IMPORTANT NOTE: This is a WORKING DRAFT of an article planned for submission to PASP or a similar journal in the near future. We are making it available at the same time as the Version 1 release (Feb 25, 2015) to help users understand the HSC. Please consider the contents TENTATIVE; various figures and discussion are still being updated or are listed as TBD. We are interested in feedback to help improve the draft before submission. Please send any comments to whitmore@stsci.edu.

Draft: February 24, 2015
Preprint typeset using LaTeX style emulateapj v. 04/17/13

VERSION 1 OF THE HUBBLE SOURCE CATALOG

Bradley C. Whitmore[1], Sahar S. Allam[1,2], Tamás Budavári[3], Stefano Casertano[1], Ronald A. Downes[1], Thomas Donaldson[1], Stephen H. Lubow[1], Lee Quick[1], Louis-Gregory Strolger[1], Geoff Wallace[1], Rick L. White[1].

Draft: February 24, 2015

ABSTRACT

The Hubble Source Catalog is designed to help optimize science from the *Hubble Space Telescope* by combining the tens of thousands of visit-based source lists in the Hubble Legacy Archive into a single master catalog. Version 1 of the Hubble Source Catalog includes WFPC2, ACS/WFC, WFC3/UVIS, and WFC3/IR photometric data generated using SExtractor software to produce the individual source lists. The current version of the catalog includes roughly 80 million detections of 30 million objects involving 112 different detector/filter combinations and about 50 thousand HST exposures. Source lists from Data Release 8 of the Hubble Legacy Archive (i.e., public data as of June 1, 2014) were matched using an algorithm developed by Budavari & Lubow (2012). The mean photometric accuracy is better than 0.10 mag and the astrometric residuals are typically within 10 mas, with a value for the mode (i.e., peak) of 2.3 mas. We describe the database design with emphasis on those aspects that enable the users to fully exploit the catalog while avoiding common misunderstanding and potential pitfalls. We provide common usage examples of the Hubble Source Catalog to illustrate some of the science capabilities and data quality characteristics.

Keywords: Catalogs; astronomical data bases:- HSC —

1. INTRODUCTION

The *Hubble Space Telescope (HST)* has been in orbit for over 25 years. In that time it has observed with a dozen different instruments, hundreds of observing modes, and roughly a thousand different filters and gratings. Selected, effectively pencil-beam observations, have been taken of only a small fraction of the total sky, with a range of exposure times from less than a second (e.g., a search for faint companion planets around very bright stars), to week-long observations of a "blank" part of the sky to observe galaxies at the edge of the universe. This diversity reflects both the promise and the challenge of the Hubble Source Catalog (HSC).

In recent times, computer-based catalogs of astronomical objects have proven to be of great benefit to astronomers (e.g., the Sloan Digital Sky Survey = SDSS). By querying such databases, astronomers are able to carry out research that would otherwise be very time-consuming or completely impractical. The HSC is designed to include the vast majority of all the objects ever observed by *Hubble* in a single master catalog. Repeat observations are common, with over 500,000 objects having more than 50 separate observations, providing a rich database for variability studies. Regions of the sky with thousands, or even tens of thousands of separate observations (e.g., the Magellanic Clouds - see Figure 1, the Virgo cluster, the Orion Nebula, ...) can be evaluated in minutes.

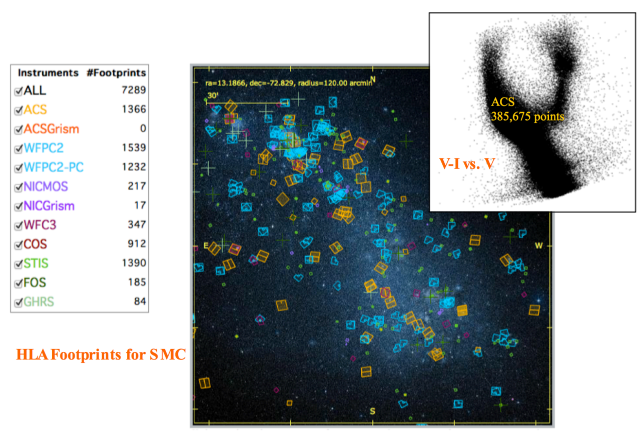

Figure 1. HLA footprints for a search of the SMC using a radius of 2 degrees. A color-magnitude diagram containing 385,675 data points, created by the HSC in less than 2 minutes, is shown in the upper right.

The basic scheduling unit for an *HST* observation is a "visit", typically lasting between a single orbit (96 min) and six or seven orbits. A visit is also a natural unit for the production of data products from the telescope. For this reason, the Hubble Legacy Archive (HLA, see

[1] Space Telescope Science Institute (STScI), 3700 San Martin Drive, Baltimore, MD 21218.
[2] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510.
[3] Center for Astrophysical Sciences, Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218.

Jenkner et al. 2006, Whitmore et al. 2008) combine data together in visit-based images, and produces source lists for each of these combined images.

In general, an astronomer is not interested in visits, but would like to retrieve all the relevant information for target observed by *Hubble*. That is the primary driver behind the production of the HSC; to combine the tens of thousands of visit-based HLA source lists into a single master catalog.

The HSC has been available as a Beta (test) version since 2012. Special purpose techniques were developed to handle the challenges of building the HSC. The pipeline, the astrometric and cross-matching algorithms, and the properties of the Beta version catalog are described in Budavari & Lubow (2012). In the current paper, we describe Version 1 of the HSC. We provide a brief update on the catalog generation methods and the catalog properties since the Beta release.

Astronomical catalog are generally produced by telescopes that conduct systematic surveys; the catalog is a key objective of the survey. The observations cover a regular geometric pattern in the sky with uniform properties, such as exposure time and filter set. The HSC is a very different catalog, as illustrated by Figure 1. Due to the diversity of Hubble observations, and accentuated by the fact that the HSC is still in an active developmental stage, the catalog can be very non-uniform and patchwork in nature in certain regions. This irregularity requires care when developing search criteria. Nevertheless, the HSC is a powerful new tool for research with *Hubble* data, even with its limitations, and will be an important reference for future telescopes, such as the James Webb Space Telescope, and survey programs such as Pan-STARRS and LSST.

The paper is organized as follows. In Section 2 we describe the data used in building Version 1 of the HSC, while in Section 3, we describe the catalog and the pipeline used to construct the catalog. In section 4, we examine the photometric and astrometric quality of the HSC. In Section 5, we describe some of the tools that can be used to query the HSC, including a database (CasJobs) interface similar to the Sloan Digital Sky Survey). Sections 6 includes advice on avoiding common misunderstanding and potential pitfalls while Section 7 describes future plans. A brief summary is included in Section 8. Appendix A provides pointers to other relevant information.

2. DATA

2.1. *Instruments and Filters*

Version 1 of the HSC includes HLA source lists from the three cameras responsible for the majority of images taken by *Hubble*, namely the Wide Field Planetary Camera 2 (WFPC2), the Wide Field Camera of the Advanced Camera for Surveys (ACS/WFC), and Wide Field Camera 3 (WFC3). Other instruments will be added in the future, including the ACS High Resolution Camera (ACS/HRC) and the Near Infrared Camera and Multi-Object Spectrometer (NICMOS). It is possible that data from other cameras (e.g., the imaging modes of the Space Telescope Imaging Spectrograph (STIS) and the Faint Object Camera (FOC) will be added at a later date.

Version 1 of the HSC is constructed using HLA Data Release 8 (DR8) images and source lists. These include public data as of June 1, 2014. Figure 1 shows the patchwork nature of the Hubble observations, with only a small fraction (TBD %) of the full sky being covered. This is a primary difference from most surveys and catalogs, and must always be kept in mind when using the HSC.

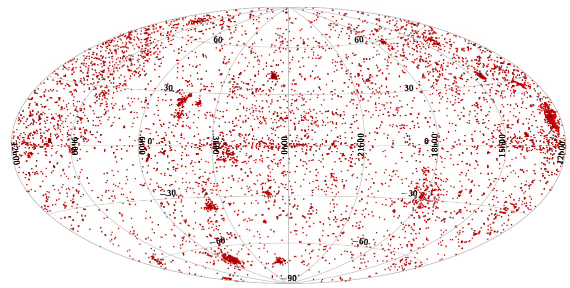

Figure 2. All HLA images used in building the HSC.

Table 1 provides some basic parameters and statistics for the different instruments used in the HSC. The majority of the images are from the WFPC2, due to its longevity (16 year) on *HST*. An important difference between the WFPC2 and the later generation ACS and WFC3 cameras is the larger pixel size (0.10 arcsec), again highlighting the diversity of *HST* data. Magnitudes based on the different instruments are reported separately. In most cases users will analyze the data for the different instruments separately, but it is also possible to combine data together in certain cases, especially if transformation equations are employed.

Table 1
Basic Instrument Statistics for the HSC

Inst.	# Filters	# Images	# Det	Area (sq deg)	pix-size	aper-size (aper1 aper2)
1	2	3	4	5	6	7
WFPC2	38	29,025	TBD	42.638	0.10″	0.10″, 0.30″
ACS	11	9,011	TBD	28.371	0.05″	0.05″, 0.15″
WFC3-UVIS	39	TBD	TBD	TBD	0.04″	0.05″, 0.15″
WFC3-IR	15	TBD	TBD	TBD	0.09″	0.15″, 0.45″

Table 2
Number of Repeat Observations in the HSC

NumImages	# of objects
≥10	8118220
≥50	475679
≥100	66196
≥150	27047
≥200	10922
≥250	2915
≥300	64
max.(329)	1

3. THE CATALOG

The SExtractor software (Bertin & Arnouts 1996) is used to produce the HLA source lists. Both aperture magnitudes (MagAper1 and MagAper2) and total magnitudes (using the MagAuto algorithm in SExtractor) are included in the HSC. DAOPHOT (Stetson 1987) source lists are also produced in the HLA, primarily for point sources. These are not used in the HSC.

The radius used for the small and large aperture measurements (MagAper1 and MagAper2) are 1 and 3 pixels for WFPC2 and ACS; 1.25 and 3.75 pixels for WFC3/UVIS; and 1.15 and 3.46 pixels for WFC3/IR. See Table 1 for the corresponding sizes in arcsec. In most cases, the detection threshold is set to three times the background noise, although it is adjusted in some regions in accordance with the source flagging (e.g., around very bright stars).

Unlike ACS and WFC3, WFPC2 source lists explicitly include a correction for Charge Transfer Efficiency (CTE) loss, based on the formulae from Dolphin (2009). Images with pixel-to-pixel corrections using the algorithm developed by Anderson & Bedin (2010) will be use to construct ACS and WFC3 source lists in the future.

3.1. *How the Catalog was Constructed*

The basic steps involved in the construction of the HLA source lists, and the subsequent construction of the HSC, are described below. More detailed descriptions of various aspects of the process are available in the HLA and HSC FAQs (see Appendix A).

The reduction pipeline, the astrometric and cross-matching algorithms, and the properties of the Beta version catalog, are described in Budavari & Lubow (2012) (see also Budavari & Szalay 2008 for a discussion of the Bayesian approach at the heart of the technique, Lubow & Budavari 2013 for more details about the pipeline, and Whitmore et al. 2008 for details about the early source list generation).

1. Combine exposures for each filter within a visit using multidrizzle (Fruchter 2009) for WFPC2 and ACS, and using astrodrizzle (http://drizzlepac.stsci.edu/) for WFC3.

2. Combine the images into a "white-light" image (i.e., combine data from different filters, but within the same visit, to provide a deeper image with a wider wavelength range). These serve as the detection image for the visit. No shifts are made for WFPC2 and ACS within a visit before combining the data. For WFC3, an early version of the "tweakreg" algorithm within the "astrodrizzle" software package was use to align the sub images within a visit, prior to combining the images for the different filters.

3. Extract SExtractor (and in the HLA, DAOPHOT also) source lists from the white-light (detection) images. Make a second-pass of the positions on the filter-based image to obtain final measurements. While both source lists are available in the HLA, only the SExtractor source lists are used in the HSC.

4. Two different magnitudes are included in the HSC; MagAper2 (aperture magnitudes - see Table 2 for sizes) and MagAuto (SExtractor estimates of the total magnitude – primarily designed for extended sources). Only the MagAper2 values are included in the HSC. However, the smaller MagAper1 magnitudes can be recovered via the Concentration Index (CI), which is the difference between MagAper1 and MagAper2.

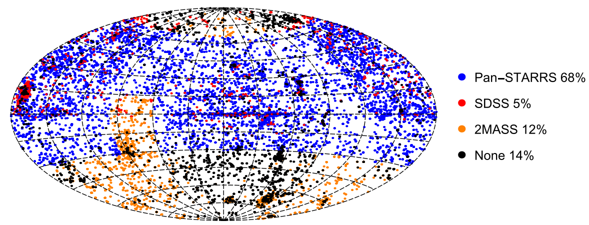

Figure 3. Sky coverage showing where Pan-STARRS, SDSS, and 2MASS provides the the astrometric "backbone" for the HSC.

5. Apply astrometric "preoffsets" based on matching with Pan-STARRS, SDSS. or 2MASS. This step is needed to reduce the typical 1 - 2 arcsec absolute astrometric accuracy for HST images to less than 0.3 arcsec. Without this step, the number of false matches in very crowded fields (e.g., globular clusters) keeps the Budavari and Lubow (2012) algorithm from converging in a reasonable time for many cases.

6. Match images and determine relative shifts needed to align the various images using the Budavari and Lubow (2012) algorithms. This reduces the relative astrometric accuracy from a few tenths of an arcsec to less than 10 mas in most cases (see Section 4.3).

7. Separate the images into linked groups (see the Budavari and Lubow 2012 paper for details), and determine mean astrometric positions for each group.

8. Readjust the absolute astrometry for each group using Pan-STARRS, SDSS, or 2MASS as the reference. The absolute astrometry for Pan-STARRS is approximately 0.1 arcsec; for SDSS and 2MASS it is about 0.2 arcsec. Hence the typical absolute astrometric accuracy for the HSC is in the range 0.1 to 0.2 arcsec (see Section 4.3 for discussion). Figure 3 shows that in 14 % of the cases, there is no overlap with Pan-STARRS, SDSS, or 2MASS. There are other reasons why absolute astrometric corrections cannot be made. Hence, only 68 % of the HSC sources have the Absolute Correction (AbsCor) flag set to yes (Y).

9. Compute photometric means and sigmas, and build the HSC summary database. Mean values for the

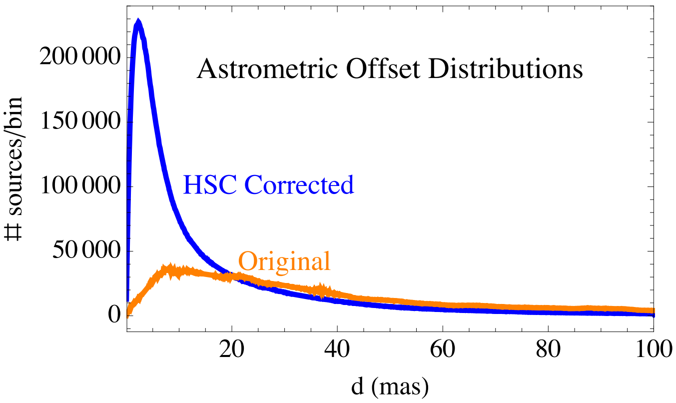

Figure 4. Improvement in astrometric accuracy before and after the Budavari & Lubow (2012) algorithms are employed. The mode (peak) for the corrected curve is 2.3 mas.

Concentration Index (CI) and the Kron Radius are also computed and included in the summary database.

10. Port the database to the Discovery Portal, HSC CasJobs, and HSC Home Page (see Section 5).

4. QUALITY ASSESMENT

A three-pronged approach is used to characterize the quality of the HSC. We first examine a few specific datasets, comparing magnitudes and positions directly for repeat measurements. The comparisons are made using first the same instrument and filter; and then using different instruments and filters.

The second approach is to compare repeat measurements for the full database. While this provides a better representation of the entire dataset, it can also be misleading since the tails of the distributions are generally caused by a small number of bad images and bad source lists.

The third approach is to produce a few well-known astronomical figures (e.g., color-magnitude diagram for the outer disk of M31 from Brown et. al. 2009) based on HSC data, and compare them with the original study.

This three-pronged approach is hierarchal in nature; 1) a spot check on the consistency and quality of the source lists for a few specific data sets, 2) a check that the entire dataset is relatively homogenous and of high quality, 3) a check that we are consistent with completely independent datasets or independent analysis techniques.

As stressed in other parts of this paper (e.g., Section 6 - caveats and warnings), it is important to keep in mind that parts of the HSC are non-uniform. Hence, researchers cannot assume that the results reported in this section represent the entire database. If uniformity is important for a specific science project, a careful examination of the data is required, including viewing the images themselves. In many cases it is possible to filter the HSC data and improve the uniformity of the data. This topic will be discussed in Section 4.1.3 (filtering out artifacts).

4.1. *Photometry*

4.1.1. *Point Source Photometry - Single Instrument/Filter Checks*

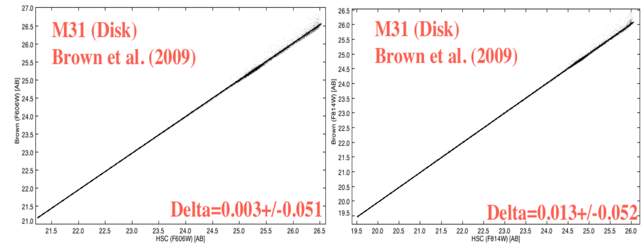

Figure 5. Comparison of HSC (ACS/WFC_F606W and ACS/WFC_F814W) magnitudes with Brown et al. (2009: PropID = 10265). See HSC Use Case # 1 for details.

Since we are primarily interested in stellar photometry in this section, aperture magnitudes (i.e., MagAper2) are used throughout.

Our first photometry check examines the Brown et al. (2009) deep ACS/WFC observations of the outer disk of M31. The observing plan for this proposal resulted in 60 separate one-orbit visits (not typical of most HST observations), hence provide an excellent opportunity for determining the true uncertainties by examining repeat measurements. However, these also highlight one of the limitations of the HSC, namely the brighter completeness limit imposed by the use of only visit-based source lists from the HLA. For this reason the deep, co-added, 60 orbit image used by Brown et al. goes roughly three magnitudes deeper (although with large photometric uncertainties at these very faint magnitudes) than the HSC. In the range of overlap, the agreement is quite good, with zeropoint differences less than 0.02 magnitudes (after corrections from ABMAG to STMAG and from aperture to total magnitudes) and mean values of the scatter around 0.05 mag (see Figure 4). More details are available in HSC Use Case # 1 (see Appendix A).

4.1.2. *Point Source Photometry - Error Estimates*

Figure 6 shows a comparison of estimated photometric errors based on single SExtractor measurements (i.e., magerr) with the true scatter in the repeat measurements (i.e., values of filter_Sigma reported in the HSC summary form). We find that the estimated values of magerr are roughly a factor of three too low. This is a known problem with the WFPC2 and ACS HLA source lists. However, we also note that the Sigma estimates for WFPC2 at bright magnitudes increase rather dramatically. This is largely due to the inclusion of a few saturated stars, but other factors may also be involved; it is still under investigation. Filtering out high values of the Concentration Index (CI) reduces, but does not eliminate the increase (new figure TBD) These problems will be rectified in the near future when the pipeline developed for the newer WFC3 source lists is used to produce the next generation of WFPC2 and ACS source lists. In the meantime we recommend the use of the filter_Sigma values based on repeat measurements when available. This is primarily an issue for the HLA rather than the HSC, since magerr values are not included in the default version of the HSC Summary table.

4.1.3. *Point Source Photometry - Cross-Instrument/Filter Checks*

The globular cluster M4 (with a search at 16:23:38.66 -26:32:10.9 r=200s) provides a good opportunity to com-

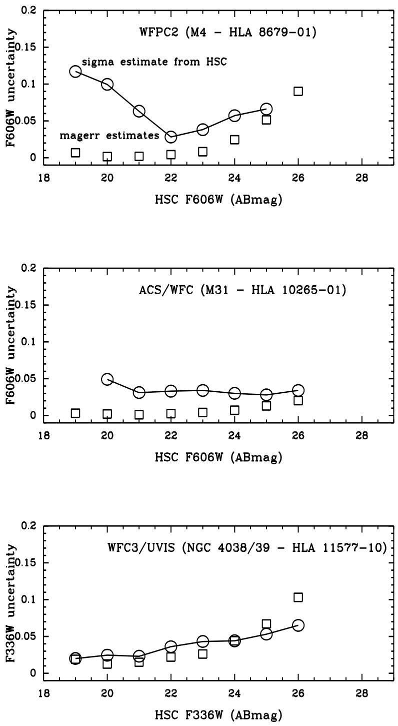

Figure 6. Comparison of Sigma values based on repeat measurements in the HSC with magerr estimates based on SExtractor estimates.

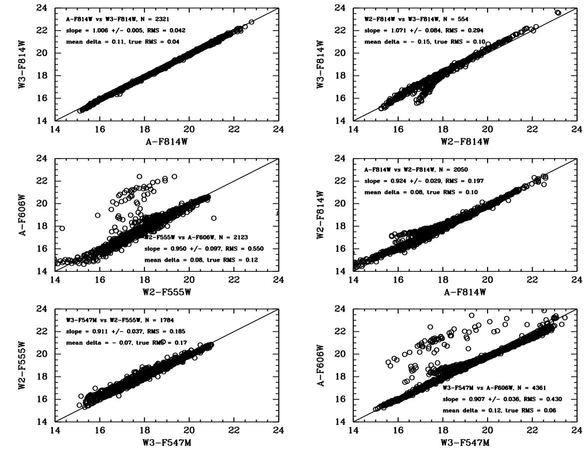

Figure 7. Comparisons of repeat measurements for similar filters in the globular cluster M4.

pare the HSC photometric system for all three instruments. Figure 7 shows comparisons in the "V" filters (i.e., WFPC2-F555W, ACS-F606W, and WFC3-F547M) and "I" filters (i.e., WFPC2-F814W, ACS-F814W, and W3-F814W).

Starting with the best case, A-F814W vs W3-F814W shows excellent results, with a slope near unity, values of RMS around 0.04 magnitudes, and essentially no outliers. The good agreement also suggests that A-F814W and W3-F814W measurements can be added together with little loss of photometric accuracy. This is not true, as we will see below, when the filter bandpasses are not as similar. Photometric transformations are necessary before combing these types of observations.

An examination of the W2-F814W vs. W3-F814W and A-F814W vs W2-F814W comparisons show that there is an issue with the WFPC2 data. The short curved lines deviating from the 1-to-1 relationship show evidence of the inclusion of a small number of slightly saturated star measurements (i.e., roughly 5 % of the data).

Much larger deviations are seen in the two panels making use of A-F606W observations, where a cloud of outliers are found several magnitudes off the 1-to-1 line. These are caused by combining data from short (20 sec) and long (1800 sec) sub exposures, defeating the flagging system which should eliminate saturated data. While these issues will be fixed in future versions of the HSC, it is also relatively easy to filter them out, as discussed below.

A careful look at Figure 7 also shows systematic deviations in the slope of the relationships, with deviations of a few tenths of a magnitude at the extremes (e.g. W3-F547M vs A-F606W). These are generally examples where the filters are not well matched (e.g., the central wavelength and width are 2180 Angstroms for the A-F606W filter but only 650 Angstroms for the W3-F547M filter). Hence sources with different colors (and hence different brightnesses since this is a globular cluster with a well defined main sequence) deviate in the two filters. A photometric transformation would need to be made before these two filters would line up. The comparison is made here in order to evaluate the RMS scatter, not to imply that the data from different instruments/filters can be added together without the loss of a few tenths of a magnitude in accuracy.

Other complications that can cause deviations are the inclusion of CTE corrections (i.e. for WFPC2 using the Dolphin 2008 formula, but no corrections for ACS and WFC3), differences in aperture corrections (typical differences between the different instruments are about 0.1 mag for the ACS, WFC3/UVIS, and WFPC2), and differences in exposure times (e.g., resulting in different completeness limits - see the transition at about the 19th magnitude in the W2-F555W vs A-F606W diagram).

4.1.4. *Filtering out Artifacts*

As stressed throughout this paper, the diverse nature of the *HST* archival database can result in a number of artifacts, some examples of which are discussed below. However, we also note that the availability of multiple observations can also provide the opportunity to catch more of these artifacts and filter them out; a circumstance that is not always possible with more limited datasets. This is one of the primary benefits of the HSC.

Figure 8 shows the same comparisons as Figure 7, but with four constraints included. These are:

- NumImages > 2 (to remove residual cosmic rays)
- CI < 1.4 (to remove extended sources and blends),
- CI_Sigma < 0.5 (to remove partially saturated stars)

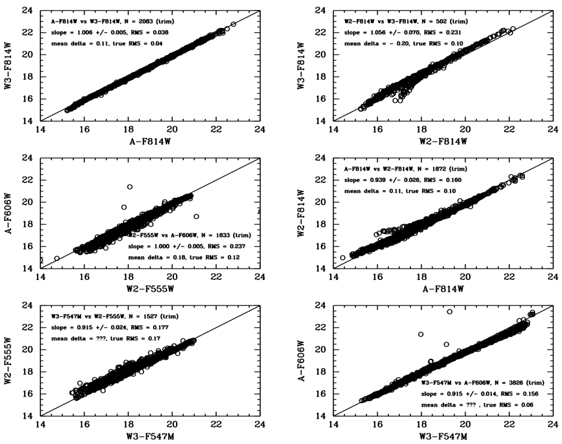

Figure 8. Same as Figure 7, but with the four constraints discussed in section 4.1.4 imposed.

- filter_Sigma < 0.2 (to remove low S/N data and saturated objects)

where CI is the Concentration Index (i.e., the difference between the small and large aperture magnitudes - see Table 1), and "Sigma" refers to the scatter in repeat measurements.

As shown in Fig. 8, the number of artifacts and discrepant points is greatly reduced, with only 3/3826 (0.1 %) in the W3-F547M vs A-F606W comparison with residuals greater than 1 mag. While the specific criteria and values may change for different datasets and scientific purposes, these four parameters can often be used to reduce the number of artifacts.

Another form of "artifact" is the non-uniformity inherent in a dataset as diverse as the Hubble archives. This is accentuated by the poorer quality of the WFPC2 and ACS HLA source lists, relative to the more recently generated WFC3 source lists. For example, Figure 9 shows that many sources are missed in regions with high background in this WFPC2 image. While the overall coverage of the HSC (i.e., the pink circles) is quite good, thanks mainly to the WFC3 images in this region, users should keep in mind that just because a given observation is missing in the HSC does not mean that it was not observed by Hubble.

More details about the comparisons discussed above, as well as other examples relevant to photometric accuracy, can be found in:
HSC Use Case # 1 - (Stellar Photometry in M31), HSC Use Case # 2 - (Globular Clusters in M87 and a Color Magnitude Diagram for the LMC),
and HSC Use Case # 5 - (White Dwarfs in the Globular Cluster M4).

See Appendix A for the location of these URLs.

4.1.5. *Extended Object Photometry*

In this section we make photometric comparisons with extended targets, such as distant galaxies. Hence, values obtained using the Source Extractor algorithm MagAuto are used throughout.

We first make a comparison with a GOODS (reference) field, using identical datasets, in order to test the effects of the different analysis techniques rather than differences in the data itself. We then compare with

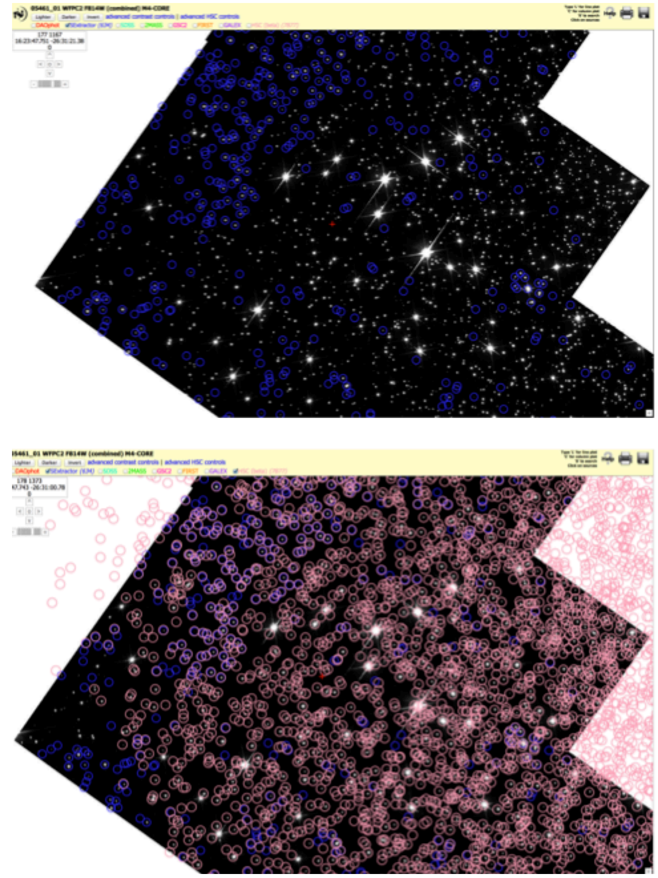

Figure 9. Top image show a WFPC2 Source Extractor source list (blue) in the globular cluster M4 (TBD - make better version of figure since can not see the blue very well). Note how nonuniform the coverage is; i.e., missing sources where the background is high. The bottom image shows the much more uniform HSC coverage. It is more uniform due to the presence of WFC3 source lists in this field.

Figure – TBD

Figure 10. TBD - Comparison of HSC and SDSS in the Hubble Deep Field.

SDSS photometry using the Hubble Deep Field (HDF). In this second example the datasets are very different (e.g., ground-based vs space-based; different photometric systems; different depth).

4.1.6. *Comparisons with CANDELS Analysis of GOODS Fields*

TBD

4.1.7. *Comparisons with SDSS Observations in the Hubble Deep Field*

The Sloan Digital Sky Survey (SDSS) has been tremendously successful, due both to the high quality, wide-field, uniform database, and to the extensive extraction and analysis tools it has made available to researchers. It has taken the field of "database astronomy" to a new level, and in many ways is the inspiration for the HSC.

A comparison between the HSC and SDSS is therefore of interest to many astronomers, and also provides an opportunity for highlighting both the similarities (e.g., agreement between photometric results; availability of CasJobs) and differences (e.g., the HSC goes deeper but with "pencil beam" coverage; the HSC can be very non-uniform in certain regions).

Figure 10 shows the overlap between the HSC and SDSS coverage of the Hubble Deep Field (HDF). There are TBD objects in common out of the TBD HSC sources in this field. The SDSS has a completeness limit around TBD while the HSC goes to TBD. Note that this comparison is with the WFPC2; similar observations with the WFC3 (which is both more sensitive and has higher quality catalogs in the HSC) would go another TBD magnitudes deeper.

Figure 11 shows the photometric comparisons between the HSC and SDSS for a wide variety of filters. We find reasonably good agreement. Both the scatter and the offsets are typically a few tenths of a magnitude. The offsets reflect the differences in photometric systems, since no transformations have been made for these comparisons. The best agreement is between A-F814W and SDSS-i. This reflects the fact that these two photometric systems are very similar, hence the transformation is nearly 1-to-1 .

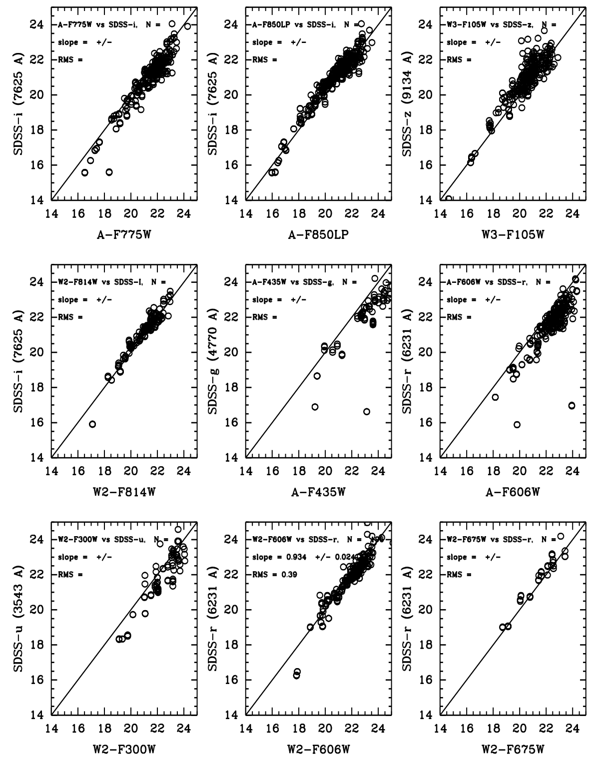

Figure 11. Comparison between HSC photometry (MagAuto) and SDSS photometry for the Hubble Deep Field (i.e., RA = 189.206, DEC = 62.2161, r=500 arcsec).

4.2. *Astrometry - Case Studies*

TBD

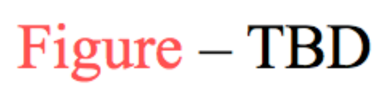

Figure 12. TBD - Astrometry case study

4.3. *Database Comparisons*

Another approach to characterizing the quality of the HSC is to make comparisons using repeat measurements from the entire database, rather than the detailed comparisons shown in Sections 4.1 and 4.2. While lacking the precision of case studies, these have the advantage of including a much larger fraction of the entire database. Hence the two approaches are complementary.

4.3.1. *Photometric Database Comparisons*

Figure 14 shows the Version 1 HSC photometric accuracy for the the entire database. The data is separated into different instruments and comparisons are made between estimates of flux in the large aperture (i.e., MagAper2) for the same filter. The x-axis is the flux difference ratio defined as abs(flux1-flux2)/max(flux1,flux2). The y-axis is the number of sources per bin (whose size is a flux difference ratio of 0.0025) that is normalized to unity at a flux difference of zero.

4.3.2. *Astrometric Database Comparisons*

Figure 15 shows a similar comparison for the entire HSC database for the relative astrometry based on repeat measurements, using the white-light detection images. The mode (peak) of the distributions for ACS and WFC3 are roughly 2 mas. The peak of the distributions for the WFPC2 and WFC3/IR occur at higher values primarily due to the larger pixels for these instruments.

4.4. *Comparisons with other Studies based on Use Cases*

4.4.1. *Comparisons: Brown et al. 2009 - Color Magnitude Diagram in the Outer Disk of M31*

4.4.2. *Comparisons: Bernard et al. 2010 - Variability in IC 1613*

4.4.3. *Comparisons: Gladders et al. 1998 - The Red Sequence in Abell 2390*

4.5. *Incompleteness*

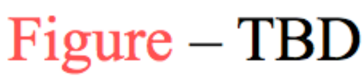

Figure 13. TBD - Astrometry case study

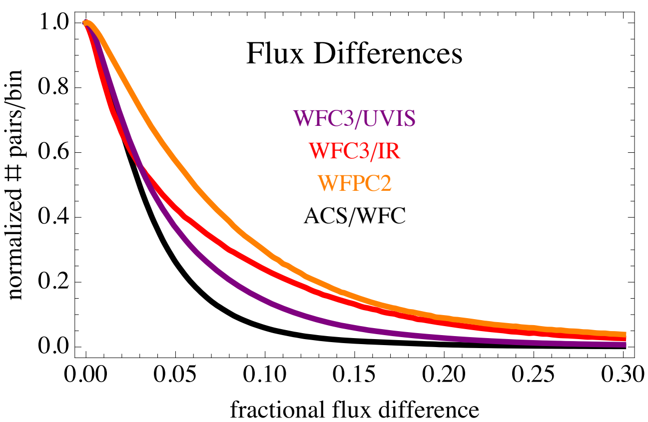

Figure 14. Photometric accuracy for Version 1 of the HSC based on repeat measurements using the entire database.

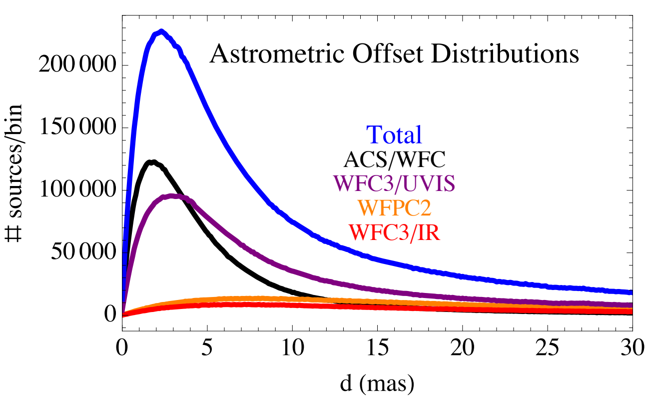

Figure 15. Astrometric accuracy for Version 1 of the HSC based on repeat measurements using the entire database.

The HSC is incomplete for a number of reasons. For example, only three of the 12 instruments flown on Hubble are included; WFPC2, ACS/WFC, and WFC3. However, these are the three instruments with the largest numbers of Hubble detections, hence it is a good starting point. Future plans call for the inclusion of NICMOS and ACS/HRC observations, and possibly others in the future (e.g., STIS imaging and FOC).

It is also important to remember that even for the three instruments included in Version 1, only about 65% of the ACS/WFC, WFPC2, WFC3 images are included in the catalog due to image quality and other issues. In addition, as will be stressed in Section 6, the quality and depth of the source lists for the three instruments is non-uniform. While this will be improved in the future, the instruments will always have different completeness thresholds due to their different quantum efficiencies (i.e., WFPC2 is much shallower than ACS and WFC3).

One final point is that in many cases the coaddition of visits to form a much deeper mosaic image can result in deeper completeness limit, as discussed in Section 4.1.1. For all these reasons, researchers should keep in mind that just because a source is not in the HSC does not mean that there is no *Hubble* observations of it. The HLA can be used to make a more complete search, but for a definitive determination (e.g., when checking for duplications when writing *HST* observing proposals), the MAST archive tools must be used (i.e., there are some *HST* observations that are not included in the HLA - see http://hla.stsci.edu/hla_faq.html#General2).

As outlined in Section 3.1, the HSC attempts to improve the astrometric accuracy of the data by cross-matching overlapping images, obtaining relative accuracies of better than 10 mas in most cases, and then using Pan-STARRS, SDSS, or 2MASS as the astrometric backbone, producing roughly 0.1 - 0.2 arcsec absolute accuracy. Unfortunately, not all HLA observations have sufficient overlap to obtain accurate cross matches. Hence, only about 68 % of the images have astrometric corrections applied to them. The Absolute Correction (AbsCor) parameter can be used to determine whether a correction has been made or not for a specific object.

5. TOOLS FOR ACCESSING THE HSC

There are three ways to access the HSC. This is partly for historical reasons, and partly to provide different types of services.

5.1. *MAST Discovery Portal (Browsing, Filtering, Plotting, and Cross Matching)*

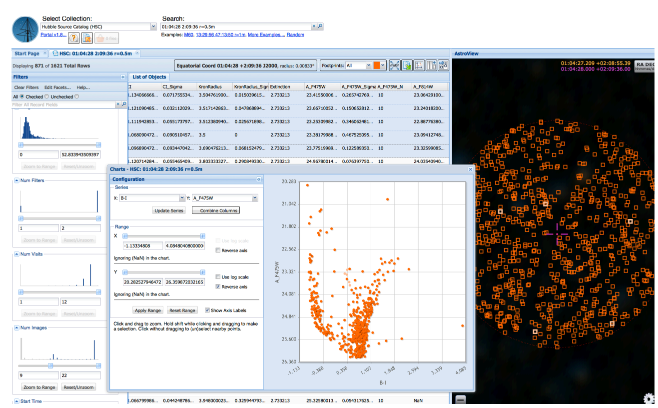

Figure 16. Screen from HSC Use Case # 3 showing various aspects of the MAST Discovery Portal.

The primary access tool for the HSC is the MAST Discovery Portal, which generally provides the best way to browse what is in the HSC, to do some quick plotting and/or cross matching with other data, and to download the needed data for further analysis. Its primary current limitation is that only 10,000 sources can be included in a given search, and only MagAper2 (aperture magnitudes) rather than MagAuto (extended photometry) are included as the default (note, however, that CasJobs can be used to obtain larger samples, and/or retrieve values of MagAuto, which can then be filtered and then uploaded into the Discovery Portal). Originally developed as part of the Virtual Observatory initiative, the Portal has been modified to include access to HSC data, and to include features needed to view HST images. It includes a wide range of tools for viewing, filtering (e.g., NumImage > # to remove residual cosmic rays and other artifacts), plotting, cross-matching, and downloading.

Figure 16 shows an example of how the Discovery Portal can be used to find variable stars in IC 1613 (from Use Case # 3).

See Appendix A for the URL for the Discovery Portal and other HSC related sites.

5.2. *CASJOBS (Advanced Search and Analysis)*

The Catalog Archive Server Jobs System (CasJobs) was developed by the Johns Hopkins University/Sloan Digital Sky Survey (JHU/SDSS) team. With their permission, MAST has used version 3.5.16 of CasJobs to construct three CasJobs-based tools for GALEX, Kepler, and the HSC. The purpose of CasJobs is to permit large queries, phrased in the Structured Query Language (SQL), to be run in either real time or in batch queues. Therefore, it does not have the limitations of only including a small subsample of the HSC, as is the case for the MAST Discovery Portal. However, it also does not have the wide variety of graphic tools available in the Discovery Portal.

Figure 1 shows an example of how CasJobs can be used to make a color magnitude diagram including 385,675 ACS sources in the Small Magellanic Cloud in less than two minutes (from Use Case # 2). Figure 17 shows an example of the query screen for CasJobs, in this case retrieving a sample of globular clusters in M87 (also from Use Case #2).

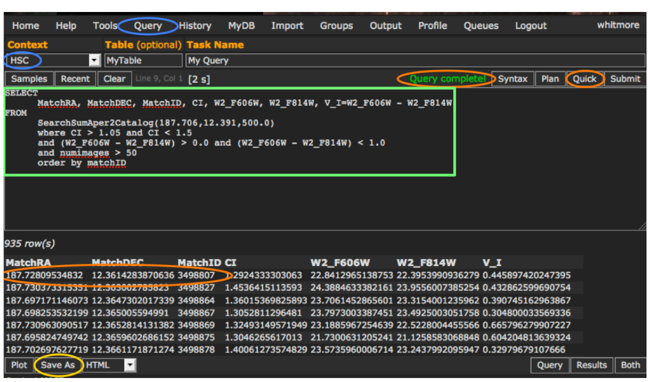

Figure 17. Example of a HSC CasJobs screen from Use Case # 2.

5.3. *HSC Home Page (Summary and Detailed Search Forms)*

The HSC Home Page represents a more basic level of sophistication. This was the original access tool (e.g., for the Beta releases), and while it may still be useful for certain very detailed searches, it has been largely superseded by the Discovery Portal and HSC CasJobs.

There are two forms-based interfaces to the HSC that follow the conventions of MAST. These are the summary search form, which allows users to obtain mean magnitudes and other information with one row per match; and the detailed search form, which includes information about each detection that went into the match. The HSC FAQ is also located at this site, providing the next level of detail beyond this paper.

Figure 18 shows an example of how the HSC Page can be used to download data from Brown et al. (2009) observations of the outer disk of M31 (from HSC Use Case # 1).

Figure 18. Example of a search using the HSC Summary Search Form.

6. CAVEATS AND WARNINGS

As stressed in many sections of this paper, the HSC is not a typical wide-area, uniform catalog such as 2MASS, SDSS, or Pan-STARRS. It is based on a diverse set of observations using pencil-beam observations of only a small fraction of the sky. While it has tremendous potential for doing science, it can also easily be misused. Users should not simply use the HSC as a database search tool. They need to:

- View the HSC overlaid on images. While the vast majority of the source lists are quite good, there are still some problem cases that can cause obvious artifacts in the HSC (e.g., see Figure 19)
- Try different selection filters (e.g., NumImages > #, WFPC2 compared to ACS compared to WFC3, etc) to see how it affects the science results.

6.0.1. *Five Things You Should Know About the HSC*

New users should keep the following in mind when using the HSC.

1. Detailed use cases and video are available for training. See Appendix A for pointers.

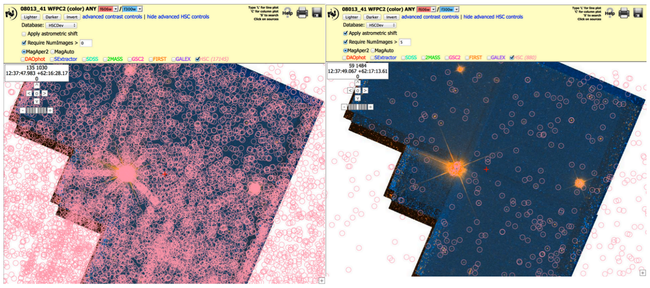

Figure 19. Example of a particularly bad WFPC2 source list (left image) showing artifacts from bright stars and edge effects. Using NumImage > 5 (right image) removes most of these artifacts.

2. Coverage can be very non-uniform (unlike surveys such as SDSS), since source lists have been combined for pointed observations from a wide range of HST instruments, filters, and exposure times. With careful selection of various parameters (e.g., NumImages included in a match), this non-uniformity can be minimized

3. WFPC2 and ACS source lists are of poorer quality than WFC3 source lists. As we have gained experience, the HLA source lists have improved. For example, many of the earlier limitations (e.g., depth, difficulty finding sources in regions of high background, edge effects, ...) have been improved in the WFC3 source lists. These improved algorithms will be included for the WFPC2 and ACS/WFC, and will be incorporated into a future release of the HSC.

4. The default is to show all HSC objects in the catalog. This may include a large number of artifacts. You can request Numimages > 1 (or more) to filter out many artifacts in the HSC. Other examples of useful selection filters are discussed in Section 4.1.3 .

5. The default is to use aperture magnitudes (i.e., Maper2) in the ABMAG system. Transformations are necessary to convert to other systems (e.g., VEGAMAG), or from one instrument to another, or to other photometric systems (e.g., Johnson-Cousins or SDSS ugriz). Aperture corrections are needed to convert aperture magnitudes to total magnitudes for stars. For extended sources MagAuto can be requested.

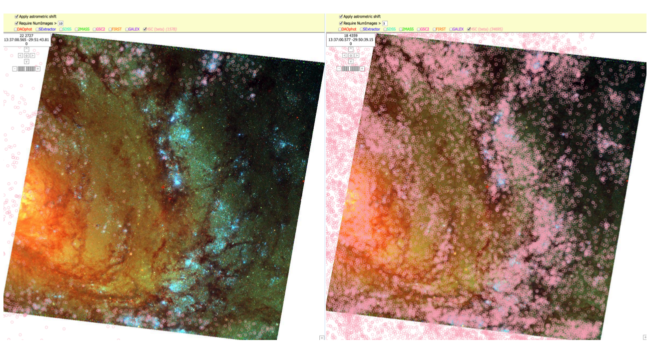

Figure 20. An example of the non-uniformities that are possible using improper search criteria, in this case NumImages > 10 (left image) rather than > 3 (right image). The small pink circles are objects in the HSC. Additional source lists from overlapping HLA images in the upper and lower parts of the galaxy (M83), images not shown here, result in various corners and linear features in the left image.

7. FUTURE PLANS

A number of improvements and enhancements for the HSC are planned for the future. In the relatively short term, the primary improvement will be to upgrade the WFPC2 and ACS source lists using the algorithms developed for the WFC3. On a longer time scale, HLA source lists will be developed for the observations taken with the ACS High Resolution Camera (ACS/HRC) and Near Infrared Camera and Multi-Object Spectrometer (NICMOS).

A more fundamental improvement planned for the future is to use the precise offsets determined for the HSC to combine the visit-based images into deeper mosaics. HLA source lists will then be obtained using these images to develop much deeper catalogs (see HSC Use Case # 1 for an example where the gain will be roughly three magnitudes), which will then be integrated into the HSC. Another important addition will be the development and integration of spectroscopic catalogs and related information with the HSC.

The tools used to access the HSC will also be enhanced in the next few years. One of the primary goals in the near term is to better integrate the three tools discussed in Section 5 (i.e., the MAST Discovery Portal, the HSC CasJobs service, and the HSC home page). Another challenge on the longer term will be to develop tools to more easily combine and compare multiwavelength data sets (e.g., with different spatial resolution) and multidimensional data-cubes (e.g., from ALMA and JWST).

We also encourage the development of value-added-projects (e.g., a ESA-based project to develop a Hubble Variable Catalog based on the HSC is already planned), determinations of transformation equations to support addition of data from different instruments, determinations of photo-Z's, etc . We expect that in many cases, the products of the value-added projects will be integrated into future version of the HSC.

8. SUMMARY

Version 1 of the Hubble Source Catalog includes WFPC2, ACS/WFC and WFC3 photometric measurements based on the SExtractor source lists from DR8 of the Hubble Legacy Archive. The current version of the catalog includes roughly 80 million detections of 30 million objects involving 112 different detector/filter combinations and about 50 thousand HST exposures. The mean photometric accuracy is better than 0.10 mag and the astrometric residuals are typically within 10 mas, with a value for the mode (i.e., peak) of 2.3 mas.

Astronomical catalogs have been a mainstay for centuries. Historical examples include the Messier, Herschel and New General Catalogs. More recent examples include 2MASS, Hipparcos , and SDSS. In many ways the Hubble Source Catalog will be unique, first and foremost because of the depth and spatial resolution of Hubble Space Telescope. In addition, the HSC will be an important reference for future telescopes, such as James Webb Space Telescope, and survey programs, such as LSST.

In this paper we have attempted to find the right balance between demonstrating the great potential of the HSC, and cautioning HSC users about potential pitfalls. The key point is that by its very nature (i.e., deep pencil-beam observations using a wide variety of instruments and observing modes), the HSC is a very different database than most other surveys that have uniform "all-sky" coverage (e.g., SDSS). While the diversity of the HSC dictates the need for caution when developing queries, it also provides the opportunity for cross checking the results in many cases, unlike many other catalogs.

Astronomers will use the HSC in different ways. At the most basic level, it provides a quick way to determine what *Hubble* observations have been taken of an object. When building their own catalogs, the HSC can be used as a consistency check. Some people will use the HSC to do feasibility checks, and to perform preliminary analysis. In other cases users will use the catalog to address their primary science goals.

A number of improvements are planned for the HSC in the future. In the short term this includes bringing the quality of the WFPC2 and ACS source lists up to the same standards at the WFC3 source lists. In the near future, source lists for ACS/HRC and NICMOS will be added to the HSC. In the longer term, deeper mosaic images will be constructed based on the combination of all the Hubble images in a region, and will serve as the detection image. The quality of both the photometric and astrometric measurements will continue to improve as known problems are fixed and new reduction techniques are incorporated.

TBD - Acknowledgments.

The Hubble Legacy Archive (HLA), is a joint project of the Space Telescope Science Institute (STScI), Space Telescope - European Coordinating Facility (ST-ECF) and the Canadian Astronomy Data Centre (CADC).

REFERENCES

Anderson, J. & Bedin, L. R. 2010, PASP, 122, 1035
Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393
Bernard et al. 2010
Brown, T. M. et al. 2009, AJ, 184, 152
Budavári, T., & Lubow, S. H. 2012, ApJ, 761, 188
Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301
Dolphin, A. E. 2009, PASP, 121, 655
Fruchter, A., & et al. 2009, HST MultiDrizzle, HST Data Handbooks,
Gladders et al. 1998
Jenkner, H. et al. 2006, ASPC, 351, 406
Lubow, S., & Budavári, T. 2013, Astronomical Data Analysis Software and Systems XXII, 475, 203
Stetson, P. B. 1987, PASP, 99, 191
Whitmore, B., Lindsay, K., & Stankiewicz, M. 2008, Astronomical Data Analysis Software and Systems XVII, 394, 481

APPENDIX

Appendix A - Access to Information

The three primary access points for the HSC are:

- Mast Discovery Portal - mast.stsci.edu
- HSC CasJobs - mastweb.stsci.edu/hcasjobs
- HSC Home Page - archive.stsci.edu/hst/hsc/

Other sources of more detailed information can be found at:

- HSC Use Cases (http://archive.stsci.edu/hst/hsc/help/HSC faq.html#use_case) - this contains several use cases that show how to use the HSC to do science analysis.
- HSC FAQ (http://archive.stsci.edu/hst/hsc/help/HSC faq.html) - this contains answers to question about the HSC.
- HLA FAQ (http://hla.stsci.edu/hla faq.html) - this contains answers to questions about the HLA itself, including about how the source catalogs used in the HSC were generated.
- Discovery Portal User's Guide (http://mast.stsci.edu/portal/Mashup/Clients/Mast/data/html/MastHelp.html) - this is the User's Guide for the Discovery Portal. Note that are video demonstrations available.
- HSC CasJobs Guide (http://mastweb.stsci.edu/hcasjobs/guide.aspx) - this contains information about CasJobs in general