ALFALFA at Ithaca June '06

Riccardo Giovanelli

Data Processing Pipeline

ALFALFA

All ALFALFA data processing takes place within the environment of IDL.

All IDL procedures have been fully developed at Cornell University.

Development of the package is virtually complete.

The package has been exported to and operates at several sites. Local groups have to provide a current IDL license.

LOVEDATA, Inc.
of Ithaca, NY

ALFALFA

The basic data unit of ALFALFA is a "**scan**", which consists of 10 minutes of data.

Since we sample the sky at 1 Hz, a scan consists of 600 spectra for each ALFA beam (plus one spare) and each polarization. Each spectrum covers 100 MHz sampled by 4096 spectral channels, so a scan consists of

2x8x600x4096 floating point fluxes + HDR

The size of a scan is ~260 Mbytes.
When converted to IDL, a scan is referred to as a "d" structure.

ALFALFA

A typical ALFALFA observing session is 3-10 hours long, consists of

20-60 scans → 0.5-2 Gbytes

First-level data processing (within days of observing) increases data archival burden by ~2.5

Gridding, signal extraction (within 6-9 months of observations) more than triples archival burden yet again.

Ultimately, ALFALFA data and data products will add up to several tens of TBytes.

For archival, we intend to take advantage of Cornell's IT Th. Center, which is organizing an interdisciplinary Cyberinfrastructure Initiative which will include NAIC.

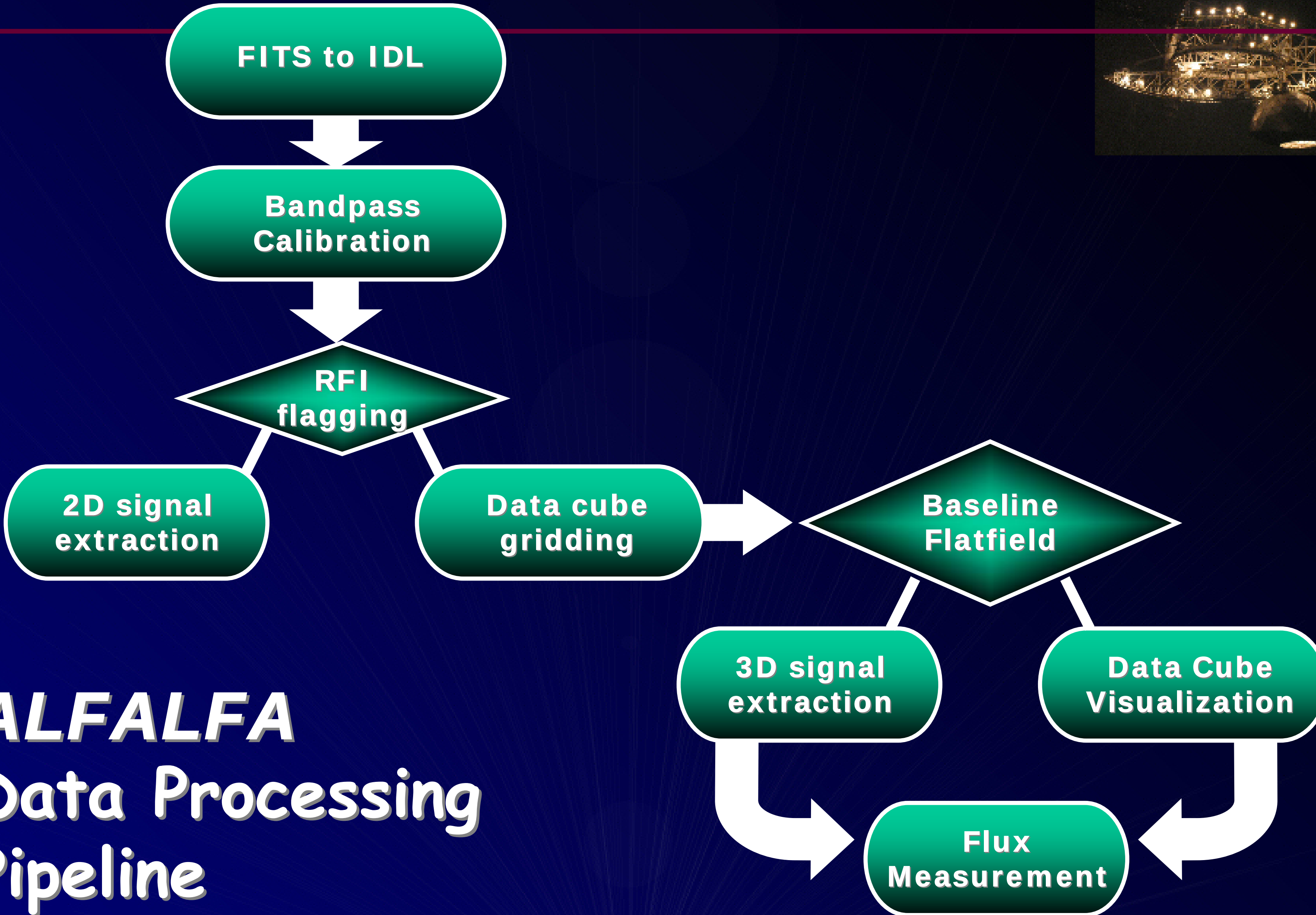
➔The basic structure carrying the bulk of the data ("d", "dred") utilized throughout the processing is based on a format defined by P. Perillat at NAIC. Other most relevant structures are

- pos, which carries the positional info of every 1-sec spectrum taken in the survey

- ncalib, which carries a history of temperature and flux calibrations

- grid, which contains all the information within, and leading to producing a 3-D cube

- src, containing data and ancillary info for each extracted source.



ALFALFA Data Processing Pipeline

ALFALFA data processing moves through the following main steps :

- 1. FITS conversion to IDL structures (FILECREATOR)
  - ➔ same day as observing session ends
- 2. IDL structures ("d") of raw data shipped to Ithaca
  - ➔ same day as observing session ends
- 3. Calibration to T_{ant} , Bandpass correction (CALIB, BPC, BPD / BPDGUI) ("dred" structures), creation of "pos" structures
  - ➔ within 2-3 weeks of data taking

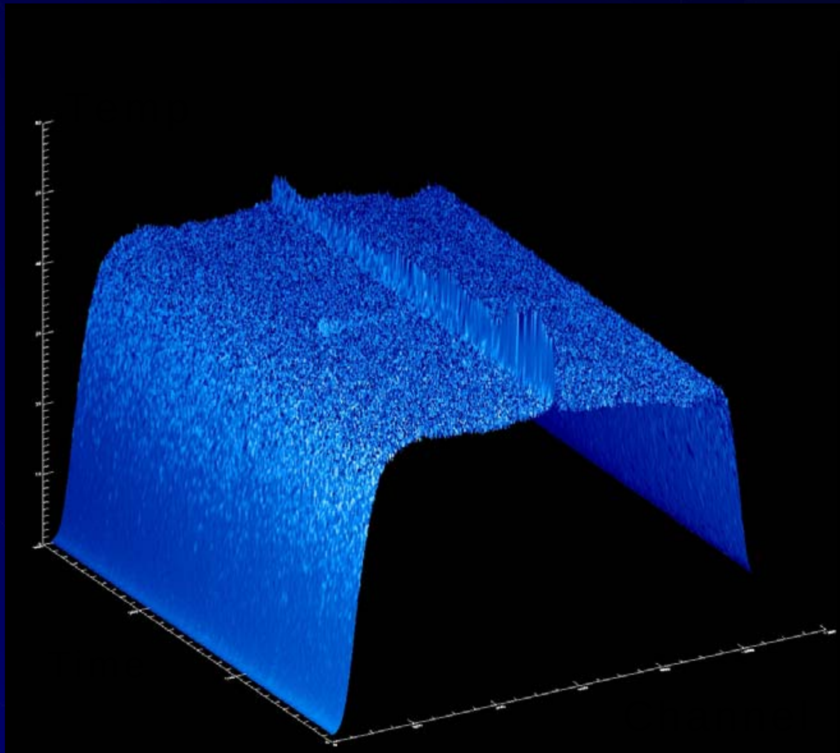
  (data processing logs accessible (requires password) at: http://caborojo.astro.cornell.edu/alfalfalog/archive.php)
- 4. RFI-flagging, visual data inspection of 2-D (frequency-time) images, data quality assessment, 2-D signal extraction (FLAGBB, REVIEWBB, INSPECT, RUN_2D_EXTRACT)
  - ➔ within one month of data taking

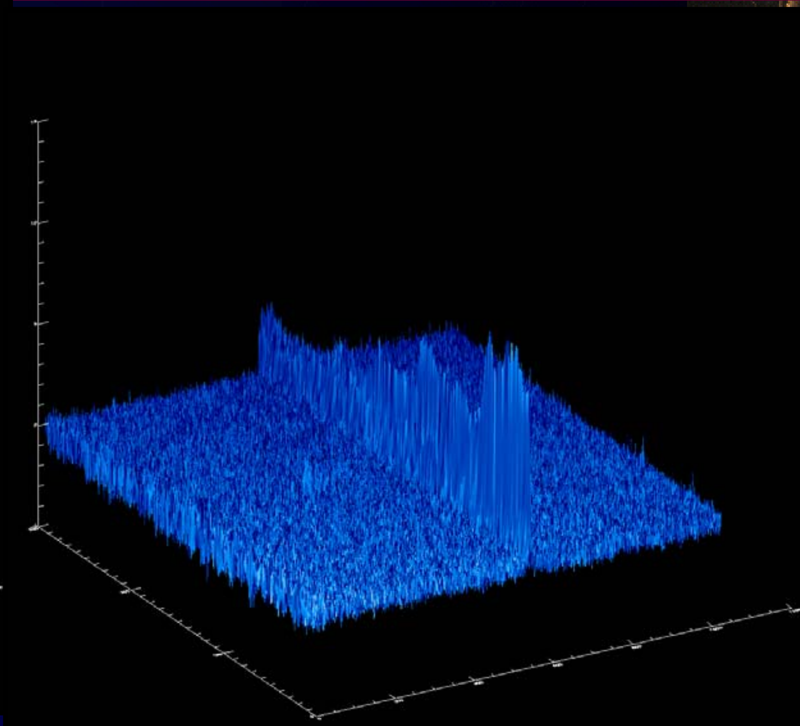
ALFALFA

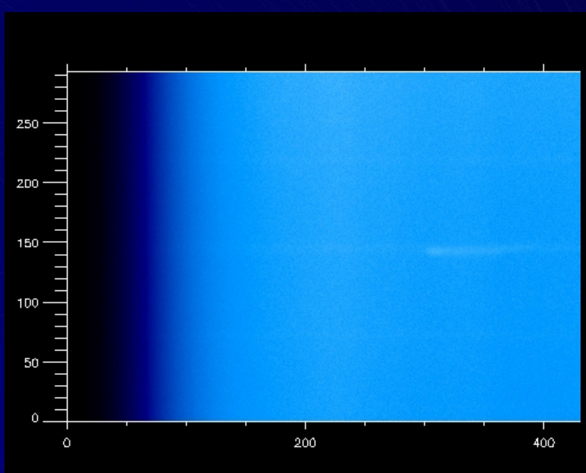
ALFALFA Data Processing Steps (continues):

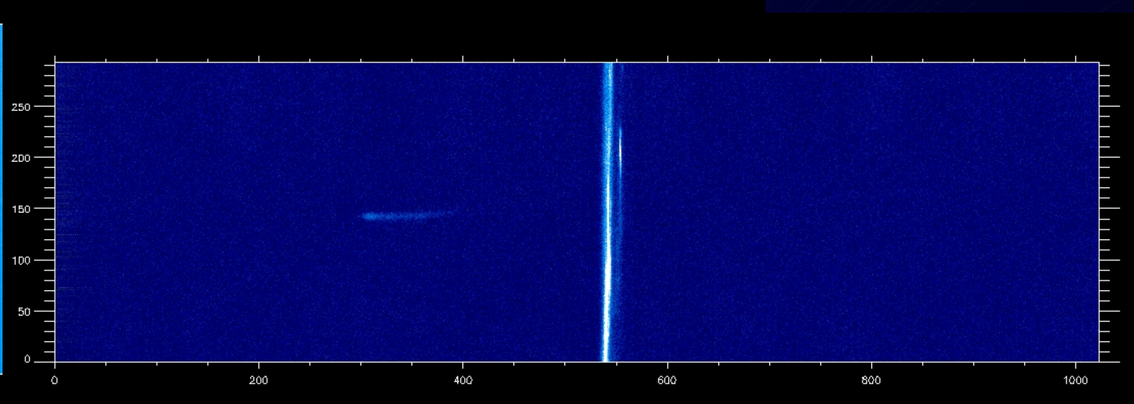
5. Production of 3-D tiles: gridding raw data into regularly sampled 3-D sets of 144x144x1024 pix (that is 2.4deg x 2.4deg x 5400 km/s) data cubes (grids). Grid centers are spaced 8 min in RA and 2 deg in Dec. Conversion to flux density scale. Four sections cover 4 different velocity ranges,per space grid (GRID_PREP). Grid management via web interface.
   Grid baselining, flatfielding, cross-calibration (GRID_BASE, GRID_FLATFIELD)
   - ➔ within 9 months of completed, 2-pass observations of given tile
6. Grid inspection (GRIDVIEW), 3-D signal extraction (EX3D_C, EX3D_D)
7. Flux Measurement, cataloging (GAL_FLUX)
   - ➔ within 12 months of completed 2-pass observations of given tile

ALFALFA





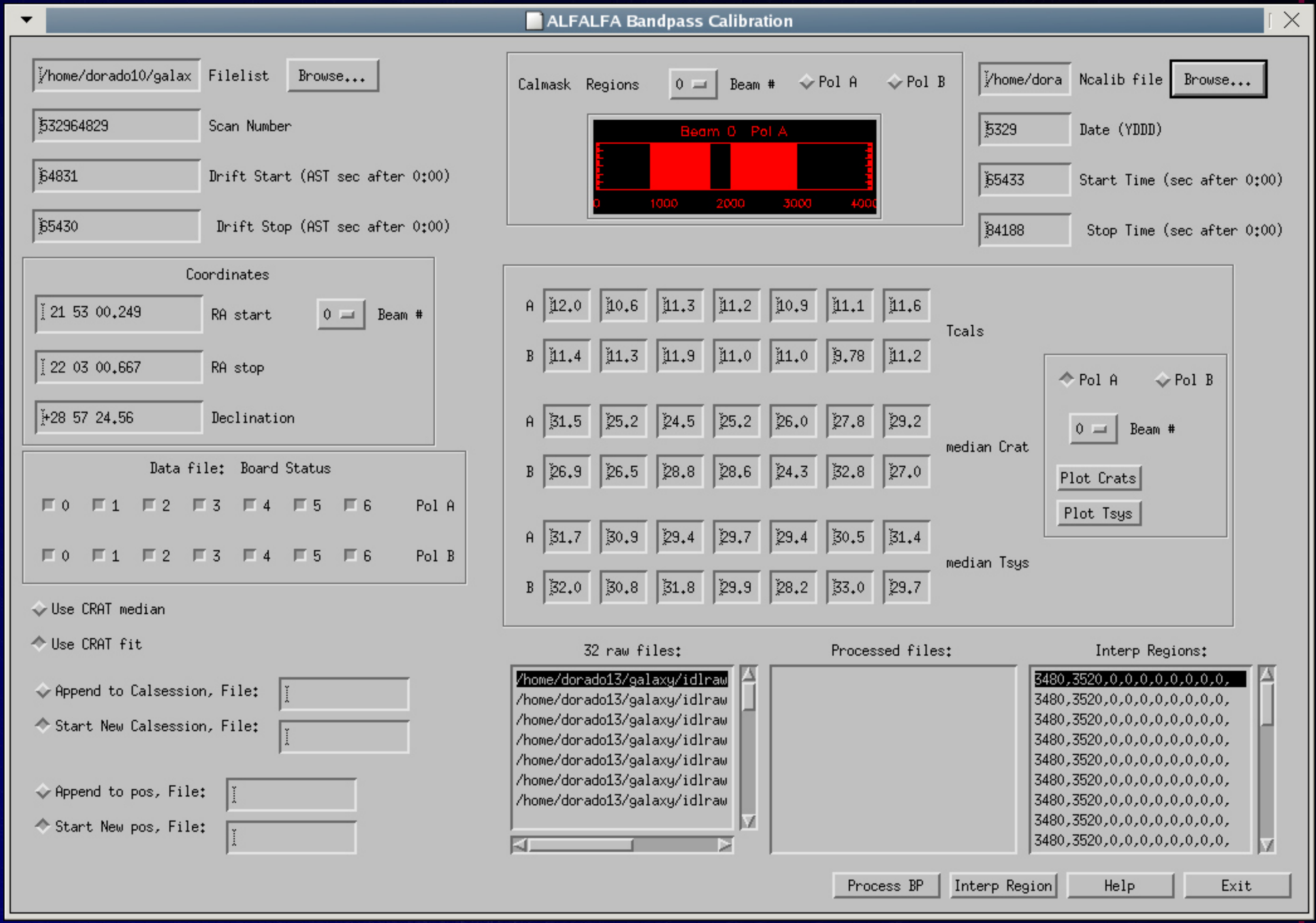

250
200
150
100
50
0
0
200
400


250
200
150
100
50
0
0
200
400
600
800
1000

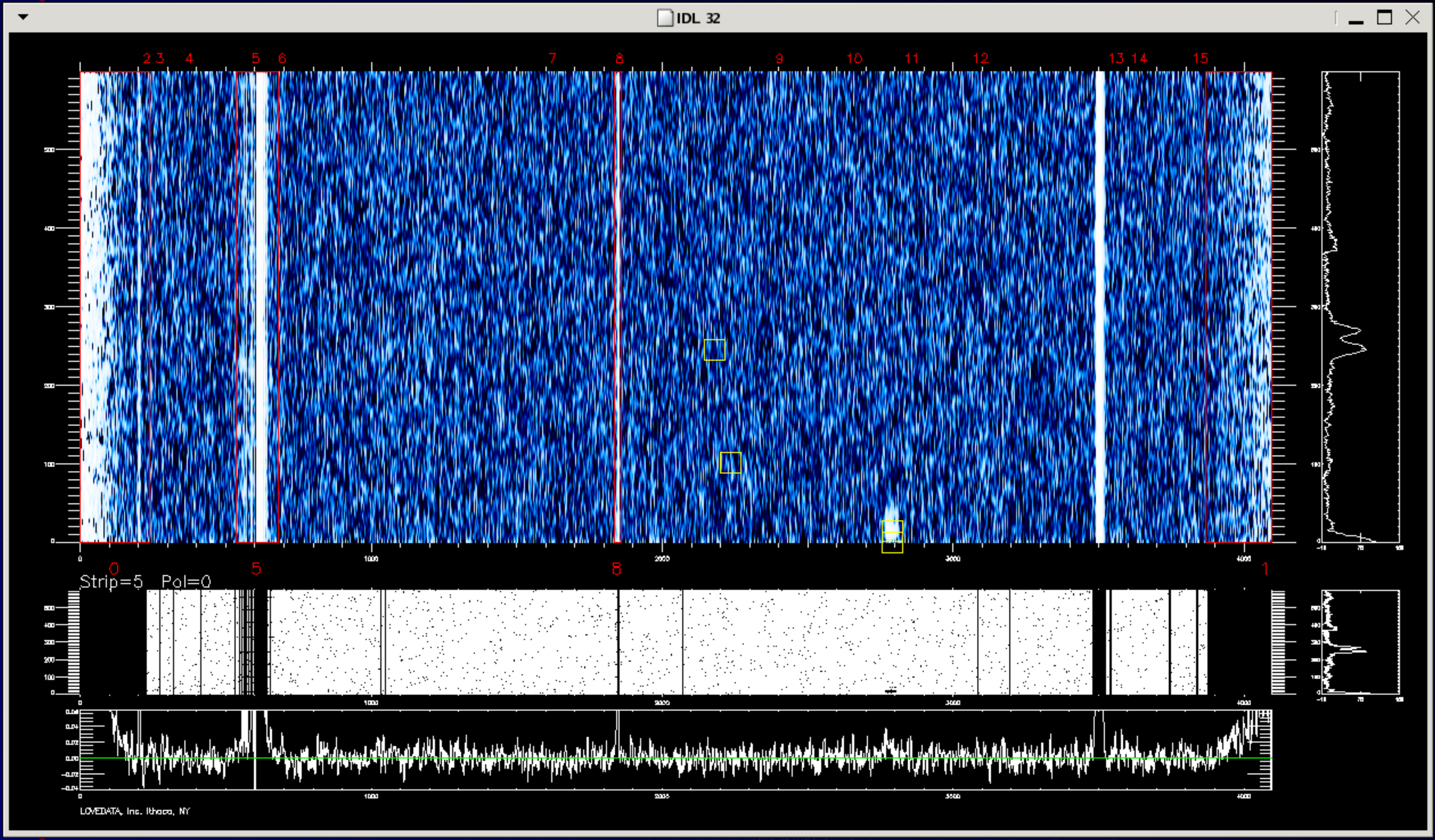
ALFALFA

Bandpass correction (BPD)

- Bandpass correction is done in the RA vs. frequency domain for each beam/polarization map.
- An entire observing session is processed at once.
- Interaction with user takes place through an easy GUI.
- Once parameters are set, processing runs as a background job.
- It takes about 0.3 hrs of processing per hour of data taking.

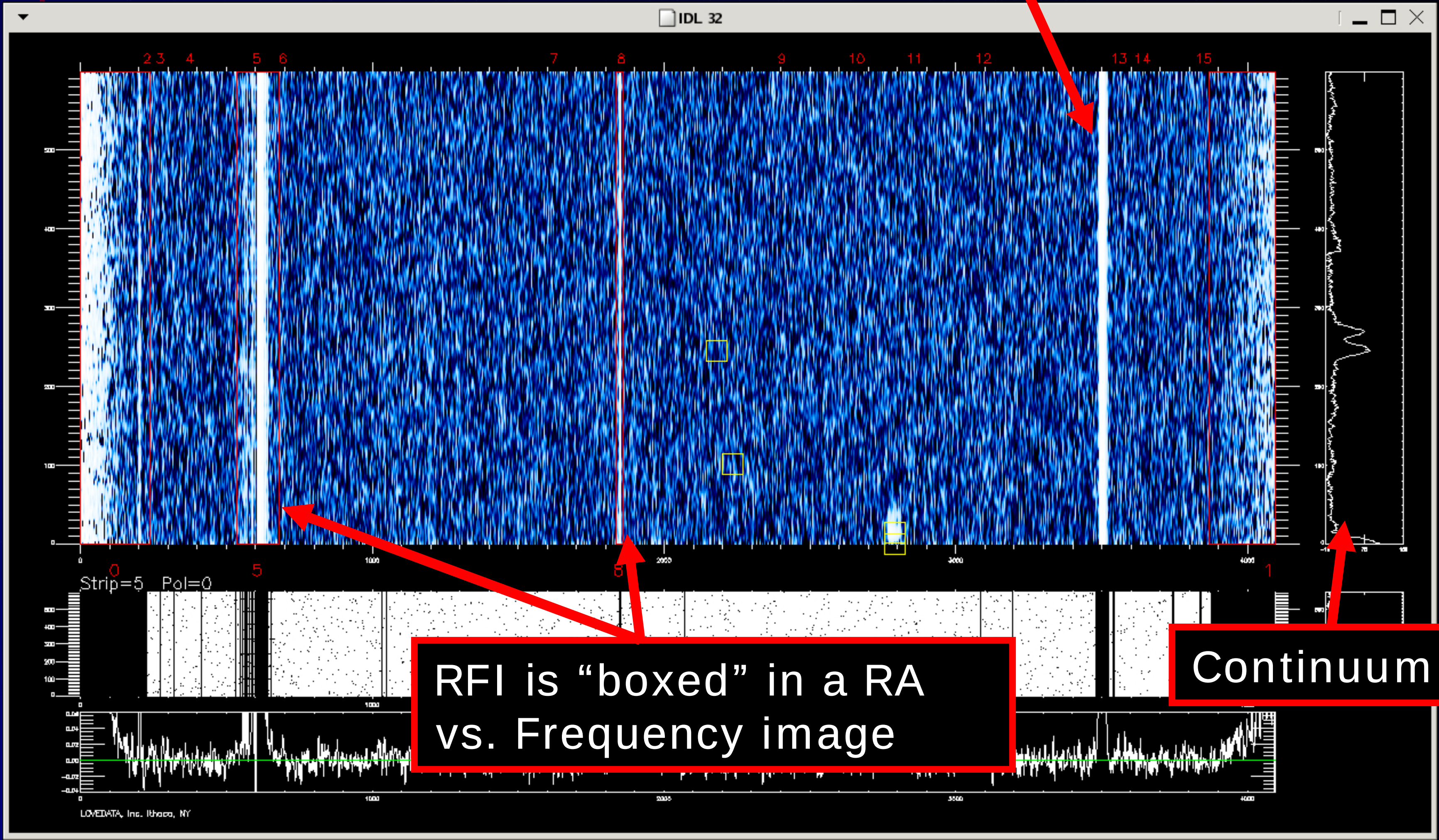


RFI-flagging: FLAGBB



ALFALFA

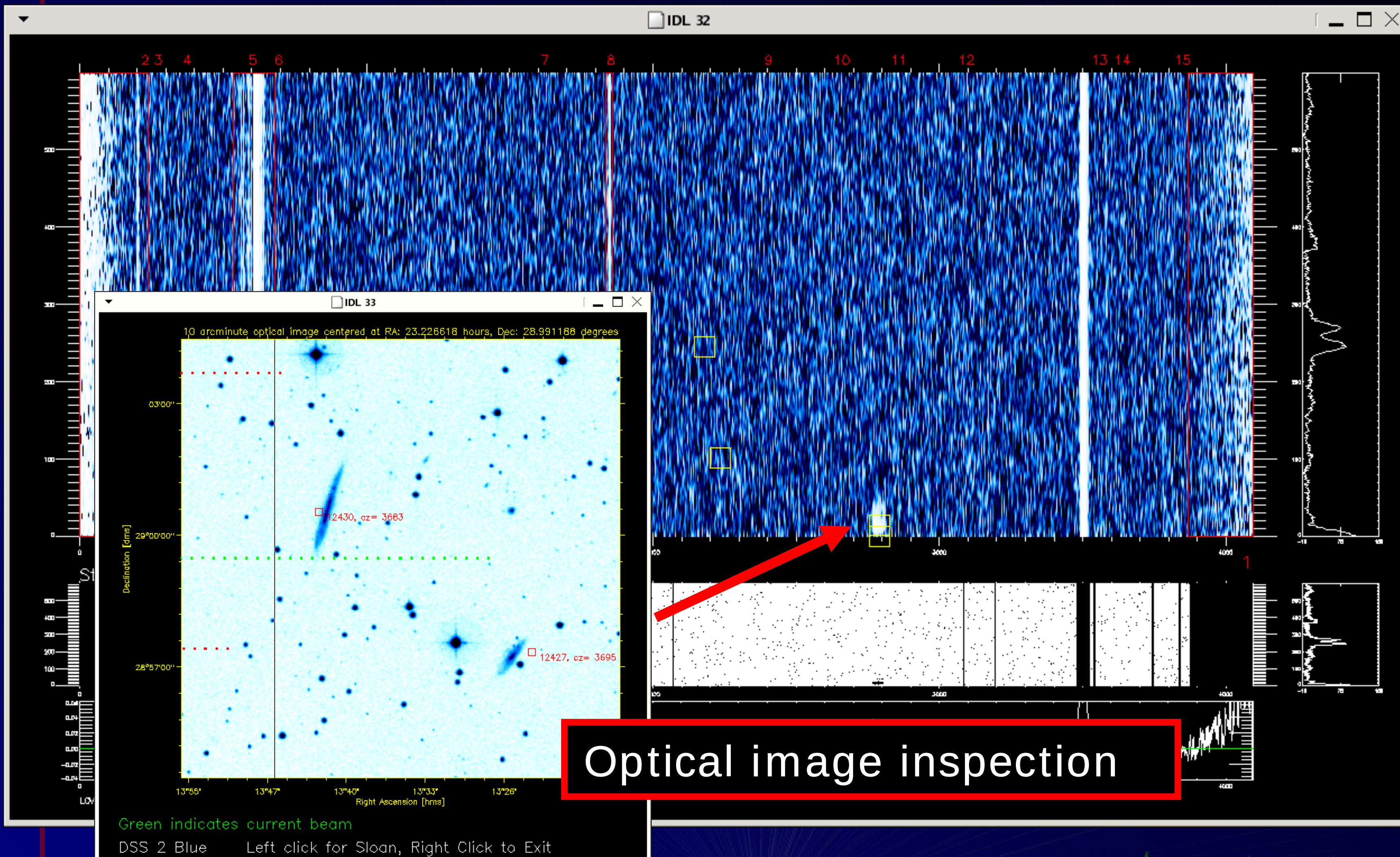
Galactic Hydrogen



RFI is “boxed” in a RA vs. Frequency image

Continuum

ALFALFA



Optical image inspection

ALFALFA

Data/RFI flagging/QA management

- Data from each observing session are “checked out” by users for bandpass calibration and RFI-flagging.
- Data quality assessment of each session is made by a senior team member.

ALFALFA LOG - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://caborojo.astro.cornell.edu/alfalfalog/archive.php

Yahoo! Mail - The best... Google SummerSchool2 - Mai... ALFALFA Team Website IDL Virtual Observator...

									0950-1002
06.01.16	63p1 +151230	0703-1105	dor13/idlraw	dor10/flagbb	ss	23	1.5	06Feb	
06.01.17	36p1 +083818	0541-1056	dor13/idlraw	dor10/flagbb	ss	31	1.2	06Feb	Early start; 'scope fre e; 3 min WAPP g
06.01.18	39p1 +092206	0723-1055	dor13/idlraw	dor10/flagbb	tb	21		06Feb	
06.01.19	40p2 +094400	0725-1107	dor13/idlraw	dor10/flagbb	tb	21		06Feb	WAPP gap 0745-0755; pow er **not** adju

February 2006

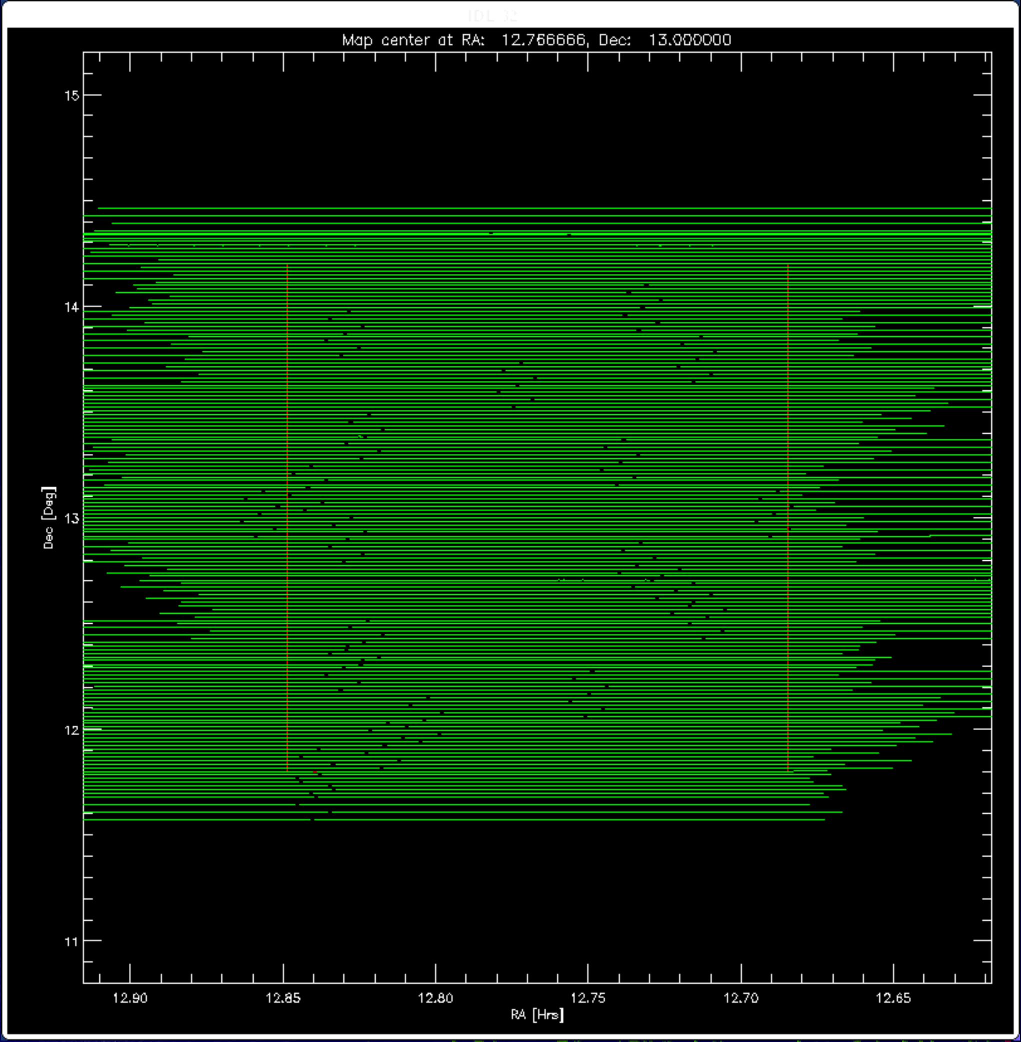
Block	DecJ	RA	Raw files disk	Level I	Who	#	QA	Backup	Notes
06.02.15	62p1 +145754 22p2 +052112	0728-1302 1305-1557	dor13/idlraw	toro2/flagbb	sb	18			2 decs; WAPP gap 1222-1 232 2 cal perio
06.02.16	57p2 +135212 59p2 +142124	0737-1139 1144-1546	dor13/idlraw	toro2/flagbb	wm	48			2 decs
06.02.17	58p2 +140638 21p2 +050636	0730-1300 1308-1550	dor13/idlraw	toro2/flagbb	mh	49			2 decs; some HVC/or/rfi ?
06.02.18	59p1 +141406 23p2 +053548	0732-1305 1313-1554	dor13/idlraw	toro2/flagbb	af	49			2 decs
06.02.22	37p1 +085254	0732-1124	dor13/idlraw	toro2/flagbb	af	23			
06.02.23	64p1 +152706	0740-1132	dor13/idlraw	toro2/flagbb	mh	23			
06.02.24	65p1 +154142	0736-1139	dor13/idlraw	toro2/flagbb	mh	24			scruffy rfi around galH I
06.02.25	53p1 +124630 45p2 +105700	0720-1324 1332-1645	dor13/idlraw	toro2/flagbb	rg	55	1.3		
06.02.26	65p2 +154900	0726-1645	dor13/idlraw						short gap 1501-1504
06.02.27	66p2 +160336	0734-1640	dor13/idlraw						
06.02.28	63p2 +151948	0744-1630	dor13/idlraw		(mh)				

Done

ALFALFA

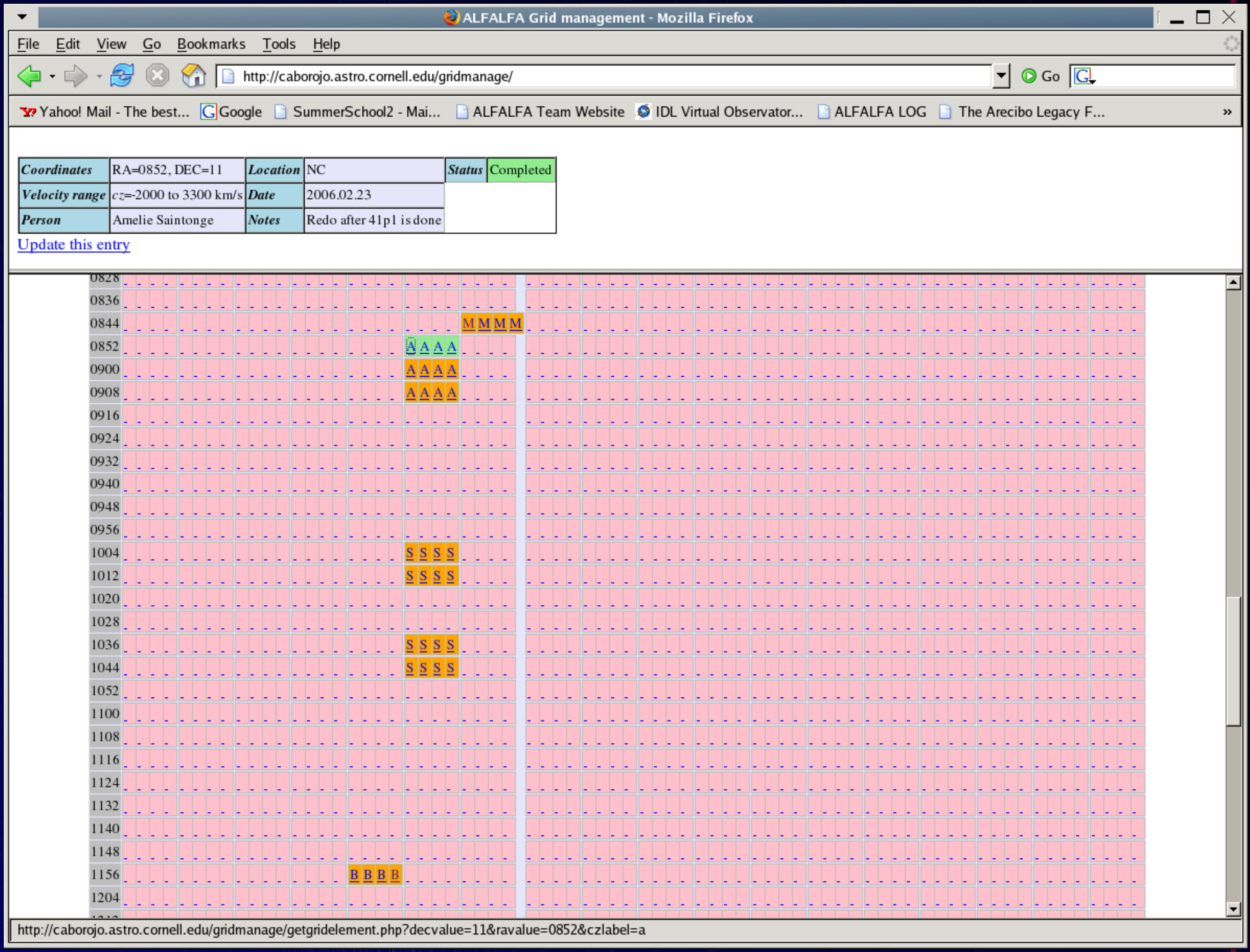
Gridding

- 2.4° x 2.4° x 5400 km/s data cubes (grids) are created via:
  - Examining "pos" structures maintained in a "masterpos"
  - For every grid point, a record is kept that describes which record, from which scan, which beam, which pol, does contribute to spectrum at that point
  - An array of "weights" is carried for each spectral value of the grid.



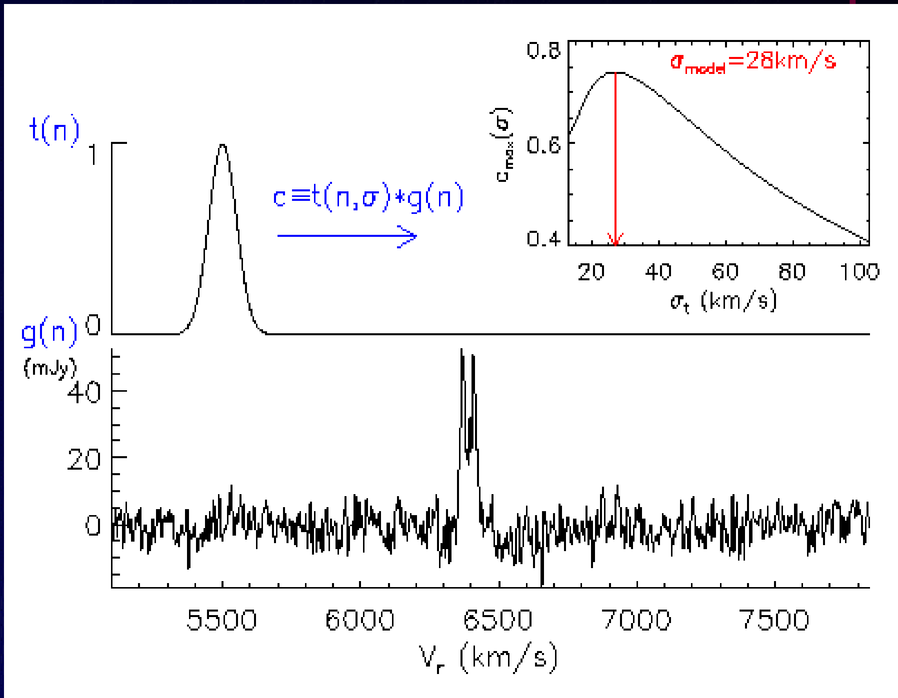
Grid management

- Grids are logged via a web interface with current status, ALFALFA user, completion dates, and file comments



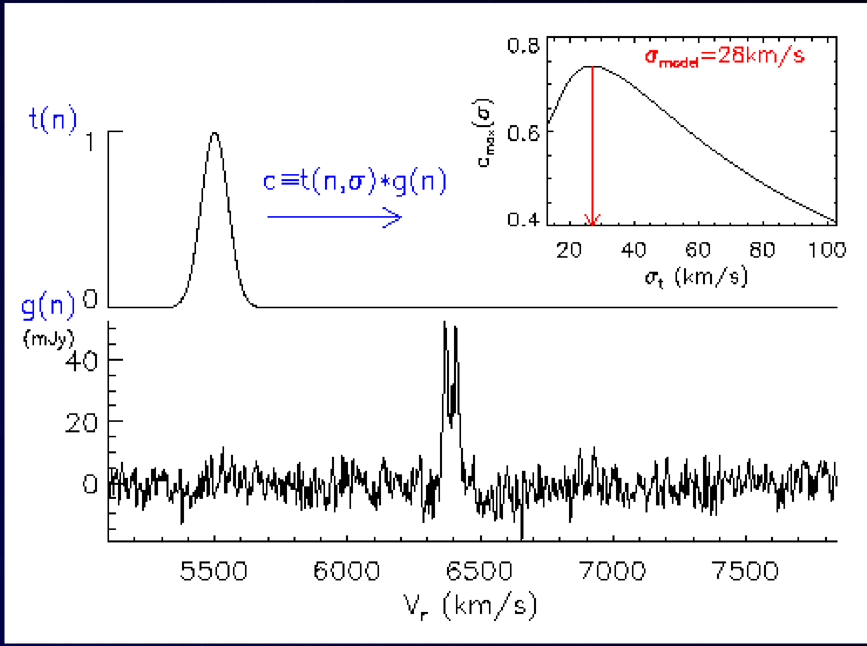
Signal Extraction

- The signals are extracted by cross-correlations of a template with the spectra (algorithm written by Amélie Saintonge)
  - More sensitive than peak-finding algorithms.
    - sensitive to total flux, not only peak flux
    - especially important for low mass systems
  - Using FFT's, cross-correlations are fast
  - A matched-filter algorithm



Signal Extraction

- The process is :
  - Repeat for a range of widths of the template
    - e.g. 10 km/s - 600km/s
  - Choose the width for which the convolution is maximized --> position of the signal
  - Calculate the amplitude of the signal from the width

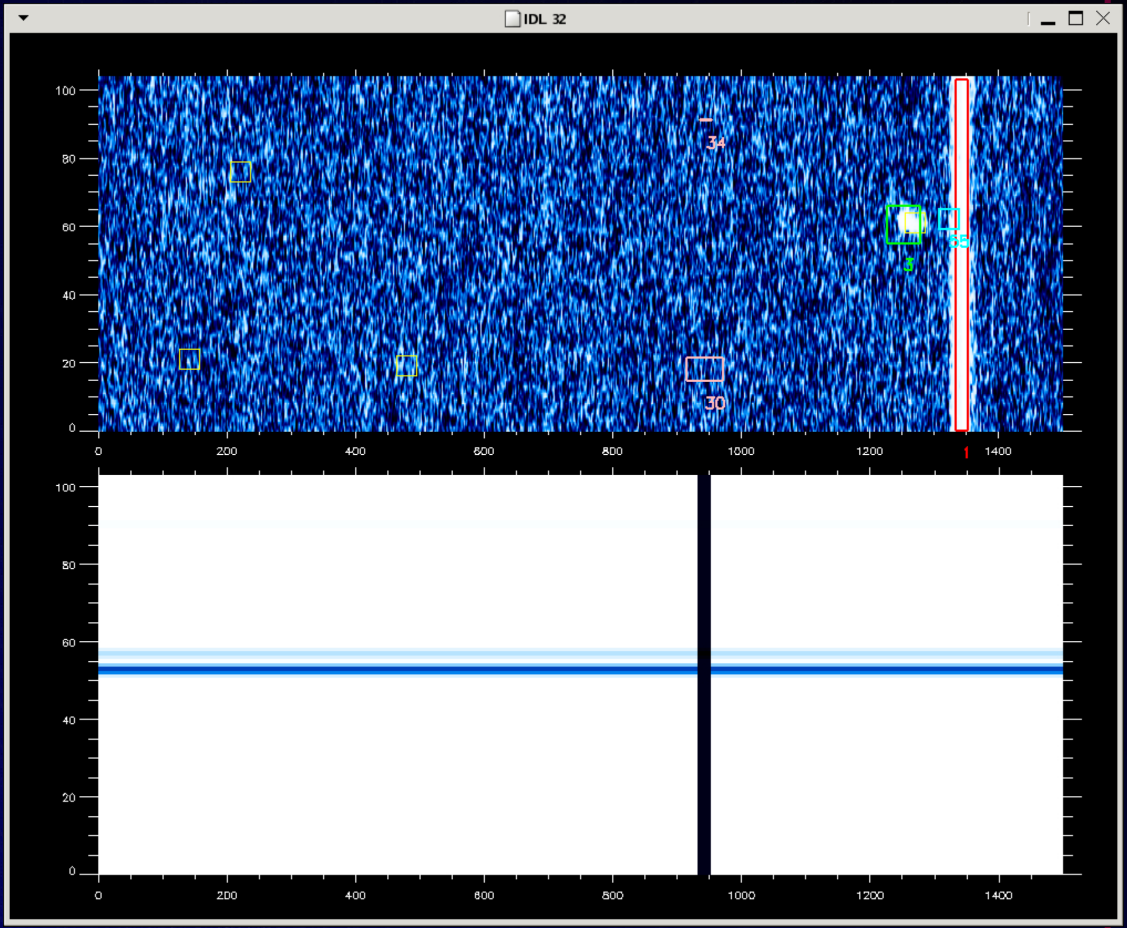


- Run on 2-D datasets -- for the 2 passes separately
- Compare catalogs; choose most probable candidates
- Run on 3-D datacubes after completion of "tile"
- Then compare catalogs --- allows digging lower into S/N

Slide: Amelie Saintonge

Signal Extraction

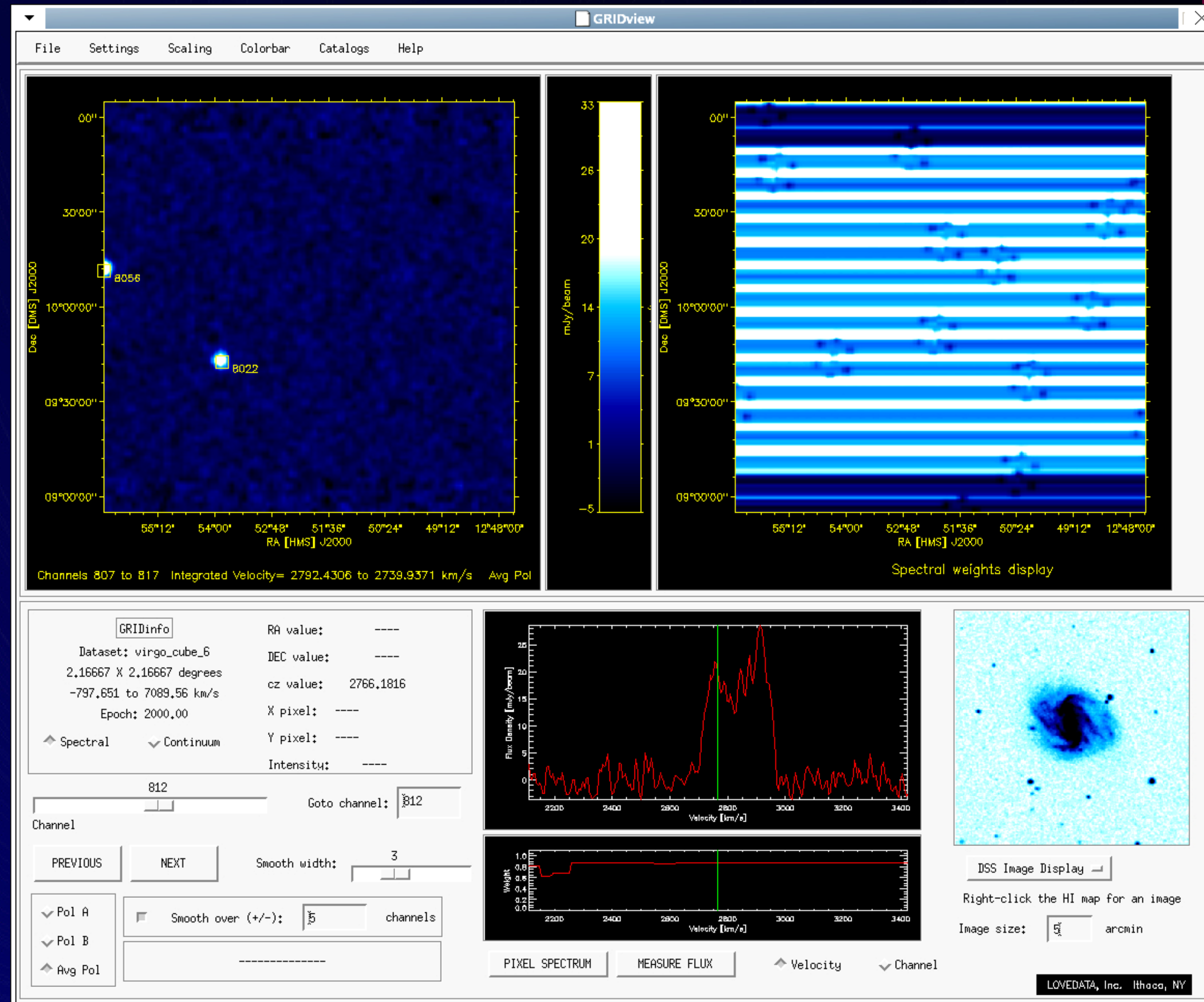
- Matched filter in the FFT domain is used for signal extraction with a specified S/N threshold level.
- Interactive program allows user to examine source detections – catalogs and preliminary source parameters are produced.



ALFALFA

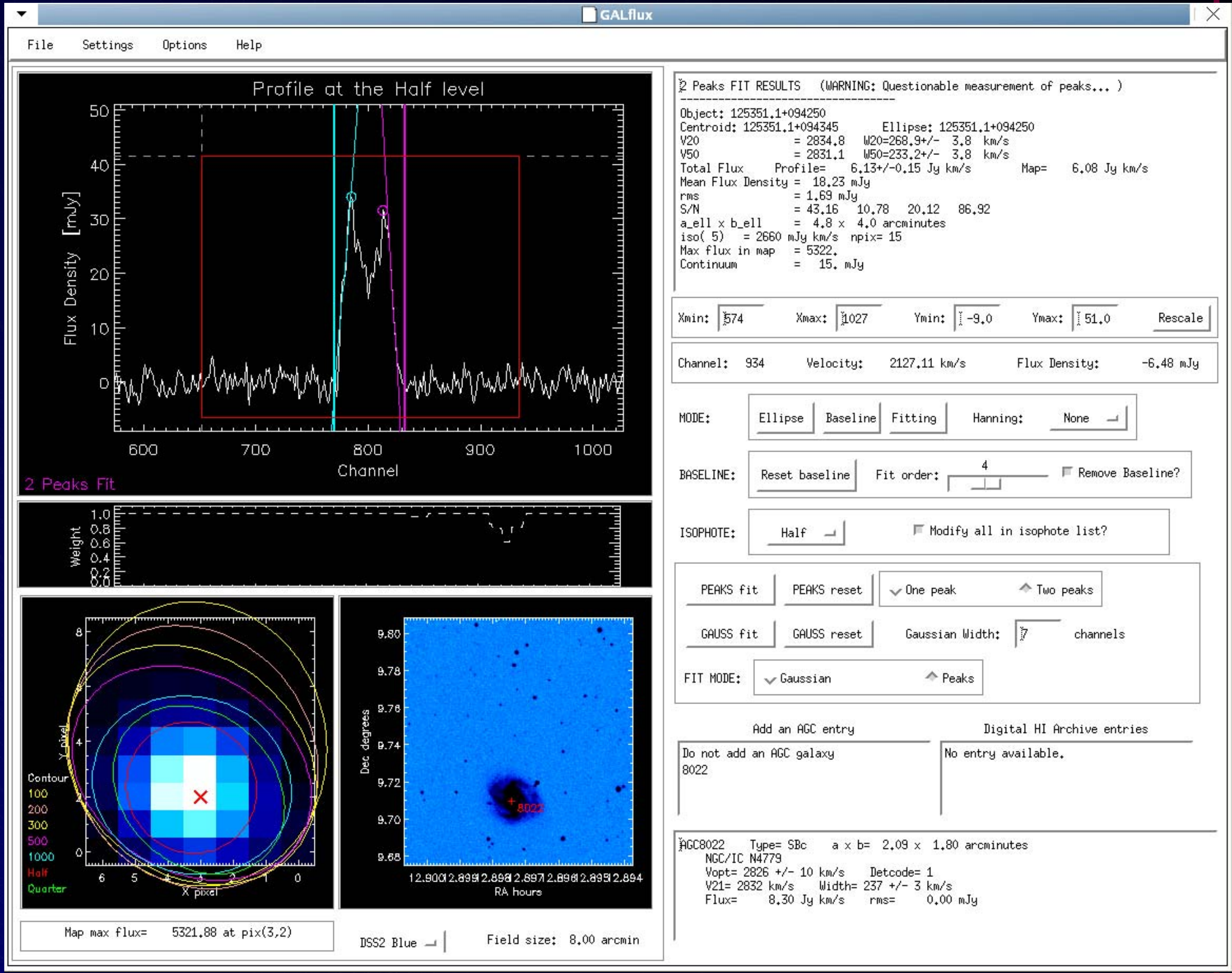
Data cube visualization

- Data cubes and corresponding 3D catalogs are examined in GRIDview.
- The upper left display is a channel map; at upper right is the corresponding weights map.
- Controls allow user to view channel or integrated maps at different velocities.
- DSS, DSS2, Sloan, NVSS images can be fetched.
- NED and other on-line catalogs – including internal ones – can be accessed and overplotted



Flux Measurement

- Centroid positions are determined
- Ellipse parameters are calculated.
- Integrated profiles are created – measurements are recorded in src (source) structures
- Data are compared with database archives.



ALFALFA

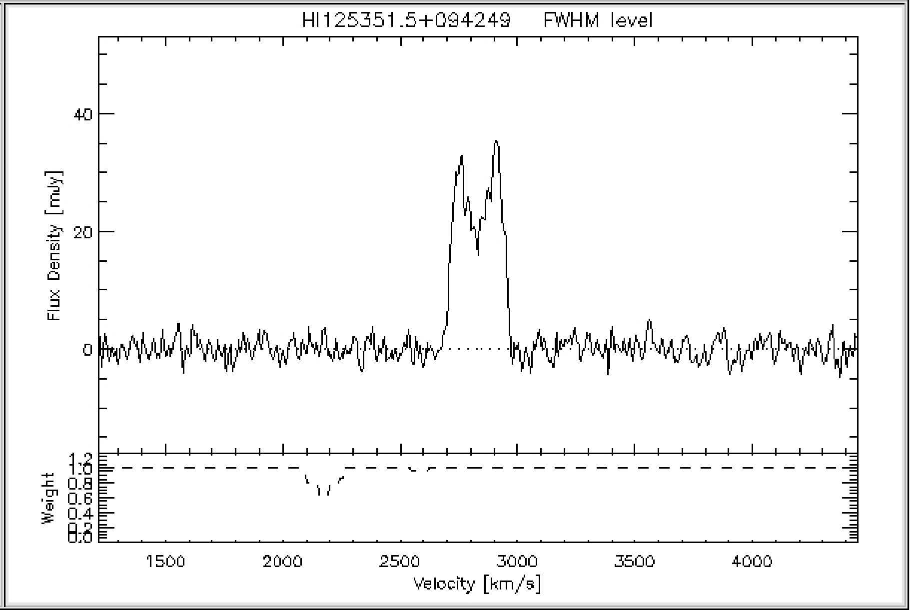
ALFALFA Catalog Object: HI125351.5+094249 - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

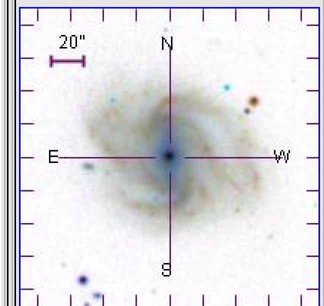
file:///home/caborojo1/bkent/idl/galflux/html/HI125351.5+094249_5.html

Yahoo! Mail - The best... Google SummerSchool2 - Mai... ALFALFA Team Website IDL Virtual Observator... ALFALFA LOG

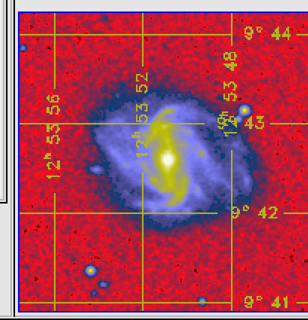
ISO 0
ISO 1
ISO 2
ISO 3
ISO 4
ISO 5
ISO 6



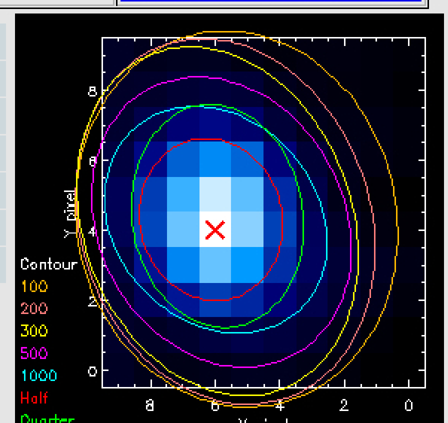
SDSS



Digital Sky Survey Blue 2



V50,W50	2831.0 233.6+/- 3.8 km/s	Centroid	125351.3+094435 [2000]
V20,W20	2834.5 269.5+/- 3.8 km/s	Opt pos	125350.8+094235 [2000]
Vcen	2832.2+/- 1.9 km/s	Cen_ell	125351.5+094249 [2000]
V,W Gauss	0.0 0.0+/- 0.0 km/s	Ellipse	4.7 x 4.4 PA=-119.
Stot(profile, P)	6.41+/- 0.16 Jy km/s	Isophote	2621. mJy km/s
Stot(profile, G)	0.00+/- 0.00 Jy km/s	Map Smax	5243. mJy km/s
rms	1.77 mJy	Map Stot	6.19+/- 0.00 Jy km/s
meanS, peakS	20.0 35.4 mJy		
S/N P	43.0 11.3 20.0 89.0		
S/N G	0.0 0.0 0.0 0.0		
Cont	15. mJy		



Done