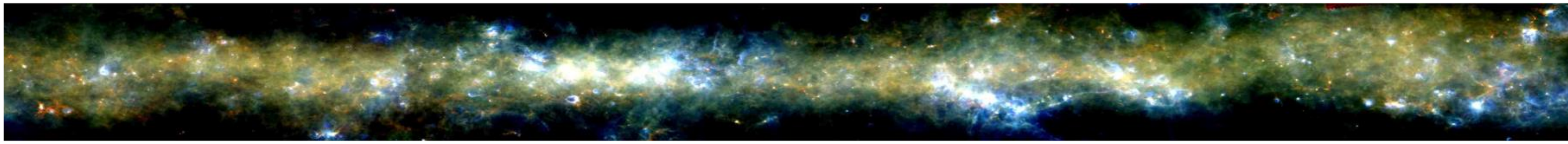
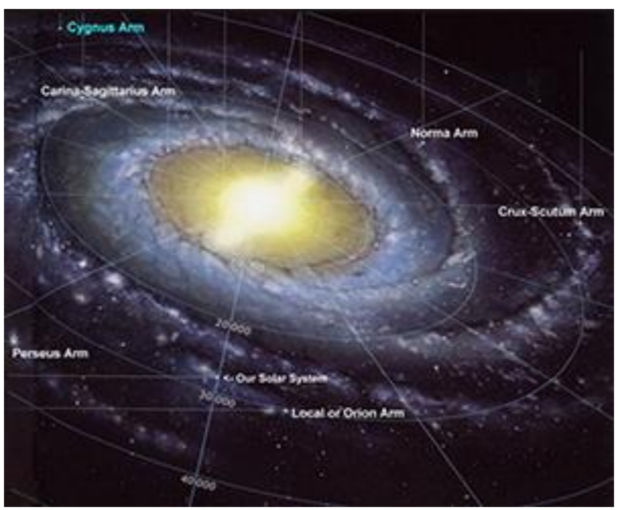


ViaLactea

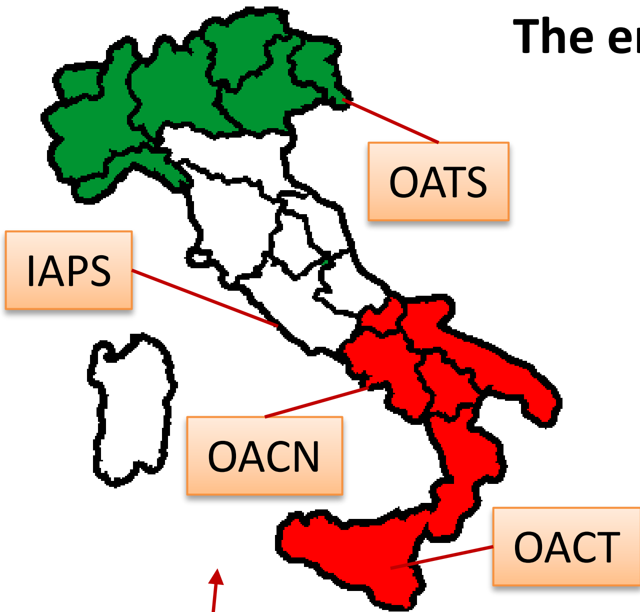
SEVENTH FRAMEWORK PROGRAMME

The Milky Way as a Star Formation Engine



ViaLactea @ INAF

The entire project is led by IAPS (*S. Molinari*)
Duration: 3 years (10/2013-10/2016)

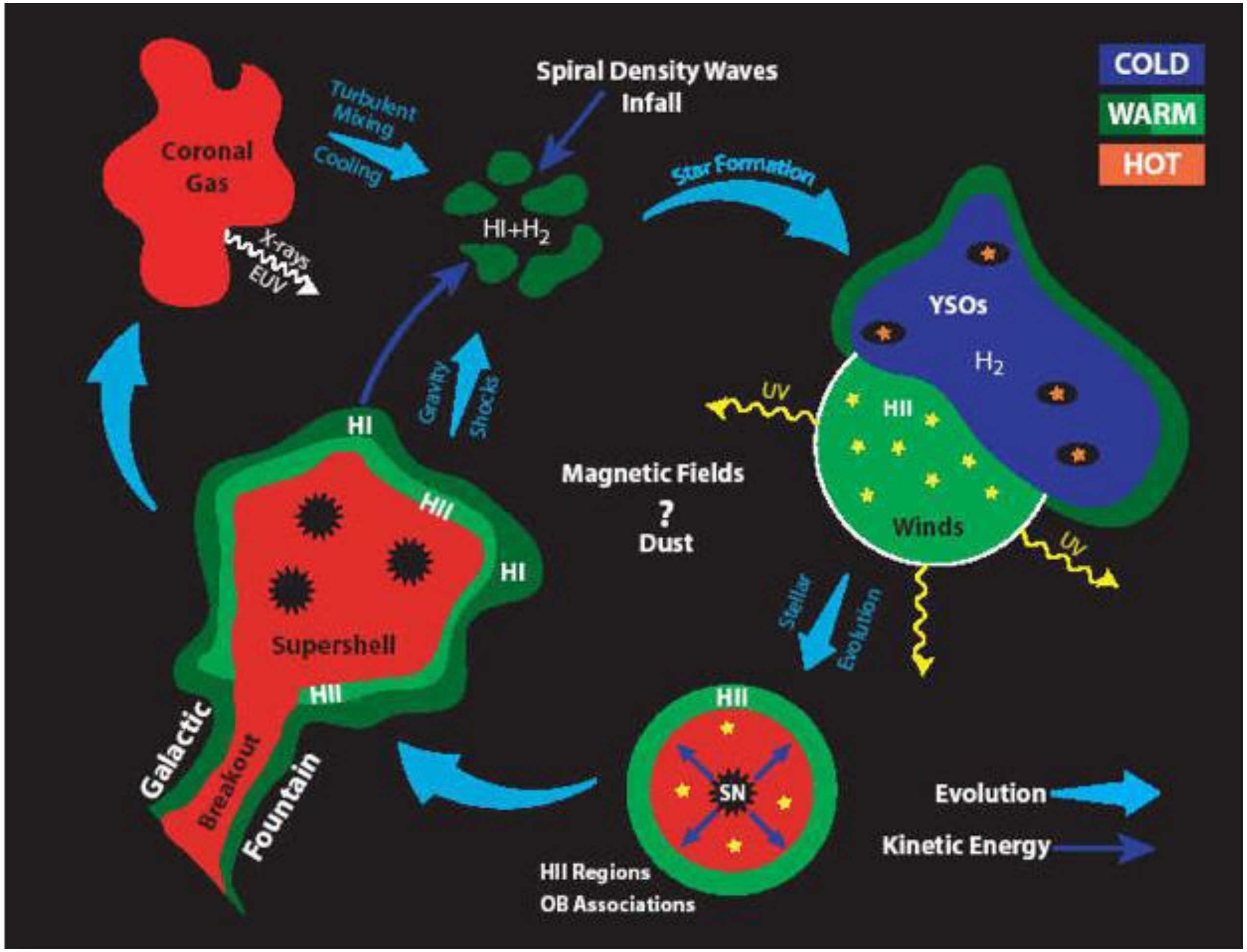


SEVENTH FRAMEWORK PROGRAMME

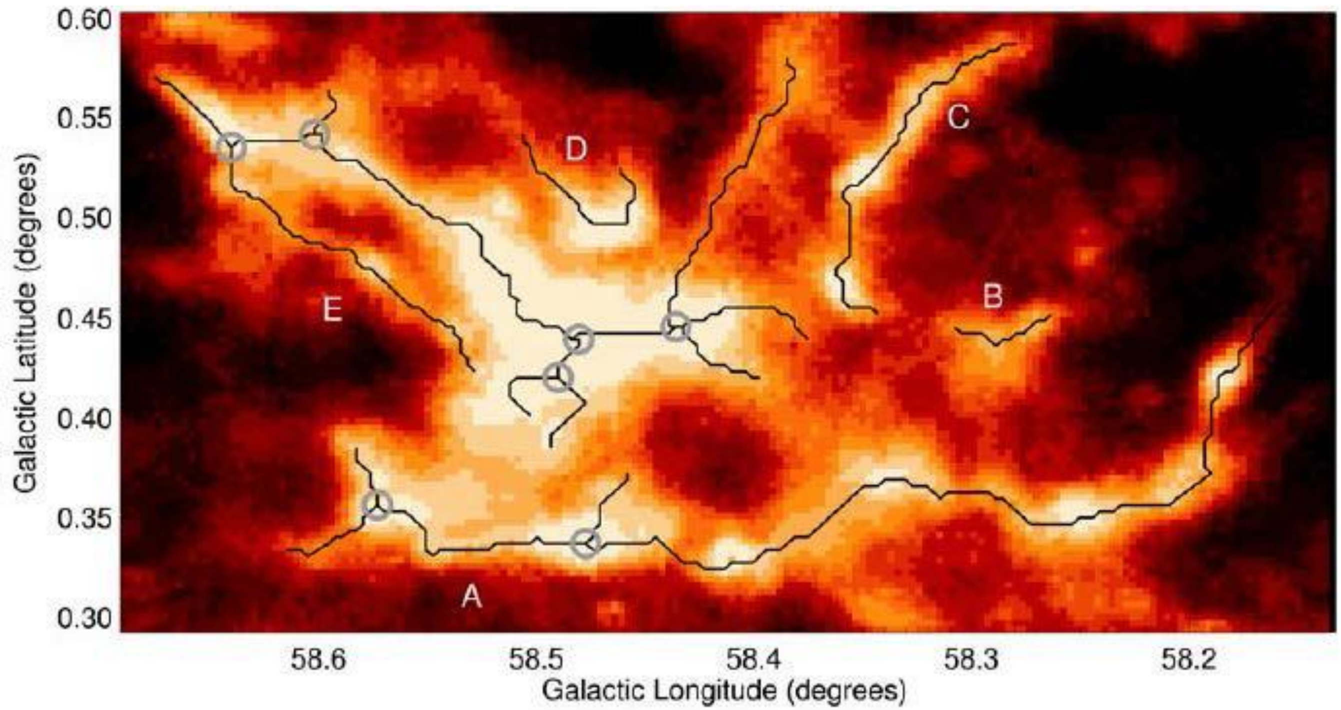
No	Name	Short name	Country
1	ISTITUTO NAZIONALE DI ASTROFISICA	INAF	Italy
2	UNIVERSITY OF LEEDS	UNIVLEEDS	United Kingdom
3	MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.	MPG	Germany
4	MAGYAR TUDOMANYOS AKADEMIA SZAMITASTECHNIKAI ES AUTOMATIZALASI KUTATO INTEZET	SZTAKI	Hungary
5	CARDIFF UNIVERSITY	CU	United Kingdom
6	UNIVERSITE D'AIX MARSEILLE	AMU	France
7	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	CNRS	France
8	NAGOYA UNIVERSITY	UON	Japan
9	THE UNIVERSITY OF EXETER	UNEXE	United Kingdom

ViaLactea Mission

The goal is to exploit the combination of all the new-generation Infrared→ Radio surveys of the Galactic Plane from space missions and ground-based facilities, using a novel data and science analysis paradigm based on 3D visual analytics and data mining framework, to build and deliver a quantitative 3D model of our VIALACTEA Galaxy as a star formation engine that will be used as a template for external galaxies and study star formation across the cosmic time



ViaLactea – Goals

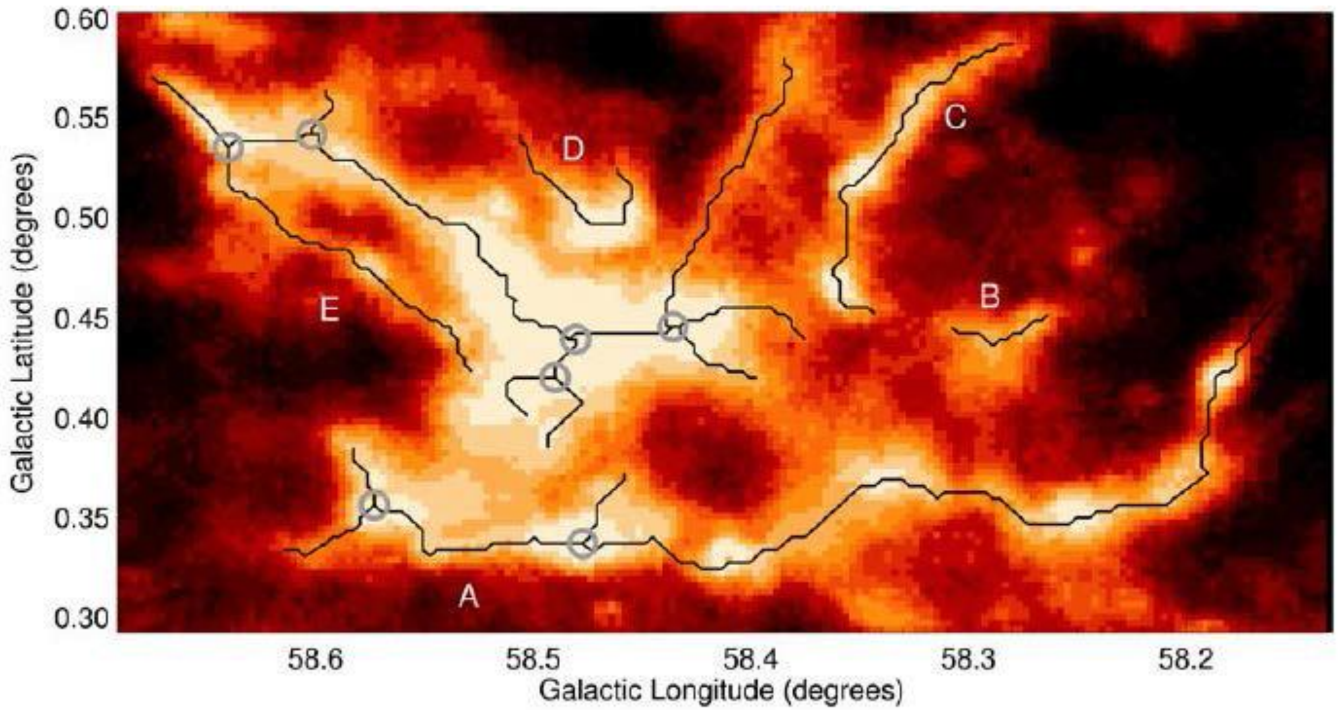


Identify the critical parameters that make star formation different:

- Spontaneous/triggered
- Filaments or no filaments ?
- Depending on the Galaxy position
- w.r.t. spiral arms
- etc.

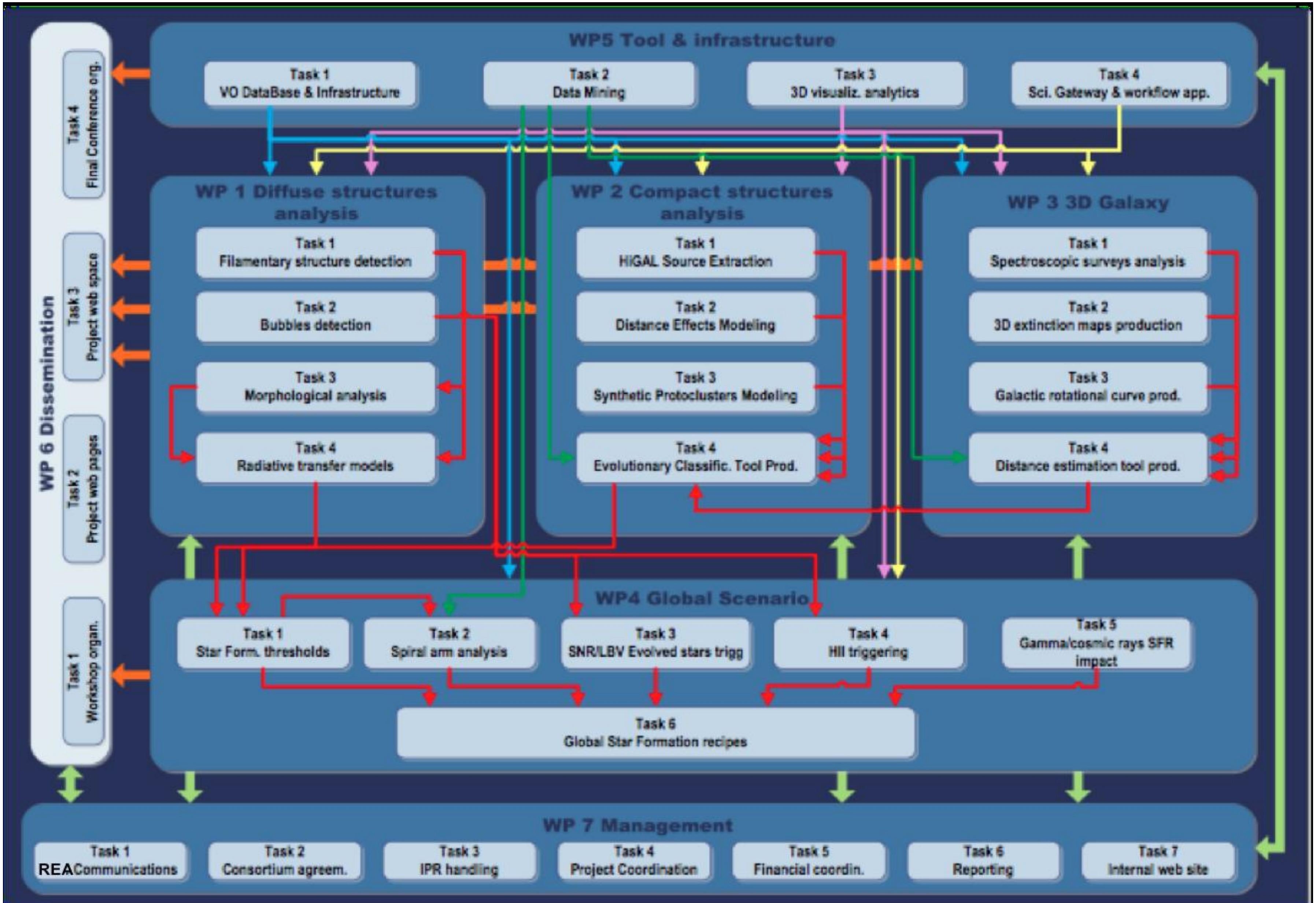
→ understand if & how the mix of the ingredients conspire to determine a global SF law

ViaLactea – Tasks

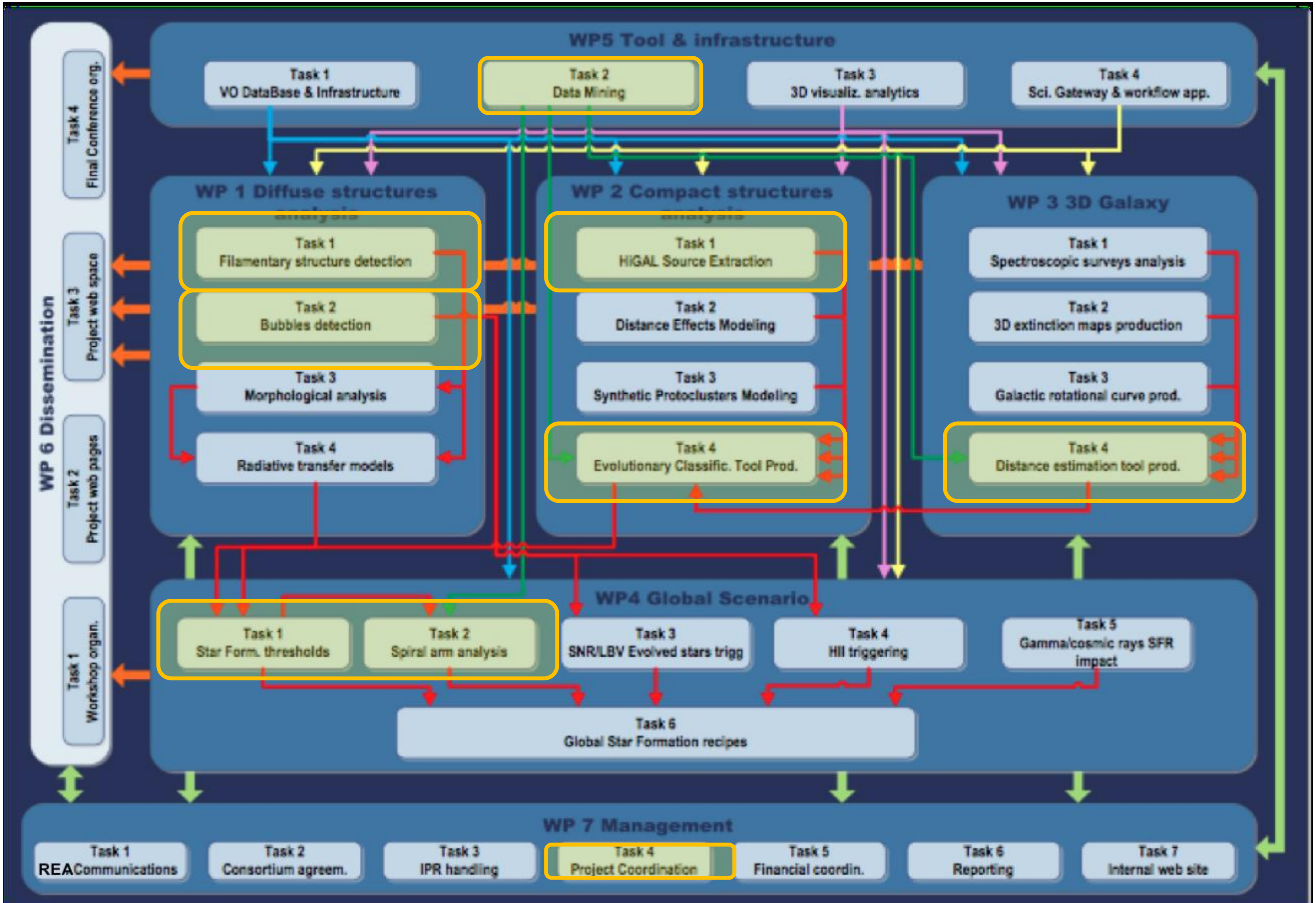


- ❑ Measure the star formation rate and history Galaxy-wide;
- ❑ Formation and fragmentation of Filamentary Molecular Clouds;
- ❑ Determining the relative importance of global vs local, spontaneous vs triggering, agents that give rise to star formation;
- ❑ Understanding star formation laws and the nature of thresholds as a function of ISM properties across a full range of galacto-centric radii metallicity and environmental conditions;
- ❑ Build bottom-up recipes and prescriptions useful for Xgal science.

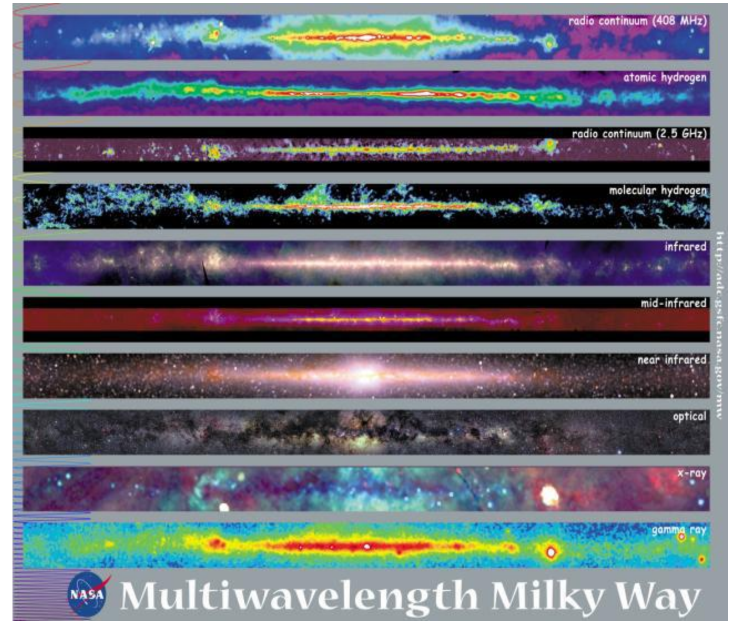
ViaLactea – Project Flow



OACN - Data Mining



ViaLactea – Data



NIR
UKIDSS: J, H, K
VISTA: K

Mid-IR
GLIMPSE: 3.6, 4.5, 5.8, 8.0 [μ]
WISE: [3 – 25] [μ]
MIPSGAL: 24 [μ]

Far-IR
Hi-GAL: 70, 160, 250, 350, 500 [μ]

Sub-mm continuum
ATLASGAL: 870 [μ]
JCMT: 870 [μ]

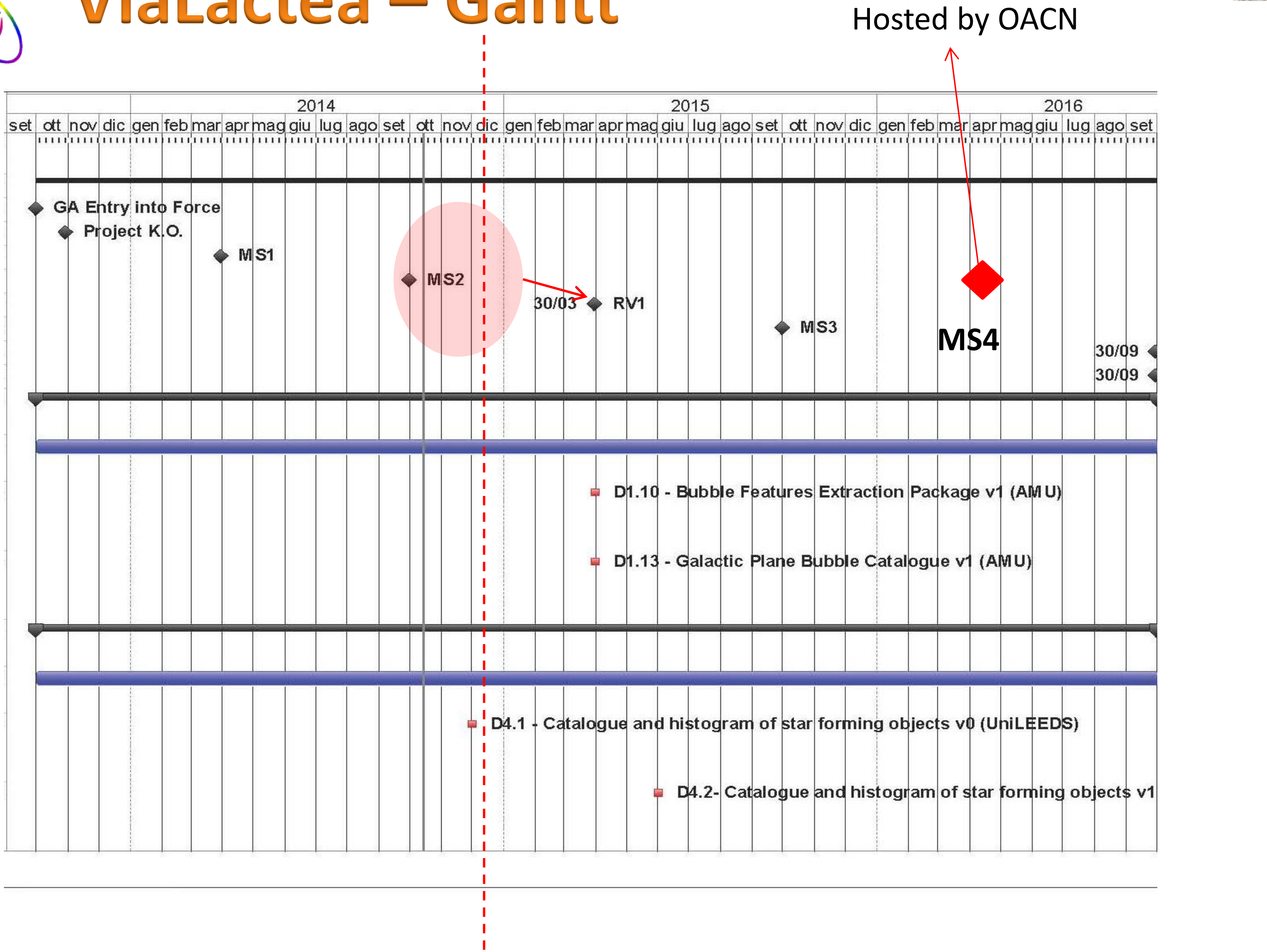
Near-IR, mid-IR and far-IR from Herschel, Spitzer and WISE
γ-ray imaging survey by AGILE and FermiLAT
Ground facilities: VISTA, JCMT, UKIRT and APEX, FCRAO, NANTEN2, VLA, Parkes, Effelsberg

Molecular and atomic line surveys
GRS: ^{13}CO at 21 cm
IGPS

Radio continuum
CORNISH: 5 [GHz] (completed by Spitzer in the mid-IR and focusing on GLIMPSE region)
MAGPIS: [0.325 – 5] [GHz] (overlaps with MIPSGAL, GLIMPSE, ATLASGAL etc…)

Molecular Masers
Methanol Multi-Beam Survey (MMB): 5 cm methanol maser emission

ViaLactea – Gantt



ViaLactea – Science vs Technology

Most of the first FTE has been spent to find a common language among members…

How astronomers see astroinformaticians

How astroinformaticians see astronomers

…with doubtful but promising results

OACN – R&D Activity report

Compact source identification through band-merging

- ✓ Q-FULLTREE method design and development done;
- ✓ Q-FULLTREE method debug and preliminary test done;
- ➢ Q-FULLTREE scientific validation in progress;
- ➢ Interaction with OACT+OATS infrastructures for integration and test in progress;

Filamentary structure Edge Detection

- ✓ FilExSeC method design and development done;
- ✓ FilExSeC method debug and preliminary test done;
- ➢ FilExSeC scientific validation with IAPS in progress;
- ❑ Started design & development of other methods;

Compact source distance estimation

- ➢ design and preliminary study in progress;
- ❑ Started preliminary interaction with IAPS for Knowledge Base definition;

Star Forming source evolutionary classification

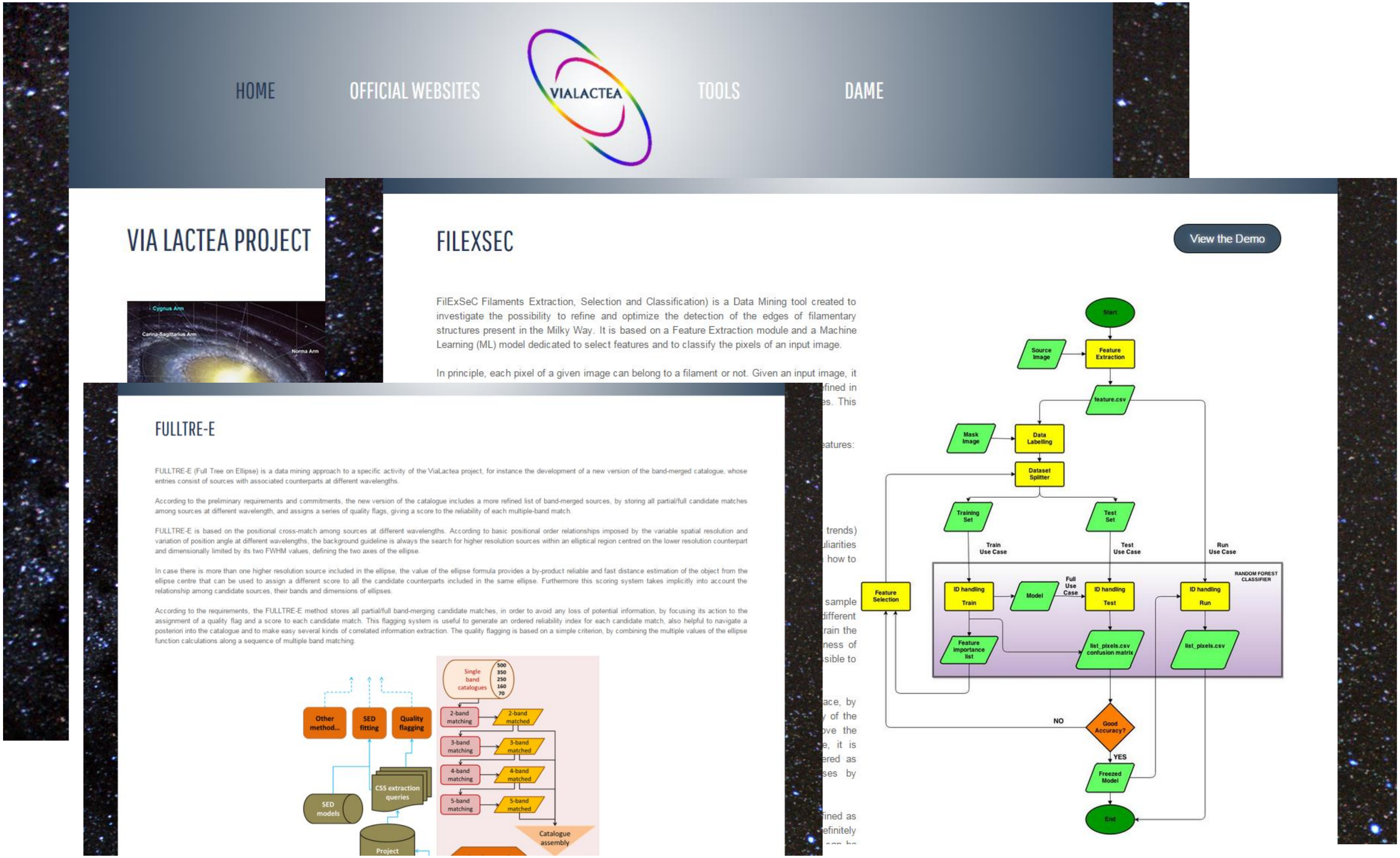
- ❑ Started design and preliminary study;
- ❑ Started interaction with IAPS for Knowledge Base definition;

Bubble structures classification

- ❑ Started interaction with OACT for Knowledge Base definition;

ViaLactea – OACN web resource

http://pc169.na.astro.it:8080



ViaLactea official Science Gateway

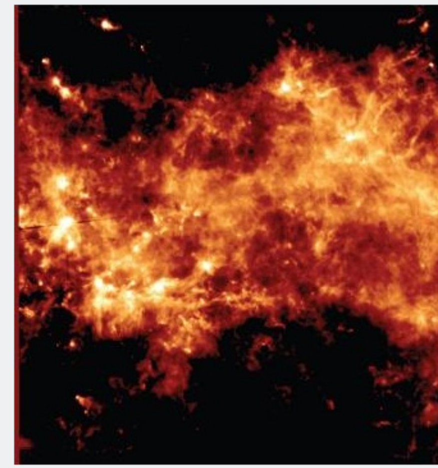
http://muoni-server-01.oact.inaf.it:8081

Sign In

VIALACTEA Project Science Gateway

Welcome | Statistics | Publications | Help

VIALACTEA Project Science Gateway > Welcome



Welcome to VIALACTEA Project Science Gateway!

The aim of the VIALACTEA project is to exploit the combination of all the new-generation Infared -> Radio surveys of the Galactic Plane from space missions and ground-based facilities, using a novel data and science analysis paradigm based on 3D visual analytics and data mining framework, to build and deliver a quantitative 3D model of our VIALACTEA Galaxy as a star formation engine that will be used as a template for external galaxies and study star formation across the cosmic time.

The main objectives of the project are:

- To boost the scientific exploitation of ESA missions space data by developing new and carefully tailored image processing tools to carry out detection and extraction of compact sources and filamentary structures, as well as more complex shape-finding to identify bubble-like features from large scale Galactic Plane imaging surveys both in the infrared continuum and in gas molecular lines.
- To combine in a VO-compatible and interoperable way the new-generation Galactic Plane surveys from space-borne missions and ground-based observatories, most of which are object of considerable European investments, to obtain a sub-arcminute resolution complete and homogeneous data coverage over the entire Galactic Plane from the infrared to the radio, that will extend the usage of already available space data and will multiply by many times the exploitation potential of the individual datasets.
- To build and visualize a new 3D representation of the Milky Way Galaxy.
- To determine, map and visualize in 3D Galaxy-wide the quantitative relationship between the physical mechanisms locally responsible for the onset, triggering and regulation of star formation, and the end products of star formation measured by the Star Formation Rate and Star Formation Efficiency.
- To bring to a common forum the scientific astronomical expertise and the e-Science technological know-how of european-leading research groups to develop the next generation data analysis tools that will enable time-effective steradiant-scale science in Galactic star formation.

Participants of the project are:

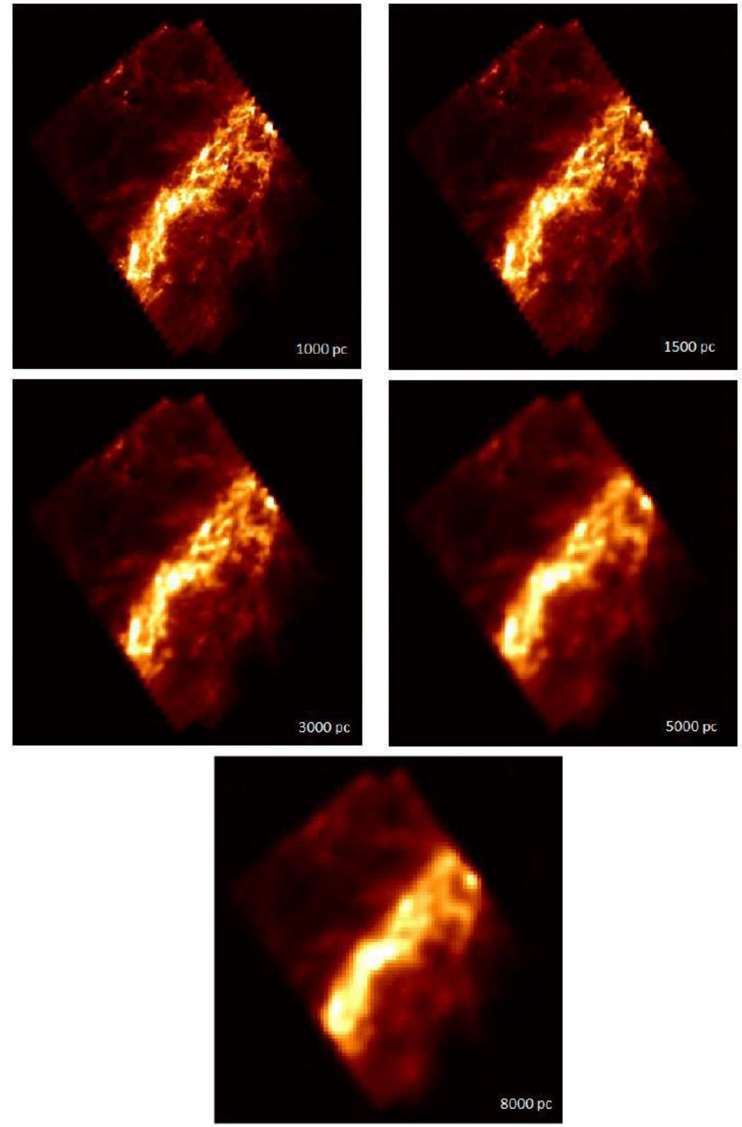
ISTITUTO NAZIONALE DI ASTROFISICA (INAF), UNIVERSITY OF LEEDS, MAX PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V., MAGYAR TUDOMANYOS AKADEMIA SZAMITASTECHIKAI ES AUTOMATIZALASI KUTATO INTEZET, CARDIFF UNIVERSITY, UNIVERSITÈ D'AIX MARSEILLE, CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, NAGOYA UNIVERSITY, THE UNIVERSITY OF EXETER

This project is funded by the European Union under the Seventh Framework Programme (FP7/2007-2013) under grant agreement no 607380 (VIALACTEA).

OACN – WP5 Activities

WP2 – Task 1
Band-merging
Q-FULLTREE

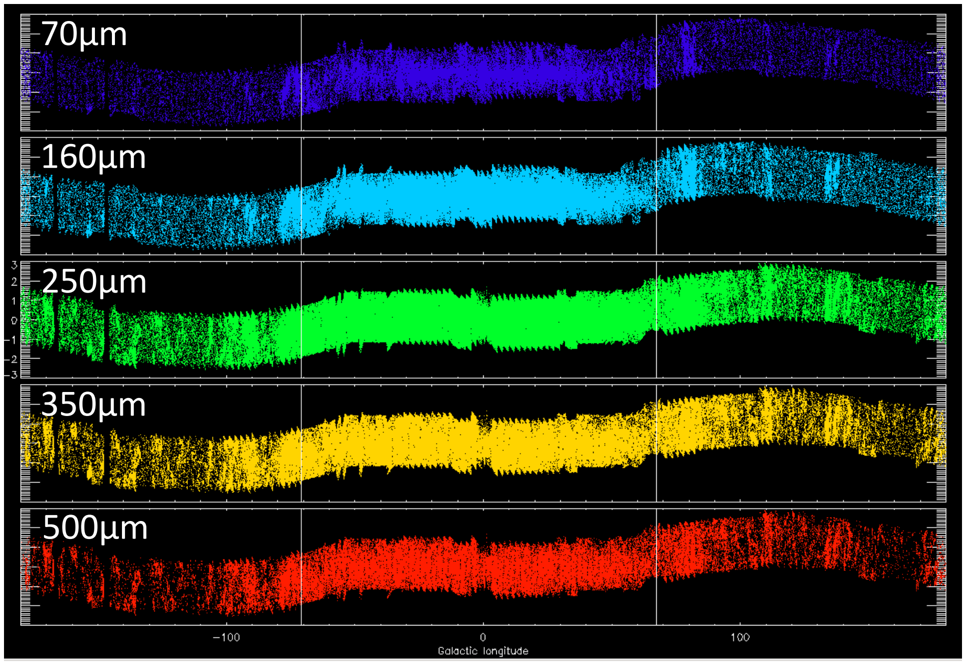
Design



WP2 – task 1 - Bandmerging

Task 1: Compact Source Extraction and band-merging

- ❑ Hi-GAL Source extraction and photometry
- ❑ Band-merging with ancillary information (from near-IR to radio)



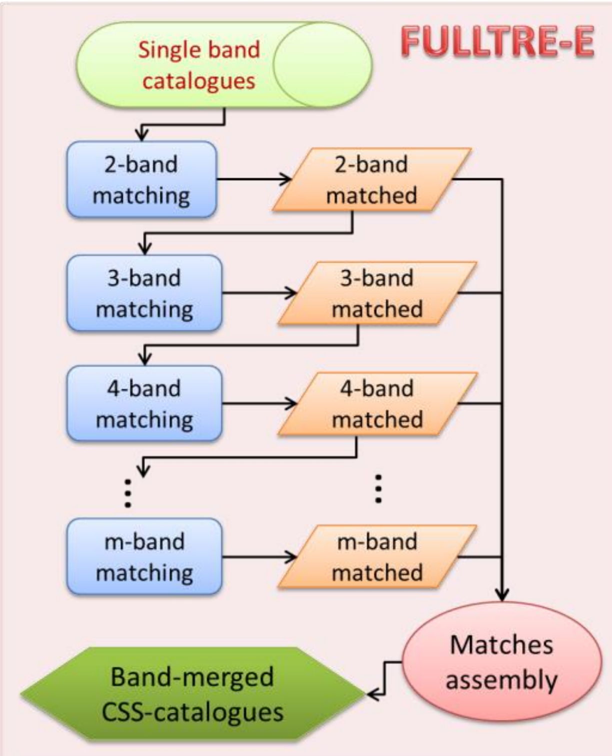
The **source extraction** with CuTEx (***Molinari et al.*, 2010a**) has been run over the entire Galactic plane.

The -71° < *l* < 67.5° portion of the HERSHEL/Hi-GAL photometry lists should be band-merged, filtered and complemented with distances and ancillary photometry : MIPSGAL, UKIDSS, WISE, MSX; ATLASGAL, BGPS …

A first result from OACN of a **band-merged catalogue** using a data-mining approach has been implemented for the Herschel bands

- ❖ Captures and maintains multiple counterpart associations;
- ❖ Topological quality flagging;
- ❖ Ingested into a VO-like database so that complex queries are possible;
- ❖ Interfaced with Visualization tools;
- ❖ Massively based on multi-threading parallelization.

WP2 – task 1 - Bandmerging



The data mining approach, named FULLTRE-E (Full Tree on Ellipse), is based on the positional cross-match among sources at different wavelengths, by always respecting the order relationship imposed by spatial resolution.

the FULLTRE-E method stores all partial/full band-merging candidate matches, in order to avoid any loss of potential information, by focusing its action to the assignment of quality flags and a score to each candidate match.

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1 \rightarrow Ell(x_{bi}, x_{bj}) \leq 1$$ **candidate match found**

a* and *b *are the two semi-axes of the ellipse (calculated upon the two given values of FWHM of the source, centre of the ellipse)*

x* and *y *are the coordinates of the higher resolution counterpart (opportunely corrected by the position angle variation)*

(x_{bi}, x_{bj}) *a match between two-band sources*

Bandmerging (FULLTRE-E Summary)

definition

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}\le 1 \rightarrow Ell(x_{bi},x_{bj})\le 1 \text{ candidate match found}$$

$$CSS=\{x_{b1},x_{b2},\dots,x_{bM}\}\equiv(x_{b1},x_{b2}),(x_{b2},x_{b3}),(x_{b1},x_{b3}),\dots,(x_{bM-1},x_{bM}).$$

Confidence Level → $CL(x_{bi},x_{bj})=1-Ell(x_{bi},x_{bj})$

scoring

number of elliptical matches NE

Theoretical NE → *TNE*$=\binom{M}{2}=\frac{M!}{(2!)[(M-2)!]}$ CL terms

Merit Score → $MS(CSS)=MS\{x_{b1},x_{b2},\dots,x_{bM}\}=\frac{NE}{TNE}\sum_{i=1}^{NE}CL_i$

quality

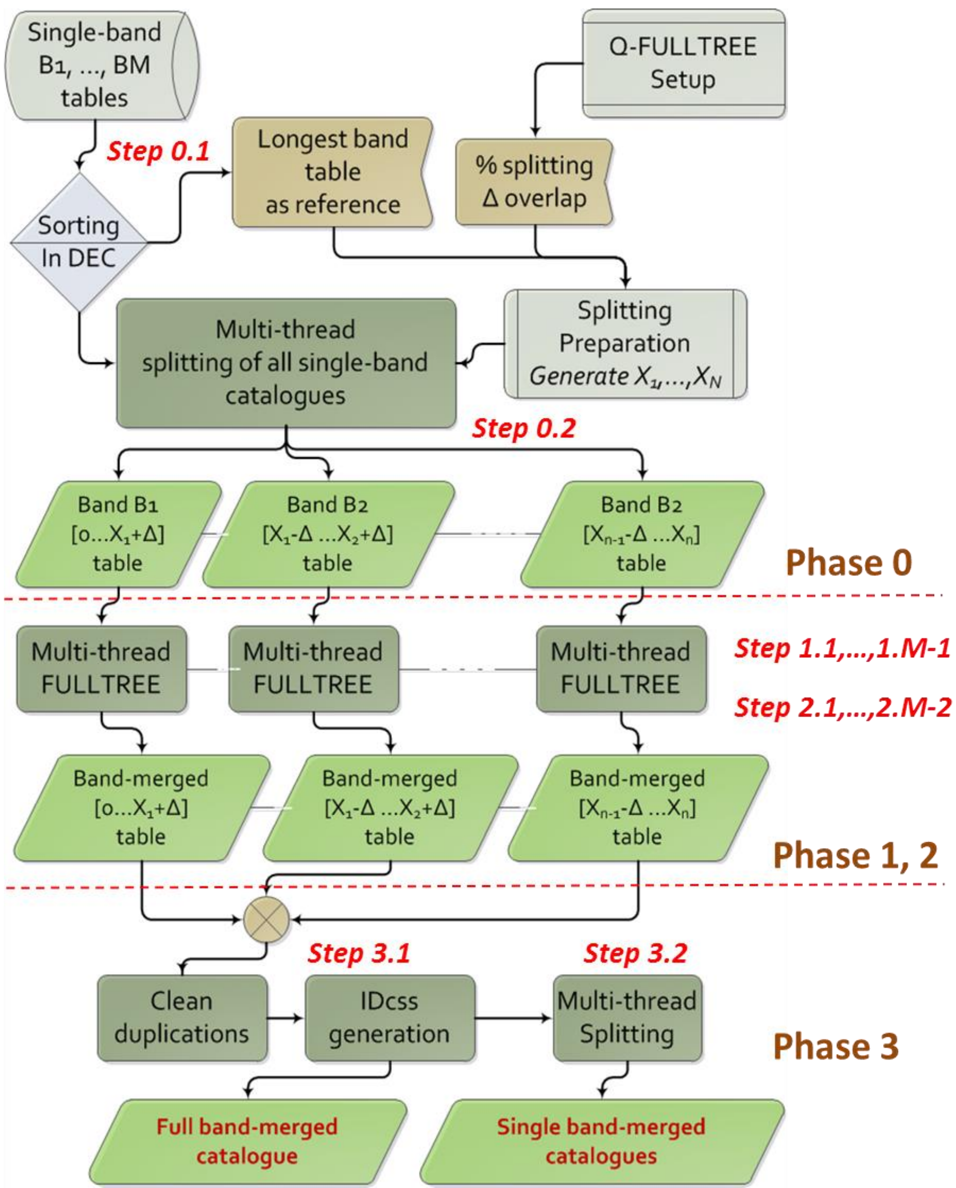
Quality Rank QR

QR = 1 to the max(MS1, MS2, MS3)
QR = 2 to the second higher value among MS1, MS2 and …
QR = 3 to the min(MS1, MS2, MS3)

Quality Fitness $QF(CSS)=\frac{MS(CSS)}{\sum_{k=1}^{N}MS(CSS_K)}$

$$\begin{cases} if\ exists\ a\ CSS_j(x_{bi})\ such\ that\ QF(CSS_j)>0.5,then\ AQF(CSS_j)=1 \\ else\ AQF(CSS_j)=0\ \forall\ j=1\dots N \end{cases}$$

Enhancing performance (Q-FULLTREE)



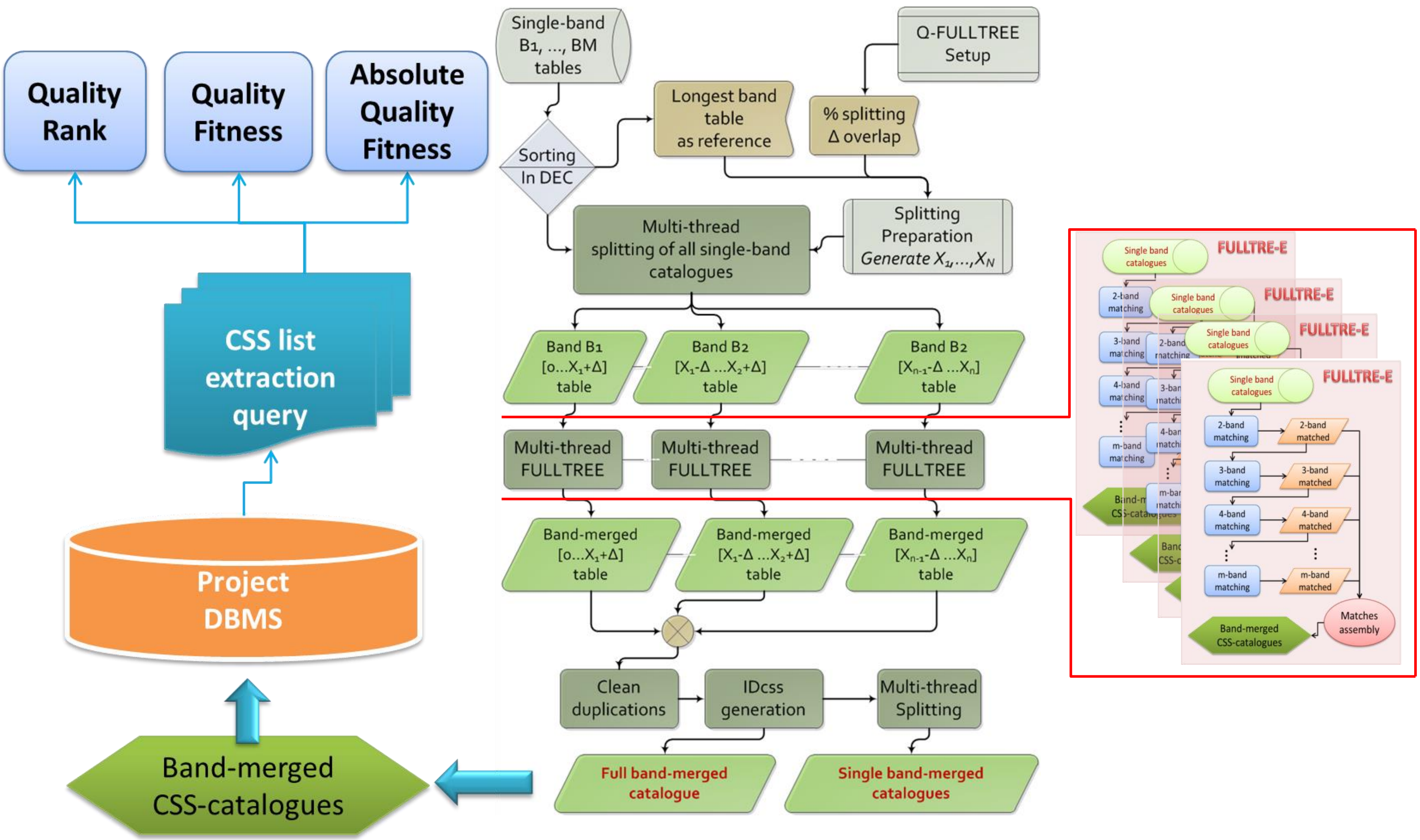
The wrapping system built around the original FULLTREE module is designed to improve the computing performance of the entire band-merging tool. Based on:

- ✓ ***Split %***: the percentage of splitting used to generate sub-tables of the input single-band catalogues;
- ✓ ***Δ overlap***: the quantity of overlap (in arcmin) around which to replicate catalogue entries in the sub-tables;
- ✓ ***Pivot band***: the reference band related to the longest single-band catalogue.

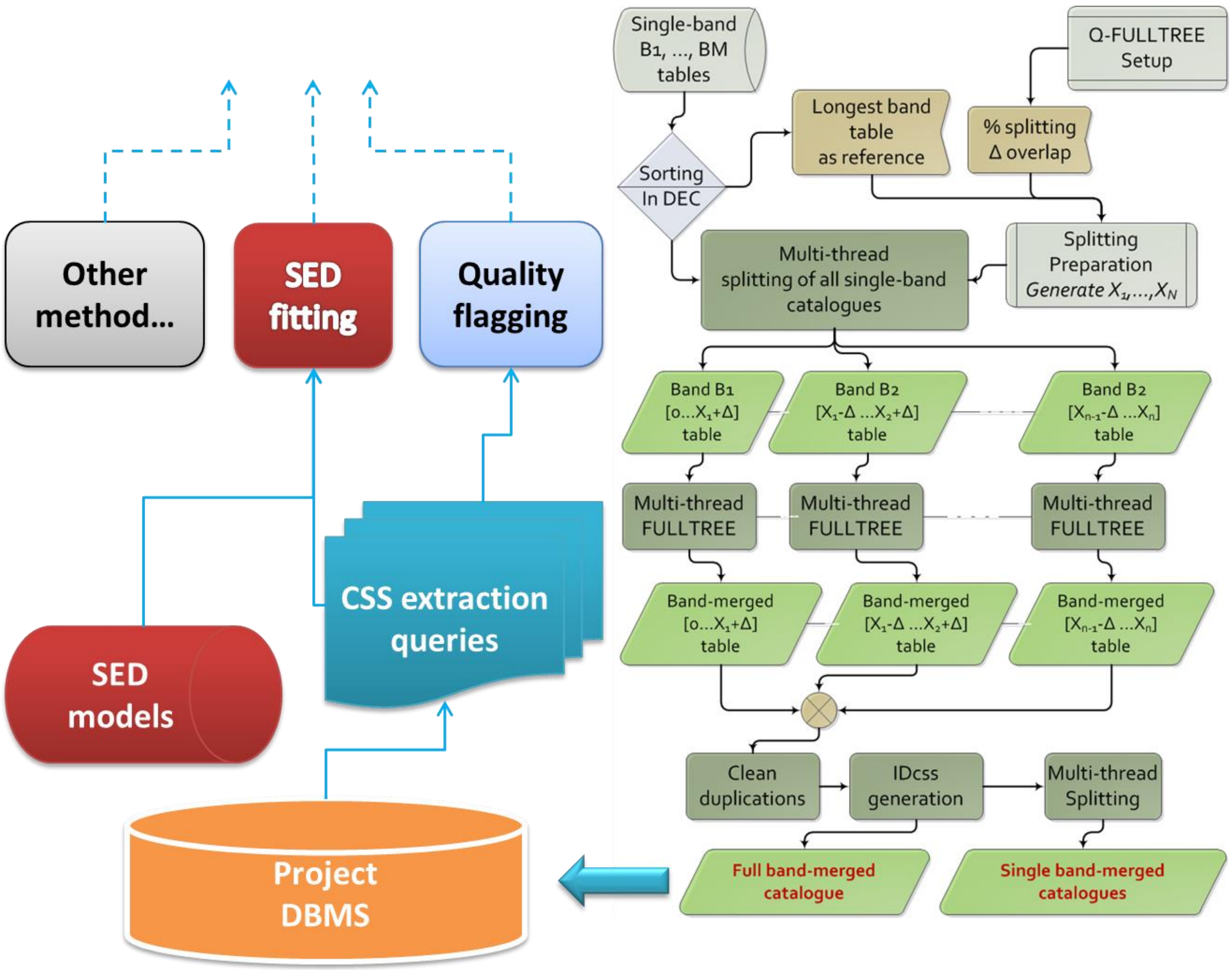
Worst gain in speedup: 200x (mostly higher)

5 bands:
on a bi-CPU 1.6GHz, 16 cores:
27 days → 3.3 hours
On a quad-CPU 2.4GHz, 32 cores:
23 days → 1.3 hours
On CT cluster (1 CPU 2.4 GHz, 12 cores):
29 days → 3,15 hours

Q-FULLTREE architecture



Bandmerging - connections



The information provided by the Q-FULLTREE output as a starting point of a workflow in which information coming from different analysis modules could be correlated to improve the overall knowledge

Bandmerging – ex. 1 (1 CSS)

FULLTREE-CSS-catalogue main bandmerged source catalogue

Table Browser for 2: band500-CSS-catalogue.csv

	ID	BANDS	ID500	RA500	DEC500	ID350	ID250	ID160	ID70	ELLIPSE500_350	ELLIPSE500_250	ELLIPSE500_160	ELLIPSE500_70	ELLIPSE350_250	ELLIPSE350_160	ELLIPSE350_70	ELLIPSE250_160	ELLIPSE250_70	ELLIPSE160_70	Merit Score
9	3-65	500-350-160	44156	261,37778	-34,37987	85671		160950		0,10382		0,18024			0,26039					2,45555
10	3-66	500-350-160	44156	261,37778	-34,37987	85671		160969		0,10382		0,14775			0,31288					2,43554
11	3-88	500-350-250	44156	261,37778	-34,37987	85671	143445			0,10382	0,48099			0,75347						1,66172
12	3-89	500-350-250	44156	261,37778	-34,37987	85671	148821			0,10382	0,16585			0,10113						2,62919
13	3-94	500-350-70	44156	261,37778	-34,37987	85671			63621	0,10382			0,27085			0,51011				2,11521
14	3-95	500-350-70	44156	261,37778	-34,37987	85671			62851	0,10382			0,39904			0,64854				1,84861
15	4-42	500-250-160-70	44156	261,37778	-34,37987		143445	160969	63621		0,48099	0,14775	0,27085				0,40939	0,13215	0,07198	4,48689
16	4-43	500-250-160-70	44156	261,37778	-34,37987		148821	160950	62851		0,16585	0,18024	0,39904				0,09382	0,42777	0,20995	4,52334
17	4-44	500-250-160-70	44156	261,37778	-34,37987		148821	160969	63621		0,16585						0,59905	0,85931	0,07198	1,53587
18	4-61	500-350-160-70	44156	261,37778	-34,37987	85671		160950	62851	0,10382		0,18024	0,39904		0,26039	0,64854			0,20995	4,19803
19	4-62	500-350-160-70	44156	261,37778	-34,37987	85671		160969	63621	0,10382		0,14775	0,27085		0,31288	0,51011			0,07198	4,5826
20	4-76	500-350-250-160	44156	261,37778	-34,37987	85671	143445	160969		0,10382	0,48099	0,14775		0,75347	0,31288		0,40939			3,7917
21	4-77	500-350-250-160	44156	261,37778	-34,37987	85671	148821	160950		0,10382	0,16585	0,18024		0,10113	0,26039		0,09382			5,09475
22	4-78	500-350-250-160	44156	261,37778	-34,37987	85671	148821	160969		0,10382				0,10113			0,59905			1,098
23	4-83	500-350-250-70	44156	261,37778	-34,37987	85671	143445		63621	0,10382	0,48099		0,27085	0,75347		0,51011		0,13215		3,7486
24	4-84	500-350-250-70	44156	261,37778	-34,37987	85671	148821		63621	0,10382	0,16585			0,10113				0,85931		1,84659
25	4-85	500-350-250-70	44156	261,37778	-34,37987	85671	148821		62851	0,10382			0,39904	0,10113		0,64854		0,42777		2,76642
26	5-71	500-350-250-160-70	44156	261,37778	-34,37987	85671	143445	160969	63621	0,10382	0,48099	0,14775	0,27085	0,75347	0,31288	0,51011	0,40939	0,13215	0,07198	6,8066
27	5-72	500-350-250-160-70	44156	261,37778	-34,37987	85671	148821	160950	62851	0,10382	0,16585	0,18024	0,39904	0,10113	0,26039	0,64854	0,09382	0,42777	0,20995	7,40946
28	5-73	500-350-250-160-70	44156	261,37778	-34,37987	85671	148821	160969	63621	0,10382				0,10113			0,59905	0,85931	0,07198	1,63236
29	3-33	500-160-70	44159	261,35282	-34,38468			160999	63628			0,15269	0,11295						0,98641	1,74795
30	3-34	500-160-70	44159	261,35282	-34,38468			160999	63631			0,15269	0,124						0,182	2,54132
31	3-35	500-160-70	44159	261,35282	-34,38468			160999	67243			0,15269	0,30228						0,0979	2,44713
32	3-46	500-250-160	44159	261,35282	-34,38468		148822	160999			0,08791	0,15269					0,04081			2,7186
33	3-51	500-250-70	44159	261,35282	-34,38468		148822		63628		0,08791		0,11295					0,67655		2,1226
34	3-52	500-250-70	44159	261,35282	-34,38468		148822		63631		0,08791		0,124					0,02401		2,76408
35	3-53	500-250-70	44159	261,35282	-34,38468		148822		67243		0,08791		0,30228					0,22907		2,38074
36	3-64	500-350-160	44159	261,35282	-34,38468	82283		160999		0,03384		0,15269			0,09883					2,71464
37	3-87	500-350-250	44159	261,35282	-34,38468	82283	148822			0,03384	0,08791			0,04732						2,83093
38	3-91	500-350-70	44159	261,35282	-34,38468	82283			63628	0,03384			0,11295			0,19333				2,65989
39	3-92	500-350-70	44159	261,35282	-34,38468	82283			63631	0,03384			0,124			0,11036				2,7318
40	3-93	500-350-70	44159	261,35282	-34,38468	82283			67243	0,03384			0,30228			0,28112				2,38277
41	4-39	500-250-160-70	44159	261,35282	-34,38468		148822	160999	63628		0,08791	0,15269	0,11295				0,04081	0,67655	0,98641	3,94269
42	4-40	500-250-160-70	44159	261,35282	-34,38468		148822	160999	63631		0,08791		0,124				0,04081	0,02401	0,182	3,7844
43	4-41	500-250-160-70	44159	261,35282	-34,38468		148822	160999	67243		0,08791		0,30228				0,04081	0,22907	0,0979	3,53503
44	4-58	500-350-160-70	44159	261,35282	-34,38468	82283		160999	63628	0,03384		0,15269	0,11295		0,09883	0,19333			0,98641	4,42196
45	4-59	500-350-160-70	44159	261,35282	-34,38468	82283		160999	63631	0,03384			0,124		0,09883	0,11036			0,182	3,70914
46	4-60	500-350-160-70	44159	261,35282	-34,38468	82283		160999	67243	0,03384			0,30228		0,09883	0,28112			0,0979	3,48836
47	4-75	500-350-250-160	44159	261,35282	-34,38468	82283	148822	160999		0,03384	0,08791	0,15269		0,04732	0,09883		0,04081			5,53861
48	4-80	500-350-250-70	44159	261,35282	-34,38468	82283	148822		63628	0,03384	0,08791		0,11295	0,04732		0,19333		0,67655		4,84811
49	4-81	500-350-250-70	44159	261,35282	-34,38468	82283	148822		63631	0,03384			0,124	0,04732		0,11036		0,02401		3,88373
50	4-82	500-350-250-70	44159	261,35282	-34,38468	82283	148822		67243	0,03384			0,30228	0,04732		0,28112		0,22907		3,42198
51	5-68	500-350-250-160-70	44159	261,35282	-34,38468	82283	148822	160999	63628	0,03384	0,08791	0,15269	0,11295	0,04732	0,09883	0,19333	0,04081	0,67655	0,98641	7,56937
52	5-69	500-350-250-160-70	44159	261,35282	-34,38468	82283	148822	160999	63631	0,03384			0,124	0,04732		0,11036	0,04081	0,02401	0,182	4,50637
53	5-70	500-350-250-160-70	44159	261,35282	-34,38468	82283	148822	160999	67243	0,03384			0,30228	0,04732		0,28112	0,04081	0,22907	0,0979	4,17737

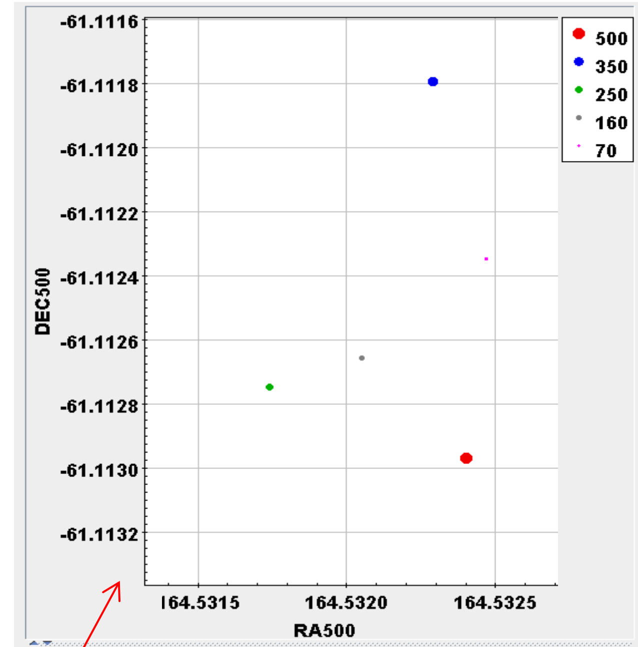


Table Browser for 2: band500-CSS-catalogue.csv

	ID	BANDS	ID500	RA500	DEC500	ID350	ID250	ID160	ID70	ELLIPSE500_350	ELLIPSE500_250	ELLIPSE500_160	ELLIPSE500_70	ELLIPSE350_250	ELLIPSE350_160	ELLIPSE350_70	ELLIPSE250_160	ELLIPSE250_70	ELLIPSE160_70	Merit Score
54	3-38	500-160-70	85295	164,5324	-61,11297			290704	122181			0,00556	0,0138						0,03477	2,94586
55	3-50	500-250-160	85295	164,5324	-61,11297		278257	290704			0,01246	0,00556					0,0114			2,97059
56	3-57	500-250-70	85295	164,5324	-61,11297		278257		122181		0,01246		0,0138					0,06956		2,90418
57	3-67	500-350-160	85295	164,5324	-61,11297	160747		290704		0,04573		0,00556			0,03048					2,91823
58	3-90	500-350-250	85295	164,5324	-61,11297	160747	278257			0,04573	0,01246			0,05057						2,89124
59	3-96	500-350-70	85295	164,5324	-61,11297	160747			122181	0,04573			0,0138			0,01445				2,92601
60	4-45	500-250-160-70	85295	164,5324	-61,11297		278257	290704	122181		0,01246	0,00556	0,0138				0,0114	0,06956	0,03477	5,85245
61	4-63	500-350-160-70	85295	164,5324	-61,11297	160747		290704	122181	0,04573		0,00556	0,0138		0,03048	0,01445			0,03477	5,85521
62	4-79	500-350-250-160	85295	164,5324	-61,11297	160747	278257	290704		0,04573	0,01246	0,00556		0,05057	0,03048		0,0114			5,84381
63	4-86	500-350-250-70	85295	164,5324	-61,11297	160747	278257		122181	0,04573	0,01246		0,0138	0,05057		0,01445		0,06956		5,79343
64	5-74	500-350-250-160-70	85295	164,5324	-61,11297	160747	278257	290704	122181	0,04573	0,01246	0,00556	0,0138	0,05057	0,03048	0,01445	0,0114	0,06956	0,03477	9,71122

Example of a full match (5 Hi-GAL bands, from 500μ to 70μ)

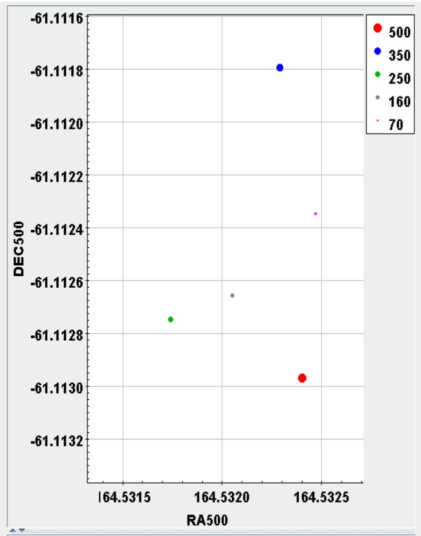
In the light-blue sub-table all the intermediate CSS found are shown

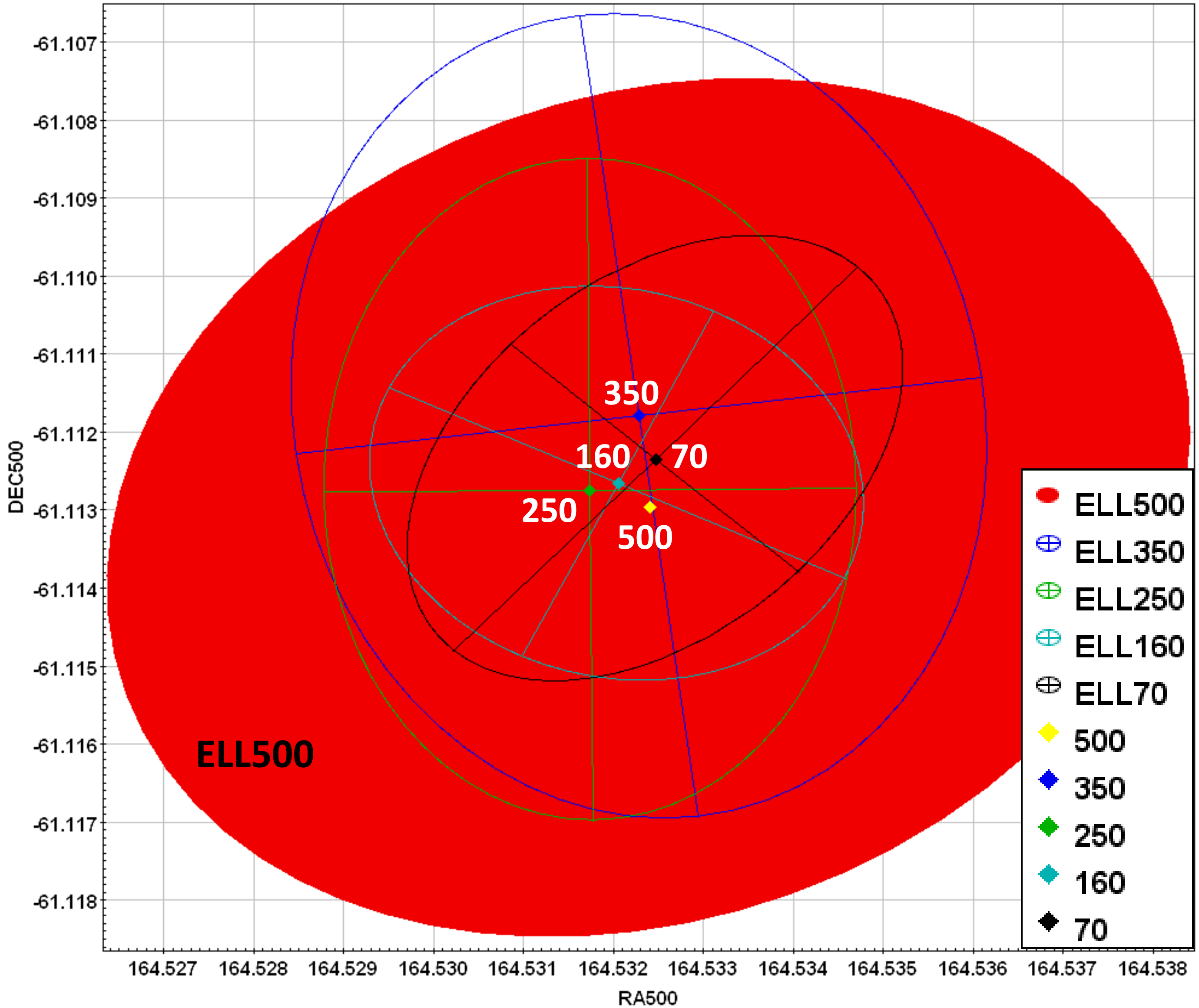
Bandmerging – ex. 1 (1 CSS)

Table Browser for 6: band500-ex2.csv

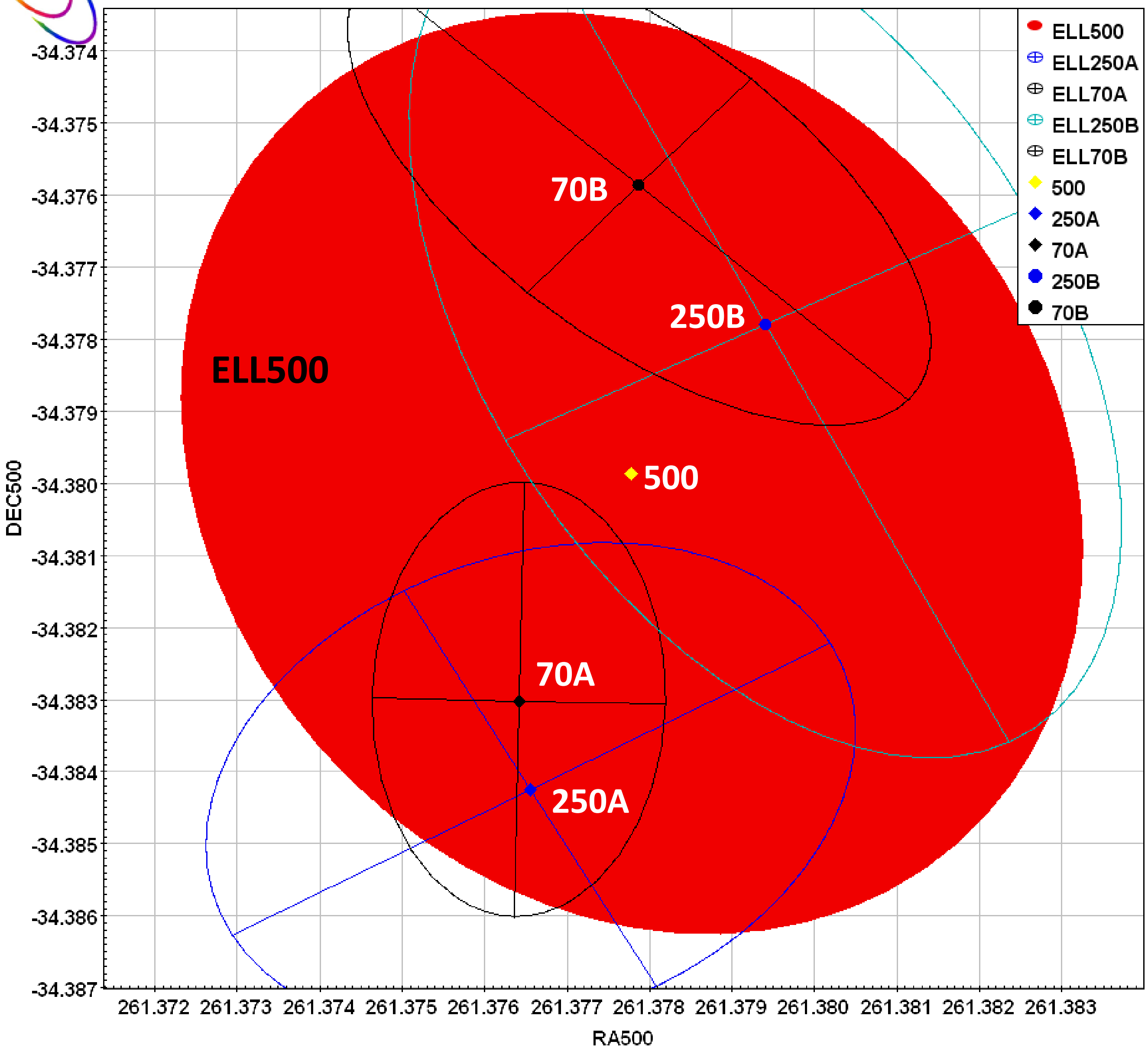
	ID	BANDS	ID500	ID350	ID250	ID160	ID70	ELLIPSE500_350	ELLIPSE500_250	ELLIPSE500_160	ELLIPSE500_70	ELLIPSE350_250	ELLIPSE350_160	ELLIPSE350_70	ELLIPSE250_160	ELLIPSE250_70	ELLIPSE160_70	Merit Score
1	5-74	500-350-250-160-70	85295	160747	278257	290704	122181	0,04573	0,01246	0,00556	0,0138	0,05057	0,03048	0,01445	0,0114	0,06956	0,03477	9,71122

MS = 9.71





Bandmerging – ex.2 (3 CSS)



Example of 3 CSS with 3-band matches

CSS_1 = 500-250A-70A

$MS_1 = 2.12$
$QR_1 = 1$
$QF_1 = 0.44$

CSS_2 = 500-250B-70A

$MS_2 = 0,65$
$QR_2 = 3$
$QF_2 = 0.14$

250B ≠ 70A

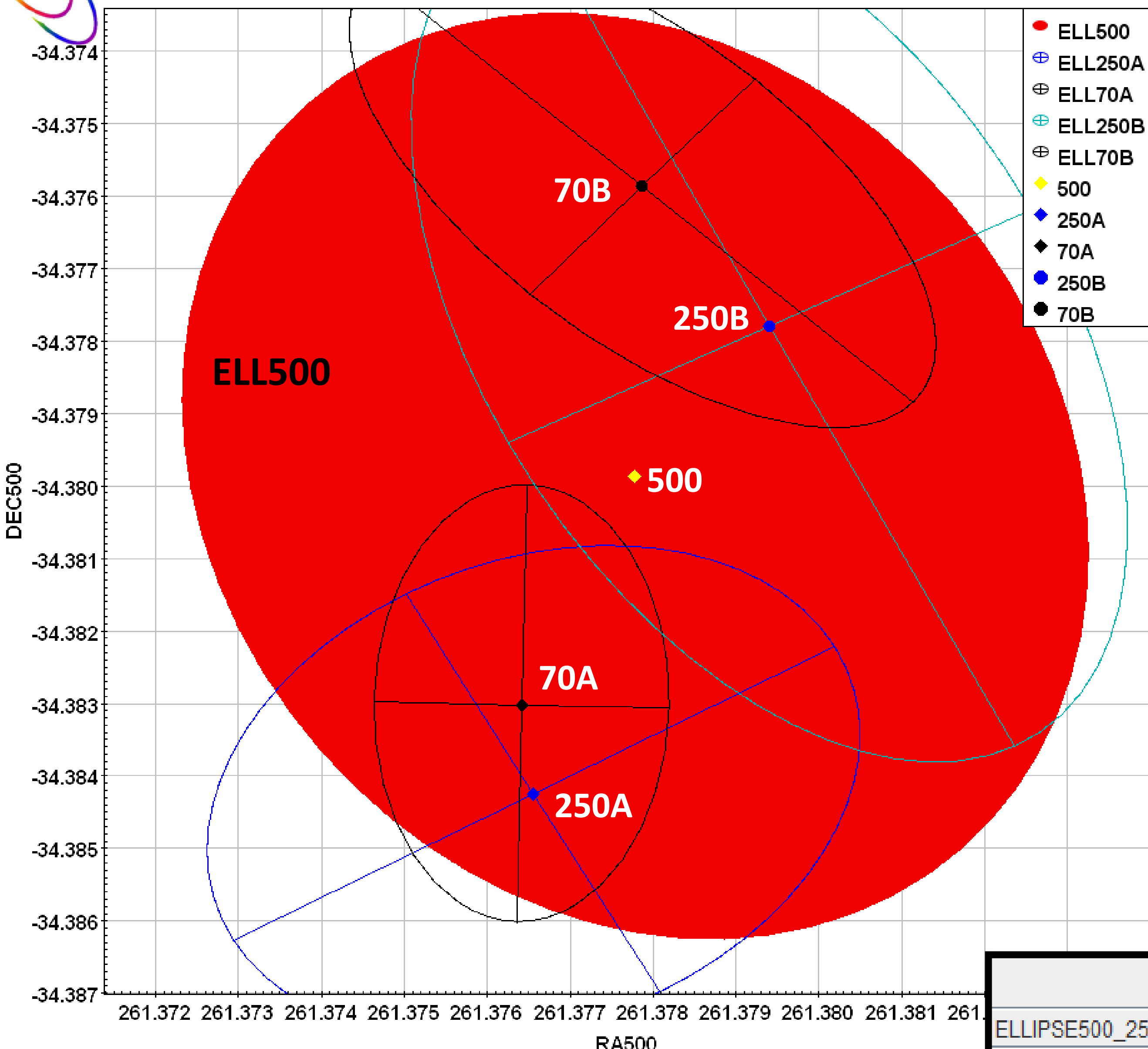
CSS_3 = 500-250B-70B

$MS_3 = 2,01$
$QR_3 = 2$
$QF_3 = 0.42$

Table Browser for 2: band500-ex1.csv

	ID	BANDS	ID500	RA500	DEC500	FWHMA500	FWHMB500	PA500	ID250	RA250	DEC250	FWHMA250	FWHMB250	PA250	ID70	RA70	DEC70	FWHMA70	FWHMB70	PA70	ELLIPSE500_250	ELLIPSE500_70	ELLIPSE250_70	Merit Score
1	3-54	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	143445	261,37656	-34,38425	22,67	29,93	209,2...	63621	261,37642	-34,38302	12,81	21,72	-91,2...	0,48099	0,27085	0,13215	2,11601
2	3-55	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	148821	261,37941	-34,3778	46,88	25,46	116,9...	63621	261,37642	-34,38302	12,81	21,72	-91,2...	0,16585	0,27085		0,64989
3	3-56	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	148821	261,37941	-34,3778	46,88	25,46	116,9...	62851	261,37787	-34,37587	31,78	14,6	-222,...	0,16585	0,39904	0,42777	2,00734

Bandmerging – ex.2 (3 CSS)



Example of 3 CSS with 3-band matches

CSS_1 = 500-250A-70A

MS_1 = 2.12
QR_1 = 1
QF_1 = 0.44

CSS_2 = 500-250B-70A

MS_2 = 0,65
QR_2 = 3
QF_2 = 0.14

250B ≠ 70A

CSS_3 = 500-250B-70B

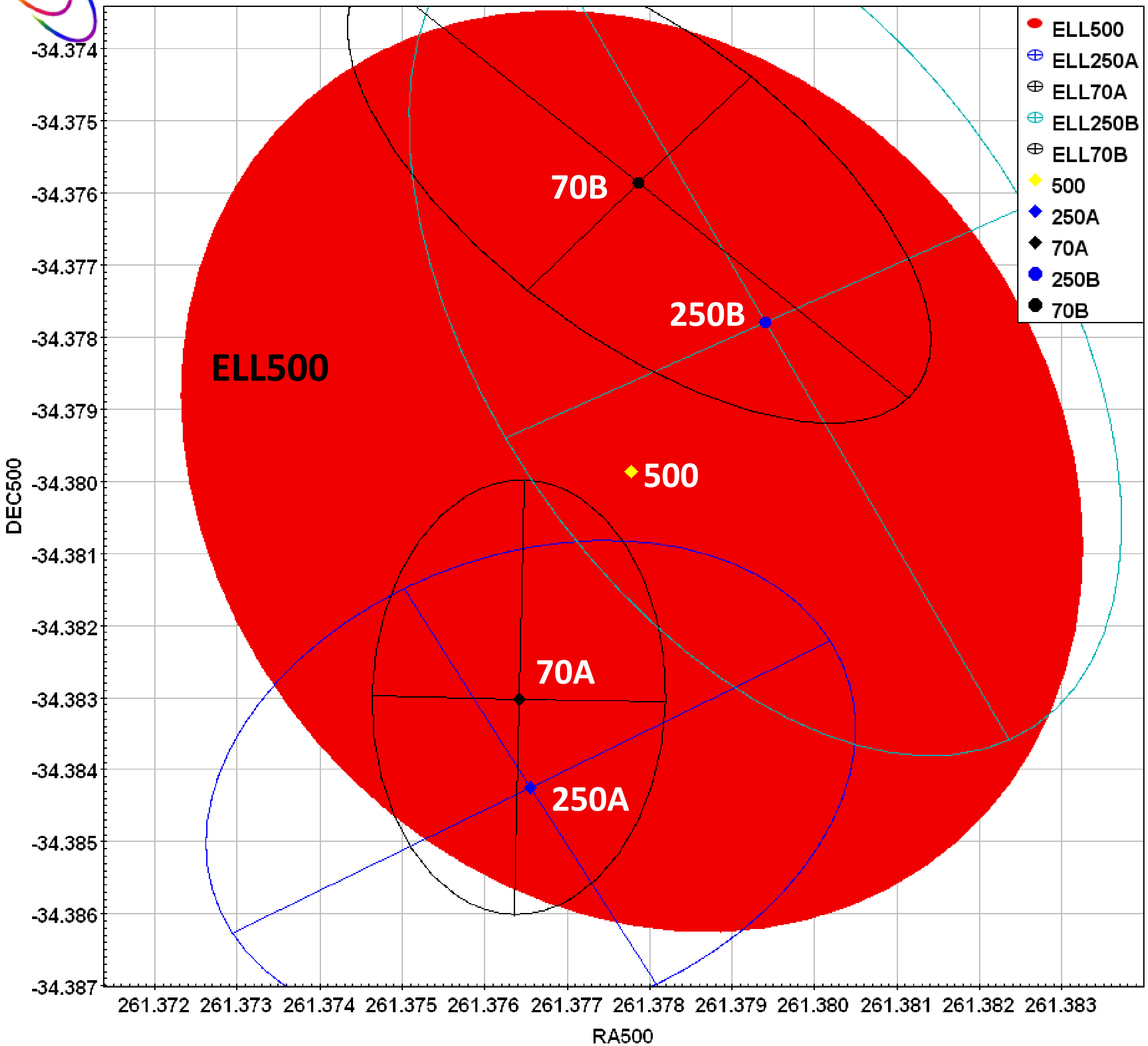
MS_3 = 2,01
QR_3 = 2
QF_3 = 0.42

Table Browser for 2: band500-ex1.csv

	ID	BANDS	ID500	RA500	DEC500	FWHMA500	FWHMB500	PA500	ID250	RA250	DEC250	FWHMA250	FWHMB250	PA250	ID70	
1	3-54	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	143445	261,37656	-34,38425	22,67	29,93	209,2...	63621	26
2	3-55	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	148821	261,37941	-34,3778	46,88	25,46	116,9...	63621	26
3	3-56	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	148821	261,37941	-34,3778	46,88	25,46	116,9...	62851	26

ELLIPSE500_250	ELLIPSE500_70	ELLIPSE250_70	Merit Score
0,48099	0,27085	0,13215	2,11601
0,16585	0,27085		0,64989
0,16585	0,39904	0,42777	2,00734

Bandmerging – ex.2 (3 CSS)



Example of 3 CSS with 3-band matches

CSS_1 = 500-250A-70A

MS_1 = 2.12
QR_1 = 1
QF_1 = 0.44

?

CSS_2 = 500-250B-70A

MS_2 = 1,65
QR_2 = 3
QF_2 = 0.14

X

250B ≠ 70A

CSS_3 = 500-250B-70B

MS_3 = 2,01
QR_3 = 2
QF_3 = 0.42

?

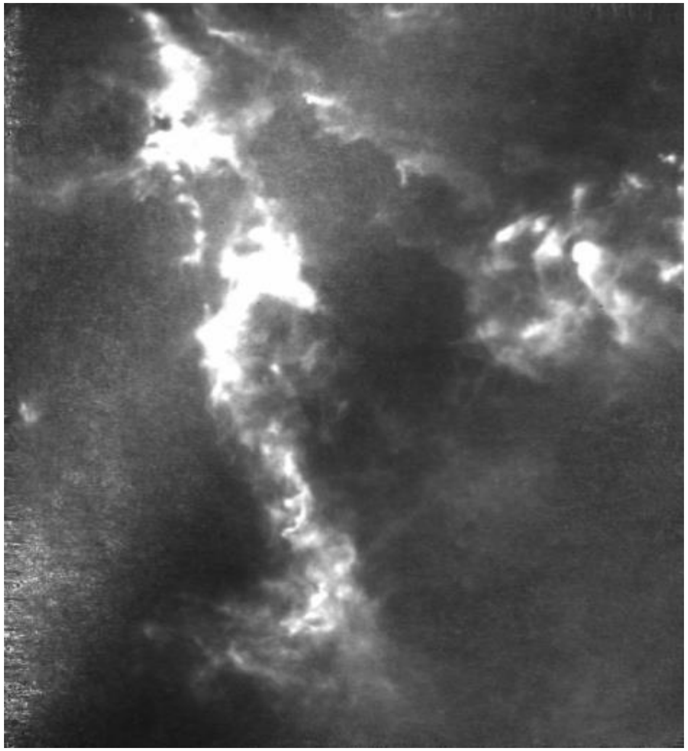
AQF = 0

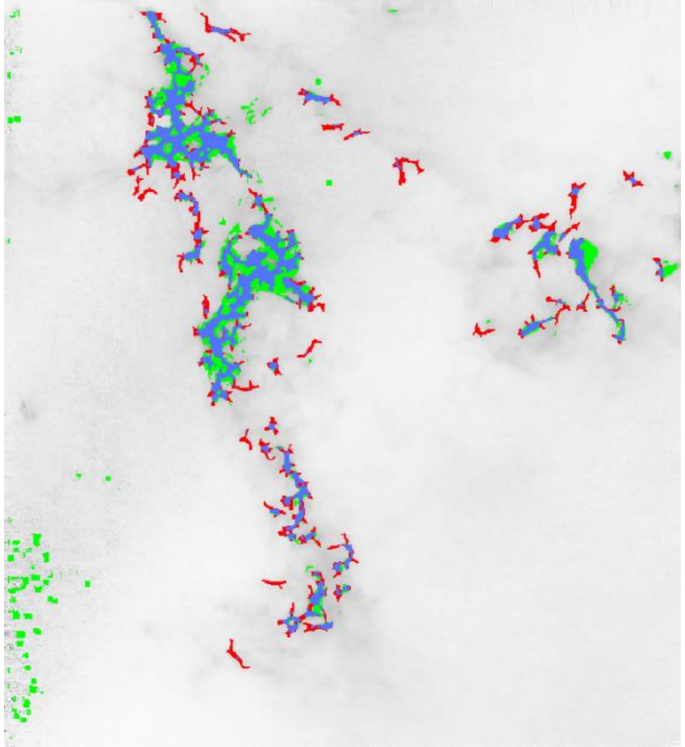
Table Browser for 2: band500-ex1.csv

	ID	BANDS	ID500	RA500	DEC500	FWHMA500	FWHMB500	PA500	ID250	RA250	DEC250	FWHMA250	FWHMB250	PA250	ID70	RA70	DEC70	FWHMA70	FWHMB70	PA70	ELLIPSE500_250	ELLIPSE500_70	ELLIPSE250_70	Merit Score
1	3-54	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	143445	261,37656	-34,38425	22,67	29,93	209,2...	63621	261,37642	-34,38302	12,81	21,72	-91,2...	0,48099	0,27085	0,13215	2,11601
~~2~~	~~3-55~~	~~500-250-70~~	~~44156~~	~~261,37778~~	~~-34,37987~~	~~47,48~~	~~37,59~~	~~-66,5...~~	~~148821~~	~~261,37941~~	~~-34,3778~~	~~46,88~~	~~25,46~~	~~116,9...~~	~~63621~~	~~261,37642~~	~~-34,38302~~	~~12,81~~	~~21,72~~	~~-91,2...~~	~~0,16585~~	~~0,27085~~		~~0,64989~~
3	3-56	500-250-70	44156	261,37778	-34,37987	47,48	37,59	-66,5...	148821	261,37941	-34,3778	46,88	25,46	116,9...	62851	261,37787	-34,37587	31,78	14,6	-222,...	0,16585	0,39904	0,42777	2,00734

OACN – WP5 Activities

WP1 – Task 1
Filament Detection Optimization
FilExSeC





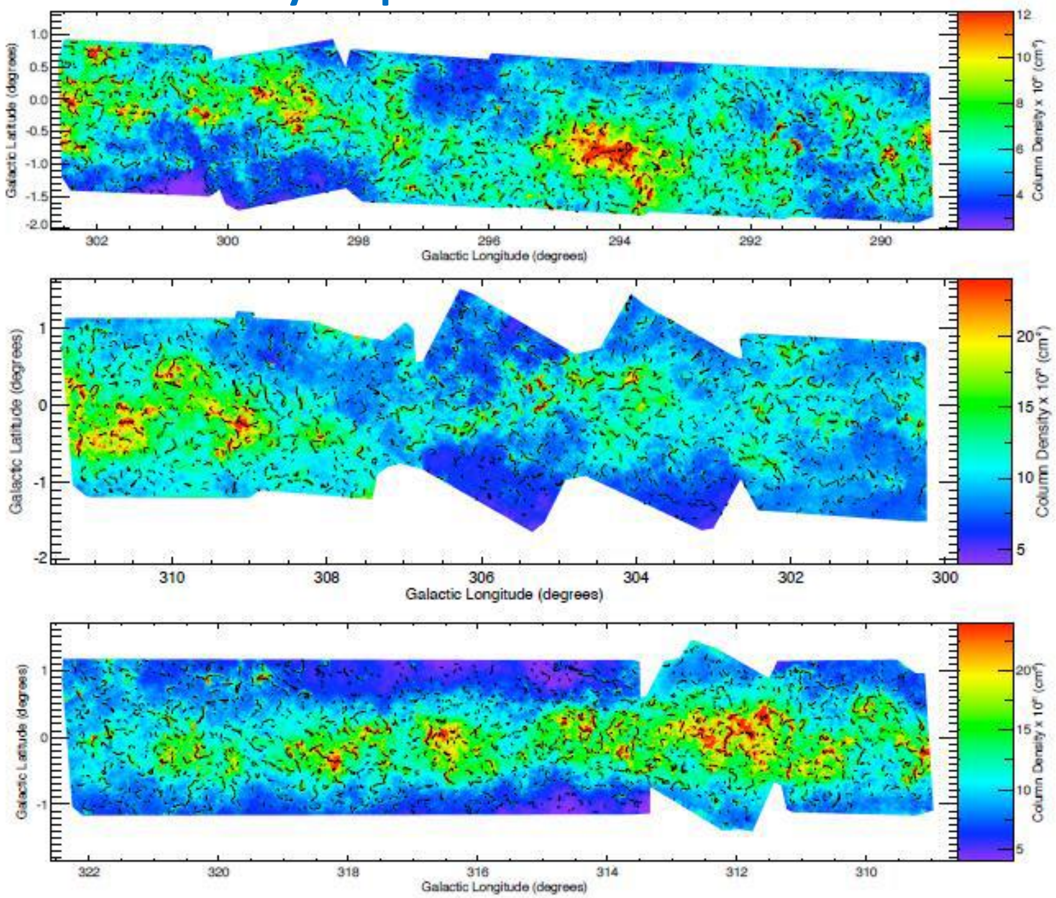
Design

WP1 – task 1 – FilExSeC

Task 1: Filamentary structure detection

- ❑ Production of column density maps of entire galactic plane
- ❑ Automated filament extraction workflow for Hi-GAL survey

Column density maps with the identified filaments



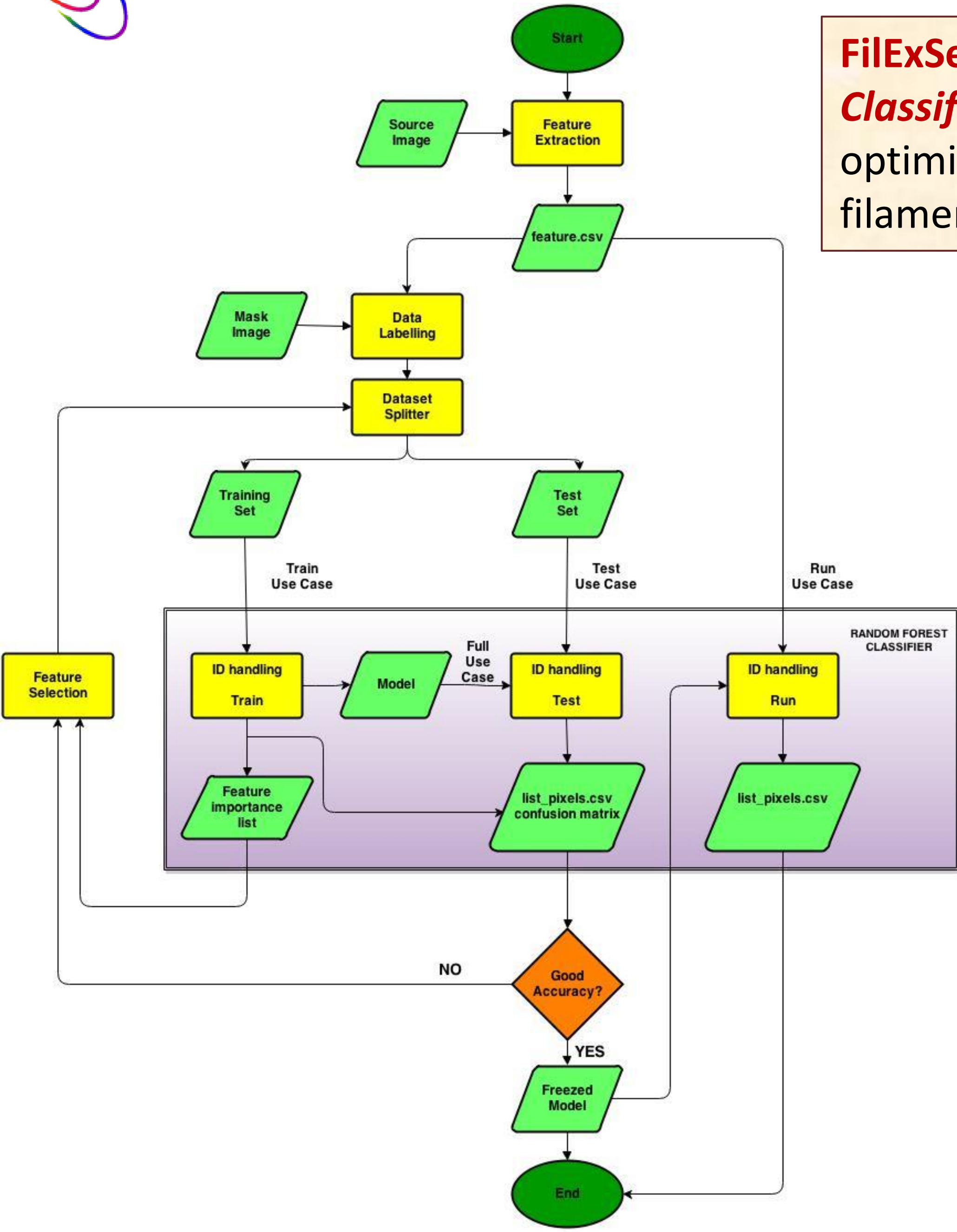
The filament extraction code was run on the column density maps covering the region between Galactic longitude 290° -- 320°, with different threshold levels equal to 2.5, 3. and 3.5 times the local standard deviation of the minimum eigenvalue (***Schisano et al.*****, 2014**)

OACN Data Mining goal:

- ❖ To refine the edges;
- ❖ To extend filament regions.

The right calculation of physical quantities related to filaments strongly depends on their dimensions, so the correct definition of edges is crucial.

WP1 – task 1 – FilExSeC



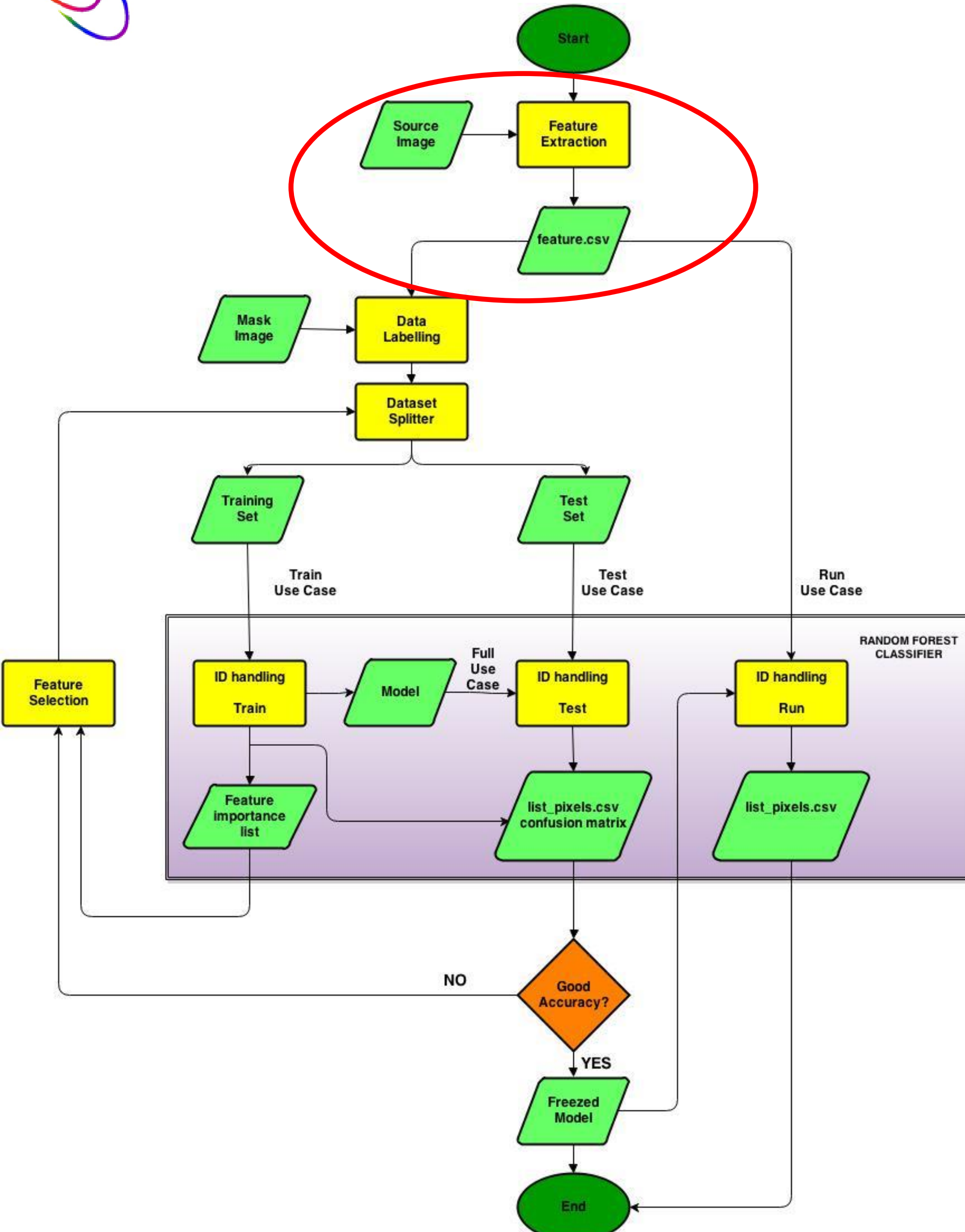
FilExSeC (*Filaments Extraction, Selection and Classification*), a data mining tool to refine and optimize the detection of the edges of filamentary structures.

The method consists in two main phases:

- ***Feature Extraction***: a set of features depending by its neighbors is associated to each pixel of the input image
- ***Classification***: image pixels are classified as filamentary or background, by using a supervised Machine Learning method, trained by these features

A further phase, ***Feature Selection***, finds the most relevant features. By reducing the initial data parameter space, it is possible indeed to improve the execution efficiency of the model, without affecting its performances.

FilExSeC (Feature Extraction)



Given an input image, it is possible to characterize each pixel by means of a set of features.

3 types of features are extracted for each pixel:

- Haralick features (***Haralick*, 1979*);**
- Haar-like features (***Viola & Jones***, 2001);
- Statistical features.

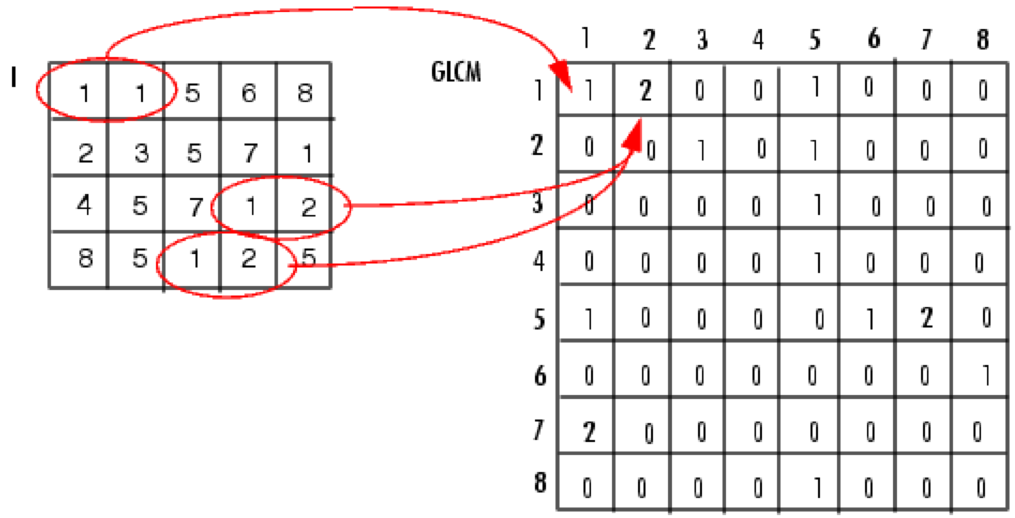
For most of the extracted features it is expected to have peculiar correlated values (or trends) for the pixels belonging to a filament, although hidden by background noise.

These peculiarities can be indeed used by a ML algorithm in order to learn how to discriminate the hidden correlation among filament pixels

FilExSeC (Features)

Haralick Features (1979)

- Based on co-occurrence matrix (GLCM)
- Element $C_{i,j}$ represents, for a fixed distance and direction, the probability to have two pixels in the image at that distance, with grey level Z_i and Z_j , respectively



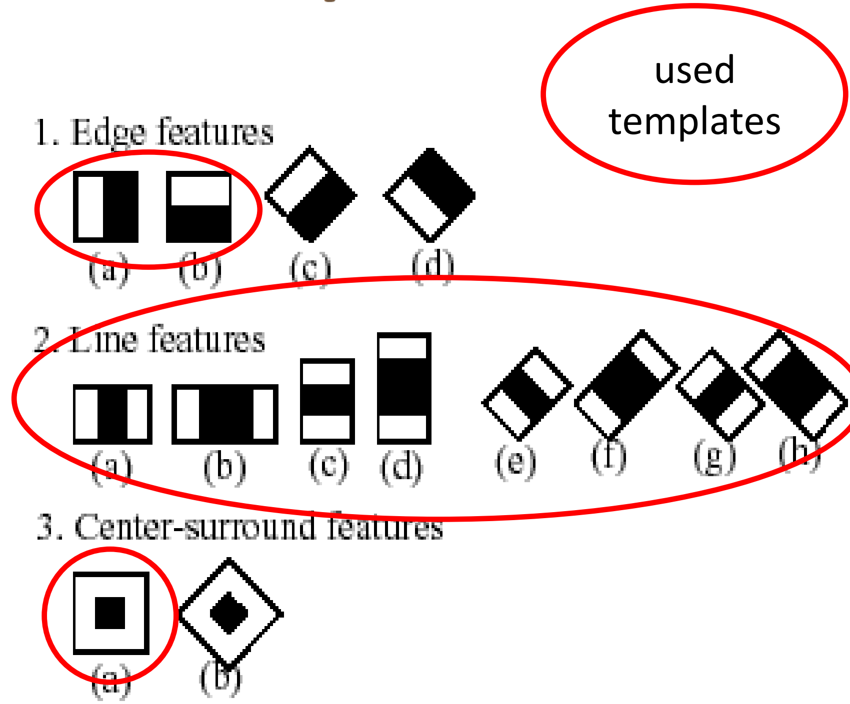
Haralick features extracted from $C_{i,j}$/(number of pairs)

Contrast	$m = \sum_i \sum_j (i-j)^2 C_{i,j}$
Energy	$\sum_i \sum_j C_{i,j}^2$
Entropy	$-\sum_i \sum_j C_{i,j} \log C_{i,j}$
Correlation	$\frac{\sum_i \sum_j (i-\mu_i)(j-\mu_j) C_{i,j}}{\sigma_i \sigma_j}$

Robert Haralick

FilExSeC (Features)

Paul Viola



Haar-like Features (2001)

- Each Haar-like variable involves 2 or 3 interconnected black and white rectangles (masks or templates)
- Values of each feature are obtained by sliding masks on the image and calculating:

$$f = \sum_{\substack{black\\rectangle}} \begin{pmatrix} grey\ level \\ image\ pixel \end{pmatrix} - \sum_{\substack{white\\rectangle}} \begin{pmatrix} grey\ level \\ image\ pixel \end{pmatrix}$$

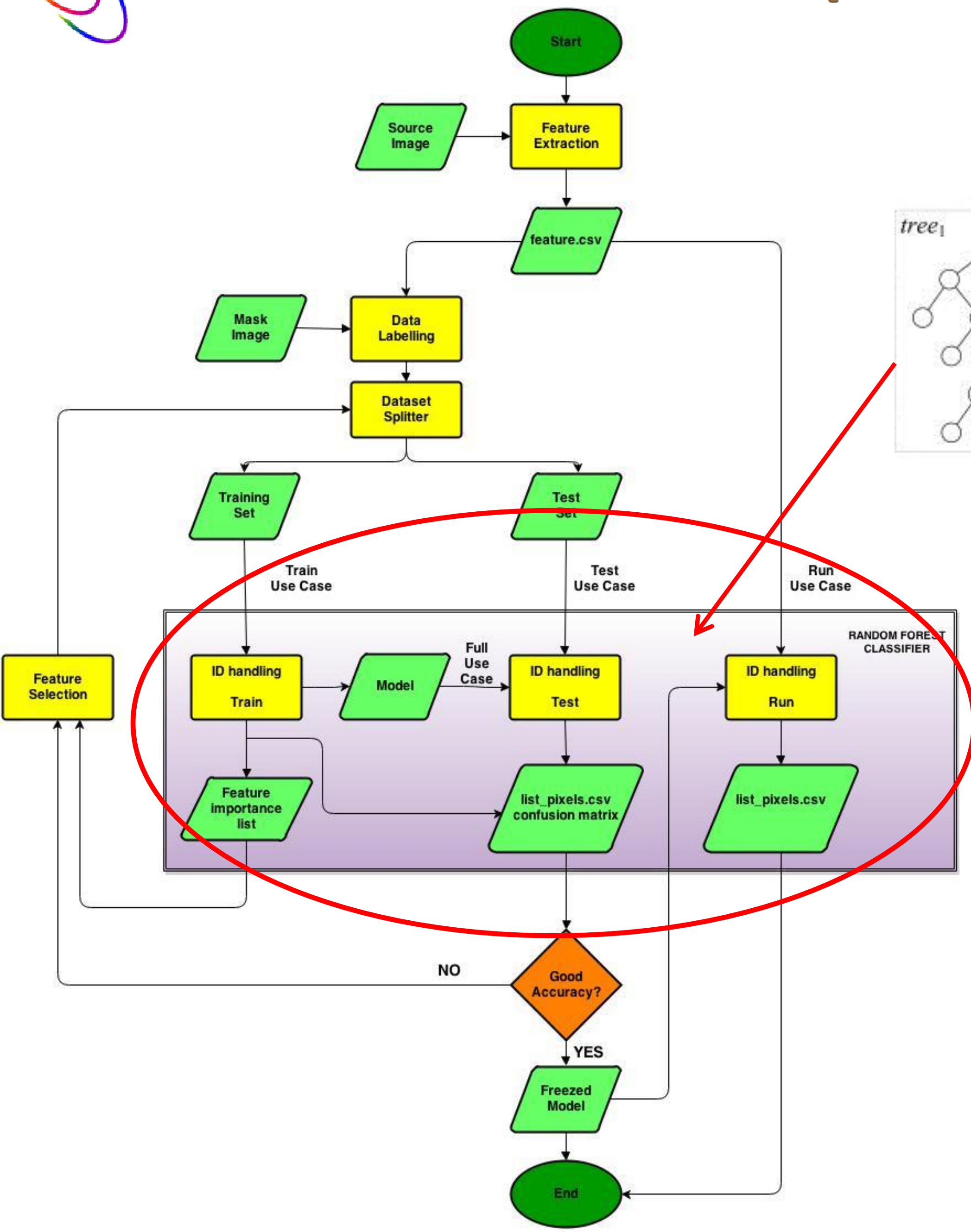
Statistical Features

For each pixel, the following features are calculated in a centered window:

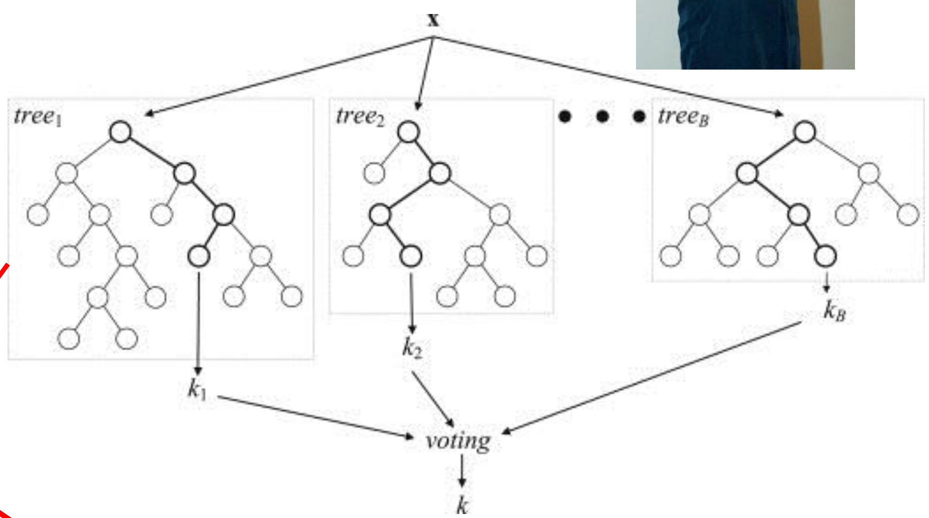
- gradients (horizontal, vertical, main diagonal, secondary diagonal)
- Mean, standard deviation, skewness, kurtosis, entropy, range

Moreover, the pixel value is considered as a Statistical Feature too

FilExSeC (Classification)



Leo Breiman



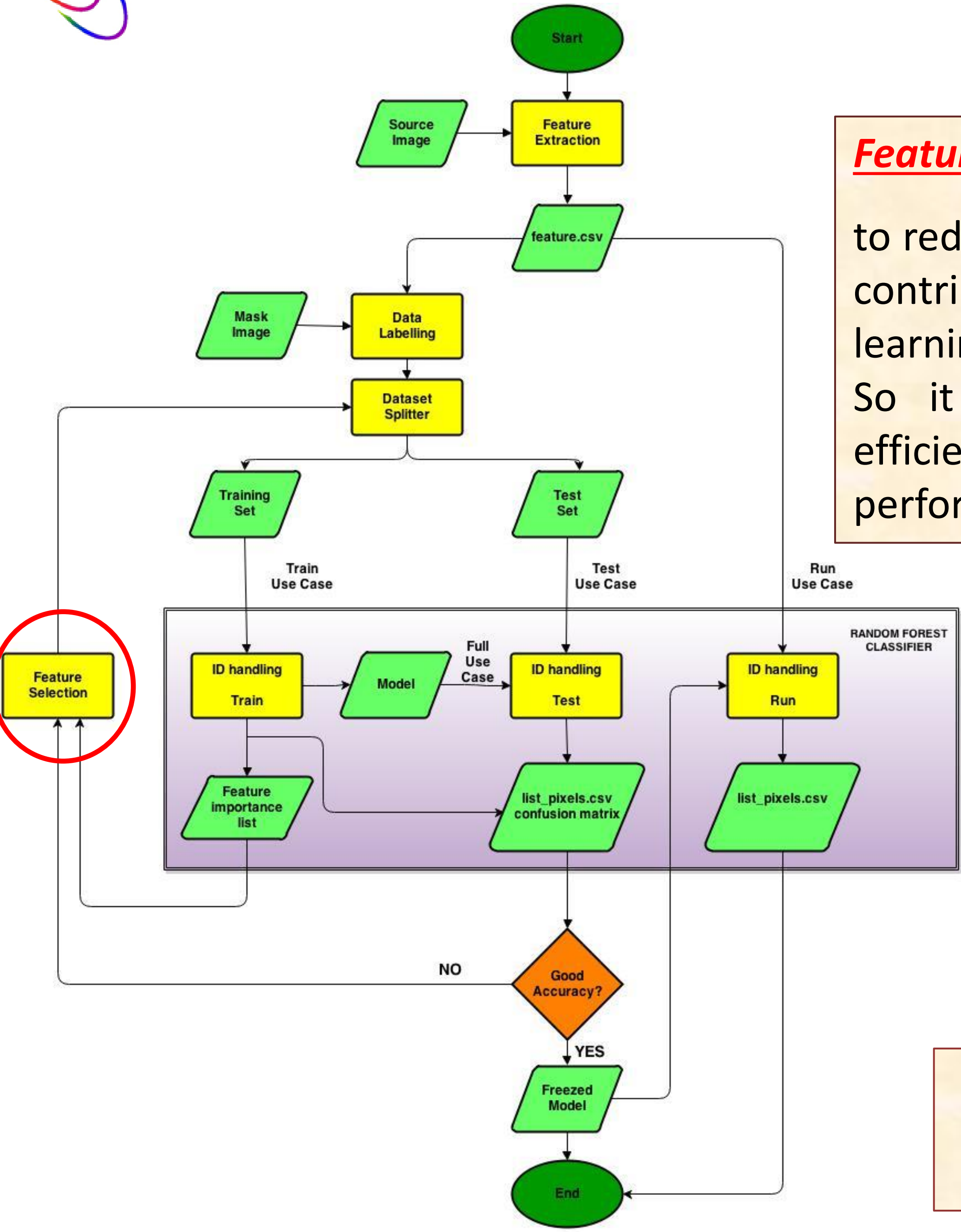
Random Forest Classifier (2001)

The classifier prefigures 3 use cases:

- **Train**: the classifier is trained to discriminate between filamentary and background pixels;
- **Test**: the classifier is evaluated by a blind test dataset;
- **Run**: the trained and validated classifier is used on new real images.

FilExSeC (Feature Selection)

Mark A. Hall



Feature Selection (Backward Elimination 1999)

to reduce the parameter space, by weighting the contribution carried by each feature to the learning capability of the classifier.
So it is possible to improve the execution efficiency of the model, without affecting its performances.

At the end of this phase, a subset of features having higher weight (defined as *importance*) in recognizing filament pixels is considered.

This subset is then used to definitely train and test the classifier with new training and testing subsets.

Tests revealed that Haralick features are useless

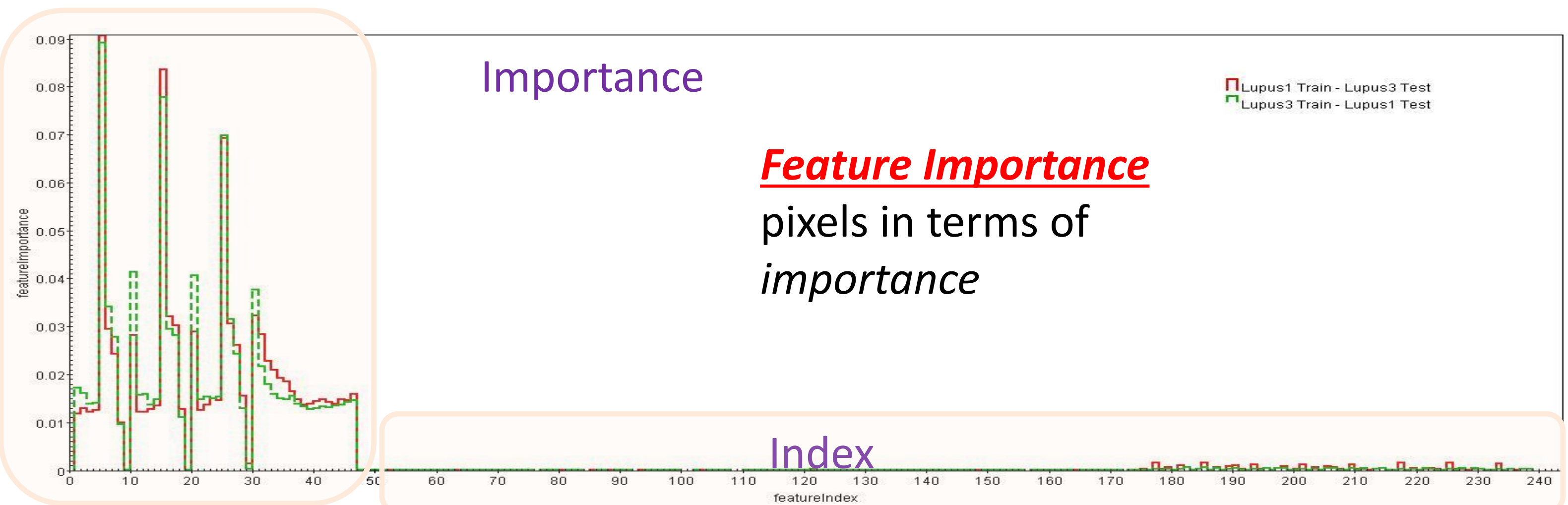
FilExSeC (products)

***Main Output*:**
A CSV table

Pixel identification id, row, column	Statistical features	Haar-like features	Haralick features	Pixel Value	Output 1 filament 0 background

Confirmed Filament Pixels	CFP	pixels correctly recognized as belonging to a filament
Undetected Filament Pixels	UFP	pixels of filaments classified as background
New Filament Pixels	NFP	background pixels classified as belonging to a filament
Confirmed Background Pixels	CBP	background pixels correctly recognized

***Confusion Matrix*:**
Pixels grouped in 4 categories



Feature Importance
pixels in terms of *importance*

Statistical + Haar-like

Haralick

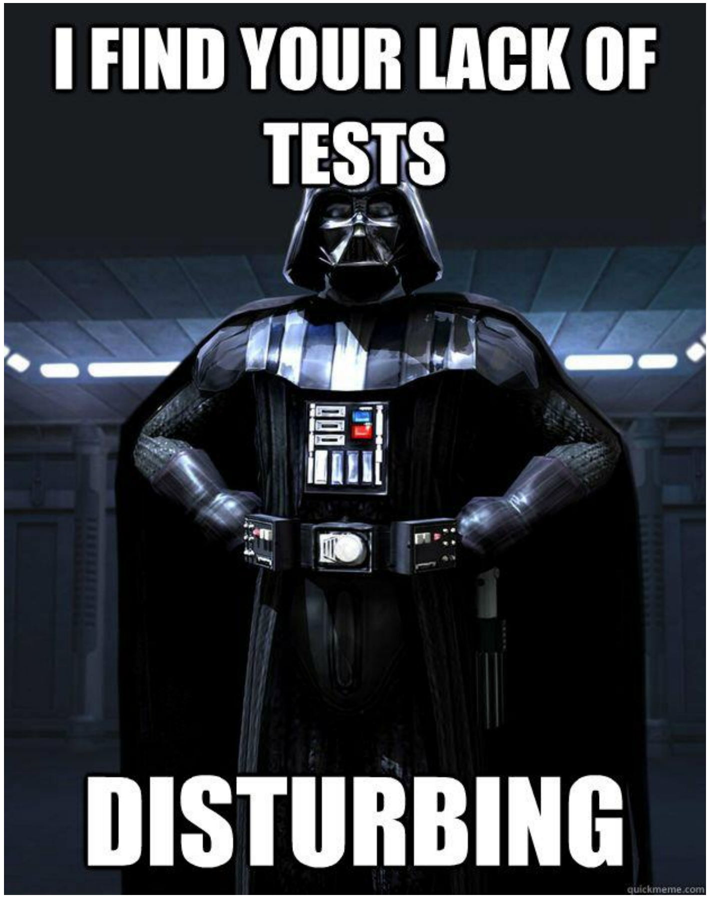
FilExSeC – Testing

In order to evaluate FilExSeC, several tests have been performed. They have been useful to:

- evaluate the robustness of the method;
- evaluate the performances of the algorithm;
- optimize the parameters settings.

Test have been performed by:

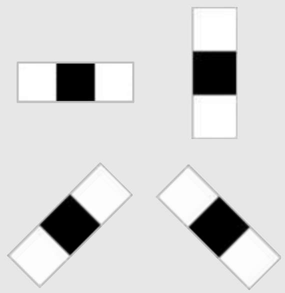
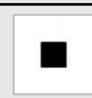
- Varying the number and type of features:
  - Haralick yes/no
  - Pixel value yes/no
  - Stats yes/no
  - Different Haar-like templates
  - Different settings of Haralick, Haar-like and Statistical windows
- Varying the configuration of the classifier
  - Different number of trees of the Random Forest (1000 or 10000)
- Using different datasets:
  - Lupus region;
  - Hi-GAL strips;
- Using different ratios for Train and Test set



FilExSeC – Tests Summary

EXP	DATA				FEATURE EXTRACTION				RANDOM FOREST				FILAMENT STATISTICS			BACKGROUND STATISTICS		
	TRAIN		TEST		HARALICK	HAAR-LIKE		STAT	SETUP				PURITY	COMPLETENESS	DICE	PURITY	COMPLETENESS	DICE
ID	name	%	name	%	windows	template	windows	windows	trees	max depth	min split	min leaf	%	%	%	%	%	%
T1a	lupusIII	15	lupusIII	85	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	84	62	72	99	99	99
T1b	lupusIII	30	lupusIII	70	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	88	65	75	99	100	99
T1c	lupusIII	60	lupusIII	40	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	92	70	79	99	100	99
T2	lupusIII	100	lupusI	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	61	67	64	99	98	99
T3a	lupusI	100	lupusIII	100	5x5, 7x7, 9x9		9x9 to 24x24		1000	none	2	1	48	29	36	98	99	98
T3b	lupusIII	100	lupusI	100	5x5, 7x7, 9x9		9x9 to 24x24		1000	none	2	1	52	24	33	97	99	98
T4	lupusIII	100	lupusI	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	10000	none	2	1	60	67	64	99	98	99
T5	lupusI	100	lupusIII	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	77	50	61	98	100	99
T6a	Hi-GAL 1+2	100	Hi-GAL 3+4	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	83,24	21,93	34,72	98,32	99,90	99,11
T6b	Hi-GAL 1+3	100	Hi-GAL 2+4	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	82,85	20,71	33,14	98,11	99,90	99,00
T6c	Hi-GAL 2+4	100	Hi-GAL 1+3	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	59,00	38,53	46,62	98,18	99,20	98,69
T6d	Hi-GAL 3+4	100	Hi-GAL 1+2	100	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	59,24	36,87	45,45	97,98	99,18	98,58
T7a	lupusI	100	lupusIII	100			9x9 to 24x24	3x3, 5x5, 7x7, 9x9	1000	none	2	1	81,32	54,66	65,37	98,54	99,59	99,07
T7b	lupusI	100	lupusIII	100			2x2 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	82,23	55,64	66,37	98,58	99,61	99,09
T7c	lupusI	100	lupusIII	100			2x2 to 24x24	3x3, 5x5, 7x7, 9x9	1000	none	2	1	82,73	56,94	67,45	98,62	99,61	99,11
T8a	Hi-GAL	60	Hi-GAL	40	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	88	38	53	98	100	99
T8b	Hi-GAL	80	Hi-GAL	20	5x5, 7x7, 9x9		9x9 to 24x24	5x5, 7x7, 9x9	1000	none	2	1	90	39	55	98	100	99
T9a	Hi-GAL	80	Hi-GAL	20			2x2 to 24x24	3x3, 5x5, 7x7, 9x9	1000	none	2	1	88,41	43,00	57,86	98,49	99,85	99,16
T9b	Hi-GAL	80	Hi-GAL	20			1x1 to 24x24	3x3, 5x5, 7x7, 9x9	1000	none	2	1	88,52	43,62	58,45	98,49	99,85	99,16
T9c	Hi-GAL	80	Hi-GAL	20			2x2 to 24x24	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	89,59	44,98	59,89	98,52	99,86	99,19
T9d	Hi-GAL	80	Hi-GAL	20			1x1 to 24x24	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	89,39	44,88	54,80	98,54	99,86	99,19
T10a	Hi-GAL 1+2	100	Hi-GAL 3+4	100		all	*	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	84,88	43,75	57,74	98,78	99,83	99,30
T10b	Hi-GAL 1+3	100	Hi-GAL 2+4	100		all	*	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	85,12	42,63	56,81	98,63	99,82	99,22
T10c	Hi-GAL 2+4	100	Hi-GAL 1+3	100		all	*	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	73,67	51,82	60,85	98,57	99,45	99,01
T10d	Hi-GAL 3+4	100	Hi-GAL 1+2	100		all	*	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	72,62	49,64	58,97	98,39	99,40	98,89
T11	lupusI	100	lupusIII	100		all	*	1x1, 3x3, 5x5, 7x7, 9x9	1000	none	2	1	85,41	63,88	73,09	98,83	99,65	99,24
T12a	HI-GAL	100	HI-GAL289	100		all	*	1x1, 3x3, 5x5, 7x7, 9x10	1000	none	2	1	71,08	54,14	61,47	96,80	98,44	97,61
T12b	HI-GAL	100	HI-GAL300	100		all	*	1x1, 3x3, 5x5, 7x7, 9x10	1000	none	2	1	52,06	71,14	60,12	98,06	95,7	96,86
T12c	HI-GAL	100	HI-GAL310	100		all	*	1x1, 3x3, 5x5, 7x7, 9x10	1000	none	2	1	52,21	72,70	60,78	98,05	95,38	96,69

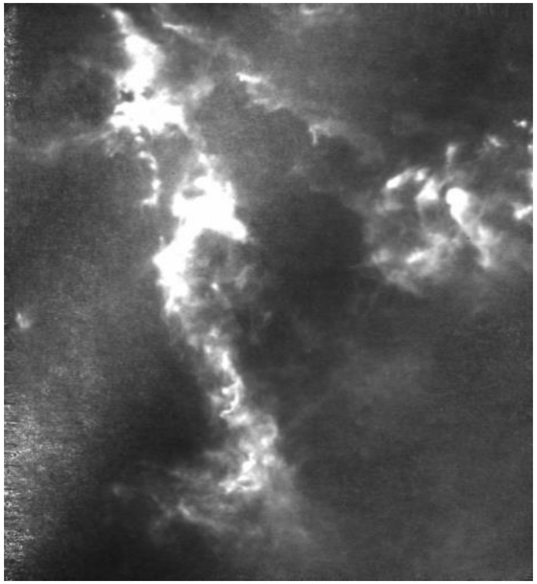
FilExSeC – Tests parameters

Type	Parameters			Features	Max. Num.
	Name	Template	Values		
Haar-like	Black rectangle dimensions		from 2x2 to 24x24	Difference between “black” and “white” rectangles	158
	Black rectangle dimensions		from 2x4 to 12x24		
	Number of black rectangles		1-2		
	Black rectangle dimensions		from 1 to 24		
Haralick	$\lvert\vec{d}\rvert$ =1,2,3,4 directions = 0°, 45°, 90°, 135° windows = 5x5 – 7x7 – 9x9			Contrast, Energy, Entropy, Correlation	192
Statistical	windows = 3x3 – 5x5 – 7x7 – 9x9			Gradients (vertical, horizontal, diagonal) Mean – Std Skewness - Kurtosis Entropy – Range	41
	windows = 1x1			Pixel Value	

FilExSeC – Tests dataset (1)

Lupus region
Sub-region I & III

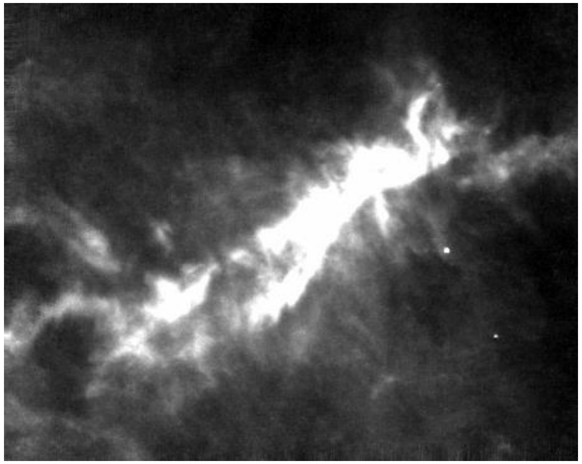
Lupus I



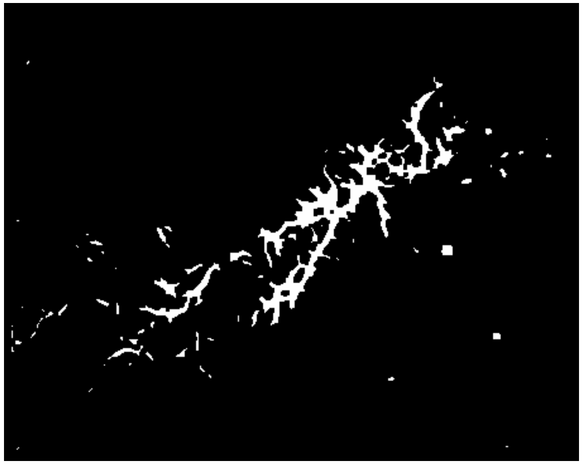
Lupus I Mask

Tests
T1 – T2 – T3 – T4 –
T5 – T7 – T11

Lupus III

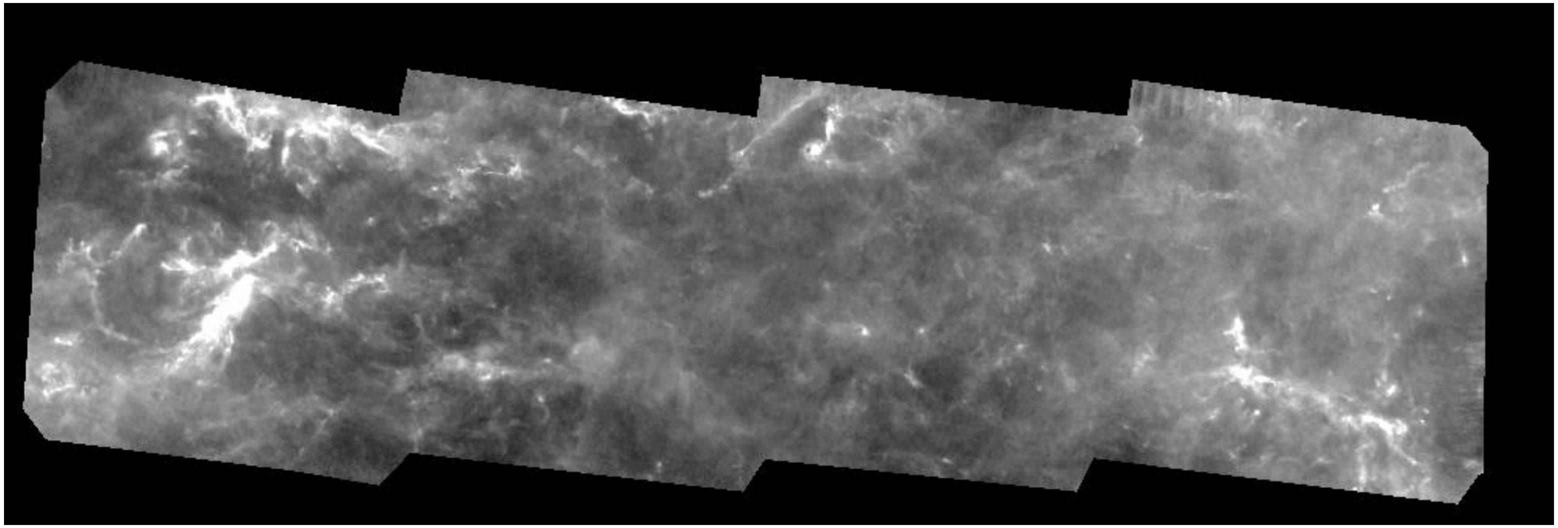


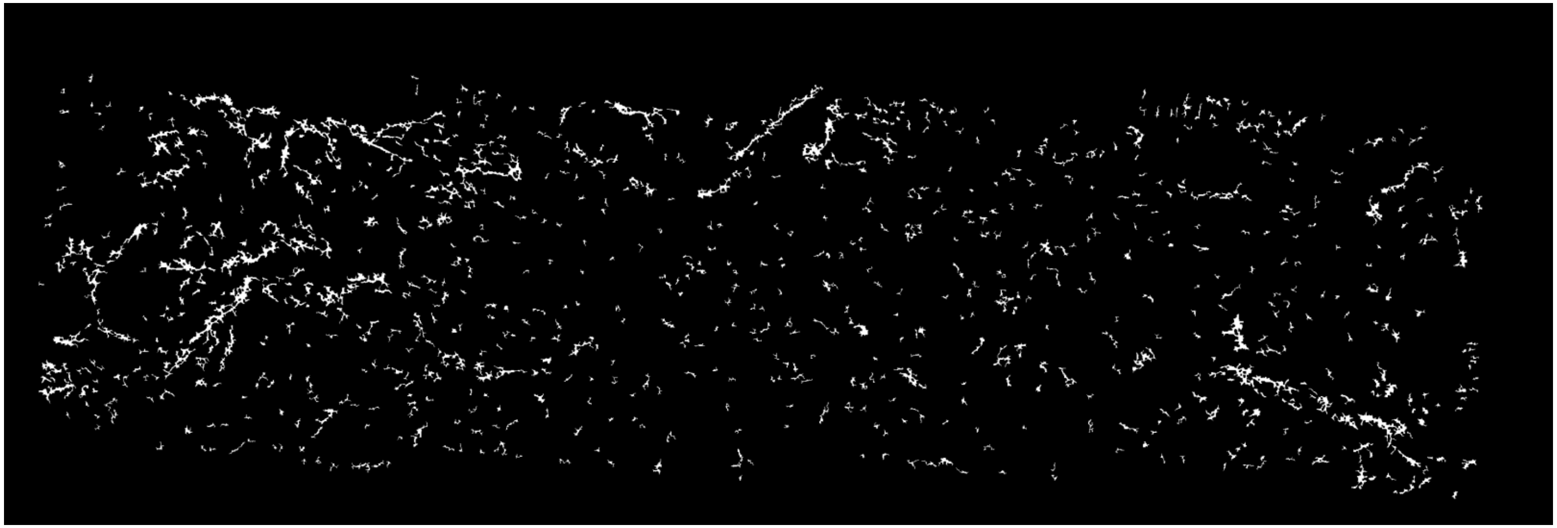
Lupus III Mask



FilExSeC – Tests dataset (2)

Column density map (250μ) of Hi-GAL 224-300 deg region: 2973x1001 pixels





Tests

T6 – T8 – T9 – T10

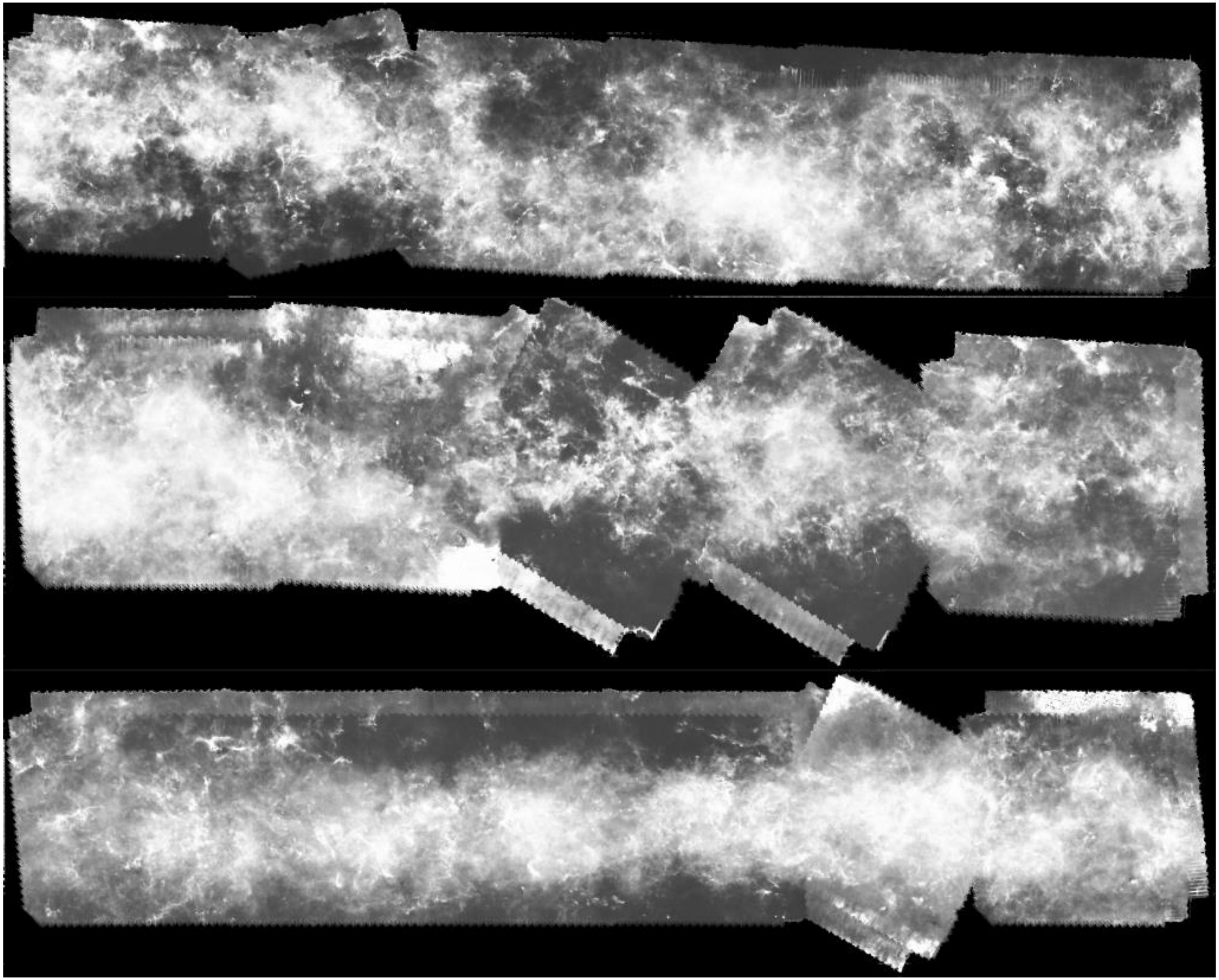
T12 (as Train set)

FilExSeC – Tests dataset (3)

Column density map (250μ) of Hi-GAL 289-320 deg region

Tests

T12 (as Test sets)



Hi-GAL 289-300

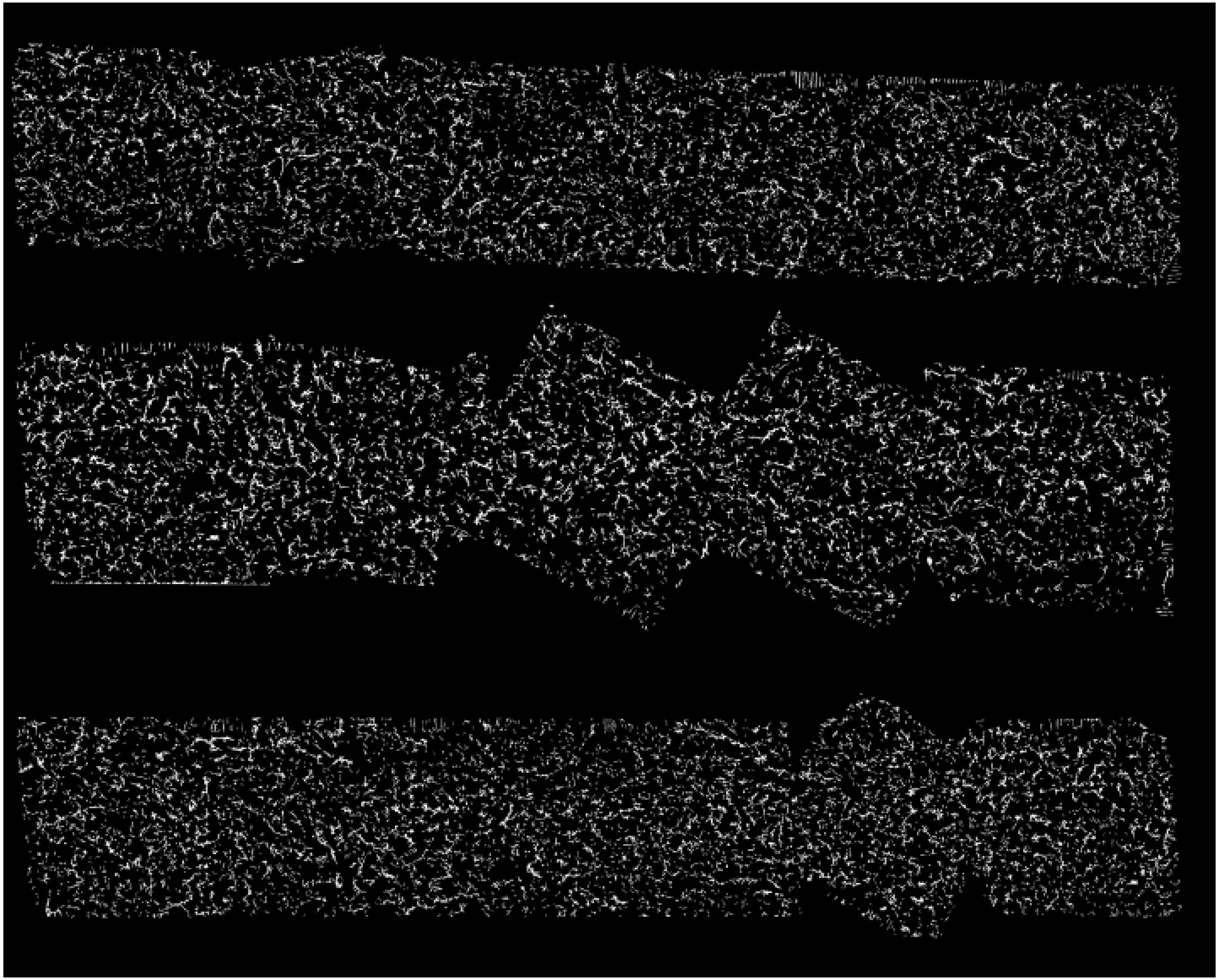
Hi-GAL 300-310

Hi-GAL 310-320

FilExSeC – Tests dataset (3)

Column density map (250µ) of Hi-GAL 289-320 deg region

Tests
T12 (as Test sets)



Hi-GAL 289-300 Mask

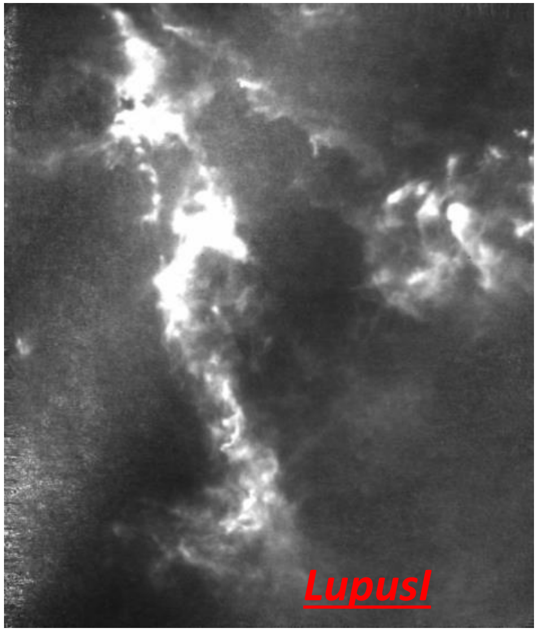
Hi-GAL 300-310 Mask

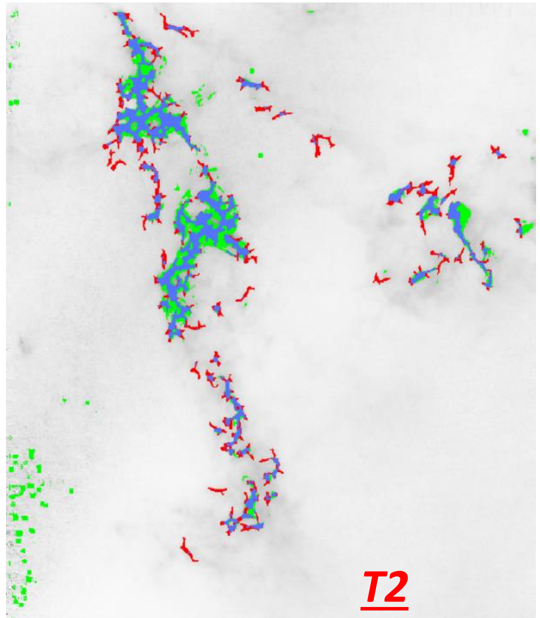
Hi-GAL 310-320 Mask

FilExSeC – Tests with Lupus (1)

Test T2: Lupus III Training – Lupus I Test
Test T5: Lupus I Training – Lupus III Test

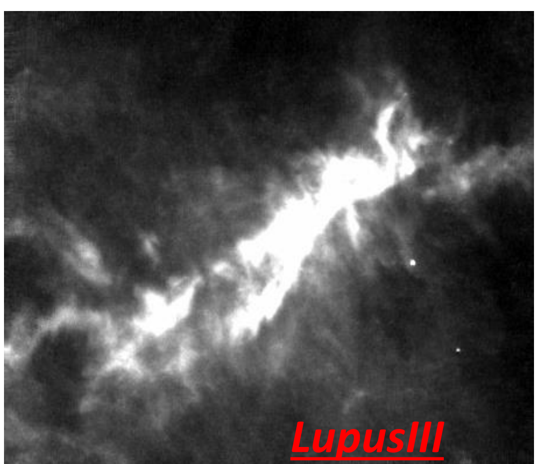
Same Features
Train/Test dataset inverted

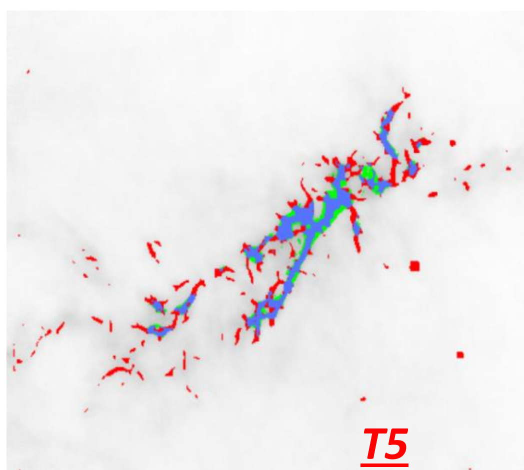




T2/T5	TEST	FIL	BG
purity	T2	61%	99%
	T5	77%	98%
completeness	T2	67%	98%
	T5	50%	100%
DICE	T2	64%	99%
	T5	61%	99%

Performances depend on the train image quality.
The results confirm the capability to extend the filament pixel regions





CFP CBP
NFP UFP

$$purity = \frac{CFP}{CFP + NFP}$$

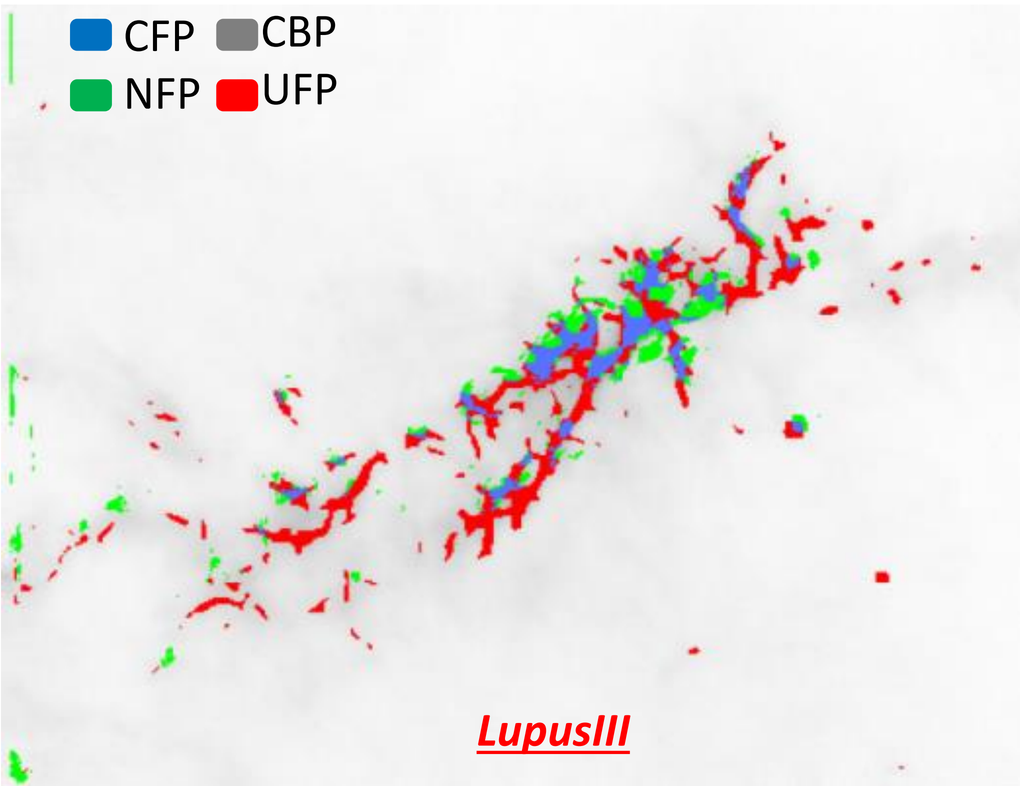
$$completeness = \frac{CFP}{CFP + UFP}$$

$$DICE = \frac{2 * purity * complet.}{purity + complet.}$$

FilExSeC – Tests with Lupus (2)

Test T3a: Lupus I Training – Lupus III Test

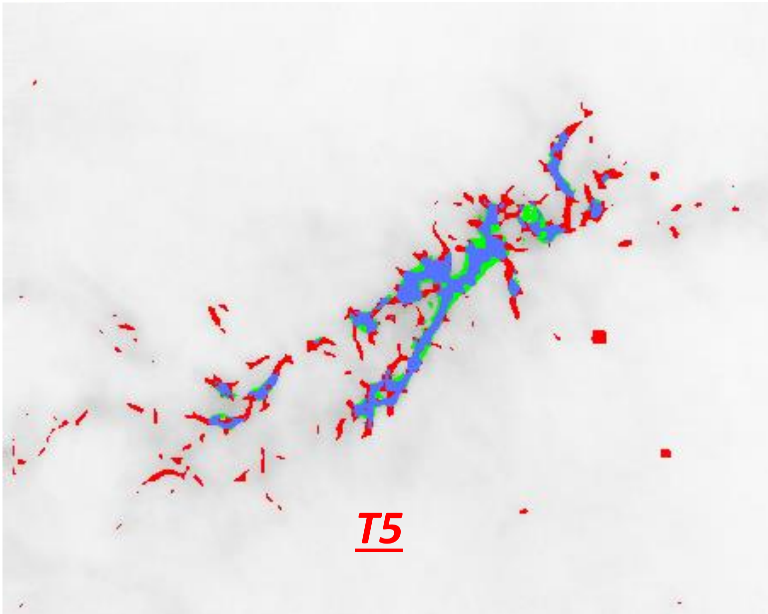
Statistical Features Excluded



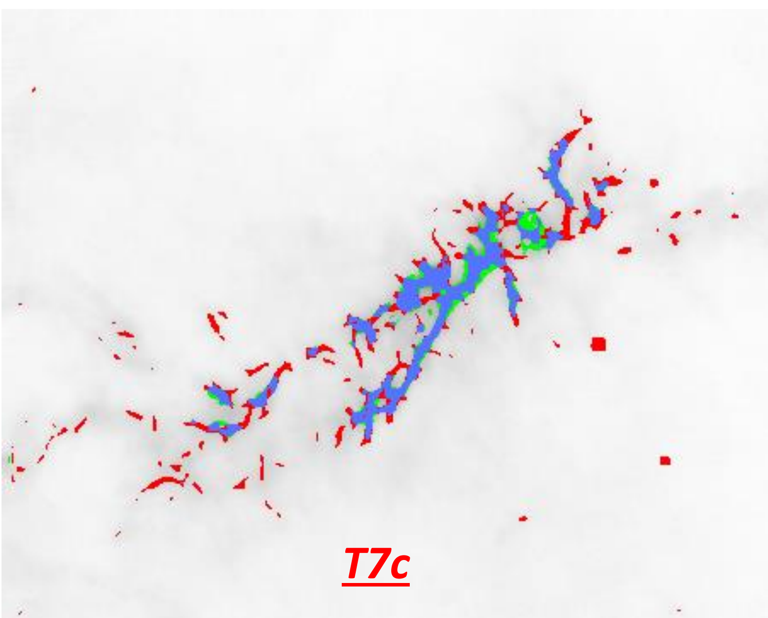
T3a	FIL	BG
Purity	48%	98%
Completeness	29%	99%
DICE	36%	98%

Without statistical features the performances strongly decrease.
Haar-like + Haralick give a lower contribute

FilExSeC – Tests with Lupus (3)

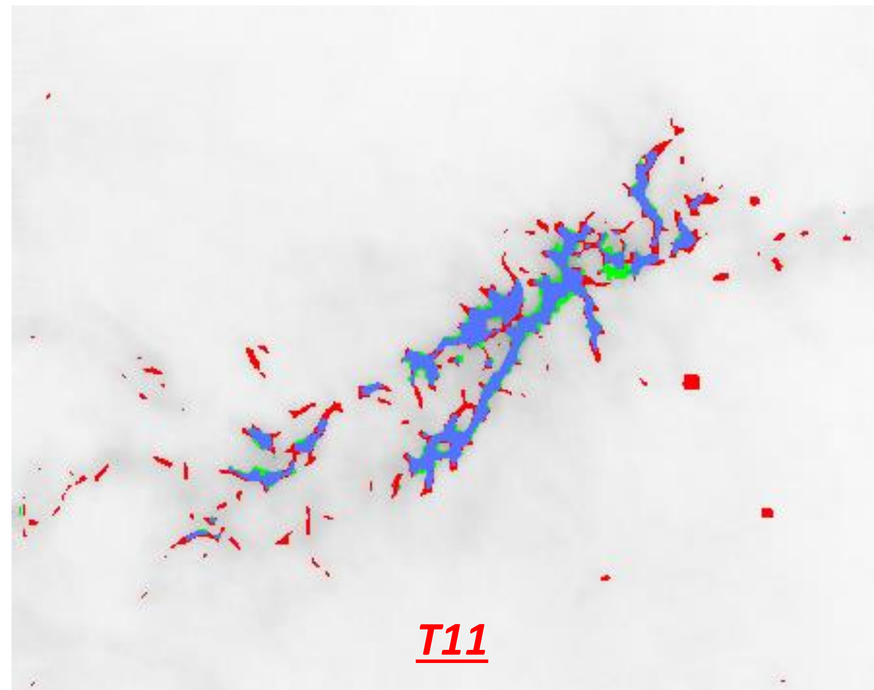


T5



T7c

FEATURE EXTRACTION					
ID	HARALICK	HAAR-LIKE tmpl	HAAR-LIKE windows	STATISTICAL windows	PIXEL
T5	Y	□■	9x9 to 24x24	5x5, 7x7, 9x9	N
T7c	N	□■	2x2 to 24x24	3x3, 5x5, 7x7, 9x9	N
T11	N	all	2x2 to 24x24	3x3, 5x5, 7x7, 9x9	Y

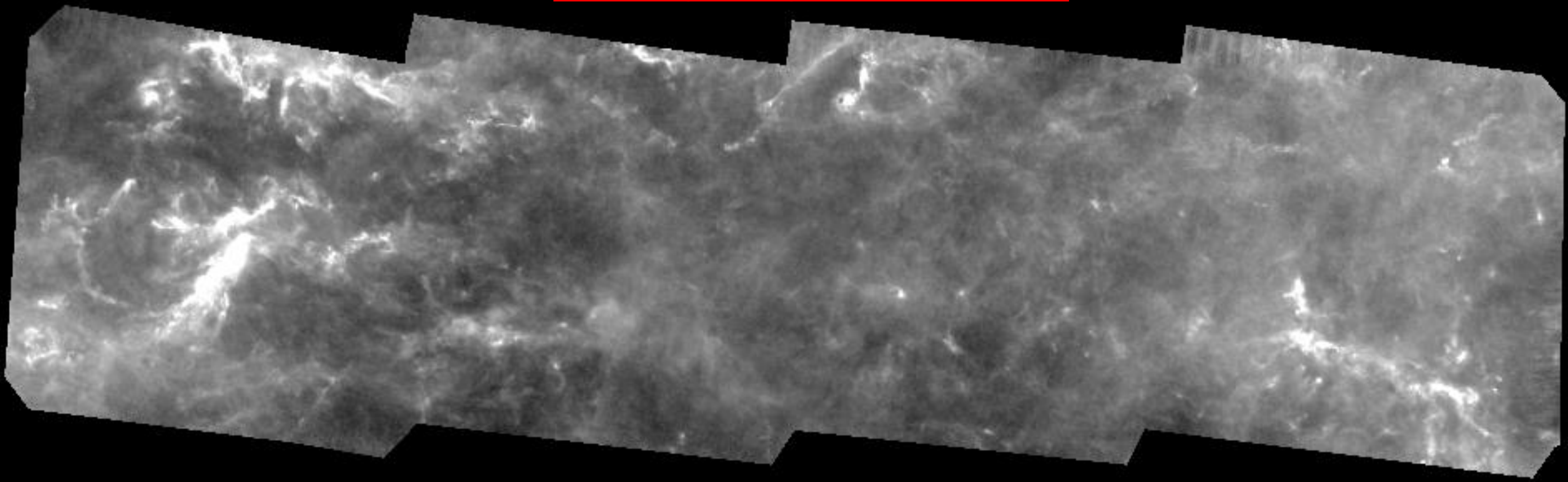


T11

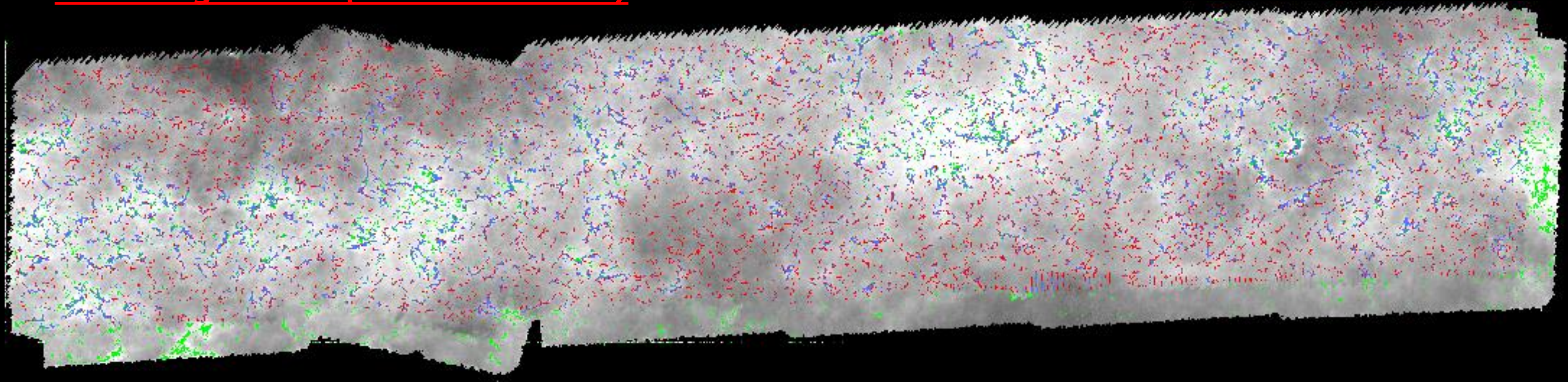
	T5		T7c		T11	
	FIL (%)	BG (%)	FIL (%)	BG (%)	FIL (%)	BG (%)
Purity	77.38	98.41	82.73	98.62	85,41	98,83
Complet.	50.42	99.52	56.94	99.61	63,88	99,65
DICE	61.06	98.96	67.45	99.11	73,09	99,24

FilExSeC – Latest tests

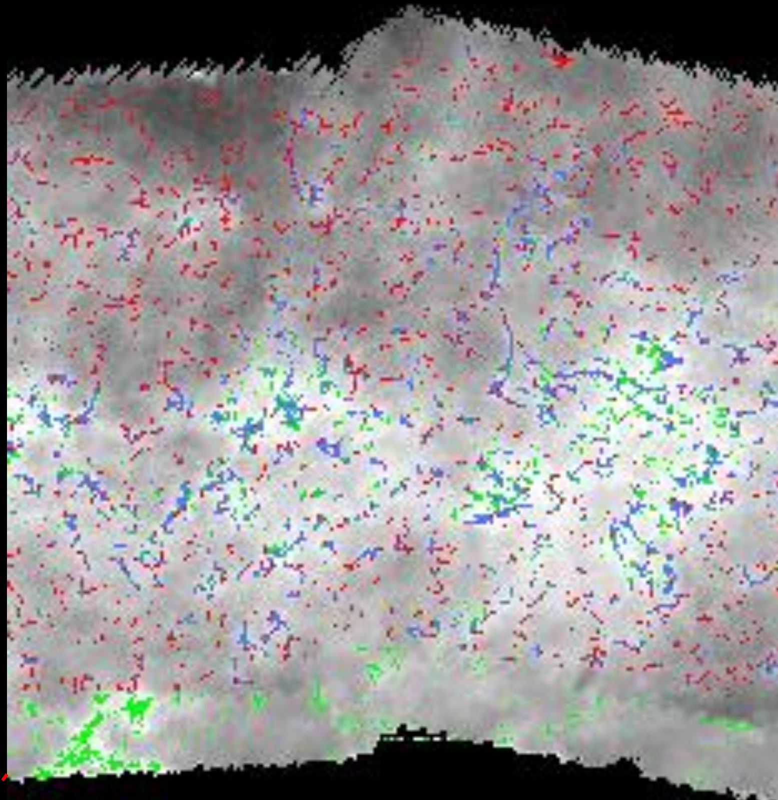
Training image (Hi-GAL 224)



Test image: T12a (Hi-GAL 289-300)

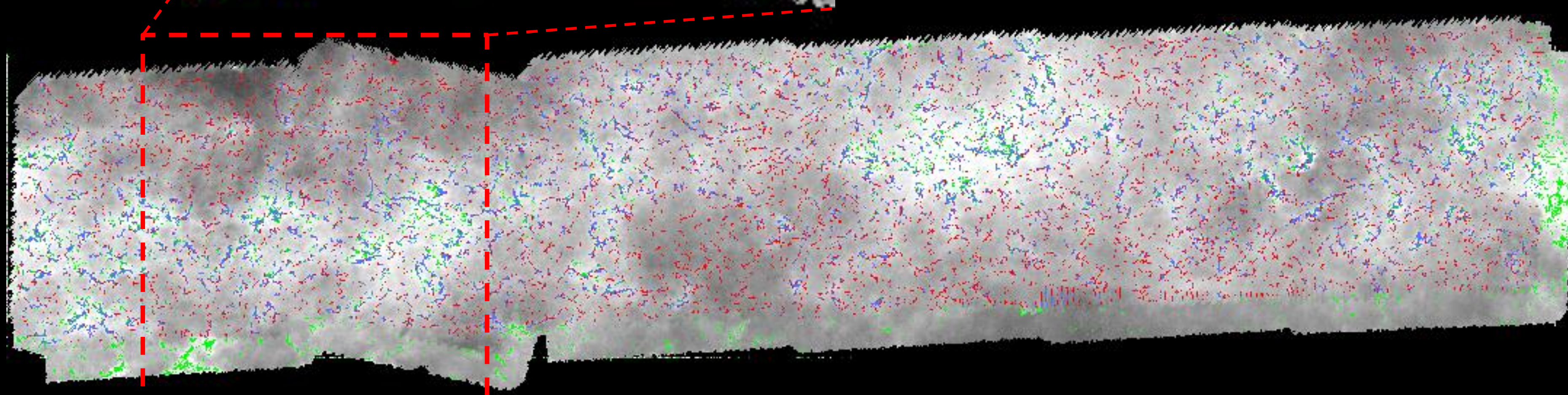


FilExSeC – Latest tests



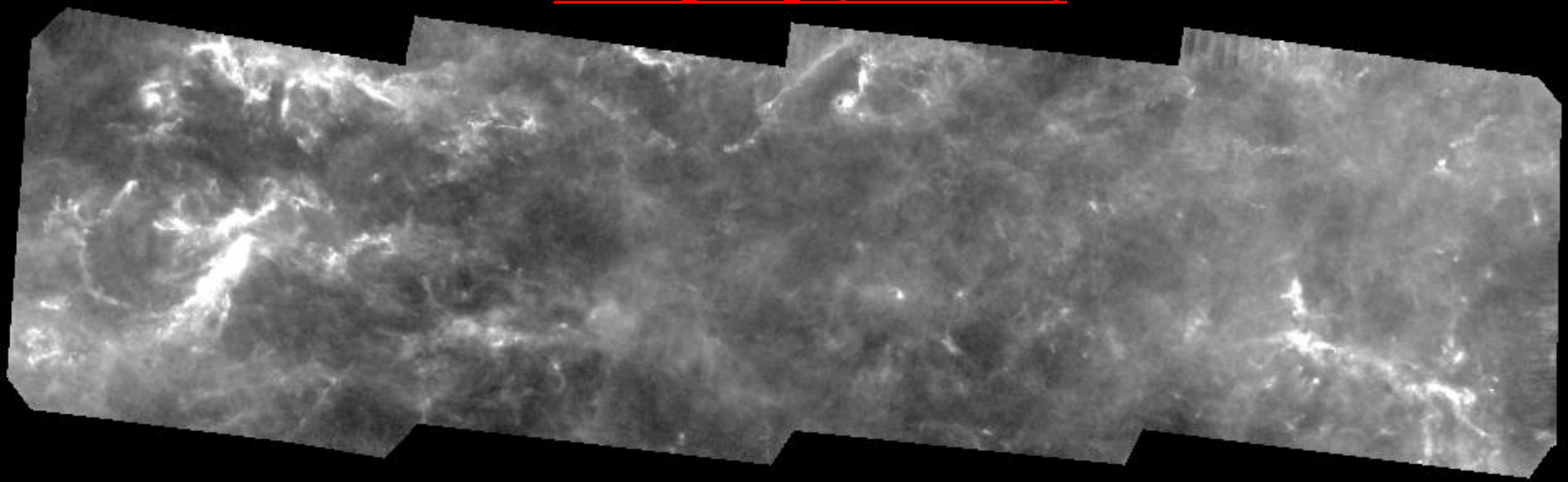
Test image: T12a (Hi-GAL 289-300)

T12a	FIL	BG
Purity	71.08%	96.80%
Completeness	54.14%	98.44%
DICE	61.47%	97.61%

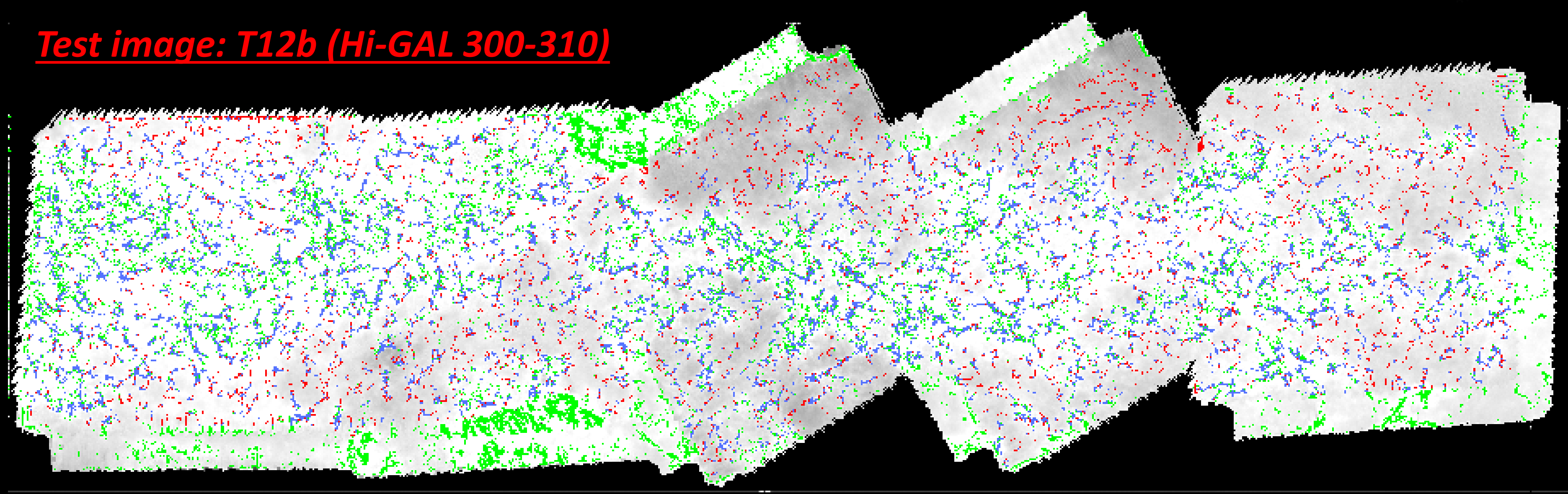


FilExSeC – Latest tests

Training image (Hi-GAL 224)



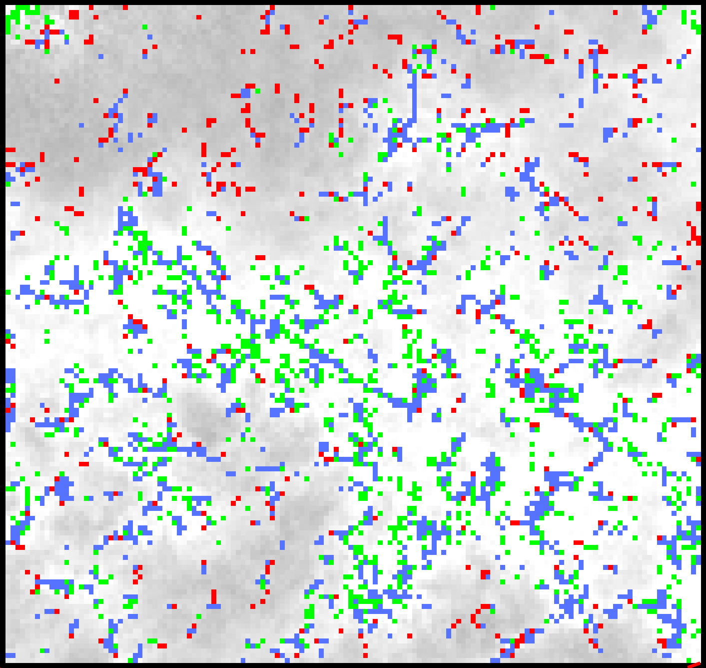
Test image: T12b (Hi-GAL 300-310)

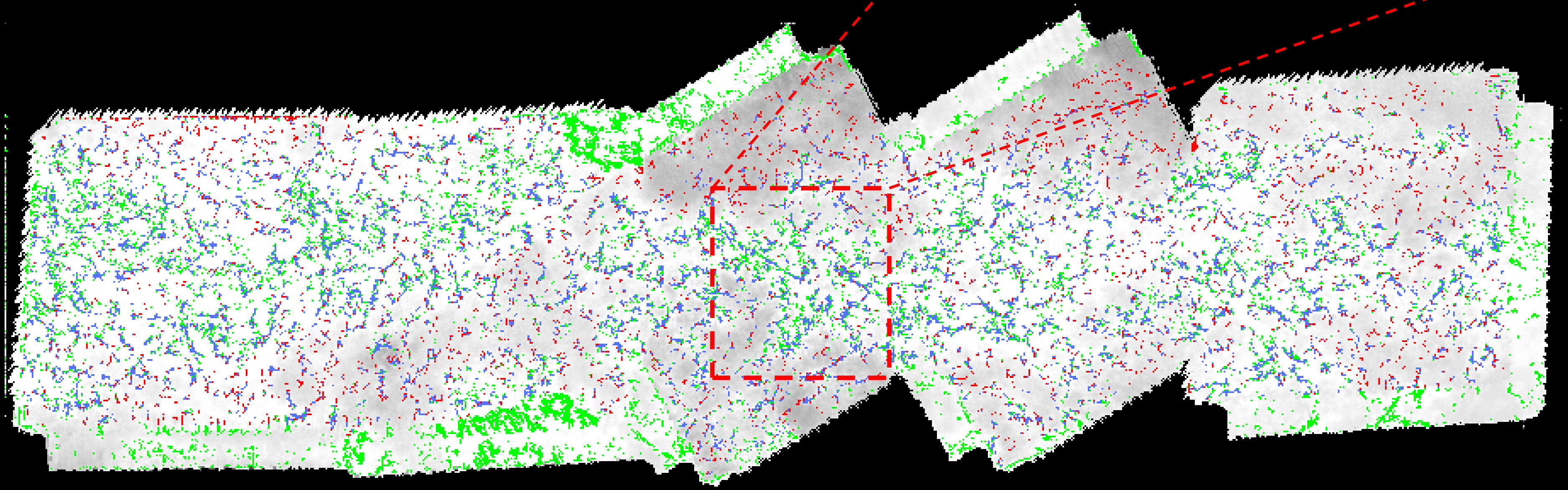


FilExSeC – Latest tests

Test image: T12b (Hi-GAL 300-310)

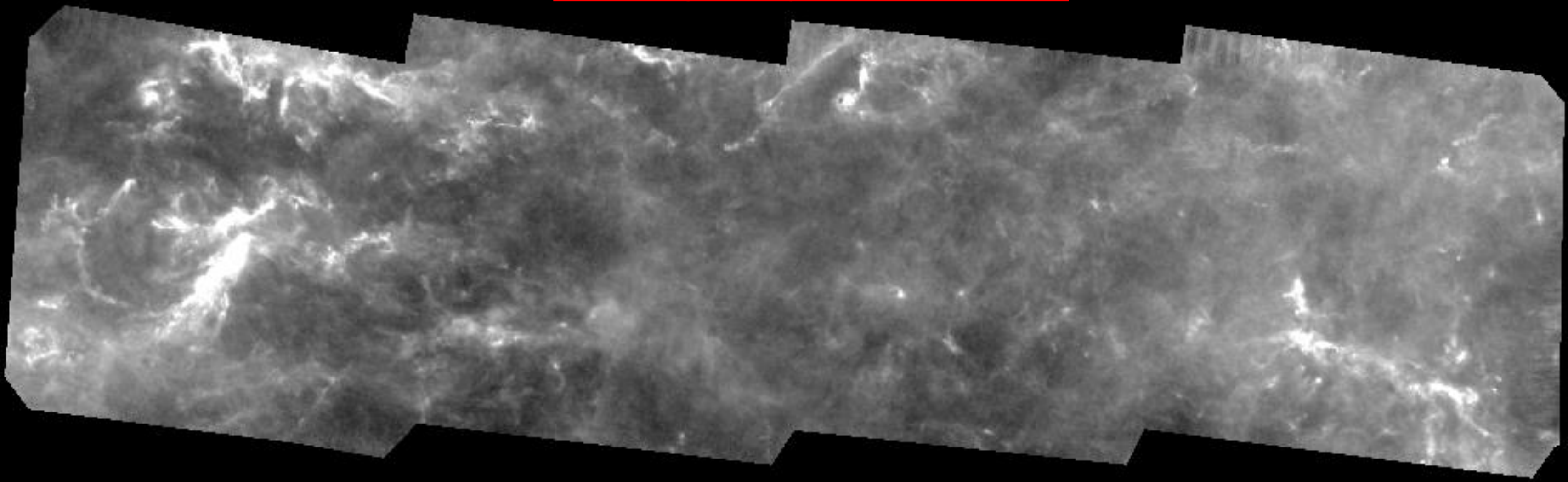
T12b	FIL	BG
Purity	52.06%	98.06%
Completeness	71.14%	95.70%
DICE	60.12%	96.86%



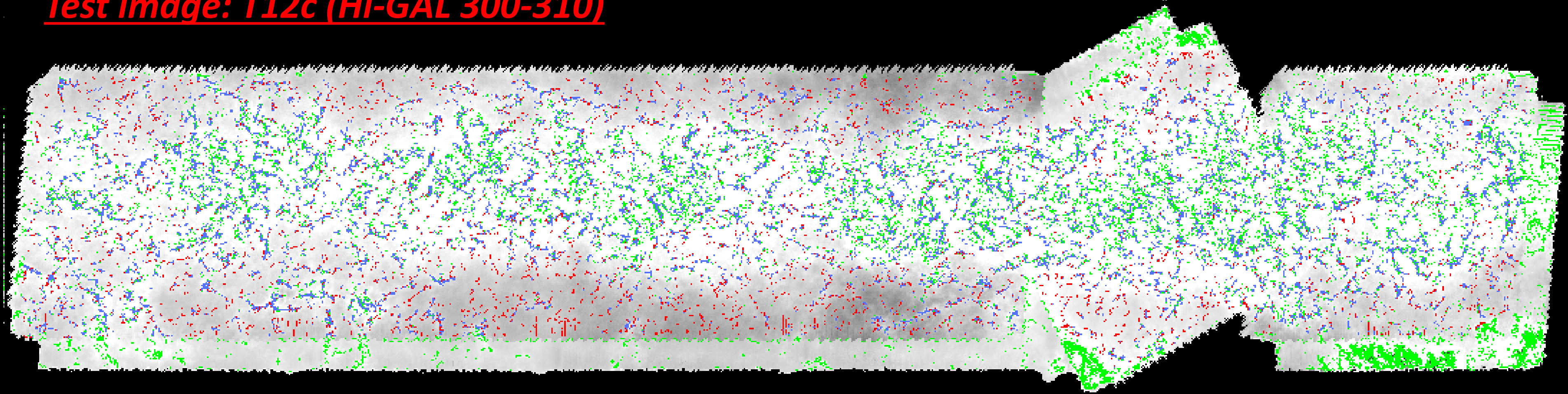


FilExSeC – Latest tests

Training image (Hi-GAL 224)

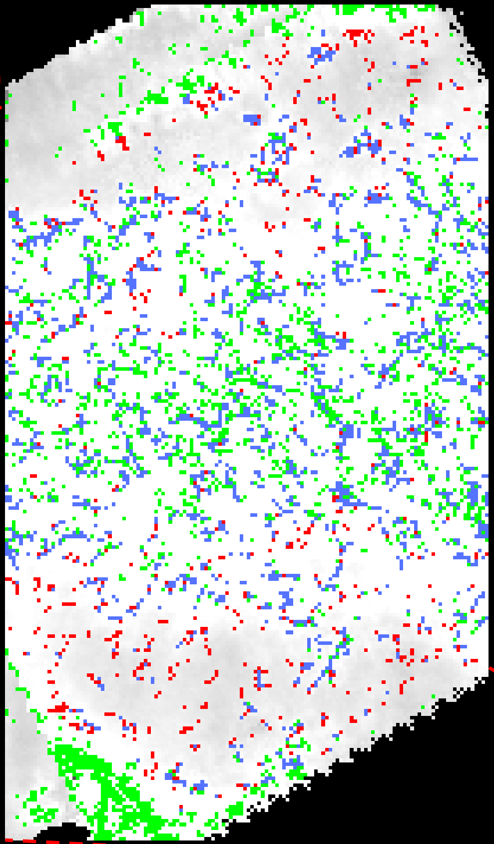


Test image: T12c (Hi-GAL 300-310)

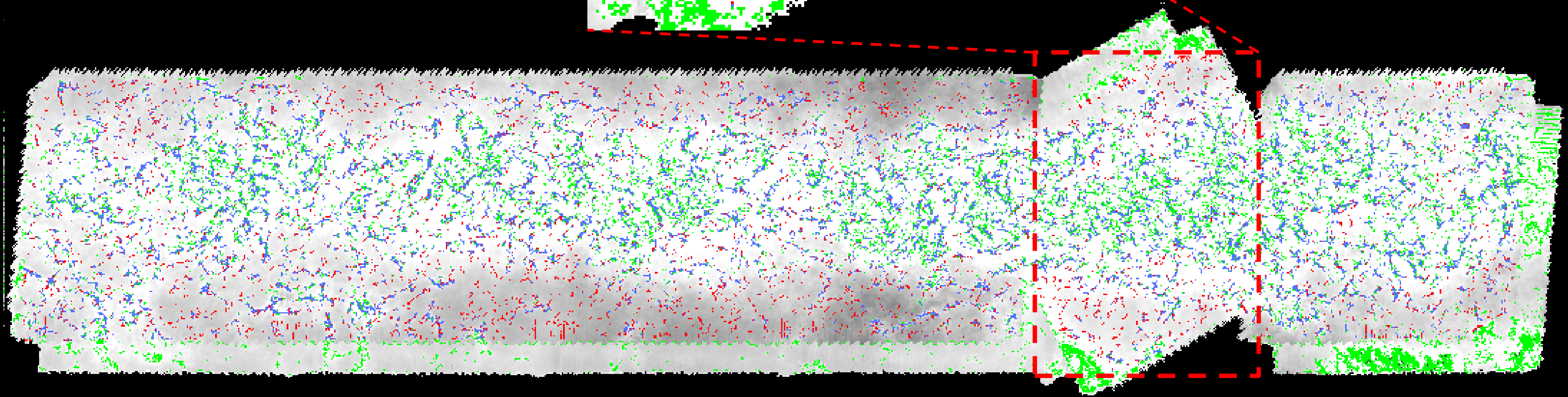


FilExSeC – Latest tests

Test image: T12c (Hi-GAL 310-320)



T12c	FIL	BG
Purity	52.21%	98.05%
Completeness	72.70%	95.38%
DICE	60.78%	96.69%



FilExSeC – Tests Results

Robustness of the method

- The features mostly contributing to classification are always the same, regardless of the datasets;
- Comparing tests on the same region by slightly varying training set, the results show extremely low % of background detection difference (always < 0.5%) and of filament (max 6.75%)

High Performances

- Global efficiency grows up by choosing a more realistic training set;
- High capability to identify additional filament pixels with respect to the traditional method. However, a certain number of very little filamentary structures was not recognized, by confusing them as background

Best Configuration

- Haralick features are useless
- All Haar-like templates needed with rectangles size up to 24x24
- Windows for Statistical features 3x3 – 5x5 – 7x7 – 9x9
- Pixel value is very important
- A low number of trees is sufficient for the best classification (1000 trees)

Next step is to verify, together with IAPS astronomers, the results obtained by FilExSeC from a physical point of view, by calculating, for example, the variation of filament's mass function by adding the new filament pixels found

FilExSeC – Alternative edge detectors

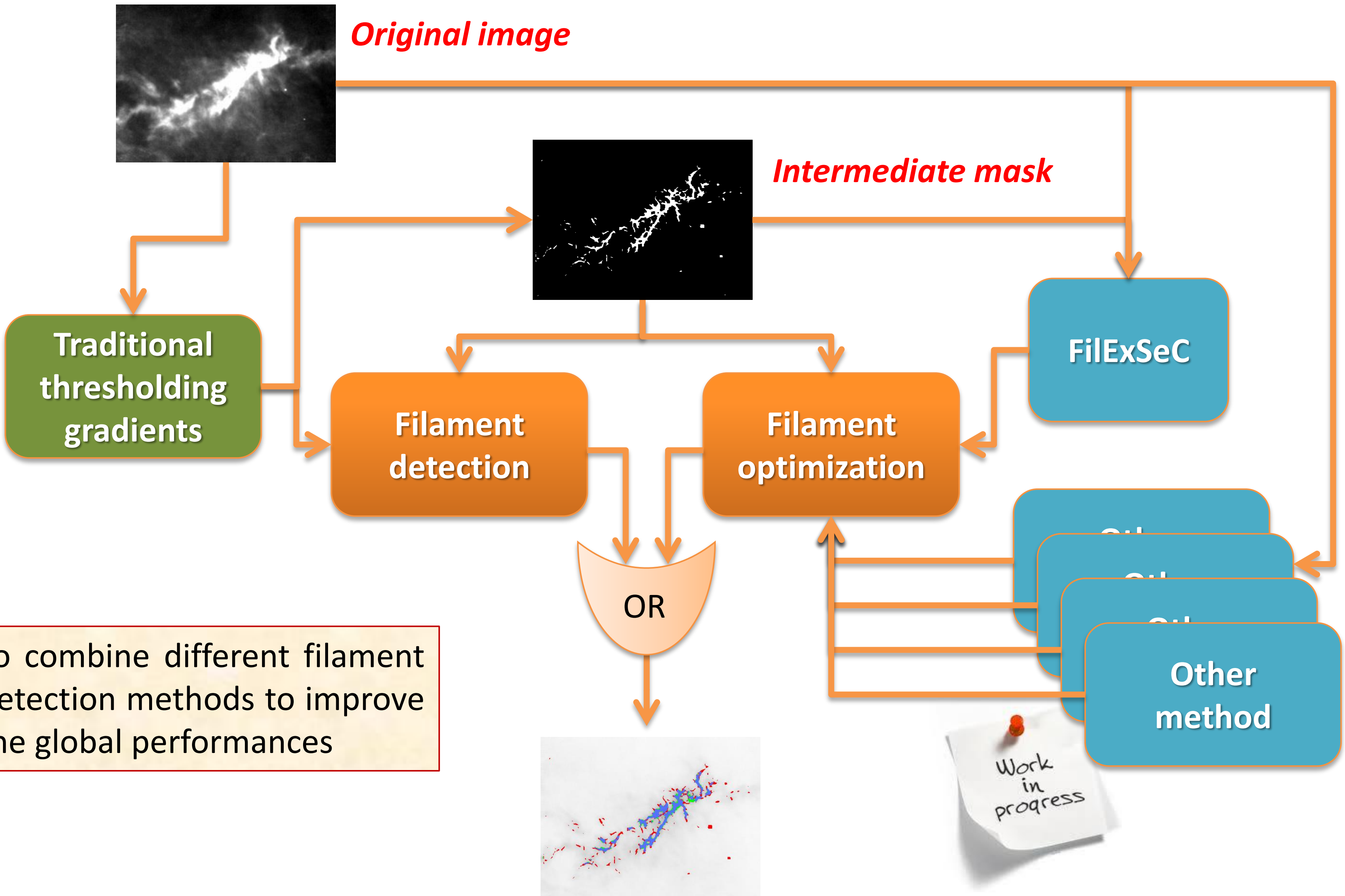
The main limitation of FilExSec is that it works with masks obtained by traditional methods. This causes a bias on our performances.

It is necessary to find a method directly working on original images without any priors

At this time, we are under investigation on new edge detectors:

- **Boosted Edge Learning** (*Dollar et al. 2006*)
- **gPb (global Probability of boundary)** (*Arbelaez et al. 2011*)
- **Beam-curve Pyramid based edge detector** (*Alpert et al. 2010*)
- **Curvelets and Wavelets** (*Starck et al. 2002* and *Mallat 1998*)
- **Fuzzy Logic Edge Detectors** (*Becerikli et al. 2005*)
- **Canny and Sobel filters enhancement** (*Canny 1986* and *Sobel 2014*)

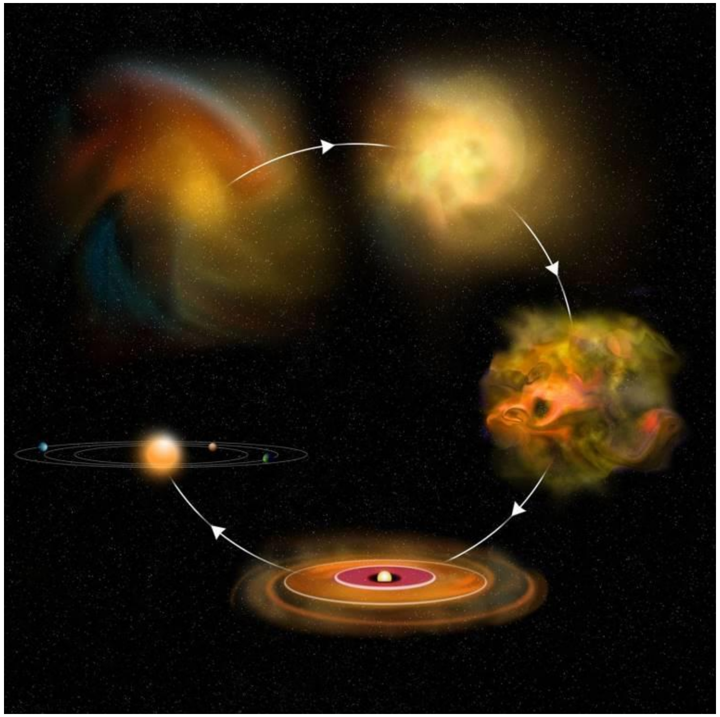
FilExSeC (connections)



To combine different filament detection methods to improve the global performances

OACN – WP5 Activities

WP2 – Task 4
Star Forming Sources
Evolutionary Classification



Design

OACN – WP5 Activities

WP2 – Task 4
Star Forming Sources
Evolutionary Classification

WORK IN PROGRESS!!!

Design

WP2 – task 4 Evolutionary classification

An evolutionary classification tool for ViaLactea, will catalogue “clumps” in terms of the evolutionary stage and mass regime of the ongoing star formation. There are two components that need to be developed at the foundation of the classification tool:

1. an evolutionary classification toolbox
2. a set of star-forming clumps in known stages of evolution to be used as a training/test-set for machine-learning algorithms… ...and adopt some kind of evolutionary scheme

Data-mining approaches to source classification

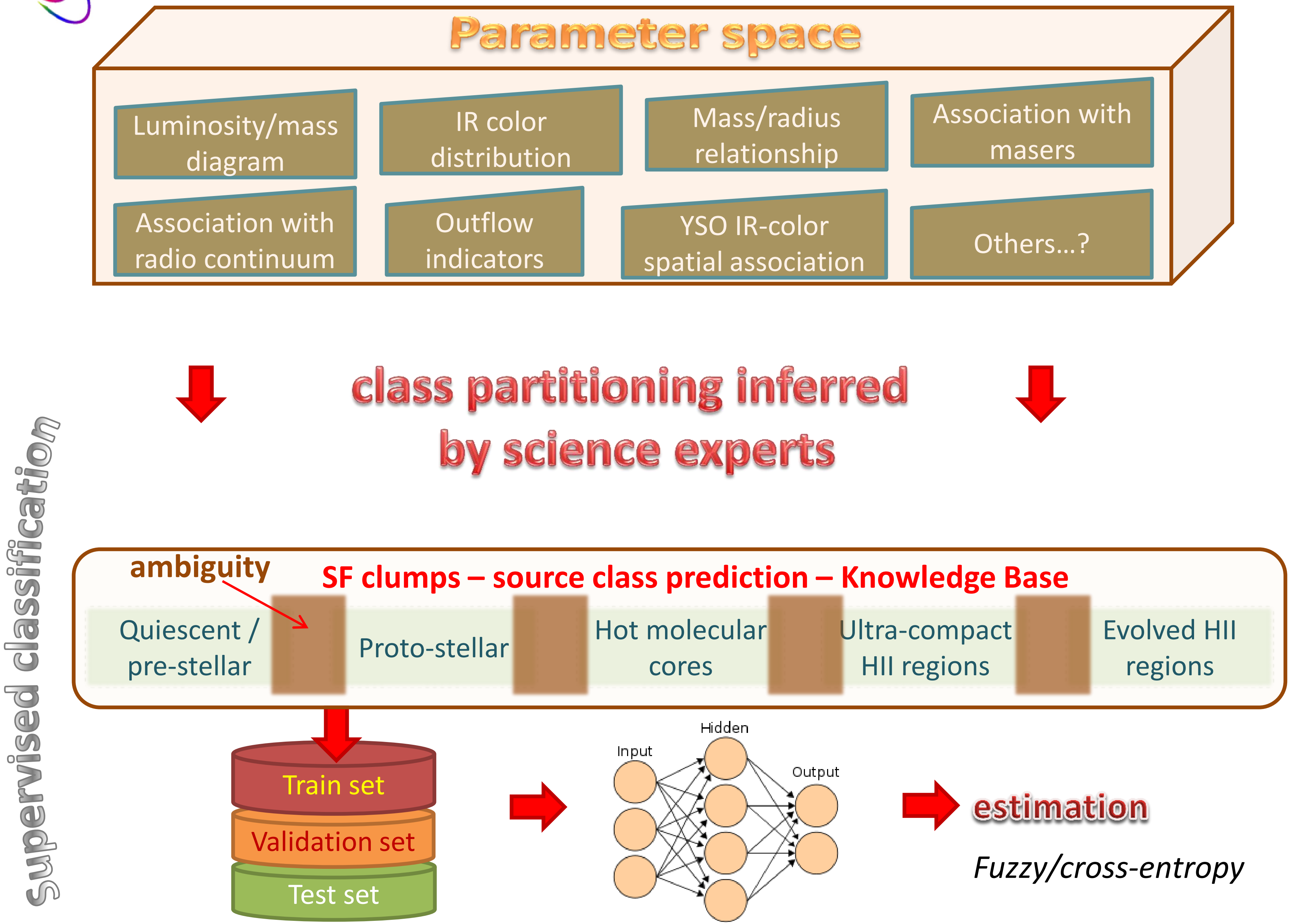
Weak Gated Classification

We know nothing about the sources evolutionary stage;
Identify over-densities in the given parameter space (e.g., built on the evolutionary toolbox, plus any other available evidence);
Data are then grouped into clusters: groups of data entries sharing common but *a priori* unknown correlations among parameter space features.

Supervised Classification

For a subsample of points, its category/class is well known;
Need order of 10^3 objects to be used as a training set;
Balanced population of classes in the training set.

SF-SEC (1st approach)



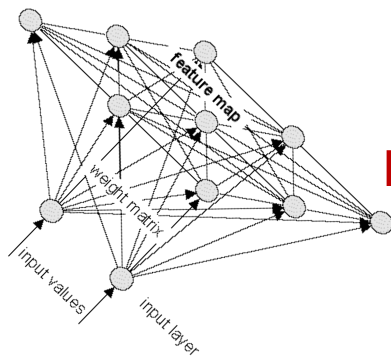
SF-SEC (2nd approach)

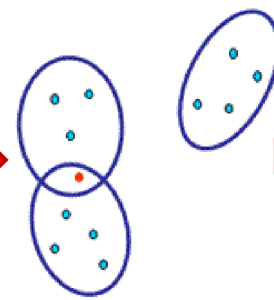
Weak gated classification

Parameter space

- Luminosity/mass diagram
- IR color distribution
- Mass/radius relationship
- Association with masers
- Association with radio continuum
- Outflow indicators
- YSO IR-color spatial association
- Others…?

unsupervised clustering



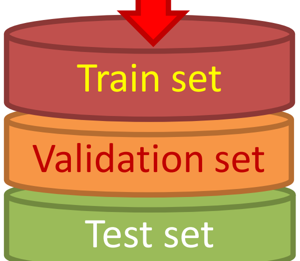


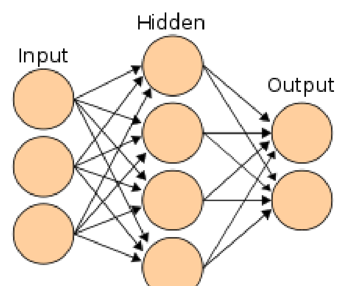
validated by science experts

ambiguity

SF clumps – source class prediction – Knowledge Base

Quiescent / pre-stellar	Proto-stellar	Hot molecular cores	Ultra-compact HII regions	Evolved HII regions





estimation

Fuzzy/cross-entropy


THANKS!
VIALACTEA