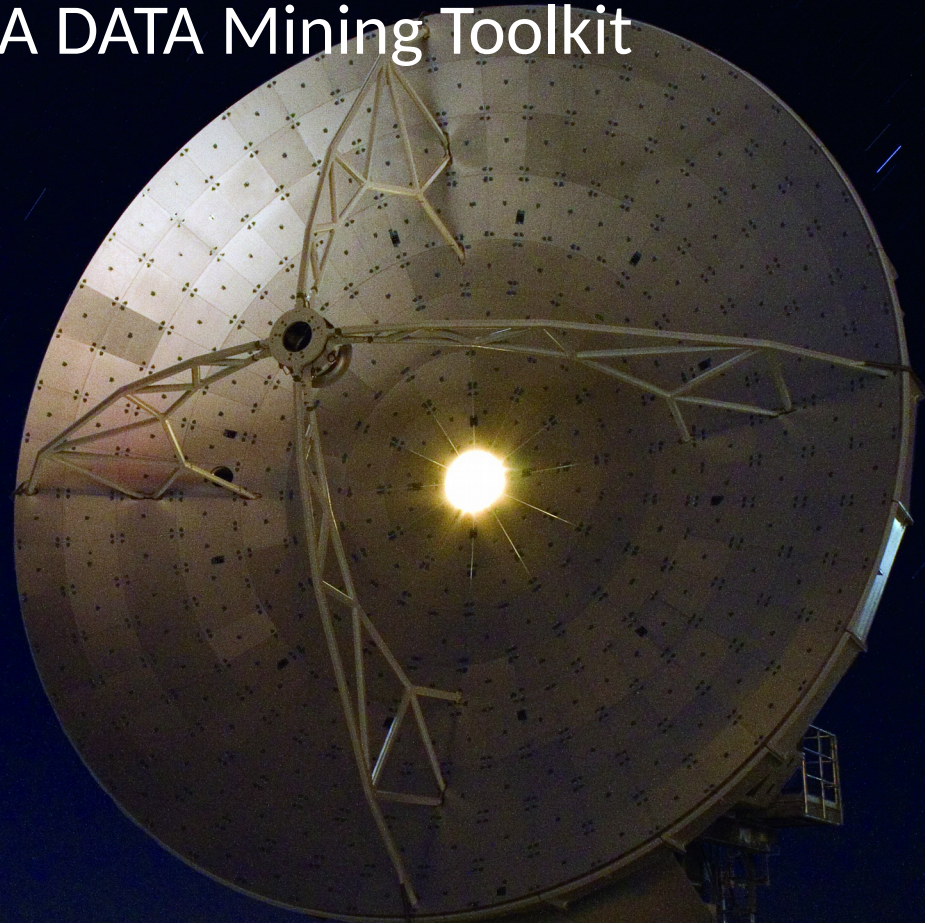


Getting the most out of your ALMA DATA with ADMIT: The ALMA DATA Mining Toolkit

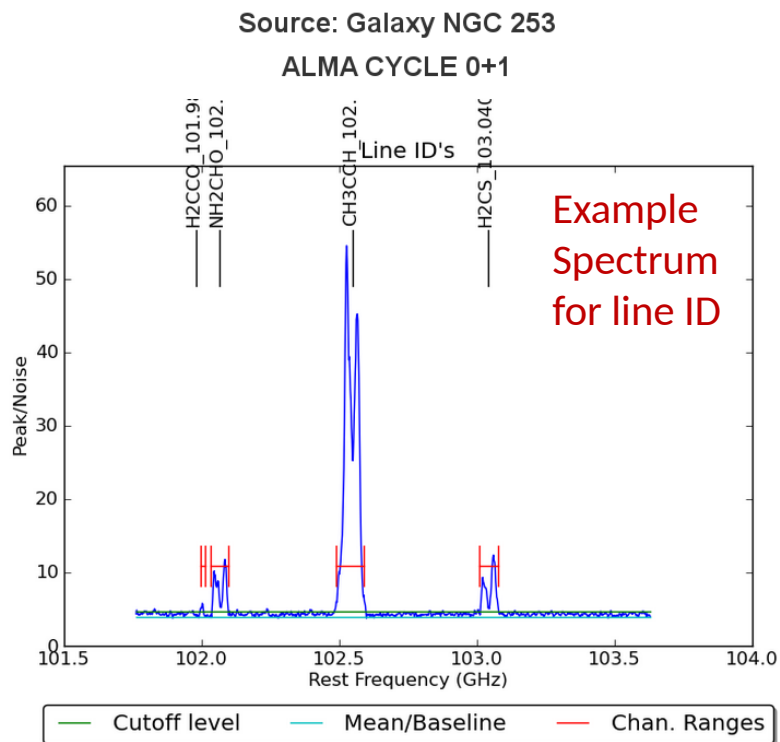


Doug Friedel, Leslie Looney, Peter Teuben, Marc Pound,
Kevin Rauch, Lee Mundy, Robert Harris, & Lisa Xu

ADMIT: ALMA Data Mining Toolkit

- ◆ Developed by University of Maryland, University of Illinois, and NRAO (PI: L. Mundy)
- **Goal #1:** First-view science data products into archive: spectra, line ID, moment maps, etc
- **Goal #2:** Python Toolkit allows user to generate their own science products from cubes.

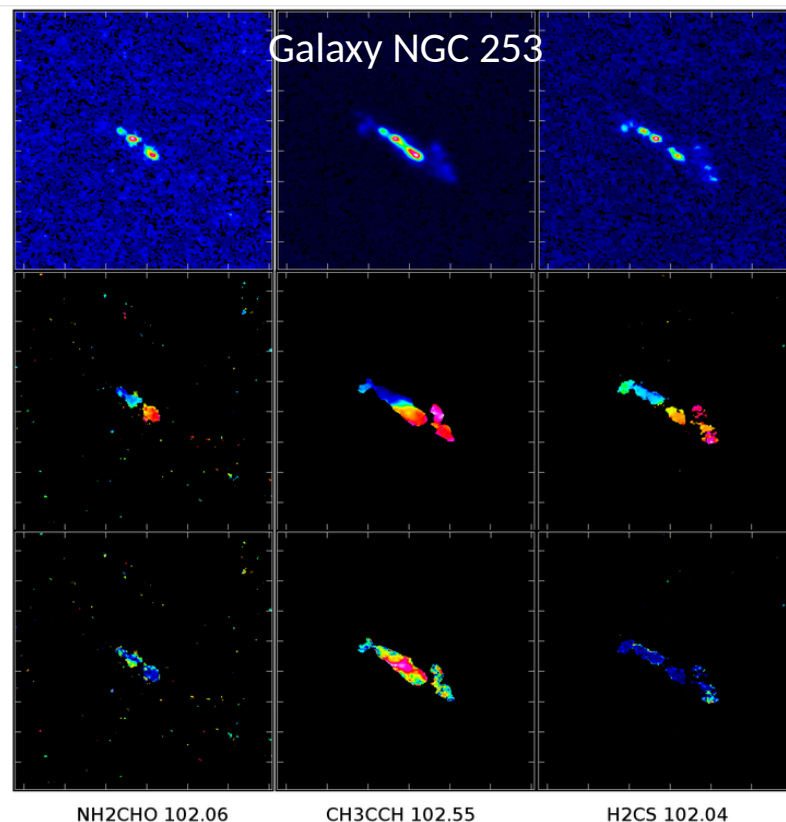
ADMIT Products shown from automated flow



Moment 0

Moment 1

Moment 2



Identification of common lines

Identified lines

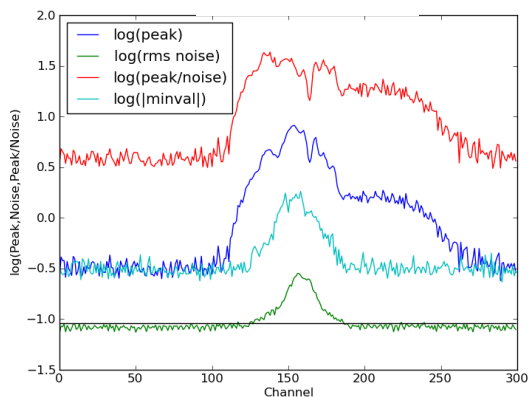
Moment maps for lines

frequency	formula	transition	velocity	fwhm	startchan	endchan
101.98143	H2CCO	5(1,4)-4(1,3)	2.938E+02	2.413E+01	828	836
102.06427	NH2CHO	5(1,5)-4(1,4)	2.382E+02	2.354E+02	783	816
102.54798	CH3CCHv=0	6(0)-5(0)	2.300E+02	2.343E+02	532	583
103.04055	H2CS	3(0,3)-2(0,2)	2.359E+02	2.332E+02	283	318

ADMIT

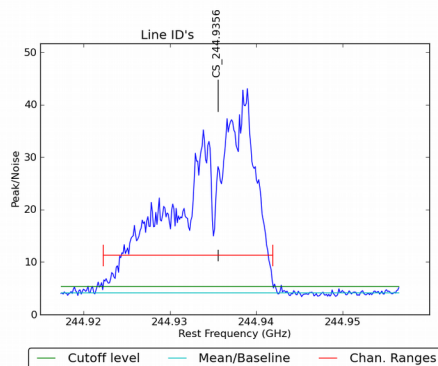
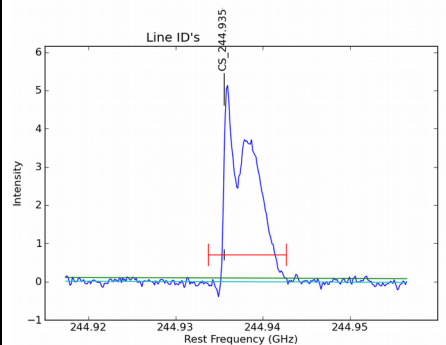
- ❖ Operates on data cubes only; cubes can be FITS or casa image format
- ❖ Compatible with CASA environment and utilizes CASA routines where possible
- ❖ Products are self-documenting with XML; compatible with future ingestion by a database

Each data cube get a full set of products
See: admit.astro.umd.edu/admit-M4 and
click on an xxx.admit directory



Spectra based on
peak flux and noise
in each channel

Blue and green
spectra highlight
impact of missing
flux



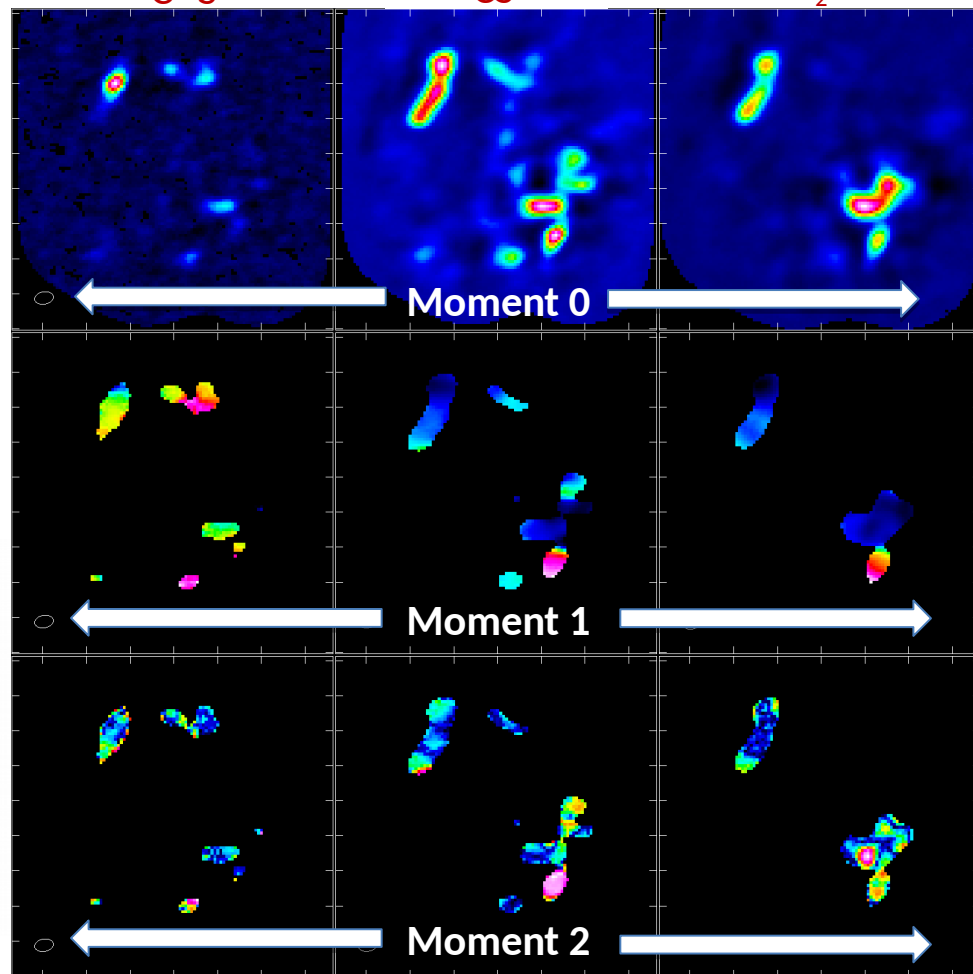
Two spectra used in line ID of CS J=5-4

Serpens Main Mosaic Image

$C^{34}S$

CS

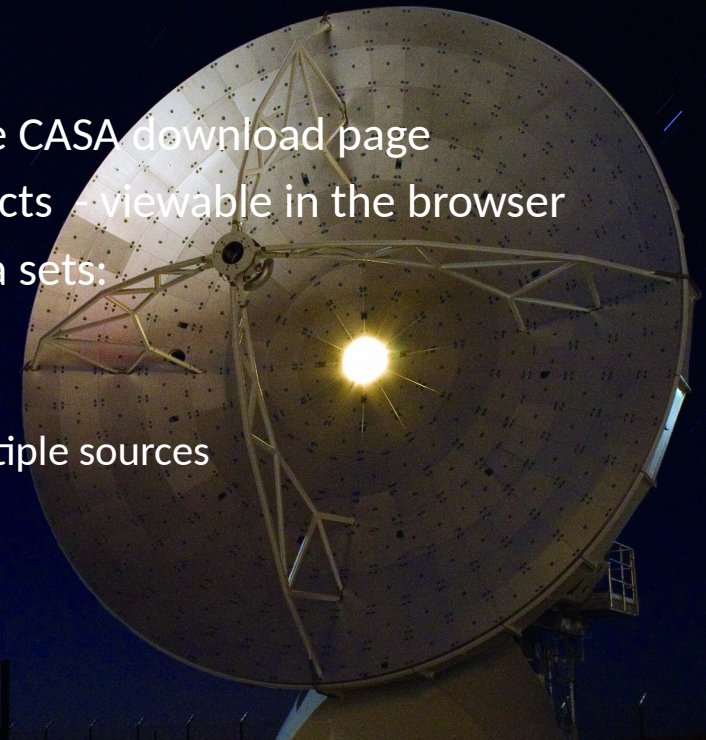
H_2CO



ADMIT

Two modes of Operation:

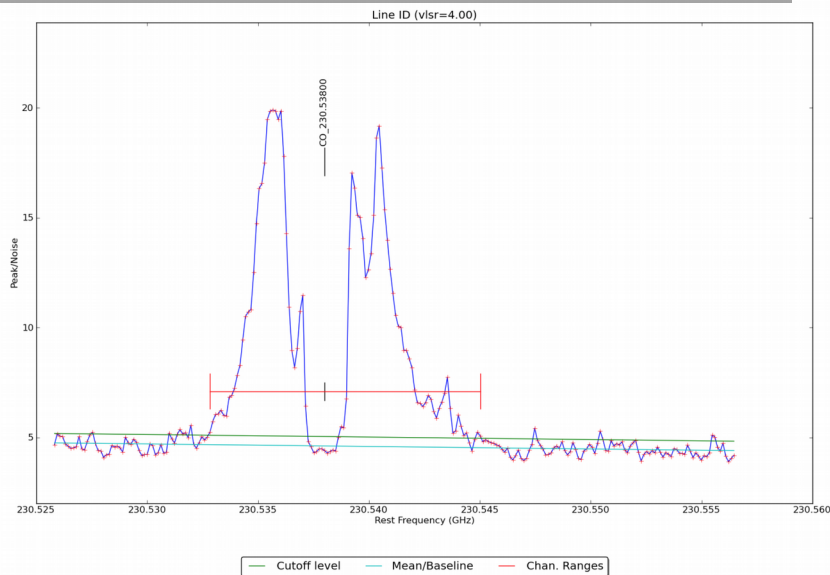
- **On-line (pipeline mode producing standard set of products):**
 - ADMIT runs after QA2 and before archive ingestion (ideally as a pre-step to the archive ingest process) – details being worked with ALMA Project
 - ALMA archive user can select to download ADMIT tarball (20-40MB)
 - XML, PNG, and HTML files; limited FITS files – details to be decided with ALMA Project
 - Browser-based view allows user to inspect products once downloaded
- **Off-line (user created data products):**
 - The ADMIT Toolkit “add-on” available from the CASA download page
 - Flow-model for creating and re-creating products – viewable in the browser
 - Environment for expanded exploration of data sets:
 - Principle component analysis of emission
 - Overlap integrals
 - Comparisons across multiple windows and multiple sources
 - New tools for examining large data cubes
 - Fine tune line ID



ADMIT

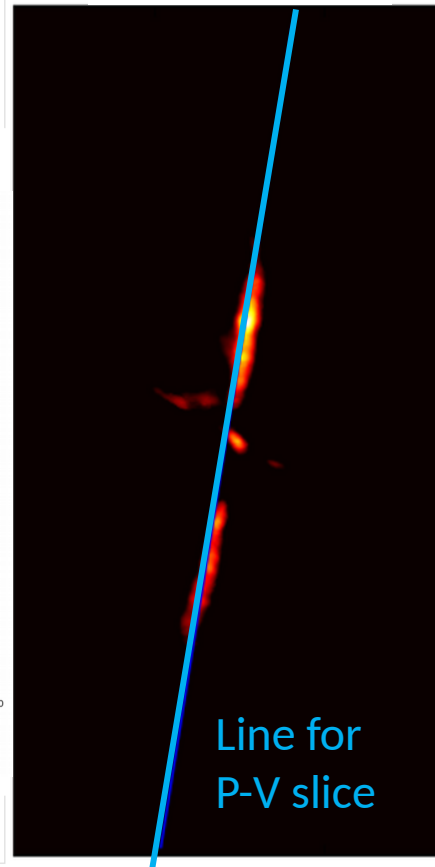
- ❖ Automated line ID allows line-based operations: moment maps, PV slices, etc
- ❖ Flow produces set products for users, which can be determine by ALMA
- ❖ Users can create their own custom products locally, which can be applied across sources

Conservative automated line ID:
do no bad ID

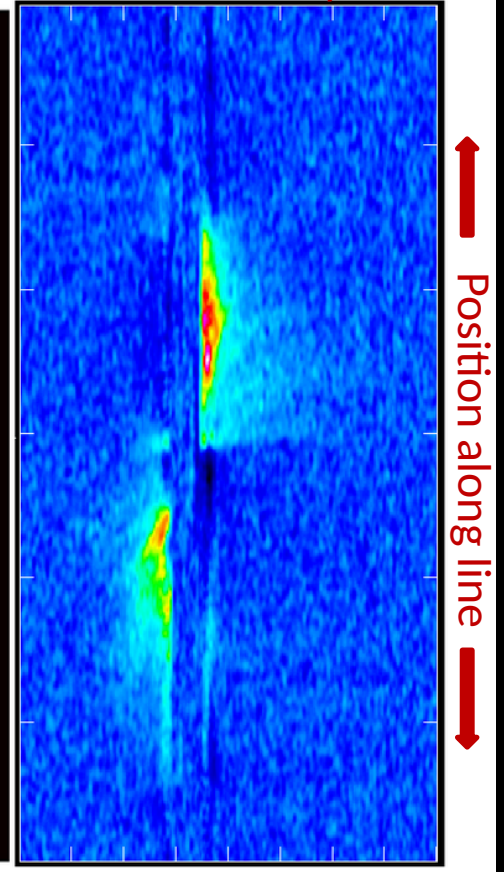


Spectrum: peak emission over noise
per channel

Moment 0



Position-Velocity Slice



Identified lines

frequency	formula	transition	velocity	fwhm	startchan	endchan
230.53800	CO	2-1	1.200E+01	3.503E+00	100	155

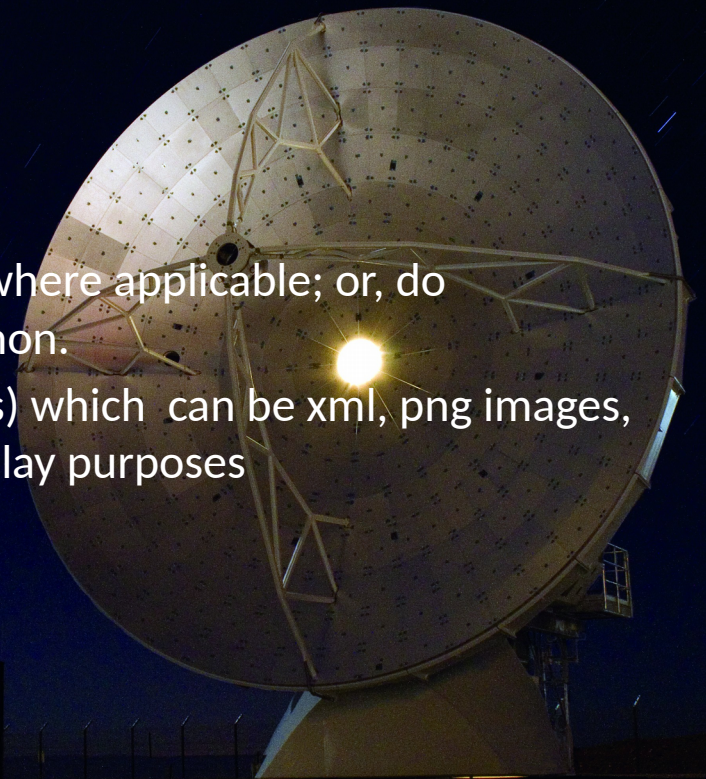
ADMIT

Flow Manager

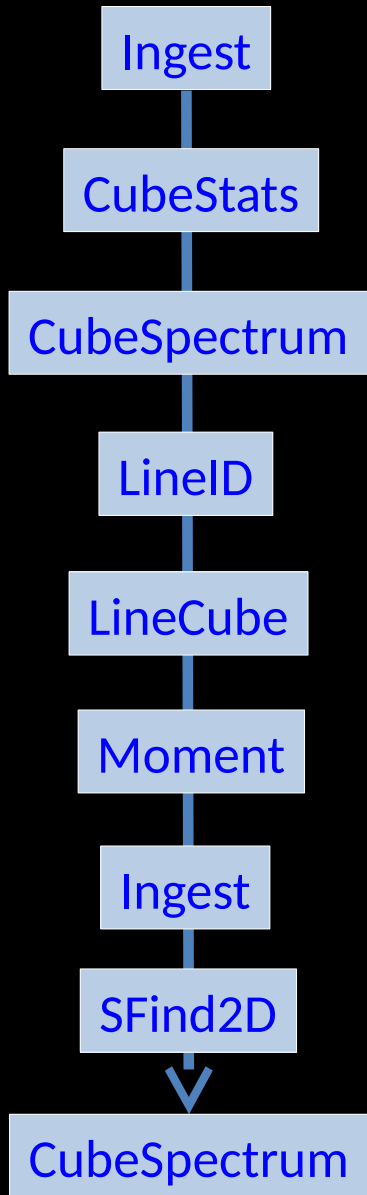
- Allows creation of sequences of ADMIT Tasks that can be run and re-run (shared) in CASA python environment – can be written out as python scripts
- Keeps a record of sequence of Tasks and products in admit.xml file

ADMIT Tasks (ATs):

- ATs are python scripts that call CASA tasks or tools where applicable; or, do appropriate calculations where needed in pure python.
- The output of tasks are “Basic Data Products” (BDPs) which can be xml, png images, and FITS files with documentation and html for display purposes
- 17 currently existing tasks



AT Flow



ADMIT

Read in full window data cube: from FITS or CASA image
Output: CASA image and xml information

Calculates statistics of data cube: RMS, Min, Max per channel, etc
Output: xml table, png's

Makes Spectra which characterize the emission -used for LineID
Output: xml table, png's

Identify lines present in data cube: where, and which transitions
Output: xml table and png's

Creates separate cubes for each line found with transition or freq labeling
Output: "N" CASA images with xml information

Creates clipped moment maps for each line (0, 1, 2... as requested)
Output: CASA images, png's, xml information

Read in continuum map
Output: CASA image and xml information

Find continuum sources to some selected depth
Output: xml table, png's

Make Spectra at each continuum position
Output: xml table, png's

ADMIT

Data Product Viewer

- Browser-based with style files similar to the ALMA calibration pipeline.
- The browser view is self-generated as ATs are run.
- The browser view is started by pointing the browser to the admit directory.

ADMIT Output for test22.admit

Flow View Form View LineID Editor Log file ADMIT documentation

TASKS LISTED IN ID NUMBER ORDER - CLICK ON TASK NAME TO SEE ITS OUTPUTS. COLORS AND ICONS INDICATE THE STATUS OF EACH TASK:
✓ TASK RAN NORMALLY (GREEN) | ▲ TASK NEEDS TO BE RE-RUN (ORANGE) | ⚙ TASK IS DISABLED (PINK)

Flow Diagram for test22.admit

- Ingest_AT (taskid=0) file=x.fits ✓
- CubeStats_AT (taskid=1) robust=medabsdevmed ppp=True ✓
- CubeSum_AT (taskid=2) numsigma=4.0 sigma=0.0912921 smooth=[] ✓
- SFind2D_AT (taskid=3) nsigma= 6.0 sigma=0.675685 region= robust=['hin', 1.5] snmax= 35.0 ✓
- CubeSpectrum_AT (taskid=4) pos=[[104, 62], (71, 111)] x.im ✓

CubeSpectrum_AT computes spectra at one or more (x,y) positions in the datacube. Alternatively, it can produce a spectrum averaged over a box.

CubeSpectrum_AT output for image x.im

Average Spectrum at centerbox[[104pix,62pix],[1pix,1pix]]

Average Spectrum at centerbox[[71pix,111pix],[1pix,1pix]]

- LineSegment_AT (taskid=5) numsigma=5.0 minchan=4 maxgap=3 segment=ADMIT smooth=[] ✓
- PVSlice_AT (taskid=6) slice=['142.80', '6.00', '28.60', '173.00'] width=5 ✓
- PVCorr_AT (taskid=7) numsigma=3.0 range=[112,256] ✓
- LineID_AT (taskid=8) numsigma=5.0 minchan=4 maxgap=3 recomb=shallow smooth=[] tier1width=0.0 csub=[0, 0] iterate=True ✓
- LineCube_AT (taskid=9) pad=5 equalize=False ✓
- Moment_AT (taskid=10) moments=[0, 1, 2] numsigma=[2.0] momOclip=2 chans=all x.CS_244.93556 ✓
- CubeSpectrum_AT (taskid=11) pos=[[71, 111)] x.CS_244.93556 ✓
- Moment_AT (taskid=12) moments=[0] numsigma=[3.0] momOclip=2 chans=all x-@1.mom ✓

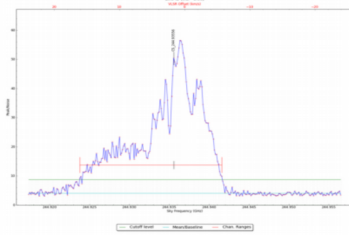
ADMIT

Data Product Viewer

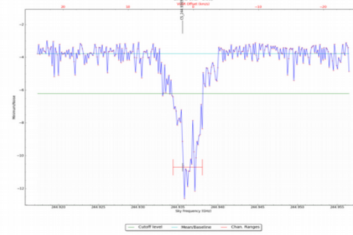
– Viewing specific task results

ADMIT Output for test22.admit

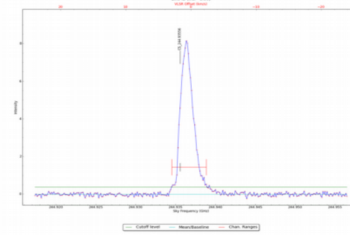
frequency	uid	formula	name	transition	velocity	El	Eu	linestrength	peakintensity	peakoffset	fwhm	startchan	endchan	peakrms	blend	force
[GHz]					[km/s]	[K]	[K]	[D^2]	[Jy/beam]	[km/s]	[km/s]					
244.93556	CS_244.93556	CS	Carbon Monosulfide	5-4	0.3196	23.511	35.266	19.1688	49.5385	-7.6804	16.1564	114	250	52.4	0	False



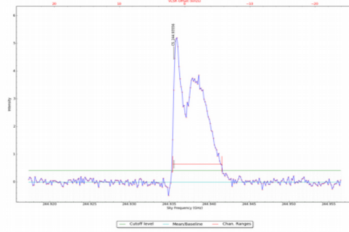
Potential lines overlaid on peak intensity plot from CubeStats_BDP.



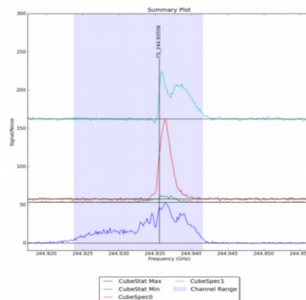
Potential lines overlaid on minimum intensity plot from CubeStats_BDP.



Identified lines overlaid on input spectrum #0.



Identified lines overlaid on input spectrum #1.



Identified lines overlaid on Signal/Noise plot of all spectra.

ADMIT

Data Product Viewer

- Form viewer allows user to change input parameters and re-run

ADMIT Task Inputs for test22.admit

Flow View **Form View** LineID Editor Log File ADMIT documentation

TASKS LISTED IN EXECUTION ORDER - CLICK ON TASK NAME TO SEE ITS OUTPUTS. COLORS AND ICONS INDICATE THE STATUS OF EACH TASK:
✓ TASK RAN NORMALLY (GREEN) | ▲ TASK NEEDS TO BE RE-RUN (ORANGE) | ⚙ TASK IS DISABLED (PINK)

- Ingest_AT (taskid=0) ✓
- CubeStats_AT (taskid=1) ✓
- CubeSum_AT (taskid=2) ✓
- SFind2D_AT (taskid=3) ✓
- CubeSpectrum_AT (taskid=4) ✓

[On-line documentation of CubeSpectrum_AT](#)

Parameters

Input BDPs

Index	Type	File name
0	SpwCube_BDP	x.im
1	CubeStats_BDP	x.cst
2	Moment_BDP	x.csm
3	None	None

Output BDPs

Index	Type	File name
0	CubeSpectrum_BDP	x.csp

Keywords

pos

xaxis

- LineSegment_AT (taskid=5) ✓
- PVSlice_AT (taskid=6) ✓
- PVCorr_AT (taskid=7) ✓

ADMIT

Data Product Viewer

- LineID editor allows user to change the line identifications, channels, etc.

ADMIT LineID Editor for test22.admit

Flow View Form View **LineID Editor** ADMIT Log ADMIT documentation

Potential lines overlaid on peak intensity plot from CubeStats_BDP.

Potential lines overlaid on minimum intensity plot from CubeStats_BDP.

Identified lines overlaid on input spectrum #0.

Identified lines overlaid on input spectrum #1.

Identified lines overlaid on Signal/Noise plot of all spectra.

accept	force	reject	frequency	uid	formula	name	transition	velocity	startchan	endchan	blend
			[GHz]					[km/s]			
<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	244.93556	CS_244.93556	CS	Carbon Monosulfide	5-4	0.3196	114	250	0

Reset Add Line Set LineID Force/Reject Keywords Overwrite LineID Output BDP

ADMIT

Data Product Viewer

- Log File view allows the user to browse the log output, which contains more details than is available through the Flow View page.

```
ADMIT Log for test22.admit
Flow View Form View LineID Editor Log file ADMIT documentation
INFO : AT.py : Setting {'basename': 'x', 'file': 'test22.fits'} for Ingest_AT
INFO : AT.py : Setting 'ppp' = True for CubeStats AT
INFO : AT.py : Setting 'numsigma' = 4.0 for CubeSum AT
INFO : AT.py : Setting 'sigma' = 99.0 for CubeSum AT
INFO : Admit.py : ADMIT run() called [flowcount 1]
INFO :
INFO :
INFO : Executing Ingest_AT - '' (V1.0.5)
INFO :
INFO :
INFO : Run using the following settings:
INFO : box : []
INFO : file : test22.fits
INFO : mask : True
INFO : basename : x
INFO : smooth : []
INFO : pb :
INFO : edge : []
INFO : vlsr : -999999.0
INFO :
TIMING : Ingest ADMIT [ 6.53000000e+00 1.46596235e+09]
TIMING : Ingest BEGIN [ 0. 0.]
INFO : Ingest AT.py : basename=x
INFO : utils.py : OSTYPE: linux
TIMING : Ingest start [ 5.00000000e-02 5.26449680e-02 1.07379297e+03 1.57281250e+02]
TIMING : Ingest fromfits [ 3.50000000e-01 8.64300013e-01 1.07567188e+03 1.60902344e+02]
INFO : Ingest AT.py : SHAPE: [180 180 300 1]
TIMING : Ingest summary-0 [ 7.00000000e-02 1.27604961e-01 1.07567188e+03 1.61277344e+02]
WARNING : Ingest AT.py : no extra mask created because input image already had one
TIMING : Ingest summary-1 [ 2.00000000e-02 1.66149139e-02 1.07567188e+03 1.61281250e+02]
TIMING : Ingest statistics [ 7.00000000e-02 8.04221630e-02 1.07567578e+03 1.62003906e+02]
INFO : Ingest AT.py : COMMONBEAM[3] {'major': {'value': 6.381334781647199, 'unit': 'arcsec'}, 'pa': {'value': -80.3289642334, 'unit': 'deg'},
INFO : Ingest AT.py : BASICS: [shape] npts min max: [180 180 300 1] 3959914 -1.821495 8.147384
INFO : Ingest AT.py : S/N (all data): 38.631754
INFO : Ingest AT.py : GOOD PIXELS: 3959914/9720000 (40.739856% good or 59.260144% bad)
WARNING : Ingest AT.py : MASKS: ['mask0']
REGRESSION : CUBE: -1.82149 8.14738 0.210899 180 180 300 59.260144
INFO : Ingest AT.py : TELESCOPE: ALMA
INFO : Ingest AT.py : OBJECT: Serpens Main
INFO : Ingest AT.py : REFFREQTYPE: LSRK
INFO : Ingest AT.py : VLSR = 8.000000 (from source catalog)
INFO : Ingest AT.py : RA Axis 1: 277.491667 -1.000000 90.000000
INFO : Ingest AT.py : DEC Axis 2: 1.229167 1.000000 90.000000
INFO : Ingest AT.py : Freq Axis 3: 244.95 -0.000130723 0
INFO : Ingest AT.py : Cube Axis 3: type=Frequency velocity increment=0.160000 km/s @ fc=244.930396 fw=-0.039217 GHz
INFO : Ingest AT.py : RESTFREQ: 244.936 244.936
INFO : Ingest AT.py : VLSRc = 6.320000 VLSRw = 48.000000 VLSR = 8.000000
TIMING : Ingest done [ 2.60000000e-01 2.63298035e-01 1.07567578e+03 1.62949219e+02]
TIMING : Ingest END [ 0.83 1.41046405]
INFO : AT.py : BDP_OUT[0] = SpwCube_BDP x.im
INFO :
INFO :
INFO :
```


ADMIT

Timeline for Science users:

- May 1, 2016: Delivery of completed software system
 - End of funded ALMA Development Project was April 30, 2016
 - Contract requires delivery of all software and documentation
 - Options for continued support will be explored with ALMA/NRAO