VALIDATION REPORT
January 30, 2004
THE GLIMPSE LEGACY SCIENCE Team

EXECUTIVE SUMMARY

The Spitzer IRAC images are of high quality and we are pleased to recommend commencement of the GLIMPSE survey using our proposed survey strategy. This strategy consists of moving the telescope in half-frame steps, giving 2 visits on the sky, each of 2 second frametimes (1.2 sec. exposures) in all four IRAC bands. This gives maximum areal coverage providing maximal science return. To accomplish the Observing Strategy Validation (OSV), ten hours of Director's Discretionary time were used to observe a representative part of the Galaxy using the GLIMPSE observing strategy, repeated five times. The goals of the validation observations were to:

1. Determine the minimum criteria to achieve reliability of at least 99.5% for the GLIMPSE Point Source Catalog (GPSC).
2. Assess if two two-second observations are sufficient to a) achieve reliable flux densities and positions of point sources in fields typical of the GLIMPSE survey area, and b) produce mosaics of high quality.
3. Determine if a 12 pixel overlap (between frames and AORs) is adequate to ensure that no gaps in the survey occur.
4. Assess the effects of instrumental artifacts (e.g., banding, column pulldown, cosmic rays, muxbleed, latency) on the survey.

Our findings are as follows:

1. To produce a Catalog with ≥99.5% reliability using our current pipeline processing and 2-visit observing strategy requires the following:
    a. The flux threshold for a source to be reliable depends on the background level. We find that in regions in which the background level in band 4 is less than 56 MJy/sr (1.9 mJy/pixel), the Band 1 and 2 fluxes are reliable to ≥99.5% down to 3-4 mJy. To achieve similar reliability in bands 3 and 4 requires background levels < 30 MJy/sr (measured in band 4) and flux densities > 30 mJy (band 3) and >4 mJy (band 4).
    b. For a source to be considered reliable it must be detected twice in one band and at least once in an adjacent band.
    c. Sources in regions affected by instrumental artifacts such as banding, column pulldown, and wings of saturated stars are excluded from the Catalog.
2. Two visits are sufficient to achieve the necessary reliability requirements, photometric accuracy, positional accuracy, and high quality mosaics.
3. Due to the accurate pointing of Spitzer (<1") and the high quality of the IRAC cameras, 12-pixel overlaps between frames and AORs are sufficient to prevent gaps in coverage of the survey, and to produce accurate photometry and mosaics.
4. Neither our Catalog nor Mosaic products are unduly affected by instrumental artifacts. Sources in affected regions are flagged and currently not included in the Catalog. We have preliminary corrections for column pulldown and banding; we see no adverse effects from latents. Cosmic rays are not a problem due to our short frametimes and source selection criteria, though we still intend to mask them.

Continuing development on both the GLIMPSE and Spitzer Science Center (SSC) pipelines will improve our products. We have already made substantial improvements to our source extraction algorithms that have not been implemented into our production pipeline yet, and are confident we will see even better results within a few months. However, even with our current pipeline processing, we can produce reliable catalogs and high-quality mosaics for delivery.

1. THE OSV OBSERVATIONS

To properly validate the observing strategy it was necessary to choose a region that displays the full range of conditions the GLIMPSE survey is likely to encounter. This means point source magnitudes that range from the detection limit to saturation, point source spatial densities ranging up to the confusion limit, and a large region that displays bright and spatially complex diffuse background emission. The luminous massive star formation region RCW49 in the southern Galactic plane meets these requirements. An area of about 1.7 x 0.6 degrees centered near RCW49 was observed by Spitzer on 23 Dec. 2003 by using precisely the same observing strategy that the GLIMPSE program proposes to use. The OSV consisted of a total of 20 AORs. Each AOR provided strips of about 1.7 x 0.17 degrees (about 40x2 frames) which were placed side-by-side with a small overlap perpendicular to the strip direction and half-frame overlaps in the strip direction. The total area imaged was about 1.7 x 0.6 degrees. The set of four AORs used to cover this area was repeated five times, each with a different overlap perpendicular to the strip direction (8, 10, 12, 15, and 20 pixels). This ensured that each point in the area was observed a minimum of 10 times, and provided a test of required overlaps. A total of 1648 exposures were taken (giving 6592 Basic Calibrated Data, BCD, frames). Each frame was exposed for 1.2 s. The total observing time was 10 hours.

The following sections describe the analysis of the OSV data. After briefly describing our pipeline processing in Section 2, we proceed to address the four goals described in the executive summary. Section 3 discusses reliability, Section 4 demonstrates that 2 visits are sufficient to achieve science goals, Section 5 discusses overlaps, and Section 6 instrumental effects.

2. GLIMPSE PIPELINE PROCESSING

We received calibrated IRAC images, dmasks (data masks), signal_to_noise fits files, and mosaics from the SSC of our OSV region on Januray 8, 2004. In preparation for running the data through our pipeline, we updated pmask (bad pixel masks) and dmask files; and created new stellar point spread functions for each IRAC band for use in our photometry routines. We modified the SSC pmask files posted 20 November 2003 by adding 4 masked pixels in band 1, 4 in band 2, the top two rows in band 3 and the first row in band 4 plus one pixel. Since the dmask does not yet contain saturated star information, we estimated the positions of saturated stars by finding clusters of high pixels. We set the saturation limit for the clusters to a conservatively low value since saturated pixels can have lower than fullwell values due to wrap-around. The dmask is updated with this information for use in our pipeline.

The OSV data were tagged with a unique project name to track frame overlap and number of visits. Some of the pipeline procedures are as follows: The SSC-provided pointing refinement keywords (using 2mass PSC) were implemented, though the pointing appears good without it. A zodiacal light level is calculated and subtracted from each BCD frame. A column pulldown correction is applied to band 1 and band 2 data, based on comparison of column median values to a running median across the frame. A preliminary banding correction is applied to band 3 and band 4 data based on parameters derived from IOC data (which need more tweaking, perhaps due to the focus change). Sources extracted in the column pulldown and banded regions are flagged. The saturated star bit in the dmask is used to create a saturated star wing area, currently a circular region surrounding the saturated star. Sources extracted in this area are also flagged. All the flags are incorporated into a GLIMPSE source quality flag associated with each source. We are continuing to tune these corrections.

The saturated pixels, bad pixels, and other severe effects are masked in the calibrated images so they are not included in future processing (source extraction, mosaicing). When the SSC dmask is functioning well, we will use that information to flag sources for latency, muxbleed, and more precise determination of saturated pixels. We plan to implement a "muxbleed fitter" (algorithm may be provided by SSC) to improve the muxbleed correction for bright sources. Currently there is no masking for cosmic rays. We plan to use the SSC Mosaicer for outlier detection and feed that information into our pipeline. With our 2 second frametimes, we expect only about 2 cosmic rays per frame. We have not masked out stray light areas yet but we have the capability of doing that manually using our Quick Look Validation Tool (QLVT). We can also mask cosmic ray streaks with the QLVT, if they are not masked by the SSC mosaicer.

Sources are extracted using a modified version of DAOPHOT (Stetson 1987, PASP, 99:191). DAOPHOT routines were modified to improve finding sources in high variable background areas, reducing the occurance of false source detection. Source lists are created listing total counts for each source, xy pixel coordinates, flux errors, and S/N estimates.

We use the SSC bandmerger to first merge the multiple detections of a source within a band and then detections between bands. The GLIMPSE source quality flag is used to cull problematic data (in regions affected by banding, column pulldown, saturated stars, etc) from the GLIMPSE Point Source Catalog (GPSC) and GLIMPSE Point Source Archive (GPSA).

3. GLIMPSE SOURCE ARCHIVE AND CATALOG: QUALITY OF A 2-PASS SURVEY

3.1 Completeness and Reliability

The GLIMPSE team will release a subset of our detected point sources as a very high-reliability (≥99.5%) Catalog. Additional sources of sufficient quality (predicted to be S/N>5) will be released in the GLIMPSE Archive. To confirm that these goals can be met, we first determined how to best calculate the completeness and reliability of a set of detected sources. Next, using the Validation data, we determined the criteria for including sources in the Catalog while maintaining very high reliability. These criteria are expected to include the number of detections in single IRAC frames, the flux density of the source, the GLIMPSE quality flag (indicating possible contamination by an image artifact), and possibly the level of diffuse background immediately around the source.

3.1.1 Method for Computing Reliability and Completeness

The simplest and most robust measure of completeness and reliability (hereafter "C&R") is to compare one's source list with a truth list of the sources that are present in the image. We refer to such a comparison as external C&R. Unfortunately, such a truth list is not in general available, and one must calculate internal C&R using observational data alone. To calculate internal C&R, one typically performs multiple observations of the same region, and uses that additional multiplicity to construct an internal truth list.The multiplicity is also used to calculate the probability of each source being cataloged using various observing strategies. . For GLIMPSE we imaged each point in the Validation region 10 times, permitting a comparison of catalog criteria using 2 or 3 exposures. The attached GLIMPSE Completeness and Reliability document presents a thorough examination of internal completeness and reliability. In simulated IRAC data, we determined that requiring 2 detections in one band and one detection in a neighboring band, and a flux density greater than a few mJy resulted in a set of sources with ≥99.5% reliability. We also established that a robust internal truth table results for a given band if at least 8 detections (out of 10 possible detections) are found in that band or an adjacent band. This criterion was determined based on simulations and was found to best match external reliability.

3.1.2 Required Number of Detections

We find that the real data are very similar to the simulated data in most respects, and that the best criterion to place on sources for Catalog inclusion is two detections in one IRAC band

and at least one detection in an adjacent band (2+1). This criterion ensures reliability ≥99.5% while simultaneously preserving a high completeness (>98%). Although this introduces some color dependence into the selection process, it provides a very high reliability, and is a very effective filter for cosmic ray hits (Section 6.3). Only sources with the most extreme colors have a chance of being rejected from the Catalog, and any color effects will be quantified in later GLIMPSE documentation. We determined that in real and simulated data alike, simply requiring two detections in one band, and no restriction on other bands (2+x) does not maintain high reliability even at fairly high flux density levels, and that more restrictive criteria than 2+1 (e.g. 2+2) do not improve the reliability at any flux density, but do reduce completeness by removing some true sources.

3.1.3 Required Flux Density

In simulated data, a Catalog containing sources detected twice in one band and at least once in a neighboring band has ≥99.5% reliability down to flux densities of 3, 3, 3 and 4 mJy in bands 1-4, respectively, and that sources with flux densities less than that should be excluded from the Catalog. In the validation data, we have found that high reliability requires placing thresholds on both the background brightness and source flux densities.

3.1.4 Required Background Level

The mid-IR sky contains bright and very highly structured diffuse emission, in particular in IRAC bands 3 & 4. While the absolute level of this background in the Galactic plane is no surprise (and in fact our simulated data accurately modeled that level based on MSX data), the complexity and power at high spatial frequencies is remarkably high. Figure 1 is a 3-color image of RCW49 showing this complex structure. Such background stresses the abilities of any source extractor to distinguish point sources from knots of diffuse emission. We are improving our point source extractor, and have already reduced the number of questionable sources in high background regions by a factor of 3 to 5 (with more improvements on the way). However, the current version of our processing pipeline produces some unreliable sources in the highest background regions, and if these are included, it is difficult to construct a high reliability catalog of sources that are only detected in bands 3 and 4.

We calculated the reliability and completeness of a Catalog for all regions in which the diffuse flux was <56 MJy/sr in band 4. Areas below this background limit comprise about 90% of the OSV region (see Figure 2, in which the contour at 56 MJy/sr is drawn), and about 97% of the total GLIMPSE survey area (based on MSX data, see Figure 3). A total of 70658 point sources were found in the selected OSV region. Figure 4 shows that with this rejection of only about 3% of the GLIMPSE surveyed area, a Catalog of sources detected 2+1 times has a reliability of ≥99.5% in bands 1 and 2 down to flux densities of ~3 mJy, satisfying the stated GLIMPSE science goals. Such a Catalog is >98% complete. To achieve reliability ≥99.5% in bands 3 and 4, different thresholds for background brightness and flux densities are required. Figure 5 shows reliabilities for sources located in regions of background brightness <30MJy/sr (true in ~2/3 of the GLIMPSE survey area). We are experimenting with different background criteria in order to include the largest possible number of sources in the Catalog without compromising high reliability.

Figure 1. An IRAC three-color composite image of the RCW49 region. IRAC bands 1, 3, and 4 are displayed as blue, green and red. Quite bright and spatially variable diffuse emission is seen over a region of about 0.33^0 x 0.18^0 centered on the RCW49 nebula. Diffuse emission is seen in all IRAC bands and is brightest in bands 3 and 4. This emission is believed to be mostly produced by PAH features that fall within IRAC bands 1, 3, and 4. Some diffuse emission could also be produced by stellar UV radiation stochastically heating small grains that reradiate in the mid-IR.

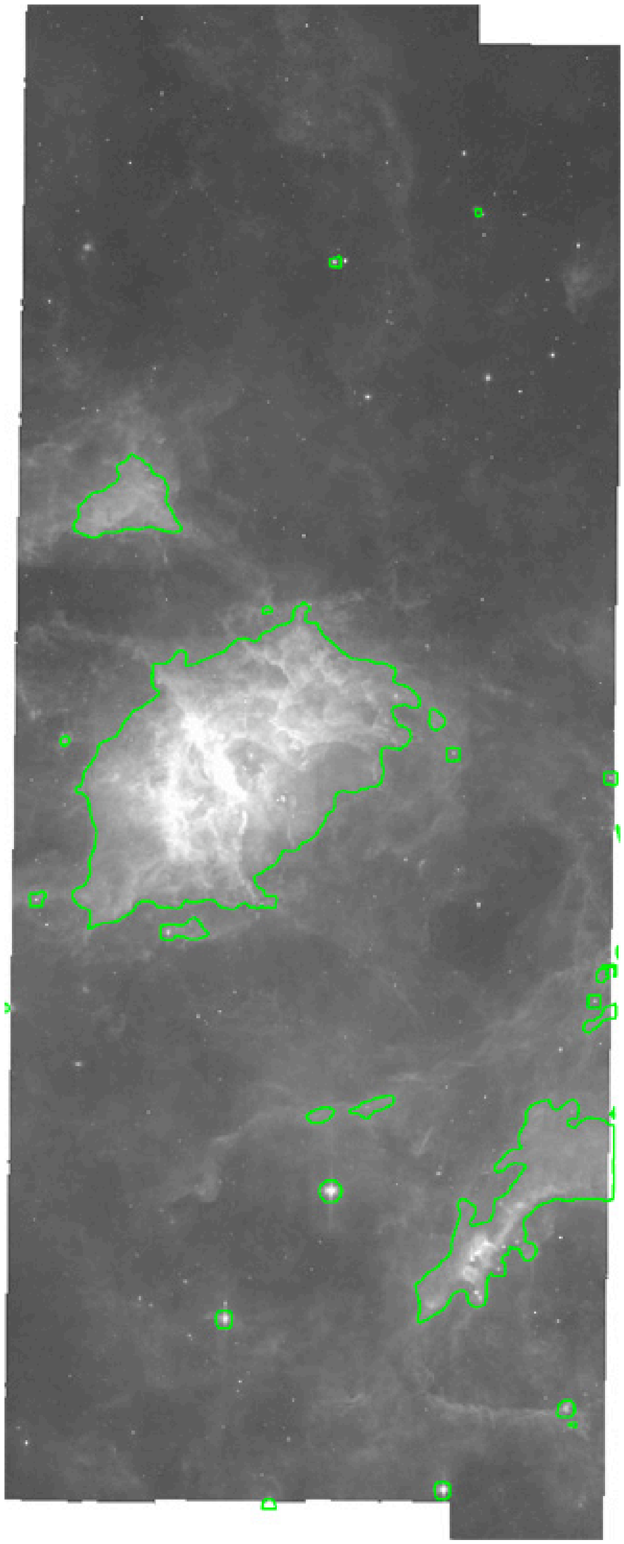

Figure 2. A mosaic of the OSV region using one pass (2 exposures) in IRAC band 4. The areas enclosed within the green contours have background levels >56 MJy/str and represent ~10% of the OSV area. Stars within these areas would currently be excluded from the GPSC. The contour marks 56MJy/sr.

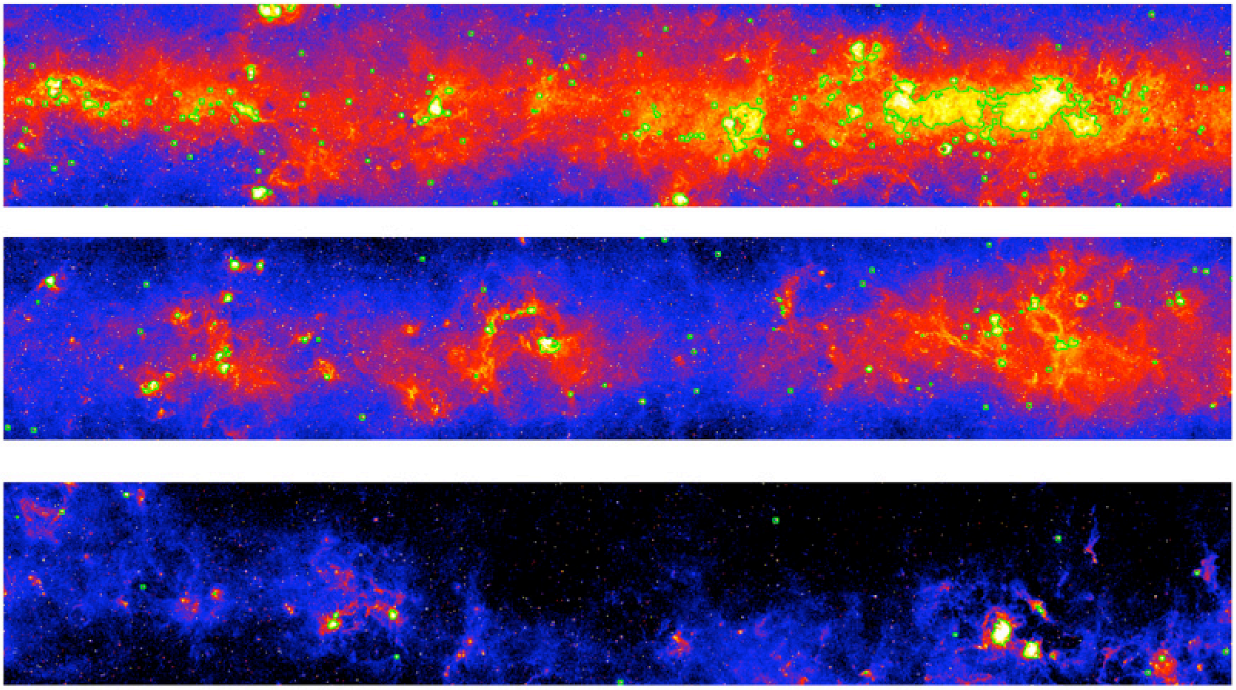

Figure 3. MSX images of the northern GLIMPSE survey region. The yellow regions, outlined in green, have background levels >56 MJy/str and represent ~3% of the GLIMPSE survey area. Stars within these areas would currently be excluded from the GPSC.

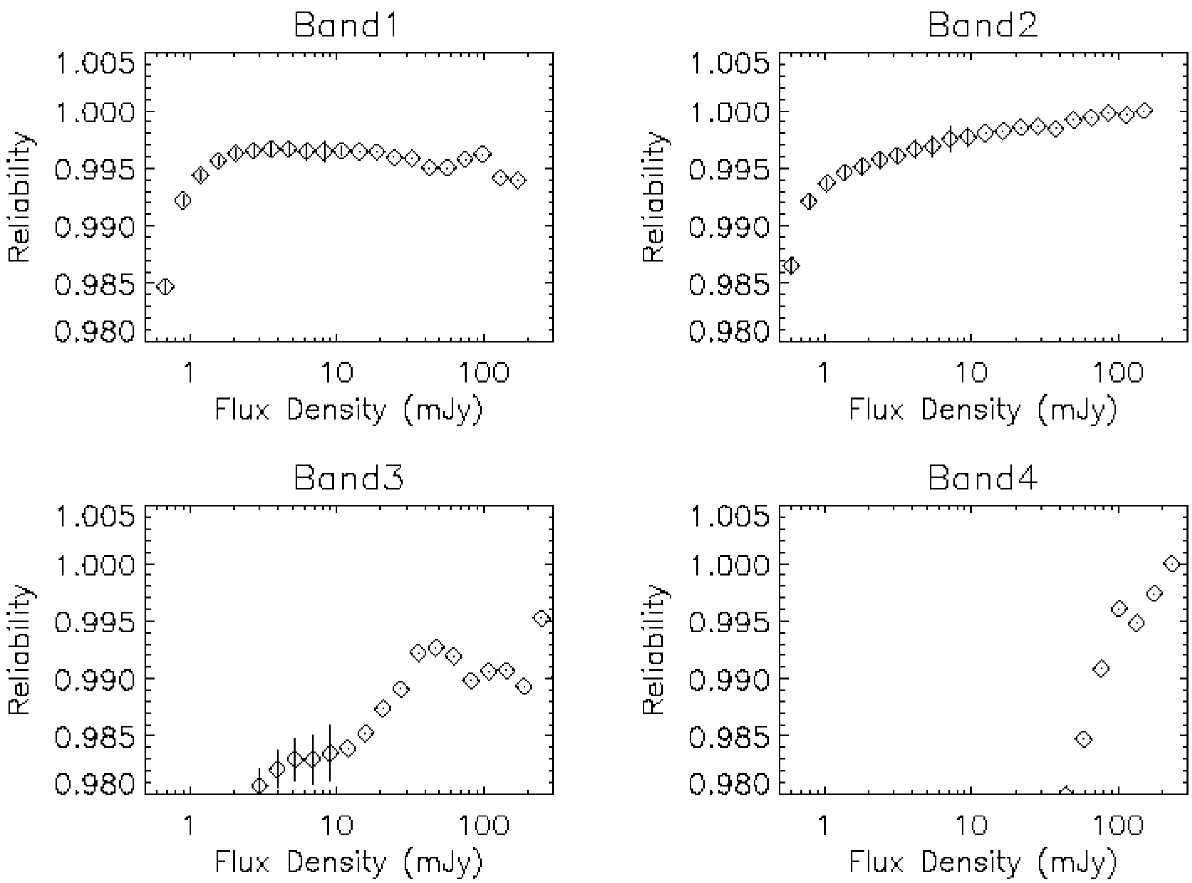

Figure 4. The calculated reliabilities for each IRAC band. These calculations included stars only in regions where the background brightness in band 4 was <56 MJy/str (~90% of the OSV region and ~ 97% of the GLIMPSE survey region). Obviously, high reliability is not achieved in bands 3 & 4 even at flux densities as high as 100 mJy at background levels as high as 56 MJy/str.

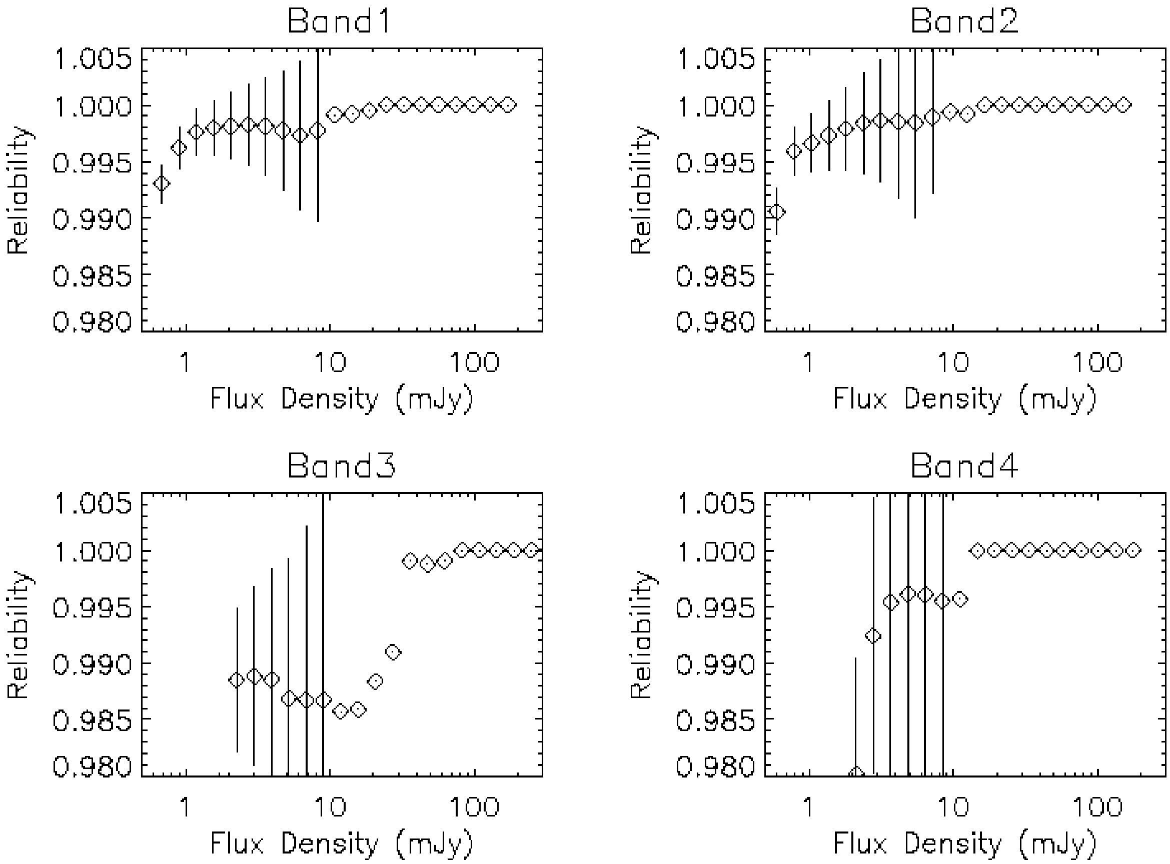

Figure 5. The calculated reliabilities for each IRAC band for stars at b>0.4 degrees. This illustrates that in low background regions (<30MJy/sr) it is possible to obtain reliabilities ≥99.5% in bands 3 and 4 for stars with flux densities ≥30 mJy (band 3) and ≥4 mJy (band 4).

3.1.5. Summary, Status, and Future Work

- We can achieve high reliability down to ~3mJy flux density consistent with the GLIMPSE science goals, for sources detected in IRAC bands 1 and 2, over about 97% of the survey area with our present pipeline processing.
- Highly structured background produces detections of questionable sources in IRAC bands 3 and 4. Sources fainter than 30 mJy in band 3 and 4 mJy in band 4 can be included in the Catalog without loss of reliability, as long as they have detections in bands 1 or 2. Since the SEDs of even reddened main sequence stars are brightest in bands 1 and 2, the vast majority of sources in our survey will be selected based on these bands. Faint very red objects (without good detections in bands 1 or 2) can be included in the Catalog for a large part of the GLIMPSE surveyed area, but more work is required in regions of high background. Several options for producing a high reliability catalog are available, including ongoing work on the extraction software.
- If we had to produce a Catalog today, it would be based only on bands 1 & 2 and bright

sources in bands 3 & 4 located in low background regions; and it would be ≥99.5% reliable.

- The GPSA will be a best effort to include point sources down to about the 5 sigma level with, of course, somewhat lower (but still high) reliability and greater uncertainties which will be indicated by flags associated with each source in the Archive. The GPSA is expected to reach flux density limits of ~ 1 mJy.

3.2. Photometric Accuracy

Flux densities for 5 calibration stars located in the OSV region have been provided by Martin Cohen. We have used these stars to check extracted GLIMPSE fluxes and point source calibration of the SSC BCD data. The flux densities of several (between 4-7) sets of two-observation averages are plotted for each IRAC band in Figure 6. The GLIMPSE extracted fluxes in each band agree to within 5% on average with the predicted fluxes given by Martin Cohen which are based on Kurucz model atmospheres convolved with the IRAC bandpass responses. This demonstrates accurate calibration on the part of the SSC pipeline and accurate flux density extraction on the part of the GLIMPSE pipeline.

3.3. Crowded –Field Photometry

The GLIMPSE point source extractor, a modified version of DAOPHOT, is robust and accurate in dense fields. In Figure 7 we show a frame from the OSV region that illustrates a typical stellar field with and without overlays of the GLIMPSE point sources (green circles) and the 2MASS point source catalog (pink pluses). The bottom panels show a relatively dense star cluster with and without the same source overlays. One can see that DAOPHOT has done a good job of finding the point sources in both regions. Our techniques for calculating fluxes in complex backgrounds will also be useful in crowded fields.

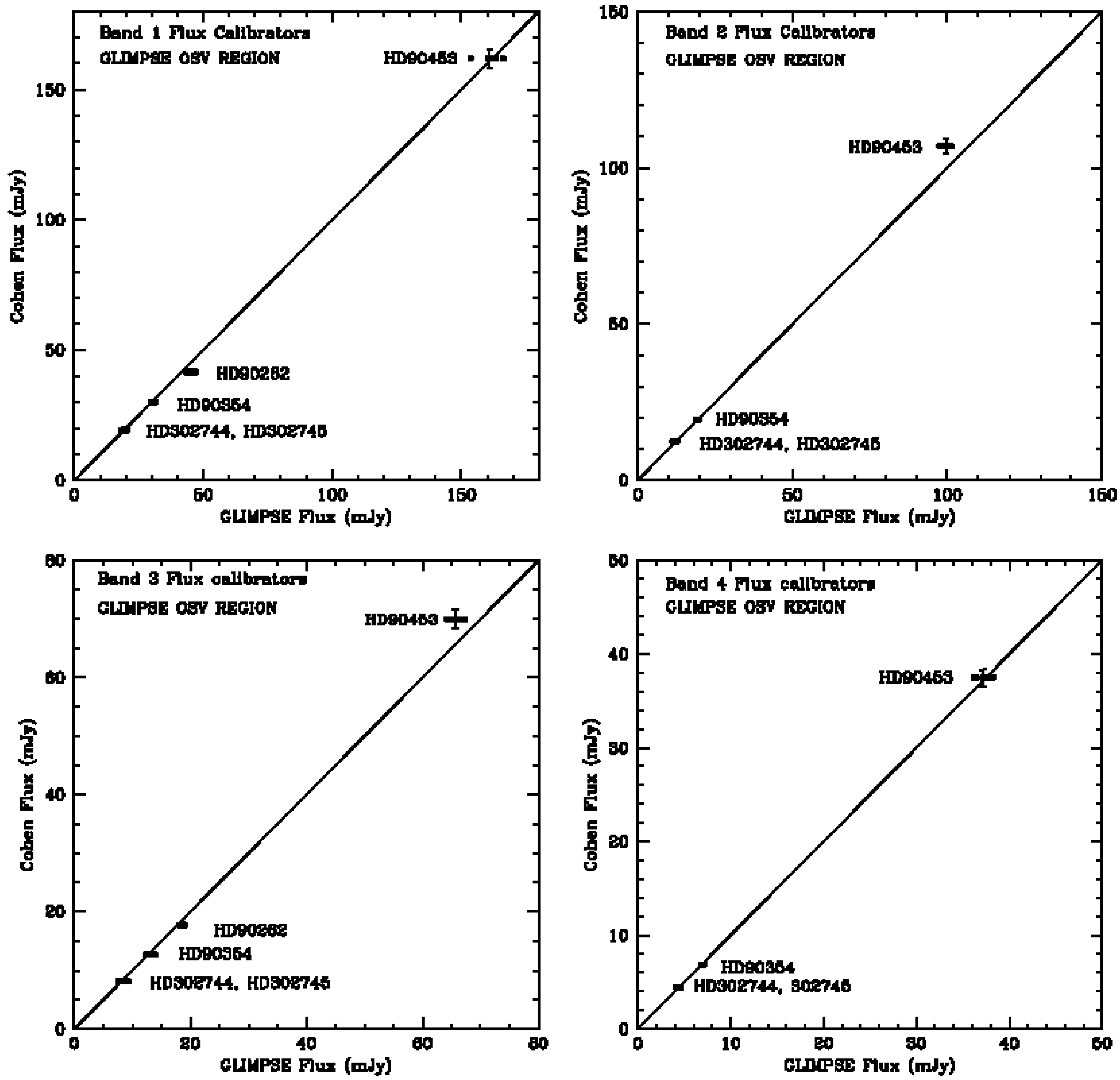

Figure 6: Comparison of GLIMPSE extracted flux densities versus those provided by Martin Cohen for a set of stars with well determined spectral and luminosity types. Cohen's flux densities are based on Kurucz model atmospheres convolved with IRAC bandpass responses. Between 4-7 observations are plotted for each star and band. On this scale, they lie nearly on top of each other in most cases. The vertical error bars are Cohen's estimated errors. In most cases, they are smaller than the symbol size.

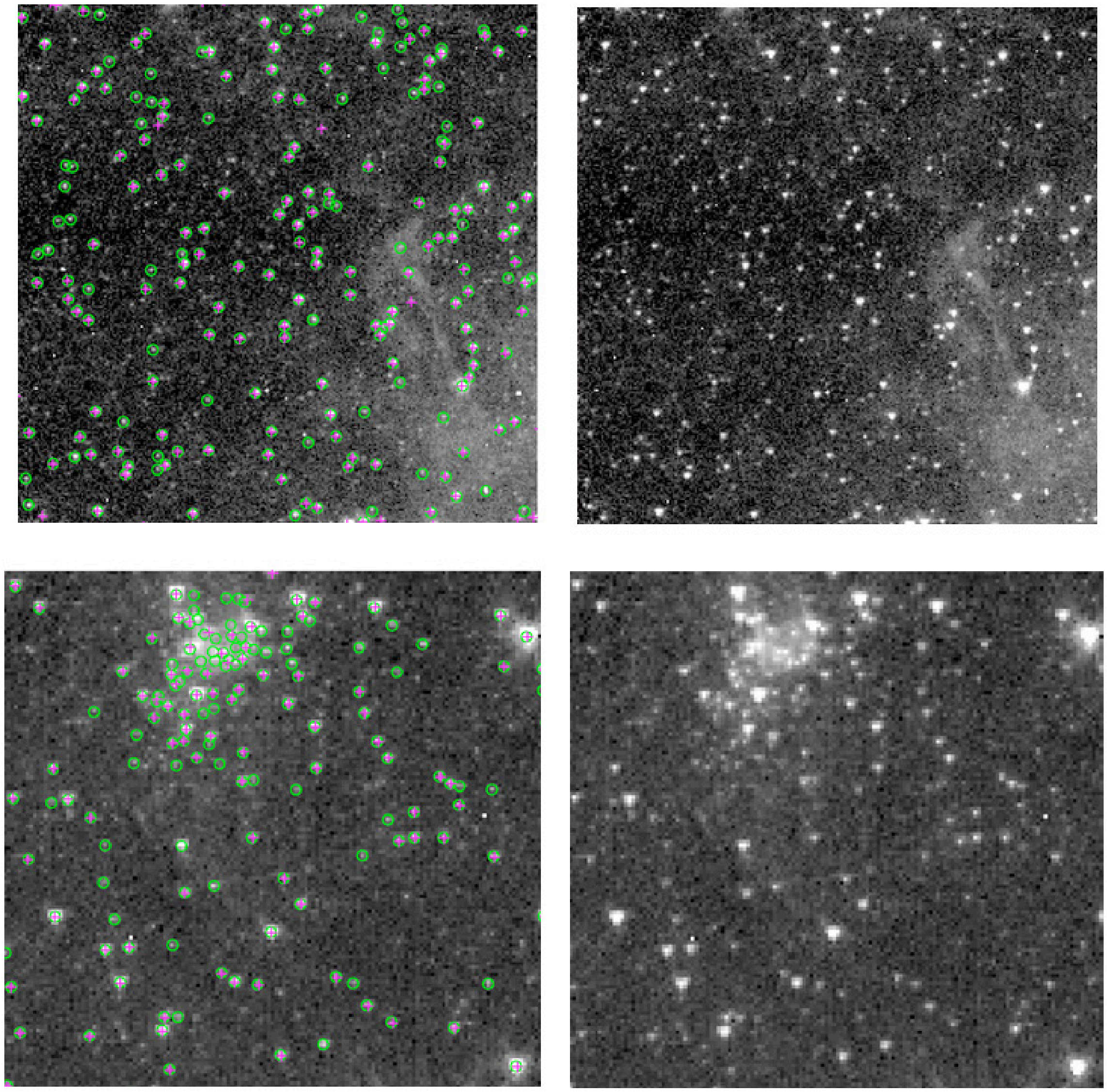

Figure 7. The top panels show a band 1 IRAC frame of a typical field in the OSV region. At left, the GLIMPSE extracted sources (green circles) and 2MASS catalog (pink pluses) are overlayed. The bottom panels show a zoom of a dense star cluster to illustrate how well GLIMPSE and 2MASS do in very crowded fields.

4. ASSESSING THE TWO-VISIT STRATEGY

Most of the science goals of the GLIMPSE survey require as large an areal coverage as possible for galactic structure determinations and statistical studies on a galactic scale. It is therefore vital to cover a particular patch of sky in as few passes as is consistent with the photometric requirements of the science. We must determine if these goals can be met with two passes. We now describe how reliability, photometric accuracy and precision, positional accuracy, and instrumental artifacts are affected by 2 visits vs. 3.

4.1 Reliability and flux limits

Figure 8 shows that there is no increase in reliability nor decrease in the flux limit at which reliability is ≥99.5% is achieved for 3 coverages versus 2 coverages. This was also found in the simulations (see attached document). The reliability is more dependent on background levels and flux density thresholds than number of visits. Multiple visits do not improve the situation because false sources are still found in the complex background. This is a systematic effect, not a statistical one.

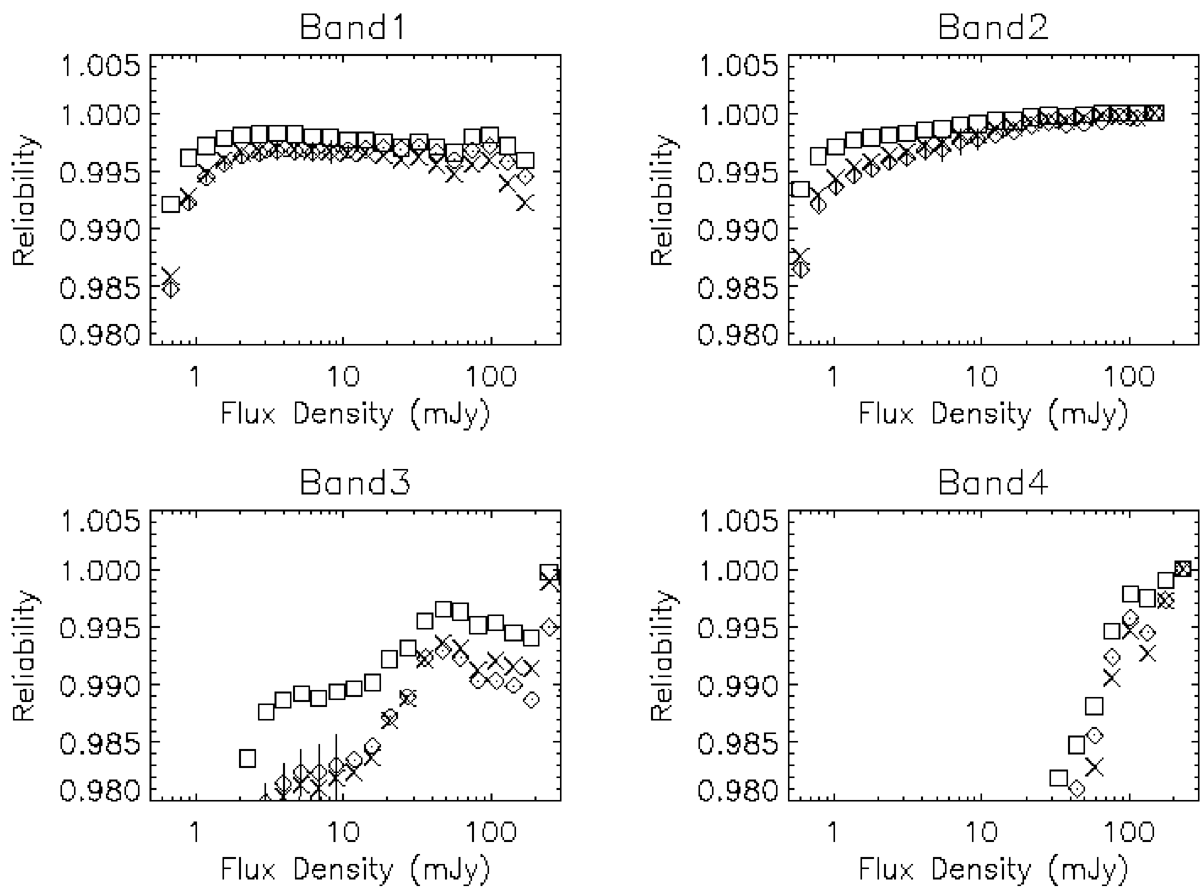

Figure 8. Reliability determinations using the same background limits (and area) as in Fig. 4. Here we compare 2 observations and 3 observations using selection criteria 2+1 (diamonds) for two observations and 3+1 (X) and 3+2 (squares) for three observations. Both 2+1 and 3+2 meet the 99.5% reliability requirement at flux densities >1 mJy, but we believe that 2+1 is the better choice because it also offers higher completeness (see attached document).

4.2 Photometric accuracy

Dithering would provide a better sampling of the point spread function which might increase photometric accuracy. However, three passes is not necessarily dithering. To do a true dither would decrease our areal coverage by a large factor which would be out of the question. Our photometric accuracy with 2 visits is good as is shown in Figure 6.

4.3 Photometric Precision

For all detected point sources in the OSV region, we calculated the flux densities that would have been measured in the 40 possible 2-pass GLIMPSE surveys which one can construct out of 10 observations on the sky. That is, the mean of each pair of observed flux densities was taken to be the flux density that would have been obtained by performing GLIMPSE once. We then calculated the RMS of the flux densities in those 40 realizations of a 2-pass survey. Finally, we divided the sources into 0.1 dex logarithmic bins, and calculated the median of the RMS flux density dispersions for the sources in each bin. This "median RMS flux density dispersion," calculated as a function of flux density, is shown in Figure 9, and provides a measure of the photometric precision of GLIMPSE.

We determined that photometric precision based on two visits is between 5-10%. We also calculated precision for a 3-pass survey by choosing combinations of 3 observations from the 10 available and taking the mean of those flux densities, to construct many realizations of a 3-pass survey. The photometric precision for a GLIMPSE-like survey with 3 visits is 4-7%. For a typical 5 mJy source, the precision would go from 9% (2 visits) to 6% (3 visits). This does not seem to be sufficient reason to go to 3 visits.

4.4. Positional accuracy

The Spitzer Space Telescope has a pointing accuracy of ~1". GLIMPSE positions based on 2 visits are accurate to about 1", satisfying our science goals. Comparisons of IRAC1 with 2MASS K shows good agreement (see Fig. 7).

4.5 Instrumental artifacts and other effects

The instrumental artifacts are discussed in more detail in Section 6. We note here that, with the exception of cosmic rays and latents, these are fixed on the sky—e.g., along the column containing a bright star—and therefore multiple passes will not remove them. Due to our short exposure times, neither cosmic rays nor latents are serious problems for us.

4.6 Mosaics

Figures 1 and 10 show that our 2-visit strategy produces high quality mosaics. Even with instrumental artifacts there will be very few gaps in the survey. There may be some gaps due to masking stray light (from bright point sources). We conservatively estimate we would mask 0.01% of our pixels.

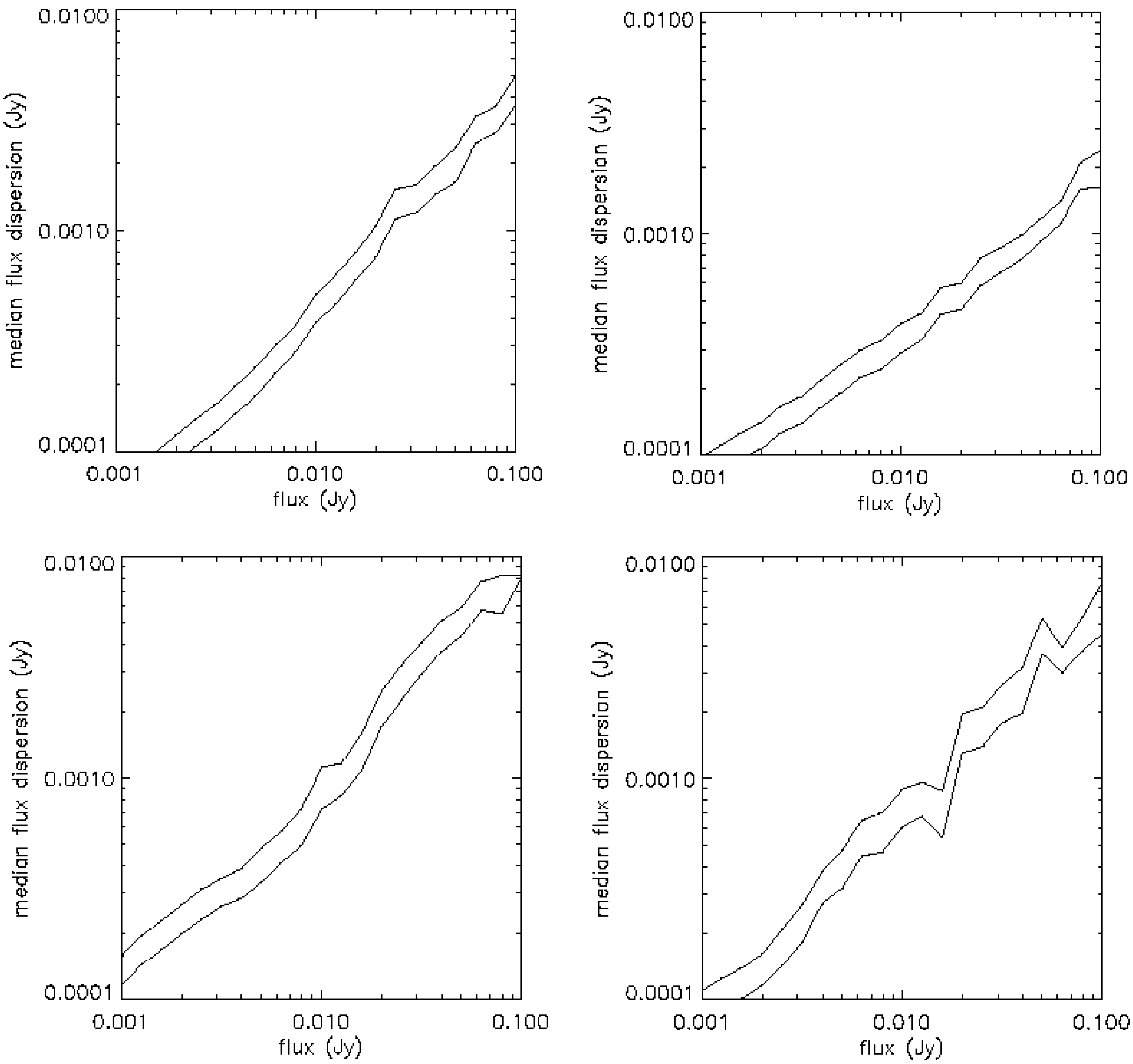

Figure 9. A measure of the photometric precision of a 2-pass survey (top line) vs a 3-pass survey (bottom). The four panels correspond to bands 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right).

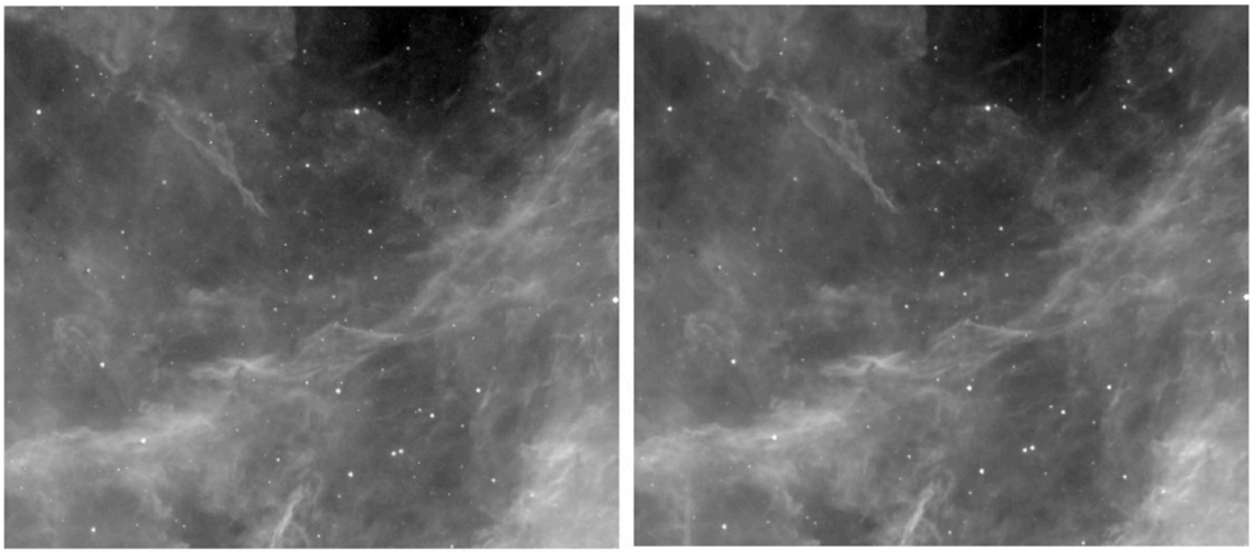

Fig. 10. These images show a 0.25x0.25 degree region in band 4. One of these is a 2-visit mosaic (the pass with 12-pixel overlaps). The other is a 10-visit mosaic. Can you guess which is which?

5. OVERLAPS BETWEEN FRAMES AND BETWEEN AORS

Our validation observations consisted of five passes, each with a different overlap between frames and AORs (8, 10, 12, 15, and 20 pixels). Our proposed overlap is 12 pixels. One of the goals of the OSV data was to determine if this overlap is sufficient.

During the period of the validation observations, the telescope roll angle was changing at a rate of 0.86^0 per day. This is faster than will occur anytime during the GLIMPSE survey, so the validation observations provide a stringent test of the possibility of missing narrow strips due to roll angle changes between AORs. Examination of mosaics of all five passes, including an 2-degree long AOR, found no gaps or missing strips. We are reasonably confident that the12-pixel overlap proposed for the GLIMPSE survey will produce a fully sampled region with no missed pieces of the sky in the survey area.

To determine quality of data near frame edges, we calculated average skies of our 1648 images in each band. To calculate these we removed the stars and computed averages and standard deviations of each pixel. We also computed medians of all the frames. We found that the first ~8 columns of band 1 are up to 10% noisier than the other columns, due to the decrease in sensitivity as shown in the flats. Band 2 has three low columns (165-167). Band 3 has pin-striping in the columns that is also seen in the darks. Band 3 has two bad rows and band 4 has one. None of these effects will impact frame overlaps with 12 pixels.

In our analysis of the photometric standards, we found no correlation between flux densities and position on the chip.

6. ASSESSMENT OF INSTRUMENTAL AND OTHER EFFECTS

6.1. Effect of Reduced Sensitivity in Bands 3 and 4

The sensitivity of band 3 is about 40% lower than expected, and that of band 4 is about 10% lower. The effect on GLIMPSE is minimal. Due to the higher backgrounds in bands 3 and 4, we reach the background limit before we reach the confusion limit. Higher sensitivity would improve our detection limits only in low-background (< 10 MJy/sr) regions.

6.2 Saturated stars

Our point source extractor finds false sources in the bright wings of saturated stars. We currently flag a circular region around saturated pixels and exclude these sources from the GPSC. SSC does not yet provide a reliable saturated star predictor, so we have made our own as described in Section 2. We are developing an algorithm to extract accurate fluxes and positions (with appropriate errors and flags) of stars located in the outer wings of bright stars.

6.3 Cosmic Rays

Simulations and OSV data show that we typically have an average of about two cosmic ray (CR) hits per frame. Assuming a CR illuminates 3 pixels on average, the probability of seeing this twice in one band and at least once in an adjacent band at the same point on the sky, assuming 2 such CR hits per frame, would be $(2x3/256^2)^2 \sim 8x10^{-9}$. Any gambler would love these odds. CRs will not affect the reliability of the Catalog or Archive. However, we will want to eliminate them from our images. Our current plan is to use the SSC mosaicer to generate cosmic ray masks using its dual outlier rejection routines. In addition, we plan to examine all of our GLIMPSE 0.33^0x0.33^0 deliverable mosaics, and will have the opportunity to flag remaining defects including cosmic rays and stray light at this time.

6.4 Other Instrumental artifacts.

Other IRAC artifacts that can affect the determination of accurate fluxes and detection of false sources are banding in IRAC bands 3&4, muxbleed and column pulldown in bands 1&2, latency, and stray light. We have made preliminary corrections for banding and pulldown in the OSV data processing, and we flag sources in these regions. We believe that the algorithms can be improved further with more experience with the data. The SSC pipeline does not fully remove muxbleed around bright stars. Our point source extractor does find sources along the leftover muxbleed. However, they are all less than 5-σ above the background, and would not appear in our Catalog or Archive. Latency does not appear to be a major problem with our very short exposures. A 0-magnitude star showed no obvious latency after two 2 sec exposures in the IOC data that we analyzed. The IRAC cameras are delivering high quality images and none of the artifacts appear to be serious enough to prevent the GLIMPSE survey from achieving its scientific goals