CHASING PHYSICS BEHIND STAR FORMATION EFFICIENCY

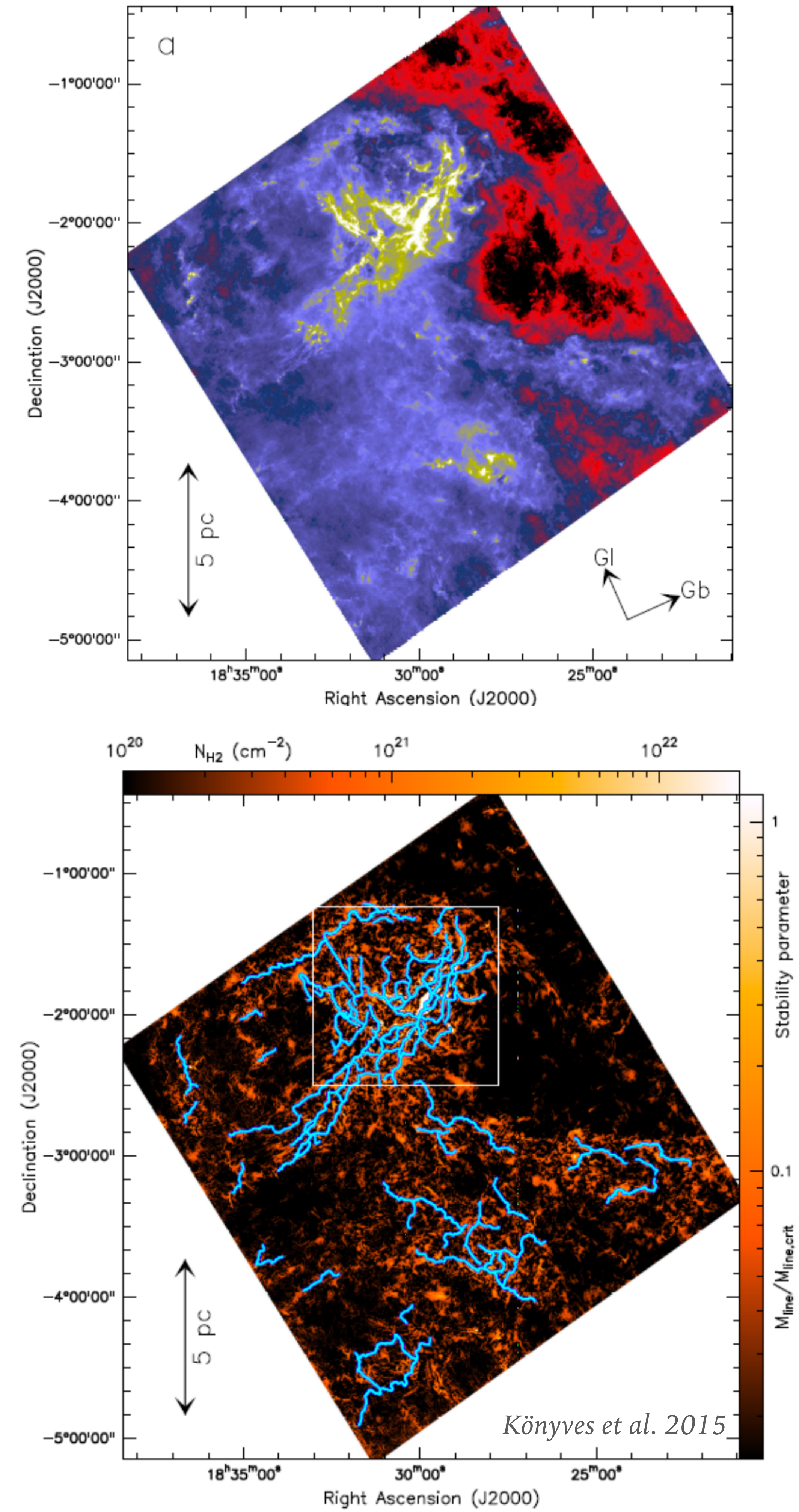
with photometry and statistics

G. Marton, Konkoly Observatory

Cold Cores meeting, Budapest, 2016

INTERSTELLAR MEDIUM

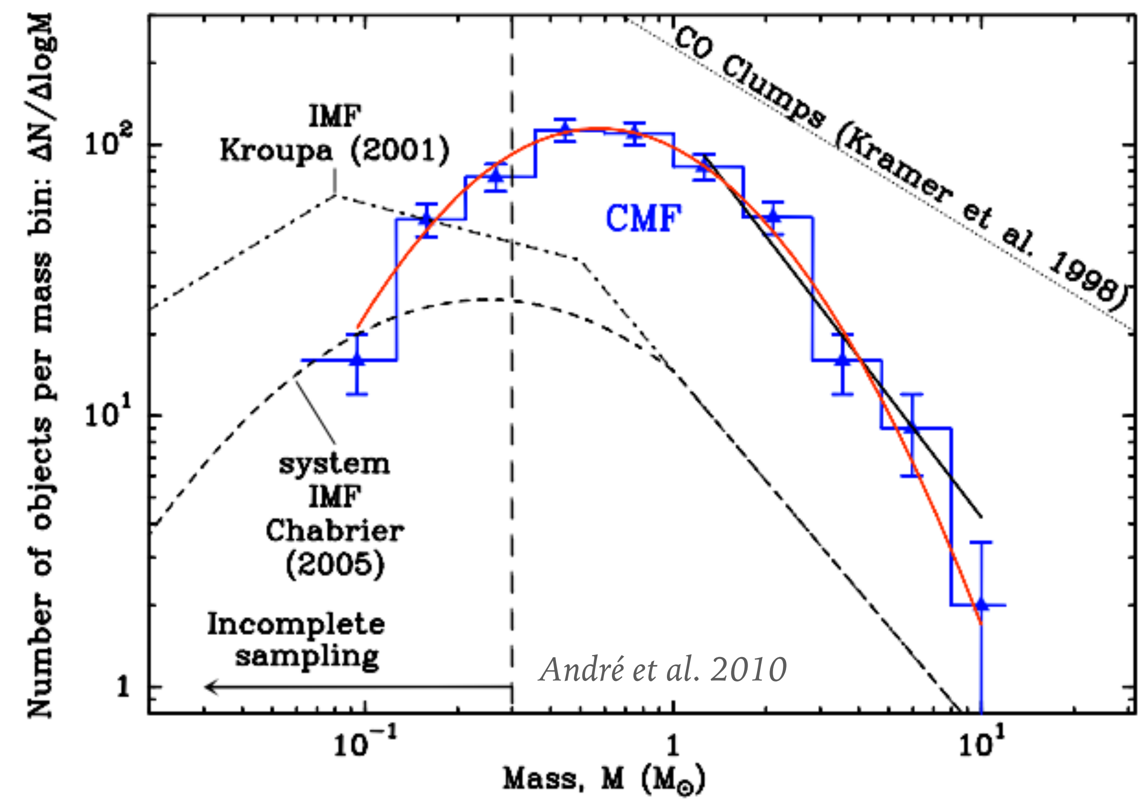
- Stars are born in the densest patches of the Interstellar Medium (ISM)
- Discrete building blocks of the hierarchical structure:
  - Giant Molecular Clouds: $M \sim 10^5\ M_\odot$, $\rho \sim 100\mathrm{cm}^{-3}$, > 10 pc
  - Clouds: $M < 10^2\ M_\odot$, $\rho > 100\mathrm{cm}^{-3}$, few pc
  - Clumps: $10 < M_\odot < 10^2$, $100\mathrm{cm}^{-3} < \rho < 10^4\mathrm{cm}^{-3}$, ~ 1 pc
  - Cores: $M < 10\ M_\odot$, $\rho > 10^4\mathrm{cm}^{-3}$, ~ 0.1 pc
- Continuous hierarchical structure of the diffuse ISM:
  - Self similar, scale invariant (fractal scaling)
  - Filaments permeate the ISM on all scales (Ph. André, PPVI)
- Dust-to-gas mass ratio: $\sim 1/100$

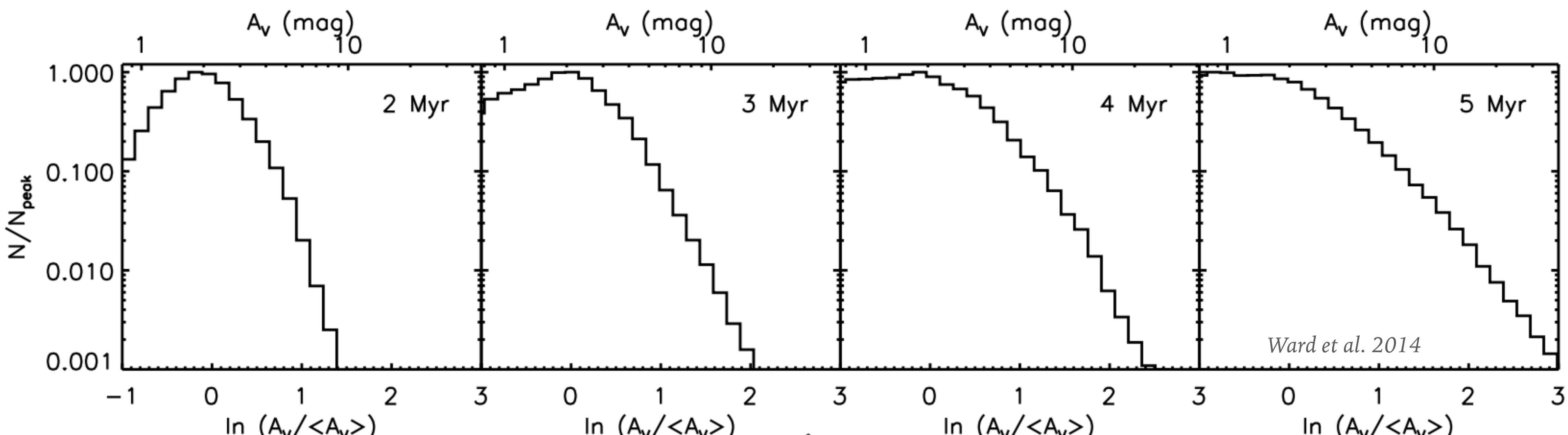


Könyves et al. 2015

FROM LARGE TO SMALL SCALES

- Initial conditions are turbulence-induced
- Pressure induced density enhancements form
- In the dense regions gravity becomes dominant
- Gravitationally bound parts collapse and form pre-stellar cores
- The infall converts potential energy into heat and the increasing density makes an obscure envelop
- Cores become dense and hot enough to start hydrogen fusion - Class 0
- Envelope and disk evolution - Class I - II - III





Time

Turbulence-induced

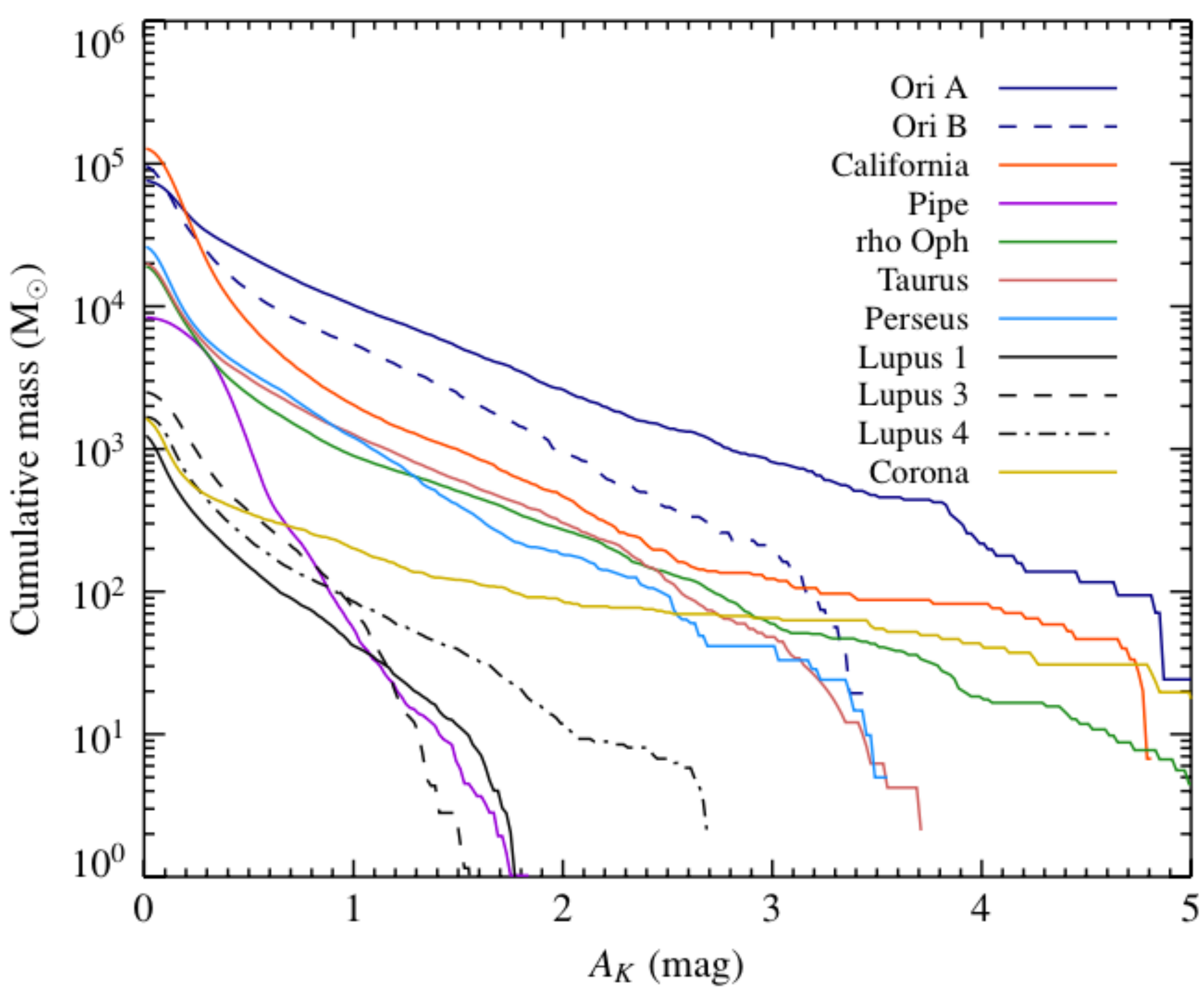
Gravity-dominated

STAR FORMATION RATE

- Pipe Nebula & ρ Ophiuchi
- Similar distance (130&120 pc)
- Similar total mass (8000 and 14000 $M_{\odot}$)
- Very different star formation efficiency (~20 & ~300 YSOs)

Pipe Nebula *C. Federrath - PPVI, Lada et al. 2010* *ρ Ophiuchi*

$$SFR_{\rho Oph} = 15 \times SFR_{Pipe}$$



- *"a linear relationship between the SFR and the mass of a cloud above a threshold extinction of $A_K \approx 0.8$ mag"*
- *"understanding the origin of the dense component of the molecular interstellar medium and how that component evolves may provide the key to the development of a predictive theory that links star formation and galaxy evolution"*

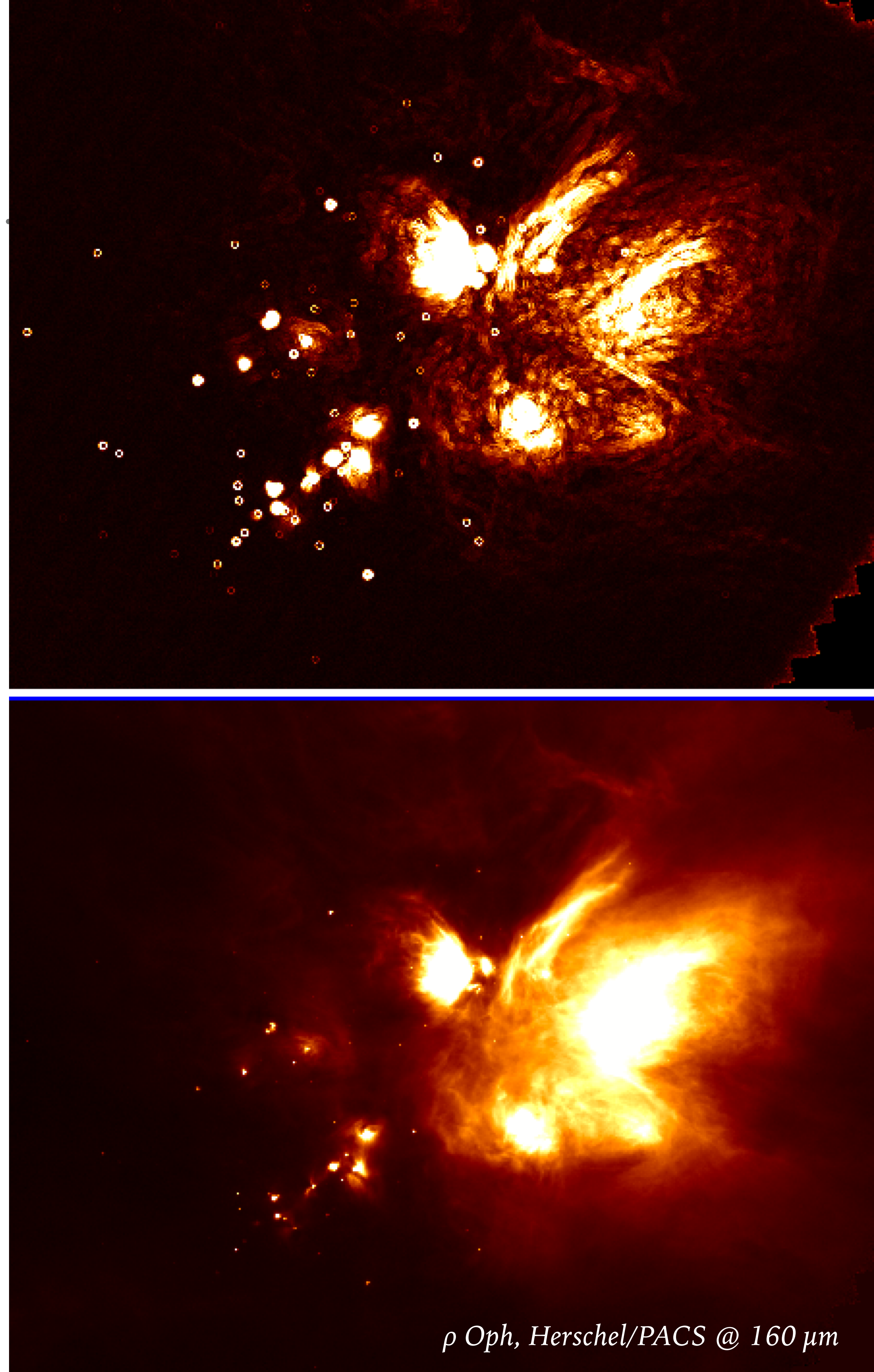
Lada et al. 2010

- *Star formation is linked to the internal cloud structure*
- *Cloud structure carries imprints of fundamental physical processes*

Kainulainen 2015 - IAU Divisional Meeting

IMAGE ANALYSIS TECHNIQUES – CONSTRAINING THE COMPLEXITY

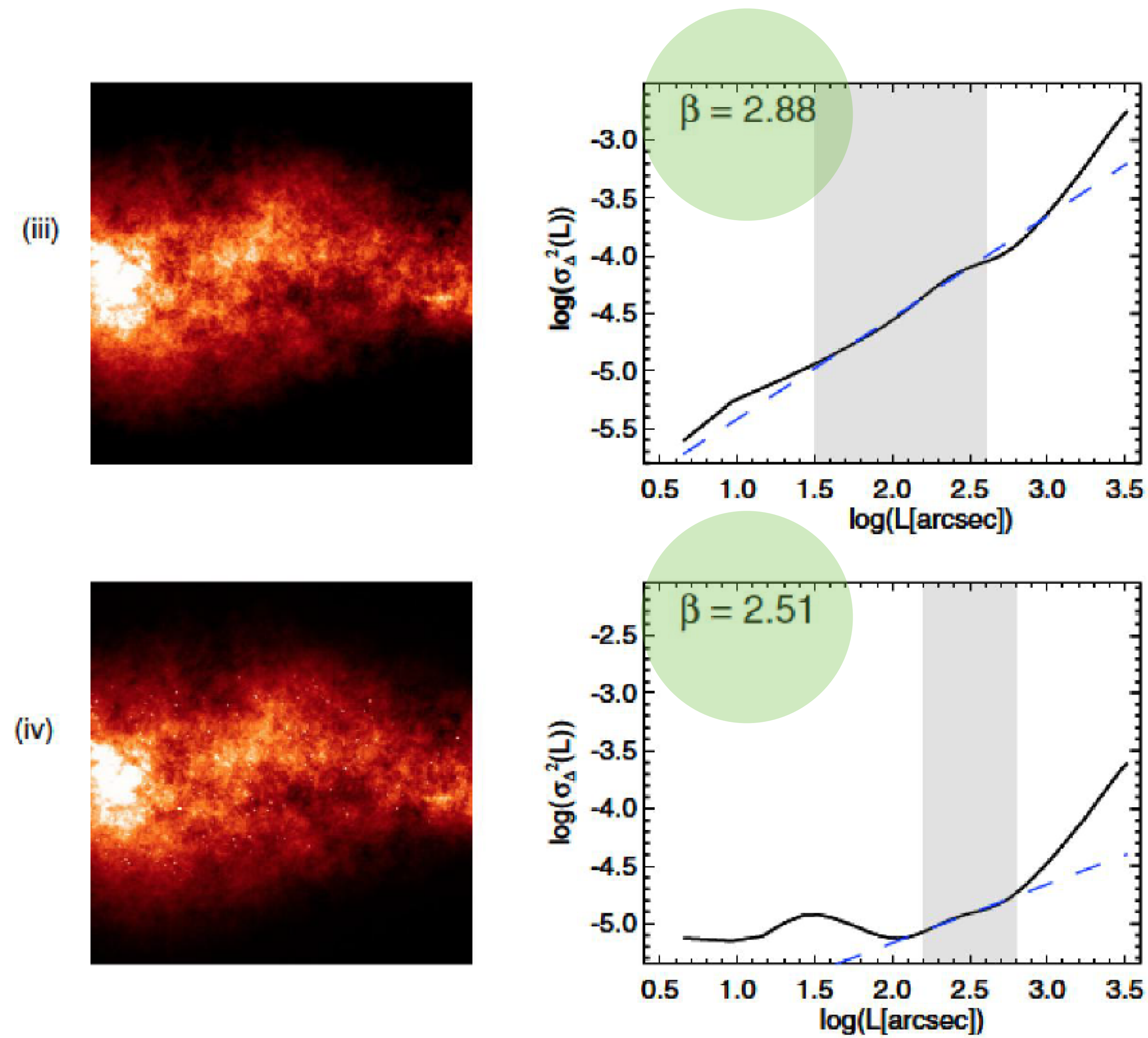
- The observed morphology can be compared to measured SFR
- **Probability Density Function** - column density N in the range [N, N+ΔN] - density threshold
- **Power spectrum** - characterises the injection of energy depending on the wavenumber
- **Tracing filaments** - morphological component separation and analysis
- **Δ-variance** - provides a measure for the amount of structure on various scales in a given image
- **Structure noise** - amplitude of variations around a given point
- **Multi-fractal spectra analysis** - multi-resolution wavelet decomposition of the fields



ρ Oph, Herschel/PACS @ 160 μm

MORPHOLOGICAL TECHNIQUES – ARE THEY COMPARABLE?

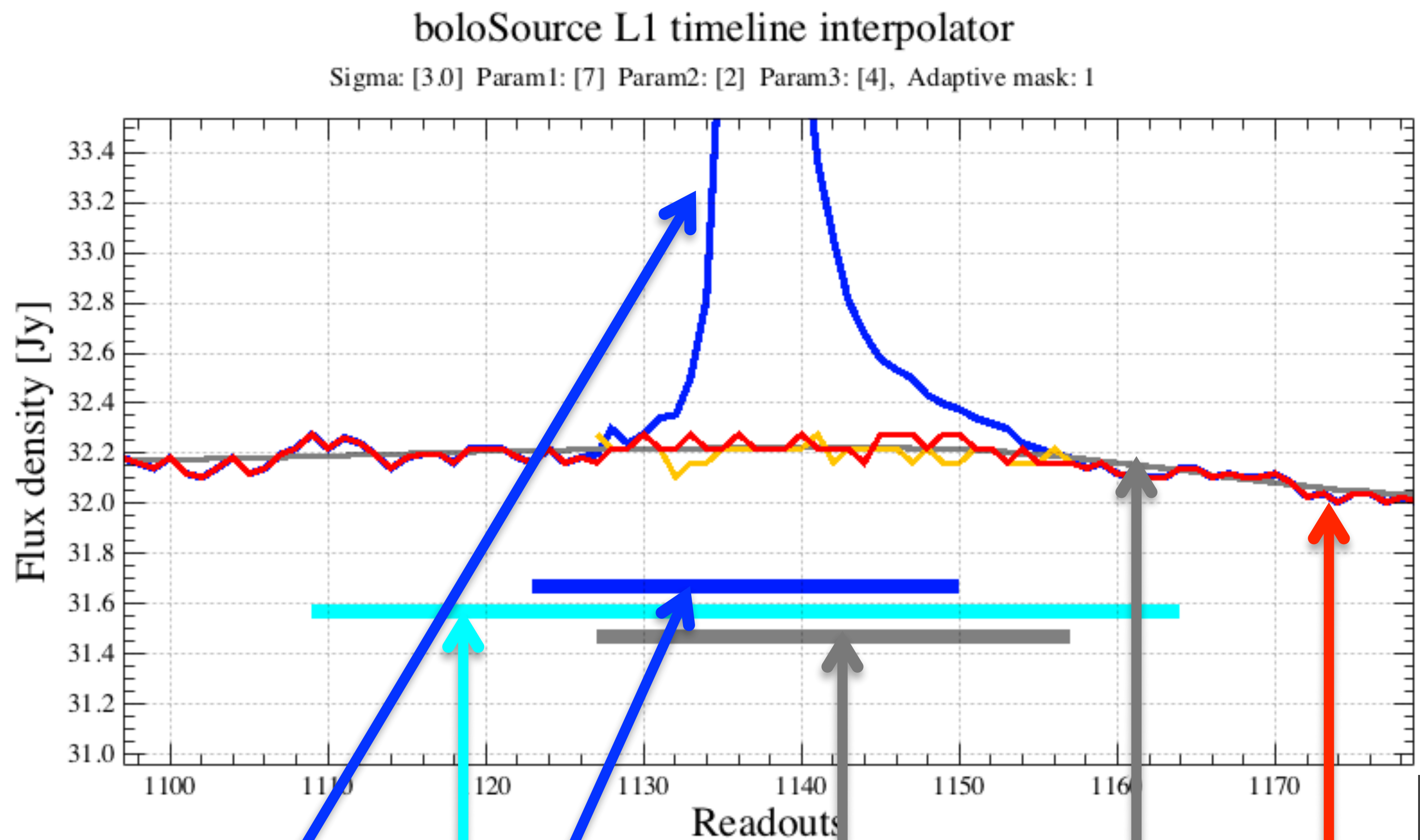
- Maps contain not only background emission, foreground and embedded sources are present
- Image analysis techniques have different sensitivity to discrete substructures
- Results at high spatial frequencies can be falsified by a few point sources
- Elia et al. 2014 - Δ-variance fitted slope is different without point sources



Elia et al. (2014)

SOURCE SUBTRACTION – HOW?

- 1D timeline reconstruction
- We are mainly interested in to subtract high-frequency components.
- In the masked part of the timeline one could interpolate with simulated noise + sky background



L1 bolometer signal in the timeline

Back projected input mask

Intermediate working mask

Interpolated signal with Monte - Carlo simulated noise spectrum

Estimated baseline

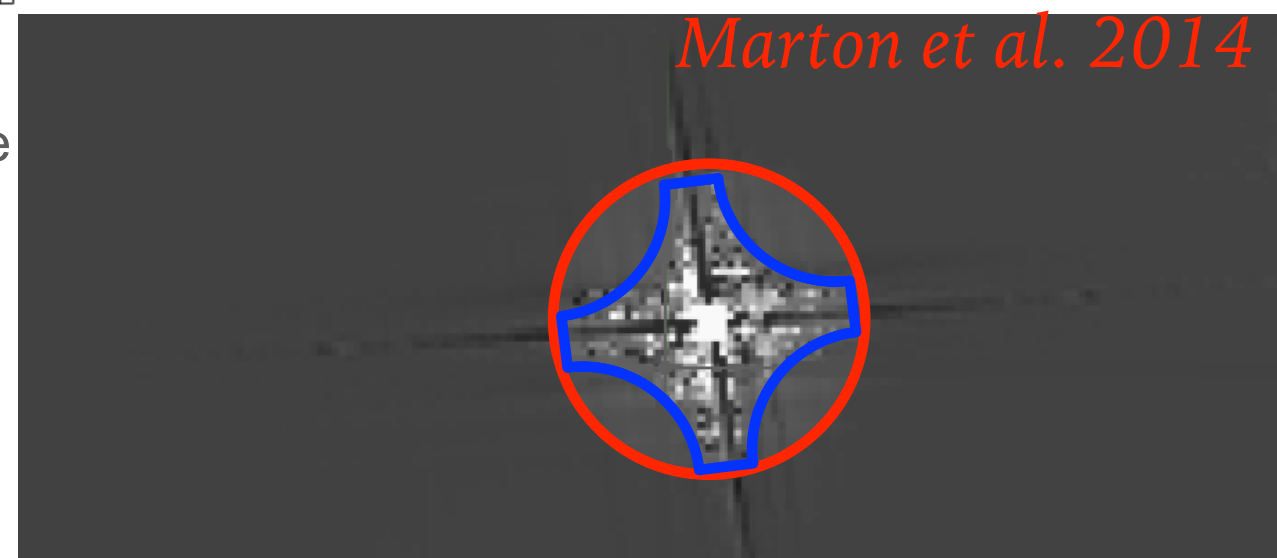
Optimized adaptive mask

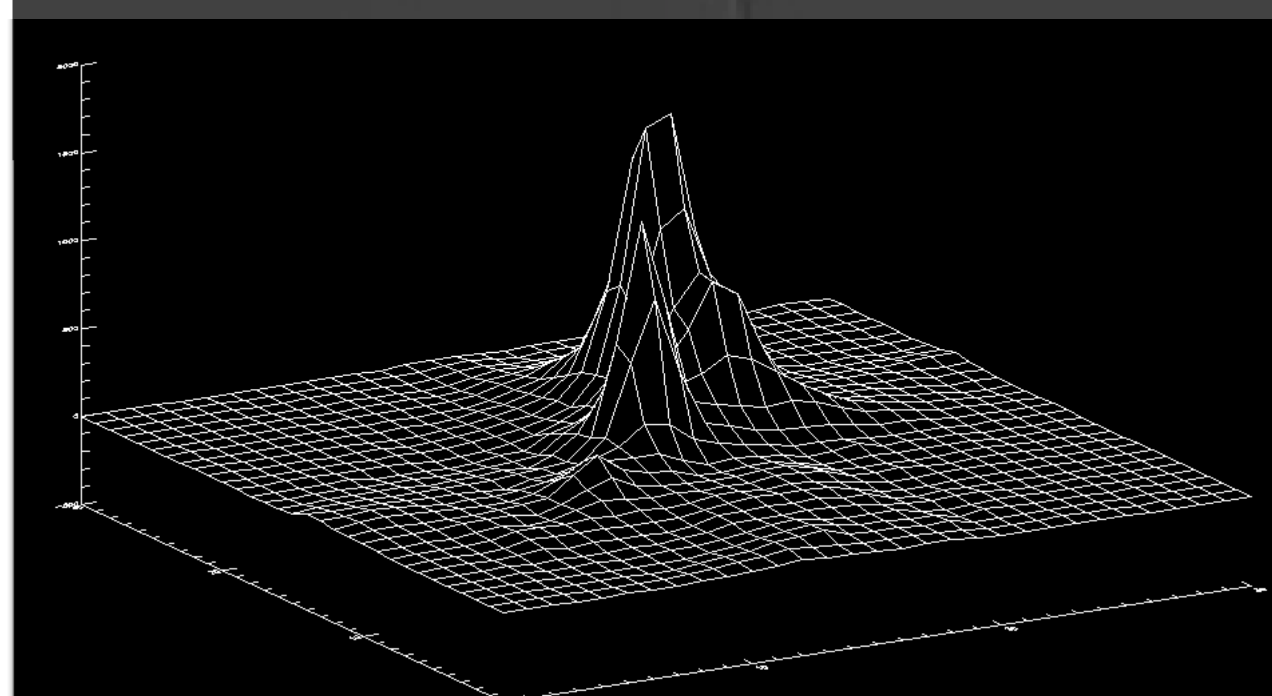
$$I_{(t)} = N^{1/f}_{(t)} + N^{D}_{(t)} + N^{det}_{(t)} + I^{S(lowfreq)}_{(t)}$$

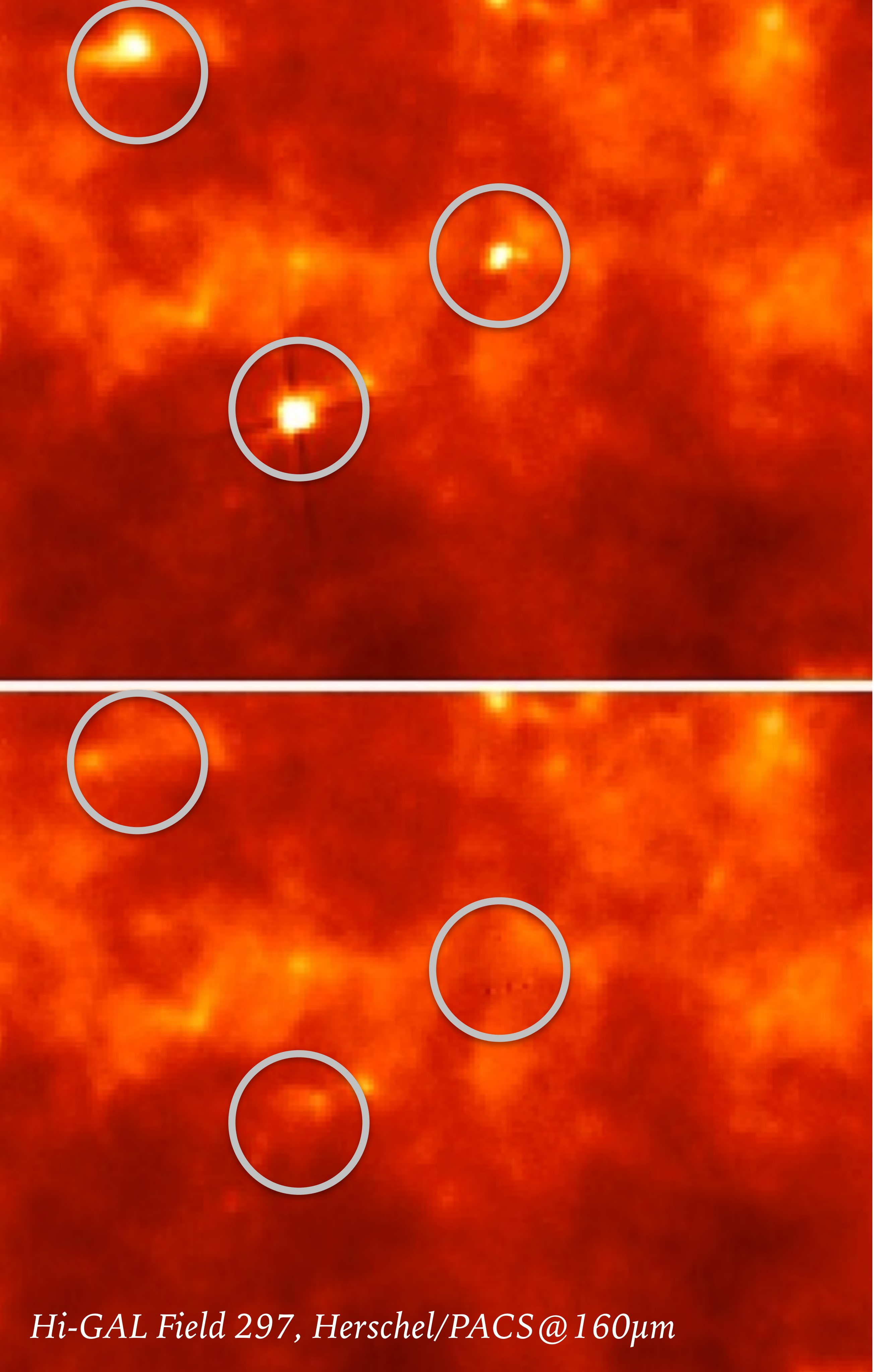
Simulated noise

Interpolated intensity in masked timeline

Baseline estimate from data







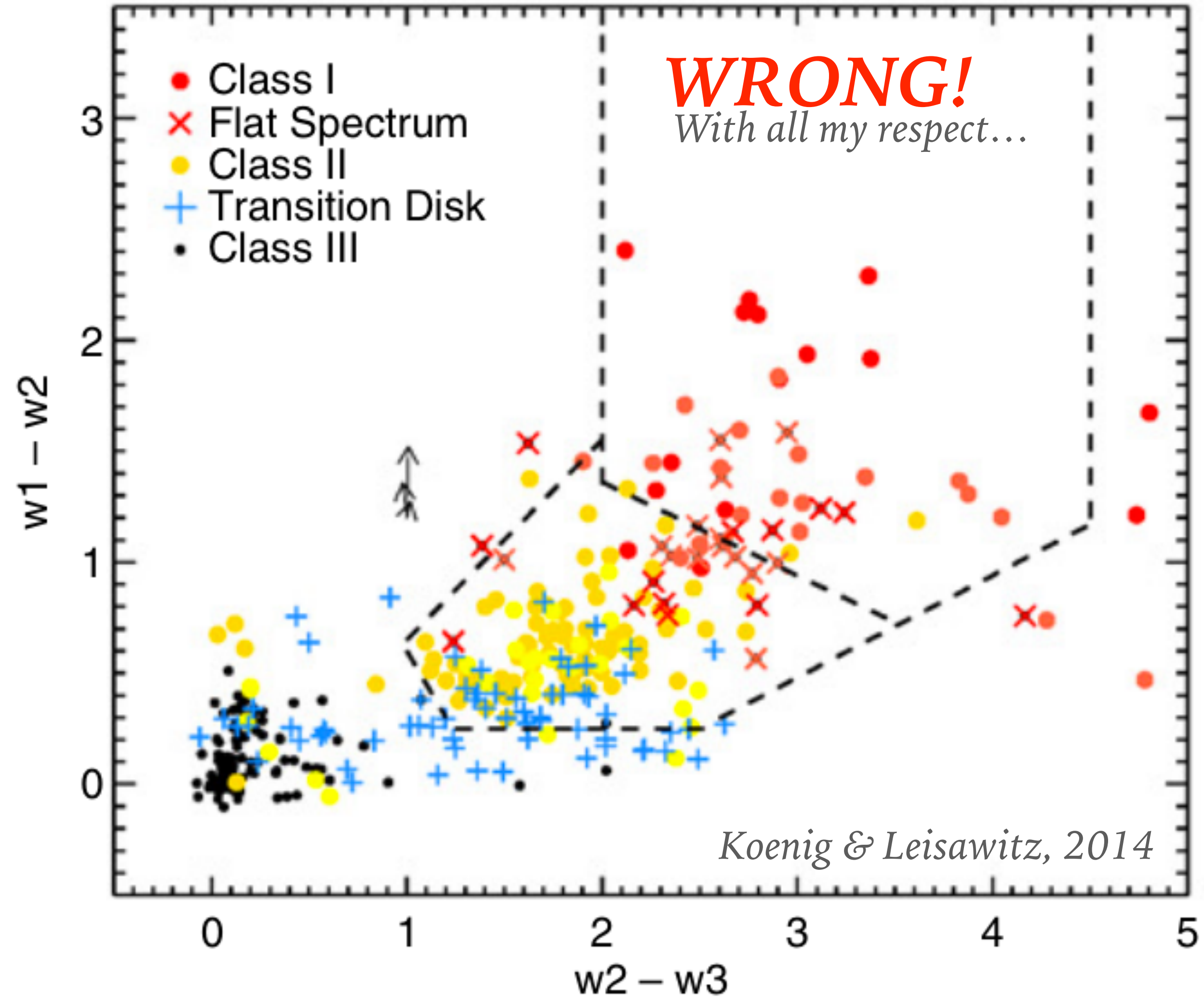
Hi-GAL Field 297, Herschel/PACS@160μm

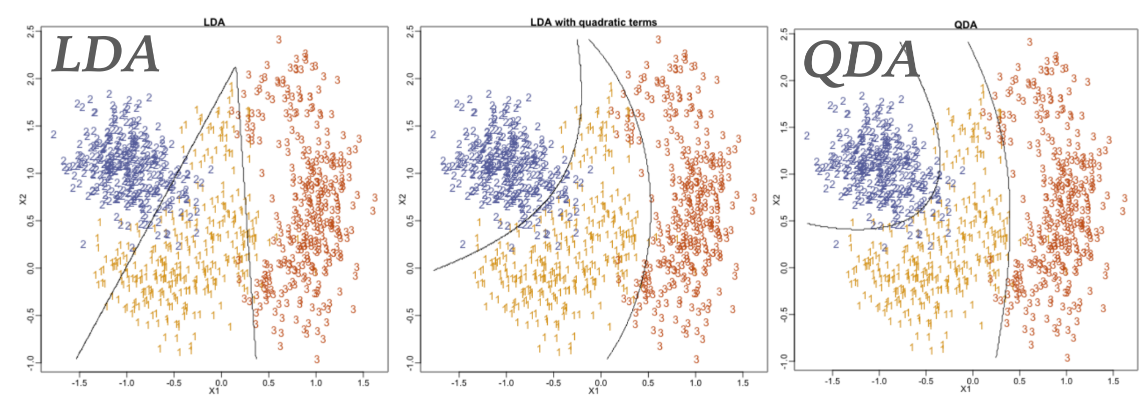
FUTURE PROJECT

- Remove identified point sources with the boloSource() task
- Use homogeneously reconstructed large-scale Herschel photometric observations of SFRs
- Analyse and compare structure analysis techniques, identify methods that can be used to parametrise the overall ISM morphology
- Provide complexity maps
- Identify the key complexity parameters that explain the link between cloud morphology and star formation rate
- Comparison with numerical simulations

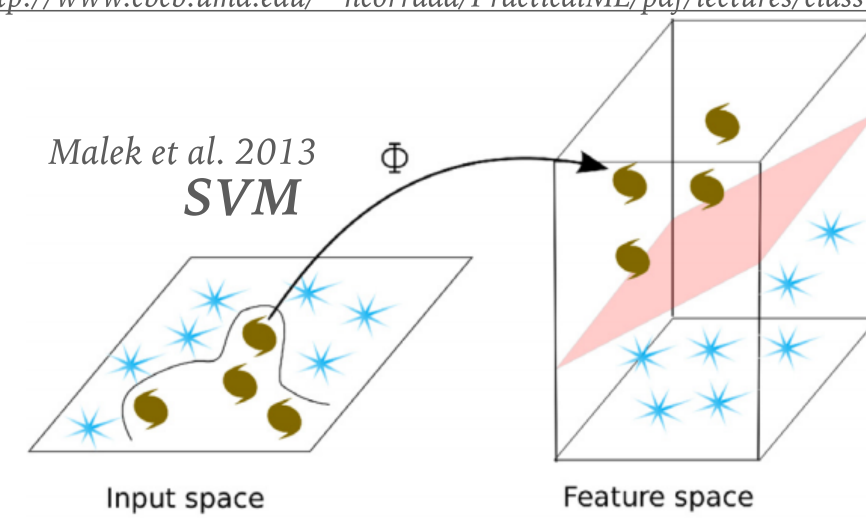
YSO SELECTIONS FOR STAR FORMATION STUDIES

- The star formation rate can be determined with a good selection of YSOs - YSO lists are subject to 30-40% uncertainty (A. Stutz, PPVI)
- QDA selection of AKARI YSO candidates (Tóth et al., 2014)
- Recent catalogues of pre- and proto-stellar objects
  - ATLASGAL catalogue of dust condensations - Csengeri et al. 2014
  - Galactic cold cores. IV - Montillaud et al. 2015
  - Planck Catalogue of Galactic Cold Clumps
  - Key Program Catalogues - Hi-GAL, HGBS, 2015-16
  - **Herschel Point Source Catalogue** - Marton et al. in prep., Schulz et al. in prep.
- **SVM selection of AllWISE YSO candidates** (Marton et al., MNRAS)





http://www.cbcb.umd.edu/~hcorrada/PracticalML/pdf/lectures/classification.pdf



Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **458**, 3479–3488 (2016) doi:10.1093/mnras/stw398
Advance Access publication 2016 February 22

An all-sky support vector machine selection of *WISE* YSO candidates

G. Marton,[1★] L. V. Tóth,[2] R. Paladini,[3] M. Kun,[1] S. Zahorecz,[2,4] P. McGehee[3] and Cs. Kiss[1]

[1] *Konkoly Observatory, Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Sciences, H-1121 Budapest, Hungary*
[2] *Department of Astronomy, Loránd Eötvös University, Pázmány P.s. 1/a, H-1117 Budapest, Hungary*
[3] *Infrared Processing Analysis Center, California Institute of Technology, 770 South Wilson Ave., Pasadena, CA 91125, USA*
[4] *European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany*

Accepted 2016 February 18. Received 2016 February 9; in original form 2015 July 29

ABSTRACT

We explored the AllWISE catalogue of the *Wide-field Infrared Survey Explorer (WISE)* mission and identified Young Stellar Object (YSO) candidates. Reliable 2MASS and *WISE* photometric data combined with *Planck* dust opacity values were used to build our data set and to find the best classification scheme. A sophisticated statistical method, the support vector machine (SVM) is used to analyse the multidimensional data space and to remove source types identified as contaminants (extragalactic sources, main-sequence stars, evolved stars and sources related to the interstellar medium). Objects listed in the SIMBAD data base are used to identify the already known sources and to train our method. A new all-sky selection of 133 980 Class I/II YSO candidates is presented. The estimated contamination was found to be well below 1 per cent based on comparison with our SIMBAD training set. We also compare our results to that of existing methods and catalogues. The SVM selection process successfully identified >90 per cent of the Class I/II YSOs based on comparison with photometric and spectroscopic YSO catalogues. Our conclusion is that by using the SVM, our classification is able to identify more known YSOs of the training sample than other methods based on colour–colour and magnitude–colour selection. The distribution of the YSO candidates well correlates with that of the *Planck* Galactic Cold Clumps in the Taurus–Auriga–Perseus–California region.

Key words: methods: data analysis – methods: statistical – stars: pre-main-sequence – stars:

INTRODUCTION - MOTIVATION

- IRAS: **~350 000** objects at 12 μm above **0.5 Jy**
- WISE: **7x10^8** objects above **1 mJy**
- Various object types - **overlapping** colours
- Many times **polygons** are used to separate object types - questionable accuracy

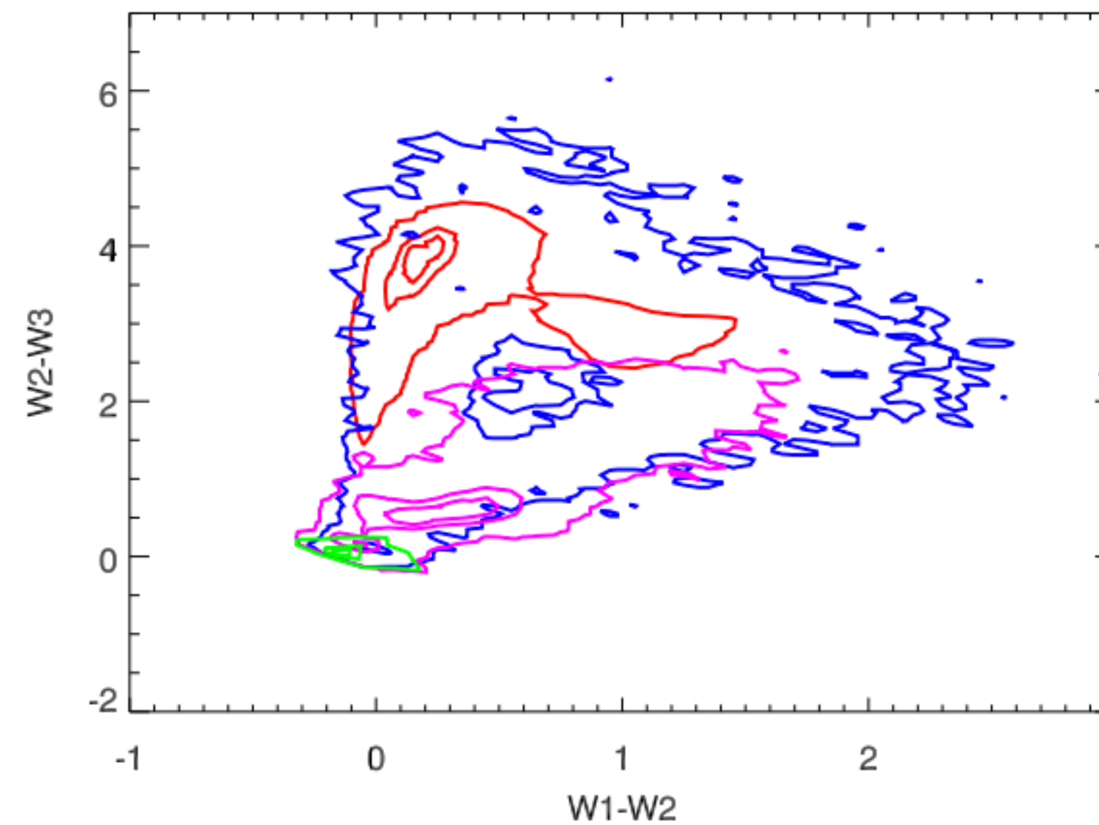
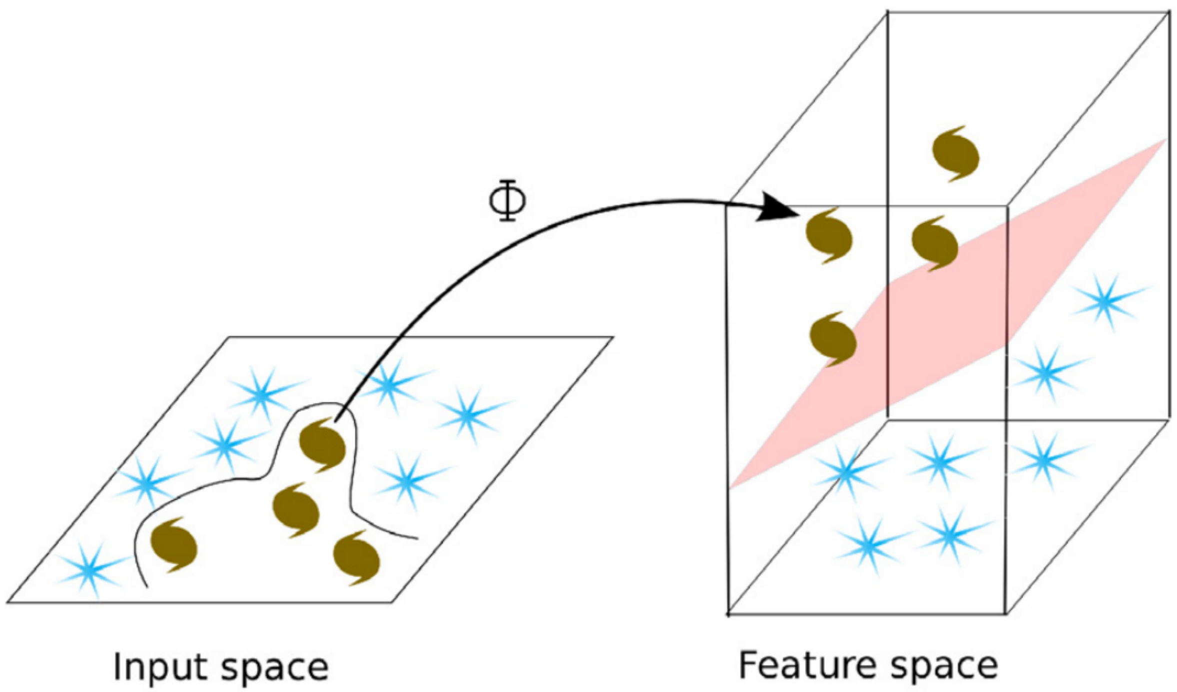


Figure 1. Object types identified with SIMBAD on the *W*1–*W*2, *W*2–*W*3 plane. The figure demonstrates that different object types are overlapping, boundaries between them are non-linear. Contour lines show the 5 per cent, 50 per cent and 75 per cent of the maximal surface density of extragalactic sources (red), field stars (green), evolved stars (magenta) and YSOs (blue). The surface density of different types was calculated in bins of 0.1 mag.

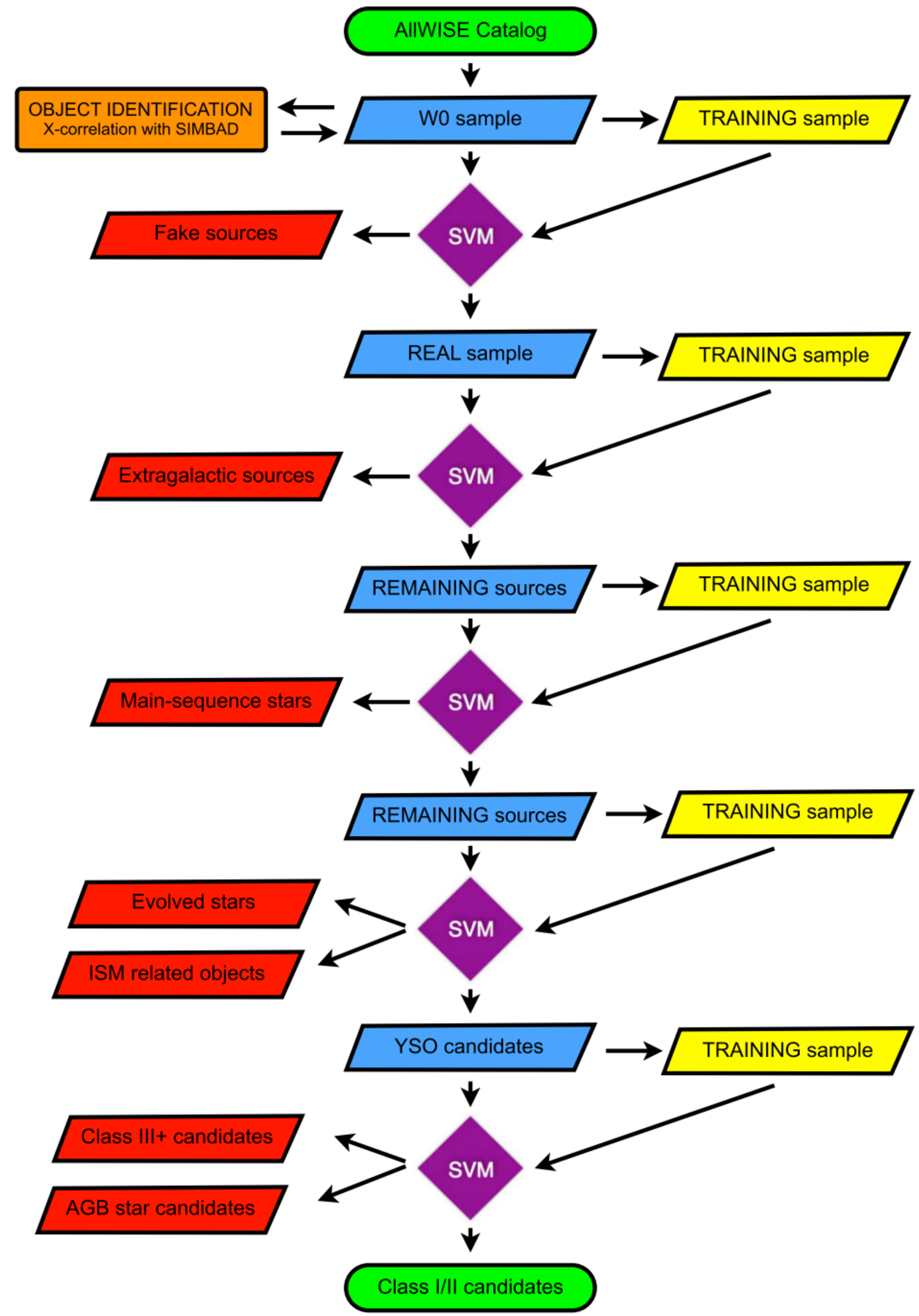
SUPPORT VECTOR MACHINE



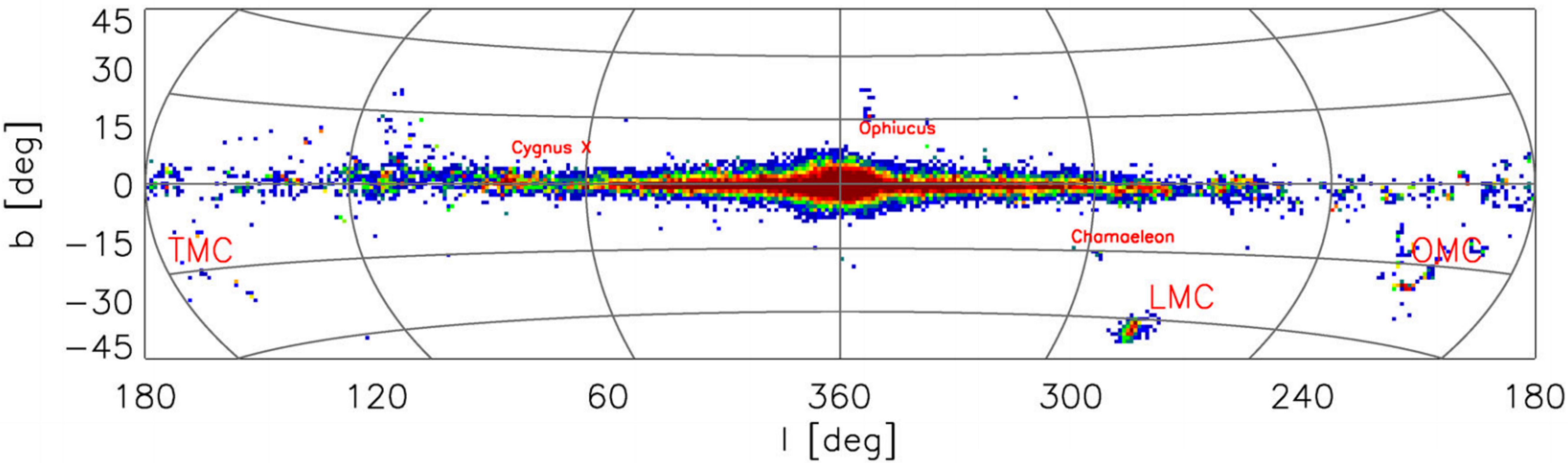
- **Supervised** machine learning algorithm
- Used for **classification** and regression analysis
- Input data are transformed into **higher dimensional** space (Malek et al, 2013)
- “MASS” and “e1071 ”library of **R** was used - www.r-project.org
- Because supervised: **training sample** is needed

CLASSIFICATION PROCESS

- Training samples based on **SIMBAD**
- Starting sample: S/N>3 in all WISE bands
- Spurious source identification - catalog positions where sources are not seen
- Multiple step classification
- Removal of extragalactic sources, MS stars, evolved stars, ISM related objects



RESULTS – WISE YSO CANDIDATES



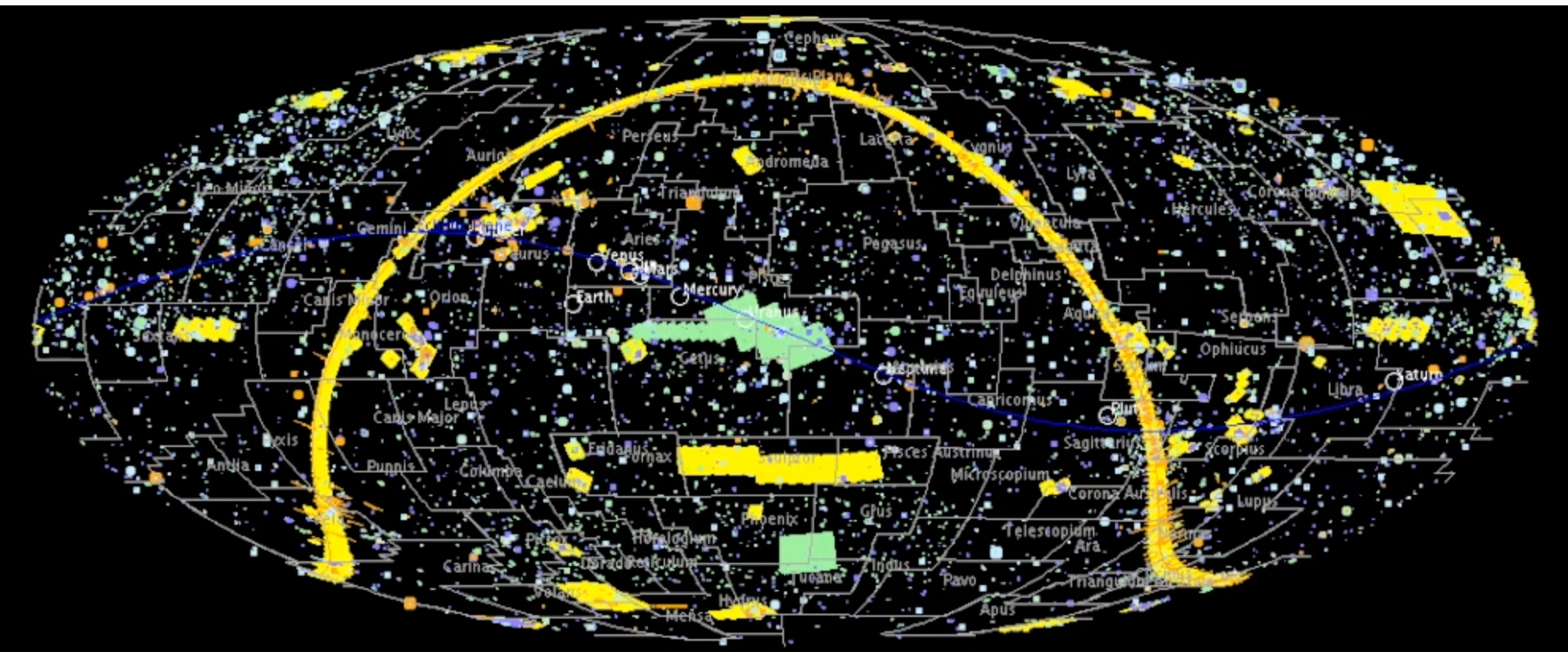
- 133 980 Class I/II YSO candidates
- 608 606 Class III or more evolved
- classification via machine learning (SVM)
- high reliability

Table 4. Number of sources of our binned SIMBAD subtypes (first column) in the initial *W0* sample (second column), in our SVM classified Class I/II candidate sample (third column) and in the YSO sample of the KL14 method (last column). KL14 successfully removes the Galactic contamination, but is less successful in identifying the known YSOs and in removing the extragalactic contamination.

Subtype	W0	SVM Class I/II	KL14 YSOs
G1	148 267	65	1398
G2	10 311	7	1188
G3	12 729	29	3303
E1	13 208	631	58
E2	3429	1118	613
E3	109	43	16
ISM1	912	183	90
ISM2	515	173	50
S	973 629	11 990	1043
Y1	9268	3705	2695
Y2	1128	637	650

HERSCHEL POINT SOURCE CATALOGUE

- Homogenous extraction from all PACS & SPIRE photometric observations
- ~10% coverage of all-sky at 6 FIR and sub-mm wavelengths (70, 100, 160, 250, 350 & 500 µm)
- To be released later this year
- So far ~8.2 million SPIRE and ~8.7 million PACS **detections** (not objects)



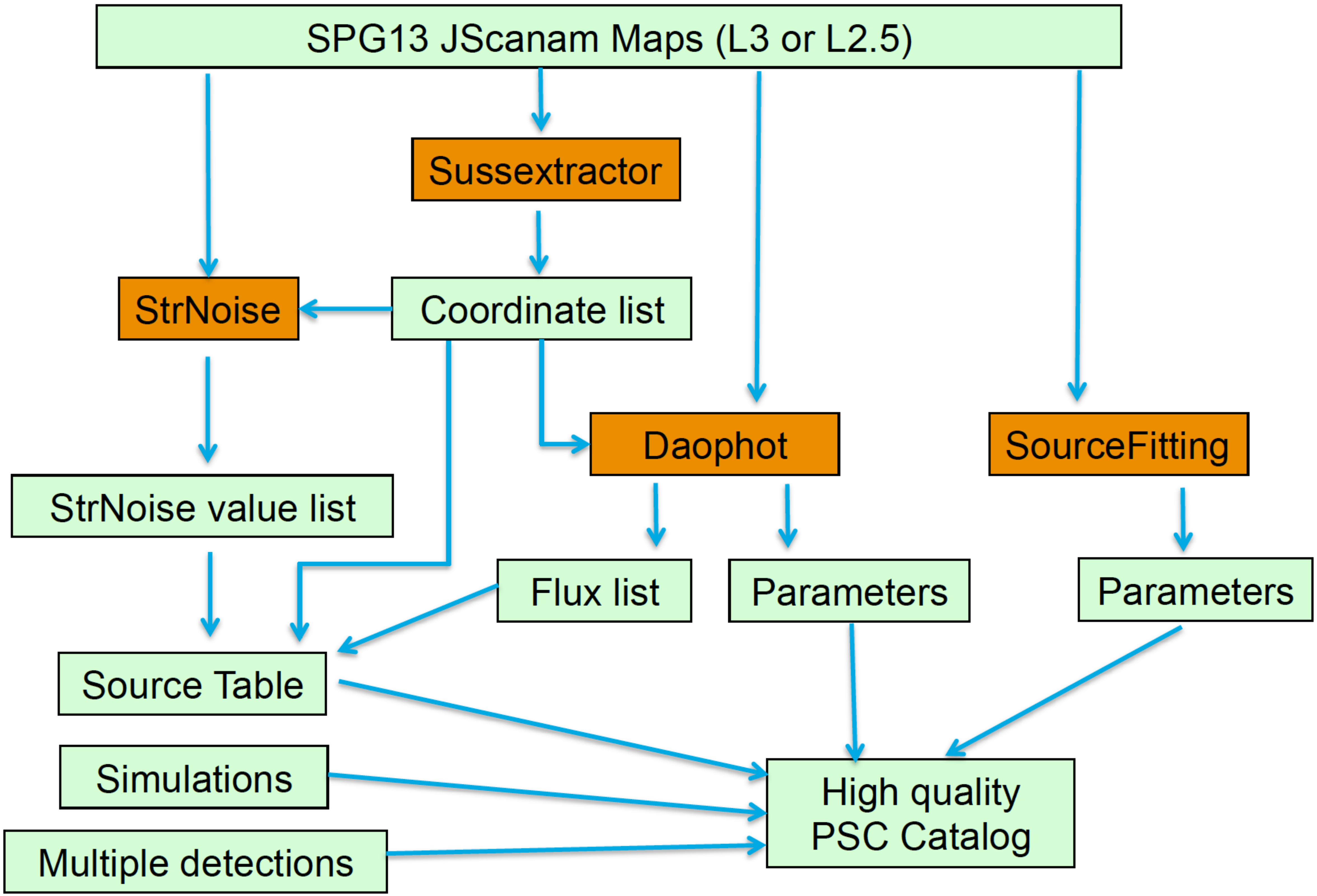
HPSC TEAM MEMBERS

Institutes	PACS	Common	SPIRE
Konkoly		Erika Varga-Verebélyi Gábor Marton Csaba Kiss	
ESA	Luca Calzoletti Bruno Altieri	Ana M. Perez Garcia Miguel Sánchez Portal	Ivan Valtchanov (Tanya Lim)
NHSC	Roberta Paladini	David Shupe John Rector	Bernhard Schulz (Kevin Xu, Nanyao Lu)
ELTE			Viktor Tóth Sándor Pintér
ICCs	Zoltán Balog		Chris Pearson Andreas Papageorgiou

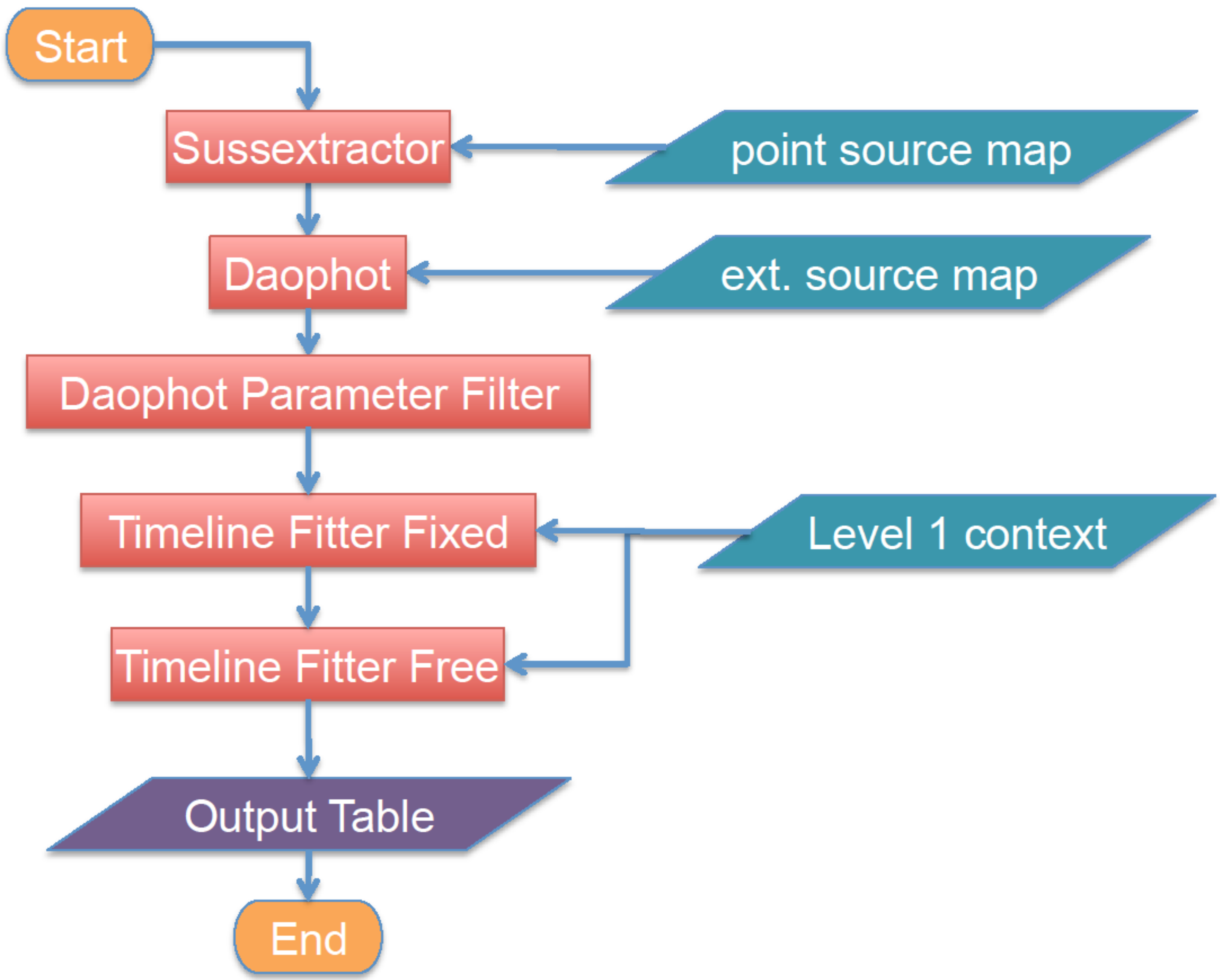
THE HPSC WILL PROVIDE:

- **Millions** of sources that have never been observed before at far-IR and sub-mm wavelengths
- Maps that have **never** been analysed for their full source content
- **Homogeneous** source extraction
- Information on the far-infrared characteristics of individual objects
- Large scale or statistical studies (like star formation rate and clustering studies) will also benefit from the catalogue
- Improvements of the pipeline processing adds knowledge to user support and ICC work

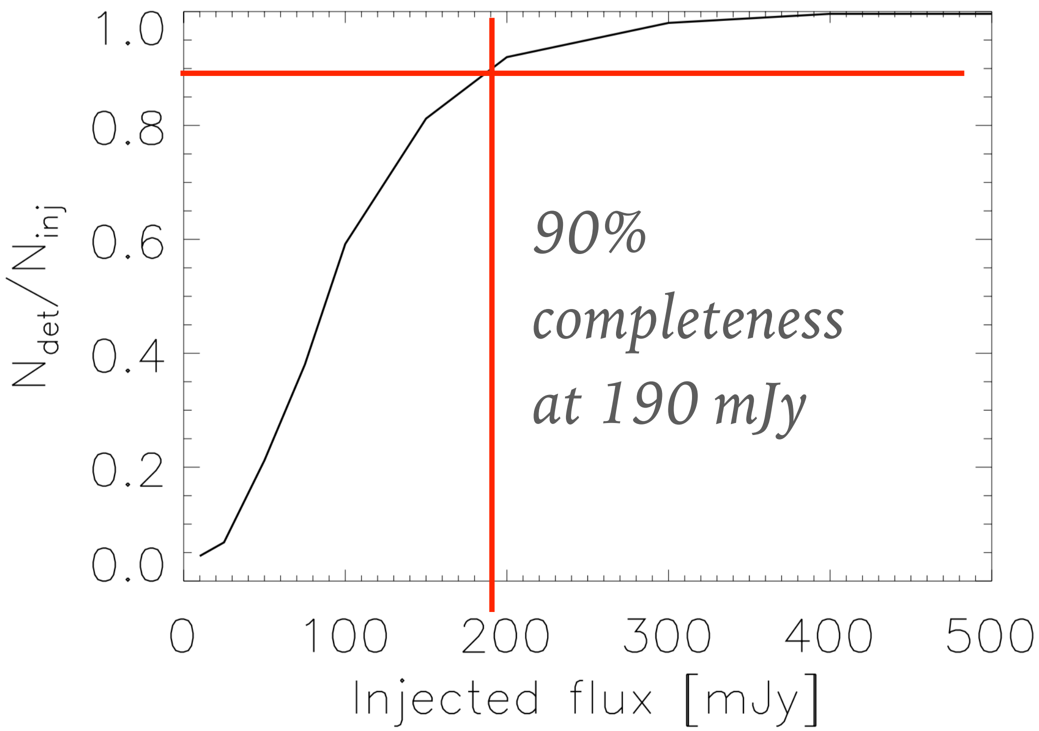
PACS PSC PIPELINE

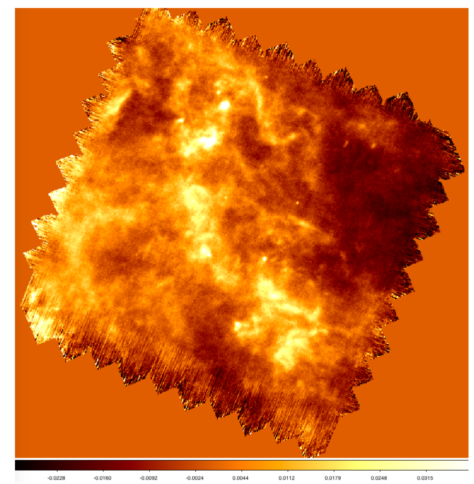


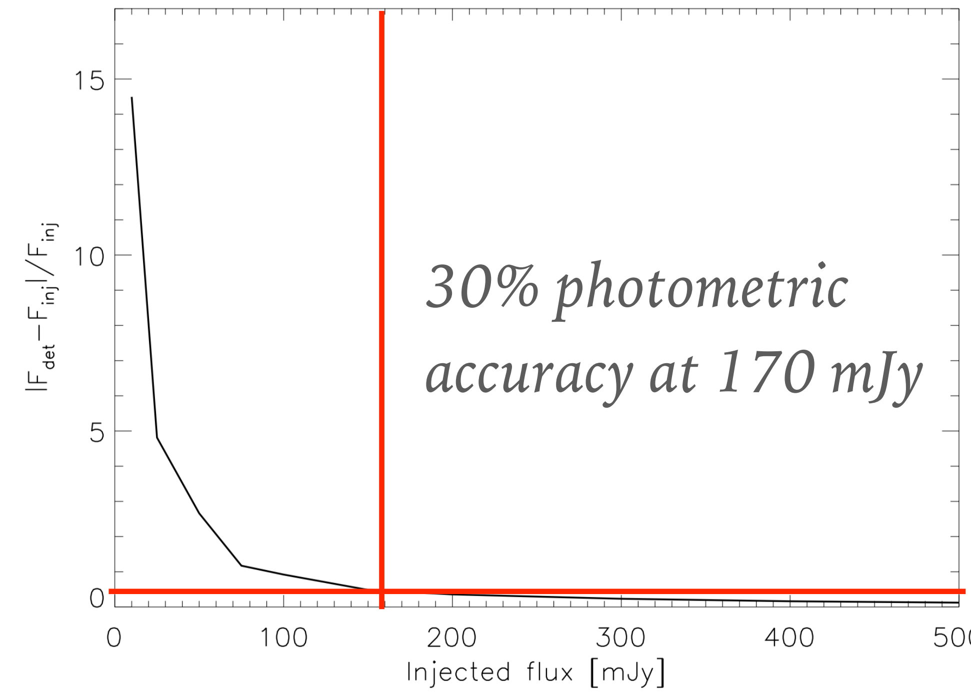
SPIRE PSC PIPELINE

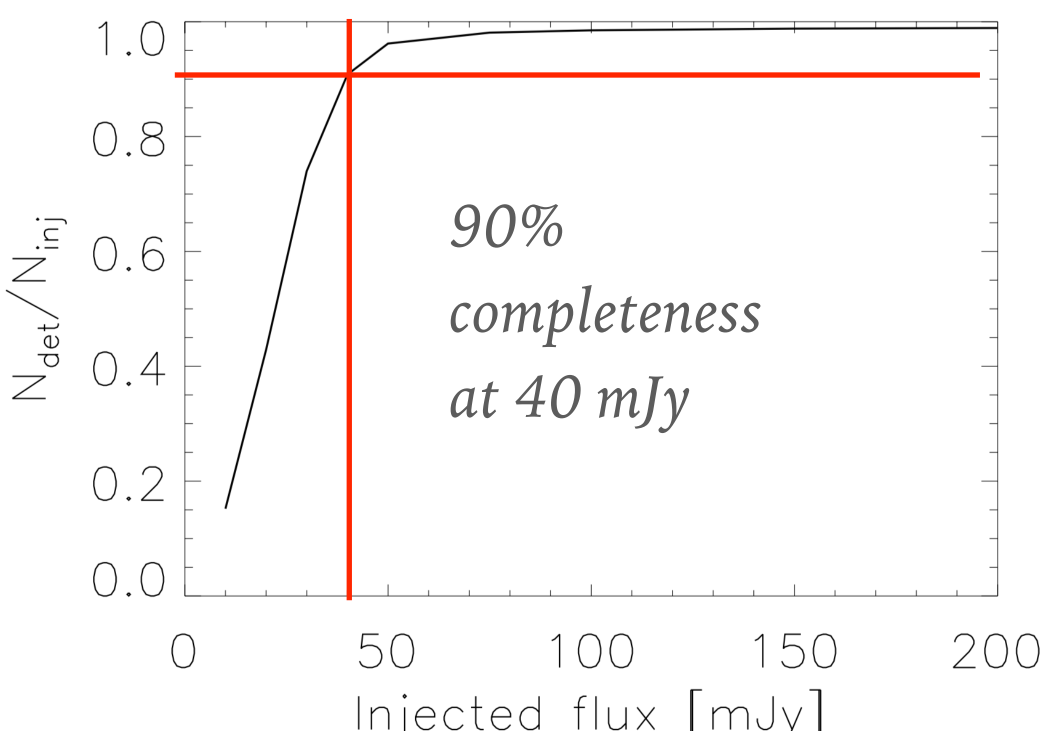


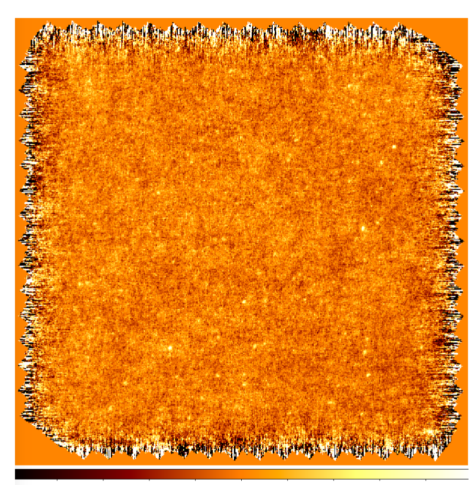
VARIETY OF ENVIRONMENTS – HOMOGENOUS EXTRACTION

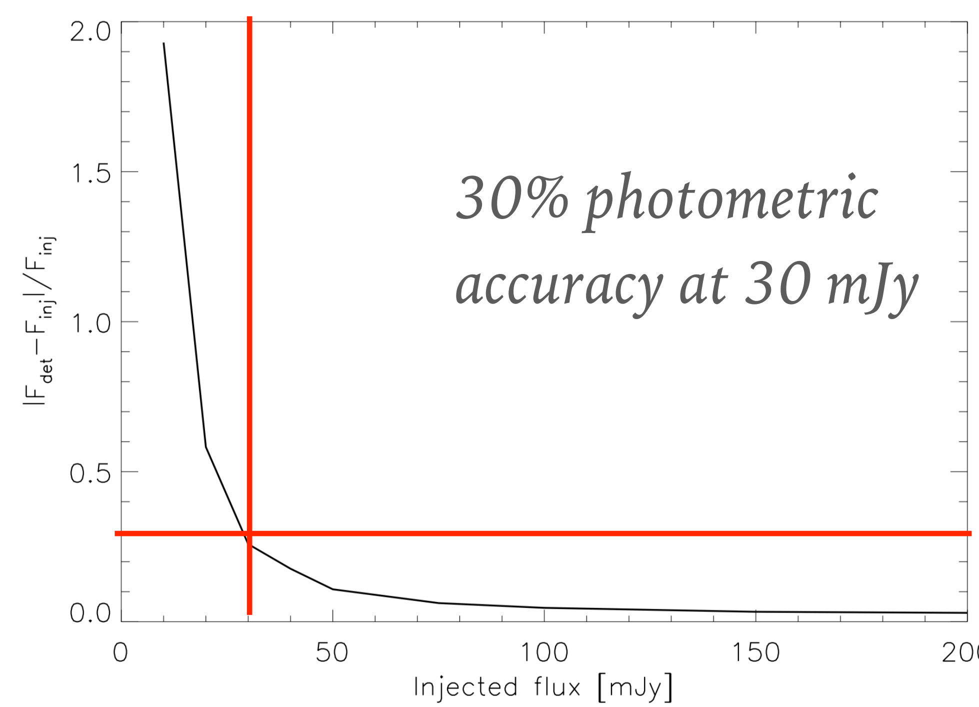






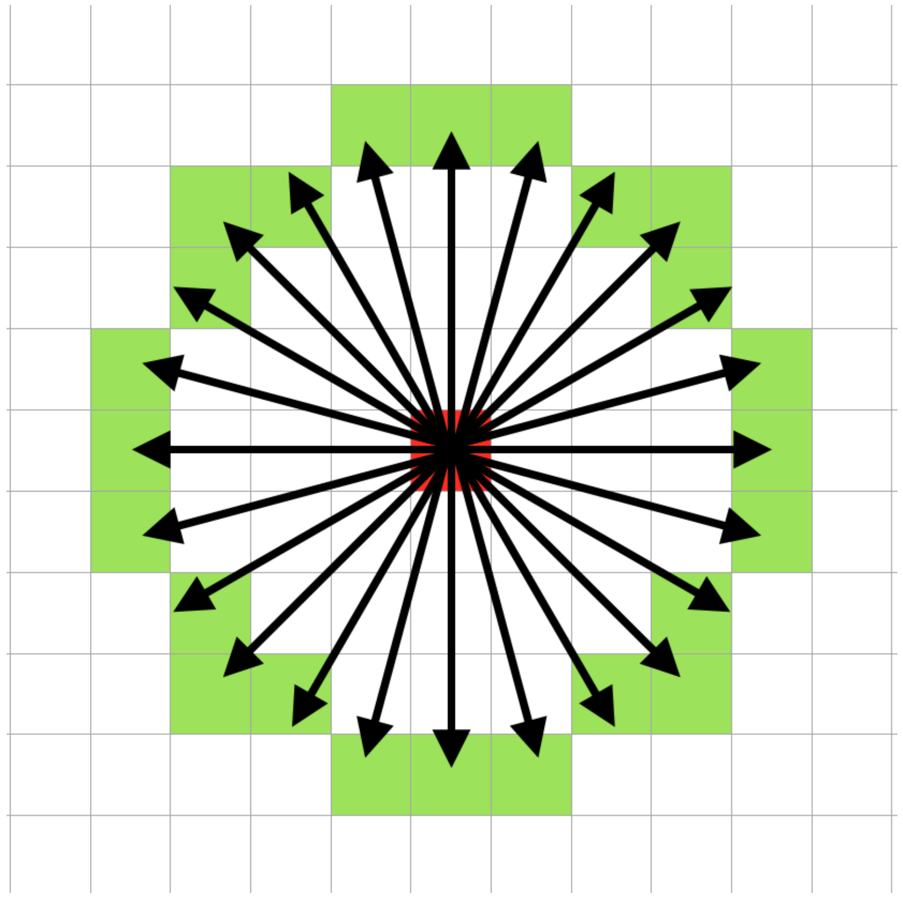






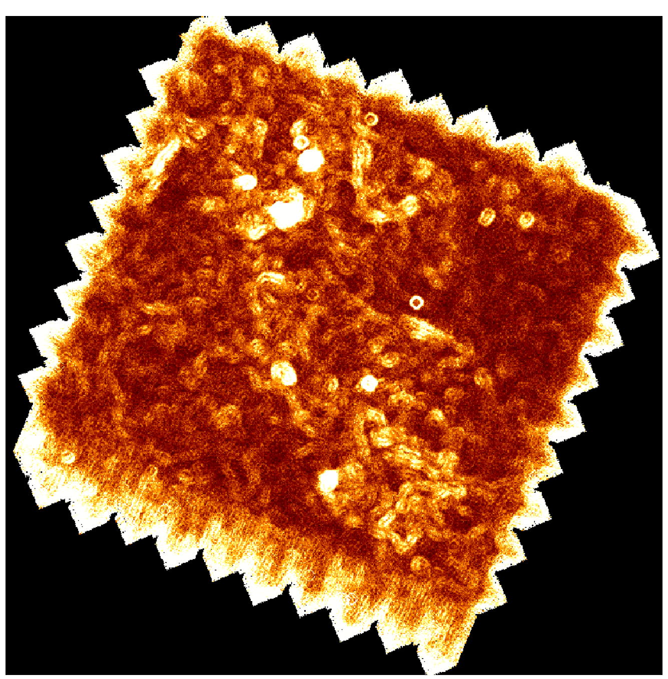
Completeness in two of our PACS 160 μm test cases, G334.65+2.67 (top) and Lockman Hole (bottom).

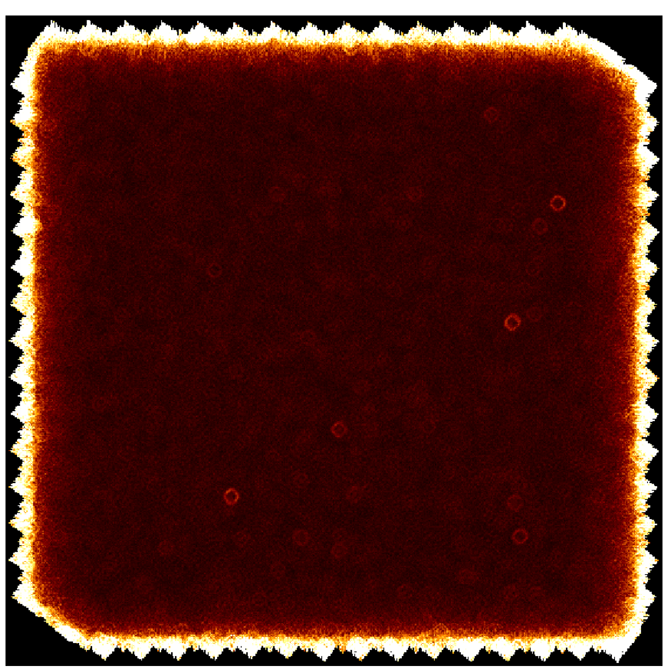
Photometric accuracy in two of our PACS test cases, G334.65+2.67 (top) and Lockman Hole (bottom).



HOMOGENOUS EXTRACTION- STRUCTURE NOISE

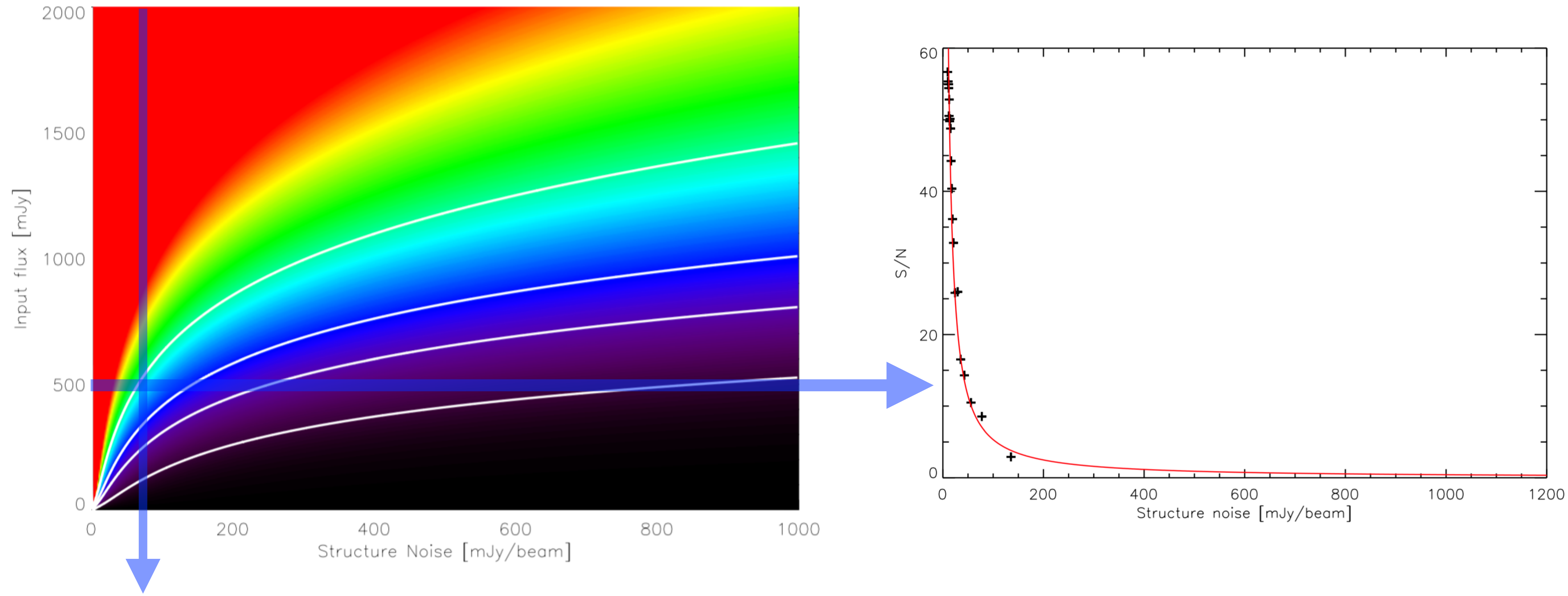
- Structure noise measures the fluctuation around a given point in the sky
- It can be translated into the fluctuation power of the neighbouring areas
- It gives a local information instead of a general (regional average) number
- We use this quantity to describe the close vicinity of each detected source

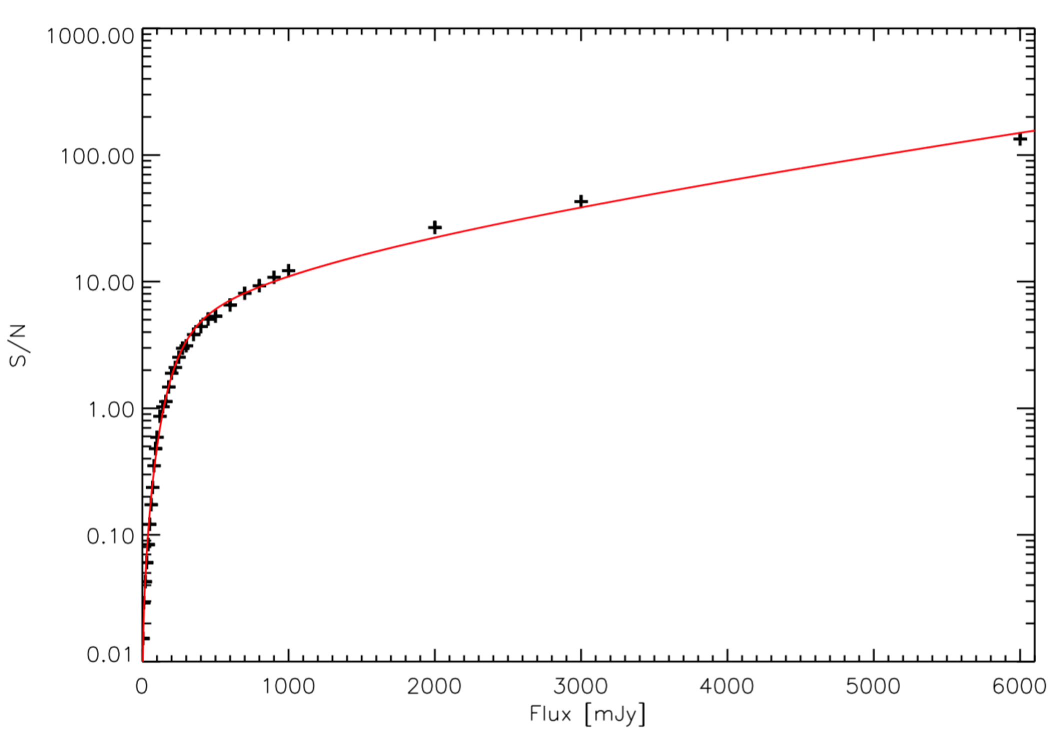




QUALITY CONTROL

S/N calibration surface PLW





- Source injection with well controlled flux
- Test fields - wide range of complexity and cover all main instrument setups
- "…for the present release, we find ourselves in the very difficult situation where it is not possible to define any combination of parameters that may offer a reliable "quality flag" for all detected sources." - Molinari et al. 2016, First Hi-GAL data release - **our plan is to offer a reliable quality flag through our S/N calibration surfaces**

SPSC STATUS

- First test of production-grade point source extraction procedure in August 2015.
- 8.2M source candidates extracted by Oct 2015 and ingested in relational Postgres database.
- A first consolidated skeleton catalog was produced end of 2015 containing only object groupings with (2.8, 2.1, 0.8)M objects/groups for 250, 350, 500 micron respectively.
- Data requires statistical analysis combined with object visualization.
- We aim for a filled-in first catalog in summer 2016.

PPSC STATUS

- First extraction from all maps was done - green maps had to be repeated, expected to finish this week.
- 8.7M source candidates extracted and ingested in relational Postgres database.
- First round of parameter analysis resulted in rejection of >50% of detections (spurious sources)
- Sources from SSO maps have to be extracted
- We aim for a first catalog by the end of summer 2016.