

Simulations of High-mass Star Formation in the Milky Way

JOÃO FERNANDO CIOTTA NEVES

*Centre for Astrophysics Research
Science & Technology Research Institute
Department of Physics, Astronomy and Mathematics
· University of Hertfordshire ·*

Thesis submitted for the Degree of Doctor of Philosophy in the
University of Hertfordshire

· September 2009 ·

Abstract

Massive star formation takes place in the dense cores of molecular clouds where the stars may be obscured at optical wavelengths. An excellent signpost of a massive young stellar object is the presence of an ultra-compact HII region (UCH II), which is a dense photo-ionised cocoon of gas surrounding the newly formed star. The aim of this project is to develop an assembly of numerical tools, *caravela*, that can simulate realistic data streams representing high-mass star forming regions in our Galaxy. The synthetic output consists in images and photometric point source catalogues, in the IRAS and Herschel wavebands. In an era when large observational surveys are increasingly important, this tool can produce simulated infrared point-source catalogues of high-mass star forming regions on a Galactic scale. The approach used is to construct a synthetic Galaxy of star-forming regions represented by SED templates. The star-forming regions are distributed randomly along a four spiral arm morphology, although a wide range of geometries can be used including rings and different numbers of spiral arms. The *caravela* code then *observes* the synthetic Galaxy to produce simulated images and point source catalogues with appropriate sensitivity and angular resolution. *caravela* was first used to model the simulated Galaxy by constraining the synthetic output to observations made by IRAS. This numerical tool will allow the user to infer physical properties of the Galactic population of high-mass star forming regions from such observations. Second, the selected model was again observed with *caravela* in Herschel mode. These are therefore predictive results for the future Herschel observations. A model with 4.0×10^4 compact proto-stars embedded in larger grey-body envelopes (with $T = 40$ K and linear size scale $l_{\text{HII}} = 5.0 \times 10^6$ AU) is the best-fit model to the IRAS observational data set studied. We found a level of contamination from low- and intermediate-mass objects of $\sim 90\%$. The modelled data set resulting from the Herschel simulation resulted in the detection of approximately twice as many Herschel objects than IRAS, which is consistent, in a limited way, with the real observed companion clump fraction (CCF) of 0.90 ± 0.07 (Thompson *et al.*, 2006) means that *on average* there were observed 2 sources per one IRAS source. Our *caravela* and the real observed CCF are therefore consistent. *caravela* was coupled with an independent diffuse emission model (Paladini *et al.*, 2007) and the resulting analysis is presented as an interesting seed for the future.

All appearances to the contrary, the only watchmaker in nature is the blind forces of physics, albeit deployed in a very special way. A true watchmaker has foresight: he designs his cogs and springs, and plans their interconnections, with a future purpose in his mind's eye. Natural selection, the blind, unconscious, automatic process which Darwin discovered, and which we now know is the explanation for the existence and apparently purposeful form of all life, has no purpose in mind. It has no mind and no mind's eye. It does not plan for the future. It has no vision, no foresight, no sight at all. If it can be said to play the role of watchmaker in nature, it is the blind watchmaker.

Richard Dawkins

This is not NAM. There are rules.

Walter Sobchak

Contents

Acknowledgments	ix
1 Context: Observational Surveys for Star Formation	1
1.1 Introduction	1
1.2 Star formation	3
1.3 High-mass star forming regions	4
1.3.1 High-mass star formation may not be a scaled-up version of low-mass star formation	13
1.3.2 The spiral structure of our Galaxy	14
1.4 Hi-GAL	16
1.5 Motivation for the project	18
1.6 Outline of this dissertation	18
2 Method: The caravela code	21
2.1 Introduction	21
2.2 Programming language and design	23
2.2.1 Random variations	25
2.3 Bird’s-eye view of caravela	27

2.4	Running <i>caravela</i>	33
2.5	The parameter file	34
2.6	A theoretical catalogue of High-mass star forming regions	39
2.6.1	The objects	39
2.6.2	3D distribution	45
2.6.3	Galactic structure model	51
2.7	Simulated observation	52
2.8	Photometry	61
2.9	Simulated observations using <i>caravela</i>	61
2.9.1	The input parameters	62
2.9.2	Simulated images	65
2.10	Summary	67
3	High-mass Star Formation as Seen by IRAS	69
3.1	Introduction	69
3.2	A model for the UCH II regions in the Galaxy	70
3.2.1	Input Parameters	71
3.2.2	Wood and Churchwell (1989) <i>total</i> IRAS sources	78
3.2.3	Infrared colours	79
3.2.4	Best-fit model for UCH II regions	92
3.2.5	2D Kolmogorov-Smirnov Test	93
3.2.6	Low- and Intermediate-mass contamination	97
3.2.7	Galactic distribution	100

3.3	Summary	102
4	Predictions for Herschel	104
4.1	Introduction	104
4.2	A model for the UCH II regions in the Galaxy	105
4.3	Herschel synthetic images	107
4.4	Results	109
4.4.1	Companion Clump Fraction interpretation	114
4.5	Summary	118
5	Conclusions and Future Work	120
5.1	Summary	120
5.2	Main results and conclusions	122
5.3	The future	125
5.3.1	caravela + the Diffuse Emission Model	127
5.3.2	caravela combined with the diffuse emission	127
5.3.3	The diffuse emission model	128
5.3.4	PACS and SPIRE co-added images	129
5.3.5	PACS and SPIRE three-colour images	131
A	The Hershel Hi-GAL Milky Way: the first <i>observational</i> science highlights	134
B	Hi-GAL: the Herschel infrared Galactic Plane Survey	135
C	caravela input parameter file	136

List of Figures

1.1	The formation of a group of high-mass stars.	12
1.2	The Russeil (2003) model for the Galaxy was adopted for <i>caravela</i> . It is a four arm model, based on a equiangular or logarithmic spiral. All the relevant equations and physical parameters of this model are described in detail in the next chapter.	15
2.1	Class diagram for <i>caravela</i> . Each arrow represents the dependence of the derived class from the base class in C++.	26
2.2	<i>caravela</i> 's main stages.	27
2.3	The figure illustrates a logarithmic spiral distribution of sources (<i>left</i> panel), and only those sources who lie inside the solid angle defined by the user (<i>right</i> panel).	48
2.4	PACS and SPIRE simulated PSFs.	59
2.5	Output from the SPIRE simulator	60
2.6	Comparison between IRAS and <i>caravela</i> images.	66
3.1	IRAS observed colour-colour diagrams	80
3.2	IRAS synthetic colour-colour diagrams.	81
3.3	IRAS synthetic colour-colour diagrams.	82

3.4	IRAS synthetic colour-colour diagrams.	83
3.5	IRAS synthetic colour-colour diagrams.	84
3.6	IRAS synthetic colour-colour diagrams.	85
3.7	IRAS synthetic colour-colour diagrams.	86
3.8	IRAS synthetic colour-colour diagrams.	87
3.9	IRAS synthetic colour-colour diagrams.	88
3.10	IRAS synthetic colour-colour diagrams.	89
3.11	IRAS synthetic colour-colour diagrams.	90
3.12	IRAS synthetic colour-colour diagrams.	91
3.13	Best-fit model	96
3.14	2D K-S test	96
3.15	Distribution of UC H II regions detected in model \mathcal{M}	98
3.16	Galactic latitude <i>vs</i> galactic longitude for the <i>caravela</i> sources.	100
4.1	Three colour <i>caravela</i> image at the PACS wavebands: $70\ \mu\text{m}$ (blue), $110\ \mu\text{m}$ (green), and $170\ \mu\text{m}$ (red).	108
4.2	Herschel predicted colour-colour diagrams.	112
4.3	Distribution of UC H II regions detected in model \mathcal{M} at the Herschel wavebands.	117
5.1	PACS and SPIRE + diffuse emission maps	130
5.2	PACS and SPIRE + diffuse emission three colour maps	132

List of Tables

2.1	Input parameters for the IRAS simulation	62
3.1	Input parameters for models $\mathcal{A}, \dots, \mathcal{G}$	73
3.2	Input parameters for models $\mathcal{H}, \dots, \mathcal{N}$	74
3.3	Input parameters for models $\mathcal{O}, \dots, \mathcal{U}$	75
3.4	IRAS Catalogues cross-matching	97
4.1	Input parameters for the Herschel simulation	106
4.2	Output parameters for the Herschel simulation	110
4.3	Herschel catalogues cross-matching	113

Acknowledgments

This dissertation was written at the Centre for Astrophysics Research, University of Hertfordshire. In particular, I like to thank Dr Mark Thompson for the great support, many discussions about star formation, comments and corrections until the last hour of work.

I would also like to show my gratitude to Dr Antonio Chrysostomou and Dr Michele Pestalozzi for the technical help and encouragement.

I am indebted to Prof. James Hough, director of the Centre for Astrophysics Research, for giving me a PPARC-equivalent studentship.

For all the collaborative work on the diffuse emission modelling, and having me working at the California Institute of Technology for a couple of weeks, I want to express my gratitude to Dr Roberta Paladini.

To everyone in *the office*, I just want to say thanks for a great working atmosphere and many discussions not about Astronomy.

I am grateful to my parents for supporting me throughout my extended education.

Last but by no means least, thanks Ana and Frederico for being always with me!

Chapter 1

Context: Observational Surveys for Star Formation

1.1 Introduction

The universe is dominated by stars. The Milky Way contains a hundred billion stars and it is nothing but one of the innumerable galaxies in the universe. Most of the elements known to mankind on Earth were produced inside a star: these heavier elements are converted by these fascinating objects, from the Big Bang hydrogen and helium.

The understanding that stars are transient objects, that they begin their lives, evolve following an evolutionary path and die, is the fruit of the last two hundred years of research. The accepted theory is that stars work by nuclear fusion, and that a constant balance must exist between fusion based outward pressure and gravity. This delicate equilibrium controls the life cycle of stars.

A Sun-like star (with regards to its mass) lives for billions of years. The observation of the current number of stars in our Galaxy implies that there is a star

formation process. As with human generations, new young stars must appear to replace the dying ones.

This work exists in the context of star formation in our Galaxy. Astronomy lives in a survey-driven era. Large unbiased observational surveys, also in the star formation field of research, are underway or being prepared. These surveys represent the most significant step forward, for observational star formation in the infrared and sub-millimetre, since the IRAS mission launched in 1983.

The aim is to create computer software capable of simulating the survey results in high-mass star forming regions in the Galaxy, producing a synthetic point source catalogue of observed regions generated from a simulated survey of a model spiral galaxy. The catalogues that are generated simulate a range of far infrared and sub-millimetre observations including IRAS, Herschel and SCUBA-2.

This numerical tool will allow the user to infer physical properties of the Galactic population of high-mass star forming regions from such observations.

The motivation behind the project is to provide a Galactic analogue to the successful simulations of large scale extragalactic surveys such as SHADES and SWIRE (van Kampen *et al.*, 2005; Coppin *et al.*, 2006; Lonsdale *et al.*, 2003). These surveys allow the predictions of observed source counts. With the survey data, these the observed source counts to constrain the cosmological deduced from the survey.

The following two questions are the essential problems under analysis here.

1. Stars must be forming constantly. What is the distribution, in the Galaxy, of the birth places of these objects? Massive stars in particular, have short lifetimes compared with low- and intermediate-mass objects, therefore they become rarer and further away thus very difficult to study.
2. What are the physical properties of high-mass star forming regions?

The *thesis*, or the statement that is put forward as a premise to be maintained and proved during the following chapters, of this project is: the developed tool can produce simulated infrared point-source catalogues of high-mass star forming regions that can be used to statistically study the physical properties of these sources.

The main contribution resulting from this work is a computer software program that can be used to address the two open questions mentioned before.

1.2 Star formation

Although the first ideas that were correct about gravity playing a crucial role in star formation were put forward in the eighteenth century (Kant, 1755; Laplace *et al.*, 1829; Beuther *et al.*, 2006), the modern theory of star formation came to light only in the second half of the last century (Larson, 1969; Shu *et al.*, 1987; Beuther, 2002). According to the standard theory, isolated stars with $m < 8M_{\odot}$ form inside cold dense cloud cores collapsing due to the gravitational pull.

In this way a central protostar is born embedded in an envelope. The protostar continues to accrete mass from its surroundings. Observations have shown that disks can be formed at this stage, and that they are responsible for the majority of the accreting activity. For $m < 8M_{\odot}$, the protostar becomes optically visible, it evolves along the Hertzsprung-Russel diagram. Its luminosity originates from gravitational contraction and not from nuclear burning (Stahler *et al.*, 2000). When the nuclear luminosity (due to the CNO cycle; p-p for lower mass stars) becomes dominant, the pre-main sequence path comes to an end. The star is now called a Zero Age Main Sequence star (ZAMS) (Lada and Kylafis, 1999).

Low- and intermediate-mass stars have masses that range from 0.1 to a few M_{\odot} . Stars with $m > 8M_{\odot}$ form in a different manner. What causes the difference

in formation processes between these two categories of stars? The short answer is the radiation pressure which massive stars exert on the surrounding envelopes from the moment their nuclear core ignites. From the cited mass cutoff ($8M_{\odot}$), the radiative pressure is too strong and prevents further accretion, therefore massive stars could not exist ... the theory has to be adapted to the observations (massive stars do exist!). The central difference between a low-mass and a high-mass star is that high-mass star forming regions produce enough ultra-violet photons to ionise its surroundings and form a HII region. The next section presents the different alternatives.

1.3 High-mass star forming regions

The study of massive stars is characterised by three words: remote, rapid, and mysterious. high-mass star forming regions statistically exist further away from the Earth than low- intermediate-star forming regions. They evolve more rapidly than their $m < 8M_{\odot}$ counterparts. Due to these issues and to the inadequacy of applying the standard star formation theory (that was developed with lower mass stars in mind), high-mass star formation is a challenging field with many mysteries to be tackled.

Why study the infrared part of the spectral energy distribution (SED)? Massive stars form in dense, cold and dark molecular clouds (i.e. high extinction) therefore clumps are only visible in infrared/sub-millimetre regime of the electromagnetic spectrum.

The large columns of dust and gas ($N(H_2) = 3 \times 10^{23} \text{ cm}^2$, Morales *et al.* (2009)) of the dense molecular clouds that lie between them and the observer, making them completely dark at visual wavelengths. Thus, the study of the environment around recently formed massive stars is *only* successful through observations at infrared,

millimeter, and radio wavelengths, where the extinction is much smaller.

Stars more massive than $10 M_{\odot}$ have a profound influence on our Galaxy, but their formation and how they are distributed remain poorly understood. The answers to how high-mass stars are born, where they exist and what their observational characteristics are, form the context for this report.

A list of the fundamental reasons why our understanding of the formation of massive stars is very incomplete include (Zinnecker and Yorke, 2007):

- high dust extinction makes it difficult to observe high-mass stars during early formation phases;
- they are rare;
- they evolve quickly;
- the theoretical problem is very hard to solve;
- massive stars rarely are formed in isolation from other stars.

The main difficulty of the study of high-mass star formation is the lack of (global) observational data. This is due to a combination of causes, the most relevant being: high-mass protostars evolve very quickly resulting in a statistically incomplete sample of well-studied sources (selection effects are also important contributors to the statistical incompleteness of high-mass star forming regions samples), and, they seem to cluster and therefore it can be difficult to resolve individual sources.

Our current understanding of this interesting and complex topic is summarised in this section. This is based chiefly on Zinnecker and Yorke (2007), Churchwell (2002) and Beuther *et al.* (2006) (and references cited in these three reviews).

The study of the formation of high-mass stars still has basic problems not solved (Barbosa and Finger, 2005, for a list of the most relevant open questions in

this field), e.g. the radiation pressure high-mass stars produce on the surrounding dust and gas is sufficiently strong to stop further accretion. So do high-mass stars form in a qualitatively different way from low-mass stars?

The first phase of the evolution of what will become a high-mass star takes place in the dense cores of giant molecular clouds. Giant molecular clouds' structure appears to be self-similar over a wide range in size and mass, i.e. the way giant molecular clouds are fragmented, in terms of sizes and masses of the its substructures (clumps), is similar, in scale, to how clumps fragment into cores (Blitz, 1991; Blitz and Williams, 1999; Williams *et al.*, 2000). High-mass stars and their lower mass companions form from over dense regions of the giant molecular clouds named as clumps.

Clumps contain most of the mass of the parent giant cloud, they are gravitationally bound and have typical masses of $\geq 300 - 500M_{\odot}$. The most massive clumps ($> 10^3 - 10^4M_{\odot}$) will form stellar clusters (Blitz, 1991; Blitz and Williams, 1999). Cores are smaller, denser, and have lower mass ($\approx 10^1 - 10^3M_{\odot}$) than clumps. These are the sites of individual star formation, and prestellar cores are identified as the first stage in the process of forming high-mass stars.

During this phase, a central protostar is not yet present inside the prestellar core therefore they are optically thick in the near infrared, they have temperatures of only $10 - 20$ K, and their SED peaks in the far infrared at $\approx 200 \mu\text{m}$ (Churchwell, 2002; Garay and Lizano, 1999). Research on prestellar cores is still immature and incomplete as these objects are rare (lifetime less than $\sim 10^3 - 10^4$ years, Parsons *et al.* (2009); Chambers *et al.* (2009)) and no unequivocal detection has been reported for a prestellar core that would unambiguously evolve into a high-mass star (Ward-Thompson *et al.*, 1994; Fuller and Myers, 1987; Churchwell, 2002; Motte *et al.*, 2008).

Hot cores can be found in the centre of an equatorial accretion disk and a massive bipolar outflow parallel to the core's rotation axis. Outflows associated with low-mass stars are ubiquitous but their masses and kinetic energy are much smaller than those linked with high-mass stars. Also, high-mass outflows are generally poorly collimated when compared with low-mass stars (Churchwell, 2002).

Although cores with diameters less than 0.1 pc, H_2 density greater than 10^7 cm^{-3} , warmer than 100 K, and containing massive protostars, are the precursors of UC HII regions, due to the rapid accretion, the protostar does not exhibit a detectable HII region at this stage. This is a rare phase as a result of the short lifetime of these objects.

The evidence that massive protostars have high accretion rates and short lifetimes can be found in the analysis of properties of outflows associated with high-mass star formation. Observations allow the determination of outflow dynamical ages and masses through mass outflow rates (typical values of $10^{-3} M_{\odot} \text{ yr}^{-1}$). The net result should be that some mass is added to the protostar at this stage of evolution, therefore, time scales for the lives of these objects are of the order of $10^4 - 10^5 \text{ yr}$ (Beuther, 2002; Zinnecker and Yorke, 2007; Zinnecker, 2007) ^a.

While the high-mass protostar is embedded in the core, it ionises the gas and forms an ultracompact HII region. UC HII regions are one of the most reliable tracers of recent massive star formation (Churchwell, 2002; Beuther *et al.*, 2006; Wood and Churchwell, 1989; Barbosa and Finger, 2005). They are dense, compact (less than 0.1 pc) aggregations of photo-ionised gas.

The ionising star of a UC HII region is on the main sequence^b, and obeys the empirical relations found for high-mass stars that no longer are inside a core, the mass-luminosity relation ($L \propto M^3$) and mass loss rate relation ($\dot{M} \propto L^{1.7}$)

^athe working assumption used to determine these properties is that high-mass stars form via accretion of surrounding material (through an equatorial disk).

^bMassive stars begin burning H while they are still accreting matter

(Churchwell, 2002).

An interesting empirical result is that UC HII only appear to have a few possible morphologies. These were classified as cometary, core-halo, shell, and irregular and multiple peaked structures (Wood and Churchwell, 1989; Kurtz *et al.*, 1994).

Observations of UC HII regions in the far infrared reveal that these objects are bright (i.e. have very high values of flux) in the far infrared wavelengths. This combined with strong obscuration in the visible (and even in the near infrared) allowed researchers to conclude that these objects are embedded in their original cloud cores (Chini *et al.*, 1986; Hoare *et al.*, 1988; Wolfire and Churchwell, 1994).

UCHII regions, due to their high luminosities, can be used to investigate the global properties of our galaxy. They play the role of beacons spread throughout the Galaxy that can be used to study the overall structure of our galaxy (Churchwell, 2002).

The Galactic population and spatial distribution of high-mass stars (and their precursors) is unknown, and it is important to ascertain the number, spatial distribution, and ages of high-mass stars. This would determine directly the current rate of massive star formation in the Galaxy^c, reveal the positions in space of high-mass star formation relative to spiral arms (and define the spiral structure of the Galaxy), give estimates of the mechanical and radiative energy and momentum contribution to molecular clouds.

While the angular position (position in the sky) of the source is relatively easy to determine with sufficient accuracy, the determination of the heliocentric distance is not straightforward. For sources inside the solar circle there is a distance ambiguity because, assuming a certain rotation curve, two sources observed in same direction but located in the two locations where their trajectory intersects the line-of-sight,

^cThis would be possible to estimate independently of uncertainties in the IMF in the high-mass star regime

have the same line-of-sight velocity (Zinnecker and Yorke, 2007). The observer is therefore unable to discriminate between these two distances (see e.g. Busfield *et al.*, 2006, for methods of breaking this ambiguity). Nonetheless heliocentric distances are needed if one is to use high-mass stars as tracers of the overall structure of the Milky Way.

How can the distribution of high-mass stars reveal the spiral structure of the Milky Way? There is empirical evidence from face on galaxies that the distribution of high mass stars and spiral arms structure coincide.

The identification of UC HII regions has been done with IR-photometry. Data from the Infrared Astronomical Satellite Point Source Catalogue (IRAS PSC) have been extensively used to estimate the distribution and number of UC HII regions in the Galaxy (Wood and Churchwell, 1989; Hughes and MacLeod, 1989; Zoonematkermani *et al.*, 1990; White *et al.*, 1991; Helfand *et al.*, 1992; Becker *et al.*, 1994). The IR fluxes at the IRAS wavebands are used to select candidate UC HII sources based on the assumed colours of a typical object. Wood and Churchwell (1989) used it most successfully and identified 1650 UC HII region candidates with a scale height of 0.6 ± 0.05 degrees in galactic latitude corresponding to 90 pc at a distance of 8.5 pc. They defined selection criteria based on 25 – 12 and 60 – 12 colours, i.e. $F_{12 \mu\text{m}}$ and $F_{25 \mu\text{m}} \geq 10 \text{ Jy}$, $\log(F_{60 \mu\text{m}}/F_{12 \mu\text{m}}) \geq 1.30$, and $\log(F_{25 \mu\text{m}}/F_{12 \mu\text{m}}) \geq 0.57$. UC HII regions appear to over-populate quadrants I and IV of the Galactic plane. This result seems to indicate that high-mass stars are more likely to be found within a few tens of parsecs from the Galactic plane and that they lie inside the solar circle (the solar circle is defined as the set of positions in the Galactic plane with the Sun's galactocentric distance).

While the IRAS was a very successful survey, studies based on the IRAS fluxes suffer from the relatively poor positional accuracy and limited sensitivity. It is clear that higher spatial resolution and more sensitivity observational surveys coupled

with relevant modelling of the problem will allow a clearer understanding of the Galactic population and distribution of high-mass stars.

The Initial Mass Function (IMF) describes the relative fractions of stars of different mass at their birth. The IMF was first measured by Salpeter (1955), and more recent observations have confirmed this result. In principle the IMF can be considered across all scales: from a star cluster or association to a large region of a galaxy or even for a whole galaxy. To a first approximation, the observed stellar IMF is universal and independent of abundance differences in our and other galaxies (Beuther *et al.*, 2006; Churchwell, 2002; Massey, 1999; Massey and Hunter, 1998; Meyer *et al.*, 2000; Kroupa, 2001). The origin and form of the IMF is a fundamental open problem in stellar Astrophysics. There is evidence that the chief mechanism determining the IMF is the mass spectrum produced from cloud fragmentation (Beuther *et al.*, 2006; Corbelli *et al.*, 2005).

As alluded before, we were assuming that massive stars are formed via accretion of the surrounding matter through an equatorial disk. There is observational evidence that support the presence of massive bipolar molecular outflows. Also the detection of equatorial accretion disks has been reported. Nonetheless, there are comparable strong arguments in favour of a paradigm change with respect to the origin of massive stars (Churchwell, 2002). Within this new scenario, massive stars form in a qualitatively different way from low-mass stars. This new scenario is called the coalescence hypothesis and was first proposed by Bonnell *et al.* (1998).

The fact that massive stars appear to occupy the centres of stellar clusters (Beuther *et al.*, 2006), led to this scenario whereby the protostellar and stellar densities within a forming massive cluster are high enough ($\sim 10^8 \text{ pc}^3$) that massive young stellar objects physically collide and merge. In this way the outward effect of radiation pressure is suppressed (Bonnell *et al.*, 1998; Bally and Zinnecker, 2005). Stahler *et al.* (2000) and Bonnell and Bate (2005) have proposed similar models for

the formation of low-mass stars.

A less drastic model for the formation of massive stars suggests that the majority of the stellar mass is accreted via competitive accretion (Bonnell *et al.*, 2004). In the competitive accretion scenario the coalescence of protostars is not required, but the mass accretion rates of the more massive cluster members depend on the number of stellar companions. Stars at the centre of the cluster have a higher accretion rate, thus end up to be more massive than other stars that live in the cluster outskirts. Competitive accretion is the idea that most of a star's mass comes not from a parent core, but from gas in the cluster-forming clump that was not originally bound to that star. In competitive accretion, most stars do not continue to accrete significantly such that their masses are set from the fragmentation process. It is the few stars which continue to accrete that become higher-mass stars.

A recent development with potential to give important contribution to the field is the successful fit of high-mass protostellar objects (HMPOs) SEDs with radiative transfer code (Fazal *et al.*, 2007). 13 high-mass protostellar objects (HMPOs) were studied and their SEDs fitted to a grid of 2-D axisymmetric radiative transfers models. Fazal *et al.* (2007) show that the models fit the observed SEDs well, supporting the accretion-based scenario of massive star formation. The envelope accretion rates were found to be $\dot{M}_{env} \approx 10^{-2.5} M_{\odot}/yr$. They concluded that it appears likely that stars with stellar masses $M_* > 20$ solar masses can form via accretion.

To summarise, the main open problem of massive star formation can be stated as follows (Beuther *et al.*, 2006): *Do high-mass stars form similarly to low-intermediate-mass stars, i.e. via accretion from the surrounding envelopes, or are fundamentally different physical processes, e.g. coalescence of protostars, taking place?*

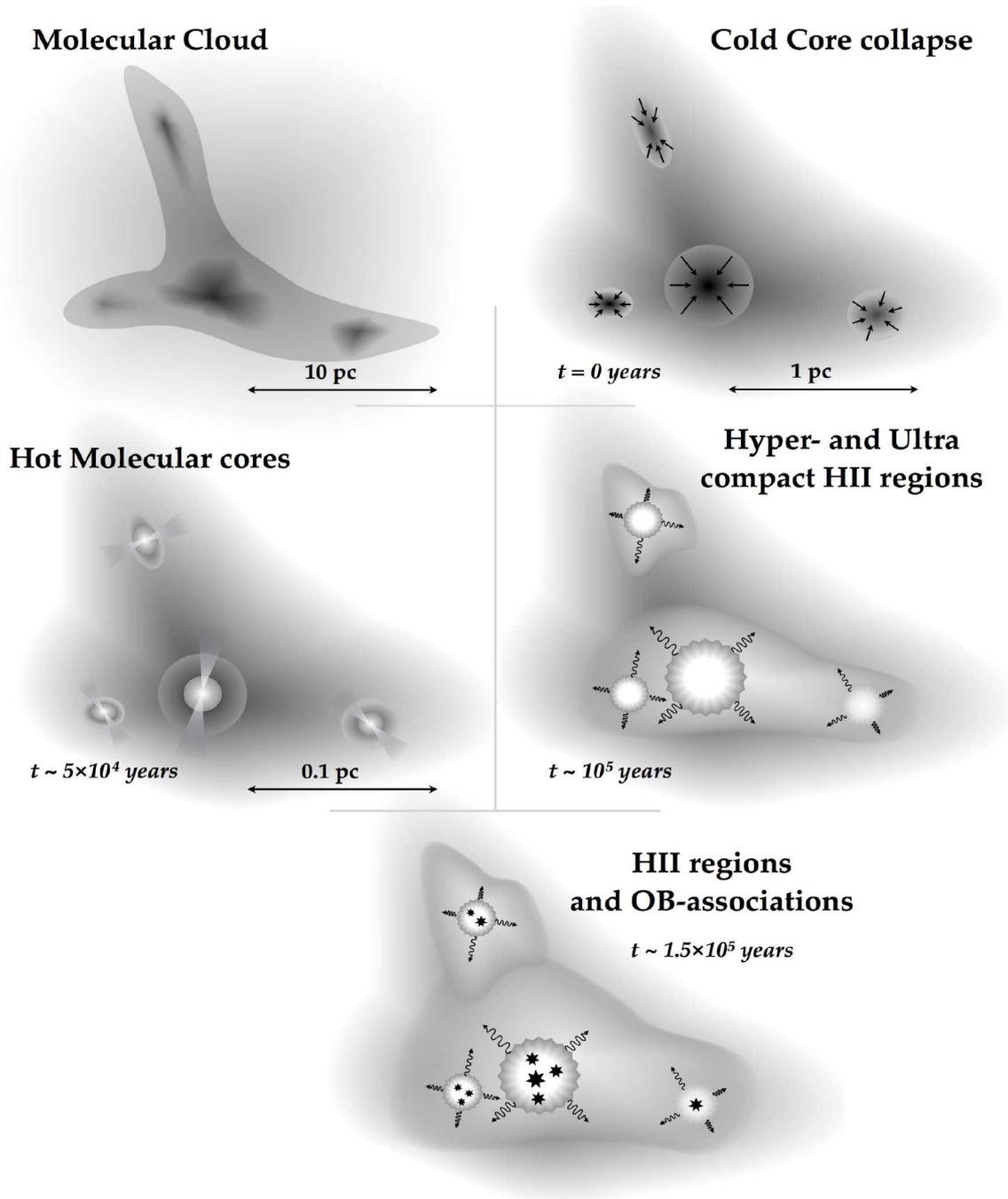


Figure 1.1: This figure illustrates the various stages during the formation of a group of high-mass stars. Further details in the text. It is adapted from Frieswijk (2008).

1.3.1 High-mass star formation may not be a scaled-up version of low-mass star formation

Here we discuss in more detail the arguments for proposing a distinct formation model for massive stars (Zinnecker and Yorke, 2007).

We use the terms massive star and high-mass star as synonymous. They refer to an OB star sufficiently massive to produce a type II supernova, i.e. $m > 8 M_{\odot}$ for solar abundances.

The $8 M_{\odot}$ limit arises from the existence, above this mass cut-off, of radiative forces on gas and dust. These forces are insignificant in the early stages of low-mass, solar-type stars, formation but, for massive stars, a substantial fraction of the luminosity is emitted in ionizing radiation^d. This radiation is responsible for new effects such as the photo evaporation of the star's accretion disk and protostellar envelope. This limits accretion and, crucially, the final stellar mass^e. In terms of timescales, the difference between low- and high-mass star formation is that low-mass stars form in a time t_{*f} short compared to the Kelvin-Helmholtz time t_{KH} , whereas high-mass stars generally have $t_{KH} \leq t_{*f}$ (Beuther *et al.*, 2006). As mentioned before, this result in the fact that low-mass stars undergo extensive pre-main sequence evolution after accretion has finished, whereas the highest mass stars can accrete a significant amount of mass while on the main sequence.

Accordingly to Beuther *et al.* (2006), observational evidence suggests that stars at least up to $30 M_{\odot}$ form via an accretion based formation scenario, i.e. a simple scaled-up version of the formation scenario for low- and intermediate-mass stars.

^dMassive stars have Kelvin times (the time required for a star to radiate away its gravitational binding energy) that are shorter than their formation times, and as a result they attain their full luminosities while still accreting from their natal clouds. As the radiation from such an embedded, massive star diffuses outward through the dusty gas in the protostellar envelope, it exerts a force that opposes gravity.

^eIn addition, the ionizing photons can photoevaporate the disks of the neighboring lower mass stars

The cited authors refer that the accretion-based formation scenario in turbulent molecular cloud cores is the more probable method to form most stars of all masses.

1.3.2 The spiral structure of our Galaxy

Why study the 3D Galactic distribution of the birth places of massive stars? The answer to this question has interest in itself, but also, it is, in principle, possible to investigate the spiral arms of our Galaxy modelling the positions of the high-mass star forming regions.

For external galaxies the distribution of star-forming regions along the spiral arms is, in general, evident from direct imaging. The situation is fundamentally different for the Milky-Way (Russeil, 2003). In our Galaxy, it is impossible to have a view of the whole Galactic plane from a direction perpendicular to it. Furthermore, the arms appear superimposed and merged together in any observation where the line of sight is parallel to the galactic plane. And the kinematic distance ambiguity makes it very hard to decompose the observed velocity structure into the true spiral structure of the Milky Way.

One possible method to investigate the grand spiral structure is to model the distribution of objects, on a galactic scale. These type of objects have to trace the spiral arms.

In this work, we are considering high-mass star forming regions as excellent tracers of the spiral structure of the Galaxy, as it is a well known observational result that HII regions and OB stars exist in the spiral arms of external galaxies (Russeil, 2003, and references therein).

The Russeil (2003) model for the Galaxy was adopted for *caravela*. It is a four arm model, based on a equiangular or logarithmic spiral (figure 1.2, page 15).

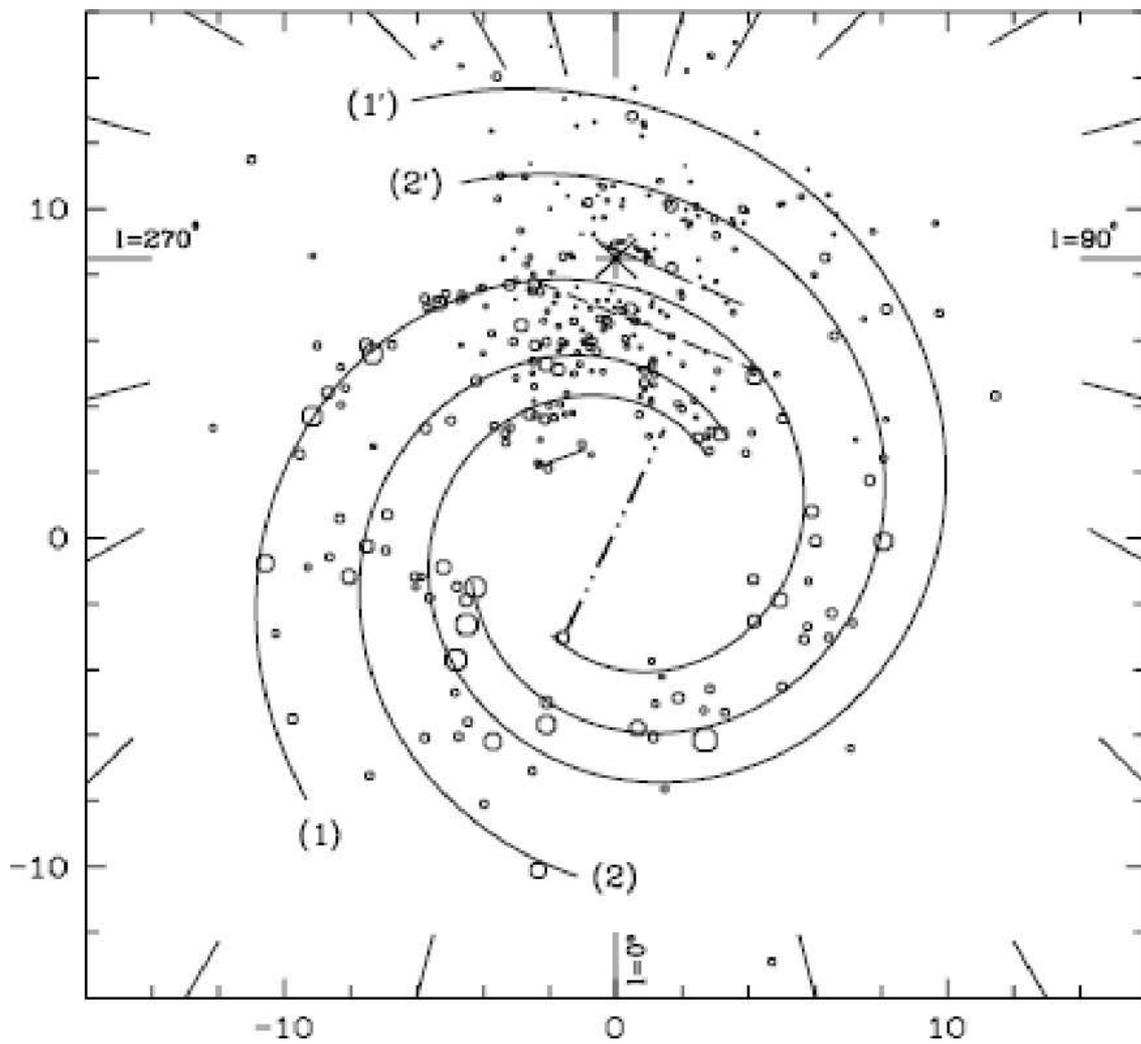


Figure 1.2: The Russel (2003) model for the Galaxy was adopted for *caravela*. It is a four arm model, based on a equiangular or logarithmic spiral. All the relevant equations and physical parameters of this model are described in detail in the next chapter.

1.4 Hi-GAL

Hi-GAL is a key-time survey of the Galactic Plane to be carried out with the Herschel Space Observatory. Herschel is the biggest far-infrared space observatory ever to be launched (with a 3.5 m mirror) (Molinari and the Hi-GAL Team, 2009).

Hi-GAL will map 240 square degrees of our Inner Galaxy, delivering a census of the following physical properties:

- temperature,
- luminosity,
- mass,
- Spectral Energy Distribution,

of star forming regions and cold diffuse structures.

Hi-GAL will map the inner Galactic Plane, i.e. (in degrees) $-60 < \ell < 60$, $-1 < b < +1$ ^f, in 5 photometric bands between $70 \mu\text{m}$ and $520 \mu\text{m}$ at a $4 - 40''$ diffraction limited spatial resolution.

The aim is to provide an homogeneous data set for a large number of objects well suited for statistical inferences on global properties (of high-mass star forming regions).

The aim of Hi-GAL is to detect the earliest phases of the formation of molecular clouds and high-mass stars and to use the optimum combination of Herschel wavelength coverage, sensitivity, mapping strategy and speed to deliver a homogeneous census of star-forming regions and cold structures in the interstellar medium. The resulting representative samples will yield the variation of source temperature, luminosity, mass and age in a wide range of Galactic environments at all scales

^fThe original proposal was to map (in degrees) $0 < \ell < 360$, $-1 < b < +1$

from massive YSOs in protoclusters to entire spiral arms, providing an evolutionary sequence for the formation of intermediate and high-mass stars.

Hi-GAL is the first dedicated project to study the early phases high-mass star formation in the Galaxy, with a legacy value similar to the IRAS mission some two decades ago.

The outcomes of Hi- GAL will consist of source lists and images to be released soon (during 2010).

The author of this work is part of the Hi-GAL team (data simulation working group). It was within the scope of the data simulation working group of Hi-GAL, that the collaborative work with Dr Roberta Paladini has emerged. The aim was to use the obvious synergies between the Paladini *et al.* (2007) diffuse emission model with *caravela*. Some results of this are presented in the final parts of this dissertation.

The author is included in the (long) authors list of (Molinari *et al.*, 2010), where the Hi-Gal Survey is described in detail.

1.5 Motivation for the project

High-mass star forming regions occur throughout the Galaxy. They are very luminous in the infrared so can be seen at large distances. Large infrared observational surveys are guaranteed to come on-line soon, e.g. Hi-GAL. For a statistical interpretation of the observational data a model is needed. One of these models is *caravela*. In theory, there is an advantage in modelling the observed data before the observation. In practice, this is what *caravela* tries to achieve in chapter 4.

The main motivation for this work is to provide a model able to be used in the interpretation of IR high-mass star forming regions survey data sets.

1.6 Outline of this dissertation

In this work, we describe how a numerical tool (*caravela*) was developed to study high-mass star forming regions. Two categories of results are presented: one using *caravela* to constrain known observations, the others apply the best-fit model found from the latter to the upcoming Herschel data sets.

This dissertation includes a description of what has been done and in what context (Chapters 2 and 1, respectively), what came out of this (Chapters 3 and 4), a discussion of these results and finally some conclusions that can be drawn (Chapter 5). I also discuss how *caravela* can be used in the future (Chapter 5). The outline of the thesis is as follows:

- We introduce in Chapter 1 the high-mass star formation research area as a very active field of research, presenting its main results, assumptions and problems. Some important differences between low-mass and high-mass star formation are presented in §1.3. A set of observational surveys[§] will make a major con-

[§]Hi-GAL is one of these surveys. It uses the photometric instruments on board the Herschel

tribution to our understanding of how high-mass stars form. The directly relevant to this work is described in §1.4.

- Chapter 2 presents *caravela*, the numerical tool developed during this project. The *caravela* code aims to produce simulated data that can be used in the interpretation of both currently available data sets (e.g. IRAS) and future observational surveys (e.g. Hi-GAL). This chapter describes the method followed during this project. At the end of chapter 2, *caravela* is put into use: a sample model tests some of the code features with physically realistic input parameters. This will give the reader a first preview of the following results chapters.
- In chapter 3, the *caravela* grid of 21 models is presented. One model is selected as being in better agreement with the IRAS point source catalogue distribution. The selection process for the best-fit model is explained and some conclusions about the *real* distribution of high-mass star forming regions are discussed.
- Chapter 4 describes how *caravela* can be used as a predictive tool. The best-fit model found in the previous chapter is *observed* through the model again, but this time *caravela* was set up to simulate the future Herschel Space Observatory.
- Furthermore, *caravela* can be successfully combined with an independent, but complementary, code^h. The latter is a simulation of the diffuse emission in the Galaxy and the combined results are presented in Chapter 4.
- We summarise the main results of this thesis in chapter 4 and outline future work. A discussion on the main conclusions of this project is part of the fifth chapter.

Space Observatory to study the plane of the Milky Way. *caravela* is being used to support the work of the Hi-GAL data simulation working group.

^hThis independent code is a well established numerical tool developed by Paladini *et al.* (2007)

- Finally, an example input parameter file for `caravela` is shown, and briefly described, in appendix C.

Chapter 2

Method: The caravela code

2.1 Introduction

In this chapter we present the tool that is behind this work: the `caravela`^a code. It is an assembly of numerical modules that simulate realistic data streams of high-mass star forming regions from a chosen Galaxy model.

We start with a bird’s-eye view of `caravela`, and then a more detailed explanation of the physical constrains follows. A few lines of code illustrate each of the main components of the program. The description follows the *sources - observation - photometry and catalogue - plots and images* logic^b.

The structure of chapter 2 is as follows. Section 2.2, page 23, accounts for the first decisions in the (complex) process of creating the `caravela` code, i.e. what programming language and design would best serve our aims. It also summarises the advantages of building the code in C++, and why `caravela` has an object oriented

^a`caravela` (in English caravel, from Greek *karabos* *horned beetle* or *light ship*.) is named after the fast Portuguese ships of the 15th-17th centuries. These ships were used to discover new countries, helping the Europeans to estimate how many new lands there were and where to put Brazil, South Africa, India and Japan on the 3D world map.

^bIn a way this is different to an observer’s viewpoint, but it is how `caravela` works and its classes are organised.

structure. In section 2.3, page 27, an overview of the whole code is presented. Section 2.4, page 33, explains how a novice user of `caravela` can start a first simulation. This is intended to be a quick-start guide to `caravela`.

Next, the parameter file is explained in some detail, and all the input parameters are described individually in section 2.5, starting at page 34. The first module of `caravela` is summarised in section 2.6 (page 39), that is the building of the theoretical catalogue. Functionally, the next step in the simulation is to implement a synthetic observation of the objects that lie in area of analysis, this is presented in section 2.7, page 52.

Section 2.8 (page 61) represents the final task that `caravela` does before quitting one specific run: to identify the sources and extract a set of their photometric properties, i.e. `caravela` generates the final point source catalogue. We present, in section 2.9, page 61, an early application of `caravela` using realistic input parameters. The reader can gain a feel for what the `caravela` output consists of and it is a prelude to the following results chapters. Section 2.10, page 67, summarises the main results from chapter 2.

The specific problem that `caravela` addresses is: what point source catalogue results from *an observation* of a distribution of high-mass star forming regions? Also given and observed distribution, can we estimate the number of high-mass star forming regions present in the Milky Way?

The aim is to create a program capable of simulating the point-source catalogues resulting from the IRAS, PACS, SPIRE and SCUBA-2 instruments, in the context of high-mass star formation.

The main physical assumptions are:

- the interstellar medium is optically thin at the relevant wavelengths,

- high-mass star formation takes place, predominately, in the spiral arms of the Galaxy,
- the high-mass star forming regions are distributed in statistical uniform way along the length of the Galaxy spiral arms.

In this chapter, the (theoretical) equations and their numerical implementation are presented together^c.

2.2 Programming language and design

From a numerical point of view, the first decisions made were on which programming language to use, and how to structure *caravela*.

I decided to use C++. This decision was based on several factors, the most relevant being that,

- C++ is a fast and flexible programming language which allows a straightforward implementation of an object orientated design,
- it is increasingly used in similar scientific research projects (e.g. the Planck mission simulation pipeline described in Reinecke *et al.*, 2006),
- there is an important amount of (freely) available scientific libraries written in C++ that are directly relevant to this project (e.g. library to create and modify FITS files Wells *et al.*, 1981; Dorman, 2001),
- there is a version of the Numerical Recipes book in C++ (Press *et al.*, 2002).

An object-orientated design is more intuitive and fun than traditional functional-style design and it has been used in similar scientific research projects (Reinecke

^cI think this approach makes the code more transparent.

et al., 2006). Figure 2.2.1, page 26, illustrates how class inheritance is explored in *caravela*, in particular how this design takes advantage of the fact that different types of astrophysical objects share a number of basic properties. An UCH II and a cold core both have a property called mass in the code, but are distinguished by the type of SED associate to them (either a radiative transfer model for population 1 or a pure grey-body curve for population 2 objects), i.e., all objects, being either UCH II or cold cores, inherit some members from a base class (called high-mass star forming regions object).^d

IDL is used in the handling of the resulting images (FITS files), and the final output catalogues are produced by *cupid*^e. IDL is used at this stage in *caravela* because it is extremely efficient in handling large 2D arrays. This proved invaluable to process operations on the super-resolution images (and subsequent final modelled images).

The two libraries most extensively used in *caravela* are the *slalib* and Numerical Recipes libraries. *Slalib* (C version) enabled *caravela* to transform 3D galactocentric coordinates in 2D sky coordinates. Several Numerical Recipes' functions form the numerical backbone of the algorithm, in particular, all the Monte Carlo sampling uses the `NR::ran2()` and `NR::gasdev()` random number generators (Press *et al.*, 2002).

The code is not designed in parallel mode per se. In order to gain efficiency when doing all-sky simulations (or multi-wavelength jobs), *caravela* was installed in the University of Hertfordshire Centre for Astrophysics Research computer cluster.

^dIn object-oriented programming, inheritance is a way to form new classes using classes that have already been defined. The new classes, known as derived classes, inherit attributes or properties of the pre-existing classes, which are referred to as base classes. It helps to structure the code better and to reuse existing code.

^e*cupid* (*CIUmP IDentification*) can be used for identifying clumps of emission in 1D, 2D or 3D data arrays. It is primarily targeted at the needs of the SCUBA2 advanced data products pipeline, it follows the Starlink pattern of being instrument-independent, so it can be used on the *caravela* synthetic images. EXTRACTOR (EXTRACTOR is SEExtractor (Source-Extractor) program re-packaged for use in the Starlink Software Environment). This is can also be used in alternative to *cupid*.

A simple form of parallel computing is then obtained by setting ten or more `caravela` jobs simultaneously.

2.2.1 Random variations

`caravela` uses random number generators extensively throughout. The random numbers generated by these stochastic functions are used to sample probability functions that represent *physical properties*, e.g. the 3D distribution of positions of the objects, or, the range in masses. Having a random probability distribution or pattern that can be analysed statistically but not predicted precisely for all the relevant physical properties, implies that results from any `caravela` simulation have to be analysed from a statistical perspective. `caravela` results are, therefore, useful in the discussion on constraining the total number of sources that will be detected by a specific instrument. Trying to use `caravela` to investigate exactly where these objects will appear in the observational image is, however, not possible.

It is important that `caravela` runs are statistically stable, i.e. that random variations are insignificant when compared to the variations produced by changing significant `caravela` parameters.

In order to study the statistical stability of `caravela`, we used the Press *et al.* (2007) test for stochastic quality: *any two different random number generators ought to produce statistically the same results when coupled to your particular applications program*, in our case `caravela`. Both while building the code and during the final stages, we tested `caravela` against this criterion by alternatively using all the distinct random number generating functions from the C++ Numerical Recipes library (Press *et al.*, 2002): `NR::ran0`, `NR::ran1`, `NR::ran2`, `NR::ran3`, and, `NR::ran4`. These functions use independent methods to simulate *true* random numbers.

The significant physical results were invariant, in a statistical sense, to the

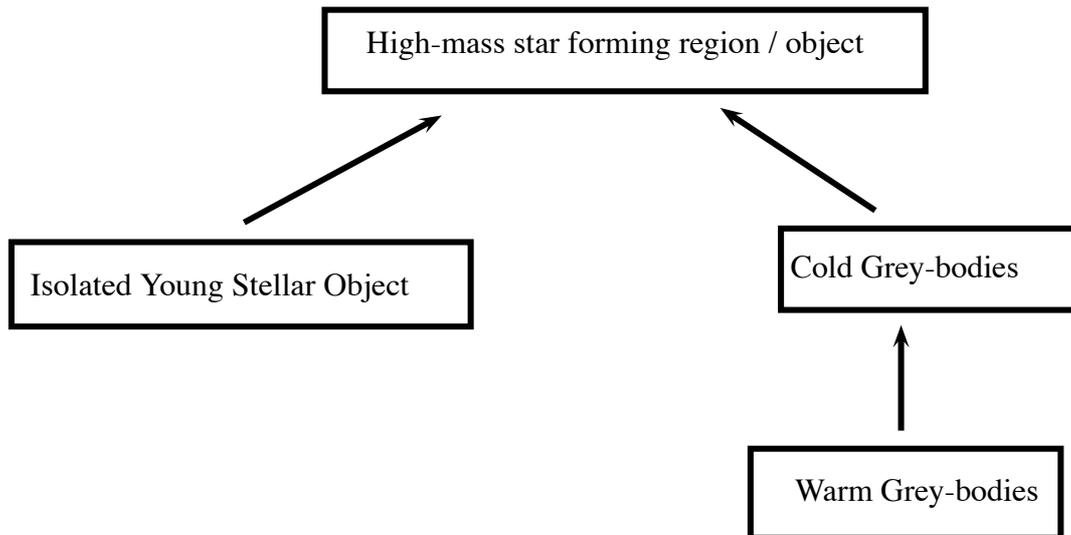


Figure 2.1: Class diagram for `caravela`. Each arrow represents the dependence of the derived class from the base class in C++.

choice of random number function. We conclude that `caravela` is stable.

`caravela` uses `NR::ran2` because this is presented as the best random function in the C++ Press *et al.* (2002) library. The authors of Press *et al.* (2002) will pay a thousand dollars to anyone who can demonstrate that `NR::ran2` is not a *perfect* computer random number generator, i.e. by finding a statistical test that `ran2` fails in a nontrivial way. `NR::ran2` has resisted to all attempts presented (Press *et al.*, 2007).

Summary

In summary, `caravela` is built using an object-oriented design in C++; IDL is used to manipulate the FITS files; and a source detection algorithm (e.g., `cupid` or `EXTRACTOR`) generates the catalogue. To make the code (more) user-friendly and easier to maintain, a C shell script controls all the `caravela` components. Also, all the input parameters are defined in a single text (.txt) file. `caravela` is statically stable.

2.3 Bird's-eye view of *caravela*

The overall structure of *caravela* is presented here. The next sections are devoted to a more detailed analysis of the code, illustrated with selected lines of code.

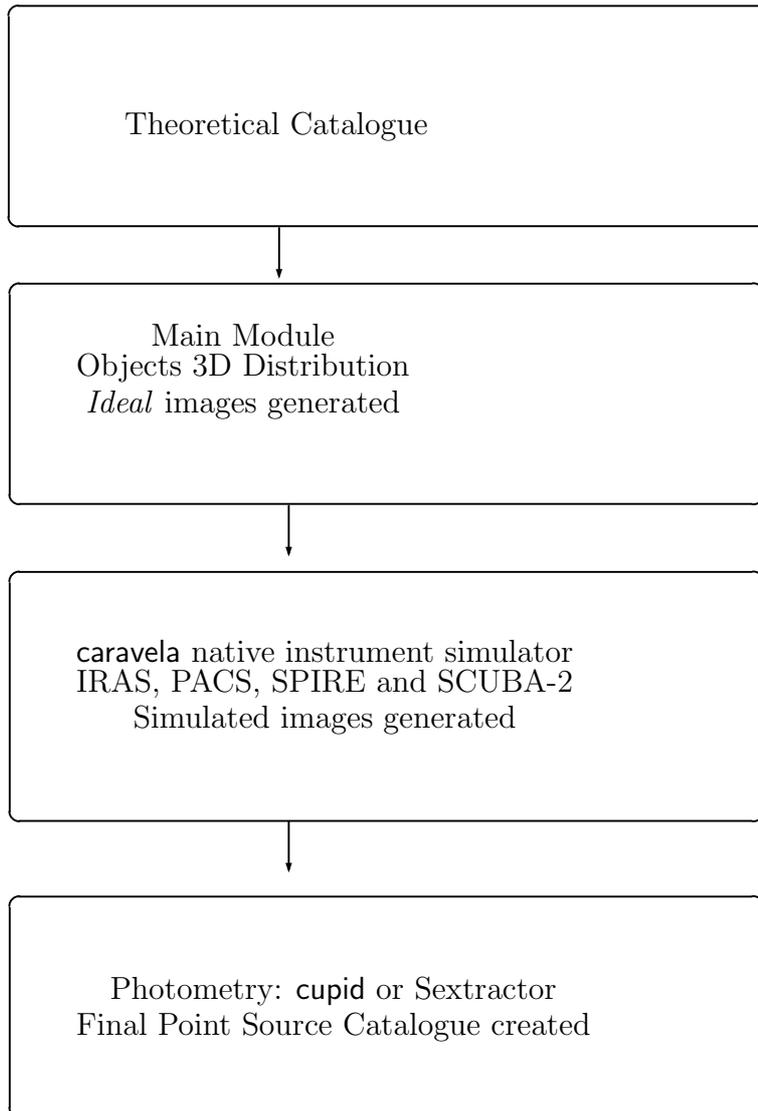


Figure 2.2: *caravela*'s main stages.

Also, given a measured catalogue, the code can be used to estimate how many high-mass star forming regions there are in the Galaxy. What is their distribution? How does the observation and data reduction, e.g. photometry on automatically identified clumps, affect the science conclusions?

The following description is translated graphically in figure 2.2, page 27.

The first step in the code is to build a catalogue of synthetic sources. These sources represent the real population of high-mass star forming regions in the model Galaxy. The input parameters are grouped in four categories:

1. Total number of sources
 - (a) number of isolated Young Stellar Objects, SEDs sampled from Robitaille *et al.* (2006)
 - (b) number of cold grey-body sources
 - (c) number of compound sources, i.e. YSO surrounded by a warm grey-body
2. Sources' sizes
 - (a) average radius
 - (b) radius standard deviation
3. 3D distribution in the Galaxy
 - (a) distribution starting radius, i.e. the inner radius of the population^f
 - (b) distribution starting radius standard deviation
 - (c) standard deviation for the height of the distribution, i.e. how thick the disk will be in the direction perpendicular to the Galactic plane
 - (d) spiral arms and/or rings
 - (e) length of the spiral arms (if present)
 - (f) width of the spiral arms (if present)
 - (g) number of spiral arms

^fUsually this can be identified with half the size of the Galactic central bar.

- (h) where do the spiral arms start (this is usually set to the same value as the distribution starting radius)
 - (i) spiral arms pitch angle
4. parameters for the spectral energy distribution (SED) for each type of source
- (a) for the isolated Young Stellar Objects: (initial) mass function exponent for radiative transfer SED sampling
 - (b) for the cold and warm grey-body sources:
 - i. grey-body equation parameters, i.e. optically thin frequency and beta
 - ii. (initial) mass function exponent (Nutter and Ward-Thompson, 2007)[§]

Three different types of objects can be selected by the user of *caravela*: compact objects, grey-body objects, grey-body *clouds* with compact objects inside. These are named populations 1, 2 and 3, respectively.

Populations 1, 2 and 3, were included as the built in populations in *caravela* because they are adequate prototypes for distinct phases of star formation evolution (described in the previous chapter). These populations do not represent all the stages of stellar formation. In practice, these populations also result from the test runs of *caravela*, i.e. during early stages of development of the code these populations seemed to give the most consistent results when compared to a number of observational test data sets.

caravela can use other sources from different populations. To this end, the user provides *caravela* with a independent theoretical catalogue.

caravela will then use this information to generate one catalogue. Monte Carlo

[§]The IMF is used in the sampling of the sources so that a more realistic set of high-mass star forming regions objects can be used. In this way the masses of the objects used obey an *observed* mass function (this refers to the input distribution of objects; the final distribution of sources detected in the end of the simulation *is* a distinct distribution) and not necessarily the same as the input distribution.

techniques are used extensively in this part of the code, so the theoretical catalogue is *one* possible realisation. The catalogue is the output from the first module of *caravela*. For each source, the catalogue will contain the following data:

1. ID
2. 3D position in Galactocentric coordinates, i.e. distance to the centre of the Milky Way, perpendicular distance to the Galactic plane, and, angle measured from the Sun's direction
3. 2D position in Galactic coordinates, i.e. ℓ and b
4. physical size
5. flux, at several wavebands^h, at the surface of the object
6. mass

As summarised in figure 2.2, page 27, when the catalogue is finally created, the main module of *caravela* reads in the file and generates a set of imagesⁱ. This is **second step**. These images are *super*-resolution images, i.e. images with a spatial resolution higher than any instrument simulated in subsequent stages of the code. The images correspond to the wavelength, central position and size^j, selected by the user. One of the issues here is to correctly account for the projection (and projection errors) from 3D Galactocentric coordinates to 2D Galactic coordinates (and then to 2D coordinates on the detector plane). *caravela* uses the *slalib* C routines.

What the main module of *caravela* does, conceptually, is to implement a 3D model of the distribution of high-mass star forming regions in the Galaxy, and

^hFor the grey-body sources, 12 μm , 25 μm , 60 μm , 70 μm , 100 μm , 110 μm , 170 μm , 250 μm , 360 μm , 520 μm . For the isolated Young Stellar Objects, IRAS bands: 12 μm , 25 μm , 60 μm , 100 μm , PACS bands: 70 μm , 110 μm , 160 μm , SPIRE: 250 μm , 360 μm , 500 μm , SCUBA-2: 450 μm , 850 μm .

ⁱThese images are in the FITS format, with World Coordinate System (WCS) header information.

^jImage central position and size selected in Galactic coordinates.

then observe it from a *super*-resolution and *ideal* sensitivity telescope, at one of the available wavelength bands. Each object's position is determined from the 3D spatial distribution^k.

All the objects' distances to the Earth are known, but not all objects will be detected (and/or not detected as one object) and their distances will not be an output from the simulated observation (the simulated observation is the next step).

The *caravela*'s next module tries to simulate the observing instrument (**third step**). The instruments available are the IRAS Space Telescope and the Herschel Space Observatory. The aim was not to build a complete simulation of these instruments, as this would be beyond the scope of the project: the simulators modules present in *caravela* are simple and only consider the instruments' spatial resolutions, wavebands and noise levels.

The output are FITS images that simulate the observed images (refer to figure 2.2, page 27). As with previous modules, the details of the native instrument simulator will be explained in the next sections. Due to the modular approach that underpins *caravela*, any other instrument simulator can, in principle, be used at this point. The instrument teams of IRAS and Herschel have simulators that are physically more realistic and efficient.

The final and fourth step missing to achieve *caravela*'s goal of producing a Point Source Catalogue, is to measure the sources' observed fluxes. The code uses aperture photometry on the images described above. Two options are available: *SExtractor* and *cupid*. Both techniques result in a photometric catalogue.

These are the fundamental steps (each is presented in more detailed in the following sections) that form *caravela*. In addition, a number of quick plots and checks are made throughout each run. E.g., these plots consist of galactic longitude

^kAll the Monte Carlo sampling uses the `NR::ran2()` and `NR::gasdev()` random number generators (Press *et al.*, 2002).

vs number of sources, and, 3D plot of the Galactocentric distribution of sources. Figure 2.3, page 48, is an illustration of one of these graphs.

`caravela` can also be set to produce more than one wavelength in one run. This can be particularly useful if one is trying to produce colour-colour diagrams¹.

Note that one run of the code, i.e. one combination of the input parameters, produces the following output (figure 2.2, page 27):

1. one theoretical catalogue
2. one (set) of *super*-resolution images
3. one (set) of instrument simulated images
4. one *observed* catalogue
5. ancillary plots, e.g. ℓ and b histograms

These output elements were designed to allow one to try to study: how the 3D distribution of sources and instrument properties affect the observed catalogues (comparing 1 with 4), what effect does the instrument simulator have on the observed images (comparing 2 and 3), the source extraction and photometry recipes (comparing 3 to 4). The flow chart presented in figure 2.2, page 27, summarises the above description.

In practice, `caravela` is run from a C shell script that sequentially compiles and runs the C++ and auxiliary IDL components of the code. All the input parameters can be changed and saved in one ASCII text file.

¹colour-colour diagrams are automatically generated in this mode.

2.4 Running caravela

A caravela run is started by typing

```
>./caravela.csh input_file.txt output_directory source_extractor_keyword
```

in the terminal window. `caravela.csh` is the overall shell script that compiles^m and runs the C++ and IDL parts of `caravela`.

`input_file.txt` is the text file containing all the input parameters needed throughout the simulation. All the output products, i.e. theoretical catalogue, observed catalogue, *super*-resolution image, real image, plots and control information files, are saved in the `output_directory` directory chosen by the user. It is useful to name this input file as `caravela_input_waveband.txt`, where `waveband` is the waveband that defines the instrument to be used, e.g. SPIRE 250 μm . An empty input file (with only typical values for the parameters less likely to be changed filled in) is part of the `caravela` distribution. The `source_extractor_keyword` can be set to `cupid_clumpfind`, `extractor`, or `no_source_extraction`ⁿ.

A *fake* parallel mode is obtained by starting more than one simulation, as described above, simultaneously in different UNIX/Linux terminals. For example, a multi-wavelength simulation^o is more efficiently achieved with the aforementioned parallelism.

Although C++ can be compiled in MSWindows, the current version of the code is limited to UNIX/Linux based terminal systems.

^mIf needed ...

ⁿThese three keywords are explained in the source extraction section

^oSimilarly a single wavelength all-sky simulation can be set by running more than one `caravela` simulation on contiguous areas of the sky.

2.5 The parameter file

The *caravela* parameter file is described in more detail here.

An example of a complete input parameter file is presented in appendix C, page 136, at the end of this work.

The first three blocks of input parameters present in a *caravela* parameter file (the file is a text file) refer to the three possible populations to be used: populations 1, 2 and 3. In the next lines, index 1 is substituted by 2 and 3 in the next two otherwise identical blocks of variables (not shown).

NUMBER_SOURCES_POP_1	0.0
OBJS_LINEAR_SIZE_SCALE_MU_POP_1_AU	5.0e5
OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_1_AU	5.0e4
Z_RANGE_POP_1_KPC	0.05
TEMPERATURE_MU_POP_1_K	30.0
TEMPERATURE_SIGMA_POP_1_K	3.0
NU_0_MU_POP_1_HZ	1.8e13
NU_0_SIGMA_POP_1_HZ	0.0
BETA_MU_POP_1	2.0
BETA_SIGMA_POP_1	0.0
SPIRAL_FLAG_POP_1	1.0
SPIRAL_N_ARMS_POP_1	4.0
SPIRAL_A_ARM_1_POP_1	3.0
SPIRAL_B_ARM_1_POP_1	0.2493
SPIRAL_LENGTH_POP_1	30.0
SPIRAL_WIDTH_SIGMA_POP_1	0.2

The first parameter is the total number of theoretical sources to be included in the model, for population 1, 2 and/or 3.

The next two parameters are the objects' linear size distribution average and standard deviation (in AU), respectively.

The next parameter determines the distribution of the sources in the galactocentric coordinate z , i.e. the height from the galactic plane.

The following six parameters completely characterise the grey-body function (used for population 2 and 3 objects). For population 1, this set of parameters is not in use. These six parameters are of the form (for each distribution): average value, standard deviation for temperature, ν , and β , respectively.

The `SPIRAL_FLAG_POP_1` is a flag variable: if equal to 1 then the correspondent population will be distributed in a spiral; if equal to 2 no spiral structure is generated and the objects are distributed in rings; if equal to 3 the spiral and ring structures coexist in the simulation.

The next five `SPIRAL_` parameters are the:

- number of spiral arms,
- equiangular spiral constant a ,
- equiangular spiral constant b ,
- spiral arm length (needed for the normalisation),
- spiral arm width,

respectively.

The same block is repeated three times.

Next, we have all the parameters that refer to the image(s) properties and simulator to be used.

IMAGE_WAVELENGTH_MICRONS	100.0
IMAGE_SIZE_X_DEGREES	1.0
IMAGE_SIZE_Y_DEGREES	0.0
IMAGE_CENTRAL_POSITION_LAT_DEGREES	0.0
IMAGE_CENTRAL_POSITION_LONG_DEGREES	40.0
INDIV_IMAGES_LINEAR_SIZE_X_DEGREES	2.0
INDIV_IMAGES_LINEAR_SIZE_Y_DEGREES	2.0
IMS_SCALE_PIXELS_PRE_DEGREE_X	180.0
IMS_SCALE_PIXELS_PRE_DEGREE_Y	180.0
PROJECTION_TYPE_FLAG	1.0
INSTRUMENT_RESOLUTION_ARCMIN	0.5

The first parameter defines the wavelength band to be used. This can be set to any of the IRAS, Herschel or SCUBA bands.

Next, the user defines the image size (horizontally and vertically), in degrees: parameters `IMAGE_SIZE_X_DEGREES` and `IMAGE_SIZE_Y_DEGREES`.

The following two parameters decide where the synthetic *caravela* telescope will be pointing to, in galactic coordinates (b, ℓ) : `IMAGE_CENTRAL_POSITION_LAT_DEGREES` and `IMAGE_CENTRAL_POSITION_LONG_DEGREES`.

The total area of the sky to be observed is set by `IMAGE_SIZE_X_DEGREES` and `IMAGE_SIZE_Y_DEGREES`. This global area is subdivided into individual adjacent tiles. Each tile, in a *caravela* simulation, is `INDIV_IMAGES_LINEAR_SIZE_X_DEGREES` \times `INDIV_IMAGES_LINEAR_SIZE_Y_DEGREES`, in size (in degrees).

`IMS_SCALE_PIXELS_PER_DEGREE_X` and `IMS_SCALE_PIXELS_PER_DEGREE_Y` define the scale of the output 2D matrices (i.e. the super-resolution images described in this chapter), in units of pixels per degrees, in the horizontal and vertical directions respectively.

`PROJECTION_TYPE_FLAG` is a flag variable that decides which projection type should be used in `caravela`. This is the projection between the 3D galactocentric coordinates and the 2D Cartesian coordinates in the CCD instrument plane. In the current version of the code, only the gnomonic or tangent plane projection is available to the user.

For the instruments where we did not have a realistic model PSF available, e.g. the IRAS 100 μm band, the super-resolution image is processed using a PSF approximated by a Gaussian profile. This normal function is defined as having a `INSTRUMENT_RESOLUTION_ARCMIN` full width half at half maximum value (in arc minutes). The user should, for the majority of the foreseeable `caravela` applications, make sure that this parameter and the `IMAGE_WAVELENGTH_MICRONS` (the first parameter from this second set of variables) are consistent. These pair of parameters define the same instrument.

The next group of input parameters define the Galactocentric rings that can be considered in each `caravela` run. These are **not** the Galactocentric object distribution rings but refer to ring images described in this chapter, i.e. the sources can be distributed either in a spiral or ring 3D distribution and the ring images, that these parameters refer to, can still be generated. The first parameter here, `RING_IMAGES` is a flag variable that if these ring images are to be generated or not (1 for no and 0 for in the positive case). The following 12 `RING_ring_number_MIN` and `RING_ring_number_MAX` type variables defined the start and end, respectively, of each of the 6 possible rings to be considered in the generation of the described ring images (in kpc).

<code>RING_IMAGES</code>	1.0
<code>RING_1_MIN</code>	0.1
<code>RING_1_MAX</code>	4.0

RING_2_MIN	4.0
RING_2_MAX	5.6
RING_3_MIN	5.6
RING_3_MAX	7.2
RING_4_MIN	7.2
RING_4_MAX	8.9
RING_5_MIN	8.9
RING_5_MAX	14.0
RING_6_MIN	14.0
RING_6_MAX	17.0

Finally, the presented summarised description of the `caravela` input parameter file ends with the mass function constants. `MF_A_1` and `MF_B_1` define the mass function two exponents: for $1.0 < m(M_{\odot}) < 2.4$ and $2.4 < m(M_{\odot})$, respectively. The former pair of parameters refers to population 1 sources. `MF_A_2` and `MF_B_2`, and `MF_A_3` and `MF_B_3` correspond to populations 2 and 3, respectively.

MF_A_1	0.3
MF_B_1	1.0
MF_A_2	1.2
MF_B_2	1.79
MF_A_3	1.2
MF_B_3	1.79

All these parameters are stored in the `caravela caravela_input_XX.txt` text file (XX is the wavelength band to be used).

2.6 A theoretical catalogue of High-mass star forming regions

This is the first section where some of the core physical and numerical features of *caravela* will be presented. *caravela*'s first step is to build a collection of objects to be observed. This set of synthetic high-mass star forming regions form the theoretical catalogue and is the starting point of the simulation.

The `caravela_cat_gen_v11.cc` file^P is responsible for reading in the input parameters and generating the output catalogue. The final catalogue is a table^Q listing all the properties of each individual source **and** the location of the source in the model Galaxy, i.e. the 3D distribution of the sources is contained in the output file.

2.6.1 The objects

Each *caravela* object represents a high-mass star forming region with the following properties:

- size^R,
- spectral energy distribution (SED),
- mass.

Three different types of objects can be selected by the user of *caravela*: compact objects, grey-body objects, grey-body *clouds* with compact objects inside. Any of these three populations can be excluded setting the correspondent input variable to nought, in the input file `caravela_input_waveband.txt`.

^P2,000 lines of C++ code.

^QThe user is able to choose between ASCII or FITS format.

^RThe size is a linear size scale of the region/object, assumed to be spherical.

Size

The physical size of each object is sampled from the size scale distribution. The distribution used here is the Gaussian distribution with centred on:

`OBJS_LINEAR_SIZE_SCALE_MU_POP_#_AU`,

and with standard deviation `OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_#_AU`.

The objects' size distribution to be used is chosen by the user: for population 2 objects, the grey-bodies, and the outer envelopes of population 3 objects, the user defines the values for `OBJS_LINEAR_SIZE_SCALE_MU_POP_#_AU`, `OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_#_AU`, in AU (as with all user definable input values, these are defined inside `caravela_input_waveband.txt`. # is the population number, i.e. either 2 or 3). `OBJS_LINEAR_SIZE_SCALE_MU_POP_#_AU` and `OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_#_AU` are the mean value and standard deviation of the size distribution for the population, respectively. The `gasdev()` (Press *et al.*, 2002, page 292) routine returns a normally distributed deviate with $\mu = 0$ and $\sigma = 1$. To generate the correct distribution, the $N(0, 1)$ normal distribution is transformed in the desired $N(\mu, \sigma)$ distribution by $\mu + \sigma \times N(0, 1) = N(\mu, \sigma)$, i.e.:

```
linear_size_scale = linear_size_scale_mu_pop_2 +
linear_size_scale_sigma_pop_2 * NR::gasdev(idum_1);
```

For population 1 and 2, the compact young stellar objects and the core of population 3 high-mass star forming regions (Robitaille *et al.*, 2006), respectively, the linear size scale assumed by `caravela` is twice the value of the maximum between the outer envelope radius and the outer disk radius^s used in the Whitney *et al.* (2003) radiative transfer models.

^sModels with no disk are treated as having outer disk radius zero.

As it will become clear next, the sources' sizes are crucial both in determining their sizes in the image, and their grey-body masses (for populations 2 and envelope of population 3 objects).

SED

caravela produces images and point-source catalogues at 10^6 distinct wavebands (from IRAS $12\ \mu\text{m}$ to SPIRE $500\ \mu\text{m}$), therefore each object is associated with a spectrum. This is the spectral energy distribution, SED, of the object. The SED is an intrinsic property of the object, regardless of where it lies in the model Galaxy.

Objects belonging to populations 2 and the outer envelopes of the composite population 3 objects have their *intrinsic* SED given by the grey-body equation:

$$F_{e\nu} = B_\nu \times \left(1 - e^{-\left(\frac{\nu}{\nu_0}\right)^\beta} \right) \quad (2.1)$$

where $F_{e\nu}$ is the emitted flux at frequency ν , B_ν at frequency ν is the Planck function, and ν_0 is the frequency where the material is optically thin, and β is called the grey-body exponent. At this point it is unimportant if the object will be resolved or unresolved by the observer, hence the physical units of $F_{e\nu}$ are units of specific intensity, i.e. $\text{W m}^{-2} \text{Hz}^{-1} \text{sr}^{-1}$. If the source will be resolved, what is being measured by the observer is the specific intensity (and this is independent of the distance to the source). However, it is the radiation flux that is being measured for an unresolved source. As the source recedes farther and farther, the energy received from the entire source will disperse throughout the diffraction pattern (the Airy disc and rings) defined by the telescope's aperture. Because the light arriving at the detector leaves the surface of the source at all angles, the detector is effectively integrating over all directions: this is just the definition of radiative flux.

[†]12 considering the two SCUBA bands, $450\ \mu\text{m}$ to $850\ \mu\text{m}$

Note that the SEDs are integrated with the corrected transmission function, i.e. Robitaille *et al.* (2006) (hence population 1 and 3 objects) radiative transfer SEDs take into account the width and transmission function of the wavelength band for IRAS and Herschel filters.

For an unresolved source, the linear size of the source R and its distance d , are important. The relation between received and emitted fluxes is:

$$F_r = F_e \times \frac{R^2}{d^2} \quad (2.2)$$

In *caravela*, the values of ν_0 and β are defined by the user (by setting NU_0_MU_POP_#_HZ, NU_0_SIGMA_POP_#_HZ, BETA_MU_POP_#, and BETA_SIGMA_POP_#, where # is either 1 or 2).

The values of d will become determined only when the spatial distribution of the objects will be made.

The SEDs of population 1 and the core of population 3 objects were obtained from radiative transfer models of high-mass star forming regions, by Robitaille *et al.* (2006)^u. There are 207100 objects in this database^v, and the objects are selected according to the mass distribution set by the user of *caravela*.

Robitaille *et al.* (2006) consist of a grid of radiation transfer models of axisymmetric young stellar objects (YSOs), covering masses from 0.1 to 50M_⊙. The models comprehend a wide range of high-mass star forming regions evolutionary stages, from early envelope infall stage to the disk-only stage. The set of models, that is integrated in *caravela* for populations 1 and 3, is made of ~ 20000 models with SED computed at ten different viewing angles, resulting in 200,000 *distinct* objects.

^uThe Herschel bands were kindly released via a personal email from the author. The other wavebands are freely available from the website.

^v20710 from ten viewing angles.

Mass

The masses of the objects are used to choose a realistic set of objects, i.e. the resulting *true* distribution of the high-mass star forming regions obeys a mass function defined by the user. The (final) observed distribution *can*, in principle, be different.

The grey-body objects (population 1 and 3) have their masses determined by (equation 1, Fontani *et al.*, 2005, assuming a gas-to-dust ratio of 100 and $\beta = 2$):

$$M(M_{\odot}) = 1.3 \times 10^{-3} \left(\pi I_{\nu} (\text{Jy sr}^{-1}) \left(\frac{R}{2} \right)^2 \right) \times \left(e^{\frac{h\nu}{kT}} - 1 \right) \left(\frac{\nu}{2.4 \text{ THz}} \right)^{-3-\beta} \quad (2.3)$$

where R is the linear size of the object in kpc. This equation is independent of the temperature since

$$I_{\nu} = B_{\nu} \times \left(1 - e^{-\left(\frac{\nu}{\nu_0}\right)^{\beta}} \right) \quad (2.4)$$

All the objects based on Robitaille *et al.* (2006) (i.e. populations 1 and 3) have their masses assigned from the radiative transfer model (mass of the dusty envelope/disk added to the mass of the central source).

The mass function used is based on the IMF found in Nutter and Ward-Thompson (2007). The authors claim to have observed a Salpeter-like (i.e. a mass function obeying the $\zeta(\log m) \propto m^{-x}$) mass function for clumps (the CMF). From Chabrier (2003), the mass function was defined by Salpeter (1955) as the number of stars N in a volume V observed at time t per logarithmic mass interval $d \log m$ (n is the stellar number density):

$$\zeta(\log m) = \frac{d \left(\frac{N}{V} \right)}{d \log m} = \frac{dn}{d \log m} \quad (2.5)$$

The mass function can also be defined as:

$$\zeta(m) = \frac{n}{m} = \frac{1}{m \ln 10} \zeta(\log m) \quad (2.6)$$

In *caravela*, the mass function is assumed to be a power-law, i.e. a Pareto function:

$$\zeta(\log m) \propto m^{-x} \quad (2.7)$$

The problem that *caravela* needs to solve is how to sample a value of mass according to the power-law mass function chosen by the user. In other words, numerically, how do we sample a set of random numbers from *any* probability distribution function (only the uniform and Gaussian are usually available in the Mathematics libraries). The technique implemented in *caravela* is the Inverse transform sampling. This method enables one to generate any number of numbers (in our case the high-mass star forming regions masses) from the inverse of the cumulative distribution (for the Pareto distribution the inverted distribution is

$$T = \frac{b}{U^{\frac{1}{a}}} \quad (2.8)$$

where a and b are normalisation constants and U is an uniform deviate between 0 and 1). Given a continuous uniform variable U in $[0, 1]$ and an invertible cumulative distribution function F , the random variable $X = F^{-1}(U)$ has distribution F .

Therefore, *caravela* uses the above equation and the NR uniform random number generator to produce a set of masses to be used. More precisely, this technique is used to calculate the number of sources in each mass bin (of the distribution). Then, the objects are generated until all the mass bins are filled with the correct number of sources. In this way, each source has a mass and SED that are physically compatible. Although the grey-bodies and the radiative transfer sources have their

masses calculated in fundamentally different ways, the former technique assures that the SED, masses and sizes are physically sound amongst them. This means that the set of selected objects obeys a physical mass function.

Numerically, equation 2.8, page 44, is:

```
Mass = (MF_b) / (pow((NR::ran2(idum_2)), (1.0/MF_a)));
```

2.6.2 3D distribution

The set of objects described in the previous sections still does not have 3D properties, i.e. the objects are not distributed in space. The distribution of the objects is configurable by the user in the following ways:

- all sources are distributed in a ring centred on the Galactic centre.
- all sources lie along a number (one, two, three or four) of spiral arms.
- a fraction of the total sources forms a ring structure and the rest form the spiral arms(s).

This flexibility could perhaps be useful to study the issue of the number (and location) of spiral arms in the Milky Way.

It has been suggested in the literature that (high mass) star formation regions could form 3D ring structures, *tori*, with kpc length scales. In this configuration, the objects are distributed in a toroidal region centred at a radius from the centre of the Galaxy, with a Gaussian width. The radius and the Gaussian thickness, both in the direction parallel to the Galactic plane and perpendicular to it, of the distribution are the relevant input parameters here. The user also has the possibility of creating several of these rings with separate populations of sources living in each ring. Alternately, all the different populations can be mixed within the ring structures.

The physical coordinates that best explore the symmetry of the problem, hence simplifying the calculations, during this stage are the galactocentric coordinates: r , ϕ and z , the distance to the centre of the Galaxy, the angle measured in the Galactic plane from the Sun's direction, and the height from the Galactic plane, respectively.

The creation of the rings of high-mass star forming regions, numerically, is simple. To each object is assigned a value of r using a Gaussian distribution centred at the ring radius, the ϕ distribution is uniform form $[0, 2\pi]$ radians and z is Gaussian distributed using the z scale height given by the user (symmetrically to the Galactic plane).

No clustering is assumed between each pair of sources, i.e. all the objects' 3D positions are independent. Note that having a random alignment of a pair of population 1 and population 2 objects is not the same as having a single population 3 object, although these alignments may occur^w. Russeil (2003) spiral arms model is included in *caravela* insofar as the spiral arms are logarithmic.

From extragalactic observations, it is accepted that massive star formations traces the spiral structure of spiral galaxies. Assuming the Milky-way is no exception, it was crucial that *caravela* could be used to study the spiral structure of the Galaxy.

There is no consensus (Benjamin *et al.*, 2008) on the characterisation of the spiral structure of our galaxy (it is easier to determine the number of spiral arms in extragalactic objects than in our galaxy). The model we adopted is the Russeil (2003) spiral model: four logarithmic spiral arms with 14 degree pitch angle.

How could we distribute the objects along the spiral arms? *caravela* needs to ensure both that:

1. there is a uniform distribution of sources along the spiral arm, i.e. if the spiral

^wThese alignments could be determined in *caravela*.

arm is forced to become a straight line there should not be any noticeable concentration of sources at any point along the straight line.

2. the spiral arm cannot flare, i.e. the width of the cross section has to be constant from the inner to the outer regions of the Galaxy.

The logarithmic (or equiangular) spiral is defined by:

$$r(\theta) = a e^{b\theta} \quad (2.9)$$

where (r, θ) are the standard 2D polar coordinates (2D is sufficient in the following discussion since the z component is trivial). r is the distance to the centre of the Galaxy and θ is the angle measured from the Sun's direction. The constants a and b have the following physical interpretation in this context: a is the galactocentric distance where the arm begins, and, b determines the spiral pitch angle. a corresponds to half the size of the Milky Way central bar.

Figure 2.3, page 48, illustrates a logarithmic spiral distribution of sources (*left* panel), and only those sources who lie inside the solid angle defined by the user (*right* panel). The latter are the source candidates to feature in the final **cupid** catalogue.

In 2D, hence working only on the Galactic plane, the Cartesian and polar coordinates exhibit the following relations:

$$x(t) = r \cos t \quad (2.10)$$

and

$$y(t) = r \sin t \quad (2.11)$$

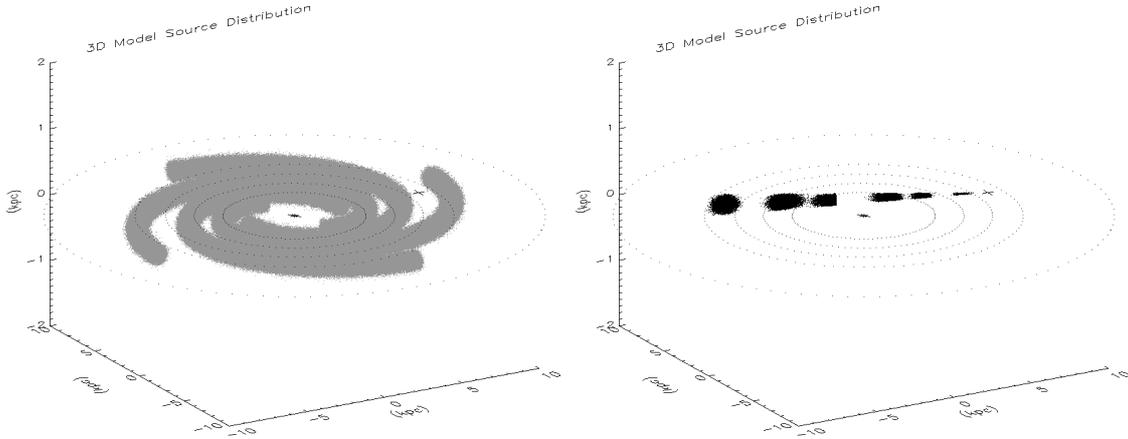


Figure 2.3: The figure illustrates a logarithmic spiral distribution of sources (*left* panel), and only those sources who lie inside the solid angle defined by the user (*right* panel).

t is a parameter used to parametrise equations of the spiral arm (let $\bar{c}(t)$ be the curve representing the equiangular spiral), i.e.:

$$\bar{c}(t) = (x(t), y(t)) = (r \cos t, r \sin t) = (a e^{bt} \cos t, a e^{bt} \sin t) \quad (2.12)$$

To tackle point 1, i.e. to ensure a uniform distribution of sources along the length of the spiral arm, the naive approach of generating uniform random numbers for t and then determining $(x(t), y(t))$ is wrong. The objects would concentrate more at the start of the arm because the parameter t does not walk along the arm at a constant speed. In other words, the curve t is not the natural parametrisation of the curve.

`caravela` implements the natural parametrisation of each spiral using the arc-length of the curve $\bar{c}(t) = (a e^{bt} \cos t, a e^{bt} \sin t)$:

$$s(t) = \int_{t_0}^t \|c'(\tau)\| d\tau \quad (2.13)$$

Note that $\|c'\|$ is the modulus of the derivative of \bar{c} , i.e. the speed at which we run along the curve. All we are doing is to find the exact parametrisation with speed one, this will mean that `caravela` can simply use a uniform random number generator, between $[0, 1]$, and multiply it by the length of the spiral arm (in kpc). The result will be a distribution along the arm with no regions of statistically significant over density of sources.

To determine the natural parametrisation we need to find $s(t)$, equation 2.13:

$$\bar{c}(t) = (a e^{bt} \cos t, a e^{bt} \sin t) \quad (2.14)$$

so,

$$\frac{d\bar{c}(t)}{dt} = \bar{c}'(t) = (a b e^{bt} \cos t - a e^{bt} \sin t, a b e^{bt} \sin t + a e^{bt} \cos t) \quad (2.15)$$

Using the Pythagoras theorem to find the modulus,

$$\|\bar{c}'(t)\| = \sqrt{(a b e^{bt} \cos t - a e^{bt} \sin t)^2 + (a b e^{bt} \sin t + a e^{bt} \cos t)^2} = a e^{bt} \sqrt{b^2 + 1} \quad (2.16)$$

Finally, we can insert this result into equation 2.13 and,

$$s(t) = \int_{t_0}^t (a e^{b\tau} \sqrt{b^2 + 1}) d\tau = \sqrt{b^2 + 1} \frac{a}{b} (e^{bt} - e^{bt_0}) \quad (2.17)$$

Solving the above equation in order to isolate t , we get,

$$t = \frac{1}{b} \ln \left(\frac{s}{\sqrt{b^2 + 1} \frac{a}{b}} + e^{bt_0} \right) \quad (2.18)$$

and this what is implemented in the code:

```
// select a point in the spiral parametrised by t.
```

```

s = 0.0 + NR::ran2(idum_3) * spiral_length_pop_1;
// now use the natural parametrisation
t = ( 1/b ) * log ( (s / (sqrt(b * b + 1)* a / b )) + exp ( b * 0.0 ) );

```

Therefore, for each random number generator `caravela` finds a x_c and y_c from the centre of the spiral.

We have solved and implemented the uniform distribution of sources along the spiral arm. The final *normalisation* (second issue listed in page 46) is to prevent the spiral arms from flaring, i.e. ensure a constant cross section along the arm. The aim is to find x, y (Cartesian coordinates in the Galactic plane) where to place each object. x and y will, in general, be different from x_c and y_c , i.e. the object will not lie exactly on the spiral curve but it will have, in general, a non-zero perpendicular distance to the curve.

At each point (x_c, y_c) , the tangent vector is $c'(t)$ determine in equation 2.15, page 49. The perpendicular vector $p(t)$ is needed give the perpendicular direction at the point. $p(t)$ is then,

$$p(t) = (-a b e^{bt} \sin t - a e^{bt} \sin t, a b e^{bt} \cos t - a e^{bt} \sin t) \quad (2.19)$$

The second normalisation is used in `caravela` to prevent flaring of the spiral arms. It can also be used to insert a controlled degree of flaring.

In summary, `caravela` uses the speed one normalisation of the spiral curve to find a point on top of the spiral arm. It then uses the direction of the perpendicular vector to determine the final (x, y) coordinate of the object. The process is repeated for each spiral arm to be built (this is defined by the user). The arms are equally spaced between them. The result is n spiral homogeneous non flaring spiral arms, extending from the two input parameters `SPIRAL_A_ARM` to `SPIRAL_LENGTH`.

2.6.3 Galactic structure model

In *caravela* the user is able to choose his favourite large scale pattern for the Milky Way.

If a spiral structure is selected, the following input parameters are needed: *number of spiral arms, spiral arms pitch angle, and, spiral arms starting position*^x.

The spiral nature of our Galaxy has been accepted for a long time, however its precise design is still widely discussed. Indeed, neither the number of arms nor their pitch angle are yet well defined. In the models presented in this work, *one* Galactic model was adopted: the Russeil (2003) best fit model. The latter is simply described as a four arm logarithmic spiral structure with 14 degrees pitch angle.

^xEach spiral is symmetric to all the others, i.e. it is rotated by 90 degrees. Also, the arms always start with direction tangent to a galactocentric circumference with radius defined by the arm starting position. As noted by M. Silva (private communication), these are limitations of the Galactic model used in *caravela*.

2.7 Simulated observation

In the context of a *caravela* run, a simulated observation is the generation of physically realistic images. In other words, *caravela* simulates what a telescope in the Sun's position would observe if it was embedded in the model Galaxy described in the previous sections.

So far, the code has produced a 3D distribution of objects (with sizes, masses and SEDs). This is one possible representation of the high-mass star forming regions in the Galaxy, obeying the input parameters given by the user in the input file. The distribution of objects, i.e. the set of synthetic sources, is read in. The input to this *caravela* module is the theoretical catalogue. Note that the user could, in principle, have generated the synthetic catalogue not using the first part of *caravela* but using *any* program of its preference (or he could test the code by giving a real observed distribution as input), so long as the Galactocentric positions of the objects are given.

The observation part of *caravela* has three main parts:

1. projection of the sources in the CCD plane and selection of sources inside each individual image,
2. generation of the pre-telescope image (this stage includes giving a morphology to the sources, i.e. an specific intensity profile),
3. instrument simulator (convolution, noise level and noise fluctuations),

The positions of the objects in the theoretical catalogue are in galactocentric coordinates. Part 1 is to correctly select the sources that exist in the solid angle defined by the image central position and image size. Both these values are defined by the user (often the user selects a large area of the sky to be observed so the image

central position and image size refer to the smaller individual images that form the requested composite images). `caravela` uses the SLALib (c version) functions to convert galactocentric coordinates to X, Y coordinates in the CCD instrument plane, and galactocentric coordinates to standard Galactic coordinates (galactic latitude and longitude, (ℓ, b)). Note that there is always an error associated with this projection. This error has a minimum value in the centre of the image and increases towards the edge of the image. `caravela` (through the SLALib functions) takes into account this error and rejects all the source with large errors (and anti-stars^y, of course).

The observation is assumed to be carried out from the adopted Earth position of $(r, \phi, z) = (8 \text{ kpc}, 0, 0)$, i.e. in the galactic plane at 8 kpc from the galactic centre. The observation is defined by the central direction and size of the field of view. Which sources lie within each observation is not trivial and involves geometrical projection between two coordinates systems. As described before, the galactic coordinates, (ℓ, b) , of the sources are known. It is, however, incorrect to define which sources are inside a specific field of view by including all the sources within a defined range in (ℓ, b) , e.g. $(\ell \pm \Delta, b \pm \Delta)$. It is more correct to map the galactic coordinates onto tangent plane coordinates. This projection is included in the program using the C version of the SlaLib package kindly sent by P.T. Wallace. In this way, the spherical coordinates (ℓ, b) are transformed into (X, Y) plane coordinates. (X, Y) result from a gnomonic projection, having the contact point between the surface of the sphere, where (ℓ, b) are defined, and the tangent plane, where (X, Y) exist, defined by the central direction of the observation, (ℓ_0, b_0) .

To convert from Galactocentric coordinates r, ϕ, z to Galactic coordinates, galactic longitude and latitude, (ℓ, b) , `caravela` uses the cosine rule to the trian-

^yAnti-stars are spurious objects resulting from projecting the real object on the incorrect hemisphere

gle defined by the Sun, the Galactic centre and point A , which is the position of the orthogonal projection, to the Galactic plane, of the source position.

Let a be the length of the side between the Sun and A , k the length of the side between the galactic centre and A , and, m is the distance from the Sun to the centre of the Galaxy. Then,

$$k^2 = a^2 + m^2 - 2 a m \cos \ell \quad (2.20)$$

and, applying the cosine rule again to the same triangle,

$$a^2 = k^2 + m^2 - 2 a m \cos \phi \quad (2.21)$$

The calculation of k is simple because, by construction, $k = r$. Now we can calculate a using equation 2.21, and, finally, ℓ and b are calculated by,

$$\ell = \arccos \left(\frac{a^2 - k^2 + m^2}{2 a m} \right) \quad (2.22)$$

$$b = \arctan \left(\frac{z}{a} \right) \quad (2.23)$$

The right ascension and declination are calculated from the galactic coordinates, applying the cited coordinate conversion library directly to (ℓ, b) .

As the sources are distributed in three dimensions, it is possible to determine the heliocentric distance to the source, d . If the source is in the galactic plane, i.e. $z = 0$, then d is,

$$d = a \quad (2.24)$$

if not, then,

$$d = \frac{z}{\sin b} \quad (2.25)$$

The determination of d is crucial to predict observable fluxes using this method.

The projection takes into account the presence of spurious anti-stars, i.e. stars that, due to periodicity of the trigonometric functions, involved in the projection, may appear to lie inside the image but are, in fact, in diametrical positions. Anti-stars are removed.

This is the inverse method of that included in the analysis of all astronomical observations, i.e. the (X, Y) observed in the detector plane is transformed to astronomical coordinates by the inverse of the projection described.

(X, Y) is expressed in pixel coordinates defining a pixel to degrees scale. The limits of the field of view to be observed are defined by the projection (gnomic or tangent plane projection) of the four vertices $(\ell_0 \pm \Delta, b_0 \pm \Delta)$, i.e. sources with projected (X, Y) within $(X_0 \pm \Delta_{\text{proj}}, Y_0 \pm \Delta_{\text{proj}})$.

It was decided that data would be generated and exchanged between modules in FITS format (Wells *et al.*, 1981), therefore the program generates both these outputs in FITS format. This allows data (tables and images) to be exchanged between modules in a standardised and machine-independent way. This is crucial due to the modular nature of the simulation, in particular during the input-output interfaces of the modules. This follows other astronomical simulations (Reinecke *et al.*, 2006) and has the great advantages over other image and table storing formats that it is widely used within the community, and there are numerous programs that are able to open and analyse FITS files.

Part 2 of the simulated observation module in `caravela` is to actually generate

the images. Up to this point, the code knows which of the sources lie inside the solid angle defined by each individual image, so it is finally possible to start to populate the 2D matrix that is the image.

Images are implemented numerically as 2D matrices, where each pixel corresponds to a matrix element, which can be identified by two indices, (i, j) . The matrix is flat, in the sense that it does not follow the curvature of the sphere where (ℓ, b) are defined, therefore a projection onto a flat plane is needed. This is implemented as described previously, so (X, Y) correspond, for each source, to (i, j) .

For each waveband, **caravela** generates a super resolution image, i.e. an image with very high spatial resolution. This high resolution is, in principle, much better than the characteristics of any instrument. These images, although expensive memory wise, are prove to be of great value for further analysis. It is possible that the user wants to use a better and more sophisticated instrument simulator (e.g. the official instrument teams of Herschel), so these images can be used as input to different simulators (and not only to **caravela**'s)

A simulator in this context is a computer program designed to provide a realistic imitation of the controls and operation instrument, i.e. map scanning direction and speed, and CCD properties.. Note that the output from **caravela** can be stopped here if different simulators want to be used or tested. This modular approach is important and it allows the SPIRE and PACS instrument teams to use the output from **caravela** at this point, and test it with their official simulators. This work already started and is ongoing under the umbrella of the Hi-GAL Simulation working group. The output from this is then fed into the map making algorithms (The Hi-GAL consortium is using madmap as main map making procedure. **caravela** alone does not consider the map-making phase of the simulator.

Each source (that its angular size is larger than the pixel size defined the user)

has a Gaussian symmetric intensity profile, in (X, Y) .

One key feature of this work is that there are no ambiguities in the determination of the distances to the sources (even for sources inside the solar circle), thus, the observed fluxes can be calculated and used to generate these images. This is a fundamental difference between *caravela*'s simulated observation and a real observation.

The super resolution images do not have any of the observational artifacts (e.g. diffraction spikes, instrumental noise and confusion) present in *caravela*'s final images. They correspond to an idealisation: images before the telescope observes them.

As a side product, the code generates a set of ring images. These are 6 images per each super resolution images, that correspond to observations of isolated galactocentric rings (these are not the rings discussed just before the spiral arms description), i.e. each ring image is observed as if only the sources with galactocentric distances within a minimum and maximum distance to the Galactic centre exist. The aim of these images is then to be coupled with the diffuse emission images created by Paladini *et al.* (2007). The inversion technique used in the cited paper is forced to assume that the diffuse emitting material is distributed in Galactocentric rings. This limitation arises from the well documented distance ambiguity that exists for sources inside the solar ring. These ring images correspond, *ring by ring*, to Paladini *et al.* (2007) images and can be combined to yield a compact sources plus diffuse emission image.

Part 3 of the observation module is to submit the super resolution images to the *caravela* instrument simulators. The instruments available are: the IRAS bands, the SPIRE and PACS bands, and the SCUBA-2 bands. Each image is convolved with the correspondent Point Spread Function (PSF) and instrument's noise level

and fluctuations are also added. For the four IRAS bands and the two SCUBA-2 bands, the convolution is extremely naive: a Gaussian kernel with full width half maximum (FWHM) of the size of the instrument resolution is applied to the ideal image. The Herschel bands errors should be reduced by a factor of square root of 2, if the user wants to simulate the parallel mode as used in Hi-GAL. This is the case since in parallel mode the same area of the sky is observed for longer and more times than in non-parallel mode.

For the SPIRE and PACS (the Herschel Space Observatory instruments) wavebands, the proper synthetic PSF used by the both instrument teams is used in *caravela*. Aesthetically this results in very interesting patterns of diffraction spikes. Physically, the confusion levels and the obstacles put by this realistic convolutions to the source extraction (and subsequent photometry) should add to the user confidence in the final results. This is illustrated in figures 2.4 and 2.5, pages 59 and 60, respectively.

The result of the *caravela* instrument simulators is the set of the code most realistic images (WCS keywords are included in the FITS headers, to allow further analysis by the user of the images using GAIA or DS9). Due to the realism of the final images, not all objects that existed inside the solid angle defined by the user (i.e. the image) from the theoretical catalogue will be retrieved during the next stage of the simulation: source detection and photometry.

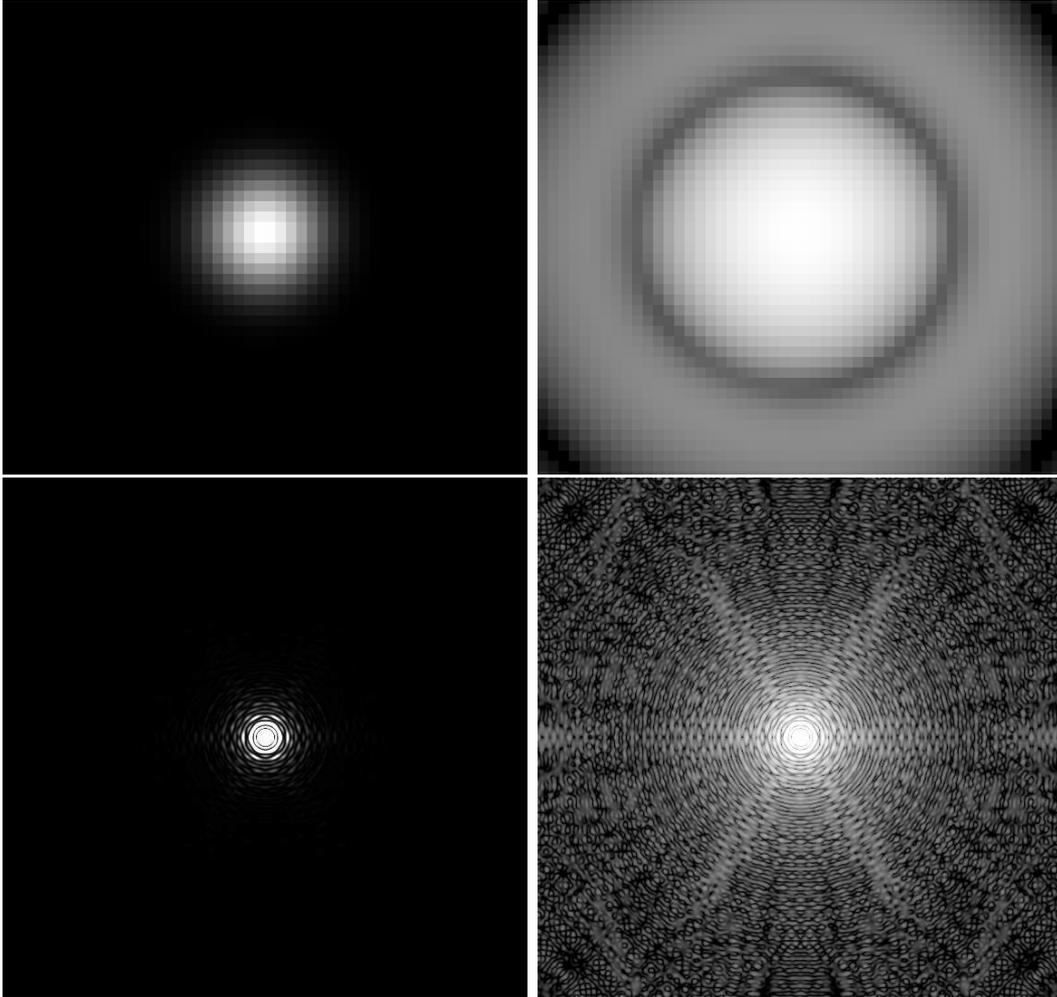


Figure 2.4: PACS and SPIRE simulated PSFs. Top panels show the simulated psf at $350\ \mu\text{m}$ used in *caravela*. This correspond to the SPIRE instrument. Images are $0.027^\circ \times 0.027^\circ$ in area (pixel size is $2''$). *Top Left*: 98% of the integrated values lie in the central region corresponding to a FWHM of $24''$ resolution. *Top Right*: The Airy rings are clearly visible as this is in logarithmic scale. Bottom panels show the simulated psf at $110\ \mu\text{m}$ used in *caravela*. This correspond to the PACS instrument. Images are $0.255^\circ \times 0.255^\circ$ in area. (pixel size is 0.000222°). *Bottom Left*: 98% of the integrated values lie in the central region corresponding to a FWHM of $7.7''$ resolution. *Bottom Right*: Diffraction spikes are clearly visible as this is in logarithmic scale.

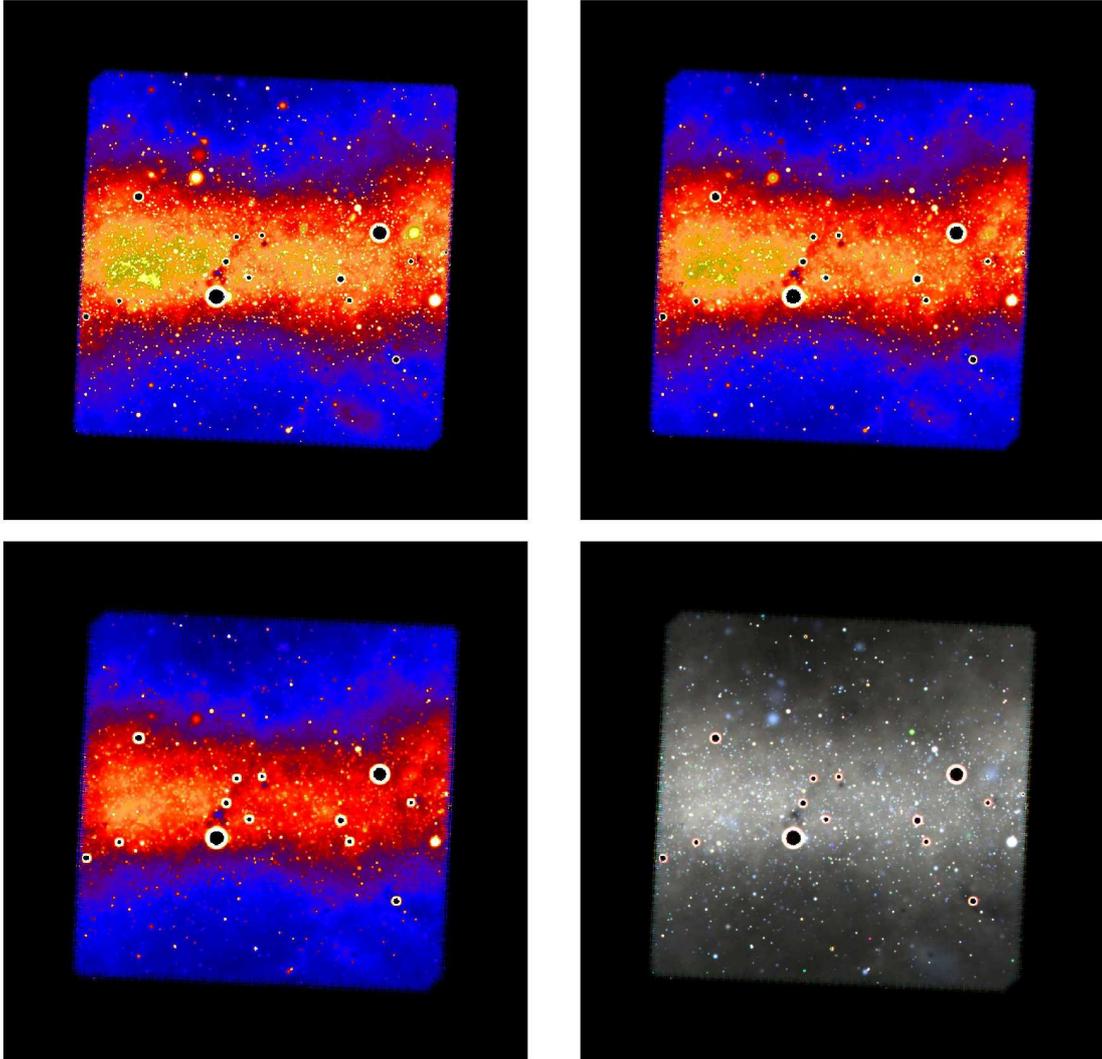


Figure 2.5: Output from the SPIRE simulator. The nominal voltage bias used was 21.2 mV. and the maps were made using madmap. The region scanned was $2^\circ \times 2^\circ$, and 2 cross scans were performed. The top left, top right, bottom left, and bottom right are the 250 μm , 350 μm , 500 μm , and a three colour combined image, respectively. See text for further details.

2.8 Photometry

The aim of `caravela` is to produce point source catalogues of high-mass star forming regions. The simulation can, at this point, produce catalogues (one per each wavelength band). The method used is aperture photometry on the final images generated.

The code implements CUPID as the starlink software package responsible for the photometry. CUPID (CIUmP IDentification) can be used for identifying clumps of emission in 1D, 2D or 3D data arrays. Whilst primarily targeted at the needs of the SCUBA2 advanced data products pipeline, it is a perfect tool to be used within `caravela`. It enables an excellent comparison between different source extraction algorithms using the `caravela` final images as test images, as well as, produces the final point source catalogue.

The main `cupid` input parameter `decision` consists of what source detection and photometry algorithm to use. As described in section 2.4, page 33, the user provides a keyword, at the instant `caravela` is set up, that defines if `clumpfind`, `fellwalker`, or `extractor` is to be used.

2.9 Simulated observations using `caravela`

The aim of this section is to test `caravela` using a realistic set of input parameters (table 2.1, page 62). These input parameters are an educated guess at the total number of sources, spectral energy distribution of sources, and 3D source distribution. The physical assumptions behind the numbers used are explained in the subsequent sections. Two images are presented as an illustration of the `caravela` output (figure 2.6, page 66).

The model described next is a sample model, i.e. `caravela` is being put to use

Table 2.1: Input parameters for the IRAS simulation

Parameter	Value	Units	Description
n_{spiral}	4		number of spiral arms
a_{spiral}	3	kpc	start of the spiral arms ^a
b_{spiral}	14	°	spiral arms pitch angle ^b
n_{I}	10^6 ^c		number of isolated compact sources
l_{I}	5.0×10^6		linear size scale
x_1	0.3		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	1.2		mass function exponent for $2.4 < m(M_{\odot})$
n_{II}	0.0		number of isolated extended sources
l_{II}	–		linear size scale
x_1	–		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	–		mass function exponent for $2.4 < m(M_{\odot})$
T_{II}	–		Grey-body temperature
β	–		Grey-body β parameter
n_{III}	0.0		number of embedded sources
l_{III}	–	AU	linear size scale for the envelope sources
x_1	–		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	–		mass function exponent for $2.4 < m(M_{\odot})$
T_{III}	–	K	Grey-body temperature for the envelope sources
β	–		Grey-body β for the envelope sources

^aThis is parameter a in equation 2.9, page 47, and corresponds to half the size of the Galactic bar used.

^bFrom Russeil (2003) best-fit model.

^c§2.9.1 for the more details.

with realistic input parameters for the first time during this work.

2.9.1 The input parameters

Table 2.1 presents the input parameters for the simulation.

The first three parameters define the 3D distribution of the objects. The objects in this simulation are distributed in $n_{\text{spiral}} = 4$ spiral arms. Each spiral arm starts at $a_{\text{spiral}} = 3$ kpc from the Galactic centre, and has a pitch angle of $b_{\text{spiral}} = 14^\circ$. These three values correspond to the best fit model for the Milky Way found by Russeil (2003). These three parameters are kept constant in all simulations presented in this work. This is because we feel that the cited model can be trusted in the context of this work. Also, the number of free parameters is reduced using in this approach.

Physical sizes, masses and SEDs

As explained in the previous chapter, the physical size distribution of the objects to be observed is one of the parameters set by the user before each *caravela* run. In this case, we have used $l_{\text{I}} = 5.0 \times 10^6$ AU (with a standard variation of 1%). This is based on an average size for the population I objects used. These objects are sampled (according to a mass function) from Robitaille *et al.* (2006) radiative transfer models.

Only population I objects are used in this simulation, therefore the SEDs are sampled from Robitaille *et al.* (2006) considering the user defined x_1 and x_2 mass function exponents for $1.0 < m(M_{\odot}) < 2.4$ and $2.4 < m(M_{\odot})$, respectively.

Galactic distribution of sources

The total number of sources used in this simulation is $n_{\text{I}} = 10^6$. As with any *caravela* run, this is a key parameter. Why $n_{\text{I}} = 10^6$?

The argument is as follows. If ρ is the number density of high-mass star forming regions in the Galaxy, and V is the total volume occupied by these regions, then the equation,

$$n_{\text{I}} = \rho \times V \quad (2.26)$$

gives the as estimate of the total number of sources.

To determine ρ we have calculated the number density of sources in the regions observed in Nutter and Ward-Thompson (2007). The authors have observed four regions in Orion: Orion AN, Orion AS, Orion BN, and Orion BS.

The volume of these four regions can be approximated as the region inside the solid angle observed considered only between the near and far distance, that is the thickness of the region being studied. The assumed near and far distances are

$d^n = 320$ pc and $d^f = 500$ pc, respectively (Nutter and Ward-Thompson, 2007). The solid angles α_i observed are also needed for the calculation, and this is by the sum of the four areas mapped in deg². For the four regions (in sr): $\alpha_1 = 7.0 \times 10^{-5}$, $\alpha_2 = 1.2 \times 10^{-4}$, $\alpha_3 = 9.7 \times 10^{-5}$, and, $\alpha_4 = 1.3 \times 10^{-4}$. This corresponds to 0.23 deg², 0.41 deg², 0.32 deg², and 0.45 deg².

The volume inside a solid angle α (in sr), cut off at the near distance d^n and far distance d^f is given by,

$$v = \int_0^{d^f} \int_0^\alpha r^2 d\Omega dr - \int_0^{d^n} \int_0^\alpha r^2 d\Omega dr = \frac{\alpha}{3} (d_f^3 - d_n^3) \quad (2.27)$$

Applying this equation to the four Orion regions, i.e. to α_i with $i = 1$ to $i = 4$: $v_1 = 2153.99$ pc³, $v_2 = 3839.72$ pc³, $v_3 = 2996.85$ pc³, $v_4 = 4214.32$ pc³. The total volume observed is $V^{\text{Orion}} = \sum_i v_i = 13204.9$ pc³.

The number of star forming regions candidates in the four volumes is $N^{\text{Orion}} = 120$ (only objects with $m > 1M_\odot$ are included). Therefore, the number density of sources for the four regions is,

$$\rho^{\text{Orion}} = 9.0 \times 10^{-3} \text{ pc}^{-3} \quad (2.28)$$

To estimate the total volume V representing *all* the regions where there is active star formation in the Milky-Way, I have determined the volume of four idealised (and equal) spiral arms. Each spiral arm is approximated by a cylinder with length $l = 30$ kpc and area of (circular) cross section given by a typical average width of a spiral arm $a = 0.05$ kpc. Then,

$$V = 4 \times l \times a \approx 10^9 \text{ pc}^3 \quad (2.29)$$

Assuming that ρ^{Orion} is representative of an average star forming region, a crude estimate of the total number of sources to be used in this *caravela* run is,

$$n_{\text{I}} = \rho \times V \approx 10^{-3}(\text{pc}^{-3}) \times 10^9(\text{pc}^3) = 10^6 \quad (2.30)$$

Considering the assumptions used, the estimate for the number of star forming regions in the Galaxy is $\approx 10^{6z}$. This is not the total number of sources *detected* (after photometry) at the end of the simulation.

2.9.2 Simulated images

IRAS wavebands

At the heart of *caravela* there is a Monte-Carlo technique, i.e. the objects properties (e.g. their positions in the 3D Galaxy model) result from stochastic distributions. That is to say that *caravela* is built to be used as a statistical tool. How many high-mass star forming regions are consistent with the real observed point source catalogues? What distribution of objects will reproduce exactly a given real image? Between the two questions, *caravela* is capable of addressing the former but is completely inadequate to answer the latter.

The *caravela* user cannot expect a one-to-one correspondence between observations and synthetic sources.

Figure 2.6, page 66, exhibits four panels comparing the synthetic *caravela* images with IRAS real images. The real and simulated images correspond to the input parameters listed in table 2.1, page 62. They represent a qualitative comparison between model and real data.

^zIf we consider the 3D porosity parameter (or volume ratio) $Q_{3\text{D}} = 0.4$ from Oey and Clarke (1997), the estimate for the total number of sources is reduced to $n_{\text{I}} = \rho \times V \times Q_{3\text{D}} \approx 10^5$

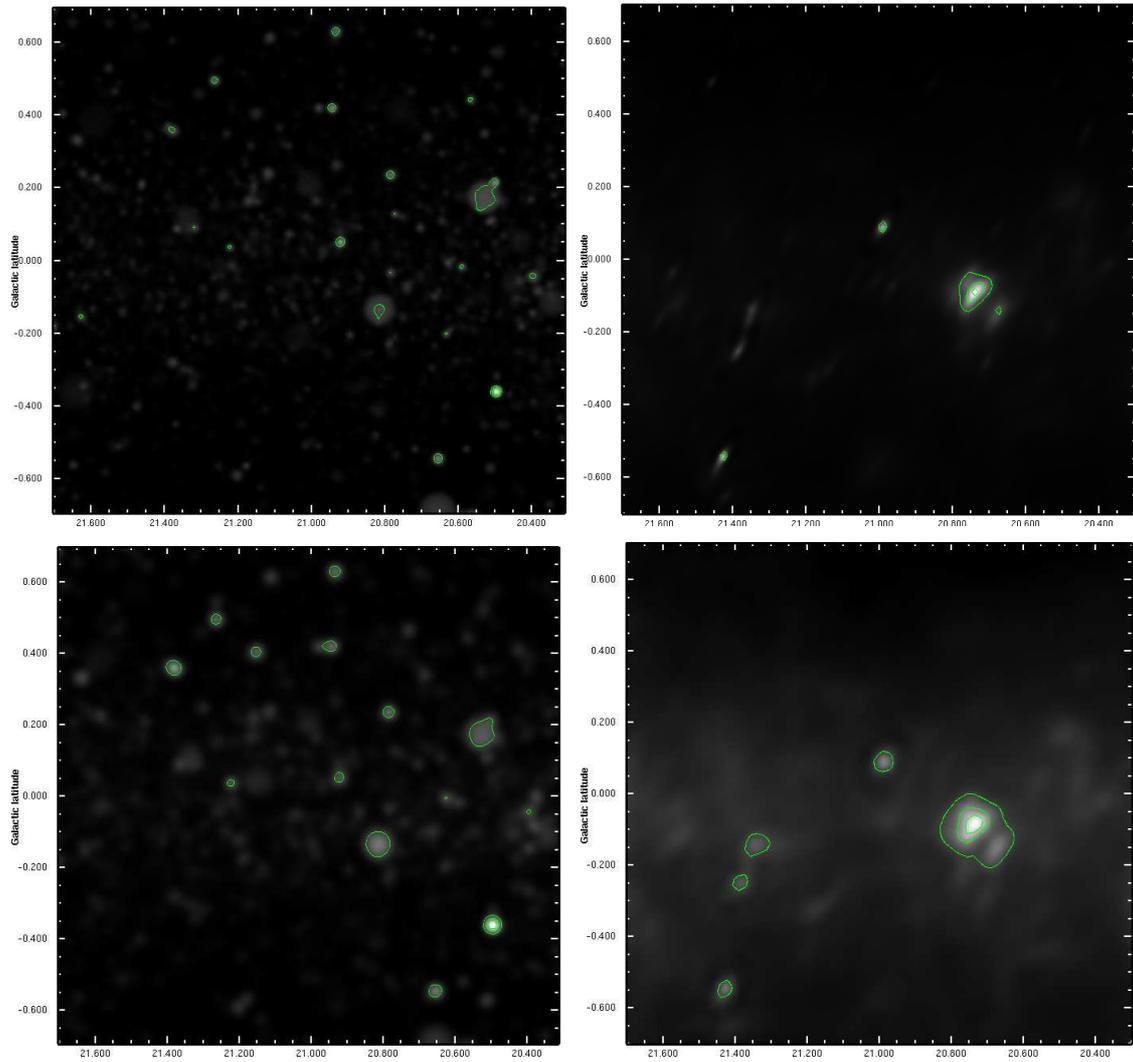


Figure 2.6: Comparison between IRAS and caravela images (3σ contours). Top panels show IRAS caravela image (left) and IRAS *real* image (right) at $60\ \mu\text{m}$. Bottom panels exhibit IRAS caravela image (left) and IRAS *real* image (right) at $100\ \mu\text{m}$.

2.10 Summary

Here the main aspects from `caravela` are put together in order to summarise the second chapter.

1. The fundamental physical assumptions underpinning `caravela` are
 - the interstellar medium is optically thin at the IRAS and Herschel wavebands,
 - high-mass star formation takes place, predominantly, in the spiral arms of the Galaxy,
 - the high-mass star forming regions are distributed in a statistically uniform way along the length of the Galaxy spiral arms (hence the two normalisations built-in the code and presented in section 2.6.2, page 45).
2. `caravela` is written in C++ supported by IDL, with an object oriented design. Two main reasons justify the use of C++: it is a fast and flexible programming language which allows a straightforward implementation of an object oriented design; and, it is increasingly used in similar scientific research projects (e.g. the Planck mission simulation pipeline Reinecke *et al.*, 2006).
3. The output from the code has two fundamental forms: images and photometric catalogues. The images are in FITS format and can be subject to independent further analysis using any FITS viewer, and immediately processed by `caravela` (e.g. through `cupid`) in order to generate *the* point source catalogue to be compared with the real observational one.
4. In practice, `caravela`'s input parameters are all defined by the user in a single parameter file (ASCII format). To start a `caravela` run the user types the command in a terminal. There is no graphical user interface at this stage.

5. The main stages of the code are:
 - building a theoretical catalogue containing both the objects intrinsic physical properties (i.e. mass, size and SED), and, their 3D spatial distribution attributes (i.e. Galactocentric 3D coordinates for each object). Three distinct populations of objects are available to the user: isolated compact objects, isolated grey-bodies, and, compound objects characterised by a central core surrounded by a grey-body extended cloud.
 - simulating an observation using the instrument selected by the user. The built-in *observatories* are IRAS, Herschel and SCUBA,
 - finally, a photometric analysis is performed on all images generated by the code. The end result is a photometric catalogue. For multi-wavelength simulations, a cross-match catalogue is also produced.

6. There are no ambiguities in the determination of the distances to the sources (even for sources inside the solar circle), thus, the observed fluxes can be calculated and used to generate these images. This is a fundamental difference between *caravela*'s simulated observation and a real observation.

7. *caravela* is capable of successfully dealing with physically realistic input parameters, e.g. total number of sources $\sim 10^6$. Assuming that ρ^{Orion} is representative of an average star forming region, a crude estimate of the total number of sources to be used in this *caravela* run is, $n_{\text{I}} = \rho \times V \approx 10^{-3}(\text{pc}^{-3}) \times 10^9(\text{pc}^3) = 10^6$. Considering the assumptions used, the estimate for the number of high-mass star forming regions in the Galaxy is $\approx 10^6$. This is not the total number of sources *detected* (after photometry) at the end of the simulation. If we consider the 3D porosity parameter (or volume ratio) $Q_{3\text{D}} = 0.4$ from Oey and Clarke (1997), the estimate for the total number of sources is reduced to $n_{\text{I}} = \rho \times V \times Q_{3\text{D}} \approx 10^5$.

Chapter 3

High-mass Star Formation as Seen by IRAS

3.1 Introduction

This chapter presents an example on how to successfully use `caravela` to reproduce existing observations from the IRAS telescope. We conclude that `caravela` can be used to gain insight into the star formation population that is contained in the IRAS point source catalogue.

The method behind the next chapters is to compare `caravela` against what is known and well studied first: the study of the distribution of UCH II regions by Wood and Churchwell (1989). The aim is to constrain a set of parameters, i.e. a model that is a more consistent against observational data sets. Then, the model will be used as an ingredient to develop a number of predictions of future observations of high-mass star forming regions in the Milky Way.

3.2 From synthetic IRAS colour-colour diagrams to a model for the UCH II regions in the Galaxy

In this section we present a more detailed example of the use of `caravela` to interpret real data. A sensible region of the parameter space is surveyed (refer to figure 3.13, page 96, corresponding to 21 independent models), and best-fit model is selected (using a 2D K-S test).

The aim is to use `caravela` to go from IRAS colour-colour diagrams to a model for the UCH II regions in the Milky Way.

This section uses an extra script to run four `caravela` models in the simple *parallel mode* described in the method chapter.

In this exercise, only the distribution of UCH II regions is being modelled. The simulated colour-colour diagrams are compared with the longstanding observational results presented in figure 1 of Wood and Churchwell (1989). Hence the total number of sources is of the order of 10^3 to 10^5 (Wood and Churchwell, 1989). This estimate for the total number of sources has two origins:

- the upper limit of 10^5 results from fine-tuning the model parameters during the run of several `caravela` models. In practice, these models formed the foundation for the models presented in tables 3.1, 3.2, and 3.3 (pages 73, 74, and 75). These 21 models are studied in detail next.
- the lower limit of 10^3 objects, corresponds to the estimation of Wood and Churchwell (1989). The authors refer to ≈ 1717 UCH II candidates using their colour-colour cut criterion.

3.2.1 Input Parameters

The 21 models studied belong to four distinct categories: population 1 only, population 2 and 3 only, population 1 and 3 only, and, finally, an ensemble of all populations.

The 3D spatial distribution for all 21 models is exactly the same as described in the previous section.

Due to the spiral nature of the `caravela` model (and the real Milky-Way) it is crucial that the models and the real data represent the same area of the sky. In this study, that is sources taken from a $2^\circ \times 2^\circ$ box in the plane of the Milky Way at 40° galactic longitude.

To compare the number of sources detected, the photometry and source detection algorithms used are important. CUPID is used in a `caravela`^a specific way when more than one waveband is analysed. Each run of `caravela` produces a point-source catalogue, but to create and understand correctly colour-colour diagrams, the photometry at each wavelength should correspond to the same source.

The process developed for `caravela` in multi-waveband mode is:

- a mask with the positions of the clumps in one single waveband image is created
- the sizes of the clumps are determined for the same image
- the `cupid::extractclumps` routine is applied to the complete set of images, using the position/size mask
- a final multi-wavelength catalogue is built with only confirmed detection at the four IRAS bands (the criteria used to validate an object was the same as the one used with real IRAS data)

^aIn fact, any multi-wavelength study uses this technique.

Tables 3.1, 3.2, and 3.3 (pages 73, 74, and 75) list the input parameters used in the 21 models studied.

Table 3.1: Input parameters for models $\mathcal{A}, \dots, \mathcal{G}$.

Parameter	Model \mathcal{A}	Model \mathcal{B}	Model \mathcal{C}	Model \mathcal{D}	Model \mathcal{E}	Model \mathcal{F}	Model \mathcal{G}	Description
n_{spiral}	4	4	4	4	4	4	4	number of spiral arms
a_{spiral}	3 kpc	3 kpc	3 kpc	3 kpc	3 kpc	3 kpc	3 kpc	start of the spiral arms ^a
b_{spiral}	14°	14°	14°	14°	14°	14°	14°	spiral arms pitch angle ^b
n_{I}	5.0×10^5	0.0	0.0	1.0×10^5	0.0	2.5×10^5	0.0	number of isolated compact sources
x_1	0.3	—	—	0.3	—	0.3	—	for $1.0 < m(M_{\odot}) < 2.4$ ^c
x_2	1.2	—	—	1.2	—	1.2	—	for $2.4 < m(M_{\odot})$ ^d
n_{II}	0.0	2.5×10^5	0.0	1.0×10^5	0.0	0.0	0.0	number of isolated extended sources
x_1	—	0.3	—	0.3	—	—	—	for $1.0 < m(M_{\odot}) < 2.4$
x_2	—	1.2	—	1.2	—	—	—	for $2.4 < m(M_{\odot})$
T_{II}	—	10.0 K	—	10.0 K	—	—	—	Grey-body temperature ^e
β	—	2	—	2	—	—	—	
n_{III}	0.0	2.5×10^5	5.0×10^5	3.0×10^5	5.0×10^5	2.5×10^5	2.0×10^3	number of compound sources ^f
x_1	—	0.3	0.3	0.3	0.3	0.3	0.3	for $1.0 < m(M_{\odot}) < 2.4$
x_2	—	1.2	1.2	1.2	1.2	1.2	1.2	for $2.4 < m(M_{\odot})$
T_{III}	—	50.0 K	10.0 K	10.0 K	50.0 K	50.0 K	50.0 K	Grey-body temperature of outer sources ^g
β	—	2	2	2	2	2	2	
N^h	479	191	514	430	524	516	2	Simulated sources detected by cupid
\mathcal{P}^i	29.85 %	70.15 %	46.49 %	28.37 %	77.09 %	57.55 %	100 %	Percentage of UCH II regions candidates
n^j	143	134	239	122	404	297	2	cupid UCH II regions candidates
2D K-S ^k test	1.37	9.96×10^{-4}	—	—	—	—	—	2D K-S test on the UCH II regions candidates

^aThis is parameter a in equation 2.9, page 47, and corresponds to half the size of the Galactic bar used.

^bFrom Russeil (2003) best-fit model.

^cCorresponds to x on equation 2.7, page 44.

^dBest fit for the mass function found by Nutter and Ward-Thompson (2007), x on equation 2.7, page 44.

^eEquation 2.1, page 41.

^fType III sources correspond to a type I source inside a type II.

^gEquation 2.1, page 41.

^hIRAS PSC observational value is 927. Wood and Churchwell (1989) observational value is 736.

ⁱPercentage of number of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 2.26 % .

^jNumber of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 21.

^k2D K-S in units of 10^{-2} .

Table 3.2: Input parameters for models $\mathcal{H}, \dots, \mathcal{N}$.

Parameter	Model \mathcal{H}	Model \mathcal{I}	Model \mathcal{J}	Model \mathcal{K}	Model \mathcal{L}	Model \mathcal{M}	Model \mathcal{N}	Description
n_{spiral}	4	4	4	4	4	4	4	number of spiral arms
a_{spiral}	3 kpc	start of the spiral arms ^a						
b_{spiral}	14°	14°	14°	14°	14°	14°	14°	spiral arms pitch angle ^b
n_{I}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	number of isolated compact sources
x_1	–	–	–	–	–	–	–	for $1.0 < m(M_{\odot}) < 2.4$ ^c
x_2	–	–	–	–	–	–	–	for $2.4 < m(M_{\odot})$ ^d
n_{II}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	number of isolated extended sources
x_1	–	–	–	–	–	–	–	for $1.0 < m(M_{\odot}) < 2.4$
x_2	–	–	–	–	–	–	–	for $2.4 < m(M_{\odot})$
T_{II}	–	–	–	–	–	–	–	Grey-body temperature ^e
β	–	–	–	–	–	–	–	
n_{III}	2.0×10^4	7.0×10^4	4.0×10^4	4.0×10^4	4.0×10^4	4.0×10^4	3.0×10^4	number of compound sources ^f
x_1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	for $1.0 < m(M_{\odot}) < 2.4$
x_2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	for $2.4 < m(M_{\odot})$
T_{III}	50.0 K	50.0 K	50.0 K	10.0 K	60.0 K	40.0 K	50.0 K	Grey-body temperature of outer sources ^g
β	2	2	2	2	2	2	2	
N^h	12	48	46	48	56	48	14	Simulated sources detected by <i>cupid</i>
\mathcal{P}^i	56.25 %	58.33 %	60.89 %	20.83 %	67.85 %	41.66 %	35.71 %	Percentage of UCH II regions candidates
n^j	5	28	28	10	38	20	5	<i>cupid</i> UCH II regions candidates
2D K-S ^k test	–	1.80	74.21	62.6	78.32	84.73	51.47	2D K-S test on the UCH II regions candidates

^aThis is parameter a in equation 2.9, page 47, and corresponds to half the size of the Galactic bar used.

^bFrom Russeil (2003) best-fit model.

^cCorresponds to x on equation 2.7, page 44.

^dBest fit for the mass function found by Nutter and Ward-Thompson (2007), x on equation 2.7, page 44.

^eEquation 2.1, page 41.

^fType III sources correspond to a type I source inside a type II.

^gEquation 2.1, page 41.

^hIRAS PSC observational value is 927. Wood and Churchwell (1989) observational value is 736.

ⁱPercentage of number of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 2.26 % .

^jNumber of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 21.

^k2D K-S in units of 10^{-2} .

Table 3.3: Input parameters for models $\mathcal{O}, \dots, \mathcal{U}$.

Parameter	Model \mathcal{O}	Model \mathcal{P}	Model \mathcal{Q}	Model \mathcal{R}	Model \mathcal{S}	Model \mathcal{T}	Model \mathcal{U}	Description
n_{spiral}	4	4	4	4	4	4	4	number of spiral arms
a_{spiral}	3 kpc	start of the spiral arms ^a						
b_{spiral}	14°	14°	14°	14°	14°	14°	14°	spiral arms pitch angle ^b
n_{I}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	number of isolated compact sources
x_1	–	–	–	–	–	–	–	for $1.0 < m(M_{\odot}) < 2.4$ ^c
x_2	–	–	–	–	–	–	–	for $2.4 < m(M_{\odot})$ ^d
n_{II}	0.0	0.0	0.0	0.0	0.0	0.0	0.0	number of isolated extended sources
x_1	–	–	–	–	–	–	–	for $1.0 < m(M_{\odot}) < 2.4$
x_2	–	–	–	–	–	–	–	for $2.4 < m(M_{\odot})$
T_{II}	–	–	–	–	–	–	–	Grey-body temperature ^e
β	–	–	–	–	–	–	–	
n_{III}	2.0×10^4	3.0×10^4	4.0×10^4	7.0×10^4	3.0×10^4	4.0×10^4	7.0×10^4	number of compound sources ^f
x_1	0.3	0.3	0.3	0.3	0.3	0.3	0.3	for $1.0 < m(M_{\odot}) < 2.4$
x_2	1.2	1.2	1.2	1.2	1.2	1.2	1.2	for $2.4 < m(M_{\odot})$
T_{III}	40.0 K	40.0 K	40.0 K	40.0 K	30.0 K	30.0 K	30.0 K	Grey-body temperature of outer sources ^g
β	2	2	2	2	2	2	2	
N^h	12	14	48	49	15	47	46	Simulated sources detected by <i>cupid</i>
\mathcal{P}^i	25.00 %	28.57 %	39.58 %	28.57 %	26.66 %	25.53 %	13.04 %	Percentage of UCH II regions candidates
n^j	3	4	19	14	4	12	6	<i>cupid</i> UCH II regions candidates
2D K-S ^k test	–	–	75.6	19.64	–	2.83	–	2D K-S test on the UCH II regions candidates

^aThis is parameter a in equation 2.9, page 47, and corresponds to half the size of the Galactic bar used.

^bFrom Russeil (2003) best-fit model.

^cCorresponds to x on equation 2.7, page 44.

^dBest fit for the mass function found by Nutter and Ward-Thompson (2007), x on equation 2.7, page 44.

^eEquation 2.1, page 41.

^fType III sources correspond to a type I source inside a type II.

^gEquation 2.1, page 41.

^hIRAS PSC observational value is 927. Wood and Churchwell (1989) observational value is 736.

ⁱPercentage of number of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 2.26 % .

^jNumber of simulated sources that satisfy the Wood and Churchwell (1989) for UCH II regions. Observational value is 21.

^k2D K-S in units of 10^{-2} .

Models \mathcal{A} to \mathcal{G}

All models \mathcal{A} to \mathcal{U} correspond to a 4 spiral arm model in agreement to Russeil (2003) best-fit model.

Model \mathcal{A} has 5.0×10^5 isolated YSOs (population 1) distributed in the model Galaxy. No population 2 or 3 sources were included. Of the 479 sources detected *in all IRAS wavelength* (927 is the IRAS PSC observational value), 143 lie within the UCH II criterion. The K-S test between the *caravela* model distribution and the real observed distribution is 1.37×10^{-2} .

Model \mathcal{B} has 2.5×10^5 single cold grey-body sources with $T = 10.0$ K randomly mixed^b with 2.5×10^5 population 3 sources. The latter are compound compact YSOs surrounded by larger warmer grey-bodies ($T = 50.0$ K). No population 1 sources were included. Of the 191 sources detected *in all IRAS wavelength* (927 is the IRAS PSC observational value), 134 lie within the UCH II criterion.

Model \mathcal{C} has 5.0×10^5 population 3 objects distributed in the model Galaxy. No population 1 or 2 sources were included. Of the 514 sources detected *in all IRAS wavelength* (927 is the IRAS PSC observational value), 239 lie within the UCH II criterion.

Model \mathcal{D} has 1.0×10^5 population 1 objects, 1.0×10^5 population 2 objects, and 3.0×10^5 population 3 objects. Therefore there are the same number of total objects as in \mathcal{A} . Of the 430 sources detected *in all IRAS wavelength* (927 is the IRAS PSC observational value), 122 lie within the UCH II criterion.

Models \mathcal{E} to \mathcal{G} all have population 3 objects. Of the 2 sources detected in model \mathcal{G} both of them lie within the UCH II colour-colour box.

We conclude that models \mathcal{A} to \mathcal{F} result in values of n (tables 3.1, 3.2, and 3.3 (pages 73, 74, and 75) *one order of magnitude* too high when compared with the

^bThere is no spatial correlation between populations.

real observational value of 21 (Wood and Churchwell, 1989).

Model \mathcal{G} , with 2.0×10^3 total number of sources (n_{III}), results in only 2 sources being detected. This led us to explore a number of models with intermediate number of sources, i.e. models \mathcal{H} to \mathcal{U} all have a total number of sources $\approx 10^4$ (this corresponds to one order of magnitude less than models \mathcal{A} to \mathcal{F} and one order of magnitude more than models \mathcal{H} to \mathcal{U} , for the total number of sources).

Model \mathcal{G} is the sole model with total number of sources $\sim 10^3$, and, analysing the output results, I concluded that it was not worthwhile pursuing this class of models further.

Models \mathcal{H} to \mathcal{N}

From the analysis of the output values of models \mathcal{A} to \mathcal{G} (and comparing them to real data), I have concentrated all the following models on population 3 objects. These compound objects seem to have the correct SED properties in this context^c.

The total number of objects has been reduced by one order of magnitude with respect to the first models. We took this decision because models with 10^5 total sources produce one order of magnitude too many sources detected, as explained in detail above.

Also, the temperature that defines the grey-body curve now varies from 10 K to 60 K. Although the reduction of the total number of sources was motivated by the analysis of values presented in the output tables (the value of n in particular), the increase in temperature was motivated by the study of the colour-colour diagrams (figures 3.2 to 3.12 (pages 81 to 91)).

This is the temperature of the grey-body envelope of population 3 objects. The grey-body curve, and hence also the object's observed infrared colours (described in

^cThis conclusion was the motivation for extending *caravela* to incorporate this class of sources.

next section), is determined by the aforementioned temperature.

The two real observed colour-colour diagrams that guided us, i.e. this change in temperature improved the Kolmogorov-Smirnov test (KS-test) between the synthetic and real IRAS data $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$.

It is worthwhile to study model \mathcal{M} in detail. In this model, 4.0×10^4 population 3 objects (with $T = 40$ K and linear size scale $l_{\text{III}} = 5.0 \times 10^6$ AU) yield the following output: 48 detected sources by CUPID, $\approx 42\%$ UCH II candidates (20), 84.73 K-S test result.

Models \mathcal{O} to \mathcal{U}

Models \mathcal{O} to \mathcal{U} are constituted exclusively by population 3 objects. The total number of sources range between 2.0×10^4 to 7.0×10^4 . We conclude that none of these models is an improvement with respect to model \mathcal{M} .

3.2.2 Wood and Churchwell (1989) *total* IRAS sources

Here we would like to address the discrepancy between the number of sources found by Wood and Churchwell (1989) in the $\ell = 40$ degrees box being studied, and the correspondent number of *caravela* sources.

Wood and Churchwell (1989) found 209 IRAS sources in the 2×2 , centred at $\ell = 40$ degrees. The *caravela* model \mathcal{M} result, N , presented in Table 3.2, page 74, reveal 48 sources detected by cupid. Nonetheless, when filtering the sources using the colour-colour cut, the *caravela* and Wood and Churchwell (1989) are in good agreement for this model: 21 Wood and Churchwell (1989) against 20 *caravela* sources.

This can be interpreted as follows: the *caravela* input catalogue has a higher percentage of UCH II candidates than the correspondent *true* IRAS set. Therefore,

the UCH II regions selection criteria applied to the *caravela* input catalogue returns $\approx 50\%$ of the sources as positive identifications: 20 out of 48.

This is consistent with the properties of the *caravela* input catalogue used here: a catalogue of young stellar objects sampled from a representative mass function.

3.2.3 Infrared colours

We present in figures 3.2 to 3.12 (pages 81 to 91) the colour-colour diagrams for models \mathcal{A} to \mathcal{U} . The colours plotted result from the multi-wavelength CUPID analysis described before. The colours plotted are, for each model, $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$ and $100\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$.

In all colour-colour panels, the black crosses are the CUPID detected sources, the red dots are the theoretical catalogue sources (i.e. the total number of sources that could be detected), the green and blue dots correspond to the two sets of real IRAS sources described in figure 3.1, page 80. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

The real observed data colours are plotted in figure 3.1, page 80.

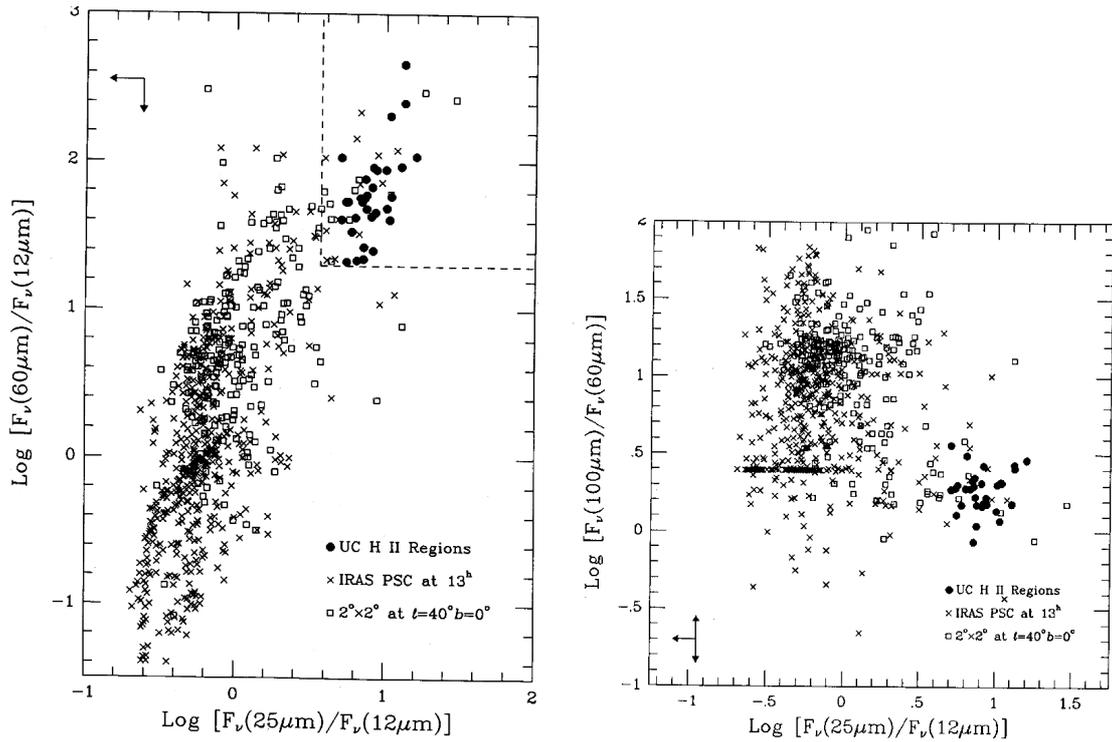


Figure 3.1: Wood and Churchwell (1989) colour-colour diagrams. IRAS sources associated with known UCH II regions are plotted as filled circles. Sources taken from a $2^\circ \times 2^\circ$ box in the plane of the Milky Way at 40° galactic longitude. Crosses and open squares are representative IRAS sources. *Left:* colour-colour diagram for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dashed lines indicate the boundary of the region used to discriminate between UCH II regions and other IRAS sources. *Right:* colour-colour diagram for $100\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$, for the same sources.

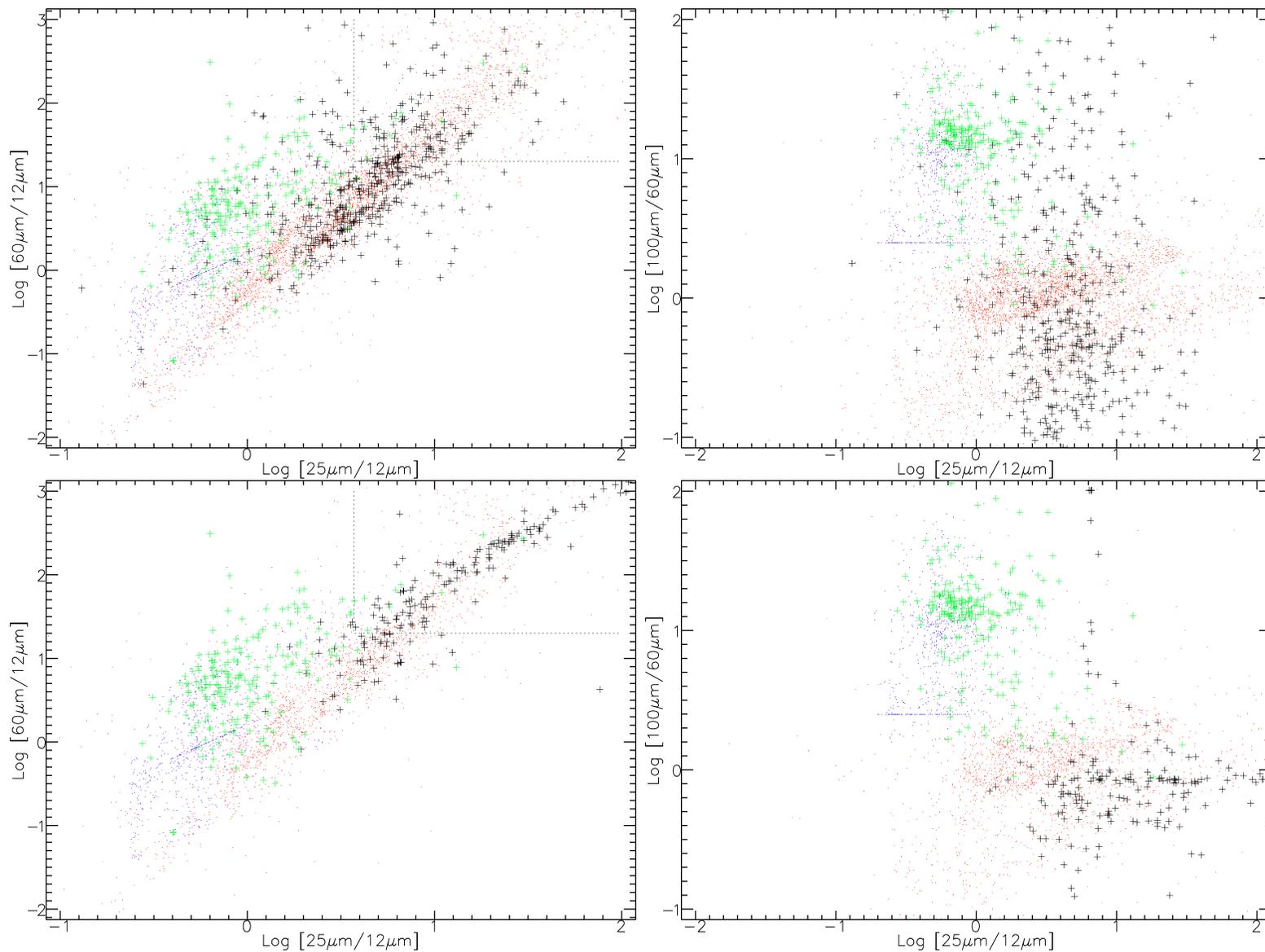


Figure 3.2: Each line corresponds to one of the four $\mathcal{A}, \dots, \mathcal{B}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. *Right column:* colour-colour diagrams for $100 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

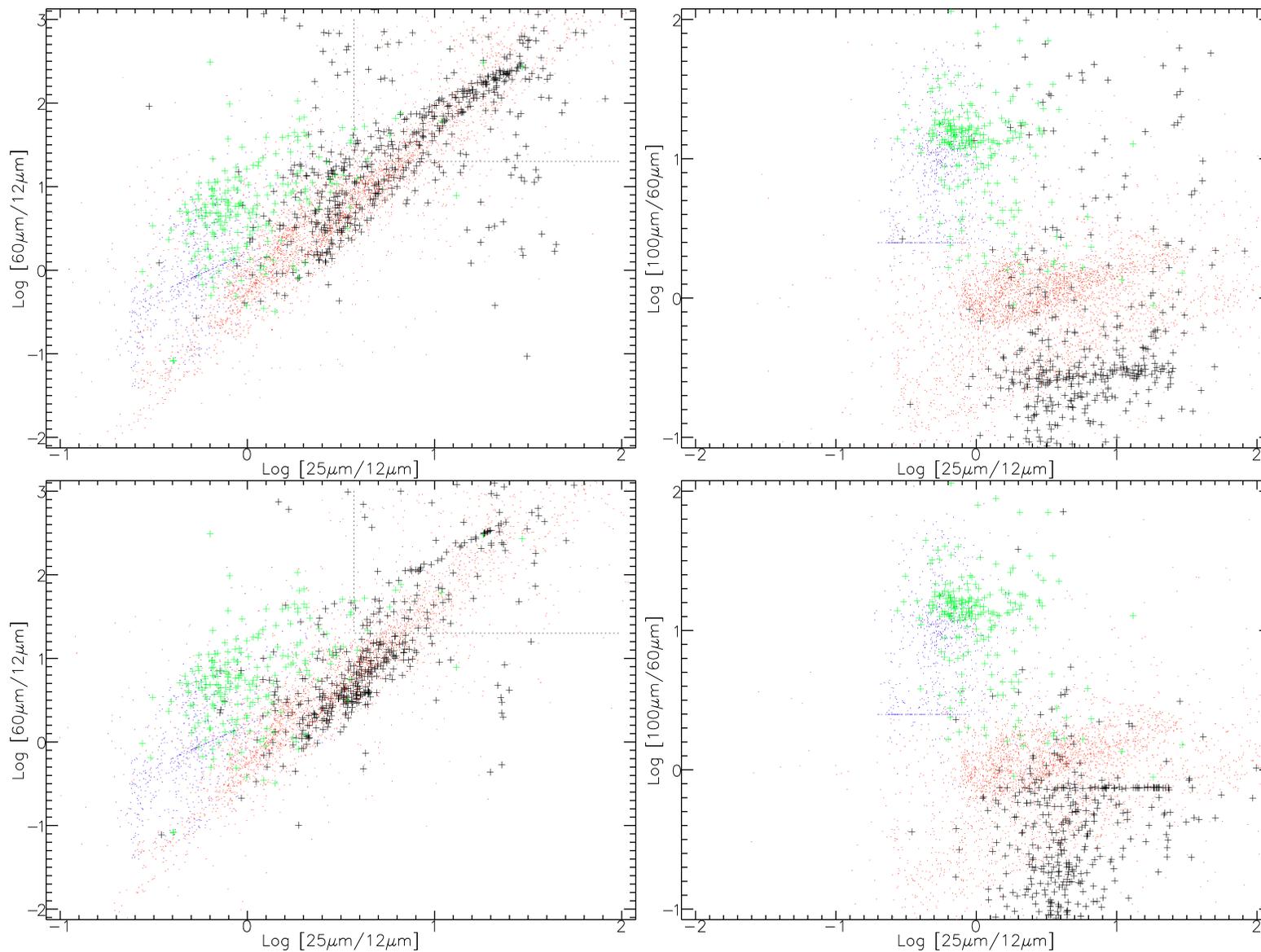


Figure 3.3: Each line corresponds to one of the four $\mathcal{C}, \dots, \mathcal{D}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. *Right column:* colour-colour diagrams for $100\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

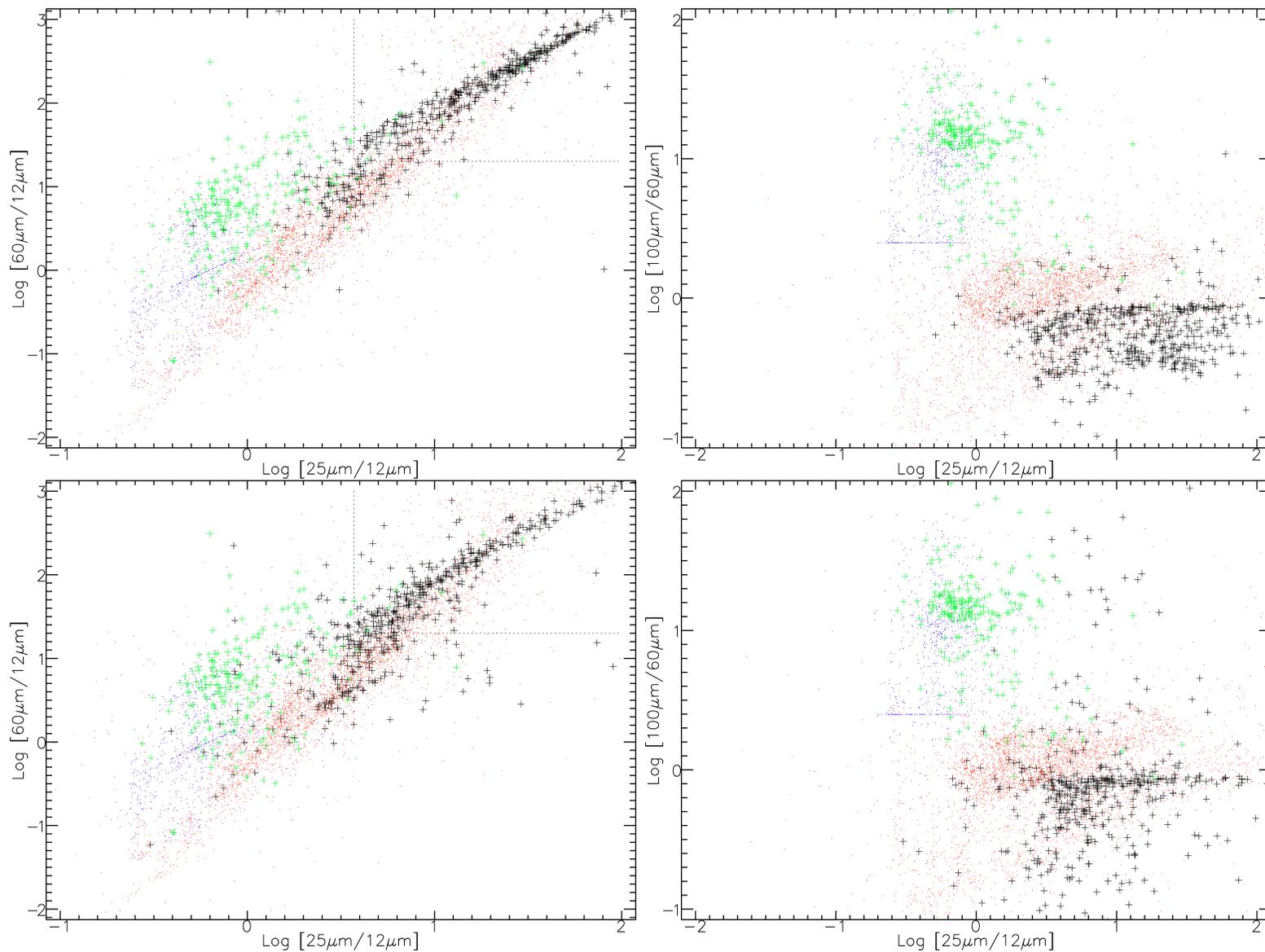


Figure 3.4: Each line corresponds to one of the four $\mathcal{E}, \dots, \mathcal{F}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. *Right column:* colour-colour diagrams for $100 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

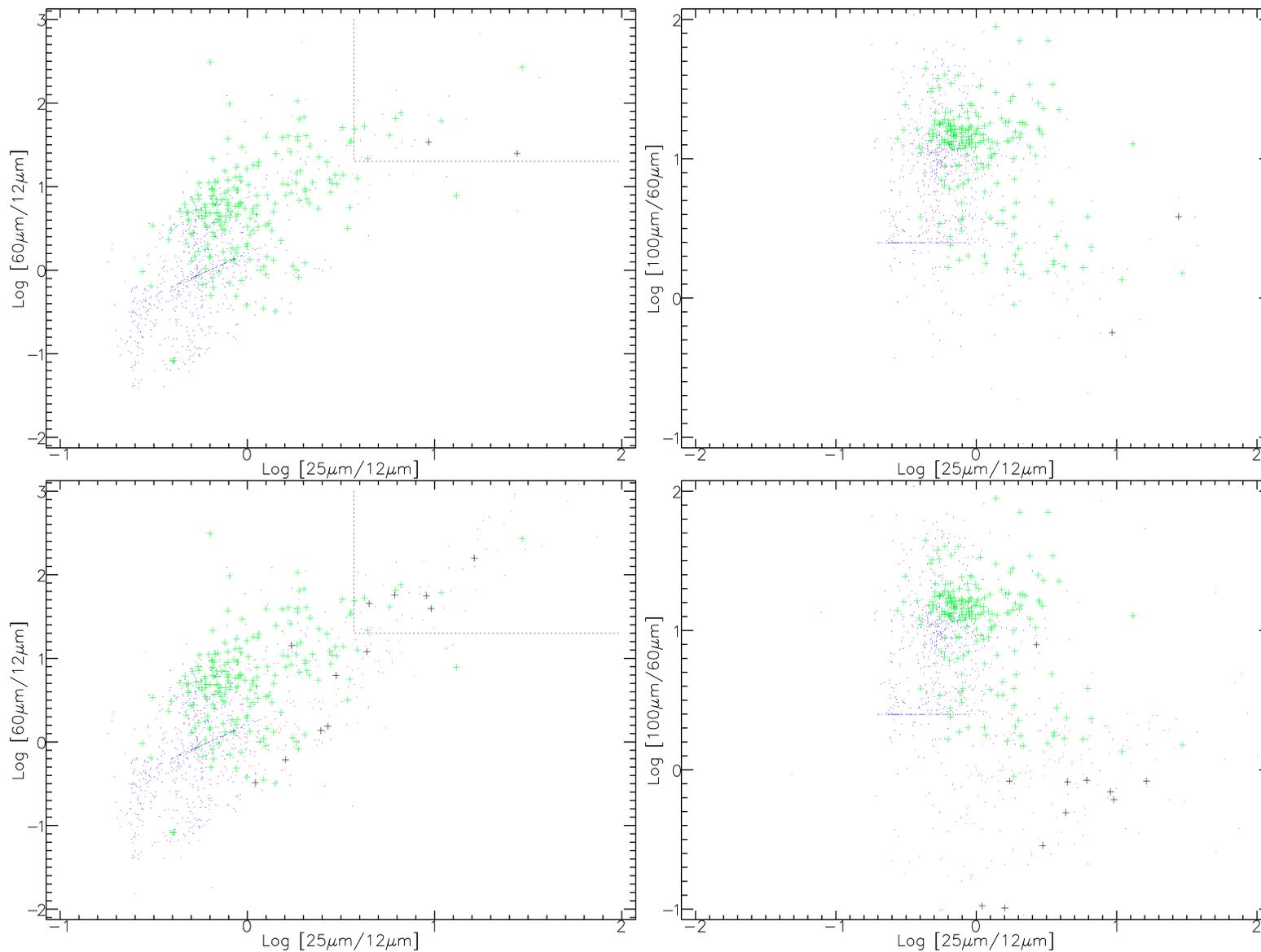


Figure 3.5: Each line corresponds to one of the four $\mathcal{G}, \dots, \mathcal{H}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. *Right column:* colour-colour diagrams for $100\mu\text{m}/60\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

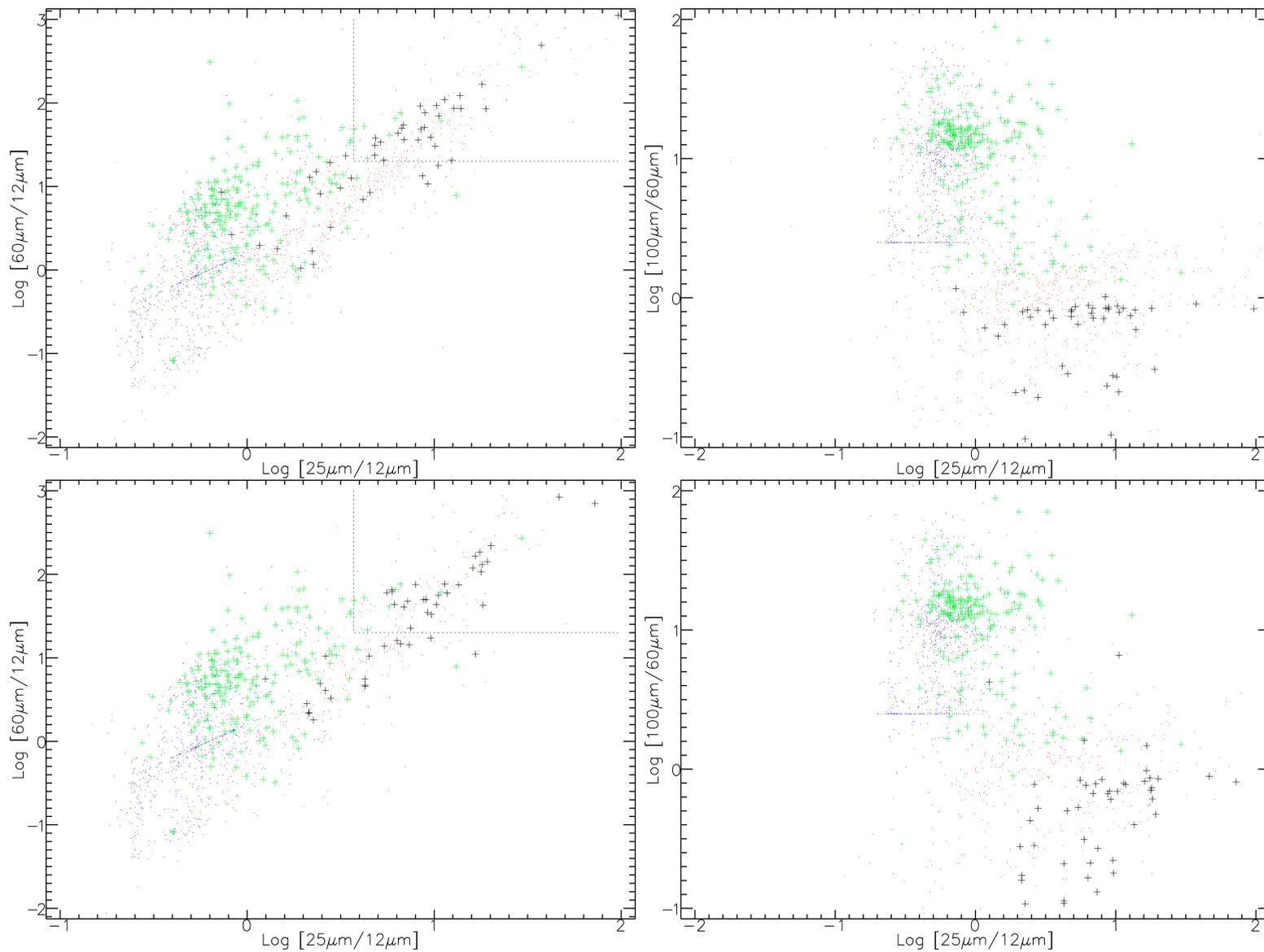


Figure 3.6: Each line corresponds to one of the four $\mathcal{I}, \dots, \mathcal{J}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. *Right column:* colour-colour diagrams for $100\mu\text{m}/60\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

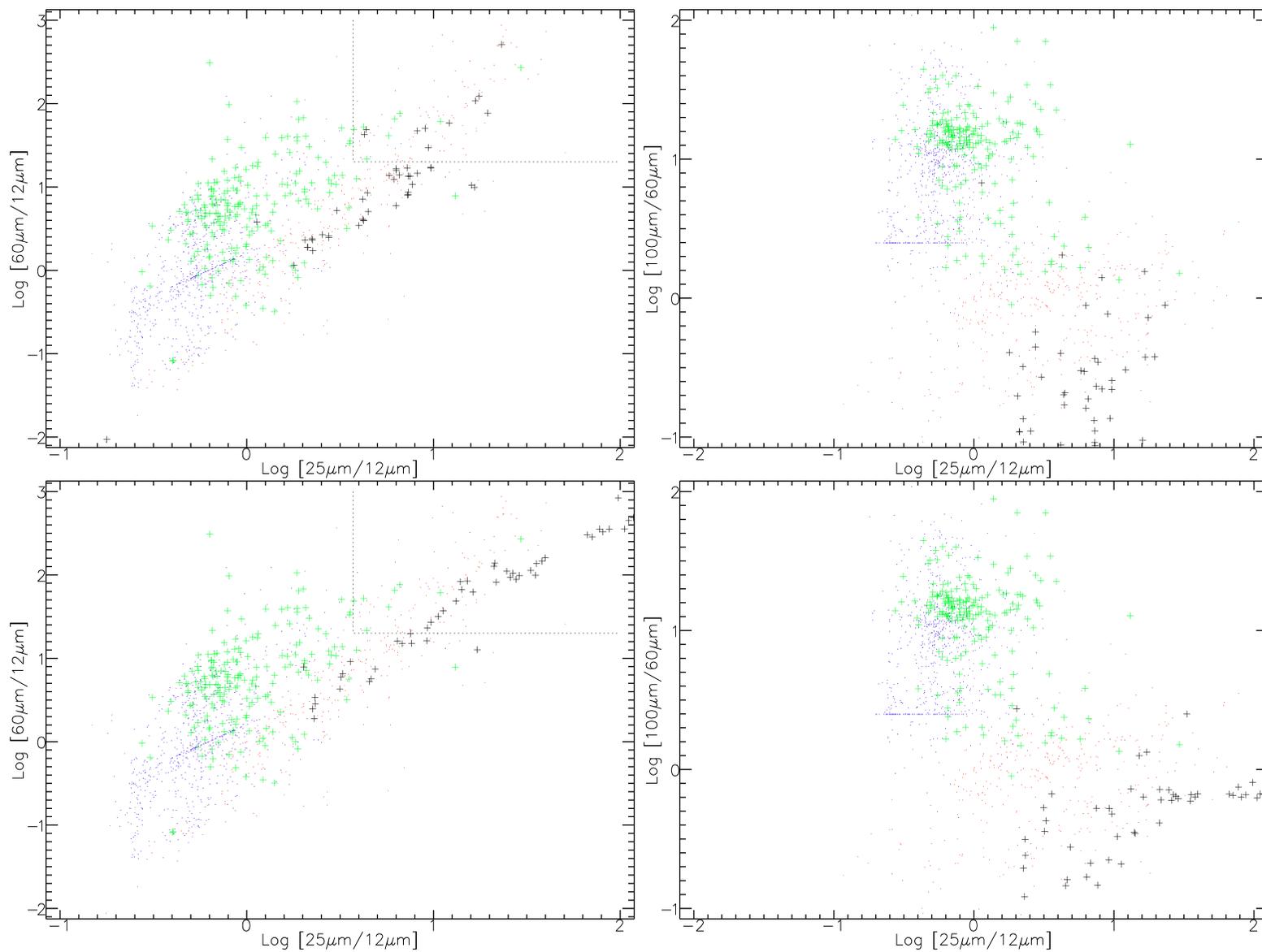


Figure 3.7: Each line corresponds to one of the four $\mathcal{K}, \dots, \mathcal{L}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. *Right column:* colour-colour diagrams for $100\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

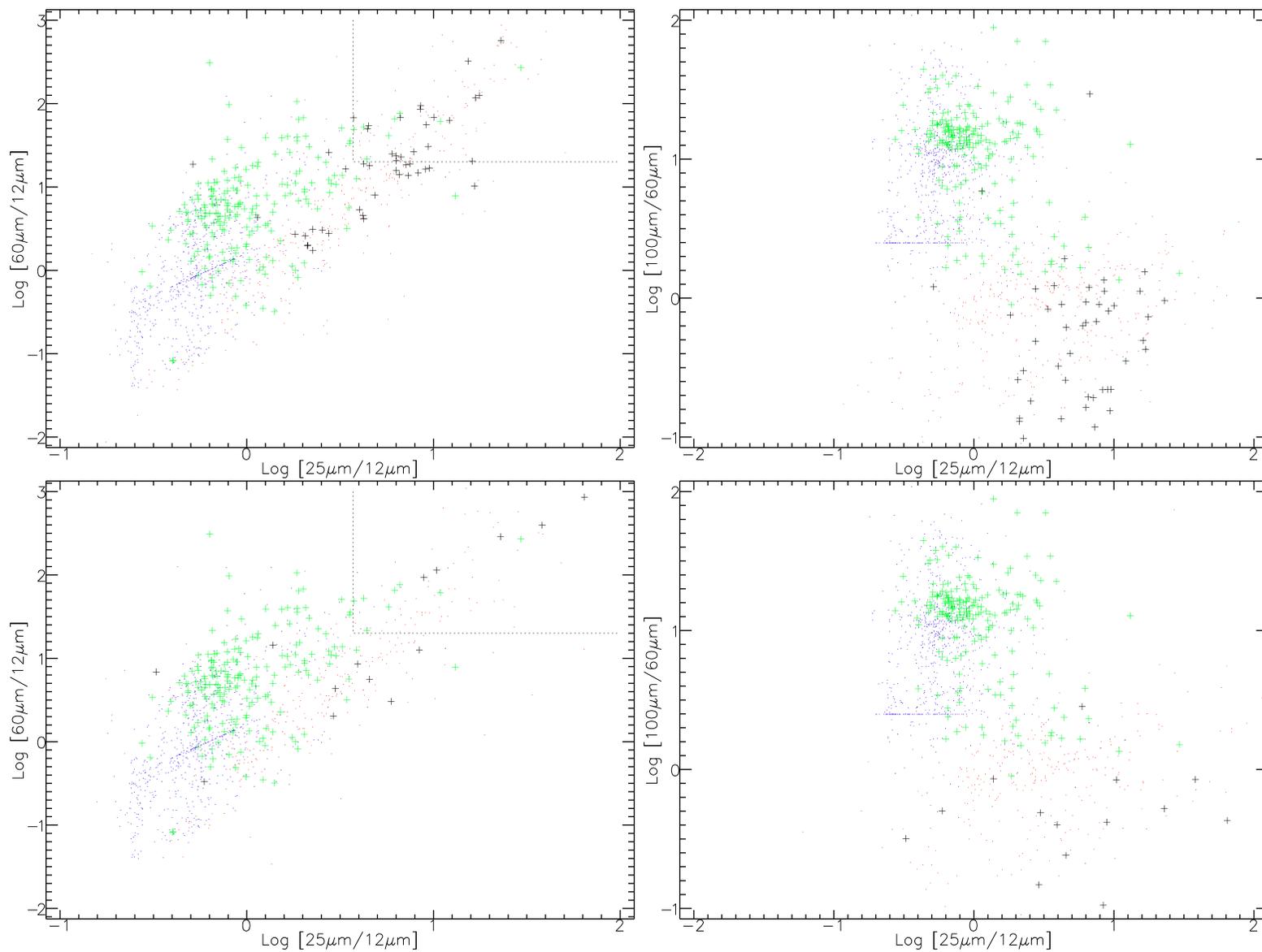


Figure 3.8: Each line corresponds to one of the four $\mathcal{M}, \dots, \mathcal{N}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. *Right column:* colour-colour diagrams for $100\ \mu\text{m}/60\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

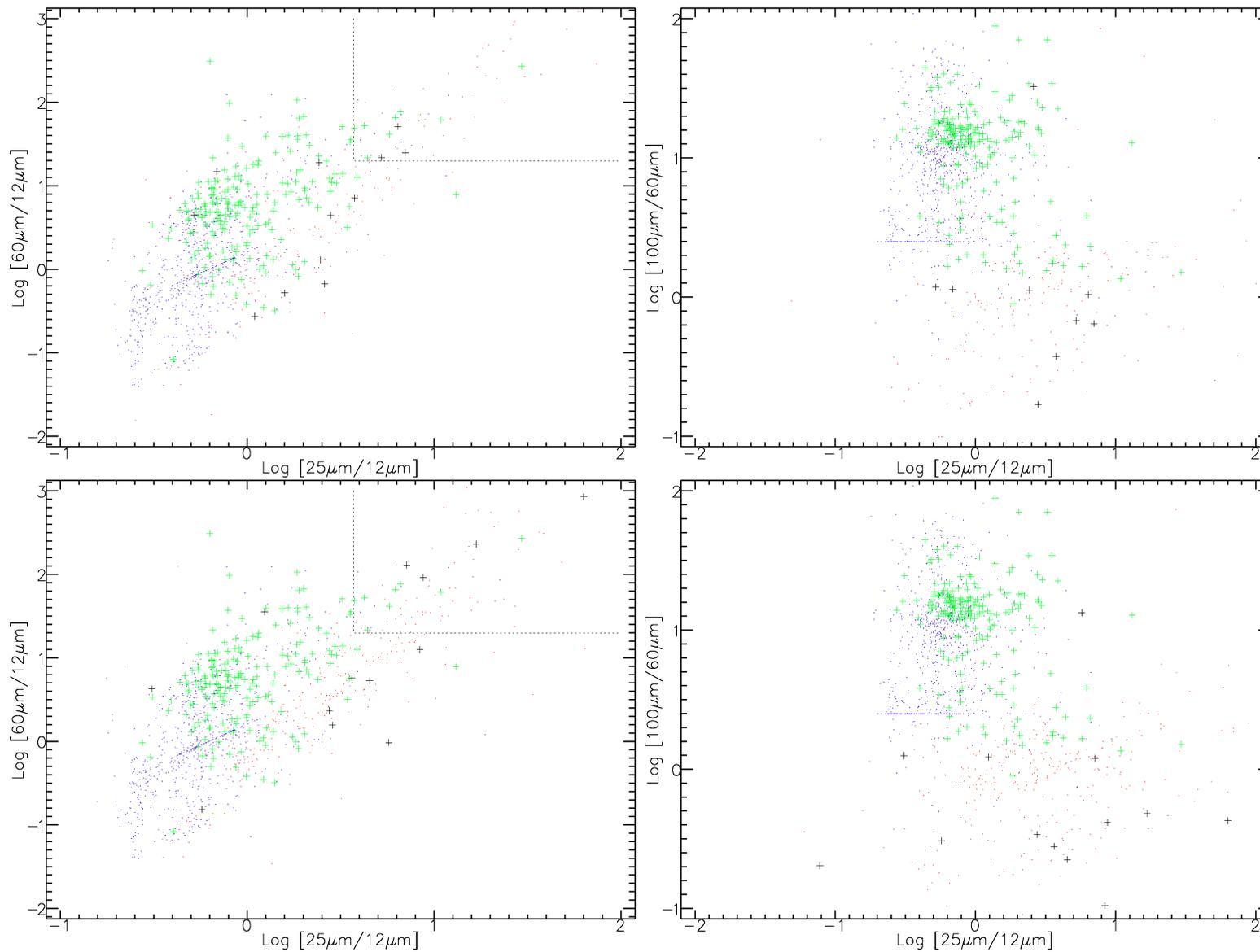


Figure 3.9: Each line corresponds to one of the four $\mathcal{O}, \dots, \mathcal{P}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. *Right column:* colour-colour diagrams for $100\mu\text{m}/60\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

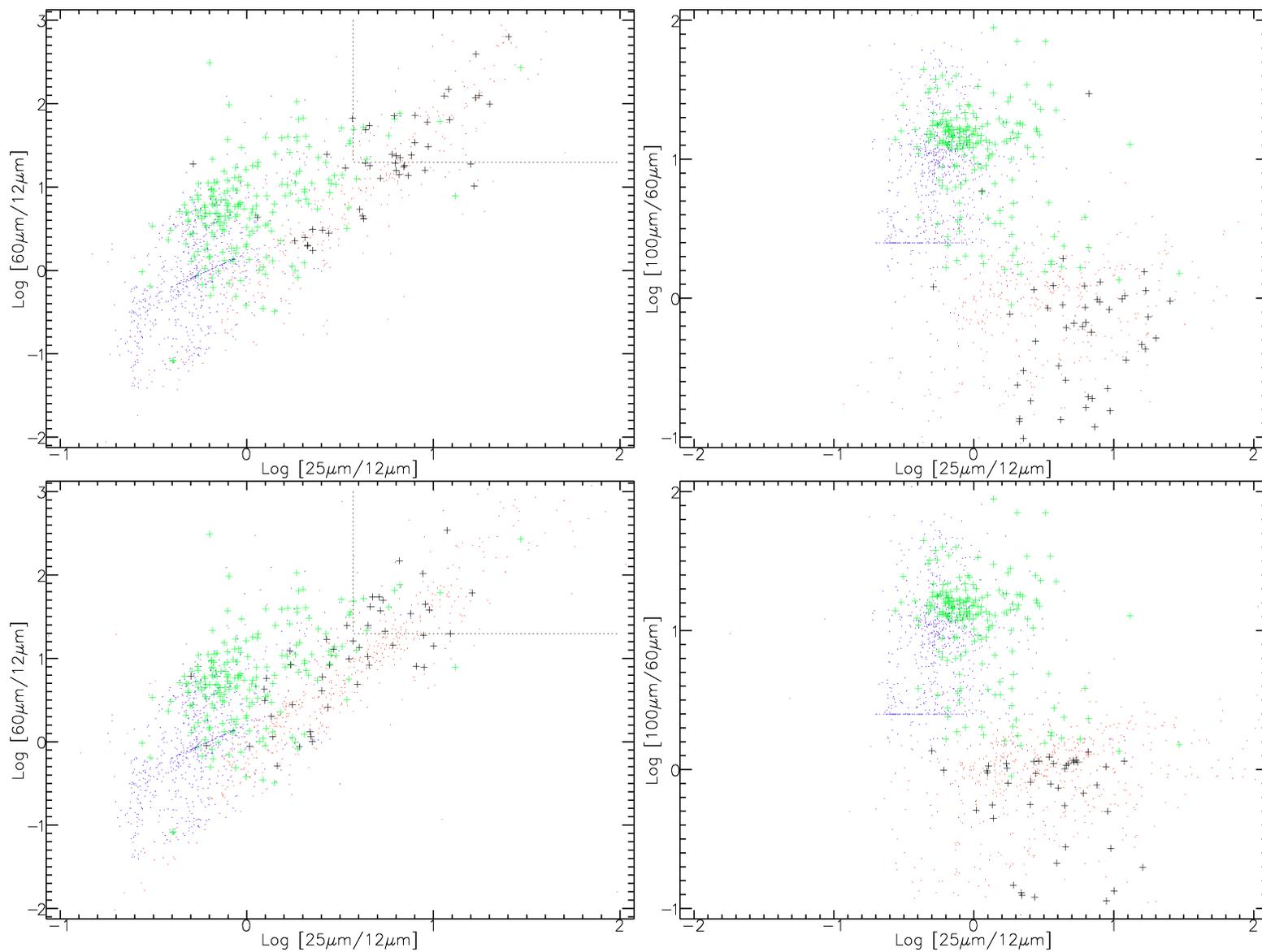


Figure 3.10: Each line corresponds to one of the four $\mathcal{Q}, \dots, \mathcal{R}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. *Right column:* colour-colour diagrams for $100\mu\text{m}/12\mu\text{m}$ vs $25\mu\text{m}/12\mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

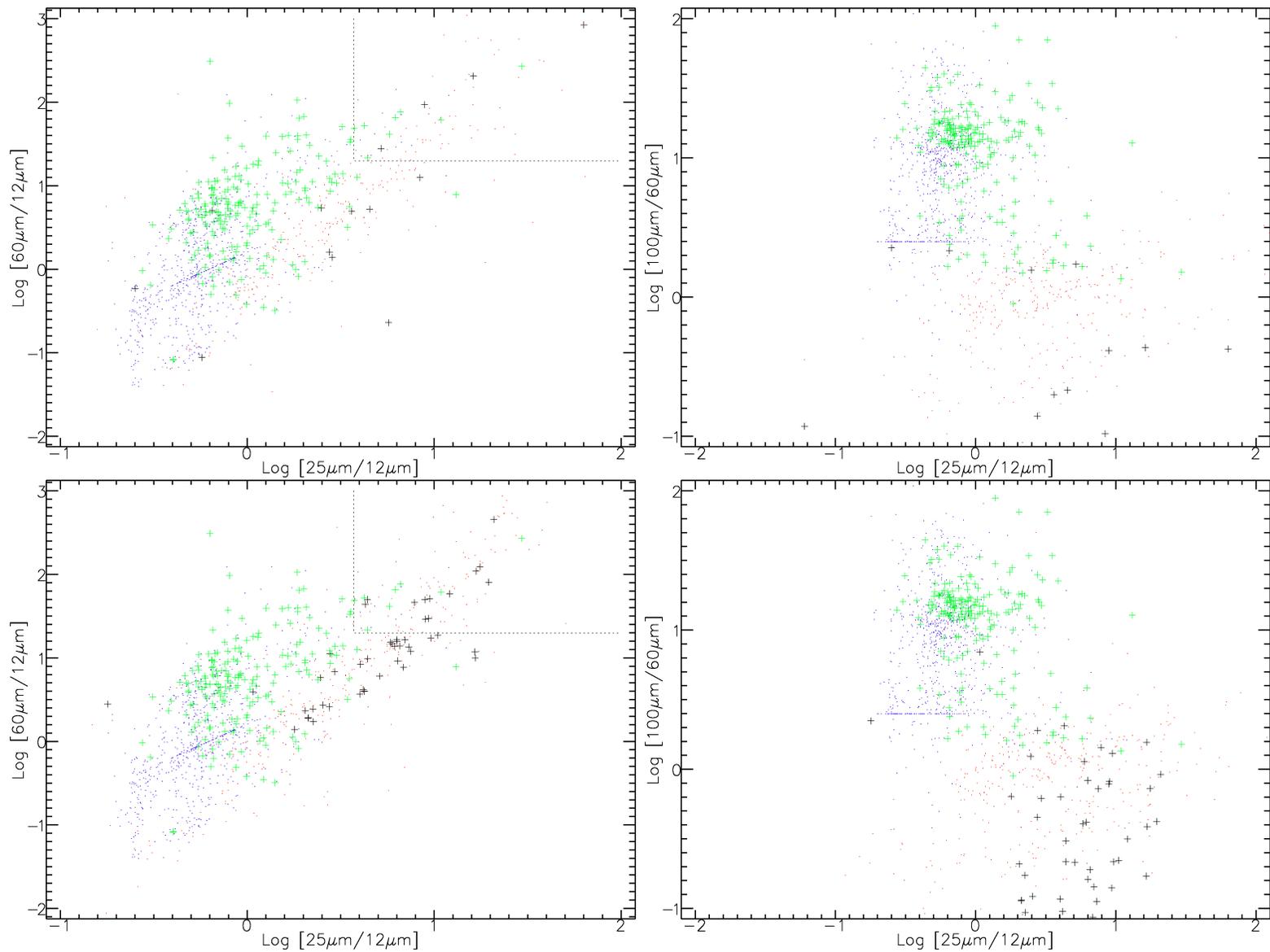


Figure 3.11: Each line corresponds to one of the four $\mathcal{S}, \dots, \mathcal{T}$ models, from top to bottom. *Left column:* colour-colour diagrams for $60 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. *Right column:* colour-colour diagrams for $100 \mu\text{m}/12 \mu\text{m}$ vs $25 \mu\text{m}/12 \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

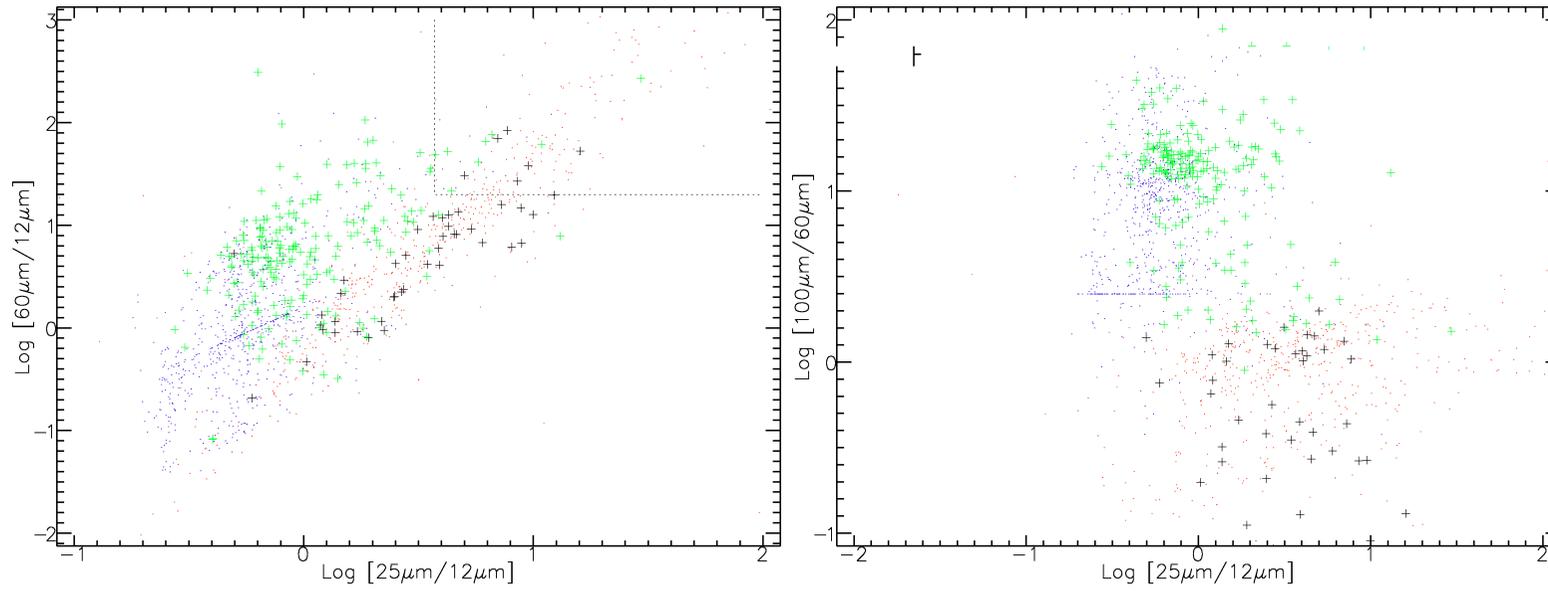


Figure 3.12: Model \mathcal{U} . *Left column:* colour-colour diagrams for $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. *Right column:* colour-colour diagrams for $100\ \mu\text{m}/60\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$. The dotted lines in the left panels represent the Wood and Churchwell (1989) UCH II selection criteria.

The dashed lines indicate the boundary^d of the region established by Wood and Churchwell (1989) to discriminate between UCH IIs and other IRAS sources. They correspond to $\log(F_{60 \mu\text{m}}/F_{12 \mu\text{m}}) > 1.30$ and $\log(F_{25 \mu\text{m}}/F_{12 \mu\text{m}}) > 0.57$.

3.2.4 The distribution of UCH ii regions in the Galaxy

Two main quantitative criteria were used to select, from models \mathcal{A} to \mathcal{U} , the one that is in better agreement with the data set. These statistical tests are:

- A two dimensional Kolmogorov-Smirnov test^e (2D KS-test) between the synthetic colours and the IRAS observed colours. The test was applied to the cumulative distribution of the $\log(25 \mu\text{m}/12 \mu\text{m})$ set of values, **and**, the cumulative distribution of the $\log(60 \mu\text{m}/12 \mu\text{m})$, hence a 2D test.
- After discovering which parameters were the key inputs to the model (i.e. a *small* change in the value of one of these variables would imply a *significant* modification of the output values), a 2D parameter space grid was built. To each coordinate point, on the temperature vs total number of sources grid, was assigned a δ value. Each δ is the distance between the synthetic output value and the real IRAS observed value in this grid. The variable chosen in this study to determine δ was the number of simulated sources that satisfy the Wood and Churchwell (1989) criterion for UCH II regions. A 2D region of optimal fit was then identified in this δ parameter space, (figure 3.13, page 96). The models close to the centre of this optimal region were the ones considered with a higher degree of consistency with the real data set.

^dThis limits are the ones still accepted and in use today for UCH II regions.

^eThe Kolmogorov-Smirnov test (KS-test) tries to determine if two data sets differ significantly. The KS-test has the advantage of making no assumption about the distribution of data. It is non-parametric and distribution free.

3.2.5 2D Kolmogorov-Smirnov Test

All the numerical tasks included in `caravela` use the algorithms from the latest editions of Numerical Recipes (Press *et al.*, 2007, 2002). The inclusion of these algorithms in `caravela` was facilitated by the adoption of C++ as the de facto scientific computer language in Press *et al.* (2007) (`caravela` is built in C++).

Here, we would like to investigate if the synthetic colours and the IRAS observed colours distributions are consistent, i.e. *are the two distribution different?* Proving that two distributions are different, or showing that they are consistent, is a ubiquitous in research. We adopted the Kolmogorov-Smirnov (or K-S) test. This test uses the a statistics to measure the overall difference between two cumulative distribution functions: the maximum value of the absolute difference between two cumulative distribution functions.

As we are comparing two 2D distributions, therefore a 2D K-S test is required. Unfortunately, cumulative probability distribution is not well-defined in more than one dimension. The function used in `caravela` to circumvent this problem is `NR::ks2d2s`. This implements an original idea from (Peacock, 1983), and is described in detail in Press *et al.* (2007). In summary, Peacock (1983) insight was that a good estimation could be achieved using the integrated probability in each of four natural quadrants around a given point.

One numerical disadvantage of the two-dimensional tests, by comparison with their one-dimensional counterparts, is that the two-dimensional tests require more operations: N^2 instead of N . In `NR::ks2d2s` used in `caravela`, two nested loops of order N take the place of an N sort for the 1D case.

Parameter space of the 21 models presented

Two more qualitative arguments were also used in this decision: the difference between the synthetic and the real number of simulated sources that satisfy the Wood and Churchwell (1989) criterion for UCH II regions.

We explored the parameter space corresponding to the 21 models presented previously in the current chapter. Graphically, the parameter space explored is presented in the left panel of figure 3.13 (page 96), where each cross is one model.

The right panel of figure 3.13 (page 96) is the $\delta = ||n^{\text{IRAS}} - n^{\text{caravela}}||$ parameter space representation.

We conclude that model \mathcal{M} is the model that is the more likely given the IRAS PSC observational data. Here I present three pieces of evidences to support this:

- model \mathcal{M} corresponds to $\approx 85\%$ 2D K-S test probability parameter (this is near the maximum of this parameter distribution for the models tested)
- model \mathcal{M} lies in the centre of the optimal region (*dark blue* region in figure 3.13, page 96) in the parameter space investigated
- the number of sources that satisfy the Wood and Churchwell (1989) criterion for UCH II regions is 20 (compared with observed 21 by IRAS)
- the $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$ and the $100\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$ colour-colour diagrams are in good agreement with the observed data (top panel of figure 3.8, page 3.8)

The first step in the search for a model that would be a good representation of the data was to find the most important input parameters. These are the parameters that, with a small change in their value, alter the output of *caravela* dramatically. This sensitivity study of the input parameters revealed that the temperature of

the envelope of the embedded sources, T_{III} , and the total number of (embedded population 3) sources in the Galaxy, n_{III} , were the parameters to explore in detail.

Figure 3.14, page 96, in a bar plot for the 2D K-S test significance output value between the observed between the synthetic colours and the IRAS observed colours: $\log(25 \mu\text{m}/12 \mu\text{m})$ and, $\log(60 \mu\text{m}/12 \mu\text{m})$.

Models \mathcal{J} to \mathcal{Q} have a *good* behaviour in the 2D K-S test, and seem to be grouped together, showing similar values.

Interestingly, *caravela*'s model \mathcal{M} (the best-fit model) yields a significantly better constraint in n than in T , i.e. the *true* value of the total number of sources can be estimate with more accuracy than the temperature of the grey-body envelope. In fact, this trend is observed when all the models are considered. The right panel in figure 3.13, page 96, and figure 3.14, page 96, illustrate this argument. The former shows that the optimal dark blue δ region appears to be stretched horizontally (the envelope temperature) and squeezed vertically (the total number axis), the latter exhibits that the 2D K-S test values for models \mathcal{J} to \mathcal{Q} are all good and similar. These are exactly the models with similar n and a spread in the temperature values.

The two figures have independent origins and there seems to be a consistency between them.

In conclusion, n is constrained to the $[1 \times 10^4, 9 \times 10^4]$ interval in total number of objects, and T lies within ≈ 30 to ≈ 50 Kelvin, for the grey-body temperature.

Note that the model elected as the best-fit one, uses exclusively population 3 sources. This category of sources, the compound central compact sources embedded in a larger grey-body envelope, were idealised for and implemented for the first time during the development of *caravela*.

The conclusion is that a total number of 4.0×10^4 population 3 objects distributed in 4 spiral arms are compatible with the IRAS PSC data. This is our best

estimate for the number of UCH II objects in the Galaxy.

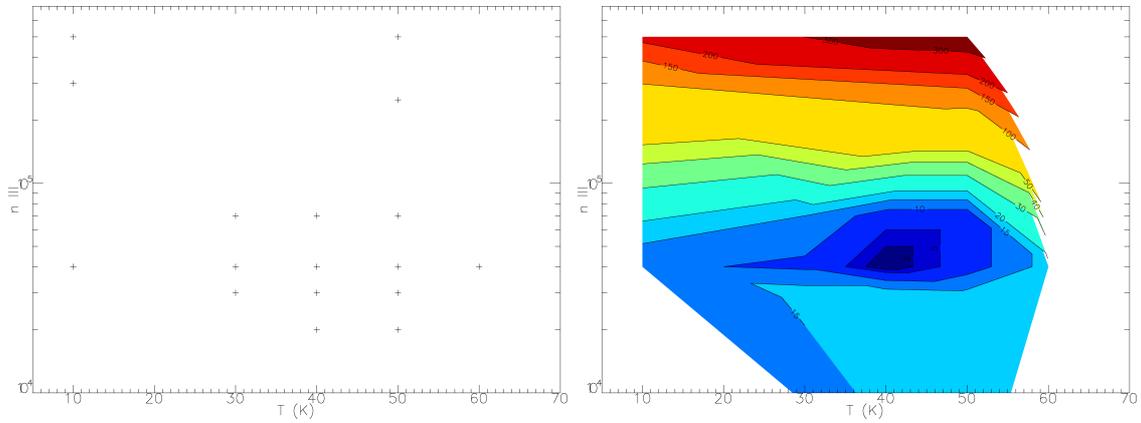


Figure 3.13: *Right* : 2D parameter space searched. Dark blue colours indicate a minimum in the δ function and hence an optimum model. Here, $\delta = ||n^{\text{IRAS}} - n^{\text{caravela}}||$. *Left* : Each cross is one of the 21 models in this parameter space.

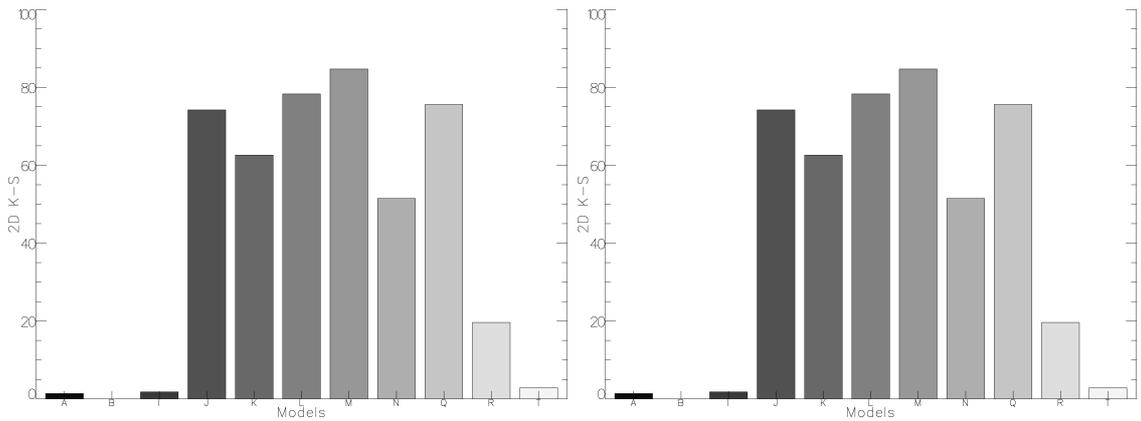


Figure 3.14: 2D K-S plot for models described in tables 3.1, 3.2, 3.3, pages 73, 74, 75.

Table 3.4: IRAS cross-matching results: Theoretical input catalogue *vs* *caravela* output observational catalogue, for model \mathcal{M} .

	n^a	α^b	β^c	n_1^d	n_2^e	$n_{1\&2}^f$	\bar{s}^g (")	\bar{m}^h (M_\odot)	σ_{mass}^i	σ_s^j (")
Input catalogue	469	$2.0'^k$	2	25^l	9	9	98.44	3.35	5.12	23.45
Output catalogue	20	$2.0'$	2	25	9	9	98.44	3.98	4.01	23.45

^aNumber of sources that satisfy the Wood and Churchwell (1989) criterion.

^bSeparation error.

^cNumber of high-mass star forming regions, i.e. number of objects with $m \geq 8 M_\odot$.

^dNumber of sources in the theoretical catalogue with a match in the *caravela cupid* output table.

^eNumber of sources in the output catalogue with a match in the theoretical input catalogue.

^fNumber of matches between catalogues

^gMean value for the distribution of separations between catalogues.

^hMean value for the distribution of masses.

ⁱStandard deviation for the distribution of masses.

^jStandard deviation for the distribution of separations.

^kThe matching algorithm used was `topcat::sky`

^lNumber of distinct sources.

3.2.6 Low- and Intermediate-mass contamination

The aim is to compare the theoretical catalogue to the *cupid* photometry final catalogue, for model \mathcal{M} , the best fit model for the UCH II regions. In this way *caravela* can be used to investigate which are the masses, and other physical properties, of the objects that satisfy the Wood and Churchwell (1989) criterion, i.e. investigate which input physical objects survive the *caravela* process thus being detected by *cupid* as an end product of *caravela*.

The two catalogues are cross-matched with $\alpha = 2'$ angular radius^f. The theoretical catalogue and the simulated observed catalogue. The theoretical catalogue corresponds to the *caravela* input objects (distribution of objects described in §2.6, page 39), and consists of the position in the sky, SEDs and the mass for each object^g. The simulated observed catalogue is the output from applying *cupid* to the four *caravela* synthetic IRAS maps ($12 \mu\text{m}$, $25 \mu\text{m}$, $60 \mu\text{m}$, and $100 \mu\text{m}$).

^fThe cross-matching of the theoretical and simulated observed catalogues was made using Tool for OPERations on Catalogues And Tables, i.e. `topcat`

^gEach object can be a type I, II or type III source.

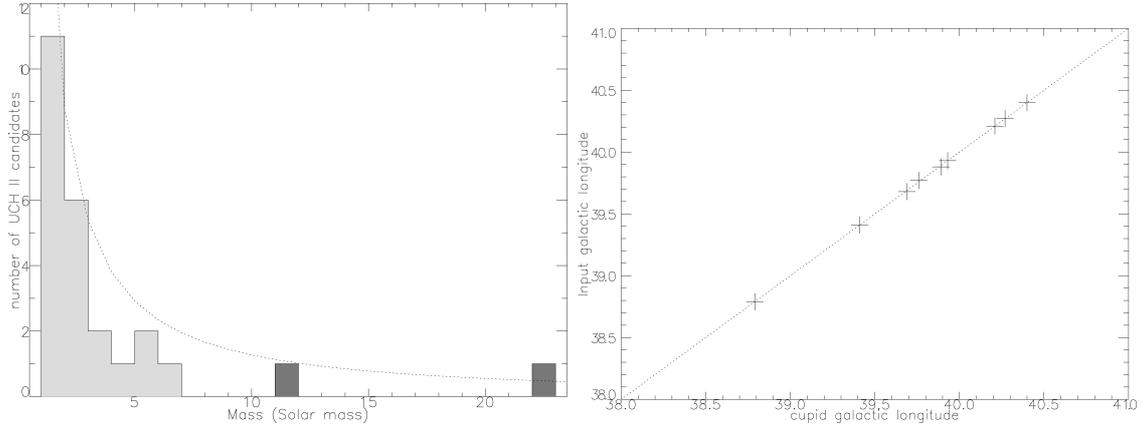


Figure 3.15: *Left:* Distribution of UCH II regions detected in model \mathcal{M} . This results from the cross-correlation between the *caravela* input catalogue and *cupid* detected clumps. Lighter bins represent the low- and intermediate-mass objects, i.e. sources cross-correlated to input objects with $m < 8 M_{\odot}$. Darker bins represent the high-mass star forming regions, i.e. sources cross-correlated to input objects with $m \geq 8 M_{\odot}$. Here *only* $\beta/n_1 = 2/25 = 8\%$ of the UCH II candidates have $m > 8 M_{\odot}$. The dotted line is the input mass function with Salpeter exponent of 1.2. *Right:* Galactic longitude for the input sources *vs* galactic longitude for the output catalogue after cross-matching of sources. The plot shows that the cross-matching is consistent (with a maximum difference of 0.02°).

Only $\beta/n_1 = 2/25 = 8\%$ of the UCH II candidates have $m > 8 M_{\odot}$, suggesting $\approx 90\%$ of contamination by low- and intermediate-mass objects.

UCH II regions have sizes of order 0.1 pc. The ionised gas within has sound speed $c_s \sim 10 \text{ km s}^{-1}$, so they should have dynamical times of order 10^4 yr. However, there are roughly 10% as many UCH II regions as there are OB stars in the Galaxy, suggesting lifetimes an order of magnitude longer (Mac Low *et al.*, 2004). This is referred to as the UCH II lifetime problem.

The level of contamination from low- and intermediate-mass objects supported by the analysis of the described models, $\sim 90\%$, *could* be a possible solution for this conundrum. In principle, we could be overestimating the number of UCH II regions in the Galaxy due to the fact that our UCH II samples are polluted with low- intermediate-mass objects.

Contamination by low-mass objects

The level of contamination by low- and intermediate-mass objects found by this *caravela* study is, nevertheless, a problem. This is the case since Kurtz *et al.* (1994) concluded that $\approx 50\%$ of the sources in their sample (based on the Wood and Churchwell (1989) criteria) were UCH II regions; and, $\approx 40\%$ were IRAS sources not confirmed to be UCH II. The rest belonged to some other class of objects.

This discrepancy is not resolved in this work and remains an interesting question to be investigated with further analysis using *caravela* and future observational data sets.

3.2.7 Galactic distribution

Figure 3.16, page 100, is the 2D plot for the galactic latitude and galactic longitude, i.e. l vs b .

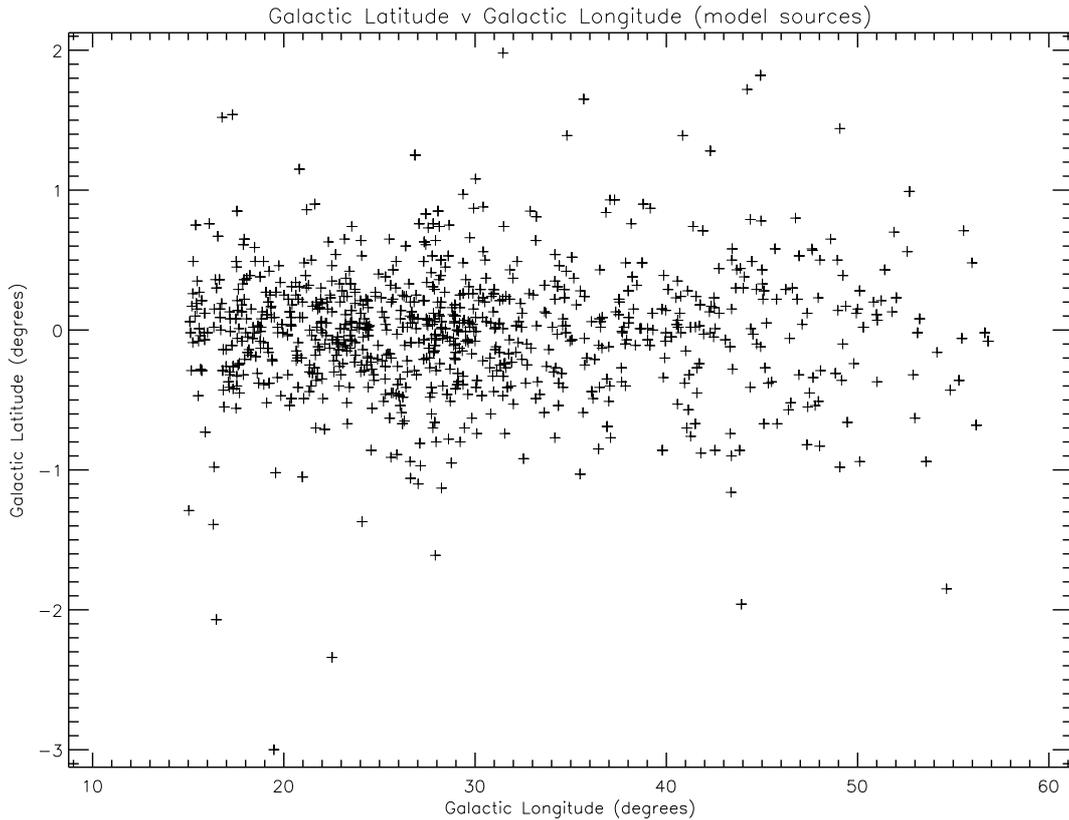


Figure 3.16: Galactic latitude *vs* galactic longitude for the *caravela* sources.

The synthetic *caravela* data sets can be compared with a number of additional observational products, e.g.:

1. mass function, Figure 3.15 (selection or input mass function, and, simulated output or *observed* mass function, against *real* observed mass function);
2. global brightness distribution.

Detailed comparisons of both the mass function and global brightness function with the IRAS PSC is nor presented here. Their implications are discussed concisely

in the Future Work section.

3.3 Summary

This chapter is an attempt to use `caravela` in order to interpret *real* observational data sets.

The following conclusions result from this effort:

1. The grid of models studied in this chapter consists of 21 models spanning a significant area of the parameter space. Model \mathcal{M} is the model that is the more likely given the IRAS PSC observational data. Here I present three evidences to support this:
 - model \mathcal{M} corresponds to $\approx 85\%$ 2D K-S test probability parameter (this is near the maximum of this parameter distribution for the models tested)
 - model \mathcal{M} lies in the centre of the optimal region (*dark blue* region in figure 3.13, page 96) in the parameter space investigated
 - the number of sources that satisfy the Wood and Churchwell (1989) criterion for UCH II regions is 20 (compared with observed 21 by IRAS)
 - the $60\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$ and the $100\ \mu\text{m}/12\ \mu\text{m}$ vs $25\ \mu\text{m}/12\ \mu\text{m}$ colour-colour diagrams are in good agreement with the observed data (top panel of figure 3.8, page 3.8)
2. `caravela`'s model \mathcal{M} (the best-fit model) yields a significantly better constraint in n than in T , i.e. the *true* value of the total number of sources can be estimated with more accuracy than the temperature of the grey-body envelope. In fact, this trend is observed when all the models are considered. In conclusion, n is constrained to the $[1 \times 10^4, 9 \times 10^4]$ interval in total number of objects, and T lies within ≈ 30 to ≈ 50 kelvin, for the grey-body temperature.
3. *We were able to conclude that a total number of 4.0×10^4 population 3 objects*

distributed in 4 spiral arms are consistent with the IRAS PSC data. This is our best estimate for the number of UCHII objects in the Galaxy.

4. There is contamination by low and intermediate mass star forming regions of the Wood and Churchwell (1989) region for UCHII regions. This effect was quantified to be of the order of $1 - (\beta/n_{1\&2}) = 1 - (2/9) \approx 77\%$ (using the same notation as in table 3.4, 97) in this *caravela* best fit simulation. If one considers all the sources in the input catalogue that have a match in the output catalogue, then $1 - (\beta/n_1) = 1 - (2/25) \approx 92\%$. This may be caused by confusion of observations due to the lack of resolution of the IRAS survey.
5. The level of contamination from low- and intermediate-mass objects supported by the analysis of the described models, $\sim 90\%$, *could* help to understand the UCHII lifetime problem. In principle, we could be overestimating the number of UCHII regions in the Galaxy due to the fact that our UCHII samples are polluted with low- intermediate-mass objects.
6. The Salpeter mass function is found in the subset of UCHII regions, for this best-fit *caravela* model studied, therefore there is no evidence that the Wood and Churchwell (1989) criterion selects high-mass star forming regions preferentially.
7. Column β in table 3.4, page 97, shows that all the high-mass star forming regions that were UCHII candidates in the input catalogue were detected in the output catalogue (2/2).

Chapter 4

The UCH II regions observed by Herschel

4.1 Introduction

In the previous chapter, a model has emerged as being in reasonably good agreement with the observations of UCH II regions. `caravela` was used to establish that model \mathcal{M} , formed by 4.0×10^4 population 3 sources distributed in 4 spiral arms, was consistent with the IRAS point source catalogue.

As seen in Chapter 2, `caravela` has the ability to generate catalogues and images simulating the Herschel Observatory photometry instruments: The Photodetector Array Camera and Spectrometer (PACS), and The Spectral and Photometric Imaging REceiver (SPIRE)^a.

Here, we *observe* the best-fit model \mathcal{M} with PACS and SPIRE, and present some results.

In particular, we aim to predict how many UCH II regions the Herschel Space

^aPACS is provided a consortium led by A. Poglitsch, MPE, Garching, Germany; SPIRE is provided by a consortium led by M. Griffin, Cardiff University, UK.

Observatory would be detected using PACS and SPIRE independently, and, with the two instruments working together in parallel mode, under the assumption that model \mathcal{M} is a realistic representation of the high-mass star forming regions in the Galaxy.

In principle the proposed exercise is straightforward: we use the same input parameter file with only the wavelength modified to $70\ \mu\text{m}$, $110\ \mu\text{m}$, $170\ \mu\text{m}$, and $250\ \mu\text{m}$, $360\ \mu\text{m}$, and $500\ \mu\text{m}$. The spatial resolution and sensitivity change accordingly. This implies six *caravela* runs in *fake* parallel computing. The code associates the chosen wavelength with the correspondent instrument, i.e. the correct SED transmission band for each object and the correct instrument (resolution, sensitivity and synthetic PSF).

4.2 A model for the UCHII regions in the Galaxy

Model \mathcal{M} *observed* by *caravela* applying only the four IRAS bands, 4.0×10^4 population 3 objects (with outer envelope modelled by a grey-body shell of $T = 40\ \text{K}$ and linear size scale $l_{\text{III}} = 5.0 \times 10^6\ \text{AU}$), resulted in the following output:

- 48 sources detected simultaneously in coherent positions in all four bands,
- $\approx 42\%$ UCHII candidates (i.e. 20 sources).

These results were then compared, in chapter 3, with the IRAS UCHII regions survey (Wood and Churchwell, 1989).

The work presented in this chapter results from submitting the identical model \mathcal{M} to the Herschel *caravela* simulator. The output of this will then be assessed against the IRAS synthetic output, and used to estimate the number of sources Herschel could detect in the near future.

The input parameters for model \mathcal{M} are described in table 4.1, page 106.

Table 4.1: Input parameters for the Herschel simulation (model \mathcal{M}^b is described in detail in section 3.2.1, page 77.)

Parameter	Value	Units	Description
n_{spiral}	4		number of spiral arms
a_{spiral}	3	kpc	start of the spiral arms ^c
b_{spiral}	14	°	spiral arms pitch angle ^d
n_{I}	0.0		number of isolated compact sources
l_{I}	–		linear size scale
x_1	–		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	–		mass function exponent for $2.4 < m(M_{\odot})$
n_{II}	0.0		number of isolated extended sources
l_{II}	–		linear size scale
x_1	–		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	–		mass function exponent for $2.4 < m(M_{\odot})$
T_{II}	–		Grey-body temperature
β	–		Grey-body β parameter
n_{III}	4.0×10^4		number of embedded sources
l_{III}	5.0×10^6	AU	linear size scale for the envelope sources
x_1	0.3		mass function exponent for $1.0 < m(M_{\odot}) < 2.4$
x_2	1.2		mass function exponent for $2.4 < m(M_{\odot})$
T_{III}	40.0	K	Grey-body temperature for the envelope sources
β	2		Grey-body β for the envelope sources

^aModel \mathcal{M}

^bModel \mathcal{M}

^cThis is parameter a in equation 2.9, page 47, and corresponds to half the size of the Galactic bar used.

^dFrom Russeil (2003) best-fit model.

4.3 Herschel synthetic images

Although the final output for each *caravela* run is a point source catalogue at the requested wavelength band (or wavelength bands, if multiple instruments are to be simulated at once), the code also produces synthetic images (and *three colour* images for multiple bands simulations). Figure 4.3, page 108, is an illustration of *caravela* output showing the three PACS wavebands combined.

The PACS and SPIRE *caravela* simulator is used, i.e. the real PSFs and noise levels were used in the simulators. Figure 4.3, page 108, is a three colour *caravela* image at the PACS wavebands: $70\ \mu\text{m}$ (blue), $110\ \mu\text{m}$ (green), and $170\ \mu\text{m}$ (red). Note the diffraction spikes visible in the bright source on the bottom left of the image. *caravela* is able to simulate these artifacts (and study the subsequent effect on the point source catalogue) because it uses the *correct* PACS synthetic PSF.

The noise levels used in the PACS and SPIRE images define how efficiently the source detections and extraction algorithm will work. This corresponds to the instrumental noise. The values built-in to the *caravela* simulator are (in MJy sr^{-1}): 19.9, 8.8, 2.9, 2.1, and 0.8, for $70\ \mu\text{m}$, $170\ \mu\text{m}$, $250\ \mu\text{m}$, $360\ \mu\text{m}$, and $500\ \mu\text{m}$, respectively. In MJy beam^{-1} the noise values used are: 17.6, 26.8, 12.8, 17.6, and, 14.9, for $70\ \mu\text{m}$, $170\ \mu\text{m}$, $250\ \mu\text{m}$, $360\ \mu\text{m}$, and $500\ \mu\text{m}$, respectively. These are the 1σ noise values.

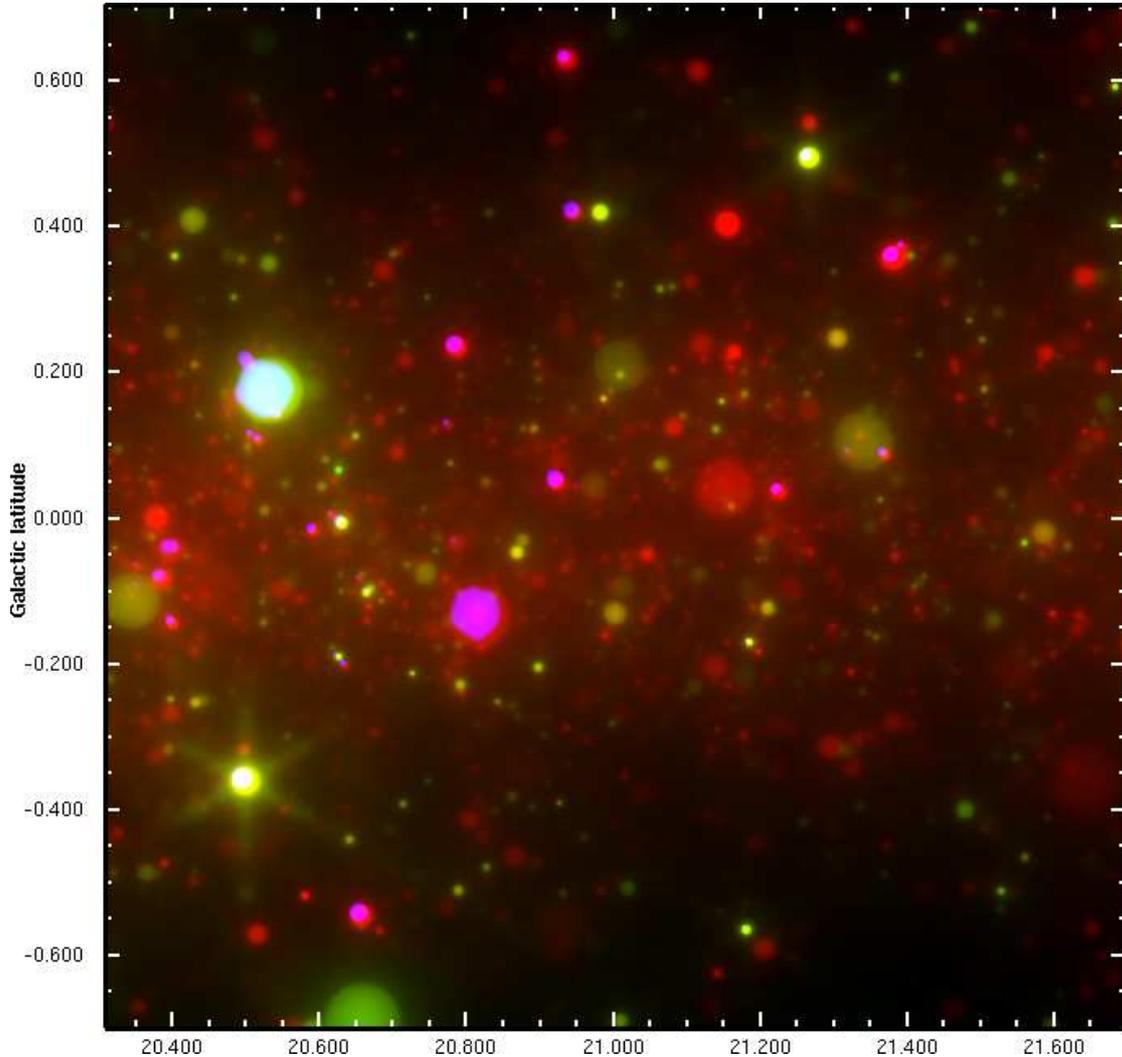


Figure 4.1: Three colour *caravela* image at the PACS wavebands: $70\ \mu\text{m}$ (blue), $110\ \mu\text{m}$ (green), and $170\ \mu\text{m}$ (red). Note the diffraction spikes visible in the bright source on the bottom left of the image. *caravela* is able to simulate these artifacts (and study the consequent effect on the point source catalogue) because it use the *correct* PACS synthetic PSF. The SED models for population 3 objects use the correct PACS transmission filters.

4.4 Results

The results described next are summarised in table 4.2, page 110, and table 4.3, page 113.

Table 4.2, page 110, presents the output from the *caravela* simulation of model \mathcal{M} using the PACS and SPIRE wavebands.

In order to be coherent with the Chapter 3 area of study, the area of the sky examined is a $2^\circ \times 2^\circ$ box centred in the plane of the Milky Way at 40° galactic longitude.

The final photometry, considering all PACS and SPIRE wavebands combined, resulted in $n_{\text{PACS}} = 46$ sources and $n_{\text{SPIRE}} = 42$ sources. n_{PACS} is the number of sources detected simultaneously and in the same positions in *all* PACS wavebands^b. n_{SPIRE} is the number of sources detected simultaneously and in the same positions in *all* SPIRE wavebands. Note that these are high-mass star forming regions candidates, selected from population 3 objects following a $x_1 = 0.3$, $x_2 = 1.2$, for $\zeta(\log m) \propto m^{-x}$, mass function.

A number of Herschel key programs, e.g. the Hi-GAL survey, will be using PACS and SPIRE in parallel mode. It is interesting to estimate the number of common objects detected at all Herschel wavelengths. This is given by n (table 4.2, page 110). In this study, $n = 40$. In order to successfully estimate the common objects detected in more than one image (i.e more than one wavelength band), *caravela* uses `cupid::extractclumps` to try to detect the same clumps on all six wavelengths. One waveband must be selected as a mask image.

Once the objects are detected and *cupid* is able to extract their photometric information, it is possible to build colour-colour diagrams for the Herschel bands. In table 4.2, page 110, the mean value and standard deviation of these colours are

^bIn reality, the PACS instrument cannot work simultaneously at its three wavelengths.

Table 4.2: Output parameters for the Herschel simulation (model \mathcal{M})

Parameter	Value	Units	Description
n_{PACS}	46		number of sources detected at all PACS wavelengths
n_{SPIRE}	42		number of sources detected at all SPIRE wavelengths
n	40 ^a		number of sources detected at all Herschel wavelengths
110 $\mu\text{m}/70 \mu\text{m}$	-1.6811149		mean value of the 110 $\mu\text{m}/70 \mu\text{m}$ colour
$\sigma_{110 \mu\text{m}/70 \mu\text{m}}$	0.80768770		standard deviation for the 110 $\mu\text{m}/70 \mu\text{m}$ colour
170 $\mu\text{m}/70 \mu\text{m}$	-0.81292277		mean value of the 170 $\mu\text{m}/70 \mu\text{m}$ colour
$\sigma_{170 \mu\text{m}/70 \mu\text{m}}$	0.32876475		σ value of the 170 $\mu\text{m}/70 \mu\text{m}$ colour
360 $\mu\text{m}/250 \mu\text{m}$	-0.43023142		mean value of the 360 $\mu\text{m}/250 \mu\text{m}$ colour
$\sigma_{360 \mu\text{m}/250 \mu\text{m}}$	0.46501149		σ value of the 360 $\mu\text{m}/250 \mu\text{m}$ colour
520 $\mu\text{m}/250 \mu\text{m}$	-1.1800373		mean value of the 520 $\mu\text{m}/250 \mu\text{m}$ colour
$\sigma_{520 \mu\text{m}/250 \mu\text{m}}$	0.24772479		σ value of the 520 $\mu\text{m}/250 \mu\text{m}$ colour

^acupid::extractclumps was used to try to detect the same clumps on all six wavelengths.

listed. The colour-colour diagrams are presented in figure 4.2, page 112.

For $\log [110 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -1.6 , with $\sigma_{110 \mu\text{m}/70 \mu\text{m}} = 0.80$. For $\log [170 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -0.8 , with $\sigma_{170 \mu\text{m}/70 \mu\text{m}} = 0.32$. For $\log [360 \mu\text{m}/250 \mu\text{m}]$, the mean value *observed* is -0.43 , with $\sigma_{360 \mu\text{m}/250 \mu\text{m}} = 0.46$. Finally, for $\log [520 \mu\text{m}/250 \mu\text{m}]$, the mean value *observed* is -1.18 , with $\sigma_{520 \mu\text{m}/250 \mu\text{m}} = 0.24$.

Figure 4.2, page 112, presents four selected colour-colour diagrams for model \mathcal{M} . This is the best-fit model using the IRAS data set (presented in the previous chapter).

The cited Herschel colour-colour diagrams may have a two-fold utility. Reveal some insight into the model \mathcal{M} (and hence the true physical distribution of high-mass star forming regions, since this the more likely model). And, produce a direct comparison between real PACS and SPIRE data and *caravela* data set (we look forward for this comparison). On the former, it is interesting to note the relatively narrow range in the $\log (360 \mu\text{m}/250 \mu\text{m})$ in Figure 4.2, page 112. One valid interpretation for this result is as follows. The SEDs for the model high-mass star

forming regions peak at $\sim 250 \mu\text{m}$ therefore a variation in temperature (T controls the shape of the SED curve) would result in a small variation in the fluxes at $250 \mu\text{m}$ and $360 \mu\text{m}$, since both values are relatively close to the peak wavelength. On the other hand, the same variation of temperature would yield a significant variation in the other colours, since their correspondent fluxes lie further away from the peak wavelength.

However, IRAS did not detect *any* cold objects such as IRDCs (Egan *et al.*, 1998; Netterfield *et al.*, 2009; Simon *et al.*, 2006). These type of objects fall below the IRAS detection threshold at $100 \mu\text{m}$. Therefore, the Herschel colour-colour diagrams that result from the described Hi-GAL *caravela* simulation, can be used as a sensitive test on the number of cold objects. *If the $\log(360 \mu\text{m}/250 \mu\text{m})$ colour would exhibit a wide range, then IRAS would have missed a significant fraction of cold sources.*

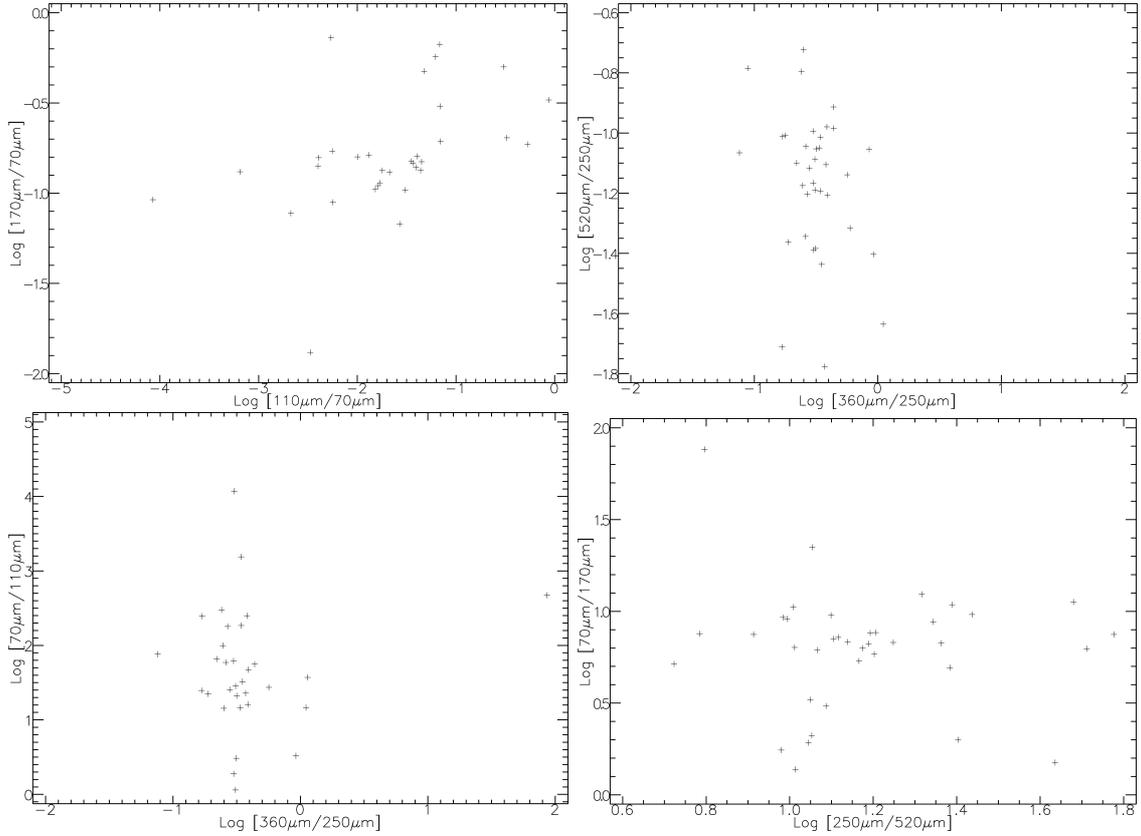


Figure 4.2: Herschel predicted colour-colour diagrams. *Clockwise*: $170\ \mu\text{m}/70\ \mu\text{m}$ vs $110\ \mu\text{m}/70\ \mu\text{m}$, $520\ \mu\text{m}/250\ \mu\text{m}$ vs $360\ \mu\text{m}/250\ \mu\text{m}$, $70\ \mu\text{m}/170\ \mu\text{m}$ vs $250\ \mu\text{m}/520\ \mu\text{m}$, and $70\ \mu\text{m}/110\ \mu\text{m}$ vs $360\ \mu\text{m}/250\ \mu\text{m}$.

The two catalogues are cross-matched^c. The theoretical catalogue and the simulated observed catalogue. Therefore,

- using the objects' positions in the sky, all the objects in the theoretical catalogue are compared with the **cupid** catalogue resulting in n_1 ,
- using the sources' positions in the sky, all the objects in the **cupid** final photometry catalogue are compared with the theoretical catalogue resulting in n_2 ,
- using the objects'/sources' coordinates, only the common objects to both cat-

^cThe cross-matching of the theoretical and simulated observed catalogues was made using Tool for OPERations on Catalogues And Tables, i.e. `topcat`

Table 4.3: Herschel cross-matching results: Theoretical input catalogue *vs* *caravela* output observational catalogue, for model \mathcal{M} .

	n	α^a	β^b	n_1^c	n_2^d	$n_{1\&2}^e$	\bar{s}^f (")	\bar{m}^g (M_\odot)	σ_{mass}^h	σ_s^i (")
Input catalogue	1720 ^j	2.0' ^k	2	20 ^l	18	18	87.51	5.54	8.71	29.37
Output catalogue	40 ^m	2.0'	2	20	18	18	87.51	5.54	8.71	29.37

^aSeparation error.

^bNumber of high-mass star forming regions candidates, i.e. number of objects with $m \geq 8 M_\odot$.

^cNumber of sources in the theoretical catalogue with a match in the *caravela cupid* output table.

^dNumber of sources in the output catalogue with a match in the theoretical input catalogue.

^eNumber of matches between catalogues

^fMean value for the distribution of separations between catalogues.

^gMean value for the distribution of masses.

^hStandard deviation for the distribution of masses.

ⁱStandard deviation for the distribution of separations.

^jNumber of objects in the region of the sky simulated.

^kThe matching algorithm used was `topcat::sky`

^lNumber of distinct sources.

^mNumber of objects detected by *cupid*.

alogues are selected, resulting in $n_{1\&2}$.

The theoretical catalogue corresponds to the *caravela* input objects (distribution of objects described in §2.6, page 39), and consists of the position in the sky, SEDs and the mass for each object^d. The simulated observed catalogue is the output from applying *cupid* to the six *caravela* synthetic PACS/SPIRE Herschel maps (70 μm , 110 μm , 170 μm , 250 μm , 350 μm and 520 μm).

Analysing table 4.3, page 113, it is interesting to note that the modelled data set resulting from the Herschel simulation presented here has $n_{1\&2} \approx 2 \times n_{1\&2}^{\text{IRAS}}$, where $n_{1\&2}$ is the number of cross matches between the theoretical catalogue for Herschel instruments and the final *cupid* list, and $n_{1\&2}^{\text{IRAS}}$ is the number of cross matches between the theoretical catalogue for IRAS *caravela* simulation and the final photometric catalogue.

In order to interpret the $n_{1\&2} \approx 2 \times n_{1\&2}^{\text{IRAS}}$ result, i.e. that Herschel at all wavelengths will detect twice as many sources than IRAS, the findings from Thompson

^dEach object can be a type I, II or type III source.

et al. (2006) can be useful. The authors studied 105 sources selected from the Wood and Churchwell (1989), and observed them with SCUBA. Thompson *et al.* (2006) quantify the likelihood of finding neighbouring clumps within a SCUBA field using the companion clump fraction (CCF). The CCF is defined as,

$$CCF = \frac{B + 2T + 3Q}{S + B + T + Q} \quad (4.1)$$

where S , B , T and Q , are the number of single, binary, triple and quadruple sources.

The observed result was (Thompson *et al.*, 2006),

$$CCF = 0.90 \pm 0.07 \quad (4.2)$$

where the error was determined by simple Poisson statistics. Is this result consistent with $n_{1\&2} \approx 2 \times n_{1\&2}^{\text{IRAS}}$?

A $CCF = 0.90 \pm 0.07$ means that *on average* there were observed 2 SCUBA sources per IRAS source. Assuming that SCUBA and Herschel are similar, then the our *caravela* and the real observed CCF are compatible.

4.4.1 Companion Clump Fraction interpretation

The comparison between the *caravela* results and the CCF from Thompson *et al.* (2006) must be interpreted with caution. It is tempting to assume that the factor of 2 discussed previously is consistent with the a $CCF = 0.90 \pm 0.07$, but this may not be the case. The CCF presented does take into account the proximity (although in 2D project terms, and not in 3D) of the *pairs* of sources. In contrast, the factor of 2 found in these *caravela* simulated images does not. It is possible that the Herschel *caravela* simulations are resulting in twice as many sources detected, on average, but these sources may not be clustered (in 2D).

In principle, a more rigorous analysis on the clustering of sources in the *caravela* images (and the detections rates for different *caravela* instruments) is possible using the input and output information from each *caravela* run. This will be described in the Future Work section.

We conclude that the number of sources detected by *caravela* in IRAS and Herschel mode is consistent with a $CCF = 0.90 \pm 0.07$ from Thompson *et al.* (2006) only in a limited way: the number of detections may be coherent but a spatial investigation is needed.

However, the factor of two resulting from the presented *caravela* simulations seems, intuitively, too low, due to the better resolution and sensitivity of Herschel when compared with IRAS. We suggest two possible reasons that could explain this:

- the sources used in *caravela* are not clustered, i.e. although line of sight alignments can occur, the objects' positions are correlated. There is no real 3D clustering function between objects,
- the values of $n_{1&2}$ and $n_{1&2}^{\text{IRAS}}$ refer to the cross-matched sources, i.e. sources detected in all wavebands for Herschel and IRAS, respectively. To achieve this one of the wavebands images was taken as a mask. For the Herschel analysis, the mask image was the $500\ \mu\text{m}$ one, so all the sources were searched in the other wavebands using this waveband as a mask.

Following these two arguments, the factor of two found in the modelling could be considered as a lower limit value, i.e. to each IRAS source will correspond two or more sources. When a similar analysis on the future Herschel data sets will be made, it will be interesting to try to understand where the adopted *caravela* model is incorrect.

We conclude that the improved spatial resolution of the Herschel Space Observatory when compared with the IRAS telescope (between $0.5'$ and $2.0'$ for IRAS and ranging between $5.2''$ and $35''$ for Herschel, from shorter to longer wavelengths) improves significantly the detection rate of common sources.

Quantitatively, *caravela* has shown here that *twice* as many sources are detected, at all 6 wavelengths, in Herschel than they were positively found with *caravela* simulating the 4 IRAS bands. This is based on the sixth column (labelled $n_{1&2}$) of table 4.3, page 113, and table 3.4, page 97, Herschel and IRAS models respectively.

In Chapter 3 (considering the four bands of IRAS), we have shown that

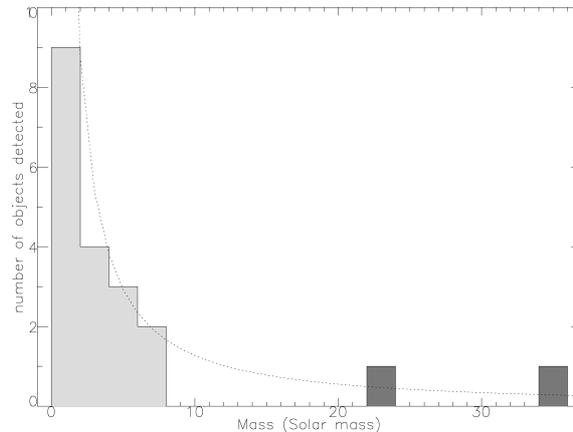


Figure 4.3: Distribution of UCH II regions detected in model \mathcal{M} at the Herschel wavebands. This results from the cross-correlation between the *caravela* input catalogue and *cupid* detected clumps. Lighter bins represent the low- and intermediate-mass objects, i.e. sources cross-correlated to input objects with $m < 8 M_{\odot}$. Darker bins represent the high-mass star forming regions, i.e. sources cross-correlated to input objects with $m \geq 8 M_{\odot}$. Here $2/20 = 10\%$ of the UCH II candidates have $m > 8 M_{\odot}$. The dotted line is the input mass function with Salpeter exponent of 1.2. The set of objects plotted correspond to n_1 (table 4.3, page 113), i.e. number of sources in the theoretical catalogue with a positive match in the *caravela cupid* final output catalogue.

$\beta/n_1 = 2/25 = 8\%$ of the UCH II candidates have $m > 8 M_{\odot}$, suggesting $\approx 90\%$ of contamination by low- and intermediate-mass objects.

Table 4.3, page 113, and figure 4.3, page 117, indicate that the Herschel simulation presented here quantifies the contamination by $m < 8 M_{\odot}$ as $\sim 90\%$ ($\beta/n_1 = 2/20 = 10\%$).

Although model \mathcal{M} is the one that is in better agreement with the data set presented in the previous chapter, it is still *limited* to the simulation of compact objects, i.e. no diffuse emission is considered. This is the final step in the complexity ladder for *caravela* and is presented in chapter 5: adding the diffuse emission (Paladini *et al.*, 2007, 2004, 2003) to the *caravela* models.

4.5 Summary

A model has emerged as being in reasonably good agreement with the observations of UCH II regions. `caravela` was used to establish that model \mathcal{M} , formed by 4.0×10^4 population 3 sources distributed in 4 spiral arms, was consistent with the IRAS point source catalogue. The methodology used in this chapter is to *observe* the best-fit model \mathcal{M} with PACS and SPIRE.

Here we present the more relevant results that emerged from the modelling described in the fourth chapter.

1. The final output for each `caravela` run is a point source catalogue at the requested wavelength band. The code also produces synthetic images (and *three colour* images for multiple bands simulations). In this chapter, a three colour `caravela` image at the PACS wavebands: $70 \mu\text{m}$ (blue), $110 \mu\text{m}$ (green), and $170 \mu\text{m}$ (red) is presented. `caravela` is able to simulate these artifacts (and study the consequent effect on the point source catalogue) because it use the *correct* PACS synthetic PSF.
2. The final photometry, considering all PACS and SPIRE wavebands combined, resulted in $n_{\text{PACS}} = 46$ sources and $n_{\text{SPIRE}} = 42$ sources. n_{PACS} is the number of sources detected simultaneously and in the same positions in *all* PACS wavebands. n_{SPIRE} is the number of sources detected simultaneously and in the same positions in *all* SPIRE wavebands.
3. In this study, $n = 40$, where n is the number of sources extracted for *all* PACS and SPIRE wavelengths.
4. The analysis of the Herschel colour-colour diagram showed that $\log [110 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -1.6 , with $\sigma_{110 \mu\text{m}/70 \mu\text{m}} = 0.80$. For $\log [170 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -0.8 , with $\sigma_{170 \mu\text{m}/70 \mu\text{m}} = 0.32$. For $\log [360 \mu\text{m}/250 \mu\text{m}]$,

- the mean value *observed* is -0.43 , with $\sigma_{360\mu\text{m}/250\mu\text{m}} = 0.46$. Finally, for $\log [520\mu\text{m}/250\mu\text{m}]$, the mean value *observed* is -1.18 , with $\sigma_{520\mu\text{m}/250\mu\text{m}} = 0.24$.
5. The cited Herschel colour-colour diagrams revealed that a relatively narrow range in the $\log (360\mu\text{m}/250\mu\text{m})$. We interpret this result as follows: the SEDs for the model high-mass star forming regions peak at $\sim 250\mu\text{m}$ therefore a variation in temperature (T controls the shape of the SED curve) would result on a small variation on the fluxes at $250\mu\text{m}$ and $360\mu\text{m}$, since both values are relatively close to the peak wavelength. On the other hand, the same variation of temperature would yield a significant variation in the other colours.
 6. the modelled data set resulting from the Herschel simulation presented here has $n_{1\&2} \approx 2 \times n_{1\&2}^{\text{IRAS}}$, where $n_{1\&2}$ is the number of cross matches between the theoretical catalogue for Herschel instruments and the final **cupid** list, and $n_{1\&2}^{\text{IRAS}}$ is the number of cross matches between the theoretical catalogue for IRAS **caravela** simulation and the final photometric catalogue.
 7. A $CCF = 0.90 \pm 0.07$ means that *on average* there were observed 2 sources per one IRAS source. Our **caravela** and the real observed CCF are consistent, *with the limitations discussed in section 4.4.1*.

Chapter 5

Conclusions and Future Work

5.1 Summary

The results presented in this dissertation are concerned with the physical properties and spatial distribution of high-mass star forming regions. These objects are far more difficult to study than their low- and intermediate-mass counterparts because they are rare and their lifetimes are short. Whereas there exist a number of low- and intermediate-mass sources at distances of a few hundreds of pc, the average distance of high-mass star forming regions is of the order of 10^3 pc. As a consequence the current knowledge of high-mass star forming regions is less complete, both on the intrinsic physical properties and their spatial distribution in the Galaxy. `caravela` was developed to help users to address this issue in a statistical manner by providing a simulated Milky Way of star formation that can be used as a benchmark to upcoming surveys.

The aim of this work is to create a computer program capable of simulating the high-mass star forming regions as observed point source catalogues. The catalogues that can currently be simulated result from the IRAS, Herschel and SCUBA-2 sur-

veys. This numerical tool has to allow the user to infer physical properties of the Galactic population of high-mass star forming regions.

The following two questions were the essential problems under analysis during the previous chapters.

1. Stars must be forming constantly. What is the distribution, in the Galaxy, of the birth places of these objects? Massive stars in particular, have short life times compared with low- and intermediate-mass objects, therefore they become rarer and further away thus very difficult to study.
2. What are the physical properties of high-mass star forming regions?

This dissertation includes a description of what has been done and in what context (chapters 2 and 1, respectively). Chapters 3 and 4 analyse the *caravela* output against known and future observational data sets, respectively.

5.2 Main results and conclusions

1. A numerical tool, *caravela*, has been built to study the distribution of high-mass star forming regions in our Galaxy. In an era when large observational surveys are increasingly important, this tool can produce simulated infrared point-source catalogues of high-mass star forming regions *on a Galactic scale*. The general properties of this population of objects can be studied using *caravela*.
2. A model with 4.0×10^4 population 3 objects (with $T = 40$ K and linear size scale $l_{\text{III}} = 5.0 \times 10^6$ AU) yield the following output: 48 detected sources by CUPID, $\approx 42\%$ UCH II candidates (20), 84.73 K-S test result, for a 2×2 degrees box. This is the best-fit model to the IRAS observational data set studied.
3. *caravela*'s model \mathcal{M} (the best-fit model) yields a significantly better constraint in n than in T , i.e. the *true* value of the total number of sources can be estimated with more accuracy than the temperature of the grey-body envelope. In fact, this trend is observed when all the models are considered.
4. In conclusion, n is constrained to the $[1 \times 10^4, 9 \times 10^4]$ interval in total number of objects, and T lies within ≈ 30 K to ≈ 50 K, for the grey-body temperature.
5. A total number of 4.0×10^4 population 3 objects distributed in 4 spiral arms are consistent with the IRAS PSC data under analysis. Population 3 objects are defined as compound sources constituted by a compact YSO surrounded by a larger grey-body object.
6. There is contamination by low and intermediate mass star forming regions of the Wood and Churchwell (1989) region for UCH II regions. This effect was quantified to be of the order of $1 - (\beta/n_{1\&2}) = 1 - (2/9) \approx 77\%$ (using the same notation as in table 3.4, 97) in this *caravela* best fit simulation. If one

considers all the sources in the input catalogue that have a match in the output catalogue, then $1 - (\beta/n_1) = 1 - (2/25) \approx 92\%$. The level of contamination from low- and intermediate-mass objects supported by the analysis of the described models, $\sim 90\%$, *may* help to understand the UCH II lifetime problem. In principle, we could be overestimating the number of UCH II regions in the Galaxy due to the fact that our UCH II samples are polluted with low- to intermediate-mass objects.

7. *caravela* was used to produce single-band and three colour synthetic images of the future Herschel Space Observatory survey Hi-GAL.
8. The final photometry, considering all PACS and SPIRE wavebands combined, resulted in $n_{\text{PACS}} = 46$ sources and $n_{\text{SPIRE}} = 42$ sources, in a 2×2 degree box centred at $\ell = 40$ degrees.
9. In this study, $n = 40$, where n is the number of sources extracted for *all* PACS and SPIRE wavelengths.
10. The analysis of the Herschel colour-colour diagram showed that $\log [110 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -1.6 , with $\sigma_{110 \mu\text{m}/70 \mu\text{m}} = 0.80$. For $\log [170 \mu\text{m}/70 \mu\text{m}]$, the mean value *observed* is -0.8 , with $\sigma_{170 \mu\text{m}/70 \mu\text{m}} = 0.32$. For $\log [360 \mu\text{m}/250 \mu\text{m}]$, the mean value *observed* is -0.43 , with $\sigma_{360 \mu\text{m}/250 \mu\text{m}} = 0.46$. Finally, for $\log [520 \mu\text{m}/250 \mu\text{m}]$, the mean value *observed* is -1.18 , with $\sigma_{520 \mu\text{m}/250 \mu\text{m}} = 0.24$.
11. The cited Herschel colour-colour diagrams revealed that there is a relatively narrow range in the $\log (360 \mu\text{m}/250 \mu\text{m})$. We interpret this result as follows: the SEDs for the model high-mass star forming regions peak at $\sim 250 \mu\text{m}$ therefore a variation in temperature (T controls the shape of the SED curve) would result on a small variation on the fluxes at $250 \mu\text{m}$ and $360 \mu\text{m}$, since

- both values are relatively close to the peak wavelength. On the other hand, the same variation of temperature would yield a significant variation in the other colours.
12. the modelled data set resulting from the Herschel simulation presented here has $n_{1\&2} \approx 2 \times n_{1\&2}^{\text{IRAS}}$, where $n_{1\&2}$ is the number of cross matches between the theoretical catalogue for Herschel instruments and the final cupid list, and $n_{1\&2}^{\text{IRAS}}$ is the number of cross matches between the theoretical catalogue for IRAS caravela simulation and the final photometric catalogue.
 13. The observed companion clump fraction (CCF) is 0.90 ± 0.07 (Thompson *et al.*, 2006) means that *on average* there are observed 2 sources per one IRAS source. Our caravela and the real observed CCF are therefore consistent, *only in a limited sense as discussed in detail in the previous chapters.*
 14. caravela can work together with a diffuse emission model successfully. The images presented make a strong case for the co-adding of caravela and diffuse emission coherent images. Some examples were presented and discussed at the end of this work.

5.3 The future

Section 5.3.1, page 127, presents a few selected synthetic images resulting from the partnership between `caravela` and a model for the diffuse emission (Paladini *et al.*, 2004). This serve to illustrate the interesting future scientific potential of this compound technique.

This work presented `caravela` first and then the code was used to reach a number of results. I believe that the best is still to come for `caravela`, i.e. more time than initially planned has been spent idealising and then developing the code. In the next paragraphs we present both what we think can be achieved using `caravela` in the future, and a set of future modifications to the code.

- Include the possibility to change the viewpoint of each observation. This would result in face-on images of the Milky Way, as well as it would enable to use `caravela` for other galaxies.
- Use `caravela` to model an all sky simulation, co-adding the diffuse background emission.
- Include a built-in clustering function for population 1, 2, and 3 objects.
- Compare the results from the `caravela` native built-in instrument simulator with the SPIRE and PACS simulators from their instruments' teams.
- Test different spiral arms configurations, i.e. altering the number of spiral arms, pitch angle, width and scale height of the arms. Investigate if Russeil (2003) best fit model is consistent with Hi-GAL or even IRAS.
- Wrap-up `caravela` in a user-friendly window-based interface. Put `caravela` freely available on the web.

- Expand the families of sources used, e.g. include a built-in brown dwarf and planetary nebulae populations.
- Extract the background from a *caravela* simulated image with diffuse emission. Then compare this with the image with only the diffuse emission, i.e. the simulated image that results solely and directly from Paladini *et al.* (2004) simulation. Subtract the images and study the resulting image in detail.
- Explore the random number parameter space in more detail, i.e. study what effect changing the stochastic seed has on the results (and conclusions).

5.3.1 `caravela` and the Diffuse Emission Model

Introduction

The aim is to demonstrate that `caravela` can be used to try to simulate a distribution of high-mass star forming regions considering the diffuse background emission, i.e. that `caravela` can be used to complement the diffuse emission model developed by Paladini *et al.* (2004)^a. We think this could be one of the main avenues to be explored in the near future, as the images generated offer a more realistic match to the real observations.

Note that this is the case where the galactocentric ring images described in chapter 2 are used.

5.3.2 `caravela` combined with the diffuse emission

This collaboration exists in the context of Hi-GAL (cf. §1.4, page 16) project, more specifically within the Data Simulation working group. This group is responsible for supplying the Map Making working group with realistic synthetic sky images so that the map making routines are tested. The Map Making working group then gives its output to the Point Source Extraction and Photometry working group. In this chapter, `caravela` maps are combined with the diffuse emission maps at all Herschel bands (PACS and SPIRE).

The results presented in this section are simply qualitative results to illustrate the case for the co-adding of `caravela` and diffuse emission models. In particular, the effect of diffuse structure on compact or point source photometry could be investigated by this approach.

i.e. the images presented make a strong case for the co-adding of `caravela` and

^aAs mentioned in Chapter 4, the diffuse emission model has been developed independently by Paladini *et al.* (2004).

diffuse emission models, but there is a lack of quantitative analysis. The synergies and complementary nature of *caravela* and Paladini *et al.* (2007) result from:

- both models can be used to investigate the distribution of high-mass star forming regions in our Galaxy,
- one of the fundamental limitations in *caravela* is the absence of the diffuse component.

As with the Herschel colour-colour diagrams (figure 4.2, page 112), it will be interesting to do a comparative analysis of the synthetic *caravela* + diffuse emission images with the future real PACS and SPIRE images, e.g. from the Hi-GAL survey. A selection of these images is displayed in the following sections.

5.3.3 The diffuse emission model

Here we describe succinctly the diffuse emission model (Paladini *et al.*, 2007).

Along the Galactic plane, the detected diffuse infrared emission is a blend of radiation arising from dust that is spread over a wide range of distances and Galactic radii. The diffuse emission model simulates this emission by applying an inversion method: the observed diffuse emission is decomposed into radial bins associated with each phase of the interstellar gas. The physical properties of each phase in each bin are determined. A core step in the technique is to employ kinematic distances to assign gas to radial bins. This is why *caravela* generates the galactocentric radius images. This may also be a limitation of the diffuse emission model: due to the inescapable use of kinematic distances only rings can be assumed/used. No spiral structure is permitted.

5.3.4 PACS and SPIRE co-added images

Here we present the maps. Figure 5.1, page 130, represents the same region of the sky at the six Herschel bands. The area observed is a 3×3 degrees square centred at $(\ell, b) = (21, 0)$. The pixel size is $2''$. For the three PACS images the correct synthetic PSF is used. These maps are integrated with the correct PACS and SPIRE transmission filters for both compact and diffuse emission models. All images correspond to 10^6 population 3 objects, but only 10^3 lie within the solid cone of the observation. The Galactic plane is the most remarkable feature in all panels. Note however, that the number of sources and the distribution of the diffuse emission is not symmetrical with respect to $\ell = 0$. There seems to be an over density of emission centred at $(\ell, b) = (20.2, 0.0)$. This is due to the fact that we are looking in a direction parallel to the spiral arm, hence the line of sight lies along the arm (this effect can be visualised in 3D in figure 2.3, page 48).

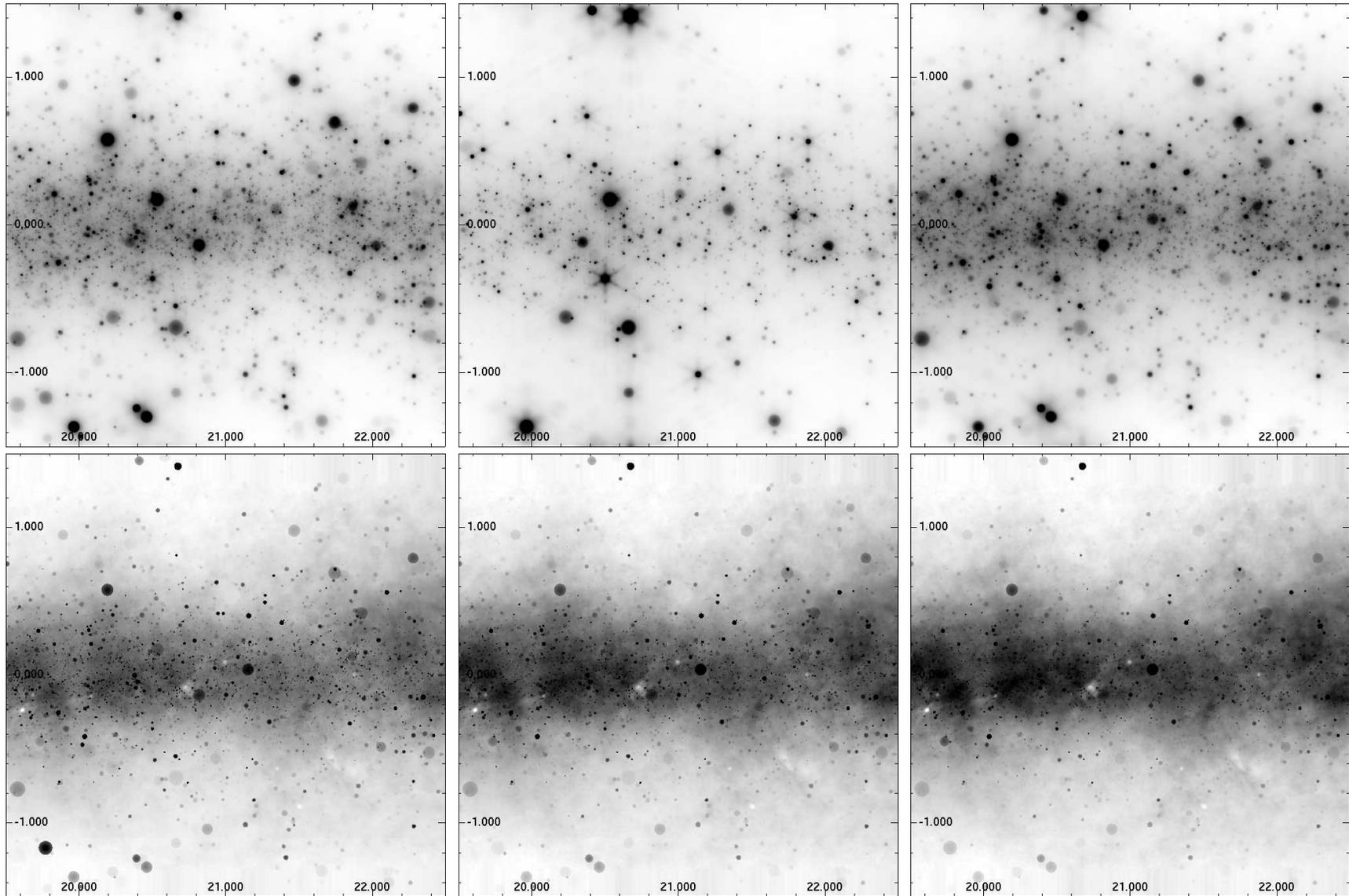


Figure 5.1: PACS and SPIRE + diffuse emission maps (Paladini *et al.*, 2007). The panels show the same region of the sky at the six Herschel bands: $70\ \mu\text{m}$, $110\ \mu\text{m}$, $170\ \mu\text{m}$, $250\ \mu\text{m}$, $360\ \mu\text{m}$, and $520\ \mu\text{m}$, from top to bottom respectively.

5.3.5 PACS and SPIRE three-colour images

Figure 5.2, page 132, present the composite *rgb* images for PACS and SPIRE, co-added with the diffuse emission model (Paladini *et al.*, 2007). These two panels are the most *realistic* output presented in this dissertation. We look forward to compare it with the real Herschel images.

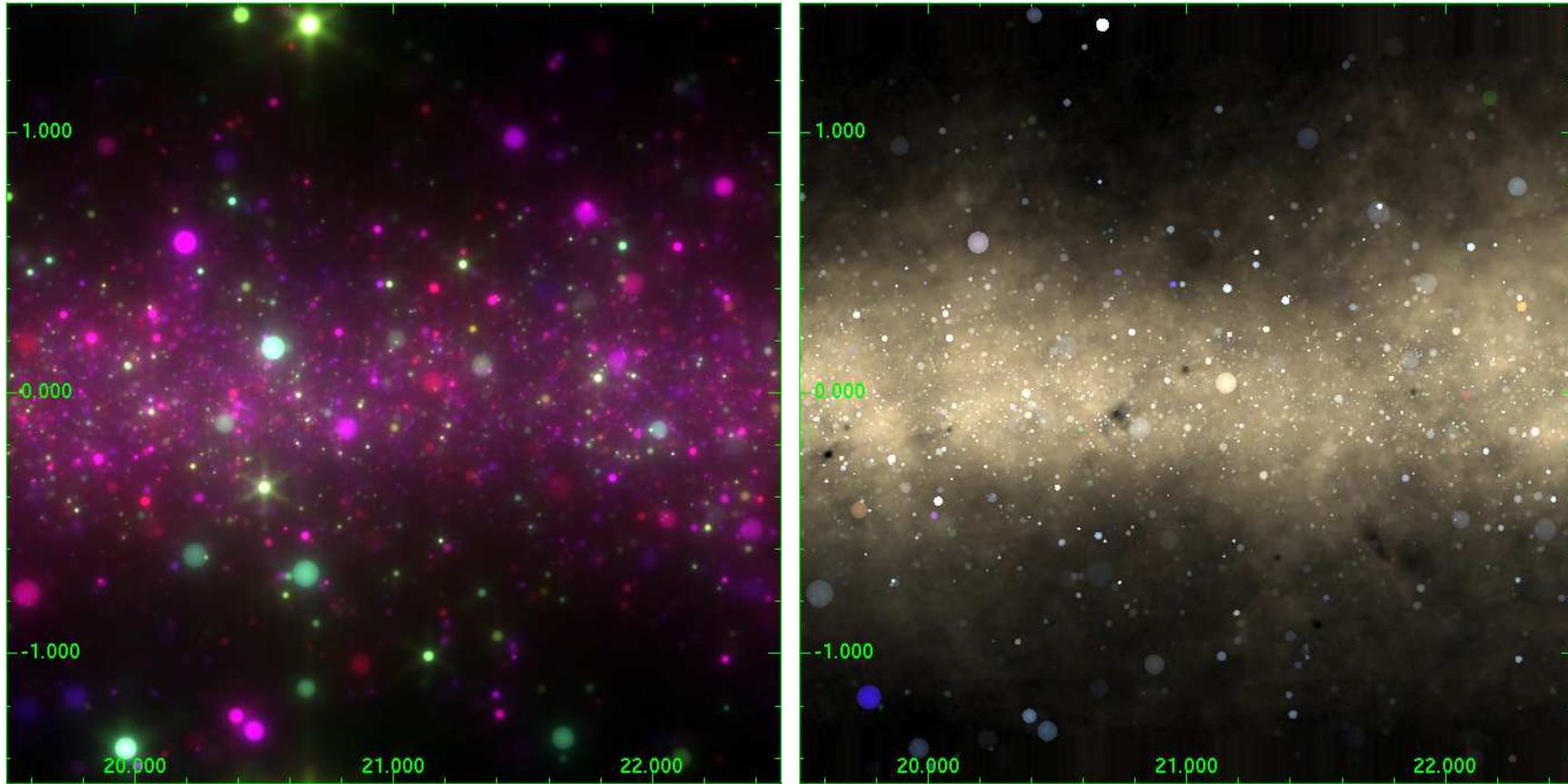


Figure 5.2: PACS and SPIRE + diffuse emission three colour maps (Paladini *et al.*, 2007). *Left*: PACS three colour image. *Right*: SPIRE three colour image.

The Grey Havens

Away from the glowing cities, a diffuse band of light emerging from a countless number of objects can be observed. These objects reach from horizon to horizon and constitute the disk of the Galaxy. *caravela* is limited (just infrared wavebands are reproduced), admittedly incomplete (only high-mass star forming regions are considered) and starts from arguable physical assumptions. It is also an attempt to understand and reproduce the Milky Way, and this gave the author of this work the energy and motivation for this project.

He drew a deep breath. ‘Well, I’m back’, he said.

Appendix A

The Herschel Hi-GAL Milky Way:
the first *observational* science
highlights

LETTER TO THE EDITOR

Clouds, filaments, and protostars: The *Herschel*[★] Hi-GAL Milky Way

S. Molinari¹, B. Swinyard², J. Bally³, M. Barlow⁴, J.-P. Bernard⁵, P. Martin⁶, T. Moore⁷, A. Noriega-Crespo⁸, R. Plume⁹, L. Testi^{10,11}, A. Zavagno¹², A. Abergel¹³, B. Ali¹⁴, L. Anderson¹², P. André¹⁵, J.-P. Baluteau¹², C. Battersby³, M. T. Beltrán¹⁰, M. Benedettini¹, N. Billot¹⁴, J. Blommaert¹⁶, S. Bontemps^{15,17}, F. Boulanger¹³, J. Brand¹⁸, C. Brunt¹⁹, M. Burton²⁰, L. Calzoletti⁵⁰, S. Carey⁸, P. Caselli²², R. Cesaroni¹⁰, J. Cernicharo²³, S. Chakrabarti⁵², A. Chrysostomou²⁵, M. Cohen²⁶, M. Compiègne²⁷, P. de Bernardis²⁸, G. de Gasperis²⁹, A. M. di Giorgio¹, D. Elia¹, F. Faustini⁵⁰, N. Flagey⁸, Y. Fukui³¹, G. A. Fuller³², K. Ganga³³, P. Garcia-Lario³⁴, J. Glenn³, P. F. Goldsmith³⁵, M. Griffin³⁶, M. Hoare²², M. Huang³⁷, D. Ikhenade⁴⁹, C. Joblin⁵, G. Joncas³⁸, M. Juvela³⁹, J. M. Kirk³⁶, G. Lagache¹³, J. Z. Li³⁷, T. L. Lim², S. D. Lord¹⁴, M. Marengo²⁴, D. J. Marshall⁵, S. Masi²⁸, F. Massi¹⁰, M. Matsuura^{4,54}, V. Minier¹⁵, M.-A. Miville-Deschênes¹³, L. A. Montier⁵, L. Morgan⁷, F. Motte¹⁵, J. C. Mottram¹⁹, T. G. Müller⁴¹, P. Natoli²⁹, J. Neves⁴⁰, L. Olmi¹⁰, R. Paladini⁸, D. Paradis⁸, H. Parsons⁴⁰, N. Peretto^{32,15}, M. Pestalozzi¹, S. Pezzuto¹, F. Piacentini²⁸, L. Piazzo⁴⁹, D. Polychroni¹, M. Pomarès¹², C. C. Popescu⁴², W. T. Reach⁸, I. Ristorcelli⁵, J.-F. Robitaille³⁸, T. Robitaille²⁴, J. A. Rodón¹², A. Roy⁶, P. Royer¹⁶, D. Russeil¹², P. Saraceno¹, M. Sauvage¹⁵, P. Schilke³⁰, E. Schisano^{1,53}, N. Schneider¹⁵, F. Schuller⁴³, B. Schulz¹⁴, B. Sibthorpe³⁶, H. A. Smith²⁴, M. D. Smith⁴⁴, L. Spinoglio¹, D. Stamatellos³⁶, F. Strafella²¹, G. S. Stringfellow³, E. Sturm⁴¹, R. Taylor⁴⁵, M. A. Thompson⁴⁰, A. Traficante²⁹, R. J. Tuffs⁴⁶, G. Umana⁴⁷, L. Valenziano⁴⁸, R. Vavrek³⁴, M. Veneziani²⁸, S. Viti⁴, C. Waelkens¹⁶, D. Ward-Thompson³⁶, G. White^{2,51}, L. A. Wilcock³⁶, F. Wyrowski⁴³, H. W. Yorke³⁵, and Q. Zhang²⁴

(Affiliations are available in the online edition)

Received 31 March 2010 / Accepted 14 April 2010

ABSTRACT

We present the first results from the science demonstration phase for the Hi-GAL survey, the *Herschel* key program that will map the inner Galactic plane of the Milky Way in 5 bands. We outline our data reduction strategy and present some science highlights on the two observed $2^\circ \times 2^\circ$ tiles approximately centered at $l = 30^\circ$ and $l = 59^\circ$. The two regions are extremely rich in intense and highly structured extended emission which shows a widespread organization in filaments. Source SEDs can be built for hundreds of objects in the two fields, and physical parameters can be extracted, for a good fraction of them where the distance could be estimated. The compact sources (which we will call cores' in the following) are found for the most part to be associated with the filaments, and the relationship to the local beam-averaged column density of the filament itself shows that a core seems to appear when a threshold around $A_V \sim 1$ is exceeded for the regions in the $l = 59^\circ$ field; a A_V value between 5 and 10 is found for the $l = 30^\circ$ field, likely due to the relatively higher distances of the sources. This outlines an exciting scenario where diffuse clouds first collapse into filaments, which later fragment to cores where the column density has reached a critical level. In spite of core L/M ratios being well in excess of a few for many sources, we find *core* surface densities between 0.03 and 0.5 g cm^{-2} . Our results are in good agreement with recent MHD numerical simulations of filaments forming from large-scale converging flows.

Key words. stars: formation – ISM: structure – ISM: clouds – Galaxy: general

1. Introduction

From the diffuse cirrus to the molecular clouds, onto the formation and death of stars, the Galactic plane is the set where all the phases of the Galaxy life-cycle can be studied in context. Dust, best observed in the infrared and in the submillimeter, cycles through all these phases and is, as such, a privileged tracer for the Galactic ecology. IRAS (Neugebauer et al. 1984) and COBE (Mather et al. 1990) were of tremendous importance in boosting the research in Galactic star formation and interstellar medium to the prominent positions they have today. As remote as they may now seem, however, these missions are only some 20 years away. Since then, a continuing explosion of Galactic plane surveys,

both in the mid-infrared at $\lambda \leq 70 \mu\text{m}$ (Omont et al. 2003; Price et al. 2001; Benjamin et al. 2003; Carey et al. 2009) and in the submillimeter at $\lambda \geq 800 \mu\text{m}$ (Schuller et al. 2009; Rosolowsky et al. 2009), are assembling a picture where the galactic plane has become accessible at sub-30'' resolution over three decades of wavelength. The exception is the critical interval between 70 and 500 μm where the bulk of the cold dust in the Galaxy emits and reaches the peak of its spectral energy distribution (SED). The Hi-GAL key program (*Herschel* Infrared GALactic plane survey) will fill this gap.

Hi-GAL is the key program (KP) of the *Herschel* satellite (Pilbratt et al. 2010) that will use 343 h observing time to carry out a 5-band photometric imaging survey at 70, 160, 250, 350, and 500 μm of a $|b| \leq 1^\circ$ -wide strip of the Milky Way Galactic plane in the longitude range $-60^\circ \leq l \leq 60^\circ$. Hi-GAL is going to be the keystone in the multiwavelength Milky Way, opening up unprecedented opportunities with a promise of breakthroughs

[★] *Herschel* is an ESA space observatory with science instruments provided by European-led Principal Investigator consortia and with important participation from NASA.

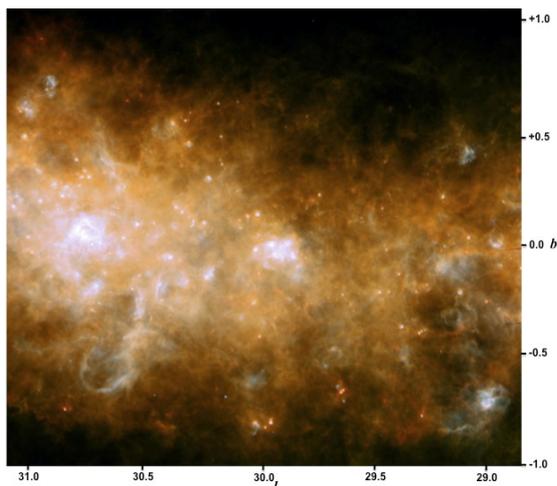


Fig. 1. Three-color image (blue 70 μm , green 160 μm , red 350 μm) of the $2^\circ \times 2^\circ$ field around $l = 30^\circ$.

in several fields of Galactic astronomy. A full description of the survey and its science goals are given elsewhere (Molinari et al. 2010b). This contribution presents the first Hi-GAL data obtained in the *Herschel* science demonstration phase (SDP) and describes a few of the main early results that will be detailed in other contributions in this volume.

2. Observations and data reduction

The *Herschel* PACS (Poglitsch et al. 2010) and SPIRE (Griffin et al. 2010) imaging cameras were used in parallel mode at $60''/\text{s}$ satellite scanning speed to obtain simultaneous 5-band coverage of two $2^\circ \times 2^\circ$ fields approximately centered at $[l, b] = [30^\circ, 0^\circ]$ and $[59^\circ, 0^\circ]$. The detailed description of the observation settings and scanning strategy adopted is given in Molinari et al. (2010b). Data reduction from archival data to Level 1 stage was carried out using the *Herschel* interactive processing environment (HIPE, Ott 2010) using, however, custom reduction scripts that considerably departed from standard processing for PACS (Poglitsch et al. 2010) and, to a lesser extent, for SPIRE (Griffin et al. 2010). Level 1 time ordered data (TODs) were exported from HIPE into FITS files. Further processing including the map generation was carried out using dedicated IDL and FORTRAN codes. Saturation conditions were reached for all detectors only in SPIRE 250 μm and 350 μm images in correspondence with the 3 brightest peaks in the $l = 30^\circ$ field. The prescribed flux correction factors for PACS (Poglitsch et al. 2010) and SPIRE (Swinyard et al. 2010) were applied to the maps since their photometric calibration was carried out using the default calibration tree in HIPE. A detailed description of the entire data processing chain, including the presentation of the maps obtained in the five bands for the two observed fields, can be found in Traficante et al. (in prep.). In the present letter we present in Figs. 1 and 2 the three-color images obtained using the 70, 160, and 350 μm data ($l = 30^\circ$ and $l = 59^\circ$, respectively).

These amazing maps convey the immediate impression of extended filamentary structures dominating the emission on all spatial scales. Measurements of the standard deviation of the signal at all wavelengths in the lowest brightness regions of the $l = 59^\circ$ field yield average values a factor two higher than the sensitivity predictions for point source sensitivity from the HSpot time estimator for all bands except at 70 μm where the predicted limit is effectively reached, confirming that the noise in our maps is dominated by the cirrus confusion at all wavelengths. A more detailed quantitative analysis is presented by Martin et al. (2010).

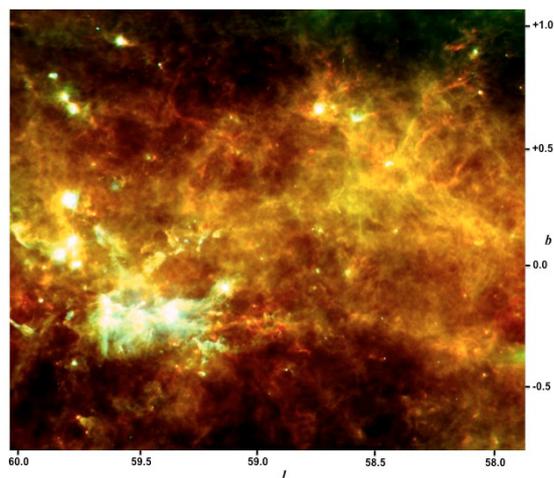


Fig. 2. Three-color image (blue 70 μm , green 160 μm , red 350 μm) of the field around $l = 59^\circ$.

3. Results and science highlights

3.1. From IRDCs to mini-starburst and their impact on the ISM

Herschel's ability to observe such large areas with unprecedented wavelength coverage and extraordinary signal dynamical range allows us to image simultaneously progenitors clouds for massive protoclusters to entire clusters of young stellar objects (YSOs) in active star forming regions, while also measuring the effect of their strong stellar winds and powerful outflows on the surrounding medium.

Infrared dark clouds (IRDCs) have received considerable attention in recent years (e.g. Rathborne et al. 2006; and Peretto & Fuller 2009) as potential sites for precursors of cluster forming sites. Found in silhouette against the bright mid-IR background, they shine in emission with *Herschel*. Peretto et al. (2010) shows how temperature effectively decreases from ambient values (20–30 K) down to $T = 8\text{--}15$ K toward the densest ($\sim 10^{23}$ cm^{-2}) peaks of these objects, resolving further temperature substructures that can be proxies for subsequent fragmentation.

At the other end of the massive star formation timeline, we find W43, visible in the left portion of Fig. 1, as an outstanding case of Galactic mini-starburst. Detailed SED construction and luminosity estimates allow us to assess the very early evolutionary stage of the most luminous and massive YSOs in the region. It is remarkable how the same images show a prominent ridge extending southward which encompasses a 70 pc-wide large cavity excavated by the W43 cluster and which possibly triggers further star formation (Bally et al. 2010). Triggered star formation in less extreme environments can also be studied in statistically significant fashion modeling the SED of the sources found in correspondence of the multitude of HII-driven bubble-like structures found in the images, as shown for the bubble N49 by Zavagno et al. (2010).

Feedbacks from massive star formation, together with the intricate relationship between the interstellar radiation field and molecular clouds, are at the origin of the observed complexity of the ISM emission structure, where temperature ranges from ~ 10 K of pre-stellar cores to the $\sim 40\text{--}50$ K of the photodissociation regions (Bernard et al. 2010).

3.2. Census of compact sources

The extraction of compact sources is quite a challenging task in these fields, which we faced using a novel approach based on the

study of the multidirectional second derivatives in the image to aid in source detection and size estimate, and subsequent constrained multi-Gaussian fitting. This approach greatly increases the dynamical range between compact sources and diffuse emission, irrespective of the local absolute value of the emission. The method is fully described elsewhere (Molinari et al. 2010a) and has been applied for this first attempt to generate source catalogs. As the thresholding for source detection is done on the *curvature* image (Molinari et al. 2010a), the S/N of the detected sources is determined a posteriori measuring the ratio of the source peak flux over the rms of the residuals after the Gaussian fit. Source catalogs were generated for the two fields and for the 5 bands and are made available in tabular form in the online version of the paper. Catalogs completeness was estimated with artificial source injection experiments, and the peak flux levels for 80% completeness for the 70, 160, 250, 350 and 500 μm photometry are [0.5, 4.1, 4.1, 3.2, 2.5] Jy/beam for the $l = 30^\circ$ field, and [0.06, 0.9, 0.7, 0.7, 0.8] Jy/beam for the $l = 59^\circ$ field. The difference is entirely compatible with the very different intensity regimes of the underlying diffuse emission in the two fields.

Estimating the source's physical properties requires that detection in the various band catalogs are merged in coherent SEDs, a process that can only be done coarsely in this early stage, but which is nonetheless useful for isolating 528 sources in the $l = 30^\circ$ field and 444 sources in the $l = 59^\circ$ field (see Elia et al. 2010). The two observed fields encompass emission from regions at very different distances. In a considerable effort, which involved a critical re-evaluation of available data and evidence, and the collection of additional data for hundreds of previously unknown objects, Russeil et al. (in prep.) provide recommended distances for a fraction of the detected sources (312 out of 528, and 91 out of 444 sources for the two fields, respectively) for which the derivation of masses and luminosities is possible. Adopting standard prescriptions for Class 0 classification ($L_{\lambda \geq 350 \mu\text{m}}/L_{\text{bol}} \geq 0.005$, André et al. 2000) results in almost the totality of sources being Class 0 (90 out of 91 sources in $l = 59^\circ$ and 306 out of 312 in $l = 30^\circ$, see Elia et al. 2010).

3.3. Filamentary star formation

The most extraordinary feature exhibited by the *Herschel* maps is the *ubiquitous* pattern of filaments in the ISM structure. This is more apparent when we enhance the contrast of the filaments using the same method (see Sect. 3.2) as used for the source detection (Molinari et al. 2010a). Here we start from the ∂^2 derivatives carried out in four directions (x , y , and the two diagonals), as for the standard detection method, and then create another image F of the same size so that the maximum curvature is selected for each pixel: $F_{ij} = \max[\partial^2 x_{ij}, \partial^2 y_{ij}, \partial^2 D_{1ij}, \partial^2 D_{2ij}]$. In this way we are following the direction of maximum curvature pixel-by-pixel for all compact features in the image. We show in Fig. 3 the result of this processing on the $l = 59^\circ$ field at 250 μm .

The maps clearly show an interconnected maze of filaments at different levels of brightness (e.g. different levels of emission intensity and curvature), and the striking aspect is that the compact sources detected at 250 μm are distributed for the most part along the brightest filaments. Interestingly, a similar scenario was also reported for Taurus by Goldsmith et al. (2008) where the physical conditions and spatial scales involved are radically different. Since the source integrated fluxes are estimated by fitting Gaussians on top of planar plateaus, the values of the local background at every wavelength are a by-product of our source extraction and, after applying the absolute correction factors as recommended by Bernard et al. (2010) and subtracting the foreground contribution estimated using

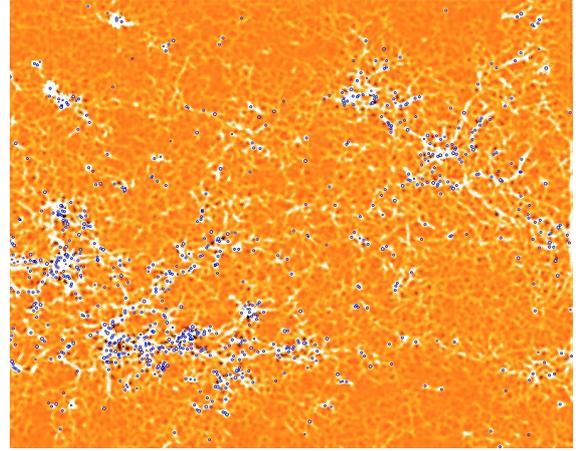


Fig. 3. Multidirectional second-derivative image (see text) of the $l = 59^\circ$ field at 250 μm . The blue circles represent the compact sources detected at 250 μm . The filamentary structure of the ISM appears at various levels of intensity, i.e. curvature, and it is striking how the detected compact objects are for the most part distributed only along the brightest filaments.

Bohlin (1975), can be used to estimate the local beam-averaged column density in the hosting filaments. The relationship between the mass of the detected cores (when the SED is reliable and the distance is known, see previous paragraph) and the local beam-averaged H_2 filament column density is reported in Fig. 4. The points for the $l = 59^\circ$ field mostly lie in the range of column densities ($10^{21} \text{ cm}^{-2} \leq N(\text{H}_2) \leq 10^{22} \text{ cm}^{-2}$) that corresponds to $1 \text{ mag} \leq A_V \leq 10 \text{ mag}$, values that are entirely reasonable for the transition regime between diffuse ISM and dense molecular clouds (Cambresy 1999; Snow & McCall 2006). Higher values of $N(\text{H}_2)$ are found for the points for the $l = 30^\circ$ field, most likely due to the larger relative distances of the sources in this latter field. The core masses are spread between 1 and $10^4 M_\odot$, with no indication of a correlation between the two quantities. The strong impression, however, is that of a threshold at $A_V \sim 1 \text{ mag}$ for the $l = 59^\circ$ field above which dense cores are found, a threshold that is evidently exceeded only in bright filaments (Fig. 3). More in particular, the $A_V \sim 1 \text{ mag}$ threshold corresponds to $\sim 17 M_\odot \text{ pc}^{-2}$ in molecular hydrogen, which is surprisingly close to the $10 M_\odot \text{ pc}^{-2}$ value that Krumholz et al. (2009) find critical for the dust content in HI clouds to efficiently shield the cloud interior from external FUV field and allow effective H_2 formation. It is tempting to relate the appearance of clumps to an extinction regime where the HI/ H_2 boundary shields the cloud interiors from interstellar FUV field, causing the photoelectric heating efficiency to drop considerably and causing in turn a drop in dust and gas temperature (Tielens & Hollenbach 1985). This threshold value seems to be of the order of $A_V \sim 5 \div 10$ for the $l = 30^\circ$ field most likely due to the relatively larger distances of sources in this field (Russeil et al., in prep.).

The ubiquitousness of dense filaments in the ISM, the high degree of association between bright filaments and cores, and the suggestion of a column density threshold for the appearance of dense cores, all appear to coherently support a formation scenario that starts with the condensation of diffuse clouds into long filaments. As the column density increases, a threshold is exceeded and denser star-forming (or potentially star forming) condensations start to appear. A preliminary association with *Spitzer* 24 μm counterparts (Elia et al. 2010) suggests that our detected sources may be a mixture of protostellar and pre-stellar objects, although more work will be needed to ascertain the composition of this mixture.

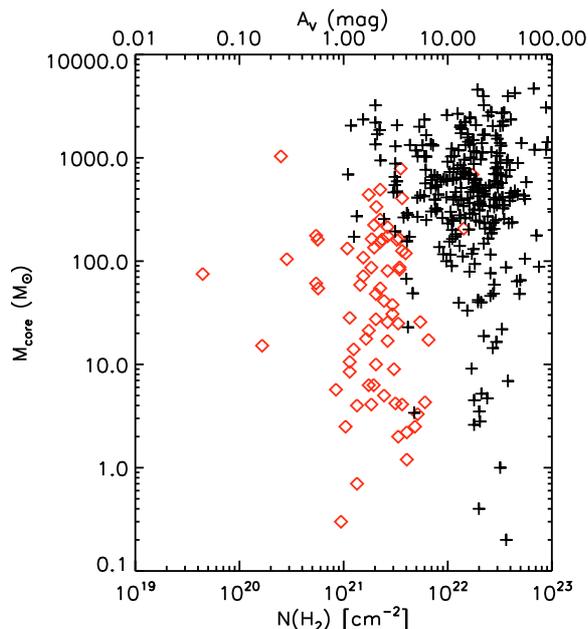


Fig. 4. Mass of detected cores as a function of the hosting filament local column density, for the cores where the distance is known and the mass could be estimated. The red diamonds represent the cores in the $l = 59^\circ$ region, while the plus signs represent the cores from the $l = 30^\circ$ region. The top X -axis represents the equivalent A_V .

Since our source extraction also yields a measure of the core sizes, we are in the position to estimate their surface density Σ . These values appear to be on average a factor 3–5 higher than the column densities of the underlying filaments as reported in Fig. 4, spanning a range of $0.03 \text{ g cm}^{-2} \leq \Sigma \leq 0.5 \text{ g cm}^{-2}$, after changing units, with a mean value of 0.1 g cm^{-2} . It is puzzling that the number of cores exceeding L/M ratios of a few (the exact number depending on the core masses), corresponding to the critical surface density threshold of 1 g cm^{-2} for the formation of massive stars (Krumholz & McKee 2008), is not consistent with very few of the cores actually exceeding that critical threshold (see also Elia et al. 2010). The difference of a factor ~ 2 between the dust opacities that we used compared to Krumholz & McKee (2008) is not sufficient to reconcile this apparent discrepancy. This result deserves more attention and needs to be confirmed in the future with more detailed and accurate analysis.

Testing of large-scale “dynamical” star formation scenarios (e.g., Hartmann et al. 2001), where filaments are formed in the post-shock regions of large HI converging flows, is one of the original science goals of Hi-GAL. It is remarkable how the predictions from recent MHD numerical simulations (Banerjee et al. 2009) of formation and subsequent fragmentation of filaments, agree with our results. Besides the morphological resemblance of these simulations with the structures we see in our *Herschel* maps (Fig. 3), there is striking agreement of their predictions with the $N(\text{H})$ regime we measured for our core-hosting filaments, as well as with the mass regime of the cores being formed.

Instability and fragmentation of dense filaments has also been investigated in the context of helical magnetic fields enclosing the filaments by Fiege & Pudritz (2000); interestingly, the models predict the formation of regularly spaced condensations at length scales that depend on the properties of the magnetic field, the velocity dispersion, and density of the filament. The predicted length scale for filament velocity dispersion of 0.5 km s^{-1} and density of 10^4 cm^{-3} is 2.8 pc, and curiously this is not at all far from the median distance of each source to its

nearest neighbor: $\sim 1.8 \text{ pc}$ for the $250 \mu\text{m}$ sources in the $l = 59^\circ$ field for the sources’ average distance (Russeil et al., in prep.). It is also interesting that the typical fragmentation length scale decreases with increasing filament density, in broad qualitative agreement with a higher spatial density of sources in the brightest filaments. This clearly deserves further investigation to be confirmed as a viable hypothesis.

4. Conclusions

The first science highlights presented in this paper, as well as in the accompanying papers in this volume and elsewhere, show that owing to its optimal use of unique *Herschel* characteristics of wavelength coverage, spatial resolution and mapping speed, the Hi-GAL survey has the potential to lead to a quantum leap in our understanding of large-scale Galactic star formation from cloud to cluster-forming clump formation and of the evolution of protoclusters and massive protostars.

The outstanding feature emerging from these first images is the impressive and ubiquitous ISM filamentary nature. Dense cores seem to appear when a certain beam-averaged column density threshold is exceeded in close spatial association with these filaments.

Acknowledgements. Data processing and map production has been possible thanks to generous support from the Italian Space Agency via contract I/038/080/0. We are indebted to Nicola Giordano for the production of the trichromatic overlay images. Data presented in this paper were also analyzed using The *Herschel* interactive processing environment (HIPE), a joint development by the *Herschel* Science Ground Segment Consortium, consisting of ESA, the NASA *Herschel* Science Center, and the HIFI, PACS, and SPIRE consortia.

References

- André, P., Ward-Thompson, D., & Barsony, M. 2000, in *Protostars and Planets IV*, ed. S. S. Russell, V. Mannings, & A. P. Boss, 59
- Bally, J., Anderson, L., Battersby, C., et al. 2010, *A&A*, 518, L90
- Banerjee, R., Vázquez-Semadeni, E., Hennebelle, P., & Klessen, R. S. 2009, *MNRAS*, 398, 1082
- Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, *PASP*, 115, 953
- Bernard, J. P., Paradis, D., Marshall, D. J., et al. 2010, *A&A*, 518, L88
- Bohlin, R. C. 1975, *ApJ*, 200, 402
- Cambresy, L. 1999, *A&A*, 345, 965
- Carey, S. J., Noriega-Crespo, A., Mizuno, D. R., et al. 2009, *PASP*, 121, 76
- Elia, D., Molinari, S., Schisano, E., et al. 2010, *A&A*, 518, L97
- Fiege, J. D., & Pudritz, R. E. 2000, *MNRAS*, 311, 105
- Goldsmith, P., Heyer, M., Narayanan, G., et al. 2008, *ApJ*, 680, 428
- Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, *A&A*, 518, L3
- Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, *ApJ*, 562, 852
- Krumholz, M., & McKee, C. F. 2008, *Nature*, 451, 1082
- Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, *ApJ*, 693, 216
- Martin, P. G., et al., G., et al., G., et al. 2010, *A&A*, 518, L105
- Mather, J. C., Hauser, M. G., Bennett, C. L., et al. 1990, in *Observatories in Earth Orbit and Beyond: Proceedings of the 123rd. ed. Y. Kondo (Dordrecht, Boston: Kluwer Academic Publishers) IAU Colloq.*, 9
- Molinari, S., Faustini, F., Schisano, E., Pestalozzi, M., & Di Giorgio, A. 2010a, *A&A*, submitted
- Molinari, S., Swinyard, B., Bally, J., et al. 2010b, *PASP*, 122, 314
- Neugebauer, G., Soifer, B. T., Beichman, C. A., et al. 1984, *Science*, 224, 14
- Omont, A., Gilmore, G. F., Alard, C., et al. 2003, *A&A*, 403, 975
- Ott, S. 2010, in *Astronomical Data Analysis Software and System XIX*, ed. M. O. Y. Mizumoto, K.-I. Morita, ASP Conf. Ser., in press
- Peretto, N., & Fuller, G. A. 2009, *A&A*, 505, 405
- Peretto, N., Fuller, G. A., Plume, R., et al. 2010, *A&A*, 518, L98
- Pilbratt, G. L., et al. 2010, *A&A*, 518, L1
- Poglitsch, A., et al., 2010, *A&A*, 518, L2
- Price, S. D., Egan, M. P., Carey, S. J., et al. 2001, *AJ*, 121, 2819
- Rathborne, J. M., Jackson, J. M., & Simon, R. 2006, *ApJ*, 641, 389
- Rosolowsky, E., Dunham, M. K., et al. 2009, *ApJS*, 188, 123
- Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, *A&A*, 504, 415
- Snow, T. P., & McCall, B. J. 2006, *ARA&A*, 44, 367
- Swinyard, B. M., et al. 2010, *A&A*, 518, L4
- Tielens, A. G. G. M., & Hollenbach, D. 1985, *ApJ*, 291, 722
- Zavagno, A., Anderson, L., Russeil, D., et al. 2010, *A&A*, 518, L101

-
- ¹ INAF-Istituto Fisica Spazio Interplanetario, via Fosso del Cavaliere 100, 00133 Roma, Italy
e-mail: sergio.molinari@ifsi-roma.inaf.it
- ² STFC, Rutherford Appleton Labs, Didcot, UK
- ³ Center for Astrophysics and Space Astronomy (CASA), Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, USA
- ⁴ Department of Physics and Astronomy, University College London, London, UK
- ⁵ Université de Toulouse, UPS, CESR, and CNRS, UMR5187, Toulouse, France
- ⁶ Department of Astronomy & Astrophysics, University of Toronto, Toronto, Canada
- ⁷ Astrophysics Research Institute, Liverpool John Moores University, UK
- ⁸ *Spitzer* Science Center, California Institute of Technology, Pasadena, CA, USA
- ⁹ Department of Physics & Astronomy, University of Calgary, Canada
- ¹⁰ INAF - Osservatorio Astrofisico di Arcetri, Firenze, Italy
- ¹¹ European Southern Observatory, Garching bei Muenchen, Germany
- ¹² LAM, Université de Provence, Marseille, France
- ¹³ Institut d'Astrophysique Spatiale, Université Paris-Sud, Orsay, France
- ¹⁴ NASA *Herschel* Science Center, Caltech, Pasadena, CA, USA
- ¹⁵ Laboratoire AIM, CEA/DSM – INSU/CNRS – Université Paris Diderot, IRFU/Sap CEA-Saclay, 91191 Gif-sur-Yvette, France
- ¹⁶ Institute for Astronomy, Katholieke Universiteit Leuven, Leuven, Belgium
- ¹⁷ LAB/CNRS, Université de Bordeaux, BP89, 33271 Floirac Cedex, France
- ¹⁸ INAF - Istituto di Radioastronomia, Bologna, Italy
- ¹⁹ School of Physics, University of Exeter, Stocker Road, Exeter, EX4 4QL, UK
- ²⁰ School of physics, University of New South Wales, Australia
- ²¹ Dipartimento di Fisica, Università del Salento, Lecce, Italy
- ²² School of Physics and Astronomy, University of Leeds, Leeds, UK
- ²³ Centro de Astrobiología, CSIC-INTA. Madrid, Spain
- ²⁴ Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, USA
- ²⁵ Joint Astronomy Center, Hilo, Hawaii
- ²⁶ Radio Astronomy Lab., UCB, Berkeley, CA, USA
- ²⁷ Canadian Institute for Theoretical Astrophysics, University of Toronto, Toronto, Canada
- ²⁸ Dipartimento di Fisica, Università di Roma 1 “La Sapienza”, Roma, Italy
- ²⁹ Dipartimento di Fisica, Università di Roma 2 “Tor Vergata”, Roma, Italy
- ³⁰ I. Physikalisches Institut der Universität zu Köln, Zùlpicher Str. 77, 50937 Köln, Germany
- ³¹ Department of Astrophysics, Nagoya University, Nagoya, Japan
- ³² Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Manchester, M13 9PL, UK
- ³³ APC/Université Paris 7 Denis Diderot/CNRS, Bâtiment Condorcet, 10, Rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France
- ³⁴ *Herschel* Science Centre, European Space Astronomy Centre, Villafranca del Castillo. Apartado de Correos 78, 28080 Madrid, Spain
- ³⁵ Jet Propulsion Laboratory, Pasadena, USA
- ³⁶ School of Physics and Astronomy, Cardiff University, Cardiff, UK
- ³⁷ National Astronomical Observatories, Chinese Academy of Sciences, Beijing, PR China
- ³⁸ Departement de Physique, Université Laval, Québec, Canada
- ³⁹ Department of Physics, University of Helsinki, Finland
- ⁴⁰ Centre for Astrophysics Research, Science and Technology Research Institute, University of Hertfordshire, Hatfield, UK
- ⁴¹ MPE-MPG, Garching bei München, Germany
- ⁴² Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK
- ⁴³ MPIfR-MPG, Bonn, Germany
- ⁴⁴ Centre for Astrophysics & Planetary Science, University of Kent, Canterbury, UK
- ⁴⁵ Center for Radio Astronomy, University of Calgary, Calgary, Canada
- ⁴⁶ Max-Planck-Institut für Kernphysik, Heidelberg, Germany
- ⁴⁷ INAF - Osservatorio Astrofisico di Catania, Catania, Italy
- ⁴⁸ INAF Istituto di Astrofisica Spaziale e Fisica Cosmica, Bologna, Italy
- ⁴⁹ Dipartimento di Scienza e Tecnica dell'Informazione e della Comunicazione, Università di Roma 1 “La Sapienza”, Roma, Italy
- ⁵⁰ ASI Science Data Center, 00044 Frascati (Roma), Italy
- ⁵¹ Department of Physics and Astronomy, The Open University, Milton Keynes, UK
- ⁵² Astronomy Department, UCB Berkeley, CA, USA
- ⁵³ Dipartimento di Fisica, Università di Napoli “Federico II”, Napoli, Italy
- ⁵⁴ Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking, Surrey RH5 6NT, UK

Appendix B

Hi-GAL: the Herschel infrared Galactic Plane Survey

Hi-GAL: the *Herschel* infrared Galactic Plane Survey

S. Molinari^{1,2}, B. Swinyard⁴⁴, J. Bally⁶, M. Barlow⁸, J.-P. Bernard⁹, P. Martin³¹, T. Moore⁴⁶, A. Noriega-Crespo¹⁹, R. Plume⁴⁸, L. Testi^{21,55}, A. Zavagno⁷
and

A. Abergel³, B. Ali⁴, P. André⁵, J.-P. Baluteau⁷, M. Benedettini², O. Berné¹⁰, N. P. Billot¹¹, J. Blommaert¹², S. Bontemps^{5,13}, F. Boulanger¹⁴, J. Brand¹⁵, C. Brunt¹⁶, M. Burton¹⁷, L. Campeggio¹⁸, S. Carey¹⁹, P. Caselli²⁰, R. Cesaroni²¹, J. Cernicharo²², S. Chakrabarti²³, A. Chrysostomou^{24,25}, C. Codella²¹, M. Cohen²⁶, M. Compiègne²⁷, C. J. Davis²⁴, P. de Bernardis²⁸, G. de Gasperis²⁹, J. Di Francesco³⁰, A. M. di Giorgio², D. Elia¹⁸, F. Faustini², J. F. Fischera³¹, Y. Fukui³², G. Fuller³³, K. Ganga³⁴, P. Garcia-Lario³⁵, M. Giard⁹, G. Giardino³⁶, J. Glenn⁶, P. Goldsmith³⁷, M. Griffin³⁸, M. Hoare²⁰, M. Huang³⁹, B. Jiang⁴⁰, C. Joblin⁹, J. Joncas⁴¹, M. Juvela⁴², J. Kirk³⁸, G. Lagache⁴³, J. Z. Li³⁹, T. L. Lim⁴⁵, S. D. Lord⁴, P. W. Lucas²⁵, B. Maiolo¹⁸, M. Marengo²³, D. Marshall⁹, S. Masi²⁸, F. Massi²¹, M. Matsuura^{8,45}, C. Meny⁹, V. Minier⁵, M.-A. Miville-Deschênes³, L. Montier⁹, F. Motte⁵, T. G. Müller⁴⁷, P. Natoli²⁹, J. Neves²⁵, L. Olmi²¹, R. Paladini¹⁹, D. Paradis¹⁰, M. Pestalozzi², S. Pezzuto², F. Piacentini³⁵, M. Pomarès⁷, C. C. Popescu⁴⁹, W. T. Reach⁴, J. Richer⁵⁰, I. Ristorcelli⁹, A. Roy³¹, P. Royer¹², D. Russeil⁷, P. Saraceno², M. Sauvage⁵, P. Schilke⁵¹, N. Schneider-Bontemps⁵, F. Schuller⁵¹, B. Schultz⁴, D. S. Shepherd⁵², B. Sibthorpe³⁸, H. A. Smith²³, M. D. Smith⁵³, L. Spinoglio², D. Stamatellos³⁸, F. Strafella¹⁸, G. Stringfellow⁶, E. Sturm⁴⁷, R. Taylor⁵⁴, M. A. Thompson²⁵, R. J. Tuffs⁵⁶, G. Umana⁵⁷, L. Valenziano⁵⁸, R. Vavrek³⁵, S. Viti⁸, C. Waelkens¹², D. Ward-Thompson³⁸, G. White^{44,59}, F. Wyrowski⁵¹, H. W. Yorke³⁷, Q. Zhang⁶⁰

ABSTRACT

Hi-GAL, the *Herschel* infrared Galactic Plane Survey, is an Open Time Key Project of the *Herschel* Space Observatory. It will make an unbiased photometric survey of the inner Galactic Plane by mapping a two-degree wide strip in the longitude range $|l| < 60^\circ$ in five wavebands between $70\mu\text{m}$ and $500\mu\text{m}$. The aim of Hi-GAL is to detect the earliest phases of the formation of molecular clouds and high-mass stars and to use the optimum combination of *Herschel* wavelength coverage, sensitivity, mapping strategy and speed to deliver a homogeneous census of star-forming regions and cold structures in the interstellar medium. The resulting representative samples will yield the variation of source temperature, luminosity, mass and age in a wide range of Galactic environments at all scales from massive YSOs in protoclusters to entire spiral arms, providing an evolutionary sequence for the formation of intermediate and high-mass stars. This information is essential to the formulation of a predictive global model of the role of environment and feedback in regulating the star-formation process. Such a model is vital to understanding star formation on galactic scales and in the early Universe. Hi-GAL will also provide a science legacy for decades to come with incalculable potential for systematic and serendipitous science in a wide range of astronomical fields, enabling the optimum use of future major facilities such as JWST and ALMA.

Subject headings: ISM–star formation–high-mass stars–IR–*Herschel*

¹PI of the Hi-GAL project,
²INAF-IFSI, Fosso del Cavaliere 100, 00133 Roma, Italy
³Institut d’Astrophysique Spatiale, Universit Paris-Sud, Orsay, France
⁴NHSC/IPAC/Caltech, USA
⁵SAP CEA Saclay, France
⁶Center for Astrophysics and Space Astronomy (CASA), Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, USA
⁷LAM-OAMP, Marseille, France
⁸University College London, UK
⁹Centre d’Etude Spatiale du Rayonnement, Toulouse, France
¹⁰CESR, CNRS et Université de Toulouse 3, France
¹¹IPAC, USA
¹²Institute for Astronomy, Leuven, Belgium
¹³Obs. Bordeaux, France
¹⁴Institut d’Astrophysique Spatiale, France
¹⁵INAF - Istituto di Radioastronomia, Bologna, Italia
¹⁶University of Exeter, UK
¹⁷University of New South Wales, Australia
¹⁸Università del Salento, Lecce, Italy
¹⁹Spitzer Science Center, California Institute of Technology, USA
²⁰School of Physics and Astronomy, University of Leeds, UK
²¹INAF - Osservatorio Astrofisico di Arcetri, Italy
²²CSIC. Department of Infrared and Molecular Astrophysics. Madrid, Spain
²³Harvard-CfA, USA
²⁴Joint Astronomy Centre, Hawaii, USA
²⁵University of Hertfordshire, UK
²⁶Univ. of California, Berkeley, USA
²⁷Canadian Institute for Theoretical Astrophysics, Toronto, Canada
²⁸Dip. Fisica, Univ. Roma 1 ”La Sapienza”, Rome, Italy
²⁹Dipartimento di Fisica, Università di Roma 2 ”Tor Vergata”, Rome, Italy
³⁰National Research Council of Canada
³¹University of Toronto, CITA, Canada
³²Nagoya University, Japan
³³Jodrell Bank Centre for Astrophysics, University of Manchester, UK
³⁴APC/Université Paris 7, France
³⁵Herschel Science Centre Community Support Group Leader at ESAC/ESA, Madrid, Spain
³⁶ESA - Research and Scientific Support Department, ESTEC, The Netherlands
³⁷Jet Propulsion Laboratory, Pasadena, USA
³⁸Cardiff University School of Physics and Astronomy, UK
³⁹National Astronomical Observatories of CAS, China
⁴⁰Department of Astronomy, Beijing Normal University,

1. Introduction

Dust is the most robust tracer of the ‘Galactic ecology’ - the cycling of material from dying stars to the ionized, atomic, and molecular phases of the ISM, into star forming cloud cores, and back into stars. While atoms, ions, and molecules are imperfect tracers because they undergo complex phase changes, chemical processing, depletion onto grains, and are subject to complex excitation conditions, dust is relatively stable in most phases of the ISM. It is optically thin in the Far Infrared (FIR) over most of the Galaxy, so that its emission and absorption simply depend on emissivity, column density and temperature. Cold dust in particular ($10\text{ K} \leq T \leq 40\text{ K}$) traces the bulk of non-stellar baryonic mass in all of the above “habitats” of the Galactic ecosystem.

Temperature and luminosity and, as their by-product, mass of cold dust measured over the entire Galactic Plane (GP), are, at sub-parsec resolution, the critical quantities needed to formulate a global predictive model of the cycling process between the Galactic ISM and star formation. This process drives the galactic ecology in normal spirals as well as the enhanced star-formation rates of starburst galaxies and mergers and a quantitative

China

⁴¹Université Laval, Canada
⁴²Observatory, University of Helsinki, Finland
⁴³IAS, Paris, France
⁴⁴STFC, Rutherford Appleton Laboratory, UK
⁴⁵National Astronomical Observatory of Japan
⁴⁶Astrophysics Research Institute, Liverpool John Moores University, UK
⁴⁷MPE Garching, Germany
⁴⁸Department of Physics & Astronomy, University of Calgary, Canada
⁴⁹University of Central Lancashire, UK
⁵⁰Cavendish Labs, Cambridge, UK
⁵¹MPIfR, Bonn, Germany
⁵²National Radio Astronomy Observatory, Socorro, USA
⁵³University of Kent, UK
⁵⁴Centre for Radio Astronomy, University of Calgary, Canada
⁵⁵European Southern Observatory, Garching, Germany
⁵⁶Infrared Astrophysics, MPI-Kernphysik, Germany
⁵⁷INAF-Osservatorio Astrofisico di Catania, Italy
⁵⁸INAF IASF-Bologna, Italy
⁵⁹The Open University
⁶⁰Smithsonian Astrophysical Observatory, USA

understanding of it is needed in order to follow the formation and evolution of galaxies throughout the cosmos. The adequate measurement of these key quantities has been beyond the capabilities of the previous mid- to far-infrared surveys of the Galactic Plane (IRAS, Neugebauer et al. 1984; MSX, Price et al. 2001; COBE/DIRBE and FIRAS, e.g. Sodroski et al. 1994; ISO, Omont et al. 2003; Spitzer, Benjamin et al. 2003; Carey et al. 2009) either due to limited wavelength coverage and/or inadequate spatial resolution leading to confusion. The balloon-borne BLAST experiment (Pascale et al. 2008) implements Herschel/SPIRE detector arrays and is providing exciting anticipations of what Herschel will do. The AKARI satellite (Murakami et al. 2007) improves over IRAS, and results from its FIR photometric mapping of the GP are eagerly awaited.

Observing the distribution and temperature of dust across the Galaxy will resolve many current debates such as the modes of formation of molecular clouds and high-mass stars.

Molecular clouds are traditionally thought to follow a “slow formation” scheme, where distributed material is accumulated by large-scale perturbations such as the passage of a spiral arm. Shielding by dust and surface reactions on grains promotes the $\text{H I} \rightarrow \text{H}_2$ transition, which in turn allows the formation of other molecules that cool the cloud. Gravity, mediated by magnetic fields, leads to star formation. In this scenario cloud lifetimes are about ~ 30 Myr (Leisawitz et al. 1989). This picture has difficulty explaining the absence of quiescent, non star forming GMCs (however, see Palla & Galli 1997) and the continuous regeneration of turbulence needed to support GMCs for many crossing times. Alternatively, a “fast formation” scenario has been proposed (Hartmann et al. 2001) in which most MCs are transient, short-lived structures (Stone et al. 1998; Padoan & Nordlund 1999) created in the post-shock regions of converging large-scale flows. Stars form on very short timescales (Elmegreen 2000). However, rapid MC formation requires rapid $\text{H I} \rightarrow \text{H}_2$ conversion (Goldsmith & Li 2005). Accelerated H_2 formation requires either high-density pre-shock conditions ($n \sim 200 \text{ cm}^{-3}$, $T \leq 100\text{K}$; Price et al. 2001), or strong turbulence (Glover & Mac Low 2007), higher than observed.

On the other hand, the formation of high-mass

stars and of the star clusters hosting them is likely the most important process that shapes the formation and evolution of galaxies. Massive stars are responsible for the global ionization of the ISM. Their energetic stellar winds and supernova blast waves direct the dynamical evolution of the ISM, shaping its morphology, energetics and chemistry, and influencing the formation of subsequent generations of stars and planetary systems. Despite their importance, remarkably little is known about how massive stars form (McKee & Tan 2003). We lack a “fundamental theory” or, rather, a galaxy-scale predictive model for star formation. One of the main limitations to this goal is the lack of statistically significant and well-characterized samples of young massive stars in the various evolutionary stages and environments on which a theory can be based. In turn, this results from the difficulty of gathering observational data on a large number of forming high-mass stars: they make up only a very small fraction of the total number of stars in the Galaxy, their early evolutionary phases of massive stars are more rapid than those of low-mass stars, they lie at large distance and form in crowded environments.

There is thus a long list of questions that the community has been addressing for some time, not finding satisfactory answers. Here is an abridged list:

- What is the temperature and density structure of the ISM? How do molecular clouds form, evolve, and how are they disrupted?
- What is the origin of the stellar initial mass function (IMF)? What is its relationship to the mass function (MF) of ISM structures and cloud cores on all scales?
- How do massive stars and clusters form and how do they evolve? What are the earliest stages of massive star formation and what are the timescales of these early phases?
- How do the Star Formation Rate (SFR) and Efficiency (SFE) vary as a function of Galactocentric distance and environmental conditions such as the intensity of the Interstellar Radiation Field (ISRF), ISM metallicity, proximity to spiral arms or the molecular ring, external triggers, and total pressure?

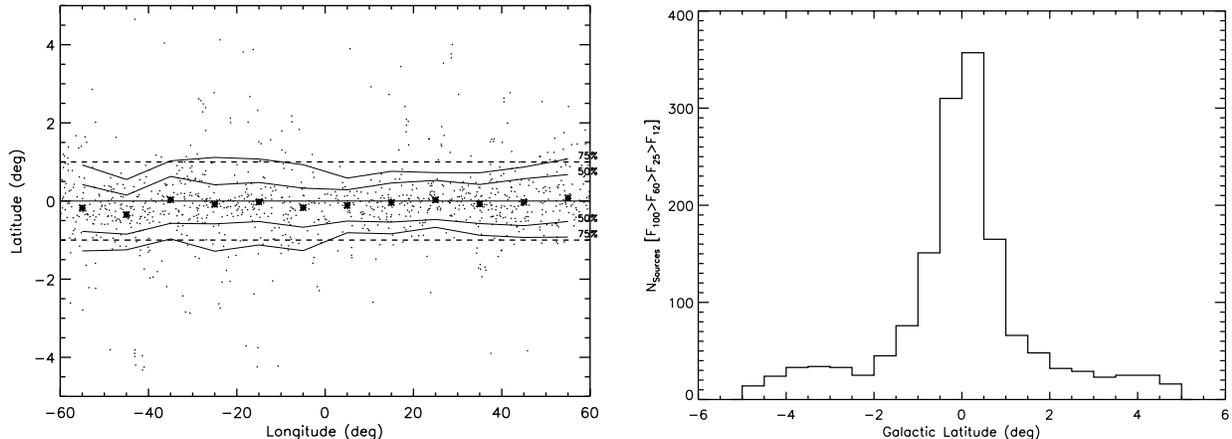


Fig. 1.— Left panel: $l - b$ plot of λ -rising SED IRAS sources; the straight thin line is the Galactic midplane. The asterisks mark the median latitude of the sources computed in $10^\circ l$ bins. The thin lines delimit the regions where 50% and 75% of $|b| \leq 5^\circ$ IRAS sources are contained. Right panel: b -distribution of the same IRAS sources in the $|l| \leq 60^\circ$ region.

- Does a threshold column density for star formation exist in our Galaxy? What determines the value of this possible threshold?
- What are the physical processes involved in triggered star formation on all scales and how does triggered star formation differ from spontaneous star formation?
- How do the local properties of the ISM and the rates of spontaneous or triggered star formation relate to the global scaling laws observed in external galaxies?

Using the Herschel telescope, the largest ever in space, Hi-GAL, the Herschel infrared Galactic Plane survey, will provide unique new data with which to address these questions. Hi-GAL will make thermal infrared maps of the Galactic Plane at a spatial resolution 30 times better than IRAS and 100 times better than DIRBE, from which a complete census of compact source luminosities, masses, and spectral energy distributions (SEDs) will be derived. Source distances are a crucial parameter in this respect, and a dedicated effort will be needed (see §4). Extraction of statistically significant samples of star-forming regions and cold ISM structures will be possible in all the environments of the Milky Way at all scales from massive Young Stellar Objects (YSOs) in individual protoclusters to complete spiral arms.

In the following we present the specific characteristics of the survey as well as some of the science outcomes that we expect to obtain with this unique project.

2. Hi-GAL Observing Strategy

The area covered by Hi-GAL ($|l| \leq 60^\circ, |b| \leq 1^\circ$) contains most of the star formation in the Galaxy, and it is the one which offers the best coverage in ancillary data which will be critical in the scientific analysis (see §4). The b distribution and extent of the survey is shown in Fig. 1 along with the $l - b$ plot of λ -rising SED IRAS sources ($F_{100} > F_{60} > F_{25} > F_{12}$) which are potential YSOs. The Hi-GAL area (thick dashed lines in that figure) represents the $|b| \leq 1^\circ$ strip centered on the midplane and contains about 80% of all potential YSOs contained in $|b| \leq 5^\circ$ strip, thus encompassing most of the potential star formation sites in the inner Galaxy.

The Herschel photometric cameras PACS (Poglitsch et al. 2008) and SPIRE (Griffin et al. 2009) will be used in parallel mode (pMode¹) to maximize survey speed and wavelength coverage. Due to the instruments wavelength multiplexing capabilities,

¹In pMode the Herschel telescope is scanning the sky in a raster fashion at constant speed while both PACS and SPIRE acquire data simultaneously

one pMode observation delivers maps at five different wavelengths: 70 and $170\mu\text{m}$ with PACS and 250, 350 and $500\mu\text{m}$ with SPIRE. Both cameras use bolometric detector arrays to map the sky by scanning the spacecraft along approximate great circles. Both instruments require their on board sub-kelvin coolers to be recycled to provide the detectors with the operating temperature required of about 0.3 K in each case. In pMode both instruments are placed into their photometric observing mode with the detectors at their correct operating temperature, i.e. both instrument coolers are recycled, and data are taken from the five arrays simultaneously as the spacecraft is scanned across the sky.

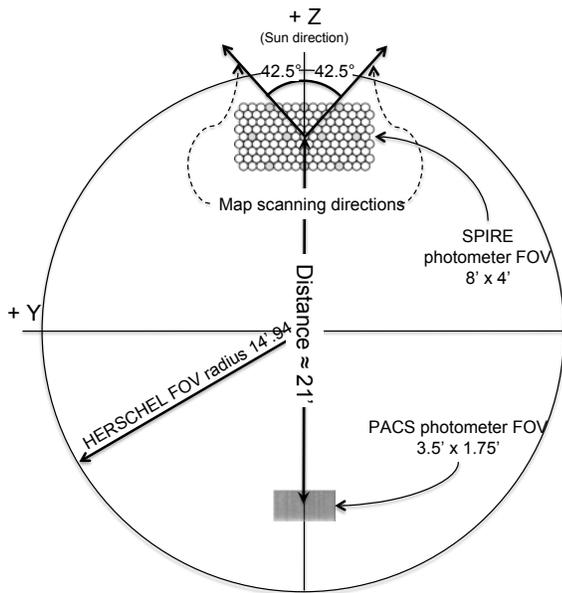


Fig. 2.— The field of view of the PACS and SPIRE instruments shown in the context of the Herschel field of view as viewed in the co-ordinate system of the spacecraft +Z refers to the axis towards the Sun. The +X axis is along the telescope bore-sight out of the page. The scan directions used to map the sky are as indicated. The different photometric channels of each instrument map the same region of the sky.

The size and separation of the fields of view of PACS and SPIRE are shown in Fig. 2 as viewed in the spacecraft co-ordinate system. Although the PACS array fully spatially samples the point spread function from the telescope it still has gaps

between the sub-arrays, and the SPIRE arrays only sparsely sample the sky. In order to make fully spatially sampled maps it is necessary to scan the SPIRE array at an angle of 42.5° with respect to its short symmetry axis. Scanning at an angle is also used for the PACS arrays to fill in for the gaps between sub-arrays. To achieve redundancy in the data and remove instrumental effects such as high-frequency detector response, slow drifts in gain or stray light, saturation and environmental (cirrus confusion) effects it is also necessary to make at least a second pass over the same region of the sky using the other scan angle at -42.5° angle which, quite conveniently, is nearly orthogonal to the first one (see fig. 3).

The distance between each scan in parallel mode is set by the size of the PACS array (being the smaller of the two), and the effective length of each leg of the raster takes into account the separation between the two fields of view. HSPOT, the Herschel-SPOT observing tool², automatically calculates these parameters to ensure that the area required is covered. The distance between scans is approximately $155''$ and the excess length of the scan beyond the required length to cover the area is typically $20'$. An example of how the sky is covered in a Parallel Mode observation used in Hi-GAL is shown in Fig. 3.

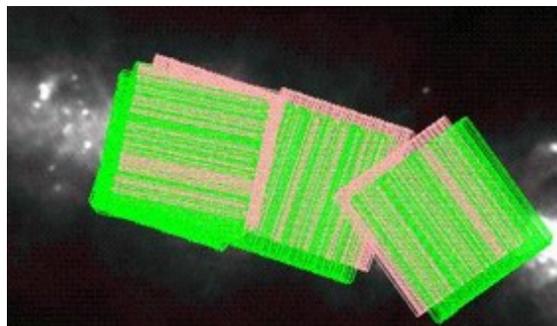


Fig. 3.— Sample AORs (Astronomical Observation Request) overlaid on the IRAS $100\mu\text{m}$ image of a portion of the Galactic plane. From left to right we show: both nominal and orthogonal $2.2^\circ \times 2.2^\circ$ pMode AORs overlaid on one another. The PACS-covered area is outlined in pink while SPIRE is green.

²ftp://ftp.sciops.esa.int/pub/hspot/HSpot_download.html

The strategy employed to cover the $-60^\circ \leq l \leq 60^\circ$, $|b| \leq 1^\circ$ survey area is to conduct observations with series of 55 $2.2^\circ \times 2.2^\circ$ square tiles with two passes over each tile with the two above mentioned scan angles (see Fig. 2). These tiles will be spaced every 2° , so that the overlap between tiles ensures that no coverage gaps are introduced by different tile orientation due to variable satellite roll angles with time; fig. 3 shows a section of the galactic plane with consecutive observing blocks overlaid providing overlapping coverage.

Given the spatial separation required for PACS in the pMode observations, the SPIRE data is heavily oversampled and we cover a greater area than required with each individual instrument than would be required using them sequentially. Although it might seem that sequential PACS and SPIRE scan mode observations would be more efficient in fact the satellite overheads, set up, calibration and pointing acquisition, *et cetera*, mean that it requires 30% more time to cover the same area sequentially compared to using the pMode.

In order to cover the maximum area in the shortest time Hi-GAL data will be taken at the maximum possible scan speed for the satellite of 60 "/sec. This implies a beam crossing time for the short wavelength, 250 μm band of SPIRE of 3 Hz well within the bandwidth available in the detectors of 5 Hz. However, although the PACS detectors have a similar response time the much smaller point spread function will be smeared out compared to that achievable with a slower scan. Additionally because of the finite data transmission bandwidth between the Herschel satellite and the ground, it is necessary to perform on-board data compression for the PACS data which are the most demanding in terms of number of pixels (2048 for the 70 μm array and 512 for the 170 μm array) at the frame acquisition rate of 40Hz. The baseline configuration for the pMode is then to average on board groups of 8 frames at 70 μm and 4 frames at 170 μm . Since the telescope is continuously scanning while acquiring, this coaddition will result in a further degradation of the Point Spread Function in the direction of the scan from its original diffraction limited shape; the effect will be more severe at 70 μm where the degradation should be of a factor two based on simulations. This loss in imaging fidelity at the shortest wavelength is considered acceptable for a survey like

Hi-GAL because, as discussed in the previous section, our main focus is toward a large-scale picture of the galaxy. Taking advantage of the orthogonal cross scan observing strategy, we may be able to recover some of the spatial resolution by careful deconvolution during post processing.

2.1. Detection of compact sources

The SPIRE digital readout electronics impose a limitation on the brightest sources that can be observed for a given offset setting (DC voltage removal) before digitization (SPIRE Instrument Users Manual, 2007). This problem can be alleviated to some extent by choosing a bias setting that gives the largest dynamic range per offset range. Simulations of the effect of bias variation show that setting a bias higher ($\sim 3x$) than the predicted nominal value will approximately double the dynamic range for most detectors under the conditions likely to be found in orbit (telescope temperature and emissivity and sky background). The same simulations show that a significant ($>10\%$) fraction of the SPIRE 250 μm array detectors will saturate on sources greater than 500 Jy. The situation is slightly more relaxed for the 350 and 500 μm arrays. We take the upper limit of detectable sources in the SPIRE bands as 500 Jy/beam assuming that a strong source instrument setting is used. This setting is required for all observations of bright regions/sources with SPIRE and is not a special Hi-GAL configuration. The saturation limits for PACS should be around 2000Jy at nominal bias, that will be used for the Hi-GAL survey.

The $1-\sigma$ sensitivities provided by HSPOT, for a single Astronomical Observation Request (AOR), are 17.6 and 26.8 mJy in the two PACS 70 and 170 μm bands, and 12.8, 17.6 and 14.9 mJy for the SPIRE bands; co-addition of the orthogonal scanning patterns will provide $\sqrt{2}$ better figures. These sensitivities result from the adopted scanning strategy designed to maximize redundancy and map fidelity especially for large scale diffuse structures. However, the limiting factor for the detectability of sources and clouds will likely be cirrus confusion. Estimates based on recent BLAST measurements (Roy et al. 2010) suggest values of the order of 75, 140, and 160 mJy in the 170, 250 and 500 μm Herschel bands for a representative region of the Galactic Plane at $l=45^\circ$; these values

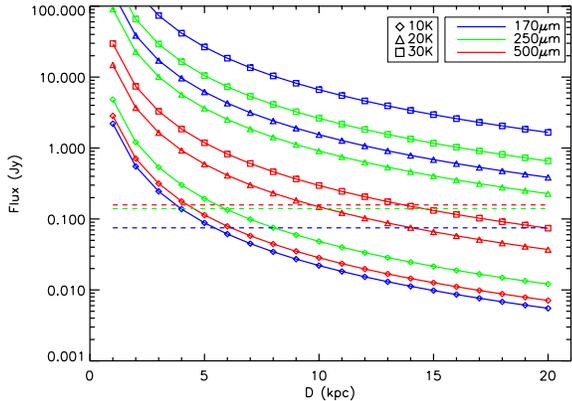


Fig. 4.— Flux expected in the 170, 250 and 500 μm Herschel bands (color coded) from a 20 M_{\odot} core (dust+gas) for different temperatures (different symbols), as a function of distance in kpc. The dashed lines represent the confusion noise expected in the three bands.

are greater than the detector sensitivities. Fig. 4 shows the expected flux from a 20 M_{\odot} envelope as a function of distance (in kpc) in each of three bands above mentioned, and for three different dust temperatures; we adopted $\beta=2$ and the dust opacity from Preibisch et al. (1993). The horizontal dashed lines (color-coded with wavelength) represent the predicted confusion noises based on BLAST images. Fig. 4 shows that we will detect the representative 20 M_{\odot} core everywhere in the Galaxy except for very cold dust ($T \leq 10\text{K}$), for which detectability is predicted to be limited within a distance of about 5 kpc. We may then conclude that cirrus confusion is not going to be a problem for the investigations of the intermediate and high-mass star formation studies which are the "core science" of this project (see §3.3 and 3.4).

2.2. Detection of extended structures

The diffraction-limited instrument beams at all wavelengths can be used to translate the confusion noises reported in the previous section into brightness units to investigate the detectability limits expected for diffuse extended structures. Calculations of the expected brightness levels from optically thin dust as a function of temperature and H column density (assuming gas/dust=100) are reported in Fig. 5a,b for $\lambda=170$ and 500 μm respec-

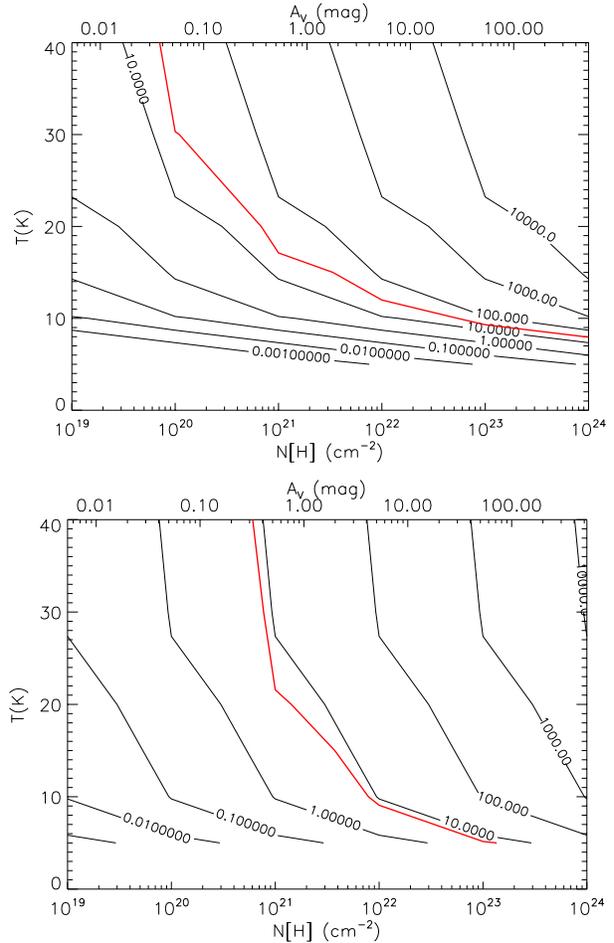


Fig. 5.— Brightness (in MJy/sr) of optically thin dust as a function of temperature and column density at 170 μm (top) and 500 μm (bottom). The red lines indicate the expected confusion noise.

tively, where the expected confusion noise levels are also reported in red lines.

The figures show that with a typical confusion noise of about 10 MJy/sr at 500 μm it will be possible to detect most Infrared Dark Clouds (IRDCs, see §3.2), where column densities are in the range 10^{23} - 10^{25}cm^{-2} . Less dense clouds with column densities of the order of few 10^{21}cm^{-2} should be easily detectable at levels of tens of MJy/sr at 170 μm at temperatures as low as $T \sim 20\text{K}$. The situation may be less simple in the regions closer to the Galactic center. However, our broad spectral coverage provides an important advantage for measuring the temperature accurately, and

for isolating structures and sources with temperature different from the standard diffuse ISM cirrus ($\sim 20\text{K}$). Besides, all of the above is based on extrapolated estimates and we will provide the definitive measurement of the cirrus confusion at Herschel wavelengths and resolution.

2.3. Checks on photometric calibration

Virtually every block of the Hi-GAL survey will contain secondary calibrators, either stars or other well known objects, ensuring accurate checks of the flux calibration of Hi-GAL data. Our baseline calibration targets will be the 400 stars used as calibrators for the Spitzer GLIMPSE-I/II surveys. These are mostly A0-5V or K0-M0III stars, although they also include ~ 60 calibrators of hot or warm dwarfs (B-G). In order to remove sources with non-photospheric FIR emission we will make predictions for the $24\mu\text{m}$ fluxes and then test for excesses in the MIPS GAL data first. Once anomalous sources are excluded we can extrapolate to the FIR and create an initial set of calibrators.

We will also be able to obtain a reliable calibration for extended sources, which is one of the more difficult parts of the nominal instrumental calibration activity. As part of the calibration scheme we will compare fluxes in the SPIRE $500\mu\text{m}$ band with fluxes from the same band of Planck-HFI in suitable locations³.

3. Hi-GAL Key Science

3.1. The Distribution of the ISM Temperature and the Intensity of the Interstellar Radiation Field

At near-infrared wavelengths, the emission from dust is produced by small particles whose abundance varies significantly, being strongly depleted by coagulation processes in the dense ISM (e.g. Bernard et al. 1993; Abergel et al. 1994; Stepnik et al. 2003). Far-infrared (FIR) emission is produced by larger grains which are more stable and dominate the total dust mass and trace all phases of the ISM. The ISM dust spectrum peaks in the FIR where the Galaxy is transparent. FIR emission is therefore a reliable tracer of the overall ISM column density structure in

our Galaxy. Other phase-independent tracers include dust absorption and gamma-ray production, where however the former can be used to sample only the nearest 1 kpc, and γ -ray surveys currently lack sensitivity and angular resolution.

Variations in the FIR emissivity (the ratio of surface brightness to column density) are dominated by the non-linear effects of dust temperature through the Planck function. Fortunately, the shape of the dust SED as measured by PACS and SPIRE will be most sensitive to temperature variations as the spectral bands sample the peak of the Big Grain emission and the contribution of Very Small Grains can be estimated from the Hi-GAL data at $70\mu\text{m}$ and MIPS GAL at $24\mu\text{m}$. The dust temperature (T_d) and its spatial variations will therefore be measured precisely. This important parameter can be used, in conjunction with complementary data from Planck, HI, CO, $\text{H}\alpha$ and γ -ray surveys, to estimate the strength and spectral shape of the InterStellar Radiation Field (ISRF), which is set by the stellar content in a given region. So far, the dust temperature in the Galactic Plane has been mapped over limited regions using IRAS (Kim et al. 1999, Douglas & Taylor 2007, and over the entire Plane at a resolution of $40'$ (Lagache et al. 1998) using DIRBE. Hi-GAL will improve with respect to the latter by a factor of about 100 in linear scales. It will trace the local radiation field on scales relevant to star formation, and provide mass estimates even at large distances. In the case of the dense medium, determining the 3-D distribution of the ISRF strength and spectral shape in a given cloud will require radiative transfer modeling. This is possible, even for complex geometries, using Monte-Carlo codes (e.g. Juvela & Padoan 2003). Using such codes, the equilibrium dust temperature T_d and the dust emission can be predicted at any 3D location in the cloud. Integration along the line-of-sight in turn allows to predict 2D emission maps.

3.2. Molecular cloud formation

About a quarter of the mass in the ISM is in molecular form (Blitz 1997) and most of that material resides in giant molecular clouds (GMCs). Since GMCs are also the dominant sites of star formation, understanding their origins and evolution is essential to our understanding of the Galactic environment.

³many of the Hi-GAL Co-Is are also Planck Consortium members

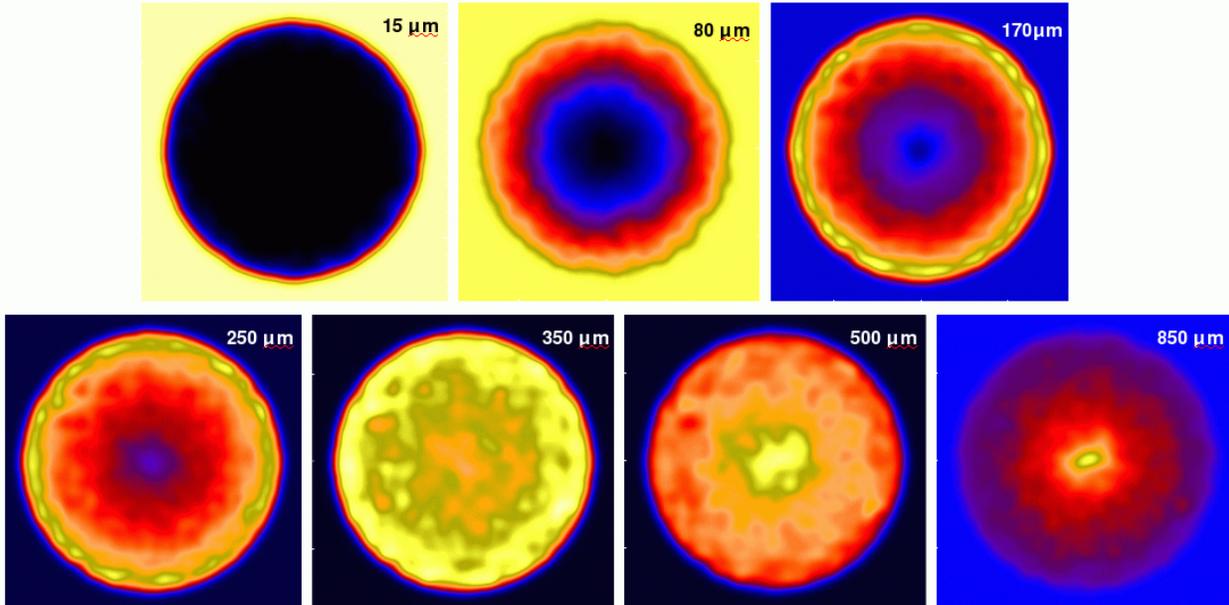


Fig. 6.— Model of a typical IRDC after Stamatellos et al. (2004) at seven different wavelengths illustrates the importance of the Herschel wavelength coverage and resolution (at 2 kpc the linear extent is $5'$) to obtain measurements of IRDCs in emission.

In combination with molecular line surveys, Hi-GAL will provide the data needed to derive basic physical properties of GMCs. We will detect and characterize cold structures in the inner GP and classify them based on star formation activity. Detection statistics for clouds with different temperatures and degree of star formation activity will provide the fraction of quiescent *vs* star-forming clouds. It will thus be possible to constrain the properties and lifetimes of GMCs in our Galaxy and to compare with the predictions of fast evolution of molecular clouds (Hartmann et al. 2001) or a more traditional slow evolution of star formation in our Galaxy (Shu et al. 1987). Variations with Galactocentric radius will determine if the *slow/fast* scenarios are mutually exclusive or reflect different initial/environmental conditions. A large-area survey like Hi-GAL will provide the needed statistical significance in all mass bins, especially at the high-mass end, and in a variety of Galactic environments.

Direct detection of cold (i.e. $T < 20$ K) dust which could be the quiescent counterparts to GMCs, is difficult (Sodroski et al. 1994; Reach et al. 1995; Lagache et al. 1998) either because

of insufficient wavelength coverage (e.g. IRAS) or inadequate spatial resolution (DIRBE, FIRAS). CO observations are problematic due to molecular freeze-out onto grains (Flower et al. 2005), or photo-chemical effects in low-metallicity environments (Bot et al. 2007). The recent detection of very cold clumps in the GP with Archeops (Désert et al. 2008), confirms the FIR and submm continuum as the best tool to trace cold ISM components. Notable examples are Infrared Dark Clouds (IRDCs) and HI Self-Absorption (HISA) clouds.

IRDCs are structures initially discovered as extinction features against the bright mid-IR Galactic background, and soon verified to exhibit properties similar to molecular clouds. Their properties ($n > 10^5 \text{ cm}^{-3}$, $N_H \sim 10^{22} - 10^{24} \text{ cm}^{-2}$ and $T < 25$ K - Egan et al. 1998; Carey et al. 2000, $R \sim 5$ pc and $M \sim 10^3 M_\odot$ Simon et al. 2006; Rathborne et al. 2006) suggest that they are the precursors of cluster-forming molecular clumps like Orion. Thus, IRDCs are ideal for the study of the pristine, undisturbed physical conditions that may produce massive stars and clusters. IRDCs have only been detected against the bright Galactic mid-IR background (mostly for $|l| \leq 30^\circ$); their

true Galactic distribution is unknown. Modeling of IRDCs in the IR and submm (Fig. 6, Stamatellos et al. 2004) proves Herschel’s unique ability to detect them and to measure their SEDs. IRDCs, with $\tau_{200\mu m} \geq 1$, are not detectable by either IRAS or Spitzer. Hi-GAL will provide a definitive inventory of cold dust and potential sites of massive star cluster formation everywhere in the inner GP.

HISAs are traced by cold HI gas seen in absorption against a background of warm HI emission (Gibson et al. 2000, 2005; Goldsmith & Li 2005); they may provide additional clues to the formation of molecular clouds. When compared with molecular tracers, these cold (relative to the ambient neutral medium) and relatively quiescent ($\Delta v \sim 1 - 3$ kms) clouds show a wide range of HI/H₂ ratios (Li & Goldsmith 2003; Klaassen et al. 2005) which suggests that they might be HI→H₂ conversion sites. Detailed studies of their FIR→mm SED shapes can help clarify this issue, providing evidence for the grain types necessary for a reasonable HI→H₂ formation timescale (Goldsmith & Li 2005).

3.3. Timeline of high-mass star formation

The paradigm for the formation of solar-type stars via accretion through a circumstellar disk (Shu et al. 1987) predicts an evolution from cores to protostars and, finally, pre-main sequence stars that is well matched with distinctive characteristics of their SEDs (Lada & Wilking 1984; Andre et al. 1993). The empirical classification of the SED of low mass YSOs has thus been used as a powerful tool to constrain theoretical models.

Higher mass stars reach the conditions for H-burning faster than the time required to assemble them, so that winds and radiative feedback will strongly influence accretion and may limit the final mass of the star (Zinnecker & Yorke 2007). However, since massive stars exist, several theories have been proposed to solve this puzzle including accretion from a disk, a very high-pressure ambient medium, "flashlight effect" (Yorke & Sonnhalter 2002; McKee & Tan 2003), competitive accretion, or coalescence (e.g. Stahler et al. 2000). Application of SED-based classification tools, and evolutionary diagnostics like the M_{env} - L_{bol} diagram which relates the bolometric luminosity of a YSO to the mass of its envelope (Molinari et al.

2008), to a large sample of luminous protostar candidates in the GP will define a timeline for the various phases of massive star formation that will constrain the theories and lead to new estimates of the SFR. Clearly the source distance is the crucial parameter here; we are collecting the information from the major molecular line surveys over the inner Galactic Plane, while planning to undertake additional surveying activities at a variety of facilities in several high-density tracers to get additional data (see §4 for more detail).

An evolutionary sequence for massive YSOs has been proposed (cold massive cloud core; Hot Molecular Core with outflow; IR-bright massive YSO; ultracompact (UC) HII region, e.g. Evans et al. 2002; Kurtz et al. 2000) but it is qualitative and based on small and possibly incomplete samples. Samples of bright and massive YSOs (Molinari et al. 1996; Sridharan et al. 2002; Hoare et al. 2004) are IRAS or MSX selected and tend to suffer from age biases and confusion which prevent firm quantitative conclusions. A phase of intense and accelerating accretion prior to H-burning ignition, that may be observable (e.g. Molinari et al. 1998) in the form of dense condensations devoid of IR as well as radio continuum emission, seems confirmed by recent large mm surveys (Beltrán et al. 2006; Hill et al. 2005). Millimeter continuum alone, however, cannot distinguish between pre-collapse condensations and rapidly accreting cores; Hi-GAL will use the full potential of Herschel wavelength coverage and spatial resolution to trace the SED peak of dust envelopes in all phases, from massive pre-stellar condensations to UCHII regions. An angular resolution of 30" or less, typical of Hi-GAL, has been proven (e.g. Molinari et al. 2008) to be the key to building accurate SEDs, deriving reliable luminosities for massive YSOs, and distinguishing embedded UCHII strongly emitting in the Mid-IR and radio from pre-UCHIIs (Fig. 7).

The abundance of high-mass YSOs per mass bin in the various evolutionary phases will provide an estimate of the duration of each phase. This timeline can be directly compared with the predictions of various models, and together with the YSO mass function, will be used to infer the SFR. As an example, using current estimates for SFR and IMF (McKee & Williams 1997; Kroupa 2001), and a 10⁵ yr period to assemble a massive star

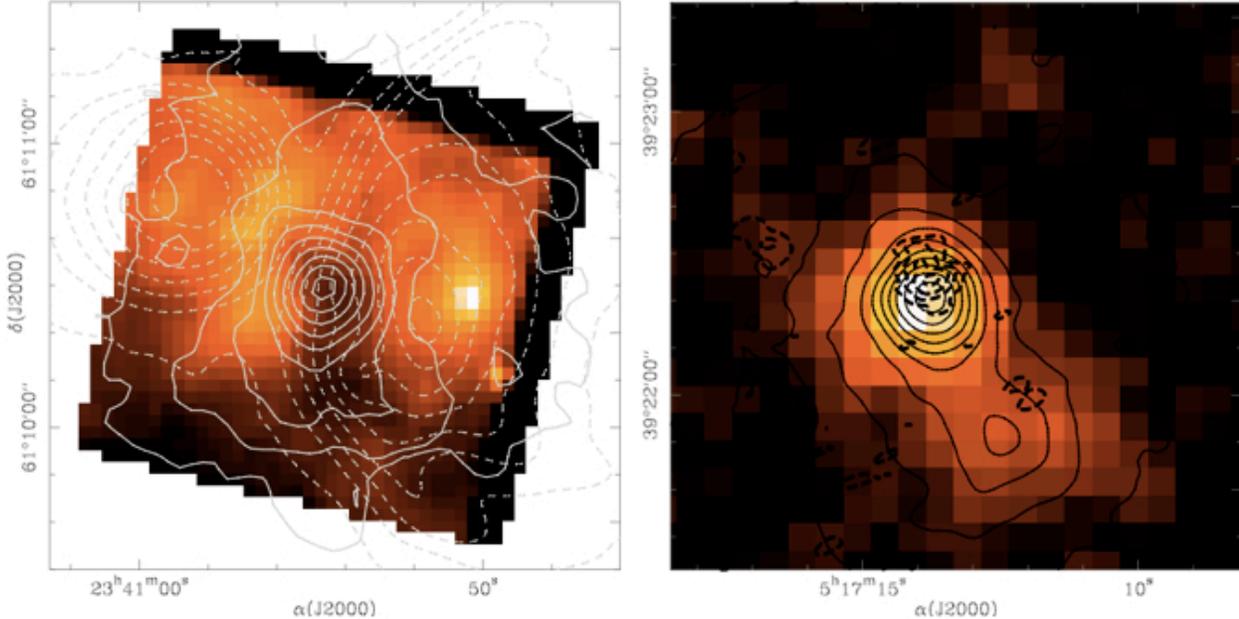


Fig. 7.— Mid-IR $21\mu\text{m}$ MSX images for a candidate precursor of a Hot Core (IRAS23385+6053, left), and a UCHII (IRAS05137+3919, right). Full and dashed contours represent the millimeter (cold dust) and radio (jet) continuum emission (from Molinari et al. (2008)). Hi-GAL will complement these with similar resolution FIR data.

(McKee & Tan 2003), we expect ≥ 1000 FIR/sub-mm objects with $M > 15M_{\odot}$ in the Galaxy, and only ~ 180 objects with $M > 50M_{\odot}$. Recent millimeter surveys in the Cyg X region (Motte et al. 2007) confirm the rarity of such massive precursors and strengthen the need for a systematic unbiased search.

3.4. Bridging the Gap between Global and local star formation

Galactic phenomenology currently invokes an indeterminate mixture of spontaneous and triggered star formation. Triggering agents include radiation pressure from OB stars (Sugitani et al. 1989), compression by expanding HII regions (Elmegreen & Lada 1977; Deharveng et al. 2005), or fragmentation of supershells by multiple supernovae in OB associations (McCray & Kafatos 1987). On larger scales we still do not know if spiral density waves actively induce star formation (e.g., Elmegreen 2002) or simply assemble star-forming regions, with local feedback and triggering becoming more important within the arms

(Sleath & Alexander 1996). The mean SFE of a galaxy can increase (up to 50 times) in starbursts (Sanders et al. 1991) and galaxy mergers due to strong feedback effects, a process observed in miniature in Galactic star-forming regions (e.g. Moore et al. 2007). Whether the IMF depends on local triggering and other environmental factors is unclear.

Hi-GAL will enable quantitative analysis based on basic observables - the luminosity functions of YSOs, the mass function of dense star-forming structures and quiescent clouds. Hi-GAL will provide the essential context of high-mass star formation, as it relates to molecular gas, HI gas, stars, HII regions, OB associations, SNRs and spiral arms. Theoretical models and numerical simulations will be tested in multiple ways. We will discover whether a local triggering agent is necessary for high-mass star formation or if a spiral arm is sufficient, clarifying the differences between spontaneous and triggered star formation. We will quantify the relationship between the interaction strength (estimated using available data from ancillary surveys) and the resulting increase in SFE

above the spontaneous rate. By locally relating the SFR to the properties of the ISM we will probe star formation thresholds as a function of environment and spatial scale, and possibly unveil the mechanism giving rise to global Schmidt-like scaling laws. We will determine the dominant physical process underlying triggering.

3.5. Serendipitous Science

The 5-band FIR images and source catalogues provided by Hi-GAL will allow research in many fields that can only be partially anticipated. A detailed description of the specific outcomes in all these fields goes beyond the scope of the present paper. These aspects will of course be the object of dedicated publications to be released in due time. In the following we list some of the possible by products of Hi-GAL :

- An input catalogue for ALMA: we expect to detect some 200-400 objects per tile, most of which will mark very cold objects to be studied at all possible wavelengths;
- complete characterization of the Galactic foreground in the Far-IR and submillimeter, critical for the correct interpretation and modeling of cosmological backgrounds;
- Dust formation and destruction in supernovae remnants
- Debris dust disks around main sequence stars, with unbiased statistics on frequency and mass as a function of star age
- Evolution of dust properties, especially around AGB stars, the factories of cosmic dust;
- Detection of detached dust shells around first ascent giant stars to investigate missing mass in AGB envelopes;
- Detection of multiple shells around AGB stars, post-AGB objects and planetary nebulae, as well as around various classes of interacting binaries;
- Detection of ejecta shells and swept-up ISM bubbles around massive stars, providing a complete census of WR and LBV stars;
- Extinction maps to aid in correcting Near-IR galactic star counts;

- Detection of Solar system objects via comparison of cross-linked rasters: in particular the detection of asteroids will be very interesting for studies of the albedo
- Nearby Low-Mass SFRs in the GP: Herschel will detect many nearby star forming regions and individual YSOs.

4. Hi-GAL and its place in the context of the Multi-Wavelength Milky Way

PACS and SPIRE are unique in tracing the peak of the Spectral Energy Distribution (SED) of cold dust and, hence, temperatures and luminosities of both star-forming complexes and the ISM, at resolutions unmatched by any previous instruments. However, the full potential of the Hi-GAL survey will be realised in the context of the other unbiased Galactic Plane surveys which are shaping our understanding of the Galactic ecosystem. A suite of surveys in the mid- and far-infrared continuum, ISOGAL (Omont et al. 2003), MSX (Price et al. 2001), GLIMPSE (Benjamin et al. 2003), MIPS GAL (Carey et al. 2005), IRAS (Neugebauer et al. 1984) and Akari (Murakami et al. 2007), has been and will be complemented by surveys in the submillimeter and millimeter spectral range including the BGPS survey with Bolocam at 1.1mm (Rosolowsky et al. 2009), the ATLASGAL survey currently underway with the LaBoCa camera at APEX (Schuller et al. 2009), and the SCUBA2 JPS survey beginning in 2010 (Fig. 8).

BLAST and AKARI have wavelength coverage and resolution not too different from PACS and SPIRE. However, BLAST has been used to map limited portions of the GP while results from AKARI photometric imaging of the GP are not yet found in the literature.

Hi-GAL will be the scientific keystone of this suite of surveys, completing the continuous coverage of the dust continuum over three orders of magnitude in wavelength at sub-30'' resolution, and allowing the measurement of dust temperatures and luminosity over the inner Galactic Plane.

An extensive plan for radio spectroscopic cross-correlation and follow up of the Hi-GAL survey has been devised. The top priority is for radio spectroscopic observations to obtain distance estimates for detected sources and structures. CO and ¹³CO data at sub-arcminute spatial resolution are

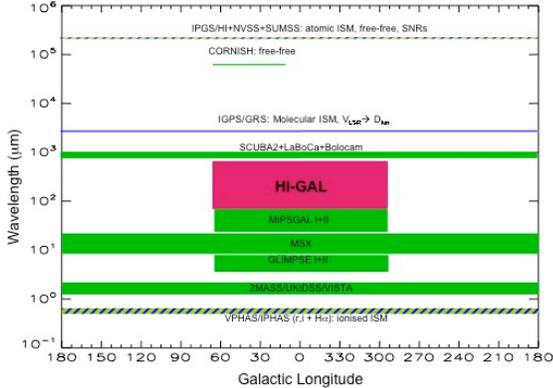


Fig. 8.— Wavelength-longitude coverage plot for photometric (green) and spectroscopic (blue) existing or planned GP surveys for the next decade. Hi-GAL (red) fills the critical gap where the ISM dust emission peaks, between MIPSAL and the SCUBA2/LaBoCa/Bolocam surveys.

available via the IGPS for the first Galactic Quadrant (QI); lower resolution CO data are available from the NANTEN survey (QIV). Should CO resolution not be sufficient, e.g. in case of very cold cores with CO depletion, we plan extensive follow-up programs (e.g. N_2H^+ , NH_3) at the Mopra 22m antenna (Australia) for QIV, and at Green Bank, Effelsberg, Medicina and Onsala for QI. HI is also available for QI and IV at sufficient resolution from the IGPS to help resolve the distance ambiguity in the inner Galaxy Busfield et al. 2006. $\text{H}\alpha$ VPHAS+IPHAS surveys will be also used.

In addition to the kinematic distance estimates, we plan to use NANTEN2, Mopra and APEX for detailed multiline studies of evolution-sensitive chemical tracers (N_2H^+ , NH_3 , CS, HCN, CH_3OH , CH_3CN , etc.) toward clouds and objects discovered by Hi-GAL. Hi-GAL catalogues will be the primary source of major future high spatial resolution follow ups in the sub-mm with ALMA; in the meantime such programs will be attempted using the SMA interferometer through Legacy class proposals.

5. Data Processing and Products

It is relatively easy to translate our scientific goals into a clear set of requirements on the data processing: we require that the dust continuum

emission be detectable, and accurately measurable, at all bands over the broadest range in signal levels (down to the confusion limit) and spatial scales. The observing strategy is carefully designed to that effect, but ensuring that this information is properly extracted from the data stream over a 240 sq.deg. area is a formidable challenge. We will use the Herschel Interactive Processing Environment (HIPE) for all those processing steps dealing with fundamental instrumental calibration and issues, but we anticipate areas where a dedicated set of specialised tools can take advantage of the homogeneous observing strategy and deliver higher quality results compared to the standard pipeline products; pointing refining, map-making, source extraction and photometry are examples.

We will make available a set of data products which will include maps and compact source catalogues at the five Hi-GAL wavelengths. These products will be made available after the end of observations (EoO) for the entire survey via incremental releases.

Improved reprocessed maps and source catalogues will be subsequently made public, together with a first release for an extended source catalogue. All public deliveries will be accompanied by an Explanatory Supplement.

In addition to this minimum set of products, quite standard for any large-scale survey like Hi-GAL, we plan to make available to the community a set of scientific value-added products which will be created during our scientific analysis, including band-merged catalogues integrated with data from continuum surveys at adjacent wavelength, color maps, source-subtracted maps. The public access to this final set of products is foreseen for EoO+42 months.

6. Conclusions

Hi-GAL is an Open Time Key Project to be performed with the 3.5m orbiting Herschel telescope, to map photometrically the inner Milky Way ($|l| < 60^\circ$, $|b| < 1^\circ$) in five wavebands between $70\mu\text{m}$ and $500\mu\text{m}$ simultaneously, using ~ 350 hours of observing time. The unique combination of survey speed, high sensitivity, high spatial resolution and wavelength coverage (across the peak of the dust emission) make Hi-GAL the first dedicated project to study the early phases

of GMC- and high-mass star formation in the Galaxy, with a legacy value similar to the IRAS mission some 20 years ago. The outcomes of Hi-GAL will consist of source lists and images to be released in due course after EoO.

We are grateful to all the people who made the building and launch of Herschel such a success. In particular ESA and the Herschel Project Scientist G. Pilbratt, and the instrument teams of PACS and SPIRE magnificently led by A. Poglitsch (MPE, Garching) and M. Griffin (Univ. of Cardiff).

Facilities: Herschel, SPIRE, PACS.

REFERENCES

- Abergel, A., Boulanger, F., Mizuno, A., & Fukui, Y. 1994, *ApJ*, 423, L59
- Andre, P., Ward-Thompson, D., & Barsony, M. 1993, *ApJ*, 406, 122
- Beltrán, M. T., Brand, J., Cesaroni, R., et al. 2006, *A&A*, 447, 221
- Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, *PASP*, 115, 953
- Bernard, J. P., Boulanger, F., & Puget, J. L. 1993, *A&A*, 277, 609
- Blitz, L. 1997, in *IAU Symposium 170. CO: 25 years of millimeter wave spectroscopy*, ed. W. B. Latter & et al., 11
- Bot, C., Boulanger, F., Rubio, M., & Rantakyro, F. 2007, *A&A*, 471, 103
- Busfield, A. L., Purcell, C. R., Hoare, M. G., et al. 2006, *MNRAS*, 366, 1096
- Carey, S. J., Feldman, P. A., Redman, R. O., et al. 2000, *ApJ*, 543, L157
- Carey, S. J., Noriega-Crespo, A., Mizuno, D. R., et al. 2009, *PASP*, 121, 76
- Carey, S. J., Noriega-Crespo, A., Price, S. D., et al. 2005, in *Bulletin of the American Astronomical Society*, Vol. 37, *Bulletin of the American Astronomical Society*, 1252–+
- Deharveng, L., Zavagno, A., & Caplan, J. 2005, *A&A*, 433, 565
- Désert, F., Macías-Pérez, J. F., Mayet, F., et al. 2008, *A&A*, 481, 411
- Douglas, K. & Taylor, A. R. 2007, *ApJ*, 659, 426
- Egan, M. P., Shipman, R. F., Price, S. D., et al. 1998, *ApJ*, 494, L199+
- Elmegreen, B. G. 2000, *ApJ*, 530, 277
- Elmegreen, B. G. 2002, *ApJ*, 577, 206
- Elmegreen, B. G. & Lada, C. J. 1977, *ApJ*, 214, 725
- Evans, N. J., Shirley, Y. L., Mueller, K. E., & Knez, C. 2002, in *Astronomical Society of the Pacific Conference Series*, Vol. 267, *Hot Star Workshop III: The Earliest Phases of Massive Star Birth*, ed. P. Crowther, 17–+
- Flower, D. R., Pineau Des Forêts, G., & Walmsley, C. M. 2005, *A&A*, 436, 933
- Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005, *ApJ*, 626, 214
- Gibson, S. J., Taylor, A. R., Higgs, L. A., et al. 2000, *ApJ*, 540, 851
- Glover, S. C. O. & Mac Low, M. 2007, *ApJ*, 659, 1317
- Goldsmith, P. F. & Li, D. 2005, *ApJ*, 622, 938
- Griffin, M., Ade, P., André, P., et al. 2009, in *EAS Publications Series*, Vol. 34, *EAS Publications Series*, ed. L. Pagani & M. Gerin, 33–42
- Hartmann, L., Ballesteros-Paredes, J., & Bergin, E. A. 2001, *ApJ*, 562, 852
- Hill, T., Burton, M. G., Minier, V., et al. 2005, *MNRAS*, 363, 405
- Hoare, M. G., Lumsden, S. L., Oudmaijer, R. D., et al. 2004, in *Astronomical Society of the Pacific Conference Series*, Vol. 317, *Milky Way Surveys: The Structure and Evolution of our Galaxy*, ed. D. Clemens, R. Shah, & T. Brainerd, 156–+
- Juvela, M. & Padoan, P. 2003, *A&A*, 397, 201
- Kim, K., Lee, J., & Koo, B. C. 1999, *ApJ*, 523, 306

- Klaassen, P. D., Plume, R., Gibson, S. J., Taylor, A. R., & Brunt, C. M. 2005, *ApJ*, 631, 1001
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Kurtz, S., Cesaroni, R., Churchwell, E., Hofner, P., & Walmsley, C. M. 2000, *Protostars and Planets IV*, 299
- Lada, C. J. & Wilking, B. A. 1984, *ApJ*, 287, 610
- Lagache, G., Abergel, A., Boulanger, F., & Puget, J. 1998, *A&A*, 333, 709
- Leisawitz, D., Bash, F. N., & Thaddeus, P. 1989, *ApJS*, 70, 731
- Li, D. & Goldsmith, P. F. 2003, *ApJ*, 585, 823
- McCray, R. & Kafatos, M. 1987, *ApJ*, 317, 190
- McKee, C. F. & Tan, J. C. 2003, *ApJ*, 585, 850
- McKee, C. F. & Williams, J. P. 1997, *ApJ*, 476, 144
- Molinari, S., Brand, J., Cesaroni, R., & Palla, F. 1996, *A&A*, 308, 573
- Molinari, S., Pezzuto, S., Cesaroni, R., et al. 2008, *A&A*, 481, 345
- Molinari, S., Testi, L., Brand, J., Cesaroni, R., & Palla, F. 1998, *ApJ*, 505, L39+
- Moore, T. J. T., Bretherton, D. E., Fujiyoshi, T., et al. 2007, *MNRAS*, 379, 663
- Motte, F., Bontemps, S., Schilke, P., et al. 2007, *A&A*, 476, 1243
- Murakami, H., Baba, H., Barthel, P., et al. 2007, *PASJ*, 59, 369
- Neugebauer, G., Soifer, B. T., Beichman, C. A., et al. 1984, *Science*, 224, 14
- Omont, A., Gilmore, G. F., Alard, C., et al. 2003, *A&A*, 403, 975
- Padoan, P. & Nordlund, Å. 1999, *ApJ*, 526, 279
- Palla, F. & Galli, D. 1997, *ApJ*, 476, L35+
- Pascale, E., Ade, P. A. R., Bock, J. J., & et al. 2008, *ApJ*, 681, 400
- Poglitsch, A., Waelkens, C., Bauer, O., & et al. 2008, in *Proceedings of the SPIE*, Vol. 7010
- Preibisch, T., Ossenkopf, V., Yorke, H., & Henning, T. 1993, *A&A*, 279, 577
- Price, S. D., Egan, M. P., Carey, S. J., Mizuno, D. R., & Kuchar, T. A. 2001, *AJ*, 121, 2819
- Rathborne, J. M., Jackson, J. M., & Simon, R. 2006, *ApJ*, 641, 389
- Reach, W. T., Dwek, E., Fixsen, D. J., et al. 1995, *ApJ*, 451, 188
- Rosolowsky, E., Dunham, M. K., Ginsburg, A., et al. 2009, *ArXiv e-prints*
- Roy, A., Ade, P., Bock, J., et al. 2010, *ApJ*, in press
- Sanders, D. B., Scoville, N. Z., & Soifer, B. T. 1991, *ApJ*, 370, 158
- Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, *A&A*, 504, 415
- Shu, F. H., Adams, F. C., & Lizano, S. 1987, *ARA&A*, 25, 23
- Simon, R., Rathborne, J. M., Shah, R. Y., Jackson, J. M., & Chambers, E. T. 2006, *ApJ*, 653, 1325
- Sleath, J. P. & Alexander, P. 1996, *MNRAS*, 283, 358
- Sodroski, T. J., Bennett, C., Boggess, N., et al. 1994, *ApJ*, 428, 638
- Sridharan, T. K., Beuther, H., Schilke, P., Menten, K. M., & Wyrowski, F. 2002, *ApJ*, 566, 931
- Stahler, S. W., Palla, F., & Ho, P. T. P. 2000, *Protostars and Planets IV*, 327
- Stamatellos, D., Whitworth, A. P., André, P., & Ward-Thompson, D. 2004, *A&A*, 420, 1009
- Stepnik, B., Abergel, A., Bernard, J., et al. 2003, *A&A*, 398, 551
- Stone, J. M., Ostriker, E. C., & Gammie, C. F. 1998, *ApJ*, 508, L99
- Sugitani, K., Fukui, Y., Mizuni, A., & Ohashi, N. 1989, *ApJ*, 342, L87

Yorke, H. W. & Sonnhalter, C. 2002, ApJ, 569,
846

Zinnecker, H. & Yorke, H. W. 2007, ARA&A, 45,
481

Appendix C

caravela input parameter file

The next lines form the parameter file used in model \mathcal{M} for the IRAS $100\mu\text{m}$ band, chapter 3. This was the best-fit model found in cited results chapter.

```
NUMBER_SOURCES_POP_1          0.0
OBJS_LINEAR_SIZE_SCALE_MU_POP_1_AU  5.0e5
OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_1_AU  5.0e4
R_AVERAGE_POP_1_KPC          4.0
R_SIGMA_POP_1_KPC            0.4
Z_RANGE_POP_1_KPC            0.05
Z_MAX_POP_1_KPC              0.0
TEMPERATURE_MU_POP_1_K        30.0
TEMPERATURE_SIGMA_POP_1_K     3.0
NU_0_MU_POP_1_HZ             1.8e13
NU_0_SIGMA_POP_1_HZ          0.0
BETA_MU_POP_1                 2.0
BETA_SIGMA_POP_1              0.0
```

SPIRAL_FLAG_POP_1	1.0
SPIRAL_N_ARMS_POP_1	4.0
SPIRAL_A_ARM_1_POP_1	3.0
SPIRAL_B_ARM_1_POP_1	0.2493
SPIRAL_LENGTH_POP_1	30.0
SPIRAL_WIDTH_SIGMA_POP_1	0.2
NUMBER_SOURCES_POP_2	0.0
OBJS_LINEAR_SIZE_SCALE_MU_POP_2_AU	5.0e5
OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_2_AU	5.0e4
R_AVERAGE_POP_2_KPC	4.0
R_SIGMA_POP_2_KPC	0.4
Z_RANGE_POP_2_KPC	0.05
Z_MAX_POP_2_KPC	0.0
TEMPERATURE_MU_POP_2_K	10.0
TEMPERATURE_SIGMA_POP_2_K	1.0
NU_0_MU_POP_2_HZ	1.8e13
NU_0_SIGMA_POP_2_HZ	0.0
BETA_MU_POP_2	2.0
BETA_SIGMA_POP_2	0.0
SPIRAL_FLAG_POP_2	1.0
SPIRAL_N_ARMS_POP_2	4.0
SPIRAL_A_ARM_1_POP_2	3.0
SPIRAL_B_ARM_1_POP_2	0.2493
SPIRAL_LENGTH_POP_2	30.0
SPIRAL_WIDTH_SIGMA_POP_2	0.2

NUMBER_SOURCES_POP_3	4.0e4
OBJS_LINEAR_SIZE_SCALE_MU_POP_3_AU	5.0e6
OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_3_AU	5.0e5
R_AVERAGE_POP_3_KPC	4.0
R_SIGMA_POP_3_KPC	0.4
Z_RANGE_POP_3_KPC	0.05
Z_MAX_POP_3_KPC	0.0
TEMPERATURE_MU_POP_3_K	40.0
TEMPERATURE_SIGMA_POP_3_K	1.0
NU_0_MU_POP_3_HZ	1.8e13
NU_0_SIGMA_POP_3_HZ	0.0
BETA_MU_POP_3	2.0
BETA_SIGMA_POP_3	0.0
SPIRAL_FLAG_POP_3	1.0
SPIRAL_N_ARMS_POP_3	4.0
SPIRAL_A_ARM_1_POP_3	3.0
SPIRAL_B_ARM_1_POP_3	0.2493
SPIRAL_LENGTH_POP_3	30.0
SPIRAL_WIDTH_SIGMA_POP_3	0.2
NUMBER_SOURCES_POP_4	4.0e4
OBJS_LINEAR_SIZE_SCALE_MU_POP_4_AU	4.123
OBJS_LINEAR_SIZE_SCALE_SIGMA_POP_4_AU	4.12e2
R_AVERAGE_POP_4_KPC	4.0
R_SIGMA_POP_4_KPC	0.4

Z_RANGE_POP_4_KPC	0.05
Z_MAX_POP_4_KPC	0.0
TEMPERATURE_MU_POP_4_K	10.0
TEMPERATURE_SIGMA_POP_4_K	1.0
NU_0_MU_POP_4_HZ	1.8e13
NU_0_SIGMA_POP_4_HZ	0.0
BETA_MU_POP_4	2.0
BETA_SIGMA_POP_4	0.0
SPIRAL_FLAG_POP_4	1.0
SPIRAL_N_ARMS_POP_4	4.0
SPIRAL_A_ARM_1_POP_4	3.0
SPIRAL_B_ARM_1_POP_4	0.2493
SPIRAL_LENGTH_POP_4	30.0
SPIRAL_WIDTH_SIGMA_POP_4	0.2
IMAGE_WAVELENGTH_MICRONS	100.0
IMAGE_SIZE_X_DEGREES	1.0
IMAGE_SIZE_Y_DEGREES	0.0
IMAGE_CENTRAL_POSITION_LAT_DEGREES	0.0
IMAGE_CENTRAL_POSITION_LONG_DEGREES	40.0
OVERLAP_FRACTION_X	1.0
OVERLAP_FRACTION_Y	1.0
INDIV_IMAGES_LINEAR_SIZE_X_DEGREES	2.0
INDIV_IMAGES_LINEAR_SIZE_Y_DEGREES	2.0
AUTO_RESOLUTION_FLAG	0.0
IMS_SCALE_PIXELS_PRE_DEGREE_X	180.0

IMS_SCALE_PIXELS_PRE_DEGREE_Y	180.0
PROJECTION_TYPE_FLAG	1.0
INSTRUMENT_RESOLUTION_ARCMIN	0.5
NOISE_SKY_LEVEL_MJY_SR	0.0
NOISE_SKY_ERROR_MJY_SR	0.04
RING_IMAGES	1.0
RING_1_MIN	0.1
RING_1_MAX	4.0
RING_2_MIN	4.0
RING_2_MAX	5.6
RING_3_MIN	5.6
RING_3_MAX	7.2
RING_4_MIN	7.2
RING_4_MAX	8.9
RING_5_MIN	8.9
RING_5_MAX	14.0
RING_6_MIN	14.0
RING_6_MAX	17.0
MF_A_1	0.3
MF_B_1	1.0
MF_A_2	1.2
MF_B_2	1.79

A description of each input parameter is present in section 2.5, page 34.

Bibliography

- Bally, J. and Zinnecker, H. (2005). The birth of high-mass stars: Accretion and/or mergers? *The Astronomical Journal*, **129**.
- Barbosa, C. L. and Finger, D. (2005). Top 10 problems on massive stars. *Astro-ph*, **0408491**.
- Becker, R. H., White, R. L., Helfand, D. J., and Zoonematkermani, S. (1994). A 5 ghz vla survey of the galactic plane. *Astrophysical Journal Supplement Series*, **91**.
- Benjamin, R. A., Whitney, B., and Churchwell, E. (2008). The Galactic and Extragalactic Legacy of GLIMPSE. In R.-R. Chary, H. I. Teplitz, & K. Sheth, editor, *Infrared Diagnostics of Galaxy Evolution*, volume 381 of *Astronomical Society of the Pacific Conference Series*, pages 109–+.
- Beuther, H. (2002). *Early stages of massive star formation*. Ph.D. thesis, AA(Max-Planck-Institut fuer Radioastronomie <EMAIL>beuther@mpifr-bonn.mpg.de</EMAIL>).
- Beuther, H., Churchwell, E., McKee, C., and Tan, J. (2006). The formation of massive stars. *Astro-ph*, **0602012**.
- Blitz, L. (1991). Star forming giant molecular clouds. *The Physics of Star Formation and Early Stellar Evolution*, **342**.

- Blitz, L. and Williams, J. P. (1999). Molecular clouds. *The Origin of Stars and Planetary Systems*, **540**.
- Bonnell, I. A. and Bate, M. (2005). Binary systems and stellar mergers in massive star formation. *Monthly Notices of the Royal Astronomical Society*.
- Bonnell, I. A., Bate, M. R., and Zinnecker, H. (1998). On the formation of massive stars. *Monthly Notices of the Royal Astronomical Society*.
- Bonnell, I. A., Vine, S. G., and Bate, M. R. (2004). Massive star formation: nurture, not nature. *Monthly Notices of the Royal Astronomical Society*, **349**.
- Busfield, A. L., Purcell, C. R., Hoare, M. G., Lumsden, S. L., Moore, T. J. T., and Oudmaijer, R. D. (2006). Kinematic distance ambiguity (Busfield+, 2006). *VizieR Online Data Catalog*, **736**, 61096–+.
- Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. *PASP*, **115**, 763–795.
- Chambers, E. T., Jackson, J. M., Rathborne, J. M., and Simon, R. (2009). Star Formation Activity of Cores within Infrared Dark Clouds. *ApJS*, **181**, 360–390.
- Chini, R., Kruegel, E., and Kreysa, E. (1986). Dust emission spectra from star-forming regions. *Astronomy & Astrophysics*, **167**.
- Churchwell, E. (2002). Ultra-compact hII regions and massive star formation. *Annual Reviews Astronomy Astrophysics*, **40**, 27–62.
- Coppin, K., Chapin, E. L., Mortier, A. M. J., Scott, S. E., Borys, C., Dunlop, J. S., Halpern, M., Hughes, D. H., Pope, A., Scott, D., Serjeant, S., Wagg, J., Alexander, D. M., Almaini, O., Aretxaga, I., Babbedge, T., Best, P. N., Blain, A., Chapman, S., Clements, D. L., Crawford, M., Dunne, L., Eales, S. A., Edge, A. C., Farrah, D., Gaztañaga, E., Gear, W. K., Granato, G. L., Greve, T. R.,

- Fox, M., Ivison, R. J., Jarvis, M. J., Jenness, T., Lacey, C., Lepage, K., Mann, R. G., Marsden, G., Martinez-Sansigre, A., Oliver, S., Page, M. J., Peacock, J. A., Pearson, C. P., Percival, W. J., Priddey, R. S., Rawlings, S., Rowan-Robinson, M., Savage, R. S., Seigar, M., Sekiguchi, K., Silva, L., Simpson, C., Smail, I., Stevens, J. A., Takagi, T., Vaccari, M., van Kampen, E., and Willott, C. J. (2006). The SCUBA Half-Degree Extragalactic Survey - II. Submillimetre maps, catalogue and number counts. *MNRAS*, **372**, 1621–1652.
- Corbelli, E., Palla, F., and Zinnecker, H. (2005). The initial mass function 50 years later. *The Initial Mass Function 50 years later book*.
- Dorman, B. (2001). *CCfits Documentation, v1.5*. HEASARC Nasa.
- Egan, M. P., Shipman, R. F., Price, S. D., Carey, S. J., Clark, F. O., and Cohen, M. (1998). A Population of Cold Cores in the Galactic Plane. *ApJ*, **494**, L199+.
- Fazal, F. M., Sridharan, T. K., Qiu, K., Whitney, B. A., Robitaille, T. P., and Zhang, Q. (2007). Spectral Energy Distributions of High Mass Proto Stellar Objects - Evidence for High Accretion Rates. *ArXiv e-prints*, **711**.
- Fontani, F., Beltrán, M. T., Brand, J., Cesaroni, R., Testi, L., Molinari, S., and Walmsley, C. M. (2005). Search for massive protostellar candidates in the southern hemisphere. I. Association with dense gas. *A&A*, **432**, 921–935.
- Frieswijk, W. (2008). *Early Stages of Clustered Star Formation*. Ph.D. thesis, Rijksuniversiteit Groningen.
- Fuller, G. A. and Myers, P. C. (1987). Dense cores in dark clouds. *IN Physical processes in interstellar clouds*.
- Garay, G. and Lizano, S. (1999). Massive stars: Their environment and formation. *The Publications of the Astronomical Society of the Pacific*.

- Helfand, D. J., Zoonematkermani, S., Becker, R. H., and White, R. L. (1992). Compact radio sources near the galactic plane. *Astrophysical Journal Supplement Series*, **80**(1).
- Hoare, M. G., Glencross, W. M., Clegg, R. E. S., and Roche, P. F. (1988). The dust content of nebulae. *Dust in the universe, Proceedings of the Conference*.
- Hughes, V. A. and MacLeod, G. C. (1989). The use of iras data to define h ii regions. *Astronomical Journal*, **97**.
- Kant, I. (1755). *Allgemeine Naturgeschichte und Theorie des Himmels*.
- Kroupa, P. (2001). On the variation of the initial mass function. *Monthly Notices of the Royal Astronomical Society*, **322**.
- Kurtz, S., Churchwell, E., and Wood, D. O. S. (1994). Ultracompact H II regions. 2: New high-resolution radio images. *ApJS*, **91**, 659–712.
- Lada, C. J. and Kylafis, N. D., editors (1999). *The Origin of Stars and Planetary Systems*.
- Laplace, P. S., Bowditch, N., and Bowditch, N. I. (1829). *Mécanique céleste*.
- Larson, R. B. (1969). Numerical calculations of the dynamics of collapsing proto-star. *MNRAS*, **145**, 271–+.
- Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., Surace, J., Shupe, D., Xu, C., Oliver, S., Padgett, D., Fang, F., Conrow, T., Franceschini, A., Gautier, N., Griffin, M., Hacking, P., Masci, F., Morrison, G., O’Linger, J., Owen, F., Pérez-Fournon, I., Pierre, M., Puetter, R., Stacey, G., Castro, S., Polletta, M. d. C., Farrah, D., Jarrett, T., Frayer, D., Siana, B., Babbedge, T., Dye, S., Fox, M., Gonzalez-Solares, E., Salaman, M., Berta, S., Condon, J. J., Dole, H., and Ser-

- jeant, S. (2003). SWIRE: The SIRTf Wide-Area Infrared Extragalactic Survey. *PASP*, **115**, 897–927.
- Mac Low, M.-M., Toraskar, J., Oishi, J. S., and Abel, T. (2004). Ultracompact H II Regions Formed by Gravitational Collapse in Expanding Shells of Larger H II Regions. In *Bulletin of the American Astronomical Society*, volume 36 of *Bulletin of the American Astronomical Society*, pages 1574–+.
- Massey, P. (1999). Massive stars in the mcs: What they tell us about the imf, stellar evolution, and upper mass cutoffs. *New Views of the Magellanic Clouds, IAU Symposium no 190*.
- Massey, P. and Hunter, D. A. (1998). Star formation in r136: A cluster of o3 stars revealed by hubble space telescope spectroscopy. *Astrophysical Journal*, **493**.
- Meyer, M. R., Adams, F. C., Hillenbrand, L. A., Carpenter, J. M., and Larson, R. B. (2000). The stellar initial mass function: Constraints from young clusters, and theoretical perspectives. *Protostars and Planets IV*.
- Molinari, S. and the Hi-GAL Team (2009). Global star formation in the Milky Way with Hi-GAL, the Herschel infrared Galactic Plane Survey . *Memorie della Societa Astronomica Italiana*, **80**, 111–+.
- Molinari, S., Swinyard, B., Bally, J., Barlow, M., Bernard, J., Martin, P., Moore, T., Noriega-Crespo, A., Plume, R., Testi, L., Zavagno, A., Abergel, A., Ali, B., André, P., Baluteau, J., Benedettini, M., Berné, O., Billot, N. P., Blommaert, J., Bontemps, S., Boulanger, F., Brand, J., Brunt, C., Burton, M., Campeggio, L., Carey, S., Caselli, P., Cesaroni, R., Cernicharo, J., Chakrabarti, S., Chrysostomou, A., Codella, C., Cohen, M., Compiegne, M., Davis, C. J., de Bernardis, P., de Gasperis, G., Di Francesco, J., di Giorgio, A. M., Elia, D., Faustini, F., Fischera, J. F., Fukui, Y., Fuller, G. A., Ganga, K., Garcia-Lario, P., Giard,

- M., Giardino, G., Glenn, J. ., Goldsmith, P., Griffin, M., Hoare, M., Huang, M., Jiang, B., Joblin, C., Joncas, G., Juvela, M., Kirk, J., Lagache, G., Li, J. Z., Lim, T. L., Lord, S. D., Lucas, P. W., Maiolo, B., Marengo, M., Marshall, D., Masi, S., Massi, F., Matsuura, M., Meny, C., Minier, V., Miville-Deschênes, M., Montier, L., Motte, F., Müller, T. G., Natoli, P., Neves, J., Olmi, L., Paladini, R., Paradis, D., Pestalozzi, M., Pezzuto, S., Piacentini, F., Pomarès, M., Popescu, C. C., Reach, W. T., Richer, J., Ristorcelli, I., Roy, A., Royer, P., Russeil, D., Saraceno, P., Sauvage, M., Schilke, P., Schneider-Bontemps, N., Schuller, F., Schultz, B., Shepherd, D. S., Sibthorpe, B., Smith, H. A., Smith, M. D., Spinoglio, L., Stamatellos, D., Strafella, F., Stringfellow, G., Sturm, E., Taylor, R., Thompson, M. A., Tuffs, R. J., Umana, G., Valenziano, L., Vavrek, R., Viti, S., Waelkens, C., Ward-Thompson, D., White, G., Wyrowski, F., Yorke, H. W., and Zhang, Q. (2010). Hi-GAL: The Herschel Infrared Galactic Plane Survey. *PASP*, **122**, 314–325.
- Morales, E. F. E., Mardones, D., Garay, G., Brooks, K. J., and Pineda, J. E. (2009). A Multiwavelength Study of Young Massive Star-Forming Regions. III. Mid-Infrared Emission. *ApJ*, **698**, 488–501.
- Motte, F., Bontemps, S., Schneider, N., Schilke, P., and Menten, K. M. (2008). Massive Infrared-Quiet Dense Cores: Unveiling the Initial Conditions of High-Mass Star Formation. In H. Beuther, H. Linz, & T. Henning, editor, *Massive Star Formation: Observations Confront Theory*, volume 387 of *Astronomical Society of the Pacific Conference Series*, pages 22–+.
- Netterfield, C. B., Ade, P. A. R., Bock, J. J., Chapin, E. L., Devlin, M. J., Griffin, M., Gundersen, J. O., Halpern, M., Hargrave, P. C., Hughes, D. H., Klein, J., Marsden, G., Martin, P. G., Maukopf, P., Olmi, L., Pascale, E., Patanchon, G., Rex, M., Roy, A., Scott, D., Semisch, C., Thomas, N., Truch, M. D. P., Tucker,

- C., Tucker, G. S., Viero, M. P., and Wiebe, D. V. (2009). BLAST: The Mass Function, Lifetimes, and Properties of Intermediate Mass Cores from a 50 Square Degree Submillimeter Galactic Survey in Vela ($l = \sim 265$). *ArXiv e-prints*.
- Nutter, D. and Ward-Thompson, D. (2007). A SCUBA survey of Orion - the low-mass end of the core mass function. *MNRAS*, **374**, 1413–1420.
- Oey, M. S. and Clarke, C. J. (1997). The superbubble size distribution in the interstellar medium of galaxies. *MNRAS*, **289**, 570–588.
- Paladini, R., Burigana, C., Davies, R. D., Maino, D., Bersanelli, M., Cappellini, B., Platania, P., and Smoot, G. (2003). A radio catalog of Galactic HII regions for applications from decimeter to millimeter wavelengths. *A&A*, **397**, 213–226.
- Paladini, R., Davies, R. D., and DeZotti, G. (2004). Spatial distribution of Galactic HII regions. *MNRAS*, **347**, 237–245.
- Paladini, R., Montier, L., Giard, M., Bernard, J. P., Dame, T. M., Ito, S., and Macias-Perez, J. F. (2007). A broadband study of galactic dust emission. *A&A*, **465**, 839–854.
- Parsons, H., Thompson, M. A., and Chrysostomou, A. (2009). Infrared dark cloud cores in the SCUBA Legacy Catalogue. *MNRAS*, pages 1235–+.
- Peacock, J. A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *MNRAS*, **202**, 615–627.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2002). *Numerical Recipes in C++*. Cambridge University Press.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA.

- Reinecke, M., Dolag, K., Hell, R., Bartelmann, M., and Enß lin, T. (2006). A simulation pipeline for the planck mission. *A&A*, **445**.
- Robitaille, T. P., Whitney, B. A., Indebetouw, R., Wood, K., and Denzmore, P. (2006). Interpreting Spectral Energy Distributions from Young Stellar Objects. I. A Grid of 200,000 YSO Model SEDs. *ApJS*, **167**, 256–285.
- Russeil, D. (2003). Star-forming complexes and the spiral structure of our galaxy. *Astronomy & Astrophysics*, **397**.
- Russeil, D. (2003). Star-forming complexes and the spiral structure of our Galaxy. *A&A*, **397**, 133–146.
- Salpeter, E. E. (1955). The luminosity function and stellar evolution. *Astrophysical Journal*, **121**.
- Salpeter, E. E. (1955). The Luminosity Function and Stellar Evolution. *ApJ*, **121**, 161–+.
- Shu, F. H., Adams, F. C., and Lizano, S. (1987). Star formation in molecular clouds - Observation and theory. *ARA&A*, **25**, 23–81.
- Simon, R., Jackson, J. M., Rathborne, J. M., and Chambers, E. T. (2006). A Catalog of Midcourse Space Experiment Infrared Dark Cloud Candidates. *ApJ*, **639**, 227–236.
- Stahler, S. W., Palla, F., and Ho, P. T. P. (2000). The formation of massive stars. *Protostars and Planets IV*.
- Thompson, M. A., Hatchell, J., Walsh, A. J., MacDonald, G. H., and Millar, T. J. (2006). A SCUBA imaging survey of ultracompact HII regions. The environments of massive star formation. *A&A*, **453**, 1003–1026.

- van Kampen, E., Percival, W. J., Crawford, M., Dunlop, J. S., Scott, S. E., Bevis, N., Oliver, S., Pearce, F., Kay, S. T., Gaztañaga, E., Hughes, D. H., and Aretxaga, I. (2005). The extragalactic submillimetre population: predictions for the SCUBA Half-Degree Extragalactic Survey (SHADES). *MNRAS*, **359**, 469–480.
- Ward-Thompson, D., Scott, P. F., Hills, R. E., and Andre, P. (1994). A submillimetre continuum survey of pre protostellar cores. *R.A.S. MONTHLY NOTICES*, **268**.
- Wells, D. C., Greisen, E. W., and Harten, R. H. (1981). Fits: A flexible image transport system. *Astronomy & Astrophysics*, **44**.
- White, R. L., Becker, R. H., and Helfand, D. J. (1991). The infrared properties of compact galactic radio sources - the young and the restless. *Astrophysical Journal*, **371**.
- Whitney, B. A., Wood, K., Bjorkman, J., and Wolff, M. J. (2003). Two-dimensional radiative transfer in protostellar envelopes. I. effects of geometry on class I sources. *The Astrophysical Journal*, **591**(591), 1049–1063.
- Williams, J. P., Blitz, L., and McKee, C. F. (2000). The structure and evolution of molecular clouds: from clumps to cores to the imf. *Protostars and Planets IV*.
- Wolfire, M. G. and Churchwell, E. (1994). Circumstellar dust emission models. *Astrophysical Journal*, **427**.
- Wood, D. O. S. and Churchwell, E. (1989). Massive stars embedded in molecular clouds - Their population and distribution in the galaxy. *ApJ*, **340**, 265–272.
- Wood, D. O. S. and Churchwell, E. (1989). Massive stars embedded in molecular clouds: Their population and distribution in the galaxy. *The Astrophysical Journal*, **340**, 265–272.

Zinnecker, H. (2007). High-Mass Stars in the Centers of Young Dense Clusters: Mass Segregation, Binary Mergers and Gamma-Ray Bursts. In N. St.-Louis & A. F. J. Moffat, editor, *Massive Stars in Interactive Binaries*, volume 367 of *Astronomical Society of the Pacific Conference Series*, pages 669–+.

Zinnecker, H. and Yorke, H. W. (2007). Toward Understanding Massive Star Formation. *ARA&A*, **45**, 481–563.

Zoonematkermani, S., Helfand, D. J., Becker, R. H., White, R. L., and Perley, R. A. (1990). A catalog of small-diameter radio sources in the galactic plane. *Astrophysical Journal Supplement Series*, **74**.

Opinionum commenta delet dies, naturae judicia confirmat.

[Time erases the comments of opinion, but it confirms the judgements of nature.]

Cicero

