

## *EVLA Memo 116*

# *Automatic CLEAN windowing*

*W. D. Cotton*

*NRAO*<sup>1</sup>

### *Abstract*

This memo explores some of the effects of poorly constrained deconvolution of synthesis images that occur when the number of independent visibility measurements is comparable to or less than the number of beam areas in the derived image. In particular, the CLEAN algorithm is examined. These effects include the artificial reduction of the off-source RMS, creation of spurious features and the reduction of emission in the actual sources. Even in cases of relatively good uv coverage, CLEAN is known to remove flux from real sources producing the “CLEAN bias”. These effects can be large enough that seriously incorrect physical inferences may be made. This memo describes an automated procedure for specifying CLEAN windows which is useful for constraining CLEAN in wide-field and snapshot imaging. The procedure is implemented in the Obit (<http://www.cv.nrao.edu/~bcotton/Obit.html>) package and examples are given. Tests using seriously under-constrained VLBA data are presented demonstrating improved results. In cases of seriously under-constrained CLEANs (many more pixels than data), this technique significantly improves the results by reducing the effects mentioned above. Simple testing of the effects of this procedure on CLEAN bias show a very significant reduction in the bias, and is a potentially very useful tool for future surveys.

---

<sup>1</sup>The National Radio Astronomy Observatory (NRAO) is operated by Associated Universities Inc. under cooperative agreement with the National Science Foundation.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>autoWindow Technique</b>	<b>4</b>
<b>3</b>	<b>Verification</b>	<b>5</b>
3.1	Test One . . . . .	5
3.2	Test Two . . . . .	10
3.3	CLEAN Bias . . . . .	13
<b>4</b>	<b>Discussion</b>	<b>16</b>
<b>5</b>	<b>Conclusions</b>	<b>16</b>
<b>6</b>	<b>Acknowledgments</b>	<b>17</b>
<b>7</b>	<b>References</b>	<b>17</b>

## List of Figures

1	Test One. <b>Left:</b> UV coverage for a single channel of the test data on a stellar SiO maser. <b>Right:</b> Dirty beam, side-lobes above 30% shown. . . . .	6
2	Test One. Color coded velocity/intensity plot of a stellar maser in the $\nu=2$ , J=1-0 transition of SiO at 42.8 GHz. The region displayed is $37 \times 38$ mas and red=7.0 km/s, blue=-7.2 km/s. The intensity proportional to square root of intensity. The restoring beam used is $400 \times 200 \mu\text{sec}$ at position angle $35^\circ$ . . . . .	7
3	Test One. <b>Left:</b> Channel image deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of $\sqrt{2}$ times 75 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner. <b>Right:</b> Channel Image deconvolved using CLEAN and a “window” covering the full field of view. Contour levels are powers of $\sqrt{2}$ times 25 mJy/beam ( $\approx 2.5\sigma$ ). . . . .	8
4	Test One. Ratio of the Full Window image to the autoWindow image clipped at 50 mJy/beam. Overlaid is the 100 mJy/beam contour of the autoWindow image. The ratio is given by the scale bar at the top. The images going into the ratio are clipped at 2.5 times the RMS noise; this causes the spurious features in Figure 3 <b>Right</b> not to appear. . . . .	9
5	Test Two. <b>Left:</b> UV coverage for a single channel of the test data on a stellar SiO maser. <b>Right:</b> Dirty beam, side-lobes above 30% shown. . . . .	10
6	Test Two. Color coded velocity/intensity plot of a stellar maser in the $\nu=2$ , J=1-0 transition of SiO at 42.8 GHz. The region displayed is $34.5 \times 38$ mas and red=23.4 km/s, blue=13.7 km/s. The intensity proportional to square root of intensity. The restoring beam used is $400 \times 170 \mu\text{sec}$ at position angle $0^\circ$ . . . . .	11
7	Test Two. <b>Left:</b> Channel image of the full data set deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of $\sqrt{2}$ times 85 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner. <b>Right:</b> The UV coverage used in the image on <b>Left</b> . . . . .	12

8	Test Two. <b>Left:</b> Channel image from the short data segment deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of $\sqrt{2}$ times 100 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner. <b>Right:</b> Channel image from the short data segment deconvolved using CLEAN and a “window” covering the full field of view. Contour levels are powers of $\sqrt{2}$ times 30 mJy/beam ( $\approx 2.5\sigma$ ). . . . .	12
9	Test Two. <b>Left:</b> Closeup of region of reference image. Contour levels are powers of $\sqrt{2}$ times 100 mJy/beam. <b>Center:</b> Like <b>Left</b> but for short data segment deconvolved using CLEAN and windows selected by the autoWindow algorithm. <b>Right:</b> Like <b>Left</b> but for short data segment deconvolved using CLEAN and a “window” covering the full field of view. . . . .	13
10	Test Two. <b>Left:</b> Ratio of the constrained CLEAN image derived from the short data segment to the reference image. Overlaid is the 100 mJy/beam contour of the reference image. The ratio is given by the scale bar at the top. <b>Right:</b> Like <b>Left</b> but the ratio of the short segment unconstrained CLEAN to the reference image. Note difference in gray scale. . . . .	14
11	Test Two. <b>Left:</b> Closeup of ratio of the constrained CLEAN image derived from the short data segment to the reference image. Overlaid is the 100 mJy/beam contour of the reference image. The ratio is given by the scale bar at the top. <b>Right:</b> Like <b>Left</b> but the ratio of the short segment unconstrained CLEAN to the reference image. . . . .	14
12	CLEAN bias test. This plot shows the average bias of each of 9 artificial sources from 220 trials from a wide–shallow survey dataset expressed as a fraction of the model flux density. Stars represent results from the unconstrained CLEAN, plus signs the autoWindow CLEAN. . . . .	15
13	CLEAN bias test. This plot shows the average bias of each of 9 artificial sources from 220 trials from a wide–shallow survey dataset expressed in units of the RMS noise determined from each field. Stars represent results from the unconstrained CLEAN, plus signs the autoWindow CLEAN. . . . .	15

# 1 Introduction

Interferometric wide-field and “snapshot” imaging are often under-constrained in the sense that there are more independent pixels (beam areas) in the image than independent samples of the visibility function. In these cases, additional constraints are needed in the deconvolution process. Unconstrained, the CLEAN deconvolution algorithm tends to remove emission from actual sources by creating many spurious but weak sources while reducing the fluctuations in the background below the actual “noise” in the image. Deep CLEANing can reduce the background fluctuations to an arbitrary level. A discussion of these effects for the NVSS survey is given by Condon et al. (1998). A generic problem with wide area surveys is the “CLEAN bias” in which the deconvolution process removes flux from real sources and scatters it around the image. In the following, it will be shown that the reduction of emission in sources can be by large factors.

The most common constraints used in deconvolution are finite support on the sky and positivity (for Stokes I). The finite support is most strongly imposed by the field of view imaged but for algorithms like CLEAN, a further restriction can be applied by specifying regions where the process is allowed to model emission, the CLEAN windows (or boxes). In general, the detailed regions of the image containing emission are not known *a priori* and the CLEAN boxes are either supplied interactively or are specified in a generous manner. Interactive specification of CLEAN boxes is a commonly used technique but does not scale to large problems and introduces biases about what the sky should look like. An *a posteriori* constraint can be imposed that rejects weak and isolated emission.

Positivity can be a powerful constraint but can be difficult to apply, especially for CLEAN deconvolution. CLEAN must be allowed to remove negative components, both to correct for over subtraction and to model small sources which are not centered on pixels.

The following describes a procedure that emulates the interactive specification of CLEAN windows but in an unbiased way. This allows automating the processing of simple cases and improves the deconvolution in cases too large to manually specify CLEAN windows.

A further use of this technique is the one pioneered by the difmap package (Pearson et al., 1994, Sheperd, 1997) used for VLBI imaging in which the self-calibration process is guided automatically rather than by human intervention. This allows batch processing of non-phase referenced VLBI data.

Subsequent sections present tests using seriously underconstrained VLBA data and tests of CLEAN bias using data from a wide area survey. The effects of constraining the deconvolution is shown to make significant improvements.

All data manipulation discussed in this report use the Obit package (<http://www.cv.nrao.edu/~bcotton/Obit.html>) and plots other than the color velocity plots are generated in AIPS.

## 2 autoWindow Technique

In the technique described here (called “autoWindow”), there are two regions of interest in each facet being imaged; the region in which CLEANing is potentially allowed and that in which it is actually allowed. In the following, these are referred to as the “outer” and “inner” windows. In the case of multi-facet imaging, the outer window is less than the whole image and is restricted to the portion of the total field of view which the facet is intended to image. Imaging distortions (the “w” term) may corrupt sources outside of the outer window. The “inner” window is the combined CLEAN window in the conventional sense.

In the Obit implementation, the initial inner window may be a set of round and rectangular regions specified when the CLEAN begins, a set of regions specified interactively, or no windows. In the latter case, the initial window will be centered on the brightest portion of the region inside

the outer window. Subsequently, a single additional window may be added at the beginning of each major CLEAN cycle. If the CLEAN is run with the residuals being displayed each major cycle of CLEAN (or iteration of self-calibration), the inner window may be edited interactively to augment or modify the decisions of the autoWindow process. The autoWindow will proceed even with user intervention.

The autoWindow process is the decision of whether or not to add a new region to the inner window and if so where. In a simple CLEAN, this procedure is run each major CYCLE before a residual image is searched for new components. In self-calibration, this procedure is performed once each iteration before the CLEAN but using the final residual from the previous iteration. The autoWindow procedure is to examine the current CLEAN residual and determine where the maximum value inside the outer window occurs. If this is inside the current inner window, the definition of the inner window need not be changed. If the maximum residual is not in the current inner window and the residual exceeds a threshold (the Obit implementation uses 5 times the background RMS in the residual), then a new box is added to the inner window. If the inner window is to be expanded, a circular box is added centered on the peak residual and with a radius corresponding to the size at which the structure function of the residual image about the maximum residual drops to 10% of the peak or 3 sigma, whichever is less (but no larger than 20 pixels).

If Stokes I, Q, U and/or V are being imaged together, the inner window derived for Stokes I is used as the initial inner window for the other polarizations. For spectral line cubes, each channel CLEAN starts with the initial window specified (generally no window). For Stokes I, which in emission is always positive<sup>2</sup>, positivity is imposed by not adding a new window if the maximum absolute residual value corresponds to a negative residual. For other Stokes parameters, both positive and negative residuals must be considered. An alternate approach is deep cleaning followed by a subsequent filtering of spurious (weak or negative and isolated) components.

Use of the autoWindow technique has implications on the depth of a CLEAN major cycle. The CLEANing must not go significantly deeper than the maximum residual inside the outer window but outside the inner window, i.e. ignored in that CLEAN cycle. The Obit implementation sets the lower limit on a major cycle CLEAN residual to the greater of the maximum ignored residual value minus 0.5 times the background RMS or 0.5 times the background RMS. This constraint is also used to avoid excessive CLEANing. This limit allows the CLEAN to proceed to the point that an actual source outside the CLEAN box (i.e. not a side-lobe) can be added to the inner window on the next cycle but not so deep as to begin putting components on its side-lobes.

### 3 Verification

One case in which there are many more pixels in the derived image than independent UV samples is that of stellar SiO masers observed with the VLBA. The synthesized beam can be as small as a few hundred microarcseconds while the maser spots occur in a ring which can be several 10's of milliarcseconds across. Observations of such objects in "snapshot" mode further complicate the analysis.

#### 3.1 Test One

In the first test, a set of VLBA test data with three 20 minute snapshots, giving about 15,000 visibility measurements, are analyzed. One of the antennas (HN) failed entirely and the source was below the horizon at MK during one observation. The UV coverage obtained and the derived dirty beam for a given channel are displayed in Figure 1.

---

<sup>2</sup>An exception is continuum-subtracted imaging of absorption line spectra for which negative values need be included.

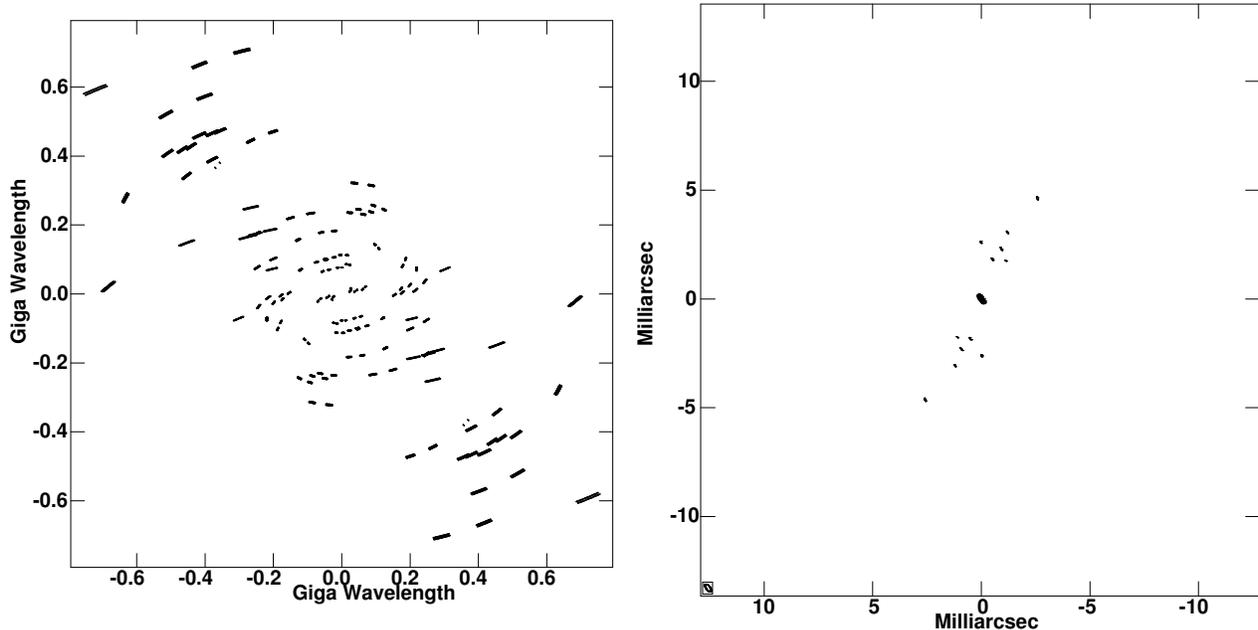


Figure 1: Test One.

**Left:** UV coverage for a single channel of the test data on a stellar SiO maser.

**Right:** Dirty beam, side-lobes above 30% shown.

A color-coded velocity image showing the regions of maser emission in the test data is shown in Figure 2. This image was derived using the autoWindow technique and was filtered to show only the “significant” emission in each channel image, significant being defined as pixels whose flux density exceeds  $5\sigma$  or 5% of the channel image peak intensity. This figure shows the region over which emission in any given spectral channel might be expected.

A comparison of a CLEAN constrained by the autoWindow technique with an unconstrained CLEAN using a full-field window is best done on a channel-by-channel basis. Such a comparison is done in Figure 3 for a single channel which shows CLEANed images obtained by the two techniques. The input data and the CLEANing parameters were identical except for the manner in which the CLEAN window was picked. The data were amplitude and phase self-calibrated with a solution interval of a scan (although the improvement in image quality obtained was marginal). There were approximately 15,000 6 second integration visibility points in the data set. Imaging used a single  $1536 \times 1536$  pixel field. The beam area in the 760 pixel radius “Full” CLEAN window corresponds to about 51,000 beam areas. There is a finite correlation between successive samples on a given baseline so it is not clear how many independent samples there are in the visibility data although clearly far fewer than the number of beam areas. CLEANing was allowed to proceed up to 2000 components but stopped if a minimum residual of 30 mJy/beam in the (inner) CLEAN window was reached.

Two features are immediately obvious from Figure 3; the “noise” is higher in the left-hand image ( $\sigma=33$  mJy/beam vs.  $\sigma=11$  mJy/beam) but there are “sources” appearing in the right-hand image that do not appear in the left-hand image, e.g. the component at (7, 6). The actual “noise” in the image can be estimated from a line-free channel. The RMS in one such channel is 8.1 mJy/beam in the Full window CLEAN and 13.2 mJy/beam in the autoWindow image. The Full window CLEAN removed 436 CLEAN components (gain 0.1) and the autoWindow CLEAN removed 16. The 40% reduction in the RMS in the image was the result of an unconstrained CLEAN modeling the noise. The RMS fluctuations in a quiet region of the image in Figure 3 are

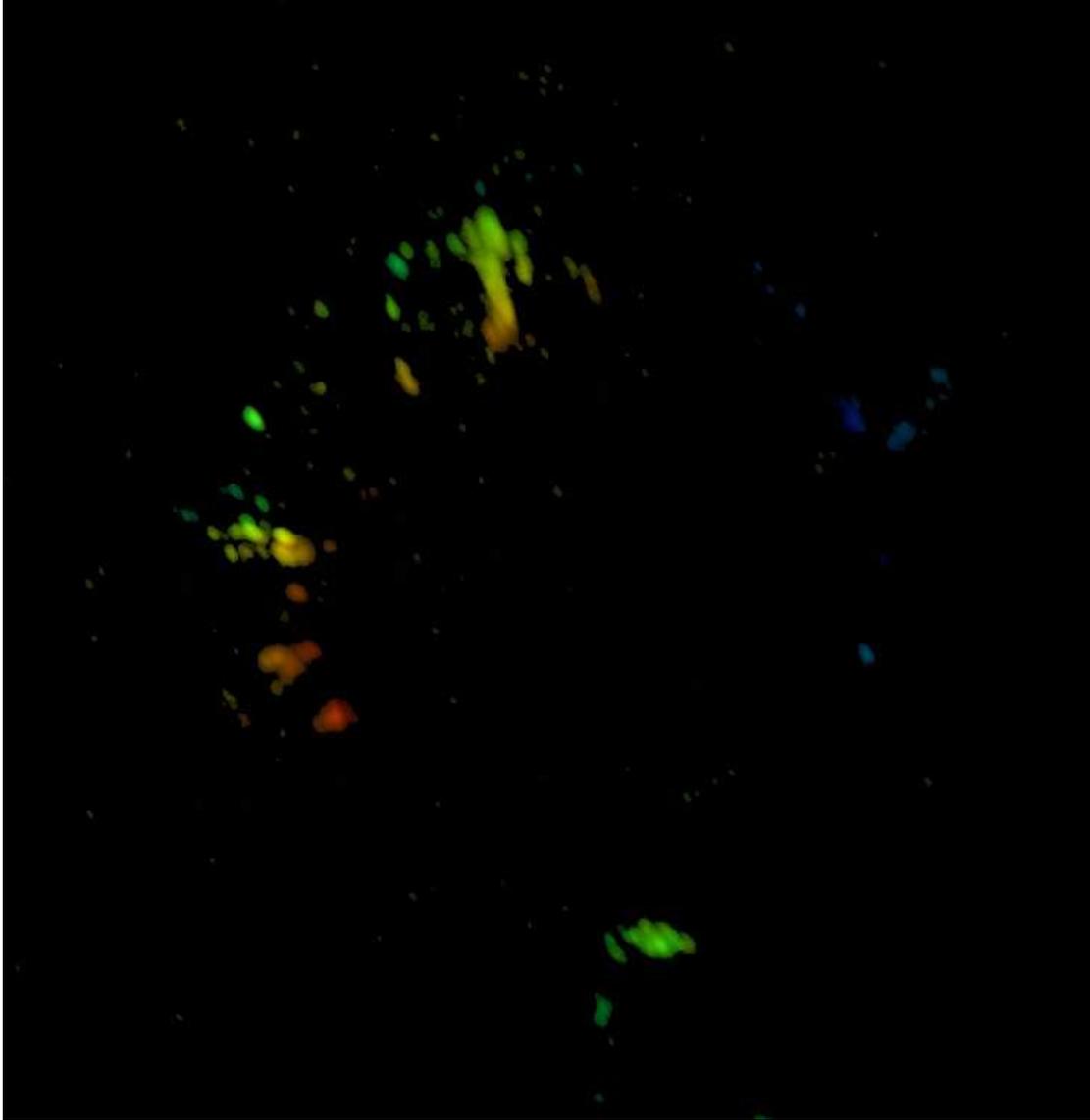


Figure 2: Test One. Color coded velocity/intensity plot of a stellar maser in the  $\nu=2$ ,  $J=1-0$  transition of SiO at 42.8 GHz. The region displayed is  $37 \times 38$  mas and red=7.0 km/s, blue=-7.2 km/s. The intensity proportional to square root of intensity. The restoring beam used is  $400 \times 200$   $\mu$ asec at position angle  $35^\circ$ .

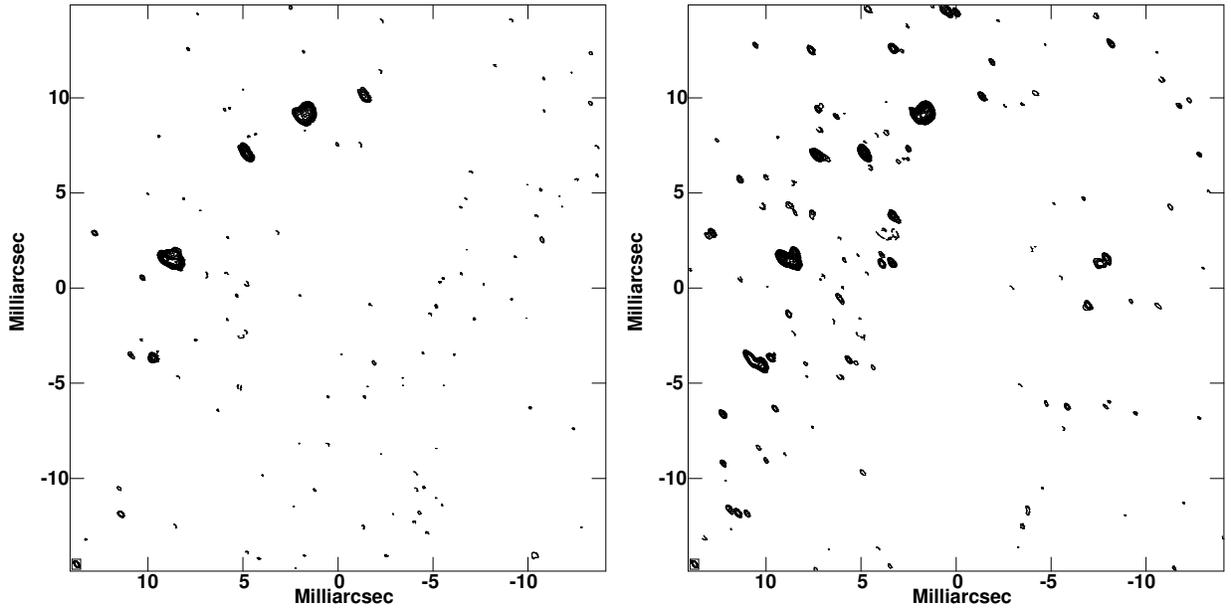


Figure 3: Test One.

**Left:** Channel image deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of  $\sqrt{2}$  times 75 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner.

**Right:** Channel Image deconvolved using CLEAN and a “window” covering the full field of view. Contour levels are powers of  $\sqrt{2}$  times 25 mJy/beam ( $\approx 2.5\sigma$ ).

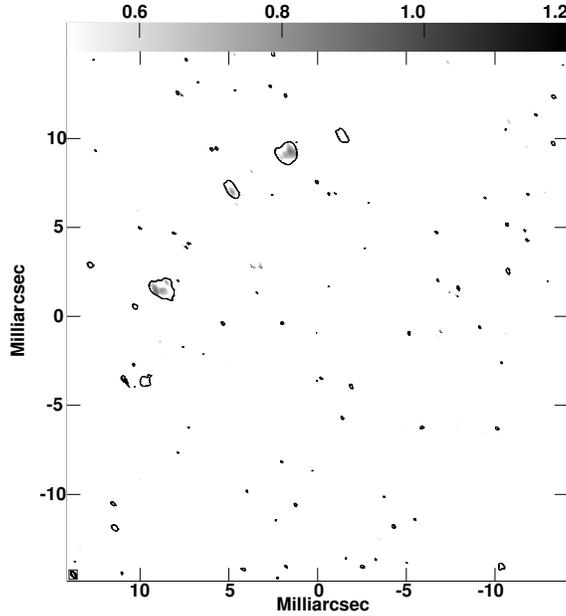


Figure 4: Test One. Ratio of the Full Window image to the autoWindow image clipped at 50 mJy/beam. Overlaid is the 100 mJy/beam contour of the autoWindow image. The ratio is given by the scale bar at the top. The images going into the ratio are clipped at 2.5 times the RMS noise; this causes the spurious features in Figure 3 **Right** not to appear.

11 mJy/beam for the Full window CLEAN and 33 mJy/beam for the autoWindow clean. Most of the “features” visible in the right-hand image in Figure 3 are well above the “noise” - the lowest contour is  $2.5 \sigma$  - but only the brightest few have counterparts in the left-hand image. Many of these weaker features are likely the side-lobes of the real sources.

The autoWindow CLEAN selected 1192 components (in 321 pixels) with 26.3 Jy total. The Full window CLEAN used the maximum of 2000 components (in 1462 pixels) but only removed 22.1 Jy of flux.

A detailed comparison of the two images in Figure 3 shows that the stronger components in common to both images are systematically fainter in the unconstrained (full windowed) CLEAN. A gray-scale of the ratio of the two images clipped at 50 mJy/beam with the 100 mJy/beam contour of the autoWindow image is shown in Figure 4. In this figure, the contours indicate the stronger sources in the autoWindow image and sources without such a contour are only in the Full window image. Empty contours have significantly stronger emission in the autoWindow than the full window image. The most notable feature of Figure 4 is that the flux densities of the stronger (i.e. real) components in the Full window image are about 80% of that in the autoWindow image. The peak brightness in the autoWindow image is 3.02 Jy/beam while the corresponding pixel in the Full Window image is 2.63 Jy/beam.

There are also several features in the autoWindow image that are significantly brighter than in the Full Window image. The feature at (-2, 10 mas) has a Full/autoWindow ratio of 0.23; the feature at (-9, -4) has a ratio of 0.32 (the extended feature seen in the Full Window image does not correspond to this feature). Neither of these features is near a major side-lobe of one of the brighter features (see Figure 1).

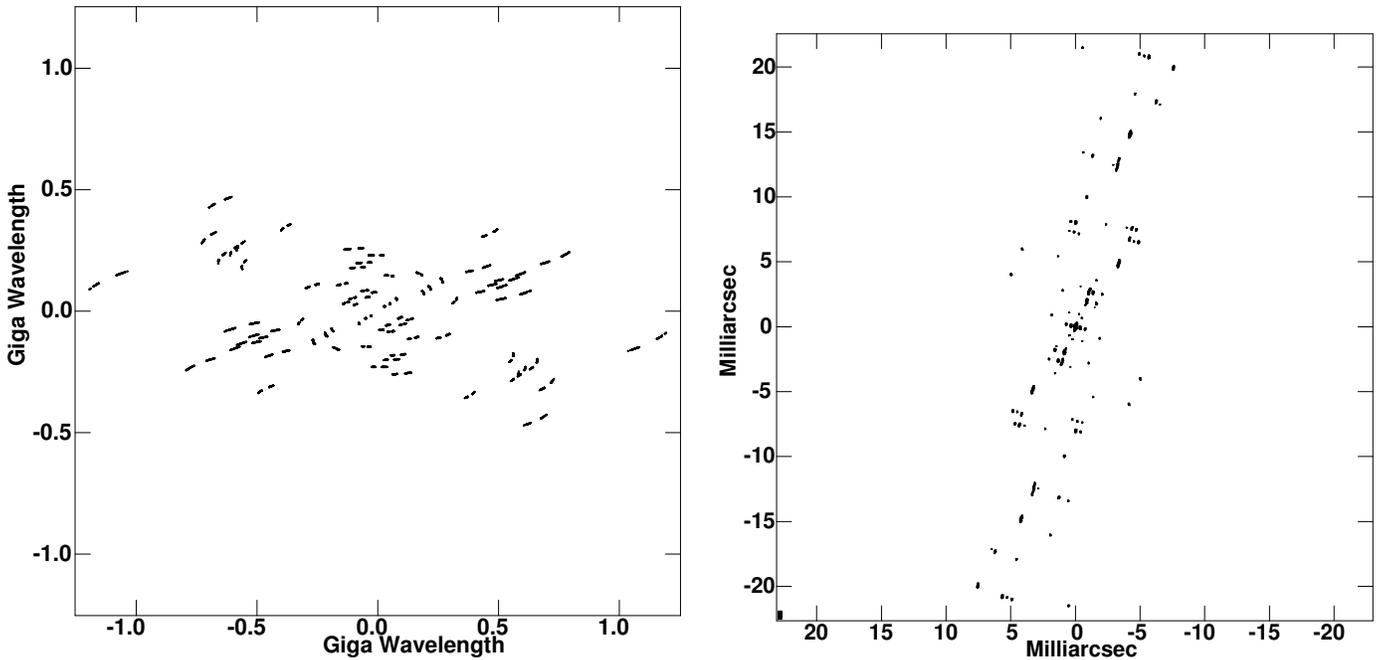


Figure 5: Test Two.

**Left:** UV coverage for a single channel of the test data on a stellar SiO maser.

**Right:** Dirty beam, side-lobes above 30% shown.

### 3.2 Test Two

In a second test, part of a more extensive data set is imaged and compared with the results from the full data set. The data is similar to Test One in that it is VLBA observations of a stellar SiO maser source; all antennas worked for at least part of the observations and the short, test section of data was three 20 min. scans over two hours during which all 10 antennas were working. The full data set contained 23,500 visibility measurements and the test section 12,700. As in Test One, the data for a single spectral channel was amplitude and phase self-calibrated using the full data set. The UV coverage for the test data and the derived dirty beam for a given channel are displayed in Figure 5.

A color-coded velocity image showing the regions of maser emission in the test data is shown in Figure 6. This image was derived using the autoWindow technique and was filtered as for Figure 2.

In this test, the image made from the full data set is closer to the true image than either of the images made from the shorter data set and can be used to help establish the reality of features. The full data set is used with the autoWindow technique to make the reference image for the following. CLEANing was allowed to proceed up to 2000 components or to a minimum of 50 mJy/beam in the (inner) CLEAN window. The single channel reference image and the uv coverage used in it are shown in Figure 7. The constrained and unconstrained CLEAN results are shown in Figure 8.

The results of this test are similar to that of Test One. The wide view of the field is given in Figure 8 and a closeup of a region with multiple maser spots is shown in Figure 9. The constrained CLEAN of the full data set CLEANed 38.9 Jy using 1089 components in 427 pixels and had an off-source RMS of 34.2 mJy/beam. The unconstrained (Full window) CLEAN on the short set of data CLEANed 17.1 Jy in 1505 components on 1206 pixels and had an off-source RMS of 12.4 mJy/beam. The constrained CLEAN of the short data segment CLEANed 34.6 Jy using 1158 components on 391 pixels and had an off-source RMS of 38.3 mJy/beam. The peak value in the reference image is 3.5 Jy and the corresponding values in the constrained and unconstrained

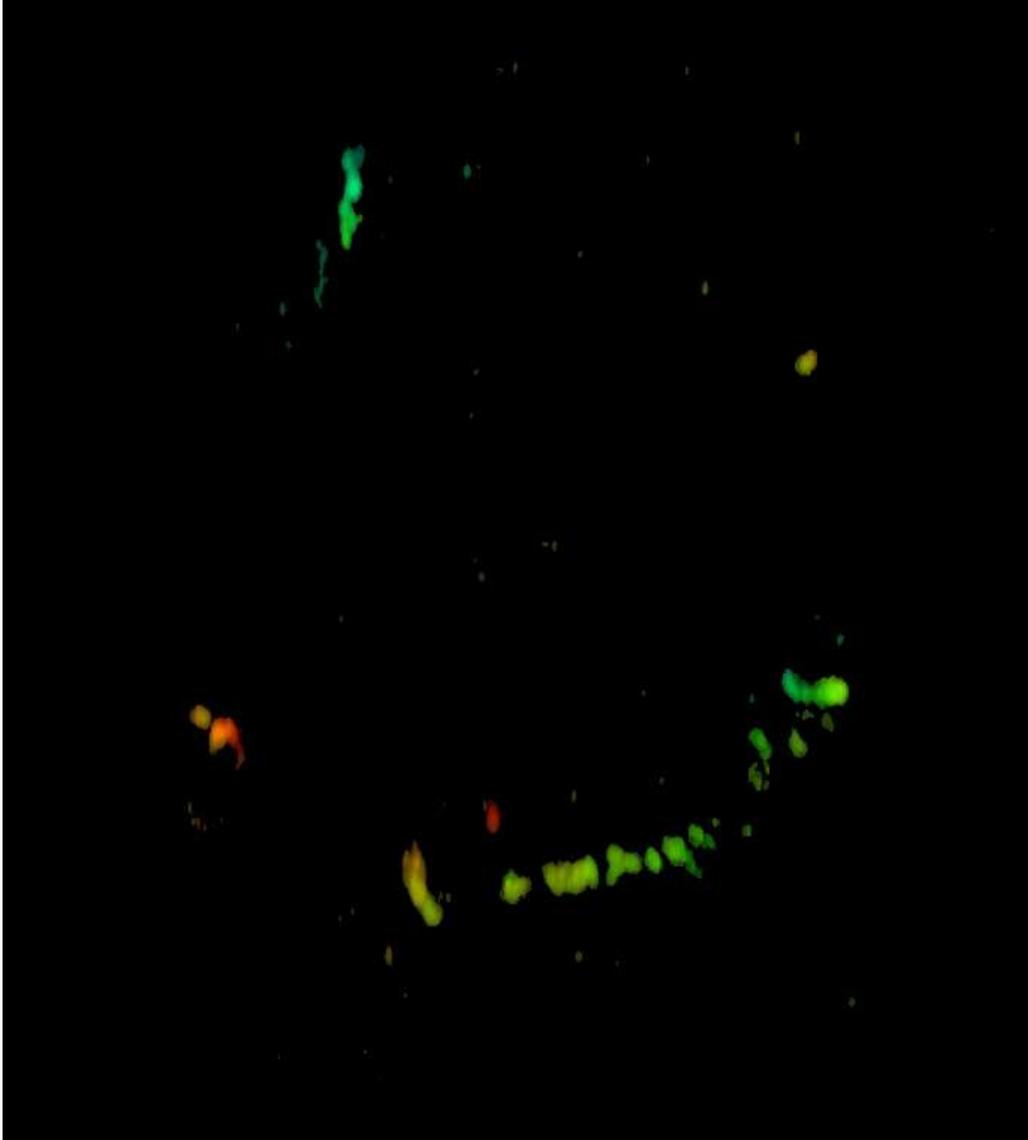


Figure 6: Test Two. Color coded velocity/intensity plot of a stellar maser in the  $\nu=2$ ,  $J=1-0$  transition of SiO at 42.8 GHz. The region displayed is  $34.5 \times 38$  mas and red=23.4 km/s, blue=13.7 km/s. The intensity proportional to square root of intensity. The restoring beam used is  $400 \times 170$   $\mu$ asec at position angle  $0^\circ$ .

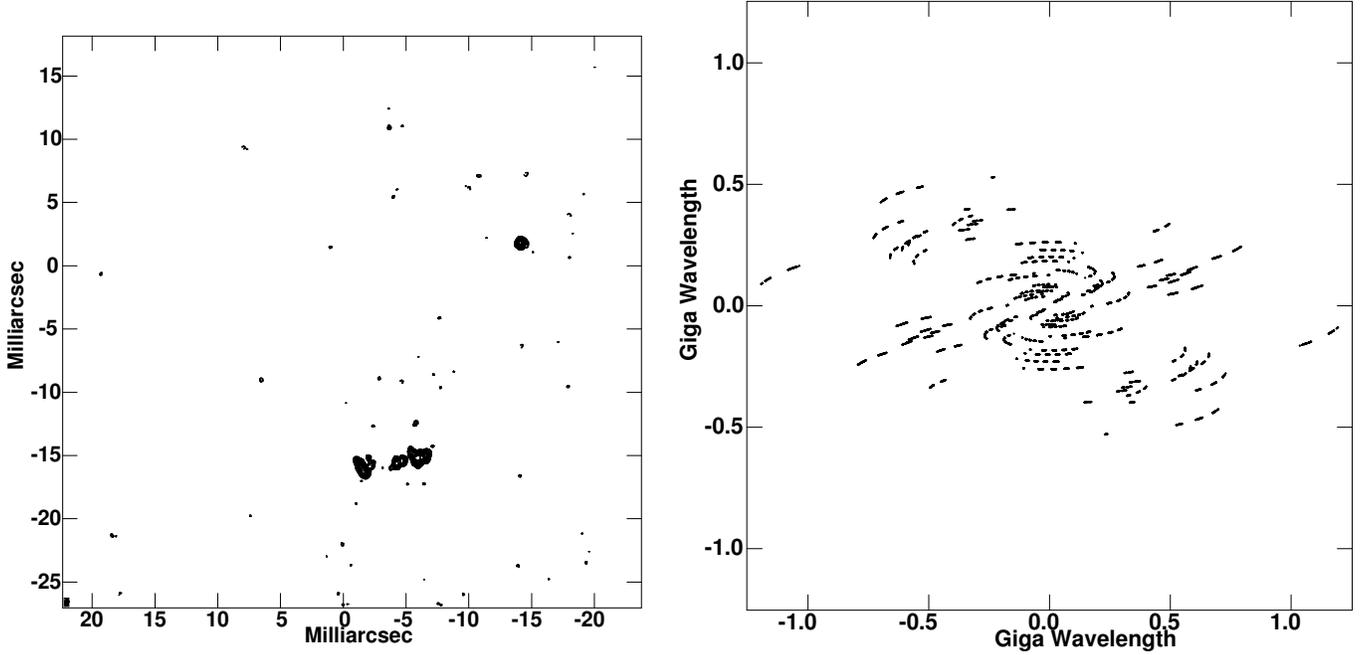


Figure 7: Test Two.

**Left:** Channel image of the full data set deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of  $\sqrt{2}$  times 85 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner.

**Right:** The UV coverage used in the image on Left.

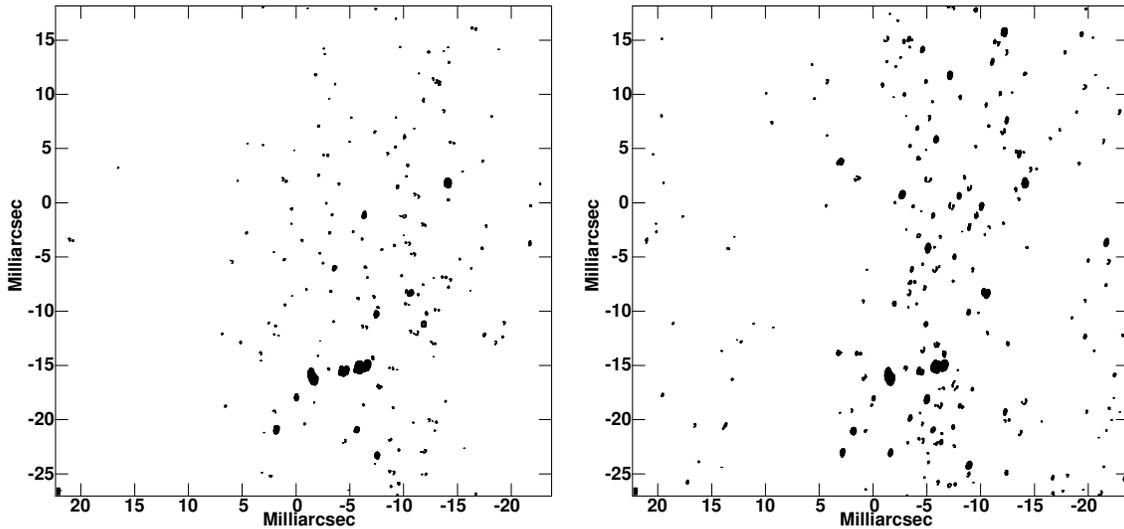


Figure 8: Test Two.

**Left:** Channel image from the short data segment deconvolved using CLEAN and windows selected by the autoWindow algorithm. Contour levels are powers of  $\sqrt{2}$  times 100 mJy/beam ( $\approx 2.5\sigma$ ). The resolution is indicated in the bottom left corner.

**Right:** Channel image from the short data segment deconvolved using CLEAN and a “window” covering the full field of view. Contour levels are powers of  $\sqrt{2}$  times 30 mJy/beam ( $\approx 2.5\sigma$ ).

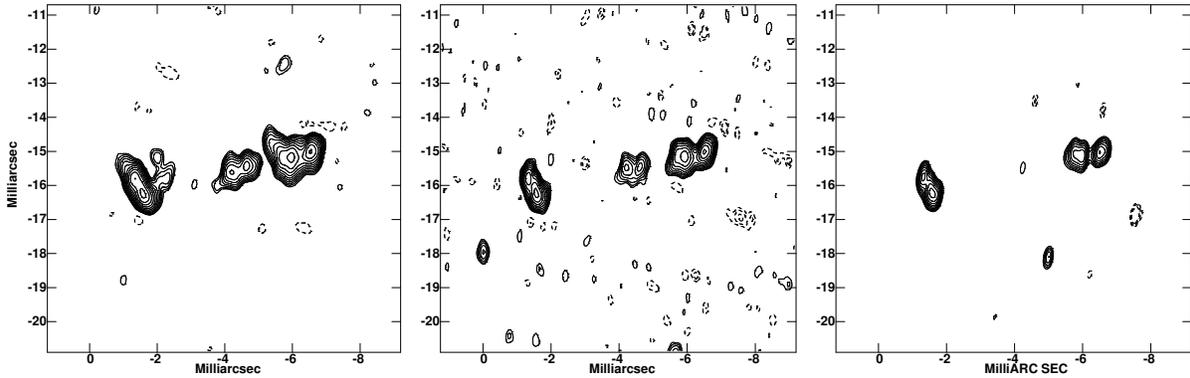


Figure 9: Test Two.

**Left:** Closeup of region of reference image. Contour levels are powers of  $\sqrt{2}$  times 100 mJy/beam. **Center:** Like **Left** but for short data segment deconvolved using CLEAN and windows selected by the autoWindow algorithm. **Right:** Like **Left** but for short data segment deconvolved using CLEAN and a “window” covering the full field of view.

CLEANs are 2.5 and 1.8 Jy, respectively.

The unconstrained CLEAN has lower RMS but more and stronger spurious sources and only recovered about half of the flux density in either of the constrained CLEANs.

In the images shown in Figure 9, the progression from left to right shows a decrease in the flux densities of the components in the reference image (presumed to be real). The ratios of the constrained and unconstrained CLEANs of the short data segment to the reference image are shown in Figure 10 with a closeup view of a region with a number of maser spots in Figure 11. In the Figure 11 **Right**, there is substantial variation in the ratio at the locations of components in the reference image with values between 10% and 50%. The ratio of the two constrained CLEANs seen in Figure 11 **Left** is somewhat better with values typically between 50% and 80% of those in the reference image. The constrained CLEAN recovered typically twice (or more) the flux density in real components as the unconstrained CLEAN.

### 3.3 CLEAN Bias

Wide area-shallow surveys are performed by spending relatively little observing time on each pointing and the resulting limited uv-coverage gives rise to relatively high sidelobes. The “CLEAN bias” generally seen in such surveys is thought to be the result of the deconvolution process putting flux (components) on sidelobes and noise bumps and effectively removing power from the real sources. The autoWindow technique should reduce this effect as it constrains CLEAN to only put components where there is plausibly real emission.

A test was performed on a dataset from such a survey by adding a set of 9 artificial sources with a range of flux densities to each pointing and then imaging the data to determine the recovered peak brightness. The models were added using Obit task UVSub. The dataset used consists of  $2 \times 50$  second snapshots in spectral line mode. The data with the models added were imaged both with and without the autoWindow option enabled using Obit task Imager and the value at the expected position of each model source determined from the image of each pointing. The dataset used had 220 pointing and the average bias values are shown in Figure 12 and 13. These figures show that the CLEAN bias was reduced by a factor between 4 and 5 by using the autoWindow technique. Figure 13 shows one of the signatures of the CLEAN bias, that the magnitude of the bias is nearly a constant multiple of the noise. The autoWindow algorithm increased the average RMS noise in the fields by 2.3%, from 381 to 389  $\mu$ Jy/beam.

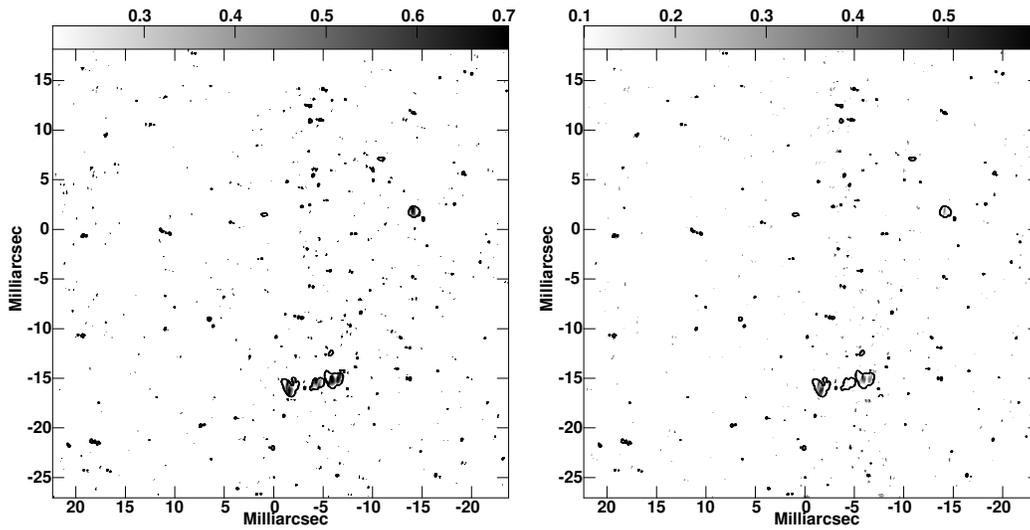


Figure 10: Test Two.

**Left:** Ratio of the constrained CLEAN image derived from the short data segment to the reference image. Overlaid is the 100 mJy/beam contour of the reference image. The ratio is given by the scale bar at the top.

**Right:** Like **Left** but the ratio of the short segment unconstrained CLEAN to the reference image. Note difference in gray scale.

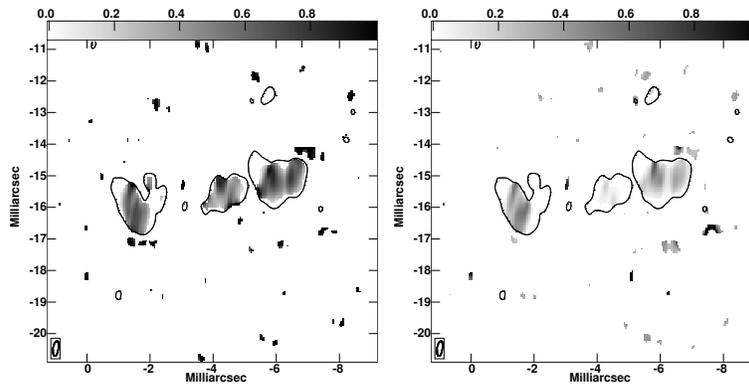


Figure 11: Test Two.

**Left:** Closeup of ratio of the constrained CLEAN image derived from the short data segment to the reference image. Overlaid is the 100 mJy/beam contour of the reference image. The ratio is given by the scale bar at the top.

**Right:** Like **Left** but the ratio of the short segment unconstrained CLEAN to the reference image.

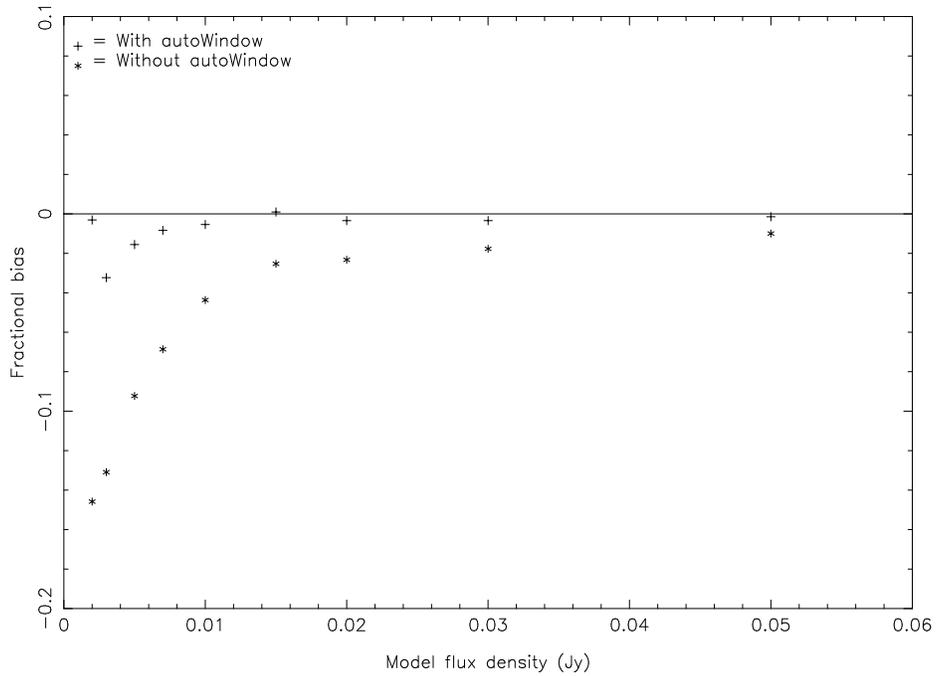


Figure 12: CLEAN bias test.

This plot shows the average bias of each of 9 artificial sources from 220 trials from a wide-shallow survey dataset expressed as a fraction of the model flux density. Stars represent results from the unconstrained CLEAN, plus signs the autoWindow CLEAN.

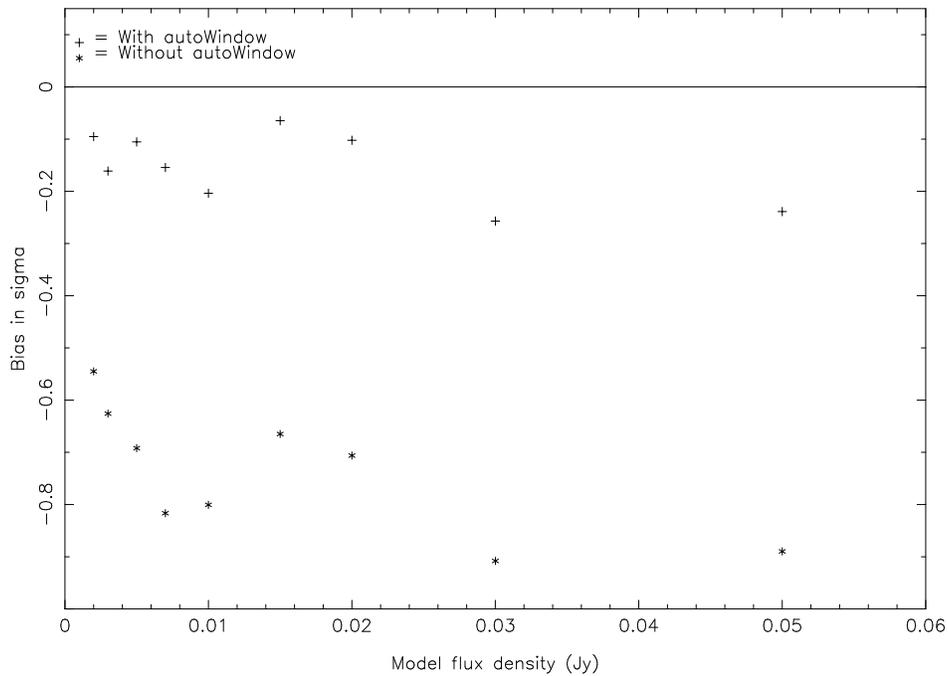


Figure 13: CLEAN bias test.

This plot shows the average bias of each of 9 artificial sources from 220 trials from a wide-shallow survey dataset expressed in units of the RMS noise determined from each field. Stars represent results from the unconstrained CLEAN, plus signs the autoWindow CLEAN.

## 4 Discussion

The CLEANs in two under-determined images (more independent beams than independent visibilities) examined display several serious pathologies: artificial reduction of “noise”, creation of spurious “sources” and a dramatic reduction of the flux density in real sources. The autoWindow technique was used to impose constraints on the CLEAN deconvolution which appear to reduce these pathologies. As expected, the autoWindow images shown in Figures 3, 8 and 9 have a higher off-source RMS fluctuation than do the full window images. However, the dynamic ranges of 70–80 obtained in the autoWindow images is plausible given the limited uv-coverage. On the other hand, the off-source RMS in the unconstrained CLEAN shown in Figure 3 is 83% of the estimate of the noise from a line-free channel, so clearly, the CLEANing has artificially reduced the level of fluctuation in the image. It should be noted that the amplitude self-calibration may have caused some changes to the flux density scale although the results with and without this self-calibration are not very different. The unconstrained (Full window) CLEAN has generated what appear to be spurious sources which may correspond to side-lobes of the stronger sources.

In the second test where a better reference image was available, both the CLEANs of the shorter data segment had peak emission in the correct locations. Spurious “sources” at a high significance in terms of peak-to-RMS are created but more so using the unconstrained CLEAN.

Finally, the flux densities of the sources from the unconstrained CLEAN appear to be systematically lower, sometimes dramatically so, than those from the more tightly constrained CLEAN. This is in agreement with the substantially larger flux density CLEANed by the constrained CLEAN. In the second test, both images from the shorter data set had substantially less flux density in the real components than the reference image but the unconstrained CLEAN had only about half of the flux density of the constrained CLEAN. Much of this missing emission appeared in the spurious sources although the unconstrained CLEAN had substantially less total flux density in its CLEAN models than the constrained CLEANs. Spurious sources created at side-lobes of the brighter components will also remove flux density from the brighter sources.

The results of the CLEAN bias tests are less dramatic as the initial problems are less severe. Even in this case, the use of the autoWindow technique significantly reduced the magnitude of the bias (by a factor of 4-5) but did not completely eliminate it.

The improvement yielded in image quality by the autoWindow technique comes from the addition of constraints to the deconvolution. The strongest of these is that the regions of the image in which CLEAN is allowed to put emission are tightly limited to those most likely to contain emission. The second constraint, for Stokes I, is positivity; CLEANing is only allowed in regions of positive brightness.

## 5 Conclusions

This memo has investigated several of the problems associated with under-determined imaging. These are cases in which the number of independent visibility measurements is comparable to or less than the number of independent beam areas in the region being imaged. In the under-determined images examined, there are serious errors which are capable of leading to incorrect interpretation of the results. The principal defects are the artificial reduction of the RMS fluctuation in the image, creation of spurious “sources”, and a dramatic reduction of the brightness of actual sources.

The autoWindow technique has been presented as a way to constrain otherwise poorly constrained deconvolutions and has been demonstrated to yield more accurate results than an unconstrained CLEAN. The method adds CLEAN boxes as the deconvolution proceeds covering areas which are likely to correspond to actual emission. This technique was compared to an unconstrained CLEAN in the deconvolution of two sets of snapshot VLBA data on stellar SiO masers. The unconstrained deconvolutions appear to create more spurious sources and reduce the RMS off-source

fluctuations below the true noise in the image. The unconstrained CLEANs produced images with substantially lower flux density than produced by the autoWindow method and with less of the flux density in true sources. Most of the CLEAN components in the unconstrained CLEAN appear to be modeling the noise rather than the sky. The autoWindow technique shows the potential for a substantial improvement over an unconstrained CLEAN in under-determined cases.

The limited testing of the autoWindow technique on survey style observations suggest that this technique very significantly reduces the effects of CLEAN bias; by a factor of 4-5 in the test presented. This technique should be generally useful in processing the data from such surveys and significantly reduce an important systematic effect.

## 6 Acknowledgments

I would like to thank Juan Uson, Eric Greisen, Michael Rupen and Ed Fomalont for stimulating discussions on this subject.

## 7 References

- Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, "The NRAO VLA Sky Survey", *Astron. J.*, **1115**, 1693–1716.
- Pearson, T. J., Sheperd, M. C., Taylor, G. B., and Myers, S. T., 1994, "Automatic Synthesis Imaging with Difmap", *BAAS*, **185**, 0808.
- Sheperd, M. C., 1997, "Difmap: an Interactive Program for Synthesis Imaging", in ASP Conf. Ser. 125: Astronomical Data Analysis Software and Systems VI, Hunt, G. and Payne, H. eds., 77.