Effects of Baseline Dependent Time Averaging of UV Data

W. D. Cotton, June 8, 2009

***Abstract*—One of the major computational problems associated with the next generation of radio astronomical interferometric arrays is the extremely large data volumes. One of contributors to these large data volumes is the need to sample data fast enough in time that there is minimal smearing of the visibilities as the UV tracks move through the UV plane. The short time scales are driven by the longest baselines in the data-set whereas the UV coverage of most arrays is highly centrally condensed. This memo explores some of the imaging effects of time averaging the data in a baseline dependent fashion to reduce the bulk of the data. Compression ratios and the level of artifacts were found to be relatively independent of the VLA configuration (with proper temporal sampling) but to depend on the undistorted field of view desired. Even with a field of view to the 1% power point of the beam, a compression factor of 5–6 only reduces the dynamic range limit near a strong source to of order $10^{5.3}$ and a reduction in the peak in the image of ≈0.3%. Upper limits on the averaging time short enough to allow self calibration (15 sec.) do not seriously reduce the compression ratios for VLA A and B configurations.**

***Index Terms*—interferometry**

I. INTRODUCTION

IN order to reach the potential sensitivity limits of the next generation of radio interferometers such as EVLA and MeerKAT, the full primary beam of the antenna elements may need to be imaged in order to remove the side-lobes of the sources there. Rapid temporal sampling of the visibility data is necessary to avoid smearing the response to sources far from the pointing center on the longest baselines as the UV location sampled moves through the UV plane. This time smearing produces artifacts which are not even approximately convolutional so will not deconvolve. The need for rapid sampling is one of several effects contributing to the large data volumes these instruments will produce.

Most interferometer configurations are relatively centrally condensed meaning there are many more short baselines than long ones. The sampling cadence needed to avoid smearing on short baselines is much slower that on long baselines. This memo explores the possibilities of baseline dependent averaging where the averaging time is chosen so as not to reduce the visibility amplitude by more than a fixed amount. Test data-sets are simulated, averaged and imaged in order to evaluate the imaging effects of the averaging. Further tests were made to determine the effects of the upper limit on the averaging time on the achieved reduction in data volume. This evaluation was implemented in the Obit [1] [1] package.

National Radio Astronomy Observatory, 520 Edgemont Rd., Charlottesville, VA, 22903 USA email: bcotton@nrao.edu

[1]http://www.cv.nrao.edu/∼bcotton/Obit.html)

II. BASELINE DEPENDENT TEMPORAL AVERAGING

A simple criterion for the maximum distance in the UV plane over which data may be averaged (Δuv_{max} in wavelengths) and not reduce the amplitude of the response to any source within a field of view FOV (radians) by more than a factor of max_fact is given by:

$$\Delta uv_{max} = \frac{Sinc^{-1}(\frac{1.0}{max\_fact})}{FOV} \quad 1)$$

where $Sinc(x) = sin(\pi x)/\pi x$. max_fact is the maximum allowed reduction of a visibility amplitude due to averaging over the uv plane. Thus, for a unit flux density point source, the minimum visibility resulting from averaging would be $1/max\_fact$. This criterion is applied in AIPS task UBAVG (max_fact = 1.01) and in the Obit task UVBlAvg and ObitUVUtil function ObitUVUtilBlAvgTF.

III. OBIT TASK UVBLAVG

Baseline dependent temporal (and independent frequency averaging) is implemented in the Obit task UVBlAvg and ObitUVUtil function ObitUVUtilBlAvgTF. It is generally advisable to apply any prior calibration and editing before data averaging so this is supported in UVBlAvg.

Baseline dependent time averaging will also result in the data samples getting out of strict time ordering. In order to write a time ordered data-set while avoiding an expensive sort of the whole data-set, UVBlAvg uses a substantial size output data buffer which is sorted into time order before being written. The on-line documentation for UVBlAvg is given in the appendix.

IV. LIMITS TO TIME AVERAGING

Avoidance of time smearing of visibilities is not the only reason to limit the integration time. In particular, if subsequent calibration is to be used to remove time dependent effects, the data should not be averaged longer than the shortest timescale to be removed. High dynamic range imaging usually involves self–calibration on at most modest timescales; Section V-B explores the effects of upper limits to the averaging time on the achieved reduction in the data volume.

V. IMAGING TESTS

Test data sets were generated using Obit Task UVSim. [2]. Test data-sets were generated for VLA “A” (35 km) and “B” (10 km) configurations. The data-set simulation used four “IFs” near 1.4 GHz separated by 128 MHz each with 4 channels each for the A configuration and 10 channels of width 0.25 MHz for the B configuration. The initial integration time for each data-set was set to the maximum value allowable to result in minimal time smearing on the longest baselines at the edge of the primary beam, 1 sec for the B configuration and 0.33 sec for the A configuration. Data consisted of 34 scans of approximately 5 min. each spread over 7 hours of time. Tests were made of two sizes of field of view by using a single test source at the distance from the pointing center of the desired field of view. For the smaller field of view, to the 45% point of the antenna power pattern, a single unit amplitude point source was added offset by 600" in both RA and Dec giving a FOV of 850" (14.1'). The larger field of view was to the 1% power point and used offsets of 1200" giving a FOV of 1700" (28.2'). No corrections were made for the primary power pattern of the antennas.

A. Compression Ratios and Imaging Artifacts

Tests were then performed in which the data were time averaged using a number of maximum amplitude distortion factors (max_fact in eq. 1) but to no longer than one minute. The averaged data-sets were then imaged using Obit task Imager centered on the expected position of the "source", using 300×300 pixel grids, and 250 iterations of CLEAN. The "autoCenter" feature was used to ensure that the source was exactly centered on a pixel.

The imaging results were evaluated by determining the "dynamic range" (= ratio of the peak flux to off–source RMS) and the peak pixel value in the image. The "compression ratio" is the ratio of the size of the unaveraged dataset to the averaged one. The results are given in Tables I and II and Figures 1 - 3. Tables I and II give the maximum amplitude distortion factor (max_fact), the VLA configuration, the number of visibilities in the averaged data set, the log of the dynamic range, the compression ratio and the peak in the image. The sequence of images derived from the B configuration 28' offset test is shown in Figures 4 & 5.

B. Upper limits on Averaging

The results given above all used an upper limit to the averaging time of one minute. This allows for large compression ratios but is rather longer than desired for self–calibration. A series of tests were performed using values of max_fact which give compression factors of 5–6 with a one minute upper limit (1.0025 for a half power FOV, 1.004 for the full beam FOV); in which a range of upper limits on the averaging time were used. The results are shown in Table III for maximum integration times of 1 min, 30 seconds and 15 seconds. For the simulated VLA A configuration tests, the upper limit had very little effect of the results of the averaging. This indicates that even on the shorter baselines, the limit on amplitude reduction limited the averaging to less than 15 seconds. The 15 seconds upper limit increased the size of averaged simulated VLA B configuration data by 3–4%. Thus, it appears that the upper limit on the averaging time can be set small enough that subsequent self–calibration is possible.

TABLE I
DATA AVERAGING RESULTS 14.1' OFFSET

factor	VLA config	num. vis	$\log_{10}$ DR	ratio	peak
1.000	A	10173475	8.99	1.00	0.999924
1.001	A	7073573	7.35	1.44	0.999833
1.002	A	2517840	6.14	4.04	0.999432
1.003	A	1204279	5.57	8.45	0.998001
1.005	A	847517	5.28	12.00	0.996111
1.007	A	722859	5.15	14.07	0.994705
1.010	A	610511	5.00	16.66	0.992634
1.015	A	504367	4.84	20.17	0.989278
1.020	A	440284	4.72	23.11	0.985964
1.030	A	364000	4.55	27.95	0.979500
1.050	A	287224	4.35	35.42	0.967090
1.070	A	246714	4.21	41.23	0.955339
1.100	A	210788	4.08	48.26	0.938647
1.150	A	177494	3.91	57.32	0.913258
1.000	B	3664791	8.77	1.00	1.000000
1.001	B	2388415	7.31	1.53	0.999988
1.002	B	822331	6.16	4.46	0.999618
1.003	B	394719	5.57	9.28	0.998230
1.005	B	279389	5.29	13.12	0.996396
1.007	B	239775	5.16	15.28	0.995043
1.010	B	204331	5.02	17.94	0.993064
1.015	B	171089	4.85	21.42	0.989906
1.020	B	151623	4.74	24.17	0.986842
1.030	B	128564	4.57	28.51	0.980969
1.050	B	106256	4.37	34.49	0.970237
1.070	B	95179	4.25	38.50	0.960279
1.100	B	85962	4.12	42.63	0.970829
1.150	B	78188	3.98	46.87	0.929677

TABLE II
DATA AVERAGING RESULTS, 28.2' OFFSET

factor	VLA config	num. vis	$\log_{10}$ DR	ratio	peak
1.000	A	10173475	9.01	1.00	0.999925
1.001	A	8821295	7.76	1.15	0.999877
1.002	A	4552068	5.90	2.23	0.999509
1.003	A	2234629	5.45	4.55	0.997840
1.005	A	1605326	5.19	6.34	0.995840
1.007	A	1376200	5.09	7.39	0.994318
1.010	A	1170652	4.95	8.69	0.992177
1.015	A	971166	4.80	10.48	0.988643
1.020	A	850755	4.69	11.96	0.985243
1.030	A	704951	4.52	14.43	0.978497
1.050	A	556110	4.32	18.29	0.965650
1.070	A	476847	4.19	21.33	0.953408
1.100	A	406311	4.05	25.04	0.936211
1.150	A	339538	3.89	29.96	0.909550
1.000	B	3664791	8.78	1.00	1.000000
1.001	B	3074679	7.72	1.19	0.999998
1.002	B	1467759	5.88	2.50	0.999647
1.003	B	731204	5.46	5.01	0.998060
1.005	B	523978	5.20	6.99	0.996073
1.007	B	450477	5.09	8.13	0.994638
1.010	B	383662	4.96	9.55	0.992536
1.015	B	318858	4.80	11.49	0.989078
1.020	B	280405	4.70	13.07	0.985792
1.030	B	234145	4.54	15.65	0.979317
1.050	B	187333	4.34	19.56	0.967073
1.070	B	162714	4.21	22.52	0.955536
1.100	B	141063	4.08	25.98	0.939698
1.150	B	121266	3.91	30.22	0.915789

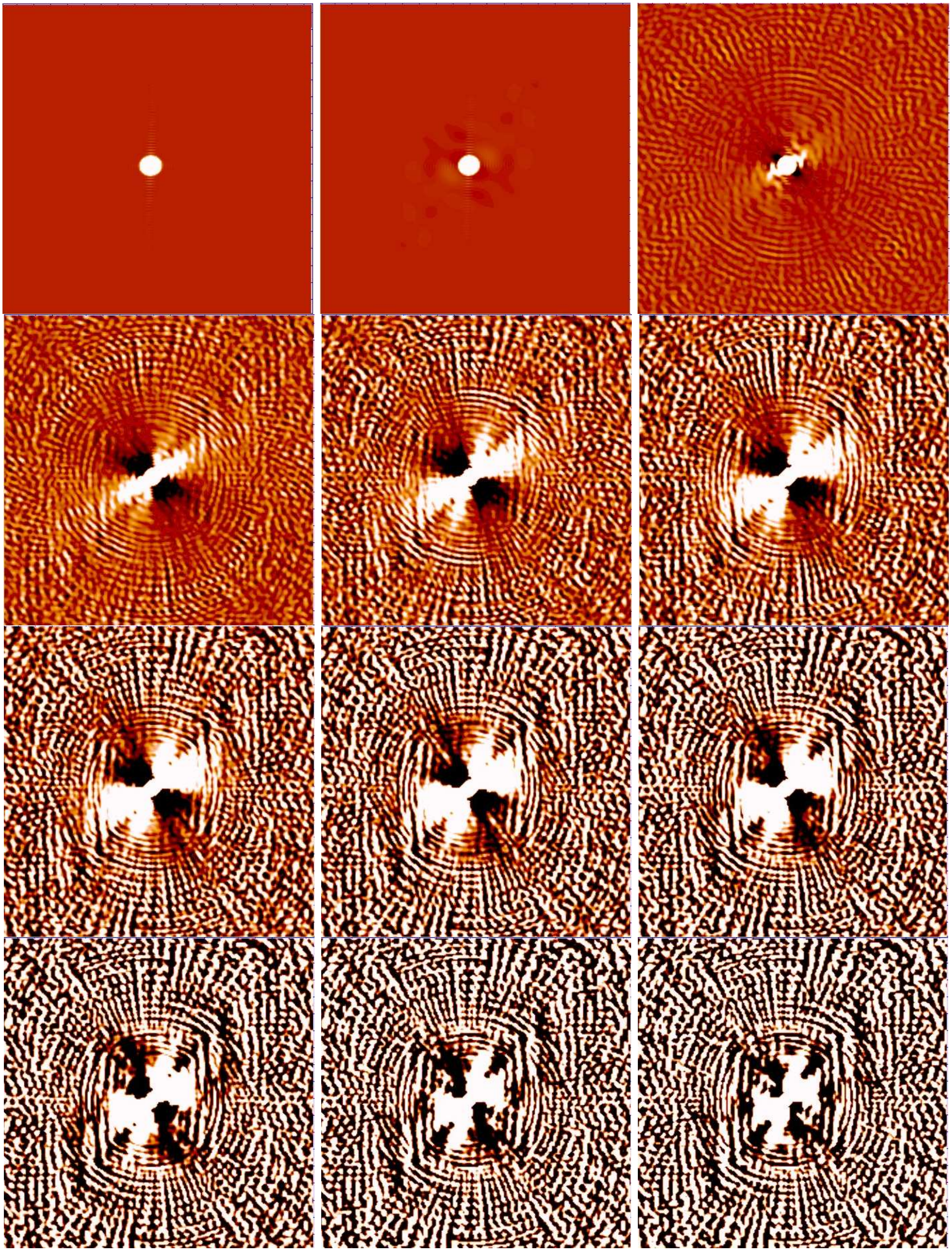

Fig. 4. The B configuration image artifacts around a source 28.2' from the pointing center for various allowed distortion factors, top row: 1.000,1.001, 1.002; second row: 1.003, 1.005, 1.007; third row: 1.01, 1.015, 1.02; bottom: 1.03, 1.05, 1.07. The range of values shown is -1.0$\times 10^{-5}$ to 1.0$\times 10^{-5}$ of the peak for the 4 $\times$ 4 arc min field centered on the peak.

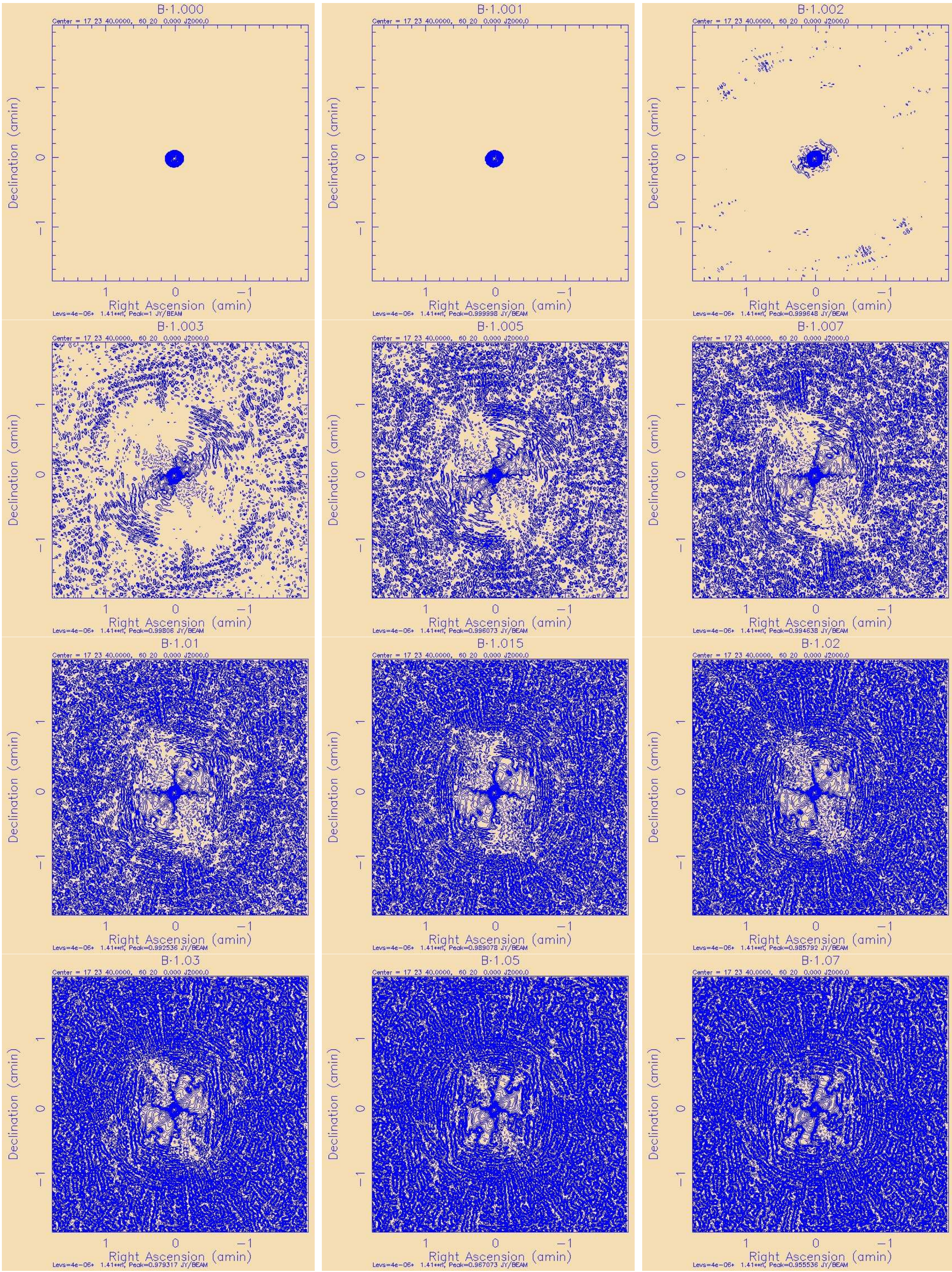

Fig. 5. The B configuration image artifacts around a source 28.2' from the pointing center for various allowed distortion factors, top row: 1.000,1.001, 1.002; second row: 1.003, 1.005, 1.007; third row: 1.01, 1.015, 1.02; bottom: 1.03, 1.05, 1.07. Contours are at powers of $\sqrt{2}$ times 4.0×10^{-6} of the peak. Negative contours are dashed.

TABLE III
DATA AVERAGING VARYING MAX. INTEGRATION

factor	VLA config	Max Int min.	FOV amin	num. vis	log_10 DR	ratio	peak
1.0025	A	1.00	14	1608297	5.79	6.33	0.998819
1.0025	A	0.50	14	1608420	5.79	6.33	0.998820
1.0025	A	0.25	14	1610040	5.79	6.32	0.998822
1.0025	B	1.00	14	524960	5.80	6.98	0.999017
1.0025	B	0.50	14	528114	5.81	6.94	0.999023
1.0025	B	0.25	14	545905	5.83	6.71	0.999044
1.0040	A	1.00	28	1789668	5.29	5.68	0.996650
1.0040	A	0.50	28	1789722	5.29	5.68	0.996651
1.0040	A	0.25	28	1791018	5.29	5.68	0.996657
1.0040	B	1.00	28	584483	5.28	6.27	0.996888
1.0040	B	0.50	28	587002	5.28	6.24	0.996902
1.0040	B	0.25	28	602350	5.30	6.08	0.996954

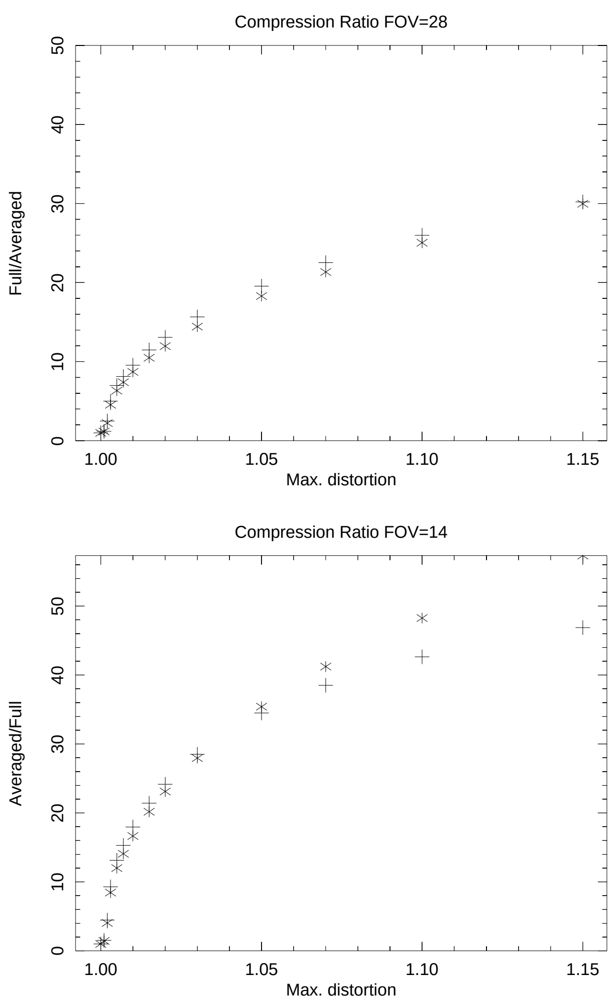

Fig. 1. The ratio of the averaged data set size to the unaveraged data set size for each test. The “*” symbols denote the VLA A configuration simulations and “+” the B configuration. The upper plot is for the 28’ FOV, lower for 14’.

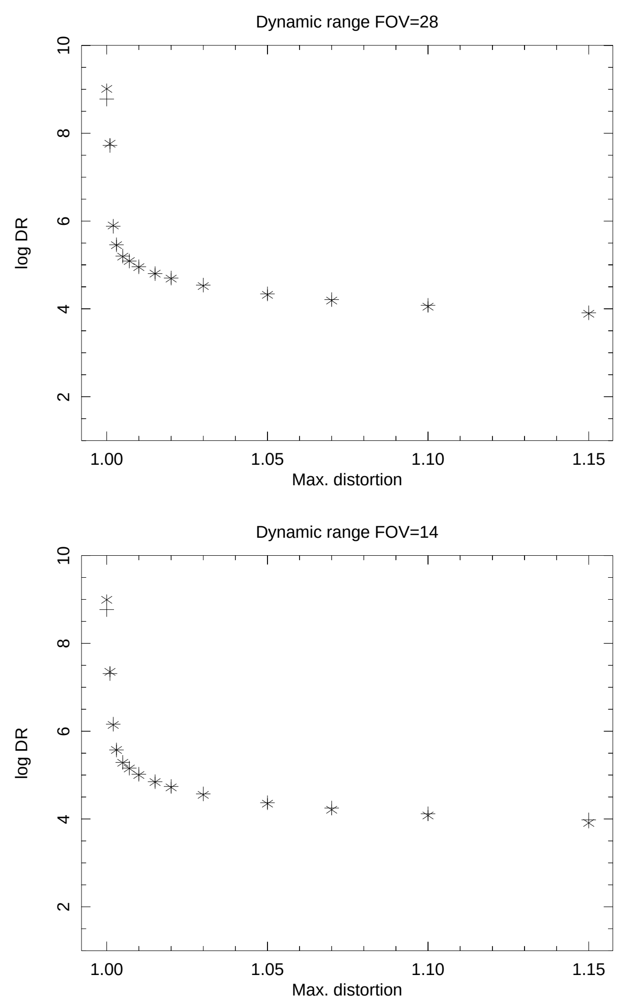

Fig. 2. Log_{10} of the dynamic range in the derived image in each test. The “*” symbols denote the VLA A configuration simulations and “+” the B configuration. The upper plot is for the 28’ FOV, lower for 14’.

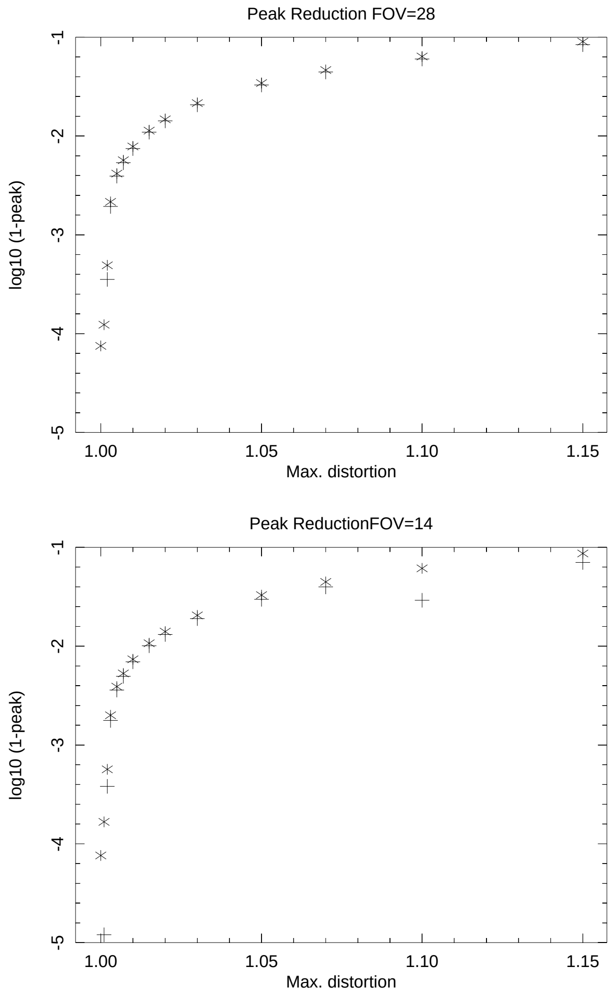

Fig. 3. Log_{10}(1.0-peak) in the derived image in each test. The "*" symbols denote the VLA A configuration simulations and "+" the B configuration. The upper plot is for the 28' FOV, lower for 14'.

VI. Discussion

Previous sections explore the results of time averaging visibility data in a baseline dependent fashion with the constraint that the averaging does not to distort the visibility amplitude by more than a given amount due to the motion through the uv plane. The time averaging cannot be of an arbitrary length; tests of the effect of an upper limit to the averaging time on the compression ratio were also performed.

The results above show that the proper integration times for minimal time smearing on the longest baselines out to the first null is $\approx$1 second for the VLA B configuration and $\approx$0.33 seconds for the A configuration. The effects of time averaging shown in Figures 1, 2 and 3 and Tables I and II show only a weak dependence on the VLA configuration but a data compression ratio that depends on the desired field of view. A baseline dependent averaging can reduce the data volume by a factor of 5–6 at the cost of degrading the dynamic range in the vicinity of bright sources at the edge of the target field of view to of order $10^{5.3}$ or more and a reduction in the peak brightness by no more than 0.2%. The artifacts are stronger very close to the strong source. In practice, other effects such as pointing errors and uncertainties in the detailed antenna power pattern may be significant factors limiting the dynamic range.

The degradation of the image is a function of the maximum allowed distortion factor for a given field of view and is nearly independent of the size scale of the array; this is as expected. The maximum amplitude distortion factor and the field of view may be tunable to the application.

It is, in principle, possible to make first order corrections for time smearing in deconvolution and self–calibration. In the calculation of the instrumental response to the sky model, these corrections can be made on a baseline basis for each CLEAN component included in the model when using the "DFT" model calculation method. This comes at the expense of an increase in the computing required and a slight increase in the noise.

Application of baseline dependent averaging needs to be taken into account in any subsequent calibration in which a solution interval is less that the maximum allowed integration time. In such cases, the full sensitivity of the array will not be included in each solution interval as the baselines with the longest averaging (generally the shortest baselines) may not be represented in each solution interval. In such cases, it will be necessary to fit for a functional form of the calibration (e.g. phase and fringe rate) over time scales longer than the maximum integration time. This also implies that it is advantageous to apply as much calibration as possible prior to averaging the data. Test results suggest that the maximum averaging time may be set short enough that this is not a major problem.

The results above show that a reduction of in data volume of order 5–6 from VLA A configuration data sampled every 0.33 seconds and VLA B configuration data sampled every 1 second can be achieved with a resulting dynamic range in the neighborhood of strong sources at the edge of the field of view of order $10^{5.3}$ and a reduction of peak amplitude by about 0.3%. The suggested value of max_fact (`maxFact` in UVBlAvg) is 1.0025 for a field of view of the half power beam-width of the primary beam and 1.004 for a field of view of the full width of the primary beam. An upper limit to the averaging time of 15 seconds has a minimal impact of the averaged data size but allows subsequent self calibration on this time scale. The results were calculated for 1.4 GHz but when expressed in terms of the beam width, should also apply to other frequencies.

In summary, judicious use of baseline dependent averaging can lead to a very significant reduction in data volume with a modest degradation in the image.

Appendix

The "help" segment of the on-line documentation for UVBlAvg is given.

References

[1] W. D. Cotton, "Obit: A Development Environment for Astronomical Algorithms," *PASP*, vol. 120, pp. 439–448, 2008.
[2] ——, "UVSim: Obit UV Data Simulator," *Obit Developement Memo Series*, vol. 7, pp. 1–8, 2009.

```
UVBlAvg
Task:  Frequency and time average UV data with averaging times
depending on time and baseline.  The averaging time is the lesser
of maxInt and the time it takes for time smearing to reduce the
visibility amplitude by maxFact.

A description of this task can be found in
ftp://ftp.cv.nrao.edu/NRAO-staff/bcotton/Obit/BLAverage.pdf

Adverbs:
  DataType..'FITS' or 'AIPS'  type of input
  inFile.....FITS input uvdata if DataType=='FITS'
  inName.....Input UV AIPS file name     Standard defaults.
  inClass....Input UV AIPS file class.   Standard defaults.
  inSeq......Input UV AIPS fileseq. #.  0 => highest.
  inDisk.....Disk drive # of input UV (FITS or AIPS). NO default
  Sources... Sources  to be processed if multisource file.
  souCode....Calibrators may be selected on the basis of the
             calibrator code given in the SU table.
                  '    ' => any calibrator code selected
                  '*   ' => any non blank code (cal. only)
                  '-CAL' => blank codes only (no calibrators)
                  anything else = calibrator code to select.
             NB: The souCode test is applied in addition to the
             other tests, i.e. Sources and Qual, in the
             selection of sources to process.
  Qual.......Only sources with a source qualifier number in the
             SU table matching Qual will be used if Qual is not
             -1.
  timeRange..Time range of the data to be processed. In order:
             Start and end times in days relative to ref. date.
             Use  dhms2day to convert from human readable form
  Stokes.....Stokes parameters to copy, "    "=> Don't convert
             Otherwise will select and/or convert..
  FreqID.....Frequency identifier to select , <=0=>any
  BChan......First channel number to image, 0=>1.  Channel numbers are 1
             relative as defined in the input data file.
  EChan......Highest channel number to to include in image,
             0 => max
  BIF........First IF to process. 0=>1
  EIF........Highest IF to process 0=> do BIF to highest.
             Note: not all data sets will have IFs.
  subA.......Sub-array number to use. 0=>all.
  Antennas...A list of the antennas to  have solutions
             determined. If any number is negative then all
             antennas listed  are NOT to be used to determine
             solutions and all others are. All 0 => use all.
  doCalib....If >0, calibrate the data using information in the
             specified Cal (CL) table for multi-source or SN table for
             single-source data.  If > 1.5, also calibrate the weights
  gainUse....Version number of the CL table to apply to multi-source
             files or the SN table for single-source files.
             0 => highest.
  doPol......If > 0 then correct data for instrumental polarization as
             represented in the AN table.  This correction is only
             useful if PCAL has been run or feed polarization parameters
             have been otherwise obtained.
  flagVer....Specifies the version of the flagging table to be applied.
             0 => highest numbered table. <0 => no flagging to be
             applied.
```

```
doBand.....If > 0 then correct the data for the shape of the
           antenna bandpasses using the BP table specified by BPVer.
           The correction has five modes:
           (a) if doBand=1 all entries for an antenna in the table
           are averaged together before correcting the data.
           (b) if doBand=2 the entry nearest in time (including
           solution weights) is used to correct the data.
           (c) if doBand=3 the table entries are interpolated in
           time (using solution weights) and the data are then
           corrected.
           (d) if doBand=4 the entry nearest in time (ignoring
           solution weights) is used to correct the data.
           (e) if doBand=5 the table entries are interpolated in
           time (ignoring solution weights) and the data are then
           corrected.
BPVer......Specifies the version of the BP table to be applied if
           doBand > 0.  0 => highest numbered table.
           <0 => no bandpass correction to be applied.
Smooth.....Specifies the type of spectral smoothing to be applied to
           a uv database . The default is not to apply any smoothing.
           The elements of Smooth are as follows:
           Smooth(1) = type of smoothing to apply: 0 => no smoothing
             To smooth before applying bandpass calibration
               1 => Hanning, 2 => Gaussian, 3 => Boxcar, 4 => Sinc
             To smooth after applying bandpass calibration
               5 => Hanning, 6 => Gaussian, 7 => Boxcar, 8 => Sinc
           Smooth(2) = the "diameter" of the function, i.e. width
             between first nulls of Hanning triangle and sinc
             function, FWHM of Gaussian, width of Boxcar. Defaults
             (if < 0.1) are 4, 2, 2 and 3 channels for Smooth(1) =
             1 - 4 and 5 - 8, resp.
           Smooth(3) = the diameter over which the convolving
             function has value - in channels.  Defaults: 1,3,1,4
             times Smooth(2) used when input Smooth(3) < net
             Smooth(2).
FOV........The radius of the field of view over which time averaging
           is not to reduce the visibility amplitude at the top of
           the observed frequency range by more than maxFact.
           Given in degrees
maxInt.....Maximum integration time (min) for time averaging.
           Note: time tags within maxInt will be averaged, if this is
           an integral number of integrations, the actual averaging
           time will be one more integration than you are expecting.
maxFact....Maximum allowable additional amplitude loss due to time
           averaging.
avgFreq....If 0 < avgFreq <= 1, then spectral channels in the
           range from BChan to EChan are averaged in blocks of
           chAvg channels subject to further selection by
           ChanSel.
           If avgFreq > 1, then all frequency channels in
           each IF (under control of ChanSel) will be averaged.
           If avgFreq > 2, then all IF's will be averaged also.
chAvg......If avgFreq = 1, the number of channels to be averaged
           together in each channel.
           0 => ALL.
```

```
  ChanSel....The channels to be averaged when avgFreq > 0 in the
             form of an array of start and stop channels plus a
             channel increment and the IF to which they apply.
             All 0 => BChan, EChan, 1, 0.
             ChanSel(4,i) gives the IF and 0 means all IFs.
             Up to 20 groups of these 4 numbers may be
             specified.  Note that the channel numbers are absolute
             numbers; they are NOT relative to BIF and BChan.
             For instance if your data had a spectral line covering
             channels 56 - 80, and you wished to exclude channels
             1 - 10 and 121 - 128 because of bandpass effects, then
             you could set avgFreq=2, and ChanSel=11, 55, 1, 0,
             81, 121, 1, 0 for a 1-IF data set.
             If you only wished to use every other channel from the
             second group then you would set ChanSel = 11, 55, 0, 0,
             81, 121, 2, 0.
  outDType..'FITS' or 'AIPS'  type of output
  outName....Output UV AIPS file name    Standard defaults.
  outClass...Output UV AIPS file class.  Standard defaults.
  outSeq.....Output UV AIPS file seq. #. 0 => highest unique.
  outDisk....Disk drive # of output UV (FITS or AIPS) NO default
             0 FITS => current directory
  Compress...If True  the output data is written in
             compressed format which can result in a substantial
             reduction in disk space needed.

VLA Beam sizes
--------------
Freq (GHz)  FOV of full beam(deg)
   0.074          5.00
   0.330          1.25
   1.4            0.25
   5.0            0.075
   8.             0.045
  15              0.025
  22              0.01667
  45              0.00833

Recommended Parameters
----------------------
maxFact
For undistorted imaging with a field of view half of the primary
beam size a maxFact=1.0025 or maxFact=1.004 for a FOV of the full
primary beam will result in a reduction in the data volume of 5-6
with a limit on the dynamic range near bright sources at the edge
of the field of view of about 10^5.3 and a reduction in the peak
brightness at the edge of the field of view of no more than 0.5%.

maxInt
A value of maxInt=0.25 (15 seconds) has minimal impact on the
size of the output data set for VLA A and B configuration data
and allows subsequent self calibration with a solution interval
of 15 seconds or longer.
```