ERROR RECOGNITION & IMAGE ANALYSIS

Gustaaf van Moorsel (NRAO)
Ed Fomalont (NRAO)

Twelfth Synthesis Imaging Workshop
2010 June 8-15
Why are these two topics – ‘Error Recognition’ and ‘Image Analysis’ in the same lecture?

-- Error recognition is used to determine defects in the (visibility) data and image during and after the ‘best’ calibration, editing, etc.

-- Image analysis describes the almost infinite ways in which useful insight, information and parameters can be extracted from the image.

Perhaps the two topics are related to the reaction one has when looking at an image after ‘good’ calibration, editing, self-calibration, etc.

If the reaction is:
Rats!!

This can’t be right. This is either the most remarkable radio source ever, or I have made an error in making the image.

Clear signs of problems:

- Image rms > expected rms
- Unnatural features in the image

How can the problems be found and corrected?

mJy scale
Great!!

After lots of work, I can finally analyze this image and get some interesting scientific results.

What were defects?

Two antennas had 10% calibration errors, and one with a 5 deg error, plus a few outlier points.

This part of the lecture.

How to find the errors and remove them.
GENERAL PROCEDURE

We assume that the data have been edited and calibrated reasonably successfully (earlier lectures). Self-calibration is sometimes necessary.

So, the first serious display of an image leads one—to inspect again and clean-up the data repeating some or all of the previous reduction steps.

removal of one type of problem can reveal next problem!

once all is well, proceed to image-analysis and obtaining scientific results from the image.

But, first a digression on data and image display. First:

Images
### Digital image

Numbers are proportional to the intensity.

Good for very slow links; rarely used anymore.
These plots are easy to reproduce and printed

Contour plots give good representation of faint emission.
Profile plots give a good representation of the bright emission and faint ripples.
TV-based displays are most useful and interactive:

Grey-scale shows faint structure, but not good for high dynamic range and somewhat unbiased view of source. Color displays more flexible; e.g. pseudo contours.
Very primitive display, but sometimes worth-while: e.g., can search on e.g. Amp > 1.0, or large Wt. Often need precise times in order to flag the data appropriately.
Visibility Amplitude versus Projected uv spacing

General trend of data. Useful for relatively strong sources.

Triple source model. Large component cause rise at short spacings.

Oscillations at longer spacings suggest close double.

(see Non-imaging lecture)
Visibility amplitude and phase versus time for various baselines

Good for determining the continuity of the data
Should be relatively smooth with time
Outliers are obvious.

DATA DISPLAYS(3)
Weights of antennas 4 with 5,6,7,8,9

All u-v data points have a weight. The weight depends on the antenna sensitivity, measured during the observations.

The amplitude calibration values also modify the weights.

Occasionally the weight of the points become very large, often caused by subtle software bugs.

A large discrepant weight causes the same image artifacts as a large discrepant visibility value.

Please check weights to make sure they are reasonable.
Errors obey Fourier transform relationship

Narrow feature in uv plane $\leftrightarrow$ wide feature in image plane
Wide feature in uv plane $\leftrightarrow$ narrow feature in image plane

- Note: easier to spot narrow features

Orientations are orthogonal

Data uv amplitude errors $\leftrightarrow$ symmetric image features
Data uv phase errors $\leftrightarrow$ asymmetric image features

An obvious defect may be hardly visible in the transformed plane
A small, almost invisible defect may become very obvious in the transformed plane
GOLDEN RULE OF FINDING ERRORS

---Obvious outlier data (u-v) points:

100 bad points in 100,000 data points gives an 0.1% image error (unless the bad data points are 1 million Jy)

LOOK at DATA to find gross problem (you’d be hard pressed to find it in the image plane other than a slight increase in noise)

---Persistent small data errors:

e.g. a 5% antenna gain calibration error is difficult to see in (u-v) data (not an obvious outlier), but will produce a 1% effect in image with specific characteristics (more later).

USE IMAGE to discover problem

---Non-Data Problems:

Perfect data but unstable algorithms. Common but difficult
Editing obvious errors in the u-v plane

---Mostly consistency checks assume that the visibility cannot change much over a small change in u-v spacing
---Also, double-check gains and phases from calibration processes. These values should be relatively stable.

See Summer school lecture notes in 2002 by Myers
See ASP Vol 180, Ekers, Lecture 15, p321
VISIBILITY AMPLITUDE PLOTS

Amp vs. uvdist shows outliers

Amp vs. time shows outliers in last scan

Amp vs. time without ant 7 should good data

(3C279 VLBA data at 43 GHz)
VISIBILITY AMPLITUDE RASTERS

(Raster scan of baseline versus time immediately shows where the bad data are)

Pixel range is 5 to 20 Jy

Bad data can be flagged with an interactive clipping control
Example Edit – plotms (2)

Fourier transform of nearly symmetric Jupiter disk

Jansky

Kilo-wavelength

bad

Butler lecture: Solar System Objects
Drop-outs at Scan Beginnings

Often the first few points of a scan are low. E.g. antenna not on source.

Software can remove these points (aips,casa 'quack')

Flag extension:
Should flag all sources in the same manner even though you cannot see dropout for weak sources.
No source structure information is detected. Noise dominated.

All you can do is remove outlier points above 0.3 Jy. Precise level not important as long as large outliers removed.
USING TVFLG (VIEWER) DISPLAY on a source

Plot amplitude rms

ANT-23 problems

quack these!

Baseline--->
RFI environment worse on short baselines

Several 'types': narrow band, wandering, wideband, ...

Wideband interference hard for automated routines

AIPS tasks FLGIT, FLAGR, and CASA flagdata, mode='rfi'

Automation is crucial for WIDAR (wide band, lots of data)
Some Questions to ask:

**Noise properties of image:**
- Is the rms noise about that expected from integration time?
- Is the rms noise much larger near bright sources?
- Are there non-random noise components (faint waves and ripples)?

**Funny looking Structure:**
- Non-physical features; stripes, rings, symmetric or anti-symmetric
- Negative features well-below 4xrms noise
- Does the image have characteristics in the dirty beam?

**Image-making parameters:**
- Is the image big enough to cover all significant emission?
- Is cell size too large or too small? ~4 points per beam okay
- Is the resolution too high to detect most of the emission?
EXAMPLE 1
Data bad over a short period of time

Results for a point source using VLA. 13-5min observation over 10 hr. Images shown after editing, calibration and deconvolution.

no errors:
max 3.24 Jy
rms 0.11 mJy

10% amp error for all antennas for 1 time period
rms 2.0 mJy

6-fold symmetric pattern due to VLA “Y”. Image has properties of dirty beam.
EXAMPLE 2

Short burst of bad data

Typical effect from one bad antenna

- 10 deg phase error for one antenna at one time
  rms 0.49 mJy

- 20% amplitude error for one antenna at one time
  rms 0.56 mJy (self-cal)

- anti-symmetric ridges

- symmetric ridges
EXAMPLE 3

Persistent errors over most of observations

NOTE: 10 deg phase error to 20% amplitude error cause similar sized artifacts

10 deg phase error for one antenna all times
rms 2.0 mJy

20% amp error for one antenna all times
rms 2.3 mJy

rings – odd symmetry

rings – even symmetry
EXAMPLE 4

**Spurious Correlator Offset Signals**

Occasionally correlators produce ghost signals or cross talk signals. Occurred during change-over from VLA to EVLA system.

Symptom: Garbage near phase center, dribbling out into image.

- Image with correlator offsets
- Image after correlation of offsets
Even if the data are perfect, image errors and uncertainties will occur because the (u-v) coverage is not adequate to map the source structure.

The extreme rise of visibility at the short spacings makes it impossible to image the extended structure. You are better of imaging the source with a cutoff below about 2 kilo-wavelengths.

Get shorter spacing or single-dish data.
DIRTY IMAGE and BEAM (point spread function)

The dirty beam has large, complicated side-lobe structure. It is often difficult to recognize any details on the dirty image. An extended source exaggerates the side-lobes. 5% in dirty beam becomes 20% for extended source.
CLEANING WINDOW SENSITIVITY

Tight Box | Middle Box | Big Box | Dirty Beam

One small clean box | One clean box around all emission | Clean entire inner map quarter

Make box as small as possible to avoid cleaning noise interacting with sidelobes
How Deep to Clean?

Under-cleaned

- Residual sidelobes dominate the noise
- Emission from second source sits atop a negative "bowl"

Over-cleaned

- Regions within clean boxes appear "mottled"

Properly cleaned

- Background is thermal noise-dominated; no "bowls" around sources.
FINDING HIDDEN BAD DATA

Chandra Deep Field South

Peak = 45 mJy, rms = 0.02 mJy

Source to NE in first Primary beam sidelobe

See Lectures Bhatnagar and Ott on Wide-field Imaging, and Perley on High Dynamic Range Imaging

Center of Field
Fourier Transform Dirty Image

Shows the u-v data as gridded just before imaging.

Diagonal lines caused by structure in field.

A few odd points are not very noticeable.
Fourier Transform Clean Image

Shows the u-v data from clean image.

Diagonal lines still present. Notice that clean does an interpolation in the u-v plane between u-v tracks.

The odd points are smeared, but still present. These produce the low level ripples.
Bad weighting of a few u-v points

After a long search through the data, about 30 points out of 300,000 points were found to have too high of a weight by a factor of 100. Effect is <1% in image.

Cause??
Sometimes in applying calibration produced an incorrect weight in the data. Not present in the original data.

These problems can sneak up on you. Beware.
Improvement of Image

Removal of low level ripple improves detectability of faint sources

Before editing

After editing
SUMMARY OF ERROR RECOGNITION

Source structure should be ‘reasonable’, the rms image noise as expected, and the background featureless. If not,

UV data
Look for outliers in u-v data using several plotting methods.
Check calibration gains and phases for instabilities.
Look at residual data (uv-data - clean components)

IMAGE plane
Do defects resemble the dirty beam?
Are defect properties related to possible data errors?
Are defects related to possible deconvolution problems?
**IMAGE ANALYSIS**

Input: Well-calibrated data-base producing a high quality image  
Output: Parameterization and interpretation of image or a set of images  

This is very open-ended  
Depends on source emission complexity  
Depends on the scientific goals  

Examples and ideas are given.  
Many software packages, besides AIPS and Casa (e.g., IDL, DS-9) are available.
IMAGE ANALYSIS OUTLINE

Multi-Resolution of radio source.
Parameter Estimation of Discrete Components
Polarization Data
Image Comparisons
Positional Registration
Different aspect of source structure can be seen at various resolutions, shown by the ellipse in the lower left corner of each box.

SAME DATA USED FOR ALL IMAGES

For example, outer components are small from SU resolution. There is no extended emission from low resolution.
Imaging and Deconvolution of Spectral Line Data:

Type of weighting in imaging

HI contours overlaid on optical images of an edge-on galaxy
PARAMETER ESTIMATION

Parameters associated with discrete components

Fitting in the image
Assume source components are Gaussian-shaped
Deep cleaning restores image intensity with Gaussian-beam
True size * Beam size = Image size, if Gaussian-shaped. Hence, estimate of true size is relatively simple.

Fitting in (u-v) plane
Better estimates for small-diameter sources
Can fit to any source model (e.g. ring, disk)
(see non-imaging analysis)

Error estimates of parameters
Simple ad-hoc error estimates
Estimates from fitting programs
IMAGE FITTING

- AIPS task: JMFIT
- Casa tool
- imfit
DIFMAP has good u-v fitting algorithm

- Fit model directly to (u-v) data
- Compare model to data

Contour display of image
Ellipses show true component size. (super-resolution?)
COMPONENT ERROR ESTIMATES

$P$ = Component Peak Flux Density
$s$ = Image rms noise

$P/s = \text{signal/noise} = S$

$B$ = Synthesized beam size
$q_i$ = Component image size

$DP = \text{Peak error} = s$

$DX = \text{Position error} = B / 2S$

$Dq_i = \text{Component image size error} = B / 2S$

$q_t = \text{True component size} = (q_i^2 - B^2)^{1/2}$

$Dq_t = \text{Minimum component size} = B / S^{1/2}$

eg. $S=100$ means can determine size of $B/10$
Comparison and Combination of Images of Many Types

FORNAX-A Radio/Optical field

Radio is red
Faint radio core
in center of
NGC1316

Optical in
blue-white

Frame size is
60’ x 40’

NATIONAL RADIO ASTRONOMY OBSERVATORY

2005 NSF Senior Review
LINEAR POLARIZATION

Multi-purpose plot

Contours: I,Q,U Pol
Grey scale: P Pol
\sqrt{Q^2+U^2} - noise
Line segments – P angle
\text{atan2}(0.5*Q/U)
COMPARISON OF RADIO/X-RAY IMAGES

Contours of radio intensity at 5 GHz
Dots represent X-ray Intensity (photons) between 0.7 and 11.0 KeV

Contours of radio intensity at 5 GHz
Color intensity represents X-ray intensity smoothed to radio resolution
Color represents hardness of X-ray (average weighted frequency)
- Blue - soft (thermal)
- Green - hard (non-thermal)
SPECTRAL LINE REPRESENTATIONS

Intensity Image
Sum of velocity
Amount of HI
Red high,
Blue low

Average velocity
Red low vel
Blue high vel
Rotation

Second moment
Velocity width
Turbulence?

New Mexico Tech
Science Engineering Research University

The University of New Mexico
Institute for Advanced Studies of the Americas Research Library
Visualizing Spectral Line Data: Channel Images

Greyscale+contour representations of individual channel images
IMAGE REGISTRATION AND ACCURACY

- Separation Accuracy of Components on One Image due to residual phase errors, regardless of signal/noise:
  Limited to 1% of resolution
  Position errors of 1:10000 for wide fields, i.e. 0.1” over 1.4 GHz PB

- Images at Different Frequencies:
  Multi-frequency. Use same calibrator for all frequencies.
  Watch out at frequencies < 2 GHz when ionosphere can produce displacement. Minimize calibrator-target separation

- Images at Different Times (different configuration):
  Use same calibrator for all observations. Daily troposphere changes can produce position changes up to 25% of the resolution.

- Radio versus non-Radio Images:
  Header-information of non-radio images often much less accurate than that for radio. For accuracy <1”, often have to align using coincident objects.
DEEP RADIO / OPTICAL COMPARISON

Grey-Scale:
Optical emission faintest is 26-mag

Contours:
Radio Emission faintest is 10 $\mu$Jy
Radio Source Alignment at Different Frequencies

Self-calibration at each frequency aligns maximum at (0,0) point
Frequency-dependent structure causes relative position of maximum to change
Fitting of image with components can often lead to proper registration

43 GHz: res = 0.3 mas
23 GHz: res = 0.6 mas
15 GHz: res = 0.8 mas
IMAGE ANALYSIS: SUMMARY

- Analyze and display data in several ways
  Adjust resolution to illuminate desired interpretation, analysis

- Parameter fitting useful, but try to obtain error estimate
  Fitting in u-v plane, image plane

- Comparison of multi-plane images tricky (Polarization and Spectral Line)
  Use different graphics packages, methods, analysis tools

- Registration of a field at different frequencies or wave-bands can be subtle.
  Often use ad hoc methods by aligning ‘known’ counterparts