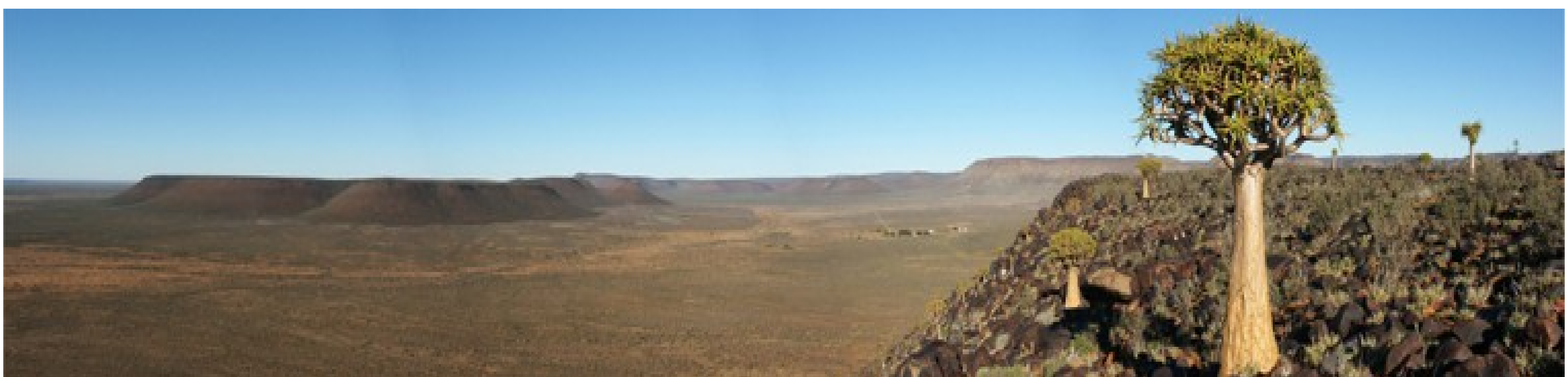
Data Analysis

Tony Foley

tony@hartrao.ac.za

science and technology

Department:
Science and Technology
REPUBLIC OF SOUTH AFRICA



Pre-analys: What, Where How

I have some project I want to study: Is radio data important ?

- Cosmology: number counts of AGNs
- Clusters
- Lensing systems
- Nearby extragalactic:
  - HI rotation curves (Dark Matter, holes, cosmic web...)
  - Supernovae, SNRs
  - Nuclei (Black Holes,...)
  - Megamasers
  - Star Forming galaxies
- Galactic
  - Masers (circumstellar matter)
  - Microquasars
  - WR winds
  - HII regions

How big is it?

Work it the relevant wavelength range and size in λ:

- The size of array/dish should correspond to $D=\lambda/\Theta$ (if λ is in meters and Θ is in radians the D is in meters). e.g

 - 1 ° (0.017 radians at 0.21m corresponds to 12m size), so a dish
 - 1 arcmin (at 21cm 700m, so an array)
 - 1 arcsecond (at 21cm 40km, so a large array)
 - 1 milliarcsecond (at 21cm 40000km, so a space VLBI array)

Think about both the largest and smallest relevant scales!

Getting the data

Find the array you want and (look for archival data or propose)

- What sensitivity will you need (array collecting area, Tsys, Bandwidth, time)

  noise $\propto$ SEFD/$\sqrt{(\text{Bandwith*Time})}$

- Is the biggest structure visible by the array?
  - Mosaic
  - Fill short (zero u,v) spacing with a dish
- What about polarization (what is needed)
- What about frequency resolution -vital for lines (what velocity or redshift and is it still observable)

What array /arrays is/are suitable...

- Propose
  - Make sure you have flux, phase and bandpass calibrators or ask !

External effects

- Ionosphere
  - Everything goes as λ^2 , so important at long wavelength (>50cm) and unimportant at short wavelength (< 20cm)
  - Rotation Measure (function of direction, time of day, state of solar cycle, .)
    - Typically <2 rad/m^2
  - Depends on where the array is (how close you are to a magnetic pole etc.)
  - Hard to follow at sunrise & sunset and often best in winter nights
  - Refraction (global and differential)
  - How to get the information
    - Internally (rotation of polarization across the band)
    - GPS
    - Ionospheric sounding
    - Model (World Data Center, Boulder Colorado)

External effects -2

- Troposphere
  - Important at short wavelength (<20cm) and unimportant at short wavelength (>50cm)
  - Depends on your weather, so very variable
  - Can be a problem for arrays >1km and large fields (>few degrees)
  - Refraction (is one global correction enough?)
  - Meteorology (pressure)
  - Water Vapour Radiometer

Internal effects

- Telescope beam:
  - Depends on how far off-axis you want to image in units of the telescope primary beam
  - Depends on how circular the beam is
  - Some telescopes have a good(-ish) model
  - Not only amplitude but also polarization cross-talk are effected
  - Easier to model when all telescopes are the same
  - Change as (if) the sky rotates with respect to the telescope axes -parallactic angle
    - Some telescope designs don't rotate with respect to sky
- How good was the pointing.. differential effects

Interference (alias RFI)

- Depends on where you are (TV stations, GSM masts, microwave ovens, WLANs, air lanes and the band you observe at)
- Has an intermittent nature
- \- but everybody gets satellites

What to do?

- Remove it in hardware
  - Reference dish pointing at main culprit (subtract)
- Try fancy removal schemes based on statistics
  - cyclostationary filters, fringe rate and delay filtering, skewness and curtosis statistics....
- Edit the data by hand...

I now have some data..

I Now have my data: what next?

- **What data reduction package is suitable**
  - **AIPS (and ParselTongue python interface) – general use**
  - **CASA (and casapy python interface)**
  - **MeqTrees (based on casa core libraries)**
  - **Miriad**
  - **DIFMAP**
  - **...**

Basic flow

1. Read in the data
2. Put know corrections in
3. Edit known problems
4. Find the calibrators and the field(s) of interest
5. Look for a bandpass (**strong)** calibrator and determine bandpass (shape and phase slope) calibrations, then apply them
6. Look for a flux density scale calibrator (strong) and apply the flux scale to the phase (point like) calibrator
7. Calculate gain and phase corrections from the phase calibrator
8. Apply interpolated gain+phase values to the field
9. Make dirty map & beam (grid+interpolate, FFT)
10. Apply a deconvolution (CLEAN or similar)

Caveats

- Inspect the corrections before you apply
  - Jumpy phased (>30º) or gain(10%) cal is useless
  - If you cant follow the trend the interpolation won't
- If you use CLEAN, start small and get large later
- If your total CLEANed flux is ≪ that the total of the field you are missing something

Slightly harder flow

1. Calculate gain and phase corrections from the sources in the field
2. Apply interpolated values to field
3. Make a new dirty map & beam (grid+interpolate, FFT)
4. Apply a deconvolution (CLEAN or similar) and get a new improved set of point sources in the field

Repeat until you hit noise floor or nothing improves
- note that you can lose absolute position information but preserve relative positions in the field (possible problems with alignment for multi-wavelength observing)

This does not cover all cases!

Some alternatives

- Model the data with something other than point sources (like shapelets) -Sarod Yatawatta
  - Sky Sources will not always fall on our pixels (image pixellation problem)
  - Slightly extended sources can be better modelled
- Subtract sources in the uv-plane (some packages do this)
- Make differences from your last map *à la* DIFMAP
- Model the whole system with a **Measurement Equation** and solve for the parameters without explicit assumptions, but:
  - Few parts of the matrix formulation commute
  - Needs far more observation points (knowns) than parameters (unknowns)
  - It is always iterative

Future concepts

The future has big telescopes (eVLA, eMERLIN, LOFAR, MeerKAT, ALMA) with correspondingly big data sets(100-1000Gbyte):

- Probably unmanagable by laptop and eyeball inspection!
- Hardly managable by server and high speed data transfer

- Maybe DFT (not FFT) by GPU and avoid gridding of u,v data
- Build better RFI removal systems and time averaging
- Automatic (pipeline) data processing
  - Your pipeline may depend on what you need
- Connect image (and data?) cubes to virtual observatory systems