemented in the new release
LLU-03	implemented in the new release
LLU-04	implemented in the new release
LLU-05	implemented in the new release
LLU-06	implemented in the new release
LLU-10	implemented in the new release
LLU-12	implemented in the new release

## References

- Cuby, J. G., Lidman, C., & Moutou, C. 2000, *ESO Messenger*, 101, 2
- Czoske, O. 2002, PhD thesis, Université Toulouse III – Paul Sabatier
- Haerdle, W. & Steiger, W. 1995, *Applied Statistics*, 44, 258, algorithm AS 296
- Kaiser, N., Wilson, G., Luppino, G., & Dahle, H. 1999, astro-ph/9907229
- Retzlaff, J., Rosati, P., Dickinson, M., Vandame, B., Rit , C., Nonino, M., Cesarsky, C., et al. 2010, *A&A*, 511, A50