

A vision for Chivo: Chilean Virtual Observatory

Diego Mardones
Universidad de Chile

THE NEED IN CHILE OF HAVING A COLLABORATIVE E-INFRASTRUCTURE SUPPORTING THE ASTRONOMICAL RESEARCH.

KAWAX 2008, by Rodolfo Barba, Paulo Cortes, Doug Geissler, Eduardo Unda, REUNA

“Propose the creation of a Chilean Virtual Observatory: ChiVO”

Programa Bicentenario de Ciencia y Tecnología



Kawax

1. Connectivity
2. Software
3. Grid Infrastructure
4. Human Capital
5. IVOA
6. Policy for data usage

Fondef: Development of an Astro-Informatic Platform for Management and Intelligent Analysis of Large-scale Data



- **Director:** Mauricio Solar-USM (AI, HPC)
- **Alternate Director:** Marcelo Mendoza-USM (DB, DM)
- **Researchers Team**
 - Diego Mardones – UChile (Astronomer)
 - Guillermo Cabrera – UChile (DM)
 - Nelson Padilla – PUC (HPC, Astronomer)
 - Karim Pichara – PUC (DM)
 - Neil Nagar – UDEC (Astronomer)
 - Ricardo Contreras – UDEC (AI)
 - Victor Parada – USACH (AI)
- **Technology Transfer:** Gerardo Rivas.

Fondef: Development of an Astro-Informatic Platform for Management and Intelligent Analysis of Large-scale Data



Partners and Associates:



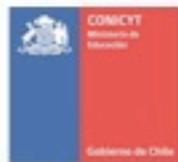
ALMA

- Technology Transfer
- Sponsor (\$80.000.000 CLP):
 - Human Resources
 - Astronomers and Astro-Engineers
 - Infrastructure
 - Data generated by observing projects.



REUNA

- Communication infrastructure
- Sponsor (\$14.218.646 CLP):
 - Development
 - High Speed Connectivity



FONDEF

- Sponsor (US\$ 600 K)

Fondef: Development of an Astro-Informatic Platform for Management and Intelligent Analysis of Large-scale Data

- January 2013 to June 2015:
 - i. ChiVO prototype
 - ii. VO-compliant model for ALMA data
 - iii. Use cases specifically tailored to ALMA

The International Virtual Observatory Alliance



ChiVO officially joined IVOA in 2013

VO Science

- From data search to data analysis:
 - Finding images off the www using e.g. aladin
 - Searching catalogues, using e.g. topcat
 - Finding available data on a given sky position and retrieving it, e.g. cone search.
 - Tools: (massive) SED fitting, time series analysis, statistics ...
- Variety of data kinds makes VO protocols particularly useful: user does not need to focus efforts on data read/access, or on specific metadata.

The Spanish Virtual Observatory

<http://svo.cab.inta-csic.es/main/index.php>

The SVO



The Spanish Virtual Observatory (SVO) officially started in June 2004. Its purpose is to coordinate the VO activities at national level and act as a contact point for the other VO initiatives. The SVO core team is hosted at Centro de Astrobiología (INTA-CSIC).

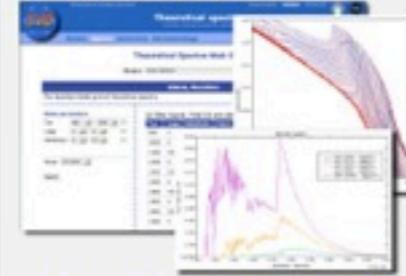
- SVO participants
- VO FAQs

The CAB Scientific Data Centre



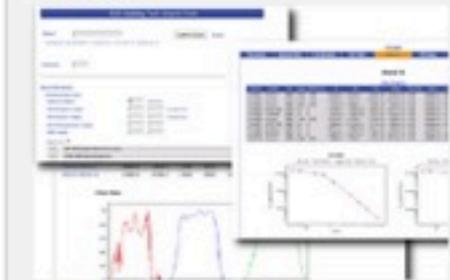
- Calar Alto
- DUNES
- GASPS
- GTC
- OMC
- X-exoplanets
- CMC-15
- Mark-I
- Census of astronomical Data Centres in Spain
- COROT
- DSS-63
- GAUDI
- INES
- Protostars
- ALHAMBRA
- Joan Oró
- Publishing data in the VO

Theoretical Data Server



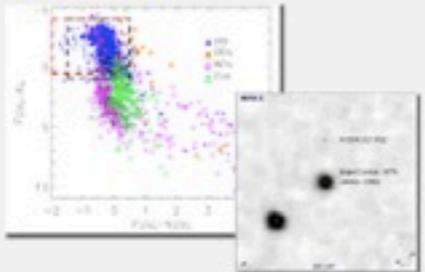
- Stellar Spectra theoretical models
- Evolutionary Synthesis Models
- Observational templates
- Synthetic Photometry
- Asteroseismology

Services



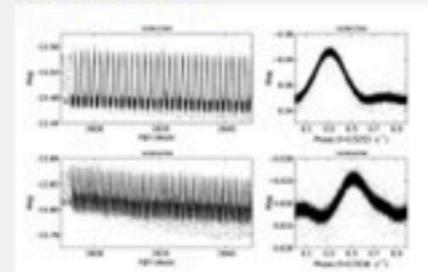
- VOSA
- VOSED
- TESELA
- Filter Profile Service

VO Science



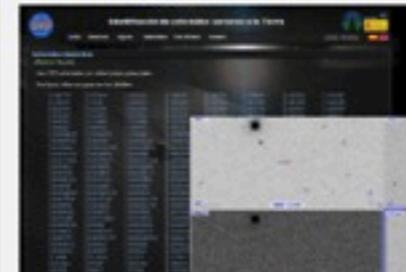
- Projects

Data Mining



- Projects

Education & Outreach



- Near Earth Asteroids Precovery
- Pro-Am collaborations
- Teaching Astronomy with the VO
- Undergraduate & graduate projects
- SVO schools and meetings

Miscellanea



- Papers
- Presentations
- Press Releases
- SVO in the media
- Job opportunities
- Summer school
- GREAT-ITN School
- CoRot Data Access scenarios
- SVO survey (Dec 2013)

ChiVO

From David Schade, chair of IVOA executive committee:

What role should CHIVO play?

Immediate Opportunities in IVOA:

- ALMA requirements for Data Discovery
 - And subsequent development of standards that satisfy those requirements
- ALMA requirements for Data cube access
 - And subsequent development of standards that satisfy those requirements
- Development and deployment of tools based on these standards

ALMA data

- Limited data in Early Science period (number of projects and range of observable parameter space)
- Delivery includes a (rich) variety of fits cubes.
 - Spatial, frequency and polarization information are defining properties of ALMA data.

ALMA Archive query



Atacama Large Millimeter/Submillimeter Array

In search of our Cosmic Origins

You are here: [Home](#) > [ALMA Data](#) > Archive Query

ALMA Science Archive Query

[Query Form](#) [Result Table](#)

[Query Help](#)

[Search](#) [Reset](#)

Position

Source name (Sesame)
Source name (ALMA)
RA Dec
Search radius

Energy

Frequency
Bandwidth
Spectral resolution
Band

Time

Observation date
Integration time

Polarisation

Polarisation type

Observation

Water vapour
Scan intent

Project

Project code
Project Title
PI Name

Options

Results view
 raw data project

Release status
 public data all data

<https://almascience.nrao.edu/aq/>

ALMA Archive query

You are here: [Home](#) > [ALMA Data](#) > Archive Query

ALMA Science Archive Query

Query Form

Result Table

Submit download request

Resu

Showing 319 rows (319 before filtering)

[More columns](#)

<input type="checkbox"/>	project_code	SOURCE_NAME	RA	DEC	BAND	Integration	RELEASE_DATE ▲	vel_resolution	frequency_support
Filter:	<input type="text"/>	<input type="text" value="m/s"/>	<input type="text"/>						
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	67.376	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	67.438	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	67.458	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	67.909	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.31	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.338	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.374	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.648	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.652	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...
<input type="checkbox"/>	2011.0.00367.S	HH46/47	08:...	-51:...	3	69.68	2013-03-30 14:5...	90.69	[100.85..100.91GHz,30.5...

Problems: either too many rows (data ids, meaningless), or no information available

ALMA Archive query

Showing 1 rows (1 before filtering)

[More columns](#)

<input type="checkbox"/>	project_code	TITLE	TYPE	PI_NAME	project_uid	RELEASE_DATE ▲
Filter:	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="checkbox"/>	2011.0.00367.S	Outflow Entrainment in HH 46/47 v0.6	S	Mardones, Diego	uid://A001/X3b/X927	2013-03-30 14:51:00.0

ChiVO includes ALMA Archive query as VO

- Big effort both to design and implement
 - Substantial effort to educate team
 - Need communication with astronomers' community, ALMA development teams, IVOA development teams

Use cases under development by Fondef

1. Source finding, including spatial information and relations between sources.
2. Image stacking, as a potential method to study samples of objects.
3. Emission line detection and characterization.
4. IA source morphological classification.

1. Source Finding

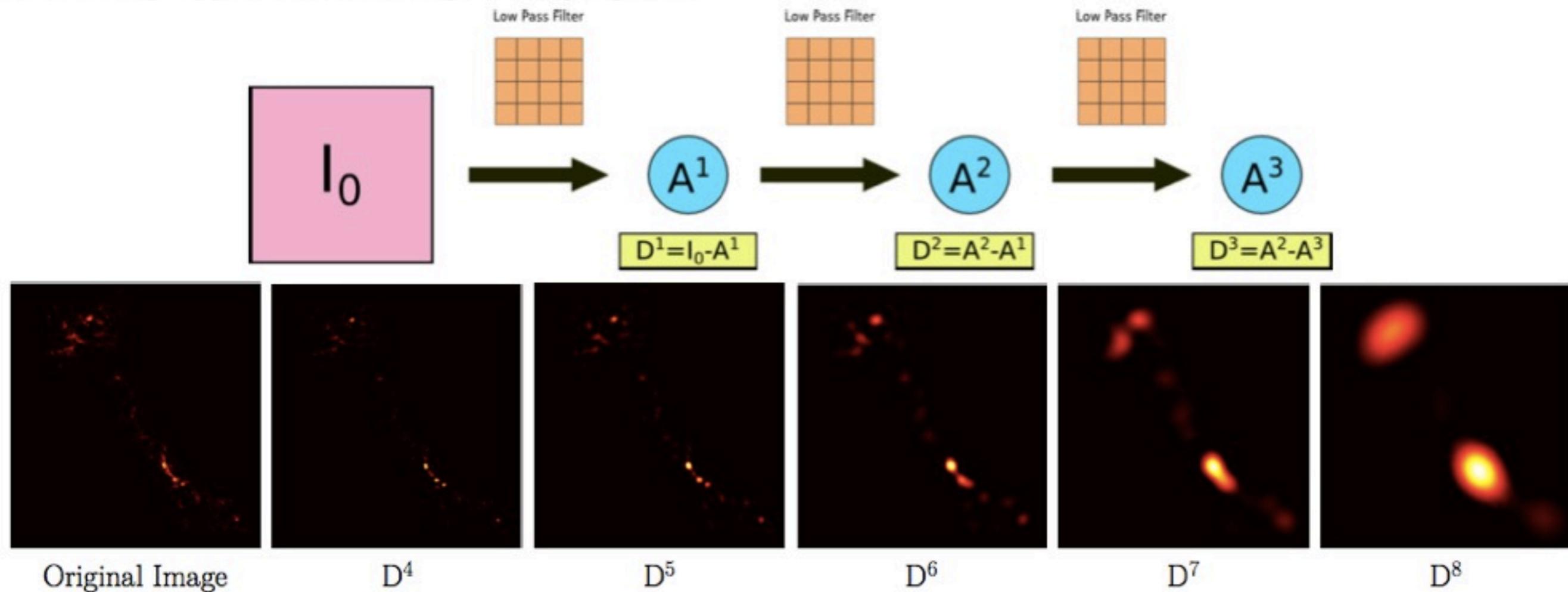
- * Detect structures for different spatial scales in 2d-images, generate catalogue for each spatial scale.
- * Related to astronomical tools such as PSF photometry, sextractor, clumpfind, gaussclumps, dendrograms.
- * Initial wavelet-based development is fast and optimal for ALMA-like data.

Thesis of Rodrigo Gregorio at UTFSM

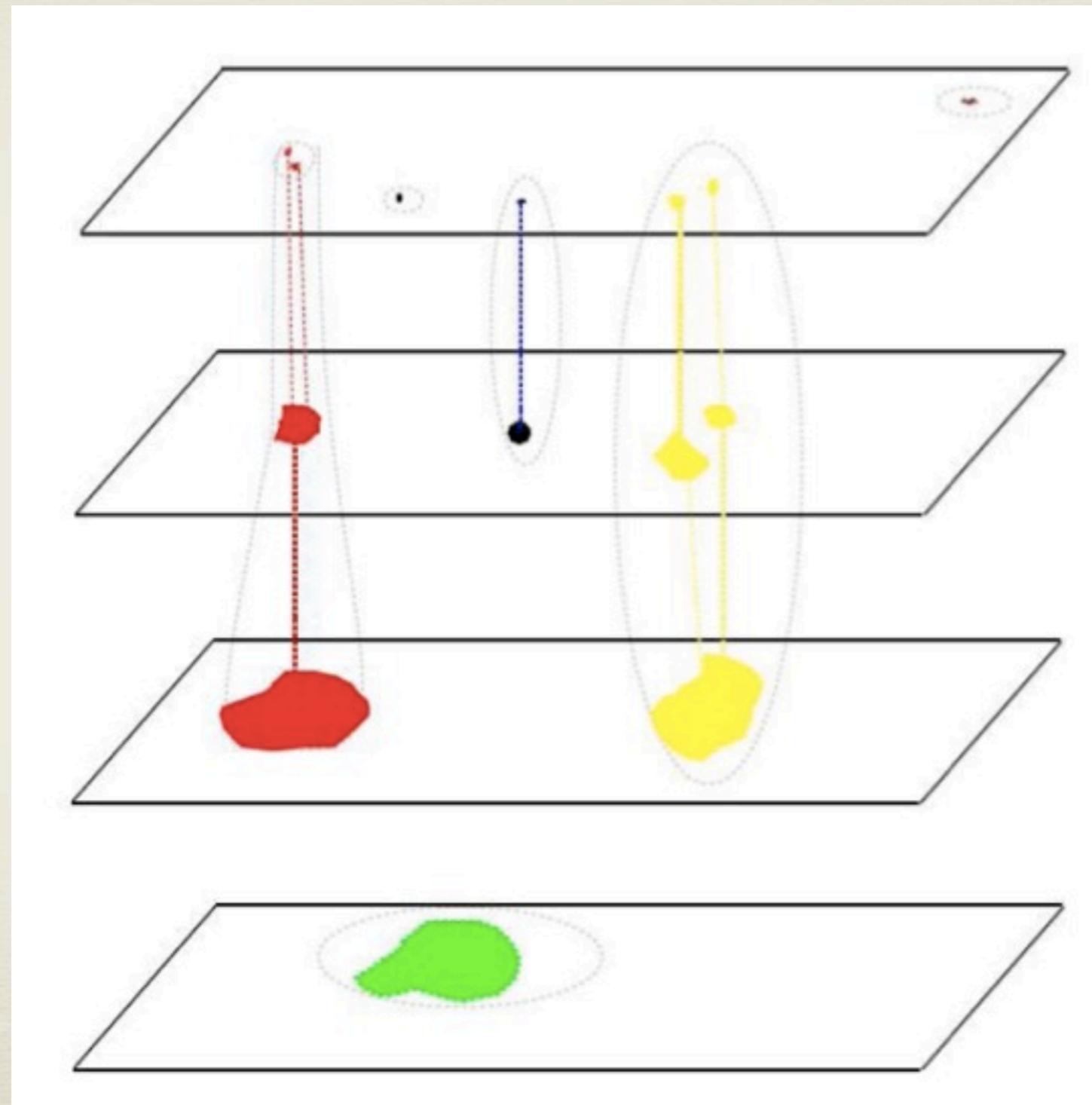
Wavelet approach:

Images at Different Scales

The proposed workflow consists in create sub-image using an low pass filter from A Trous wavelet. Each sub-image applies a detection algorithm by separate.

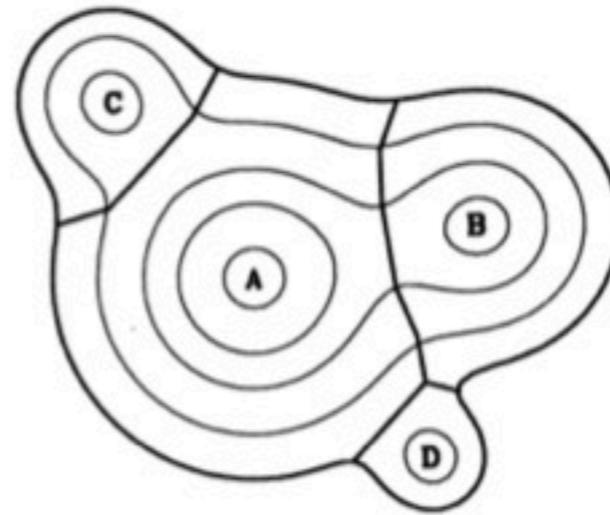


Wavelets: multiscale vision model

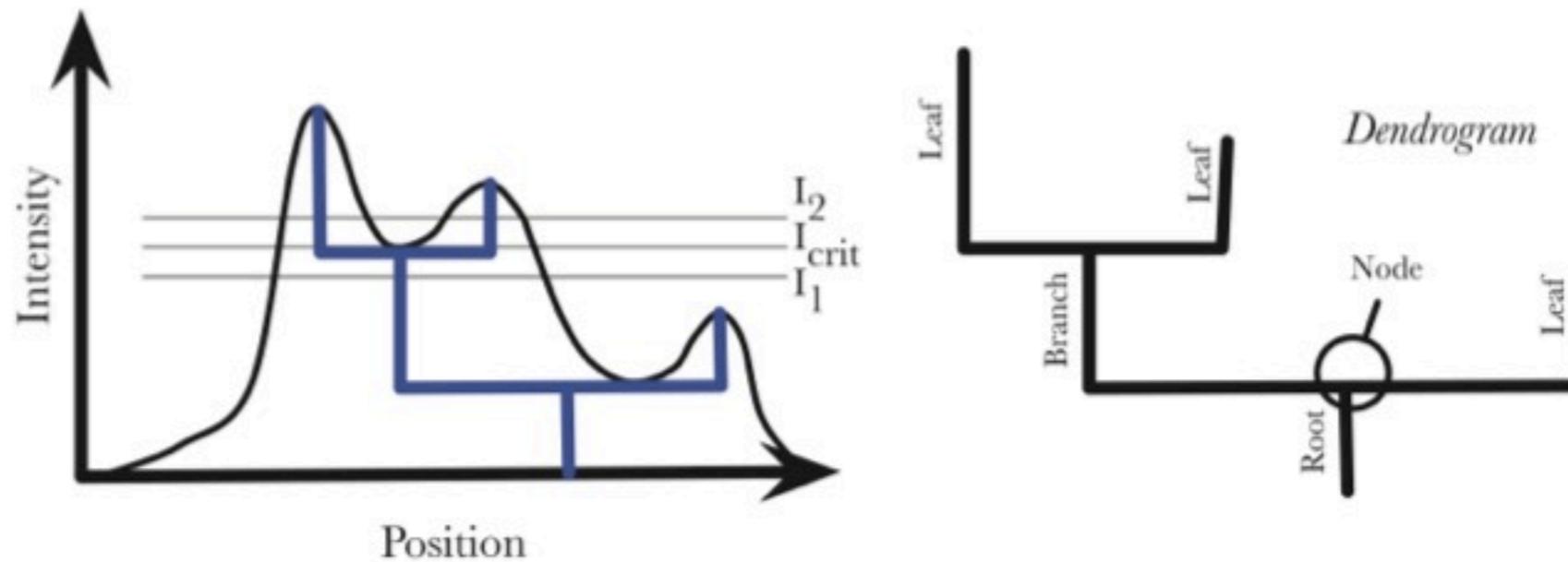


Stark & Murtagh (2006)

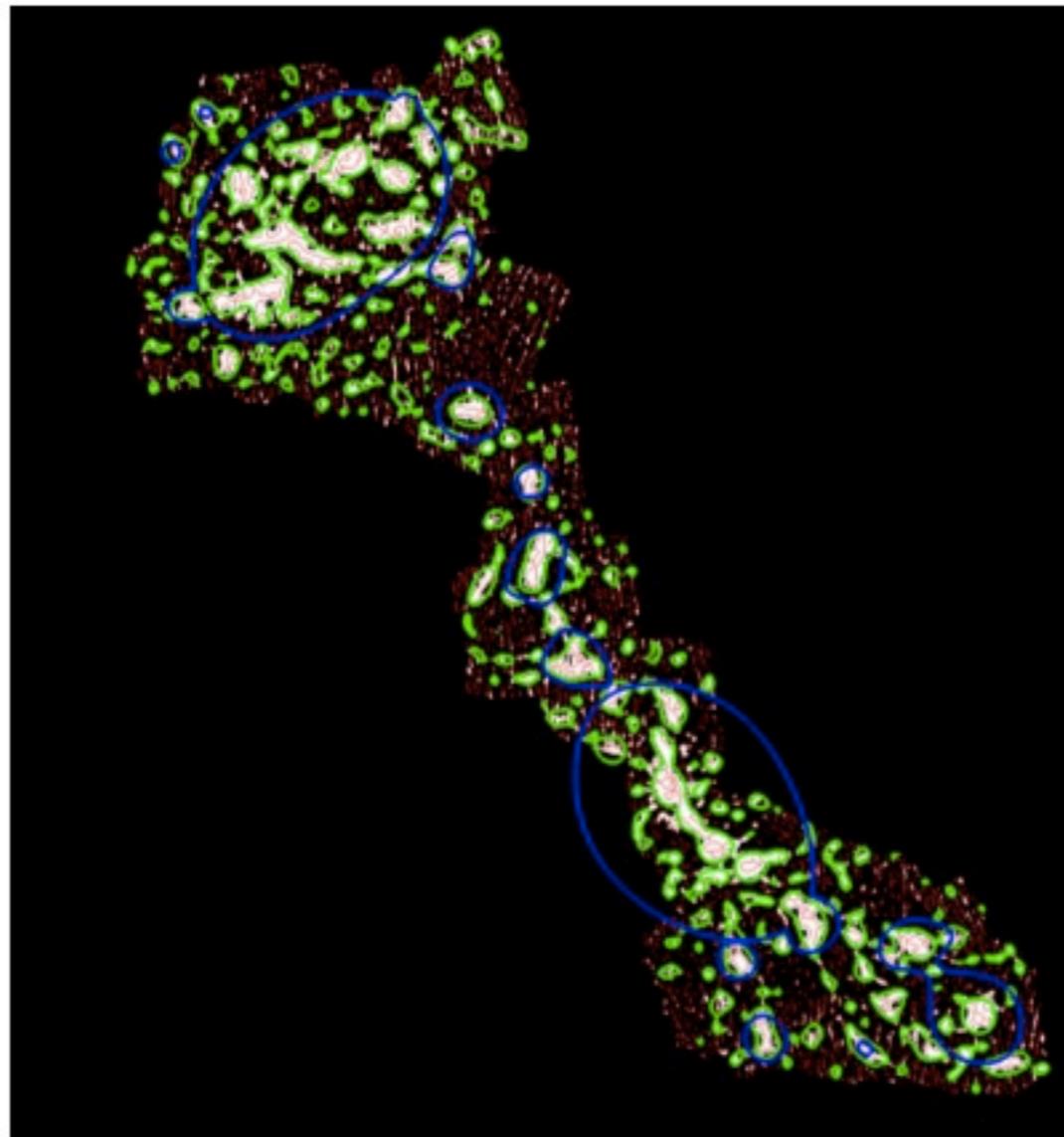
Detection Algorithms



Example of clumpfind structure.



Example of construction method for dendrograms .



Red = $D^1 + D^2$

Green = $D^3 + D^4$

Blue = $D^5 + D^6 + D^7$

Development:

1. Implement 2d wavelet+detection algorithm as stand-alone, within CASA, within ChiVO.
2. Implement HPC and cloud computing solutions.
3. Explore other wavelet & detection algorithms.
4. Explore 1d and 3d algorithms; multi-wavelength solutions
5. Combine with other modules: imaging, statistics, ...
6. Explore direct implementation from uv data

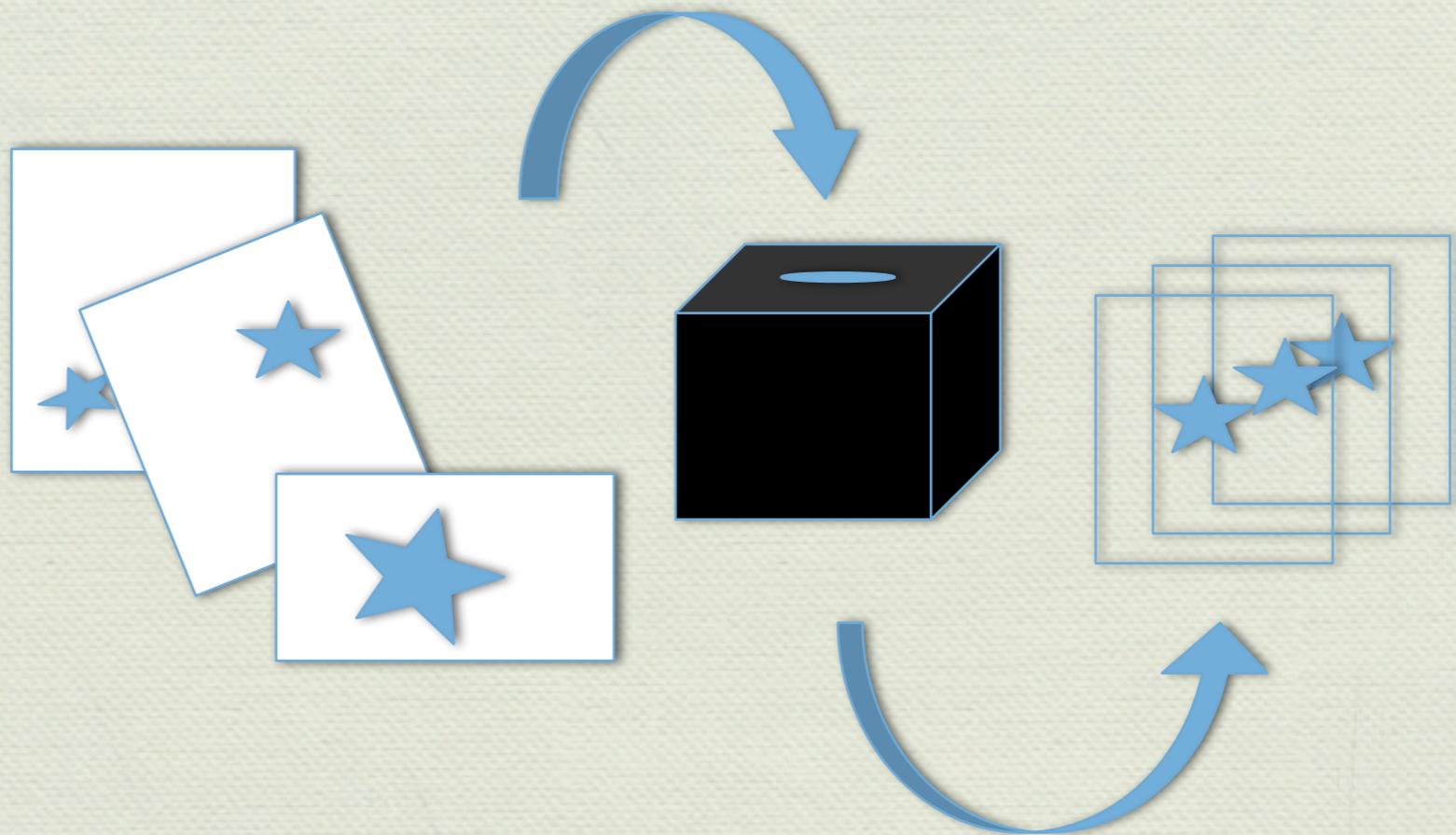
2. Source Stacking

- ◆ Put multiple images on top of each other
- ◆ Related to 1-source statistics, multiple-data per source, multiple-source distribution.

Thesis by Rodrigo Jara at U. de Santiago

Stacking:

- ◆ Align
- ◆ Scale
- ◆ Rotate
- ◆ De-project
- ◆ Statistics



Development

- ◆ Input a collection of 2d-images of the same “kind”
- ◆ Trivial implementation for 3d-cubes
- ◆ Extension to 1d datasets: align & expand
- ◆ Implement as stand-alone, within CASA, ChiVO
- ◆ Implement HPC & Cloud solutions

3. EMISSION LINE DETECTION AND CHARACTERIZATION

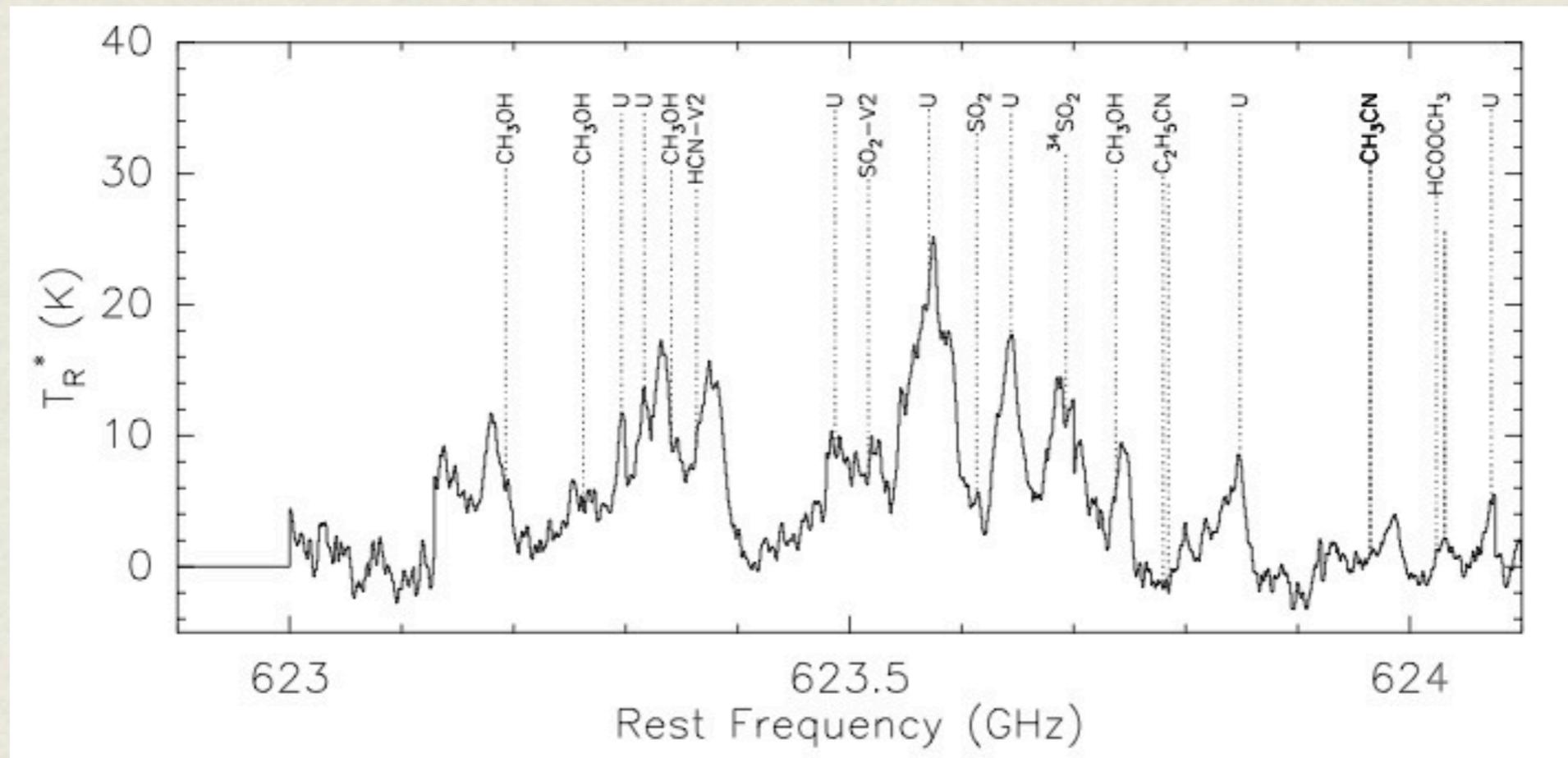
- Detect multiple emission lines.
- Characterize lines by intrinsic shape & origin.
- Use data-mining and AI techniques.

Thesis of: Andrés Riveros (PUC)

Nicolás Miranda (U de Chile)

Alejandro Barrientos, Jorge Sepulveda (UTFSM)

Abundant Spectral Lines



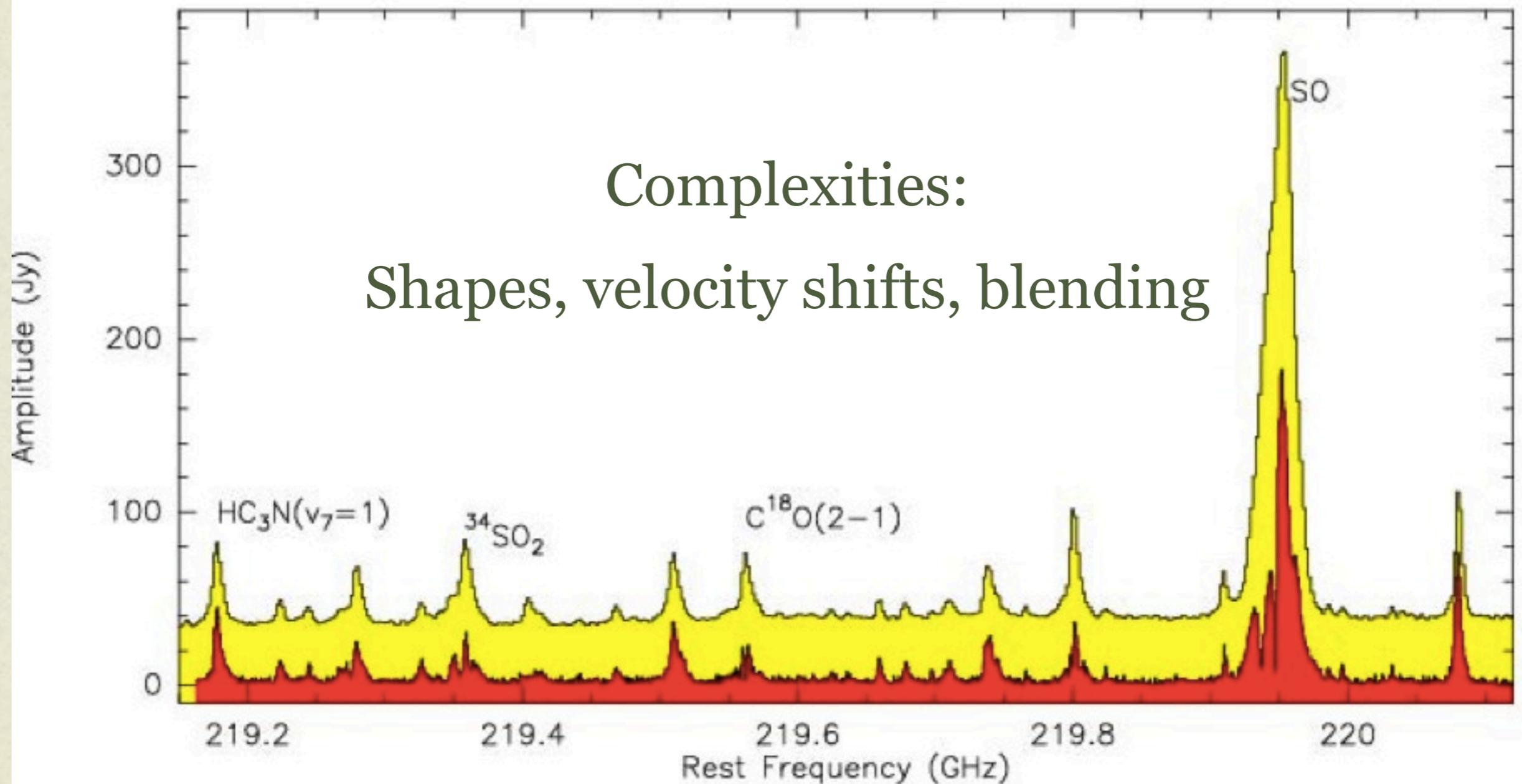
Orion: Schilke et al 2001

Abundant Spectral Lines

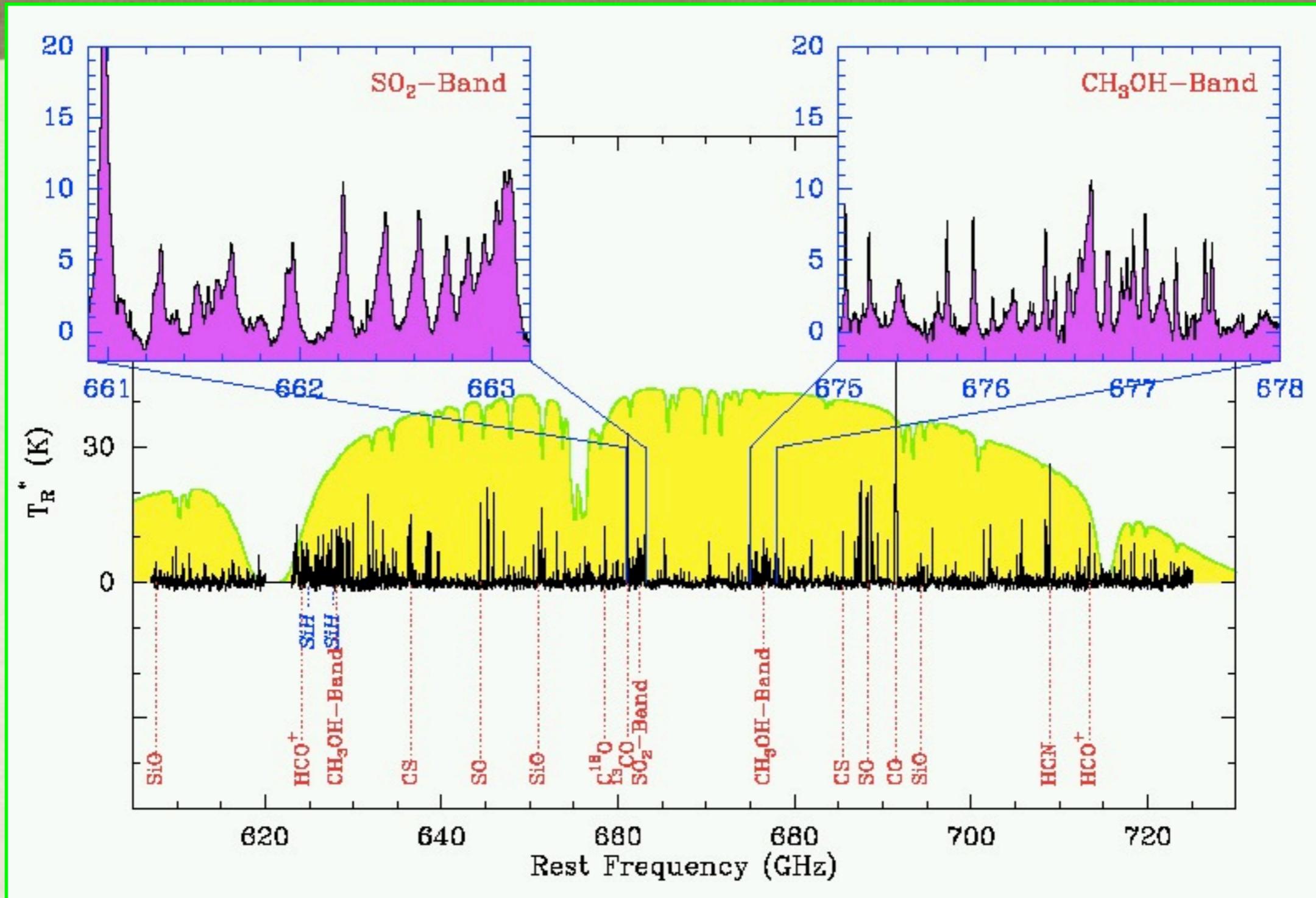
yellow: single dish (30m); red: OVRO 20m baseline

Complexities:

Shapes, velocity shifts, blending



Abundant Spectral Lines



Rich potentially unused information: chemistry, kinematics, spatial distribution

Abundant Spectral Lines

- Existing tools: Splatalogue, Xclass, Cassis, Madcuba, ... use simple assumptions about e.g. line shapes, molecular abundances, gas temperatures
- Proposed tools: (i) Purely statistical descriptors of lines, (ii) Full Data Mining algorithm
 - i. “Full” catalogue of line ids including Doppler shifts.
 - ii. Automatic discovery of “correlations” between lines.

Use cases under development by Fondef

1. Each tool is both tuned to ALMA data and is VO-complaint.
2. Each tool is complementary to efforts elsewhere.
3. Each tool is to have fully working prototypes within this year.
 - But further (data) complexity can lead to years of future development.

Implementing ChiVO

- Need dedicated computers and network in order to make it feasible.
- Project to implement archive computers at UTFSM funded by CASSACA (2014-2016)
 - Comparable to ALMA ARCs in capabilities, start with ~256 TB, goal of >1 PB
 - Initial investment ~US\$1 million
 - Goal: host 1-2 cycles of full alma data archive (500 TB/yr), and permanent public reduced data products (1/100 of raw data).

Implementing ChiVO

- Most likely solution: container at UTFSM in San Joaquin (Tier 2 data center).
- connectivity to NLHPC computers at the CMM and astro-PUC for further processing.
- Long term and multi-observatory data archive capabilities require a different large-scale national data-center (Ibsen, Liello, Vera)

ChiVO and Chilean Astronomers:

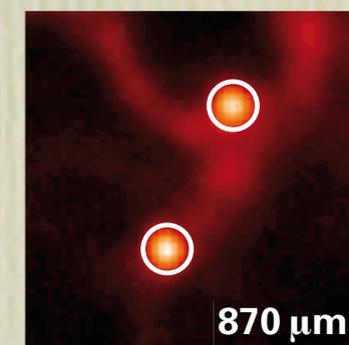
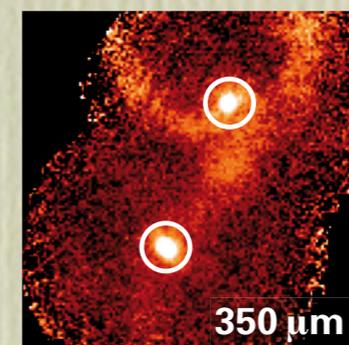
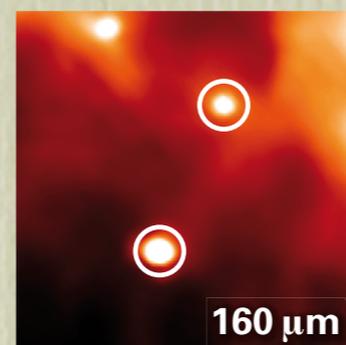
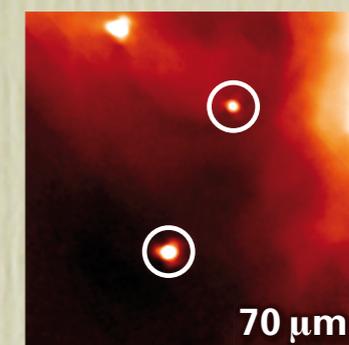
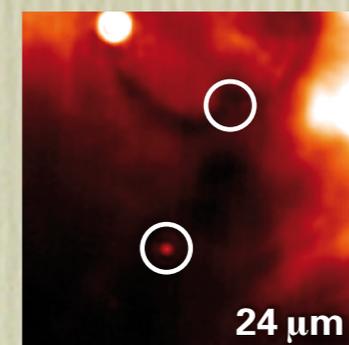
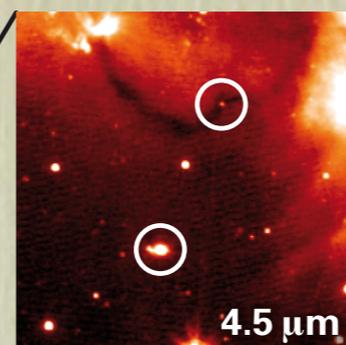
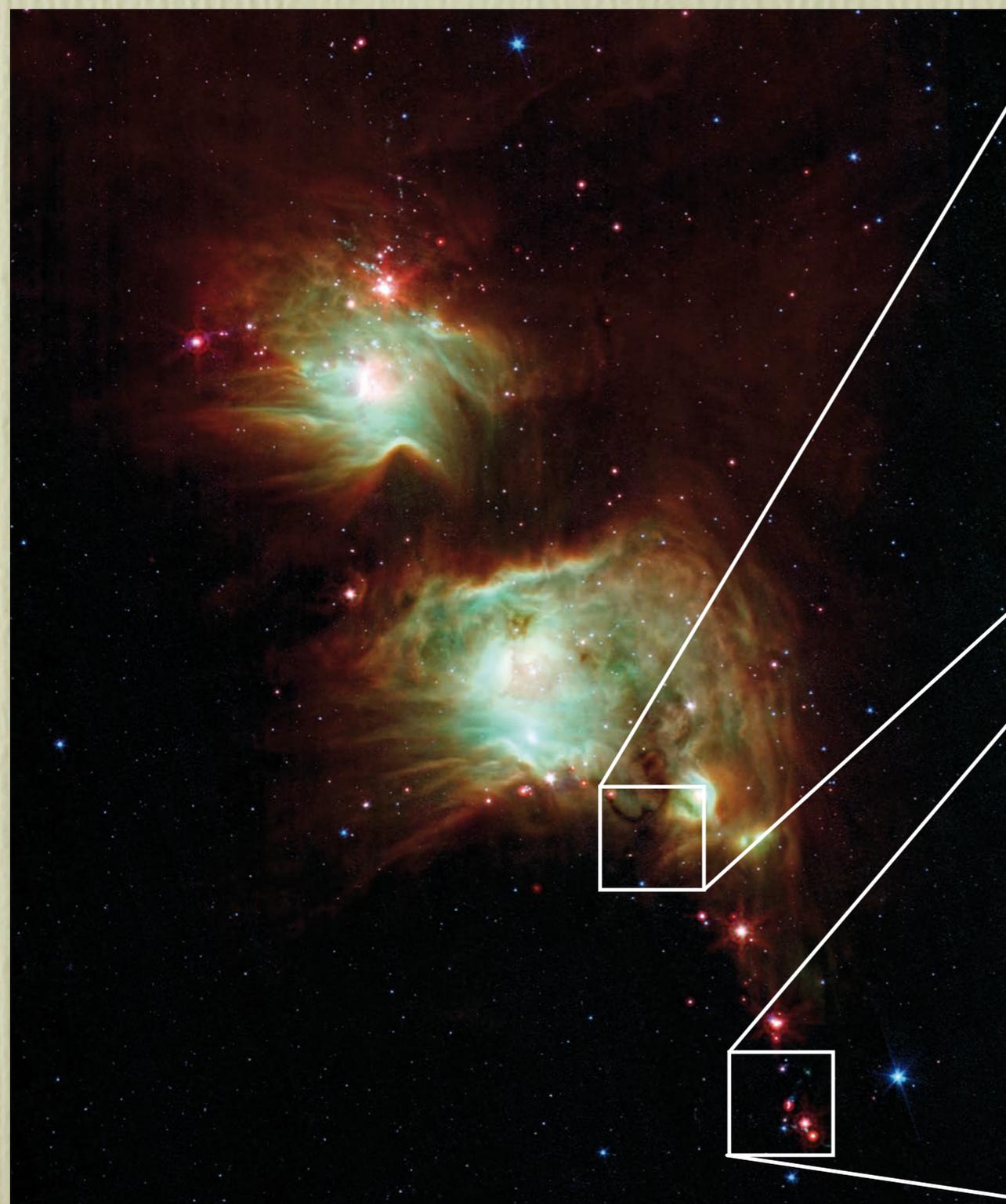
- Chilean astronomy: **large amount of data**
- Enormous variety of data (and administration)
 - Most Chilean astronomers specialize in both of the above
- Big challenges in data access and preservation
- Large value added for historic data and/or long term projects

ChiVO and Chilean Astronomers:

- Example: Orion A Vista survey
 - 20 square degrees in JHK, 33 hours observed
 - sensitivity to all stars and BDs
 - Final data is 3 images of ~16 GB each (multi-scale view)
 - Careful data reduction achieved twice the angular resolution of the standard ESO pipeline
 - Catalogue of ~10 million objects
 - includes ~20,000 YSOs, ~300 Spitzer protostars
 - binaries, spatial distribution, galaxies
 - 3 PhD thesis projects (Vienna and UChile)

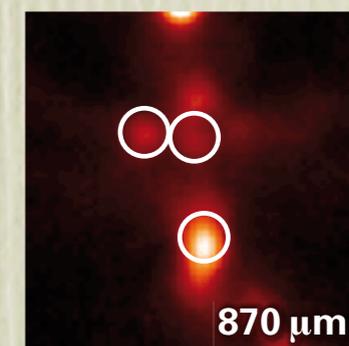
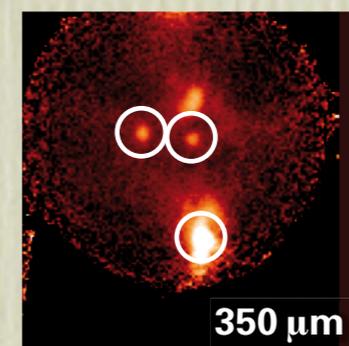
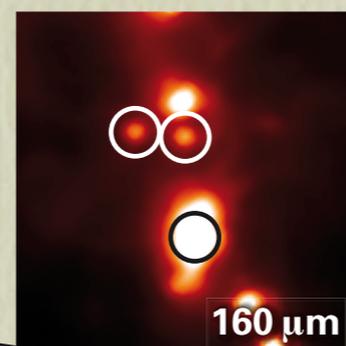
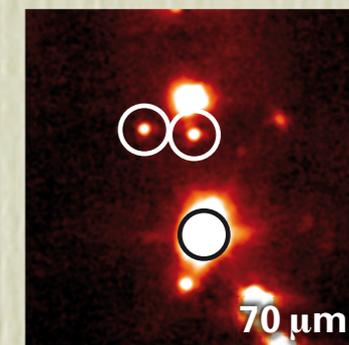
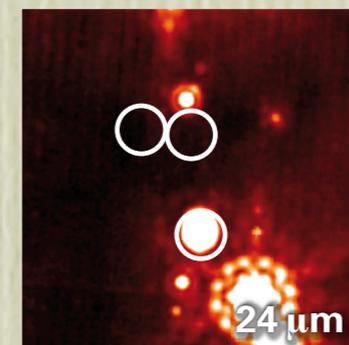
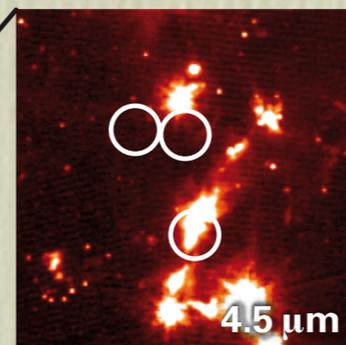
Orion A





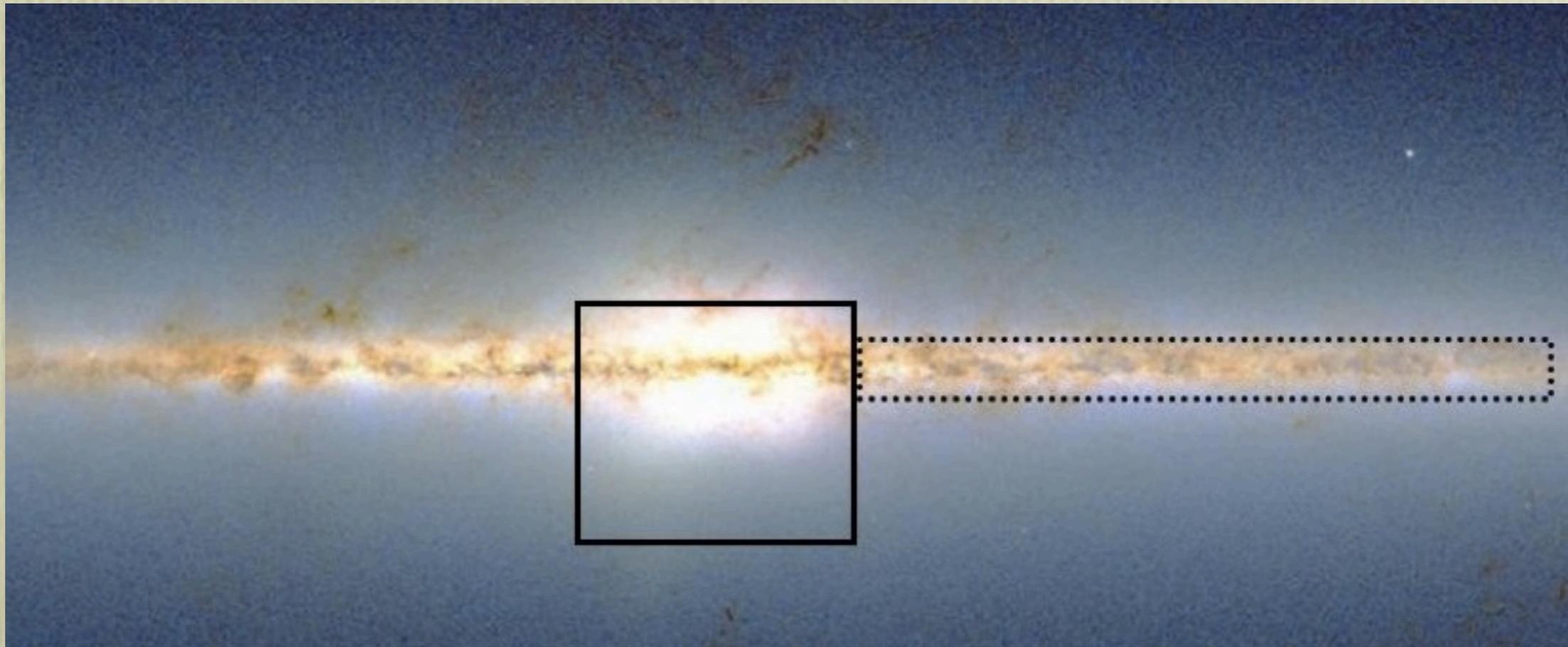
NGC2068 093005

NGC2068 091016



ChiVO and Chilean Astronomers:

- Example: VVV survey
 - Galactic plane and galactic center: 500 square deg.
 - Time domain: 200 nights, 5 years
- Reduced data is not available to the public: relies on ESO archives for long term access.



Editorial

Ten Simple Rules for the Care and Feeding of Scientific Data

Alyssa Goodman¹, Alberto Pepe^{1*}, Alexander W. Blocker¹, Christine L. Borgman², Kyle Cranmer³, Merce Crosas¹, Rosanne Di Stefano¹, Yolanda Gil⁴, Paul Groth⁵, Margaret Hedstrom⁶, David W. Hogg³, Vinay Kashyap¹, Ashish Mahabal⁷, Aneta Siemiginowska¹, Aleksandra Slavkovic⁸

1. Love your data and help others love it too
2. Share your data online with permanent id
3. Do science with data reuse in mind
4. Publish workflow as context
5. Link data to publications
6. Publish your code

Editorial

Ten Simple Rules for the Care and Feeding of Scientific Data

Alyssa Goodman¹, Alberto Pepe^{1*}, Alexander W. Blocker¹, Christine L. Borgman², Kyle Cranmer³, Merce Crosas¹, Rosanne Di Stefano¹, Yolanda Gil⁴, Paul Groth⁵, Margaret Hedstrom⁶, David W. Hogg³, Vinay Kashyap¹, Ashish Mahabal⁷, Aneta Siemiginowska¹, Aleksandra Slavkovic⁸

10. and more people. **You, as a scientist, need to help organize your discipline and your institution to move more quickly toward a world of open, discoverable, reproducible data and research. One important step is to *advocate* for hiring data specialists and for the overall support of institutional programs that improve data sharing.** Make sure not

ChiVO and Chilean Astronomers:

- Could incorporate local survey work: Araucaria, SN, VVV, extrasolar planets, Atlasgal, Hat-pi, ...
- Could serve as portal for LSST archives, Gemini data, CCAT, ...
- Could serve simulations community
- CANFAR: Canadian Network For Astronomy Research is a good model for potential development.

The Canadian Advanced Network For Astronomical Research



The screenshot shows the CANFAR website header with the logo and navigation menu. Below the header, there are three main content blocks: 'Store', 'Process', and 'Know', each with a description and a 'Learn more' button.

CANFAR
Canadian Advanced Network for Astronomical Research

Portal ▾ Documentation ▾ Login

Store
Keep your data close to the processing nodes, and have it public, private or accessible to a restricted set of collaborators.
[Learn more](#)

Process
Build your data processing pipeline, interactively analyze results, and launch thousands of batch jobs on multiple clusters, using exactly the same environment.
[Learn more](#)

Know
Learn about the Canadian Advanced Network for Astronomical Research, read documentation, and contribute to its building pieces.
[Learn more](#)

www.canfar.net

Focus on cloud computing and high-speed network

Potential key role for NLHPC

Niche for multidisciplinary efforts

ChiVO and Chilean Astronomers:

Examples of possible roles hosted by different institutions:

- Access to ALMA archive UTFSM
- IR data center PUC
- Time series tools MAS
- Apogee ULS
- Astrostatistics UV
- Cloud processing NLHPC

Possible Scientific Goals for Chilean Long Term Telescope Programs

to be held on **Sept 10, 2014** at the dependencies of the Departamento de Astronomía of the Universidad de Chile (Cerro Calán).

- Aims:
 - identify potential themes for long term programs
 - find synergies between efforts of separate groups
 - assess needed structures at the Chilean TACs that better accommodate large programs.
- Outcome: recommendations regarding the structures that are deemed necessary in the call for proposals to best accommodate the needs of the community regarding large/long term programs.
- cntac: What do you want and what are you willing to give?
 - Proposal: large programs required to follow Alyssa's 10 rules

ChiVO and Chilean Astronomers:

- By mid 2015 Fondef should turn ChiVO to the astronomy community. Need to establish a board-like structure to ChiVO.
- All chileans welcome to participate in development, to contribute data, tools or ideas
- IVOA: meetings twice a year, welcome participants in working groups, email groups, newsletter, ...

Summary:

- ChiVO will provide a working VO interface to public ALMA data within a year.
- Opportunity to generate shared infrastructure associated with:
 - Large and varied data
 - Data Mining and analysis tools
 - Science focus and efforts should be on ideas and not on how to implement them