

Weather vanes and runaways: Bow shocks across the HR diagram

William J. Henney & Jorge A. Tarango Yong

Instituto de Radioastronomía y Astrofísica, Universidad Nacional Autónoma de México, Apartado Postal 3-72, 58090 Morelia, Michoacán, México

13 October 2017

ABSTRACT

Stellar bow shocks are the result of the supersonic interaction between a stellar wind and its environment. Some of these are "runaways": high-velocity stars that have been ejected from a star cluster. Others are "weather vanes", where it is the local interstellar medium itself that is moving, perhaps as the result of a champagne flow of ionized gas from a nearby H II region. We recently proposed a new two-dimensional classification scheme for the shape of such bow shocks, which we here apply to three very different observational datasets. These are: (1) mid-infrared arcs around hot main-sequence stars ($N = 227$); (2) far-infrared arcs around luminous cool stars ($N = 7$); and (3) emission-line arcs around proplyds and other young stars in the Orion Nebula ($N = 18$). We find significant differences between the three datasets: cool star bow shocks have markedly more closed wings than hot star bow shocks, while differences in the shape of the apex region are only marginally significant. The Orion Nebula arcs, on the other hand, have both significantly more open wings and significantly flatter apices than the hot star bow shocks. We discuss the implications of these differences for understanding the physics of the bow shock interaction in the three classes of sources.

1 INTRODUCTION

2 COMPARISON WITH OBSERVATIONS

Placing various classes of objects on the $R_{90}-R_c$ plane:

- LL arcs
- runaway O stars
- AGB stars

2.1 Mid-infrared arcs around early-type stars

The most extensive observational sample of stellar bow shock nebulae to date is a catalog of 709 arcs (Kobulnicky et al. 2016) detected in mid-infrared surveys of the Galactic Plane by the *Spitzer Space Telescope* (SST, Werner et al. 2004) and *Wide-field Infrared Survey Explorer* (WISE, Wright et al. 2010). These sources are believed to be powered by the winds of early-type stars, which are either moving supersonically through the interstellar medium (runaway stars, Gvaramadze & Bomans 2008), or are interacting with a local bulk flow, such as the champagne flow from a nearby H II region (weather vanes, Povich et al. 2008).

2.1.1 Automatic tracing and fitting of bow shocks

In order to study the shapes of these bow shocks, we downloaded data from the NASA/IPAC Infrared Science Archive archive¹ and extracted 4' square images in the 24 μm bandpass of the Multiband Imaging Photometer for *Spitzer* (MIPS) centered on each of the 471 Kobulnicky et al. (2016) sources that are covered by the MIPSGAL (Carey et al. 2009) survey, which includes most of the sources with Galactic longitude within $\pm 60^\circ$ of the Galactic center.

We developed a methodology for automatically tracing the arcs as follows:

1. Calculate arrays of celestial coordinates, C , for each pixel of the image.
2. Using the central source coordinates, C_0 and nominal bowshock radius, R_0 from Kobulnicky et al. (2016), construct a pixel mask that includes only those pixels with separations from the source that satisfy $\frac{1}{2}R_0 \leq |C - C_0| \leq 3R_0$. This mask will be used for all subsequent operations, which serves to help avoid confusion from the star itself and other bright sources in the field of view.
3. Define a “step-back” point, C_1 , which is located at a separation $2R_0$ from the source, but in the opposite direction from the apex of the bow shock. That is, along a position angle 180° from the nominal position angle, PA_0 , of the bow shock axis. This point is at one end of the orange line shown superimposed on the bow shock images in Figure 1.
4. Looping over a grid of 50 position angles, PA_k , within $\pm 60^\circ$ of PA_0 , estimate the location of the arc along rays cast from the step-back point, using two different methods:
 - (a) The pixel with the peak brightness, with coordinates $C_{k,\text{peak}}$ (red dots in Fig. 1).
 - (b) The mean brightness-weighted separation from C_1 , with coordinates $C_{k,\text{mean}}$ (light blue dots in Fig. 1).

For each PA_k in the grid, the calculation is performed over only those pixels that satisfy $|\text{PA}(C, C_1) - \text{PA}_k| < \frac{1}{2}\delta\theta$, where $\delta\theta = 120/50 = 2.4^\circ$, which defines a thin radial wedge from C_1 . The results are shown as red and blue dots superimposed on the

images in Figure 1. Each of the two methods, “peak” and “mean”, works better in some objects and worse in others (according to the subjective judgment of “correctly” tracing the bow shock shape). We therefore take the average by amalgamating all the $C_{k,\text{peak}}$ and $C_{k,\text{mean}}$ points into a single set, C_k , for the following steps.

5. For each of the points C_k , determine the radial separation from the central source, $R_k = |C_k - C_0|$ and the angle from the bow shock axis about the central source $\theta_k = \text{PA}(C_k, C_0) - \text{PA}_0$. These are plotted in the upper left panels of Figure 1. Note that, even though the rays are cast from the step-back point C_1 within $\pm 60^\circ$ of PA_0 , the angles θ_k are measured from the source, C_0 , which is closer to the bow shock than C_1 and therefore $|\theta_k|$ can be much larger than 60° .
6. Make our own estimate of the axial size, R_0 , of the bow shock by calculating the mean of R_k over all points C_k with $|\theta_k| \leq 10^\circ$. Note that this is distinct from the nominal value of R_0 given in the Kobulnicky et al. (2016) catalog, which was “measured by eye”. We denote by $\epsilon(R_0)$ the standard deviation of the R_k that go into calculating R_0 .
7. Estimate the radius of curvature, R_c , by fitting a circle to all those points within $\pm 45^\circ$ of the nominal axis ($|\theta_k| < 45^\circ$), but after excluding any point with $R_k < \frac{1}{2}R_m$ or $R_k > 2R_m$, where R_m is the median R_k for $|\theta_k| < 45^\circ$.
8. Determine two separate estimates, R_{90+} and R_{90-} , of the perpendicular radius, R_{90} , by taking the mean of R_k over all points C_k with $|\theta_k - 90^\circ| \leq 10^\circ$ for R_{90+} , and with $|\theta_k + 90^\circ| \leq 10^\circ$ for R_{90-} . The average of the two standard deviations of the R_k that contribute to R_{90+} and R_{90-} is denoted by $\epsilon(R_0)$.

2.1.2 Subjective evaluation of the fit quality

After these automatic steps, we subjectively evaluate the results by giving a star rating to each source:

0 stars The fitting algorithm failed for some reason.

1 star The fit was formally successful, but the results for R_c or R_0 are far removed from what a human would predict by looking at the image. For example, in the smallest bowshocks, which are only marginally resolved by Spitzer’s 6” beam, the dispersion in R_k can be a significant fraction of R_0 , in which case our algorithm tends to erroneously favor $R_c < R_0$.

2 stars The fit results are not totally outlandish, but nonetheless some problem is apparent that casts doubt on their reliability. For example, a double-shell structure to the bow shock that leads to large differences between the “peak” and “mean” methods, or point sources near to the bow shock that interfere with the tracing procedure.

3 stars A good fit, but where the dispersion in R_k and/or the asymmetry in the bow shock reduces the precision in the determination of R_c and R_0 , giving subjectively estimated uncertainties around the 20% level. An example of a 3-star fit is shown in Figure 1a.

4 stars A high quality fit, with subjectively estimated uncertainties in R_c and R_0 around the 10% level. An example of a 4-star fit is shown in Figure 1b.

5 stars The highest-quality fit, usually corresponding to large, sharply defined bow shocks, whose shape is determined with high precision. An example of a 5-star fit is shown in Figure 1c.

¹ http://irsa.ipac.caltech.edu/docs/program_interface/api_images.html

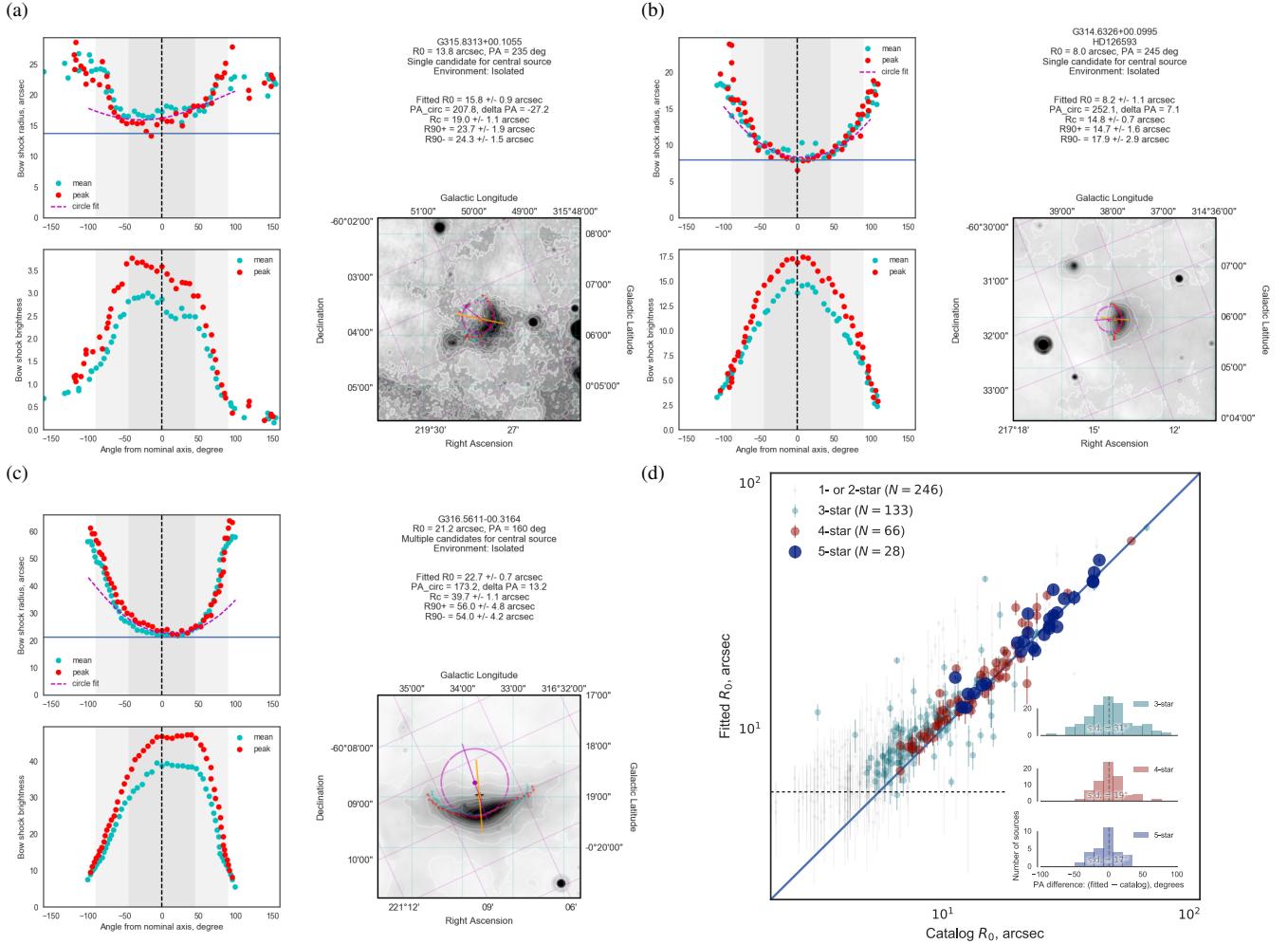


Figure 1. Examples of typical fits to the bow shock shapes of MIPSGAL sources with different star ratings: (a) K510, 3-star rating; (b) K506, 4-star rating; (c) K517, 5-star rating. Right panels of parts (a)–(c) show a 4' square 24 μm image, centered on each source. Contours are ten linearly spaced levels between the median brightness of the entire image and the maximum brightness of the bow shock arc. Grids of galactic coordinates (light blue lines, parallel to the box sides) and equatorial coordinates (tilted magenta lines) are shown. The stellar source and the bow shock axis, as determined by Kobulnicky et al. (2016) are indicated by an orange star and an orange line, respectively, where the line extends from $-2R_0$ to $+2R_0$. The automatically traced arc shapes using the “mean” and “peak” methods (see text) are shown by blue and red dots, respectively. The magenta circle shows the fit to the arc points within $\pm 45^\circ$ of the nominal bowshock axis, with the magenta dot showing the center of curvature and the magenta line showing the fitted bow shock axis, which is the line passing through the source and the center of curvature. Left panels of parts (a)–(c) show the radius measured from the source (upper panel) and brightness (lower panel) of the arc points, plotted as a function of angle θ from the nominal bow shock axis, and with the same color coding as used on the image. Angular ranges of $\theta = \pm 45^\circ$ and $\pm 90^\circ$ are shown by gray shaded boxes. In the upper panel, the R_0 value tabulated by Kobulnicky et al. (2016) is shown by a horizontal blue line. (d) Comparison of the bow shock sizes (scatter plot) and position angles (inset histograms) determined from our fits with those tabulated by Kobulnicky et al. (2016) for the MIPSGAL sources. The horizontal dotted line on the main plot shows the MIPS 24 μm point spread function FWHM of 5.5''. The standard deviation (s.d.) of the position angle differences is shown on each inset histogram.

Figure 1d compares the bow shock size, R_0 , determined by our fits (vertical axis) with the corresponding value given in the Kobulnicky et al. (2016) catalog (horizontal axis). For most sources with 3-star or higher rating, the two estimates agree to within $\pm 20\%$, but there are a small number of sources with a discrepancy of more than a factor of two. In all cases that we checked, we believe that our estimates of R_0 are more accurate than those in the catalog. It is apparent that the star ratings are correlated with the bow shock size, with larger bow shocks tending to receive higher ratings, although there is considerable overlap. In particular, most of the 1- and 2-star sources are close to the resolution limit of the MIPSGAL 24 μm images (6'', indicated by the dotted horizontal line in te figure).

In the following analysis, only those sources with a 3-star or

higher rating are used. These comprise approximately half (227 out of 471) of all the MIPSGAL arc sources. In some cases of poor and failed fits, there is nothing apparently “wrong” with the source itself, and it is likely that minor tweaks to the methodology would improve matters, but we have elected not to do so, in order to maintain a uniform methodology across all sources.

The inset of Figure 1d shows histograms of the difference between the position angle, PA_0 determined by our fits and that listed in Kobulnicky et al. (2016). Although observational uncertainties undoubtedly contribute in part, the differences are mainly due to real asymmetries in the bow shocks, especially for the 4- and 5-star sources. The Kobulnicky et al. catalog PA_0 values are mostly sensitive to the orientation of the bow shock wings, whereas our

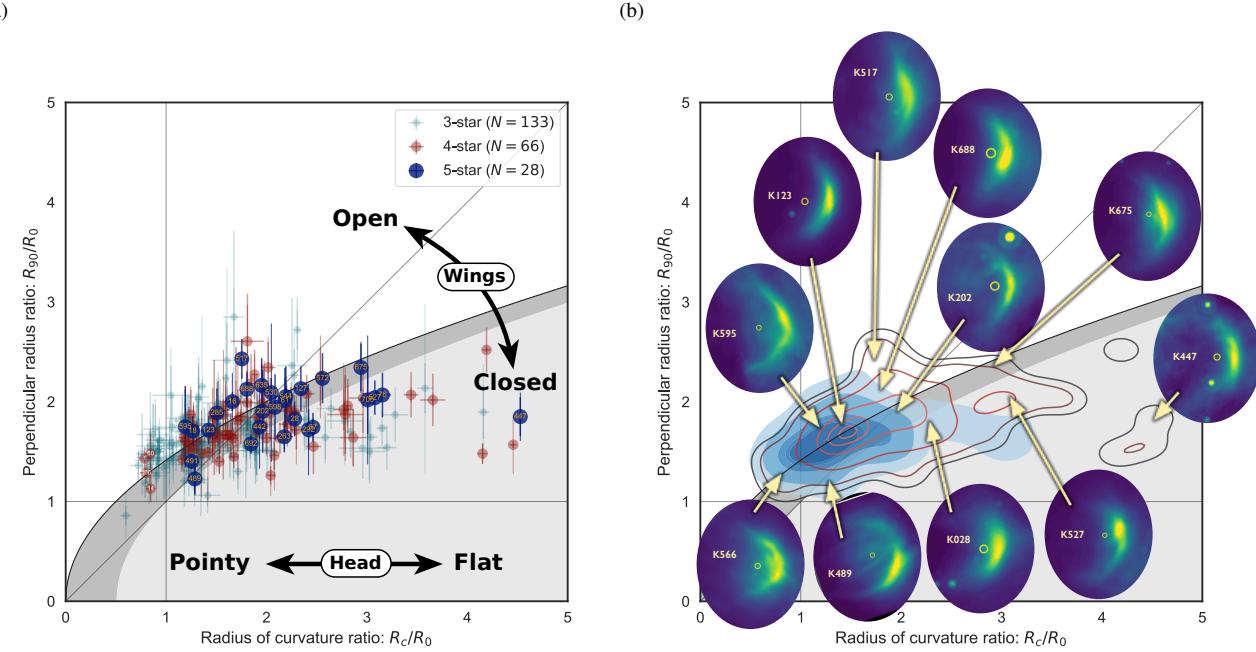


Figure 2. MIPSGAL sources on the bow shock shape diagnostic diagram of dimensionless radius of curvature versus perpendicular radius. The regions corresponding to different classes of quadrics are shown by shading (see § ??): oblate spheroids (light gray background); prolate spheroids (darker gray background); paraboloids (curved black line); hyperboloids (white background). (a) Individual sources with bow shock fit quality rating of 3-star or above. All 5-star sources plus those 4-star sources with $R_c/R_0 < 1$ are labelled with their Kobulnicky et al. (2016) catalog number. Horizontal error bars do not directly reflect the uncertainty in R_c/R_0 but are instead simply the standard deviation from the circle fit of bowshock points R_k within $\pm 45^\circ$ of the axis. Values on the vertical axis represent the average of R_{90+} and R_{90-} , with thin vertical error bars showing the difference between R_{90+} and R_{90-} , and thick vertical error bars showing the rms dispersion of R_k about these values for bow shock points within $\pm 10^\circ$ of the $+90^\circ$ and -90° directions. (b) Kernel density estimator (KDE) of the distribution for 3-star sources (blue, filled contours) and 4- plus 5-star sources (orange/brown, unfilled contours). The KDE uses an anisotropic gaussian kernel with bandwidths of 0.18×0.12 . Thumbnail images of representative 4- and 5-star sources at different points on the R_c-R_{90} plane are also shown. The angular scale of each image is indicated by a yellow circle of diameter $7.5''$, centered on the stellar source.

fitted PA_0 values are determined by the point in the bow shock head that is closest to the stellar source. For this reason, we use the catalog PA_0 values for defining the axis when measuring R_{90+} and R_{90-} . On the other hand, the fitted values of PA_0 are better correlated with the position of the bow shock's brightness peak, as is apparent in the lower left panels of Figure 1a and c.

2.1.3 OB bow shock shapes on the diagnostic plane

The derived bowshock shapes of all the 3-, 4-, and 5-star sources are shown in Figure 2 on the plane of R_c/R_0 versus R_{90}/R_0 . Panel a shows the individual points superimposed on the theoretical results for quadrics of revolution (see figure caption and § ??), while panel b shows contours of the kernel density estimator (KDE, see Leiva-Murillo & Artés-Rodríguez 2012; Scott 2015) of the two-dimensional distribution of points on the R_c-R_{90} plane. The horizontal axis corresponds to the shape of the head of the bow shock near its apex, ranging from sharper, pointier shapes with $R_c/R_0 < 1$ to flatter, snubber shapes with $R_c/R_0 \gg 1$, where it must be understood that all judgments of sharpness/flatness are with respect to the axial separation, R_0 , between the source and the bow shock apex. The vertical axis corresponds to the shape of the bow shock wings, ranging from closed ‘‘C’’ shapes for smaller values of R_{90}/R_0 to open ‘‘V’’ shapes for larger values of R_{90}/R_0 . The boundary between closed and open corresponds to the paraboloids, and is shown by the solid curved line that divides the shaded from the unshaded regions of the graph.

The KDE contours in Figure 2b show that the distribution of 3-star sources is very similar to that of 4- and 5-star sources, although the higher-rated sources are shifted slightly to the upper right. Possible reasons for this are discussed in § 2.1.5 below. The bulk of the sources are concentrated around the paraboloid line, with $1 < R_c/R_0 < 2.5$, and $1.2 < R_{90}/R_0 < 2$. But significant minorities are found in three other regions: (1) a clump with $R_c/R_0 \lesssim 1$; (2) a vertical spur towards higher R_{90} at $R_c/R_0 \approx 2$; and (3) a broad horizontal tail towards higher R_c at $R_{90}/R_0 \lesssim 2$.

2.1.4 Correlation between bow shock size and other parameters

In Figure 3 we show the distributions over all MIPSGAL bow shock sources of the bow shock size, Galactic coordinates, extinction-corrected stellar source magnitude, and dust extinction. For the bow shock size, R_0 , we use the results from our model fitting rather than the values given in the Kobulnicky et al. (2016) catalog, but the distribution is very similar, as can be seen by comparing the top-left plot of Figure 3 with Kobulnicky et al.’s Figure 8.

The catalog gives the K -band extinction, A_K , derived using the method of Majewski et al. (2011), but that assumes an intrinsic color of $(H - [4.5\mu\text{m}])_0 = +0.08$ magnitudes, which is too red if the sources are assumed to be OB stars. We therefore re-derive A_K from the catalog magnitudes combined with the Indebetouw et al. (2005) reddening law, but assuming $(H - [4.5\mu\text{m}])_0 = -0.1$ magnitudes, which is more typical of early type stars. This does not make very much difference (compare the top-right plot of our

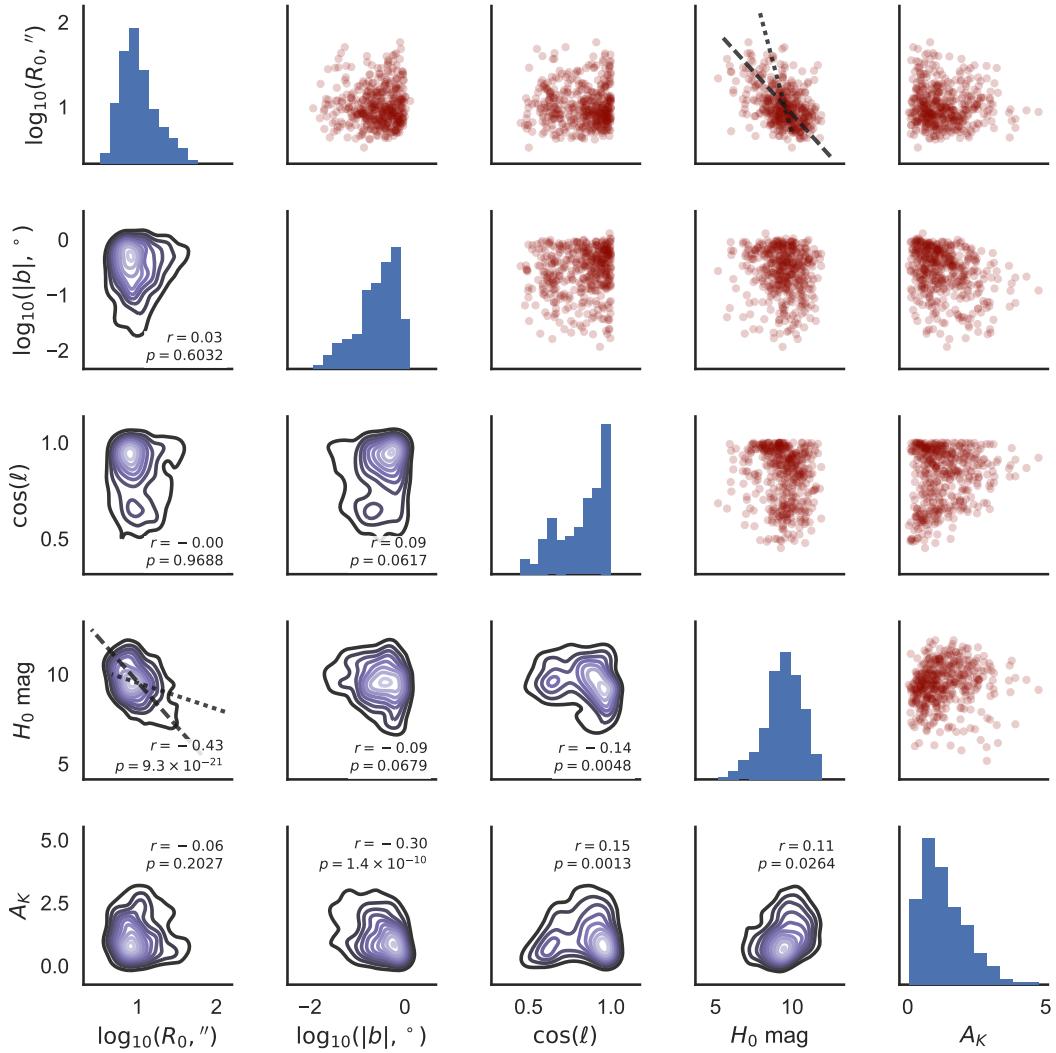


Figure 3. Matrix of pair plots that illustrate distributions of and correlations between the non-shape parameters of all MIPSGAL bow shock sources from Kobulnicky et al. (2016). Plots on the leading diagonal show histograms of the following parameters: bow shock angular size, $\log_{10} R_0$; Galactic latitude, $\log_{10} |b|$; Galactic longitude, $\cos \ell$; extinction-corrected H -band magnitude of the stellar source, H_0 ; K -band extinction, A_K . Scatter plots in the upper triangle show the joint distribution of each pair of parameters. These are repeated in the lower triangle but showing the KDEs of the joint distributions, which are annotated with the Pearson linear correlation coefficient, r , for each pair. The straight lines shown superimposed on the plots of stellar magnitude versus bowshock size correspond to toy model results for the same star at a sequence of distances (dashed lines) and a sequence of stellar luminosities at a fixed distance (dotted lines). See text for details.

Fig. 3 with Kobulnicky et al.’s Fig. 9), but it does eliminate some of the apparent negative extinctions that are found in the catalog. The same reddening law gives $A_H = 1.55A_K$, and this is used to derive extinction-corrected H -band apparent magnitudes, H_0 .

The most significant linear correlation between any pair of parameters in Figure 3 is that between bow shock size and stellar source brightness: H_0 versus $\log_{10} R_0$, with correlation coefficient $r = -0.43$. The distribution of H_0 depends on the absolute magnitude, M_H , and the distance, d , to the source. It is likely that variation in d is the more important of the two because M_H changes relatively little for main-sequence OB stars, ranging from $M_H \approx -4$ (early-O) to $M_H \approx -1.5$ (mid-B). This is because part of the increase in bolometric luminosity, L , as one ascends the main sequence is offset by an increase in the effective temperature, T_{eff} , which shifts the peak of the stellar spectrum farther away from the H band, resulting in $L_H \propto L/T_{\text{eff}}^3 \sim L^{0.53}$, where the last step uses the upper

main-sequence mass–luminosity and mass–radius scalings from Eker et al. (2015). It is true that evolved OB supergiants can be much brighter, reaching $M_H \approx -7$, but such stars are expected to be rare. Assuming a B2V star ($M_H = -2$), then the observed range $H_0 = 5\text{--}12$ corresponds to distances $d = 100\text{--}6300$ pc, and the histogram peak at $H_0 \approx 9.5$ corresponds to $d \approx 2000$ pc, which is all perfectly reasonable.

Turning now to the distribution of bow shock angular size, R_0 , this will also be affected by distance to at least some degree, since for a constant physical size the angular size will vary as $R_0 \propto d^{-1}$. For instance, if we assume that the physical size of all bow shocks is 0.1 pc and the absolute magnitude of all stars is $M_H = -2$, as above, then we find the relation $H_0 = 14.57 - 5 \log_{10} R_0$ if R_0 is measured in arcseconds. This is shown as a dashed line on the relevant panels of Figure 3 for values of R_0 that correspond to $d = 300$ pc to 8000 pc. It can be seen that this relation is in excellent agreement with the

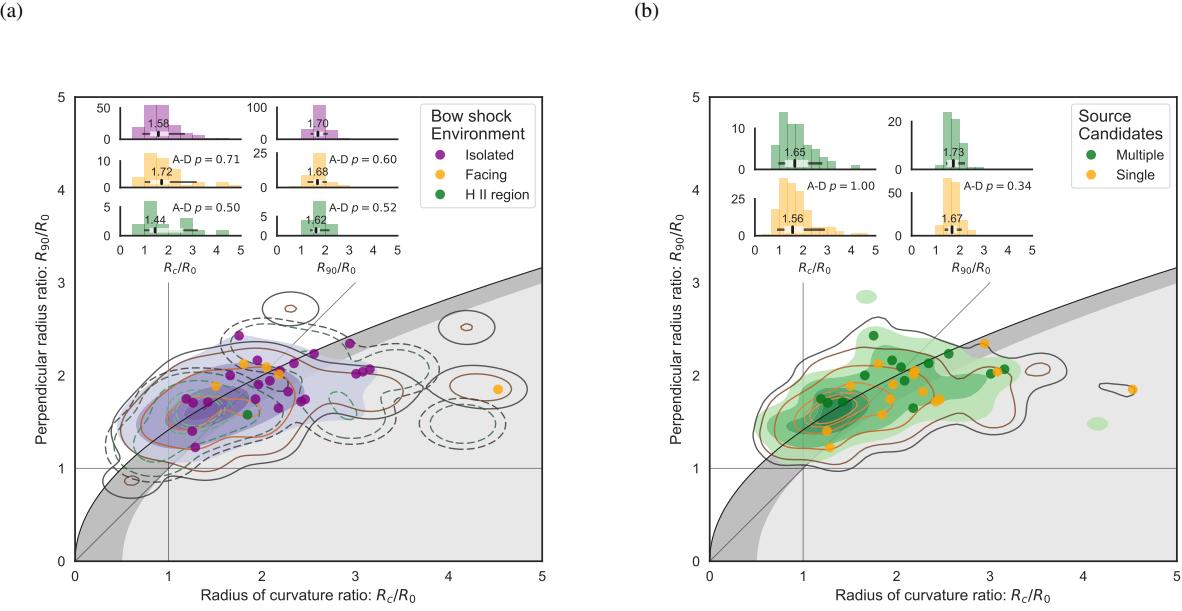


Figure 4. Example comparisons between the distribution of bowshock shapes when the sources are divided into two or more sub-samples according to the value of a categorical parameter. Contours show the KDE of the distribution of each sub-sample for all 3-, 4-, and 5-star sources, while filled circle symbols show 5-star sources only. Inset histograms show the marginal distributions on the two shape axes. (a) Source environment, divided into three sub-samples: “Isolated” (purple symbols and purple filled contours), “Facing H II region or 8 μm bright-rimmed cloud” (orange symbols and orange-brown hollow continuous contours), and “Within H II region” (green symbols and green hollow dashed contours). (b) Uncertainty in stellar source identification, divided into two sub-samples: “Multiple candidates for stellar source” (green symbols and green filled contours) and “Single candidate for stellar source” (orange symbols and orange-brown hollow continuous contours).

linear trend in the data. On the other hand, the correlation coefficient of $r = -0.43$ means in broad terms that only a fraction $r^2 \approx 20\%$ of the total variance in H_0 is “explained” by changes in R_0 , and vice versa, implying that one or both of H_0 and R_0 is only a very imperfect proxy for d . We have already seen that the spread in H_0 probably *is* mostly due to a spread in distance, rather than a spread in H -band stellar luminosity. If this is true, it follows that it is R_0 that depends only weakly on d and is more influenced by other factors.

One such factor is the stellar/environmental momentum-loss ratio, β , between the two supersonic flows that form the bow shock. All other things being equal, we have $R_0 \propto \beta^{1/2}$ for $\beta \ll 1$, as is typically the case. If the environmental flow is constant and the OB star wind has mass-loss rate \dot{M} and terminal velocity V_∞ , then $\beta \propto \dot{M}V_\infty$. Empirical and theoretical studies of hot star winds (e.g., Puls et al. 1996) imply $\dot{M}V_\infty \sim L^{1.88}R^{-1/2} \sim L^{1.80} \sim L_H^{3.40}$, where the final two steps apply only to main-sequence stars and again use the relations of Eker et al. (2015). If we assume as above that a B2V star with $M_H = -2$ has a bow shock physical size of 0.1 pc, and consider a sequence of stars with varying H -band luminosities but all at a fixed distance of $d = 2000$ pc, then we find the relation $H_0 = 8.02 - 1.49 \log_{10} R_0$. This is shown as a dotted line in the relevant panels of Figure 3 for the absolute magnitude range $M_H = -3.6$ (O6V) to $M_H = -1.5$ (B5V). It can be seen that this relation does not match the linear trend in the data, and predicts a much larger spread in R_0 over a narrow range in H_0 than is observed. This could mean one of two things: first, it may be that the range of stellar luminosities is significantly narrower than we have supposed, implying that B stars vastly outnumber O stars among the sources. Alternatively, there may be a positive correlation between the stellar luminosity and the momentum of the environmental flow, with the result that β varies less steeply with L_H than we have assumed.

That could arise if more luminous stars were preferentially found in denser environments, or, in the case of runaways, if more luminous stars tended to be faster moving.

A third factor that may influence R_0 is the inclination, i , of the bow shock axis with respect to the plane of the sky. Figure ??b shows that for $R_c > 1$, then the projected R'_0 becomes larger than the true R_0 as $|i|$ increases. This is further illustrated in Figure ?? for simple quadric shapes (spheroids, paraboloids, hyperboloids) and in Figure ?? for the wilkinoid, cantoids, and ancantoids. It can be seen that the effect is relatively modest for bow shocks with projected shapes falling in the range $1 < R_c/R_0 < 3$ and $R_{90}/R_0 < 3$, as the vast majority of our observed sources do. The increase in R_0 is no more than a factor of 2 to 3 for moderate inclinations of 30 to 60°, although it can reach a factor of 5 to 10 for some shapes at more extreme inclinations.

2.1.5 Correlation between bow shock shape and other parameters

We now investigate if the bow shock shapes of the MIPSGAL sources are correlated with any other parameters via the following methodology:

- For each of the parameters in the Kobulnicky et al. (2016) catalog, we divide the sources into two or more sub-samples, according to the value of the parameter. For quantitative parameters, such as those discussed in the previous section, we use two sub-samples of equal size, with membership determined by whether the parameter is larger or smaller than the median value. But we also use categorical parameters, such as the type of bow shock environment, or the presence/absence of 8 μm emission, in which case the sub-samples are of unequal size.
- We plot the KDEs of the two sub-samples separately on the

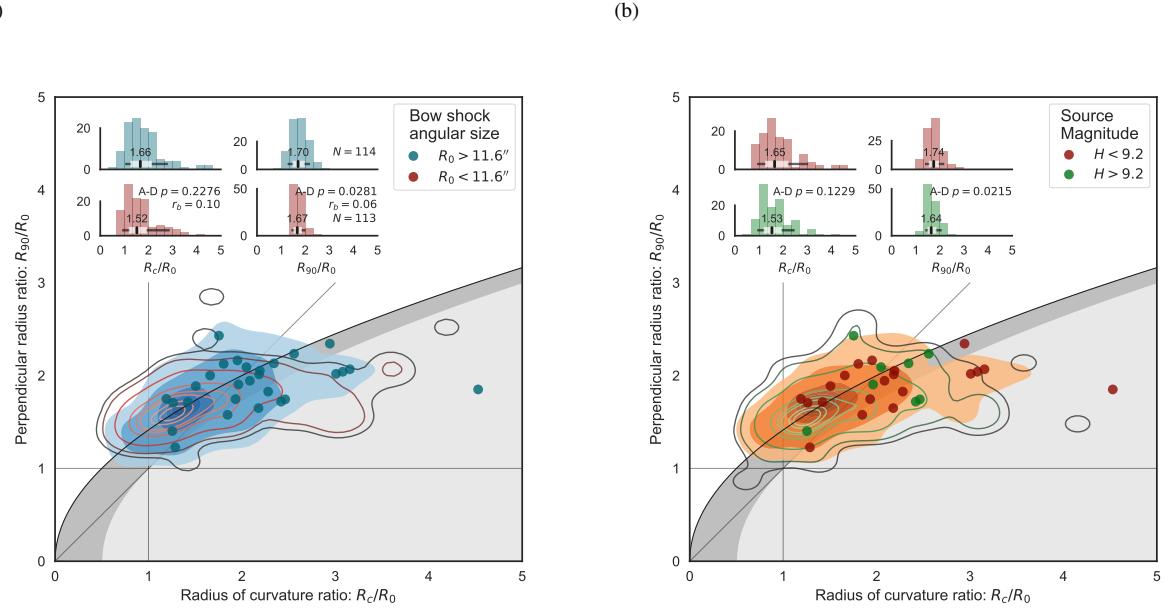


Figure 5. As Fig. 4 but for median splits of continuous parameters. (a) Bow shock angular size, R_0 , divided into two equal-sized sub-samples: large (blue symbols and blue filled contours) and small (red symbols and red hollow contours). (b) Extinction-corrected H -band magnitude of the stellar source, divided into two equal-sized sub-samples: bright (red symbols and orange filled contours) and faint (green symbols and green hollow contours).

R_c – R_{90} plane (see Figs. 5 and 4) to check for any obvious differences.

- We check if there is any statistically significant difference between the bow shock shapes of the sub-samples by applying three different non-parametric tests to R_c/R_0 and R_{90}/R_0 separately. The principal test used is the two-sample Anderson–Darling test (Anderson & Darling 1952; Scholz & Stephens 1987; Makarov & Simonova 2017), which is a general test of the *null hypothesis* that the two sub-samples are drawn from the same distribution. It returns a p -value, which is the estimated probability that the observed difference between the two sub-samples would be as large as it is purely by chance if they were all drawn from the same distribution. We consider two different thresholds for significance: $p < 0.05$ (approximately 2σ for a normal distribution) and $p < 0.003$ (approximately 3σ). We show in Appendix A that the more stringent $p < 0.003$ condition is required in order to confidently reject the null hypothesis and declare a “significant” difference between the two sub-samples, but we also consider the less strict threshold of $p < 0.05$ as an indicator of “possible” difference. We supplement the general-purpose Anderson–Darling test with two tests that probe specific features of the sample distributions: the Mann–Whitney–Wilcoxon U test (Mann & Whitney 1947), which is sensitive to differences in the central value (e.g., median) and the Brown–Forsythe test (Brown & Forsythe 1974), which is sensitive to differences in the variance, or width, of the distribution (see Appendix A for details).

As shown in detail in Table A1, there is remarkably little variation in the bow shock shape distributions as a function of most of the other parameters. Two examples in which there is *no* significant shape difference between the sub-samples are shown in Figure 4. This lack of difference is interesting because in both examples there are a priori grounds to suspect that differences might exist. The first example (Fig. 4a) is the bow shock environment, which was categorized in Kobulnicky et al. (2016) as “Isolated” (I), “Facing a

large $\mathrm{H}\,\mathrm{\alpha}$ region” (FH), “Facing a $8\,\mu\mathrm{m}$ bright-rimmed cloud” (FB), and “Within a giant $\mathrm{H}\,\mathrm{\alpha}$ region” (H), and where we have merged the FH and FB categories, labelled “Facing” in the figure.² The shapes might be expected to vary with environment because the FB and FH categories should be associated with “weather vane” interactions (Povich et al. 2008) between the stellar wind and a divergent photoevaporation flow. This is expected to give a more open bow shock than in the “runaway” case of interaction of a moving star with a static environment. In the thin-shell approximation, the predicted shapes are a cantoid for weather vanes and a wilkinoid for runaways, see § ???. The fact that no such difference is detected could be explained in one of two ways: (i) the momentum ratio β for the weather vanes could be small, since the cantoid becomes indistinguishable from the wilkinoid as $\beta \rightarrow 0$, or (ii) many of the bow shocks classified as “Isolated” might also be weather vanes rather than runaways.

The second example (Fig. 4b) divides the sources according to whether or not Kobulnicky et al. (2016) judged there to be multiple candidates for the identity of the central star that drives the bow shock. If the central star were to be misidentified in a significant number of sources, then the measured value of R_0 for those sources would be erroneous, which would increase the scatter in both R_c/R_0 and R_{90}/R_0 . The fact that no significant difference is seen in the shape distributions between sources with/without multiple candidates implies that such mistakes in identification of the central star must be rare.

We also tested all the other source parameters listed in Kobulnicky et al.’s catalog, finding no significant shape differences for sources with/without $8\,\mu\mathrm{m}$ emission, with low/high extinction, closer/farther from the Galactic plane, or closer/farther from the Galactic center. Details are given in Table A1. In all these cases, differences in mean or median R_{90}/R_0 less than 0.06 and in R_c/R_0

² Similar results are also found for the FB and FH categories separately.

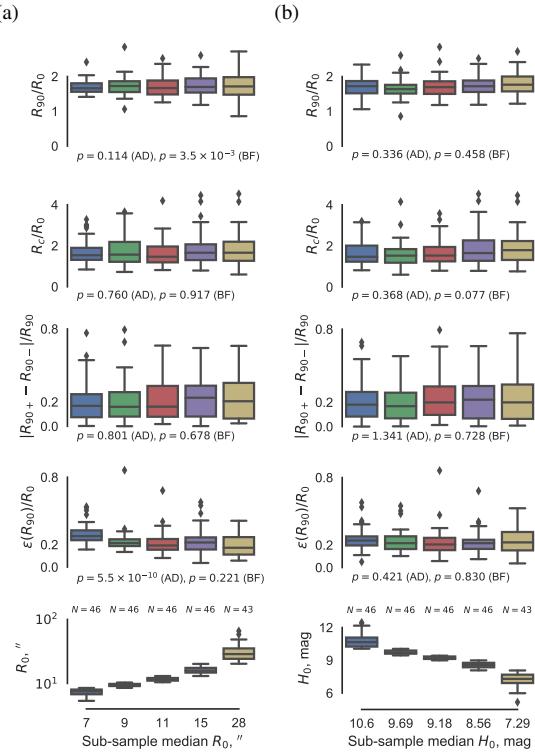


Figure 6. (a) Box plots of the distributions of bow shock shape parameters, R_c/R_0 and R_{90}/R_0 , together with a measure of the tail asymmetry, $|R_{90+} - R_{90-}|/R_{90}$, after partitioning on the bow shock size, R_0 . All 3-, 4-, and 5-star sources are sorted according to R_0 and divided into 5 non-overlapping sub-samples of roughly equal size, each labelled by their median value of R_0 in arcseconds, as illustrated in the lower panel. The colored boxes show the interquartile range, with the median indicated by a horizontal line, the 1st-to-9th interdecile range by error bars, and outliers by diamonds. A 5-sample Anderson–Darling test and Brown–Forsythe test is performed for each dependent variable, with resultant p -value given at the bottom of each panel. It is apparent that the dispersion in R_{90}/R_0 (upper panel) increases systematically with R_0 , although the central value is roughly constant. No clear systematic changes are apparent in R_c/R_0 (middle panel). (b) As (a), but partitioning on the extinction-corrected H -band magnitude, H_0 . Note that sub-samples are plotted in order of *decreasing* magnitude to allow comparison between (a) and (b), given the negative correlation between H_0 and R_0 (see Fig. 3).

less than 0.16 are found, corresponding to rank biserial correlation coefficients (a non-parametric dimensionless measure of the difference between two samples, see Appendix A) of $r_b < 0.15$, which are not significant even at the 2σ level.

The only parameters that *do* show a possible correlation with the bowshock shape are the bow shock angular size, R_0 , and the extinction-corrected magnitude of the central star, H_0 , which are illustrated in Figure 5. In both cases, the general purpose Anderson–Darling test indicates differences in the sub-sample distributions of R_{90}/R_0 at the 2σ level. It is apparent from the histograms shown in the right-hand inset graphs of Figure 5a that in the case of the small/large R_0 sub-samples it is the width rather than the central tendency of the distributions that is different, which is confirmed by the more specific Brown–Forsythe test, which indicates a difference between the sub-sample dispersions at the 3σ level. In the case of the faint/bright H_0 sub-samples (Fig. 7b), it is less clear what feature of the distributions differ.

In order to investigate these effects in more detail and look

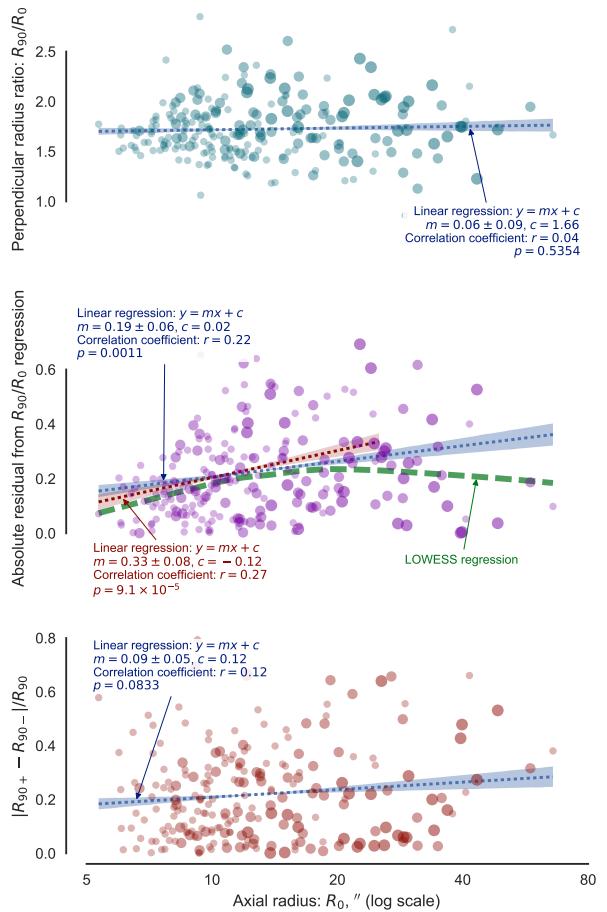


Figure 7. Regression analysis of bow shock wings shape versus size. The top panel shows a linear fit to R_{90}/R_0 versus $\log_{10} R_0$, which is essentially flat and is consistent with no correlation between the two quantities. On the other hand, in the central panel, which shows the absolute values of the residuals from this fit, it is apparent that the dispersion about the average value of the bow shock shape parameter R_{90}/R_0 increases with bow shock size. A linear fit to all the data is shown by the blue dotted line, while a locally weighted regression (LOWESS; Cleveland 1979; Cleveland & Devlin 1988) is shown by the heavy green dashed line. The negative curvature of the LOWESS curve suggests that the slope of the linear fit is a compromise between a steeper slope for smaller bow shocks and a flat (or negative) slope for the largest bow shocks. We therefore repeat the linear fit, but after excluding sources with $R_0 > 25''$, with results shown by the red dotted line. The bottom panel shows the bow shock wings asymmetry parameter, $|R_{90+} - R_{90-}|/R_{90}$, together with a linear fit. Symbol size indicates the star rating of each source, from 3-star (smallest) to 5-star (largest).

for systematic trends, we divide the independent parameter (R_0 or H_0) into 5 rather than 2 equal-sized sub-samples, with results shown as box plots in Figure 6. A systematic increase with R_0 in the dispersion of R_{90}/R_0 (upper panel of Fig. 6a) is apparent from both the interquartile range (colored boxes) and interdecile range (error bars). As a check on whether observational uncertainties might be contributing to this trend, the third and fourth rows of box plots show the statistics for, respectively, the fractional asymmetry of the bowshock wings and the standard deviation, $\epsilon(R_{90})$, of the individual points on the bow shock that go into determining R_{90} (see step 8 of the tracing/fitting methodology described in § 2.1.1). It can be seen

that neither of these quantities tends to increase with R_0 , and in fact there is a significant *decrease* in $\epsilon(R_{90})$ with R_0 . This implies that the increase with angular size of the diversity of bow shock wing shapes is real, and not due to observational uncertainties. On the other hand, the fact that no clear trends are evident as a function of source magnitude H_0 (Fig. 6b) leads us to suspect that the $p < 0.05$ result obtained for the 2-sample Anderson–Darling test was a *false positive*, which is also supported by the analysis in Appendix A and Figure A1.

So, the only correlation that survives scrutiny is the increase in dispersion of R_{90}/R_0 with increasing R_0 , which is finally presented in the greatest detail in Figure 7, where we eschew any binning and perform linear regressions of the individual sources as a function of $\log_{10} R_0$. As expected from the previous 2-sample and 5-sample analyses, the regression of R_{90}/R_0 is almost flat, with a low correlation coefficient, whereas a regression of the absolute residuals to this fit against $\log_{10} R_0$ shows a significant positive slope. This increase in dispersion of shape with angular size is only evident for $R_0 < 30''$, becoming flat at the largest sizes (see middle panel of Fig. 7).

As mentioned in § 2.1.3 there is also a shape difference between the 3-star sources and the 4/5-star sources (see Fig. 2b). The p -values of statistical tests (see Table A1) indicate that this is much more significant than any correlation with the other parameters discussed in this section ($p < 10^{-4}$ for R_{90}/R_0 and $p < 10^{-5}$ for R_c/R_0). This means that it cannot be simply due to the tendency of the higher-rated sources to have larger angular sizes. However, the subjective nature of the star ratings makes this result hard to interpret.

2.2 Far-infrared arcs around late-type stars

We obtain a second sample of bow shocks from a far-infrared survey (Cox et al. 2012) of circumstellar shells around asymptotic giant branch (AGB) stars and cool supergiants, obtained as part of the *Herschel* MESS (Mass-loss of Evolved StarS) program (Groenewegen et al. 2011). The survey sources were divided into four classes by Cox et al., according to the overall shape of the extended emission at 70 μm : “fermata ⚜”, or arcs (Class I); “eyes ⚞” (Class II); “rings ⚠” (Class III), and “irregular ⚡” (Class IV). Of these, only the Class I sources clearly correspond to bow shocks, which represent 22 out of 50 total sources detected with extended emission. For our sample, we select the 16 Class I sources where the shape parameters can be reliably measured, and which are shown in Figures 8 and 9. The remaining six Class I sources (Fig. 1 of Cox et al. 2012) are too asymmetrical or irregular to reliably determine R_c and R_{90} , and are therefore excluded from our sample.

In most sources, the bow shock shell is most clearly visible in the 70 μm band, with the exception being CW Leo where the 180 μm band is used instead. In several sources, the shell is split into multiple filaments, with the clearest example being α Ori (Betelgeuse), shown in Figure 8(a). For such sources, we take the outer envelope of the filaments as the bow shock arc. Given this complication, and the fact that the angular resolution relative to the bow shock size is much better than in the MIPSGAL sources, we judge that the arc tracing is best performed by eye and this is carried out using the SAOImage DS9 FITS viewer (Joye & Mandel 2003), with results shown as small red open circles in Figures 8 and 9. Subsequently, the bow shock parameters R_0 , R_{90} , and R_c are determined by circle fits, as in steps 5–8 of § 2.1.1, with results shown in magenta on the figures.

Figure 10 compares the distributions of bow shock shapes between the Herschel (RSG/AGB) and MIPSGAL (OB star) samples.

The Anderson–Darling test gives a highly significant difference between the R_{90}/R_0 ratio distributions of the two samples, but only a marginally significant difference between the R_c/R_0 ratio distributions (detailed results are shown in Table A1). The Herschel sources show considerably smaller R_{90} ratio (median $R_{90}/R_0 \approx 1.4$, implying bow shock wings that are more closed than in the MIPSGAL sources (median $R_{90}/R_0 \approx 1.7$). For the radius of curvature ratio, there is only a slight difference in average values: median $R_c/R_0 \approx 1.5$ for Herschel versus ≈ 1.6 for MIPSGAL, which is not statistically significant (rank biserial $p = 0.29$). On the other hand, the dispersion in R_c/R_0 is four times smaller for the Herschel sample, which is marginally significant (Brown–Forsythe $p = 0.015$). In particular, the MIPSGAL sample shows a significant minority of very flat-nosed shapes ($R_c/R_0 > 2$), but these are absent in the Herschel sample. The same can be seen directly by comparing the RSG/AGB bow shock shapes in Figures 8 and 9 with the representative OB bow shocks of Figure 2b. The MIPSGAL sources K489 and K123 have shapes that are similar to specific Herschel sources (UU Aur and R Hya, respectively), while sources such as K447 (extreme flat head) or K517 (extreme open wings) have no analog among the Herschel sources.

2.3 Stationary emission line arcs in M42

Mention future JWST observations with 0.85'' resolution at 22 μm .

3 CONCLUSIONS

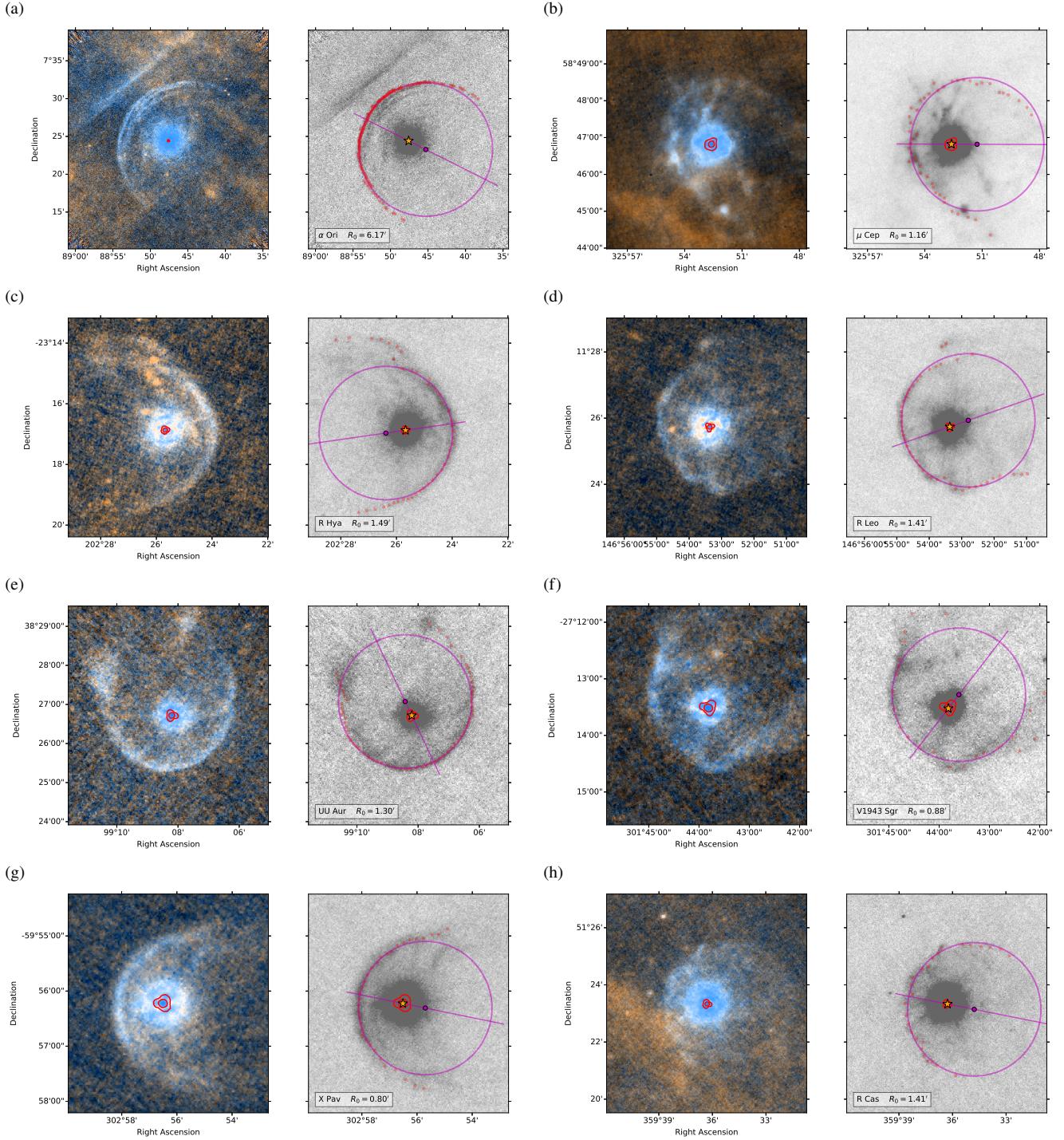


Figure 8. RSG and AGB shells observed with Herschel. Selected Class I sources (“fermata-like”) from the MESS survey Cox et al. (2012), where the arc structure is particularly clear and symmetrical. Left panels show PACS 70 μm surface brightness (blue) and 160 μm (orange). Right panels show tracing of the bow shock arc (red symbols) and circle fit (magenta lines and symbols) superimposed on a low-contrast image of the 70 μm surface brightness. (a) α Ori. (b) μ Cep. (c) R Hya. (d) R Leo. (e) UU Aur. (f) V1943 Sgr. (g) X Pav. (h) R Cas.

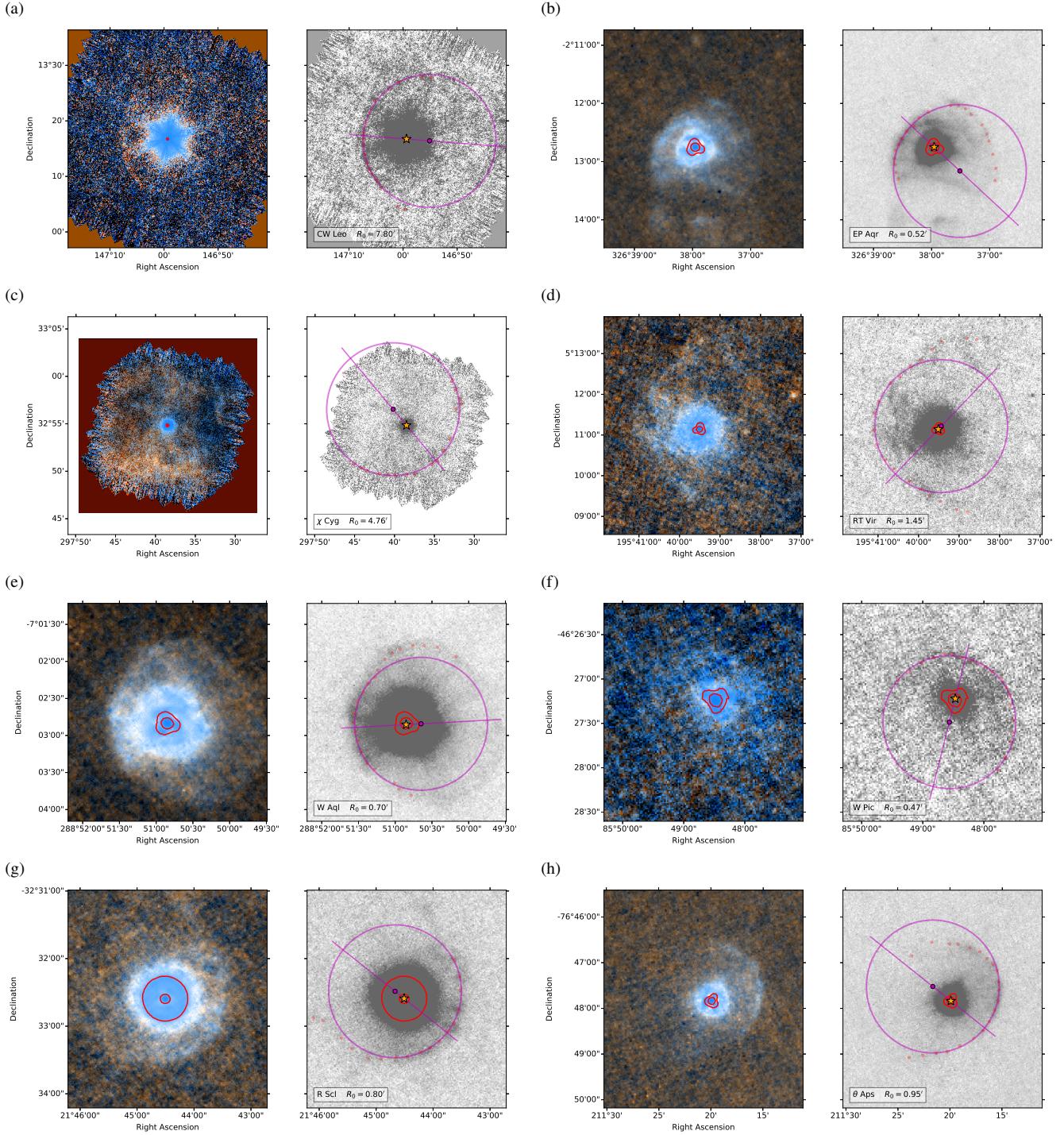


Figure 9. As Fig 8, but for a further 8 Class I sources from Cox et al. (2012) where the arc structure is more diffuse, weak, and/or asymmetric. (a) CW Leo. For this source only, the right-hand panel shows a grayscale image of the 160 μ m rather than 70 μ m emission. (b) EP Aqr. (c) χ Cyg. (d) RT Vir. (e) W Aql. (f) W Pic. (g) R Scl. (h) θ Aps.

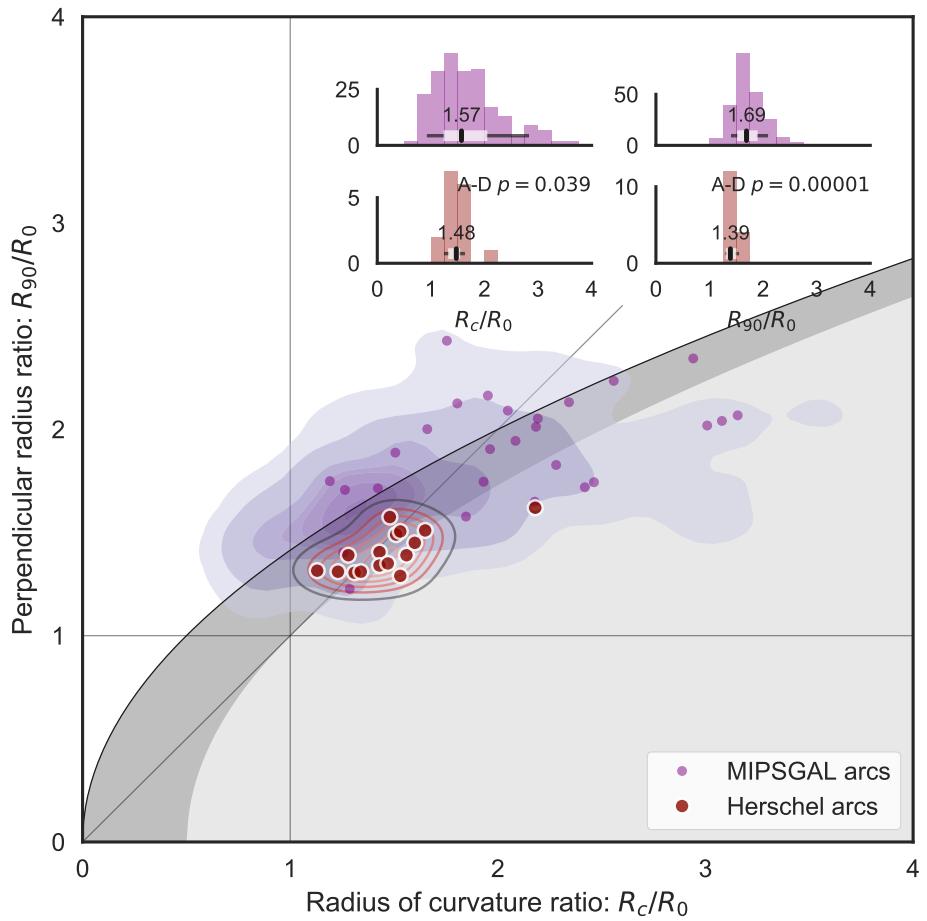


Figure 10. Comparison of bow shock shapes between Herschel RSG/AGB arcs (large red symbols and hollow red contours) and MIPSGAL OB star arcs (small purple symbols and filled purple contours). Other details of the plots are as in Figs. 2 and 4, except that the diagnostic plots are zoomed in slightly to show more detail, and a smaller KDE smoothing bandwidth of 0.12×0.08 is used.

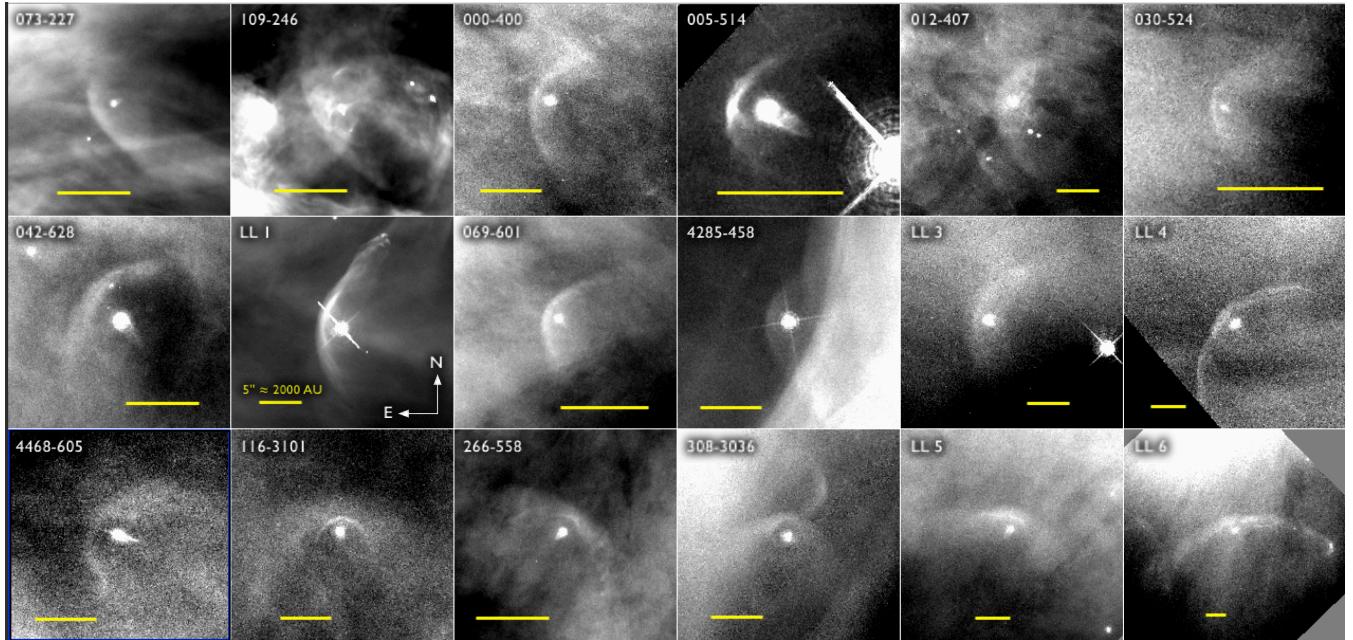


Figure 11. Stationary bow shock arcs in the Orion Nebula.

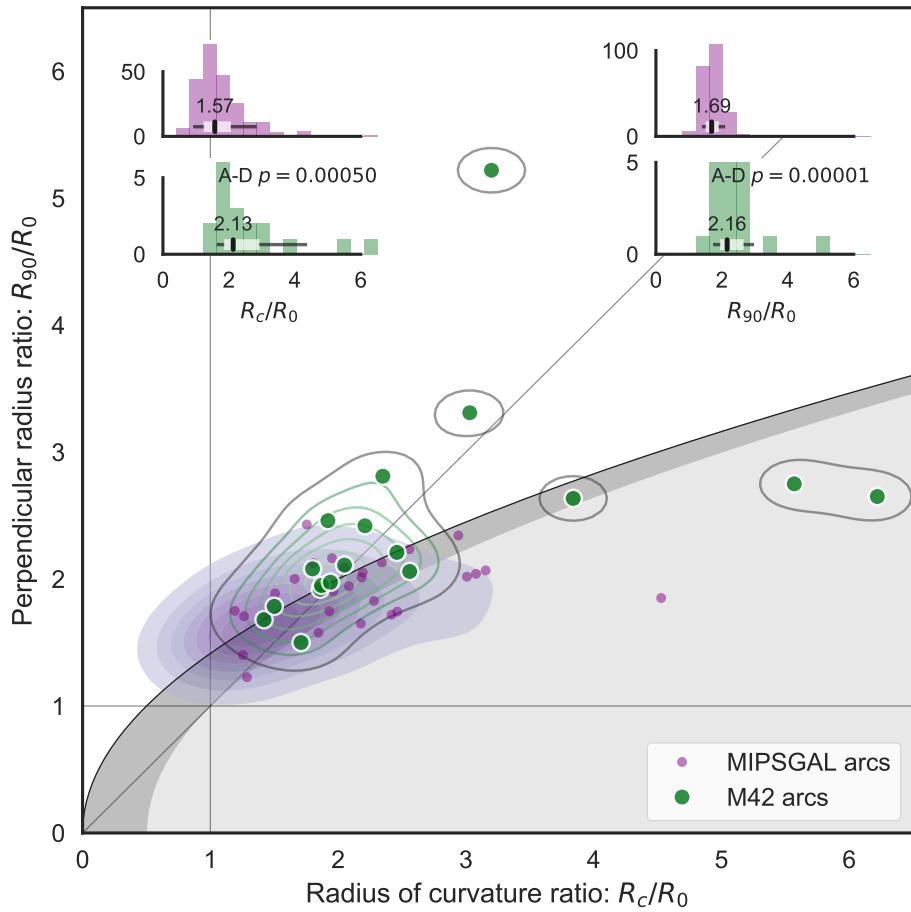


Figure 12. Comparison of Orion with OB stars. Other details of the plots are as in Figs. 2 and 4, except that the diagnostic plots are zoomed out to encompass larger ratio values, and a larger KDE smoothing bandwidth of 0.3×0.2 is used.

REFERENCES

- Anderson T. W., Darling D. A., 1952, Ann. Math. Statist., 23, 193
- Brown M. B., Forsythe A. B., 1974, Journal of the American Statistical Association, 69, 364
- Carey S. J., et al., 2009, PASP, 121, 76
- Cleveland W. S., 1979, Journal of the American Statistical Association, 74, 829
- Cleveland W. S., Devlin S. J., 1988, Journal of the American Statistical Association, 83, 596
- Cohen J., 1988, Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Routledge
- Colquhoun D., 2014, Royal Society Open Science, 1
- Cox N. L. J., et al., 2012, A&A, 537, A35
- Cureton E. E., 1956, Psychometrika, 21, 287
- Delaney H. D., Vargha A., 2002, Psychological Methods, 7, 485
- Eker Z., et al., 2015, AJ, 149, 131
- Goodman S. N., 1999a, Ann Intern Med, 130, 995
- Goodman S. N., 1999b, Ann Intern Med, 130, 1005
- Groenewegen M. A. T., et al., 2011, A&A, 526, A162
- Gvaramadze V. V., Bomans D. J., 2008, A&A, 490, 1071
- Head M. L., Holman L., Lanfear R., Kahn A. T., Jennions M. D., 2015, PLOS Biology, 13, 1
- Indebetouw R., et al., 2005, ApJ, 619, 931
- Joye W. A., Mandel E., 2003, in Payne H. E., Jedrzejewski R. I., Hook R. N., eds, Astronomical Society of the Pacific Conference Series Vol. 295, Astronomical Data Analysis Software and Systems XII. p. 489
- Kobulnicky H. A., et al., 2016, ApJS, 227, 18
- Leiva-Murillo J. M., Artés-Rodríguez A., 2012, Pattern Recogn. Lett., 33, 1717
- Majewski S. R., Zasowski G., Nidever D. L., 2011, ApJ, 739, 25
- Makarov A. A., Simonova G. I., 2017, Journal of Mathematical Sciences, 221, 580
- Mann H. B., Whitney D. R., 1947, The Annals of Mathematical Statistics, 18, 50
- Povich M. S., Benjamin R. A., Whitney B. A., Babler B. L., Indebetouw R., Meade M. R., Churchwell E., 2008, ApJ, 689, 242
- Puls J., et al., 1996, A&A, 305, 171
- Ruscio J., 2008, Psychol Methods, 13, 19
- Scholz F. W., Stephens M. A., 1987, Journal of the American Statistical Association, 82, 918
- Scott D. W., 2015, Multivariate density estimation: theory, practice, and visualization, 2nd edn. John Wiley & Sons
- Sellke T., Bayarri M. J., Berger J. O., 2001, The American Statistician, 55, 62
- Werner M. W., et al., 2004, ApJS, 154, 1
- Wright E. L., et al., 2010, AJ, 140, 1868

Table A1. Results of all statistical tests performed on observed bow shock shape parameters. Significant correlations are shown in **bold**, marginally significant correlations in *italic*

Comparison (1)	Dependent Variable (2)	Mean $\langle A \rangle$ (3)	Std. Dev. σ_A (5)	Obs. Disp. $\langle \epsilon_A \rangle$ (7)	s.e.m. $(\sigma/\sqrt{n})_A$ (9)	Effect sizes r_b (11)	Cohen d (12)	σ_A/σ_B (13)	Non-parametric test p -values Anderson–Darling (14)	Rank biserial (15)	Brown–Forsythe (16)
<i>Median split of continuous independent variables</i>											
Faint/bright	R_{90}/R_0	1.68	1.77	0.27	0.32	0.23	0.24	0.025	0.030	0.17	0.31
H magnitude	$\Delta R_{90}/R_{90}$	0.18	0.20	0.16	0.16			0.015	0.015	0.070	0.093
$n_A = n_B = 113$	R_c/R_0	1.66	1.92	0.63	1.05	0.097	0.078	0.059	0.098	0.12	0.30
Low/high bow shock size, R_0	R_{90}/R_0	1.71	1.74	0.25	0.34	0.26	0.21	0.024	0.031	0.061	0.11
$\Delta R_{90}/R_{90}$	0.18	0.20	0.16	0.17			0.015	0.015	0.091	0.18	
$n_A = n_B = 113$	R_c/R_0	1.77	1.80	0.98	0.76	0.11	0.062	0.092	0.071	0.10	0.043
Low/high extinction, A_K	R_{90}/R_0	1.70	1.74	0.27	0.32	0.23	0.24	0.025	0.030	0.040	0.13
$\Delta R_{90}/R_{90}$	0.19	0.20	0.14	0.18			0.013	0.017	-0.039	0.057	
$n_A = n_B = 113$	R_c/R_0	1.73	1.85	0.82	0.92	0.091	0.085	0.077	0.086	0.082	0.14
Low/high $ b $	R_{90}/R_0	1.72	1.72	0.33	0.26	0.23	0.23	0.031	0.024	0.020	0.0080
$\Delta R_{90}/R_{90}$	0.19	0.19	0.16	0.16			0.015	0.015	0.0090	0.021	
$n_A = n_B = 113$	R_c/R_0	1.71	1.86	0.73	0.99	0.085	0.091	0.068	0.092	0.069	0.18
High/low $\cos \ell$	R_{90}/R_0	1.73	1.71	0.28	0.32	0.24	0.22	0.024	0.034	-0.049	-0.093
$\Delta R_{90}/R_{90}$	0.18	0.20	0.16	0.17			0.013	0.018	0.054	0.12	
$n_A, n_B = 137, 90$	R_c/R_0	1.81	1.75	0.95	0.74	0.090	0.084	0.081	0.078	0.	-0.064
<i>Categorical independent variables</i>											
Environment:	R_{90}/R_0	1.74	1.69	0.28	0.34	0.24	0.22	0.022	0.053	-0.070	-0.14
Isolated vs Facing	$\Delta R_{90}/R_{90}$	0.19	0.20	0.16	0.17			0.012	0.027	-0.019	0.034
$n_A, n_B = 170, 41$	R_c/R_0	1.76	1.85	0.85	0.90	0.087	0.083	0.066	0.14	0.042	0.11
Environment:	R_{90}/R_0	1.74	1.68	0.28	0.31	0.24	0.23	0.022	0.077	-0.13	-0.19
$\Delta R_{90}/R_{90}$	0.19	0.18	0.16	0.14			0.012	0.034	-0.048	-0.095	
$n_A, n_B = 170, 16$	R_c/R_0	1.76	1.91	0.85	0.96	0.087	0.11	0.066	0.24	0.024	0.17
Single/multiple source candidate	R_{90}/R_0	1.71	1.76	0.29	0.32	0.23	0.24	0.022	0.041	0.074	0.18
$\Delta R_{90}/R_{90}$	0.18	0.21	0.16	0.16			0.013	0.021	0.093	0.14	
$n_A, n_B = 167, 60$	R_c/R_0	1.77	1.83	0.83	0.99	0.090	0.080	0.064	0.13	0.027	0.076
With/without 8 μ m emission	R_{90}/R_0	1.73	1.72	0.29	0.30	0.22	0.24	0.043	0.022	-0.042	-0.044
$\Delta R_{90}/R_{90}$	0.20	0.19	0.21	0.15			0.031	0.011	0.021	-0.11	
$n_A, n_B = 45, 182$	R_c/R_0	1.71	1.80	0.60	0.93	0.091	0.087	0.089	0.069	-0.012	0.10
3-star vs (4+5)-star	R_{90}/R_0	1.66	1.81	0.29	0.29	0.25	0.22	0.025	0.029	0.33	0.53
$\Delta R_{90}/R_{90}$	0.19	0.19	0.16	0.16			0.014	0.017	-0.0070	0.0090	
$n_A, n_B = 133, 94$	R_c/R_0	1.63	2.00	0.91	0.76	0.11	0.061	0.079	0.079	0.39	0.43
<i>Intercomparison with other datasets</i>											
MIPS vs Orion	R_{90}/R_0	1.72	2.42	0.30	0.81			0.020	0.19	0.70	1.9
$n_A, n_B = 227, 18$	$\Delta R_{90}/R_{90}$	0.19	0.70	0.16	0.53			0.011	0.15	0.69	2.6
	R_c/R_0	1.78	2.64	0.87	1.30			0.058	0.31	0.52	0.94
MIPS vs RSG	R_{90}/R_0	1.72	1.41	0.30	0.10			0.020	0.025	-0.72	-1.1
$n_A, n_B = 227, 16$	$\Delta R_{90}/R_{90}$	0.19	0.15	0.16	0.08			0.011	0.024	-0.072	-0.27
	R_c/R_0	1.78	1.48	0.87	0.23			0.058	0.057	-0.16	-0.36

Table A1 – continued Results of all statistical tests performed on observed bow shock shape parameters.

Description of columns: (Col. 1) How the two A/B source sub-samples are defined, also giving the size of each sub-sample, n_A and n_B . (Col. 2) Dependent variable whose distribution is compared between the two sub-samples. (Cols. 3–6) Mean and standard deviation, σ , of the dependent variable for each of the two sub-samples. (Cols. 7–8) Mean over each sub-sample of the observational dispersion (ϵ , standard deviation) of radii that contribute to the dependent variable for each individual source, as in steps 6 and 8 of § 2.1.1. Note that in the case of R_c , this is $\epsilon(R_0)$, and so is not a direct measure of the observational uncertainty in R_c . (Cols. 9–10) “Standard error of the mean” (s.e.m.) of the dependent variable for each of the two sub-samples. (Cols. 11–13) Standardized “effect sizes”, which are dimensionless measures of the difference in the distribution of the dependent variable between the two sub-samples. (Col. 11) Rank biserial correlation coefficient (Cureton 1956), which is obtained by considering all $n_A n_B$ pair-wise comparisons of the dependent variable between a source in sub-sample A and a source in sub-sample B. It is the difference between the fraction of such comparisons “won” by sub-sample A and those “won” by sub-sample B, and thus may vary between -1 and +1. (Col. 12) Cohen’s d , which is a dimensionless mean difference: $d = (\langle A \rangle - \langle B \rangle)/\sigma_{\text{pool}}$, where $\sigma_{\text{pool}} = (n_A \sigma_A^2 + n_B \sigma_B^2)^{1/2}/\sqrt{n_A + n_B}$ is the pooled standard deviation. (Col. 13) Ratio of standard deviations between the two sub-samples. (Cols. 14–16) Probabilities (p -values) of the two sub-samples being as different as observed if they were to be drawn from the same population, according to three different non-parametric tests. (Col. 14) Anderson–Darling 2-sample test, which is a general test of similarity between two distributions that is designed to retain sensitivity to differences in the tails of the distributions. (Col. 15) Mann–Whitney–Wilcoxon U test (Mann & Whitney 1947), which is sensitive (Col. 16) Brown–Forsythe test for equality of variance (Brown & Forsythe 1974)

APPENDIX A: DISTRIBUTION OF P-VALUES FOR ALL CORRELATIONS TESTED

Results from all the statistical tests discussed in § 2.1.5 are given in Table A1.

The p -values are the probability of finding a difference between two populations as large as (or larger than) what is observed, *given* that there is no difference in the underlying distribution from which the two populations are drawn (that is, given that the null hypothesis is true). However, what we really want to know is something else: the probability that the null hypothesis is true, *given* the observations. That is, the probability, α , of a *false positive*, also known as the *Type I error rate*. The common mistake of conflating these two definitions is known as the “ p -value fallacy” (Goodman 1999a), or “the error of the transposed conditional”, as discussed in detail by Colquhoun (2014). It is possible to derive α from p using Bayes’ theorem (e.g., Goodman 1999b), but that requires an estimate of the prior probability of the null hypothesis, independent of the observations. Alternatively, it is also possible to find a lower bound on α from a frequentist approach (Sellke et al. 2001):

$$\alpha(p) \geq \left[1 - (e p \ln p)^{-1} \right]^{-1} \quad \text{valid for } p < 1/e. \quad (\text{A1})$$

This is the approach we adopt here, which also numerically coincides with the Bayesian approach for the case where the prior probability of the null hypothesis is 0.5. The reason that this is only a lower limit for α is that if we have overwhelming a priori evidence that the null hypothesis is true (for instance, from previous empirical studies, or because it follows from a well-supported theory), then a Bayesian

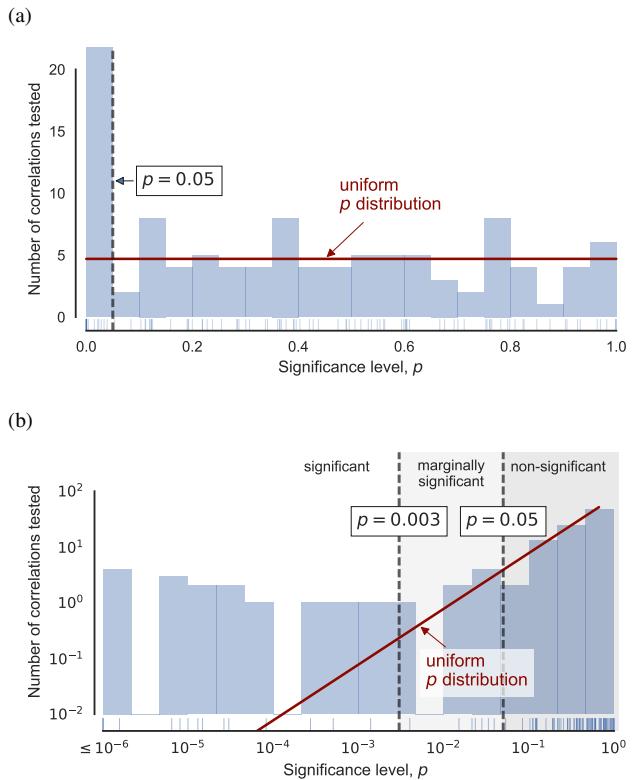


Figure A1. Histogram of p -values for all non-parametric 2-sample tests listed in Table A1. (a) Uniformly spaced linear bins and linear vertical axis. (b) Uniformly spaced logarithmic bins and logarithmic vertical axis, with all values $p \leq 10^{-6}$ included in the leftmost bin. Short thin vertical lines above the horizontal axis show the individual value. The thick vertical dashed lines show the traditional threshold values for significance: $p = 0.003$ ($\approx 3\sigma$) and $p = 0.05$ ($\approx 2\sigma$). The red solid line shows the uniform distribution of p -values that would be expected if the null hypothesis were always true, that is, if no significant correlations existed.

calculation would give a much higher value of α than (A1) does. In our case, however, we have no strong reasons for favoring any of the null hypotheses, so it is reasonable to assume α is close to the lower limit given in (A1).

In order to choose a threshold p -value that counts as a “*significant*” result, one then needs to balance the risks of false positives against the risks of *false negatives*. The false negative probability, β , also known as *Type II error rate*, is the probability of failing to reject an untrue null hypothesis. That is, in the context of this paper, it is the probability of failing to detect a real difference between two sub-samples, or a real correlation between two variables. The complementary probability, $1 - \beta$, is known as the *statistical power* or sensitivity of the test. The value of β depends on three factors:

1. The *effect size*, which is a measure of the magnitude of the difference in a dependent variable between two sub-samples, or the degree of correlation between two continuous variables. For the two sub-sample case, it is common to use a standardised mean difference, such as Cohen’s d statistic (Cohen 1988): $d = (\bar{X}_A - \bar{X}_B)/s$, where \bar{X}_A , \bar{X}_B are the means of the dependent variable X for samples A and B, while s is the pooled standard deviation of X . For the case of two continuous variables, the Pearson linear correlation coefficient, r , can be used. In both

Table A2. Lower bounds on the Type I error rate, α

Correlation	<i>p</i>	α
<i>Quantitative</i>		
R_c vs R_0	0.0054	0.0712
R_{90} vs R_0	0.0001	0.0025
R_c vs H_0	0.1229	0.4119
R_{90} vs H_0	0.0215	0.1833
R_c vs A_K	0.19	0.4617
R_{90} vs A_K	0.63	1
R_c vs $ b $	0.19	0.4617
R_{90} vs $ b $	0.31	0.4967
R_c vs $\cos(\ell)$	1.00	1
R_{90} vs $\cos(\ell)$	0.36	0.4999
<i>Categorical</i>		
R_c vs Facing	0.71	1
R_c vs H II	0.50	1
R_{90} vs Facing	0.60	1
R_{90} vs H II	0.52	1
R_c vs Multiple	1.00	1
R_{90} vs Multiple	0.34	0.4993
R_c vs 8 μm	0.82	1
R_{90} vs 8 μm	0.6	1
<i>Other datasets</i>		
R_c vs Herschel	0.074	0.3437
R_{90} vs Herschel	0.00052	0.0106
R_c vs M42	0.00048	0.0099
R_{90} vs M42	0.0000105	0.0003

cases, rules of thumb have been developed (Ruscio 2008) for classifying an effect as “large” ($d > 0.8$, $r > 0.4$) or “small” ($d < 0.2$, $r < 0.1$). Alternatively, non-parametric statistics can be used, such as the A measure of stochastic superiority (Delaney & Vargha 2002).

Obviously, this depends on the *effect size*, which is the

All astronomical data analysis is *post hoc* analysis, since the universe was not set up to test a particular hypothesis (as far as we know). It is therefore important to guard against the “multiple comparisons problem”, whereby seemingly significant correlations are found where none really exist, simply by virtue of the large number of tests that were carried out.

Under the more conservative Holm–Bonferroni method, only comparisons with $p < 0.001$ would be significant.

The p-curve (Head et al. 2015)