

The Swiss Army Knife of Astronomical Data

Francesco D'Eugenio
francesco.deugenio@ugent.be

The Swiss Army Knife of Astronomical Data

By three methods we may learn wisdom: **First**, by reflection, which is noblest; **second**, by imitation, which is easiest; and **third** by experience, which is the bitterest.

Overview

- I. Storage (how to store the data)
- II. Procurement (how to get the data)
- III. Analysis (how to interrogate the data)

I. Data storage

- State of the art: HDF5, VOTable, ASDF
- Outdated: FITS
- Really, really outdated: plain text (csv, etc.)

I. Data storage

- State of the art: HDF5, VOTable, ASDF
- Outdated: FITS
- Really, really outdated: plain text (csv, etc.)

the best is also the least used!

the worst are amongst the most common.

I. Data storage

- State of the art: **HDF5**, **VOTable**, **ASDF**
- Outdated: **FITS**
- Really, really outdated: **plain text** (csv, etc.)

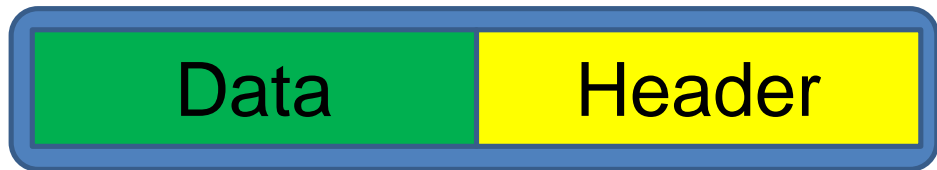
the best is also the least used!

the worst are amongst the most common.

I. Data storage: FITS files

- interface (quick): topcat, ds9, QFitsView
- interface (analysis): cfitsio, astlib, astropy
- tables & images
- multiple extensions
- headers → WCS

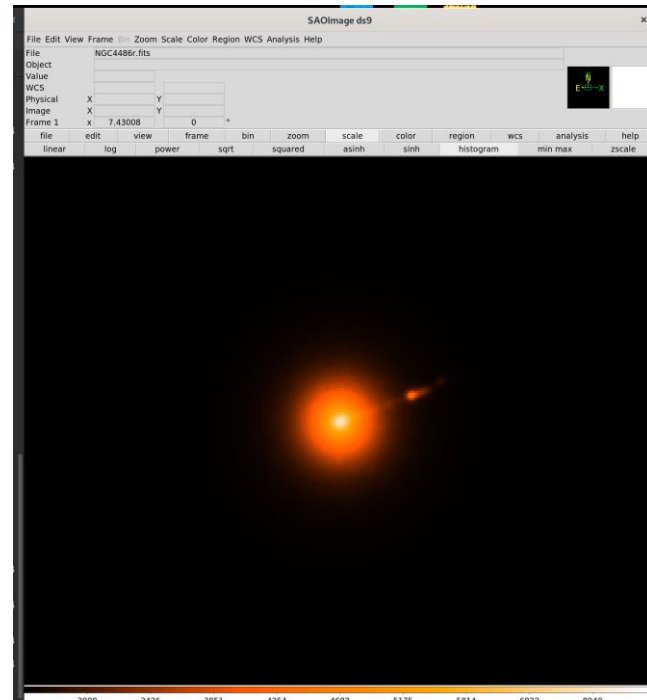
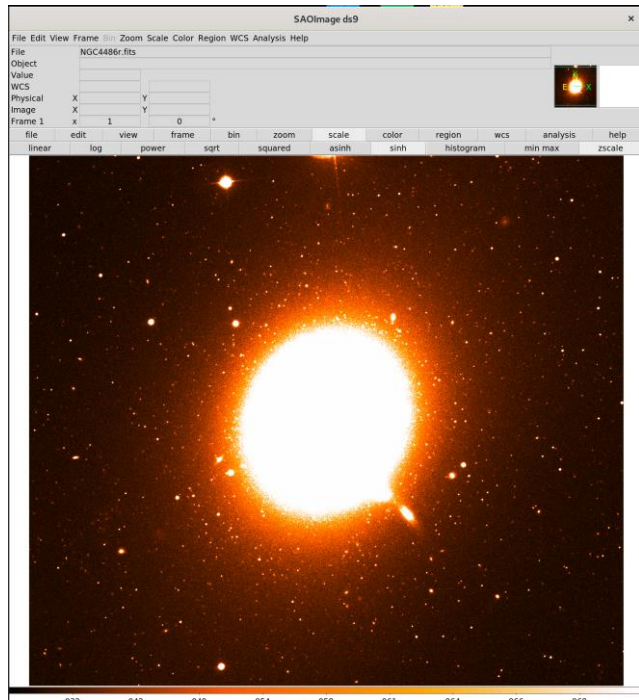
I. Data storage: FITS files



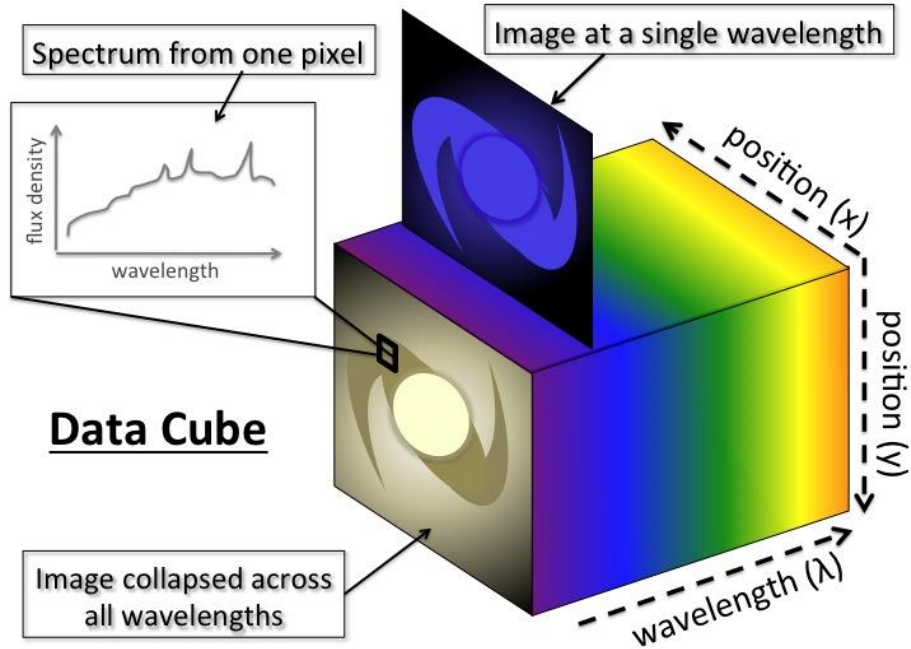
a FITS file
(single extension)

contains either an “image” or a
table (also called binary table)

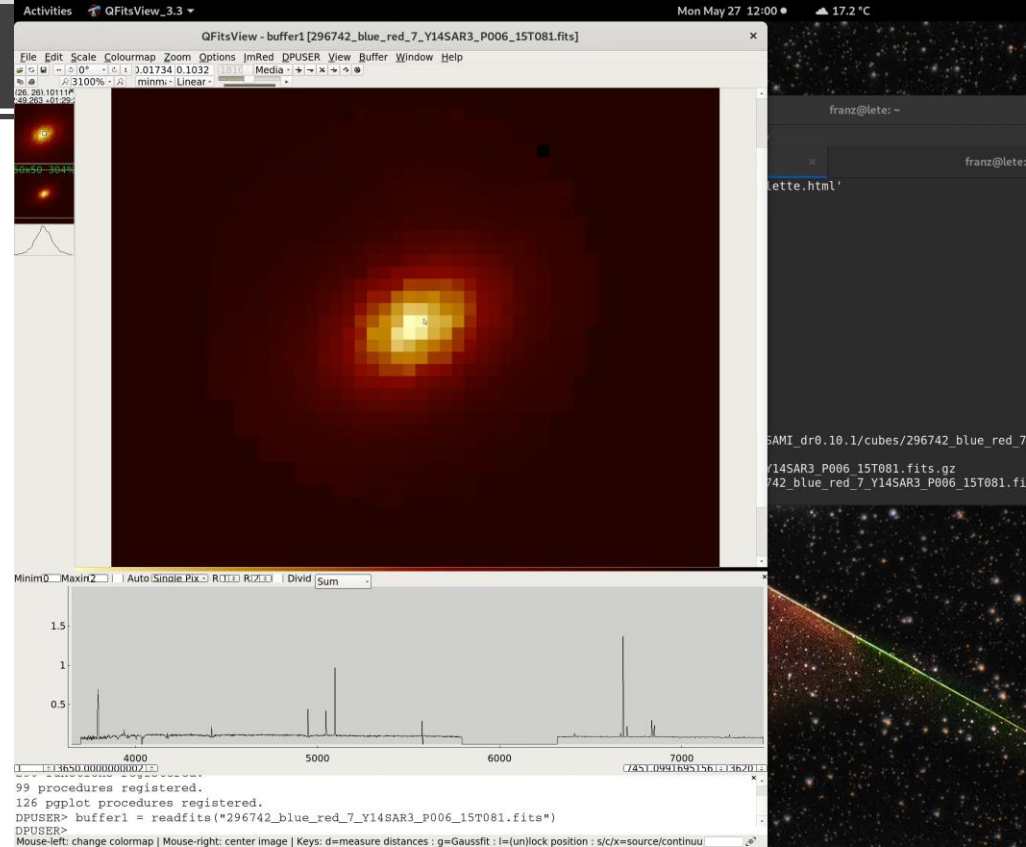
FITS Images



FITS "Images" (Cubes)



@ChrisHarrison



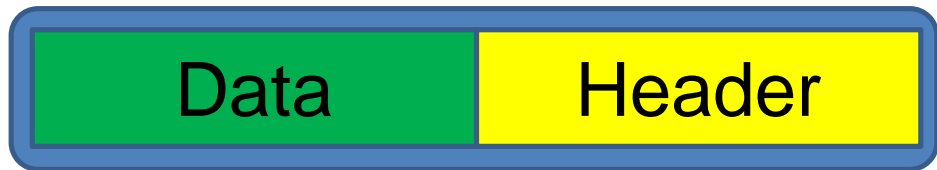
FITS Tables

```
<Table masked=True length=497>
  CATID      phot      mag      n_comp_r
  int64      bytes4    float64  int64
  -----
      7839      GAMA      15.003      4
      8487      GAMA      14.068      7
     15481      GAMA      14.888      5
     16525      GAMA      16.892      3
     16526      GAMA      16.816      4
     16926      GAMA      16.395      4
      ...
9403800763      VST  15.0276728      8
9403800813      VST  13.8577576      8
9403800833      VST  15.4592133      5
9403800911      VST  15.5672979      5
9403801002      VST  15.1227217      7
9403801010      VST  14.4940319      9
```

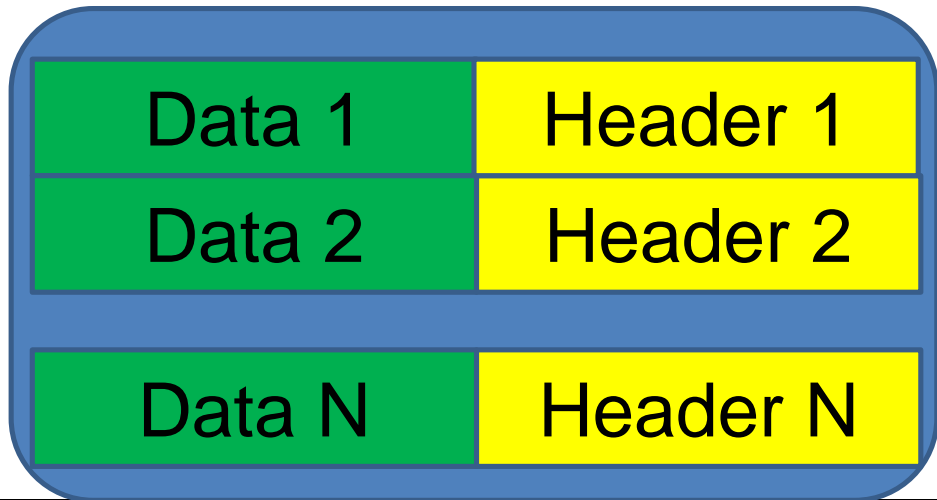
FITS Tables

```
<Table masked=True length=497>
  GATID      phot      mag      n_comp_r
  int64      bytes4    float64  int64
-----
      7839      GAMA      15.003      4
      8487      GAMA      14.068      7
     15481      GAMA      14.888      5
     16525      GAMA      16.892      3
     16526      GAMA      16.816      4
     16926      GAMA      16.395      4
      ...
9403800763      VST     15.0276728      8
9403800813      VST     13.8577576      8
9403800833      VST     15.4592133      5
9403800911      VST     15.5672979      5
9403801002      VST     15.1227217      7
9403801010      VST     14.4940319      9
```

FITS structure: multiple extensions



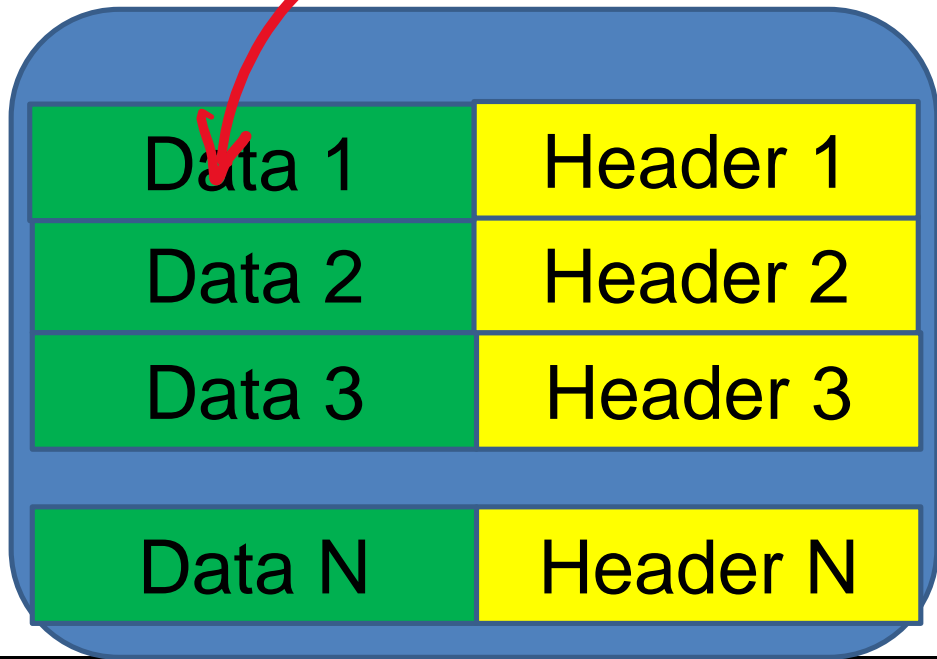
a FITS file
(single extension)



a FITS file
(multiple extensions)

FITS structure: multiple extensions

empty!



a FITS file
(multiple extensions)

...done right.

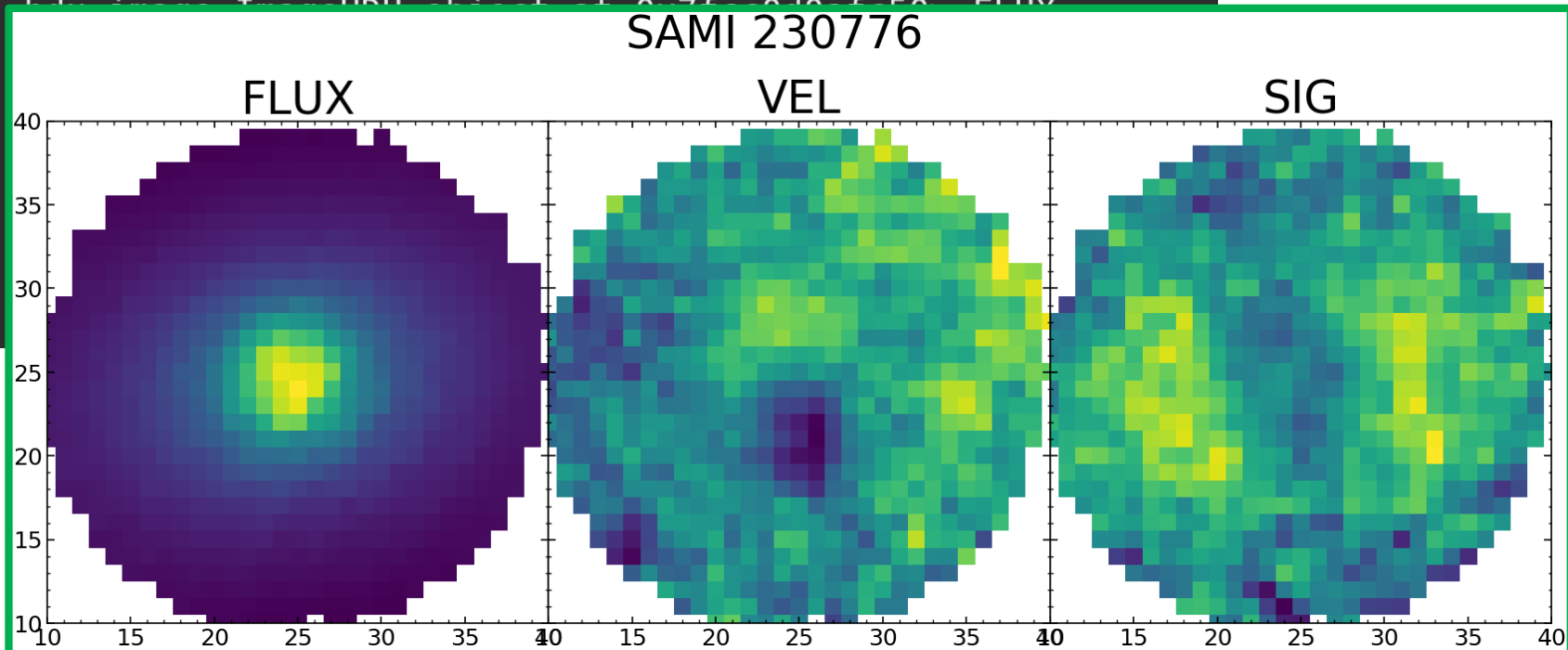
Why use FITS extensions?

```
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d0afc50> FLUX  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d42e860> FLUX_ERR  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d42edd8> VEL  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d424470> VEL_ERR  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d424a90> SIG  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d424f98> SIG_ERR  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d42b588> FORMAL_SNR  
<astropy.io.fits.hdu.image.ImageHDU object at 0x7fcc0d42bef0> QC
```

A way to group together different FITS files

What are FITS extensions?

```
<astropy.io.fits
<astropy.io.fits
<astropy.io.fits
<astropy.io.fits
<astropy.io.fits
<astropy.io.fits
<astropy.io.fits
```



What are FITS headers?

```
SIMPLE = T / conforms to FITS standard
BITPIX = 8 / array data type
NAXIS = 0 / number of array dimensions
EXTEND = T
CRPIX1 = 25.5 / Pixel coordinate of reference point
CRPIX2 = 25.5 / Pixel coordinate of reference point
CDELT1 = -0.00013888888888889 / [deg] Coordinate increment at reference point
CDELT2 = 0.00013888888888889 / [deg] Coordinate increment at reference point
CTYPE1 = 'RA--TAN' / Right ascension, gnomonic projection
CTYPE2 = 'DEC--TAN' / Declination, gnomonic projection
CRVAL1 = 181.112792 / [deg] Coordinate value at reference point
CRVAL2 = 1.895972 / [deg] Coordinate value at reference point
EQUINOX = 2000.0 / [yr] Equinox of equatorial coordinates
WCS_SRC = 'Nominal' / WCS Source
CATID = '230776' / Object ID
HGCUBING = '0627572adcf5+' / Hg changeset ID for cubing code
INSTRUME = 'AAOMEGA-SAMI' / Instrument in use
SPECID B = 'BL' / Spectrograph ID blue
GRATID B = '580V' / Disperser ID blue
SPECID R = 'RD' / Spectrograph ID red
GRATID R = '1000R' / Disperser ID red
PLATEID = 'Y13SAR1_P014_12T001_15T029' / Plate ID (from config file)
LABEL = 'Run 10 galaxy plate 1' / Configuration File label
IFUPROBE = '7' / ID number(s) of the SAMI IFU probe(s)
A_POLY = 12 / Order of additive polynomial in pPXF fit
N_MOM = 2 / Number of moments fit with pPXF
Z_IN = 0.01973 / Initial guess redshift input to pPXF
Z_PPXF = 0.01975 / Best fitting redshift from ppxf
AP_SIG = 295.4755098343882 / Aperture velocity dispersion
AP_SERR = 1.598230857424561 / Error on aperture velocity dispersion
AP_VEL = 5.261451588145117 / Measured aperture velocity
AP_VERR = 1.60067683112066 / Error on measured aperture velocity
APER = 'Circular 2 arcsec' / Aperture used to extract central spectrum
ORIG_CUB = '230776_blue_red_7_Y13SAR1_P014_12T001.fits' / Original data cube
PRODUCT = 'Stellar Kinematics (2 moments)' / Data Product
DATE = '2018-6-21' / Date this data product was generated
AUTHOR = 'Jesse van de Sande'
CONTACT = 'Jesse van de Sande <jesse.vandesande@sydney.edu.au>'
VERSION = 'v02' / Version of this data product
SAMI_VER = 'v00.10' / Version of SAMI cubes to generate product
```

```
CRPIX1 = 25.5 / Pixel coordinate of reference point
CRPIX2 = 25.5 / Pixel coordinate of reference point
CDELT1 = -0.00013888888888889 / [deg] Coordinate increment at reference point
CDELT2 = 0.00013888888888889 / [deg] Coordinate increment at reference point
CTYPE1 = 'RA--TAN' / Right ascension, gnomonic projection
CTYPE2 = 'DEC--TAN' / Declination, gnomonic projection
CRVAL1 = 181.112792 / [deg] Coordinate value at reference point
CRVAL2 = 1.895972 / [deg] Coordinate value at reference point
EQUINOX = 2000.0 / [yr] Equinox of equatorial coordinates
WCS_SRC = 'Nominal' / WCS Source
```

The place for “everything else”

Want to know more?

https://fits.gsfc.nasa.gov/fits_standard.html

<http://docs.astropy.org/en/stable/io/fits/>



II. Data procurement

- Observations
- Archives
- Various

II. Data procurement: observations

- Guaranteed time or competitive proposals
- Proposals have generally biannual cadence

Proposal tools

- Exposure time calculator (telescope website)
- Observability charts (iobserve, staralt)
- Proposal template (Latex)

Proposal success criteria

Compelling science case	Somewhat arbitrary
Quantitative time justification	Objective
Quantitative sample size	Objective
Quantitative expected results	Objective
Convincing telescope justification	Somewhat arbitrary
Informative, pretty figures	Absolutely objective
Robustness	Objective

Moon: on or off?

- Off for blue wavelengths and/or faint targets
- Moon has the same spectrum as the Sun → bluer than the sky background
- At red enough wavelength (>800 nm), sky emission dominated by other processes

II. Data procurement: various

- dedicated website
- email / usb stick / dropbox

→ difficult to obtain, unless searching for something specific.

II. Data procurement: archives

- CDS – Strasbourg astronomical data center (e.g. VizieR)
- SciServer (e.g. casjobs)
- Telescope archives (e.g. archive.eso.org, HST legacy)

Telescope archives

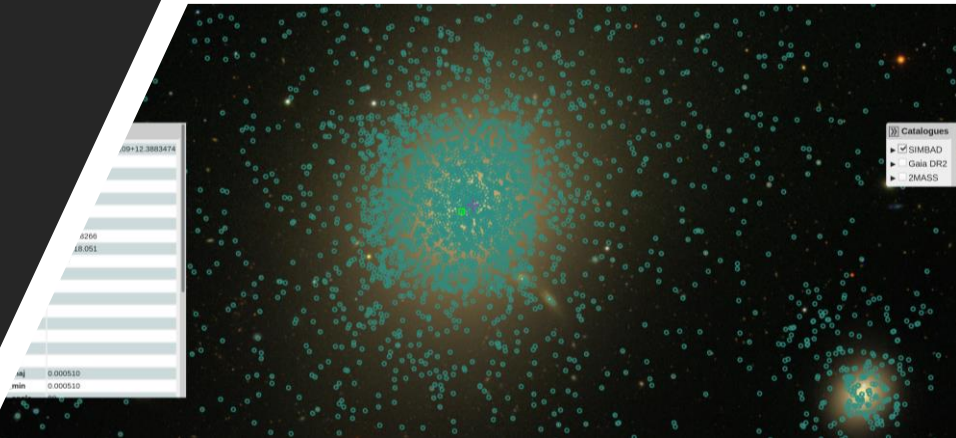
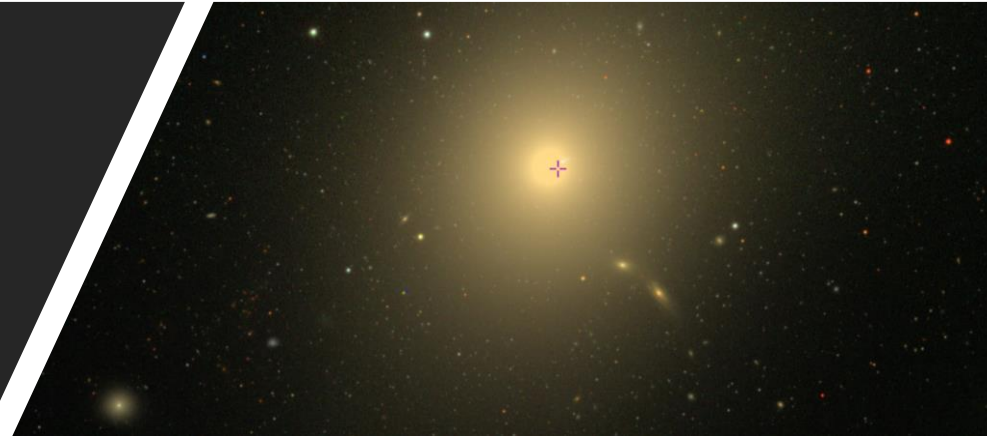
- ESO archive
- Keck archive
- HST Legacy archive
- Gemini archive
- Subaru ...and many more

CDS: Aladin Sky Atlas

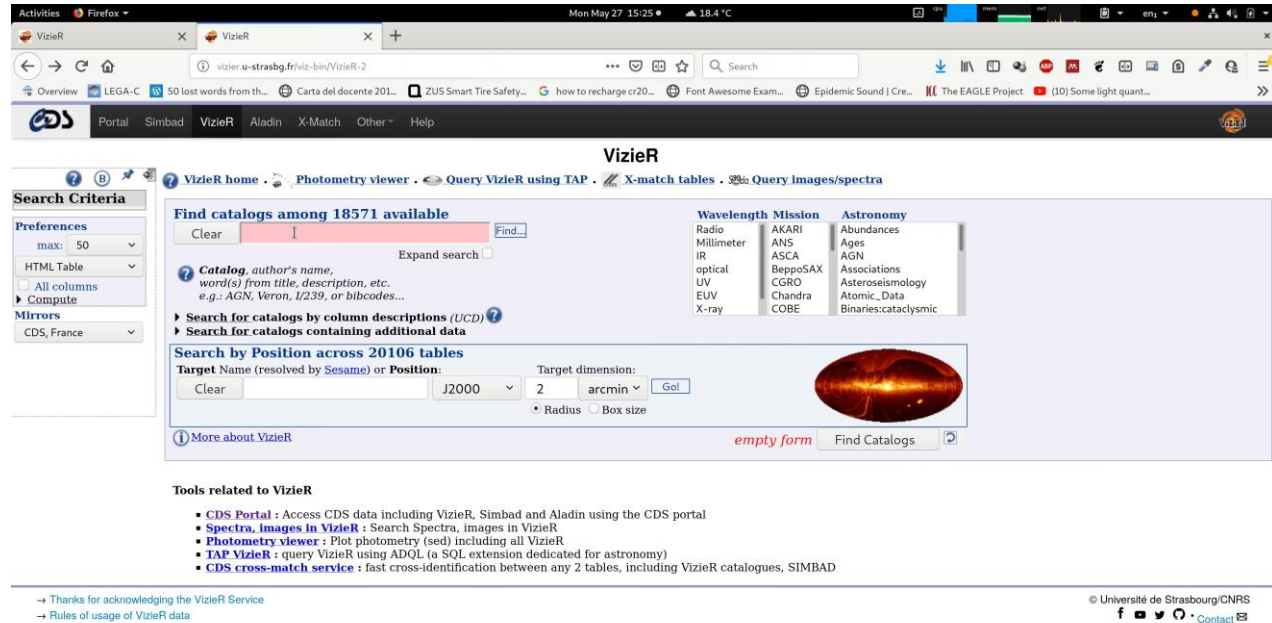
Useful to view the data, and to search data around a particular sky location

There is a “lite” version that runs in the browser, and a desktop version, which allows to download catalogues, over plot.

See here, looking for all the sources around M87 (these are mostly globular clusters)



CDS: VizieR “interactive” query



The screenshot shows the VizieR web interface in a Firefox browser window. The page title is "VizieR" and the URL is "vizier.u-strasbg.fr/viz-bin/VizieR-2". The interface includes a navigation bar with "Portal", "Simbad", "VizieR", "Aladin", "X-Match", "Other", and "Help". The main content area is titled "VizieR" and contains several search options:

- Find catalogs among 18571 available**: A search box with a "Find..." button and an "Expand search" link.
- Search by Position across 20106 tables**: A search box with a "Go!" button and a "Target dimension" dropdown set to "2 arcmin".
- Search for catalogs by column descriptions (UCD)**: A section with a "Find..." button and a "Search for catalogs containing additional data" link.

On the right side, there is a table with three columns: "Wavelength", "Mission", and "Astronomy".

Wavelength	Mission	Astronomy
Radio	AKARI	Abundances
Millimeter	ANS	Ages
IR	ASCA	AGN
optical	BeppoSAX	Associations
UV	CGRO	Asteroseismology
EUV	Chandra	Atomic_Data
X-ray	COBE	Binaries:cataclysmic

Below the search options, there is a section titled "Tools related to VizieR" with a list of links:

- CDS Portal**: Access CDS data including VizieR, Simbad and Aladin using the CDS portal
- Spectra, images in VizieR**: Search Spectra, images in VizieR
- Photometry viewer**: Plot photometry (sed) including all VizieR
- TAP VizieR**: query VizieR using ADQL (a SQL extension dedicated for astronomy)
- CDS cross-match service**: fast cross-identification between any 2 tables, including VizieR catalogues, SIMBAD

At the bottom of the page, there is a footer with the text "© Université de Strasbourg/CNRS" and social media icons for Facebook, YouTube, Twitter, and LinkedIn.

Useful to look what's around.

CDS: VizieR SQL-like query

Why?

- easy to reproduce
- easy to pass around
- pre-select the data

How?

- know table/column names
(schema browser)
- learn the SQL language (the basics)

CDS: VizieR SQL-like query

Structured Query Language – the most common query language.

Astronomers also made their own version, called Astronomy Data Query Language ADQL

A working example with SDSS



The screenshot displays the SDSS Query / CasJobs web interface. At the top left is a logo with three interlocking gears. The title "SDSS Query / CasJobs" is centered at the top. Below the title is a navigation menu with the following items: Help, Tools, Query, History, MyDB, Import, Groups, Output, Schema Browser, Queues, and SkyServer. The main content area is titled "Context" and contains a form with the following fields: "MyScratch Table (optional)" set to "default", "Task Name" set to "My Query", and "Context" set to "DR15". Below the form are buttons for "Samples", "Recent", and "Clear". The main area displays a SQL query with line numbers 1 through 19. The query is as follows:

```
1 SELECT TOP 5000
2   s.ra, s.dec,
3
4   s.z, s.zErr,
5   s.velDisp, s.velDispErr,
6   s.snMedian,
7
8   l.Sigma_Ha_6562 as sigHa, l.Sigma_Ha_6562_Err as sigHaErr,
9   l.Sigma_Hb_4861 as sigHb, l.Sigma_Hb_4861_Err as sigHbErr,
10
11  mpa.lgm_tot_p50 as lgm, (mpa.lgm_tot_p84-mpa.lgm_tot_p16)/2. as lgm_Err,
12  mpa.sfr_tot_p50 as lsfr, (mpa.sfr_tot_p84-mpa.sfr_tot_p16)/2. as lsfr_Err
13
14 INTO mydb.methclass
15 FROM SpecObj AS s
16   JOIN emissionLinesPort as l ON s.specobjid=l.specobjid
17   JOIN galSpecExtra as mpa on s.specobjid=mpa.specobjid
18
19 WHERE mpa.lgm_tot_p50 > 9 AND s.velDispErr > 0 AND l.Sigma_Ha_6562_err > 0
```

A working example with SDSS

```
SDSS Query / CasJobs
Help Tools Query History MyDB Import Groups Output Schema Browser Queues SkyServer
Context MyScratch Table (optional) Task Name
DR15 default MyTable My Query
Samples Recent Clear
1 SELECT TOP 5000
2   s.ra, s.dec,
3
4   s.z, s.zErr,
5   s.velDisp, s.velDispErr,
6   s.snMedian,
7
8   l.Sigma_Ha_6562 as sigHa, l.Sigma_Ha_6562_Err as sigHaErr,
9   l.Sigma_Hb_4861 as sigHb, l.Sigma_Hb_4861_Err as sigHbErr,
10
11  mpa.lgm_tot_p50 as lgm, (mpa.lgm_tot_p84-mpa.lgm_tot_p16)/2. as lgm_Err,
12  mpa.sfr_tot_p50 as lsfr, (mpa.sfr_tot_p84-mpa.sfr_tot_p16)/2. as lsfr_Err
13
14  INTO mydb.methclass
15 FROM SpecObj AS s
16 JOIN emissionLinesPort as l ON s.specobjid=l.specobjid
17 JOIN galSpecExtra as mpa on s.specobjid=mpa.specobjid
18
19 WHERE mpa.lgm_tot_p50 > 9 AND s.velDispErr > 0 AND l.Sigma_Ha_6562_err > 0
```


How do I know table/column names?

The screenshot shows the SkyServer DR15 Schema Browser interface. At the top, there is a navigation bar with links for Home, Data, Schema, Education, Astronomy, SDSS, Contact Us, Download, Site Search, Help, and History (marked as NEW!). The main heading is "Schema Browser". On the left side, there is a search box with a "Go" button and a list of categories: Tables, Views, Functions, Procedures, Constants, and Indices. The main content area contains several paragraphs of text explaining the database structure, including sections on Tables, Views, Indices, Functions, and Procedures. The text describes how data is organized in tables and views, and provides information on how to search for and access specific data.

SLOAN DIGITAL SKY SURVEY
SkyServer DR15

Home Data Schema Education Astronomy SDSS Contact Us Download Site Search Help History **NEW!**

Schema Browser

Glossary
Algorithms

Search for

- ▣ Tables
- ▣ Views
- ▣ Functions
- ▣ Procedures
- ▣ Constants
- ▣ Indices

The data in the database is contained in **Tables**, organized in columns and rows. We have defined **Views** over the tables. These represent special subsets of the original table.

Most of the tables also have one or more **Indices** defined on them to speed up searches on them. Please see the **Archive Intro** Help page for more information on the types of Indices.


Functions and stored **Procedures** take a number of parameters, and execute a previously defined sequence of commands. Usually, their names are prefixed by *f* or *sp*, like in *fPhotoStatus* or *spGetFiberList*.

The table **SDSSConstants** contains most of the parameters relevant to the SDSS. Their values can be displayed by clicking on the link in the left hand panel.

The table **DataConstants** contains most of the bit-flags and enumerated quantities relevant to the SDSS. Their values " can be displayed by clicking on the link in the left hand panel. There are several access functions to make interpretations and the " back and forth conversions easier. They are displayed when you look" at the individual enumerated fields."

Click on the **[+]** sign in front of the categories on the left to get an expanded view, on the **[-]** to collapse the list, on the name to get a summary, and on the individual items to get their detailed description.

Enter a keyword into the search box on the top, and press Enter or the **Go** button to search through the descriptions and units stored in the database.

The symbol , appearing at various places in the documentation indicates a link to look up the values of various flags and enumerated constants.

Schema Browser

Glossary Algorithms

Search for Go

TABLE galSpecExtra

Estimated physical parameters for all galaxies in the MPA-JHU spectroscopic catalogue.

The estimates for stellar mass are derived using the methodology described in Kauffmann et al (2003), applied to photometric data as described in Salim et al (2007). The star formation rates are derived as discussed in Brinchmann et al (2004), but the aperture corrections are done by estimating SFRs from SED fits to the photometry outside the fiber following the methodology in Salim et al (2007).

name	type	length	unit	ucd	description
specObjID	numeric	13		ID_CATALOG	Unique ID
bptclass	smallint	2			Emission line classification based on the BPT diagram using the methodology described in Brinchmann et al (2004). -1 means unclassifiable, 1 is star-forming, 2 means low S/N star-forming, 3 is composite, 4 AGN (excluding liners) and 5 is a low S/N LINER.
oh_p2p5	real	4			The 2.5 percentile of the Oxygen abundance derived using Charlot & Longhetti models. The values are reported as 12 + Log O/H. See Tremonti et al (2004) and Brinchmann et al (2004) for details.
oh_p16	real	4			The 16 percentile of the Oxygen abundance derived using Charlot & Longhetti models. The values are reported as 12 + Log O/H. See Tremonti et al (2004) and Brinchmann et al (2004) for details.

Example from SDSS Schema:

Search for keywords
Search by table/view

SLOAN DIGITAL SKY SURVEY

SkyServer DR15

Home Data Schema Education Astronomy SDSS Contact Us Download Site Search Help History **NEW!**

Schema Browser

Glossary Algorithms


SEARCH FOR ...mass...

Columns

Search for Go

tablename	name	unit	description
apogeeObject	alt_id		Alternate name for non-2MASS objects
apogeeObject	j	mag	2MASS J-band magnitude
apogeeObject	j_err	mag	2MASS J-band magnitude error
apogeeObject	k	mag	2MASS Ks-band magnitude
apogeeObject	k_err	mag	2MASS Ks-band magnitude error
apogeeObject	tmass_a		Source of 2MASS optical counterpart
apogeeObject	tmass_ccfg		2MASS contamination flag
apogeeObject	tmass_extkey		2MASS Extended Source Catalog ID of associated source

Our query:



The screenshot shows the SDSS Query / CasJobs web interface. At the top left is a logo with three interlocking gears. The title "SDSS Query / CasJobs" is in the top right. Below the title is a navigation menu with items: Help, Tools, Query, History, MyDB, Import, Groups, Output, Schema Browser, Queues, and SkyServer. The main area is titled "Context" and contains three dropdown menus: "Context" (set to DR15), "MyScratch Table (optional)" (set to default), and "Task Name" (set to MyTable). Below these are buttons for "Samples", "Recent", and "Clear". The main content area displays a SQL query with line numbers 1 through 19. The query selects the top 5000 rows from a table, listing various astronomical parameters and joining several tables.

```
1 SELECT TOP 5000
2   s.ra, s.dec,
3
4   s.z, s.zErr,
5   s.velDisp, s.velDispErr,
6   s.snMedian,
7
8   l.Sigma_Ha_6562 as sigHa, l.Sigma_Ha_6562_Err as sigHaErr,
9   l.Sigma_Hb_4861 as sigHb, l.Sigma_Hb_4861_Err as sigHbErr,
10
11  mpa.lgm_tot_p50 as lgm, (mpa.lgm_tot_p84-mpa.lgm_tot_p16)/2. as lgm_Err,
12  mpa.sfr_tot_p50 as lsfr, (mpa.sfr_tot_p84-mpa.sfr_tot_p16)/2. as lsfr_Err
13
14 INTO mydb.methclass
15 FROM SpecObj AS s
16   JOIN emissionLinesPort as l ON s.specobjid=l.specobjid
17   JOIN galSpecExtra as mpa on s.specobjid=mpa.specobjid
18
19 WHERE mpa.lgm_tot_p50 > 9 AND s.velDispErr > 0 AND l.Sigma_Ha_6562_err > 0
```

methclass.fit

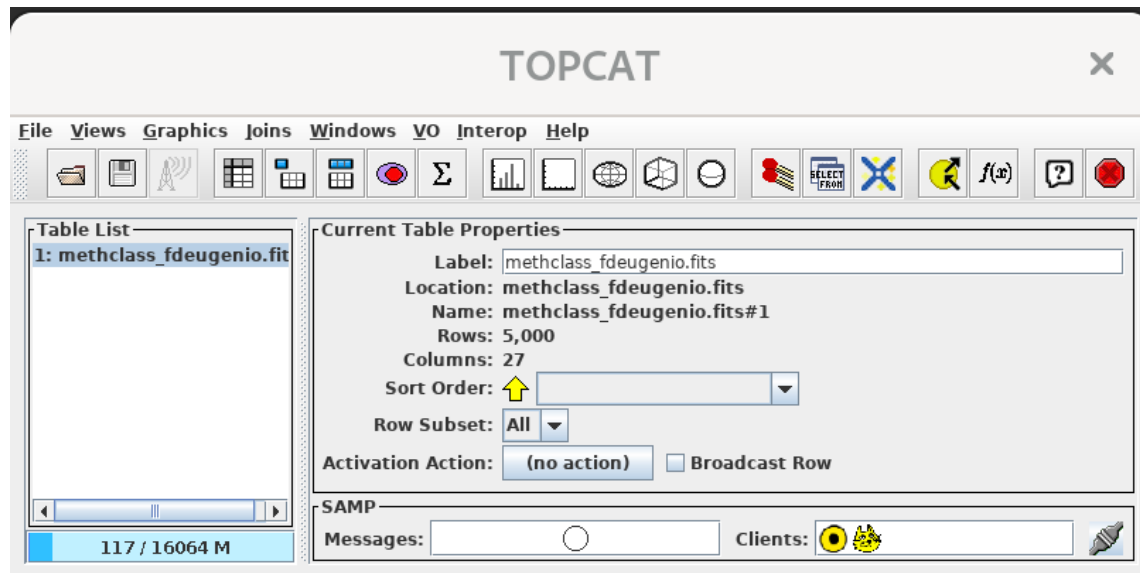
II. Data procurement: Summary

- Archives (catalogues) (SQL, Schema)
- Archives (images/spectra)
- Observations

III. Quick analysis tools

- What is it? A test that takes < 30'
- Why? Evaluate data quality, feasibility
- What tools?
 algorithms Plotting & robust fitting

Topcat: see the data



Topcat:
see the data

<date/time>

TOPCAT(1): Table Browser



Window Subsets Help

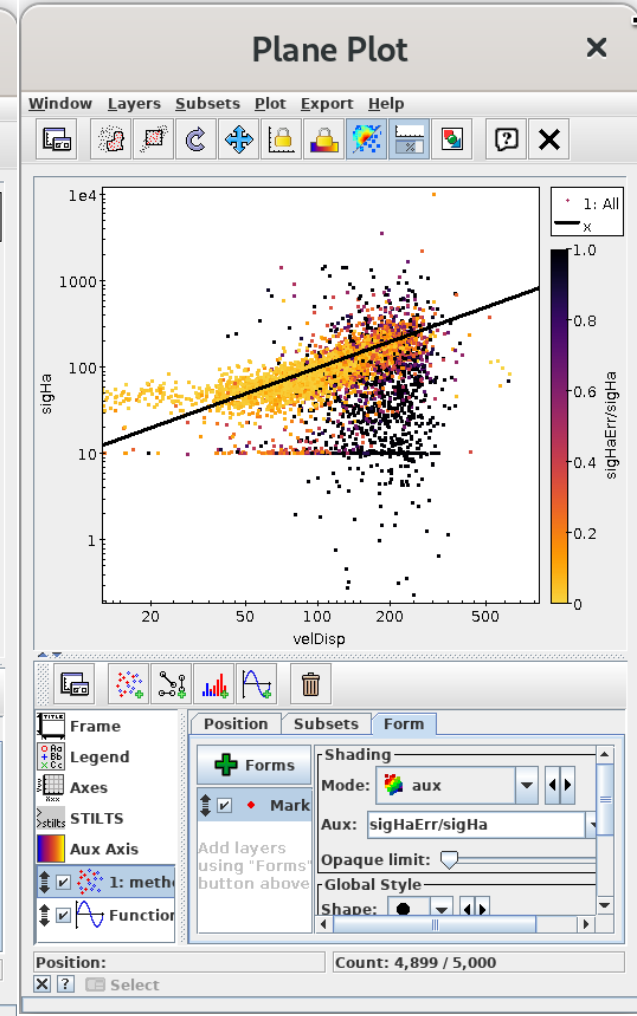
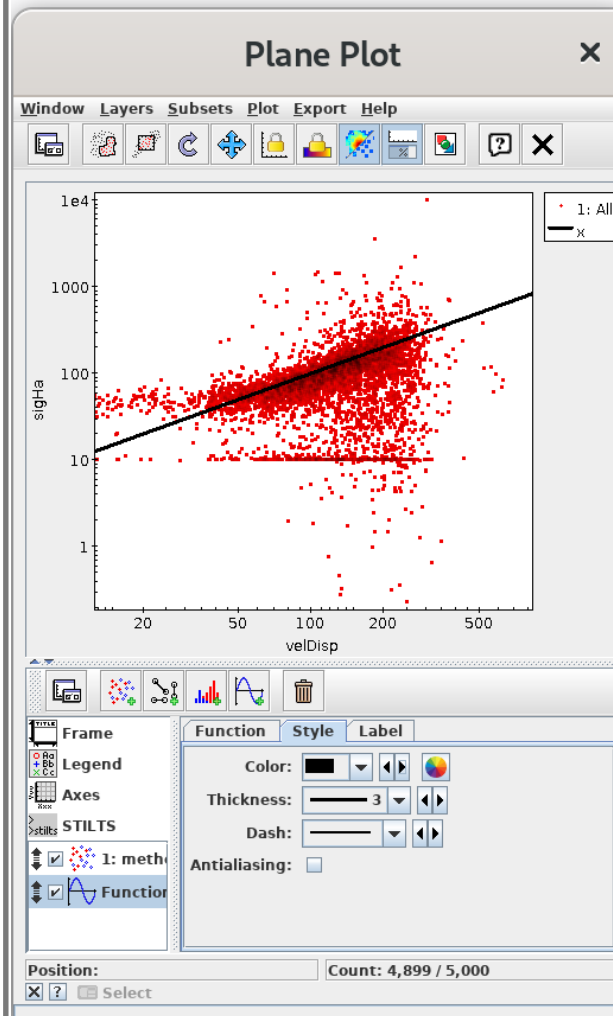


Table Browser for 1: methclass fdeugenio.fit

	ra	dec	z	zErr	velDisp	velDispErr	snMedian	
1	146.07944	-0.68096	0.067273	2.26460e-5	139.946	9.52048	23.6345	20
2	146.22225	-0.37589	0.067406	1.47889e-5	86.9503	6.60209	23.7421	5
3	145.68114	-0.86722	0.067595	1.34051e-5	151.046	5.14736	28.138	5
4	145.34424	-0.3685	0.204814	3.85925e-5	218.041	13.3036	10.7861	110
5	145.33123	-0.54207	0.153353	1.79960e-5	143.079	25.6172	8.06431	12
6	146.25244	1.11958	0.120132	3.69646e-5	184.509	12.289	13.4383	21
7	146.84222	0.74084	0.020131	6.66649e-6	63.9882	12.4257	14.5219	1
8	147.12735	0.50446	0.086057	1.67476e-5	97.1918	9.38074	13.0875	6
9	148.31755	-1.05055	0.139379	2.90861e-5	160.118	10.2306	12.9407	21
10	148.79165	-0.02102	0.085213	1.10301e-5	121.832	20.8426	10.0832	10
11	148.50575	-0.11386	0.127313	2.84980e-5	137.955	10.5404	13.7249	8
12	147.69322	0.13054	0.069145	9.18849e-6	135.601	16.5603	11.3991	8
13	147.91036	-0.57614	0.138029	3.85532e-5	167.963	13.4732	11.3162	79
14	147.46847	-0.45819	0.328609	7.38732e-5	228.161	26.9164	5.10561	129
15	147.16557	-0.89133	0.167423	3.99187e-5	208.962	16.7285	12.2723	5
16	146.78847	-0.31066	0.056614	1.13770e-5	72.7607	33.1433	4.30038	2
17	147.74002	0.61567	0.163001	3.09153e-5	230.855	11.5988	14.833	19
18	147.86563	0.502	0.080631	1.24978e-5	199.665	7.60252	24.6108	18
19	147.94533	0.33048	0.084409	1.60110e-5	185.044	6.34636	25.8529	14
20	150.37774	-0.13016	0.089852	2.85727e-5	127.846	9.56837	15.1371	16
21	150.51385	-0.2138	0.045162	7.81269e-6	22.5576	41.9494	7.7782	5
22	149.83076	-0.40297	0.072408	2.25239e-5	122.357	8.01675	17.7386	
23	149.19869	-1.05644	0.093009	3.29696e-5	144.224	12.2589	12.5113	4
24	149.42268	-0.31072	0.087942	1.10970e-5	65.7446	10.2425	15.2187	3
25	149.47209	-0.36201	0.168546	2.97094e-5	202.903	10.6778	15.3426	19
26	149.23315	-0.39416	0.089092	1.72371e-5	107.664	8.82474	15.1249	10
27	148.24123	-0.79525	0.088637	2.41463e-5	127.856	7.84762	16.4083	15
28	149.007	-0.24103	0.083715	2.04446e-5	177.642	7.8413	22.0525	18
29	149.05682	0.75387	0.034631	9.25833e-6	24.957	59.3203	7.51928	3

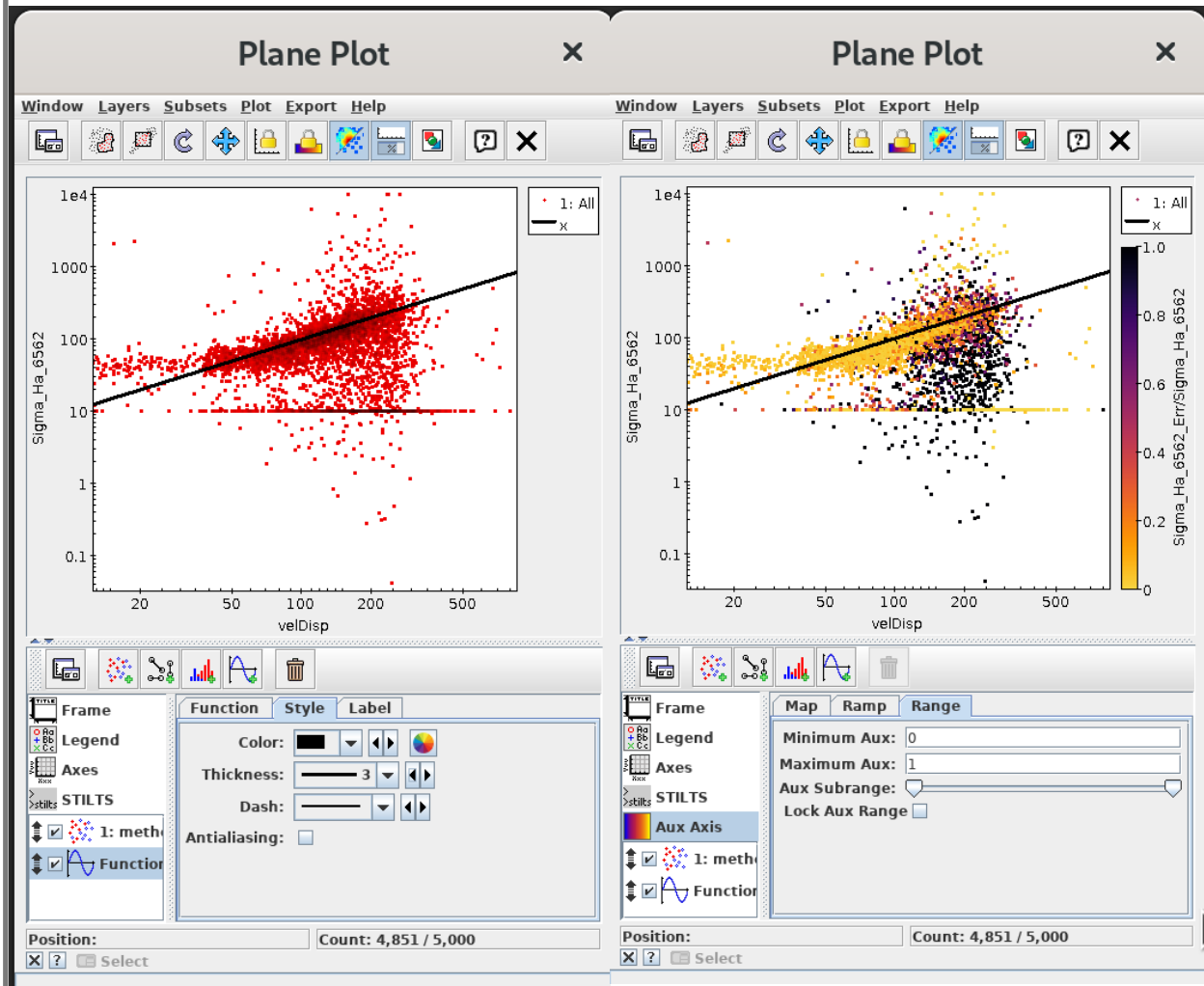
Visually inspecting the data

identify outliers
identify problems with the data



Visually inspecting the data

identify outliers
identify problems with the data



Aladin v10.0

File Edit Image Catalog Overlay Coverage Tool View Interop Help

Available data Command Frame ICRS Projection Altoff

■ DSS ■ PanSTARRS ■ SDSS ■ 2MASS ■ GALEX ■ Gaia ■ Simbad ■ NED +

- ▼ Collections → 23124
 - ▼ Image → 409
 - ▼ Gamma-ray → 23
 - ▼ X-ray → 36
 - ▼ UV → 27
 - ▼ Optical → 84
 - ▼ HST → 28
 - ▼ Skymapper → 7
 - ▼ SDSS → 7
 - SDSS9 band u
 - SDSS9 band g
 - SDSS9 color (alt)**
 - SDSS9 color
 - SDSS9 band r
 - SDSS9 band i
 - SDSS9 band z
 - ▼ DES → 1
 - ▼ CFHTLS → 12
 - ▼ Swift → 6

SDSS9 color-alt

360° x 180°

select from -- all collections --

col sort view scan filter
 grid study wink north fdr multiview match

select
 pan
 dist
 phot
 draw
 tag
 moc
 spect
 filter
 cross
 copy
 rgb
 assoc
 crop
 cont
 pixel
 prop
 del

methclass_fdeuge
 methclass_fdeuge
 methclass_fdeuge
 CDS/P/DES-DR1/

epoch
 size
 dens.
 opac.
 zoom

000.0000

00:00:00.00 +00:00:00.0
360° x 180°

Search

Plane Plot

Window Layers Subsets Plot Export Help

Frame Legend Axes STILTS
 1: meth

Range	Grid	Labels	Font
Coords			
<input checked="" type="checkbox"/> X Log	<input checked="" type="checkbox"/> Y Log	<input type="checkbox"/> X Flip	<input type="checkbox"/> Y Flip
Navigation			
<input type="checkbox"/> Aspect Lock			

Position: Count: 4,867 / 5,000

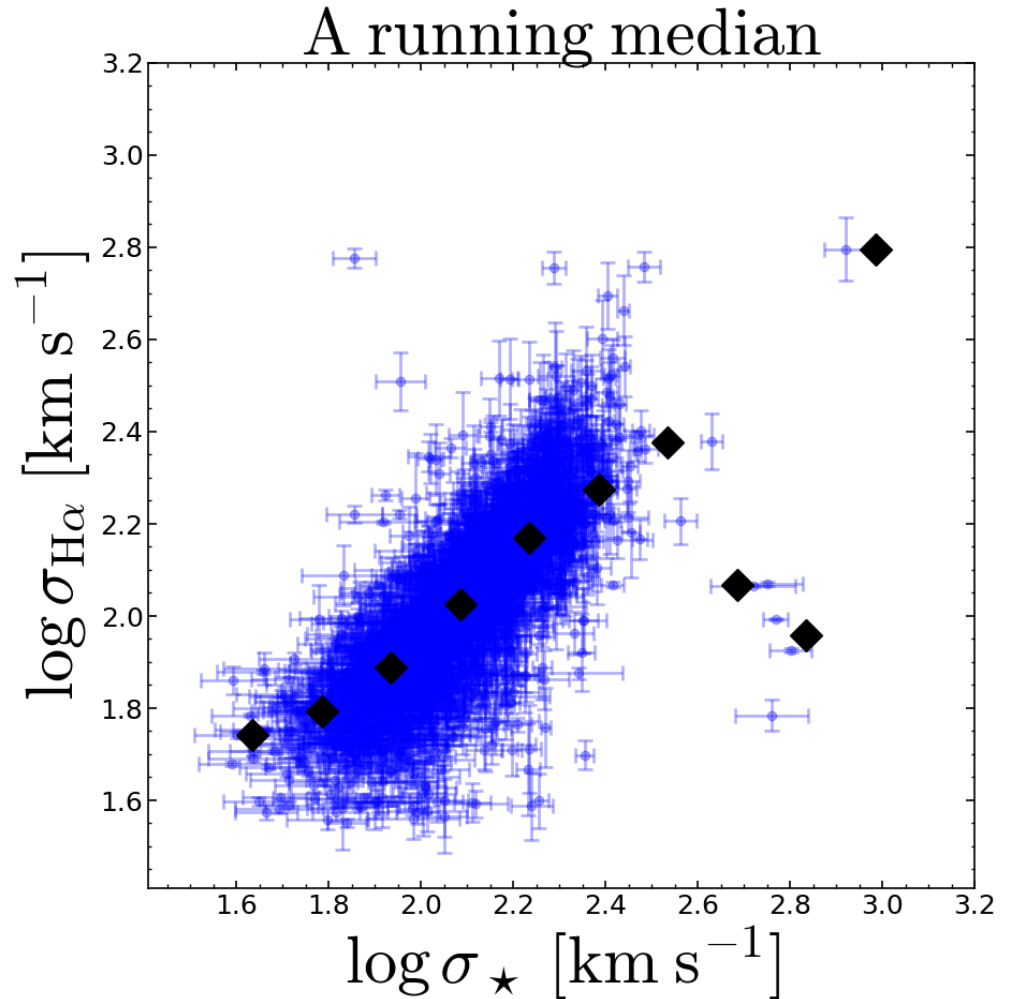
Select

A running median can help...

in python:

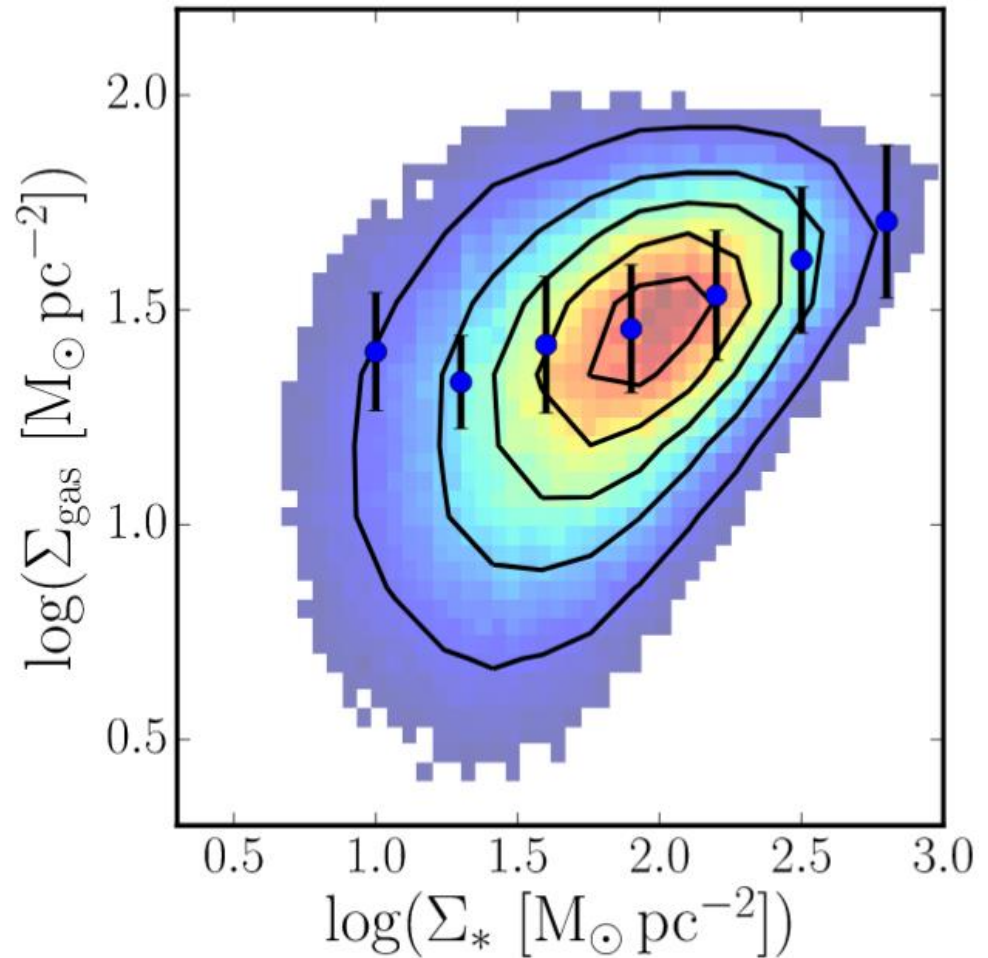
```
>>> binned_statistic(sigstar, sigHa,  
bins=..., statistic=np.median)
```

```
python3 methclass_running_median.py
```



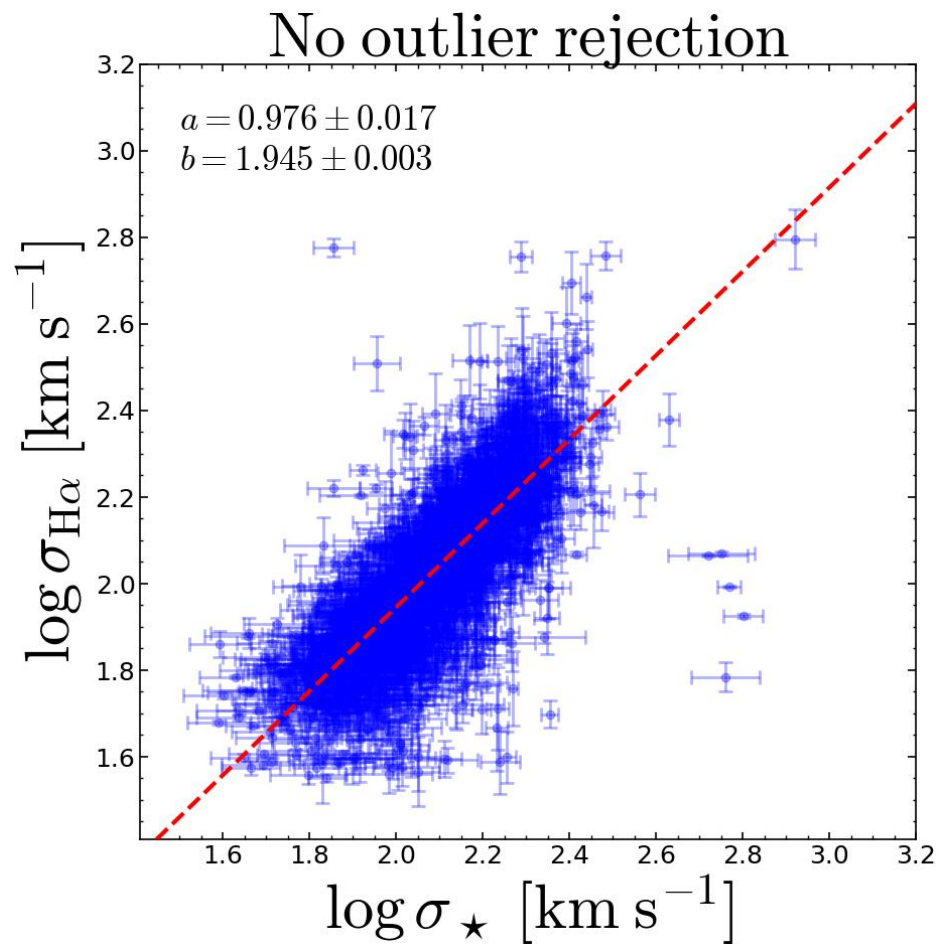
A running median
can help...

—
...but be sensible!



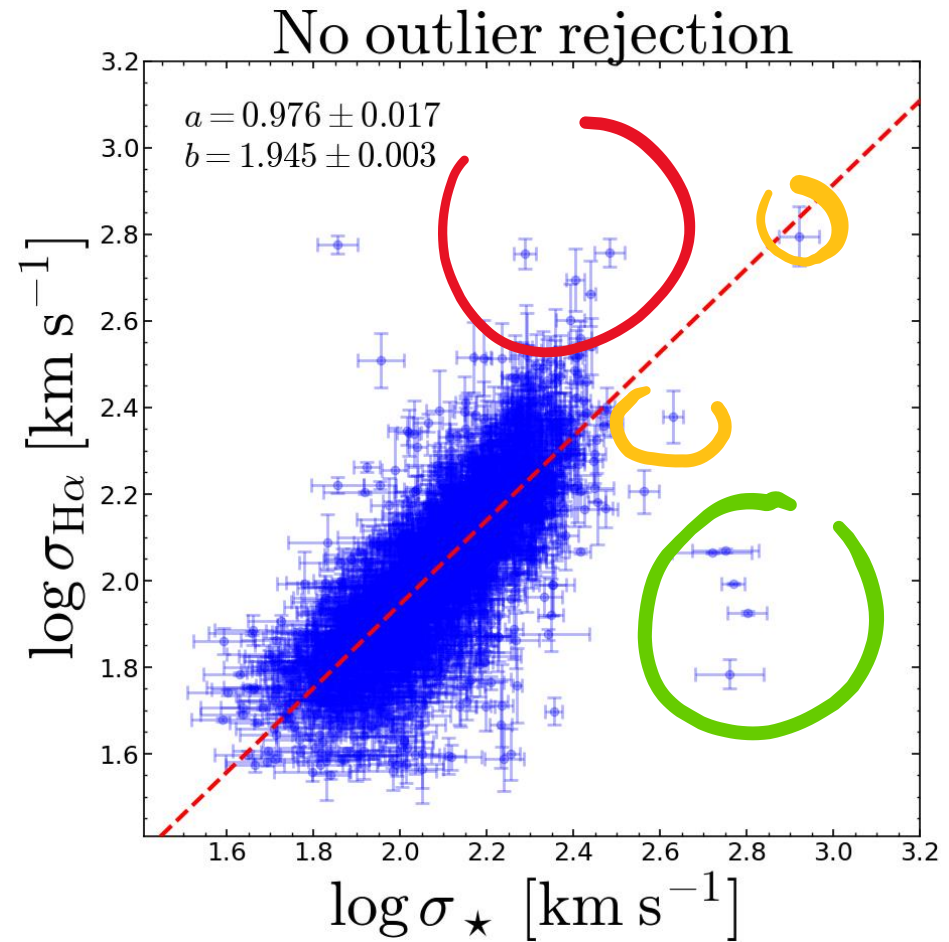
Least Squares

Something's wrong...



Need to reject
outliers!

Difficult and often lengthy
process
Labour intensive



Least Trimmed Squares (LTS)

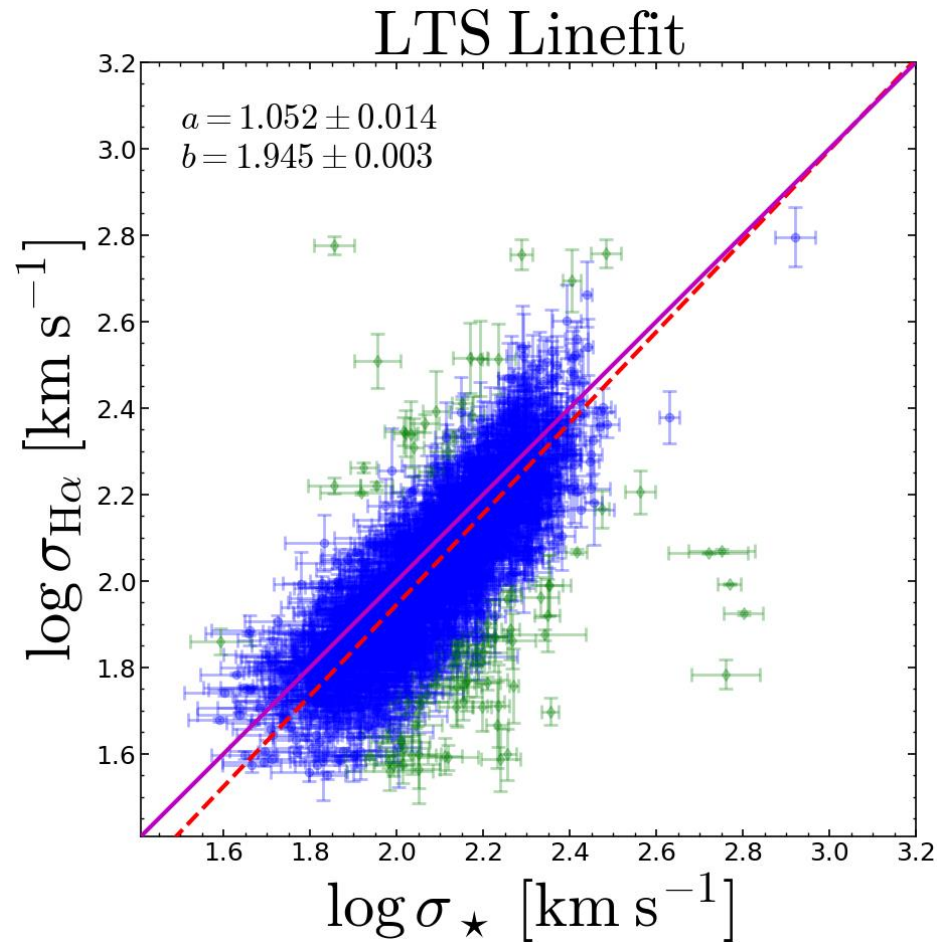
Least squares minimization +
sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

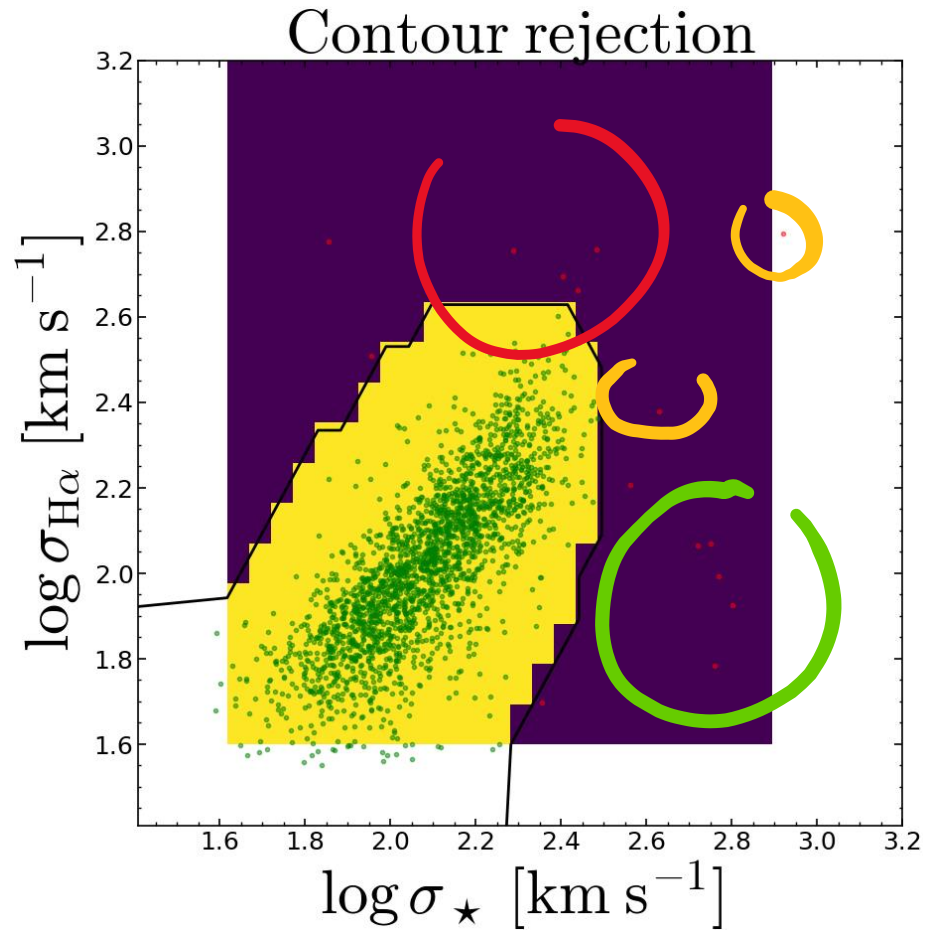
slope = $b = 1.08 \pm 0.02$

python3 methclass_lts_linefit.py



Contour-based rejection

Quick to implement
Can be recycled for Bayesian approach



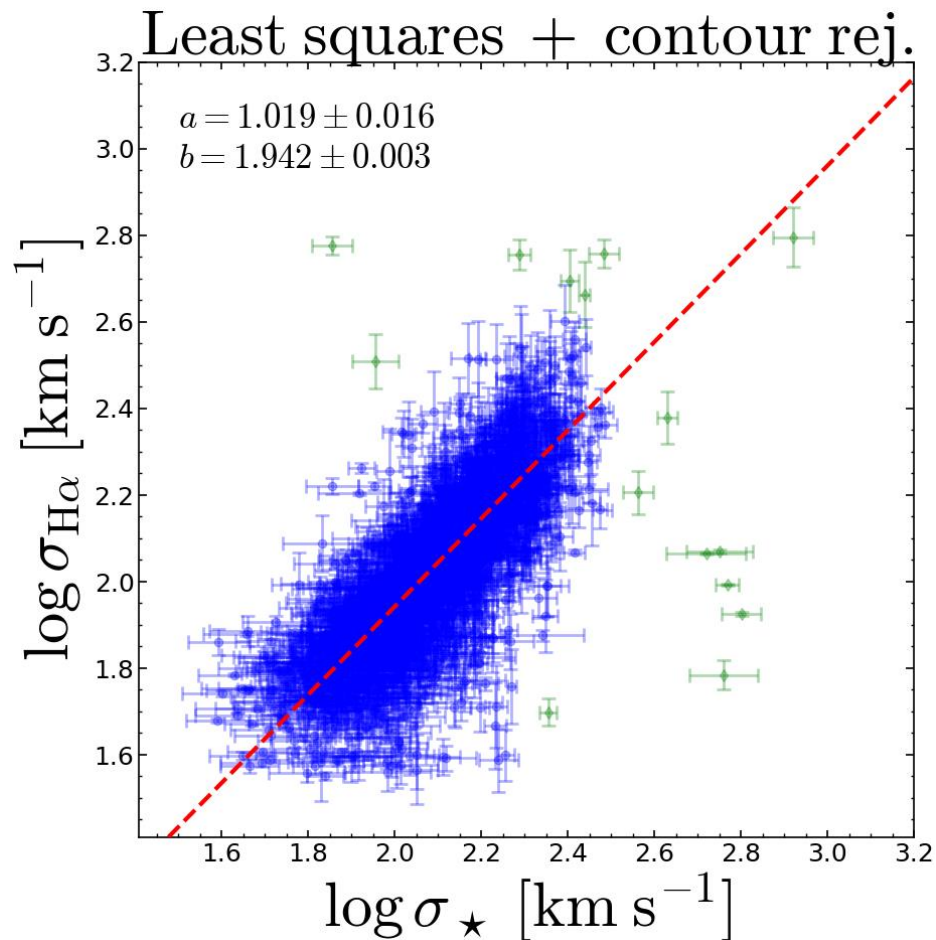
Applying contour rejection.

Least squares minimization +
sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



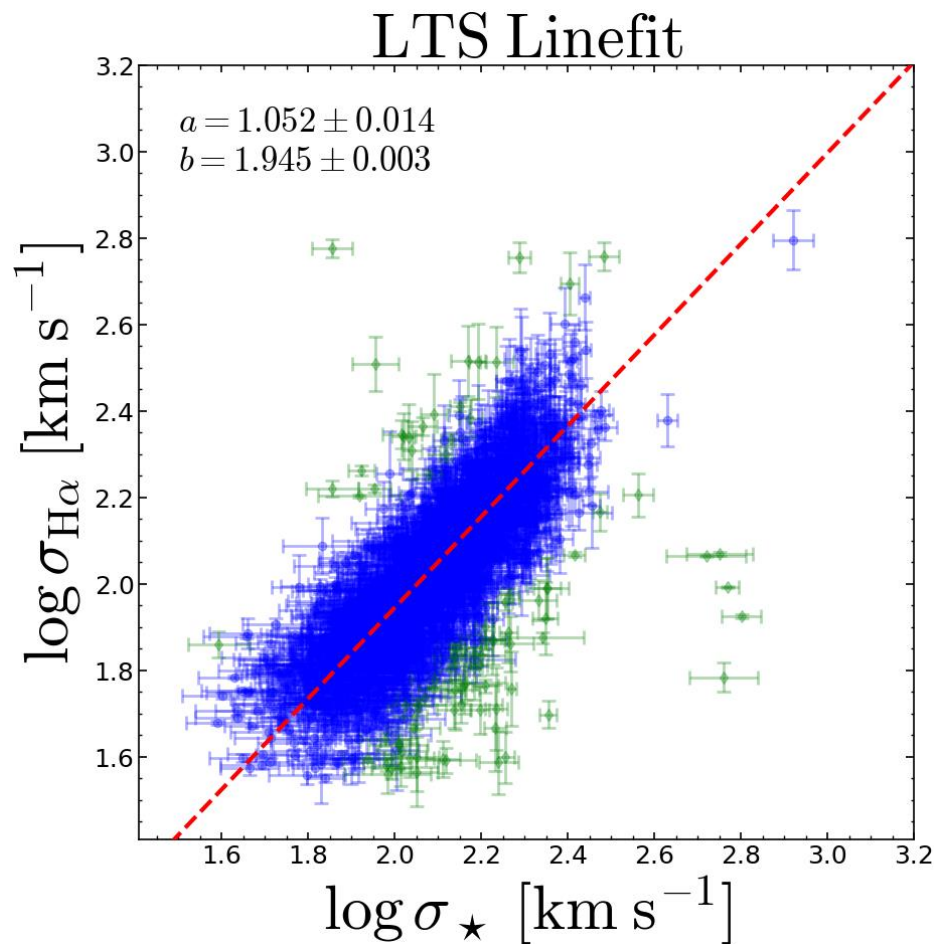
Towards a probabilistic model

Least squares minimization + sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



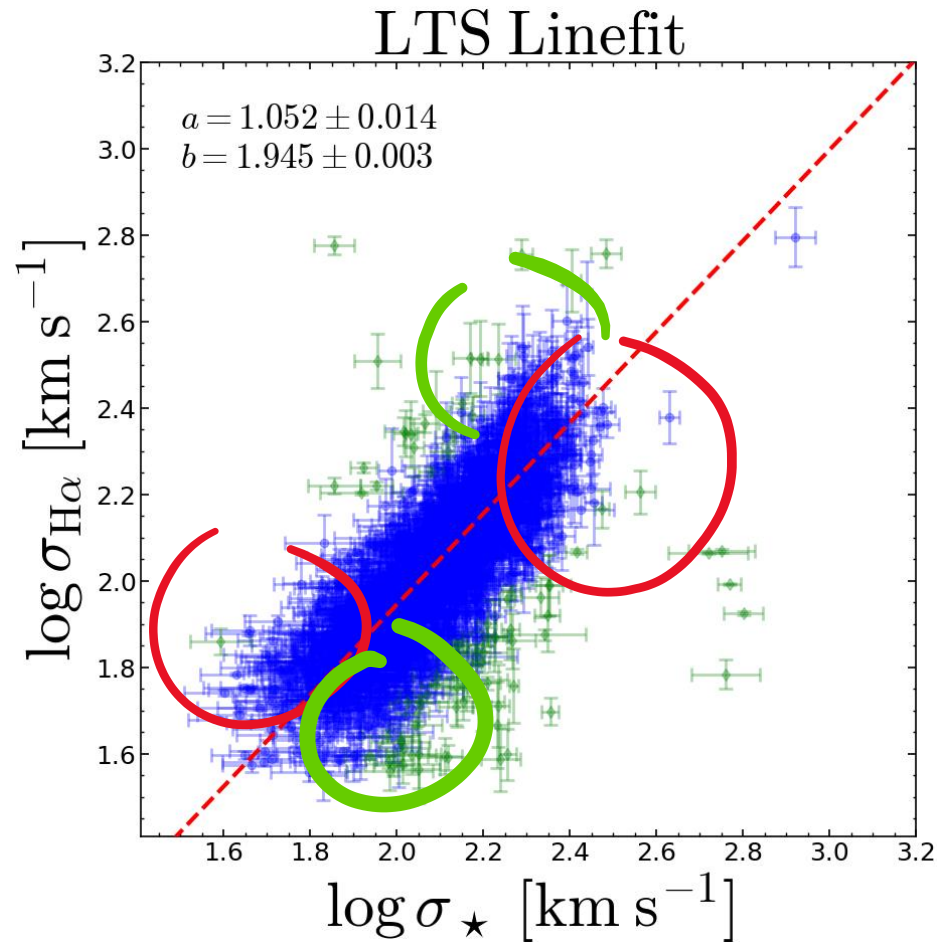
Least Trimmed Squares (LTS)

Least squares minimization +
sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



How to construct an adequate model?

create a probabilistic model

$$p(model|data) \propto p(model) \times p(data|model)$$

ALL algorithms we have seen so far assume a model

...know thy model

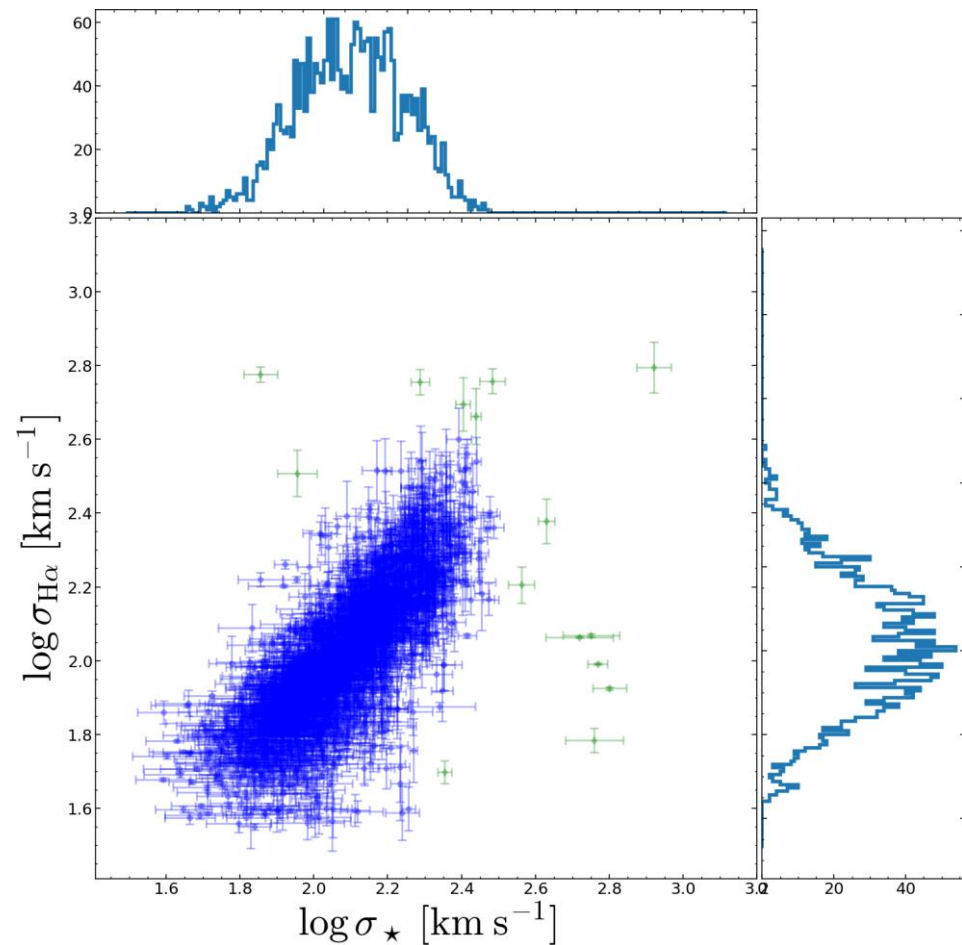
Towards a probabilistic model

Least squares minimization +
sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



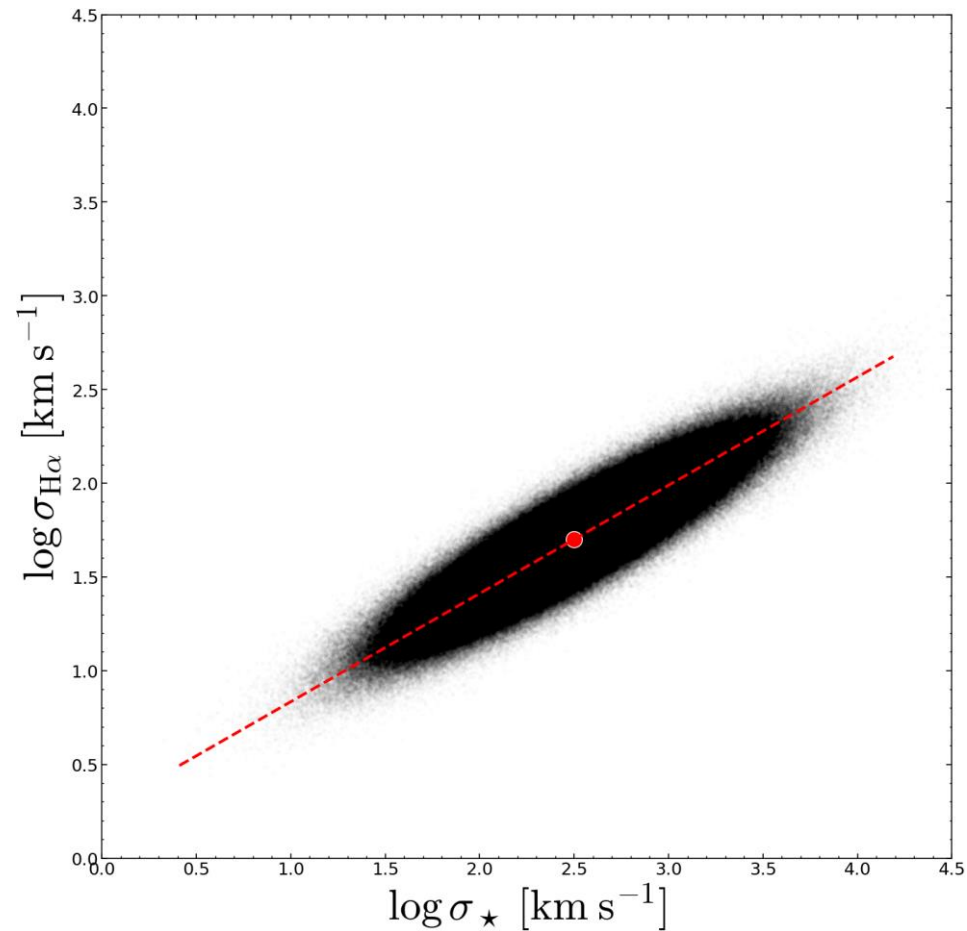
Towards a probabilistic model

Least squares minimization +
sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



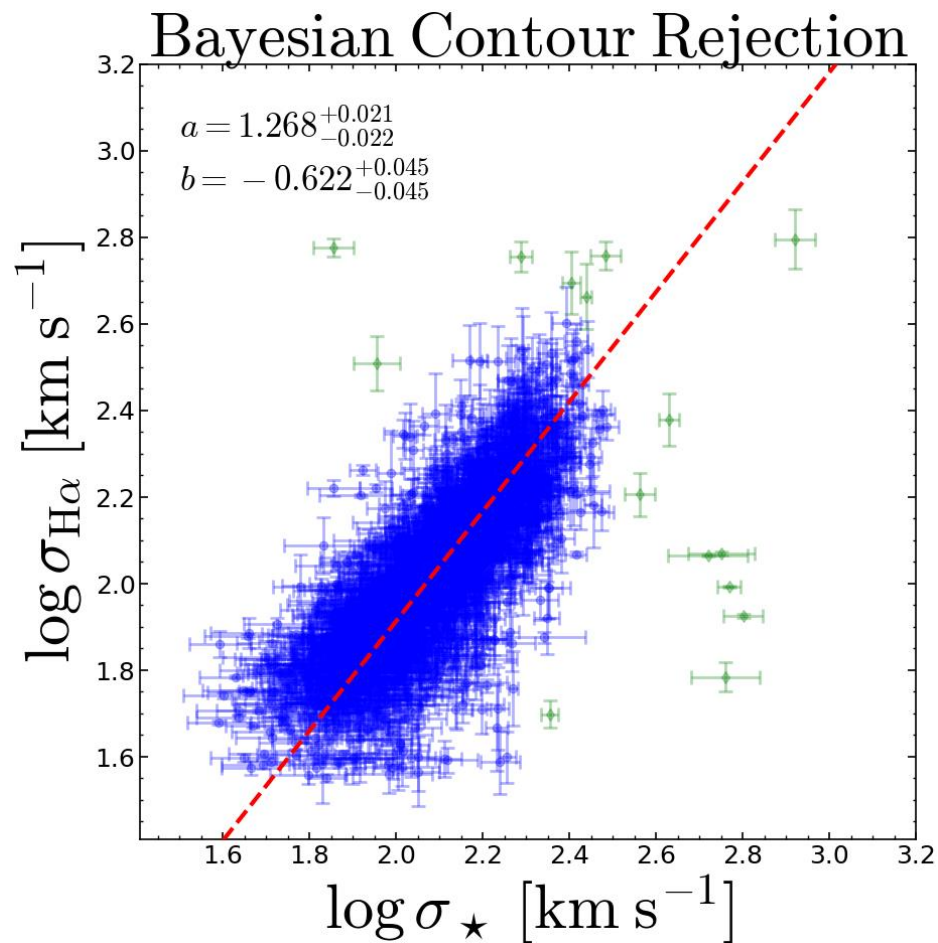
Towards a probabilistic model

Least squares minimization + sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



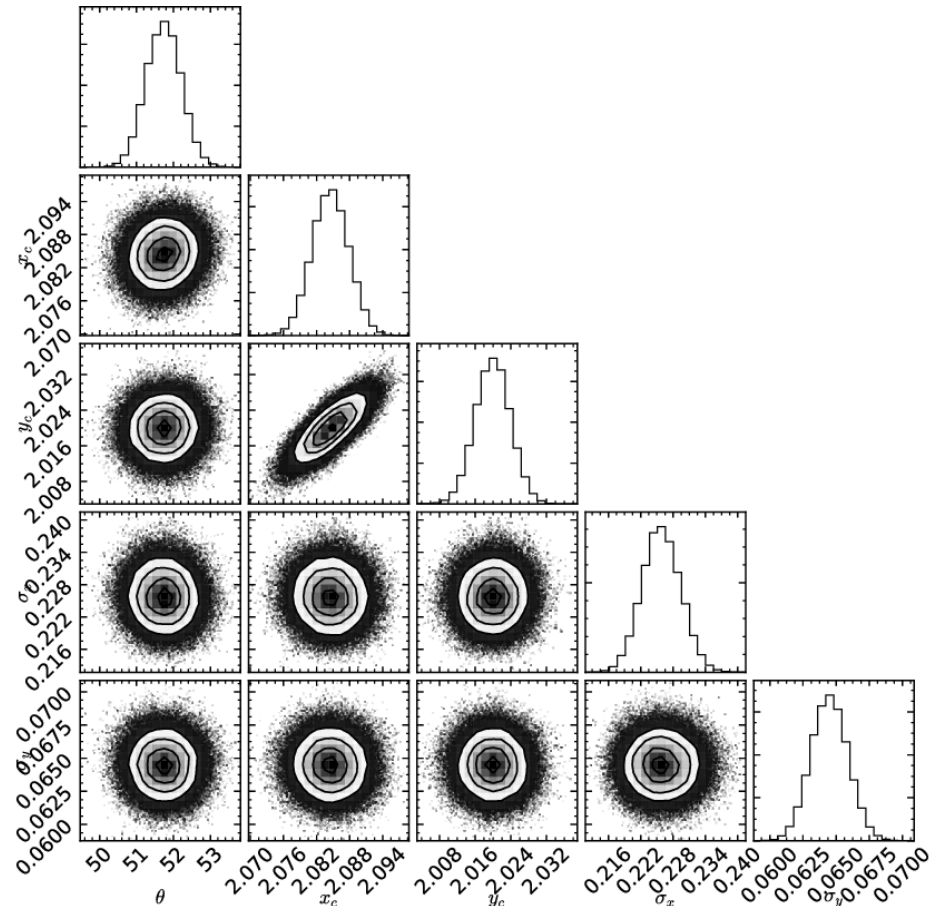
Towards a probabilistic model

Least squares minimization + sigma clipping but:

- proceeds inside-out
- guaranteed to converge

For the record:

slope = $b = 1.08 \pm 0.02$



Locally wEighted Scatter-plot Smoother = LOESS

