

DASTWAR: A Tool for Completeness Estimation in Magnitude-size Plane *

*Source of DASTWAR software can be requested by sending e-mail to authors.

Ali Koohpae¹ · Mehdi Khakian Ghomi²

¹ Department of Energy Engineering and Physics,
Amirkabir University of Technology, Tehran, Iran
email: akoohpae@gmail.com

² Department of Energy Engineering and Physics,
Amirkabir University of Technology, Tehran, Iran

Abstract. Today, great observatories around the world, devote a substantial amount of observing time to sky surveys. The resulted images are inputs of source finder modules. These modules search for the target objects and provide us with source catalogues. We sought to quantify the ability of detection tools in recovering faint galaxies regularly encountered in deep surveys. Our approach was based on completeness estimation in magnitude - size plane. The adopted method was incorporating artificial galaxies. We improvised a software that estimates completeness in a given interval of magnitude and size. The software generates artificial galaxies and iteratively inserts them to the source finder modules input image. Evaluating the ratio of the number of detected to the number of inserted artificial galaxies provides us with means to estimate completeness. Completeness estimation is helpful in selecting unbiased samples.

Keywords: Galaxies: structure, galaxies: size, magnitude

1 Introduction

The past two decades have seen the growing number of imaging surveys of the extragalactic sky. Deep field optical/NIR imaging surveys such as COSMOS [19], HUDF [1], CANDELS [16] and GOODS [14] have become the frontier of astronomical studies in various topics. Moreover, within the next few years, imaging surveys with unprecedented depth and area (e.g. LSST [18] and Euclid [21]) will take place.

Any imaging survey is restricted and biased in its sampling of the galaxy population by a number of selection effects (e.g. [10, 11, 17, 4]). The visibility of a particular galaxy depends both on its intrinsic properties (e.g. brightness, light profile, apparent scale size) and the nature of the survey imaging data (e.g. exposure time and sky brightness) [4].

When CCD is used in acquisition of imaging data, outcome images will be of digital type. Acquired images, after passing the process of data reduction, will be given to source finder modules. These modules will identify the sources targeted by the survey and provide us with their photometric and structural properties.

In the case of galaxies, surface brightness is seen to be a key factor in their detectability by source finder modules. Generally, galaxies with lower surface brightness are harder to detect [14, 2, 4]. However, detection is a complex process and surface brightness is not the only factor in determining the detectability of a particular galaxy. For instance, blending with other sources, image artefacts as well as structural properties such as morphological

type and position angle can be mentioned as factors playing a role in their detectability [3]. There is no strict low surface brightness threshold, above which, the detectability of galaxies is assured. We expect that, as the surface brightness of a particular galaxy decreases, the probability of its detection also reduces. Completeness parameter has been defined to quantify the probability of detection [29, 20, 31, 7]. This parameter is defined as the ratio of the number of detected objects to the total number of objects present in the image.

The prospect of forthcoming imaging surveys with unprecedented depth and area testifies to the significance of automated and efficient modules to evaluate their completeness. In this paper, we describe a software package which is improvised to estimate completeness of galaxy detection as a function of apparent magnitude and half-light radius. This paper is organized as follows. In section 2, the methodology of the software is described. Section 3 is devoted to the study of the usage and efficiency of software. Ultimately, summary will be presented in section 4.

2 Methods

We based the completeness estimation procedure on incorporation of artificial galaxies. The core of **DASTWAR**¹ is an IRAF script written in IRAF command language. The software generates artificial galaxies and iteratively inserts them to the source finder modules input image (cf. [29, 20, 14, 4, 6]). Next, it utilizes the source finder module to detect the inserted artificial galaxies. By comparing the extracted catalog to the catalog of artificial galaxies inserted to the input image, completeness would be estimated. Completeness is defined as the ratio of the number of extracted artificial galaxies to the number of artificial galaxies present in the image. **DASTWAR** performs completeness estimation as a function of apparent magnitude and half-light radius of artificial galaxies (e.g. [14, 2, 4]). Inserting artificial galaxies to the observed image preserves any observational artifacts and sky noise when quantifying the probability of detection [6].

The software makes use of **IRAF artdada** package for generating artificial galaxies². This package has been widely used to simulate galaxies in deep images (e.g. [14, 4]). Simulated galaxies are of either early-type or late-type morphology, respectively obeying de Vaucouleurs [9] and Exponential [12] surface brightness laws. The package enables the generation of artificial galaxies in a given bin of apparent magnitude and half-light radius. The software makes use of **SExtractor** [3] as source finder module. **SExtractor** is the standard detecting tool in extracting galaxies based on deep optical/NIR images. **SExtractor** isolates sources in the image given as input, and carries out photometric and structural measurements. Also, a catalogue of detected sources along with their photometric and structural parameters is returned at the end.

The workflow of the procedure is depicted in Fig. 1. The procedure starts with obtaining the values of the input parameters and the input image, which are provided by the user (see Table 1). Next, the software initiates generating the simulated images. A simulated image is a modified version of the input image. This modified version is constructed by inserting the artificial galaxies to the input image. The software utilizes **IRAF artdada** package to produce artificial galaxies according to the prescriptions of the user indicated by values of input parameters. Among the parameters which could be set by the user are the number of artificial galaxies to insert, their apparent magnitude and half-light radius tolerance, their morphological distribution and their inclination tolerance (see Table 1). It should be noted that, the software would not modify the input image before inserting artificial galaxies.

¹Dastwar (pronounced Dastoor in present day persian) is the persian word for adviser.

²<ftp://iraf.noao.edu/iraf/docs/glos210b.ps.Z>

Figure 1: Workflow of the software is illustrated (see text).

Hence, the user has to decide whether to mask real sources or to leave them intact in the input image.

A two-dimensional point spread function, provided by the user, will be convolved with artificial galaxies before inserting them to the input image. Artificial galaxies will be uniformly distributed throughout a subregion of the input image defined by the user. Also their apparent magnitude and half-light radius will be uniformly distributed in the apparent magnitude and half-light radius tolerances indicated by the user. For subsequent referencing, properties of the inserted artificial galaxies, including their positions, apparent magnitudes and half-light radii will be saved in a catalog.

The process of generating the simulated images is iterative. The outcome of each iteration is a single simulated image (see Fig. 2). The total number of iterations will be set by giving value to the appropriate parameter. Willing to end up with an adequate completeness estimation necessitates balance between the number of artificial galaxies in each simulated image and the total number of iterations. When sufficient number of simulated images are generated, software moves forward to the next step.

As a result of complex observing strategies which are at work in deep imaging surveys (e.g. dithering [22]), yielded images are not generally associated with flat edges. Also, these images normally result from stacking a number of slide images on top of one another. In consequence, the edges of the obtained images are usually indented and a set of pixels in the image array are seen to have zero value. In such an image, the entire area is not covered by data. If user notifies the software of the partial data-coverage in the input image, a mask image will be created. It is an image with the same width and height as the input image, in which the partial area covered by data is indicated. This mask image will be multiplied with each of the simulated images. In this way, the analogy of the data-covered area in the input image and the simulated images is assured. If user has not warned about the partial data-coverage in the input image, software leaps to the next step.

Now, all is at hand to start source detection. For this task, **DASTWAR** makes use of **SExtractor** [3]. **SExtractor** will be executed on each of the simulated images based on the input parameters set by the user. Hence, for each simulated image, we will be provided by a catalog of detected sources along with their photometric and structural properties.

Amongst the measured quantities for each detected source is the pixel position of the center. These positions will be used to crossmatch the **SExtractor** provided catalog with the catalog of artificial galaxies inserted to each simulated image. The radius of crossmatch will be set by the user. An artificial galaxy is designated as recovered if centroid of a unique detected source falls within its circle of crossmatch. For each simulated image, recovered galaxies will be listed in a new catalogue. Each line of this catalogue represents an artificial galaxy which was successfully detected by **SExtractor**.

At this point, for each of the simulated images, two catalogs are at hand. The first one is the catalog of inserted artificial galaxies and the second one is the catalog of recovered artificial galaxies. By accumulating the catalogs of inserted artificial galaxies into one catalog, we end up in the master catalog of artificial galaxies. This master catalog enlists all of the artificial galaxies inserted to the set of simulated images. In the same manner master catalog of recovered artificial galaxies is constructed. The latter catalog embraces the list of all artificial galaxies which are already detected by the source finder module. Comparing these two master catalogs enable us to quantify the degree of completeness. **DASTWAR** estimates completeness as a function of artificial galaxies' apparent magnitude and half-light radius.

Table 1: Input parameters of DASTWAR are listed in this table. IRAF artdada package parameters are marked with star.

Parameter	Data Type	Description
nrun	int	Number of iterations
ngal	int	Number of galaxies generated in each iteration
xcormin	real	Minimum X coordinate of artificial galaxies
ycormin	real	Minimum Y coordinate of artificial galaxies
xcormax	real	Maximum X coordinate of artificial galaxies
ycormax	real	Maximum Y coordinate of artificial galaxies
magmin	real	Upper magnitude limit for artificial galaxies
magmax	real	Lower magnitude limit for artificial galaxies
minrad	real	Minimum half-light ratio of artificial galaxies
maxrad	real	Maximum half-light ratio to artificial galaxies
*efrac	real	Fraction of early-type galaxies
*axisrat	real	Minimum axis ratio for early-type galaxies
*srefrac	real	Late-type/early-type radius at a given magnitude
*abs	real	Absorption in edge-on late-type galaxies
inpimage	char	Input image name
wimage	char	Weight image name
psf	char	PSF image name
*poinoi	bool	Add Poisson noise?
*rad	real	Seeing radius/scale (pixels)
*psfar	real	Star/PSF axial ratio
*psfpa	real	Star/PSF position angle
*magzp	real	Magnitude zero point
*ccdgain	real	Gain
*ccdreadnoise	real	CCD Read noise
seconfig	char	Name of SExtractor configuration file
sennw	char	Name of SExtractor Neural network/weights file
separam	char	Name of SExtractor Parameter file
seconv	char	Name of SExtractor convolution kernel
dx_gal	int	X-range for coverage tests (artificial galaxies)
dy_gal	int	Y-range for coverage tests (artificial galaxies)
dx_sex	int	X-range for coverage tests (detected sources)
dy_sex	int	Y-range for coverage tests (detected sources)
magbin	int	Number of magnitude bins for completeness estimation
sizebin	int	Number of size bins for completeness estimation
covcheck	bool	Apply coverage checks?
crop	bool	Apply cropping?
clean	int	Delete additional files made?
pixsize	real	Image pixel scale (arcsec/pix)
maxdist_arcs	real	Cross-match radius (arcsec)
compoutput	char	Completeness matrix file name
*background	real	Default background
*nxc	int	Number of PSF centers per pixel in X
*nyc	int	Number of PSF centers per pixel in Y
*nxsub	int	Number of pixel subsamples in X
*nysub	int	Number of pixel subsamples in Y
*nxgsub	int	Number of galaxy pixel subsamples in X
*nygsub	int	Number of galaxy pixel subsamples in Y
*dyrange	real	Profile intensity dynamic range
*psfrange	real	PSF convolution dynamic range

Figure 2: Five examples of simulated images are shown in this figure. In the uppermost left mosaic, we have shown the input image and the remaining mosaics illustrate examples of simulated images. The input image is a 512×512 pixels cutout of v2.0 images of HST/ACS in southern GOODS field which is acquired in F850LP band.

Table 2: Five lines of DASTWAR’s output file are given as example.

Label 1	Label 2	Magnitude	Half-light radius	Completeness	Detected	Inserted
31	10	20.95	0.9189	0.842105	32	38
31	11	21.05	0.9189	0.932203	55	59
31	12	21.15	0.9189	0.886364	39	44
31	13	21.25	0.9189	0.903846	47	52
31	14	21.35	0.9189	0.857143	48	56

As was noted earlier, galaxies will be uniformly distributed throughout a subregion of the input image that is defined by user. Occasionally, an inserted artificial galaxy would reside in a position too close to the edge of the image or edge of the data-covered area. In such instances, the light profile of the artificial galaxy may become cropped; a phenomenon which usually results in its erroneous detection. When user warns DASTWAR of the possibility of existence of such sources, software attempts to identify them. This is done by defining a rectangular mask for each artificial galaxy, width and length of which is to be determined by the user. Center of this mask will be coincided to the center of each of the inserted artificial galaxies. Inspecting values of the pixels residing inside the mask would characterize the distance between the object and the edges. When an object is identified as being too close to the edges, it will be marked with edge-grazing flag.

The software proceeds to compute completeness as a function of apparent magnitude and half-light radius. Completeness is defined as the ratio of the number of detected artificial galaxies to the number of inserted galaxies not marked with edge-grazing flag. Completeness is estimated as a function of artificial galaxies apparent magnitude and half-light radius. Accordingly, inserted and recovered artificial galaxies are enumerated in bins of apparent magnitude and half-light radius. The plane of apparent magnitude and half-light radius is divided to two dimensional bins. Number of these bins will be determined by user and completeness will be assessed specifically in each bin.

The output of the software is a text file, each line of which provides the result of completeness estimation for each of the two dimensional bins. The first two columns contain two labels which uniquely designate every two dimensional bin. In the third and fourth columns magnitude and half-light radius of the center of two dimensional bin are given respectively. Completeness value for the two dimensional bin is written in the fifth column. Finally, in sixth and seventh column, the number of detected and inserted artificial galaxies for each bin are given.

3 Example

In this section, we intend to demonstrate the usage and efficiency of the software. The inspection is based on Hubble Space Telescopes data acquired during GOODS ⁴ survey [14].

⁴Great Observatories Origins Deep Survey

The input image given to the software for completeness estimation has been selected from southern GOODS field and covers nearly 25 square arcminutes. We have used images taken in F850LP band which is the band normally utilized for detection [8]. We based our study on version v2.0 of HST/ACS data [13].

The number of iterations was set to 165 and in each iteration 1500 artificial galaxies generated. In total, 247500 artificial galaxies were used in the procedure. Artificial galaxies uniformly populate the apparent magnitude range $20 \leq m \leq 30$ (AB magnitude system measured in F850LP band) and half-light radius range $0.01 \leq r_{50} \leq 1.5$ arcseconds. The F850LP-band magnitude zero point was set to 24.862 which is obtained from this URL ⁶ [30]. The apparent magnitude range was divided to 100 bins while the half-light radius range was intersected to 50 bins. The average number of inserted galaxies in each two-dimensional bin is 50.

The fraction of early-type galaxies to late-type galaxies was set to one. Also, the ratio between half-flux scale radii of late-type and early-type galaxies at a given magnitude was equal to one. For early-type galaxies the axial ratio (b/a) was randomly selected in the range $0.3 \leq b/a \leq 1.0$. For late-type galaxies, inclinations range uniformly between 0 and 90 degrees. We did not apply internal absorption correction.

As was noted in §2, before being inserted to input image, surface brightness profile of the artificial galaxy is to be convolved with appropriate point spread function. For the present study, point spread function is inferred from detailed examination of surface brightness profiles of spectroscopically confirmed stars. Using the updated version (v2) of the GOODS-MUSIC catalogue [15, 28], we selected 138 objects with spectroscopic redshift quality flag < 2 and $z_{\text{spec}} = 0$ as stars (cf. [5]). Moreover, we included 63 stars in GOODS-South field identified by [23] based on low resolution spectra acquired in PEARS survey [24]. Of 172 unique stars thus spotted throughout the southern GOODS field, 45 stars reside within the region covered in our input image. Surface brightness profile of 42 stars of the selected sample were used to construct point spread function. For construction of point spread function, we utilized IRAF DAOPHOT package.

When convolved with point spread function, artificial galaxies inserted to the input image without additional Poisson noise. As a result, for bright objects, the noise is slightly underestimated while for faint objects this shortcut does not affect the results as the background completely dominates (cf. [14]). We used the **SExtractor** configuration files which were optimized for detection based on v2.0 of HST/ACS images of southern GOODS field in F850LP band and are publicly available through this URL ⁷

The crossmatch radius used to isolate the recovered artificial galaxies was 0.15 arcsec. Recalling the value of 0.03 arcsec/pix for pixel scale, 0.15 arcsec is equivalent to 5 pixels in the input image. The crossmatch radius was made conservatively small to ensure that the chance of erroneous matches to existing objects is negligible (cf. [4]). The width and length of the rectangular mask defined to identify edge-grazed galaxies were equally set to 3 pixels.

The resulting distribution of completeness in the plane of apparent magnitude and half-light radius is shown in Fig. 3. Contours of constant completeness are illustrated. It is seen that an increase in apparent magnitude in constant half light radius is associated with a decrease in completeness. The same behavior is seen when half-light radius is increased in fixed apparent magnitude. Such a trend introduces surface brightness as a key factor in detectability of galaxies.

Moreover, it should be noted that constant completeness contours tend to fainter apparent magnitudes as half-light radius becomes smaller. For instance, at $r_{50} = 1.5$ arcsec,

⁶http://archive.stsci.edu/pub/hlsp/goods/v2/h_goods_v2.0_rdm.html

⁷http://archive.stsci.edu/pub/hlsp/goods/catlog_r2/h_r2.0z_readme.html

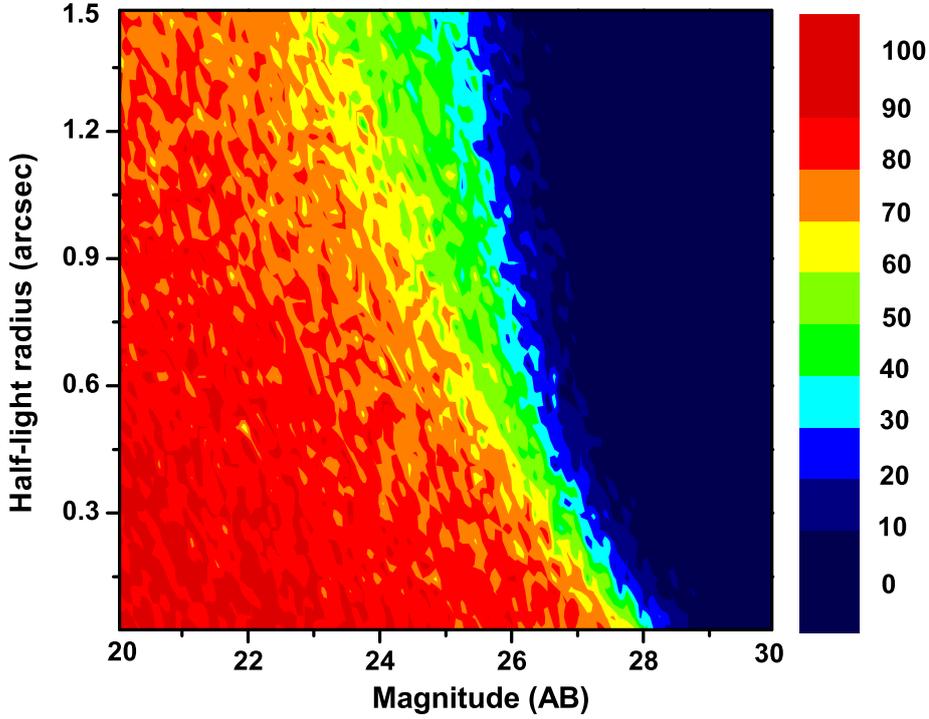


Figure 3: Contour plot representing completeness values in a plane of F850LP magnitude and half-light radius (in units of arcsec.) is shown.

the 70% completeness contour is located at $m \sim 23$ while at $r_{50} = 0.15$ arcsec the same contour is seen at $m \sim 27$. This fact also testifies to the crucial role of surface brightness in detectability of galaxies. It is also seen that the space between the adjacent contours are not uniform. In high surface brightness areas, the contours are more apart compared to low surface brightness regions. This fact reflects that, in spite of its importance, surface brightness is not the only factor that influences the detectability of galaxies.

In a similar study which dates back to 2004 [14], Giavalisco et al. assessed completeness limits based on version v0.5 of HST/ACS images acquired in southern GOODS field in F850LP band. They adopted an analogous method for estimation of completeness as a function of apparent magnitude and half-light radius (see Fig. 4 in [14]). In their study, IRAF `artdata` package was utilized for generating artificial galaxies and `SExtractor` was used as source finder module. Their artificial galaxies uniformly populated the magnitude range $20m28$ (AB magnitude system measured in F850LP band) and the range of galaxy half-light radius was $0.01 \leq r_{50} \leq 1.5$ arcseconds. Morphologically, half of their generated galaxies were of early-type morphology and the remaining half were of late-type morphology. Early-type galaxies had a uniform axial ratio distribution in the range $0.3 \leq b/a \leq 0.9$. Internal absorption was ignored and no additional Poisson noise was at work.

The methodological homology noticed between the two studies enables the comparison of

results and interpretation of the differences in terms of differences in the inputs. Comparison between the two completeness distributions reveals the shifting of completeness contours toward fainter apparent magnitudes in our distribution. Hence, our study implies a higher completeness at a given apparent magnitude in constant half-light radius.

Such a difference can be attributed to the difference in depth of the input images used for the two studies. Our study was based on version v2.0 [13], a significant improvement upon the previous v1.0 release, which is itself an improved version of v0.5 release of GOODS reduced HST/ACS images. Version v1.0 data release provided data acquired as part of the original GOODS HST/ACS program [14]. Version 2.0 augments this with additional data acquired on the two GOODS fields during the search for high redshift Type Ia supernovae carried out during Cycles 12 and 13 (Program ID 9727, P.I. Saul Perlmutter, and 9728, 10339,10340, P.I. Adam Riess [27, 26, 25]. As a result of the additional data, the v2.0 mosaics offer roughly twice the exposure time in the F850LP band compared to version v1.0 images.

4 Summary

Within the next few years, imaging surveys with unprecedented depth and area will revolutionize our vision of the extragalactic sky. Nevertheless, any imaging survey is restricted and biased in its sampling of the galaxy population. Completeness parameter, which quantifies the probability of detection, has proved to be a useful and conventional parameter in assessing the bias in sampling of galaxies. Given the prospect of forthcoming imaging surveys, automated and efficient modules to evaluate their completeness are demanded.

Throughout this paper, we described a software package, named **DASTWAR**, which was improvised to estimate completeness of galaxy detection as a function of apparent magnitude and half-light radius parameters. The software generates artificial galaxies and iteratively inserts them into the input image and then utilizes source finder module to detect them. Comparing the extracted catalog with the catalog of artificial galaxies inserted to input image, yields completeness in the magnitude-size plane (see section 2).

In order to demonstrate the efficiency of the software, we utilized it for completeness estimation on the basis of version v2.0 HST/ACS data in southern GOODS field. In total, 247500 artificial galaxies were generated and used in the procedure. Distribution of completeness values in the magnitude-size plane shows that an increase in apparent magnitude in constant half-light radius is associated with a decrease in completeness and the same trend is noticed when half-light radius is increased in fixed apparent magnitude. We interpreted the mentioned trend as an evidence of pivotal role of surface brightness in determining detectability of galaxies.

We also compared the resulted completeness distribution with the corresponding distribution given by [14]. Comparison between the two completeness distributions revealed the shifting of completeness contours toward fainter apparent magnitudes in our distribution. Such a difference was expected, given the substantially higher depth of v2.0 compared to v0.5 HST/ACS data in southern GOODS field. We envisage that the improvised software would be effective in estimating completeness and helpful in quantifying the biases in sampling of the galaxy population.

Acknowledgment

The cornerstone of the software was laid down when A. K. had been collaborating with School of Astronomy of Institute for Studies in Basic Sciences (IPM), Tehran, Iran. He would like to express his gratitude to Reza Mansouri, Bahram Mobasher, Habib Khosroshahi, Saeed Tavasoli and administrators of the school for their financial and scientific support.

References

- [1] Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, *AJ*, 132, 1729
- [2] Bouwens, R. J., Illingworth, G. D., Blakeslee, J. P., Broadhurst, T. J., & Franx, M. 2004, *ApJL*, 611, L1
- [3] Bertin, E., & Arnouts, S. 1996, *A&AS*, 117, 393
- [4] Cameron, E., & Driver, S. P. 2007, *MNRAS*, 377, 523
- [5] Cameron, E., Carollo, C. M., Oesch, P. A., et al. 2011, *ApJ*, 743, 146
- [6] Chiu, I., Desai, S., & Liu, J. 2016, *Astronomy and Computing*, 16, 79
- [7] Chiu, I., Dietrich, J. P., Mohr, J., et al. 2016, *MNRAS*, 457, 3050
- [8] Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, *ApJ*, 724, 425
- [9] de Vaucouleurs, G. 1948, *Annales d'Astrophysique*, 11, 247
- [10] Disney, M. J. 1976, *Nature*, 263, 573
- [11] Disney, M.; Phillipps, S., 1983, *MNRAS*, 205, 1253
- [12] Freeman, K. C. 1970, *ApJ*, 160, 811
- [13] Giavalisco, M., and the GOODS Team, 2008, in preparation
- [14] Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, *ApJL*, 600, L93
- [15] Grazian, A., Fontana, A., de Santis, C., et al. 2006, *A&A*, 449, 951
- [16] Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, *ApJS*, 197, 35
- [17] Impey, C., & Bothun, G. 1997, *ARA&A*, 35, 267
- [18] Ivezić, Z. et al., 2008, arXiv:0805.2366
- [19] Koekemoer, A. M., Aussel, H., Calzetti, D., et al. 2007, *ApJS*, 172, 196
- [20] La Barbera, F., de Carvalho, R. R., Kohl-Moreira, J. L., et al. 2008, *PASP*, 120, 681
- [21] Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
- [22] Mutchler, M., Published online at <http://www.stsci.edu/institute/conference/cal10/proceedings>, p.69
- [23] Pirzkal, N., Burgasser, A. J., Malhotra, S., et al. 2009, *ApJ*, 695, 1591

- [24] Pirzkal, N., Xu, C., Malhotra, S., et al. 2004, ApJS, 154, 501
- [25] Riess, A. G., Strolger, L.-G., Casertano, S., et al. 2007, ApJ, 659, 98
- [26] Riess, A. G., Li, W., Stetson, P. B., et al. 2005, ApJ, 627, 579
- [27] Riess, A. G., Strolger, L.-G., Tonry, J., et al. 2004, ApJ, 607, 665
- [28] Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751
- [29] Stetson, P. B. 1987, PASP, 99, 191
- [30] Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27
- [31] Zenteno, A., Song, J., Desai, S., et al. 2011, ApJ, 734, 3