OU-MER Data Challenge N.1

Photometric catalogs on PSF-matched images

Document Version: V2.0 – 2013-12-18

M. Castellano, S. Pilo, A. Fontana, M. Kuemmel, H. Dole

1 Introduction

The Data Challenge N.1 (DC1 hereafter) is the first of a planned series of data analysis efforts, coordinated among the various Euclid OUs, that have the goal of testing the steps of the SGS data processing.
This first Data Challenge is based on a first version of matched aperture photometry catalogs, with the aim of starting the DC interaction among OUs and testing basic approaches to the data processing.

In this case OU-MER will release a set of data (images and catalogs) that simulates data obtained by Euclid, although on a very small scale and with some simplifications.

The basic idea is to use the images obtained by CANDELS in 8 different filters to mimic the EUCLID observations in the VIS and NIR bands, as well as in 4 bands from the ground. Objects from these images will be extracted with standard processing tools to obtain multiwavelength catalogs that can be used to test photometric redshifts tools.

In DC1 the first version of two set of different multiwavelength catalogs are released:

- S1: optimal PSF-matched dataset (Sect. 3.1.1).

- S2: homogeneous PSF-matched dataset (Sect. 3.1.2).

2 The simulated data

2.1 The concept

The simulations are based on data obtained by the CANDELS survey, that uses HST instruments to map five small sky regions. In this first release, we have used CANDELS data on the GOODS-S field.

The key advantages of this approach are:
- simulations are based on real images of extragalactic fields. This guarantees real galaxy properties (shapes, colors, clustering);
- Images have been taken from HST in all bands, and they are both much deeper (at least 2mags) and have higher resolution than Euclid images.

This allows us to have a reference "truth" to compare with the Euclid simulated data.

- The available HST images sample the entire wavelength range over which Euclid will operate, including the ancillary ground based images. At each wavelength, colors and morphologies of galaxies are therefore correct.

The shortcomings or known limitations of these simulations are the following:

- the survey area is very small – the overall CANDELS survey is only 800 sq. arcmin, and the data used here cover only 100 sqarcmin. In addition to the obvious statistical limitations, the CANDELS fields are deliberately chosen to be devoid of bright stars and galaxies, and therefore do not include the full range of objects sampled in a typical Euclid field.
- the PSF adopted are not exactly identical to the expected Euclid/ground based PSFs, they are just simple analytical function, chosen in order to produce average FWHMs similar to the predicted ones. Note also that the PSF is constant across the field (albeit different in each band);
- the simulations do not include instrumental defects and observational mishaps like inhomogeneous S/N, fluctuations or errors in the ZP and so on. In this sense, they are idealized versions of the final data.
- the band-passes of the CANDELS images are not identical to those of Euclid, although they encompass a similar wavelength range. In particular the Euclid VIS filter is significantly wider in wavelength than the I814 used here to mimic the VIS image. For this reason we dubbed the latter VIS* to remark that it is significantly different than In this respect the output of the simulations and of the relevant photo-z are only broadly indicative of the future Euclid data.

2.2 Properties of simulated images

Images have been generated by taking the original CANDELS images, rebinning and smoothing them to the expected pixel-scale and seeing, and adding random Gaussian noise in order to match the expected depth (Fig. 1, more details on the simulation software are given in Appendix 1). At the present stage the simulation software does not add any photon-noise component.
Since original images are nearly one order of magnitude deeper than the final simulated ones, the final noise is essentially uncorrelated. Image properties have been chosen according to the specifications reported in table (1), i.e. the values predicted in the Red Book, for the Euclid bands, and from DES for the ground based images (J. Mohr, priv. Comm.).

image properties	VIS	NIR	G-ext	R-ext	I-ext	Z-ext
seeing (arcsec)	0.2	0.3	0.83	0.79	0.79	0.78
depth (AB mag)	24.5 @ 10σ	24.0 @ 5σ	25.2 @ 10σ	24.8 @ 10σ	24.0 @ 10σ	23.4 @ 10σ
pixel scale (arcsec/pixel)	0.1	0.3	0.3	0.3	0.3	0.3

1. Image properties in the various bands.

Referring to table (1), reported depths are computed using total magnitudes of the sources on VIS and NIR images and magnitude in 1.6*FWHM for EXT images.

Each set of images in a specific band has been simulated from a related image from CANDELS dataset that has been acquired in a similar filter, according to the following scheme:

I814 −> VIS*
Y105 −> Ynir
J125 −> J nir
H160 −> H nir
B435 −> G ext
V606 −> R ext
I775 −> I ext
Z850 -> Zext

The original CANDELS images have been masked before simulation with a mask that selects the common area of all original images.

We have obtained ten different independent realizations of the full data set. They have been obtained by adding independent noise on the original CANDELS images. This implies that each object has been measured independently up to ten times, and up to ten different photometric redshifts can be derived for it. It is therefore possible to derive uncertainties by computing the r.m.s. among different individual measurements.

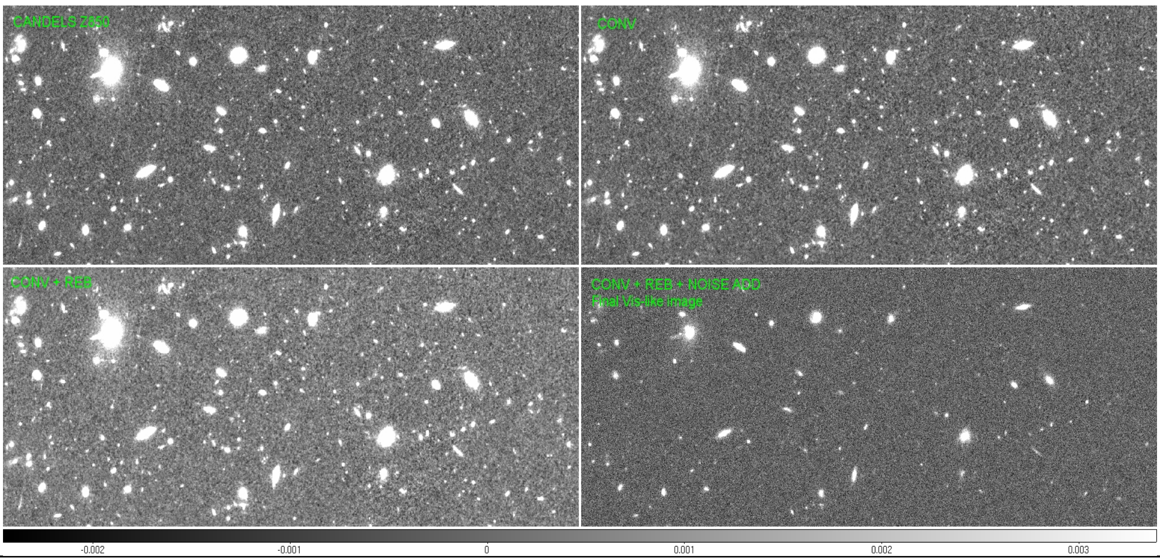

Figure 1: the various steps of the image creation. From top to bottom, and left to right: the original image, the image after convolution to the Euclid PSF, the image after rebinning to the Euclid pixelscale, the final image with noise added.

3 The catalogs

3.1 The method: PSF-matched aperture photometry

In this first version, catalogues have been extracted from the simulated images using SExtractor 2.5.0. Detection and total flux determination has been performed on the VIS*-like image. Fluxes in the other bands have been computed in apertures on PSF-matched images, according to two different strategies which are described in Sect. 3.1.1 and Sect. 3.1.2 considering fluxes in physical units.

The conversion between counts (C), as measured by SExtractor, and fluxes (F) in μJys has been performed following:

F=C x 10^(0.4*(23.9-ZP))

where the zeropoints (ZP) are quoted in the Sect. 3.3 of this document.

The zeropoints determination of HST bands is described in:

http://www.stsci.edu/hst/acs/analysis/zeropoints
http://www.stsci.edu/hst/wfc3/phot_zp_lbn

for ACS and WFC3 bands respectively.

3.1.1 S1: optimal PSF-matching

Fluxes in the various bands have been computed as follows. For each image, the VIS image (i.e. the "detection" image) has been convolved with a specific kernel (different for each image) in order to have the same PSF of the "measure" image. Fluxes on both images have been computed in circular apertures of *n* times the FWHM of the "measure" image.

Fluxes in each band have to be intended as total flux measured according to the following equation:

$$F_{tot}(Mband)=\left(\frac{F(Mband)}{F(Dband_{M-matched})}\right)_{aper}\times F_{tot}(Dband)$$

where *Mband* is the measure band, *Dband* is the detection band, and $Dband_{M\text{-}matched}$ is the VIS* one matched to *Mband*. Total VIS* fluxes (Ftot) have been measured on the original VIS* image *Dband* from SExtractor FLUX_AUTO counts.

For consistency, flux uncertainties have also been scaled on the basis of the ratio between the total *Dband* flux and the $Dband_{M\text{-}matched}$ one measured in aperture:

$$errF_{tot}(Mband)=errF(Mband)_{aper}\times\frac{F_{tot}(Dband)}{F(Dband_{M-matched})_{aper}}$$

Three different versions of the catalogues have been obtained for *n*=1,2,3. We remark that the FWHM is different in each band. This procedure does not degrade the information contained in the images with better seeing. In the case of Euclid, this is particularly important for the NIR images that have FWHM significantly better than ground-based ones.

3.1.2 S2: homogeneous PSF-matching

The S2 catalogues are built with a simpler procedure, namely all images are smoothed to the *g-ext* band, which is the one with the worst PSF.

Total VIS* fluxes have been measured on the original VIS* image from SExtractor FLUX_AUTO counts. Fluxes on the PSF-matched images have been computed in circular apertures of *n* times the FWHM of the *g-ext* one. The flux in each band is scaled to total flux according to the following equation:

$$F_{tot}(Mband) = \left(\frac{F(Mband_{g-matched})}{F(Dband_{g-matched})} \right)_{aper} \times F_{tot}(Dband)$$

where $Mband_{g\text{-}matched}$ is the measure band (matched to the *g-ext*), $Dband_{g\text{-}matched}$ is the VIS* matched to the *g-ext*, and *Dband* is the detection band where total VIS* fluxes are measured. Three different versions of the catalogues have been obtained with *n*=1,2,3.

As above, flux uncertainties have also been scaled on the basis of the ratio between the total *Dband* flux and the $Dband_{g\text{-}matched}$ one measured in aperture:

$$errF_{tot}(Mband) = errF(Mband_{g-matched})_{aper} \times \frac{F_{tot}(Dband)}{F(Dband_{g-matched})_{aper}}$$

3.2 The output catalogues

We have obtained full multiwavelength S1 and S2 catalogues for each of the 10 different realizations of the image sets. We note that since also the detection image (VIS*) is different in each realization, the list of objects detected in each realization will also be different.

All the S1 and S2 catalogues have been finally added into three single files, one for each aperture adopted. E.g.:

EUSIM_1FWHM_S1_v2.cat
EUSIM_2FWHM_S1_v2.cat
EUSIM_3FWHM_S1_v2.cat

Each catalogue contains 10 different realizations of the data set. Only objects with S/N(VIS*)>5 have been included, (S/N computed from the total VIS* flux and relevant uncertainty). In addition, the catalogues have been cleaned by removing objects with problematic photometry (mostly located at the borders). After the S/N cut and the culling procedure the average number of detected objects in each realization is ~ 6000 objects, of which ~1200 have spectroscopic redshift and ~200 are AGNs (see below).
The total number of sources is about 60000 in each catalogue (catalogues in different FWHM have slightly different length after the culling procedure). Note that sources share the same ID number in the S1 and S2 catalogs since the detection is done on the VIS* image in both cases.

The catalogues contain total fluxes and relevant uncertainties (μJy) in the eight bands plus additional information on each object as described below.

Catalogue format

The catalogue contains the following columns preceded by a header in “Topcat” format:

ID
FLUX_G
FLUX_R
FLUX_I
FLUX_VIS
FLUX_Z
FLUX_Y
FLUX_J
FLUX_H
FLUXERR_G
FLUXERR_R
FLUXERR_I
FLUXERR_VIS
FLUXERR_Z
FLUXERR_Y
FLUXERR_J
FLUXERR_H
ZSPEC
AGN_FLAG
SUPER_ID

ID

The ID column reports the progressive identification number of the objects.
The first realization of each one of the three catalogues has the original extraction ID numbers.
The ID number for each object in each of the following nine realizations has been incremented by a factor of 10000 at a time (e.g. the object with ID number 23 in the fifth realization is put at ID 40023 in the general catalogue) .
On the basis of SExtractor extraction flag, and/or after inspection, a small fraction of objects (typically 8-10 %) have been removed from the initial SExtractor catalogue because of unreliable photometry (i.e. objects falling at the edges of the images, saturated stars etc.): for this reason some consecutive numbers in the catalogue are missing from the ID list.

Fluxes

Fluxes in each band have to be interpreted as the total flux measured as described above. They are measured in μJy.
Due to noise fluctuations negative flux values can be found when objects are undetected in one or more bands.

Flux Errors

The flux measurement error corresponds to the uncertainty in the Ftot flux for the VIS* band, and to the uncertainty in F(Mband) computed as described in Sect. 3.1.1 (S1) and Sect. 3.1.2 (S2).

SPECZ
A column with spectroscopic redshift of the objects is also included.
We performed a cross-correlation between the EUCLID simulated catalogue and the CANDELS one using a very small aperture (0.15 arcsec) to ensure accurate and unambiguous association between the two. The spec-z column report the spectroscopic redshift for every source appropriately matched to CANDELS sources provided with reliable spectroscopic determination. Only spectroscopic redshifts with quality flag=1 (highest reliability) have been considered.
If no univocal cross-correlation with a CANDELS source is found, or if the corresponding CANDELS source has no spectroscopic redshift, the value of the specz column is set to -1.0.

AGN_FLAG

This column report a three digit number with the following meaning:

1 = detected, 0 = not detected with order: |X-ray|IR|Radio|

Where:

XRAY - Chandra 4Ms Sample of Xue et al. (2011) cross-matched to CANDELS
IR - Sample selected using method of Donley et al. (2012). The infrared selected AGN sample has not yet publicly released, for this reason the CANDELS photometry of these objects has not been included in the companion EUSIM_v2.CANDELS_PHOT.cat.
Radio - Sample of Padovani et al. (2011)

Examples: 100 - X-ray detected only

110 - X-ray and IR detected

011 - IR and Radio detected

The value is set to “999” if the object has no univocal cross-correlation with objects in the CANDELS catalogue.

SUPER_ID
The SUPER ID uniquely identifies a given source among different realizations. As for the assignment to a spectroscopic redshift value, a cross-correlation between catalogues has been performed in small apertures to ensure univocal association. When no robust cross-correlation is found the SUPER_ID is set to “-1”: this may typically happen in two cases: either two nearby sources that are deblended in the CANDELS data and not in Euclid, due to lower S/N, or a single source in CANDELS that is erroneously deblended in the shallower Euclid data. Typically a bright galaxy is expected to have 10 different entries in the catalogue (i.e. 10 different IDs all with the same SUPER_ID), while fainter galaxies may have less than 10 if they are detected in all bands. A companion catalogue (EUSIM_v2.CANDELS_PHOT.cat) is also realesed containing the official CANDELS GOODS-South photometry of the objects identified with a SUPER_ID and having public SPECZ and AGN_FLAG information.

3.3 Data release

The following data products are available for download at

http://euclid.roe.ac.uk/projects/sgw/wiki/Data_Challenge

EUSIM.tar.gz:

- catalogues (EUSIM_xxFWHM_Syy_v2.cat, where xx=1,2,3, yy=1,2) described in the present document.
- companion CANDELS photometric catalogue: EUSIM_v2.CANDELS_PHOT.cat. Only objects with fully public spectroscopic and AGN information are included. The first column of the catalogue report the SUPER_ID enabling cross-identification with EUCLID simulated data. The remaining columns are extracted from the official CANDELS GOODS-SOUTH catalogue and are described in detail in:
Guo, Y. et al. ApJS 2013 (http://adsabs.harvard.edu/abs/2013ApJS..207...24G).

FILTERS.tar.gz : total system throughput curves (ascii format) for the adopted filters: wavelength (Angstroms), transmission. The name indicates the simulated euclid band and the native instrument band (e.g. Gext_ACSf435w.txt, Jnir_WFC3f125w.txt etc).

One realization of the images (plus rms maps) rebinned to the VIS* pixel scale (0.1”).

The zeropoints (AB magnitudes) are:

g	25.673
r	26.486
i	25.654
VIS*	25.943
z	24.862
Y	26.27
J	26.25
H	25.96

3.4 Terms of usage

Data products from DC1 release shall not be distributed outside the Euclid Consortium. DC1 products are meant for internal use of OU-MER and OU-PHZ: they can be shared with other OUs, presented and used in publications upon approval from OU-MER.

For any request please contact: euclid-oumer@usm.lmu.de.

4 Validations

We show in the following some plots produced to check the fluxes given in the catalogues.

Comparison with CANDELS fluxes
The following plots (Fig. 2 and 3 for S1 and S2 respectively) present a comparison between flux ratios of one of the ten realization of the 3 FWHM EUCLID catalogue

and the CANDELS catalogue. The abscissa is the flux measured in the CANDELS I814 band (µJy) .
The flux ratios for all bands are on average consistent with the corresponding flux ratios estimated (for the same objects) on the much deeper CANDELS images, with a scatter that obviously increases at faint fluxes. Similar results are obtained for the S1 realizations with 1FWHM and 2FWHM aperture photometry. <u>We caution that the present version of S2 catalogues in 2FWHM and 1FWHM apertures shows a systematic positive offset at the ~5% and 10% level (respectively) in the NIR-based flux ratios with respect to CANDELS photometry</u>. The origin of the discrepancy is still under investigation, with a possible explanation being the problematics in matching the undersampled NIR PSF to the much larger EXT one.

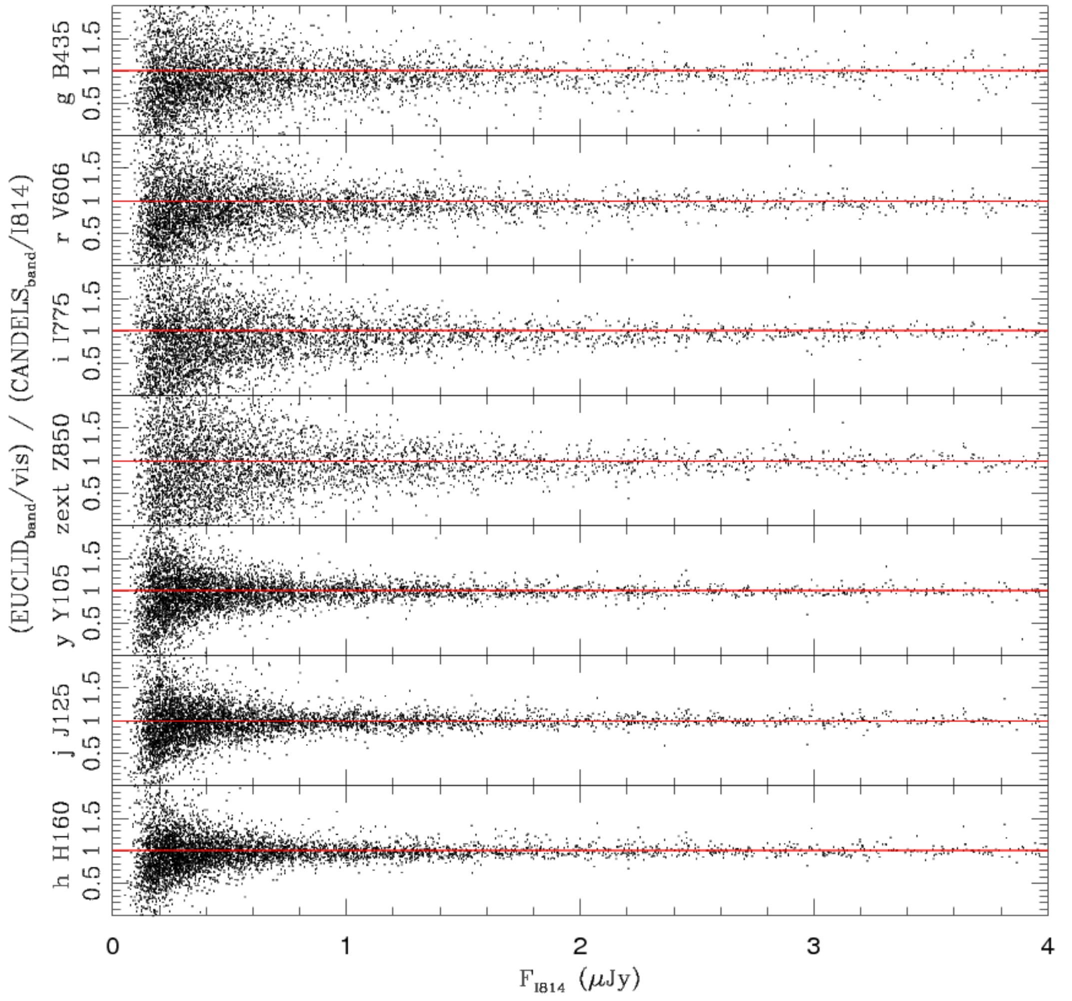

Figure 2: ratio between flux ratios estimated in the S1 Euclid data (3 FWHM apertures) and the same ratios measured on original CANDELS data.

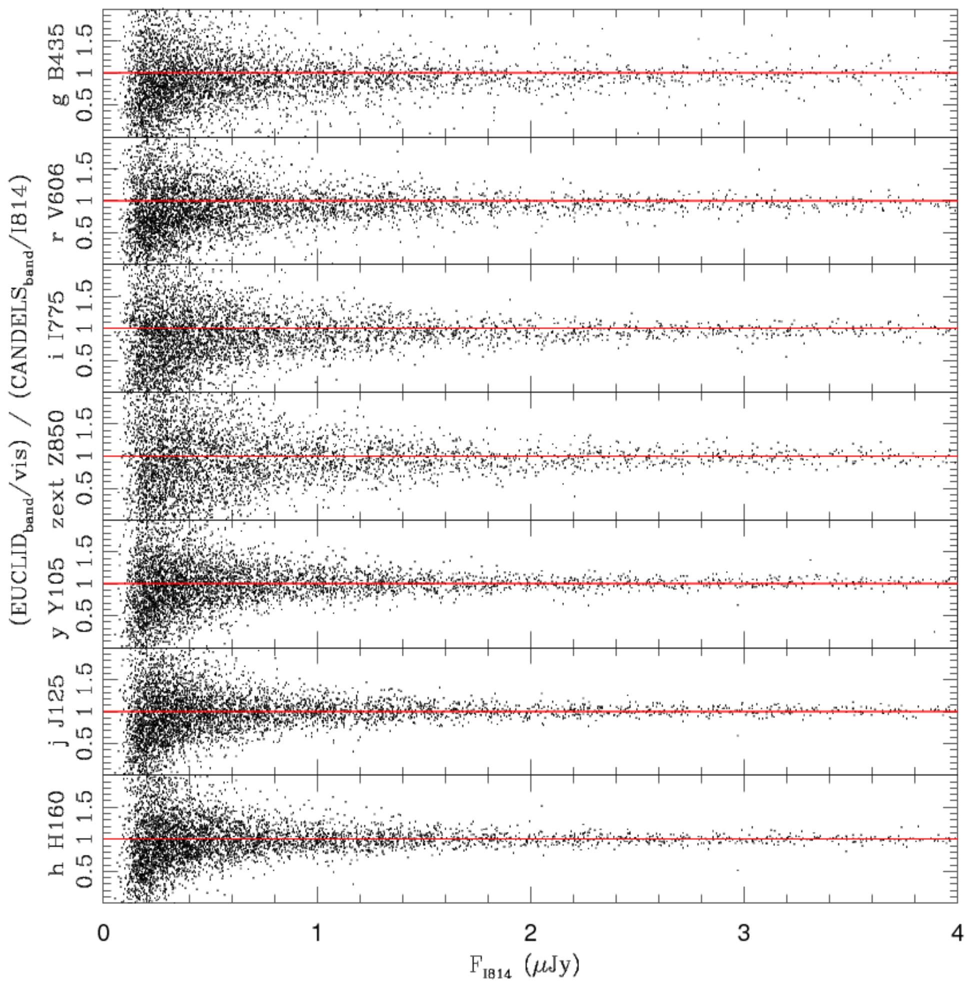

Figure 3: ratio between flux ratios estimated in the S2 Euclid data (3 FWHM apertures) and the same ratios measured on original CANDELS data.

Flux RMS

The following plots (Fig. 4 and 5 for S1 and S2 respectively) show the rms value of the distribution of the colours (expressed in fluxes) for each object over the ten realizations of the 1,2,3 FWHM EUCLID catalogues.
The abscissa value is the average VIS* flux (μJy) for each object over the 10 realizations.
It is clearly seen that the measures with the 3FWHM apertures are noisier than the others, since they extend to outer regions of galaxies where S/N is poor. In addition, NIR-based colors in the S2 catalog are nosier than in the S1 one, as expected given the much larger apertures used.

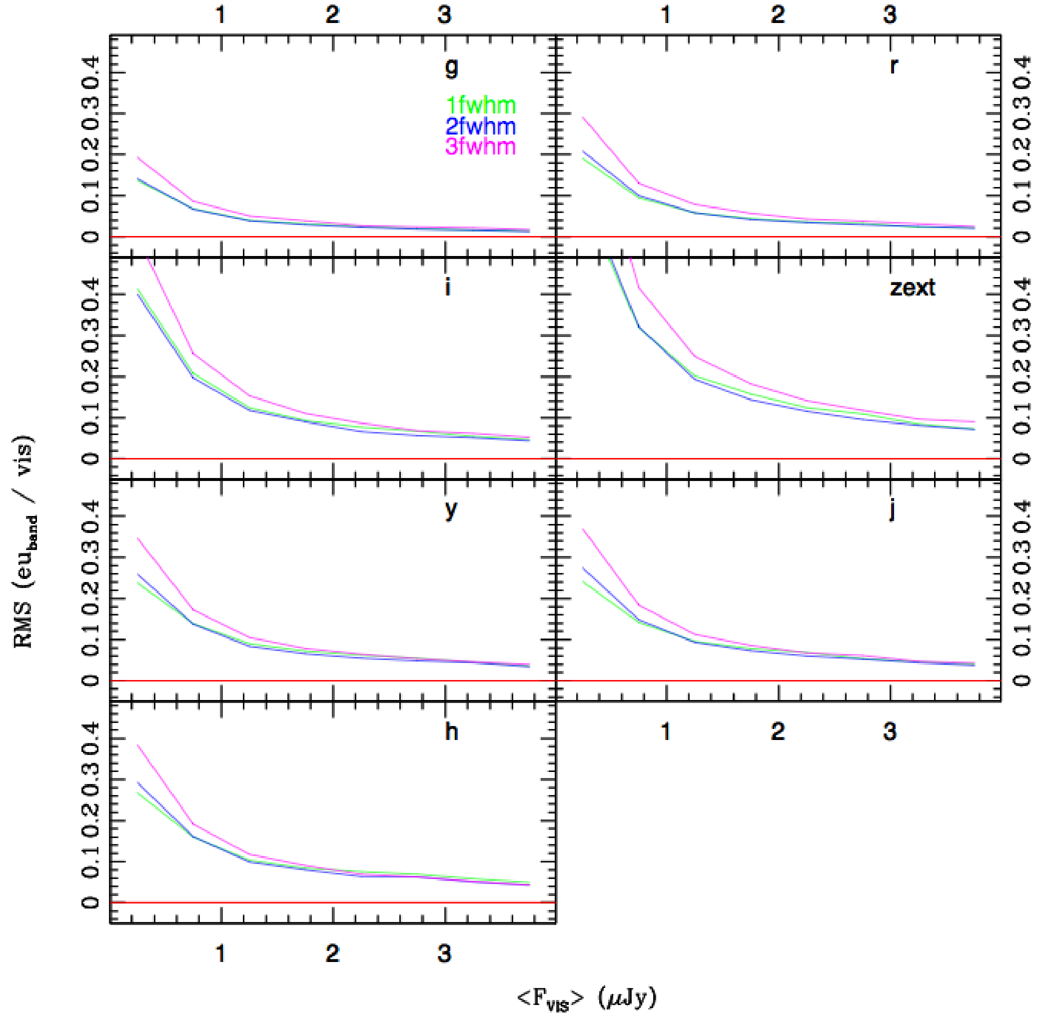

Figure 4: average RMS of the measured colors in various bands in the S1 catalogues as a function of the VIS* flux. RMS is computed for each galaxy over its ten realizations.

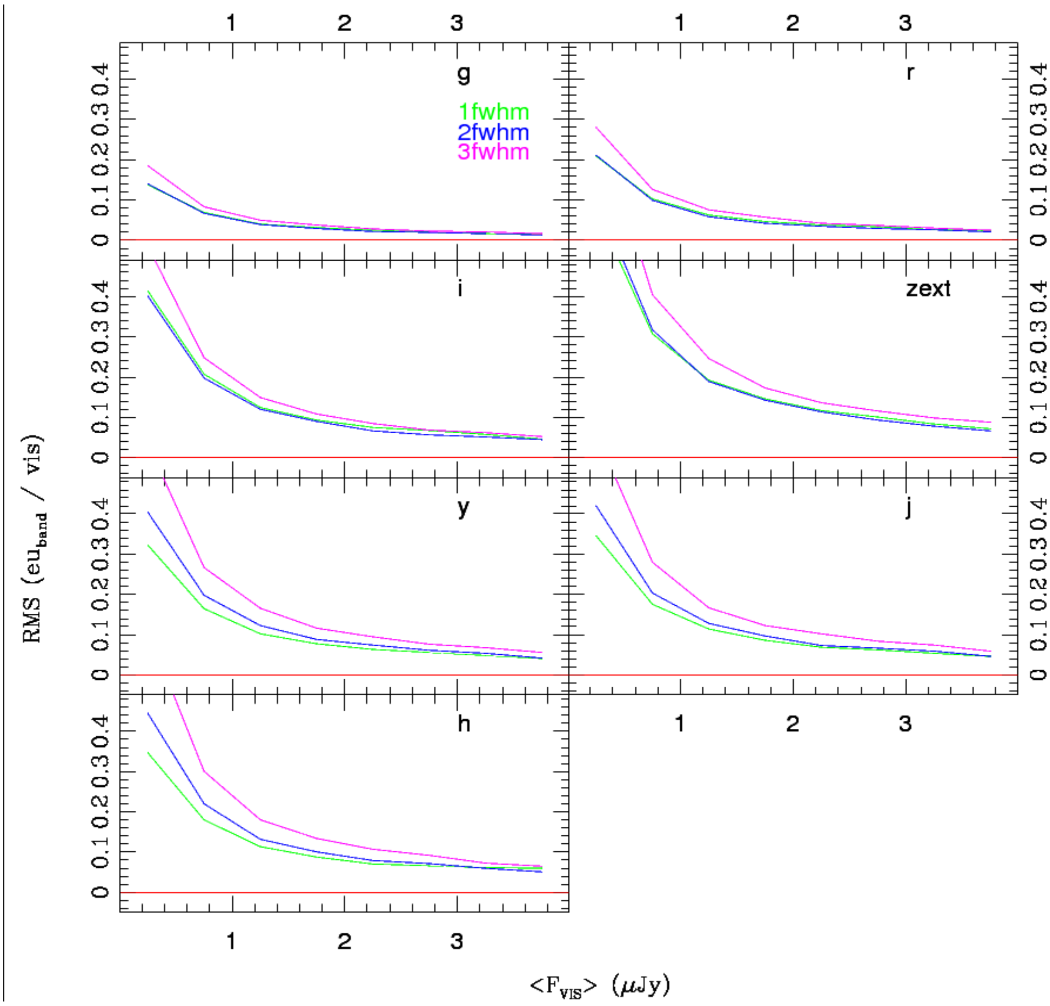

Figure 5: average RMS of the measured colors in various bands in the S2 catalogues as a function of the VIS* flux. RMS is computed for each galaxy over its ten realizations.

Appendix 1: the simulation software

Our software (simpipe.py) takes as input an high-resolution image (from CANDELS GOODS field in this specific case) and manipulates it in order to obtain a simulated one with the desired features (i.e. a VIS-like, NIR-like or EXT-like image).
Figure (6) reproduces a scheme of the pipeline, a series of four processing steps (marked with the numbers 1-4 in the figure) that are exectued in the sequence:
• Mkkernel −> generates an analytical kernel according to the input image PSF and the PSF required for the simulated one.
• Convolve −> operates the convolution from the input image to the convolved one according to the previously generated kernel.
• Swarp −> performs the rebinning of the convolved image to required pixel scale.
• Mknoise −> Gaussian noise is added in each pixel to reproduce the desired depth in output.

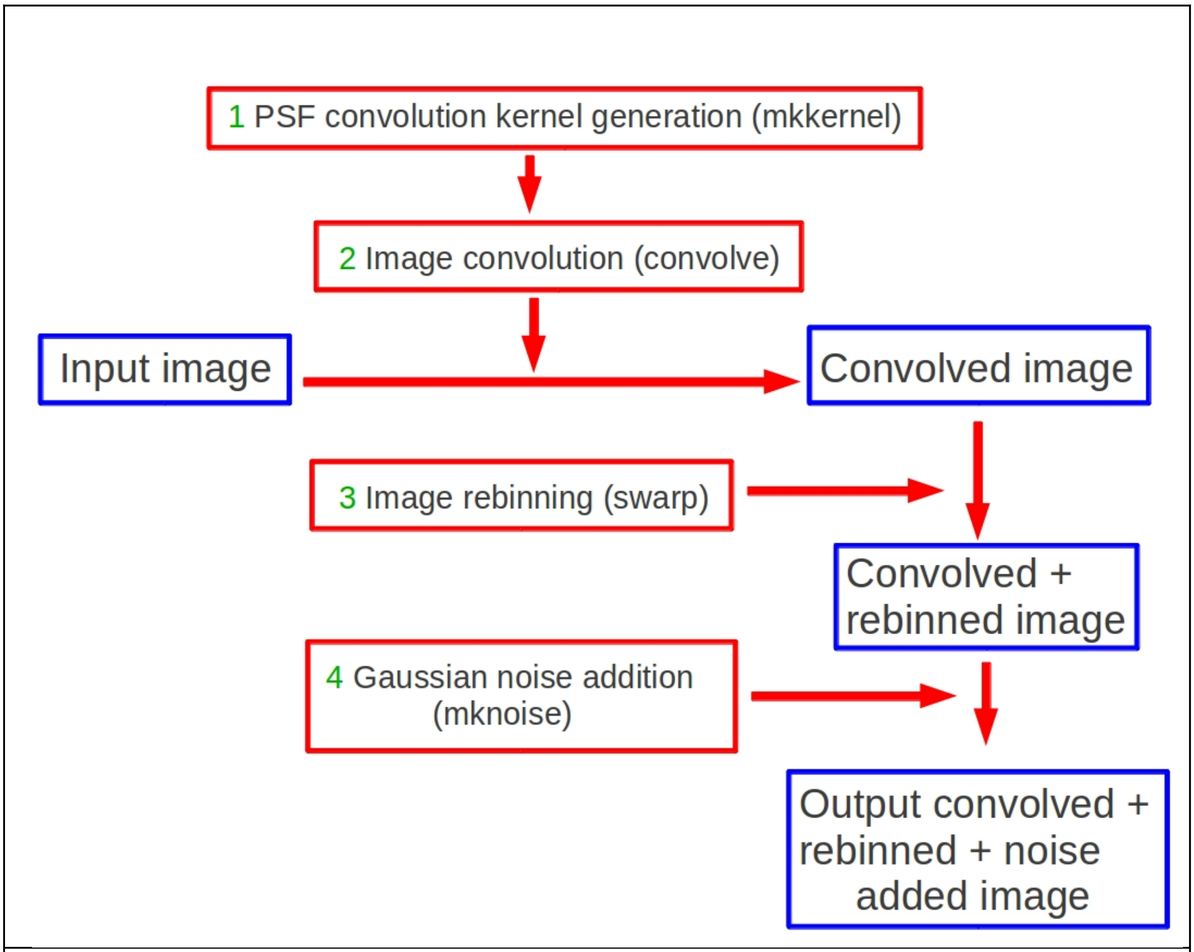

Figure 6: Pipeline scheme

The original CANDELS images have a non-uniform depth, and particular care has been devoted to the noise addition: Gaussian noise is addded on a pixel-to-pixel

basis according to a scaling factor that takes into account the pixel-to-pixel variation of the original image depth.
Resulting rms map is an image with constant value in the portion covered by the field and has a constant value of 10^{16} outside. Rms map value is the result of the following equation:

$$rms_{out} = \frac{10^{[0.4\times(ZP-m_l)]}}{[(S/N)\times\sqrt{(}\pi)\times(\frac{nFWHM}{2\times pxs})]}$$

where *ml* is the reference magnitude (at the given S/N) measured in *n* times the PSF FWHM. *ZP* and *pxs* are the zeropoint and the pixelscale of the image respectively.

Work is in progress to include in the simulation pipeline also realistic photon-noise together with the Gaussian sky noise component.