Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **466,** 3682–3705 (2017)
Advance Access publication 2016 December 23

doi:10.1093/mnras/stw3353

Distance biases in the estimation of the physical properties of Hi-GAL compact sources – I. Clump properties and the identification of high-mass star-forming candidates

Adriano Baldeschi,[1,2]★ D. Elia,[1] S. Molinari,[1] S. Pezzuto,[1] E. Schisano,[1] M. Gatti,[3] A. Serra,[4] M. Merello,[1] M. Benedettini,[1] A. M. Di Giorgio[1] and J. S. Liu[1]

[1]*Istituto di Astrofisica e Planetologia Spaziali – INAF, Via Fosso del Cavaliere 100, I-00133 Roma, Italy*
[2]*Dipartimento di Fisica, 'Sapienza' Università di Roma, Piazzale Aldo Moro 5, I-00185 Roma, Italy*
[3]*Institut de Fisica d'Altes Energies (IFAE), Edifici Cn, Universitat Autonoma de Barcelona (UAB), E-08193 Bellaterra (Barcelona), Spain*
[4]*IBM Italia – Via Sciangai 53, I-00144 Roma, Italy*

Accepted 2016 December 21. Received 2016 December 20; in original form 2016 July 21

ABSTRACT

The degradation of spatial resolution in star-forming regions, observed at large distances ($d \gtrsim 1$ kpc) with *Herschel*, can lead to estimates of the physical parameters of the detected compact sources (clumps), which do not necessarily mirror the properties of the original population of cores. This paper aims at quantifying the bias introduced in the estimation of these parameters by the distance effect. To do so, we consider *Herschel* maps of nearby star-forming regions taken from the *Herschel* Gould Belt survey, and simulate the effect of increased distance to understand what amount of information is lost when a distant star-forming region is observed with *Herschel* resolution. In the maps displaced to different distances we extract compact sources, and we derive their physical parameters as if they were original *Herschel* infrared Galactic Plane Survey maps of the extracted source samples. In this way, we are able to discuss how the main physical properties change with distance. In particular, we discuss the ability of clumps to form massive stars: we estimate the fraction of distant sources that are classified as high-mass stars-forming objects due to their position in the mass versus radius diagram, that are only 'false positives'. We also give a threshold for high-mass star formation $M > 1282 \left(\frac{r}{[\mathrm{pc}]}\right)^{1.42} \mathrm{M}_{\odot}$. In conclusion, this paper provides the astronomer dealing with *Herschel* maps of distant star-forming regions with a set of prescriptions to partially recover the character of the core population in unresolved clumps.

Key words: methods: statistical – stars: formation – ISM: clouds – infrared: ISM.

1 INTRODUCTION

The impact of massive stars in the Milky Way is predominant with respect to low-mass stars: they produce most of the heavy elements and energize the interstellar medium (ISM) through the emission of ultraviolet photons. Therefore, understanding massive star formation (MSF) is one of the most important goals of modern astrophysics (see e.g. Tan et al. 2014). The *Herschel* infrared Galactic Plane Survey (Hi-GAL; Molinari et al. 2010), based on photometric observations in five bands between 70 and 500 μm, was designed to study the early phases of star formation across the Galactic plane, with particular interest in the high-mass regime (Elia et al. 2010; Beltrán et al. 2013; Olmi et al. 2013; Veneziani et al. 2013; Molinari et al. 2014).

Star-forming regions observed in Hi-GAL span a wide range of heliocentric distances (1 kpc $\lesssim d \lesssim$ 15 kpc; Russeil et al. 2011; Elia et al. 2016), therefore the physical size of the detected compact sources (i.e. not resolved or poorly resolved) could correspond to quite different types of structures, depending on the combination of their angular size and distance.

The smallest and densest structures in the ISM, considered as the last product of cloud fragmentation, then progenitors of single stars or multiple systems, are called dense cores ($D \lesssim 0.2$ pc; Bergin & Tafalla 2007), while larger unresolved overdensities in giant molecular clouds that host these cores are called clumps (0.2 pc $\lesssim D \lesssim$ 3 pc). In few cases, very distant unresolved Hi-GAL sources may have a diameter $D > 3$ pc (Elia et al. 2016), even fulfilling the definition of cloud. Correspondingly, other distance-dependent source

★ E-mail: adriano.baldeschi@iaps.inaf.it

© 2016 The Authors
Published by Oxford University Press on behalf of the Royal Astronomical Society

properties, as mass and luminosity, are found to span a wide range of values, from typical conditions of a core to those of an entire cloud.

Unfortunately, for distant sources *Herschel* is not able to resolve the internal structure and the contained population of cores (e.g. Elia et al. 2013), therefore only global and/or averaged parameters can be quoted to describe the physical conditions of the source and the characteristics of possible star formation ongoing in its interior.

Observations at higher resolution would be needed (for example by means of sub-mm interferometry) to fully resolve any individual clump identifying all its single components, but this would require very large observing programmes with different facilities, to cover the entire Galactic plane.

To overcome empirically the lack of spatial resolution, we pursued a completely different approach, namely to consider *Herschel* maps of nearby star-forming regions and degrade their spatial resolution to simulate the view of the same region if located at a larger heliocentric distance. We implemented this idea using some nearby ($d < 0.5$ kpc) molecular clouds observed in the *Herschel* Gould Belt survey (HGBS; André et al. 2010), where the compact sources correspond to dense cores. Obviously, these maps do not represent the 'reality', in turn being those regions located at a heliocentric distance of some hundred pc. However they constitute the closest, then best resolved, available view of star-forming regions on which to base our analysis. 'Moving away' these regions to farther distances, we aim to understand not only how a region would appear if it were placed at a distance farther than the actual, but mainly at linking the physical properties of the compact sources detected in the new maps with those of the underlying source populations present in the original maps. In this way, we can evaluate the degree of information lost as a function of distance, in other words the distance bias affecting the estimation of Hi-GAL compact source physical properties. To reach this goal, we probed a set of different simulated distances for each considered region, and at each distance we repeated the typical procedures applied to Hi-GAL maps for extracting the compact sources (e.g. Elia et al. 2013), treating the simulated maps as a completely new data set, with no reminiscence either to the original map or to those 'moved' at other distances.

The first paper is organized as follows: in Section 2.1, we present the regions of the Gould Belt survey that we use in this paper; in Section 2.2, we describe the procedure of 'moving away' the regions; and in Section 2.3, we report how the detection and the photometry of the sources in all the produced maps has been carried out. In Sections 3.1 and 3.2, we describe how the number of detected sources and the fraction starless-to-protostellar changes with distance. In Section 3.3, we present the distribution of the size of the detected objects as a function of distance. Section 3.4 shows a procedure to associate the sources detected at different distances. Section 3.5 describes uncertainties in the average temperature for protostellar and prestellar objects due to distance effect. Section 4 describes the distance bias in the mass versus radius (MR) relation. A summary of main conclusions is reported in Section 6. In a second paper, we will discuss the effects of the distance bias on the luminosity versus mass diagram and on the derived star formation rate.

2 OBSERVATIONS AND METHODOLOGY

2.1 Observations and data reduction

The observations used in this paper were taken from the *Herschel* (Pilbratt et al. 2010) Gould Belt survey (André et al. 2010) for the study of nearby star-forming regions. A full description of the HGBS is given by Könyves et al. (2015). For this paper, we concentrated on a few regions among those observed in the HGBS, namely Orion A, Perseus, Serpens and Lupus III and IV.

We selected these regions both because they are close, so that we can reasonably assume that most of the cores we detected are not blended, and because they cover a range of different core masses: Orion A is a high-mass star-forming region, Perseus is an intermediate-to-low mass star-forming region, while Lupus and Serpens are forming low-mass stars. Other similar *Herschel* programmes, such as HOBYS (Motte et al. 2010) and Hi-GAL itself, observed farther regions, for which confusion is an issue: for example, in the Hi-GAL clump catalogue of the inner Galaxy (Elia et al. 2016), only 166 sources out of 36 644 having a distance estimate are located at $d < 500$ pc. Furthermore, each HGBS region has the advantage of being self-consistent in distance, which is a basic requirement for our kind of analysis. Instead, the aforementioned nearby Hi-GAL sources are generally found to be mixed, in the same maps, with sources belonging to other distance components, which would represent an irreparable contamination of our data set. On the other hand, the disadvantage of using HGBS data consists of a limited statistics due to limited map size at the furthest simulated distances.

HGBS observations were taken at 60 arcsec/s in parallel mode with the two cameras PACS (Poglitsch et al. 2010) and SPIRE (Griffin et al. 2010): the observed wavelengths were 70 and 160 μm for PACS, and 250, 350 and 500 μm for SPIRE.

Maps were generated using the Unimap (Piazzo et al. 2015) map maker for both instruments. The area considered for our work is that common to the PACS and SPIRE fields of view.

We assumed the following distances to the selected regions: 150 and 200 pc (Comerón 2008) for Lupus IV and III, respectively; 235 pc (Hirota et al. 2008) for Perseus; 230 pc for Serpens (Eiroa, Djupvik & Casali 2008) and 415 pc for Orion A (Menten et al. 2007). The distance of the Serpens molecular cloud has been a matter of controversy: some authors place the Serpens molecular cloud at approximately 400 pc. In Appendix C we discuss, briefly, how the physical properties of the Serpens molecular cloud change if we assume a distance of 436 pc (Ortiz-León et al. 2016). These regions will be the subject of dedicated papers to be published by the HGBS consortium: in this respect, we stress that in this paper we are not interested in deriving the physical properties of the sources at the nominal distances, but we want to derive how the intrinsic properties of the sources change with the distance. For this reason, we do not provide any catalogue of cores.

2.2 The simulation of increased distance

The methodology adopted in this paper simulates the view, through *Herschel*, of a region of the Gould Belt as if it were placed at a distance larger than the actual one. Of course, this procedure implies a loss of spatial resolution and detail since the angular size of the moved maps (MMs) decreases. The pipeline adopted to obtain an MM is the following:

(i) Rescaling/rebinning the original map.
(ii) Convolving the new rescaled map with the point spread function (PSF) of the instrument at the given wavelength.
(iii) Adding white Gaussian noise to the map.

In detail,
(i) A structure of spatial size L placed at distance d_0 subtends an angle $\varphi_0 = \frac{L}{d_0}$, while if the same object is moved to a distance

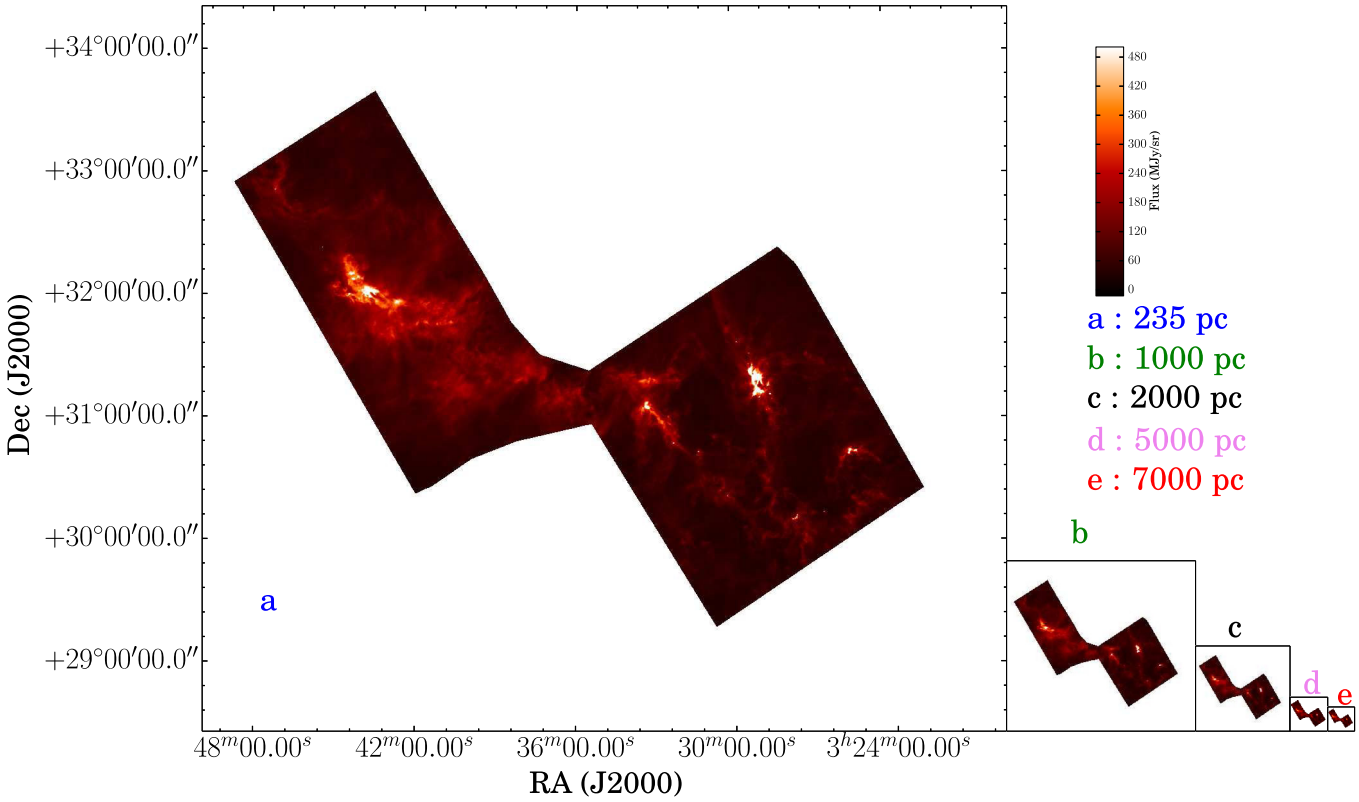

Figure 1. Original (a) and moved (b–e) maps of Perseus at 250 μm. Here, only a few of the simulated distances are shown.

$d_1 > d_0$ its angular dimension becomes $\varphi_1 = \frac{L}{d_1} < \varphi_0$. Therefore $\varphi_1 = \varphi_0 \frac{d_0}{d_1}$, so an image rebinned by a factor of $\frac{d_0}{d_1}$ mimics the movement of the region from d_0 to d_1.

(ii) To reproduce more realistically the effect of a region observed with *Herschel*, the rescaled map must be reconvolved with the PSF of the instrument. However, one must take into account the fact that the original map already results from a convolution of the sky with the PSF, i.e. a kernel that can be approximated with a Gaussian of width, $\theta_{\rm beam}$, which is equal to 8.4, 13.5, 18.2, 24.9 and 36.3 arcsec at 70, 160, 250, 350 and 500 μm, respectively (Molinari et al. 2016). So the width of the kernel we use to reconvolve the map is

$$\theta_{\rm conv} = \sqrt{\theta_{\rm beam}^2 \left(1 - \left(\frac{d_0}{d_1}\right)^2\right)}. \quad (1)$$

(iii) The noise in the maps can be well modelled as a combination of correlated noise (which is strongly attenuated by the map making algorithm; see Piazzo et al. 2015) and white noise. The map rescaling, of course, reduces white noise with respect to the original map by a factor of $\sqrt{d_0/d_1}$. The sample standard deviations of the noise in the original and rescaled map are s_N and $s_N(\sqrt{d_0/d_1})$, respectively, then to restore the noise level of the original map one has to add a white noise image to the rebinned map. In this noise image, each pixel is the realization of a Gaussian process with 0 mean and a standard deviation of $s_N\sqrt{1 - d_0/d_1}$ (to keep in all the simulated maps the same white noise level of the original one). The s_N was estimated in a box of the original map where no sources and quite low diffuse emission are found, and therefore where the signal is essentially due to statistical fluctuation.

This procedure is applied to every map at each band. We decide to 'move' the maps of each region for each band to the following virtual distances: 0.75, 1, 1.5, 2, 3, 5 and 7 kpc. Fig. 1 shows the original and MMs of the Perseus nebula at 250 μm, while in Appendix A the maps of the remaining regions are shown. Fig. 2 displays how the MMs of Perseus lose detail at increasing distance, and correspondingly, how sources resolved at the original distance become unresolved at large distances.

2.3 Source extraction and catalogue compilation

The detection and photometry of compact sources on the original and MM is carried out with CUTEX (Curvature Thresholding Extractor, presented in Molinari et al. 2011). This algorithm detects the sources as local maxima in the second derivative images and then fits an elliptical Gaussian to the source brightness profile to estimate the integrated flux. The main output parameters of the fit are: the peak position, the minimum and maximum FWHM ($\phi_{\rm min}$, $\phi_{\rm max}$) of the fitting ellipse, the peak flux and the integrated flux. Since in this paper we intend to treat the moved region at each simulated distance as an independent data set, to be analysed according to typical procedure applied to Hi-GAL fields (e.g. Elia et al. 2016), we run CUTEX on all the maps at each simulated distance for each region. Depending on the brightness level of the region, we adopted a CUTEX setup more suitable for the inner Galaxy (Molinari et al. 2016) or for the outer Galaxy (Merello et al. in preparation). Regions in the outer Galaxy have in general a lower median flux per map with respect to the inner part, due to much lower background emission. Moreover, the detection of sources in the outer Galaxy is more sensitive to pixel-to-pixel noise. For maps in the outer Galaxy, the CUTEX setup also includes the smoothing of the PACS image and a lower detection threshold for SPIRE and PACS 160 μm. We therefore made a check on our set of HGBS maps to ascertain which ones require a

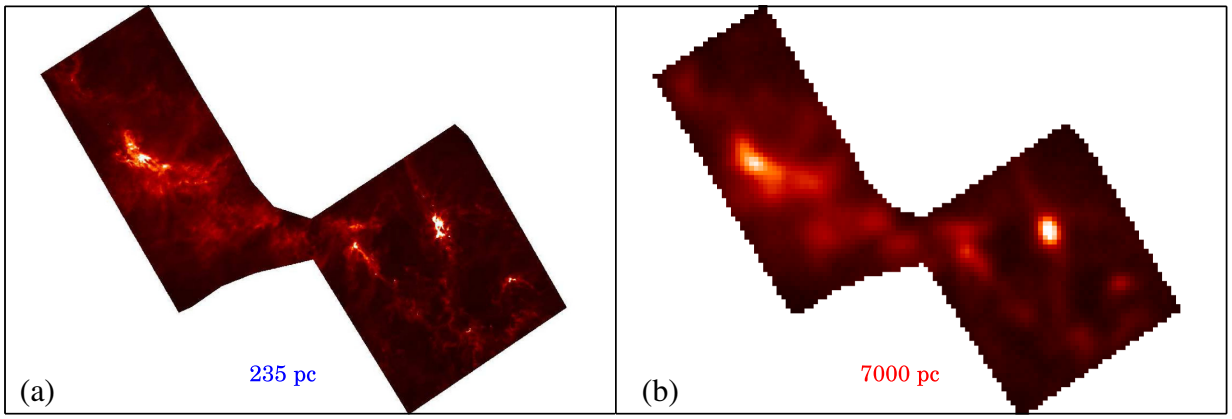

Figure 2. Perseus map at 250 μm as it appears at the original distance (panel a) is the same as panel (a) of Fig. 1 and how it would look like at the distance of 7 kpc (panel b). This latter panel has been enlarged to make the figure have the same size of panel (a). The loss of spatial resolution at 7 kpc is evident.

CUTEX setup similar to the Hi-GAL one for the outer Galaxy. Orion A and Perseus have a larger median flux (∼70 MJy sr^{-1}) respect to the others. Furthermore, Serpens, Lupus III and IV have a median absolute deviation comparable (∼11 MJy sr^{-1}) with the outer Galaxy regions such as the $2^\circ \times 2^\circ$ field centred at $l \sim 160^\circ$, showing the lack of an extended background.

Therefore, we used the inner Galaxy setup of CUTEX for Orion A and Perseus while we used the outer Galaxy setup for Serpens, Lupus III and Lupus IV. The fluxes measured with CUTEX are then corrected with the procedure described in Pezzuto et al. (2012) to take into account for the fact that the instrumental PSF is not a Gaussian, while CUTEX performs a Gaussian fit.

After the source detection and flux measurement in all five *Herschel* bands, we select the good compact source candidates (band-merging) by applying the procedure described in Elia et al. (2010, 2013) as well as Elia et al. (2016). The band merging makes it possible to build the spectral energy distribution (SED) of the sources. An SED eligible for a grey-body fit must satisfy the following criteria: (1) at least three consecutive fluxes between 160 and 500 μm, (2) showing no dips (negative second derivative) and (3) not peaking at 500 μm.

An additional issue in our case is that the absolute astrometry of the MMs has no physical meaning because the rescaling of the image shrinks their size. In the band merging procedure we take care of this issue, to ensure that the coordinates of the detected sources, at different wavelengths, are consistent with each other.

We address this issue as follows: the function that we used to execute the band merging works with equatorial coordinates. For this reason, we had to perform the band merging considering the pixel coordinates and the angular extent of the objects in the MMs, then rescaled them to the original map and then finally convert them into the equatorial system. With these quantities it is finally possible to perform the band merging.

This procedure is repeated for all maps at each wavelength. Once the SEDs are built, it is possible to perform a modified blackbody (hereafter grey-body) fit (e.g. Elia et al. 2010) described by the equation:

$$F_\nu = \frac{M}{d^2} k_0 \left(\frac{\nu}{\nu_0}\right)^\beta B_\nu(T), \tag{2}$$

where F_ν is the flux at frequency ν, M is the mass of the source located at the distance d and k_0 is the opacity at the frequency ν_0. We adopt $k_0 = 0.1\,\mathrm{cm^2\,g^{-1}}$ at $\nu_0 = 1000$ GHz (i.e. $\lambda_0 = 300$ μm; Beckwith et al. 1990); $B_\nu(T)$ is the Planck function at the temperature T. We fixed β to 2 as in Elia et al. (2013). The flux at 70 μm, where present, is not considered for the fit since it is mostly due to the protostellar content of a clump, rather than its large-scale envelope emitting as a grey-body (Elia et al. 2013). In this way, we obtain estimates of temperature and mass for each source in both the original and MMs.

3 DISCUSSION OF DISTANCE BIAS

3.1 Number of sources as a function of distance

The amount of objects detected by CUTEX in *Herschel* maps is expected to decrease with increasing simulated distance (for each band). In this section, we analyse this effect from a quantitative point of view. Note that, since in this section we study each band separately, not all the sources discussed here constitute a regular SED as those considered in the following sections.

The decrease in the number of objects with distance (see Fig. 3) is due to two main combining effects. The first one is related to the decrease of the flux with distance ($F_\lambda \propto d^{-2}$): if the flux goes below the sensitivity limit at a given band, the source is not detected any more. The second one is due to blending of sources that are close to each other in the original map and hence are not resolved any more at larger distances. The blending effect may also prevent losing some sources (since the flux is an additive quantity) that would be undetected when their flux would be below the sensitivity limit.

In principle, since the flux decreases quadratically with distance and since in a log–log plot, data show a remarkable linear correlation, we fit a power law to the data:

$$N_\mathrm{s}(d) \propto d^{-\delta}, \tag{3}$$

where N_s is the number of sources at distance d. The best-fitting power laws with the corresponding values of the δ exponent are shown in Fig. 3 and reported in Table 1. This exponent is smaller for the 70- and 160-μm bands and is found generally to be between 1 and 1.9. If we assume that the decreasing of sources is due to the decrease of the flux with distance, and assuming a power-law relation between the number of detected sources and the flux ($N_\mathrm{s} \propto F_\lambda^{-\eta}$), it turns out that $N_\mathrm{s}(d) \propto d^{2\eta-2}$. To get the value of η, we fit the distribution of the fluxes, for the different regions at 70 μm. We find that $2\eta - 2$ is equal to 1.6, 1.0 and 1.0 for Orion A, Serpens and Perseus, respectively. Their sample is too small for the Lupus III and IV. We did this analysis only at 70 μm because in the other bands the effects of the blending are more prominent. These values

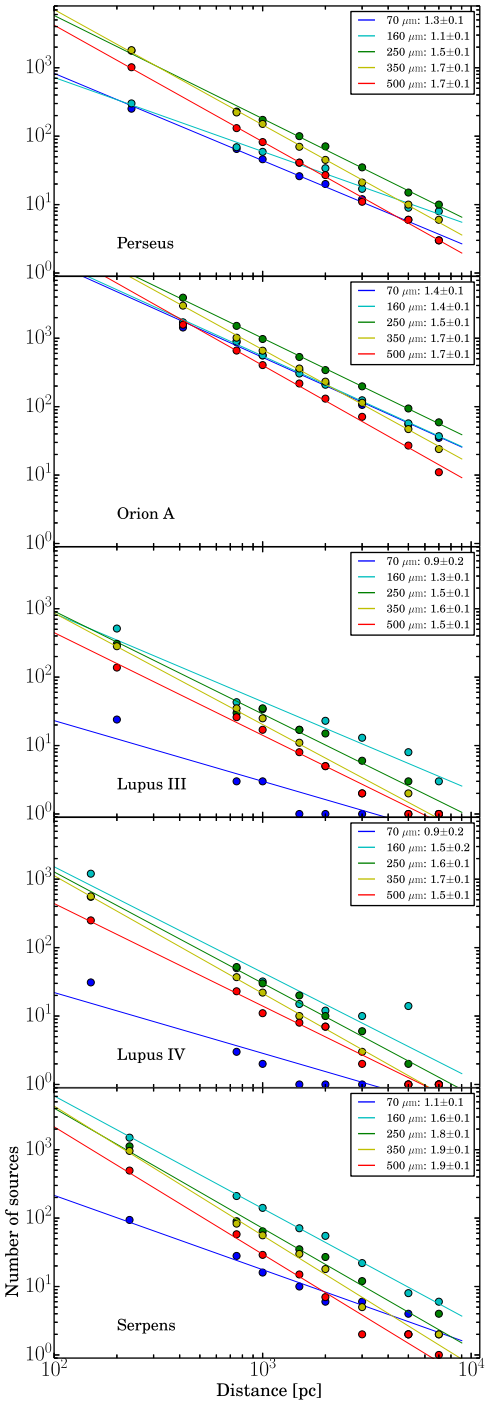

Figure 3. Number of sources detected by CUTEX at various bands for each region as a function of the virtual distance at which the map is moved. The power-law exponents δ estimated through the best fit are reported in the upper right corner of each panel.

are in good agreement with the values of δ in Table 1, considering that we make the simple assumption that the distribution of the fluxes is a power law.

Knowing the slope δ can turn out to be very interesting for practical purposes, as shown by the following example: suppose we detect N objects at a certain wavelength in an Hi-GAL region located at a distance $d > 1$ kpc; it is possible to obtain an estimate of the number of 'real' sources (at a specific band) that one would expect

Table 1. Values of the slope δ of the power-law relation between the number of sources and the distance. Uncertainty on the values of δ is almost always 0.1 (see Fig. 3). The values of δ for the 70 -μm band for Lupus III and Lupus IV are not very reliable since the number of detected sources is very low.

	70 μm	160 μm	250 μm	350 μm	500 μm
Perseus	1.3	1.1	1.5	1.7	1.7
Orion	1.4	1.4	1.5	1.7	1.7
Lupus III	0.9	1.3	1.5	1.6	1.5
Lupus IV	0.9	1.5	1.6	1.7	1.5
Serpens	1.1	1.6	1.8	1.9	1.9

to observe if the cloud was located at a distance $d_0 < d$, using the formula

$$N_0 = N \left(\frac{d}{d_0} \right)^{\delta}. \quad (4)$$

A critical point is represented by the choice of d_0, which can lead to quite different values of N_0. We suggest using values in the range 200 pc $< d_0 <$ 400 pc since this corresponds to compact source scales typical of cores. For example, let us consider a star-forming region located at $d = 5000$ pc, whose Hi-GAL map at 250 μm contains 20 sources: how many sources we would see if the region was located at $d_0 = 300$ pc? Using equation (4), we can give a rough estimate of this number assuming $\delta = 1.6$ (see Table 1) namely $N_0 = 1802$. Applying instead equation (4) to a larger d_0 (e.g. $d_0 > 1000$ pc) would imply to deal with clumps also at the original distance (i.e. with still unresolved structures).

Certainly equation (4) requires to confirmation by independent observational evidence, such as interferometric measurements aimed at exploring the real degree of fragmentation of clumps.

3.2 Starless and protostellar fraction versus distance

After assembling the SEDs, the sources are classified as follows: if a 70 -μm counterpart is present, the source is classified as protostellar candidate (e.g. Giannini et al. 2012), hereafter protostellar, otherwise it is considered a starless object. The starless sources can be further subdivided into bound (objects that can form stars and which are usually called prestellar) or unbound (transient objects that will not form stars) whether their mass is larger or smaller, respectively, than the mass given by the relation of Larson (1981):

$$M_{\rm Lars} = 460 \left(\frac{r}{[\rm pc]} \right)^{1.9} \rm M_{\odot}. \quad (5)$$

We define the starless and protostellar fraction as $n_{\rm sl}/n$ and $n_{\rm pro}/n$, respectively, where $n_{\rm sl}$ and $n_{\rm pro}$ are the number of starless and of protostellar sources, respectively, and $n = n_{\rm sl} + n_{\rm pro}$ is the total number of sources. In Fig. 4, we show $n_{\rm sl}/n$ and $n_{\rm pro}/n$ while in Fig. 5 we show $n_{\rm sl}$ and $n_{\rm pro}$ as a function of distance, respectively. The general trend for $n_{\rm pro}/n$ in Orion A, Perseus and Serpens is to increase with distance until it reaches a plateau. A different behaviour is found in Lupus III, where $n_{\rm pro}/n$ decreases and in Lupus IV where $n_{\rm pro} = 0$ to all moved distances due to the complete lack of detected protostellar sources.

The increase of $n_{\rm pro}/n$ with distance is explained with the fact that a source detected at a larger distance is an unresolved object that probably contains multiple cores (protostellar but also starless). Suppose there were an unresolved source, classified as protostellar, detected at a simulated distance d_1, and actually containing a certain amount of sources at the original distance d_0, and suppose that one of them is a protostellar and the remaining ones are starless: in that

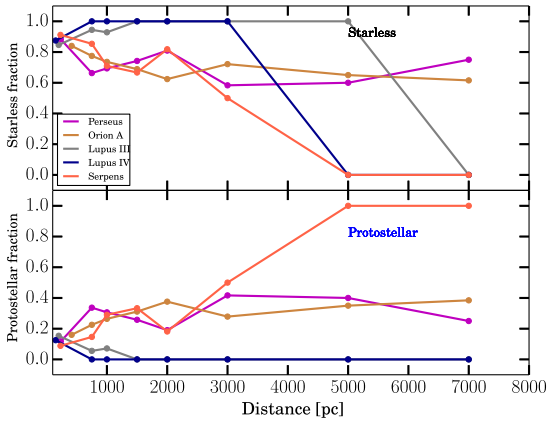

Figure 4. Fraction of starless and protostellar objects ($n_{\rm sl}/n$ and $n_{\rm pro}/n$) as a function of the distance for the five considered regions, respectively.

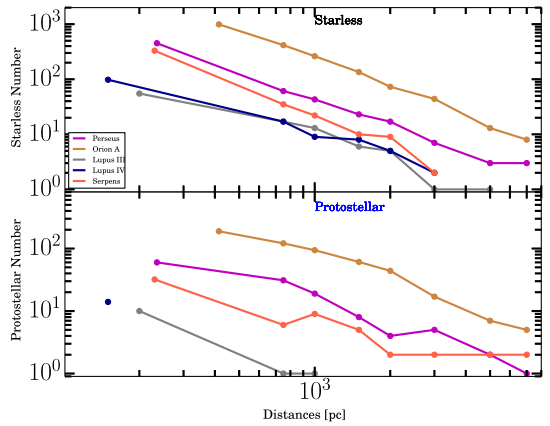

Figure 5. Number of starless and protostellar objects ($n_{\rm sl}$ and $n_{\rm pro}$) as a function of the distance.

case one would assign a protostellar character to a source which actually contains also a certain number of prestellar cores. This simple example suggests to us that in principle $n_{\rm pro}/n$ is expected to increase with distance, as long as $n_{\rm pro}/n$ reaches a plateau where the number of detected objects is very low due to the sensitivity effect. The decrease of $n_{\rm pro}/n$ with distance found for Lupus III and IV may instead be due to a weak emission at 70 μm at the original distance, which leads to a decrease in the number of protostellar at the larger virtual distances (Fig. 5).

Ragan et al. (2016) found that the behaviour of $n_{\rm pro}/n$ as a function of the heliocentric distance, in the Hi-GAL catalogue (Molinari et al. 2016), is in agreement with that we derived from Fig. 4. Indeed, they found that $n_{\rm pro}/n$ grows up to 2 kpc and then $n_{\rm pro}/n$ reaches a plateau. In their fig. 4 they found that $n_{\rm pro}/n \lesssim 0.2$ up to 2 kpc, but of the order of 0.2 for larger distances. This significantly supports the result we obtained here.

3.3 Size distribution versus distance

A common feature of molecular clouds is the hierarchical structure they show, containing bright agglomerates called clumps, that in turn are formed by smaller condensations called dense cores (see Section 1). In Fig. 6, the distribution of the radii (at 250 μm) of the detected sources is shown, for the Perseus and Orion A maps, at different distances. We do not show the same for the Lupus III, Lupus IV and Serpens due to poor statistics. As one can see from

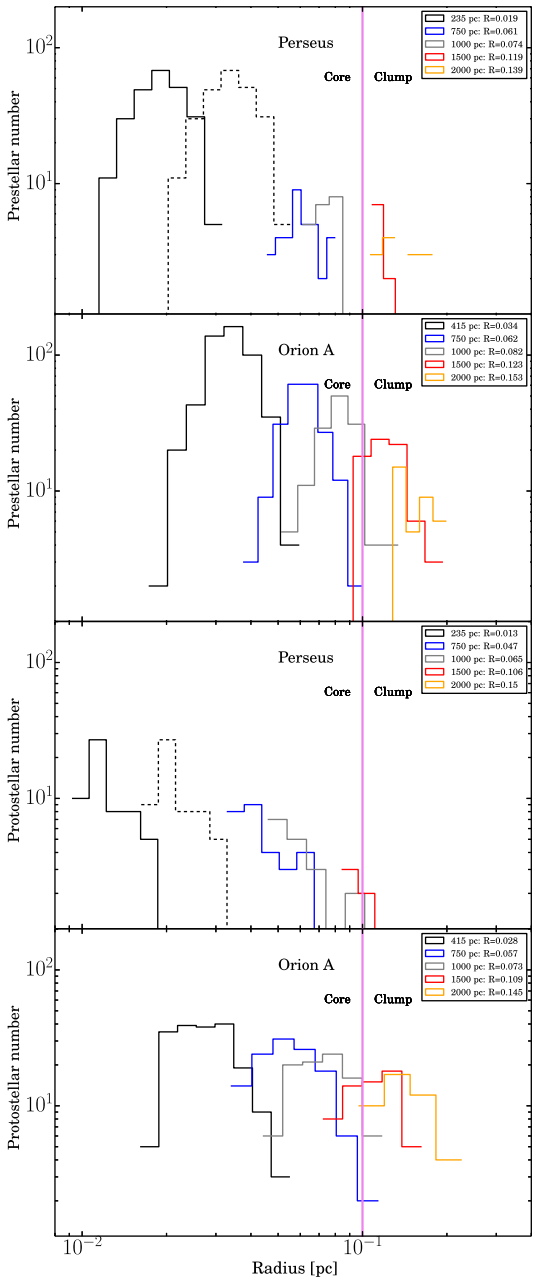

Figure 6. Radii distribution of the protostellar and prestellar sources for different distances for the Orion A and Perseus regions. The vertical magenta line represents the separation between the cores/clumps according to the classification of Bergin & Tafalla (2007), namely $R_{\rm sep} = 0.1$ pc. The values of the mean radius for all regions at each simulated distance are also reported. Dashed histogram represents the distribution of the radii, in the original map, for Perseus if it was located at an original distance of 415 pc (same as Orion A) instead of 235 pc (see Appendix C).

these figures, the physical radius increases on average, and hence the fraction of cores in the overall population of detected compact sources decreases with distance. As it emerges from this figure, up to $d = 1000$ pc, most of the detected sources are classified as cores, while at larger distances they are generally classified as clumps. Of course, the threshold $r \lesssim 0.1$ pc is not so strict and the transition between core and clump definition is not so sharp. Furthermore, a difference is found between the prestellar and protostellar source distribution: the former is characterized on average by larger radii

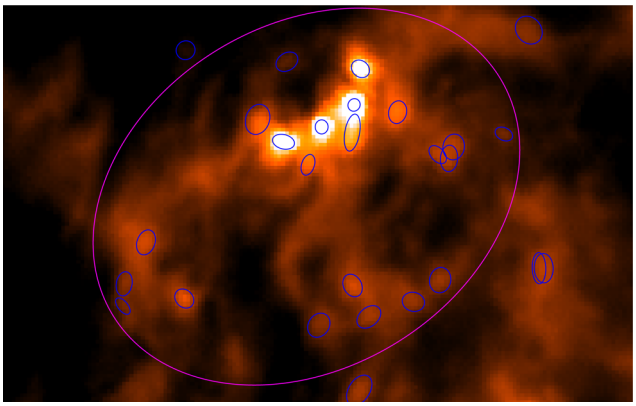

Figure 7. Example of association between a source detected in an MM and the population of cores present in the original map. A portion of the original Perseus map at 250 μm is shown, with blue ellipses representing the sources detected with CUTEX. The magenta ellipse represents a source detected in the MM at 5 kpc and projected on to the original map.

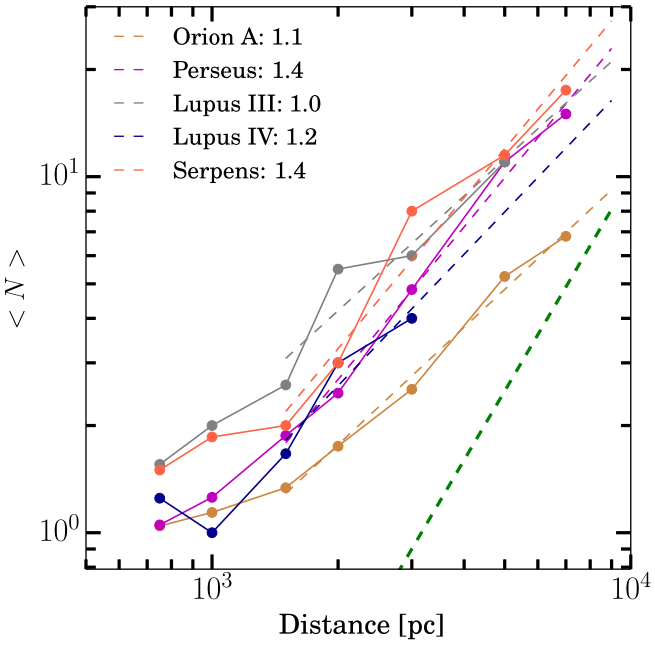

Figure 8. Average number of cores that are contained within a clump at the moved distances for each region. The best power-law fit is plotted as a dashed line, and the corresponding exponent, estimated from $d \geq 1500$ pc, is also reported. Green dashed thick line: a reference power law with exponent 2.

than the latter, as found, e.g. by Giannini et al. (2012). In Fig. 6, the behaviour of the average radius for each of the two populations is also reported (top-right corner): that of prestellar objects $\langle R_{\rm pre}\rangle$ is larger than $\langle R_{\rm pro}\rangle$ for protostellar sources, but this gap appears to get smaller at larger distances. If we define q such that $\langle R_{\rm pro}\rangle = q\langle R_{\rm pre}\rangle$, q is found to be 0.68, 0.77, 0.88, 0.89 and 1.08 for the Perseus region at 235, 750, 1000, 1500 and 2000 pc, respectively, while q is 0.82, 0.92, 0.89, 0.89 and 0.95 for the Orion A region at 415, 750, 1000, 1500 and 2000 pc, respectively.

3.4 Association of the 'moved' sources with the original ones

In Section 3.3, we have seen that for $d \gtrsim 1.5$ kpc the source sizes are such that the objects are classified as clumps. For further analysis, it becomes important to identify and count the cores present in the original map that are contained in clumps found in the MM. This information can be used, for example, to estimate which fraction of mass of such clumps comes from the contained cores, and what comes from the diffuse 'inter-core' material. To associate the sources detected at a given distance d with the original population of objects, we projected the ellipse, found at 250 μm at distance d, back to the original distance d_0, and consider the sources falling within such ellipse: an example is shown in Fig. 7. At this point, it is possible to associate the sources of the MM and those of the original one in two ways:

(1) By doing band-specific associations, including those without regular SEDs. For example, suppose to consider an object, detected at 250 μm at 5 kpc and then count the sources that, in the map corresponding to the original distance and to the same band, lie inside the area occupied by this object.

(2) Association using only sources with regular SEDs. As an example, let us suppose that there is a clump with a regular SED at 5 kpc and then count the number of cores with a regular SED that are contained in the clump, in the original map. This association between a clump and the contained cores is very useful because one can decompose an unresolved object (clump) into its smaller components (cores). Therefore, from a practical point of view, this corresponds to observing an unresolved object with a higher resolution and hence revealing its internal structure.

The former approach will be used in Section 3.4.1 where we discuss the contribution of the diffuse emission separately for the various different wavelengths, while the latter will be used both in the following and in Section 3.4.2, where we will discuss the relation between the physical properties of the moved sources and the original core population.

Fig. 8 shows the average number of cores $\langle N\rangle$, at the original distance, that are merged into one single source with a regular SED, at the moved distances. As one can see from this figure, $\langle N\rangle$ increases slowly, with distance, up to 1500 pc and then tends to increase faster. We fit the data with a power law $\langle N\rangle \propto d^{\zeta}$, starting from 1500 pc and we find values for ζ between 1 and 1.5 for the different regions. The slower increase of $\langle N\rangle$ at smaller distances is simply due to the fact that the sizes of the objects in the MMs are more similar to the sizes of the Original cores.

Moreover, the values of ζ are always smaller than 2, that is the expected value if the distribution of cores in the maps was uniform. Since, on the other hand, the actual core distribution is far from being uniform, but rather it is clustered, it is reasonable to find values of $\zeta < 2$.

The values of $\langle N\rangle$ depend both on the original average surface density Σ (measured as the number of sources per pc^2) of the core population, and on the minimum spatial scale present in the original maps. The latter effect explains, for example, why the values of $\langle N\rangle$ in the Orion A are smaller than in the other regions: the spatial detail corresponding to the *Herschel* resolution in Orion A is coarser than in Lupus, Serpens and Perseus due to the larger distance. The former effect, is instead, related to Σ which is 93.0, 56.4, 50.4, 39.7, 129.8 pc^{-2} for Orion A, Perseus, Lupus III, Lupus IV and Serpens, respectively. Considering two regions with comparable original distance, namely Serpens and Perseus, this effect can be appreciated, since Σ is larger in Serpens and, correspondingly, larger $\langle N\rangle$ is systematically found for Serpens than for Perseus.

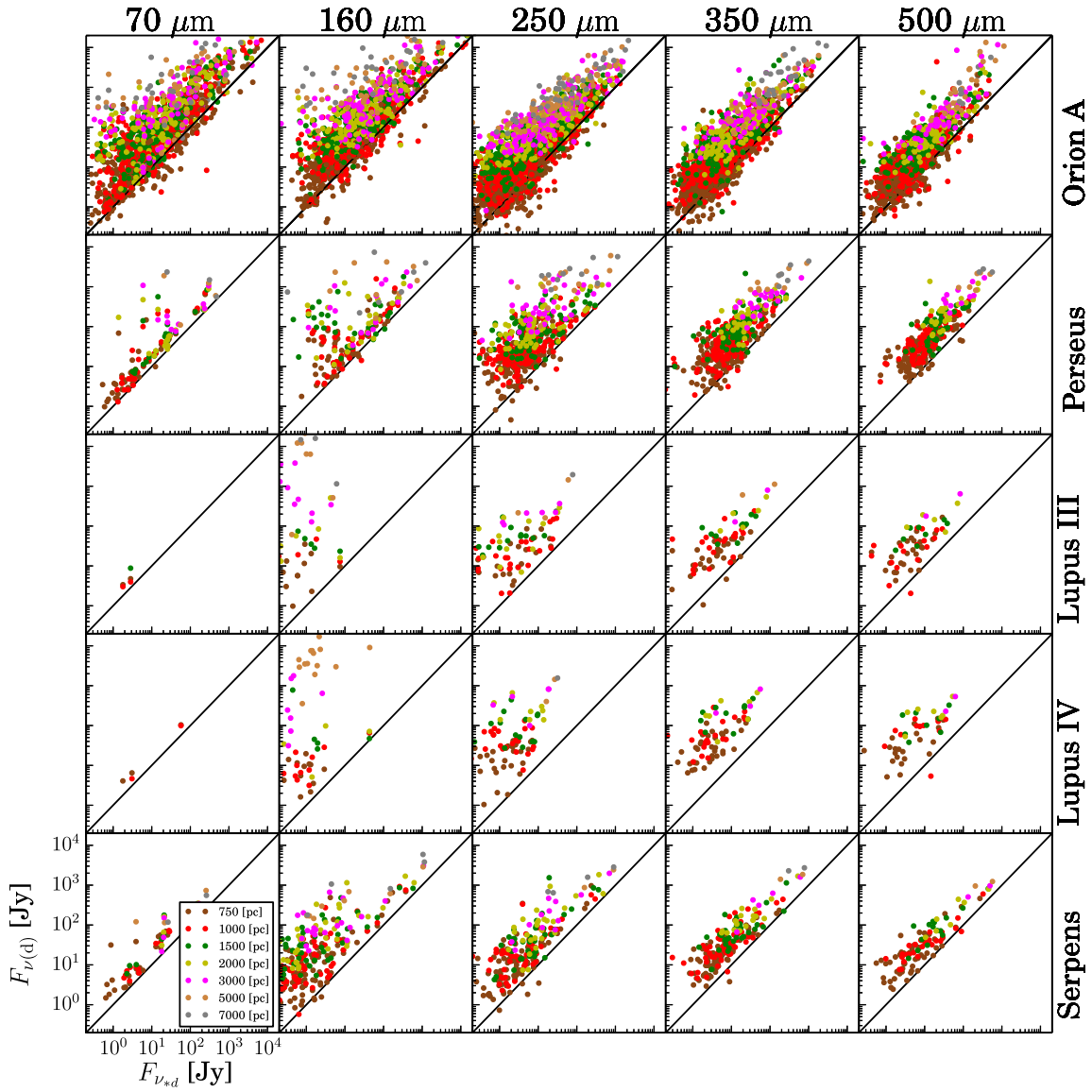

Figure 9. Flux of a source in the MM ($F_{\nu(\mathrm{d})}$) versus the sum of the flux $F_{\nu_{*\mathrm{d}}}$ of the sources in the original map that are within the moved source. The columns correspond to five different Herschel bands, while the rows correspond to the five considered regions.

3.4.1 Diffuse emission

A first comparison between the properties of the sources detected in the MM and those of the sources at the original distance that are within the rescaled object can be carried out looking at the total fluxes. We perform such analysis separately at each wavelength, therefore we also include detections that do not contribute to build regular SEDs. Therefore, we associate sources through the method (1) described in Section 3.4. Let $F_{\lambda(\mathrm{d})}$ the flux of a source detected in an MM at a wavelength λ and at distance d rescaled to the original map, and $F_{\lambda_{*\mathrm{d}}} = \sum_{\mathrm{i}=1}^{n} f_{\lambda_{\mathrm{i}}}$ the sum of the flux of the sources that are within the moved source at the original distance. Fig. 9 shows $F_{\lambda(\mathrm{d})}$ versus $F_{\lambda_{*\mathrm{d}}}$ for different regions at each wavelength: it appears that, as expected, the contribution of the diffuse emission $F_{\lambda(\mathrm{d})} - F_{\lambda_{*\mathrm{d}}}$ increases with distance. This is not surprising since the objects in the map are patchily distributed (see Fig. 7) and since the size of the sources increases with d (see Section 3.3). As the radius becomes larger, the contribution of the diffuse inter-core emission, which is expected to be more uniformly distributed, increases quite regularly with the increasing area of the source (in addition to this, an increasing amount of background emission is included in such flux estimate, as shown at the end of this section through Fig. 11). On the contrary, the contribution of emission from cores increases with distance depending on the degree of clustering.

Fig. 10 shows the average fraction of the diffuse emission $\epsilon_{\lambda(\mathrm{d})} = \left\langle (F_{\lambda(\mathrm{d})} - F_{\lambda_{*\mathrm{d}}})/F_{\lambda(\mathrm{d})} \right\rangle$ versus distance. As one can see from this figure, the contribution of the diffuse emission at 70 μm is lower than at larger wavelengths for all the considered regions; this is due to the fact that the 70-μm emission is typically associated with protostellar activity (e.g. Dunham et al. 2008; Elia et al. 2013) and is therefore more concentrated in compact structures than the emission in the other bands, which appears instead arranged in a diffuse network of cold filaments. Furthermore, $\epsilon_{\lambda(\mathrm{d})}$ increases with distance up to a certain point, which depends on the region, after which it tends to reach a plateau. In conclusion, this analysis suggests that the contribution of the diffuse emission for large distances ($d \gtrsim 1$ kpc), i.e. a typical case for Hi-GAL sources, goes from

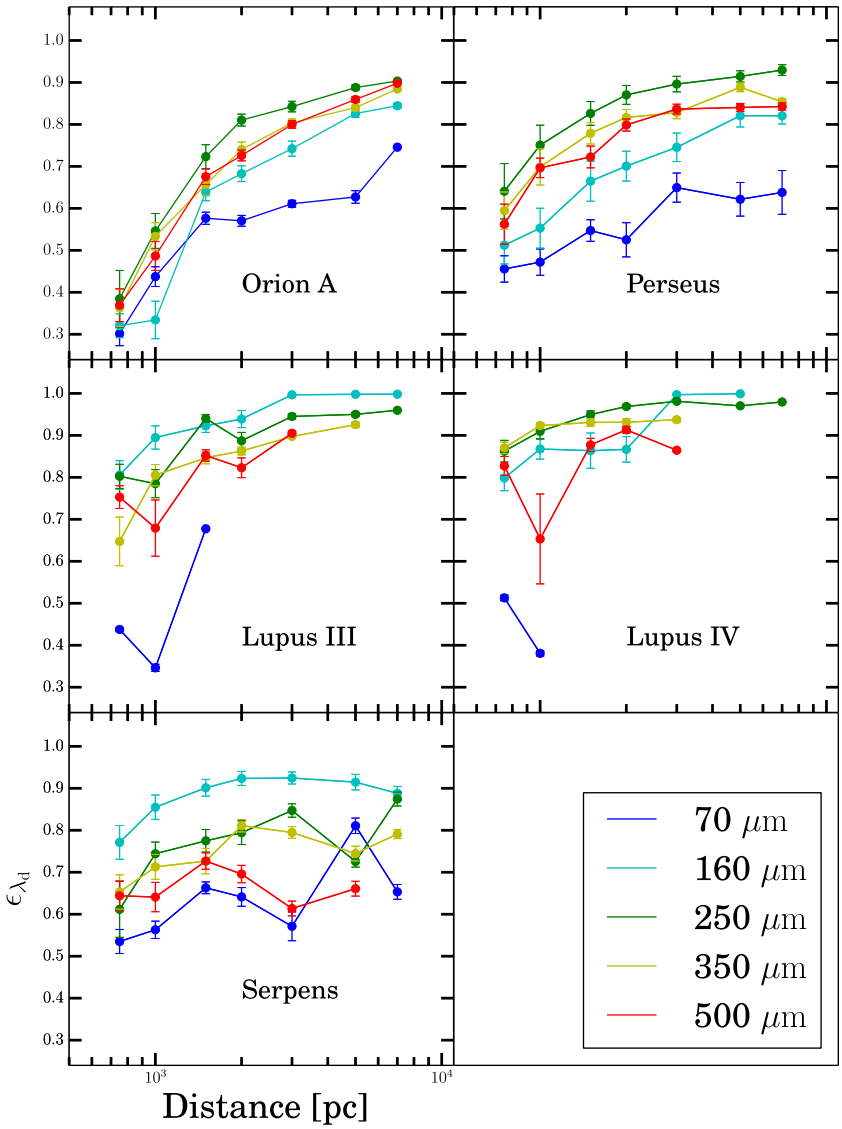

Figure 10. Mean fraction of diffuse emission $\epsilon_{\lambda(d)} = \langle (F_{\lambda(d)} - F_{\lambda_{*d}})/F_{\lambda(d)} \rangle$ versus virtual distance.

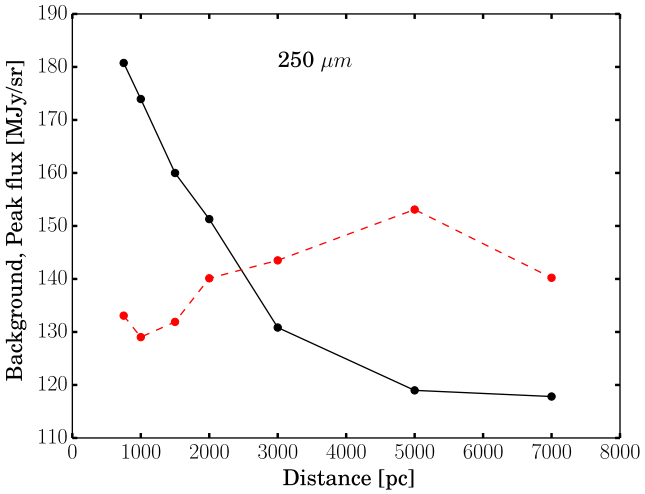

Figure 11. Red dashed: mean peak flux F_p at 250 μm for all the regions merged as a function of distance. Black solid: same as red dashed but for the background emission.

50 per cent up to 95 per cent (depending on the region) except for the 70-μm band where it goes from 50 per cent up to 80 per cent (see Fig. 10). These large values of $\epsilon_{\lambda(d)}$ suggest that most of the clump emission is due to the diffuse inter-core emission when the clump is located far away ($d \geq 1$ kpc). This has to be taken into account to distinguish the whole clump mass from the fraction of it contained in denser substructures, possibly involved in star formation processes. Such behaviour cannot be simply explained with the increasing physical size of the source, but it must be taken into account how the background level estimate changes with distance. Indeed, we expect that the background level for the sources detected in the original map is higher than that generally found in the MMs, since the background, in the original map, is due to the inter-core emission, while in the MM it is the weaker cirrus emission on which the entire clump lies.

Since CUTEX derives the flux of a source by fitting a 2D Gaussian with the formula $F_\lambda = 4.53\, a\, b\, F_p$, where a and b are the semi-axes at half-maximum of the elliptical Gaussian and F_p is the peak flux measured in MJy sr^{-1}, with increasing distance, a and b are constrained to have physical plausible size as in general is done

with the extraction tool, so the physical area we use to scale F_λ increases quadratically, on average. Moreover, the background level is generally expected to be estimated lower and lower (see above), producing an increase of F_p. To show that, we computed the mean value of the background and of F_p for all the regions merged together as a function of distance: in Fig. 11, the background is found on average to become smaller with d, and correspondingly F_p becomes larger.

Therefore, the derived values of $\epsilon_{\lambda(d)}$ may be explained roughly with the following considerations: the value of $\epsilon_{\lambda(d)}$ for one source can be modelled by the formula

$$\epsilon_{\lambda(d)} = 1 - \frac{\sum_{i=1}^{N} a_i b_i F_{p_i}}{ab F_p}, \quad (6)$$

where a_i, b_i and F_{p_i} are the parameters of the objects in the original map that are contained within a larger object at the moved distance d having a, b and F_p parameters in turn. The product ab is proportional to the square of the distance, and we roughly assume that the peak flux of the sources at the original distance is the same for all sources. The peak flux F_p is weakly dependent on distance (Fig. 11): on one hand, the dramatic drop of the background emission at increasing distance should be expected to produce correspondingly, by subtraction, a strong increase of F_p. However, this is not observed mostly due to the fact that the distance increase implies the averaging of boxes of pixels implied by the map rebinning we impose to simulate the distance effect. In particular, we find that F_p at the largest probed distance is at most 1.5 times the average peak flux for smaller distances, therefore adopting this value we get that $\epsilon_{\lambda(d)} = 1 - \left(\frac{d_0}{d}\right)^2 N(d)/1.5$, where N is the number of contained sources within the moved source. For example, if we consider the Perseus region observed at 250 μm at 5000 pc assuming $N(d) = 22$ (as in Fig. 7), we get $\epsilon_{\lambda(d)} = 0.97$, which is in good agreement, despite the naivety of the model, with the result of Fig. 10, namely $\epsilon_{\lambda(d)} = 0.9$.

3.4.2 Physical properties of the 'moved' clump and those of the original core population

In this section, we consider the mass of those sources of the MM for which we performed the grey-body fit, to understand how these quantities for a moved source mirror those of the original core population rather than those of the diffuse material. We define with M_d the mass of the moved detected sources (protostellar and prestellar) at distance d, and with $M_{*_d} = \sum_{i=1}^{n} m_{i_d}$ the sum of the masses of all the original sources (protostellar and starless) that lie inside the source when it is reported in the original map. Fig. 12 shows the average values of M_d versus M_{*_d} at distance d: we find that $\langle M_d \rangle$ is always larger than $\langle M_{*_d} \rangle$. A similar conclusion can be obtained if one uses the median statistics, instead of the mean: the median mass of the clumps at large distances is by far larger than the sum of the masses of the contained cores, as well.

We want to know the effect of distance on the determination of the core formation efficiency. The average core formation efficiency, that we define as $\langle \mathrm{CFE} \rangle = \langle M_{*_d}/M_d \rangle$, can be derived from the values plotted in Fig. 12 (panel a). In Fig. 13, we show $\langle \mathrm{CFE} \rangle$ as a function of distance for each of the considered regions. It can be seen that the $\langle \mathrm{CFE} \rangle$ changes from region to region and tends to decrease with d for distances below 1500 pc while, at large distances, it becomes independent of d. The decrease of the $\langle \mathrm{CFE} \rangle$ with d below 1500 pc is due to the fact that the size of the clumps and the contained cores are quite similar and hence the fraction

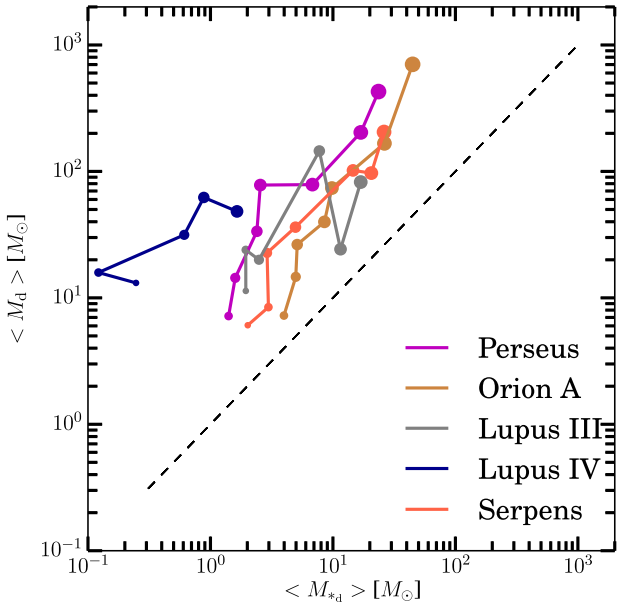

Figure 12. Average mass M_d of the moved sources for different distances, versus the average of the sum of the masses of the sources in the original map that are inside the rescaled sources. The size of the filled circles increases with distance. Black dashed line is the bisector of the plot.

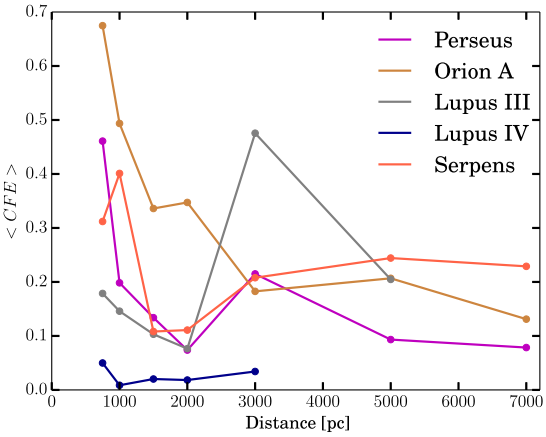

Figure 13. Average core formation efficiency versus distance for each of the considered regions.

of diffuse material remains low. In fact, this effect is particularly prominent in Orion A, because it is located at 415 pc while we do not see this effect in the Lupus IV, which is located at 150 pc. The values of the $\langle \mathrm{CFE} \rangle$ at large distances are between 1 per cent and 20 per cent, depending on the region.

3.5 Mean temperature versus distance

Let us consider now the effects of the distance on the mean temperature of the sources detected in each region at all virtual distances, comparing it with the temperature of the original population of cores that fall within each clump. We fit the average temperature with a power law $\langle T \rangle_d \propto d^\xi$. The uncertainties of $\langle T_d \rangle$ are given by $\frac{S_{T_d}}{\sqrt{n}}$, where S_{T_d} is the sample standard deviation of the temperature at distance d and n is the sample size. Fig. 14 shows $\langle T_d \rangle$ for each d together with the best power-law fit and the values of ξ. We find that $\langle T \rangle_d$ increases slowly with distance for the prestellar objects,

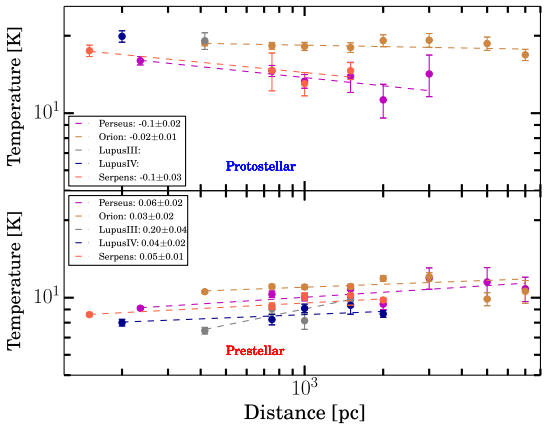

Figure 14. Mean temperature of the prestellar (bottom) and protostellar (top) sources as a function of distance. The best fit is also plotted with the relative power-law exponent.

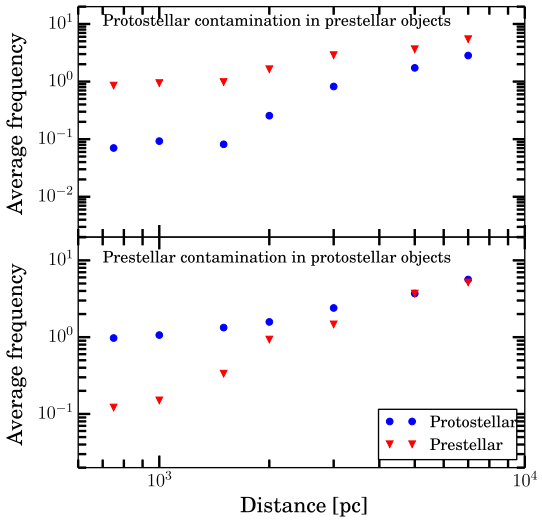

Figure 15. Upper panel: average protostellar contamination in prestellar clumps. Lower panel: average prestellar contamination in protostellar clumps. At each distance all the regions are merged together.

while it decreases slowly for the protostellar ones. The values of ξ for the prestellar objects (see Fig. 14, lower panel) are 0.06, 0.03, 0.2, 0.04 and 0.05 for Perseus, Orion A, Lupus III, Lupus IV and Serpens, respectively. These slopes are very shallow but are positive anyway. An opposite trend (i.e. temperature weakly decreasing with distance) is found for the protostellar objects: the values of ξ are -0.1, -0.02 and -0.1 for Perseus, Orion A and Serpens, respectively. The statistics are not sufficient at large distance to perform the fit for the Lupus III and IV regions. It appears unlikely that both of those behaviours are due by chance, since we systematically find these trends in all the considered regions.

We are quite confident that the decreasing behaviour of temperature for the protostellar sources and the increasing one for the prestellar ones is due to the effect of confusion and can be possibly explained by taking into account two main factors likely to be concomitant in most cases.

The first is related to the fact that the protostellar objects are typically warmer than the prestellar ones (e.g. Elia et al. 2013). A protostellar clump detected at a large distance is probably an unresolved object containing in turn some smaller objects that can be starless or protostellar. If inside the unresolved protostellar clump there are some prestellar ones, this can decrease the global average temperature. In the lower panel of Fig. 15, we show the prestellar contamination of sources, namely the average number of prestellar cores contained in protostellar clumps (for all the regions merged together). The prestellar contamination, as expected, is close to 0 up to 1000 pc and then starts to get larger, in particular between 5000 and 7000 pc the average number of original prestellar cores contained in the protostellar moved clumps becomes the same as the original protostellar objects. On the other hand, if we take an unresolved prestellar clump at large distance this can contain some protostellar cores, whose flux at 70 μm goes below the sensitivity threshold (and hence not detected), that can contribute anyway to increase the average temperature. Indeed, although the signature of the presence of a protostellar component would remain undetected, in such case the effect of this component on the remaining wavelengths ($\lambda > 160\,\mu$m) would remain still observable as an unnaturally high temperature for a prestellar source. In the upper panel of Fig. 15, we show the average protostellar contamination in the prestellar clumps: the protostellar contamination is close to 0 at low distances, while it becomes larger at increasing distances. The second effect responsible for the temperature decrease for the protostellar sources and increase for the prestellar ones is related to the presence of the diffuse inter-core material, which is known for being typically warmer than the prestellar sources and colder than the protostellar ones (Elia et al. 2013). This might homogenize the temperature of the protostellar and prestellar as the physical radius of the clumps increases with distance.

4 MASS–RADIUS RELATION

The MR relation for cores/clumps is a useful tool for checking the conditions for MSF. Indeed several authors (e.g. Krumholz & McKee 2008; Kauffmann & Pillai 2010) use such diagram to identify thresholds in column density supposedly required for the formation of stars with $M > 10\,\mathrm{M}_\odot$.

Here, we want to investigate how the predictive ability of the MR plot is affected by the distance effects. In Figs 16–20, the MR plot is shown for the different regions. The MR diagram for sources found in the original map is reported in the panel (a) of each of these figures. With open red and green circles we indicate bound and unbound starless objects, respectively, while with open blue circles we indicate protostellar objects. The green dashed line represents Larson's relation reported in equation (5), while the filled sky-blue and pink area of the diagram corresponds to the conditions compatible with possible MSF, according to two different prescriptions, respectively: the first (which includes the second) corresponds to the criterion of Kauffmann & Pillai (2010, hereafter KP), namely $M(r) \geq 870 \left(\frac{r}{[\mathrm{pc}]}\right)^{1.33} \mathrm{M}_\odot$, which is an empirical limit for MSF based on observations of infrared dark clouds; the latter corresponds to the criterion of Krumholz & McKee (2008, hereafter KM), which is a more demanding threshold on column density of $1\,\mathrm{g\,cm^{-2}}$ for MSF, based on theoretical calculations. This theoretical limit can be also expressed as $M(r) \geq 15042 \left(\frac{r}{[\mathrm{pc}]}\right)^2 \mathrm{M}_\odot$. We assume that these MSF thresholds break at masses lower than $M = 20\,\mathrm{M}_\odot$, since the typical values of the core-to-star conversion efficiency are 0.5–0.33 (Alves, Lombardi & Lada 2007; KP) and hence it is not reasonable that a core with a mass lower than $20\,\mathrm{M}_\odot$ can form a $10\,\mathrm{M}_\odot$ star.

In the panels from (b) to (i) of Figs 16–20, the diagram is built for the objects detected in the MM of different regions for various distances, and compared with the case of the original map

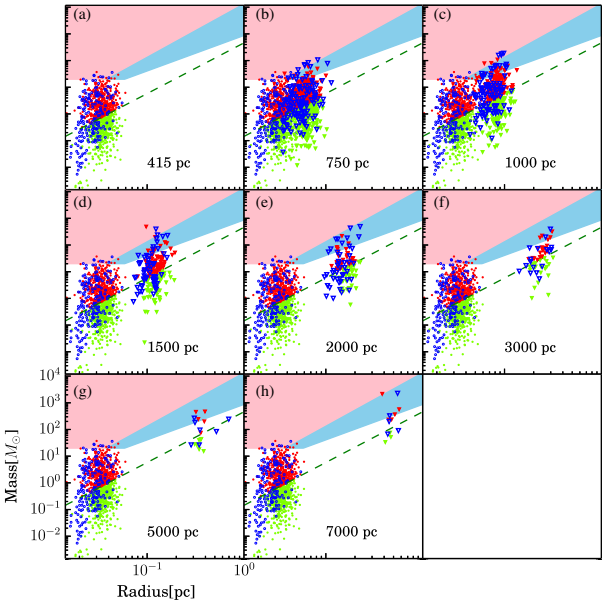

Figure 16. MR relation for the Orion A region. Red, blue and green circles are prestellar, protostellar and starless unbound sources at the original distance and are shown in all the panels for reference. Larger triangles represent the objects detected in the MMs at different distances (panels b–i). Dashed green line is the Larson's mass (see equation 5). Pink and sky-blue zones correspond to thresholds for compatibility with MSF provided by KM and KP, respectively.

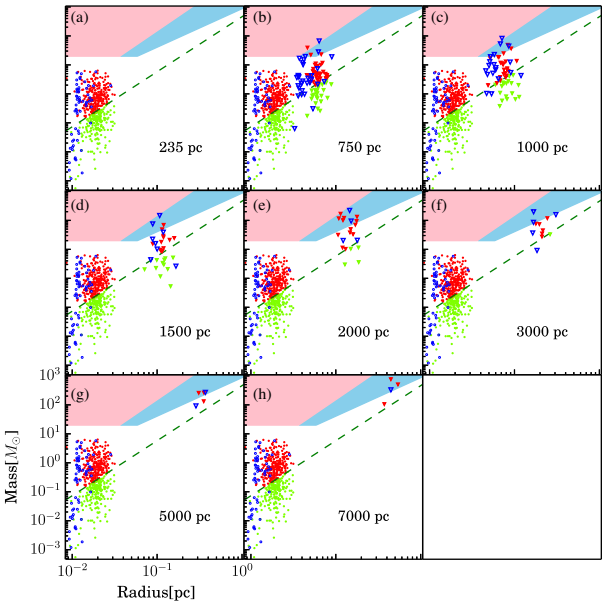

Figure 17. The same as Fig. 16 but for the Perseus region.

(panel a). Fig. 16, in particular, displays the MR diagram for the Orion A region; this nebula is known to be an MSF molecular cloud (see e.g. Genzel & Stutzki 1989; Polychroni et al. 2013), and this is consistent with the fact that several sources lie inside the MSF zone of the plot. If the Orion A map is moved to larger distances (panels from b to i), this region can still be considered an MSF nebula, based on this plot. This implies that apparently the intrinsic character of an MSF region is still conserved if this region is observed at larger distances. This remark is based on the case of a single region. However, from panel (a) of Fig. 16 one can deduce that for another MSF

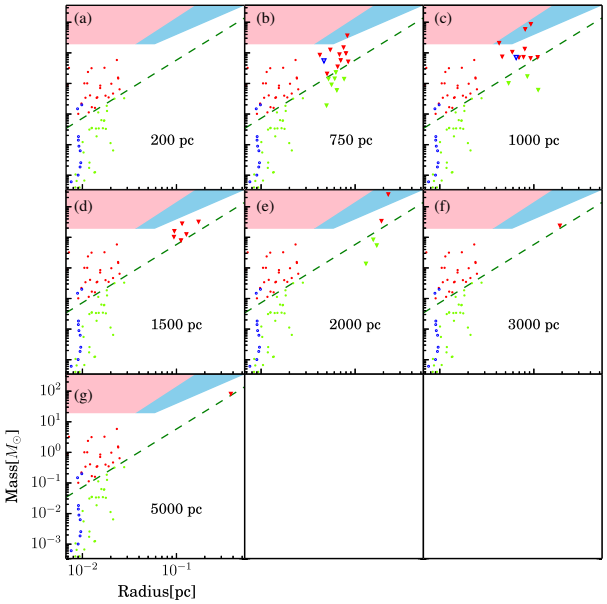

Figure 18. The same as Fig. 16 but for the Lupus III region. The panel for 7 kpc is missing since no sources are detected at this virtual distance.

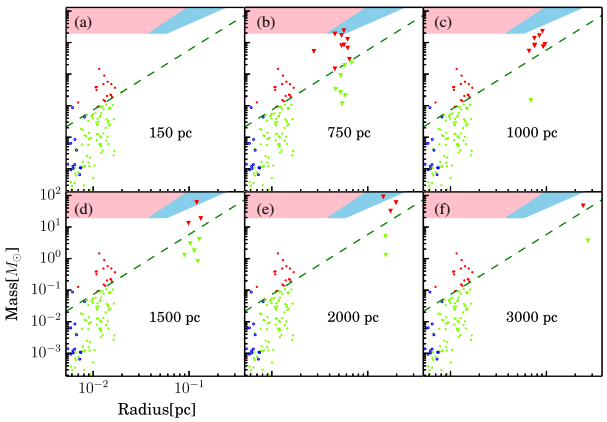

Figure 19. The same as Fig. 16 but for the Lupus IV region. The panels for 5 and 7 kpc are missing since no sources are detected at these virtual distances.

region containing cores denser and more massive than those found in Orion A, clumps extracted in the MMs would likely continue to populate even more so the KP region of the plot. In Figs 17–20, the same analysis is repeated for the other regions. These regions are known for not being MSF, and this is consistent with the fact that there are no sources inside the blue and pink zone in panel (a) of all of these figures. However, at larger distances (panels b–i) sources are found inside the zone of the plot compatible with both MSF prescriptions, and in particular in the KP zone; therefore, clearly, the distance is biasing the character one assigns to a region regarding its ability to form massive stars.

The Perseus region (Fig. 17) at the virtual distances (panels b–i) would be always classified as MSF according the KP prescription. The Serpens nebula (Fig. 20) at the moved distances would be classified as MSF, except at 1000 and 5000 pc. The Lupus III and IV regions (Figs 18 and 19) are characterized by a regime of somewhat lower masses compared with other regions (Benedettini et al. 2015). The Lupus III, at the virtual distances, would be classified as MSF at 750, 1000 and 2000 pc, while it is not classified as MSF at

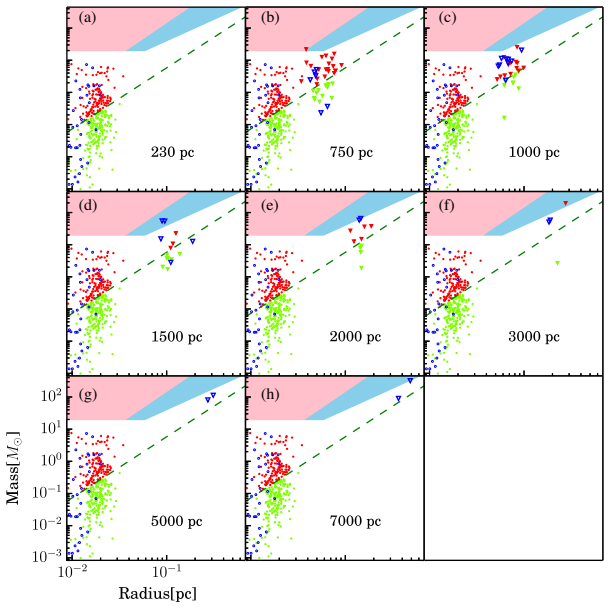

Figure 20. The same as Fig. 16 but for the Serpens region.

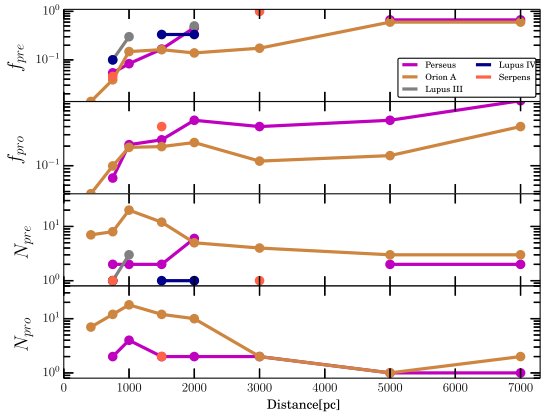

Figure 21. Fraction of prestellar $f_{\rm pre}$ and protostellar $f_{\rm pro}$ objects above the KP zone in the MR plot as a function of distance, with also the absolute number of prestellar ($N_{\rm pre}$) and protostellar ($N_{\rm pro}$).

1500–3000 and 5000 pc. The Lupus IV nebula is classified as MSF at 750, 1500 and 2000 pc, while it would be classified as non-MSF at 1000 and 3000 pc.

At this point, we want to quantify how the fraction of supposed MSF objects varies with distance: we define, at each distance, the fraction of prestellar and protostellar objects inside the MSF region (one of the two zones of the plot where MSF is possible according to KP and KM prescriptions), as $f_{\rm pre} = N_{\rm pre}/N_{\rm PRE}$, $f_{\rm pro} = N_{\rm pro}/N_{\rm PRO}$, where $N_{\rm pre}$ and $N_{\rm pro}$ are the number of prestellar and protostellar inside the MSF region, while $N_{\rm PRE}$ and $N_{\rm PRO}$ are the total number of prestellar and protostellar. Looking at Figs 16–20, it appears that $f_{\rm pre}$ and $f_{\rm pro}$ increase with distances for the KP zone, while they seem to decrease for the KM one. This statement can be made more quantitative through Fig. 21, that shows $f_{\rm pre}$ and $f_{\rm pro}$ as a function of distance for the KP zone. This figure highlights a clear trend of $f_{\rm pre}$ and $f_{\rm pro}$ to increase in the KP zone up to ∼1000–2000 pc (depending on the region) and then to reach a plateau (Fig. 21). The former of these trends is quite steep for both $f_{\rm pre}$ and $f_{\rm pro}$. The gap $f_{\rm pre}(d) - f_{\rm pre}(d_0)$ (and similarly for $f_{\rm pro}$) is found to be larger than the largest error bar associated with these points, indicating that the increase of this fraction is statistically significant.[1] For example, for Orion A, $f_{\rm pro} = 0.03$ at d_0 and $f_{\rm pro} = 0.2$ at $d = 1000$ pc, while the largest uncertainty among $f_{\rm pro}$ values at $d = d_0$ is 0.06. As for the plateau behaviour for the two regions with enough statistics, namely Orion A and Perseus, the average value of these fractions is ∼37 per cent and ∼60 per cent for the prestellar sources, respectively, and ∼22 per cent and ∼60 per cent for the protostellar ones, respectively. We do not show the trend of $f_{\rm pre}$ and $f_{\rm pro}$ for the KM zone because of the poor statistics. With respect to the KP criterion, the ratios $f_{\rm pre}$ and $f_{\rm pro}$ are found to increase for distances up to 1000 pc. This is mostly due to the sharp break of the KP relation we impose at $M < 20\,{\rm M}_\odot$ (corresponding to $r_{\rm break} = 0.06$ pc): indeed, cores in the original map at the smallest probed distances typically have radius smaller than $r_{\rm break}$. At the virtual distance of 1000 pc, all detected sources have, instead, $r > r_{\rm break}$. Therefore some of them, classified as not-MSF in the original map, can be more easily found inside the KP area of the diagram (see Figs 16–20). This turns out in the behaviour of Fig. 21 (first two panels starting from the top), where a plateau in the $f_{\rm pre}$ and $f_{\rm pro}$ versus d relation is reached at $d = 1000$ pc.

To understand the effects of distance on the MR relation, we fit the data for prestellar and protostellar objects, with a power law $M \propto r^\alpha$ in the following way: for each distance d we calculate the average values of the mass $\langle M \rangle_{\rm d}$ and radius $\langle r \rangle_{\rm d}$ of the sources in the sample and the corresponding standard error. The fit is shown in Fig. 22 with the corresponding α exponent evaluated separately for prestellar and protostellar objects. They range mostly between the exponent of the KP relation, 1.33, and the one of the KM prescription, 2.0; this means that if one decides to use the KM criterion for inferring conditions for MSF, one probably loses sources with increasing distance, while on the contrary one tends to get some false positives (FPs) if the KP criterion is adopted.

To examine in more detail how the properties of a far source mirror those of the contained core population, now let us consider the relation between the clumps detected at the largest probed distance and the 'contained' sources at all smaller distances, adopting the procedure described in Section 3.4. Let $O_{\rm d_{max}}$ a source detected at the largest distance $d_{\rm max}$ and $O_{\rm d}$ the sources at distance $d < d_{\rm max}$ that fall within $O_{\rm d_{max}}$. Figs 23–27 contain an MR diagram for each $O_{\rm d_{max}}$ and the corresponding $O_{\rm d}$. In detail, in Fig. 23 we display four $O_{\rm d_{max}}$ cases found in the Perseus nebula: three of them are classified as MSF according to KP criterion (panels b, c and d) and one of them as low-mass star forming (panel a). The corresponding $O_{\rm d}$ are classified in some cases as MSF, while in other cases as low-mass star forming; in particular, looking at the sources detected at the original distance, none of them is found to be MSF.

Fig. 24 contains the cases found in Orion A. Again, at the moved distances, many sources are classified as MSF according to the KP criterion. In panels (b) and (d), also some $O_{\rm d}$ found at the original distance are classified as MSF. This suggests that the character of an intrinsically MSF clump (or of an entire MSF region as Orion A) is preserved under the moving procedure we apply, while for a low-mass star-forming region like Perseus a spurious classification as MSF can be introduced as an effect of distance bias.

This can be observed also in the behaviour of $O_{\rm d}$ in the other low-mass star-forming regions we considered, namely Lupus III, IV and Serpens (Figs 25, 26 and 27): also in these cases some $O_{\rm d}$ at intermediate distances are found to lie in the KP zone of the MR diagram, although they do not contain sources in the original

[1] Error bars are estimated by assuming a counting statistics.

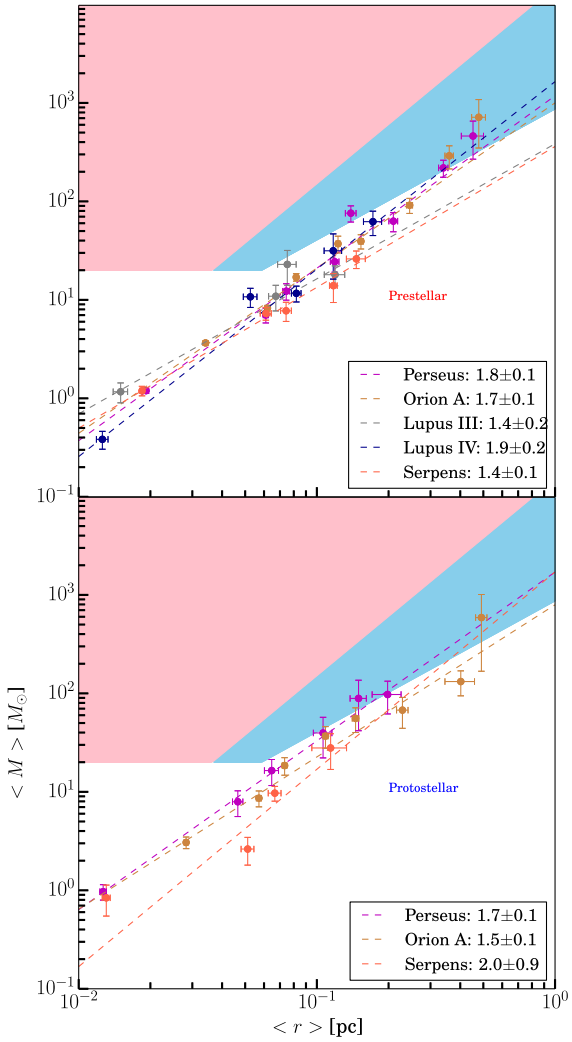

Figure 22. MR plot for the mean mass and the mean radius of the sources found at each simulated distance. The best power-law fit of these points is shown, and the corresponding power-law exponent, estimated through best fit, is reported.

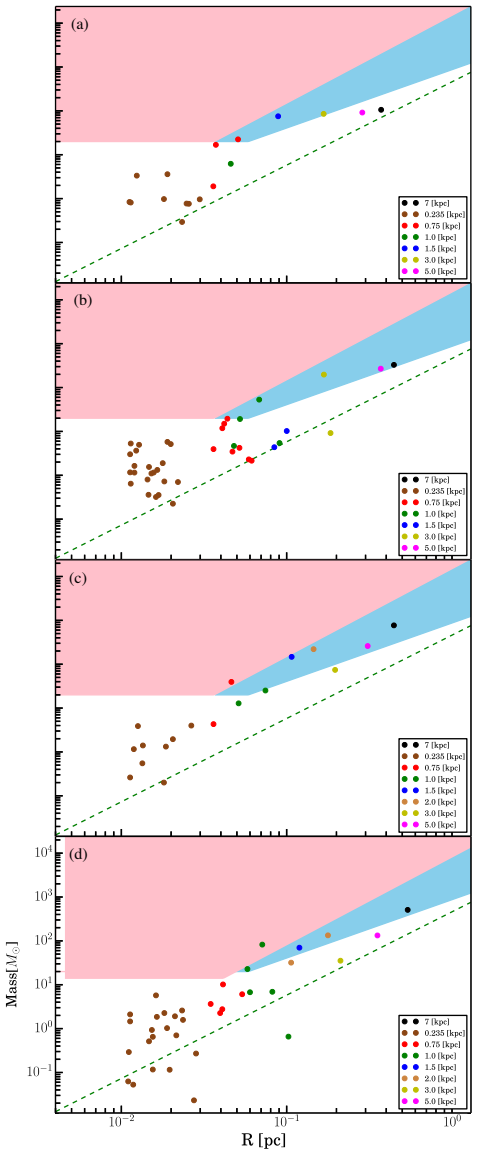

Figure 23. MR plots of sources found in the Perseus maps moved to various distances. Each plot is built starting from one of the four sources found at the largest distance (7 kpc), and contains all the sources found at shorter distances and contained in it.

map with this property. In Figs 25 and 26, we notice an apparently unnatural behaviour: a $O_{\rm d}$ source has a mass larger than that of the corresponding $O_{\rm d_{max}}$ containing it.

Many factors can contribute to such cases occurring: multiplicities in source detection which appear/disappear at different distances and intrinsic fluctuations in CUTEX flux extraction can change the shape of the SEDs of a source observed at two different distances. A shift of the SED peak results in a different temperature estimate, and in turn in a change of the mass that can be very relevant at very small temperatures, a shift of few K may lead to an order of magnitude change in mass (see e.g. Elia & Pezzuto 2016).

In conclusion, the analysis of the MR diagram, carried out by means of Figs 16–27 suggests that (1) the region that is recognized as MSF at the nominal distance (Orion A) is still classified as MSF at each of the moved distances; (2) the regions that are not recognized as MSF at the original distance (Perseus, Serpens, Lupus III and IV) are classified as MSF for most of the virtual distances; (3) the fraction of objects fulfilling the KP relation increases for each region up to 1000 pc and then reaches a plateau and (4) the mean value of the clump mass at a certain distance is generally related to the mean value of the radius with a power law with an index larger than 1.33 (KP criterion) and smaller than 2 (KM criterion) for all regions. Therefore, we find a trend to introduce MSF objects, on average, if one adopts the KP prescription and to lose MSF objects if one chooses the KM one.

For these reasons, it is important to estimate the fraction of misclassified objects according to the KP prescription. To increase statistics, all regions are merged together in this analysis. We define as FP the objects, detected at the moved distance, that are classified as MSF although they do not contain MSF cores at the original distance; with true negative (TN) the objects that are not classified as MSF at the moved distance and do not have MSF association with the objects at the original distance. The true positive (TP) are objects that are classified as MSF at the moved distances and have associations with MSF objects at the true distance. Finally, the false negative (FN) are the objects that are not classified as MSF at the moved distances and are associated with MSF at the original distance. Given these definitions, our attention has to be focused on

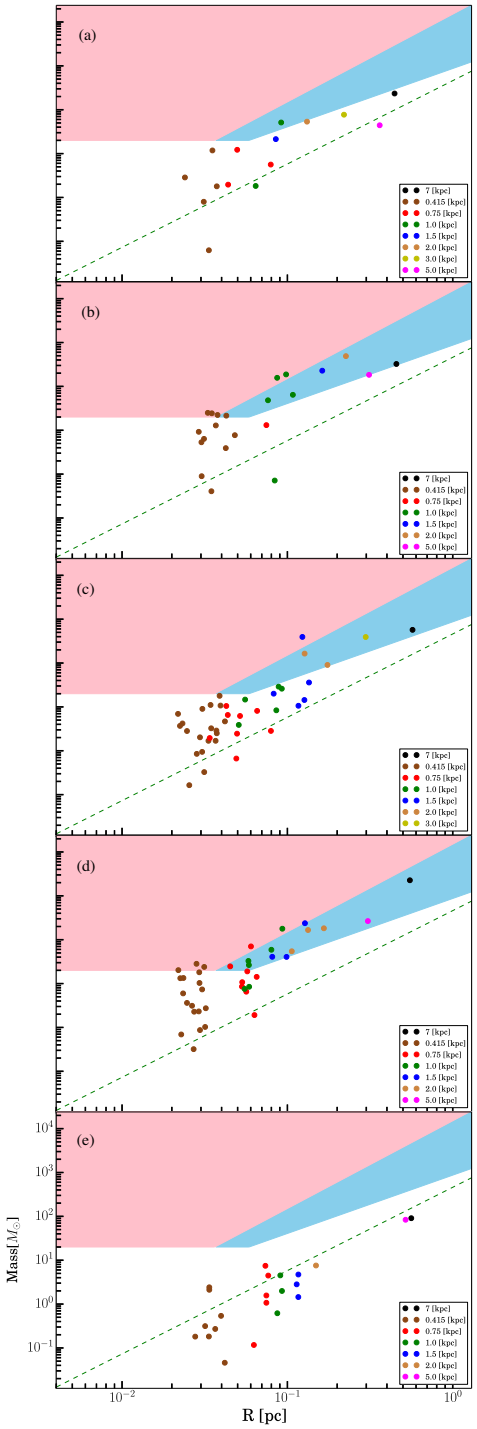

Figure 24. Same as Fig. 23 but for the Orion A. Here, we show only 5 out of 10 cases of associations between the sources found at shorter distances.

the fraction of FP with respect to the population from which they originate (and to which they should ideally belong), namely that of TN. Therefore, we define $f_{\rm kauff} = n_{\rm FP}/(n_{\rm FP} + n_{\rm TN})$, where $n_{\rm FP}$ is the number FP at distance d and $n_{\rm TN}$ is the number of TN at distance d; the $f_{\rm kauff}$ fraction as a function of distance gives an indication of the rate of misclassification of the objects introduced by distance.

As one can see in Fig. 28, $f_{\rm kauff}$ increases between 750 and 1000 pc, remains almost constant between 1000 and 3000 pc and then shows fluctuations above 3000 pc. Such fluctuations are due

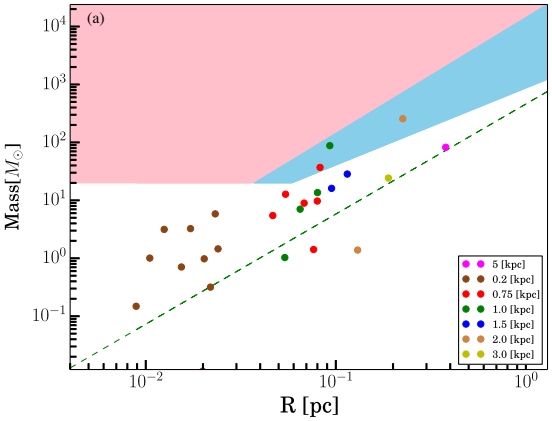

Figure 25. Same as Fig. 23 but for the Lupus III region.

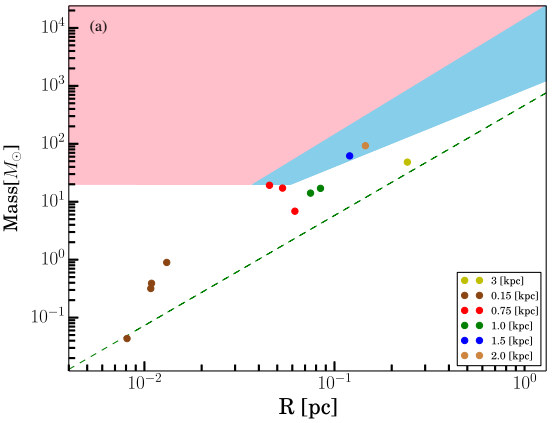

Figure 26. Same as Fig. 23 but for the Lupus IV region.

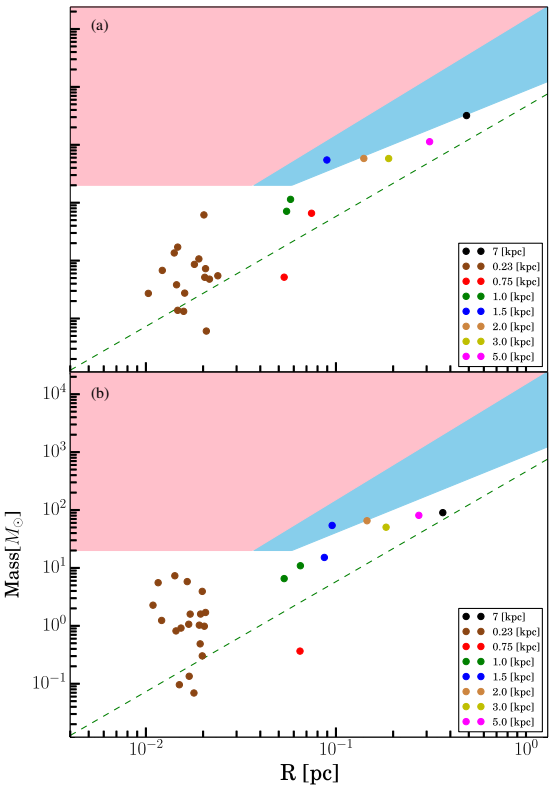

Figure 27. Same as Fig. 23 but for the Serpens region.

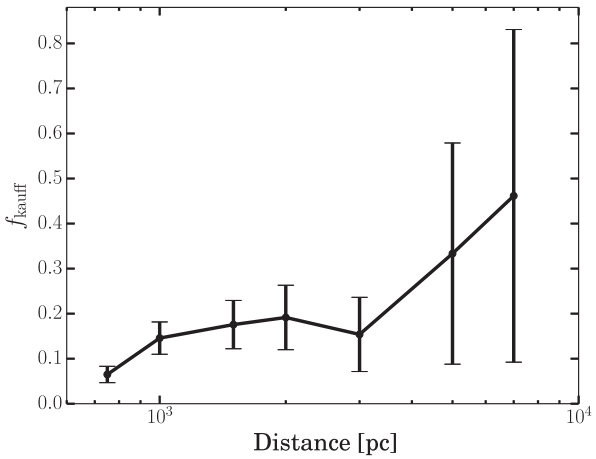

Figure 28. Black thick: fraction of objects $f_{\rm kauff} = n_{\rm FP}/(n_{\rm FP} + n_{\rm TN})$ that are classified as MSF at the moved distances, according to the KP relation, while in reality they are only FP.

to the lack of statistics starting from that distance (indeed the error bars, estimates from the Poisson statistics, are very large). The gap between 750 and 1000 pc is simply due to the fact that many sources found at 750 pc lie at $r < r_{\rm break}$ (see above). From 1000 up to 3000 pc $f_{\rm kauff}$ is of the order of 20 per cent while at larger distance it is larger, but the error bars are very large in turn. Therefore, we can reasonably say that above 1000 pc $f_{\rm kauff}$ slightly increases with distance although the trend is hard to quantify above 3000 pc. We can plot some lines, which are parallel to the KP relation (see Fig. 29), whose analytical form can be expressed as

$$M_{\rm k} > k \left(\frac{r}{[\rm pc]}\right)^{1.33} {\rm M}_\odot, \tag{7}$$

and we can count the number of FPs that lie above these lines. Calling $n_{\rm M_k}$ the number of FP fulfilling equation (7), the relative fraction of FP will be

$$p(k) = n_{M_{\rm k}}/(n_{\rm FP} + n_{\rm TN}). \tag{8}$$

We have to make a distinction between the objects detected at distances lower and larger than 4000 pc, respectively, since as we have seen in Fig. 28 for distance above 4000 pc $f_{\rm kauff}$ tends to be larger. A list of several values of $p(k)$ is reported in Table 2 for the two different cases ($d < 4000$ and $d > 4000$ pc). From these values, we can infer that the fraction of misclassified objects $p(k)$ decreases with increasing k, hence for a clump, classified as MSF, with a high value of k indicates a lower probability of dealing with an FP and the other way around for a lower value of k. The values reported in Table 2 may be very useful to the *Herschel* astronomer. Suppose, for example, we find an Hi-GAL clump having mass M_0 and radius r_0 and within the MSF zone, the coefficient k can be found as

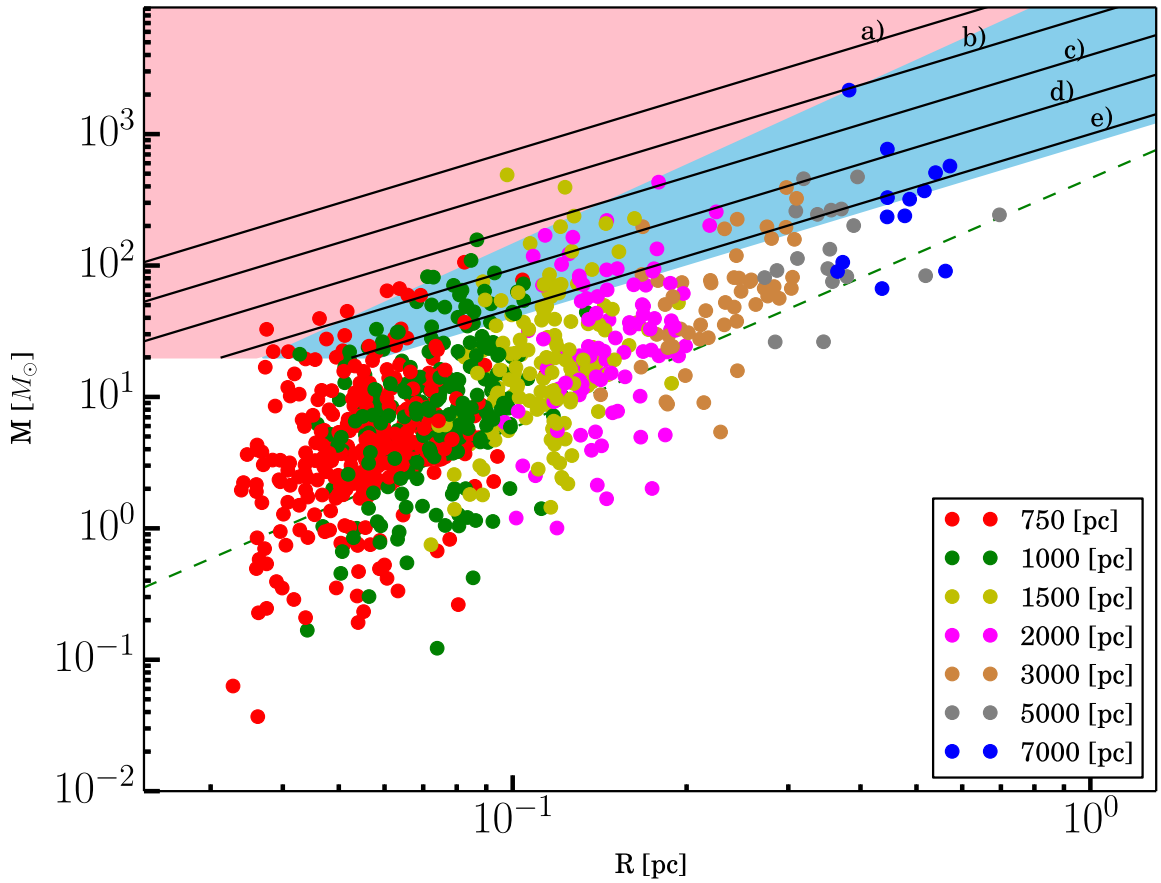

Figure 29. MR plot containing all the sources (solid circles) detected at each virtual distance being associated with sources detected in the original map that are below the MSF zone (FP and TN). Black solid lines are lines parallel to the KP relation $M = k\, r^{1.33}$. The (a) line is for $k = 16000$ and delimits the probability of getting an FP above this line, (b) is for $k = 8000$, (c) to $k = 4000$, (d) to $k = 2000$ and (e) $k = 1000$. The $k = 870$ line coincides with the KP relation. The probability values $p(k)$ (equation 8) are shown in Table 2. Dashed green line is the so-called Larson's third law (see equation 5). Pink and sky-blue zones correspond to thresholds of KM and KP, respectively.

Table 2. Some values of $p(k)$ (equation 8), where $k = M/r^{1.33}$ and $p(k)$ is the fraction of FP fulfilling equation (7). The mass is measured in $M_\odot$ and r in pc. The values of k are chosen such that $p(k)$, in the second column, decreases in steps of 0.01. The values in the third and fifth columns are the error bars associated with the values in the second and fourth columns, respectively. The second and third columns are valid for $d < 4$ kpc, while the fourth and fifth ones are valid for $d > 4$ kpc.

		$d < 4$ kpc	$d > 4$ kpc	
k	$p(k)$	$\sigma_{p(k)}$	$p(k)$	$\sigma_{p(k)}$
870	0.13	0.02	0.39	0.18
922	0.12	0.02	0.35	0.17
1000	0.11	0.01	0.32	0.16
1096	0.10	0.01	0.22	0.12
1122	0.09	0.01	0.22	0.12
1188	0.08	0.01	0.19	0.11
1245	0.07	0.01	0.13	0.09
1332	0.06	0.01	0.13	0.09
1508	0.05	0.01	0.13	0.09
1638	0.04	0.01	0.10	0.07
1968	0.03	0.01	0.10	0.07
2263	0.02	0.01	0.03	0.04

$k = M_0/R_0^{1.33}$. Looking at Table 2 one can get the corresponding value of $p(k)$ and hence estimate how likely it is that the clump is an FP. In Section 5, we provide a practical example of the use of Table 2 for a more correct interpretation of Hi-GAL data.

4.1 A new prescription for high-mass star formation

As pointed out in Section 4, the exponent of the $\langle M\rangle_{\rm d}$ versus $\langle r\rangle_{\rm d}$ power-law relation is smaller than 2.0 (KM) and larger than 1.33 (KP), so that distance effects can bias the fraction of sources fulfilling the two aforementioned criteria and the character itself of star formation in the considered region. A new prescription for identifying compatibility with high-mass star formation is therefore needed. The prescription we provide here is

$$M > 1282 \left(\frac{r}{[\rm pc]}\right)^{1.42} {\rm M}_\odot \qquad (9)$$

(see Appendix B for the derivation of the way it was derived). It is interesting to make a comparison of our formula with that of KP in terms of FP and TP above 1 kpc, since about 99 per cent of the Hi-GAL sources lie in that range of distances. Fig. 30 displays the fraction of FP as a function of distance for both thresholds. The fraction of FP (Fig. 30) generated by the KP prescription is larger than ours at any distance; in particular it is larger by few per cent between 1000 and 3000 pc, while it gets larger by 20–30 per cent at larger distances. The number of TP and FN is comparable for both prescriptions.[2] Note that equation (9) keeps the fraction of FP almost constant, unlike the KP prescription. The main differences between the two are found in the fraction of FP at the largest probed distances, namely 5000 and 7000 pc. Note also that using the KM prescription, one would get a smaller fraction of FP, but would also get a very low number of TP and a higher number of FN (4 and 30, respectively).

Therefore, to minimize the occurrence of FP at all distances, we suggest to use the prescription presented here. In particular, this choice is critical at $d \geq 5000$ pc, where the majority of Hi-GAL sources (73 per cent) provided with a distance estimate are found to lie (Elia et al. 2016).

[2] 24 TP and 10 FN for our prescription and 26 TP and 8 FN for the KP one.

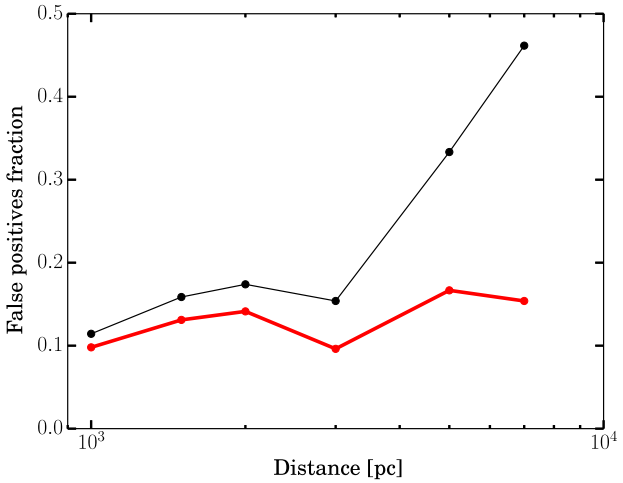

Figure 30. Black line: fraction of objects $f_{\rm kauff} = n_{\rm FP}/(n_{\rm FP} + n_{\rm TN})$ that are classified as MSF at the moved distances, according to the KP relation, whereas they are actually FP. Red line: same as the black one but for our prescription for MSF (equation 9).

5 COMPARISON WITH PREVIOUS LITERATURE

The results obtained in this paper can be used to discuss some results found in the literature on clump properties. For example, Wienen et al. (2015) discuss the MR diagram for the sources detected with the APEX Telescope Large Area Survey of the whole inner Galactic plane at 870 μm (ATLASGAL). These authors found that a relevant fraction (namely 92 per cent) of the clumps are potentially forming massive stars because they fulfil the KP relation. This fraction is notably larger than the one found for the Hi-GAL sources (Elia et al. 2016), namely 71 per cent and 65 per cent for the protostellar and prestellar, respectively. This relevant discrepancy may not be only explained with the Hi-GAL better sensitivity, but most likely it is implicit in the KP relation itself. We have seen, indeed, at large distances the KP relation overestimates the number of high-mass star-forming candidates due to a shallow slope of 1.33. In the MR plot of Wienen et al. (2015, their fig. 23), they found that for large values of the radius (∼1 pc), and hence at large distances, all the sources fulfil the KP relation, while for smaller values of the radius (∼0.1 pc) this fraction is smaller than 1. This effect is likely to be due to the distance bias that we discuss in our paper. Even in the Hi-GAL catalogue, we observe the same effect at large distances in the MR plot, but it is less evident because the sources are more scattered due to the spread in temperature of the Hi-GAL sources. The sources of Wienen et al. (2015) are less scattered in the MR plot, because the masses were derived using only two fixed values of the temperature.

It is noteworthy that masses in ATLASGAL literature are also derived through a different dust opacity; this leads to rescale the value of the coefficient in the KP relation, to 580. Therefore to properly rescale equation (9) to make it comparable with ATLASGAL data, we obtain

$$M_{\rm A} > 855 \left(\frac{r}{[\rm pc]}\right)^{1.42} {\rm M}_\odot. \qquad (10)$$

We apply equation (10) to the data of table 3 of Wienen et al. (2015) to get the fraction of high-mass star-forming candidates.

We find that this fraction is 89 per cent, while if we use the KP relation we get a 94 per cent fraction.[3] Note that Wienen et al. (2015) consider as a high-mass star-forming candidate a clump with a mass larger than 650 $M_\odot$ that fulfils the KP relation. By using such a more demanding prescription they found that 71 per cent of the sources of the catalogue entries are massive star-forming candidates.

We can also make the same test on the data of Ellsworth-Bowers et al. (2015) from the Bolocam Galactic Plane Survey (Ginsburg et al. 2013), again using equation (10), since they used the same dust opacity of the ATLASGAL collaboration. We get 25 per cent of the sources classified as potentially forming massive protostar against 36 per cent found using the KP relation.

Similarly, we can also apply the prescription of Table 2 to the Hi-GAL data. As already mentioned, the most comprehensive catalogue of compact objects, at present, is provided in Elia et al. (2016). In that catalogue we find that, in the inner Galaxy, there are 62 438 objects, with 35 029 of them having a known distance and classified as prestellar and protostellar. We find 23 733 objects that fulfil the KP prescription. This means that 68 per cent of the sources might be able to form a massive star according to the KP prescription. One can apply now the values given in Table 2 to give an estimate of the fraction of FP. Clearly for each source we get a different value of $p(k)$ (equation 8); taking the average of the values of $p(k)$ for all the sources we get $\langle p(k) \rangle \sim 7$ per cent. Note, however, that the values given in Table 2 were derived for $d < 7$ kpc while in the Elia et al. (2016) catalogue 63 per cent of the objects are found to lie at larger distances.

Finally, we apply equation (9) to the Hi-GAL catalogue to discriminate the fraction of high-mass star-forming candidates. We find that according to our prescription, the fraction of such objects is 52 per cent and 64 per cent for the prestellar and protostellar clumps, respectively.

6 CONCLUSIONS

Distance bias increasingly affects estimates of the physical parameters (mass, temperature and radius) of far-infrared sources. This is particularly critical for the Hi-GAL survey, which observed a large area of the sky underlying a wide range of heliocentric distances. In this paper, using the information taken from nearby star-forming regions, we have shown how this bias influences the estimation of these quantities. The main results of our work are as follows.

(i) We present an original pipeline to virtually 'move' the maps at larger distances.

(ii) The number of sources detected with CUTEX decreases with distance, for each band, as a power law with exponent between 1.1 and 1.9. The smallest values for these exponent are found in the 70-μm band.

(iii) The protostellar fraction $n_{\rm pro}/n$ increases with distance until it reaches a plateau (above 1 kpc).

(iv) The effect of the confusion is to increase with distance the physical radius of the compact sources; we show that the sources are classified, on average, as cores up to 1 kpc and as clumps at larger distances.

(v) The contribution of the diffuse (inter-core) emission to the flux of a source increases, with respect to the original population, with distance. This is due to both the increasing physical area of the source, and to the background that gets lower at larger distances. The smallest effect is systematically found at 70 μm.

(vi) We found that the average core formation efficiency, for distances above 1500 pc, depends on the region and can go from few per cent up to 20 per cent.

(vii) The average temperature derived from SED fits increases quite weakly with distance for the prestellar objects, whereas it decreases slowly for the protostellar objects.

(viii) In the MR diagram, the fraction of sources classified as compatible with high-mass star formation increases with distance if one considers the Kauffmann–Pillai (KP) prescription, whereas it decreases for the Krumholz–McKee (KM) prescription. This happens because the mean value of the mass is found to be related to the mean value of the radius with a power law with an exponent larger than 1.33 (KP criterion) and smaller than 2 (KM criterion) for all the investigated regions. Therefore, adopting the KM prescription to check the presence of MSF clumps one tends to lose genuine candidates with distance, while adopting the KP criterion one tends to gain FPs at increasing distances.

(ix) We show that the fraction of FP, defined as the number of FP (clumps that are classified as MSF at a moved distance, according the KP prescription, but do not have associations with MSF objects at the original distance) over the total number sources, is 13 ± 2 per cent, keeping almost constant between 1000 and 4000 pc. Above this distance, the fraction of 'false' high-mass star-forming clumps according to the KP criterion climbs up to almost 40 per cent, but has large associated uncertainties due to small statistics.

(x) We estimated how likely it is that a clump classified as MSF is actually an FP, as a function of its position with respect to the parametrized area $M_{\rm k} > k \left(\frac{r}{[\rm pc]}\right)^{1.33} M_\odot$ in the MR plot. A dichotomy is found for distances shorter and larger of 4000 pc. As an example, a probability of 10 per cent is achieved for $k = 1096$ for $d < 4000$ pc, and at $k = 1969$ for $d > 4000$ pc.

(xi) We derive a new prescription to discriminate possible MSF clumps: $M > 1282.4 \left(\frac{r}{[\rm pc]}\right)^{1.422} M_\odot$ that appears to produce a smaller amount of FPs than the KP relation and preserved the same rates of TP.

(xii) We applied our prescription to discriminate high-mass star-forming candidates in the Hi-GAL data set: we found that the fraction of high-mass star-forming objects is 52 per cent and 64 per cent, for prestellar and protostellar sources, respectively.

(xiii) Taking into account a recently derived distance estimate for the Serpens region of 436 pc (Ortiz-León et al. 2016) instead of the one used throughout this paper ($d_0 = 230$ pc), we notice (see Appendix C) that the MR relation for a few cores changes so that they get classifiable as high-mass star-forming candidates. This might lead, in turn, to classify this region as compatible with episodes of MSF.

ACKNOWLEDGEMENTS

We are grateful to the referee for very useful comments that helped to greatly improve the paper. This research has made use of data from the Herschel Gould Belt survey project (http://gouldbelt-herschel.cea.fr). The HGBS is a Herschel Key Project jointly carried out by SPIRE Specialist Astronomy Group 3 (SAG3), scientists of several institutes in the PACS Consortium (CEA Saclay, INAF-IAPS Rome and INAF-Arcetri, KU Leuven, MPIA Heidelberg), and scientists of the Herschel Science Center (HSC). AB, DE, SM, SP, ES, MB, AD, SL and MM's research activity is supported by the VIALACTEA Project, a Collaborative

[3] Note that source radii quoted by Wienen et al. (2015) are not beam deconvolved. Using, instead, deconvolved radii would increase, in principle, both the fractions reported here.

Project under Framework Programme 7 of the European Union funded under contract #607380, that is hereby acknowledged.

REFERENCES

Alves J., Lombardi M., Lada C. J., 2007, A&A, 462, L17
André P. et al., 2010, A&A, 518, L102
Beckwith S. V. W., Sargent A. I., Chini R. S., Guesten R., 1990, AJ, 99, 924
Beltrán M. T. et al., 2013, A&A, 552, A123
Benedettini M. et al., 2015, MNRAS, 453, 2036
Bergin E. A., Tafalla M., 2007, ARA&A, 45, 339
Comerón F., 2008, in Reipurth B., ed., Handbook of Star Forming Regions, Volume II: The Southern Sky ASP Monograph Publications, Vol. 5. The Lupus Clouds, p. 295
Dunham M. M., Crapsi A., Evans II N. J., Bourke T. L., Huard T. L., Myers P. C., Kauffmann J., 2008, ApJS, 179, 249
Eiroa C., Djupvik A. A., Casali M. M., 2008, Handbook of Star Forming Regions, Volume II: The Serpens Molecular Cloud, p. 693
Elia D., Pezzuto S., 2016, MNRAS, 461, 1328
Elia D. et al., 2010, A&A, 518, L97
Elia D. et al., 2013, ApJ, 772, 45
Elia D. et al., 2016, MNRAS, in press
Ellsworth-Bowers T. P. et al., 2015, ApJ, 805, 157
Genzel R., Stutzki J., 1989, ARA&A, 27, 41
Giannini T. et al., 2012, A&A, 539, A156
Ginsburg A. et al., 2013, ApJS, 208, 14
Griffin M. J. et al., 2010, A&A, 518, L3
Hirota T. et al., 2008, PASJ, 60, 961
Kauffmann J., Pillai T., 2010, ApJ, 723, L7 (KP)
Könyves V. et al., 2015, A&A, 584, A91
Krumholz M. R., McKee C. F., 2008, Nature, 451, 1082 (KM)
Larson R. B., 1981, MNRAS, 194, 809
Menten K. M., Reid M. J., Forbrich J., Brunthaler A., 2007, A&A, 474, 515
Molinari S. et al., 2010, A&A, 518, L100
Molinari S., Schisano E., Faustini F., Pestalozzi M., di Giorgio A. M., Liu S., 2011, A&A, 530, A133
Molinari S. et al., 2014, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 125
Molinari S. et al., 2016, A&A, 591, A149
Motte F. et al., 2010, A&A, 518, L77
Olmi L. et al., 2013, A&A, 551, A111
Ortiz-León G. N. et al., 2016, preprint (arXiv:1610.03128)
Pezzuto S. et al., 2012, A&A, 547, A54
Piazzo L., Calzoletti L., Faustini F., Pestalozzi M., Pezzuto S., Elia D., di Giorgio A., Molinari S., 2015, MNRAS, 447, 1471
Pilbratt G. L. et al., 2010, A&A, 518, L1
Poglitsch A. et al., 2010, A&A, 518, L2
Polychroni D. et al., 2013, ApJ, 777, L33
Ragan S. E., Moore T. J. T., Eden D. J., Hoare M. G., Elia D., Molinari S., 2016, MNRAS, 462, 3123
Russeil D. et al., 2011, A&A, 526, A151
Straižys V., Černis K., Bartašiutė S., 2003, A&A, 405, 585
Tan J. C., Beltrán M. T., Caselli P., Fontani F., Fuente A., Krumholz M. R., McKee C. F., Stolte A., 2014, Protostars and Planets VI. Univ. Arizona Press, Tucson, AZ, p. 149
Veneziani M. et al., 2013, A&A, 549, A130
Wienen M. et al., 2015, A&A, 579, A91

APPENDIX A: PLOT OF THE PERSEUS REGION

In this appendix, we show the MM at 70, 160, 350 and 500 μm for the Perseus region described in Section 2.2 and the maps for the other regions at 250 μm only.

APPENDIX B: DERIVATION OF A NEW HIGH-MASS STAR-FORMING PRESCRIPTION

In this appendix, we show how the prescription, expressed by equation (9), to discriminate high-mass star-forming candidates is derived.

We consider all the objects (both prestellar and protostellar) detected in all the regions moved at $d > 1000$ pc. We define a potentially high-mass star-forming candidate $O_{\rm hm}$ (as defined in Section 3.4) as an object detected at a moved distance that is associated with, at least, a high-mass star-forming core (according to the KP relation) detected at the original distance. Similarly, we define a low-mass star-forming candidate $O_{\rm lm}$ as an object detected at a moved distance that is associated with low-mass star-forming cores detected at the original distance.

We search for a prescription in the form $M > A\,r^a$. For a given (A, a) pair, we define as a true positive (TP) an $O_{\rm hm}$ whose mass is larger than M, a false negative (FN) an $O_{\rm hm}$ whose mass is smaller than M, a true negative (TN) an $O_{\rm lm}$ whose mass is less than M and

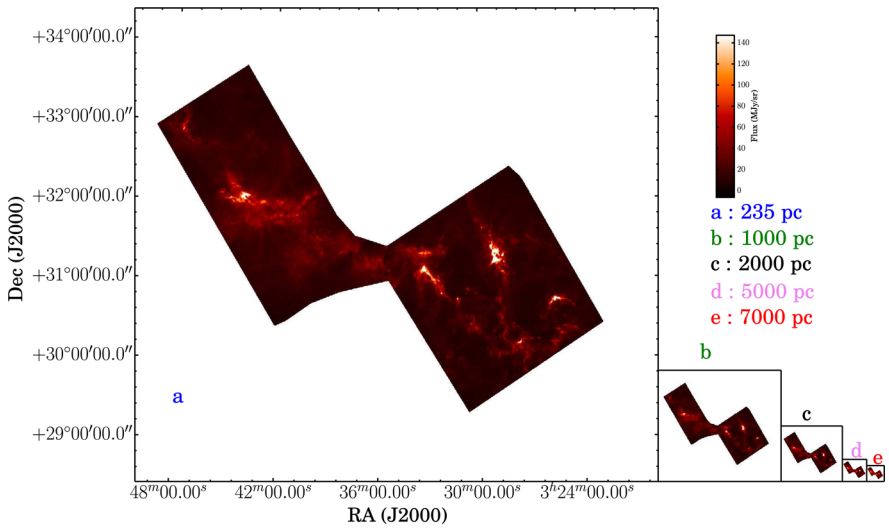

Figure A1. Original and moved Perseus maps at 500 μm.

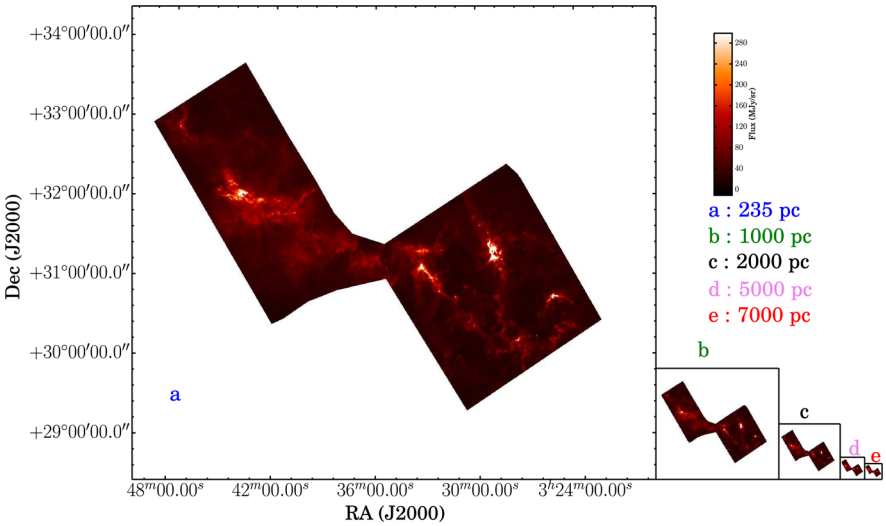

Figure A2. Original and moved Perseus maps at 350 μm.

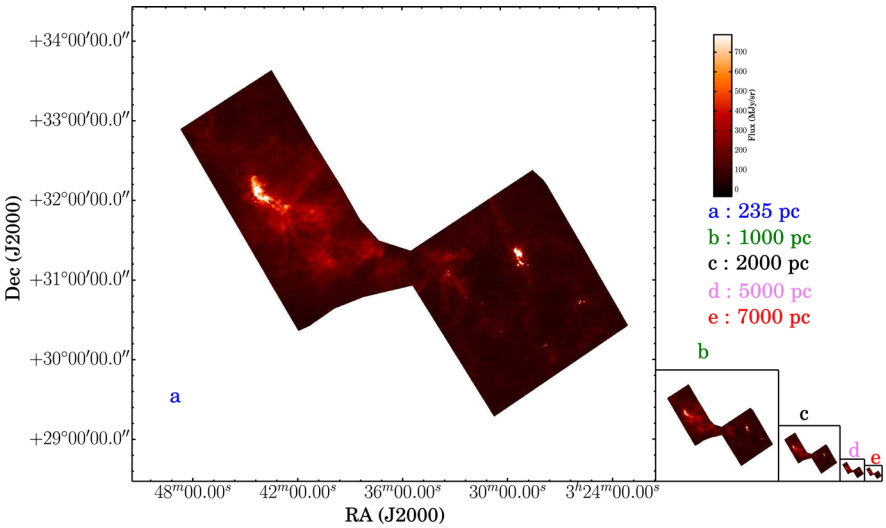

Figure A3. Original and moved Perseus maps at 160 μm.

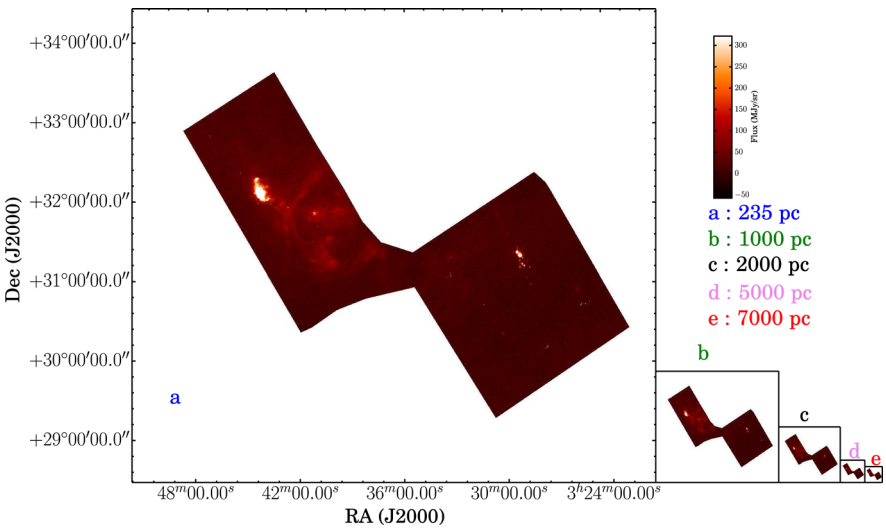

Figure A4. Original and moved Perseus maps at 70 μm.

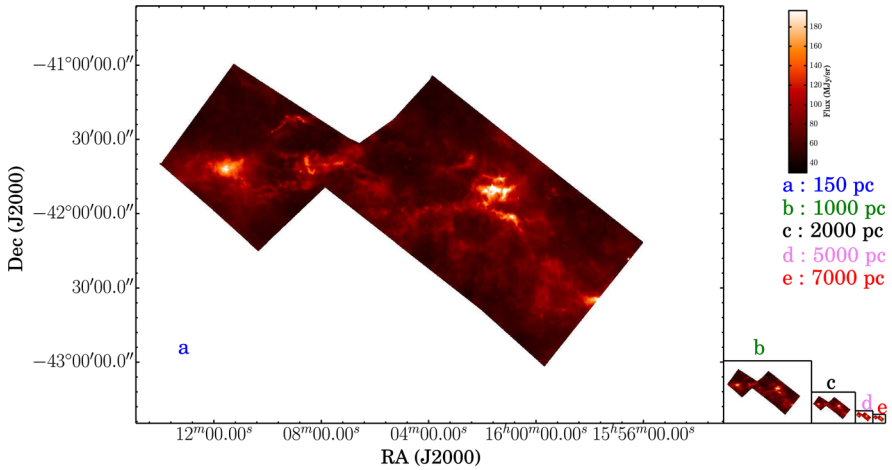

Figure A5. Original and moved Lupus IV maps at 250 μm.

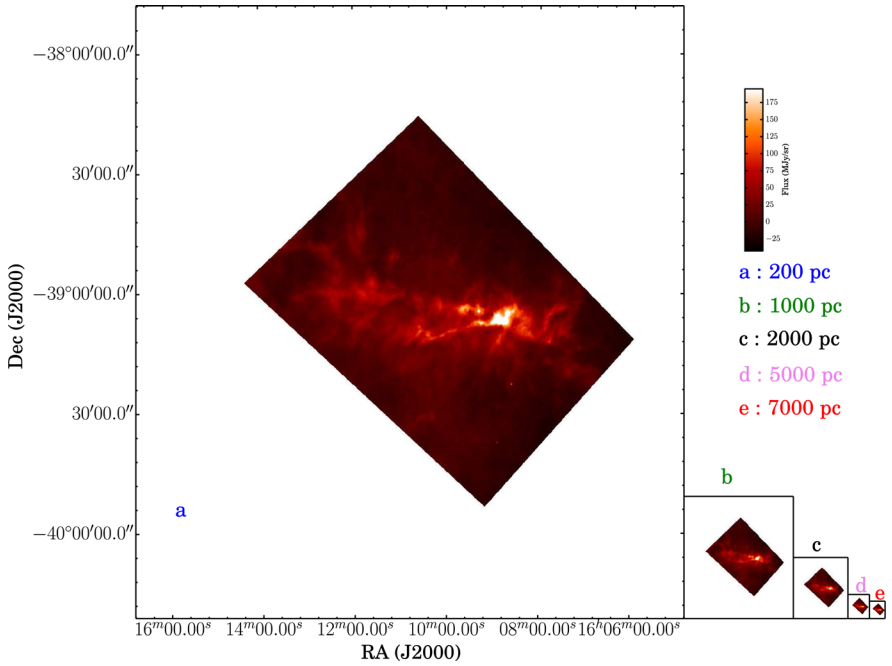

Figure A6. Original and moved Lupus III maps at 250 μm.

a false positive (FP) an $O_{\rm lm}$ whose mass is larger than M. We search (A, a) pairs that maximize the ratio $f = (N_{\rm TP} + N_{\rm TN})/(N_{\rm FP} + N_{\rm FN})$, where $N_{\rm TP}$, $N_{\rm TN}$, $N_{\rm FP}$ and $N_{\rm FN}$ are the numbers of TP, TN, FP and FN, respectively. To do that, we build a grid of values of a and A and then take those maximizing f. The grid was built exploring values of a between 1.33 and 2, in 30 steps, and values of A between 870 and 1800, in 200 steps. Then we impose the further constraint that the fraction of FP should be roughly constant with distance. It turns out that the best values are $A = 1282$ and $a = 1.42$. It is also possible to provide an uncertainty on the value of A: we keep a fixed and we vary A as long as the total fraction of FP changes by 5 per cent. For $A = 1282$ the fraction of FP is ∼9 per cent while ∼4 per cent and ∼14 per cent are achieved for $A = 2000$ and $A = 940$, respectively.

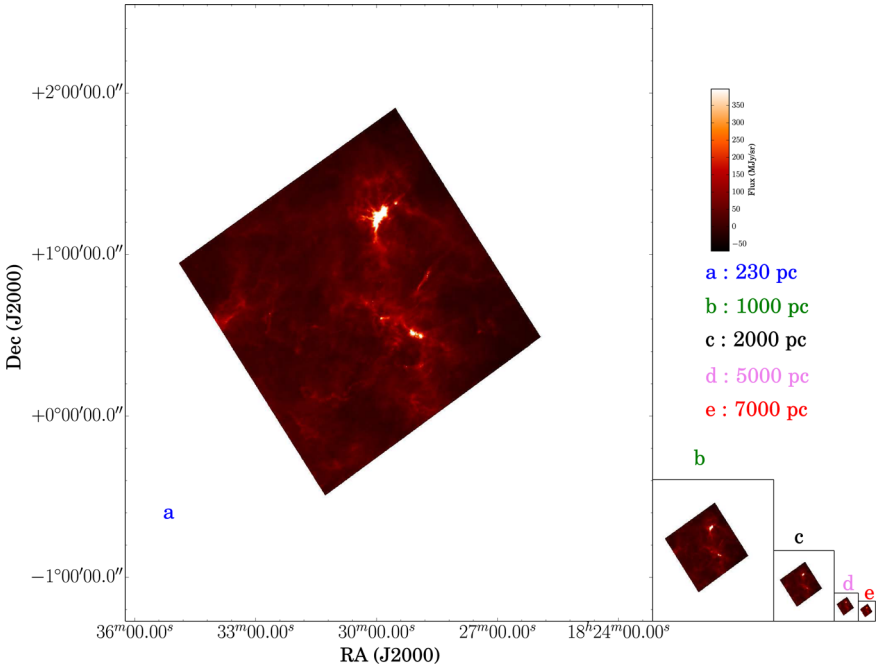

Figure A7. Original and moved Serpens maps at 250 μm.

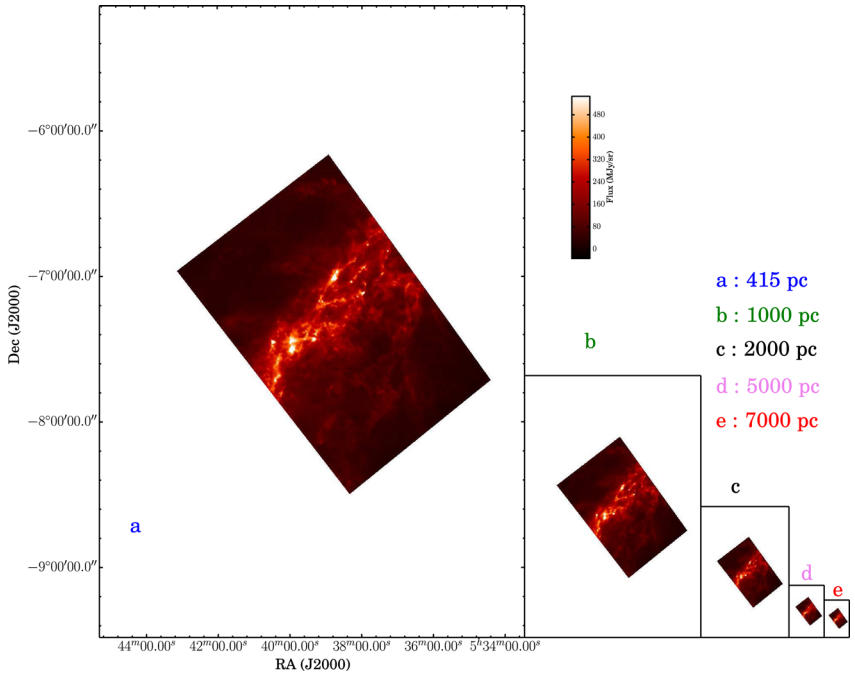

Figure A8. Original and moved Orion A centre maps at 250 μm.

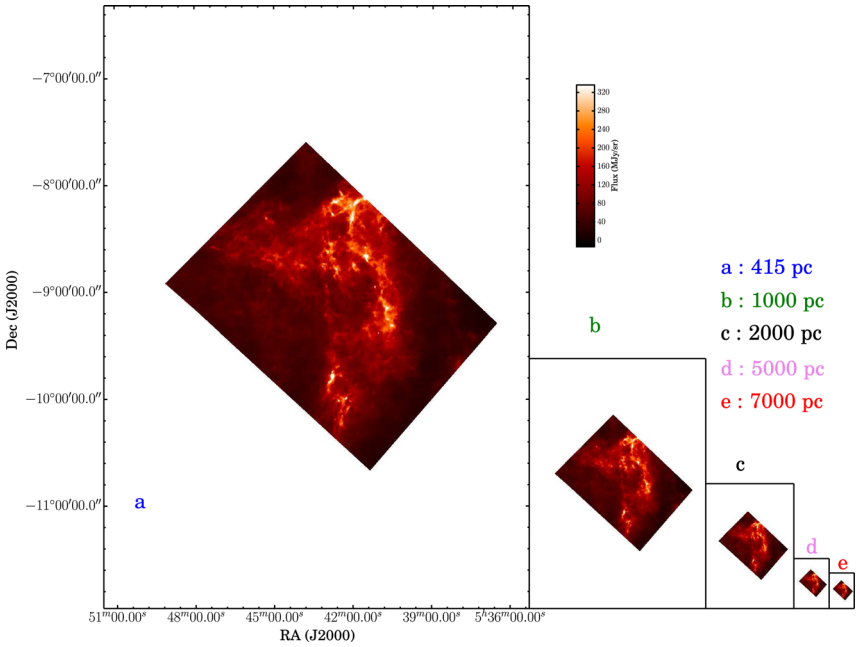

Figure A9. Original and moved Orion A south maps at 250 μm.

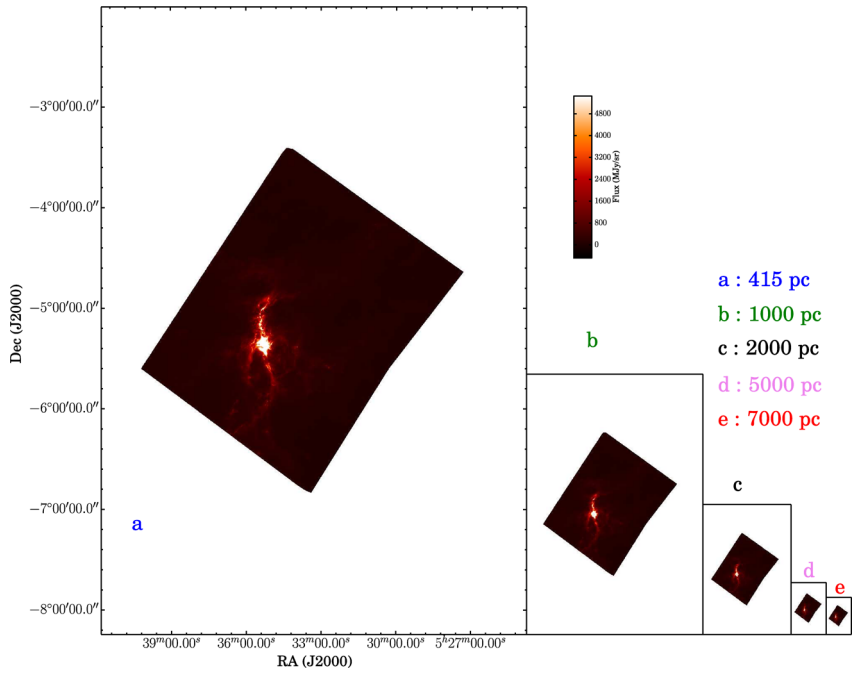

Figure A10. Original and moved Orion A north maps at 250 μm.

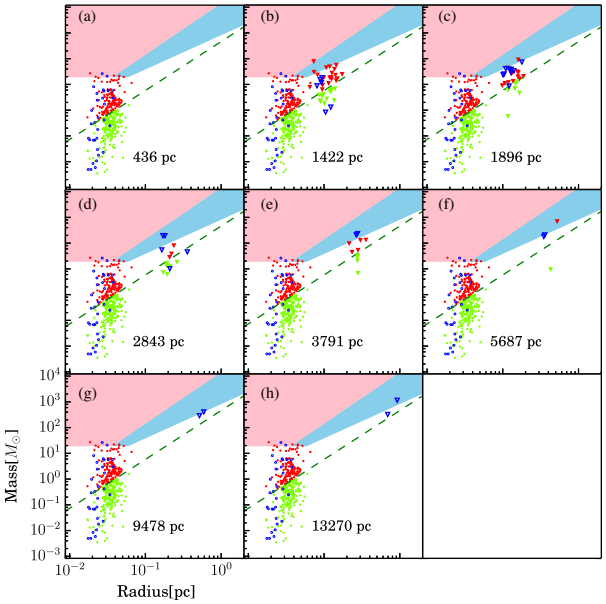

Figure C1. The same as Fig. 20 but starting from $d_0 = 436$ pc instead of 230 pc (simulated distances are scaled accordingly, see text).

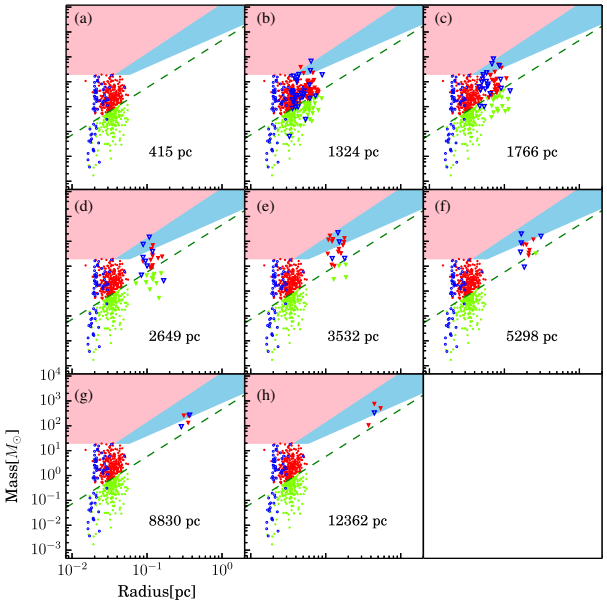

Figure C2. The same as Fig. 17 but imposing $d_0 = 415$ pc instead of 235 pc for Perseus, to make a direct comparison with Orion A (Fig. 16). Simulated distances are scaled accordingly, see text.

APPENDIX C: ANOTHER DETERMINATION OF THE DISTANCE OF THE SERPENS CLOUD

The distance of the Serpens molecular cloud has been a matter of debate. In literature a broad spectrum of distances can be found; for example, Straižys, Černis & Bartašiūtė (2003) and Eiroa et al. (2008) estimate a distance of 225 and 230 pc, respectively, while recently Ortiz-León et al. (2016) place the Serpens cloud at 436 pc.

Here, we want to check how the MR diagram, that we have shown in Section 4, is affected if we adopt a distance of 436 pc instead of 230 pc (Fig. C1). Since the radius and the mass of the sources scale linearly and quadratically with distance, respectively, to get the new values of r and M, we simply multiply these parameters by 436/230 and $(436/230)^2$, respectively. Obviously, also the previously simulated distances must be multiplied by 436/230 leading to new simulated distances of 1422, 1896, 2843, 3791, 5687, 9478 and 13 270 pc.

Note that the Serpens cloud, with this new d_0, might be considered as a potentially MSF region, since in this case at the original distance there are six sources fulfilling the KP relation. Note that also at the moved distances this cloud remains classifiable as an MSF region. Therefore, the behaviour of this cloud in our analysis, if we assume $d_0 = 436$ pc, is very similar to the Orion A molecular cloud, which is a genuine high-mass star-forming region (e.g. Genzel & Stutzki 1989).

It is also interesting to apply the same technique developed in this section to make a further virtual comparative check between the two regions of our sample with the richest source statistics. In practice, we want to understand how the properties of the Perseus cloud would change if it was located at the same distance of Orion A (415 pc) instead of 235 pc. We rescale the mass and the radius with the same procedure that we used above for Serpens data. Fig. C2 displays the MR plot for the Perseus nebula assuming $d_0 = 415$ pc. As one can see from Figs C2 and 6, the radius of the sources at the original distance for Orion A and Perseus would be quite similar, while the masses in Orion A would be still larger than in Perseus. This suggest that, while the size distributions tend to be similar when reported at the same distance (this is somehow trivial since compact sources span the same range of angular sizes), the behaviour of the corresponding masses seems to be more related to the intrinsic nature of the specific region.

This paper has been typeset from a TEX/LATEX file prepared by the author.