Mem. S.A.It. Vol. 86, 589

Memorie della

The interstellar distance toolbox: deriving distances to star forming regions

D.J. Marshall[1], L.A. Montier[2,3], I. Ristorcelli[2,3], M. Juvela[4], I.A. Grenier[1], P. McGehee[5], J. Montillaud[6], R. Paladini[5], V.-M. Pelkonen[4], and Q. Remy[1], on behalf of the Planck collaboration

[1] Laboratoire AIM, IRFU/Service d'Astrophysique - CEA/DSM - CNRS - Université Paris Diderot, Bât. 709, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France
e-mail: douglas.marshall@cea.fr
[2] CNRS, IRAP, 9 Av. colonel Roche, BP 44346, F-31028 Toulouse cedex 4, France
[3] Université de Toulouse, UPS-OMP, IRAP, F-31028 Toulouse cedex 4, France
[4] Department of Physics, P.O.Box 64, FI-00014, University of Helsinki, Finland
[5] Infrared Processing and Analysis Center, California Institute of Technology, 1200, East California Boulevard, Pasadena, CA 91125, USA
[6] Institut UTINAM, CNRS UMR 6213, OSU THETA, Université de Franche-Comté, 41 bis avenue de l'Observatoire, 25000 Besançon, France

Abstract. Fully characterising interstellar clouds requires reliable distance estimates. The study of a large number of clouds is therefore quite challenging, as the different methods do not always agree and furthermore are only able to provide an inhomogeneous sampling of the cloud population. Distances to cold dark clouds are particularly interesting as these constitute the population of pre-stellar cores. The *Planck* Catalogue of Galactic Cold Clumps (PGCC) is an all-sky catalogue of Galactic cold clump candidates detected by *Planck*. The PGCC catalogue is an observational catalogue consisting exclusively of Galactic cold sources, thus potentially sites of future star formation. The three highest *Planck* bands (857, 545, 353 GHz) have been combined with IRAS data at 3 THz to perform a multi-frequency detection of sources colder than their local environment. It contains 13188 Galactic sources spread across the entire sky. Using seven independent methods, reliable distance estimates have been obtained for 5574 sources, which allows us to derive their physical properties such as their mass, physical size, mean density and luminosity. The PGCC sources are located mainly in the solar neighbourhood, up to a distance of 10.5 kpc towards the Galactic centre, and range from low-mass cores to large molecular clouds. Modifications to existing methods along with entirely new ones open up the possibility to re-analyse the catalogue with more robust distance information.

Key words. ISM: dust, extinction – ISM: structure – Galaxy: structure – Galaxy: local interstellar matter – Submillimetre: ISM – ISM: clouds Stars: formation

1. Introduction

The all-sky *Planck* mission has opened up the possibility of carrying out comprehensive investigations of the Galactic emission components. With its high sensitivity and wide wavelength coverage, *Planck* has provided all-sky maps of the thermal dust emission and, in particular, of the emission arising from cold dust. Because cold dust is mainly associated with dense regions within molecular clouds, these observations are relevant for studies of the early phases of star formation, in particular to explore how star formation depends on the initial conditions of the parent molecular cloud. To this end, it is necessary to investigate the spatial distribution and physical properties of dense clumps in different Galactic environments, and this objective can be attained only by extended surveys, which can cover the full range of scales encompassed by the star formation process, i.e. from a subparsec to several kpc.

Distances to the observed objects are crucial to understand the context in which we are observing them, estimate their physical properties (mass, size, density) and thus provide a basis for choosing the most appropriate sources to follow-up. Kinematic distances have often been used to determine distances to objects where a radial velocity can be attributed. By assuming circular rotation for the gas with a simple rotation curve, a distance can be attributed to the source. However the inner Galaxy is dogged by the kinematic distance ambiguity and departures from circular rotation, especially near spiral arms, can not only complicate the analysis but provide misleading distances (Gomez 2006; Reid et al. 2009).

A new category of distance determinations has been developed over the last few years, referred to as extinction distances. Marshall et al. (2009) derived three dimensional extinction distributions towards Infrared Dark Clouds (IRDCs) and inferred their distances, and Stead & Hoare (2010) performed extinction measurement towards Giant Molecular Clouds (GMCs) and compared these to the distance extinction relationships presented in Marshall et al. (2006). Foster el al. (2012) compare these methods along with kinematic distances to reliable maser parallaxes. Although the sample size is small, they conclude that extinction distances are more reliable than kinematic distances. More recently Schlafly et al. (2014) reported distances to local interstellar clouds using the method described in Green et al. (2014) and observations from the PanStarrs survey (Kaiser et al. 2010).

2. Data

2.1. Planck

This PGCC catalogue is based on 15.5 months of observing time of the *Planck* mission, which corresponds to two full-sky surveys. The *Planck* data are combined with the IRIS all-sky data (Miville-Deschenes & Lagache 2005), i.e. a reprocessed version of the *IRAS* data. As described in Planck Collaboration XXII (2013) and Planck Collaboration XXIII (2011), the IRIS 3 THz (100 μm) data have been chosen because they allow us to complement the *Planck* data. In fact:

i) 3 THz is a very good tracer of Galactic warm dust;
ii) the emission from small grains does not contribute substantially at this frequency;
iii) the IRIS and *Planck* data angular resolutions are very similar ($\sim$4.5′).

Note that the *IRAS* survey coverage presents two gaps, which in total accounts for 2% of the whole sky. In IRIS data, these gaps were filled in by using lower angular resolution *DIRBE* data. Due to the discrepancy in resolution between IRIS and *DIRBE*, these regions have been excluded from our analysis. Furthermore, sources detected by our algorithm close to the location of the gaps were carefully examined, since they might be contaminated by noisy features in the IRIS 3 THz map.

2.2. 2MASS

In order to derive distances to clouds we use stellar reddening from the Two Micron All Sky Survey (2MASS). This ground based survey uniformly scanned the entire sky in three near-infrared bands (J, H & K_s). Amongst its final products is the point source catalogue (PSC) which includes point sources brighter

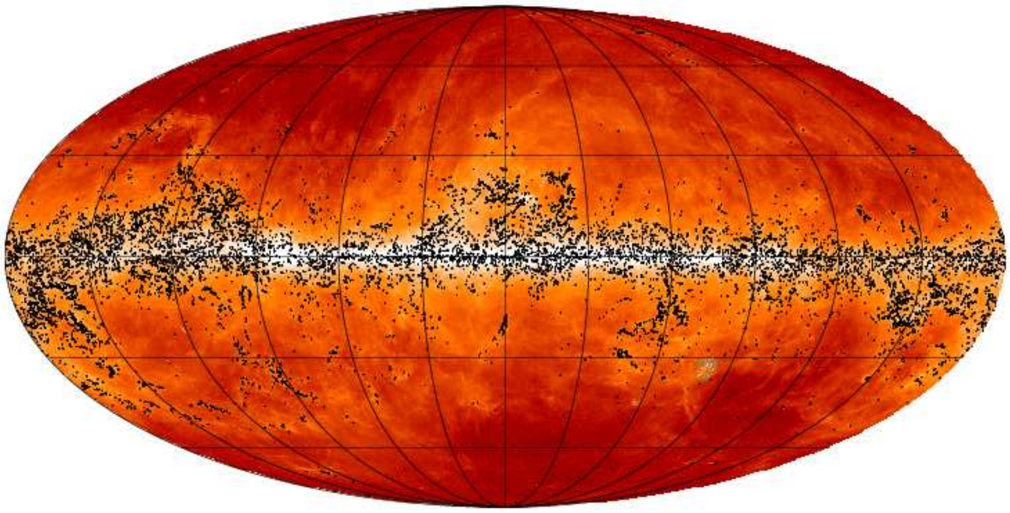

Fig. 1. All-sky distribution of the PGCC cold clumps (black dots). The background map is the 857 GHz *Planck* band.

than about 1 mJy in each band, with signal-to-noise ratio (SNR) greater than 10. Each star has accurate photometry and astrometry as defined in the level 1 requirements (Skrutskie et al. 2006).

3. Cold clump detection method

To detect cold sources in the combined *Planck* and IRIS 3 THz maps, we applied the detection algorithm presented in Planck Collaboration XXIII (2011), and described in Montier et al. (2010). This algorithm is based on a multi-frequency approach which exploits the specific colour properties of this type of source. The detection is performed independently at 857, 545 and 353 GHz using the *cold residual* maps, which are built by subtracting a *warm* component from each frequency map. This *warm* component is estimated separately in each pixel by extrapolating a *warm* template, i.e. the IRIS 3 THz map, to a given *Planck* frequency ν, using the local average background colour estimated at 3 THz and ν, and computed in an annulus from $5'$ to $15'$ centred on each pixel. The catalogues obtained in each of the three *Planck* bands, 857, 545 and 353 GHz, are then merged by requiring:

i) a detection in each band on the *cold residual* maps;
ii) a signal-to-noise ratio (SNR) greater than 4 in all bands;
iii) a maximum distance between the centres of the three detections of $5'$.

These criteria assure cross-band detection consistency as well as source compactness.

We emphasise that our method differs from classical detection algorithms which typically perform the detection directly on frequency maps, as for instance is the case for the *Planck* Catalogue of Compact Sources (PCCS, Planck Collaboration XXVIII 2013). Our method allows a detection in *temperature*: cold sources show a positive signal in the *cold residual* maps, while warm sources show a negative signal. More precisely, this technique allows us to enhance sources having a temperature lower than the local background. This does not automatically imply that the detected sources are intrinsically cold: for example, a source could be detected as *cold* simply because it is seen against a very warm background (or foreground). This is the typical case of objects located along the line of sight towards active star forming regions. In building the catalogue, this effect has been taken into account.

4. Distances to interstellar clouds

In the PGCC we have included distance estimates for a number of the sources using four different methods:

i) cross-checking with kinematic distance estimates already available,
ii) using the optical or near-infrared extinction due to the PGCC sources as an indicator of their distance,
iii) associations with known molecular complexes,
iv) estimates from the literature.

The existing kinematic distances are taken from the IRDC surveys of Simon et al. (2006) and Jackson et al. (2008). In total, 497 sources in the PGCC have an IRDC counterpart with an associated kinematic distance. As for the molecular complexes, a simple inspection of the all-sky distribution of cold clumps (Fig. 1) suggests that it follows the distribution of known molecular complexes at intermediate latitude. We have checked the presence of a given PGCC source inside a molecular cloud, by using a mask generated from the all-sky CO *Planck* map (Planck Collaboration XXII 2013). This method has been applied to 11 molecular complexes lying outside the Galactic plane, which allows us to minimise the effect of confusion. Following this procedure, we have obtained 1895 distance estimates.

The extinction distances are derived using three dimensional extinction determinations, using either optical (Sloan Digital Sky Survey, SDSS) or near infrared (2MASS) observations. Along a line of sight to an interstellar cloud, the extinction is expected to rise sharply at the cloud distance. The method applied to the optical data relies on a star by star analysis of the $g - i$ colour excess. A single cloud is assumed, and the distance where a jump in colour excess exists is retained as the cloud distance.

For the near infrared technique, a Galactic model (Robin et al. 2003, 2012, 2014) is used to create a simulated catalogue of stars towards the cloud and this is compared with the observed stars from 2MASS. Assuming that the Galactic model accurately describes the stellar content of the Milky Way, the differences between model and observations will be solely due to interstellar extinction. The question then is how to determine what is the most probable three dimensional distribution of dust along the line of sight.

Marshall et al. (2009) used a Genetic Algorithm to deduce the line of sight extinction, where the aim was to determine the distance to and mass of IRDCs. The observed stellar colour distribution was compared to the modelled distribution via colour histograms, where a χ^2 test was performed to quantify the similarity of the two histograms. However, the method did not intrinsically supply an estimate of the associated uncertainty and it was susceptible to converge quickly on local maxima. For the PGCC we have used a Markov Chain Monte Carlo (MCMC) technique which is able to return robust uncertainty estimates and is less likely to be trapped by local maxima.

In the final catalogue we have assigned to each source a unique distance value, corresponding to the distance estimate with the highest confidence level among the available estimates. Furthermore, we have attributed a distance quality flag (DIST_QUALITY) which can take one of the following values:

0. "No estimate": no distance estimate is available;
1. "Consistent": multiple estimates exist and are within 1σ
2. "Single": only one estimate exists
3. "Inconsistent": multiple estimates exist and are not within 1σ
4. "Upper limit": only an upper limit is available.

In total 5574 source have a distance estimate, of which 4655 have either DIST_QUALITY = 1 or 2.

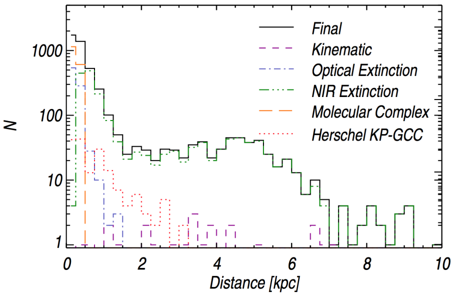

Fig. 2. Distance distribution per method used: kinematic, optical extinction, near-infrared extinction, molecular complex association, and Herschel HKP-GCC.

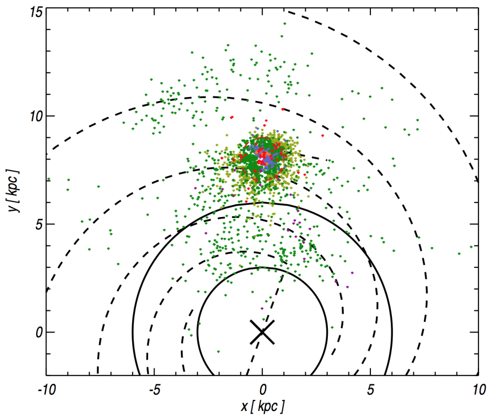

Fig. 3. Distribution of the PGCC sources as seen from the North Galactic pole. Only distance estimates with a DIST_QUALITY flag equal to 1 or 2 are plotted. Different methods have been used to derive the distances: kinematic (purple), optical extinction (blue), near-infrared extinction (green), molecular complex association (orange), and Herschel HKP-GCC (red). We also show the distribution of the distance upper-limits (DIST_QUALITY=4) provided by the near-infrared extinction method (light green). The red dashed circle shows the 1 kpc radius around the Sun. Black dashed lines represent the spiral arms (pitch angle of 28° following Vallée 2008) and the local bar. The black circles, centred on the cross, provide an indication of the location of the molecular ring.

5. Characteristics of PGCC clouds using derived distances

For the Planck Galactic cold clumps we have derived: temperature, column density and, when distance estimates are available, size, mass, mean density and luminosity.

5.1. Galactic distribution

Referring to Fig. 2, it can be seen that the near-infrared extinction and kinematic methods allows us to probe distant regions (from 1 kpc to 9 kpc) across the Galactic plane. On the contrary, the optical extinction and molecular complex associations methods are applicable in the nearby Galaxy only (up to 1 kpc and 0.5 kpc, respectively).

As seen from the North Galactic pole (Fig. 3), the complementarity of the different methods shows clearly. About 88% of the sources with a reliable distance estimate (DIST_QUALITY=1 or 2) lie within 2 kpc from the Sun. Therefore, the PGCC catalogue mainly probes the solar vicinity. It is interesting to notice that the distribution of the PGCC sources at larger distance follows at first order the Galactic arms and the molecular ring. This is especially significant for the Perseus arm towards the outer Galaxy and for the Scutum-Centaurus and Norma arms in the inner Galaxy. We conclude this section by emphasising that, due to the variety of distance estimators, any statistical analysis involving distances or the related quantities, will be affected by severe biases which, given the fact that the catalogue is not flux density complete, are very hard to quantify.

5.2. Cloud properties

The source temperatures and the local warm background temperatures have been estimated using the flux density measurements and their corresponding uncertainties in the IRIS 3 THz band and the Planck 857, 545, and 353 GHz channels. The fits of the spectral energy distributions (SED)s have been performed assuming that the emission can be described as a modified black body:

$$F_\nu = F_{\nu,0} B_\nu(T)(\nu/\nu_0)^\beta \quad (1)$$

where F_ν is the observed flux density, β is the emissivity spectral index (either fixed at $\beta = 2$ or allowed to vary), T is the fitted colour temperature, and $F_{\nu,0}$ is the fitted flux density at a reference wavelength ν_0. In Eq. 1 we have assumed that the observed emission is optically thin at frequencies $\nu \le 3$ THz, that the emissivity spectral index is constant within the fitted wavelength range, and that the source is isothermal.

The temperature estimates of the sources correspond to the temperature of the clump only after removal of the warm background, while the colour temperature is usually associated with the total emission on the line-of-sight. Sources where reliable flux densities were obtained in all four bands are given a Flux Quality (FQ) of 1. Where no IRIS flux is available, they are given FQ=2. This latter group of sources are typically characterised by low flux densities and extremely cold temperatures, and have no detectable counterparts in the infrared. They are potentially interesting very cold clump candidates.

The resulting clump temperature distribution is shown in Fig. 4 for the two categories of sources, with FQ=1 and 2. When the emissivity spectral index β is allowed to vary (solid lines), the temperature of the Planck cold clump candidates with FQ=1 ranges from 8.6 to 30 K, and actually from 10.5 to 19.9 K when excluding the 2% extreme percentiles, with a peak of the distribution at about 14.5K. Likewise, when we take $\beta = 2$, the distribution is narrower (ranging from 6 to 22.5K, and from 11.1 to 16.8 K when excluding the 2% extreme percentiles) but still peaks around 14 K.

For sources with a reliable distance estimate, the mass of the clump is given by:

$$M = \frac{S_\nu D^2}{\kappa^2 B_\nu(T)} \tag{2}$$

where S_ν is the flux density measured at 857GHz, integrated over the solid angle Ω, D is the distance, κ_ν is the dust opacity and $B_\nu(T)$ is the Planck function for a dust temperature T.

The mass distribution shown in Fig. 4, ranges from a few 10^{-2} to almost $10^5 M_\odot$, probing a large variety of objects, from cores to giant molecular clouds. This mass distribution is

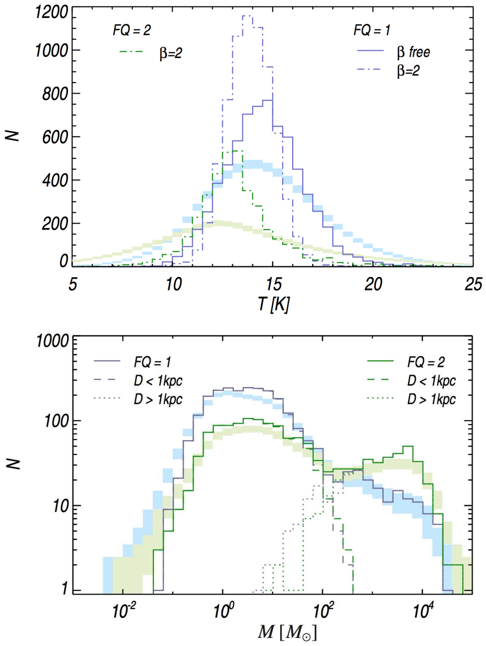

Fig. 4. Distribution of the physical properties of the PGCC objects with FQ=1 (blue) and FQ=2 (green). The coloured shaded regions provide an estimate of the impact of the individual uncertainties on the statistical distribution at a 1σ dispersion level around the mean value. The temperature distribution (**TOP**) is shown for two temperature estimates, i.e., using a free emissivity spectral index β (solid line) or a fixed $\beta = 2$ (dot-dashed line). The distribution of the physical mass of the clumps (**BOTTOM**) is provided for sources with reliable distance estimates (DIST_QUALITY=1 or 2). These distributions are also shown separately for sources at solar distances smaller or larger than 1 kpc from the Sun, in dashed and dotted lines, respectively.

biased by our distance sample, which we know is highly heterogeneous. Except for less than 40% of the cases where the distance estimates are obtained from molecular complex association providing highly reliable estimates, the uncertainty on the computed mass is mainly dominated by the uncertainty on the distance, which is about two to three times larger than the uncertainties on the temperature and flux density involved in the calculation Note, however, that we have almost reached the theoretical sensitivity limit of Planck to low-mass cold cores, which is about 0.03 $M_\odot$ for a cold source located at 100 pc and with a column density of 10^{20} cm^{-2}

6. Future improvements using Bayesian methods

Several authors have recently proposed Bayesian methods that infer the three dimensional distribution of dust. Green et al. (2014) proposed a Bayesian approach to derive a 3D map of extinction based on the Pan-STARRS 1 (PS1) survey. They derive the absolute magnitude of stars using their colours and metallicity, following the method described by Ivezic et al. (2008). They also assume priors on the stellar populations on the basis of the model of Juric et al. (2008). Schlafly et al. (2014) use this method to stars that lie along lines of sight towards local molecular clouds. They fit the stellar observations using a simple dust screen model to find the distance to each cloud. Typical uncertainties are 5% (stat.) and 10%(sys). As the method presently uses optical observations, the cloud distances are limited to off plane clouds, although the method itself does not set any such limitation.

The distances to clouds at high latitude in the PGCC (in our case this would translate to $|b| > 10°$) were not satisfactorily recovered using the near infrared extinction technique. The lower stellar density, especially of foreground stars, is the main limiting factor for these lines of sight. In order to improve the distance estimates above the plane, we have modified the algorithm in a number of ways compared to Marshall et al. (2009) and Planck Collaboration XXVIII (2015).

We have implemented our own Bayesian method based on near infrared observations and the Galactic model. We use an MCMC method which explores the parameter space and returns an estimate of the probability distribution function for each parameter. In this study, the parameter space has been reduced by describing the problem differently. The problem is described using only four parameters:

Diffuse extinction disc The diffuse extinction is assumed to come from a smooth double exponential (Robin et al. 2003). A single parameter controls the solar neighbourhood normalisation

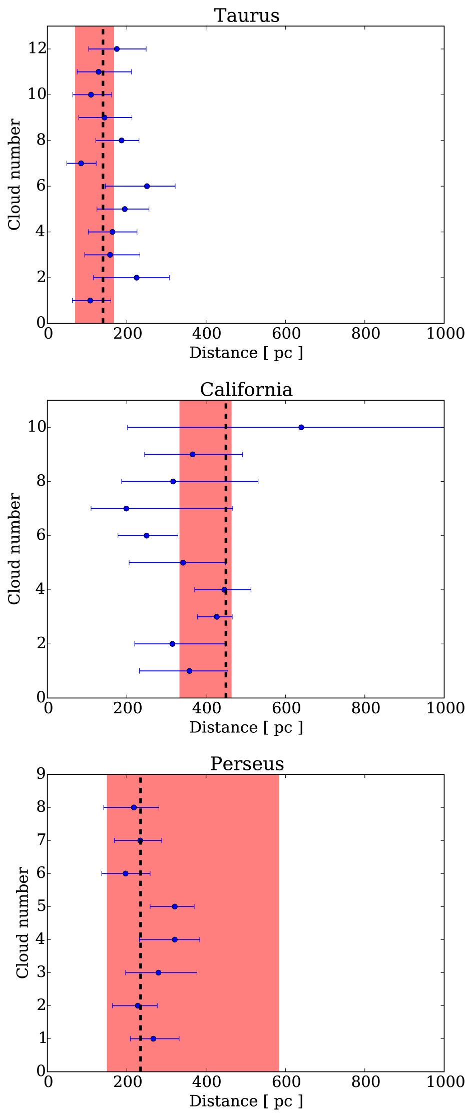

Fig. 5. Comparison of distance estimates to clouds in three molecular complexes: Taurus (Top), California (Middle) and Perseus (Bottom). Each point with associated error bars is an estimate from our new Bayesian method, whereas the dashed line and the shaded zone shows the literature distance value and the range of distances reported in Schlafly et al. (2014).

Distance to cloud Only one cloud is assumed to lie along the line of sight. Its distance is constrained so that the cloud altitude lies below 2 scale heights of the dust disc, or 250 pc (Marshall et al. 2006)

Extinction to the cloud The total extinction along the line of sight is estimated from the 95th percentile of observed stellar colour. The cloud's extinction can take any value up to this value.

Dispersion of extinction Any solid angle towards an interstellar cloud will contain unresolved structure. Here we assume that the distribution follows a lognormal and use one parameter to describe its dispersion.

This description is obviously only valid for lines of sight with only one cloud, although a diffuse component can also be present. More clouds can in theory be added, but care must be taken to avoid degeneracies between them. We have therefore limited our method to one cloud for the present study.

Furthermore, the likelihood of the model parameters is estimated using the full colour magnitude diagram (CMD). The comparison between model and observations is performed by binning the CMD in two dimensions and using a binned likelihood analysis assuming that the number of stars in each bin follows a Poisson distribution.

This new method has been applied to a number of molecular clouds in the Taurus, California and Perseus cloud complexes. Schlafly et al. (2014) have recently published extinction distances to a number of lines of sight in these and other regions. In Fig. 5 we compare our estimates to theirs. As we have not targeted exactly the same lines of sight as them, we report their results as the full range of distances they found in each region, including the associated uncertainties. Our estimates are in very good agreement with theirs. Compared to the previous near infrared extinction technique, which could usually only provide an upper limit distance to clouds within 1 kpc, the current one provides much more precise measurement of distance to local clouds.

7. Conclusions

The highest frequency bands of the Planck-HFI instrument provide an extremely powerful tracer of Galactic cold dust. By combining data from the three highest frequency bands from *Planck* with IRAS 3THz data, we have conducted a multi-frequency compact source detection, and generated the *Planck* Catalogue of Galactic Cold Clumps (PGCC). A first version of this catalogue was released in 2011, i.e., the Early Cold Core catalogue (ECC), together with the Early Release Compact Source Catalogue (ERCSC). At that time, 915 sources, selected for their low temperature and high S/N, were made publicly available to the astronomical community. With the present work, we are releasing the whole PGCC catalogue, containing 13188 Galactic sources and 54 cold sources located in the LMC and SMC.

We have applied the CoCoCoDeT algorithm (Montier et al. 2010) to the *Planck* 857, 545, and 353 GHz maps and to the IRAS 3 THz data. The combined use of these maps allow the separation of the cold and warm emission components, hence the identification of sources colder than their local environment.

We have combined seven independent methods to assign a distance estimate to 5574 sources. In the PGCC catalogue, for each source we quote all the available distances derived from the different methods, however, only the most reliable estimate is used to compute other source physical quantities such as mass and luminosity. Distance estimates from different methods have been compared and validated. Accordingly, we have assigned, to each clump, a DIST_QUALITY flag: 464 sources have consistent distance estimates (DIST_QUALITY=1); 4191 sources have only one estimate (DIST_QUALITY=2); 255 have incompatible estimates (DIST_QUALITY=3); 664 sources have only distance upper limits (DIST_QUALITY=4). More details on this analysis can be found in the on-line version of the PGCC catalogue. The 4655 sources with an accurate distance estimate (DIST_QUALITY=1 or 2) are mainly located in the solar neighbourhood, with about 85% of sources at less than 1 kpc, and 91.4% within 3 kpc from the Sun.

Furthermore, we have shown that newer Bayesian methods are showing great promise in their ability to provide robust distance estimates, even to nearby clouds that have been problematic in the past. This is a necessary step in building a homogeneous distance sample with which a meaningful analysis of the bulk cloud properties can be done.

Finally, we believe that the PGCC catalogue, covering the whole sky, hence probing wildly different environments, represents a real goldmine for investigations of the early phases of star formation. These include, but are not limited to:

i) studies of the evolution from molecular clouds to cores and the influence of the local conditions;
ii) analysis of the extreme cold sources, such as the most massive clumps or those located at relatively high latitude;
iii) characterisation of the dust emission law in dense regions and the role of the environment.

Acknowledgements. The text and figures describing the *Planck* cold clump catalogue has been reproduced with permission from Astronomy & Astrophysics, ©ESO.

References

Green, G.M., et al. 2014, ApJ, 783, 114
Gomez, G.C. 2006, AJ, 132, 2376
Ivezic, Z., Sesar, B., Juric, M., et al. 2008, ApJ, 684, 287
Jackson, J.M., Finn, S.C., Rathborne, J.M., et al. 2008, ApJ, 680, 349
Juric, M., Ivezic, Z., Brooks, A., et al. 2008, ApJ, 673, 864
Juvela, M., et al. 2010, A&A, 518, L93
Juvela, M., et al. 2011, A&A, 527, A111
Juvela, M., et al. 2012, A&A, 541, A12
Juvela, M., Ristorcelli, I., Marshall, D. J., et al. 2015, A&A, 584, A93
Foster, J. B., Stead, J.J., Benjamin, R.A., et al., 2012, ApJ, 751, 157
Kaiser, N., Burgett, W., Chambers, K., et al. 2010, Proc. SPIE, 7733, 7733E
Marshall, D.J., et al., 2006, A&A, 453, 635
Marshall, D.J. et al., 2009, ApJ, 706, 727
Miville-Deschenes, M. & Lagache, G. 2005, ApJS, 157, 302
Montier, L.A., et al. 2010, A&A, 522, A83
Planck Collaboration 2011, A&A, 536, A22
Planck Collaboration 2011, A&A, 536, A23
Planck Collaboration 2013, A&A, 571, A8
Planck Collaboration 2014, A&A, 571, A13
Planck Collaboration 2014, A&A, 571, A28
Planck Collaboration 2015, arXiv:1502.01599
Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009, ApJ, 700, 137
Robin, A.C., et al. 2003, A&A, 409, 523
Robin, A.C., et al. 2012. A&A538, A106
Robin, A.C., Reylé, C., Fliri, J., et al. 2014, A&A, 569, A13
Schlafly, E.F. et al. 2014, ApJ, 786, 29
Simon, R., Rathborne, J.M., Shah, R., et al. 2006, ApJ, 653, 1325
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Stead, J. J. & Hoare, M. G. 2010, MNRAS, 407, 923
Valle, J.P. 2008, AJ, 135, 1301