The Milky Way Project First Data Release: a bubblier Galactic disc

R. J. Simpson, M. S. Povich, S. Kendrew, C. J. Lintott, E. Bressert, K. Arvidsson, C. Cyganowski, S. Maddison, K. Schawinski, R. Sherman, A. M. Smith and G. Wolf-Chase

1 Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH
2 Department of Astronomy and Astrophysics, The Pennsylvania State University, 525 Davey Laboratory, University Park, PA 16802, USA
3 Max Planck Institut für Astronomie, Königstuhl 17, 69117 Heidelberg, Germany
4 Astronomy Department, Adler Planetarium, 1300 S. Lake Shore Drive, Chicago, IL 60605, USA
5 ESO, Karl-Schwarzschild-Strasse 2, D85748, Garching, Germany
6 School of Physics, University of Exeter, Stocker Road, Exeter EX4 4QL
7 Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
8 Centre for Astrophysics & Supercomputing, Swinburne University, H39, PO Box 218, Hawthorn, VIC 3122, Australia
9 Department of Physics, Yale University, PO Box 208121, New Haven, CT 06520, USA
10 Yale Center for Astronomy and Astrophysics, Yale University, PO Box 208121, New Haven, CT 06520, USA
11 Department of Astronomy & Astrophysics, University of Chicago, 5640 S. Ellis Ave., Chicago, IL 60637, USA

ABSTRACT

We present a new catalogue of 5106 infrared bubbles created through visual classification via the online citizen science website ‘The Milky Way Project’. Bubbles in the new catalogue have been independently measured by at least five individuals, producing consensus parameters for their position, radius, thickness, eccentricity and position angle. Citizen scientists – volunteers recruited online and taking part in this research – have independently rediscovered the locations of at least 86 per cent of three widely used catalogues of bubbles and H II regions whilst finding an order of magnitude more objects. 29 per cent of the Milky Way Project catalogue bubbles lie on the rim of a larger bubble, or have smaller bubbles located within them, opening up the possibility of better statistical studies of triggered star formation. Also outlined is the creation of a ‘heat map’ of star formation activity in the Galactic plane. This online resource provides a crowd-sourced map of bubbles and arcs in the Milky Way, and will enable better statistical analysis of Galactic star formation sites.

Key words: stars: formation – dust, extinction – H II regions – infrared: ISM.

1 INTRODUCTION

H II regions ionized by young O- and B-type stars provide the most readily observable tracers of star formation in the Milky Way and other galaxies. Ionized gas produces strong emission in optical and infrared (IR) recombination lines, forbidden lines and thermal (free–free) radio continuum. Dust mixed with ionized gas and heated by the hard radiation field makes H II regions bright sources of thermal IR emission.

In the Milky Way, the spatial morphology of individual H II regions can generally be resolved, revealing complex structures which are often shaped by the newly formed stars (Anderson et al. 2011). One particularly common and informative morphology is the presence of a rounded ‘bubble’ of emission from gas excited presumably by a central source.

Churchwell et al. (2006, 2007, hereafter CP06, CWP07) catalogued nearly 600 IR ‘bubbles’ in the inner 130° of the Galactic plane by visually searching for ring-shaped structures (complete or partially broken) in 3.6–8.0 μm images from the Galactic Legacy Infrared Mid-Plane Survey Extraordinaire (GLIMPSE; Benjamin et al. 2003; Churchwell et al. 2006). In the GLIMPSE-I survey area, CP06 catalogued 322 bubbles, of which 25 per cent corresponded to previously catalogued radio H II regions and 75 per cent were attributed to late B-type stars with insufficient ionizing luminosity to produce radio-bright H II regions. Bania et al. (2010) later selected 24 μm diffuse emission sources from SpitzerMultiband Imaging Photometer for Spitzer Inner Galactic Plane (MIPSGAL) survey images (Carey et al. 2009) with spatially coincident 21 cm continuum emission for a radio recombination line survey with the Green Bank Telescope, and discovered several hundred new Galactic H II regions in the region −16° ≤ l ≤ 67° and |b| ≤ 1°, including those associated with 65 of the CP06 bubbles.
GLIMPSE images are particularly helpful in identifying bubbles as emission from polycyclic aromatic hydrocarbons (PAHs) at 8 μm demarcates the bubble rims in Spitzer Space Telescope images, while a second peak, arc or torus of 24 μm emission from warm dust is frequently observed inside the bubble with a morphology that closely traces the radio continuum emission (CP06; Watson et al. 2008; Watson, Hanspal & Mengistu 2010). This qualitative pattern of bright 8 μm emission shells surrounding regions of bright 24 μm/radio emission is also characteristic of giant Galactic H II regions and extragalactic star-forming regions (e.g. Povich et al. 2007; Bendo et al. 2008; Relaño & Kennicutt 2009; Flagey et al. 2011).

Nevertheless, existing surveys are not sufficient to establish the true relationship between bubbles and H II regions. Both CP06 and CWP07 stressed that their bubble catalogues were far from complete, particularly with regard to bubbles with large (> 10 arcmin) and small (<2 arcmin) angular diameters, and the majority of previously catalogued, bright radio H II regions (Paladin et al. 2003) in the GLIMPSE survey area were not associated with catalogued bubbles. None of the large bubbles identified by Rahman & Murray (2010) were included in the CP06 and CWP07 catalogues, nor was the large bubble associated with the well-studied giant H II region M17 identified by Povich et al. (2009). These large bubbles were missing from the catalogues because they are generally faint, broken and confused with smaller, brighter H II regions.

In this paper, we present a new catalogue of 5106 IR bubbles identified via visual inspection of the GLIMPSE and MIPSGAL survey images as part of the citizen science Milky Way Project1 (MWP). Our catalogue expands the CP06 and CPW07 catalogues by nearly a factor of ten, and represents a far more complete sample of Galactic H II regions. Three key advances in the bubble selection process enabled the identification of bubbles that were missed by CP06 or CWP07. (1) We enlisted >35,000 volunteers to examine the images rather than relying on only a handful of experts; (2) we incorporated MIPSGAL 24 μm data, which greatly facilitates the identification of bubbles compared to relying primarily on 8.0 μm data; and (3) we relaxed the criteria of what defines a ‘bubble’ to include more incomplete shells and arc-shaped structures. The remainder of this paper is organized as follows: in Sections 2 and 3 we describe the MWP and the construction of our catalogue. Some initial results from the catalogue are presented in Section 4 and discussed in Section 5. We summarize our conclusions and directions for future work in Section 6.

2 THE MILKY WAY PROJECT

Galaxy Zoo (Lintott et al. 2008, 2011) and the larger suite of Zooniverse2 projects (Smith et al. 2011) have successfully built a large community of volunteers3 eager to participate in scientific activities.

The Zooniverse has shown that enlisting ‘citizen scientists’ via the internet is a powerful way to analyse large amounts of data. Human brains excel at pattern recognition tasks, and most people will reach a level of accuracy as high as any expert after a brief introduction. By enlisting citizen scientists, researchers can extend visual classification to large samples of images, having each image examined by a large number of independent classifiers. This allows researchers to tap into the ‘wisdom of the crowd’ effect where the consensus of a group of non-experts is often more accurate than the testimony of a single expert.

Another advantage of enlisting human classifiers is their ability to recognize unusual objects which computer search algorithms may be unable to spot. This has been shown by the serendipitous discovery of Hanny’s Voorwerp (Lintott et al. 2009) and the case of the Galaxy Zoo ‘Green Peas’ (Cardamone et al. 2009). Spitzer GLIMPSE data are ideally suited to classification by citizen scientists as the amount of data is large and the images contain complex, overlapping structures that are impossible to disentangle using automated algorithms. The task of recognizing bubbles may eventually be handled by advanced machine-learning algorithms (e.g. Beaumont, Williams & Goodman 2011), but in the meantime the community of Zooniverse users are keen to contribute to astronomy and science (Raddick et al. 2010).

The MWP is the ninth online citizen science project created using the Zooniverse Application Programming Interface (API) toolset. The Zooniverse API is the core software supporting the activities of all Zooniverse citizen science projects. Built originally for Galaxy Zoo 2, the software is now being used by 11 different projects. The Zooniverse API is designed primarily as a tool for serving up a large collection of ‘assets’ (for example, images or video) to an interface, and collecting back user-generated interactions with these assets.

The assets in the MWP are multiband, false-colour JPEG images, created by gridding the Spitzer GLIMPSE and MIPSGAL mosaics into smaller images at three different zoom levels. The highest zoom level provides users with tiles of 0.3 × 0.15, and at a resolution of 800 × 400 pixels these tiles nearly reproduce the 1.2 arcsec pixel scale of the GLIMPSE survey images. Larger tile sizes of 0.75 × 0.375 and 1.5 × 0.75 were also generated. The tiles were plotted in an overlapping grid to allow all parts of the inner Galactic plane (l ≤ 65°, b ≤ 1°) to be viewed by the MWP users, at all zoom levels. To provide an optimal representation of the dynamic range within each tile, each of three single-band images was independently scaled to a square-root stretch function (with the faintest 5 per cent of image pixels clipped to black and the brightest 0.2 per cent clipped to white), assigned to a colour channel (red = 24 μm, green = 8.0 μm, blue = 4.5 μm), and finally composited into a three-colour image. The MIPSGAL 24 μm mosaics frequently saturate in regions of bright nebulosity, and saturated 24 μm pixels were set to maximum red to preserve the visual appeal of the images and to avoid presenting MWP users with saturation artefacts. The resulting composite images allow visual identification of both bright and faint features within a given image tile.

The MWP user interface (see Fig. 1) was built using Flash, based upon the pre-existing Moon Zoo interface (Joy et al. 2011).

1 http://www.milkywayproject.org
2 http://www.zooniverse.org
3 Over 480,000 registered volunteers at the time of writing.

© 2012 The Authors, MNRAS 424, 2442–2460
Monthly Notices of the Royal Astronomical Society © 2012 RAS

Figure 1. Screenshot of the MWP user interface. Colour figure available online.
Volunteers are primarily encouraged to draw ellipses on to the image to mark the locations of bubbles. A short, online tutorial shows how to use the tool, and examples of prominent bubbles are given. As a secondary task, users can also mark rectangular areas of interest, which can be labelled as small bubbles, green knots, dark nebulae, star clusters, galaxies, fuzzy red objects or ‘other’. Examples of these are also given in a tutorial on the website, and these are discussed further in Section 3.4. Users can add as many annotations as they wish before submitting the image, at which point they are given another image for annotation. Each image’s annotations are stored in a database as a classification, and users can see the images they have classified in a part of the site called ‘My Galaxy’. Users can only classify a given image once.

When marking bubbles, users place a circular annulus that can be scaled in size and stretched into an elliptical annulus. As they first draw out an object, they are able to control the position and size of the bubble. Once the bubble has been drawn they can edit these initial parameters as well as the bubble’s ellipticity, annular thickness and rotation. In this way users can attempt to match the bubbles they see in each image to give an accurate representation. Users are also able to mark regions of the annulus where there is no obvious emission – as in the case of broken or partial bubbles. These ‘cut-outs’ are created by erasing (and then refilling if necessary) segments of the annulus. Of all the bubbles drawn, 75 per cent have thicknesses other than the default or minimum values, 50 per cent are non-circular (of which 56 per cent have been rotated) and only 12 per cent were drawn with cut-outs.

When marking another area of interest on an image (e.g. star clusters, green knots, etc.) users simply draw rectangles. Since these objects are secondary to the main bubble-finding task, the site was designed so that they should be simple and quick to mark. Simple rectangles allow us to record the positions and approximate sizes of any interesting objects.

Citizen scientists can discuss and share objects and images via the ‘Talk’ interface5 where more heavily discussed objects trend upwards as they do, for example, in a news aggregator. The ‘Talk’ web application is open source6 and was developed by the Zooniverse team for general use on its projects, including the MWP. Through the use of ‘Talk’, interesting objects float to the top of discussion and are identified as interesting to the MWP scientists. Thus, by harnessing the social nature of the MWP, we can extract additional information from the classification process.

Principally the objects highlighted by volunteers using ‘Talk’ are visually interesting, unusual or defy classification in the primary interface. One example of such a feature is the ‘yellowballs’ – named by users because of their compact, rounded, yellow appearance in MWP images. It is believed that these represent a type of ultracompact H II region and they are the subject of a future paper currently being prepared. They were an unintended consequence of our colour scheme, and thus no flag was provided in the main interface to enable volunteers to mark their presence.

2.1 Users’ favourite bubbles

Users are able to mark images as ‘favourites’ as they classify on the MWP. These images are often particularly beautiful, interesting or unusual. Ten of the most favoured images are shown in Fig. 2. The list includes many notable and well-known objects that have been ‘rediscovered’ by the MWP users, such as the Eagle Nebula (b), the Trifid Nebula (c) and the Galactic Centre (h). Users not only select beautiful images but also those that contain interesting objects. Image i contains the massive star cluster Westerlund 1 (Westerlund 1961), which is much discussed on Milky Way Talk (see Section 2).

3 CATALOGUE CONSTRUCTION

3.1 MWP user statistics and scoring

By 2011 October, over 35 000 people had logged in to milkywayproject.org. 45 per cent of those users classified at least one of the 12 263 MWP images; 25 per cent classified five or more images and 5.7 per cent classified 50 or more. Fig. 3 shows the distribution of users with the number of classifications they perform over the lifetime of their involvement with the project.

The top five nations visiting the MWP have been the United States (42 per cent), the United Kingdom (20 per cent), Canada (4.3 per cent), Poland (3.8 per cent) and Germany (3.8 per cent). The remaining 26.1 per cent of project visitors have come from 173 other countries. The MWP site is predominately accessed by English-speaking countries with a high level of internet connectivity; in addition, it has been translated into Polish and hence that country also provides a large proportion of visitors.

In order to assist in the data-reduction process, users are given scores according to how experienced they are at drawing bubbles. We treat the first 10 bubbles a user draws as practice drawings and these are not included in the final reduction. Users begin with a score of 0 and are given scores according to the number of precision bubbles they have drawn (see Table 1). Precision bubbles are those drawn using the full toolset, meaning they have to have adjusted the ellipticity, the thickness and the rotation. This is done to ensure that users’ scores reflect their ability to draw bubbles well. While only precision bubbles are used to score volunteers, all bubbles are drawn as included in the data reduction. The scores are used as weights when averaging the bubble drawings to produce the catalogue.

3.2 Combining the user-drawn bubbles

Combining all bubble drawings, at all zoom levels, the MWP has created a data base of 520 120 user-drawn bubbles as of 2011 October 31. To identify bubbles with at least five user classifications for inclusion in our final catalogue, the data reduction process (shown as a flow chart in Fig. 4) begins by splitting the data set into 2’ × 2’ boxes and treating each box in turn.

If a box contains five or more bubbles with a maximum outer ellipse that is between a half- and a whole box, then a simple clustering algorithm picks out groups of these bubbles with dispersions in their positions of less than a quarter of the box size (i.e. less than the radius of the smallest bubble under consideration). If a cluster contains at least five bubble drawings, it is saved for additional processing and inclusion in the catalogue, and the bubble drawings are removed from the working list. Bubble drawings that are not clustered enough, or numerous enough, remain on the working list for potential inclusion in a later iteration. The box is then split into four, and the process repeats until no more boxes containing five or more bubbles are found, or until the box size falls below the smallest bubbles drawn on the MWP – the ellipse-drawing tool has a lower size limit.
Figure 2. 10 MWP images most favoured by volunteers, in no particular order. Coordinates are image centres; image sizes are indicated by the zoom level (zoom). Zoom levels 1, 2 and 3 refer to images of 1.5 × 1°, 0.75 × 0.375 and 0.3 × 0.15, respectively. Colour figure available online.
Table 1. A user’s score given according to the number of precision bubbles he/she has drawn. Precision bubbles are those that require the use of multiple modifications to the default bubble parameters (see the text for details). As such the number of precision bubbles drawn is a proxy for the care users are willing to take, and thus their experience with the tool.

| Precision bubbles | Score |
|-------------------|-------|
| 1                 | 1     |
| 5                 | 2     |
| 20                | 4     |
| 50                | 6     |
| 100               | 8     |
| 500               | 10    |

of a diameter of 20 pixels (0.45 arcmin at the highest image zoom level).

The same process is also run on an offset grid where the initial boxes are displaced by 1° in both galactic latitude and longitude. This catches bubbles that may fall on box boundaries. The two resultant lists of bubble groups are combined later on by clustering bubbles that fall within 0.5 radii separation from each other and which both have radii within 50 per cent of each other.

Each resulting cluster of bubbles marked ≥5 times is combined into a single ‘clean’ bubble using a weighted mean, where the weighting is provided by the score of the user that drew each bubble (see Section 3.1). The bubble’s mean size, position, angle (in degrees from North, in galactic coordinates) and thickness are all determined in this way (see Fig. 5). The cleaned bubble catalogue is given in Table 2.

The bubble’s ‘hit rate’ is the ratio of the number of qualifying bubbles drawn to the number of times the bubble was seen by users on the MWP website. A clean bubble produced from a cluster of five user drawings placed on to an asset that was seen 50 times would have a hit rate of 0.1. In cases where a bubble could be marked in more than one asset, for example across different zoom levels, the total view counts are summed, such that a cluster of five bubbles drawn on to two assets of 50 views each would have a hit rate of 0.05. The hit rate gives a measure of consensus among users that a bubble is present in the data.

The bubble’s dispersion is also calculated as the spread in coordinates of the individual classifications ($\sqrt{\sigma_x^2 + \sigma_y^2}$, where $\sigma$ is the variance in the coordinate value).

Only bubbles that were seen 50 times or more, and which have hit rates of 0.1 or more, are included in the final catalogue. This ensures that each final, cleaned bubble is a combination of at least five individual users’ drawings and was drawn by volunteers at least 10 per cent of the time when displayed on the website.

3.3 Selection effects

It is not known how many bubbles exist in the Galaxy; hence, it is impossible to quantify the completeness of the MWP catalogue. There will be bubbles that are either not visible in the data used on the MWP or not seen as bubbles. Distant bubbles may be obscured by foreground extinction. Faint bubbles may be masked by bright Galactic background emission or confused with brighter nebular structures. Fragmented or highly distorted bubbles present at high inclination angles may not appear as bubbles to the observer.

The MWP’s ‘citizen science’ approach creates its own biases, whilst overcoming others experienced in similar studies. By comparison with CP06, this study has many thousands of times more eyes scanning each section of the sky, and each section is broken down to an optimal colour stretch, thus improving the chances of seeing bubble-like structures. The majority of the MWP volunteers have no professional bias or expectation as to what constitutes a ‘good’ bubble. MWP volunteers may experience measurement fatigue when classifying assets with many bubbles. They may also suffer bright neighbour bias and fail to draw quite obvious bubbles that are adjacent to very prominent or beautiful examples.

3.4 Small bubbles and other objects

In addition to marking elliptical bubbles on images, users are also encouraged to mark the locations of other interesting objects. Users can mark areas using a simple rectangle and are asked to label
Figure 5. Example of raw user drawings and reduced, cleaned result using a sample MWP image. A GLIMPSE-only colour sample is included to illustrate the differences in the appearance of images inspected by CP06 and the MWP users. The image shown is centred at $l = 18.8, b = -0.125$, with size $1.5 \times 0.75$. In image (c) all user-drawn bubbles are placed, from all zoom levels, with an opacity of 2.5 per cent. In image (d) reduced bubbles are placed with an opacity two times their hit rate, such that bubbles with hit rates $\geq 50$ per cent are drawn as solid white bubbles. Colour figure available online.

4 RESULTS

4.1 Catalogue description

The final, reduced catalogue contains 5106 visually identified bubbles. These are split into a catalogue of 3744 large bubbles drawn by users as ellipses, and a catalogue of 1362 small bubbles drawn by users at the highest zoom level images in the MWP. These bubbles are plotted in Galactic coordinates in Fig. 6. Each bubble in both lists has been drawn by at least five different individuals, and the listed parameters have been obtained from a weighted average based on each user’s score (see Section 3.1). The complete catalogue can be accessed at http://data.milkywayproject.org. Table 2 gives the large bubbles, in order of hit rate. The large-bubble catalogue includes values for the position in galactic longitude and latitude; radius and thickness, position angle (given in degree from North), eccentricity; the hit rate (described in Section 3.2) and a hierarchy flag indicating whether bubbles are (1) identified as having further smaller bubbles within their boundary, or (2) located on the rim of a larger bubble. The radius and thickness values represent effective values calculated from geometric means of the inner and outer diameters in both axes of the ellipse, defined as

$$R_{\text{eff}} = \frac{(R_{\text{in}}r_{\text{in}})^{0.5} + (R_{\text{out}}r_{\text{out}})^{0.5}}{2}$$

(1)

$$T_{\text{eff}} = (R_{\text{out}}r_{\text{out}})^{0.5} - (R_{\text{in}}r_{\text{in}})^{0.5},$$

(2)

where $R_{\text{in}}, R_{\text{out}}$ are the inner and outer semimajor axes, and $r_{\text{in}}, r_{\text{out}}$ the inner and outer semiminor axes, respectively. The minimum effective radius imposed by the limitations of the drawing tool (see Section 3.4) is 0.27 arcmin. The eccentricity is calculated using

$$e = \frac{(R_{\text{in}}^2 - R_{\text{out}}^2)^{1/2}}{R_{\text{in}}}$$

(3)

with symbols as above. The MWP drawing tools do not allow for different eccentricities in inner and outer sizes of the bubbles.

4.2 Cross-matching with existing catalogues

For each bubble and small bubble produced by the MWP, cross-matching was performed with GLIMPSE bubbles (i.e. the CP06 and CWP07 catalogues) and the Paladini et al. (2003) and Anderson et al. (2011) catalogues of H II regions. Sources are

them as either a small bubble, green knot, dark nebula, star cluster, galaxy, fuzzy red object or other.

The ellipse-drawing tool of the MWP has a lower size limit of an inner diameter of 20 pixels (0.45 arcmin at the highest image zoom level). The corresponding minimum outer diameter is 0.64 arcmin. The small bubble category allows users to mark bubbles which are too small to draw in detail but which can still be clearly made out. These small bubbles are reduced in a similar fashion to the more complex ellipses. To produce the catalogue of small bubbles listed in Table 3 we use only the small bubbles drawn by users at the highest zoom level (this is the vast majority of those drawn). Small bubbles marked at lower zoom levels are equivalent to larger bubbles at the higher zoom levels. By rounding their locations to the nearest 20 pixels, the drawings are clustered. Our catalogue of 1362 small bubbles is given in Table 3. Each of these small bubbles was drawn by at least five users and was drawn by at least 10 per cent of the volunteers who saw it – as with the main, large-bubble catalogue.

Catalogues of green knots, dark nebulae, star clusters, galaxies, fuzzy red objects and objects in the other category are currently being prepared for later publication.
marked as coincident when the central coordinate of the catalogue object lies within the radius of the MWP bubble. 12 percent of MWP bubbles matched GLIMPSE bubbles, and 86 percent of GLIMPSE bubbles were rediscovered by the MWP. Similarly 10 and 7 percent of MWP bubbles are coincident with Paladini et al. (2003) and Anderson et al. (2011) Hα regions, respectively. The MWP finds 86 percent of the available Paladini sources and 96 percent of the Anderson sources. From the Anderson et al. (2011) sample, kinematic distance ambiguities have been resolved for a subset of 266 Hα regions in the first Galactic quadrant (Anderson & Bania 2009). 185 of these, 70 percent, were recovered by the MWP. The presence of 24 μm emission coincident with 20 cm emission was a selection criterion for Anderson et al. (2011) and so there may be some overlap in selection methods with the MWP.
**Table 3.** MWP small bubbles. Shown here are the 50 bubbles from the small-bubble catalogue with the highest hit rates. The complete small-bubble catalogue contains 1362 visually identified bubbles. Where cross-correlation is possible, identifiers from CP06 and CWP07 are given. Hierarchy flags denote (1) bubbles identified as having smaller bubbles on their rim and (2) bubbles located on the rim of a larger bubble. The complete bubble catalogue can be found online at http://data.milkywayproject.org and as supporting information with the electronic version of the paper.

| MWP ID | Churchwell ID | $l$ (°) | $b$ (°) | Mean radius (arcmin) | Hit rate | Hierarchy flag |
|--------|---------------|---------|---------|----------------------|----------|---------------|
| MWP1G331470–001400S | – | 331.47 | −0.14 | 0.37 | 0.54 | – |
| MWP1G018180–001100S | – | 018.18 | −0.11 | 0.34 | 0.50 | – |
| MWP1G017690–000900S | – | 017.69 | −0.09 | 0.42 | 0.49 | – |
| MWP1G336010–003700S | – | 336.01 | −0.37 | 0.37 | 0.45 | – |
| MWP1G012200–001600S | – | 012.20 | −0.16 | 0.38 | 0.42 | – |
| MWP1G013780–004900S | – | 013.78 | +0.49 | 0.42 | 0.42 | 2 |
| MWP1G024920–000800S | – | 024.92 | +0.08 | 0.47 | 0.42 | – |
| MWP1G338870–000200S | – | 338.87 | +0.02 | 0.45 | 0.42 | – |
| MWP1G011020–003700S | – | 011.02 | −0.37 | 0.41 | 0.41 | – |
| MWP1G345510–001600S | – | 345.51 | +0.16 | 0.40 | 0.41 | 2 |
| MWP1G039430–001900S | – | 039.43 | −0.19 | 0.41 | 0.40 | – |
| MWP1G339770–000100S | – | 339.77 | +0.01 | 0.43 | 0.40 | 2 |
| MWP1G346040–000500S | – | 346.04 | +0.05 | 0.34 | 0.40 | – |
| MWP1G331400–001800S | – | 331.40 | −0.18 | 0.44 | 0.39 | 2 |
| MWP1G004060–000100S | – | 004.06 | −0.01 | 0.38 | 0.38 | – |
| MWP1G008910–001700S | – | 008.91 | +0.17 | 0.38 | 0.38 | 2 |
| MWP1G012810–003100S | – | 012.81 | −0.31 | 0.44 | 0.38 | 2 |
| MWP1G052980–006200S | – | 052.98 | −0.62 | 0.43 | 0.38 | 2 |
| MWP1G339780–000200S | – | 339.78 | +0.02 | 0.34 | 0.38 | 2 |
| MWP1G343290–001600S | – | 343.29 | +0.16 | 0.37 | 0.38 | 2 |
| MWP1G357970–001700S | – | 357.97 | −0.17 | 0.37 | 0.38 | 2 |
| MWP1G010990–003700S | – | 010.99 | −0.37 | 0.45 | 0.37 | – |
| MWP1G018850–004800S | N24 | 018.85 | −0.48 | 0.41 | 0.37 | 2 |
| MWP1G049570–002700S | – | 049.57 | −0.27 | 0.44 | 0.37 | 2 |
| MWP1G301760–002600S | – | 301.76 | +0.26 | 0.41 | 0.37 | – |
| MWP1G331140–001100S | – | 331.14 | +0.11 | 0.34 | 0.37 | – |
| MWP1G334350–006600S | – | 334.35 | −0.66 | 0.34 | 0.37 | – |
| MWP1G352760–003500S | – | 352.76 | −0.35 | 0.39 | 0.37 | – |
| MWP1G307500–008200S | – | 307.50 | −0.82 | 0.49 | 0.36 | 2 |
| MWP1G332670–003300S | – | 332.67 | −0.33 | 0.38 | 0.36 | – |
| MWP1G339240–000100S | – | 339.24 | +0.01 | 0.47 | 0.36 | 2 |
| MWP1G005100–000000S | – | 005.10 | +0.00 | 0.49 | 0.35 | – |
| MWP1G014210–001100S | – | 014.21 | −0.11 | 0.41 | 0.35 | – |
| MWP1G032400–003300S | – | 032.40 | −0.33 | 0.40 | 0.35 | – |
| MWP1G035720–009300S | – | 035.72 | −0.93 | 0.42 | 0.35 | – |
| MWP1G335940–002900S | – | 335.94 | −0.29 | 0.46 | 0.35 | 2 |
| MWP1G339700–003800S | – | 339.70 | +0.30 | 0.37 | 0.35 | – |
| MWP1G002170–000100S | – | 002.17 | +0.01 | 0.39 | 0.34 | – |
| MWP1G005630–002900S | – | 005.63 | −0.29 | 0.42 | 0.34 | – |
| MWP1G009970–002100S | – | 009.97 | −0.21 | 0.41 | 0.34 | – |
| MWP1G018170–003400S | – | 018.17 | +0.34 | 0.46 | 0.34 | 2 |
| MWP1G025820–001900S | – | 025.82 | −0.19 | 0.36 | 0.34 | – |
| MWP1G027610–000300S | – | 027.61 | +0.03 | 0.45 | 0.34 | 2 |
| MWP1G046140–001400S | – | 046.14 | −0.14 | 0.39 | 0.34 | – |
| MWP1G029770–007000S | – | 297.72 | −0.70 | 0.46 | 0.34 | – |
| MWP1G311630–002600S | – | 311.63 | −0.26 | 0.34 | 0.34 | – |
| MWP1G355470–002000S | – | 355.47 | +0.20 | 0.38 | 0.34 | – |
| MWP1G342900–000900S | – | 342.90 | −0.09 | 0.50 | 0.34 | – |
| MWP1G345480–002200S | – | 345.48 | −0.22 | 0.36 | 0.34 | – |
| MWP1G388900–000800S | – | 358.89 | +0.08 | 0.34 | 0.34 | – |

Mizuno et al. (2010) catalogued 416 disc- and ring-like structures seen in the MIPS GAL 24 μm images and suggested that the majority of these objects were produced by evolved stars. The Mizuno et al. (2010) catalogue is dominated by small sources with radii < 0.2 arcsec, and so these objects should be unlikely to overlap with the main MWP bubble catalogue. In fact we rediscover only 9 per cent of the sources in this catalogue and they constitute less than 1 per cent of the combined MWP bubble catalogues. Of the 1093 small MWP bubbles with a mean width of less than 1 arcmin, only five correspond to objects in the Mizuno et al. (2010) catalogue (for...
Figure 6. All 5106 MWP bubbles plotted in Galactic coordinates. The MWP large-bubble catalogue is marked with outer radii as grey ellipses, and small bubbles are shown as green crosses.

reference, the IDs of these bubbles are MWP1G031730+007000S, MWP1G314360+004900S, MWP1G319220+001600S, MWP1G334110+003800S, MWP1G358770+001100S).

The above crossover fractions (summarized in Table 4) show that the MWP has excellent overlap with existing bubble catalogues and is also more complete, in terms of locating H II regions, than the two Churchwell studies. The lack of agreement between the MWP and Mizuno et al. (2010) is evidence that the small-bubble catalogue is not contaminated by small, 24 µm disc- and ring-like structures.
Table 4. Crossover between the MWP and relevant catalogues of bubbles (CP06; CWP07), H II regions (Paladini et al. 2003; Anderson et al. 2011) and MIPSGAL ring-like structures (Mizuno et al. 2010).

| Catalogue                | Fraction of MWP bubbles matched | Fraction rediscovered |
|--------------------------|---------------------------------|-----------------------|
| CP06 and CWP07           | 0.12                            | 0.85                  |
| Paladini et al. (2003)   | 0.10                            | 0.86                  |
| Anderson et al. (2011)   | 0.07                            | 0.96                  |
| Mizuno et al. (2010)     | 0.01                            | 0.09                  |

4.3 Errors

A number of quality control measures, such as the user weighting scheme, were adopted from the outset; these are described in Section 3.1. To assess the performance of the processing procedure, we examine the four bubbles in the catalogue with the highest classification scores, i.e. those drawn by the largest number of users. These bubbles were identified by an average of 243.2 users (see Figs 7–10). Three of these were previously identified as bubbles and H II regions by other authors (CP06; Lockman et al. 1996; Kuchar & Clark 1997; Misanovic et al. 2002). Bubble MWP1G303056+01645 (Fig. 10) has no associated CP06 bubble.

In bubbles MWP1G309059+001661 (Fig. 7) and MWP1G303056+001645 (Fig. 10) the computed weighted averages for the bubble position and size parameters, indicated with the dashed lines, accurately follow the distribution of individual classifications, and the distribution is well approximated by a Gaussian distribution. In these cases, each user classification can be reasonably considered to be an independent measurement of the same quantity. In other cases, such as with MWP1G304463−00217 (Fig. 9), there are skews or multiple peaks in the distribution of classifications, indicating a relative lack of consensus among the users over the location, size, shape or multiplicity of the bubble. Indeed, multiple peaks in the distribution may suggest the presence of more than one bubble at a given location.

These variations can be attributed to the definition of our clustering algorithm. Using a tighter clustering threshold will do better at separating out closely spaced bubbles, but will also artificially fragment single bubbles into multiples where the bubble’s positional coordinates have a high uncertainty value. Our data reduction could be modified to track the level of dispersion in bubble-drawing clusters and dynamically split or reject clusters based on this value. We aim to address this issue in a subsequent data release.

To assess the reliability of the data presented in a quantitative way, two reliability metrics were defined.

(i) Hit rate: the fraction of all users that were presented with the region of sky, who drew a bubble at this location (described in Section 3.2). This is a measure of the reliability of the existence of a bubble.

(ii) Dispersion: the spread in coordinates of the individual classifications ($\sqrt{\sigma_x^2 + \sigma_y^2}$, where $\sigma$ is the variance in the coordinate value). This describes the uncertainty on a bubble’s location.

Note that these metrics do not offer any insight into the physical nature of the bubbles – i.e. how likely a feature with the form of a bubble is to actually be an H II region – rather they reflect the level of consensus gathered from the users that a bubble appears at this location.

4.4 Completeness

To investigate the completeness of our catalogue, we compare the number of classifications over time with the number of new large bubbles found by volunteers. A new bubble’s discovery is defined as the earliest date a given bubble, which ended up the large-bubble
catalogue, was drawn by a user. These quantities are plotted in Fig. 11, binned by a 2-week period.

The plot shows how the number of new bubbles found has declined at a faster rate than the overall number of classifications. A declining bubble discovery rate implies an increase in multiple classifications for pre-existing bubbles.

If we assume the latest discovery rate of approximately one new large bubble per 5000 classifications to be maintained, an additional 500 000 classifications would yield only an additional 100 large bubbles or roughly 3 per cent of the current catalogue (3744 large bubbles). This is a conservative estimate, as the rate of bubble discoveries continues to decline a year into the project. As the project continues to gather classifications, more information will become available about the bubble discovery rate over time.

The data processing algorithm also plays a role in the completeness of the catalogue. Specifically, the clustering threshold chosen in our data processing algorithm may merge bubbles that closer inspection suggests to consist of several bubble components. Figs 7–10 illustrate how the dispersion is an informative metric for the possible multiplicity of a given bubble. We can use this quantity to assess how many bubbles may have been ‘lost’ in data processing.

Fig. 12 shows the cumulative distribution function of the dispersion of all large bubbles, normalized to their effective radii. If we posit that a dispersion on a bubble’s central position covering much of the bubble’s effective radius is highly suggestive of underlying multiplicity, we can estimate how many bubbles may be lost. While 89 per cent of large bubbles have a dispersion $<0.5R_{\text{eff}}$, 2.4 per cent (88 bubbles) show a dispersion $>0.75R_{\text{eff}}$. Thus in the worst-case scenario where we assume that each of these is likely to host at least one additional bubble around its rim, we can estimate that around 80–100 bubbles were artificially merged by the algorithm, or around 2–3 per cent. Combining these numbers, we estimate the current catalogue to be complete to $>94$ per cent.

The completeness of the small-bubble catalogue is harder to assess quantitatively, as the classification tool is much coarser than the ellipse-drawing tool for the large bubbles.

4.5 Bubble catalogue properties

The longitudinal distribution of the MWP bubbles is shown in Fig. 13. This figure shows a broad rise and fall either side of the Galactic Centre, with the number of bubbles beginning low near the Galactic Centre, rising and then diminishing towards the edge of the survey at $l \pm 65^\circ$. The notable lack of bubbles around $l = 60^\circ$ is partly due to reaching the survey’s edge at $l = 65^\circ$ and also due to an absence of any filaments or bubbles around $l = 58^\circ$. Fig. 13 marks several notable Galactic line-of-sight features in red, adopted from Beuther et al. (2011). Many of the rises and falls in the number of bubbles across the range of Galactic longitude appear to derive from the large-scale structure of the Milky Way.

The drop in bubble count near the Galactic Centre may be a physical effect but could also arise due to confusion from background emission towards the centre of the Milky Way. Beuther et al. (2011) studied the distribution of submm clumps from ATLASGAL – the APEX Telescope Large Area Survey of the GALaxy (Schuller et al. 2009). They find a large peak in the number of sources towards the Galactic Centre and note that this is in contrast with current surveys of HII regions (e.g. Anderson et al. 2011) and recent surveys of H2O and CH3OH masers (Green et al. 2011; Walsh et al. 2011).

The distribution of large bubbles with latitude is shown in Fig. 14. We divide the GLIMPSE/MIPSGAL survey area by longitude into northern ($l = 0^\circ$ to $65^\circ$) and southern ($l = 295^\circ$ to $360^\circ$) regions to facilitate comparison of possible morphological, positional and size differences with CP06 and CWP07. The same profile is shown in fig. 5 of CP06 and the plots share the same asymmetry towards lower latitudes.
Figure 9. Errors for bubble MWP1G304463−000217. This bubble has a hit rate of 0.395 and a dispersion of 0.70 arcmin. See Fig. 7 for more information. Colour figure available online.

Figure 10. Errors for bubble MWP1G303056+001645. This bubble has a hit rate of 0.355 and a dispersion of 0.59 arcmin. See Fig. 7 for more information. Colour figure available online.

Figs 15, 16 and 17 show the distributions of the MWP bubble radii, thicknesses and eccentricities, respectively. For reference, the properties of the 591 combined GLIMPSE (CP06 and CWP07) bubble catalogues are also shown, along with the distributions for the 540 MWP large bubbles that were cross-matched to those catalogues.

The MWP bubbles display a decreasing power-law distribution with increasing angular size (Fig. 15). This is similar to the
distribution for the GLIMPSE catalogues. Figs 16 and 18 show that the MWP bubbles are generally thicker than their GLIMPSE counterparts, which is to be expected given the inclusion of additional wavelength data and the chosen MWP colour combination (see Fig. 5).

The MWP eccentricities (Fig. 17) peak at $\sim 0.35$, compared to a value of $\sim 0.7$ for the previous GLIMPSE bubbles. The MWP finds an order of magnitude more bubbles than CP06 and CWP07 combined – and many more smaller bubbles. The difference between the catalogues could result from our averaging the parameters of multiple ellipses to create the MWP large bubbles, i.e. merging a large number of elliptic annuli tends towards circularity as the number and variety increase. The limitations of our interface may also mean that smaller bubbles are harder to draw with precision.

4.6 Bubble distances

Cross-matching with Anderson & Bania (2009) provides distances to 185 of the larger MWP bubbles (cross-matching as described in Section 4.2). Fig. 19 shows these 185 bubbles plotted in terms of their diameters against (a) their distance from the Sun and (b) their distance from the Galactic Centre. Fig. 19(a) shows a slight tendency for more distant bubbles to be larger – most likely a reflection of the selection effect whereby only very large distant bubbles are easily seen in the GLIMPSE/MIPSGAL images used in the MWP. Fig. 19(b) shows little correlation other than reflecting the fact that fewer bubbles are seen at greater distances from the Galactic Centre – this is to be expected given the longitudinal range of the MWP and the confusion effect where nearer bubbles and dust obscure more distant ones, looking towards the Galactic Centre.

Fig. 20 shows that larger bubbles tend to have thinner shells relative to their diameters. This could be the effect of material cooling and condensing as the bubble expands. The apparent relationship warrants further investigation, but we note caution that this relationship may be biased by the effect of a minimum size and thickness of a bubble in the MWP drawing tool.

4.7 ‘Heat maps’

In addition to the reduced bubble catalogue, a crowd-sourced ‘heat map’ of bubble drawings has also been produced. This simple map reflects the full range of classifications placed on to the MWP images. All 520 120 bubbles drawn by all users are placed on to the sky with an opacity of 2.5 per cent meaning that 40 individuals need to have drawn over the same region for it to become fully opaque (white). Examples of these images are shown in Figs 5 and 21. These MWP ‘heat maps’ are available at http://data.milkywayproject.org as FITS and DS9 region files.

The MWP ‘heat maps’ allow the bubble drawings to be explored without them needing to be reduced to elliptical annuli. Rather, the ‘heat maps’ allow contours of overlapping classifications to be drawn over regions of the Galactic plane reflecting levels of agreement between independent classifiers. In most cases the structures outlined in these maps are photodissociation regions (PDRs) traced by 8 $\mu$m emission, but more fundamentally they are regions that multiple volunteers agree reflect the rims of bubbles.

5 DISCUSSION

CP06 and CWP07 performed visual classification, using a handful of experts, of GLIMPSE data only. The MWP images include data from the MIPSGAL 24 $\mu$m survey, which can enhance the shape and definition of bubble-like structures (see Fig. 5). Most likely the biggest difference between the MWP and GLIMPSE studies is that tens of thousands of classifiers are involved in the MWP and most begin with no grounding in the astrophysical phenomena they are told to locate. This results in a set of unbiased classifications, sourced from multiple independent classifiers who are able to explore and annotate the entire data set without fatigue.

5.1 The nature of MWP bubbles

Bubbles were identified in the MWP based on ring- or arc-like shapes observed by multiple independent classifiers in IR images of the Galactic plane. Many astrophysical phenomena can give rise to such features. Aside from young, massive stars, several classes of evolved objects can produce bubbles in the interstellar medium (ISM), including supernova remnants (SNRs), planetary nebulae (PNe), Wolf–Rayet (W-R) stars and (post-)asymptotic giant branch (AGB) stars. SNRs are known to produce IR-detectable shell and bubble features; however, Reach et al. (2006) found only a small number of IR counterparts in GLIMPSE images of the radio-detected SNR. We therefore expect the contamination by SNR to be low. CP06 reports no contamination of their bubble catalogue by...
Figure 13. Histogram showing the distribution of combined MWP small- and large-bubble catalogues with Galactic longitude \(l\). Notable line-of-site features (including Galactic spiral arms) marked as dashed boxes.

Figure 14. Distribution of MWP large bubbles with \(b\) (see fig. 5 of CP06). In addition to the overall distribution (drawn as a solid line), the dashed line shows just southern bubbles \((l = 295^\circ \text{ to } 360^\circ)\) and the dotted line northern bubbles \((l = 0^\circ \text{ to } 65^\circ)\).

PNe or W-R stars; their presence in the MWP catalogue is assumed to be small.

Ground-based mid-IR imaging of (post-)AGB candidates by Lagadec et al. (2011) and Meixner et al. (1999) shows that these objects may be surrounded by circumstellar dust shells emitting at mid-IR wavelengths following episodes of mass loss. Morphologies are varied and often complex, and a small number of objects were found to display toroidal or shell-like morphologies. In the Lagadec et al. (2011) sample, 59 of 93 objects are however unresolved even at high spatial resolutions, and the remainder are typically \(<10\) arcsec in size. Any bubbles associated with AGB or post-AGB objects are therefore likely to be found in the small-bubble catalogue, and many will simply appear as highly reddened point sources in the GLIMPSE images.

Further detailed cross-matching with known catalogues of these objects, and possibly observational follow-up, is required for providing accurate identification of the full MWP bubble sample. The large number of bubbles suggests that new identifications of or associations with different types of objects are highly likely. We note for example the association of a large bubble from the CP06 catalogue (S174, identified in MWP as MWP1G299701+000151) with the symbiotic star BI Crucis reported by McCollum et al. (2008).

Many lines of evidence suggest that the MWP bubbles, like the GLIMPSE bubbles, predominantly trace massive star formation. The distribution of bubbles with Galactic latitude reflects a low scale height (Fig. 14), similar to molecular clouds and the known Galactic OB population (see CP06).

A bubble is produced around a massive star when an H II region, driven by thermal overpressure, stellar winds, radiation pressure or a combination of these feedback mechanisms, expands into the surrounding cold ISM, sweeping up gas and dust into a dense shell surrounding a low-density, evacuated cavity (Weaver et al. 1977; Garcia-Segura & Franco 1996; Arthur et al. 2011; Draine 2011). The relative contributions of different feedback mechanisms likely depend on the properties of the driving star(s), with the most massive, early O-type stars combining powerful stellar winds with high UV luminosities producing ‘wind-blown bubbles,’ while lower mass,
Figure 16. Distribution of the thicknesses of the 3744 MWP large bubbles shown as a dot–dashed blue line. The 591 bubbles from the combined CP06 and CWP07 catalogues are shown as a dashed line. The subset of 540 MWP bubbles cross-matched to those from CP06 and CWP07 are shown as a solid line. The minimum thickness a user can draw is \( \approx 0.2 \) arcmin.

Figure 17. Distribution of the eccentricities of the 3744 MWP large bubbles shown as a dot–dashed blue line. The 591 bubbles from the combined CP06 and CWP07 catalogues are shown as a dashed line. The subset of 540 MWP bubbles cross-matched to those from CP06 and CWP07 are shown as a solid line.

late-O and B dwarfs give rise to ‘classical’ H\( \text{\upshape II} \) regions powered by UV photons alone (CWP07; Watson et al. 2008), Castor, McCray & Weaver (1975) and Weaver et al. (1977) derived analytic solutions for the expansion of stellar wind-blown bubbles into a uniform low-density medium. More recent modelling efforts have included the effects of ionizing radiation and the stellar winds (e.g. Capriotti & Kozminski 2001; Draine 2011).

The wind-blown bubbles around massive stars produced by these models display the following general structure.

(i) An inner cavity cleared rapidly by a freely flowing hypersonic stellar wind.

(ii) A high-temperature region of shocked stellar wind material \( (T > 10^6 \text{ K}) \).

(iii) A shell of shocked, photo-ionized gas \( (T \sim 10^4 \text{ K}) \).

(iv) A shell of non-shocked, ionized gas \( (T \sim 10^4 \text{ K}) \).

(v) An outer shell of neutral material.

Figure 18. Distribution of the ratio of thickness to outer diameter for the 3744 MWP large bubbles shown as a dot–dashed blue line. The 591 bubbles from the combined CP06 and CWP07 catalogues are shown as a dashed line. The subset of 540 MWP bubbles cross-matched to those from CP06 and CWP07 are shown as a solid line.

The bright PAH emission in the PDRs surrounding H\( \text{\upshape II} \) regions produces the bright 8 \( \mu \text{m} \) bubble rims, while dust mixed with the ionized gas and heated by the hard radiation field produces 24 \( \mu \text{m} \) emission interior to the bubbles (Povich et al. 2007; Watson et al. 2008; Everett & Churchwell 2010; Watson et al. 2010). These considerations guided our choice of multiband colour combination for the MWP images, which visually bias identification towards H\( \text{\upshape II} \) regions. This is consistent with the cross-matching results in Section 4.2.

Watson et al. (2008) proposed that wind-blown bubbles with cleared, central cavities tend to produce a toroidal or arc-shaped morphology in the 24 \( \mu \text{m} \) emission, while this emission is more likely to be centrally peaked in classical H\( \text{\upshape II} \) regions. Draine (2011) found that radiation pressure alone can produce cleared central cavities in H\( \text{\upshape II} \) regions, but suggested cases where winds must also play a role. A wide range of 24 \( \mu \text{m} \) morphologies are apparent among the MWP bubbles.

The models described above consider the effects of a single star on its surrounding uniform-density medium. Massive stars form preferentially in clusters within molecular clouds with highly non-uniform densities. Freyer, Hensler & Yorke (2003) performed two-dimensional radiation hydrodynamic simulations to study the combined influence of a massive star’s stellar wind and its ionizing radiation on the evolution of the circumstellar cloud material. Mellema et al. (2006) simulated the evolution of an H\( \text{\upshape II} \) region surrounding a single O star, focusing on the interaction of the ionization front with the turbulent cloud medium. More recently, Arthur et al. (2011) performed magnetohydrodynamic (MHD) simulations of H\( \text{\upshape II} \) regions around OB stars expanding into a turbulent, magnetized medium. All of these simulations show strong inhomogeneities around the boundary of the expanding shell, reproducing the clumps, globules and filaments observed in H\( \text{\upshape II} \) regions at optical or IR wavelengths and seen in the MWP bubbles.

5.2 Multiple bubbles and potentially triggered star formation

A major application of the CP06 and CPW07 bubble catalogues has been triggering studies (e.g. Deharveng et al. 2008; Watson et al. 2008, 2010; Zavagno et al. 2010, and references therein), as the Churchwell catalogues provided a large sample of (mostly...
Figure 19. Bubble diameters against (a) distance from the Sun and (b) distance from the Galactic Centre, for 185 MWP bubbles cross-matched with Anderson & Bania (2009).

Figure 20. Variation of the thickness–diameter ratio of large MWP bubbles (Table 2) with their physical size, for 185 MWP bubbles cross-matched with Anderson & Bania (2009).

Previously unknown regions of potential triggering by massive stars. In any individual source, difficulties with establishing cause and effect are significant, and the Churchwell catalogues provided a basis for triggering investigations to move beyond case studies of individual objects to statistics.

Bubbles serve as laboratories to test theories of sequential, massive star formation triggered by massive star winds and radiation pressure (Elmegreen & Lada 1977; Whitworth et al. 1994). CP06 and CWP07 noted that a (small) fraction of bubbles exhibited hierarchical structure, meaning that one or more small ‘daughter’ bubbles were found on the rims of, or projected inside, larger ‘parent’ bubbles. The prevalence of triggering is a key unresolved question in the study of massive star formation, with important implications for extragalactic studies as well as detailed star formation physics. Quantifying triggering allows the formulation of galaxy-scale star formation rate prescriptions for use in simulations, particularly if the physical conditions in triggered regions affect the IMF, as suggested by Whitworth et al. (1994) and Dale et al. (2009).

Rahman & Murray (2010), however, observed that Galactic giant H II regions were arranged on the rims of very large (up to 100 pc
Figure 21. Examples of sites showing potential evidence of triggering. In each case the ‘heat map’ and reduced data are shown overlaid on the MWP image. Coordinates are image centres; image sizes are indicated by the zoom level (zoom). Zoom levels 1, 2 and 3 refer to images of 1° × 1°, 0.75° × 0.375 and 0.3° × 0.15, respectively. Colour figure available online.

Figure 21. Examples of sites showing potential evidence of triggering. In each case the ‘heat map’ and reduced data are shown overlaid on the MWP image. Coordinates are image centres; image sizes are indicated by the zoom level (zoom). Zoom levels 1, 2 and 3 refer to images of 1° × 1°, 0.75° × 0.375 and 0.3° × 0.15, respectively. Colour figure available online.

diameter), mid-IR bubble structures associated with the most luminous sources of free–free emission in the Galaxy, suggesting large-scale triggering driven by very massive clusters and OB associations.

However, whether these structural features are a direct result of the expanding shell’s interaction with the molecular cloud, or whether the clearing of material simply reveals the underlying turbulent cloud structure, is not yet clear. Beaumont & Williams (2010) proposed that the three-dimensional geometry of bubbles in the CP06 catalogue, when traced by molecular gas, resembles flattened rings rather than spherical shells, and they noted that this geometry could reduce the efficiency of triggering.

From results of smoothed particle hydrodynamic (SPH) simulations of the ionizing radiation from massive stars or small protoclusters, Dale & Bonnell (2011) argue that ISM bubbles are features of the turbulent nature of the molecular clouds rather than shells created by feedback from massive stars and clusters. Their simulation involves multiple clusters within a very massive molecular cloud. In this regard their simulation is more analogous to a region such as the Carina Nebula than to the kind of bubbles predominantly seen in the MWP, where the ionization is dominated by one or two stars.

Most studies to date have asked the question: ‘what percentage of bubbles show evidence for triggered star formation?’ Understanding Galactic-scale star formation in the Milky Way, however, requires answering not this question but its corollary: ‘what fraction of all (massive) star formation is triggered?’ Thompson et al. (2011) have recently made the first effort to address this question, using the Red MSX Source (RMS) data base of massive YSOs and the bubbles from CP06 and CWP07, but the major caveat of their analysis is that the Churchwell bubble catalogues are incomplete. By identifying an order of magnitude more bubbles, and providing a reliability indicator in the form of the hit rate, the MWP bubble catalogue greatly ameliorates this problem.

In addition, Oey et al. (2005) have suggested that the most convincing candidates for triggered star formation are regions where three-generation hierarchies can be established. The MWP has identified a much larger number of broken, old bubbles and small bubbles than the CP06 and CWP07 catalogues. As such, it is an excellent data set for searching for multiple hierarchies. Indeed, 29 percent
of MWP bubbles have hierarchy flags of 1 or 2, indicating bubbles that are located on or within larger bubbles and bubbles that have smaller bubbles situated on or within them.

Additional analysis to assess triggering related to MWP bubbles (using multiwavelength data sets) is ongoing and will be the subject of a subsequent paper.

6 SUMMARY AND FUTURE WORK

A new catalogue of 5106 IR bubbles has been created through visual classification via the MWP website. Bubbles in the new catalogue have been independently measured by at least five individuals, producing averaged parameters for their position, radius, thickness, eccentricity and position angle. Citizen scientists have independently rediscovered the locations of 86 per cent of the CP06 and producing averaged parameters for their position, radius, thickness, eccentricity and position angle. Citizen scientists have independently rediscovered the locations of 86 per cent of the CP06 and CPW07 bubble catalogues and 96 per cent of the Anderson et al. (2011) HII region catalogue, whilst finding an order of magnitude more objects.

The MWP bubble catalogue constitutes a resource that, in combination with other recent and ongoing Galactic Plane surveys of star formation tracers [including the Methanol Multibeam Survey (MMB), the Bolocam Galactic Plane Survey (BGPS), ATLASGAL and Herschel Infrared Galactic Plane Survey (HiGAL)], has the potential to provide sufficient statistics to address the question of how prevalent, and important, triggered star formation really is. In addition, we hope that this new resource will complement these surveys as a tracer of massive star formation on Galactic scales.

Also outlined is the creation of a ‘heat map’ of star formation activity in the Galactic plane. This online resource provides a crowdsourced map of bubbles and arcs in the Milky Way, and should enable better statistical analysis of nearby star formation sites.

Additional papers are currently being prepared to outline catalogues of ‘green knots’, dark nebulae, star clusters, galaxies and ‘fuzzy red objects’ that have also been created by the MWP’s community of citizens scientists. Similarly, we anticipate a second, refined bubble catalogue incorporating not only better data reduction techniques but also 100 000s of more bubble drawings by volunteers.

ACKNOWLEDGMENTS

This publication has been made possible by the participation of more than 35 000 volunteers on the Milky Way Project. Their contributions are acknowledged individually at http://www.milkywayproject.org/authors. We would like to thank Bob Benjamin for his helpful suggestions.

The Milky Way Project and RJS. were supported by the Leverhulme Trust. The ‘Talk’ discussion tool used in the MWP was developed at the Adler Planetarium with support from the National Science Foundation CDI grant: DRL-0941610. MSP was supported by a National Science Foundation Astronomy & Astrophysics Postdoctoral Fellowship under award AST-0901646.

CJC is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1003134. Support for the work of KS was provided by NASA through a Einstein Postdoctoral Fellowship grant number PF9-00069 issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-03060.

This work is based on observations made with the Spitzer Space Telescope, which is operated by the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

REFERENCES

Anderson L. D., Bania T. M., 2009, ApJ, 690, 706
Anderson L. D., Bania T. M., Balser D. S., Rood R. T., 2011, VizieR Online Data Catalog, 2194, 40032
Arthur S. J., Henney W. J., Mellema G., de Colle F., Vázquez-Semadeni E., 2011, MNRAS, 414, 1747
Bania T. M., Anderson L. D., Balser D. S., Rood R. T., 2010, ApJ, 718, L106
Beaumont C. N., Williams J. P., 2010, ApJ, 709, 791
Beaumont C. N., Williams J. P., Goodman A. A., 2011, preprint (arXiv:1107.5584)
Bendo G. J. et al., 2008, MNRAS, 389, 629
Benjamin R. A. et al., 2003, PASP, 115, 953
Beuther H. et al., 2011, ApJ, in press (arXiv:1112.4609)
Capiotti E. R., Kozminski J. F., 2001, PASP, 113, 677
Cardamone C. et al., 2009, MNRAS, 399, 1191
Carey S. J. et al., 2009, PASP, 121, 76
Castor J., McCray R., Weaver R., 1975, ApJ, 200, L107
Churchwell E. et al., 2006, ApJ, 649, 759 (CP06)
Churchwell E. et al., 2007, ApJ, 670, 428 (CPW07)
Churchwell E. et al., 2009, PASP, 121, 213
Dale J. E., Bonnell I., 2011, MNRAS, 414, 321
Dale J. E., Wünsch R., Whitworth A., Palouš J., 2009, MNRAS, 398, 1537
Deharveng L., Lefloch B., Kurtz S., Nadeau D., Pomarèès M., Caplan J., Zavagno A., 2008, A&A, 482, 585
Draine B. T., 2011, ApJ, 732, 100
Elmegreen B. G., Lada C. J., 1977, ApJ, 214, 725
Everett J. E., Churchwell E., 2010, ApJ, 713, 592
Flagay N., Boulanger F., Noriega-Crespo A., Paladini R., Montmerle T., Carey S. J., Gagné M., Shenoy S., 2011, A&A, 531, A51
Freyer T., Hensler G., Yorke H. W., 2003, ApJ, 594, 888
García-Segura G., Franco J., 1996, ApJ, 469, 171
Green J. A. et al., 2011, ApJ, 733, 27
Joy K. et al., 2011, Astron. Geophys., 52, 020000
Kuchar T. A., Clark F. O., 1997, ApJ, 488, 224
Lagadec E. et al., 2011, MNRAS, 417, 32
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Lintott C. J. et al., 2009, MNRAS, 399, 129
Lintott C. et al., 2011, MNRAS, 410, 166
Lockman F. J., Pisano D. J., Howard G. J., 1996, ApJ, 472, 173
McCollum B., Bruhweiler F. C., Wahlgren G. M., Eriksson M., Verner E., 2008, ApJ, 682, 1087
Meixner M. et al., 1999, ApJS, 122, 221
Mellema G., Arthur S. J., Henney W. J., Iliev I. T., Shapiro P. R., 2006, ApJ, 647, 397
Misanovic Z., Cram L., Green A., 2002, MNRAS, 335, 114
Mizuno D. R. et al., 2010, AJ, 139, 1542
Oey M. S., Watson A. M., Kern K., Walth G. L., 2005, AJ, 129, 393
Paladini R., Burgiana C., Davies R. D., Maino D., Bersanelli M., Cappellini B., Platania P., Smoot G., 2003, A&A, 397, 213
Povich M. S. et al., 2007, ApJ, 660, 346
Povich M. S. et al., 2009, ApJ, 696, 1278
Raddick M. J., Bracey G., Gay P. L., Lintott C. J., Murray P., Schawinski K., Szalay A. S., Vandenberg J., 2010, Astron. Education Rev., 9, 010103
Rahman M., Murray N., 2010, ApJ, 719, 1104
Reach W. T. et al., 2006, AJ, 131, 1479
Relaño M., Kennicutt R. C., Jr, 2009, ApJ, 699, 1125
Schuller F. et al., 2009, A&A, 504, 415
Smith A. M. et al., 2011, MNRAS, 412, 1309
Thompson M. A., Urquhart J. S., Moore T. J. T., Morgan L. K., 2011, MNRAS, 414, 1747
Watson C. et al., 2008, ApJ, 681, 1341
Watson C., Hanspal U., Mengistu A., 2010, ApJ, 716, 1478

© 2012 The Authors, MNRAS 424, 2442–2460
Monthly Notices of the Royal Astronomical Society © 2012 RAS
SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper.

Table 2. The large-bubble catalogue.
Table 3. The small-bubble catalogue.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

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