The missing light of the Hubble Ultra Deep Field

Alejandro Borlaff, Ignacio Trujillo, Javier Román, John E. Beckman, M. Carmen Eliche-Moral, Raúl Infante-Sáinz, Alejandro Lumbreras-Calle, Rodrigo Takuro Sato Martín de Almagro, Carlos Gómez-Guijarro, María Cebrián, et al.

▶ **To cite this version:**

Alejandro Borlaff, Ignacio Trujillo, Javier Román, John E. Beckman, M. Carmen Eliche-Moral, et al.. The missing light of the Hubble Ultra Deep Field. Astronomy and Astrophysics - A&A, 2019, 621, 10.1051/0004-6361/201834312 . insu-03711440

HAL Id: insu-03711440
https://insu.hal.science/insu-03711440

Submitted on 1 Jul 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A&A 621, A133 (2019)
https://doi.org/10.1051/0004-6361/201834312
© ESO 2019

Astronomy & Astrophysics

The missing light of the *Hubble* Ultra Deep Field⋆

Alejandro Borlaff[1,2], Ignacio Trujillo[1,2], Javier Román[1,2], John E. Beckman[1,2,3], M. Carmen Eliche-Moral[1], Raúl Infante-Sáinz[1,2], Alejandro Lumbreras-Calle[1,2], Rodrigo Takuro Sato Martín de Almagro[4], Carlos Gómez-Guijarro[5], María Cebrián[1,2], Antonio Dorta[1,2], Nicolás Cardiel[6], Mohammad Akhlaghi[7], and Cristina Martínez-Lombilla[1,2]

[1] Instituto de Astrofísica de Canarias, C/ Vía Láctea, 38200 La Laguna, Tenerife, Spain
e-mail: asborlaff@iac.es
[2] Facultad de Física, Universidad de La Laguna, Avda. Astrofísico Fco. Sánchez s/n, 38200 La Laguna, Tenerife, Spain
[3] Consejo Superior de Investigaciones Científicas, Spain
[4] Instituto de Ciencias Matemáticas, Consejo Superior de Investigaciones Científicas, Madrid, Spain
[5] Cosmic Dawn Center (DAWN), Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[6] Departamento de Astrofísica y CC. de la Atmósfera, Universidad Complutense de Madrid, 28040 Madrid, Spain
[7] Univ. Lyon, Univ. Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, 69230 Saint-Genis-Laval, France

Received 24 September 2018 / Accepted 12 November 2018

ABSTRACT

Context. The *Hubble* Ultra Deep field (HUDF) is the deepest region ever observed with the *Hubble* Space Telescope. With the main objective of unveiling the nature of galaxies up to $z \sim 7-8$, the observing and reduction strategy have focused on the properties of small and unresolved objects, rather than the outskirts of the largest objects, which are usually over-subtracted.
Aims. We aim to create a new set of WFC3 IR mosaics of the HUDF using novel techniques to preserve the properties of the low surface brightness regions.
Methods. We created ABYSS: a pipeline that optimises the estimate and modelling of low-level systematic effects to obtain a robust background subtraction. We have improved four key points in the reduction: 1) creation of new absolute sky flat fields, 2) extended persistence models, 3) dedicated sky background subtraction and 4) robust co-adding.
Results. The new mosaics successfully recover the low surface brightness structure removed on the previous HUDF published reductions. The amount of light recovered with a mean surface brightness dimmer than $\overline{\mu} = 26$ mag arcsec^{-2} is equivalent to a $m = 19$ mag source when compared to the XDF and a $m = 20$ mag compared to the HUDF12.
Conclusions. We present a set of techniques to reduce ultra-deep images ($\mu > 32.5$ mag arcsec^{-2}, 3σ in 10×10 arcsec boxes), that successfully allow us to detect the low surface brightness structure of extended sources on ultra deep surveys. The developed procedures are applicable to HST, JWST, EUCLID and many other space and ground-based observatories.

Key words. techniques: image processing – techniques: photometric – methods: observational – galaxies: evolution – galaxies: structure – galaxies: high-redshift

1. Introduction

The *Hubble* Ultra Deep field is a 11 arcmin^2 region of the sky located in the southern hemisphere ($\alpha = 3$h 32m 39.0s, $\delta = -27°$ 47′ 29.1″, J2000), in the Fornax Constellation. Included inside the *Chandra* Deep Field South (Giacconi et al. 2002) and the GOODS-South Field (Giavalisco et al. 2003), it was observed by Beckwith et al. (2006) during 10^6 s of the *Hubble* Space Telescope Director's Discretionary Time with the Advanced Camera for Surveys (ACS), becoming the deepest image of the sky ever obtained. The authors divided the exposure time between the available filters (F435W, F606W, F775W and F850LP) with the main objective of creating a robust sample of galaxies between $4 < z < 7$ by using the Lyman break dropout method (Steidel & Hamilton 1992; Steidel et al. 1996a,b).

Since then, an increasingly number of follow up projects with different telescopes has continued the observations of this field, including observation at radio wavelengths (VLA, Rujopakarn et al. 2016; ALMA, Dunlop et al. 2016; Aravena et al. 2016a,b; Walter et al. 2016), in the infrared (*Spitzer* IRAC, Labbe et al. 2015), optical and near-infrared (near IR) spectroscopy (VLT MUSE, Bacon et al. 2017), ultraviolet (WFC3 UVIS, Teplitz et al. 2013), and X-rays (*Chandra*, Xue et al. 2011; Luo et al. 2016; *XMM-Newton*, Comastri et al. 2011). In addition to these, the replacement in 2009 of the Wide Field and Planetary Camera 2 by the Wide Field Camera 3 (WFC3) during the *Hubble* Space Telescope Servicing Mission 4 (STS-125) allowed astronomers to continue the exploration of this field with a deep, high-resolution survey in near IR. With the main objective of finding the earliest sources in the Universe, HUDF09 (Oesch et al. 2009; Bouwens et al. 2009) and the *Hubble* Ultra Deep Field 2012 (HUDF12 hereafter, Koekemoer et al. 2012) have increased the number of filters and exposure time in the HUDF. In addition, the eXtreme deep field (XDF hereafter, Illingworth et al. 2013), reprocessed

⋆ The final ABYSS WFC3 IR HUDF mosaics are only available at http://www.iac.es/proyecto/abyss/ and at the CDS via anonymous ftp to cdsarc.u-strasbg.fr (130.79.128.5) or via http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/621/A133

all the HST ACS and WFC3 IR available data in the HUDF. The addition of four new bands in WFC3 IR (F105W, F125W, F140W and F160W) has permitted the detection of galaxies out to $z \sim 9-10$. Additionally, the extraordinary depth of the HUDF12 allows the study of galaxy stellar halos with surface brightness profiles down to $\mu_{\rm lim} \sim 31\,{\rm mag\,arcsec^{-2}}$ (Buitrago et al. 2017).

Detecting extended sources in the low-surface brightness regime is an extremely challenging task. Systematic effects such as sky background, persistence, or the PSF may dominate the light distribution of the science images. Aggressive sky background subtraction may be a tempting solution to get rid of the diffuse light, whether it is caused by real astronomical sources or by artificial sky background gradients. Nevertheless, such approach have a major setback. It removes the possibility of using the sciences mosaics to study the outskirts of the largest objects, distorting the photometry of the structures up to $2-3\,{\rm mag\,arcsec^{-2}}$ brighter than the limiting magnitude. Due to this, is common to find over-subtracted zones with negative fluxes around large galaxies in the majority of the surveys (i.e, see the Hyper Suprime-Cam survey[1], Aihara et al. 2018). It is for this reason that careful sky-subtraction is a crucial step to preserve the properties of the low-surface brightness features of extended sources. This effect is clearly visible around the largest objects of the XDF mosaics (Illingworth et al. 2013). We must note that the main objective of the XDF project was not the study of the stellar haloes of the nearest galaxies of the HUDF ($z < 1$) but to identify unresolved sources across a redshift range from $z \sim 4$ to $z \sim 12$ with aperture photometry. Nonetheless, the over subtraction of the diffuse outskirts of nearby galaxies (which cover a large fraction of the total field-of-view) can significantly affect the photometry of the background high-z objects.

In order to overcome the increasing challenge of studying the low surface brightness Universe, a number of recent observational studies has shown the way to reach and surpass the frontier of $\mu_{\rm lim} \sim 30\,{\rm mag\,arcsec^{-2}}$ (see Trujillo & Fliri 2016). The variety of systematic problems that significantly affect the background level of the images depends on the required imaging depth, such as fringing (Wong 2010), ghosts (Yang et al. 2002), gain differences between chip amplifiers, scattered light (Fowler et al. 2017) and even the effect of the point-spread function (PSF, Sandin 2014, 2015). As a consequence of this, the use of robust statistical tools for the study of the outskirts of galaxies is mandatory for these scientific objectives.

The low surface brightness Universe is one of the key fields for cosmology and to unveil the origin and evolution of galaxies. In particular, the processes that give rise to the galaxy discs are not completely clear. Apparent exponential discs (Type-I) on first inspection may on closer examination suffer from a number of deviations such as truncations and down-bending profiles (Type-II) or the opposite phenomenon: anti-truncations (Type-III discs, Erwin et al. 2005, 2007; Pohlen & Trujillo 2006; Gutierrez et al. 2011). There are various scenarios which entail a variety of possible mechanisms probing the variety of galactic discs, such as different types of mergers and gravitational interactions (Laurikainen & Salo 2001; Penarrubia et al. 2006; Younger et al. 2007; Kazantzidis et al. 2009; Borlaff et al. 2014), internal and stellar formation related processes (Roškar et al. 2007; Herpich et al. 2015a,b; Elmegreen & Struck 2016; Struck & Elmegreen 2016, 2017) or the presence of different components on the structure of disc galaxies (Comerón et al. 2012, 2014). Until now, there have been few studies that have tried to study the detailed shape of galaxy discs along significant cosmological times (Azzollini et al. 2008a,b; Trujillo & Bakos 2013; Maltby et al. 2014; Borlaff et al. 2017, 2018). The main reason for this is that PSF-effects and cosmological dimming make difficult to study the outskirts of galaxies at increasing redshifts, where these effects become critical (Sandin 2014, 2015; Trujillo & Fliri 2016; Borlaff et al. 2017). Moreover, while near IR imaging allows us to explore higher z ranges, it includes additional problems such as the high sky background variability or persistence effects, which may contaminate the science images. Consequently, there is an increasing need for very deep observations and surveys as redshift increases.

The outskirts of some nearest galaxies show traces of their formation mechanisms, such as tidal tails, haloes, plumes and satellites. Simulation-based studies predict that a hypothetical survey with a limiting magnitude fainter than $\sim 30\,{\rm mag\,arcsec^{-2}}$ would detect up to a dozen of accretion features around Milky Way-type galaxies (Johnston et al. 2008; Cooper et al. 2009). Apparently isolated galaxies show large tidal tails, warped discs and other asymmetric features in sufficiently deep images, as well as large numbers of satellites (Schweizer & Seitzer 1990; Martinez-Delgado et al. 2008a,b; Chonis et al. 2011). In fact, volume-limited samples of nearby galaxies detect that almost 14% of the galaxies present diffuse features compatible with minor merger events at a limiting magnitude of $28\,{\rm mag\,arcsec^{-2}}$ (Morales et al. 2018). An increasing number of low surface brightness explorations have been performed on individual galaxies and deep fields, but there is an increasing need for telescope-dedicated large-scale low surface brightness surveys (Horton et al. 2016; Valls-Gabaud & MESSIER Collaboration 2017, see The Australian Space Eye and The MESSIER surveyor). Direct imaging of the stellar haloes and their structure is one of the major test for ΛCDM scenario of galaxy formation (Abadi et al. 2005; Bullock & Johnston 2005). Although the study of the streams of the stellar haloes has been tested for a limited sample of galaxies in the Local Group using star counts (Ibata et al. 2009, 2013; McConnachie et al. 2009; Tanaka et al. 2011; Peacock et al. 2014), this method is limited to a maximum distance of 16 Mpc with HST (Zackrisson et al. 2012).

It is for these reasons that the main objective for this paper is to explore the capabilities of HST performing a specific reduction that intends to optimise the limiting depth around extended objects and pave the way for the exploration of the outskirts of galaxies, discs, satellites, tidal streams and their stellar haloes. The proposed techniques are directly applicable to its successor in the infrared – the *James Webb* Space Telescope (JWST) – and similar missions such as EUCLID, for deep integrated photometry of extended sources beyond the Local Universe.

The paper is structured as follows. The full methodology used for the reduction is described in detail in Sect. 2. The results are presented and discussed in Sect. 3. The final conclusions can be found in Sect. 4. We assume a concordance cosmology ($\Omega_{\rm M} = 0.3, \Omega_\Lambda = 0.7, H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, see Spergel et al. 2006). All magnitudes are in the AB system (Oke 1971) unless otherwise noted.

2. Methods

In this section we provide an outline of the reduction process that we have followed to create a version of the HUDF

[1] Hyper Suprime-Cam Subaru Data Release 1: https://hsc-release.mtk.nao.ac.jp/doc/index.php/known-problems-in-dr1/

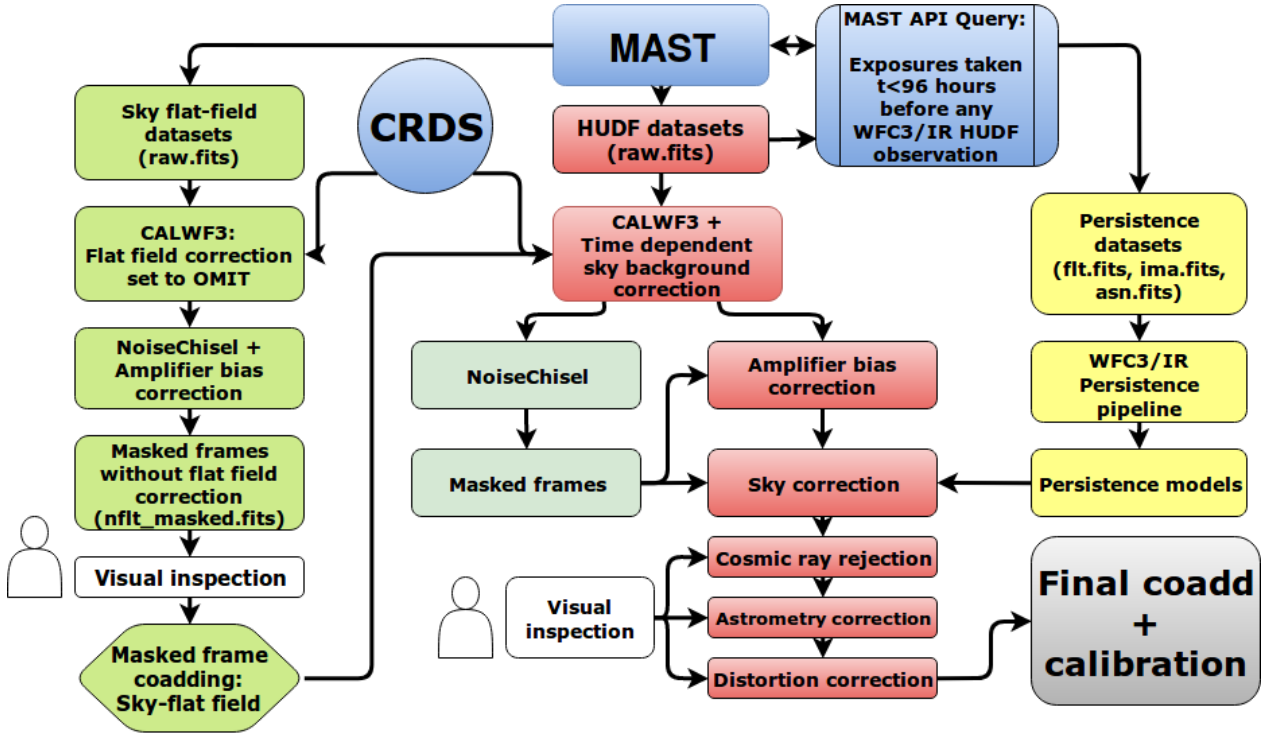

Fig. 1. Flowchart for the ABYSS HUDF reduction pipeline. Each major step in the pipeline discussed in the text is shown. The full process can be divided in three major branches: green steps: sky flat field creation. Yellow steps: modelling of the extended persistence arrays. Red steps: main reduction of the HUDF exposures. Blue steps: required queries to the MAST or the CRDS database.

WFC3 IR mosaics optimised for the study of the low surface brightness Universe (the ABYSS[2] pipeline, hereafter). The full process (represented in the flowchart of Fig. 1) can be divided into three main branches:

- Creation of sky flat fields for the four filters. This process is fully described in Sect. 2.4.
- Creation of a catalogue of all WFC3 IR datasets that may affect our mosaics (including calibration exposures) to generate a set of improved persistence models for each exposure of the HUDF. We detail this process in Sect. 2.5.
- Download and reduction of all the WFC3 IR datasets that include observations using the F105W, F125W, F140W and F160W filters on the HUDF.

For each individual exposure, we conduct the following steps. We refer the reader to the corresponding sections for a detailed description:

1. Preliminary calibration (bias, dark, flat field), including a time variation sky background correction (see Sects. 2.1 and 2.6).
2. Create masked frames for each exposure (Sects. 2.2 and 2.7).
3. Perform amplifier level correction (Sect. 2.3).
4. Perform sky correction for each readout, using the individual masks and robust statistical estimators (Sect. 2.8).
5. Perform cosmic rays rejection and alignment of each exposure (Sect. 2.9).
6. Transformation of the corrected frames into a geometric distortion corrected frame, combination and calibration of the final mosaic (Sect. 2.10).

[2] ABYSS: a low surface brightness dedicated reduction for the HUDF WFC3 IR mosaics: `http://www.iac.es/proyecto/abyss/`

2.1. Initial data and preliminary calibration

For the present work, we started downloading all the individual exposures of the WFC3 (symbolised here by `raw.fits`) within $r < 3$ arcmin radius of the original HUDF centre coordinates ($\alpha = 3\mathrm{h}\ 32\mathrm{m}\ 39.0\,\mathrm{s}$, $\delta = -27^\circ\ 47'\ 29.1''$, J2000). In order to do that, we make use of The Barbara A. Mikulski Archive for Space Telescopes (MAST)[3]. MAST allows us to perform queries with multiple constraints, such as sky-coordinates (α, δ), telescope, instrument, detector, filter, proposal ID and observation date. Given the large number of files to download, we used the MAST Python API[4]. The MAST API allows us for programmatic queries and implement them easily into any Python code. We included MAST API subroutines into our pipeline to automatically download required files from MAST when needed.

Subsequently, we made use of the HST Calibration Reference Data System[5] (CRDS, Greenfield & Miller 2016) to download the best reference files available for each individual exposure of the HUDF. The CRDS is publicly available and can be installed as part of AstroConda[6]. For HST, CRDS has a command line tool that assigns and automatically downloads the best reference files to the FITS headers of the `raw.fits`

[3] MAST is managed by Space Telescope Science Institute (STScI) and is publicly available at `https://mast.stsci.edu/portal/Mashup/Clients/Mast/Portal.html`

[4] MAST Python API: `https://mast.stsci.edu/api/v0/`

[5] HST Calibration Reference Data System: `https://hst-crds.stsci.edu/`

[6] AstroConda is a free Conda channel maintained by the Space Telescope Science Institute (STScI). It provides tools and utilities required to process and analyse data from the *Hubble* Space Telescope (HST), *James Webb* Space Telescope (JWST), and others: `http://AstroConda.readthedocs.io/`

files. We processed all our raw.fits files using the Space Telescope Science Data Analysis System (STSDAS) task calwf3. calwf3 is available as part of the stsci_python package on AstroConda, and it corrects for instrumental effects and generates calibrated frames. In addition, calwf3 can process multiple readouts for the same exposure to create improved data products.

For WFC3 IR observations, it is possible to sample the signal multiple times as an exposure accumulates, before the end of the exposure. This allows to 1) to record the signal of a pixel before it saturates, 2) to perform a better cosmic-ray rejection and 3) to reduce the net effective read noise. This observation mode is named MULTIACCUM, and is the default observation mode for WFC3 IR. Each raw.fits file contains the information of each readout (up to 16) of the chip during a certain exposure. The individual processing steps that calwf3 performs for each readout are:

1. Flagging of known bad pixels in the data quality (DQ) array.
2. Identification of pixels in the initial read that contain detectable source signal.
3. Subtraction of bias drifts using the reference pixels.
4. Subtraction of the zeroth (first) read.
5. Estimation of the noise model for each pixel and record in the error (ERR) array.
6. Photometric non-linearity and saturated pixels correction.
7. Subtraction of dark image.
8. Calculation of the photometric header keyword values for flux conversion.
9. Conversion of the data from counts to count rates.

After this, calwf3 uses each readout of the MULTIACCUM mode to create a single image for each individual exposure. This is done by analysing the count differences between readout as a function of time (up-the-ramp fitting). This new array represents the best-fit count rate for each pixel. Finally, the pipeline performs flat fielding and gain conversion (transforming from counts to count-rates). The final result is an flt.fits file, which contains the science best-fit count-rate array (SCI), the error array (ERR), a data-quality array (which flags bad-pixels and cosmic rays), the number of samples (SAMP) and integration time (TIME) arrays. Each one of the calwf3 steps can be omitted or performed by switching their corresponding keywords in the headers of the raw.fits files of each exposure. For those frames that are going to be used here to calculate our dedicated sky-flat, we set the corresponding keyword (FLATCORR) to "OMIT" prior to their processing.

2.2. Image masking

In order to create the sky flat field models, estimate the sky level of each exposure and obtain the final catalogue, we require an accurate masking of the individual sources in the images. For this task we used Gnuastro's[7] program (version 0.5) for detection and segmentation: NoiseChisel (Akhlaghi & Ichikawa 2015). NoiseChisel is a recently developed free software to detect signal in astronomical images based on erosion of pixels. This method is non-parametric, allowing to efficiently detect astronomical objects with irregular morphologies that are immersed in noise. In addition to this, NoiseChisel algorithm is based on the properties of the image noise (not only on those of the signal dominated pixels) reducing the input from the user and thus it is extremely robust against the sky level value or the properties of the sources to mask, making it an excellent choice for highly different exposures, such as the ones analysed in this paper. NoiseChisel accepts FITS files as input, and returns a multi-extension FITS file that contains: 1) the input image, 2) the segmentation map with the detection labels for each object, 3) a secondary segmentation map for the clumps inside each detection, 4) NoiseChisel estimation of the final sky value on each pixel, and 5) the standard deviation for each pixel. The final results are based on version 0.5 of Gnuastro. The masking of each individual exposure is performed as follows:

1. In order to properly detect sources in the input images with NoiseChisel they have to be flat field corrected. Otherwise, we would not have a similar sensitivity across the detector, and our results would be biased. We perform preliminary calibration of the raw.fits files, including flat field calibration. During the creation of our own sky flat fields, we have to correct the images before masking. For this step, we use the official MAST flat fields. After the sky flat fields are created, if the input image is a HUDF exposure, we use our own sky-flat fields.
2. We create the segmentation maps for each flt.fits image using NoiseChisel with default configuration (tile size equal to 50×50 pixels).
3. If the images are being masked for the generation of the sky flat fields, we multiply back the images by its corresponding official MAST flat fields. The result of this process is a properly masked flat field uncorrected exposure.

In Fig. 2 we show an example of the masking process of a F160W exposure, included in the HUDF field (ib5x2elbq, PID 11563). The top left panel represents the flt.fits file after preliminary calibration through calwf3, but before flat field correction (see Sect. 2.1). The top right panel shows the same flt.fits file after flat field correction. After this step, we analyse the flat field corrected flt.fits image with NoiseChisel, which produces a segmentation map for each exposure. We identify all the pixels that are included as part of a source obtaining a masked, flat field uncorrected flt.fits file, ready to be combined for the creation of a sky flat field. We show this final masked frame in the bottom-left panel of Fig. 2. In order to illustrate the accuracy of the segmentation map results, we compare in the bottom panels the NoiseChisel mask with the flt.fits image convolved by a Gaussian kernel with $\sigma = 5$ pixels. It is clearly visible that this process accurately masks objects with very different ranges of size, even if part of them (or the whole object) is not visible in the non-convolved frame. We refer the reader to Sect. 2.8 for a systematic benchmark analysis of the masking procedure effect on the sky background subtraction and its comparison with other methods.

2.3. Amplifier relative gain correction

The WFC3 IR detector contains 1024×1024 pixels, which are divided into four quadrants of 512×512 pixels. There is a border of 5 non-illuminated pixels around the edges of the detector, which are used to provide constant-voltage reference values for the detector. Due to this, the light exposed area of the detector includes 1014×1014, divided in four 507×507 pixels. Although the default pipeline corrects for differences in gain between the four different sections of the WFC3 IR detector, it is common to see residual differences in the flat field corrected images and specially in the combined mosaics. The amplitude and sign of these differences between amplifiers may vary from dataset to dataset,

[7] GNU Astronomy Utilities (ascl.net/1801.009) project is part of the Free Software Foundation and is freely available at: https://www.gnu.org/software/gnuastro/

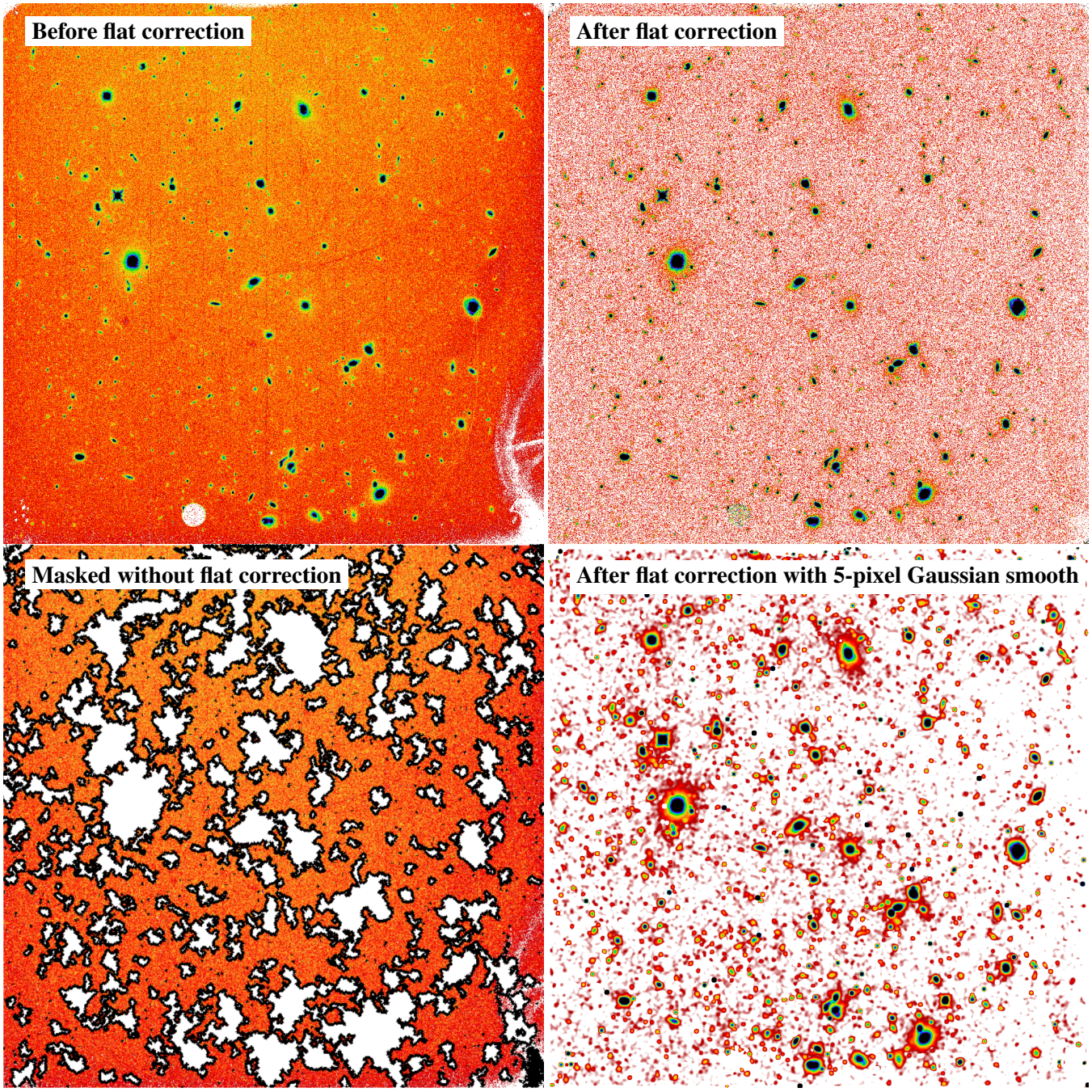

Fig. 2. Example of the masking process. *Top left panel*: pre-calibrated frame, without flat field correction. *Top right panel*: pre-calibrated frame, with flat field correction applied. *Bottom left panel*: masked frame without flat field correction. *Bottom right panel*: flat field corrected frame with Gaussian smoothing. Notice that NoiseChisel (Akhlaghi & Ichikawa 2015) efficiently detects objects with very different shapes and sizes, masking the outskirts and diffuse regions that are barely visible even in the convolved frame.

and can produce significant effects in both the final images and the sky flat fields.

In order to correct for this effect we have to calculate the median differences in flux between each section of the chip and compensate them. We cannot simply use the median values for the full area covered by each section, because that would result in a wrong estimate of the median bias if there are large scale gradients across the detector. Instead, we calculate the ratios between columns or rows of pixels at equal distances to the frontier between chips. Then we estimate the weighted median of the resulting array of ratios, where the weights are equal to the inverse distance of each pixel to the frontier of the chip. By doing this, we ensure that the differences between pixels which contribute the most are the closest ones to the frontier, but we still make use of all the information available in the image. We set as reference the A section of the chip (top left) and correct the remaining sections (B, C and D) to it. The weighted sky-levels for each frontier between chips are labelled with two letters, being the first letter the one to which the pixels of the sample belong (see Fig. 3). The correction factors are:

$$\delta AB = < \frac{\overrightarrow{AB}}{\overleftarrow{BA}} >, \quad (1)$$

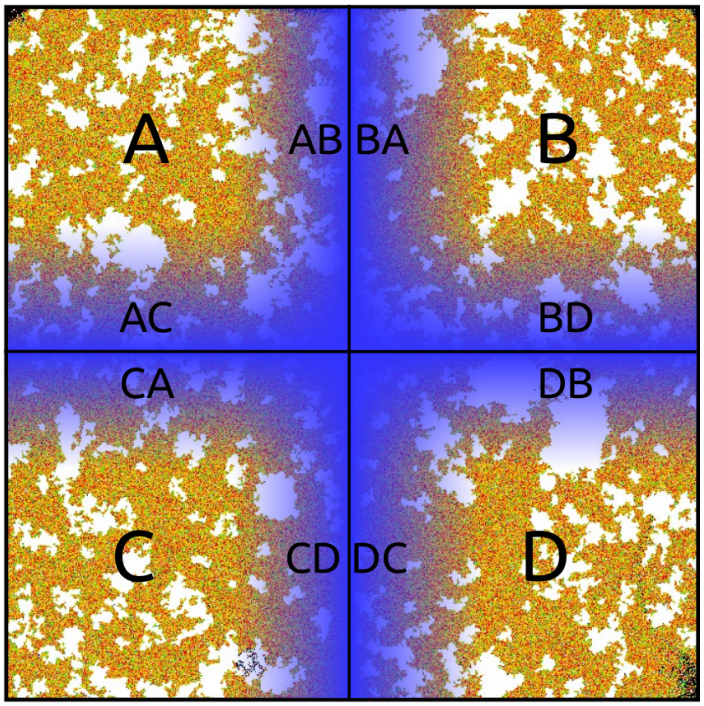

Fig. 3. Amplifier gain correction diagram. We label each one of the four 507 × 507 pixel sections of the WFC3 IR detector as A, B, C, D (see figure). We represent with a blue transparent gradient the weights (bluer represents higher weights) applied to the corresponding pixels of each region, labelled accordingly. The white regions in the background are masked pixels, either because they are part of a light source or are affected by persistence. We do not include those pixels in the amplifier gain correction estimation.

$$\delta AC = < \frac{\overrightarrow{AC}}{\overleftarrow{CA}} >, \tag{2}$$

$$\delta CD = < \frac{\overrightarrow{CD}}{\overleftarrow{DC}} >, \tag{3}$$

$$\delta BD = < \frac{\overrightarrow{BD}}{\overleftarrow{DB}} >, \tag{4}$$

where the $\rightarrow$ and $\leftarrow$ mean the original and axis reversed arrays, respectively, and the <> represents the weighted mean by the distance of each pixel to the frontier between chips. By using these, we can calculate the δAD correction factor as follows:

$$\delta AD1 = \delta AB \cdot \delta BD, \tag{5}$$

$$\delta AD2 = \delta AC \cdot \delta CD, \tag{6}$$

$$\delta AD = \frac{\delta AD1 + \delta AD2}{2}, \tag{7}$$

In addition, as in the case of sky-correction, we must avoid contamination by the objects in the field of view and persistence. In order to avoid such effects, we masked all the pixels flagged as part of sources by NoiseChisel (see Sect. 2.2). Additionally, for the HUDF field images, we also flagged all those pixels affected by persistence according to our custom-improved models (see Sect. 2.5). Finally, we applied this amplifier gain correction to all the exposures used in this study before the final sky-subtraction.

2.4. Sky flat fielding

In order to measure the relative sensitivity of the pixels of a detector (flat field), the optimal process would be to observe a uniform external source of light. Although this is certainly not possible in most cases, there are several strategies to reproduce these conditions or to compensate for any possible inhomogeneities in the illumination. In ground-based observations, the combination of several out-of focus dome images is the easiest way to estimate the flat field. Observations of the twilight sky at the start and the end of the night (twilight flats) are a reasonable alternative to avoid the possible inhomogeneities and gradients caused by the dome of the telescope. However, neither dome flats nor twilight flats are free of gradients due to non-perfect flat illumination, and the latter suffer from time variation of the sky background level. In space, twilight flats are possible using dark Earth limb observations (see e.g., HST Cycle 17 WFC3 calibration proposal 11917) or the moonlit illuminated earth (Bohlin et al. 2008). For WFC3, high signal-to noise (S/N) ground-based flat fields (Bushouse 2008) were generated prior to launch in the CASTLE HST simulator (hereafter CASTLE LP-flats). Nevertheless, the CASTLE simulator is not a perfect replica of the optical path and conditions of HST. Observations of the star clusters Omega-Cen and 47 Tuc (proposals CAL-11453 and CAL-11928) demonstrated that the CASTLE LP-flats were not able to correct the large scale structure in WFC3 IR channel (Pirzkal et al. 2011).

An alternative to these strategies is to take advantage of the observations previously accumulated. Sky background in the near IR is high enough, even in space, to use it as a uniform source of light. However, diffuse objects and the extended point spread function can introduce severe biases in the median images. It is because of this, that careful masking of any source in the field-of view is necessary to avoid contamination in the final calibration images. In addition to this, not all the images are suitable for this analysis, because the flux of the sky background is relatively low (~0.3−1.0 $e^- s^{-1} pix^{-1}$, see Pirzkal et al. 2011). Thus, creating accurate sky-flat fields requires the combination of a large number of exposures. The standard method to create sky-flats from observations is to calculate the median of the masked science images. In ground-based observations, a master-flat per night is created by using the corresponding normalised science images of that night. The reason of this is to avoid any unwanted effects due to slight changes of focus, weather conditions, or vignetting. The high stability of space telescopes compared to the ground-based ones permits the making of sky-flat fields using images that have been taken over longer time periods. Moreover, the use of sky-flat fields have the advantage that they are measuring the relative sensitivity between pixels at the same intensity level of the images that have to be corrected. This accounts for any possible spatial differences in the linearity across the detector as a function of the input intensity.

In Pirzkal et al. (2011), the authors used the observations taken between 2009 and 2010 with WFC3 IR to create a second order correction to the CASTLE LP-flats (delta sky flat or SD-flats). In order to do that, they first identified a large number of datasets with exposure times longer than 300 s. Secondly, they aggresively masked all the sources in the images with SExtractor (Bertin & Arnouts 1996) in the field of view. This step was repeated for several hundreds of images per filter. Finally the masks were normalised and combined to the CASTLE-LP flats, creating the new SD-flats. Although the SD-flats for the F160W included nearly a thousand masked images, some of the other filters did not have enough data to create a reliable flat field, which was the case for F105W, F110W and F140W, where

Table 1. Summary of the exposures used for the sky flat field analysis.

Sky-flat (1)	Period (2)	No. of exposures (3)	Proposal IDs (4)
F105W A	55000–56500	719	11563, 11584, 11738, 12060, 12065, 12067, 12068, 12069, 12099, 12100, 12101, 12102, 12103, 12104, 12184, 12286, 12442, 12451, 12452, 12453, 12454, 12455, 12456, 12457, 12458, 12459, 12460, 12461, 12496, 12498, 12553, 12590, 12787, 12788, 12789, 12790, 12791, 12949, 13063
F105W B	56500–58000	747	13386, 13420, 13459, 13495, 13496, 13641, 13677, 13687, 13718, 13767, 13779, 13790, 13792, 14037, 14038, 14096, 14122, 14227, 14327, 14808
F125W	55000–56500	1811	11144, 11149, 11189, 11359, 11520, 11557, 11563, 11678, 11700, 11702, 12025, 12028, 12036, 12060, 12061, 12062, 12063, 12064, 12065, 12066, 12067, 12068, 12069, 12099, 12100, 12101, 12102, 12103, 12265, 12286, 12329, 12440, 12443, 12444, 12445, 12451, 12452, 12453, 12454, 12459, 12460, 12461, 12572, 12590, 12616, 12960, 13063
F140W	55000–56500	875	11359, 11600, 11696, 12067, 12068, 12099, 12100, 12101, 12102, 12103, 12166, 12177, 12190, 12203, 12217, 12328, 12330, 12452, 12458, 12459, 12461, 12471, 12498, 12544, 12547, 12568, 12726, 12896, 13063
F160W	55000–56500	1727	11142, 11149, 11189, 11359, 11520, 11563, 11584, 11647, 11663, 11694, 11696, 11700, 11702, 11735, 11738, 11838, 11840, 12028, 12036, 12055, 12060, 12061, 12062, 12063, 12064, 12065, 12066, 12067, 12068, 12069, 12072, 12075, 12099, 12100, 12101, 12102, 12104, 12167, 12194, 12195, 12197, 12224, 12265, 12267, 12283, 12286, 12292, 12307, 12329, 12378, 12440, 12443, 12444, 12445, 12447, 12451, 12452, 12453, 12454, 12459, 12461, 12498, 12502, 12590, 12613, 12616, 12686, 12709, 12764, 12866, 12990, 13063

Notes. Columns: (1) Flat field identifier. (2) Exposure date limits used for the selection of the datasets for each flat field. (3) Number of exposures used for each flat field. (4) Proposal IDs of the exposures used. We remark that not all the images of a given proposal ID were used in its corresponding flat field. A careful visual inspection of each one of the images was carried out by the authors, before image combination.

only ∼100 datasets were used. For these filters, the S/N was too low and the masking process left no available data in some regions of the detector, creating holes were the corresponding SD-flat had no data. Finally, in order to reach a reliable final SD-flat solution, the authors opted to combine the ∼2000 datasets from all the filters into one SD-flat (grey SD-flat, hereafter). In addition to this, they smoothed the grey SD-flat by using a σ = 10 pixel kernel before combination with the CASTLE LP-flats. The authors did not find any significant dependence on the filter or variation of the flat fields with time between 2009 and 2010. They successfully tested the improved flat field with 32 F160W datasets from the HUDF proposal 11563, removing a large scale cross-like structure in the background. These combined CASTLE LP-flats + SD-flat fields were included into the calibration database system in December 7th, 2010, until present date.

In this paper, we follow a similar procedure as the presented in Pirzkal et al. (2011). We generate a sky flat field per filter by combination of exposures between 55000 (June 2009) to 56500 (July 2013). Additionally, we create a second flat for the F105W filter considering observations between 56500 and 58000 (April 2017). Due to the small size of the dithering pattern, and the fact that all images were taken with very similar rotation angles, we cannot create a sky-flat field just by using the exposures from the HUDF observation programmes. To tackle this problem, we selected the deep observations as the AEGIS, COSMOS, GOODS-N, GOODS-S, and UDS fields. We also included multiple exposures from other fields that were taken in the 96 h before the HUDF exposures. We list the amount of exposures per filter, time-period, and the proposal IDs of all the images used for each flat field in Table 1.

In Fig. 4 we present an example of our results. In the left panel we show our absolute sky-flat field for the F160W filter, according to the robust median of the masked datasets from Table 1. In the right panel we show the ratio between our sky flat field and its corresponding SD-flat from MAST. Notice that both panels have different colour scales, and that the right panel has been convolved with a 3 pixel wide Gaussian kernel in order to enhance the differences between flats. We do not find differences larger than ∼1%, but there is a significant coherent large scale variation, in addition to residuals on the bottom-right corner of the detector (wagon-wheel).

2.5. Persistence correction

A known effect that affects HgCdTe IR array detectors (as is the case of the WFC3 IR) is persistence. Persistence shows up as an afterglow on the pixels that were exposed to a bright source of light in a previous exposure. This charges arise from imperfections in the photo-diodes of IR detectors (Smith et al. 2008a,b). Neither non-destructive read-out nor resets can change significantly the rate at which this charge is released. These bright sources then re-appear in the following exposures as ghost images in the same regions of the detector. The intensity of this effect decays with time and eventually becomes negligible. Even so, persistence is an important effect to take into account as it may create false detections in science images. The effect of persistence and thus the accuracy of its correction becomes more challenging as we move towards lower surface brightness ranges.

Persistence is an intrinsic effect of the WFC3 IR detector and thus it cannot be avoided, but it can be partially corrected

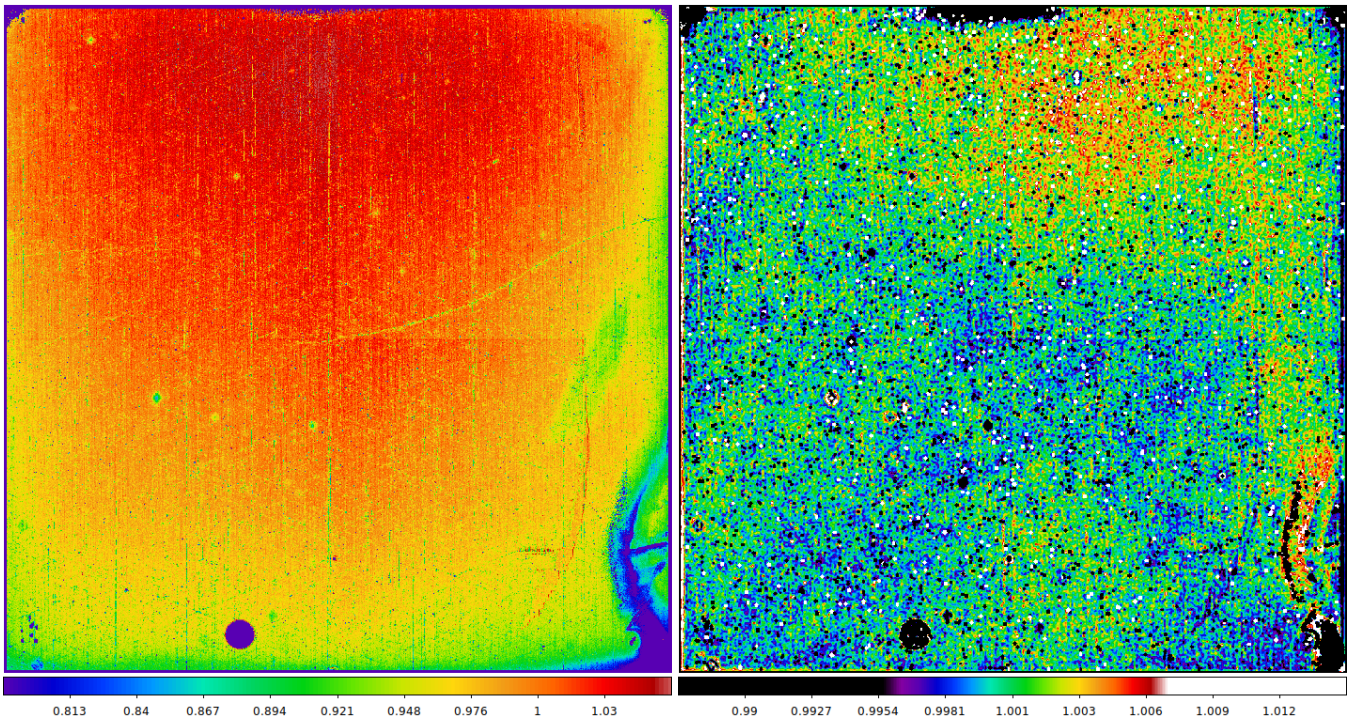

Fig. 4. Sky flat field analysis for the F160W filter. *Left panel*: absolute sky flat field measured for the June 2009 to July 2013 period, using the datasets corresponding to the proposal IDs listed in Table 1. *Right panel*: ratio of our sky flat field for the F160W to its corresponding MAST SD flat field `uc721145i`. Notice the different colour scale, spanning only a ∼1% for the flat field ratio panel. The *right panel* was smoothed with a 3 pixel wide Gaussian kernel to enhance the differences between different regions.

afterwards. STScI MAST WFC3 Persistence Project[8] provides the necessary tools to check if a certain exposure was affected by persistence and a complete set of models to correct the effects. The current method of persistence correction of WFC3 IR consists in modelling the number of electrons that would be created by persistence in each pixel by all the previous exposures (up to a certain time) that were taken before the one to correct (Long et al. 2012). These models are pre-calculated for all WFC3 IR exposures and publicly available through MAST. Nevertheless, these models only take into account the exposures that were taken up to 16 h before for the creation of the persistence model (Dr. Knox Long, priv. comm.).

In Long et al. (2015) the authors published an improved pipeline to create persistence models for WFC3 IR. The initial models predicted persistence only based on the observed flux and the time between exposures (Long et al. 2012). Nevertheless, in Long et al. (2013a,b), the authors carried out a set of experiments where they demonstrated that the persistence depended also with the amount of exposures taken and the time that a pixel remained filled with charge. Based on these findings, the authors developed a more accurate prediction of persistence in the IR channel of WFC3. This new pipeline was included into the set of software tools used to estimate the persistence in all WFC3 IR images and also was made publicly available[9].

In this paper, we make use of the last update to date (git commit hash b0b9cbeaf7, 26 Jan 2016) of the prediction software made by STScI in order to create a set of dedicated persistence models for each individual exposure of the HUDF. We increase the lookback time from 16 h to 96 h to take into account the longest time that our computer resources permit us to create the persistence models, using the default observational "A-gamma" model (Long et al. 2015). To do that, we downloaded all the exposures taken with WFC3 IR 96 h before each exposure of the HUDF. Finally, we run the persistence pipeline for each exposure in the HUDF field. The software creates the persistence models and automatically stores them in their corresponding FITS files.

We show two examples of the improvement of the persistence models in Fig. 5. The left column shows the improved persistence models for the exposures of the HUDF `ib5x22b8q` and `icxt25byq`. The right column shows their corresponding MAST official persistence models taking into account the previous 16 h. In the first case `ib5x22b8q`, the persistence model (top-left panel) is dominated by a large gradient produced by a set of calibration observations (Cycle 17, CAL WFC3 category, 11915 – IR Internal Flat Fields, PI: Bryan Hilbert)[10]. These observations were meant to create a new set of flat fields for WFC3 IR using the internal tungsten flat field lamp to illuminate the detector. As a consequence of this, the camera was flooded with persistence just 18 h before the observations of the HUDF with the F105W filter. Unfortunately, the results of the proposal 11915 were never published (Ben Sunnquist, WFC3 Help Desk, priv. comm.). Many of the HUDF exposures of the F105W filter were left unusable because of this calibration program, since we are interested on the lowest surface brightness limit, and we cannot rely on the persistence correction to fully correct the images. This effect does not appear in the official MAST persistence arrays (top-right panel), due to the 16 h limit for the lookback calculation of persistence, and thus, were not taken into account in the previous versions of the HUDF.

[8] WFC3 Persistence Project: https://archive.stsci.edu/prepds/persist/

[9] STScI Persistence repository: https://github.com/kslong/Persistence/wiki

[10] A detailed overview of proposal 11915 can be found in: http://www.stsci.edu/hst/phase2-public/11915.pdf

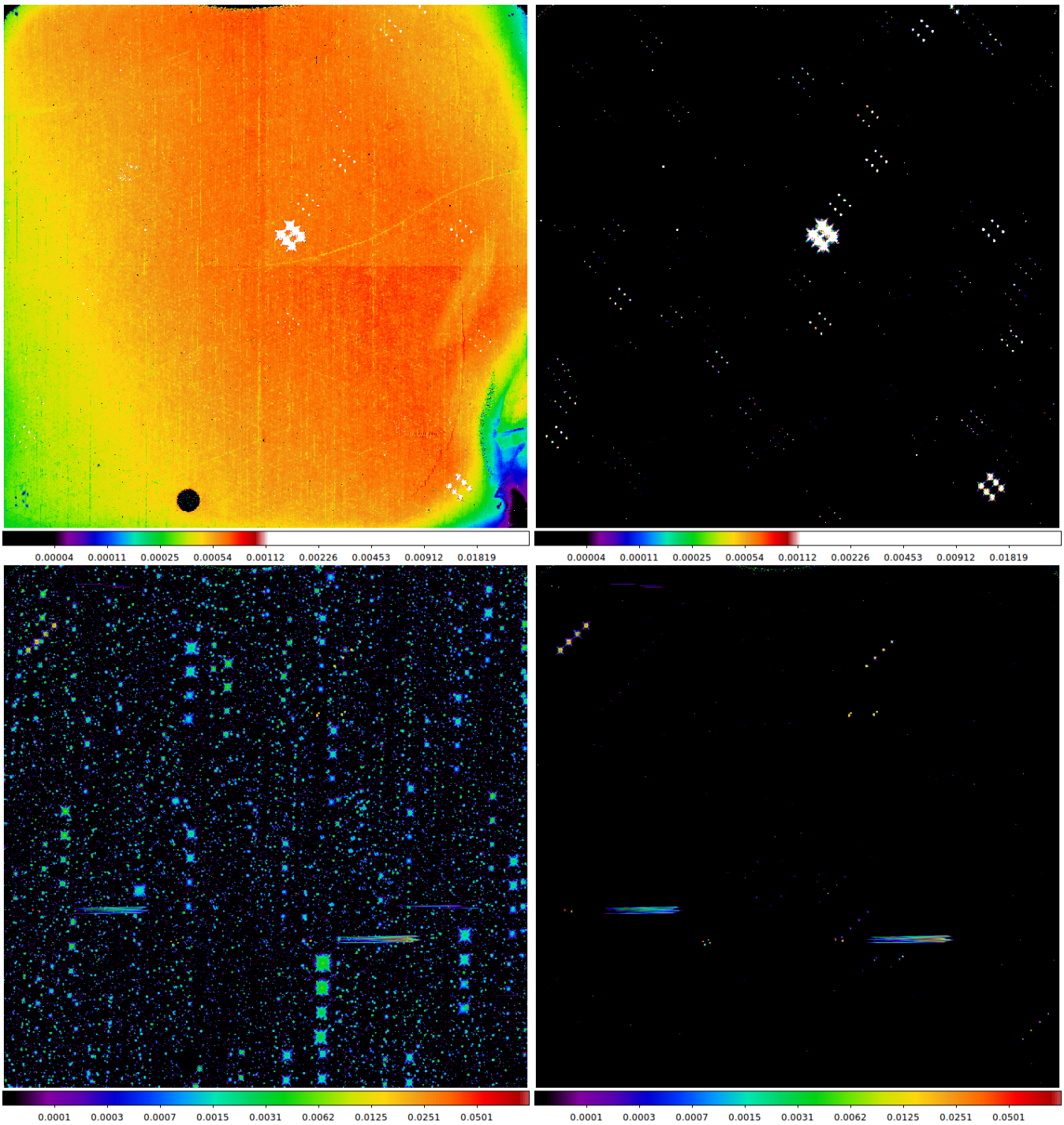

Fig. 5. Comparison between the WFC3 IR persistence models calculated taking into account 96 h of previous observations (*left column*) and the persistence models available at MAST, with a limit of 16 h of lookback time (*right column*). *Top panels*: persistence models for dataset `ib5x22b8q`. *Bottom panels*: persistence models for dataset `icxt25byq`. The images corresponding to the same dataset are at the same colour scale. Note on the *top left panel* a background of persistence covering the complete field of view caused by the flat calibration lamp in the previous hours to the observation of the HUDF, not detected on the persistence model available at MAST (*top right panel*).

Nevertheless, large scale gradients can hardly affect the results for high-z, unresolved objects. In that case, the observers must pay more attention to flag any point-like source of light that may be caused by persistence. In the bottom panels of Fig. 5 we show another example (`icxt25byq`) of the improvement of our persistence models. We identify two major causes of the persistence contamination, proposals 14074 (Opening the Window on Galaxy Assembly: Ages and Structural Parameters of Globular Clusters Towards the Galactic Bulge, PI: Roger Cohen) and the WFC3 G102 grism observations of proposal 14227 (Cycle 23 proposal 14227: The CANDELS Lyman-alpha Emission At Reionization Experiment, PI: Casey Papovich), which ended just ∼3 min before the start the HUDF exposure `icxt25byq`. The former one (which ended 40 h before the HUDF exposure, hence, was not taken into account for persistence) is responsible for the star-like sources

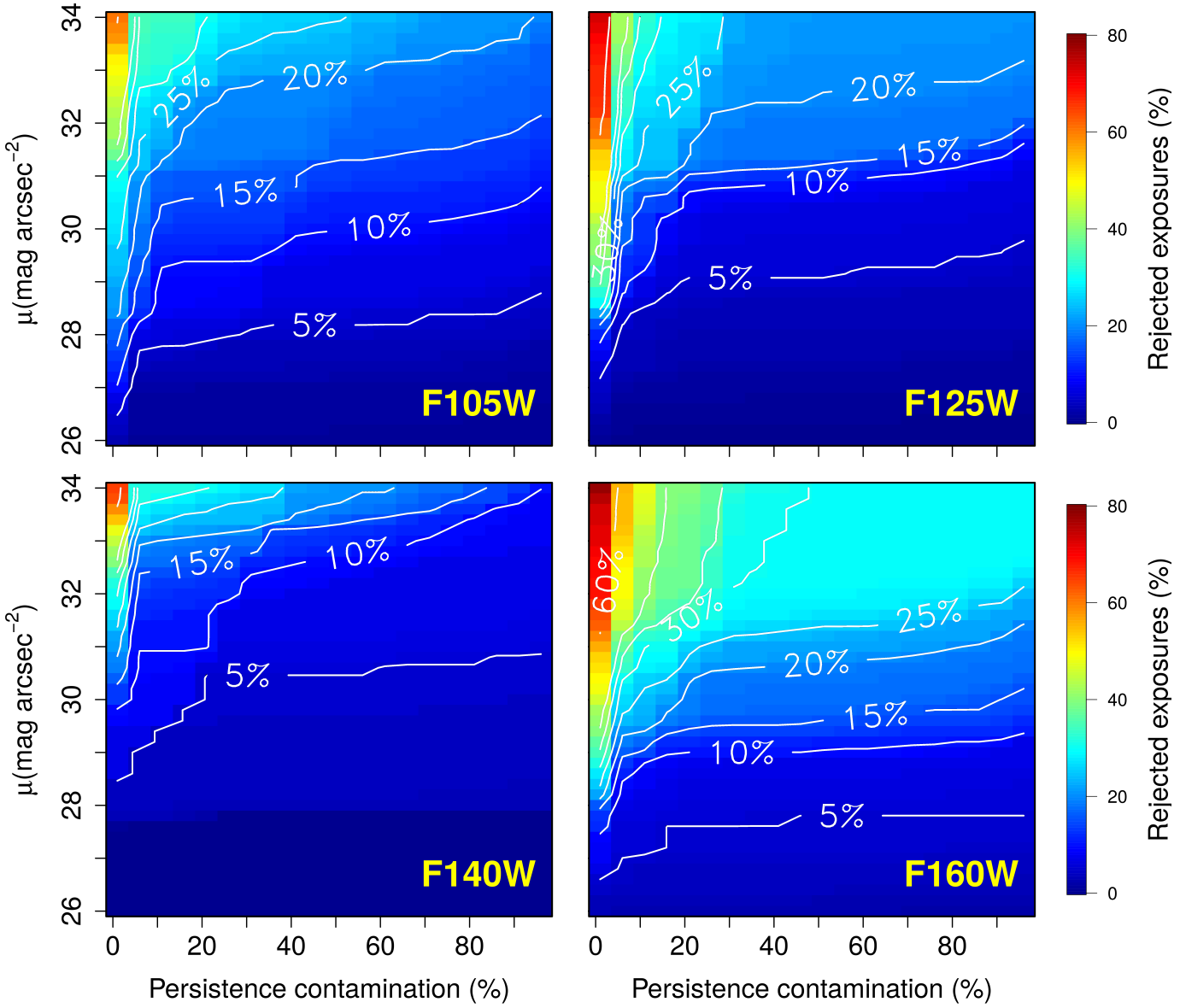

Fig. 6. Analysis of persistence contamination on the WFC3 IR HUDF exposures. In colour bins and white labelled contours, we represent the fraction of exposures that present a certain level of surface brightness contamination by persistence or higher (vertical axis) as a function of the fraction of the total amount of pixels that are affected (horizontal axis), for the F105W (*top left*), F125W (*top right*), F140W (*bottom left*) and F160W (*bottom right*) bands.

across the detector, and the latter one for the two elongated sources.

In Fig. 6 we analyse the persistence contamination level of the WFC3 IR exposures of the HUDF. We measured the amount of pixels that were affected by persistence as a function of the surface brightness contamination that this effect creates. The expected surface brightness level created only by persistence is estimated using the persistence flux predicted by the models. We then determine the fraction of each image that is affected as a function of the surface brightness persistence contamination level. We found that the F160W and F105W are the filters most affected by this issue. For F160W band, ∼17% of the images present at least half of the image contaminated at surface brightness of $\mu = 30\,\mathrm{mag\,arcsec^{-2}}$. The filter less affected by persistence is F140W, with less than ∼5% of the images with that level of contamination on at least half of the image. Given that the persistence models are an approximation to the real level of persistence, and they do not fully correct for all the contamination, the general recommendation by STScI is to use them to correct the images and flag those pixels affected up to a certain level, that has to be chosen depending on the science target. After correcting all the images with our improved persistence models, we chose a conservative compromise between losing exposure time and possible contamination by the persistence residuals. We reject all the images that present more than 50% of the pixels affected at surface brightness brighter than $\mu = 30\,\mathrm{mag\,arcsec^{-2}}$ (this removes a 10%, 8%, 5% and 17% of the total amount of exposures from the F105W, F125W, F140W and F160W filter respectively). Finally, we flag any pixel in the valid images that presents a persistence surface brightness larger than $\mu = 30\,\mathrm{mag\,arcsec^{-2}}$.

We conclude that our persistence models are more robust and include many potential persistence sources that were not taken into account in the previous versions of the HUDF mosaics. The

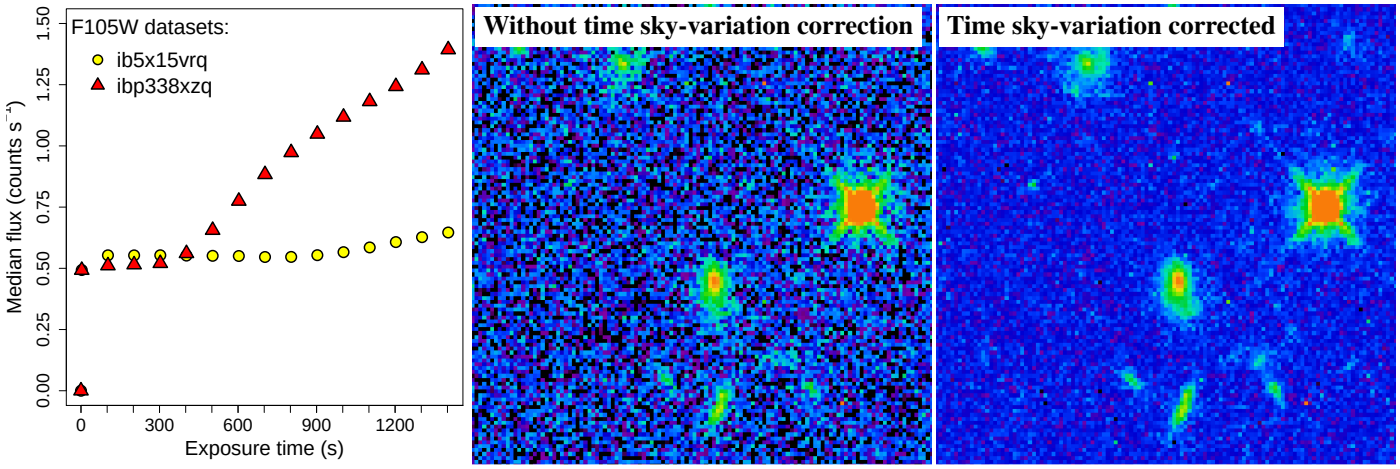

Fig. 7. Example of the effects of time varying sky background in two different datasets of WFC3 IR in the F105W band. *Left panel*: median background flux of `ib5x15vrq` (yellow circles) and `ibp338xzq` (red triangles). While the dataset `ib5x15vrq` shows a nearly constant flux on all the different readouts (∼0.50 counts s^{-1}), `ibp338xzq` is clearly affected by time variation of the sky background, increasing from ∼0.50 counts s^{-1} to ∼1.3 counts s^{-1} after 400 s from the start of the exposure. *Middle panel*: sky-corrected `ibp338xzq flt.fits` image without time-varying sky background correction applied. *Right panel*: sky-corrected `ibp338xzq flt.fits` image accounting for the time-varying sky background correction. Both images are set to the same colour scale.

reason for this is that we used a four times larger lookback time for the persistence calculation than the official MAST persistence models and the last WFC3 Persistence pipeline as presented in Long et al. (2015).

2.6. Time-dependent sky background variation

During long exposures, sky background can vary noticeably, introducing a non-linear component to the count rates calculated by `calwf3`. This causes non-Gaussian properties in the `flt.fits` frames of many exposures, with a severe impact in the depth of the final product. This problematic effect is specially common in F105W filter observations. The reason for this is the presence of atmospheric He 10 830 Å emission line in the upper atmosphere (Brammer et al. 2014). This emission line falls into the F105W and F110W filters and both HST WFC3 IR grisms. The effect is strongest at low Earth limb angles and under direct sunlight, and usually negligible in the Earth's shadow, but sometimes can be strong even when observing at 40° above the Earth limb. In the worst case scenario, the He 10 830 Å line fully dominates the sky background emission (see Dressel 2012, Fig. 7.13). Observations with long exposure times will approach closer to the Earth's limb and will be potentially more contaminated by Earth's scattered light and atmospheric emission.

The pixels of those exposures affected by time-dependent sky background will be wrongly classified by `calwf3` as cosmic rays, thus impeding any type of alignment (see Sect. 2.9). In order to correct for this effect, we follow a similar procedure as in the Sect. 3.2 of Koekemoer et al. (2012). We take advantage of the flexibility of `calwf3` to stop and re-start the calibration process at any point, and we subtract the sky background of each independent readout for each exposure (typically 16 readout per exposure), before the cosmic ray rejection process (steps 7 and 8 of Sect. 2.1). We perform this correction in four steps:

1. We run `calwf3`, stopping the procedure before the "up-the-ramp" fitting and the cosmic rays identification (Sect. 2.1).
2. We individually estimate and subtract the sky background emission from each readout of the intermediate `ima.fits` files.
3. We resume `calwf3`, obtaining a first approximation to the `flt.fits` file.
4. We use this first `flt.fits` combined image to create an object mask, using NoiseChisel (see Sect. 2.2).
5. We run again steps 1–3, using this mask to have a better determination of the sky background on each readout. The output is the final `flt.fits` file.

The final sky-background level for each exposure is calculated as the median sky value of the half of the readouts with less differences between them (first or second half). This constant sky level will be subtracted in a later step. By doing this, we successfully corrected most of the exposures affected by this effect. We must note that this correction is only valid if the time variation of sky background is flat, this is, it does not include any variation of a large scale gradient across the image. We visually inspected each dataset and rejected those exposures that were clearly affected by sky background gradients. We show in Fig. 7 an example of an exposure affected by time variation of the sky background. The left panel shows the median flux for two different exposures as a function of the readout time. In an ideal case, the background count rate should be zero or at least constant, but exposure `ibp338xzq` shows a clear increase on the sky background flux starting at ∼400 s from the start of the exposure. Running the standard `calwf3` pipeline leads to the image shown in the central panel of Fig. 7, where the sky background has a high-noise level which is not well represented by a Gaussian approximation. We present in the right panel the same exposure corrected by the procedure described above. We stress that both images are calculated from the same `raw.fits` file, and are represented with the same colour scale.

In conclusion, this procedure allows us to successfully correct for the effect of time-variation of the sky background, which is a common issue in HST WFC3 IR exposures.

2.7. Extended data quality array

In order to avoid systematic biases due to the presence of defects in some regions of the detector, we created a manual data quality mask to flag those regions were the flat field cannot fully correct the differences in sensitivity. Besides the effect of those regions

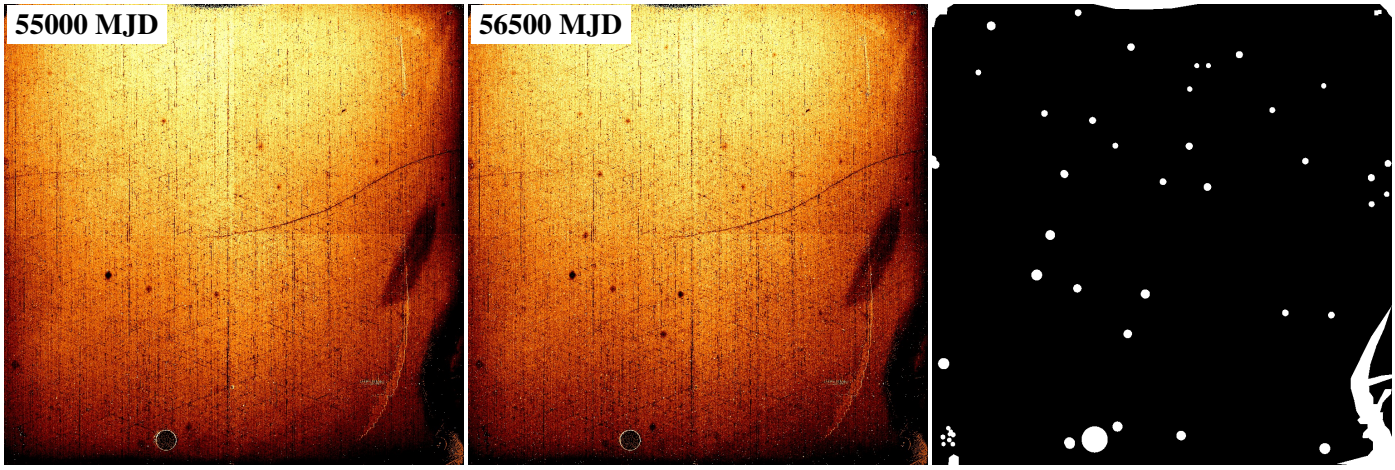

Fig. 8. Example of identification using two extreme flat fields, calculated for June 2009 (55000 MJD, *left panel*) and July 2013 (56500 MJD, *central panel*). *Right panel*: final data quality array. We represent in white those regions of the detector labelled as bad pixels. We note that this data quality array has to be combined with the individual data quality arrays from each individual exposure, containing the cosmic rays masks, the pixels affected by persistence, and the default time-dependent WFC3 IR blobs.

of the detector known to have lower sensitivity (wagon wheel, death star), we must pay attention to the presence of the WFC3 IR blobs. These blobs are regions that appear sporadically and cumulatively over the detector and present 5–10% less sensitivity on the detector (Pirzkal et al. 2010). Blobs are not related to any damage on the detector, but rather to particles that stick to the mirror of the Channel Select Mechanism (Calvi 2014). Their effect is strong enough to become a problem in the final mosaics. One of the most important problems is that the WFC3 IR blobs appear as a function of time. Although there is a continuous monitoring of the presence of new IR blobs (Pirzkal & Hilbert 2012), their appearance cannot be predicted.

To avoid including frames affected by blobs that were not flagged yet, we conservatively decided to mask manually those pixels with the largest blobs regardless of the epoch when the exposure was taken. We created a set of 16 sky-flat fields per filter (every 100 days, since 55000 to 56500 MJD) as a function of time in order to visually track the appearance of blobs (see Sect. 2.4). We show an example of the time-dependent flat fields used for the selection process in the left and central panels of Fig. 8. Notice that in the central panel, there are several additional IR blobs visible in the sky flat field. We carefully flagged any region with clear time variation according to our time-dependent flat fields and added them to our master data quality array, which is shown in the right panel of Fig. 8. We also flagged those regions with low sensitivity or none at all, which are clearly seen in the flat fields (wagon wheel, death star, the unbounded pixels of the top corners and middle strip, and the seven point-like damaged regions on the low left regions, see Dressel 2012, Chapter 5.7.7). Finally, we flag in the individual data quality (DQ) extension of all `flt.fits` files all the pixels included in our extended data quality array and mask from the individual images before sky-subtraction and co-addition.

2.8. Sky-subtraction

In this Section we describe the methods used to remove the sky background from the individual exposures and the final mosaics of the HUDF. We divide this task in two parts: flat sky background estimation (constant across the detector, Sect. 2.8.1) and removal of large scale residual gradients with a two-dimensional sky background (Sect. 2.8.2).

2.8.1. Flat sky-subtraction

A key step in the reduction process of astronomical images is the matching of the sky background level before the final co-adding of the individual frames. Because the near IR background is brighter than in the visible and its time variation is higher, this step can be particularly complicated even for space-based observations. Furthermore, for observations of extended sources or very crowded fields – such as the case of WFC3 IR observations of the HUDF – there are many low-surface brightness features that are buried deep under the 1σ noise levels of the individual images, only detectable in the final mosaics. Extended discs, stellar halos, and diffuse objects are extremely hard to mask and they contribute to bias the distribution of the pixels selected for sky background determination towards higher values. This leads to a common tendency to over-subtract in the reduction of deep images. Moreover, a biased determination of the sky background level of the different images before co-adding can greatly affect the final mosaic by including additional noise and thus distorting low surface brightness features of the final image. In this section we describe our flat sky background determination method and compare the results with other standard methods including the method applied by default in Astrodrizzle.

We first select those pixels that will be used to calculate a flat sky background level by using the masks created with NoiseChisel (see Sect. 2.2). In addition, we remove (set as NaN) from the sample those pixels that are flagged as bad pixels in the DQ array of each image, including those in the extended DQ array (see Sect. 2.7). We also remove those pixels affected by persistence, according to the custom models calculated in Sect. 2.5, flagging all pixels that present a persistence surface brightness brighter than $\mu = 30\,\mathrm{mag\,arcsec^{-2}}$. Then, we calculate the probability density distribution of the median value of the sky background level by using random re-sampling with replacement (bootstrapping). One of the main benefits of the re-sampling methods is to avoid any assumptions of normality on the sample and hence obtain a more accurate distribution for certain statistics. In addition, the shape of the resulting

probability distributions gives us information about the presence of outliers and irregularities. This method is much accurate than the sigma clipping methods applied by default by Astrodrizzle (Koekemoer 2002), as the main source of bias for a masked array are not the brightest pixels in the image, but the unmasked outskirts of the largest and most diffuse objects in the field of view. Those contaminated pixels are well below the 1σ interval of the sky background probability distribution. Careful masking based on the noise-based non-parametric algorithms such as NoiseChisel combined with robust statistic methods such as bootstrapping provides much more accurate estimations of the sky level.

We present a systematic comparison of multiple sky-subtraction methods on Fig. 9. In order to make a reasonable experiment, we used the simulated images of the Illustris Proyect[11] (Vogelsberger et al. 2014; Nelson et al. 2015). The Illustris project is a large cosmological simulation of galaxy formation, which simulates a volume of $(106.5\,\mathrm{Mpc})^3$, since $z = 127$ to $z = 0$. The project made available a set of model deep images[12] simulating the observations of HST ACS and WFC3 detectors, JWST MIRI and WFIRST. We took a section of the same size of WFC3 exposure frames (1014×1014) of the WFC3 F160W Illustris Field B. We added Poissonian noise with a standard deviation equal to 1 Jy. This choice is arbitrary and only for testing purposes. In the bottom panel of Fig. 9 we compare the following sky-subtraction methods:

1. Mean of the unmasked image.
2. Median of the unmasked image.
3. 3σ sigma clipped (50 iterations) mean of the unmasked image.
4. 3σ sigma clipped (50 iterations) median of the unmasked image.
5. Median of the SExtractor masked image (using 1σ detection limit for sources).
6. Bootstrapping median of the SExtractor masked image (using 1σ detection limit for sources).
7. Median of the Noisechisel masked image (default configuration).
8. Bootstrapping median of the Noisechisel masked image.

We remark that the sky background of the simulated image is equal to zero. This means that the methods that measure sky background levels closer to zero can be classified as better than those that predict higher levels. We found several interesting results: 1) every single tested sky-correction method systematically overestimates the true sky level, which for this simulation equals to zero Jy (see also Ji et al. 2018, for a similar result based a 2D fitting modelling method). 2) NoiseChisel is extremely efficient to remove the outer and dim regions of the simulated galaxies. The difference with respect to sigma clipping (or median sky methods is larger than one order of magnitude. 3) Bootstrapping does not add a significant improvement to the estimation of the median value, although it is still less biased than a simple median (only 2.5% less biased), even with NoiseChisel masked images. 4) As expected, sigma clipping is not a reliable method for sky-subtraction. The reason to this is that the main bias contributors to the sample of pixels are precisely those pixels which are well below the 1σ level, and thus those are not masked in the process. Finally, we choose the best subtraction method from those analysed (Bootstrapping + NoiseChisel masking) to estimate the final sky value of each exposure before co-addition.

[11] The Illustris Simulation: http://www.illustris-project.org/
[12] The Illustris ultra deep fields are publicly available at https://archive.stsci.edu/prepds/illustris/index.html

2.8.2. Two-dimensional sky-subtraction and diffuse light gradients

As we increase the depth of astronomical images, any analysis is less affected by the statistical uncertainties of the sky noise and become more dominated by systematic biases (PSF wings, sky gradients, or diffuse scatter light contamination). As a consequence of this, measuring the structure of objects with small but resolved angular sizes in highly crowded fields (such as the HUDF) is a increasingly challenging task. Even in an ideal case without any noise or residual sky background gradients, the structure of small objects is influenced by the presence of other objects in the field of view. For example, most of the objects with small angular sizes from Illustris simulation are on top of the extended wings of a larger close companion (see Fig. 9).

Besides the scatter light caused by real sources, no astronomical image is absolutely free of artificial sky gradients, and their effects can only be partially corrected. One of the most common methods for tackling with this problem is the subtraction of a two-dimensional sky background, by using $n-$degree polynomial fits, bicubic-spline interpolation on a mesh grid (i.e, SExtractor), or multiple subtractions of median filtered masked frames with different grid sizes (see Sects. 4.1 and 3.4 from Koekemoer et al. 2012; Illingworth et al. 2013, respectively). Most, if not all of these methods have to assume some kind of minimum spatial scale for the variation of the sky background (i.e, mesh grid for SExtractor). Any sky background variations larger than this minimum spatial scale will be fitted and subtracted from the corrected image, at least partially, regardless of their astronomical or instrumental nature. Correspondingly, smaller scales will contain an unknown fraction of artificial sky-gradient residuals and real astronomical diffuse light. Therefore, such methods cannot subtract a two-dimensional sky background that fully preserves the outskirts at all angular scales.

As stated before, highly aggressive sky background subtraction methods allow us to obtain very flat final mosaics on small spatial scales. While this can be an asset for photometry of unresolved objects, HUDF contains objects of very different sizes (from ~20 arcsec to a fraction of arcsec in diameter). Such methods present the disadvantage of removing the outer parts of the largest objects, which are the primary objective of this paper. In Illingworth et al. (2013) and Koekemoer et al. (2012), the authors decided to calculate a non-flat sky background correction in order to remove residual background features. The chosen sky background spatial scale ranged from 100 pixels (13 arcsec, HUDF12) to 39 pixels (5 arcsec, XDF). The process consisted in the subtraction of two-dimensional sky background arrays, after masking the individual images. In the case of the XDF, the sky background of all the masked images was stacked and subtracted from the individual images. Masking the frames previous to sky background correction partially reduces the over subtraction of the source on the images. Nevertheless –while detectable on the final mosaics– the outskirts of the largest objects extend far beyond the 1σ limit of the individual images, and thus usually get included in the two-dimensional sky background fit, biasing the final result.

Since our primary objective is to recover the outskirts of galaxies in the HUDF at lower redshifts ($z \sim 0.6-1$), we have to be more careful with our sky-subtraction techniques. (Akhlaghi & Ichikawa 2015, see Appendix A.6) show that even extensively used model-based methods to calculate the background, such as SExtractor (Bertin & Arnouts 1996), fail to interpolate the sky level for the corners of the detector, unless the observers perform a careful and individual setting of the mesh

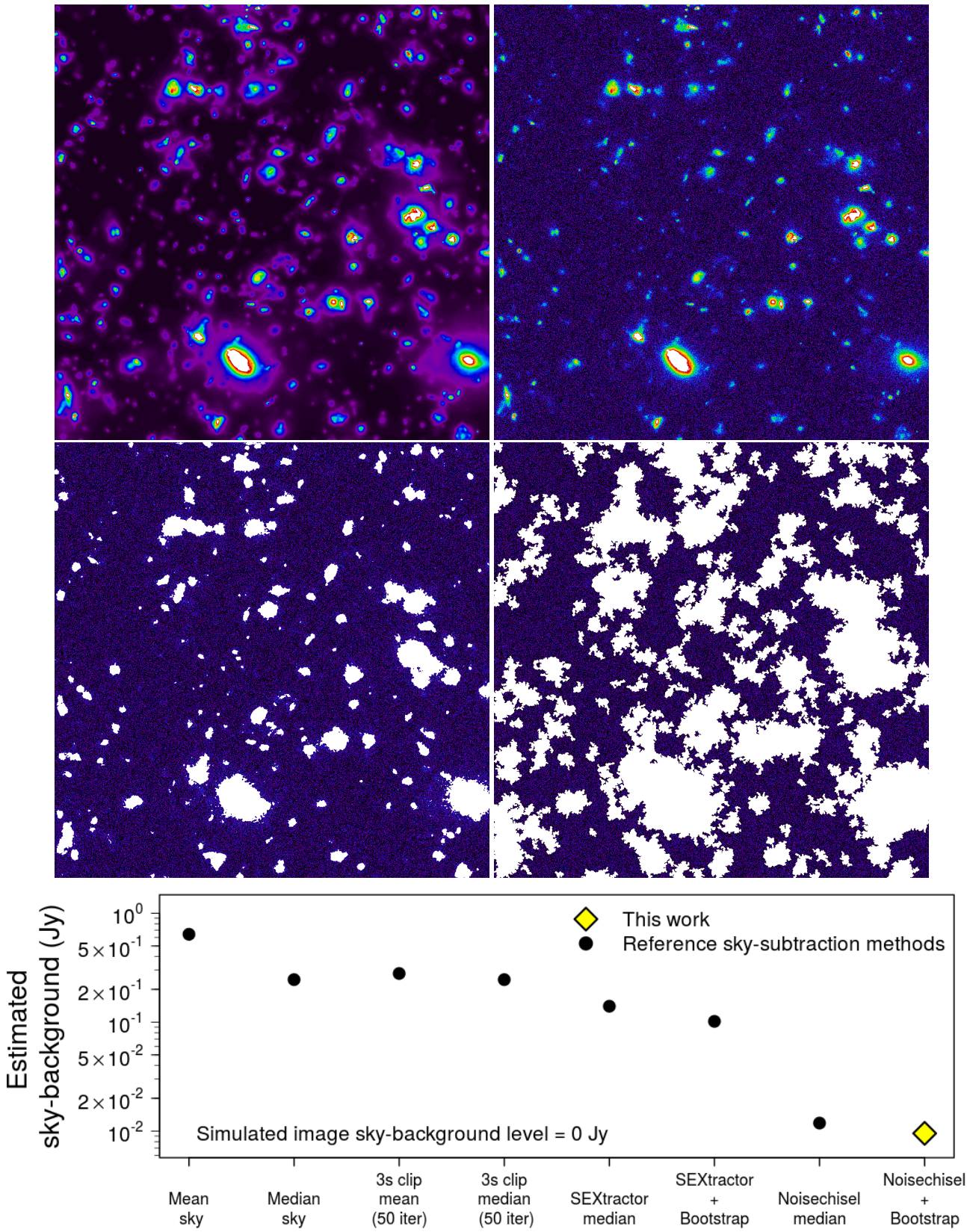

Fig. 9. Comparison of sky-subtraction methods, over a section of the Illustris simulation images. *Top-left panel*: simulated Illustris Deep Field B section. *Top-right panel*: same image with a 1 Jy wide Poissonian noise added. *Middle-right panel*: simulated Illustris Deep Field B with noise masked with SExtractor (1σ threshold level). *Middle-left panel*: same image masked with NoiseChisel (default configuration). *Bottom panel*: estimated sky background as a function of different methods. Black dots: reference methods. Yellow diamond: method applied in the present paper. Notice that all methods overestimate the true sky background level, which is equal to zero in the simulated image. See the legend and axis for details.

size (BACK_SIZE) for each image. This process may depend on the objects and their position on the detector. For this paper, we set the parameters of NoiseChisel checking the spatial distribution of the valid tiles (a subset or a section of the input array) for sky background estimation. We refer the reader to Gnuastro tutorial 2.3 – Detecting large extended targets[13] for a detailed tutorial on detecting the low surface brightness wings of extended sources on astronomical images and using NoiseChisel to avoid systematic biases on sky subtraction. In our configuration, NoiseChisel calculates the sky background over rectangular sections of the image called tiles, removing from the analysis those tiles which present significant differences between the mode and the median. The distance between valid tiles depends on the S/N and the distribution of the sources on the image, being less valid tiles (or none) on more crowded regions of the image. For the individual images we set the tile size to 100 pixels, with the smoothwidth parameter fixed to 5 and a mode-median quantile difference equal to 0.005. The needed number of neighbour tiles for interpolation was set to 5. We modified those parameters until no valid tiles were accepted by the program near the most extended objects. As a final step, we subtracted a final sky background gradient for each mosaic, using again a tile size of 100 pixels with smoothwidth equal to 5 and the mode-median quantile difference equal to 0.01.

We conclude that as a consequence of protecting the outskirts of the largest objects in our reduction process, we cannot neglect the presence of artificial sky-gradient residuals with spatial scales of several hundreds of pixels or smaller. This is an inevitable consequence of the reduction process that has to be corrected on a case-by-case basis, taking into account the intensity, shape, angular size, and environment of each object to study. In Sect. 3 we provide some examples to illustrate the benefits of our dedicated mosaics and discuss the required corrections to analyse diffuse light contaminated objects.

2.9. Image alignment

The astrometric solution of the HST images is precise up to a fraction of arcsec (typically 2–5 pixels, Gonzaga et al. 2012). As a consequence, when comparing images from different visits, it is usual to see that they are not exactly aligned. In order to exploit the full resolving capabilities of WFC3, we need to carefully re-align the images of different visits to a single reference world coordinate system solution (WCS hereafter). To perform this correction we use Tweakreg[14]. Tweakreg is a task part of Drizzlepac and it allows the user to align sets of images to each other and/or an external astrometric reference frame or image. In this paper, we perform this correction following the method presented on Lucas & Hilbert (2015) and Lucas (2015):

1. We generate four reference catalogues (F105W, F125W, F140W and F160W) for the XDF final mosaics using SExtractor (Bertin & Arnouts 1996). We iteratively increased the threshold level (DETECT_THRESH and ANALYSIS_THRESH) from 1σ to 3σ in order to remove spurious sources that may affect the final solution.
2. We group the individual exposures by visits, using AstroDrizzle to correct for cosmic rays. We use AstroDrizzle to analyse each group of images from the same visit and filter and substitute the identified cosmic rays and flagged pixels with the blotted median of the group (see Chapter 4.2.7 of Gonzaga et al. 2012).
3. We generate one catalogue per exposure to correct using the cosmic ray corrected frames.
4. We use Tweakreg to find the necessary shift and scaling for each exposure. The program uses the cosmic rays cleaned image, their corresponding source catalogue and the reference catalogue to its filter, and it matches the astrometry of the individual images to the XDF mosaic.
5. We repeat the previous process, visually checking for misalignments during each iteration and modifying the sigma level of the catalogues and the searching radius of Tweakreg.
6. Finally, we copy the astrometric solution to the flt.fits files.

One of the outputs from Tweakreg is the residuals of the positions for each object and its reference object according to the final astrometric solution for each dataset. In order to measure the average precision that we achieve on our images, we measure the median dispersion of the residuals on all our datasets. The results are shown in Fig. 10. The median dispersion of the residuals is ~0.2 pixels on both directions of the array, which corresponds to ~0.025 arcsec. We conclude that our images are sufficiently well-aligned to use a final pixel scale of the mosaics is equal to 0.06 arcsec, which is the same used by HUDF12 and XDF teams.

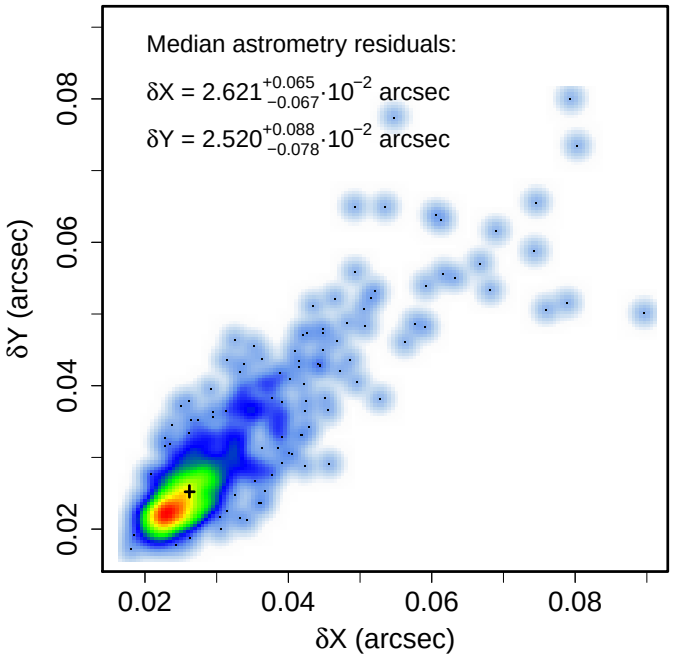

Fig. 10. Colour coded density two-dimensional distribution of the astrometry dispersion of the residuals on the HUDF dataset, according to Tweakreg. The black cross represents the median value for the astrometry residuals, in arcsec. Consult the values on the panel.

2.10. Image combination: BootIma

Image combination is one of the most common tasks for astronomers. Avoiding systematic biases at this step can be a challenging task. Ground and space based surveys are subdued to a large number of time dependent conditions, either external, such as the effect of cosmic rays, the variation of the sky background or stray-light contamination, that may affect in the

[13] Gnuastro 2.3 Tutorial – Detecting large extended targets: https://www.gnu.org/s/gnuastro/manual/html_node/Detecting-large-extended-targets.html
[14] Tweakreg is publicly available at: http://drizzlepac.readthedocs.io/en/deployment/tweakreg.html

form of gradients, or internal, such as variations on the sensitivity of the detector or the position of the different objects over the detector. All these effects produce large variations on the final mosaics, and their typical time-scale can vary between the length of the exposure time to years. The most common methods for co-adding are different types of robust median. Most (if not all) HST images rely on the imcombine task of AstroDrizzle (Koekemoer 2002). AstroDrizzle allows the user to use different methods of image combination. The recommended method for a large number of images (more than ten) is "imedian", which corresponds to the median value for each pixel, with the exception of those regions were all pixels were flagged as bad. In that case, the algorithm returns the last value of the stack. This will prevent to have holes in the centre of stars, for example.

Nevertheless, this type of processes does not provide accurate uncertainties for the final mosaics, which are extremely useful for many scientific objectives (that is, measuring the local sky noise, performing accurate photometry, or two-dimensional decomposition). For that objective, we created BootIma. BootIma is a set of programs written in Python and HTCondor[15] to perform image combination through Bootstrapping. We use this program for the flat field combination and to create the final mosaics. The main task of BootIma is the robust combination of a large number of images and estimation of their uncertainties. Such task is computationally expensive, and we optimised it to work in parallel processors and specially, HTCondor. We can summarise in two the main reasons for using bootstrapping to calculate the median images:

1. It estimates a robust measurement of the confidence intervals, which are provided with the final mosaics and flats obtained with this task, which are extremely useful in the case of deep imaging studies.
2. It allows us to combine images using weights. This is particularly useful for the HUDF, where multiple observing programs used different exposure times. This also means that the observing sample does not present a pure Gaussian distribution and has to be treated with non-parametric methods.

A previous step to image combination is the distortion correction. The images produced by WFC3 are affected by geometric distortion, caused by the tilt of the image surface with respect to the path of light (Kozhurina-Platais et al. 2012). If the distortion correction is not applied or accurate enough, the image combination will produce blurred images and distort the PSF. AstroDrizzle drizzling code corrects for this geometric distortion using the calibration `IDCTAB` files, with a precision of 0.1 pixels. After geometric transformation, AstroDrizzle reassigns each pixel to a new and undistorted pixel grid (Fruchter & Hook 1998). This process is called drizzling. The new pixel grid can take advantage of the dithering in order to reconstruct a final image with smaller pixel scale by oversampling the input data. We refer the reader to Chapter 6.3 of Gonzaga et al. (2012) for a detailed description of this process. The fraction by which each input pixel is shrunk before being drizzled onto the input image is controlled by the `final_pixfrac` parameter, which we set to 0.8, following the prescriptions for the HUDF12 and the XDF. We choose the Gaussian kernel to distribute the flux of each input pixel onto the new pixel grid. Finally, we obtain a new set of images, corrected for geometric distortion, with a pixel scale of 0.06 arcsec, and aligned into a common pixel grid, ready for image combination.

[15] HTCondor is an open-source high-throughput management system for computing-intensive jobs: `https://research.cs.wisc.edu/htcondor/index.html`

BootIma performs the median image as follows: first, we register exposure time and date for all the input images. Images are re-ordered as a function of time, to simplify visual inspection. Second, we create an `hdf5` datacube, where we store all the individual image arrays, along with their exposure time and date. Once the master `hdf5` file is created, we calculate a robust median value along the date direction using a Python-HTCondor program built for this purpose. We use the weight images created with Astrodrizzle for each image weights during random re-sampling, as they take into account the exposure time and the fraction of counts that correspond to each pixel in the new drizzle grid for each pixel from each original pixel grid. Notice that our datacube contains "holes" (that is, NaN values due to persistence masking and bad pixels). This prevents us from using random re-sampling directly over the full set of images, because all bootstrapping simulations would not have the same amount of pixels. Each pixel of the image has a different sample size, and thus requires independent analysis. Finally, we reconstruct the final image into a single frame.

We conclude that our method allows us to obtain precise error frames, extremely useful for many science cases while performing robust combination of the images into the final mosaics. In the FITS files that contain the ABYSS images, we included two additional FITS extensions, that contain the $+1\sigma$ and -1σ uncertainty intervals for each pixel in the field of view.

3. Results and discussion

3.1. General properties

In this section we detail the morphological and photometric properties of the newly reduced mosaics, which we present in the colour image of Fig. 11. We compare the results from our images to the original HUDF12 (Koekemoer et al. 2012) and the XDF (Illingworth et al. 2013) mosaics.

3.1.1. Recovering the low-surface brightness structure

In Figs. 12 and 13 we perform a comparison of the low surface brightness properties of the three different versions of the mosaics compared in this paper: HUDF12, XDF and our dedicated version of the HUDF mosaics which we called ABYSS. We perform the same analysis for the four available filters. In order to reduce the noise and enhance the differences between our mosaics and the previous versions in the low-surface brightness regions, we have re-sampled the image binning the data to 0.6 arcsec boxes (10×10 pixels).

Visual inspection of the binned difference frames reveals that the XDF mosaics present large over-subtracted regions, which are centred over the regions where the most extended objects are. They are easily identifiable in the difference intensity images that we present on Figs. 12 (for F105W and F125W) and 13 (for F140W and F160W). Comparing the light distribution of the largest objects from the ABYSS mosaics and the morphology of the difference intensity images for the XDF we find that they create almost a mirror image of the intensity mosaics. Moreover, the distribution of the residuals for the XDF is strikingly similar in the four filters, being more clear in the F105W and more noisy in the F140W (this is an expected effect due to the relative depth of the mosaics).

Although, in principle, this effect may be caused both by over-subtraction in the regions around the largest objects of the XDF or by under-subtraction in our ABYSS version, because we also applied a two-dimensional sky background correction

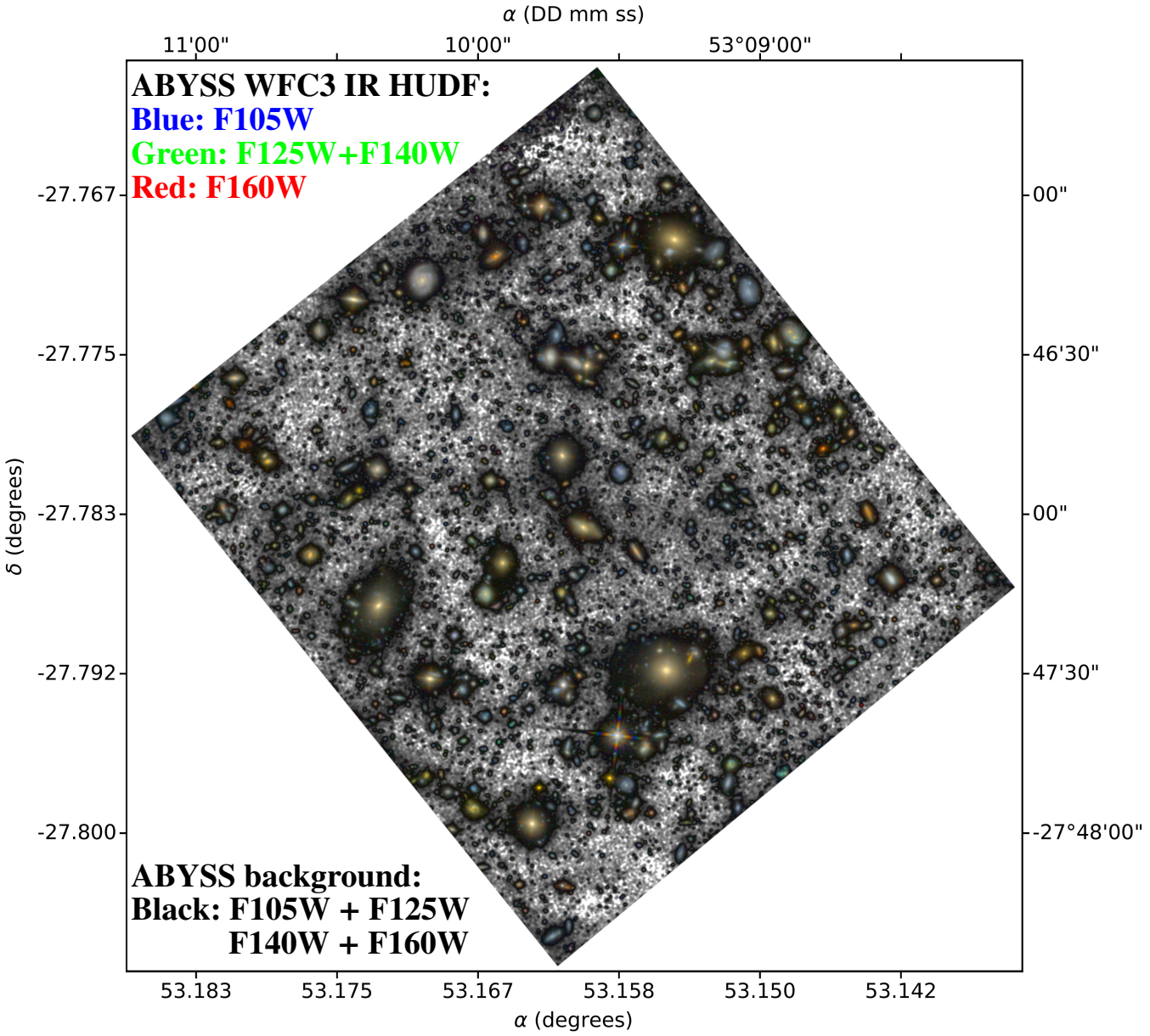

Fig. 11. Luminance-RGB image showing the full depth of the mosaics of the ABYSS version of the HUDF WFC3 IR. The high S/N parts of the mosaics are represented with colours (red: F160W, green: mean of F125W and F140W bands, blue: F105W). The low S/N regions are represented in as a black and white background (black regions are brighter than white regions) according to the mean image of the four mosaics (F105W, F125W, F140W, F160W) of the mosaics of the ABYSS (covering HUDF, deep WFC3 IR region).

to our mosaics, as Illingworth et al. (2013, see Sect. 2.8.2), it is highly unlikely that the correction that we applied generate under-subtraction around the largest objects. First, because the typical scale-length of our gradients is one third or half of the field-of-view, much larger than the size of the largest objects on the HUDF, while the variations that we detect on the binned difference frames are much smaller. And secondly and most importantly, two-dimensional sky-subtraction tends to fit the extended light of the galaxies, over subtracting these regions, not the opposite effect.

Interestingly, we found that the HUDF12 presents a similar effect at a lower scale. The distribution of residuals for the F105W and the F140W filters are similar for the XDF and the HUDF12 mosaics when comparing with our mosaics, although the differences are less intense. Inspecting the difference intensity images corresponding to those mosaics (right panels of Figs. 12 and 13), we can easily see how the extended light around one of the largest elliptical galaxies, HUDF-5 (α = 53.15545, δ = −27.79150, Buitrago et al. 2017), was over subtracted, similarly as in the case of the XDF mosaics. Nevertheless, the west region of the F125W of the HUDF12 is less over subtracted than the rest, presenting a lower level of residual intensity in Fig. 12. We found a similar effect in the southeast region of the HUDF12 F160W mosaic (see Fig. 13), where the HUDF12 predicts more flux than our own mosaics. Neither this effect nor the discussed on F125W was found when comparing to the XDF, suggesting that the issue is due to the HUDF12 mosaics.

As an illustrative exercise, we calculated the equivalent integrated magnitude of the light recovered on the low surface brightness regions of the ABYSS HUDF mosaics, compared to the previous versions. In order to do that, first we calculate the

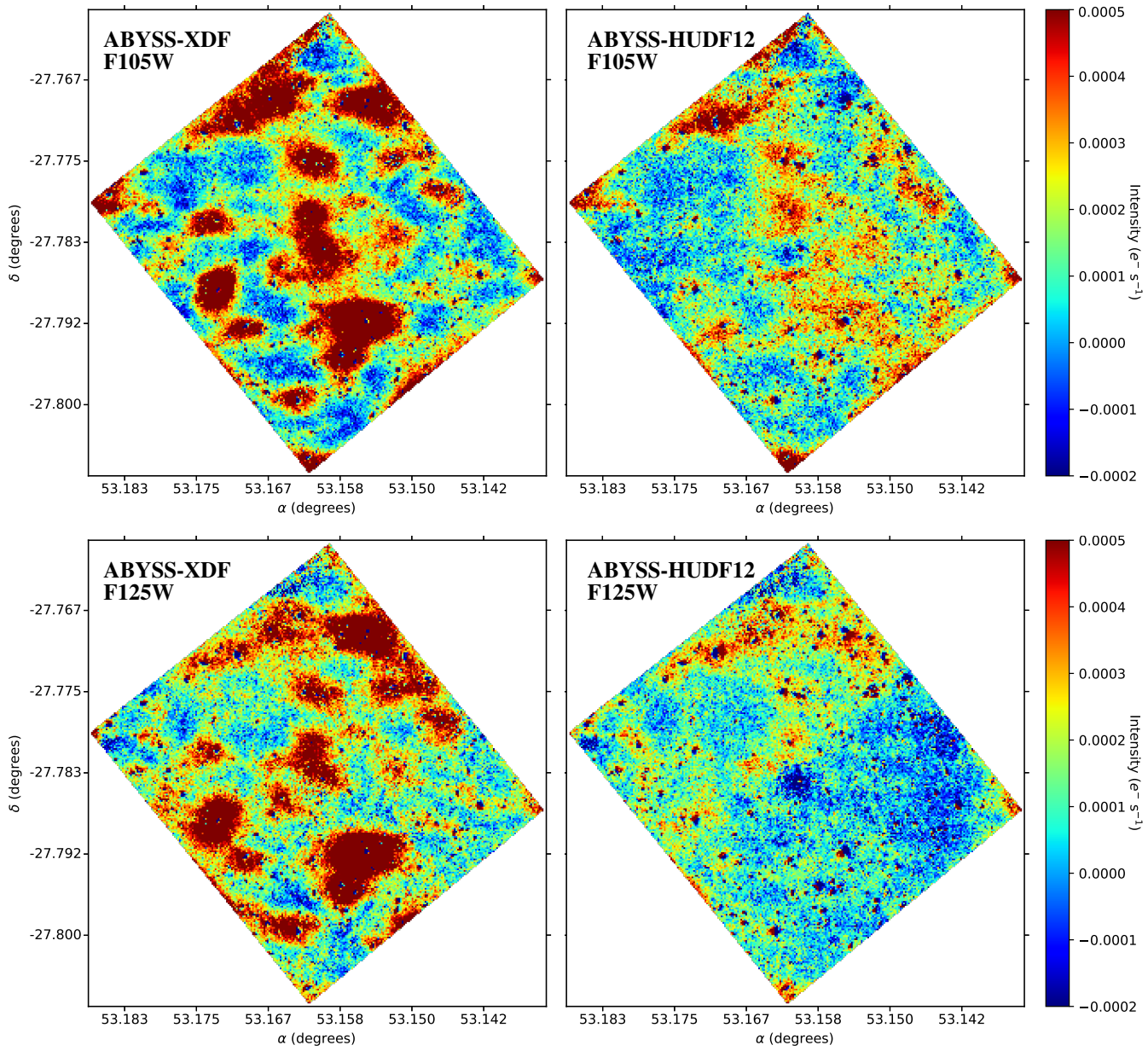

Fig. 12. Comparison of the ABYSS mosaics of the HUDF WFC3 IR and the previous releases. Each plot represents the intensity difference between our version of the mosaics (ABYSS) and the reference versions (*left panels*: XDF, *right panels*: HUDF12), for the F105W filter (*top panels*) and the F125W filter (*bottom panels*). See the colour bar for reference.

mean surface brightness on 1.2 arcsec boxes (20 pixels). Due to the fact that small differences in magnitude at the brightest cores of the galaxies may dominate over the differences on the dimmest regions of the images, we select only the regions were the mean surface brightness is lower than $\overline{\mu} = 26\,\mathrm{mag\,arcsec^{-2}}$ for each individual filter. Finally, we integrate the difference in flux between our mosaics and the reference images (XDF and HUDF12). We show the resulting equivalent magnitudes in Fig. 14. We found that the amount of recovered light is equivalent to a ∼19 mag source when comparing to the XDF and to a ∼20 mag for the HUDF12. This is comparable to the brightness of some of the largest objects in the HUDF. This result is nearly constant for all the filters, although we found that the integrated magnitude for the F125W and F160W images of the HUDF12 is dimmer, in agreement with the results found on the difference maps (see Figs. 12 and 13). We must remark that this result highly depends on the size of the image. In a larger area, the total magnitude missing will be larger. In this sense, the values provided here should serve only to compare the equivalent effect of over-subtraction of low surface brightness features on brighter sources.

We conclude that our images successfully recover most of the extended light around the largest objects from the HUDF WFC3 IR. The recovered light is easily identified as positive counts when subtracting the previous versions of the mosaics from the ABYSS HUDF WFC3 IR images around these objects.

3.1.2. Photometric consistency with previous HUDF WFC3 IR images

In this section we study the photometric properties of the final mosaics, comparing with the results from the previous

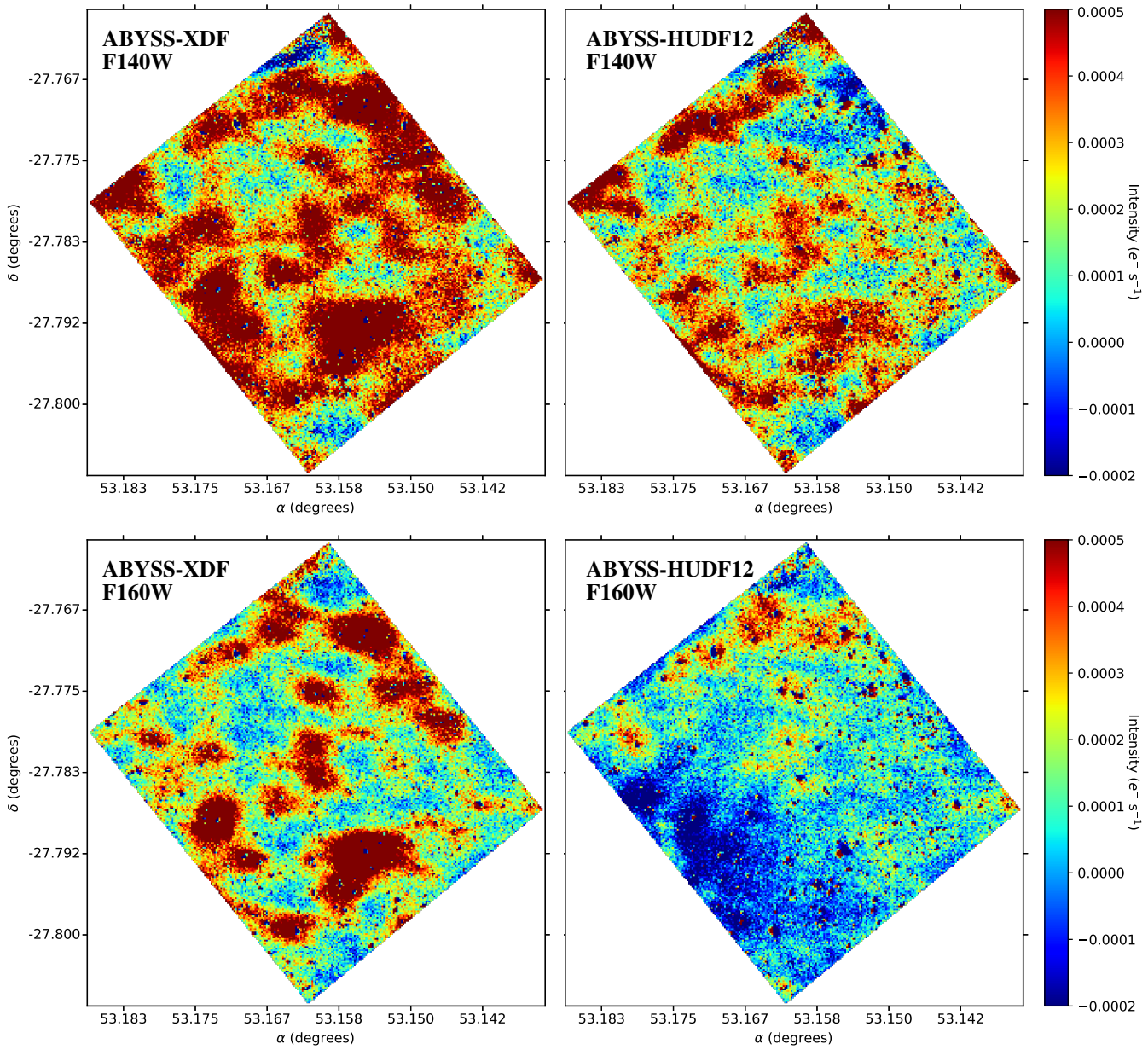

Fig. 13. Comparison of the ABYSS mosaics of the HUDF WFC3 IR and the previous releases. Each plot represents the intensity difference between our ABYSS mosaics of the HUDF WFC3 IR and the previous reductions (*left panels*: XDF, *right panels*: HUDF12), for the F140W filter (*top panels*) and the F160W filter (*bottom panels*). See the colour bar for reference.

reductions of the WFC3 IR mosaics for the HUDF. In order to do that, we measure and compare the magnitude of ∼2500 objects on the HUDF and identified using NoiseChisel and measured using MakeCatalogue (both part of Gnuastro, see Akhlaghi 2016 for a discussion about separate detection and catalogue production on astronomical surveys). We created a catalogue per filter and reduction set (12 final catalogues in total). We provide the number of objects identified on the mosaics created for this test on the legend of each panel in Fig. 15.

The NoiseChisel segmentation maps are highly dependent on the shape and amount of extended light of the objects. Because of this reason, we used one mosaic per filter to calculate the segmentation maps. In this case, we used the mosaics from the HUDF12. Notice that the exact photometry of the objects depends on this choice, but not the conclusions from this test. Thus, the segmentation maps are fixed for the three versions used to compare and independent for each one of the four filters. Finally, MakeCatalogue calculates the integrated magnitude, correcting for the local sky background for each object.

In Fig. 15 we compare the differences in magnitude between our reduction and the HUDF12 and the XDF mosaics. We found that the median differences between the magnitudes are $\Delta m_{\mathrm{HUDF12-ABYSS}} = 0.026^{+0.016}_{-0.013}$ mag when comparing to the HUDF12 versions of the mosaics, and slightly larger ($\Delta m_{\mathrm{XDF-ABYSS}} = 0.067^{+0.006}_{-0.005}$ mag) when comparing with the XDF mosaics. These differences are negligible and (most importantly) not systematic. The objects in our filters are not systematically brighter or dimmer than in the previous releases. In addition, we found that they are compatible with the differences that we found when comparing the XDF with the HUDF12 mosaics using the same method and apertures ($\Delta m_{\mathrm{F105W}} = 0.071^{+0.006}_{-0.006}$, $\Delta m_{\mathrm{F125W}} = 0.047^{+0.004}_{-0.005}$, $\Delta m_{\mathrm{F140W}} = -0.030^{+0.003}_{-0.003}$,

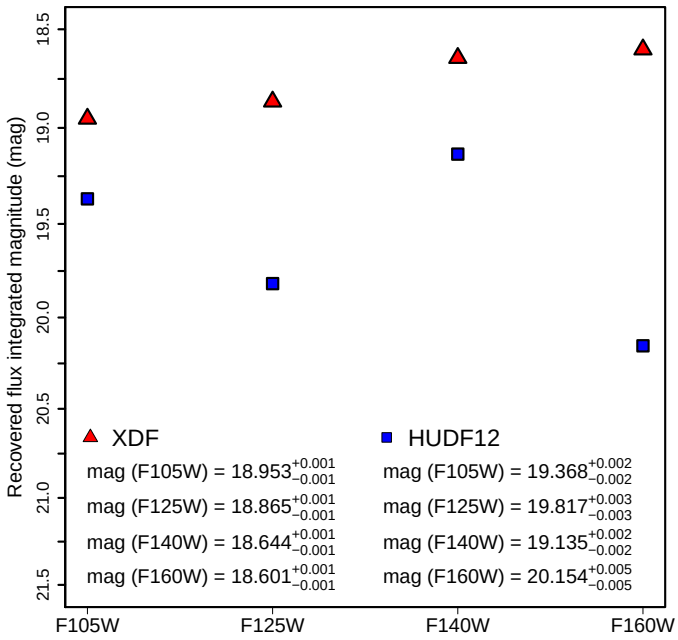

Fig. 14. Missing light of the HUDF measured as the integrated magnitude of the regions with $\overline{\mu} > 26$ mag arcsec^{-2}. These are calculated with the difference images between the ABYSS mosaics and the XDF (red triangles) and the HUDF12 (blue squares). See the legend for the values and uncertainties of the different bands.

$\Delta m_{\rm F160W} = 0.021^{+0.005}_{-0.004}$). We conclude that the photometric analysis in fixed apertures does not reveal any systematic bias or significant differences when comparing to the results and differences between the previous versions of the HUDF mosaics.

3.1.3. Surface brightness limiting magnitude maps

In Fig. 16 we present the surface brightness limiting magnitude maps, calculated using the standard deviation estimation of NoiseChisel for the F105W, F125W, F140W, and F160W mosaics. We find that despite removing a large amount of pixels because of persistence and gradient corrections, the relative depth between filters has not notably changed. We summarise the results in Table 2. In our version, as well as in the previous releases, the deepest filter is the F105W band ($\mu_{\rm lim} = 32.89^{+0.01}_{-0.02}$ mag arcsec^{-2}, 3σ in 10×10 arcsec boxes), followed by F140W ($\mu_{\rm lim} = 32.54^{+0.03}_{-0.01}$ mag arcsec^{-2}), F125W ($\mu_{\rm lim} = 32.58^{+0.01}_{-0.02}$ mag arcsec^{-2}), and F160W ($\mu_{\rm lim} = 32.52^{+0.01}_{-0.02}$ mag arcsec^{-2}). The relative differences between the surface brightness limiting magnitudes for the same filters of different versions are very small. We note that even after the extremely conservative selection criteria that we have carried out to select the images and the valid pixels (mainly due to persistence effects), our final mosaics present compatible limiting magnitudes than previous HUDF releases. The limiting magnitude for the ABYSS F105W image is 0.01 mag arcsec^{-2} deeper than in the XDF and 0.04 mag arcsec^{-2} deeper when comparing with the F105W HUDF12 image. The F160W image is the most affected presenting a limiting magnitude 0.15 mag arcsec^{-2} brighter than in the XDF and HUDF12. This is because the F160W images are more affected by persistence effects (see Sect. 2.5). Finally, the the F125W presents a compatible depth with XDF and HUDF12, while our F140W mosaic is slightly deeper (~0.06 mag arcsec^{-2}) than the previous reductions (see Fig. 16).

Nevertheless, the above limiting surface brightness magnitude should be understood as the formal limiting magnitude and not the effective ones from the images. The effective limiting magnitudes are affected by systematic effects (such as sky over-subtraction) not included in this measurement of the pixel noise. The surface brightness limiting magnitude is position dependent over the field of view, with differences of almost 0.5 mag arcsec^{-2}. It is notable the presence of a shallower region on our mosaics on the north corner, due to the effect of conservative masking of the WFC3 IR chip cosmetic defect called the "wagon wheel" (see Sect. 2.7). We find that this region can present a surface brightness limiting magnitude ~1 mag arcsec^{-2} brighter than the rest of the mosaics. Any analysis including data from this region should be done carefully.

The side effect of our reduction process to avoid systematic biases (in particular, removal of persistence contamination) is that the F160W mosaic present slightly brighter formal limiting surface brightness than the previous versions of the HUDF, if we only take into account the pixel noise. Nevertheless, as shown in Sect. 3.1.1, those previous reductions were dominated by systematic biases, much larger than the relative differences found in surface brightness limiting magnitude. As a result of this, our mosaics contain much more information on the outskirts of the extended objects of the HUDF than any previous version of the data. We will comment on this in Sect. 3.2. We find that the conservative persistence masking and removal of gradients that we have performed do not affect significantly the depth of our mosaics, which present surface brightness limiting magnitudes similar to those from the previous versions of the HUDF, while reducing the systematic biases.

3.2. Surface brightness profiles

In this section we study the surface brightness profiles of several objects (see Table 3) from the HUDF, comparing the results from our reduction and the previous releases of the mosaics. First, in Sect. 3.2.1 we focus our attention on those objects with large angular sizes, which are more sensitive to over-subtraction and the primary target of the reduction process applied on this work. We start with HUDF-5, the largest one from our sample ($z = 0.607$, following the nomenclature from Buitrago et al. 2017) and one of the objects most affected by the sky over-subtraction of the previous versions, continuing with HUDF-1 ($z = 0.618$), HUDF-2 ($z = 0.619$), and ABYSS-1 ($z = 0.622$), a spiral galaxy from Elmegreen et al. (2005). Finally, we will study whether there are significant changes on the structure of the objects with smaller ($R_{\rm lim} < 10$ arcsec) angular sizes in Sect. 3.2.2. For this, we have selected two additional spiral galaxies from Elmegreen et al. (2005): ABYSS-2 ($z = 0.607$) and ABYSS-3 ($z = 1.28$). We show the cutouts of the targets in Fig. 17. The main selection criteria for the spiral galaxies was the relative isolation to avoid light contamination from nearby sources. We refer to Table 3 for the sky coordinates, main properties, and IAU ID.

In order to properly analyse the surface brightness profiles of the target objects we first performed a careful manual masking of all the nearby objects. These masks were created using the sum of all the four filters. For consistency, we use the same masks for every filter and version of the HUDF. Secondly, we analysed the surface brightness profiles of each galaxy using concentric elliptical apertures. We take into account both the uncertainties of the intensity dispersion along the elliptical aperture and the intrinsic uncertainty of each pixel (sky noise) using bootstrapping and Monte Carlo simulations. The sky noise is measured on each

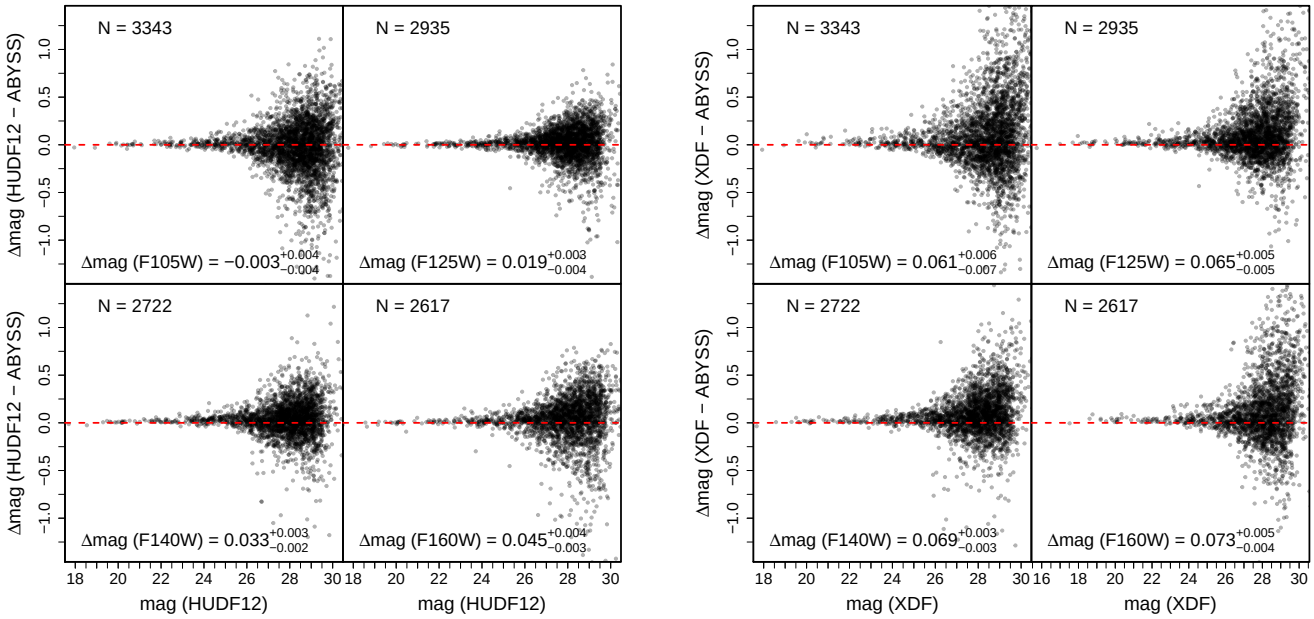

Fig. 15. Comparison of the HUDF photometry of several sources for our ABYSS reduction and the previous releases. Each panel shows the magnitude differences with their corresponding reference catalogue as a function of the object magnitude. *Left panel*: ABYSS vs. HUDF12. *Right panel*: ABYSS vs. XDF. The magnitudes are measured using fixed non-parametric apertures in each object and filter calculated with Gnuastro using the HUDF12 segmentation maps.

mosaic independently, using the standard deviation maps used for the limiting magnitude analysis (see Sect. 3.1.3). Finally, to improve the sky estimation we subtract the local sky level using the median value of all the valid appertures that are beyond a certain limit (Pohlen & Trujillo 2006), which is set at 140 kpc for HUDF-2 and HUDF-5, following the method described on Buitrago et al. (2017). For the three spiral galaxies (ABYSS-1, ABYSS-2 and ABYSS-3) and HUDF-1 (notably smaller than HUDF-2 and HUDF-5), we measure the local sky level beyond 80 kpc, which is well beyond the visual limit of the objects. The geometric parameters for the profiles (position angle and axis ratio) were calculated using SExtractor. We detail the position angles, axis ratios, and the local sky region limit for each object in Table 3. We remark that in order to do a fair comparison we apply the same masks, geometric parameters, elliptical aperture algorithm, and local sky level region to all the mosaics, regardless of their wavelength and version. For the sake of simplicity, we do not apply any type of PSF correction. The main objective of the surface brightness profile analysis that we present here is to illustrate the differences on the low surface brightness regions of the different mosaics. The differences of the PSF between the three versions studied are minimal, as they are created from similar datasets. Thus, any attempt to correct the individual images from PSF effect would introduce additional uncertainties. We will study the shape of the surface brightness profiles of the galaxies on the HUDF according to the new mosaics in a forthcoming paper (Borlaff et al., in prep.).

3.2.1. Surface brightness profiles of large objects

HUDF-5 (Buitrago et al. 2017) is an elliptical galaxy ($z = 0.667$, $\log_{10}(M/M_\odot) = 11.19^{+0.09}_{-0.05}$) and one of the brightest objects from the HUDF. It presents a noticeable shell envelope, and is one of the objects most affected by the aggressive sky subtraction from the previous releases. In Figs. 18 and 19 we represent the intensity images of the six targets in the F105W filter (which is the deepest mosaic in our reduction) in three different panels, for the ABYSS, HUDF12, and XDF version of the mosaics. On top of each panel we represent with black contours their respective $\mu = 29$ mag arcsec^{-2} isophotes. We observe that: 1) the $\mu = 29$ mag arcsec^{-2} isophotal contours are more extended in our reduction than in the previous versions of the mosaics, specially when compared to the XDF, and 2) the new diffuse light tends to appear around the largest objects on the field-of-view, while the objects with small angular size almost do not change their respective contours. This is a expected result, as aggressive sky-subtraction tends to affect mostly to the envelopes of the largest objects.

Nevertheless, we note that the intensity images are not a direct proxy of the shape of the surface brightness profile, as they are not corrected by the small sky residuals that appear on the surface brightness profiles (local sky background). The same effect shown on Figs. 18 and 19 could be caused by sky background under-subtraction on our mosaics. In order to confirm the validity of these findings, we must analyse the surface brightness profiles corrected by the local sky background of the target galaxy. In Fig. 20 we show the results of the surface brightness analysis for HUDF-5. We found several interesting results:

1. Even after applying a local sky level correction, the XDF mosaics present a significant over-subtraction when comparing to the HUDF12 or our dedicated mosaics. This effect is noticeable at surface magnitudes fainter than $\mu \sim 26{-}27$ mag arcsec^{-2}.
2. Our dedicated mosaics recover a significant amount of light when compared to the HUDF12 mosaics and the XDF. The maximum differences in surface brightness range from $\Delta\mu = 0.5{-}1.25$ mag arcsec^{-2}, being higher for the deeper images (F105W and F160W).
3. Interestingly, the XDF version of the F140W mosaic appears to be less over-subtracted than the rest of the filters, obtaining a limiting radius at $R \sim 10$ arcsec, while the rest of the filters reaches only $R \sim 8$ arcsec.

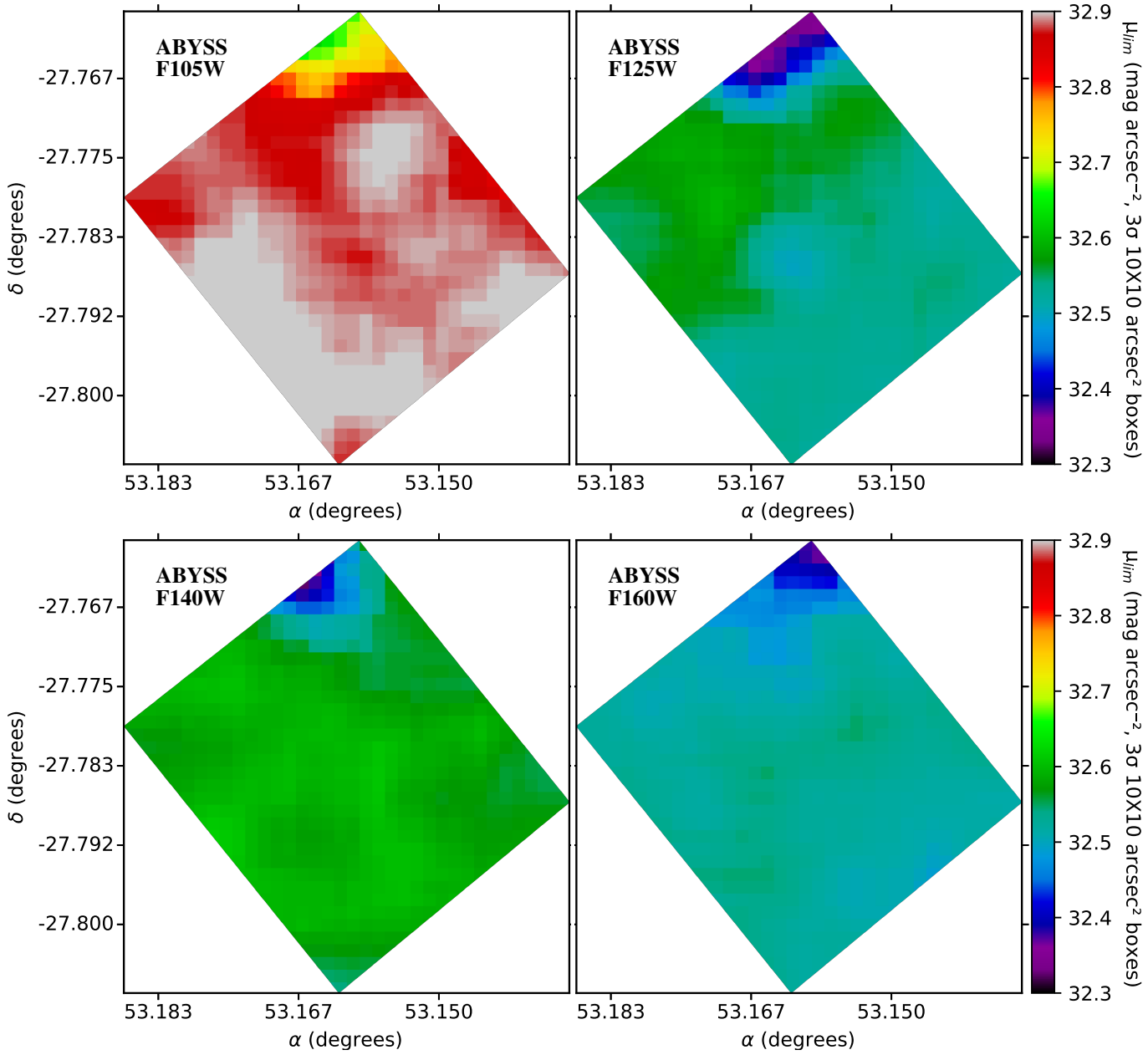

Fig. 16. Surface brightness limiting magnitude maps for the ABYSS mosaics of HUDF. *From top to bottom and left to right*: F105W, F125W, F140W, and F160W bands. Each panel represents in colour scale the surface brightness limiting magnitude for each mosaic, measured as the 3σ upper limit of the sky level on 10×10 arcsec^2 boxes (Trujillo & Fliri 2016), as a function of the position on the mosaics. All the panels are at the same scale (see the colour bar on the *right panels* for reference).

4. For the F105W, F125W, and F160W images, the surface brightness profiles of ABYSS extend much further out (up to $R \sim 25$ arcsec in the F105W and F160W bands) than in the HUDF12 and specially than the XDF mosaics, where the limit is almost one third smaller than the limiting radius of the ABYSS mosaics.

The general agreement up to a certain radius of the HUDF12 and our reduction, contrary to the results obtained with the XDF, is a quantitative proof of that the latter mosaics present a high and systematic over-subtraction of the outskirts around the most extended objects, such as HUDF-5. The over-subtraction can be as high as $\Delta\mu \sim 1.5$ mag arcsec^{-2} at a surface brightness magnitude of $\mu \sim 28$ mag arcsec^{-2} for the XDF mosaics even after applying local sky correction. We obtain a similar result for the HUDF12 mosaics at $\mu \sim 29$ mag arcsec^{-2}. This result demonstrates that the depth of previous reductions was dominated by systematic biases rather than the sky noise. We also found interesting that the surface brightness profile of the shallower mosaic (F140W) of XDF presents $S/N > 3$ to larger galactocentric radius than the rest of the filters, which are much deeper. In addition, visual inspection of the surface brightness profiles from our images reveals that they tend to reach the sky level smoothly, following the general shape of the surface brightness profile, rather than showing a sharp down-bending profile at the outskirts as observed on the XDF profiles.

In Fig. A.1 we represent the surface brightness profiles of HUDF-1, another one of the elliptical galaxies studied in Buitrago et al. (2017). We found a very similar result as in the

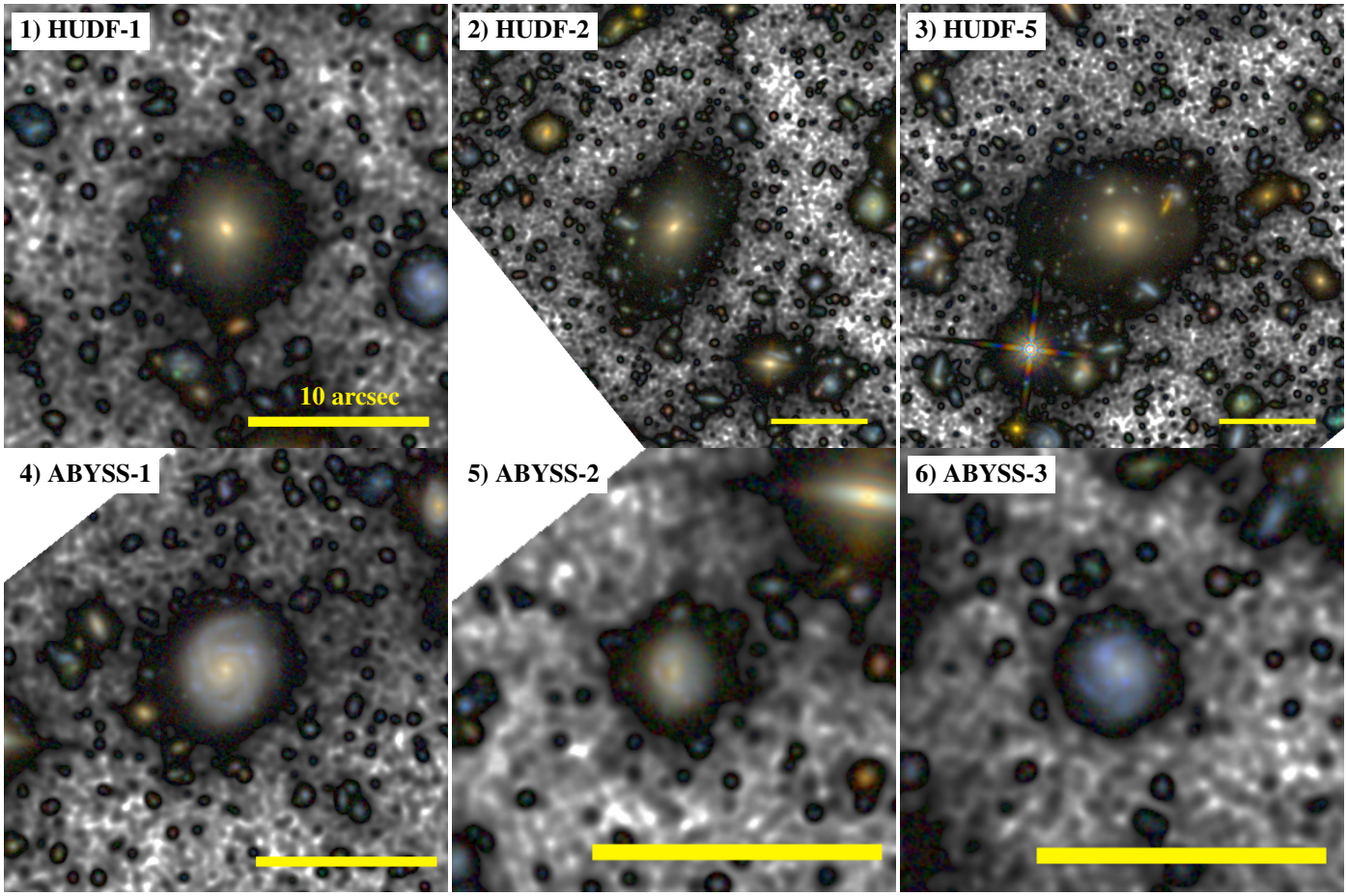

Fig. 17. Luminance-RGB images of the selected targets for the surface brightness profile analysis. *Top row, from left to right*: 1) HUDF-1 (24 × 24 arcsec), 2) HUDF-2 (45 × 45 arcsec), 3) HUDF-5 (45 × 45 arcsec). *Bottom row, from left to right*: 1) ABYSS-1 (24 × 24 arcsec), 2) ABYSS-2 (15 × 15 arcsec), 3) ABYSS-3 (15 × 15 arcsec). The yellow segment represents 10 arcsec in all images. The high S/N parts of the mosaics are represented with colours (red: F160W, green: mean of F125W and F140W bands, blue: F105W). The low S/N regions are represented as a black and white background (black regions are brighter than white regions) according to the mean image of the four mosaics (F105W, F125W, F140W, F160W).

Table 2. Summary of the data used to create the ABYSS HUDF WFC3 IR mosaics and surface brightness limiting magnitude comparison with previous mosaics.

Filter	No. exposures	Exposure time (s)	$\mu_{\rm lim}$ ABYSS (mag arcsec^{-2})	$\mu_{\rm lim}$ XDF (mag arcsec^{-2})	$\mu_{\rm lim}$ HUDF12 (mag arcsec^{-2})
(1)	(2)	(3)	(4)	(5)	(6)
F105W	264	312072	$32.888^{+0.013}_{-0.018}$	$32.877^{+0.017}_{-0.017}$	$32.846^{+0.021}_{-0.017}$
F125W	234	175837	$32.539^{+0.031}_{-0.014}$	$32.517^{+0.026}_{-0.027}$	$32.503^{+0.026}_{-0.024}$
F140W	103	86352	$32.585^{+0.012}_{-0.024}$	$32.517^{+0.014}_{-0.017}$	$32.521^{+0.013}_{-0.012}$
F160W	271	234696	$32.522^{+0.013}_{-0.016}$	$32.677^{+0.018}_{-0.020}$	$32.660^{+0.018}_{-0.019}$

Notes. Columns: (1) Filter identifier. (2) Number of exposures included in the final mosaic. (3) Total exposure time. (4) Median surface brightness limiting magnitude (3σ measured on 10×10 arcsec^2 boxes) for the ABYSS mosaics. (5) Median surface brightness limiting magnitude for the XDF mosaics, measured in the same way. (6) Median surface brightness limiting magnitude for the HUDF12 mosaics, measured in the same way. The surface brightness limits shown here refer to the pixel noise of the images and do not account for systematic effects.

case of HUDF-5. The surface brightness profiles of the ABYSS mosaics recover a significant extension that has been completely removed in the case of XDF and HUDF12. Again, we find that the XDF is the most affected by over subtraction, with $\Delta\mu \sim 1.5$ mag arcsec^{-2} with respect to the ABYSS profile at a surface brightness magnitude of $\mu \sim 29$ mag arcsec^{-2}. We recover also a significant amount of light compared to the HUDF12, reaching $\Delta\mu \sim 1.5$ mag arcsec^{-2} at the limiting radius in the F105W and $\Delta\mu \sim 1$ mag arcsec^{-2} in the rest of the images. The XDF profiles show a down-bending break that is not that strong in the HUDF12, and it is completely removed in our version of the mosaics, where the surface brightness profile reaches the sky level smoothly, following a nearly exponential shape.

Table 3. Selected targets for the surface brightness profile comparison of the HUDF WFC3 IR mosaics.

ID (1)	IAU ID (2)	MUSE ID (3)	α (4) (degrees)	δ (5) (degrees)	z (6)	PA (7) (degrees)	b/a (8)	Sky background limiting radius (9) (kpc)
HUDF-1	J033237.30–274729.3	5	53.16164	−27.78025	0.618	−14.18	0.878	80
HUDF-2	J033241.40–274717.1	870	53.17254	−27.78812	0.619	−30.61	0.599	140
HUDF-5	J033237.30–274729.3	862	53.15545	−27.79150	0.667	75.18	0.900	140
ABYSS-1	J033240.78–274615.6	1	53.16993	−27.77106	0.622	−15.82	0.884	80
ABYSS-2	J033242.25–274625.3	916	53.17606	−27.77371	1.288	−3.31	0.890	80
ABYSS-3	J033237.96–274651.9	7	53.15815	−27.78109	0.620	64.69	0.860	80

Notes. Columns: (1) ID. (2) IAU ID. (3) MUSE catalogue ID. (4) Right ascension (degrees). (5) Declination (degrees). (6) Spectroscopic z from the MUSE HUDF catalogue (Bacon et al. 2017). (7) Position angle (degrees, anti-clockwise from north). (8) Axis ratio of the minor over the major axis. (9) Minimum galactocentric radius used to calculate the local sky background (kpc).

We show the surface brightness profile analysis of HUDF-2 in Fig. A.2. For this object we detect a similar over subtraction of the XDF profiles. The XDF F160W surface brightness profile reaches $S/N = 3$ at half the extension of the ABYSS and HUDF12 profiles. Interestingly, we found that the ABYSS and HUDF12 profile agree remarkably well for this object, reaching the limiting S/N at very similar radius ($R_{\rm lim} \sim 13$ arcsec, with the exception of F125W and F160W which extends to $R \sim 15.5$−16 arcsec). In F160W, we observe that the HUDF12 surface brightness profile is $\Delta\mu \sim 0.3$ mag arcsec^{-2} brighter than the ABYSS between $R = 4$ and $R = 13$ arcsec. For the rest of the filters, the ABYSS surface brightness profiles are $\Delta\mu \sim 0.3$ brighter than the HUDF12 ones beyond $R > 11$ arcsec. This result agrees with the observed difference maps (see Sect. 3.1.1), where we detected more flux on the southern section of the F160W HUDF12 image than in our F160W ABYSS mosaic.

We continue the analysis with the spiral galaxy ABYSS-1 (see Table 3, and bottom left panel of Fig. 17). In Fig. A.3 we compare the surface brightness profiles based on our own reduction of the HUDF mosaics (ABYSS), the HUDF12, and the XDF, for the four different filters (F105W, F125W, F140W and F160W). We summarise the results in several points:

1. As in the case of HUDF-5, the surface brightness profiles from our dedicated mosaics present higher intensity and S/N on the outskirts, up to radius of $R \sim 7$−7.5 arcsec. Similarly to the previous case, the maximum differences in surface brightness range from $\Delta\mu = 0.5$−1.5 mag arcsec^{-2}.
2. The F140W band surface brightness profiles for the ABYSS and XDF version of the mosaics present a more similar shape than in the previous cases.
3. The surface brightness profiles of the four filters of the ABYSS mosaics and the F140W from XDF suggest the presence of an extended component clearly detectable at $R > 5.5$ arcsec, which could be due to PSF effects or an extended stellar halo. The effect is less clear in the F125W image, which reaches $S/N = 3$ at $R \sim 6$ arcsec.
4. There is a clear discrepancy between different filters on the shapes of the profiles for the XDF version of the mosaics, being the F140W the most extended of all XDF surface brightness profiles and the only approaching our own reduction.

The presence of an extended component detectable beyond $R = 5$ arcsec in these surface brightness profile does not necessarily imply that it is associated with a physical component (that is, a stellar halo). In order to identify the true nature of such extended light it would be necessary to estimate the amount of PSF scattered light from the nearby objects and the ABYSS-1 galaxy itself, in a similar way as done in Trujillo & Fliri (2016). As demonstrated in that work, the scattered light of the sources creates a background of light that contaminates the images and the surface brightness profiles. Assuming that the observer knows the behaviour of the PSF, it is possible to model and correct it, recovering the true shape of the objects to a certain degree (see Borlaff et al. 2017, for a similar work with HST ACS data on GOODS-N). Nevertheless, such field of scattered light can be easily mistaken with the sky background. Removing it from the final mosaics in the reduction process neglects the possibility of recovering the information at the very low surface brightness limits.

A detailed analysis of the new low surface brightness structures is well beyond the scope of the present paper and will be addressed in a forthcoming paper (Borlaff et al., in prep.). We conclude that the surface brightness profiles that we present are a valid benchmark between the different reduction processes, as they use the same masks and surface brightness profile analysis and local sky background correction procedure. In addition, we found that our reduction process allows us to recover up to $\Delta\mu \sim 1$−1.5 mag arcsec^{-2} on the surface brightness profiles when compared to the previous versions of the HUDF WFC3 IR mosaics and almost twice the radial size when compared to the XDF surface brightness profiles, even after local sky background correction.

3.2.2. Surface brightness profiles of small objects

In this section we will analyse the properties of the surface brightness profiles of ABYSS-2 and ABYSS-3 (see Table 3), two spiral galaxies with smaller angular sizes compared to the previous targets. The main objective is to do a consistency test and determine if there is any significant differences on the surface brightness profiles of such objects when comparing to the previous versions of the mosaics.

In Fig. A.4 we show the surface brightness profiles of the spiral galaxy ABYSS-2. Given that the aggressive sky background correction applied on the previous versions of the HUDF affected specially to the most extended objects, the surface brightness profiles of more compact objects should be more similar than the large objects, regardless of the version of the mosaics. Indeed, we find that the structure of the surface brightness profiles is more similar in the case of ABYSS-2 than in the previous ones, especially in comparison with HUDF-5. The surface brightness differences are reduced to $\Delta\mu \sim 0.5$ mag arcsec^{-2} at the limiting radius, and the relative differences between the HUDF12 and the XDF are notably smaller. Moreover, we still detect that the F140W profile of XDF has a more similar shape to

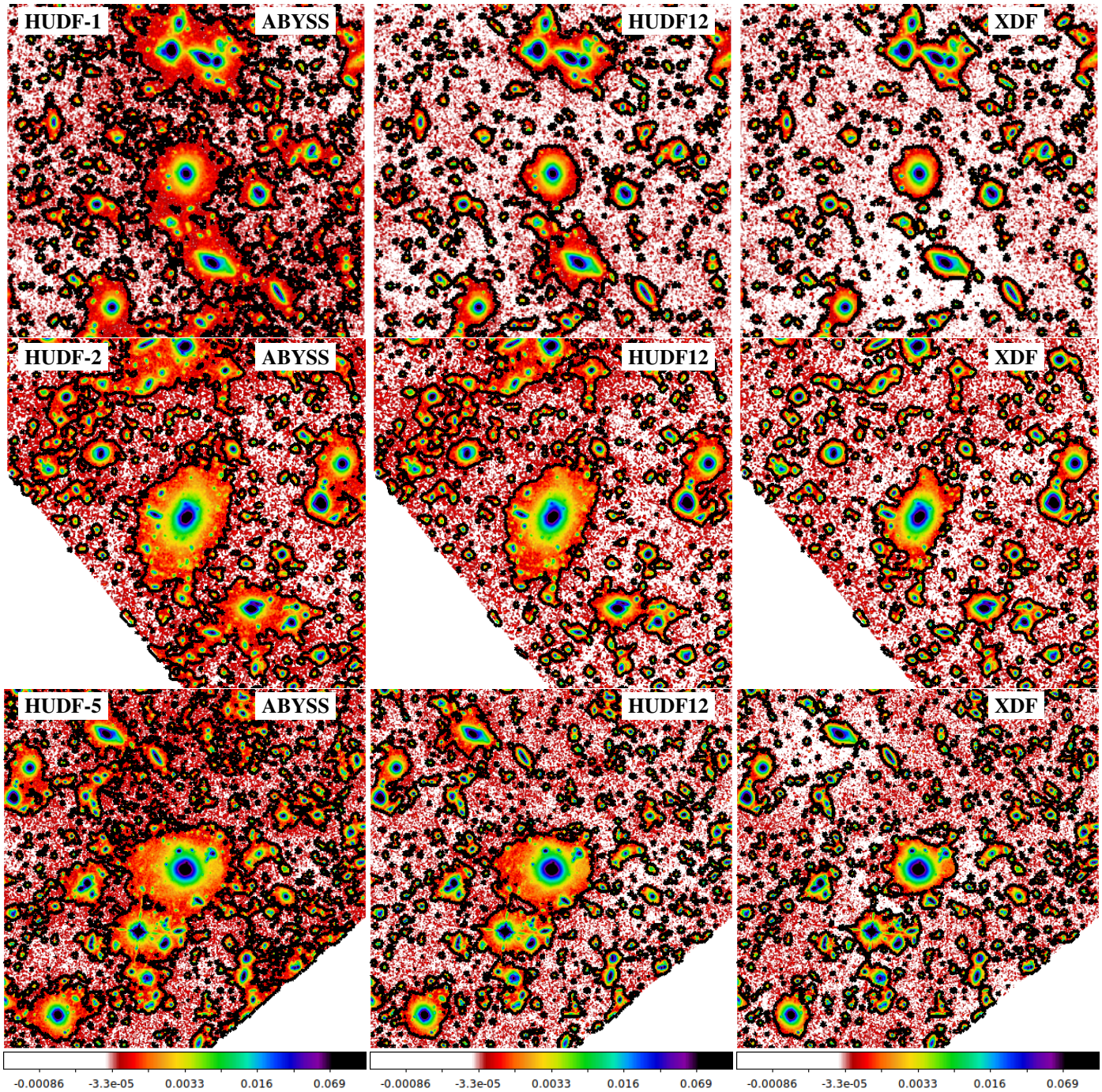

Fig. 18. F105W intensity images of HUDF-1 (*top row*, 54 × 54 arcsec), HUDF-2 (*central row*, 54 × 54 arcsec), and HUDF-5 (*bottom row*, 72 × 72 arcsec; Buitrago et al. 2017) for our version of the HUDF mosaics (ABYSS, *left column*), the HUDF12 mosaics (Koekemoer et al. 2012, *central column*), and the XDF mosaics (Illingworth et al. 2013, *right column*). The black contours represent the $\mu_{\rm F105W} = 29$ mag arcsec^{-2} isophote. All images are at the same colour scale.

our surface brightness profiles of ABYSS mosaics, suggesting that the over-subtraction could be less aggressive in this filter than in the rest. In addition to these results, we detect a tail of extended light in the ABYSS mosaics (possibly due to PSF scattered light) which dominates the surface brightness profile from $R \sim 2$–2.5 arcsec and it is not visible on the previous versions on the HUDF. This result is compatible with the analysis of ABYSS-1.

Finally, we analyse the surface brightness profile of ABYSS-3 (see Fig. A.5). This is the smallest object of all the six selected galaxies. The surface brightness profile shows a clear exponential decline without deviations (with small signs of a down-bending at $R \sim 1.5$ arcsec) until the limiting radius. We find no signs of a bulge at the inner regions. For this object we find a remarkably good agreement between the three versions of the mosaics, obtaining very similar limiting radii for all of them on each band ($R_{\rm lim} \sim 2.8$–3.2 arcsec) and no significant differences on the surface brightness profiles along all the visible radius. We do not find signs of extended scattered light, in contrast to the profiles of ABYSS-2.

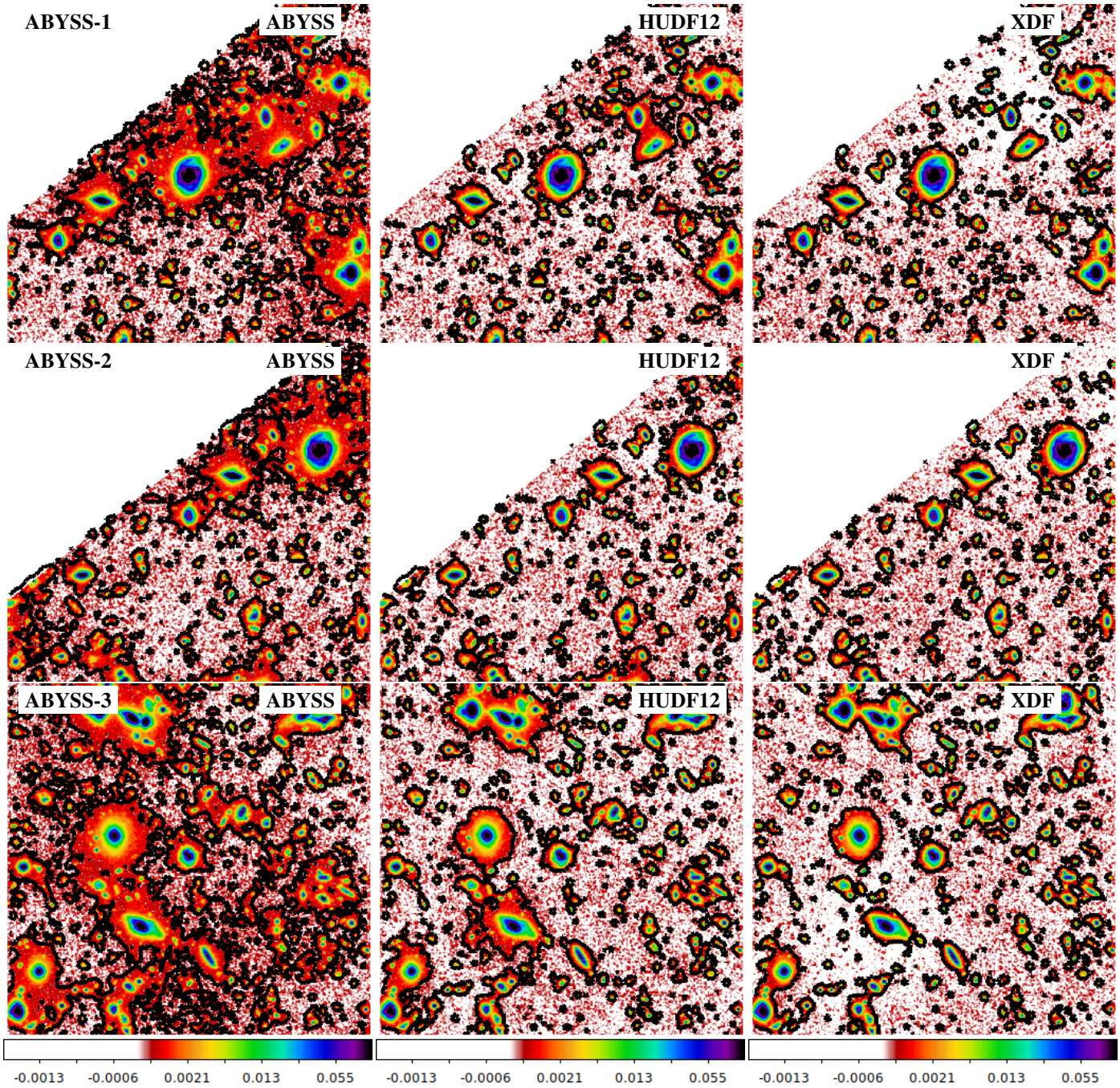

Fig. 19. F105W intensity images of ABYSS-1 (*top row*), ABYSS-2 (*central row*), and ABYSS-3 (*bottom row*) for our version of the HUDF mosaics (ABYSS, *left column*), the HUDF12 mosaics (Koekemoer et al. 2012, *central column*), and the XDF mosaics (Illingworth et al. 2013, *right column*). The field of view is 54×54 arcsec on all the images. The black contours represent the $\mu_{\rm F105W} = 29$ mag arcsec^{-2} isophote. All images are at the same colour scale.

Therefore, the photometry and structure of objects with relatively small angular sizes is similar and compatible to that observed in the previous versions of the HUDF. This analysis provides a consistency test, demonstrating that our mosaics preserve the properties of the small objects while recovering the extended light from the largest sources in the field of view. The results provided here demonstrate that the cause of the differences in flux between the XDF and HUDF12 is a systematic bias caused by over-subtraction of the sky background. We show that our reduction pipeline provides a viable process to reduce systematic biases prior to the co-adding of the final mosaics. This process reduces the need for sky background subtraction, while preserving the structure of the brightest galaxies and their extended envelopes, maintaining the limiting magnitude of the faintest objects at the same time.

4. Conclusions

The low surface brightness Universe is the next frontier for many studies in galaxy evolution and cosmology. Many observational and theoretical works demonstrate that there are very extended and complex structures larger than the visible size of

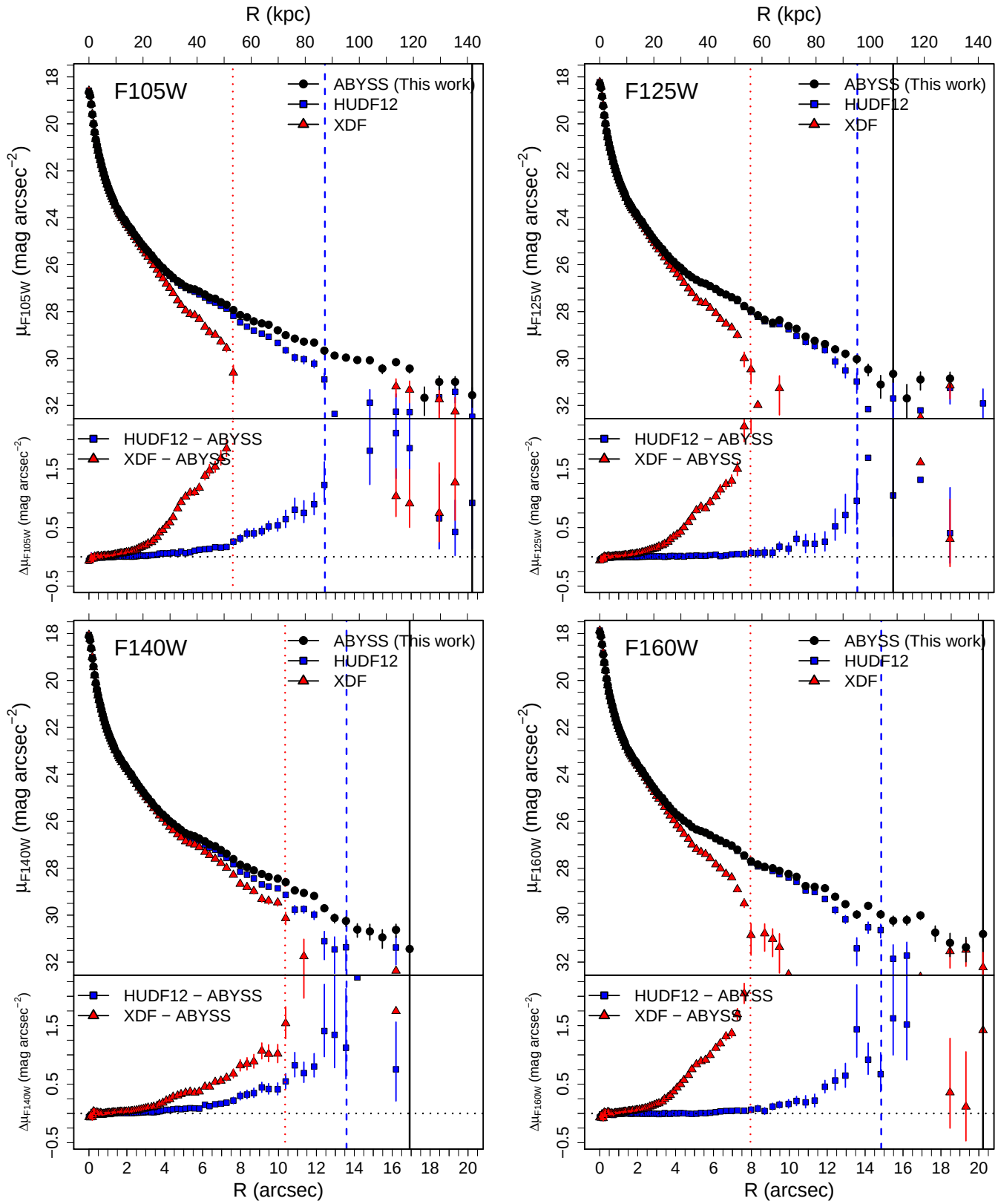

Fig. 20. Comparison of the surface brightness profiles of the elliptical galaxy HUDF-5 (α = 53.15545, δ = −27.79150, Buitrago et al. 2017) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*) and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares) and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.

their host galaxies below the limiting magnitude of most current surveys. Moreover, the cosmological dimming substantially limits our capabilities to study the structure of extended objects at high redshift, most of which are only accessible through space-based observations. It is then mandatory to improve the reduction techniques of the cosmological deep fields from HST and other space telescopes. In this paper we test a number of corrections to improve the low surface brightness limits of the HUDF WFC3 IR mosaics. We have obtained a dedicated version of the images which we named ABYSS which we made publicly available for the benefit and use of all the astronomical community. We found several interesting results:

1. The XDF version of the HUDF WFC3 IR mosaics is dominated by a systematic bias in the form of a significant over-subtraction of the sky background around the objects with large angular size. A similar result (to a lesser extent) is obtained for the HUDF12. We successfully recover a significant amount of over-subtracted diffuse light around the largest objects of the HUDF, not detected by the previous versions of the mosaics. The integrated magnitude of the recovered light is equivalent to a $m \sim 19$ mag object for the XDF and $m \sim 20$ mag for the HUDF12 mosaics, comparable to the brightest galaxies on the image.
2. A significant fraction of the images of the HUDF are (at least partially) affected by persistence effects at the very low surface brightness regime, biasing the sky background estimation. The cause of this is the observation of bright sources (astronomical or calibration runs) in the previous hours to the scheduled HUDF observations.
3. We propose and test a sky background correction method, based on careful masking using noise-based, non-parametric methods as Gnuastro NoiseChisel to detect and flag the extended envelopes of the sources on the field of view. We demonstrate that this method can improve the sky background determination more than one order of magnitude.
4. We studied the surface brightness profiles of six objects in the HUDF, demonstrating that our reduction pipeline can preserve the properties of the smallest sources in the HUDF, while recovering the low surface brightness structures of the outskirts of the largest galaxies.

Systematic biases can dominate over the sky noise. It is for this reason that the noise level of the mosaics is not always a good proxy of the real depth of astronomical images. The measured value of the intensity dispersion of the sky dominated pixels it is not very sensitive to the effects of sky background over-subtraction. Hypothetically, an image reduced with a sky-subtraction process that completely fits and subtracts the sources on the field of view might obtain a similar surface brightness limiting magnitude (measured as the standard deviation of sky-noise level) as a reduction that preserves the outskirts of the largest sources. The latter image clearly contains more information and therefore is deeper than the over subtracted version, despite of having the same formal surface brightness limiting magnitude. We have shown that, despite the small loss of exposure time and the corresponding slight increase of surface brightness limiting magnitudes, our mosaics contain more signal and information about the outskirts of galaxies than previous versions of the WFC3 IR HUDF.

Most, if not all of these problems will be found on the deep observations to be performed by HST near IR successor, the JWST. In particular, persistence will be present in its three image detectors (NIRCam, NIRISS, and MIRI). Mitigation of persistence effects on JWST deep cosmological surveys (Leisenring et al. 2016) have to be based mainly on the following observational strategy: 1) careful scheduling of the previous observations and 2) large dithering patterns to avoid observing regions of the sky with those regions that are affected by persistence on all the exposures. In addition, deep and wide survey dedicated missions, (i.e., EUCLID, MESSIER) can improve their results using these techniques, which would enable a new generation of low surface brightness studies based on their legacy data.

In this paper we have reviewed many of these systematic effects and proposed solutions to them, applying the methods on the deepest image of the Universe ever taken, the HUDF, creating our own version called ABYSS: a low surface brightness dedicated reduction for the HUDF WFC3 IR mosaics. We make the results and the calibration files publicly available to the community - as well as the ABYSS pipeline[16], hoping to promote further analysis and improvements to the proposed reduction methods.

Acknowledgements. The authors would like to thank the XDF and HUDF12 teams for their extraordinary work in which this contribution is based on, truly standing on the shoulders of giants. This work would not have been possible without the kind assistance of all the members of the STScI Help Desk. We specially thank Knox Long by its work on persistence effects and its extraordinary support during this work. A.B. acknowledges support from the International Space Science Institute (ISSI) as part of the Exploring the Ultra-Low Surface Brightness Universe International team. Some/all of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes (MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. I.T. and C.M.L acknowledges the support from the SUNDIAL EU Network and from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No 721463 and the SUNDIAL ITN network. C.M.L. acknowledges financial support from the Spanish Ministry of Economy and Competitiveness (MINECO) under grant number AYA2016-76219-P. N.C acknowledged support from the Spanish Programa Nacional de Astronimía y Astrofísica under grant AYA2016-75808-R. C.G.G. acknowledges support from the European Research Council (ERC) Consolidator Grant funding scheme (project ConTExt, grant number 648179). Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. This work was partly done using GNU Astronomy Utilities Gnuastro version 0.5. Gnuastro is a generic package for astronomical data manipulation and analysis which was primarily created and developed for research funded by the Monbukagakusho (Japanese government) scholarship and ERC advanced grant 339659-MUSICOS. Support for this work was provided by the Spanish Ministerio de Economía y Competitividad (MINECO; grant AYA 2016-77237-C3-I-P). This research made use of Astropy (http://www.astropy.org) a community-developed core Python package for Astronomy (Astropy Collaboration 2013; Price-Whelan et al. 2018).

References

Abadi, M. G., Navarro, J. F., & Steinmetz, M. 2005, MNRAS, 365, 747
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8
Akhlaghi, M. 2016, ArXiv e-prints [arXiv:1611.06387]
Akhlaghi, M., & Ichikawa, T. 2015, ApJS, 220, 1
Aravena, M., Decarli, R., Walter, F., et al. 2016a, ApJ, 833, A71
Aravena, M., Decarli, R., Walter, F., et al. 2016b, ApJ, 833, A68
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Azzollini, R., Trujillo, I., & Beckman, J. E. 2008a, ApJ, 679, L69
Azzollini, R., Trujillo, I., & Beckman, J. E. 2008b, ApJ, 684, 1026
Bacon, R., Conseil, S., Mary, D., et al. 2017, A&A, 608, A1
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bohlin, R. C., Mack, J., & Biretta, J. 2008, Instrument Science Report WFPC2 2008–01, 7
Borlaff, A., Eliche-Moral, M. C., Rodríguez-Pérez, C., et al. 2014, A&A, 570, A103

[16] ABYSS is a low surface brightness dedicated reduction for the HUDF WFC3 IR mosaics. It is freely and publicly available at http://www.iac.es/proyecto/abyss/

Borlaff, A., Eliche-Moral, M. C., Beckman, J. E., et al. 2017, A&A, 604, A119
Borlaff, A., Eliche-Moral, M. C., Beckman, J. E., et al. 2018, A&A, 615, A26
Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2009, ApJ, 709, L133
Brammer, G., Pirzkal, N., McCullough, P., & MacKenty, J. 2014, Instrument Science Report WFC3 2014–03, 14
Buitrago, F., Trujillo, I., Curtis-Lake, E., et al. 2017, MNRAS, 466, 4888
Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931
Bushouse, H. 2008, WFC3 Instrument Science Report 2008–28, 14
Calvi, V. 2014, PhD Thesis, Università degli Studi di Padova, Italy
Chonis, T. S., Martinez-Delgado, D., Gabany, R. J., et al. 2011, ApJ, 142, A166
Comastri, A., Ranalli, P., Iwasawa, K., et al. 2011, A&A, 526, L9
Comerón, S., Elmegreen, B. G., Salo, H., et al. 2012, ApJ, 759, A98
Comerón, S., Elmegreen, B. G., Salo, H., et al. 2014, A&A, 571, A58
Cooper, A. P., Cole, S., Frenk, C. S., et al. 2009, MNRAS, 406, 744
Dressel, L. 2012, Wide Field Camera 3, HST Instrument Handbook (Baltimore: STScI)
Dunlop, J. S., McLure, R. J., Biggs, A. D., et al. 2016, MNRAS, 466, 861
Elmegreen, B. G., & Struck, C. 2016, ApJ, 830, A115
Elmegreen, D. M., Elmegreen, B. G., Rubin, D. S., & Schaffer, M. A. 2005, ApJ, 631, 85
Erwin, P., Beckman, J. E., & Pohlen, M. 2005, ApJ, 626, L81
Erwin, P., Pohlen, M., & Beckman, J. E. 2007, ApJ, 135, 20
Fowler, J., Markwardt, L., Bourque, M., & Anderson, J. 2017, Instrument Science Report WFC3 2017–02 (v.1), 14
Fruchter, A. S., & Hook, R. N. 1998, PASP, 114, 144
Giacconi, R., Zirm, A., Wang, J., et al. 2002, ApJS, 139, 369
Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2003, ApJ, 600, L93
Gonzaga, S., Hack, W., Fruchter, A., & Mack, J. 2012, The DrizzlePac Handbook, HST Data Handbook
Greenfield, P., & Miller, T. 2016, Astron. Comput., 16, 41
Gutierrez, L., Erwin, P., Aladro, R., & Beckman, J. E. 2011, AJ, 142, A145
Herpich, J., Stinson, G. S., Dutton, A. A., et al. 2015a, MNRAS, 448, L99
Herpich, J., Stinson, G. S., Rix, H.-W., Martig, M., & Dutton, A. A. 2015b, MNRAS, 470, 4941
Horton, A., Spitler, L., Mathers, N., et al. 2016, Proc. SPIE, 9904, 99041Q
Ibata, R., Mouhcine, M., & Rejkuba, M. 2009, MNRAS, 395, 126
Ibata, R. A., Lewis, G. F., McConnachie, A. W., et al. 2013, ApJ, 780, A128
Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, A6
Ji, I., Hasan, I., Schmidt, S. J., & Tyson, J. A. 2018, PASP, 130, 084504
Johnston, K. V., Bullock, J. S., Sharma, S., et al. 2008, ApJ, 689, 936
Kazantzidis, S., Zentner, A. R., Kravtsov, A. V., Bullock, J. S., & Debattista, V. P. 2009, ApJ, 700, 1896
Koekemoer, A. M., Gonzaga, S., Fruchter, A., et al. 2002, HST Dither Handbook Version 2.0 (Baltimore: STScI)
Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2012, ApJS, 209, A3
Kozhurina-Platais, V., Dulude, M., Dahlen, T., & Cox, C. 2012, Instrument Science Report WFC 2012–07, 13
Labbe, I., Oesch, P. A., Illingworth, G. D., et al. 2015, ApJS, 221, A23
Laurikainen, E., & Salo, H. 2001, MNRAS, 324, 685
Leisenring, J. M., Rieke, M., Misselt, K., & Robberto, M. 2016, Proc. SPIE, 9915, 99152N
Long, K. S., Baggett, S. M., MacKenty, J. W., & Riess, A. G. 2012, Proc. SPIE, 8442, 84421W
Long, K. S., Baggett, S., & Mackenty, J. W. 2013a, WFC3 Instrument Science Report 2013–07
Long, K. S., Baggett, S. M., & MacKenty, J. W. 2013b, WFC3 Instrument Science Report 2013–06, 10
Long, K. S., Baggett, S. M., & MacKenty, J. W. 2015, WFC3 Instrument Science Report 2015–15, 18
Lucas, R. A. 2015, Instrument Science Report ACS 2015–05, 27
Lucas, R. A., & Hilbert, B. 2015, Instrument Science Report ACS 2015–04, 10
Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2016, ApJS, 228, A2
Maltby, D. T., Aragón-Salamanca, A., Gray, M. E., et al. 2014, MNRAS, 447, 1506
Martinez-Delgado, D., Gabany, R. J., Penarrubia, J., et al. 2008a, ApJ, 692, 955
Martinez-Delgado, D., Penarrubia, J., Gabany, R. J., et al. 2008b, ApJ, 689, 184
McConnachie, A. W., Irwin, M. J., Ibata, R. A., et al. 2009, Nature, 461, 66
Morales, G., Martínez-Delgado, D., Grebel, E. K., et al. 2018, A&A, 614, A143
Nelson, D., Pillepich, A., Genel, S., et al. 2015, Astron. Comput., 13, 12
Oesch, P. A., Bouwens, R. J., Carollo, C. M., et al. 2009, ApJ, 709, L21
Oke, J. B. 1971, AJ, 170, 193
Peacock, M. B., Strader, J., Romanowsky, A. J., & Brodie, J. P. 2014, ApJ, 800, A13
Penarrubia, J., McConnachie, A., & Babul, A. 2006, ApJ, 650, L33
Pirzkal, N., & Hilbert, B. 2012, Instrument Science Report WFC3 2012–15, 10
Pirzkal, N., Viana, A., & Rajan, A. 2010, Instrument Science Report WFC3 2010–06, 16
Pirzkal, N., Mack, J., Dahlen, T., & Sabbi, E. 2011, Instrument Science Report ISR WFC3 2011–11, 20
Pohlen, M., & Trujillo, I. 2006, A&A, 454, 759
Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123
Roškar, R., Debattista, V. P., Stinson, G. S., et al. 2007, ApJ, 675, L65
Rujopakarn, W., Dunlop, J. S., Rieke, G. H., et al. 2016, ApJ, 833, A12
Sandin, C. 2014, A&A, 567, A97
Sandin, C. 2015, A&A, 577, A106
Schweizer, F., & Seitzer, P. 1990, PASP, 102, 615
Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. 2008a, Int. Soc. Opt. Photonics, 7021, 70210J
Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. 2008b, Int. Soc. Opt. Photonics, 7021, 70210K
Spergel, D. N., Bean, R., Doré, O., et al. 2006, ApJS, 170, 377
Steidel, C. C., & Hamilton, D. 1992, AJ, 104, 941
Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996a, AJ, 112, 352
Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996b, ApJ, 462, L17
Struck, C., & Elmegreen, B. G. 2016, MNRAS, 464, 1482
Struck, C., & Elmegreen, B. G. 2017, MNRAS, 469, 1157
Tanaka, M., Chiba, M., Komiyama, Y., Guhathakurta, P., & Kalirai, J. S. 2011, ApJ, 738, A150
Teplitz, H. I., Rafelski, M., Kurczynski, P., et al. 2013, AJ, 146, A6
Trujillo, I., & Bakos, J. 2013, MNRAS, 431, 1121
Trujillo, I., & Fliri, J. 2016, ApJ, 823, 123
Valls-Gabaud, D., & MESSIER Collaboration 2017, IAU Symp., 321, 199
Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Nature, 509, 177
Walter, F., Decarli, R., Aravena, M., et al. 2016, ApJ, 833, A67
Wong, M. H. 2010, WFC3 Instrument Science Report 2010–04, 18
Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, ApJS, 195, A10
Yang, B., Zhu, J., & Song, Y.-Y. 2002, Chin. J. Astron. Astrophys., 2, 474
Younger, J. D., Cox, T. J., Seth, A. C., & Hernquist, L. 2007, ApJ, 670, 269
Zackrisson, E., de Jong, R. S., & Micheva, G. 2012, MNRAS, 421, 190

Appendix A: Surface brightness profiles of the objects commented in Sect. 3.2

In this appendix we show the surface brightness profiles for the five objects (HUDF-1, HUDF-2, ABYSS-1, ABYSS-2, and ABYSS-3, see Table 3) analysed on Sect. 3.2, comparing the results for the ABYSS, XDF, and HUDF12 versions on the mosaics.

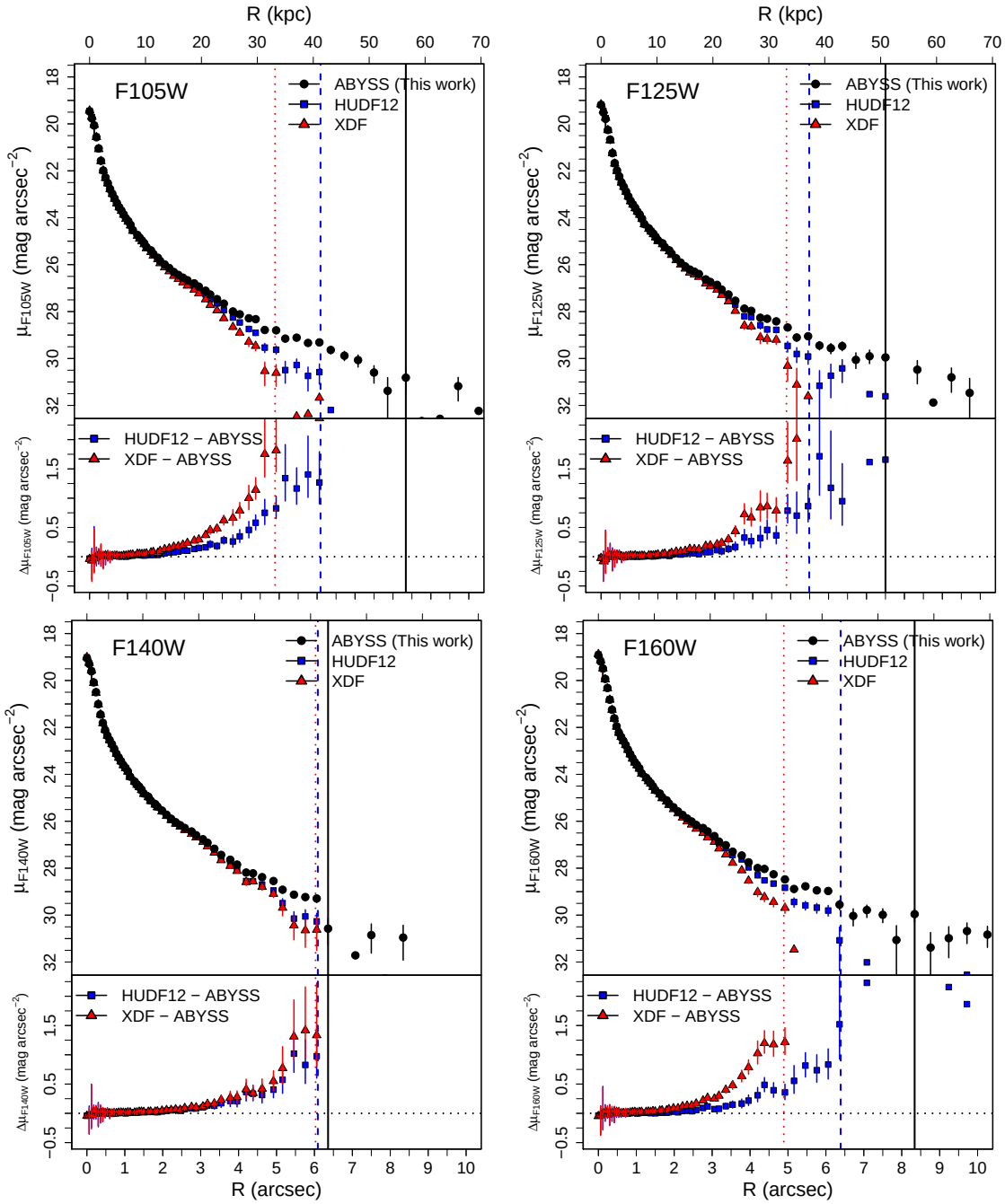

Fig. A.1. Comparison of the surface brightness profiles of the elliptical galaxy HUDF-1 (α = 53.16164, δ = −27.78025, Buitrago et al. 2017) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*), and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares), and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed, and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.

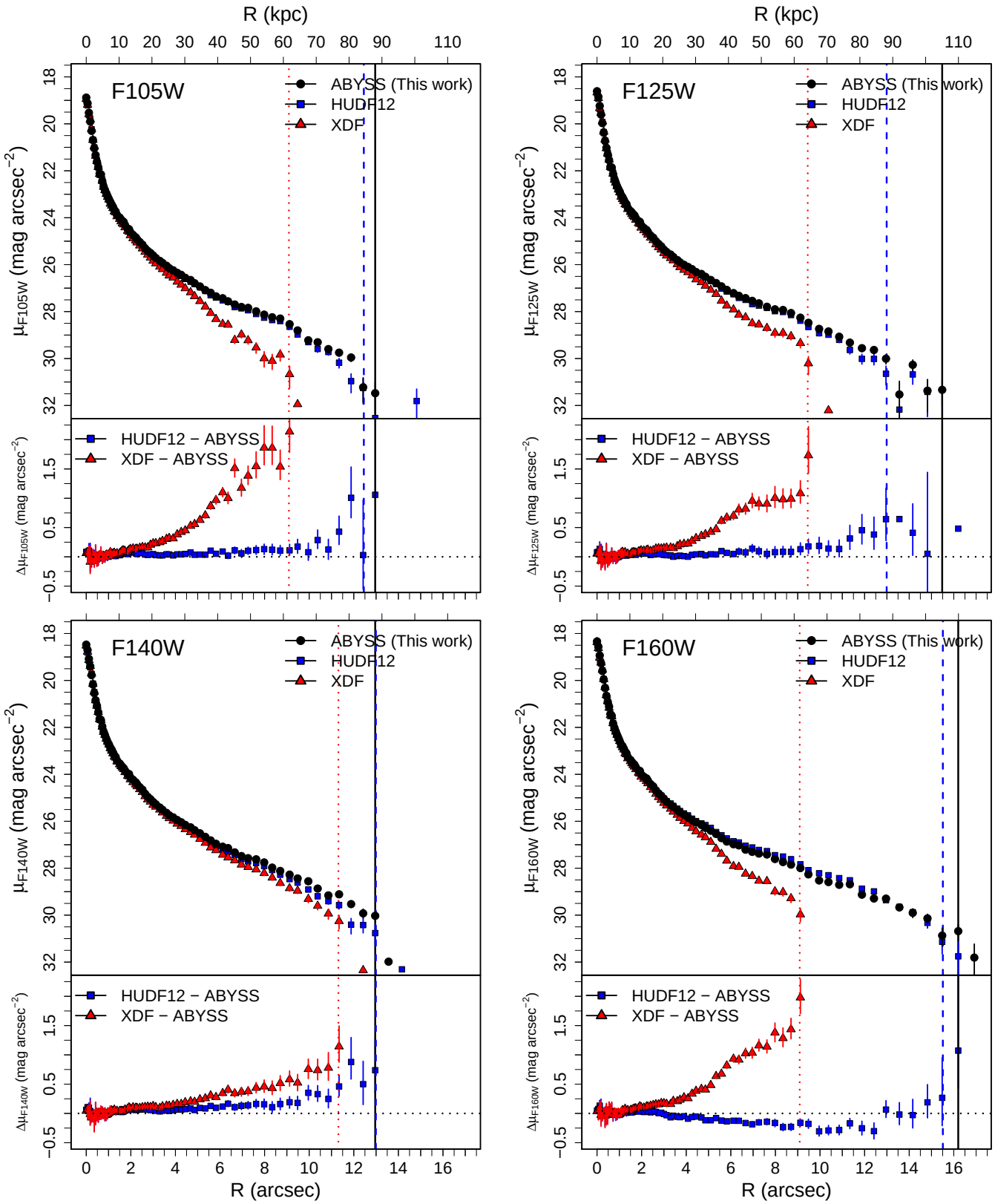

Fig. A.2. Comparison of the surface brightness profiles of the elliptical galaxy HUDF-2 (α = 53.17254, δ = −27.78812, Buitrago et al. 2017) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*), and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares) and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.

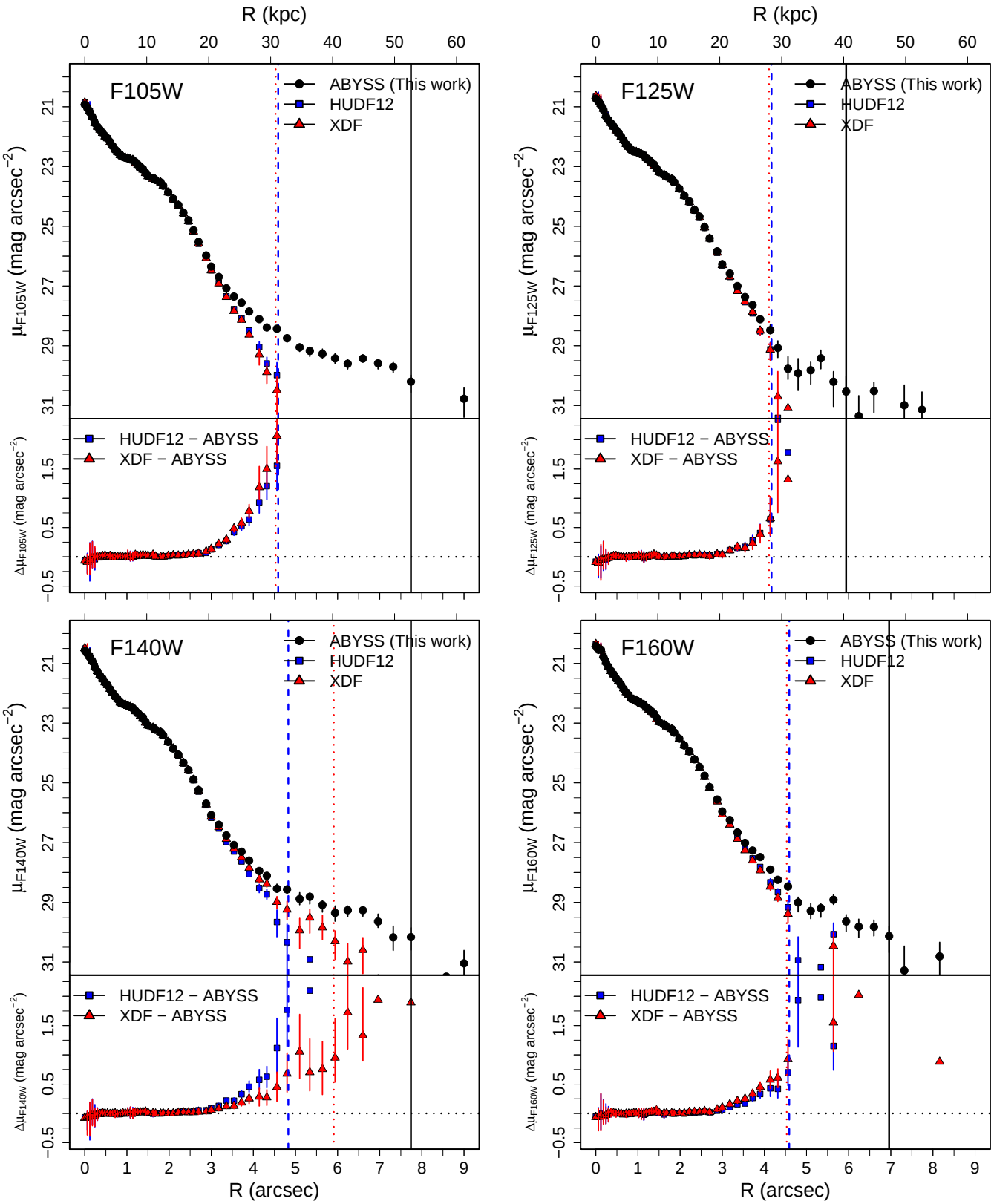

Fig. A.3. Comparison of the surface brightness profiles of the spiral galaxy ABYSS-1 (α = 53.16993, δ = −27.77106, Elmegreen et al. 2005) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*), and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares), and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed, and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.

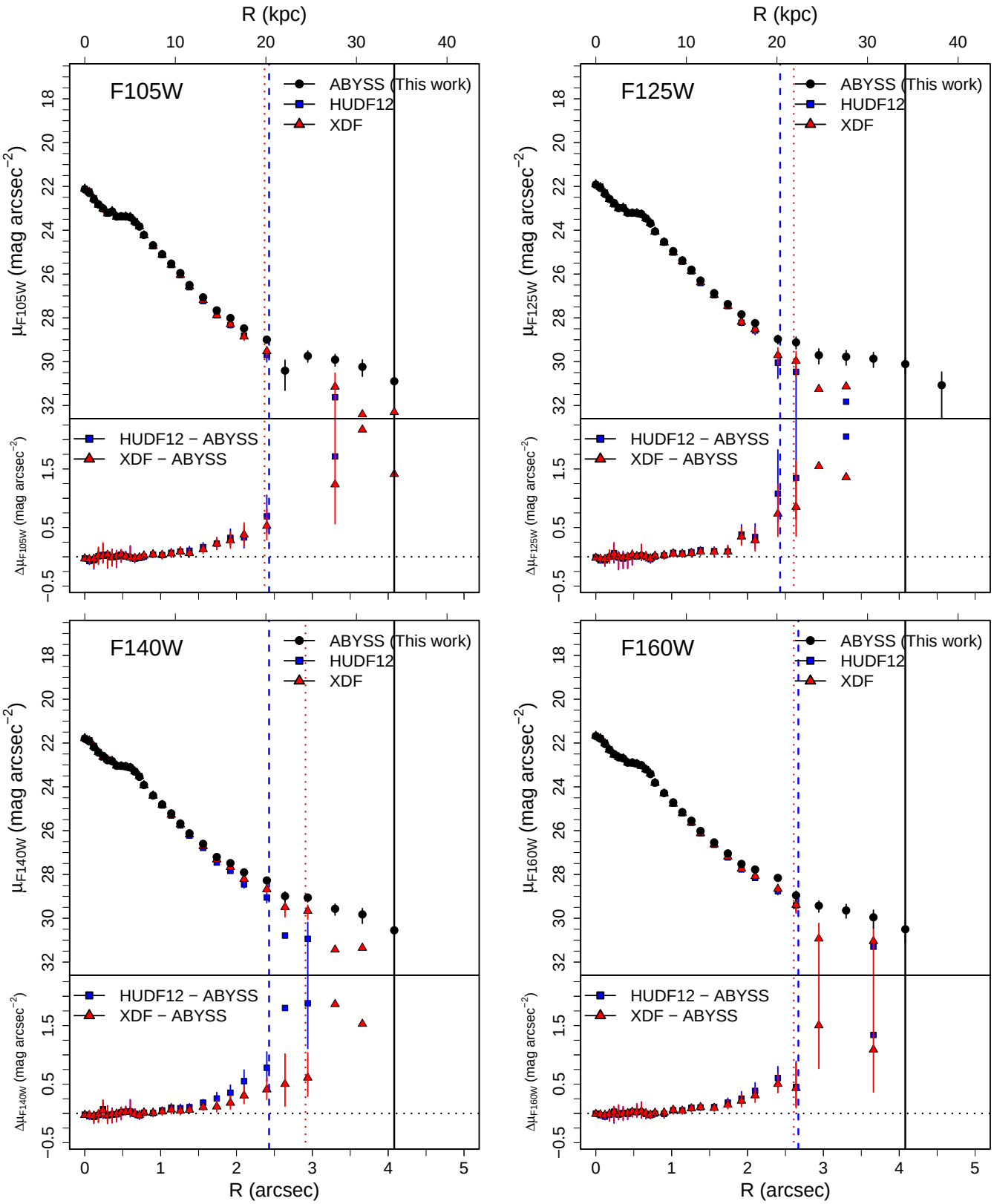

Fig. A.4. Comparison of the surface brightness profiles of the spiral galaxy ABYSS-2 (α = 53.17606, δ = −27.77371, Elmegreen et al. 2005) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*), and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares), and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed, and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.

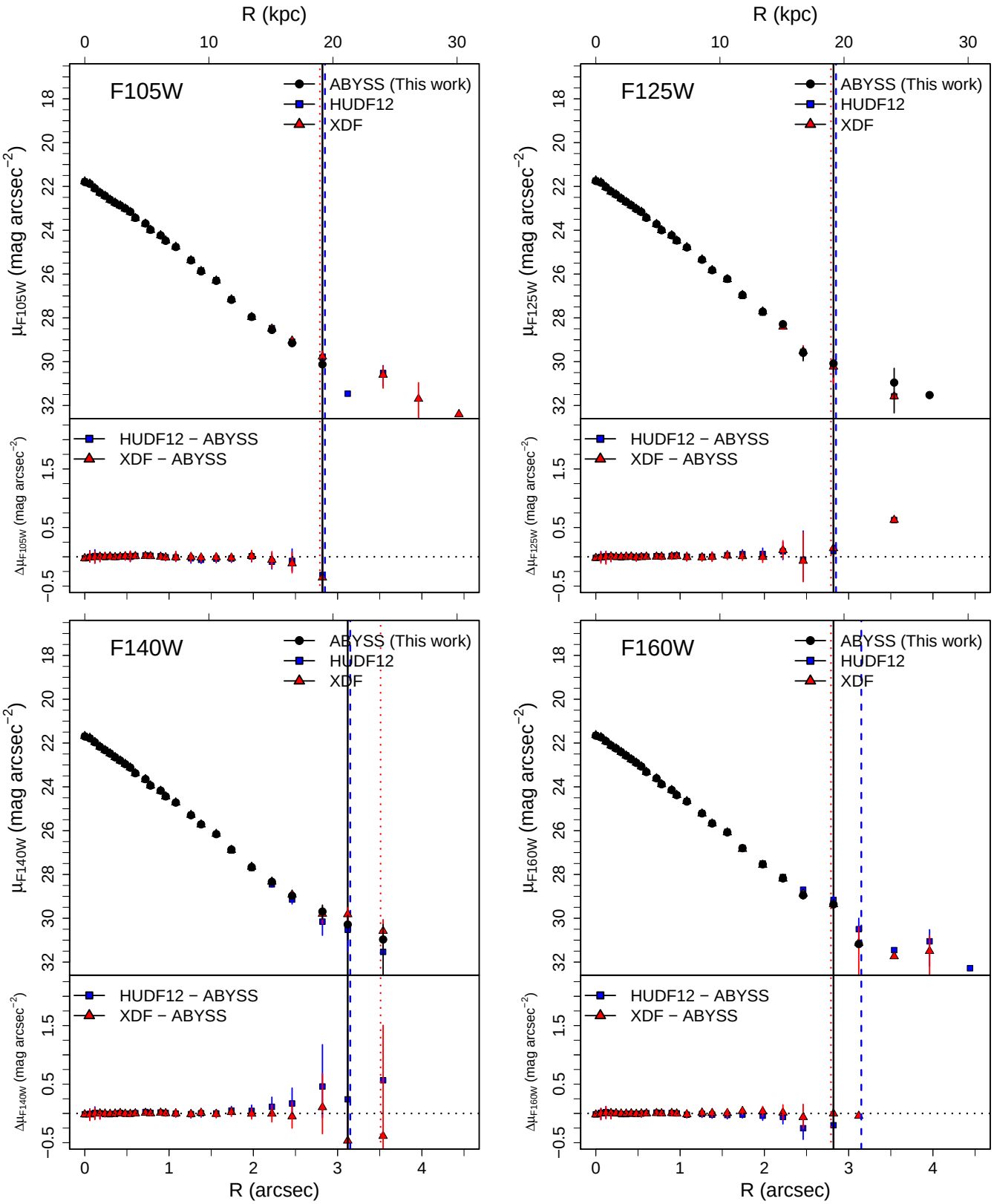

Fig. A.5. Comparison of the surface brightness profiles of the spiral galaxy ABYSS-3 (α = 53.15815, δ = −27.78109, Elmegreen et al. 2005) for the F105W (*top left panel*), F125W (*top right panel*), F140W (*bottom left panel*), and F160W filters (*bottom right panel*), using our own reduction of the HUDF WFC3 mosaics (ABYSS, black dots), the HUDF12 (Koekemoer et al. 2012, blue squares), and the XDF (Illingworth et al. 2013, red triangles). The top plot of each panel shows the surface brightness profile for each reduction. Black solid, blue dashed, and red dotted lines represent the elliptic aperture with largest semi-major axis that presents an S/N higher than 3 over the sky-level. The bottom plot represents the difference in magnitude of each previous reduction with the ABYSS version of the mosaics as a function of galactocentric radius. Consult the legend on the figure.