Polarimetric imaging of large fields in radio astronomy

R.J. Sault[1], D.C.-J. Bock[2], and A.R. Duncan[3]

[1] Australia Telescope National Facility, CSIRO, P.O. Box 76, Epping, N.S.W. 1710, Australia
[2] Radio Astronomy Laboratory, University of California at Berkeley, 601 Campbell Hall, Berkeley, CA 94720, U.S.A.
[3] Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany

Abstract. We introduce and compare schemes to form wide-field polarimetric images from radio-interferometric mosaic and single-dish observations. We concentrate on two schemes based on maximum entropy principles. One scheme deconvolves the four Stokes parameters jointly (simultaneously), the other deconvolves each Stokes parameter separately. We find that the two schemes do not produce substantially different results. Issues of wide-field polarimetric purity are also considered.

Key words: methods: data analysis — techniques: interferometric

1. Introduction

In radio interferometry, "mosaicing" is the practice of imaging a field much larger than the antenna primary beam response by combining data from multiple different pointings. Mosaicing is fundamentally more than a simple "cut and paste" merging of images from different pointings: mosaicing can recover information on shorter spacings than that possible with single pointing observations. Mosaicing provides constraints on the short spacings, which allows more reliable imaging of large-scale structure. A detailed argument on the advantages of mosaicing is given by Cornwell (1988).

Mosaicing, however, is not usually sensitive to the shortest spacings. Consequently, mosaiced images are often complemented with single-dish observations to allow the Fourier plane to be fully sampled.

Although there are some examples of mosaicing polarimetric data (Wieringa et al. 1993; Dickey 1997; Beck et al. 1998; Duncan et al. 1998), these do not combine mosaic and single-dish data. This is despite many Galactic objects being both large and significantly polarized. Thus a combination of mosaiced and single-dish data is just as

Send offprint requests to: R.J. Sault

relevant in polarimetry as in total intensity. Additionally, the published cases do not appear to have processed the different Stokes parameters jointly and some of them do not process the individual pointings jointly.

In this note, we describe and compare two maximum entropy schemes for mosaicing and combining mosaic and single-dish data which we have used to image large-scale polarized emission. Both schemes jointly process the available pointings and single-dish data. These schemes differ in whether the four Stokes parameters are separately or jointly deconvolved. We also briefly compare these with a CLEAN-based scheme. For the data we present, we find no benefit in the joint deconvolution of the Stokes parameters over separate deconvolution, although we note some cases where this may not be the case. Issues of wide-field polarimetric purity are also briefly considered.

2. A separate approach

The mosaicing schemes that we have used follow the approach of Sault et al. (1996) in which the task of forming a final image is broken into two steps. First, dirty linearly mosaiced images of the relevant Stokes parameters are formed from the interferometric visibility data. Second, this dirty mosaiced image is 'deconvolved' to form the final images. The simplest approach to deconvolving the different Stokes parameters is to deconvolve them separately (i.e. independently of each other). In this case, the use of the maximum entropy method is well established in mosaic deconvolution – see Cornwell (1988) and Sault et al. (1996) for descriptions. In particular, for a particular Stokes parameter, the maximum entropy process finds that solution image S which maximizes

$$J = H - \alpha \chi^2. \tag{1}$$

Here H is the entropy measure of the solution image, χ^2 gives a data constraint, and α is a Lagrange multiplier. Following Sault et al. (1996), we use

$$\chi^2 = \sum_i^{N_{\mathrm{pix}}} (D[S_i] - S_{\mathrm{D},i})^2 / \sigma_i^2 - N_{\mathrm{pix}} \tag{2}$$

where S_i is the value of the ith pixel of the solution (here we use S to denote any Stokes parameter), N_{pix} is the number of pixels in the mosaiced image, S_{D} is the dirty linearly mosaiced image, and D is the linear operator which converts an estimate of S into a dirty mosaiced image (the operator D involves first, for each pointing, applying a primary beam and convolving with a dirty beam, and second performing a linear mosaicing operation).

This approach is readily augmented to include single-dish data by adding a second χ^2 constraint (and second Lagrange multiplier). This constraint is similarly defined to Eq. (2), measuring squared difference between the single-dish image and the smoothed solution image.

The entropy measure most commonly advocated in deconvolution problems is

$$H = -\sum_i^{N_{\mathrm{pix}}} S_i \log(S_i/M_i e), \tag{3}$$

where M_i is the so-called "default image" (in the absence of a data constraint, the maximum entropy solution is $S_i = M_i$). This measure is clearly inappropriate for polarized quantities as the pixel values, S_i, need to be positive. However, as Narayan & Nityananda (1986) pragmatically note, there are many functional forms which can produce good "entropy measures" even if these have no basis in information theory. We have used a measure suggested by T.J. Cornwell (private communication), which has the form

$$H = -\sum_i^{N_{\mathrm{pix}}} \log\left(\cosh\left(\frac{S_i}{M_i}\right)\right), \tag{4}$$

and which he has called the maximum emptiness criterion. In the absence of data constraints, the corresponding solution image is $S_i = 0$. The rationale for this entropy measure is that, for $|S_i| \gg |M_i|$,

$$-\log\left(\cosh\left(\frac{S_i}{M_i}\right)\right) \approx -\left|\frac{S_i}{M_i}\right| + \log 2, \tag{5}$$

and so by choosing M_i less than, or comparable to the noise level, maximizing H is approximately equivalent to minimizing the L_1-norm of the solution image. As is well known (e.g. Press et al. 1986), minimum L_1-norm solutions have the property of not giving undue weight to "outliers". In our context, the measure will prefer a solution image which is mostly empty, but it will not give undue weight to attempting to eliminate some regions of significant emission (i.e. outliers from emptiness).

3. A joint approach

As the maximum entropy procedure is a non-linear one, the separate approach is not optimum: because of the non-linearity, it is desirable to include all possible data during the deconvolution. That is, it is desirable to perform a joint deconvolution of all the Stokes parameters. In doing so, we can make sure of the positivity of the total intensity and that the fractional polarized intensity is less than 1. Furthermore, it is interesting to note that polarimetry can also provide constraints on large-scale structure of the total intensity: when the total intensity is very broad and smooth, and the polarized emission is varying on a much smaller scale size, it is possible for an interferometer to resolve out the total intensity and not the polarized emission. The result is instances where the emission is *apparently* more than 100% polarized. Such extreme cases certainly exist – see Wieringa et al. (1993) for excellent examples. In these cases, a joint algorithm will constrain the total intensity to be consistent with the polarized emission – i.e. more of the large-scale total intensity structure will be recovered.

An undesirable aspect of a joint algorithm is that unmodelled errors in one image can be propagated into the other images. For example, it is common for thermal noise to be the limit to polarization images, whereas larger unmodelled systematic errors are the limit in total intensity. In this case, joint deconvolution can allow the larger errors in total intensity to couple through to the polarized images.

A measure of the joint entropy of a collection of Stokes images was first formulated by Ponsonby (1973), and elaborated by Gull & Skilling (1984) and Narayan & Nityananda (1986). Despite being well described for many years, it appears that the only previous application of this measure is by Holdaway & Wardle (1990) in analysing VLBI data. Following them, we use the entropy measure

$$H = -\sum_i I_i \left(\log\left(\frac{2I_i}{M_i e}\right) + \frac{1+p_i}{2} \log\left(\frac{1+p_i}{2}\right) \right. $$
$$\left. + \frac{1-p_i}{2} \log\left(\frac{1-p_i}{2}\right) \right), \tag{6}$$

where I_i and p_i are the total intensity and fractional polarization of the ith pixel respectively, and M_i is the default image (note here we use a different scaling to Holdaway & Wardle). In this entropy measure, the two terms containing the fractional polarization, p, will always contribute to reducing the total entropy, and the minimum reduction is achieved when p is zero. That is, the entropy measure tends to minimize the fractional polarization. In particular, in the absence of data constraints, the maximum entropy solution is $I_i = M_i$ and $p_i = 0$, and in the absence of polarization data, this measure reduces to the total intensity case, Eq. (3). This entropy measure naturally ensures that the total intensity must be positive and that the fractional polarized intensity must be less than 1.

Again following Holdaway & Wardle, we have formulated the deconvolution step as maximizing the objective

function

$$J = H - \alpha\chi_I^2 - \beta\chi_P^2. \tag{7}$$

Here χ_I^2 and χ_P^2 are two data constraints, with Lagrange multiplies α and β, corresponding to one constraint each for total and polarized intensity mosaiced images. These are defined similarly to Eq. (2), except that χ_P^2 sums over all the polarized images. Single-dish χ^2 constraints can be added in the same way.

4. Implementations

The details of the maximum entropy solver that we have used is derived from the algorithm of Cornwell & Evans (1985) with modifications by Sault (1990). It is based on a modified Newton-Raphson iterative scheme: the algorithm proceeds by evaluating first- and second-order derivatives of the objective function, and, from these, deducing changes to make to the images and the Lagrange multipliers. For reference, we give the derivatives of the entropy measure for the joint approach in an appendix (this corrects some apparent errors in Holdaway & Wardle 1990). In computing these changes, the matrix of second derivatives (the Hessian matrix) needs to be inverted. The size of this matrix is $(N_{pix} \cdot N_{pol} + N_{con}) \times (N_{pix} \cdot N_{pol} + N_{con})$, where N_{pix} is the number of pixels per image, N_{pol} is the number of polarizations (i.e. 1 for separate, and up to 4 for joint deconvolution) and N_{con} is the number of Lagrange multipliers (i.e. 1, 2 or 3). To make this inversion feasible, we follow Cornwell & Evans and approximate the point-spread function in the Hessian as a delta function. Additionally, in the joint deconvolution, we ignore the mixed second-order derivatives of H (i.e. $\partial^2 H/\partial I \, \partial Q$, etc.). As with Sault (1990), this makes the Hessian a sufficiently simple form that inversion is straightforward and the changes can be determined.

5. Off-axis polarimetric purity

The off-axis polarimetric response of an antenna is an important consideration in polarimetric mosaicing. The main concern is the leakage of total intensity into the polarized intensities as a function of position in the primary beam. Understanding off-axis response is as important in polarimetric mosaicing as understanding the total intensity primary beam response is for traditional mosaicing. Appropriate design of the antenna system can produce good off-axis polarimetric purity, and, indeed, this is the case with the data presented below. Additionally, the process of mosaicing will tend to average down the impure response, as a pixel is the result of a number of pointings, and so the off-axis response is averaged over a number of different positions in the primary beam. Furthermore, if there is significant parallactic angle rotation during an observation (e.g. long observation with an alt-az mount), this too will tend to average down the impure response.

However, off-axis impurity can still be a significant consideration in many instances. If so, software techniques can be used to reduce it – see Cotton (1994) and Sault & Ehle (1996b) for some approaches.

6. An example

As an example of these approaches, we present both interferometric and single-dish polarimetric observations of Vela-X, which is the plerionic component of the nearby Vela supernova remnant – see Bock et al. (1998) for a description. For the interferometric observations, we have used the Australia Telescope Compact Array (ATCA). We have imaged a region of approximately $2° \times 3°$, using 35 pointings at a 20 cm wavelength. The pointings were on a hexagonal grid, with a distance between pointings of $11'.4$ (the hexagonal grid Nyquist spacing is $19'.6$ – the diameter of the ATCA antennas is 22 m). A fairly complete Fourier plane $(u - v)$ coverage was observed from 30 m to 750 m spacings. The data were collected using the "pulsar bin mode" of the ATCA's correlator, and images were made using de-dispersed data from the "off" portion of the Vela pulsar's cycle. We excluded the pulsar emission as its extreme variability would have limited the dynamic range of our images. Normal polarimetric calibration (i.e. at the field centre) was done on the visibility data. The single-dish observations were made with the 64 m Parkes radio telescope as part of a larger survey. The region was covered twice, once scanning in Galactic longitude and once in Galactic latitude. Apart from the observing frequency, the observing mode was similar to that used by Duncan et al. (1995, 1997). In particular, the polarimetric calibration was done in a similar way. Given the single-dish observing technique and given Parkes observations of some bright, unpolarized sources, we estimate that the leakage of total intensity into the linear polarizations is less than 0.5%.

Note that interferometric and single-dish observations both had good sensitivity to the spacing between about 20 and 40 m. Using this overlap annulus in the Fourier plane, we verified that the flux calibrations of the interferometric and single-dish data were consistent.

We have tried a number of ways to deconvolve these data. First, we have separately deconvolved the interferometric Stokes images using a mosaiced Steer CLEAN algorithm (Steer et al. 1984; Sault et al. 1996). Second we have used a maximum entropy algorithm to deconvolve the Stokes images separately, using either the normal entropy measure for total intensity or the maximum emptiness measure for the polarized intensities. These maximum entropy deconvolutions were done both with and without a single-dish data χ^2 constraint. Finally we have performed a joint deconvolution of the four Stokes images, both with
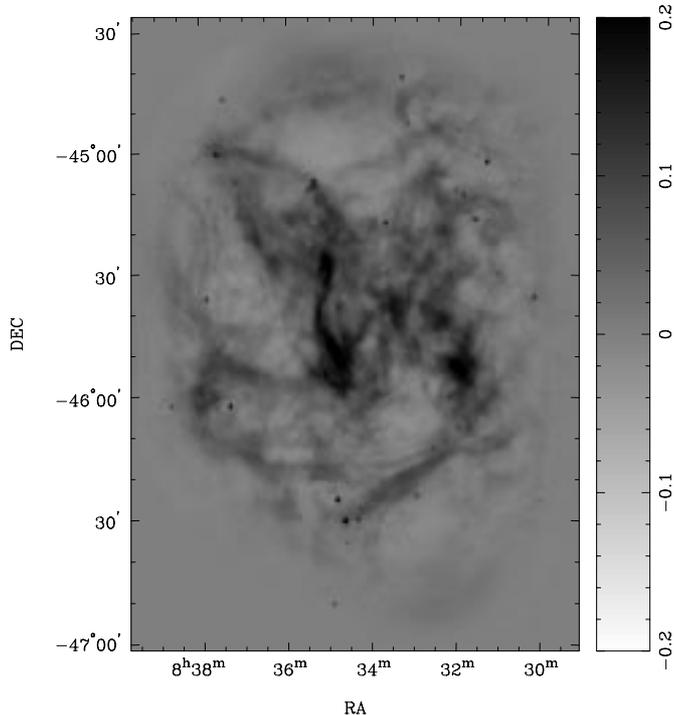
Fig. 1. Total intensity image of a portion of the Vela-X region, resulting from the joint deconvolution with single-dish constraint

and without a total-intensity single-dish χ^2 constraint. We have not done a joint deconvolution with single-dish polarized-intensity data constraints. Flat default and scale images were used in the maximum entropy deconvolutions. All the images that we present are after the solution images have been convolved with a Gaussian beam and have had residuals folded back in.

In total intensity, the only significant differences between the various deconvolved images were in the structures larger than that well-sampled by the interferometric observations. The Steer deconvolution grossly underestimated the large-scale structure, whereas the maximum entropy deconvolution without single-dish constraint overestimated it (this overestimate, however, is as much to do with the default image as the maximum entropy algorithm itself). Of course, the deconvolution with single-dish constraint, by its very nature, will produce the correct large-scale structure. Apart from the large-scale structure, the images were very similar, except that the Steer CLEAN image showed faint features that were obviously artifacts. There was no significant difference between the corresponding joint and separate deconvolutions of total intensity. Figure 1 gives the total intensity image resulting from the joint deconvolution including the single-dish data.

In the linearly polarized images, we find the differences between the various deconvolutions to be comparatively minor. In terms of features that were plainly artifacts, the Steer CLEAN images showed the most, whereas the separate approach showed the least. Including

the total intensity single-dish constraint in the joint deconvolution had negligible effect on the polarized intensity images. Similarly, including the single-dish polarized-intensity data had little effect on the result. This indicates that there is little structure larger than that which we sample with the interferometry. Figure 2 show the joint and separate deconvolutions of Q and U images, at a greyscale saturation of twice that in Fig. 1. The maximum fractional linear polarization in some regions is about 60%. Such a large fractional polarization is not inconsistent with previous Vela observations.

In the processing we have assumed that the ATCA has no off-axis polarimetric impurity, and that the primary beam response was circularly symmetric (note that the ATCA feed design at 20 cm means that the off-axis polarimetric response is substantially better than the 13 cm response investigated by Sault & Ehle 1996b). To justify these assumptions, we have measured the ATCA response of an unpolarized point source at many points in the primary beam. These measurements showed a small level of leakage of total intensity into linear polarization, but were unable to detect leakage into circular polarization. Using the approach of Sault & Ehle, we made a model of the off-axis response and then used this model to simulate our Vela observations. The simulations show that in the central region of the mosaiced image the expected leakage of total intensity into linear polarization is $\approx 0.05\%$, with this degrading to a maximum of $\approx 0.5\%$ near the edges of the image. This degradation near the edges is caused by there being less averaging from different pointings. This analysis also shows that the expected error in total intensity, caused by the deviation of the primary beam from circular symmetry, was 0.03% in the centre of the image, degrading to 0.3% towards the edges.

Some degree of confirmation that the off-axis polarimetric impurity is insignificant is that the Stokes-V image reached thermal noise (integrating the Stokes-V image over large regions did reveal some structure above the noise, but this was consistent with the errors in our polarimetric calibration). We also note that *all* the point sources in the field appear unpolarized and that there are no apparent artifacts at the locations of these point sources in the polarized images. It is difficult to imagine such a result if there was significant polarimetric impurity. Our interpretation of this is that the polarimetric impurity is negligible and that the point sources are background ones which are being Faraday depolarized by Vela and the Galaxy.

7. Conclusions

We have presented some approaches to forming wide-field polarimetric images. One approach does the deconvolutions of the Stokes images separately, and the other does a joint deconvolution of all Stokes images. In our example, we found that including polarized-intensity single-dish data added little to the deconvolution process
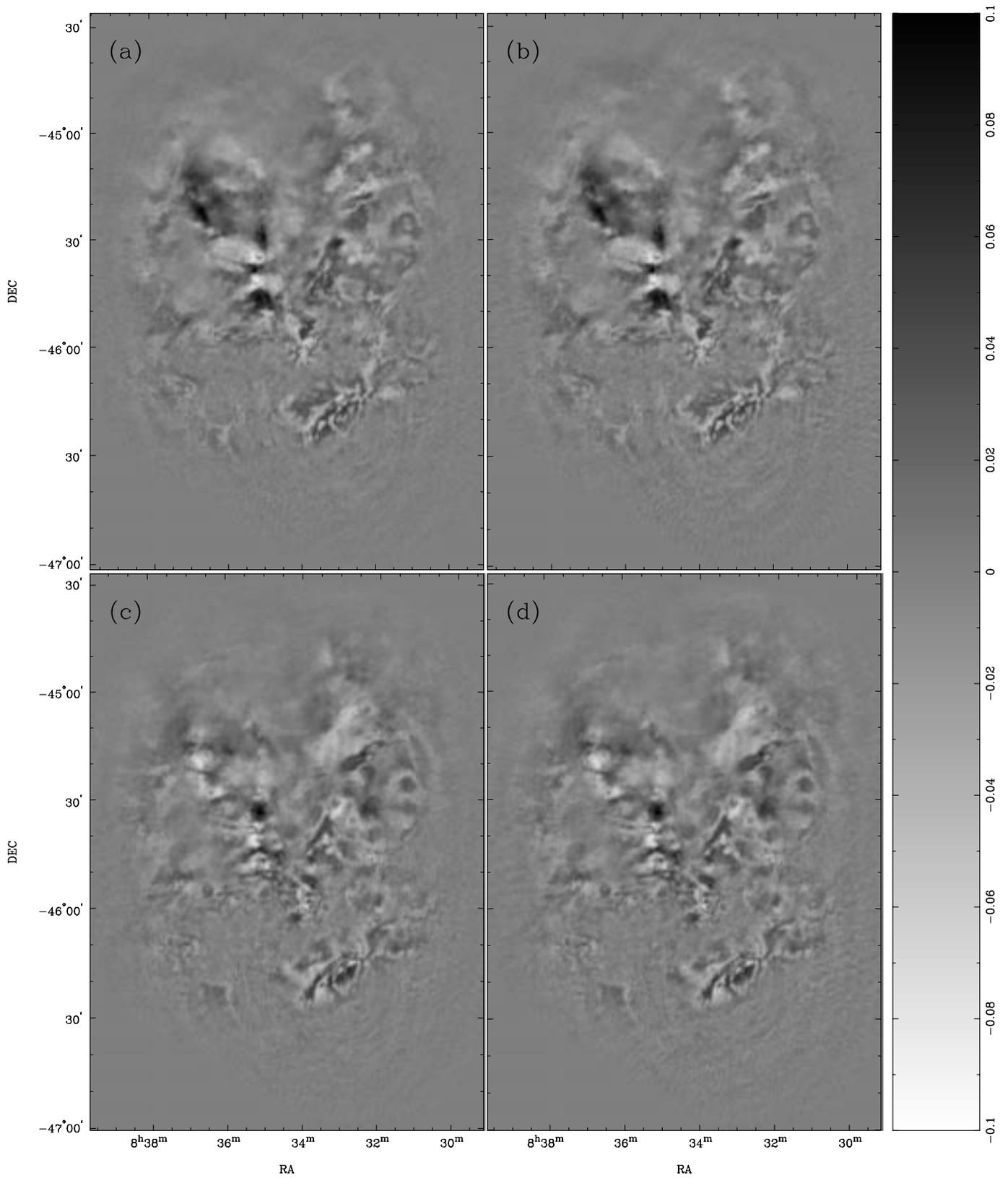
Fig. 2. Deconvolved Q and U images of a portion of the Vela-X region: **a)** Q using separate deconvolution (maximum emptiness criterion); **b)** Q using joint deconvolution; **c)** U using separate deconvolution (maximum emptiness criterion); **d)** U using joint deconvolution

over the interferometer data. Also we found that the joint deconvolution approach produced images which were mildly inferior to the separate deconvolution of the Stokes images. However, the relative merits of the joint versus separate approaches will depend very much on the data. The separate approach avoids error propagation from total intensity to the polarized intensity images, whereas the joint approach would be preferable when it can provide extra information in the deconvolution process. Certainly in instances where the Fourier-plane coverage differs between the Stokes images, a joint deconvolution would be advantageous. Indeed, where Stokes visibilities could not be formed (because instantaneously fewer than the four possible polarization cross-correlations were measured), a joint deconvolution is the only possible approach. In our example, we note that the Fourier-plane coverage of the four Stokes parameters was identical, and essentially complete between the shortest and longest spacing. With good and identical Fourier plane coverage, a joint approach is probably less advantageous.

It is interesting to consider the importance of the data constraints implicit in the joint deconvolution: that the total intensity must be positive and that the fractional polarization must be less than 100%. As we noted above, in the runs without the single-dish constraint, the maximum entropy deconvolutions substantially overestimated the total intensity in the spacings that were shorter than those well constrained by the interferometer data. That is, although the deconvolutions were plausible (i.e. positive valued), they were no more "correct" than the Steer CLEAN algorithm. While the CLEAN algorithm left a negative bowl, the maximum entropy results contained a "positive bowl". This should not be surprising – positivity is a fairly weak constraint in that it provides only a lower bound. Similarly, the constraint that the fractional polarization must be less than 100% is also weak, as it too only provides a bound. Even though the fractional polarized emission was comparatively high at 60%, the bound constraint was not a particularly strong one.

Appendix A: Entropy derivatives

As the maximization process is achieved using a Newton-Raphson iteration, we must evaluate the first- and second-order derivatives of the objective function with respect to pixel values and Lagrange multipliers. For the joint entropy measure,

$$\frac{\partial H}{\partial I} = -\frac{1}{2}\log\left(\frac{I^2}{M^2}(1-p^2)\right), \tag{A1}$$

$$\frac{\partial^2 H}{\partial I^2} = -\frac{1}{I}\frac{1}{1-p^2}, \tag{A2}$$

$$\frac{\partial H}{\partial Q} = -\frac{Q}{I}T, \tag{A3}$$

$$\frac{\partial^2 H}{\partial Q^2} = -\frac{1}{I}\left(T + \frac{Q^2}{I^2}\frac{1}{p^2}\left(\frac{1}{1-p^2} - T\right)\right), \tag{A4}$$

where

$$T = \frac{1}{2p}\log\left(\frac{1+p}{1-p}\right). \tag{A5}$$

Derivatives in U and V are similar. As p approaches 0, we use the following limits for the polarized derivatives:

$$\frac{\partial H}{\partial Q} \approx -\frac{Q}{I}\left(1 + \frac{1}{3}p^2\right), \tag{A6}$$

$$\frac{\partial^2 H}{\partial Q^2} \approx -\frac{1}{I}\left(\left(1 + \frac{1}{3}p^2\right) + \frac{Q^2}{I^2}\left(\frac{1}{3} + \frac{2}{5}p^2\right)\right). \tag{A7}$$

References

Beck R., Berkhuijsen E.M., Hoernes P., 1998, A&AS 129, 329
Bock D.C.-J., Turtle A.J., Green A.J., 1998, AJ 116, 1886
Cotton W.D., 1994, AIPS Memo No. 86, National Radio Astronomy Observatory
Cornwell T.J., 1988, A&A 202, 316
Cornwell T.J., Evans K.F., 1985, A&A 143, 77
Dickey J., 1997, ApJ 488, 258
Duncan A.R., Stewart R.T., Haynes R.F., Jones K.L., 1995, MNRAS 277, 36
Duncan A.R., Haynes R.F., Jones K.L., Stewart R.T., 1997, MNRAS 291, 279
Duncan A.R., Haynes R.F., Reich W., Reich P., Gray A.D., 1998, MNRAS 299, 942
Gull S.F., Skilling J., 1984, in: Indirect Imaging, Roberts J.A. (ed.). Cambridge University Press, p. 267
Holdaway M.A., Wardle J.F.C., 1990, in: Proceedings of the Meeting on digital image synthesis and inverse optics, Proc. SPIE 1351, p. 714
Narayan R., Nityananda R., 1986, ARA&A 24, 127
Ponsonby J.E.B., 1973, MNRAS 163, 359
Press W.H., Flannery B.P., Teukolsky S.A, Vetterling W.T., 1986, Numerical Recipes. Cambridge Univ. Press, Cambridge
Sault R.J., 1990, ApJ 354, L61
Sault R.J., Ehle M., 1996, ATNF Technical Document Series No. 39.3/088, Australia Telescope National Facility
Sault R.J., Staveley-Smith L., Brouw W.N., 1996, A&AS 120, 375
Steer D.G., Dewdney P.E., Ito M.R., 1984, A&A 137, 159
Wieringa W.H., de Bruyn A.G., Jansen D., Brouw W.N., Katgert P., 1993, A&A 268, 215