Report comparing source extractors: SUSSEXtractor in DAOPHOT, SUSSEXtractor in HIPE and DAOPHOT in HIPE

Document number: SPIRE-SUX-REP-003256

Anthony Smith, University of Sussex

27th October 2010

1 Introduction

This document is a test report, comparing three source extraction modules: SUSSEXtractor and DAOPHOT in HIPE, and an implementation of the SUSSEXtractor algorithm in IDL. This IDL implementation of SUSSEXtractor was written by Richard Savage as a prototype of the Java implementation contained in HIPE. The SUSSEXtractor algorithm itself is described in [AD01].

The purpose of this report is twofold:

1. To verify that all of the algorithms product satisfactory results on simulated data, and to recommend changes where the algorithms are deficient

2. To verify that the HIPE implementation of SUSSEXtractor gives the same results as the IDL prototype, and to recommend further tests to be developed in case of any differences (recognizing that the IDL prototype itself may be at fault in some cases).

1.1 Applicable Documents

AD01 Bayesian Methods of Astronomical Source Extraction, Savage & Oliver (2007), ApJ, 661, 1339

1.2 Build and images used for the tests

The HIPE source extractors were tested using continuous integration build 6.0.309 (with a fix applied for HCSS-11567).

For the tests, simulated SPIRE images were created:

1. Artificial image of a sparse grid of point sources, each with the same Gaussian profile, and with various flux densities between 10 and 1000 mJy. The spacing of the grid is 50 pixels, with (6, 10, 14) arcsec pixels for (PSW, PMW, PLW) respectively, but the sources are placed at random within the pixel so the pixel centre does not generally correspond to the source position.

2. Artificial image of a crowded field with no noise

3. Artificial image of a crowded field with Gaussian random noise

For each test, an error map was created with the same value in each pixel. Throughout, the point sources are Gaussian with FWHM (18.15, 25.15, 36.3) arcsec for (PSW, PMW, PLW) respectively, with a Gaussian approximation used for the beam area.

2 Results for a sparse grid of point sources

In the first test, SUSSEXtractor and DAOPHOT in HIPE are run with their default settings, and SUSSEXtractor in IDL is run with corresponding settings (tophat prior and no background fit). (The test was also run with a Jeffreys prior, rather than a tophat prior, producing results with no noticeable difference.)

2.1 Flux density accuracy

A comparison between the true and measured flux densities is shown in Figures 1–3. From the top panels it can be seen that all three methods perform well, with a close correlation between the true and measured flux densities.

For the IDL SUSSEXtractor, this seems to perform well, the measured flux density apparently being an unbiased estimator of the true flux density. The dominant source of error is multiplicative, and is likely to be related to the position of the source within the pixel (see below).

For the HIPE SUSSEXtractor, this performs similarly to the IDL SUSSEXtractor, except that a linear offset seems to have been introduced to the flux densities, producing a significant underestimation of the flux density at faint flux densities. This difference should be investigated, since the prototype is clearly performing better in this case (see Section 5).

For DAOPHOT in HIPE, the recovered flux density is consistently between 92 and 94% of the true flux density. This is likely to show the need for aperture corrections to be applied (see Section 5).

2.2 Positional accuracy

Figure 4 shows the accuracy of the positions of the recovered sources. It can be seen that the IDL prototype gives positions accurate to the nearest pixel (but with a systematic offset of unknown cause). However, the HIPE source extractors refine the position to a much greater accuracy.

Figure 5 shows the ratio of the measured to the true flux density, as a function of the distance between the true source position and the centre of the pixel. For the IDL SUSSEXtractor there is no obvious trend with distance from the centre of the pixel, but for the HIPE SUSSEXtractor (neglecting the problem with faint sources noted above) there is clear trend with distance from the pixel centre, such that sources further from the pixel centre have their flux densities underestimated by a larger factor than those near the pixel centre (see Section 5). For DAOPHOT, it can clearly be seen that the required aperture corrections have a (small) dependence on distance from the pixel centre.

3 Results for a crowded field with no noise

The purpose of this test is to investigate the effects of confusion noise on the extraction of bright point sources.

In this test, the same 81 source positions from the previous test are used, but each source is given the same flux density, of 100 mJy. The images are then populated with 10 000 sources, placed randomly, with flux densities between 1 and 30 mJy (drawn from a uniform distribution in log-flux density). The mean intensity of the images is subtracted, to simulate default SPIRE processing, in which a baseline is subtracted from the timelines.

The source extraction is then performed for the three algorithms as in the previous test, and the extracted sources closest to the 81 bright sources are found.

Figure 6 shows the ratio of the measured to the true flux density for these 81 sources, as a function of the distance between the true source position and the centre of the pixel. All three algorithms behave very similarly in the presence of significant confusion, giving a similarly-wide scatter and no obvious trend with the true source's distance from the pixel centre.

Figure 7 shows the accuracy of the measured positions of these sources. There are some outliers for SUSSEXtractor (IDL), in which (presumably) the true source has not been detected, and a neighbouring source has been chosen instead. For SUSSEXtractor (HIPE) and

DAOPHOT, comparing with Figure 4, the presence of significant confusion makes the estimation of the source position much worse, slightly more so for SUSSEXtractor than for DAOPHOT, but the positions are still good to within approximately 0.5 pixels.

A very low threshold was used for the source extraction, yielding for (PSW, PMW, PLW), (340, 340, 340) sources for SUSSEXtractor (IDL), (2930, 3843, 3670) sources for SUSSEXtractor (HIPE) and (6192, 7544, 7307) for DAOPHOT. Assuming that these are all genuine detections, this suggests that DAOPHOT and SUSSEXtractor (HIPE) are both performing well at detecting faint sources, with DAOPHOT performing better, and that the IDL SUSSEXtractor prototype is not performing well at finding faint sources. This last point may explain the outliers for SUSSEXtractor (IDL) in Figure 7.

4 Results for a crowded field with Gaussian random noise

The purpose of this test is to investigate the combined effect of confusion and Gaussian random noise.

The images from Section 3 were used, with Gaussian random noise added to the pixels, having $\sigma = 2.0$ mJy/pixel. (The peak of the 100 mJy sources is approximately 10 mJy/pixel.)

Figures 8 and 9 show the results of the source extraction in these cases, as in Figures 6 and 7, respectively.

The effect of this noise is noticeable, but generally the flux densities are no worse than before. However, the positional accuracy is made worse by the presence of strong pixel-to-pixel noise, but most positions are still good to within approximately 0.8 pixels.

5 Conclusions

Tests have been performed on SUSSEXtractor (in IDL and in HIPE) and on DAOPHOT (in HIPE), using the default settings in HIPE. The main conclusions are as follows:

1. In general, all of the algorithms perform well, finding the sources and giving reasonable estimates of their flux densities and positions.

2. In the absence of confusion, some systematic biases are evident for SUSSEXtractor (HIPE) and DAOPHOT (see recommendations below).

3. There are noticeable differences in the results of the two SUSSEXtractor implementations (IDL and HIPE). These need to be checked in the HIPE implementation, with clear tests verifying that the algorithm performs as it should (see recommendations below).

The recommendations of this report (with associated SCRs) are as follows:

1. The SUSSEXtractor IDL prototype performs better than the HIPE SUSSEXtractor for faint sources in Fig. 2. The cause of this should be investigated and the HIPE SUSSEXtractor should be modified to improve its performance [SCR HCSS-11571].

2. The SUSSEXtractor IDL prototype gives a good estimate of the source flux density, irrespective of the position of the centre of the source within the pixel, whereas the HIPE version of SUSSEXtractor gives an underestimate that is systematically worse for sources further from the pixel centre. This should be investigated [SCR HCSS-11572].

3. DAOPHOT in HIPE requires aperture corrections to be applied, as is clear from Fig. 3 [SCR HCSS-11573].

4. Tests for SUSSEXtractor: what should it do in ideal situations? [SCR HCSS-8346]

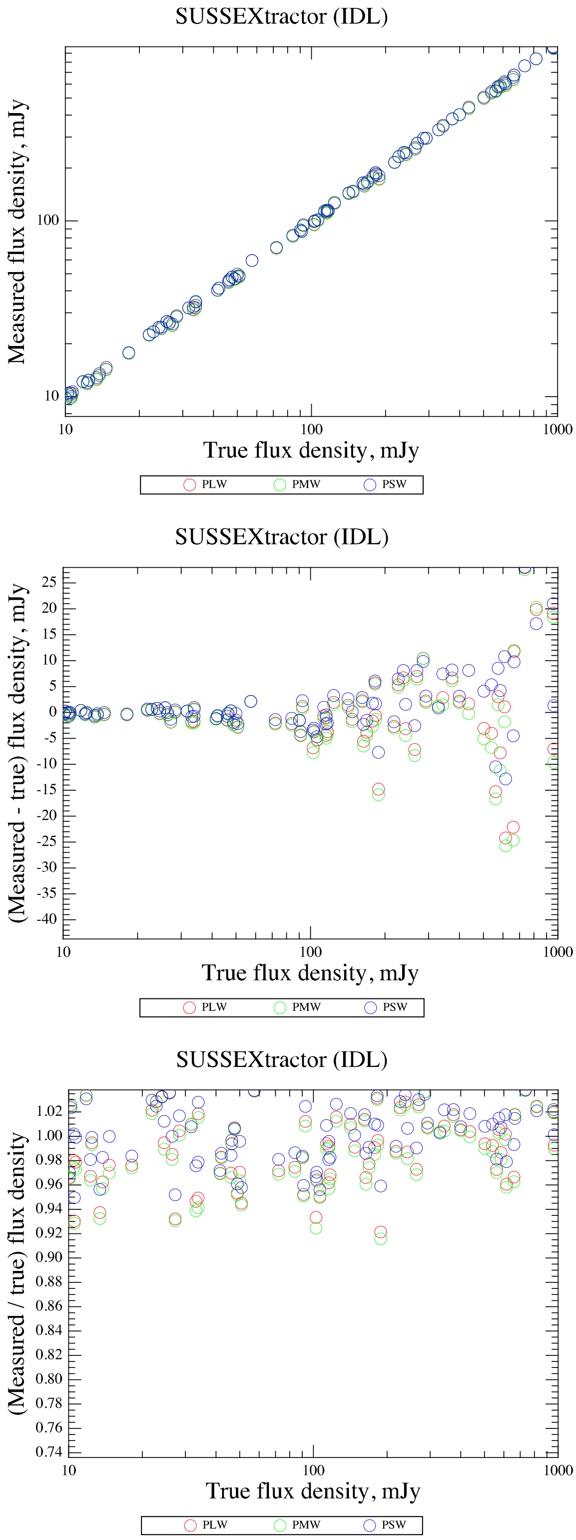

Figure 1: Comparison between the measured and recovered flux for the SUSSEXtractor IDL prototype.

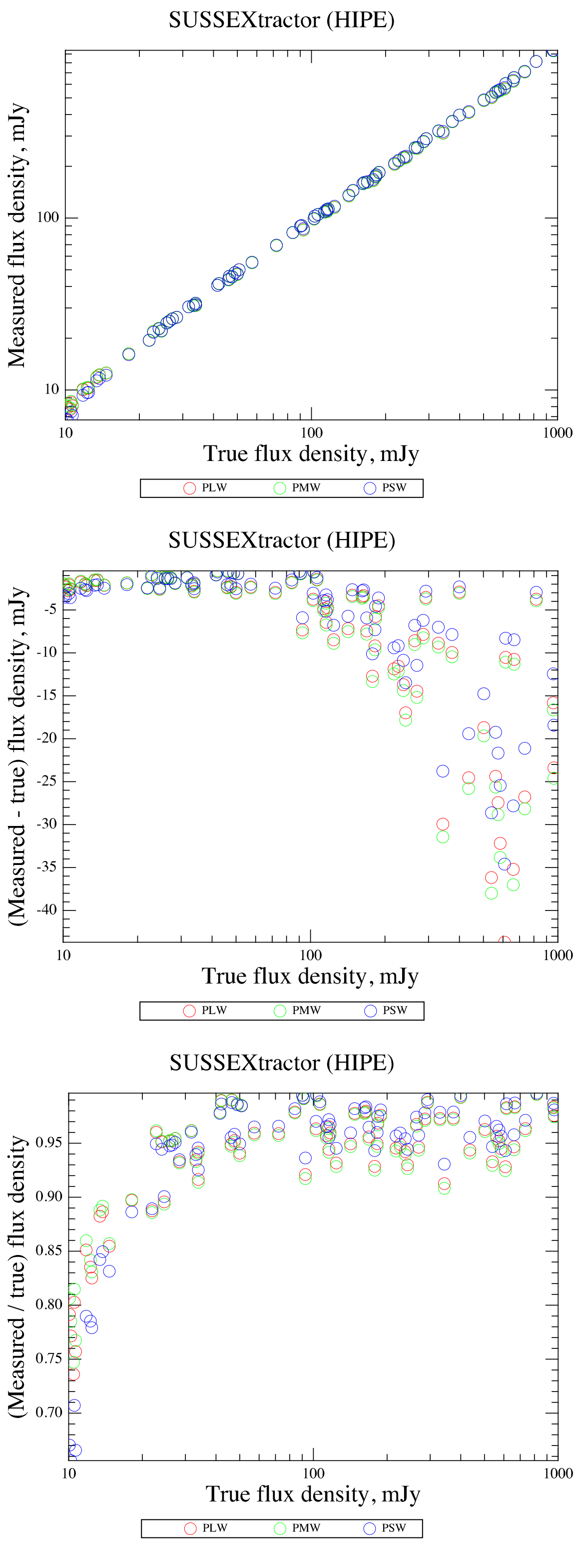

Figure 2: Comparison between the measured and recovered flux for the HIPE SUSSEXtractor.

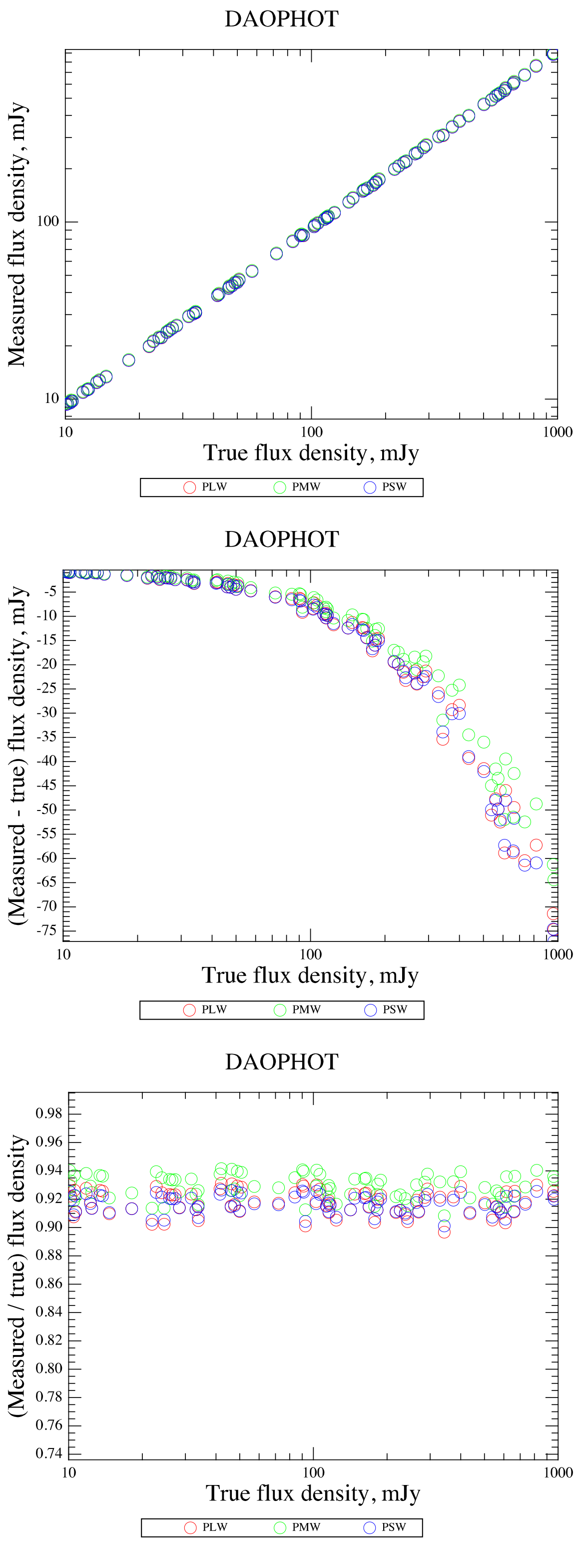

Figure 3: Comparison between the measured and recovered flux for the HIPE DAOPHOT source extractor.

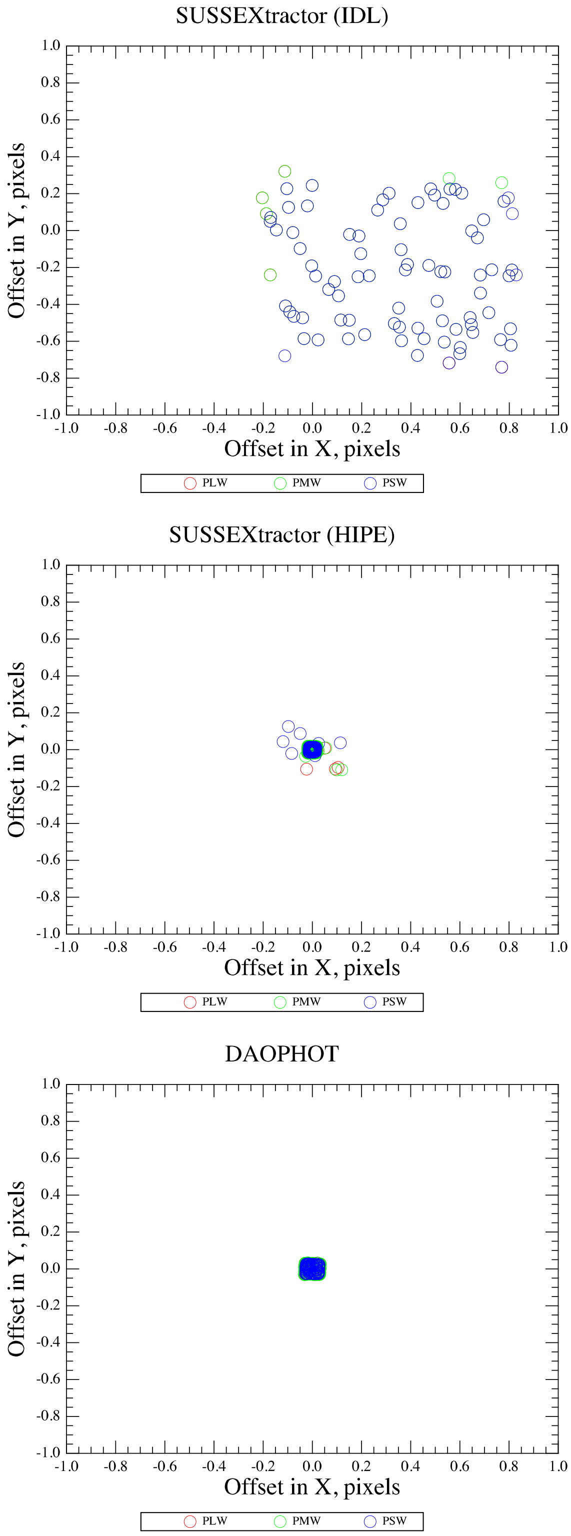

Figure 4: Distance between the measured position of a source and its true position, in pixel co-ordinates.

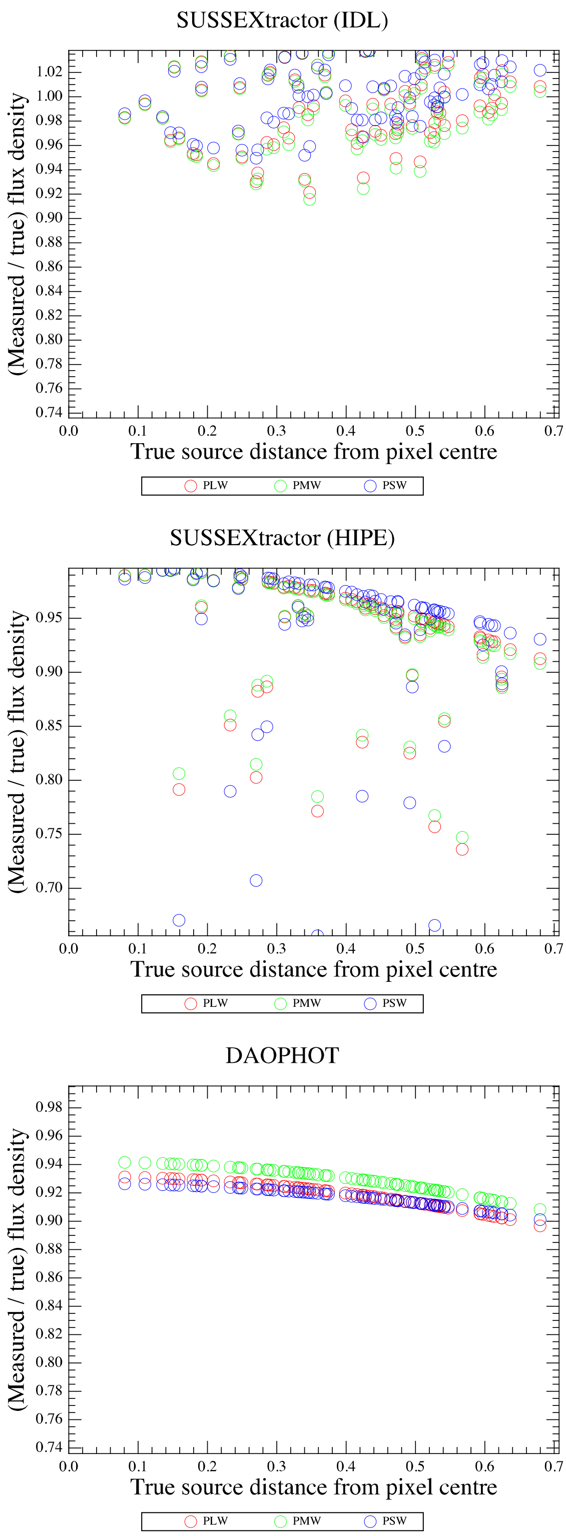

Figure 5: The ratio between measured and true flux density, as a function of the distance between the true source position and the centre of the pixel in which the centre of the source lies.

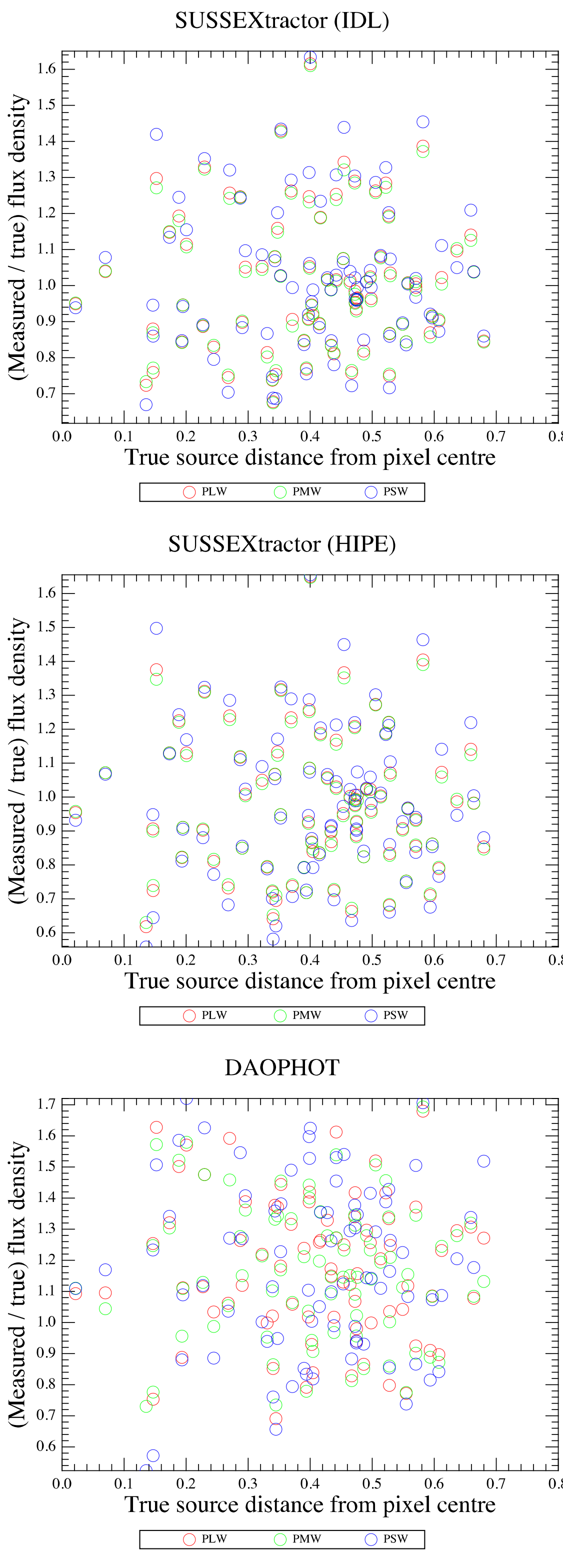

Figure 6: For a crowded field with no noise, the ratio between measured and true flux density, as a function of the distance between the true source position and the centre of the pixel in which the centre of the source lies.

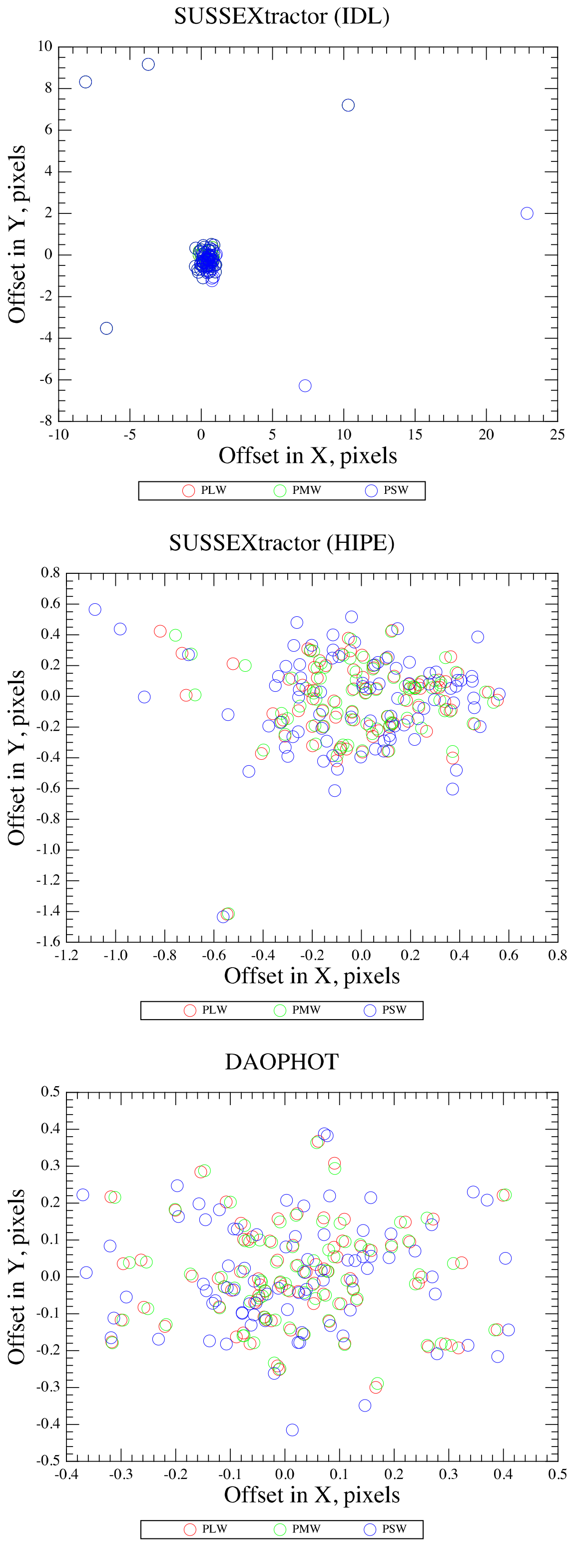

Figure 7: For a crowded field with no noise, the distance between the measured position of a source and its true position, in pixel co-ordinates.

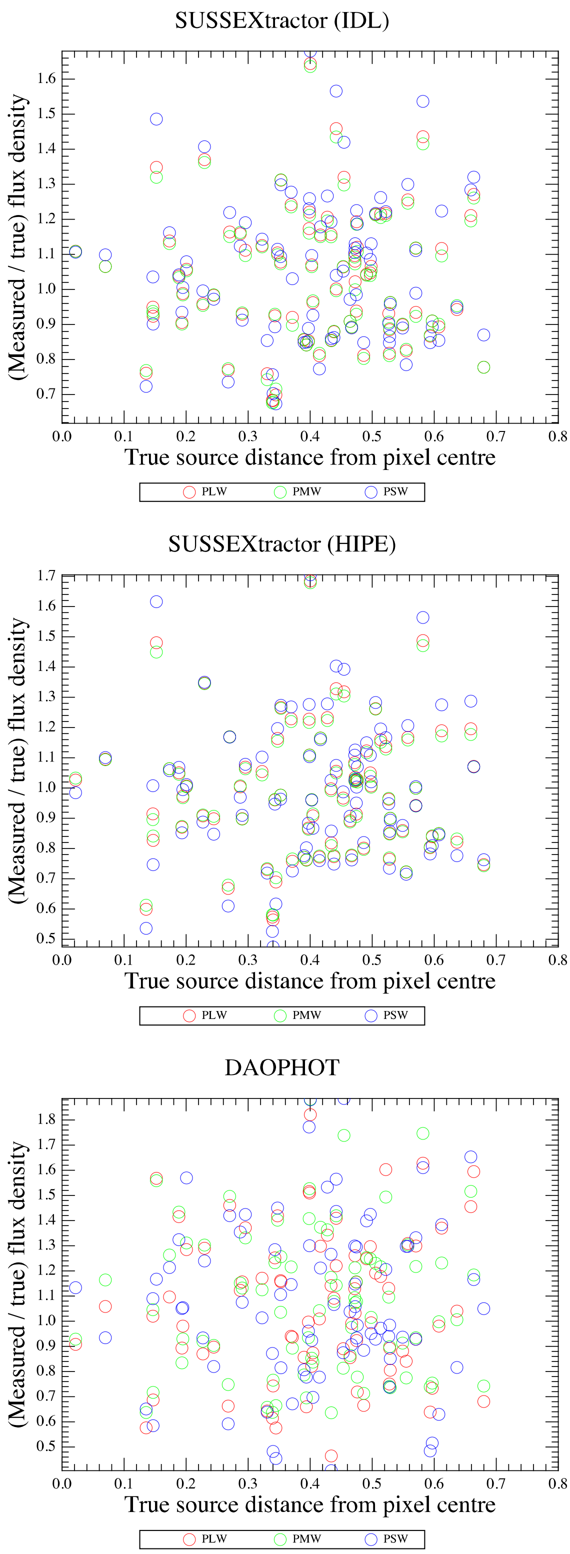

Figure 8: For a crowded field with Gaussian random noise, the ratio between measured and true flux density, as a function of the distance between the true source position and the centre of the pixel in which the centre of the source lies.

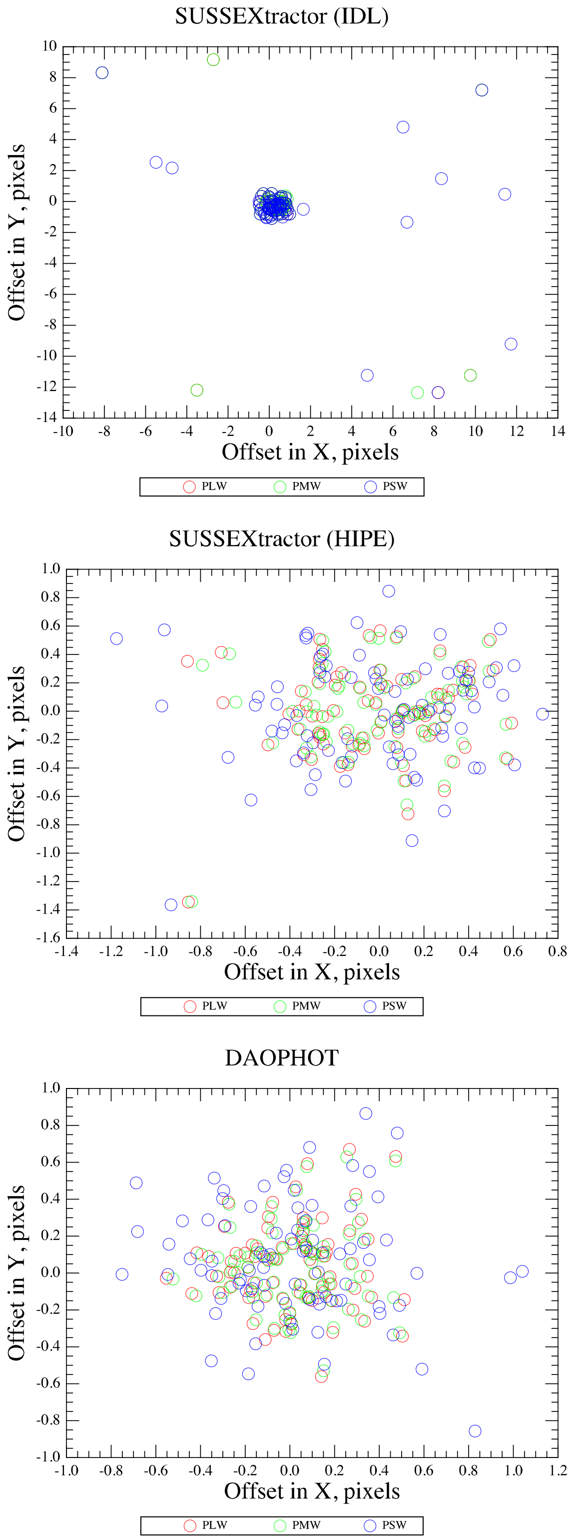

Figure 9: For a crowded field with Gaussian random noise, the distance between the measured position of a source and its true position, in pixel co-ordinates.