





The Formation of a Stellar Association in the NGC 7000/IC 5070 Complex: Results from Kinematic Analysis of Stars and Gas

Michael A. Kuhn¹ , Lynne A. Hillenbrand¹, John M. Carpenter² , and Angel Rodrigo Avelar Menendez¹

¹Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA; mkuhn@astro.caltech.edu

²Joint ALMA Observatory, Alonso de Córdova 3107 Vitacura, Santiago, Chile

Received 2020 May 6; revised 2020 June 15; accepted 2020 June 24; published 2020 August 20

Abstract

We examine the clustering and kinematics of young stellar objects (YSOs) in the North America/Pelican Nebulae, as revealed by Gaia astrometry, in relation to the structure and motions of the molecular gas, as indicated in molecular-line maps. The Gaia parallaxes and proper motions allow us to significantly refine previously published lists of YSOs, demonstrating that many of the objects previously thought to form a distributed population turn out to be nonmembers. The members are subdivided into at least six spatio-kinematic groups, each of which is associated with its own molecular cloud component or components. Three of the groups are expanding, with velocity gradients of $0.3\text{--}0.5\text{ km s}^{-1}\text{ pc}^{-1}$, up to maximum velocities of $\sim 8\text{ km s}^{-1}$ away from the groups' centers. The two known O-type stars associated with the region, 2MASS J20555125+4352246 and HD 199579, are rapidly escaping one of these groups, following the same position–velocity relation as the low-mass stars. We calculate that a combination of gas expulsion and tidal forces from the clumpy distribution of molecular gas could impart the observed velocity gradients within the groups. However, on a global scale, the relative motions of the groups do not appear either divergent or convergent. The velocity dispersion of the whole system is consistent with the kinetic energy gained due to gravitational collapse of the complex. Most of the stellar population has ages similar to the freefall timescales for the natal clouds. Thus, we suggest the nearly freefall collapse of a turbulent molecular cloud as the most likely scenario for star formation in this complex.

Unified Astronomy Thesaurus concepts: [Star formation \(1569\)](#); [Star forming regions \(1565\)](#); [Stellar associations \(1582\)](#); [Stellar kinematics \(1608\)](#); [Space astrometry \(1541\)](#); [Young stellar objects \(1834\)](#); [T Tauri stars \(1681\)](#); [Giant molecular clouds \(653\)](#); [Millimeter astronomy \(1061\)](#); [OB stars \(1141\)](#)

Supporting material: figure sets, machine-readable tables

1. Introduction

The way that stars and gas are distributed in a star-forming region can provide useful constraints on the conditions in which the stars were formed (e.g., Elmegreen 2002; Parker et al. 2014; Feigelson 2018; Gouliermis 2018). For the nearest star-forming regions, astrometric measurements by ESA's Gaia spacecraft (Gaia Collaboration et al. 2016) can provide a multidimensional picture of how the young stars are clustered. Spatial and kinematic clustering has already been examined in several of the major star-forming regions within 1 kpc, including Orion (Großschedl et al. 2018; Kounkel et al. 2018; Getman et al. 2019), Taurus (Luhman 2018; Fleming et al. 2019; Galli et al. 2019), ρ Oph (Cánovas et al. 2019), and Serpens (Herczeg et al. 2019). We aim to do a similar analysis for North America and Pelican Nebulae (hereafter NAP). Given the correlation between stars and gas within many star-forming regions (e.g., Tobin et al. 2009; Gutermuth et al. 2011; Lada et al. 2013), better understanding of the clustering of the stars can help decipher the complex velocity structures seen in radio molecular-line maps of star-forming clouds (e.g., Larson 1981; Heyer & Dame 2015).

The NAP region contains a molecular cloud complex (Bally & Scoville 1980), an H II region (W80; Westerhout 1958), and a population of young stellar objects (YSOs), many of which have been extensively studied (Reipurth & Schneider 2008). This complex is fairly extended, with a diameter of $\sim 3^\circ$ ($\approx 40\text{ pc}$) encompassing several sites of active star formation. The North America (NGC 7000) and Pelican (IC 5070) Nebulae make up the east/west components of the H II region, which is bisected in projection on the sky by a dark lane known as L933/935

(Lynds 1962), giving rise to the characteristic shape of the nebulae as seen in optical light. The YSOs are predominantly located in the dark lane (Bally et al. 2014), with the southeastern portion known as the “Gulf of Mexico” and the northern part of the cloud containing the “Atlantic Ocean” and the “Pelican” regions.³ Many of the first-identified NAP YSOs were optically visible emission-line stars (e.g., Herbig 1958). More recently, Spitzer has uncovered thousands of embedded stars and protostars in the “Gulf of Mexico,” “Pelican,” and “Pelican's Hat” (Guieu et al. 2009; Rebull et al. 2011). Cambrésy et al. (2002) identified several dense clusters of stars in the NAP region based on near-infrared star counts. Dozens of outflows from YSOs attest to ongoing star formation throughout the complex (Bally & Reipurth 2003; Bally et al. 2014).

Another intriguing aspect of this star-forming region is its primary ionizing source, 2MASS J20555125+4352246 (dubbed the Bajamar Star; Maíz Apellániz et al. 2016). This source was discovered by Comerón & Pasquali (2005) lying to the northwest of the “Gulf of Mexico” behind $\sim 9.6\text{ mag}$ of optical extinction. The star was recently classified as an O3.5(f^*)+O8: spectroscopic binary (Maíz Apellániz et al. 2016), which would make the primary the nearest known massive star with spectral type earlier than O4.⁴ Another O star, HD 199579, is projected on the North America Nebula. This star, with spectral type O6.5 V (f)z (Sota et al. 2011), has generally been regarded as being

³ The boundaries of the named subregions are given by Rebull et al. (2011) and Zhang et al. (2014).

⁴ As a point of comparison, the nearest O4 star, ζ Pup, is less than half as far as the Bajamar Star (Howarth & van Leeuwen 2019).

Table 1
Summary of Candidate NAP Members from the Literature

Reference (1)	Method (2)	Number of Candidates (3)	Reference Notes (4)	Gaia Matches (5)	Membership	
					Nonmem. (6)	Mem. (7)
Merrill & Burwell (1949)	H α	5	^a	60%	3	0
Herbig (1958)	H α	68		40%	6	21
Welin (1973)	H α	142	^b	89%	113	13
Cohen & Kuhi (1979)	H α	21	^c	57%	1	11
Bally & Scoville (1980)	IR	11		36%	3	1
Marcy (1980)	H α	7		57%	3	1
Ogura et al. (2002)	H α	32		47%	1	14
Comerón & Pasquali (2005)	Spec.	8	^d	88%	5	2
Laugalys et al. (2006)	Phot./H α	430	^e	84%	340	21
Witham et al. (2008)	H α	39	^a	82%	7	25
Straizys & Laugalys (2008)	CMD	5		60%	3	0
Corbally et al. (2009)	Spec.	34	^f	62%	18	3
Guieu et al. (2009)	IRE	1,657		24%	125	272
Rebull et al. (2011)	IRE	1,329	^g	29%	140	252
Armond et al. (2011)	H α /Clust.	54		22%	1	11
Damiani et al. (2017)	X-ray	721		19%	46	93
	H α	123		66%	22	59
	IRE	179		51%	58	34
Combined sample	All	3,473		35%	814	395

Notes. Column (1): references for published lists of candidate members of the NAP region. Column (2): method for selection, including H α emission, spectral classification, placement on the CMD, brightness in the infrared (IR) or IR excess (IRE), spatial clustering, and X-ray emission. Column (3): number of candidates in the paper. Column (4): notes about the reference. Column (5): percentage of sources with Gaia DR2 counterparts that pass our quality criteria. Column (6): number of objects that we reject as members. Column (7): number of objects whose membership we have validated. The final row gives the statistics of the combined sample; the numbers of sources (Columns (3), (6), and (7)) in preceding rows do not sum to the values in this row because many objects have been repeatedly selected by multiple studies.

^a Objects were cataloged as H α emission stars in the referenced papers but later upgraded to potential NAP members by Rebull et al. (2011).

^b We use the corrected list of 142 objects available from <ftp://ftp.lowell.edu/pub/bas/starcats/welin.cyg>, rather than the originally published list of 141 objects.

^c The H α sample was enlarged by including stars spatially associated with known H α objects.

^d In addition to the Bajamar Star, we consider all sources from Comerón & Pasquali (2005) not classified as AGB stars to be candidate members.

^e We include all 430 stars in the direction of L935. However, a subset of 41 stars are flagged as having H α emission, 20 of which are classified as nonmembers and 4 as members by our analysis.

^f We include all objects for which spectra are provided, including 19 stars with emission lines and 15 without. We classify three emission-line stars as members and nine as nonmembers, as well as nine non-emission-line stars as nonmembers.

^g This combines 1286 YSO candidates identified in a Spitzer/MIPS-based search with 43 new IRAC-only candidates.

too far from the center of the H II region to be a main ionizing source (e.g., Herbig 1958). Whether HD 199579 is a member of the NAP association has been uncertain (Wendker et al. 1983).

The NAP complex is in the Orion-Cygnus arm of the Galaxy, positioned ahead of our Sun in the direction of Galactic rotation at a distance of about 795 ± 25 pc (Section 5). In projection on the sky, NAP is located adjacent to the massive star-forming regions of Cygnus X, but it is generally thought that NAP is nearer than Cygnus X. A study of molecular clouds in the solar neighborhood by Zucker et al. (2020) and Alves et al. (2020) has confirmed that this complex is part of a string of star-forming regions stretching from the Orion Molecular Clouds to Cygnus X.

The NAP region provides a useful laboratory for studying both star formation in individual clouds and the evolution of the whole complex. Section 2 presents the YSO catalogs, Gaia data, radio data, and spectroscopy used for this study. Section 3 describes the methods used to identify stellar members and reject contaminants. Section 4 describes the properties of multiple stellar groups. Section 5 computes distance estimates for the components of the association. Section 6 describes the structure of the molecular clouds and their relation to the stars. Section 7

provides an analysis of the stellar kinematics. Section 8 identifies new candidate members identified from the Gaia catalogs. Section 9 discusses the formation and dynamical evolution of the system. Finally, Section 10 summarizes the main conclusions.

2. Data

2.1. Initial Stellar Catalog

We searched for lists of candidate YSOs from published studies of the NAP region using the “YSO Corral” (Hillenbrand & Baliber 2015), a curated database for YSOs.

The previous studies have used a variety of techniques to identify candidate members, as listed in Table 1. These included selection based on H α emission, infrared excess, X-ray emission, placement on color–magnitude diagrams (CMDs), and spatial clustering. H α emission and infrared excess are mostly sensitive to disks or accretion, while X-ray emission can be used to detect pre–main-sequence stars both with and without disks. Each of these methods yields different types of contaminants and rates of contamination, so we aim to reassess memberships of their candidate members using astrometry from Gaia, as described below. Contaminants to YSO searches in the Galactic plane include both field stars and extragalactic sources; in particular,

dusty (post-)asymptotic giant branch (AGB) stars and star-forming galaxies may be major sources of contaminants for selection based on infrared excess (Robitaille et al. 2008), while active galactic nuclei and foreground active M dwarfs may dominate X-ray contaminants (Broos et al. 2013).

2.2. Astrometric Data

We obtained stellar astrometry from Gaia’s second data release (DR2; Gaia Collaboration et al. 2018), which provides five-parameter astrometric solutions for 1.3 billion objects to as faint as $G \approx 21$ mag (Lindegren et al. 2018). These solutions provide source positions in R.A. (α) and decl. (δ), parallax (ϖ), and proper motion⁵ (μ_{α^*} , μ_{δ}). For DR2, most stars in the direction of NAP were observed during ~ 15 visibility periods; typical formal uncertainties for a $G = 15$ mag star are ~ 0.03 mas in parallax and ~ 0.05 mas yr⁻¹ in proper motion. In addition to the formal uncertainty, ~ 0.04 mas and ~ 0.07 mas yr⁻¹ correlated systematic uncertainties affect these measurements (Lindegren et al. 2018). There have been a variety of efforts to estimate the systematic zero-point offsets (e.g., Leung & Bovy 2019, and references therein); in most cases we try to work with the observed quantities (i.e., ϖ), but when distances are required we apply a 0.0523 mas parallax correction estimated in the aforementioned paper.

We use the Gaia catalog both for cross-matching to the previously identified YSO candidates and to search for new member candidates. For cross-matching, we use a match radius of 1'' and select the nearest source as a match, yielding 1939 matches out of ~ 3500 candidates. This match radius is large enough that changes in position from proper motion are unlikely to affect whether a counterpart is found. The relatively low match rate is because, due to extinction, the NAP population extends to faint optical magnitudes, far below the Gaia detection limits. To estimate the rate of false matches, we artificially shifted the DR2 coordinates 2' north and ran a cross-match using these coordinates. We found 52 matches to the shifted coordinates, suggesting an incorrect match rate of up to 3%. This rate should be regarded as an upper limit on match errors, because a true counterpart, if it exists, would likely take precedence over a field star if both lie within the match radius.

We apply cuts to the Gaia data to ensure that the measurements are sufficiently precise to be useful for distinguishing between members and nonmembers. We follow the recommendations⁶ of the Gaia Data Processing and Analysis Consortium and use only sources with renormalized unit weight error (RUWE) less than 1.4. The Gaia catalog also tabulates excess noise in the astrometric fit. Higher values of excess noise (e.g., > 1 mas; Kuhn et al. 2019) could indicate some acceleration of a source (e.g., due to a binary) that might affect the astrometric solution, so we only include sources with excess noise ≤ 1.0 mas. And, finally, we require `astrometric_sigma5d_max` ≤ 0.5 mas to limit the maximum statistical measurement uncertainties on parallaxes and proper motions. These quality cuts only slightly affect the peak of the parallax distribution (shifting it by the equivalent of 4 pc), but the disadvantage of including all sources would be that the modes of the distribution become broader and source classifications become less certain. We also removed objects with $\varpi > 4$ mas because they are clearly in the

foreground and interfere with the mixture model analysis in Section 3.

2.3. Millimeter Observations

We obtained ¹³CO $J = 1-0$ (110.201354 GHz) observations of the molecular clouds associated with the NAP region using the Five College Radio Astronomy Observatory (FCRAO) 14 m telescope in 1998 June. The CO map covers most of the region of interest in this study.

The observations were obtained with the Second Quabbin Optical Imaging Array (SEQUOIA; Erickson et al. 1999). The array contains 16 pixel elements arranged in a 4×4 grid with separation of 88'' on the sky. At the time of the observations, 12 of the pixels were functional. The spectrometer contained 512 channels with a bandwidth of 40 MHz, which provided a channel spacing of 0.21 km s⁻¹. The FCRAO telescope has an FWHM angular resolution of 47'' at the ¹³CO $J = 1-0$ frequency. The forward scattering and spillover efficiency of the telescope at the observed frequency was $\eta_{\text{FSS}} = 0.7$, while the main-beam efficiency was $\eta_{\text{mb}} = 0.45$ (Heyer & Terebey 1998). The data were obtained in position-switching mode with spectra that were sampled every 44'' on the sky, or approximately the FWHM resolution. The typical rms noise in the spectra is $\Delta T_A^* = 0.22$ K per channel. The maps used for our analysis are in units of T_R^* , calculated from the atmosphere-corrected antenna temperature T_A^* by dividing by a correction factor (Kutner & Ulich 1981). Here, we use the correction factor η_{FSS} since the ¹³CO emission is usually more extended than an arcminute in the NAP region.

Our ¹³CO data are similar in sensitivity and spatial resolution to a ¹³CO map published by Zhang et al. (2014) using the Purple Mountain Observatory Delingha 13.7 m telescope. They also provide ¹²CO and C¹⁸O maps.

2.4. Spectroscopic Observation of the Ionizing Source

Finally, high-dispersion optical spectra of the Bajamar Star were taken on 2019 December 1 and 2020 January 3 (UT) for the purposes of confirming the existing spectral typing that was based on low-resolution spectra and assessing the source radial velocity. These data were taken with the Keck I telescope and HIRES (Vogt et al. 1994) and cover $\sim 4800-9200$ Å at $R \approx 37,000$.

3. Validation of Membership with Astrometry

We can evaluate membership of candidate YSOs by observing whether they are at the same distance and moving with the same kinematics as the rest of the members in the NAP region. In Figures 1–3, we use the following color scheme: gray—stars at the wrong distance; green—stars at a similar distance but with the wrong proper motion; magenta—probable astrometric members; goldenrod—stars with uncertain classification.

The distribution of parallaxes of YSO candidates (Figure 1, left panel) has at least two modes, as well as a tail of objects with high parallax. The expected distance to NAP based on previous measurements is 500–1000 pc (e.g., Reipurth & Schneider 2008), which coincides with the second peak on the histogram (magenta mark). The leftmost peak can be attributed to contaminants with small or zero parallax, including giant stars and extragalactic sources. The tail of objects to the right includes many bright objects with high-precision parallax measurements, so these are clearly foreground stars.

⁵ We follow the convention of Perryman et al. (1997) and define $\mu_{\alpha^*} \equiv \mu_{\alpha} \cos \delta$. This quantity is called μ_{PMRA} in the Gaia DR2 tables.

⁶ <https://www.cosmos.esa.int/web/gaia/dr2-known-issues>

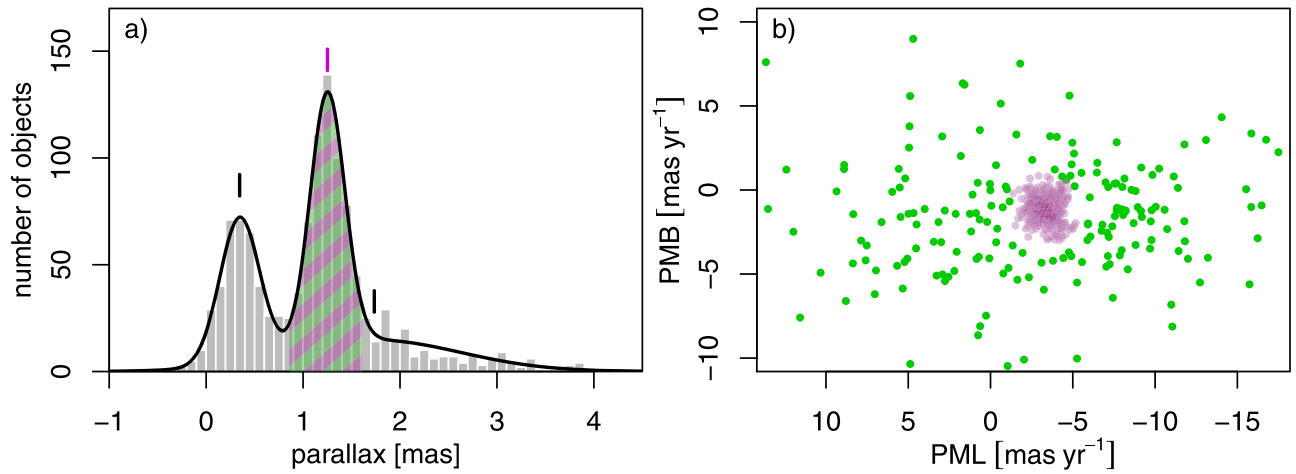


Figure 1. Illustration of the two-step process for using Gaia astrometry to validate YSO candidate membership. (a) Distribution of parallaxes (histogram) fit with a Gaussian mixture model (black line). The centers of the three Gaussians are indicated by the vertical marks above the curve. Objects with a probability $>50\%$ of being a member of the middle component (associated with NAP) are selected as indicated by the striped green/magenta region. (b) Proper motions μ_{l^*} (PML) and μ_b (PMB) for the objects selected in panel (a). The distribution of these sources is fit with a two-dimensional Gaussian mixture model, and the objects with a probability $>50\%$ of being a member of the middle density peak are colored magenta. These compose our sample of 395 Gaia-validated NAP members.

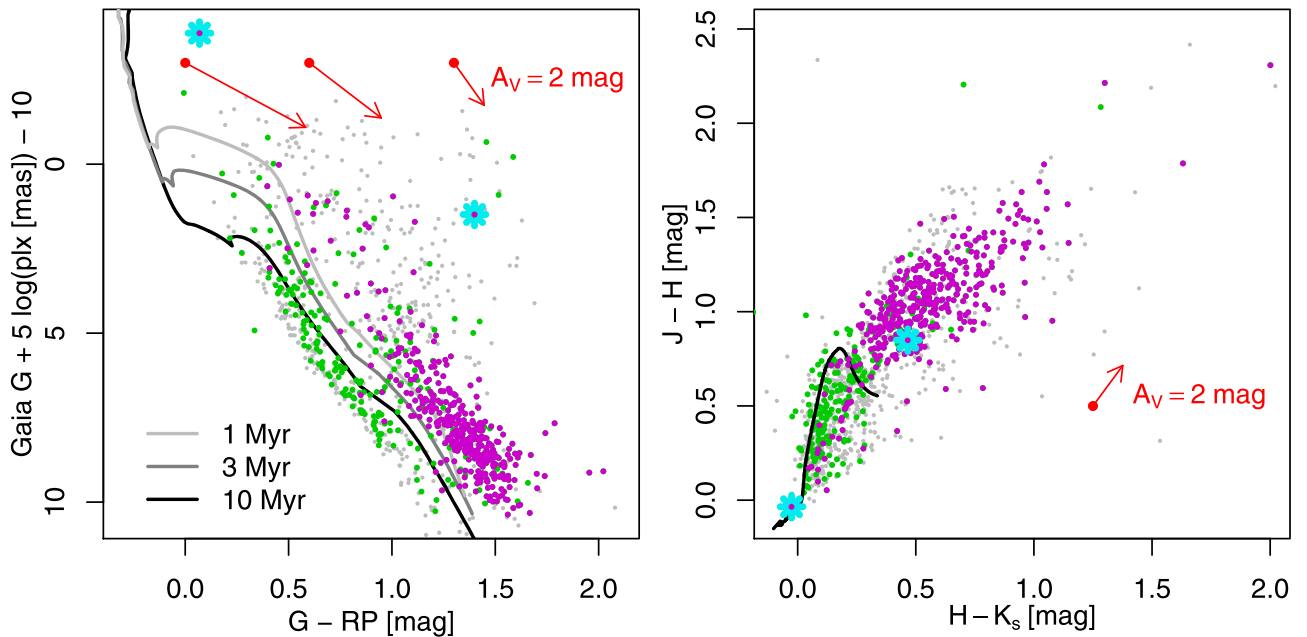


Figure 2. Left: Gaia CMD for YSO candidates classified in this work. Members are magenta points, sources excluded based on parallax are gray points, and sources excluded based on proper motion are green points. Only sources with $\varpi/\sigma_\varpi > 3.0$ are included on the diagram. The curves indicate unreddened PARSEC isochrones for 1, 3, and 10 Myr. The red arrows show approximate reddening vectors for $A_V = 2$ mag for pre-main-sequence stars with unreddened colors of $G - RP = 0, 0.6,$ and 1.3 mag. The two O stars stars are marked with the cyan asterisks; the Bajamar Star is the one with the very red color. Right: 2MASS color-color diagram using the same symbols as the left panel. The black curve indicates the locus of unreddened stellar colors from the PARSEC models, whose shape is similar for young pre-main-sequence stars but extends ~ 0.15 mag redder in $J - H$ around and beyond the peak (Herczeg & Hillenbrand 2015).

To quantitatively divide stars based on parallax, we model the parallax distribution using a Gaussian mixture model, for which we use the Bayesian Information Criterion (BIC; Schwarz 1978) to determine the number of components. For Gaussian mixture models the BIC can be viewed as an approximation for the Bayes factor, which is useful for the statistical problem of model selection (Everitt et al. 2011). We find that the distribution can be well modeled with three Gaussians, as marked in Figure 1, with $\Delta\text{BIC} = 10$ compared to the next best model, implying strong preference for this solution. From low parallax to high parallax, the first component models the peak that we associated with background contaminants, the second component

corresponds to the group of stars associated with NAP and has a mean of $\varpi_0 = 1.25$ mas, and the third component has a slightly higher mean parallax but a much broader distribution that appears to approximate the shape of the high-parallax tail. Objects that lie between $0.86 \text{ mas} < \varpi < 1.61 \text{ mas}$ (green and magenta striped region) have a $\geq 50\%$ probability of belonging to the second component; all other sources are classified as nonmembers.

In the parallax selected sample, cross-contamination from field stars that happen to have similar parallaxes to the NAP region is inevitable. To reduce this contamination, we perform a second classification step, this time using proper motion. Figure 1 (right panel) reveals a clump of sources with similar

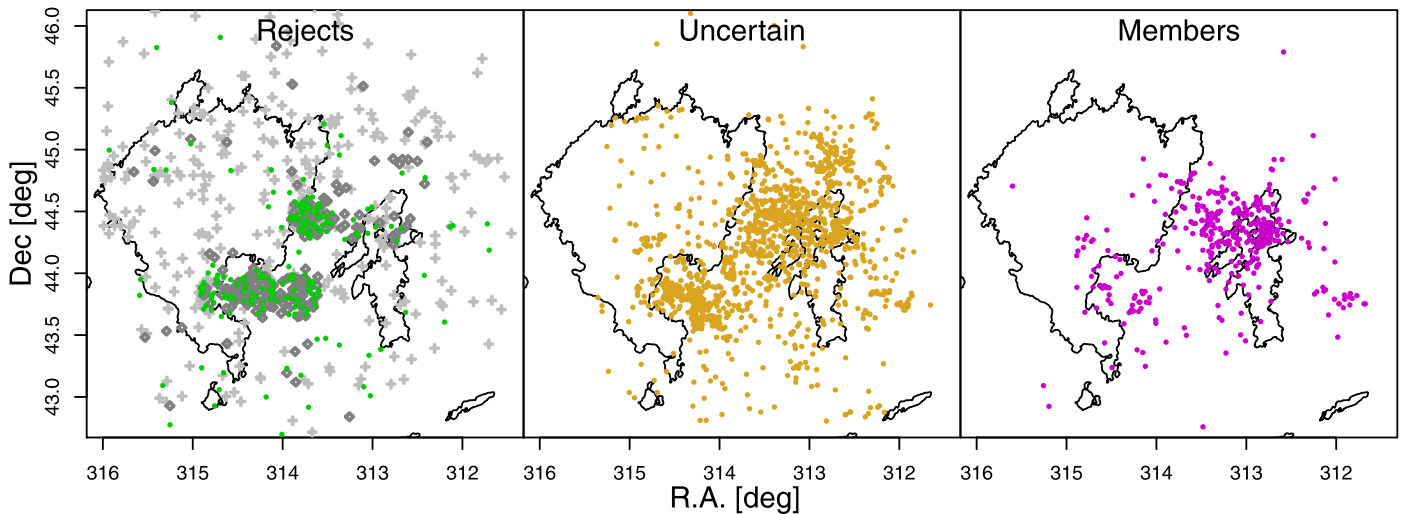


Figure 3. Spatial distributions of YSO candidates stratified by their classification as contaminants or members. The left panel shows the distributions of different classes of contaminant: objects rejected by parallax in gray (background to the NAP as light-gray plus signs and foreground as dark-gray diamonds) and objects rejected by proper motion in green. The middle panel shows objects where Gaia-based classification is uncertain because the source does not exist in the Gaia catalog or it does not meet our quality criteria. The right panel shows objects classified as bona fide members. In black we show contours of the optical nebulae from the DSS image. Much of the spatial clustering in the middle panel is reminiscent of structure in the right panel, indicating that our techniques are selecting only a portion of the true young star population, but there is also a broad spatial distribution reminiscent of the left panel, indicating a mix of contaminants in the uncertain population too.

motions and a halo of objects with significantly different motions. We show these motions in Galactic coordinates (μ_{l^*} , μ_b) to emphasize that the proper-motion dispersion in the halo is largely parallel to the Galactic plane, thus representing orbital motions of field stars in the Galaxy that are dynamically hotter than newly formed stars and thus have larger dispersions in dynamical phase space. We subdivide the sources in proper-motion space using another Gaussian mixture model, finding two components ($\Delta\text{BIC} = 7$), corresponding to the clump (smaller dispersion) and the halo (larger dispersion). We classify objects with $\geq 50\%$ probability of belonging to the first component as members (magenta points), while the others are nonmembers (green points). The members have mean proper motions of $\mu_{l^*} = -3.35 \text{ mas yr}^{-1}$ and $\mu_b = -1.15 \text{ mas yr}^{-1}$; other parameters of the classifier are given in Appendix A. The mixture model suggests a 3% residual contamination rate among the stars classified as members. We note that selection based on proper motion comes at the cost of omitting stars with high proper motions that could have been ejected from the star-forming region.

Figure 2 (left) shows the classified sources on a diagram of Gaia absolute magnitude versus color. When compared to theoretical PARSEC isochrones (Bressan et al. 2012), the members nearly all lie above the 3 Myr isochrone, and many of them lie above the 1 Myr isochrone. In contrast, the objects identified as nonmembers are scattered throughout the diagram, with many lying below the 10 Myr isochrone, while others are located in the region of post-main-sequence giant stars, and others lie in similar locations to the pre-main-sequence stars. This diagram confirms that nearly all of the astrometrically validated members are pre-main-sequence stars with ages $< 3 \text{ Myr}$, but that some rejected objects that could also be young. The effect of reddening (red arrows) varies with color owing to the large width of Gaia’s G and G_{RP} bands, so we show three approximate vectors for different $G - RP$ colors assuming a typical pre-main-sequence star spectrum. From the $J - H$ versus $H - K_s$ color-color diagram (Figure 2, right) using photometry from the Two Micron All Sky Survey (2MASS;

Skrutskie et al. 2006), we estimate that typical reddening is $E(J - H) \sim 0.1\text{--}0.6$, corresponding to $A_V \sim 1\text{--}6 \text{ mag}$. Both O stars are labeled in the diagrams, but HD 199579 is one of the brightest, bluest sources, while the Bajamar Star is significantly dimmer and redder owing to its high extinction.

3.1. Members and Contaminants

Out of the sample of YSO candidates, 395 objects are confirmed as members, while 814 objects are reclassified as nonmembers, and 2264 sources either do not have a Gaia counterpart or do not meet our quality criteria and therefore have uncertain classifications. The numbers of candidates in each category are tabulated in Table 1 for each of the published lists of YSO candidates. Table 2 provides a list of the astrometric members.

It is not surprising that nearly all historical studies have some contamination. For instance, out of the 68 objects from the pioneering study of the NAP region by Herbig (1958), Gaia provides sufficiently high quality measurements for 27 of them. Of these stars, our analysis has validated the membership of 21 objects, while rejecting membership for the other 6. The rejected members include LkH α 132, LkH α 133, LkH α 147, LkH α 183, LkH α 192, and LkH α 193, which all have parallax values that are too small to be members of the region, suggesting that they are background Be stars.

For some categories of YSO candidate, Gaia’s requirement that a source must be sufficiently bright in the optical may induce a selection bias toward higher contamination rates. For example, out of the 1286 Spitzer/MIPS sources classified as YSOs by Rebull et al. (2011) on the basis of infrared excess, only 30% met our Gaia criteria for inclusion in the analysis. Of those 30%, 131 were rejected and 251 were validated. However, this ratio is not representative of the whole sample of 1286 MIPS candidates because the YSOs with the most prominent infrared excesses tend to be very young and deeply embedded, hence not optically visible.

The spatial distribution of candidates classified as members, nonmembers, or uncertain membership is shown in Figure 3.

Table 2
Astrometrically Validated Members

Gaia DR2	α	δ	Group
(1)	(ICRS)	(ICRS)	(3)
(1)	(2)	(3)	(3)
2163138601938577664	20 51 19.8	+44 23 06	D
2163137742945112960	20 51 20.6	+44 20 32	C
2163138705017719296	20 51 21.3	+44 24 05	D
2163136123738466688	20 51 12.0	+44 18 47	C
2163137742945115136	20 51 22.6	+44 21 07	C
2163148772421081728	20 51 22.8	+44 33 42	D
2163149665774282368	20 51 23.6	+44 35 22	D
2066862546309607552	20 51 26.4	+43 53 12	D
2162947424350344448	20 51 24.4	+44 13 04	C
2163135883219217280	20 51 24.6	+44 17 54	D

Note. Gaia DR2 counterparts for the 395 previously published YSO candidates that passed our astrometric membership criteria. Column (1): source designation in the DR2 catalog. Columns (2)–(3): source positions are truncated, not rounded. We encourage readers to obtain the full-precision Gaia astrometry for these sources directly from the Gaia Archive (<https://gea.esac.esa.int/archive/>). Column (4): kinematic group assignment for each star.

(This table is available in its entirety in machine-readable form.)

These points are overplotted on an outline of the optical nebulosity making up the North America and Pelican components of the nebula. Nonmembers are widely dispersed throughout the whole region, while uncertain members are more tightly clustered, and validated members are the most tightly clustered. This suggests that the spatial distribution of stars is dominated by clustered groups, rather than distributed stars. Although a few members are located a degree or more away from the main groups, there is no significant nonclustered population scattered throughout the entire $\sim 3^\circ$ diameter region.

Our sample of 395 objects is clearly incomplete, missing both objects that were not selected in previous studies (e.g., pre-main-sequence stars without disks) and objects that could not be evaluated by Gaia. There is also evidence that our classifier has rejected a few legitimate YSOs. From a sample of 41 high-confidence YSOs studied by Findeisen et al. (2013), 4 were rejected by the classifier, likely spuriously. Altogether, it is likely that the NAP region contains at least several thousand stellar members. This assertion is based on Spitzer observations that have collectively identified ~ 2000 YSO candidates (Guieu et al. 2009; Rebull et al. 2011), most of which are not included in our study because they are inaccessible to Gaia. Furthermore, in Section 8 we identify >1000 new candidate members based on Gaia astrometric data. Strictly speaking, our membership classification is for the Gaia source, so the chance alignment of a mid-infrared YSO with a Gaia field star would likely result in rejection of the source.

3.2. Properties of the Ionizing Sources

Both the Bajamar Star (the primary ionizing source) and HD 199579 pass our Gaia membership criteria in Section 3. However, the parallaxes and proper motions of both of these O stars place them near the edges of parameter space for objects identified as members. For HD 199579 ($\varpi = 1.06 \pm 0.06$ mas, $\mu_{\ell^*} = -1.4 \pm 0.1$ mas yr $^{-1}$, and $\mu_b = -1.6 \pm 0.1$ mas yr $^{-1}$), its parallax places it behind most other cluster members. It lies on the northeast extreme of the NAP region and is moving in this direction relative to the rest of the NAP members.

The Bajamar Star ($\varpi = 1.47 \pm 0.08$ mas, $\mu_{\ell^*} = -3.4 \pm 0.1$ mas yr $^{-1}$, and $\mu_b = -2.9 \pm 0.2$ mas yr $^{-1}$) has a parallax measurement that is 2σ greater than the median parallax of the system, and the μ_{ℓ^*} proper motion is more negative than that of most other members. The Gaia data do not indicate any obvious problems with the astrometry; the RUWE is 1.04, which is within the recommended <1.4 range. The source has a small, but highly statistically significant, excess noise of 0.346 mas, which could be explained if it is a binary system. In Section 5 we argue that the star is unlikely to be in front of the complex, so the discrepant parallax measurement is likely to be the result of statistical measurement uncertainty.

From our HIRES spectrum of the Bajamar Star, we find an O4–O6 spectral type. This estimate is based on the presence of He II $\lambda 5412$ with $W_\lambda = 1.1 \text{ \AA}$ and the barely discernible He I $\lambda 4922$ ($W_\lambda < 0.18 \text{ \AA}$), He I $\lambda 5015$ ($W_\lambda < 0.13 \text{ \AA}$), and He I $\lambda 5047$ ($W_\lambda < 0.08 \text{ \AA}$).

We can also confirm the statement in Maíz Apellániz et al. (2016) that the different absorption-line species have different radial velocities. However, it is not clear based on our spectra that the source is a spectroscopic binary. While there is evidence for asymmetries in the line profiles, especially H I and He I, a high-velocity wind, as is expected from such a massive star, would produce similar features. Furthermore, a cross-correlation analysis of our HIRES spectrum does not reveal indications of two peaks, though the lines are extremely broad owing to the rapid rotation and possibly other effects, and it certainly would be possible to mask two sets of broad lines within the profiles.

Using τ Sco as a radial velocity standard, we derive $v_{\text{helio}} = 11.6 \pm 16.5 \text{ km s}^{-1}$ for our first observation and $v_{\text{helio}} = -17.1 \pm 16.8 \text{ km s}^{-1}$ for our second observation, both based on the single He II $\lambda 5412$ line. Although the errors are large, this is our best estimate of the stellar radial velocity. The velocities derived from the only other measurable and noninterstellar lines in the first spectrum are -100 km s^{-1} from a single He I line and -90 and -35 from H I (H α and H β respectively), very different from the He II velocity. The second spectrum does show a radial velocity shift. Measured directly, rather than going through a narrow-line star, the shift in the strong He II $\lambda 5412$ line is by $28 \pm 2 \text{ km s}^{-1}$. This certainly suggests binarity. However, an order containing He I $\lambda 5876$ and C IV $\lambda \lambda 5801, 5811$ shows a velocity difference of around -40 km s^{-1} , while He II $\lambda \lambda 6406, 6527$ suggest an $8\text{--}10 \text{ km s}^{-1}$ difference and both He I $\lambda 6678$ and He II $\lambda 6683$ are closer to $+40 \text{ km s}^{-1}$. In H α and H β the shift between the spectra is about $15 \pm 3 \text{ km s}^{-1}$. It remains unclear whether we should interpret the measurements as a single star having the same velocity as the molecular gas and strong wind that manifests in confusing absorption velocities, or as a binary with the more massive star producing much of the He II absorption and both stars contributing to the relatively blueshifted metal, He I, and H I profiles.

4. Kinematic Groups

The distribution of Gaia-validated NAP members is neither smooth nor centrally concentrated, but has a more complicated, clumpy structure that is apparent both in the stars' spatial positions (Figure 4) and in their distributions in proper-motion space (Figures 5 and 6). Spatial clustering and subclustering of young stars have been appreciated for decades (e.g., Carpenter 2000; Allen et al. 2007), but only recently have the kinematic

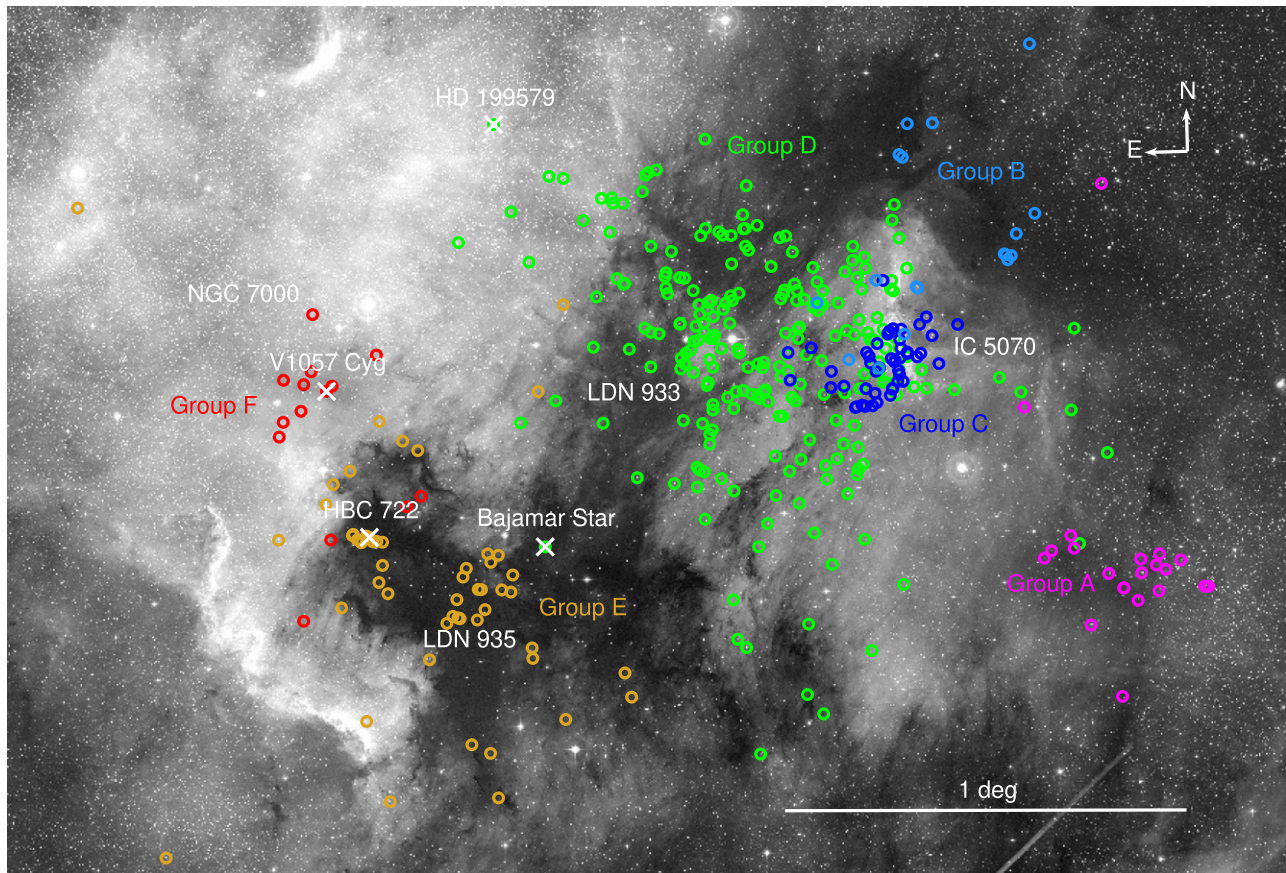


Figure 4. Members, as indicated in the right panel of Figure 3, are overlaid on a DSS red-band image of the region and color-coded based on the kinematic group they belong to. Note that several outlying members are cropped out of the image (see Figure 3) to show enough detail in the central region. Several interesting stars are indicated, including the two O stars and two well-studied outbursting stars. The nebulous regions are those that were outlined in Figure 3—on the left the nebosity resembles the shape of North America, and on the right it resembles a pelican with its beak facing the Atlantic Coast of North America. The naming of subregions (e.g., Gulf of Mexico, Atlantic, Pelican’s Neck) has historically followed these analogies.

data achieved comparable fidelity (e.g., Da Rio et al. 2017; González & Alfaro 2017). This analysis is now possible in a larger number of star-forming regions with Gaia DR2+ astrometry.

4.1. Cluster Analysis Algorithm

When analyzing the stellar population of the NAP region, it is convenient to divide the stars into subclusters. As for all cluster analysis problems, multiple possible strategies can achieve this, with no one approach being “best” (Everitt et al. 2011). We find that a Gaussian mixture model with hierarchical combination of clusters, outlined below, gives reasonable results that are astronomically useful.

We use a Gaussian mixture model analysis implemented by `mclust` (Scrucca et al. 2016) in R to identify groups of stars in four-dimensional position–proper-motion space. The analysis is performed on the variables l , b , μ_{l^*} , and μ_b , each of which is normalized by subtracting the mean value and dividing by the standard deviation. The program then selects the number of clusters, the free parameters of the Gaussian components, and the values of the parameters via minimization of the BIC. For modeling the clustering of stars, the mixture model approach has several benefits, which include the ability to identify clusters that differ in both compactness and number of members, the ability to assign stars from the dense cluster core as well as the less dense cluster wings to the same cluster,

and the ability to deal with overlapping clusters. On the other hand, stars clusters are typically not well fit with Gaussian distributions, which can lead to multiple Gaussian components used to fit a single cluster (Kuhn & Feigelson 2019). For the NAP stars, `mclust` selects a model with eight ellipsoidal components with equal shape and orientation, where several of the Gaussian components are significantly overlapping or nearly concentric.

The `mclust` package includes the function `clustCombi` to deal with such cases by hierarchically merging multiple Gaussian components into single clusters using an entropy criterion (Baudry et al. 2010). For our case, we find a large change in normalized entropy going from six to seven clusters, so we use this threshold to cut the hierarchical tree at six groups. The stars assigned to each of these six groups, labeled A through F, are overlaid on the optical DSS image of the NAP region in Figure 4.

4.2. Groups in Position–Proper-motion Space

The results of the cluster analysis algorithm can be qualitatively checked for reasonableness by examination of the group assignments in Figures 4–6. In any one of the projections, the edges of some groups overlap. However, the stars that overlap in one projection tend to be different from the stars that overlap in a different projection, and the groups are mostly well separated.

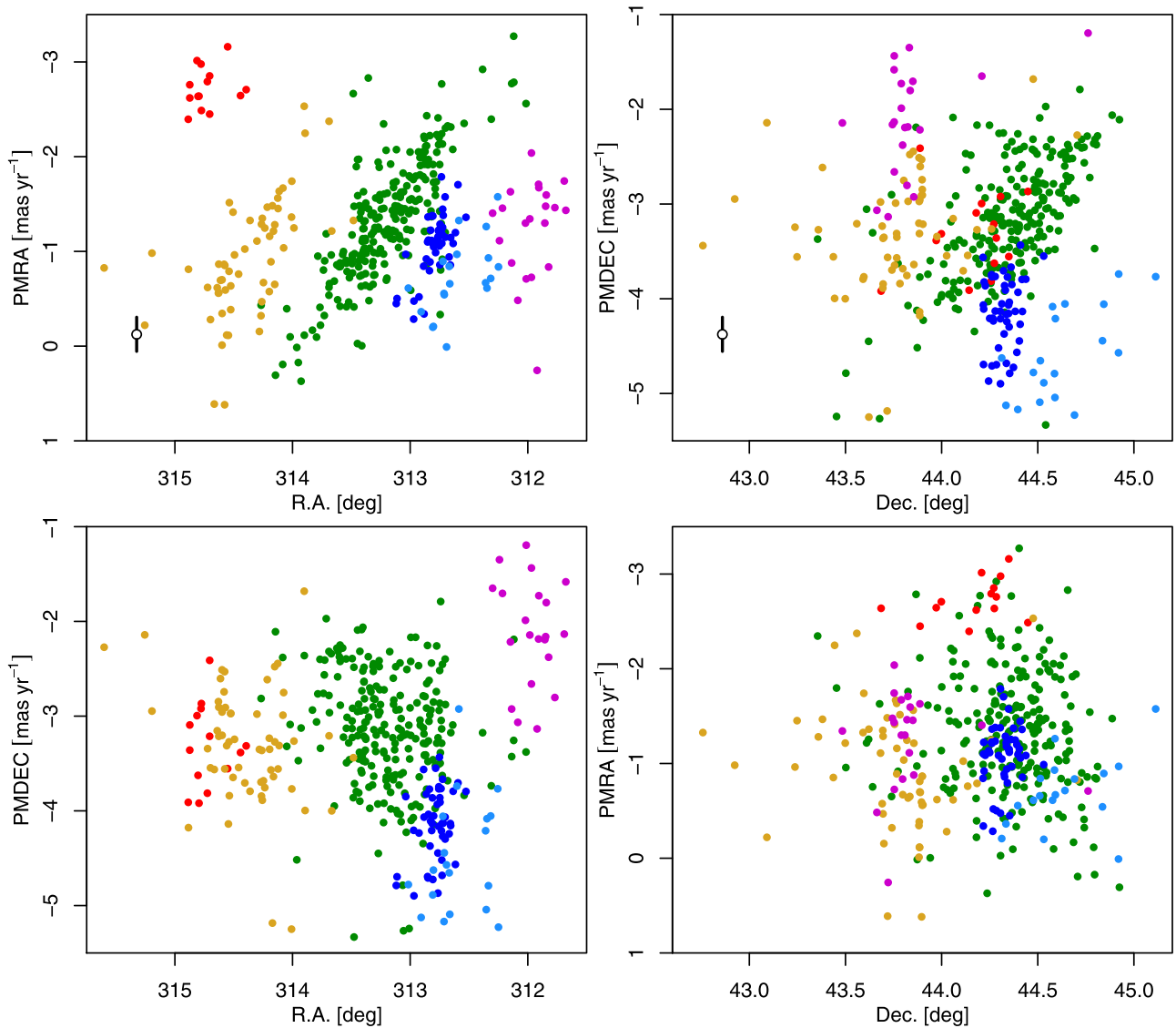


Figure 5. Plots of position coordinates α and δ vs. proper-motion coordinates μ_{α^*} (PMRA) and μ_{δ} (PMDEC). NAP members are color-coded by group using the same hues as Figure 4. The black points in the top panels with error bars indicate the median formal uncertainties in μ_{α^*} and μ_{δ} .

In projection on the sky (Figure 4), Group A is fairly isolated in the southwest corner of the region. Groups B, C, and D form a conglomeration in the northwestern portion of the nebula, more or less corresponding to the regions known as the “Pelican’s Hat,” the “Pelican’s Neck,” and the “Atlantic,” respectively. Groups E and F lie in the southeastern/eastern region—Group E is composed of several clumps of stars on the dark “Gulf of Mexico” cloud, while Group F is a single clump to the northeast of the “Gulf of Mexico” superimposed on an area with bright nebulosity.

The plot of α versus μ_{α^*} (Figure 5, top left) shows the clearest separation of the stars into distinct groups. In the center of this diagram, Group D—the largest group—forms a diagonal swath from lower left to upper right. This group is clearly separate from Group E to its left, which follows a parallel diagonal track. On the right, Group C is located adjacent to Group D but is much more tightly clumped in α and μ_{α^*} . On the right side of the diagram, Group B appears partially entangled with Group C, and Group A appears partially entangled with Group B. However, both Groups A and B are looser than Group C. The spatial separation of A from the rest

of the groups makes it clear that A is distinct. However, whether B could be considered part of the same group as C is less clear. Group B contains only a few stars from our sample; however, these stars coincide with the deeply embedded “Pelican’s Hat” cluster of stars discovered by Spitzer (Guieu et al. 2009; Rebull et al. 2011), most of which are unseen by Gaia.

In the plot of δ versus μ_{δ} (Figure 5, top right), the diagonal swath formed by Group D can still be seen, but the other groups appear distributed in a more vertical direction rather than diagonal. This plot enhances the separation between Groups A, B, and C. Groups C and D are less cleanly separated, but the stars in Group C do not continue the same diagonal trend as Group D. The bottom panels in Figure 5 show the other combinations of position and proper motion. Finally, on the plot of μ_{α^*} versus μ_{δ} (Figure 6), Groups D and E are in the center, Group A is at the top extreme, Groups B and C are at the bottom extreme, and Group E is at the rightmost extreme.

Intriguingly, the Bajamar Star is assigned to Group D even though it is spatially closer to stars from Group E in the “Gulf of Mexico.” This is a result of the proper motion of the Bajamar

Table 3
Astrometric Properties of Stellar Groups

Group	α_0 (ICRS)	δ_0 (ICRS)	N_{samp} (stars)	$\mu_{\alpha^*,0}$ (mas yr ⁻¹)	$\mu_{\delta,0}$ (mas yr ⁻¹)	ϖ_0 (mas)	r_{hm} (pc)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A	20 47 50	+43 47 27	21	-1.51 ± 0.21	-2.11 ± 0.14	1.23 ± 0.02	1.1
B	20 50 20	+44 37 58	19	-0.57 ± 0.29	-4.64 ± 0.18	1.21 ± 0.03	1.8
C	20 51 10	+44 19 25	47	-1.11 ± 0.19	-4.13 ± 0.17	1.23 ± 0.01	1.1
D	20 52 50	+44 22 13	235	-1.19 ± 0.11	-3.13 ± 0.09	1.21 ± 0.01	1.9
E	20 57 30	+43 46 36	59	-0.70 ± 0.14	-3.24 ± 0.15	1.27 ± 0.02	1.3
F	20 58 50	+44 09 43	14	-2.68 ± 0.06	-3.35 ± 0.25	1.06 ± 0.02	2.0

Note. Column (1): name of the stellar group. Columns (2)–(3): approximate coordinates of the group center. Column (4): number of Gaia-validated members assigned to each group. Columns (5)–(6): mean proper motions of the group. Column (7): mean parallax of the group. Column (8): characteristic radius for the stellar group, defined as the median distance of group members, in projection, from the group center. All values are in the Gaia DR2 system, with no correction for zero-point offset.

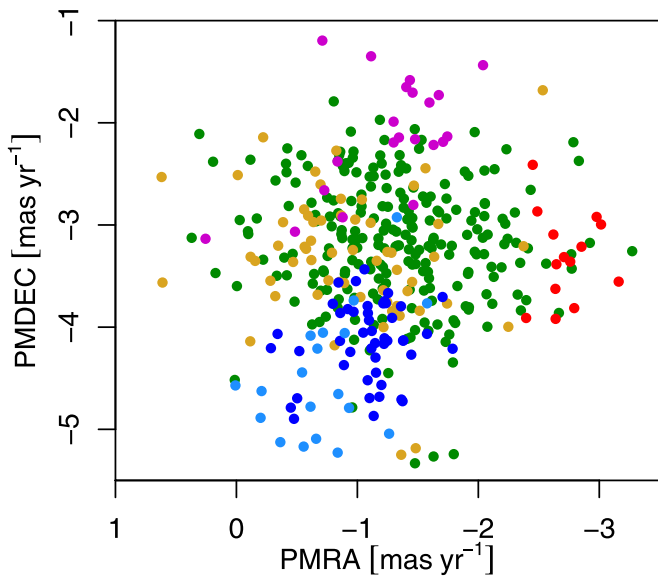


Figure 6. Scatter plot showing μ_{α^*} vs. μ_{δ} . NAP members are color-coded by group as in Figures 4 and 5.

Star, which is rapidly moving toward Group E away from the center of Group D. This proper motion does not match the proper motions of nearby stars in Group E; however, it does fit with the proper-motion gradient seen for stars in Group D (Figure 5).

Our division of stars into kinematic groups is meant to be useful for understanding the structure of the NAP region, but it is not the full description of the cluster structure. For example, nearby groups may have formed from related star formation events at different locations on the same cloud, and thus, from an astrophysical perspective, it is ambiguous whether these should be considered to be the same group. Furthermore, if the region were nearer (e.g., at the distance of Taurus; Luhman 2018; Fleming et al. 2019), we might be able to detect finer velocity differences with Gaia that could be used to subdivide the groups further, but if the region were more distant (e.g., at the distance of the Carina Nebula; Kuhn et al. 2019), the measurement uncertainties in the proper motion would be poorer and, thus, it might not have been possible to distinguish the groups.

5. Three-dimensional Structure of the Stellar Population

The region is sufficiently close that differences in distance to the groups begin to become detectable by Gaia. However, these

differences are subtle enough that care must be taken in estimating mean parallaxes of each group. We calculate the mean parallaxes using a maximum likelihood method that takes into account both the heteroscedastic measurement errors and the truncation of the sample at $\varpi_{\text{min}} = 0.86$ mas and $\varpi_{\text{max}} = 1.61$ mas (Section 3). The log-likelihood as a function of mean parallax, $\bar{\varpi}$, is given by

$$\mathcal{L}(\bar{\varpi}, \sigma^2) = \sum_i \ln \frac{\phi(\varpi_i; \bar{\varpi}, \sigma_i^2 + \sigma^2)}{\int_{\varpi_{\text{min}}}^{\varpi_{\text{max}}} \phi(\varpi; \bar{\varpi}, \sigma_i^2 + \sigma^2) d\varpi}, \quad (1)$$

where ϕ denotes a Gaussian distribution and ϖ_i are the measured parallaxes with uncertainties σ_i . This model allows stellar groups to have an intrinsic depth, characterized by the parameter σ . We use the Broyden–Fletcher–Goldfarb–Shanno algorithm to find the value of $\bar{\varpi}$ that maximizes this likelihood, and we use the Hessian matrix calculated by the R function `optim` to estimate confidence intervals.

The mean parallaxes and their confidence intervals are reported in Table 3 and plotted in Figure 7. All groups in the northern and western parts of the NAP (Groups A–D) have statistically indistinguishable parallax measurements. Of these, the value for Group D is the most precise, with $\varpi_0 \approx 1.21$ mas (corresponding to ~ 795 pc). The parallax of Group F differs the most significantly, with a value of $\varpi_0 = 1.06 \pm 0.02$ mas, placing it behind the other groups by ~ 130 pc. Group E has a mean parallax of 1.27 ± 0.02 mas, which would place it in front of the other groups by ~ 35 pc, but this difference is only significant at slightly over the 2σ level, meaning that the parallax data are not absolutely conclusive. The model also suggests nonzero depths to the groups in parallax ranging from 0.04 to 0.12 mas. These seem unphysically large when translated to physical sizes (25–80 pc) and may result from outliers with small error bars (Figure 7). It is possible that some of these outliers could be misclassified stars, either spurious members or members with a misassigned group.

Uncertainty in our estimate of the absolute distance would be dominated by the systematic ~ 0.04 mas correlated uncertainty in parallax (Lindgren et al. 2018). This would translate to a shift of ± 25 pc.

Extinction and nebulosity seen in optical images provide additional evidence to support the order in distance suggested by parallax measurements. Group F is superimposed on a bright region of the nebula, even though the stars in this group appear reddened on CMDs, suggesting that the H α nebulosity is in front of the group. The stars in Group E are associated

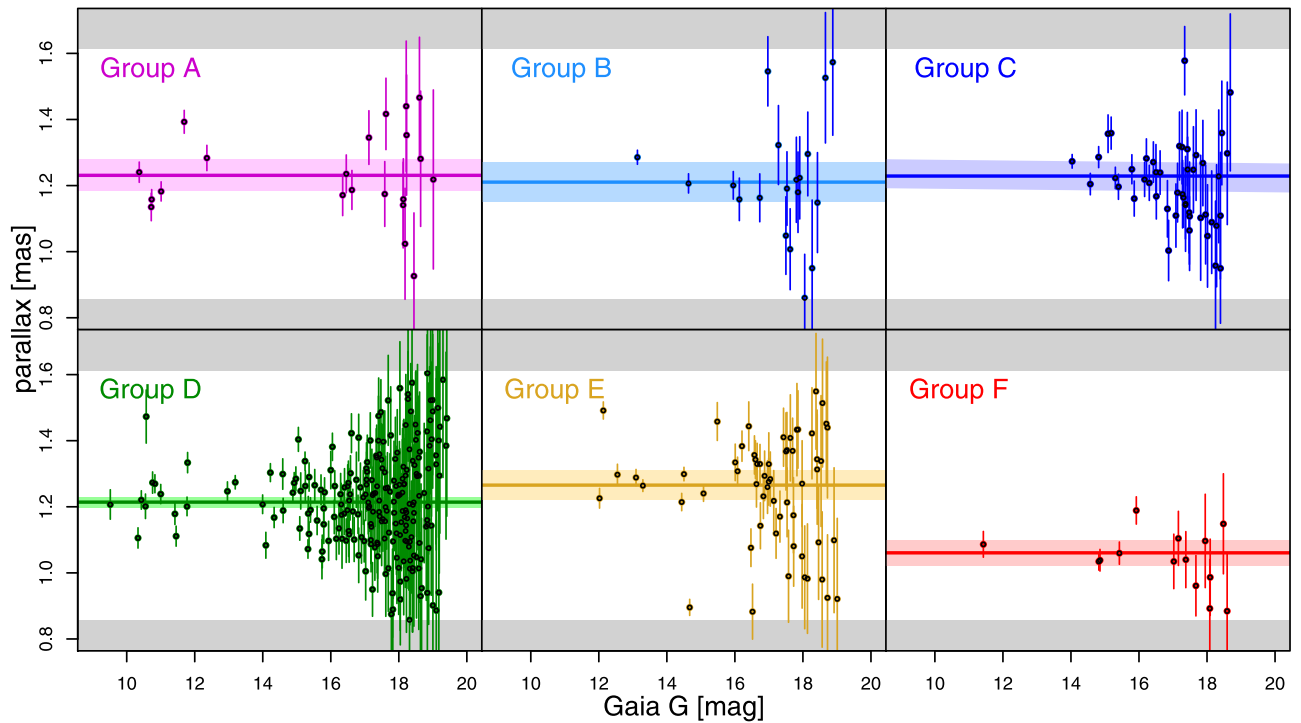


Figure 7. Mean parallax (colored horizontal line) and scatter plot of each star’s measured parallax vs. magnitude (points) for stars in each group. Uncertainty on the mean parallax is indicated by the colored band around the line, which shows the 2σ confidence interval. The gray regions at the top and bottom of each graph indicate the range of parallaxes that were excluded by the selection cut made when identifying members.

with cloud clumps in the “Gulf of Mexico” region, which are visible as dark dust lanes in the optical images, suggesting that both the cloud and group are in front of the H II region.

The Bajamar Star is obscured by ~ 10 mag of cloud near the periphery of the “Gulf of Mexico.” Thus, it is implausible that this star could be nearer than the rest of the complex as its measured parallax indicates. More likely, the Bajamar Star is at a distance similar to that of the complex, and its slightly larger measured parallax is a ~ 2 standard deviation statistical error (expected to occur with a probability of 1 in 20), with the binarity of the system possibly playing some systematic role in the offset. Furthermore, although heavily obscured, this star cannot be entirely embedded within the cloud because clear lines of sight are needed for this star to illuminate the H II region and the bright rim clouds. The “Gulf of Mexico” is superimposed between the Bajamar Star and the southeast bright rim cloud (“Pacific Coast of Mexico”), requiring the cloud to be in front of the line of sight between this star and this rim.

The order in distance of the other components is more difficult to ascertain given the data. The cloud in the “Atlantic” region (possibly associated with Group D) is dark in the optical image, suggesting that it is mostly in front of the H II region. Group C sits atop the northwest bright rim (“Pelican’s Neck”). Thus, a clear line of sight from the Bajamar Star to this bright rim would put the “Atlantic Cloud” and Group D in front of the “Pelican’s Neck Cloud” and Group C. Both the optically dark “Pelican’s Hat” cloud, associated with Group B, and the cloud around Group A can be seen in absorption, suggesting that they are also mostly in front of the H II region.

Other group properties listed in Table 3 include average proper motions, approximate centers of the groups, and the radii that contain half the groups’ stars. To calculate the proper motion, we used the same strategy as Kuhn et al. (2019) and

calculate the weighted median with the conventional $1/\text{error}^2$ weights. To estimate the group centers, we calculate the σ -clipped mean α and δ values, with a 3 standard deviation threshold to avoid strong influences from outlying stars.

6. Structure of the Molecular Cloud

The ^{13}CO map provides kinematic information about the system that is complementary to the stellar proper motions measured by Gaia. Figure 8 shows the integrated ^{13}CO emission over a velocity range of $v_{\text{lsr}} = -17$ to 10 km s^{-1} , which encompasses nearly all the line emission associated with the star-forming region. Dominant clouds are located in the “Gulf of Mexico,” “Pelican,” “Pelican Hat,” and “Atlantic” regions. These areas tend to have gas at different velocities, with differences ranging from several to tens of kilometers per second. Figure 9 shows integrated intensity maps and first-moment maps for three adjoining velocity ranges. Clouds in the southern “Gulf of Mexico” region tend to have more positive velocities, while gas in the centrally located “Atlantic” region forms a ~ 15 pc long filamentary structure with more negative velocities, and the eastern “Pelican” and “Pelican Hat” regions contain gas with both positive and negative velocities.

An early radio study of the region by Bally & Scoville (1980) suggested that the complex is in the process of expansion. Zhang et al. (2014) also attribute the cloud morphology in the “Pelican” region to an expanding shell around the H II region. We find that some parts of the cloud have kinematics consistent with expansion, but that the expansion is not global. The shell-like structure is most distinct at -1 km s^{-1} in the northwest. In contrast, in the south, the “Gulf of Mexico” cloud appears to be infalling, not outward-moving. The ~ 15 pc long filament at -5 km s^{-1} , which stretches from the north to the south, appears to be moving

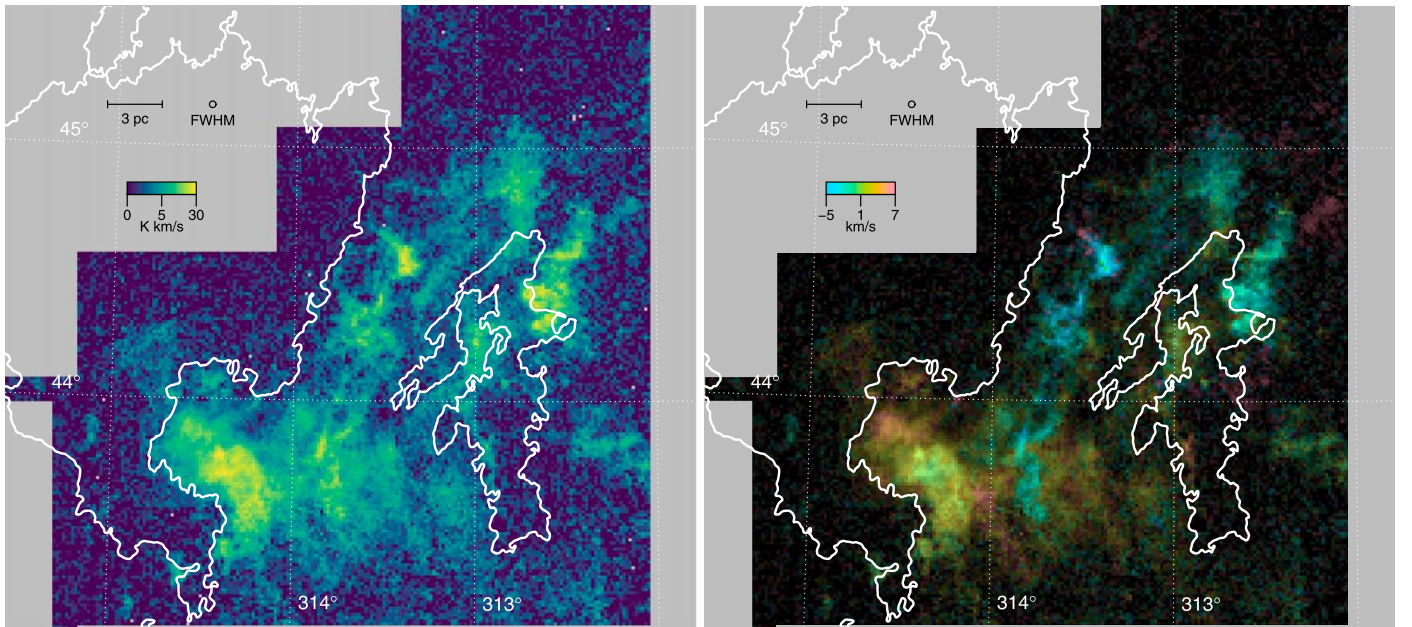


Figure 8. Left: integrated $^{13}\text{CO } J = 1-0$ emission over the velocity range $v_{\text{lsr}} = -17$ to 10 km s^{-1} . Right: the color scale for the ^{13}CO map uses intensity to indicate integrated emission and hue to indicate the first moment. The FWHM of the observations and the physical scale of the map are indicated in the upper left corner. We also include the outline of the optical emission from the DSS image (white contours).

away from the H II and toward us with a fairly coherent radial velocity across its whole length.

To investigate the molecular cloud structure in more detail, in Figure 10 we subdivide the system into individual clouds with contours at 6 K km s^{-1} from the integrated map. This threshold picks out most of the main cloud components, which are labeled 1 through 12 both in the figure and in Table 4. We also include a Cloud 13 defined by contours at 3 K km s^{-1} that is associated with stellar Group F but does not meet the 6 K km s^{-1} threshold.

From the brightness of the $^{13}\text{CO } J = 1-0$ emission line, it is possible to estimate cloud masses if we make certain assumptions. To do so, we must accept that, as in most CO studies, there are implied systematic effects on the results. For example, the ^{13}CO emission is less sensitive to diffuse gas traced by ^{12}CO but cannot trace dense gas where the ^{13}CO line becomes too optically thick. To calculate column densities, we assume local thermodynamic equilibrium, an excitation temperature, and the ratio of ^{13}CO to H_2 . Zhang et al. (2014) use a ^{12}CO map from the Purple Mountain Observatory Delingha 13.7 m telescope to estimate excitation temperature throughout the cloud complex. They find temperatures ranging from 5 to 25 K, with the bulk of the cloud around $\sim 10-15 \text{ K}$. This includes the massive ‘‘Gulf of Mexico’’ region, which is mostly $\sim 14 \text{ K}$. However, several areas have higher temperatures (20–25 K), particularly some edges of the clouds in the north of the NAP complex, including the cloud behind the bright rim in the ‘‘Pelican’s Neck’’ and a cloud filament in the ‘‘Atlantic’’ region. Consistent with the higher inferred temperatures, these regions appear geometrically oriented so they would be illuminated by the Bajamar Star. We note that the measured ^{12}CO excitation temperatures would be representative of the cloud surfaces, but the cloud interiors may not be the same.

We follow the method outlined by Mangum & Shirley (2015) using the coefficients from the JPL Molecular Spectroscopy Database.⁷ Assuming a ^{13}CO excitation temperature of

$T_{\text{ex}} = 14 \text{ K}$ (as justified above), we estimate the optical depth of the $^{13}\text{CO } J = 1-0$ transition with the equation

$$\begin{aligned} \tau_{J=1}^{13}(v) &= -\ln \left\{ 1 - \frac{T_R^{13,1}(v)}{5.29 \text{ K}} \left[\frac{1}{\exp(5.29 \text{ K}/T_{\text{ex}}) - 1} - 0.17 \right]^{-1} \right\}, \end{aligned} \quad (2)$$

using T_R^* as an approximation for the radiation temperature T_R in this equation. This equation is applied to each pixel, channel by channel, in the emission-line data cube. From the distribution of derived optical depths, $\sim 95\%$ of the ^{13}CO gas by mass has optical depths $\tau < 0.5$, while $\sim 98\%$ has optical depth $\tau < 1$. Although ^{13}CO is not sensitive to the highest-density parts of the cloud, and thus cannot account for mass in these regions, the shape of the distribution suggests that much of the cloud is only moderately optically thick. The ^{13}CO column density would then be

$$\begin{aligned} N_{\text{tot}}(^{13}\text{CO}) &\approx (6.6 \times 10^{14} \text{ cm}^2) \left(\frac{T_{\text{ex}}}{2.64 \text{ K}} + \frac{1}{3} \right) \\ &\times \left[1 - \exp\left(-\frac{5.29 \text{ K}}{T_{\text{ex}}}\right) \right]^{-1} \int \frac{\tau_{J=1}^{13}(v) dv}{\text{km s}^{-1}}. \end{aligned} \quad (3)$$

Finally, we assume that $N(\text{H}_2) \approx 3.8 \times 10^5 N(^{13}\text{CO})$ (Bolatto et al. 2013) to calculate $N(\text{H}_2)$ column density.

Integrating column density over the entire FCRAO ^{13}CO map yields a total mass of $7.2 \times 10^4 M_{\odot}$. Masses of individual clouds are given in Table 4. Uncertainty in these quantities is likely to be dominated by systematic errors in the assumptions, particularly the ^{13}CO -to- H_2 ratio (Bolatto et al. 2013). In addition, clumpiness in the clouds that is not resolved by the FCRAO beam could cause us to systematically underestimate the amount of gas in regions with high optical depth.

⁷ <https://spec.jpl.nasa.gov>

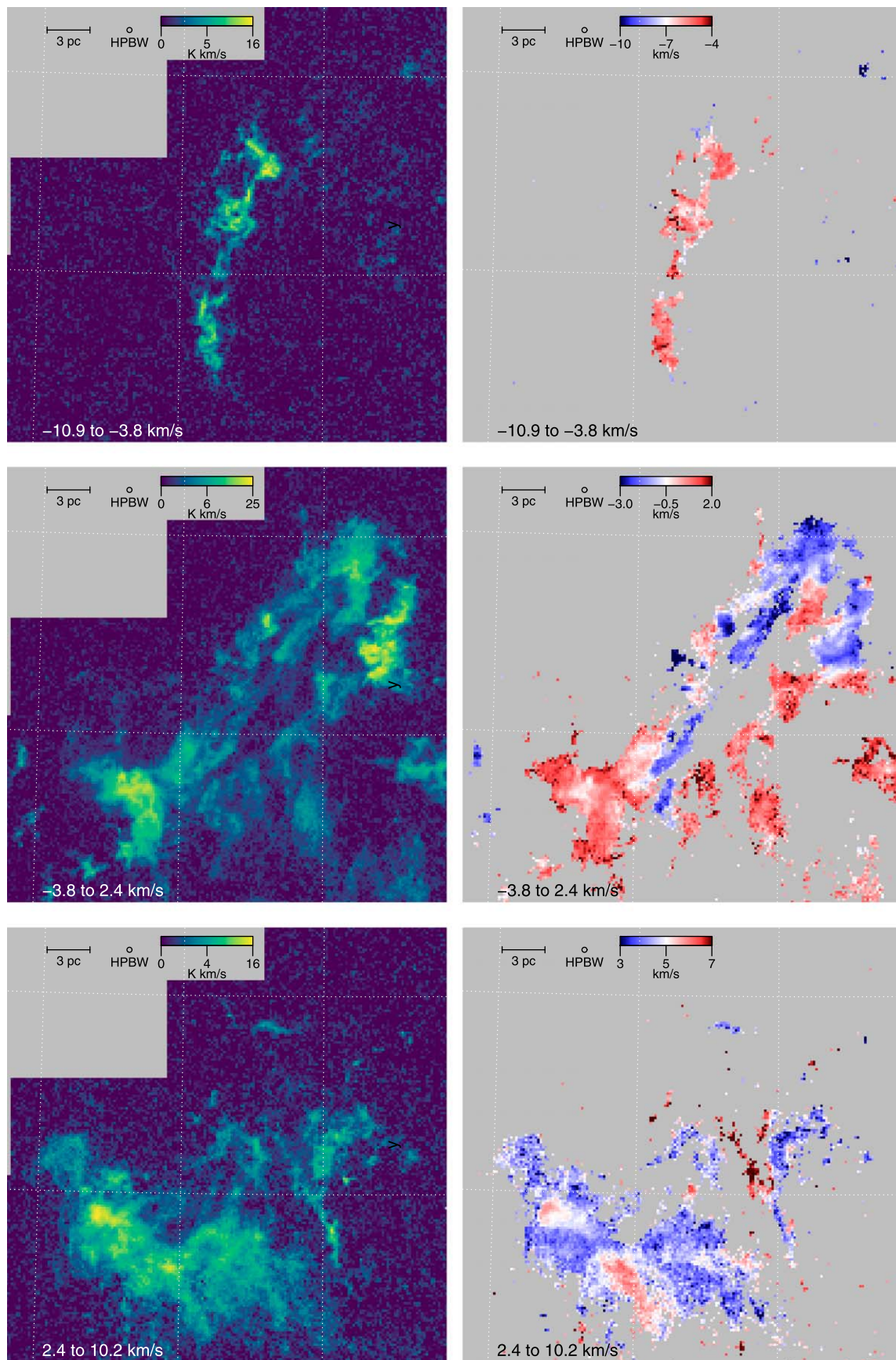


Figure 9. Integrated intensity maps (left panels) and first-moment maps (right panels) over three velocity ranges, from top to bottom in order of increasing velocity. These three ranges highlight different aspects of the cloud’s structure. The top panel (-10.9 to -3.8 km s^{-1}) reveals a ~ 15 pc long filamentary structure; the middle panel (-3.8 to 2.4 km s^{-1}) reveals structures spanning the whole region, including a vaguely shell-like structure in the north; and the bottom panel (2.4 – 10.2 km s^{-1}) reveals a large cloud complex concentrated in the “Gulf or Mexico” region.

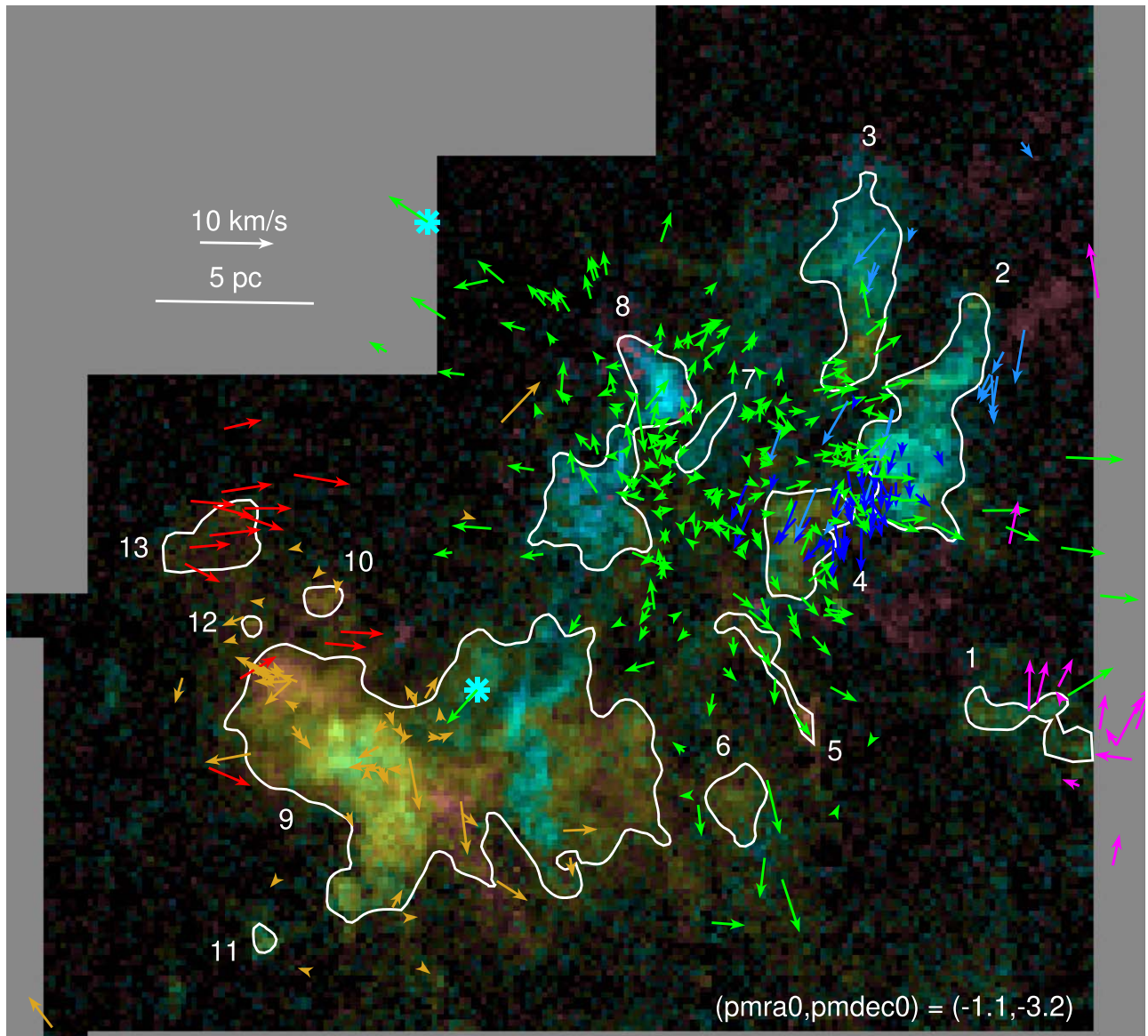


Figure 10. Map of star positions and velocities, indicated by arrows, superimposed on the ^{13}CO map. Arrows are drawn relative to the median proper motions of the system, which we define to be $(\mu_{\alpha^*}, \mu_{\delta})_0 = (-1.1, -3.2)$ mas yr^{-1} . Arrows are color-coded by the stellar groups A–F, and the outlines of the molecular clouds 1–13 are also shown. The two O stars are indicated by the cyan asterisks.

6.1. Relation of Clouds to Stellar Groups

There is spatial correspondence between the stellar groups and the clouds (Figure 10). Groups A and F are associated with the minor clouds 1 and 13, respectively. Group E is embedded in the “Gulf of Mexico,” which consists of a major cloud component (Cloud 9) and several minor cloud components (Clouds 6, 10, 11, and 13) in the ^{13}CO map. Group C lies on a bright rim cloud at the southeastern edge of Cloud 2. The stars in Group B are distributed in the northern part of Cloud 2, as well as Cloud 3.

Group D, the most spatially extended stellar group, has the most complicated relationship to the ^{13}CO clouds. This group is projected over a region that spans several clouds. However, Cloud 8 is the dominant cloud in this part of the NAP region and lies near the center of Group D. The stars selected in our Gaia analysis do not lie on top of Cloud 8, but instead the densest parts of Group D are next to the cloud. This indicates either that Group D is partially embedded in Cloud 8, which would imply that Cloud 8 is the remainder of the natal cloud that formed this

group, or that Group D lies behind the cloud. Stars from this group are also superimposed on other nearby minor clouds (Clouds 5, 6, and 7), which may either be smaller remnants of the natal cloud or chance superpositions. The edges of Group D also intersect the edges of Cloud 2, Cloud 3, and even (in the case of the Bajamar Star) Cloud 9.

If the stellar groups and clouds are associated, then it is likely that they have similar mean motions. This enables us to associate mean proper motions with clouds and radial velocities with stars, creating a three-dimensional picture of the velocities in this region.

6.2. Velocity Structure of Clouds

Line profiles for each of the clouds (Figure 11) were constructed by spatially integrating the ^{13}CO gas detected within the contours on Figure 10. Most of these profiles display several distinct peaks. In Table 4, we report basic statistics of the velocity distributions, including the first (mean) and square root of the second (standard deviation) moments, the mode, and

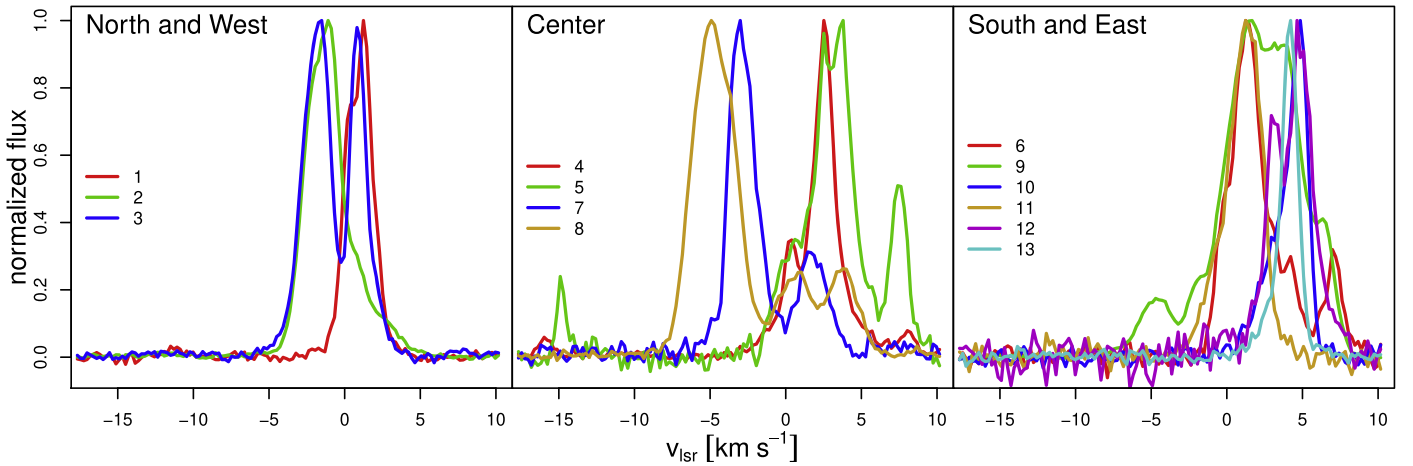


Figure 11. Spatially integrated line flux as a function of v_{lsr} for each of the clouds in Figure 10. The three panels show velocity profiles for clouds in three different areas of the NAP region. Molecular gas spans the velocity range -15 to 10 km s^{-1} , and several of the clouds have multimodal structure.

Table 4
Clouds Properties from the ^{13}CO Line Map

Desig.	α (ICRS)	δ (ICRS)	$M_{^{13}\text{CO}}$ (M_{\odot})	v_{lsr} Distribution				r_{cloud} (pc)	Optical (10)	Group (11)
				Mean (km s^{-1})	Std. dev. (km s^{-1})	Mode (km s^{-1})	FWHM (km s^{-1})			
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)		
1	20 48 58	+43 48 10	330	0.8	1.9	1.2	1.9	1.1	Dark	A
2	20 50 16	+44 29 40	2100	-1.1	2.2	-1.1	2.5	1.7	BRC ^a /Dark	B,C
3	20 51 12	+44 48 00	1100	-0.9	2.4	-1.5	1.9	1.3	Dark/Neb.	B
4	20 51 42	+44 12 30	610	1.7	4.0	2.5	1.3	1.0	Dark/Neb.	D
5	20 52 16	+43 54 20	170	2.8	4.4	3.8	2.5	1.2	Neb.	D
6	20 52 37	+43 37 10	220	2.2	2.7	1.2	2.3	0.7	Dark	E
7	20 53 01	+44 27 30	120	-1.7	3.5	-3.0	1.9	0.7	Dark	D
8	20 54 12	+44 24 40	1500	-2.8	3.7	-4.9	3.2	2.2	Dark	D
9	20 56 13	+43 41 40	8000	2.0	3.0	1.7	5.3	3.5	Dark	E
10	20 57 48	+44 04 00	80	4.4	1.5	4.9	1.5	0.4	Dark	E
11	20 58 26	+43 18 00	50	0.6	2.0	1.2	2.5	0.2	Dark	E
12	20 58 42	+44 00 20	30	4.2	1.4	4.6	2.8	0.1	Dark	E
13	20 59 13	+44 12 20	250	3.7	2.2	4.2	1.1	0.9	Neb.	F

Notes. Column (1): cloud designation. Columns (2)–(3): coordinates of the cloud center. Column (4): ^{13}CO mass estimate of the cloud. Column (5): mean velocity of the cloud. Column (6): standard deviation of the velocity distribution. Column (7): mode of the velocity distribution. Column (8): Δv_{FWHM} for the most prominent mode. Column (9): characteristic size of the cloud, defined as the radius that contains half the integrated emission. Column (10): appearance of the cloud (or nebular material in the direction of the cloud) in the optical image. Column (11): group of stars associated with the cloud.

^a BRC = bright rim cloud.

the FWHM of the tallest mode. The mode and FWHM (Δv_{FWHM}) may be more characteristic of the kinematics of the clouds because these statistics are less susceptible to merging components with very different velocities that may be from distinct clouds. Thus, we base our discussion of cloud dynamics (Section 6.3) on the latter quantities. We also estimate the characteristic size of each cloud, defined as the projected radius that contains half the integrated emission.

Optical depth can affect line profiles, flattening the peaks and effectively broadening the lines (Hacar et al. 2016). To more accurately resolve the velocity structure, we integrated over the optical depths given by Equation (2) rather than T_R^* (see Goldsmith & Langer 1999). This results in line profiles that are more sharply peaked with FWHM that are $\sim 20\%$ narrower.

The interpretation of the multimodal structure becomes clearer with the position–velocity diagrams in Figure 12. Four of the

major clouds (Clouds 2, 3, 8 and 9) are shown as examples, while the diagrams for the other clouds are accessible via the online figure set. Even some of the minor clouds (e.g., Clouds 5, 6, and 7) have complicated velocity structures.

In Cloud 2, v_{lsr} increases by several kilometers per second toward the southern end of the cloud, where the optically bright rim and stellar group are located. It is plausible that this could be an effect of the expansion of the H II region pushing this end of the cloud away.

For Clouds 3 and 8, multiple components are visible, corresponding to the multiple peaks in the line profiles. In Cloud 3, there are two velocity components separated by ~ 5 km s^{-1} , while in Cloud 8 there are three components, also separated by ~ 5 km s^{-1} from each other. The distinct components in Cloud 3 are also mostly separated along a north–south axis, possibly indicating that this cloud is composed of multiple subclouds.

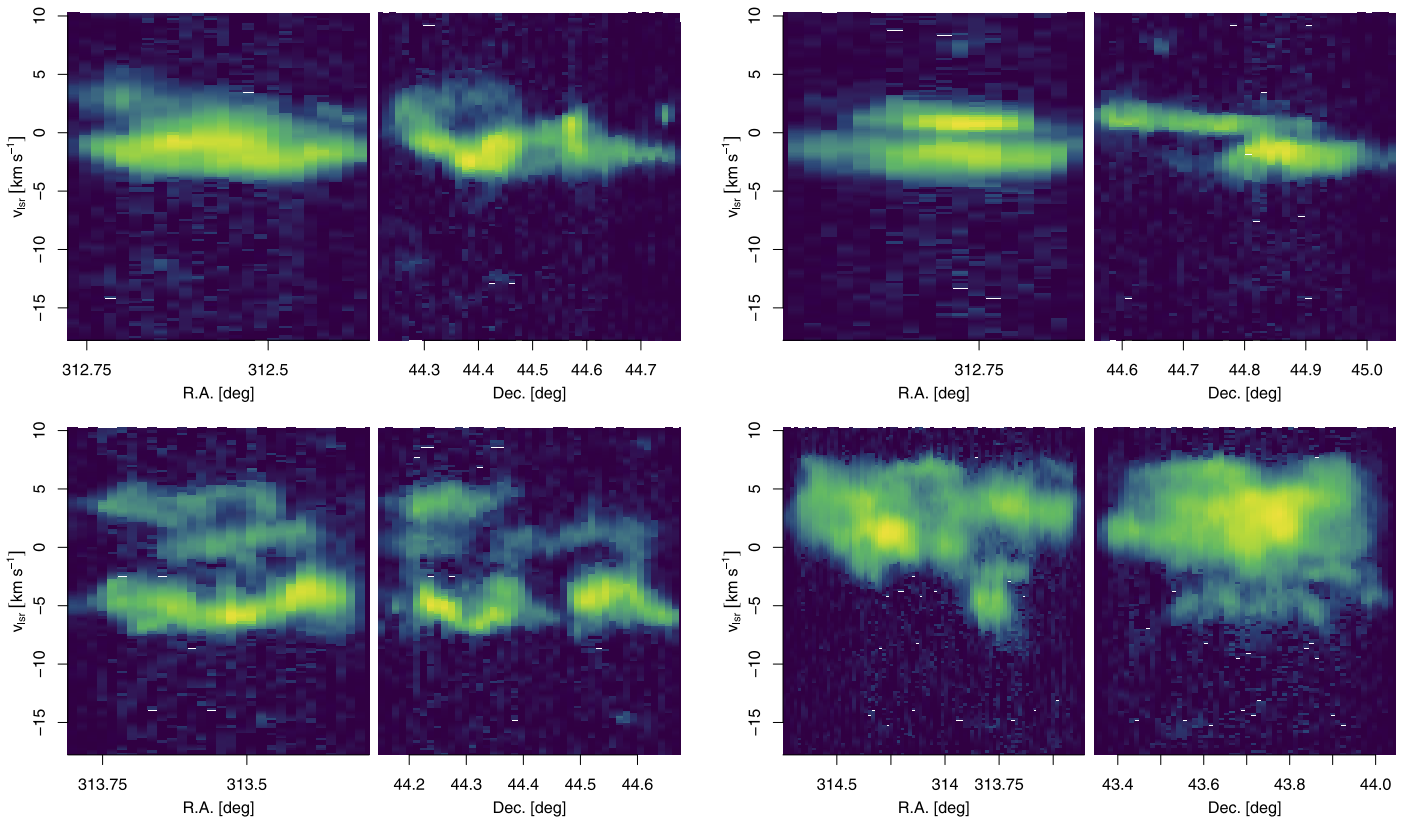


Figure 12. Position–velocity diagrams of $^{13}\text{CO } J = 1\text{--}0$ emission are shown for Clouds 2, 3, 8, and 9 (top left, top right, bottom left, and bottom right, respectively). (The complete figure set (13 images) is available.)

However, for Cloud 8 the main component is the dense northern end of the filament at $v_{\text{lsr}} \approx -5 \text{ km s}^{-1}$ (Figure 9), while the other components are related to overlapping filamentary structures at higher v_{lsr} . In our analysis, we tentatively associate the stars in Group D with the -5 km s^{-1} gas because this is the dominant component around which the stars are conglomeraed, but our data are insufficient to make a definitive conclusion (Appendix C).

The position–velocity diagram for Cloud 9 in Figure 12 shows that it has a clumpy structure, with clumps spanning the velocity range from ~ 0 to 7 km s^{-1} . A small clump at -5 km s^{-1} is probably not part of this cloud, but is instead a continuation of the long north–south filament seen in Figure 9 (top panels). Overall, Cloud 9 has the largest positive v_{lsr} of the complex, which, in combination with our evidence that the “Gulf of Mexico” and Group E are in front of the rest of the NAP region, suggests that Cloud 9 is infalling. However, the part of the cloud brightest in ^{13}CO emission also has a fairly modest velocity of $\sim 1 \text{ km s}^{-1}$. This ^{13}CO bright patch is fairly near the Bajamar Star and shows up as having a higher temperature in the ^{12}CO map from Zhang et al. (2014).

6.3. Dynamical Properties of the Clouds

From the kinematic measurements of the cloud in Table 4, several dynamical quantities can be calculated that are useful for understanding star formation in this region. For these calculations, we use the line widths for the most prominent peaks in Figure 12 to avoid combining unrelated velocity components. The dynamical quantities are given in Table 5.

If we assume that the kinetic temperature T_{kin} of the gas equals the excitation temperature measured for ^{12}CO ($\sim 14 \text{ K}$)

Table 5
Cloud Dynamical Properties

Cloud	$\log \bar{\rho}$ (g cm^{-3})	τ_{cross} (Myr)	τ_{ff} (Myr)	\mathcal{M}	M_{dyn} ($10^3 M_{\odot}$)	α_{BM92}
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	−20.40	1.3	1.1	3.6	0.8	2.5
2	−20.17	1.6	0.8	4.7	2.2	1.1
3	−20.10	1.6	0.8	3.6	1.0	0.9
4	−20.01	1.8	0.7	2.5	0.3	0.6
5	−20.80	1.1	1.7	4.7	1.6	9.3
6	−19.98	0.7	0.7	4.4	0.8	3.5
7	−20.24	0.9	0.9	3.6	0.5	4.3
8	−20.65	1.6	1.4	6.1	4.7	3.2
9	−20.52	1.5	1.2	10.1	20.6	2.6
10	−19.67	0.6	0.5	2.8	0.2	2.2
11	−19.01	0.2	0.2	4.7	0.3	5.4
12	−18.38	0.1	0.1	5.3	0.2	6.4
13	−20.26	1.9	0.9	2.1	0.2	0.9

Note. Column (1): cloud designation. Column (2): mean density of the cloud, defined as $\bar{\rho} = M_{^{13}\text{CO}} / (4/3)\pi r_{\text{cloud}}^3$. Column (3): crossing timescale, defined as $\tau_{\text{cross}} = r_{\text{cloud}} / \Delta v$. Column (4): freefall timescale defined using $\bar{\rho}$ from Column (3). Column (5): Mach number. Column (6): dynamical mass of the cloud assuming $\alpha = 1$ (Equation (9)). Column (7): Bertoldi & McKee (1992) estimate of the virial parameter, using $M_{^{13}\text{CO}}$ for the mass, r_{cloud} for the radius, and σ_{nl} for the velocity dispersion.

and that the mean molecular mass is $\bar{m} = 2.33 \text{ amu}$, the one-dimensional sound speed in the molecular clouds would be $c_s \approx 0.22 \text{ km s}^{-1}$. For each of the molecular clouds delimited in our analysis, the line width of the tallest mode is larger than

can be accounted for by purely thermal broadening. The nonthermal component is

$$\sigma_{\text{nt}} = \sqrt{\left(\frac{\Delta v_{\text{FWHM}}}{2.35}\right)^2 - \frac{kT_{\text{kin}}}{m_{\text{obs}}}}, \quad (4)$$

where $m_{\text{obs}} = 29$ amu is the mass of the observed molecule, ^{13}CO . The Mach number is defined to be $\mathcal{M} = \sigma_{\text{nt}}/c_s$. For most of the individual clouds, Mach numbers span $2 \lesssim \mathcal{M} \lesssim 5$. Exceptions are Clouds 8 and 9—the most massive clouds—which have larger velocity dispersions of $\mathcal{M} \approx 6$ and 10, respectively. The lowest velocity dispersion is found for Cloud 13 with $\mathcal{M} \approx 2$; this also happens to be the cloud with the lowest surface density. This range of Mach numbers is fairly typical for star-forming clouds, such as the Orion B cloud ($\mathcal{M} \sim 6$; Orkisz et al. 2017) or Taurus ($\mathcal{M} < 2$ –3; Hacar et al. 2016). The position–velocity plots (Figure 12) show that some of this nonthermal broadening in the NAP clouds is due to cloud-scale velocity structure. In Cloud 8, the -5 km s^{-1} component appears curved in position–velocity space, while, in Cloud 9, the large velocity dispersion is composed of multiple clumps with different velocities.

Several dynamically important quantities that can be estimated for each cloud include the mean density $\bar{\rho}$, freefall timescale τ_{ff} , and crossing timescale τ_{cross} defined as

$$\bar{\rho} = M^{13\text{CO}} / (4/3\pi)r_{\text{cloud}}^2 \quad (5)$$

$$\tau_{\text{ff}} = \sqrt{3\pi/32G\bar{\rho}} \quad (6)$$

$$\tau_{\text{cross}} = r_{\text{cloud}} / \sigma_{\text{nt}}. \quad (7)$$

Both the crossing timescales and the freefall timescales range from 0.1 to 2 Myr. For the main clouds with ongoing star formation, the typical freefall timescale is $\tau_{\text{ff}} \sim 1$ Myr. We note that these timescales are similar to the ~ 1 Myr ages inferred for the pre–main-sequence stars.

The velocity dispersions can be used to estimate cloud dynamical masses. However, these calculations depend on the virial ratio of the clouds, $\alpha = 2T/|W|$, where T is kinetic energy and W is potential energy. Previous studies typically use the definition from Bertoldi & McKee (1992) that assumes a spherical, uniform density cloud, giving

$$\alpha_{\text{BM92}} = \frac{5\sigma_{\text{nt}}^2 r_{\text{cloud}}}{GM_{\text{cloud}}}. \quad (8)$$

However, the true value of α depends on the three-dimensional distribution of mass within the cloud (Singh et al. 2019; Guszejnov et al. 2020), with corrections for centrally concentrated mass distributions and filamentary structure typically being a factor of several (Bertoldi & McKee 1992). If we let $\alpha_{\text{BM92}} = 1$, we obtain a dynamical mass estimate

$$M_{\text{dyn}} = 5 r_{\text{cloud}} \sigma_{\text{nt}}^2 / G. \quad (9)$$

In Figure 13, we plot $\log M_{\text{dyn}}$ versus $\log M^{13\text{CO}}$ for the 13 clouds we have identified in the NAP region. These quantities are strongly correlated ($p < 0.001$ from Kendall’s τ test), with an approximately proportional relationship over a dynamic range of ~ 200 in cloud mass. However, the dynamical masses are systematically ~ 0.4 dex higher, with ~ 0.4 dex scatter in the relation. The similarity in masses derived by the two methods to within a factor of ~ 3 helps corroborate our ^{13}CO -based mass

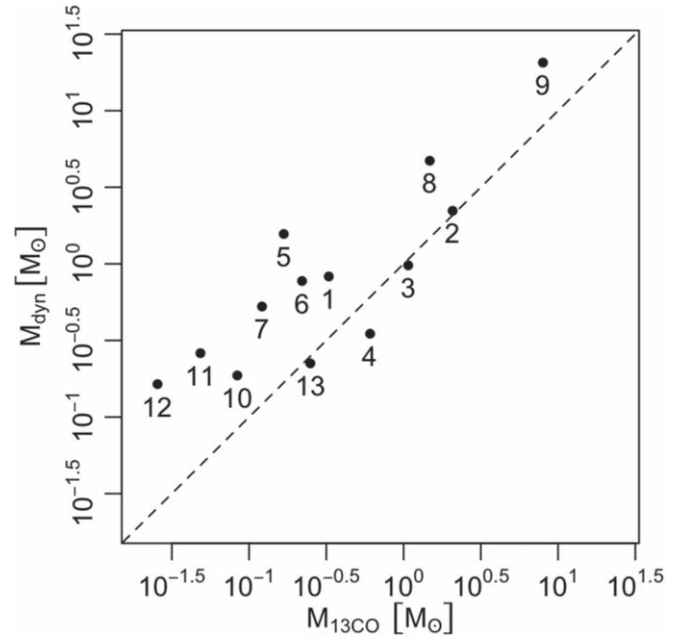


Figure 13. Dynamical cloud masses vs. mass estimates from ^{13}CO column density. The dashed line indicates equivalence between these quantities, and the points are labeled by the cloud they represent. Systematic uncertainties on ^{13}CO -based mass estimates may be a factor of several as discussed in the text.

estimates in Table 4. Vázquez-Semadeni et al. (2019) point out that, observationally, it is nearly impossible to distinguish between the freefall and virial velocities, which only differ by a factor of 2 in mass or a factor of $\sqrt{2}$ in velocity. Table 5 provides both M_{dyn} , calculated from the velocity dispersions using Equation (9), and α_{BM92} , calculated from both the velocity dispersions and the integrated ^{13}CO cloud masses using Equation (8).

7. Stellar Kinematics

For the analysis of stellar kinematics, we convert the observed astrometric quantities of parallax and proper motion into physical velocities with units of km s^{-1} . For this we establish a Cartesian coordinate system with positions (x, y, z) and velocities (v_x, v_y, v_z) , where we select the origin to be located at the center of the system that we are analyzing and moving with the median velocity of the stars in the system. Given the nature of the data, we are most interested in examining kinematics in the two-dimensional xy plane. We define x to be parallel to α at the origin and y to be parallel to δ at the origin. The transformations, including orthographic projections, corrections for perspective expansion,⁸ and shifts to the rest frame of the system, are described by Kuhn et al. (2019, their Equations (1)–(4)).

We are also interested in the component of a star’s projected velocity in the direction outward/inward relative to the center of the group (v_{out}), as well as the perpendicular azimuthal component (v_{az}). These are defined by

$$v_{\text{out}} = \mathbf{v} \cdot \hat{\mathbf{r}} \quad (10)$$

$$v_{\text{az}} = \mathbf{v} \cdot \hat{\boldsymbol{\phi}}, \quad (11)$$

⁸ To compute the perspective expansion correction, we assume that the stellar group is moving with the radial velocity of the most significant of the cloud components associated with each group (Table 4) and transformed to heliocentric coordinates.

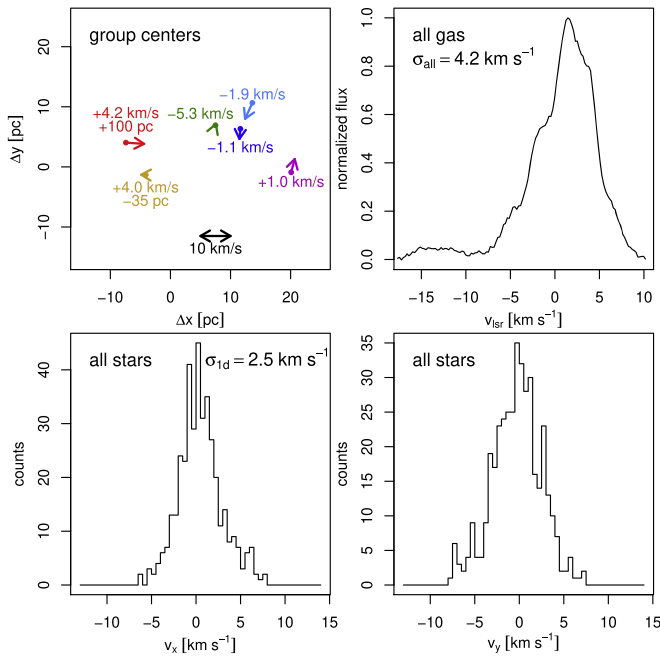


Figure 14. Diagrams illustrating global motions in the NAP region. Top left: relative 3D motions of the group centers (reference frame $\mu_{\alpha^*,0} = -1.1$ mas yr $^{-1}$, $\mu_{\delta,0} = -3.2$ mas yr $^{-1}$). Velocities in the plane of the sky from Gaia are shown by the arrows, while velocities from the molecular gas along the line of sight (v_{lsr}) and offsets in distance are indicated by the labels. Top right: spatially integrated $^{13}\text{CO } J = 0-1$ line flux for the whole NAP complex. The velocity dispersion is 4.2 km s^{-1} . Bottom panels: velocity distributions (v_x and v_y) for the entire NAP stellar population. The total velocity dispersion for stars is $\sigma_{\text{1D}} = 2.5 \pm 0.1 \text{ km s}^{-1}$. The scale in km s^{-1} is the same for all three plots of velocity distributions.

where \hat{r} and $\hat{\phi}$ are the radial and azimuthal unit vectors relative to the group center in the xy plane. Uncertainties in these velocities are calculated by propagation of the Gaia astrometric errors (see procedure in Kuhn et al. 2019).

7.1. Global Kinematics

In Figure 10, we indicate the motions of NAP members using arrows, which, as in the previous figures, are color-coded by stellar group. The figure shows that stars near each other tend to have similar velocities, but the different groups appear to have fairly random motions relative to one another.

Figure 14 (top left) shows the relative 3D motions of the group centers inferred from both the mean proper motions of stars and the radial velocities of the associated clouds. This diagram shows no clear pattern of either convergence or divergence. Stars in Group A tend to be moving mostly north and slightly west—in a direction tangential to the NAP system center. Stars in both Groups B and C are moving southeast, inward toward the center of the system. Group E is drifting slowly east, away from the center of the NAP complex in the xy plane, but it is likely moving rapidly toward the system center in the third dimension. Stars in Group F are moving west—toward the center of the region. Finally, the center of Group D appears to be nearly stationary in the xy plane in this reference frame. However, most of the individual stars in this group are directed away from its own center as can be seen in Figure 10.

The Bajamar Star is not stationary relative to the center-of-mass reference frame, but is instead rapidly traveling southeast away from Group D and toward Group E. The other O star, HD 199579, is traveling northeast away from the NAP

complex. In this global reference frame,⁹ both stars have velocities $>6 \text{ km s}^{-1}$. Such speeds would classify both objects as “walkaways” (Renzo et al. 2019; Schoettler et al. 2020), although neither object has traveled outside the star-forming region.

7.2. Kinematics within Groups

In Kuhn et al. (2019), we presented several methods for characterizing stellar kinematics and testing for expansion, contraction, or rotation of a stellar system. Here, we apply these methods to each of the stellar groups.

To characterize total velocity dispersion in a group, we fit the (v_x, v_y) distribution with a bivariate Gaussian, taking into account measurement uncertainties, which have the effect of artificially broadening the velocity dispersion.¹⁰ The stellar velocities and the Gaussian fit are shown in Figure 15 for each of the six groups. We summarize these distributions using the quantity σ_{1D} (Table 6), defined as the square root of half the trace of the Gaussian’s covariance matrix (Kuhn et al. 2019, their Equation (17)). For the six groups, σ_{1D} ranges from 1 to 2 km s^{-1} , being smallest for Groups B and F and largest for D and E. Group F is a small group in a small cloud, so it is unsurprising that this group has a small velocity dispersion. In contrast, Groups D and E are large groups, located in a massive cloud (E) or in a partially dispersed cloud (D), so it is unsurprising that they have larger velocity dispersions. However, it is notable that Group C, spatially adjacent to Group D, has a much lower velocity dispersion. In many cases, the velocity dispersions show signs of being anisotropic, similar to the results for other star-forming regions (Kuhn et al. 2019).

We use several plots that are analogous to those in Kuhn et al. (2019), to investigate evidence for expansion. Figure 16 shows Group D, the largest group in our sample where expansion is most evident. In this group, the fastest-moving stars, with velocities $>5 \text{ km s}^{-1}$, are located near the edges of the region, and they are generally directed away from the center of the group. The top right panel uses both arrow direction and hue to indicate the direction of motion, with color saturation indicating statistical uncertainty on direction. This use of color allows patterns of motions to be identifiable as a gradient in the color of the marks, in this case showing that stars at different positions in the group are oriented in different directions, all preferentially away from the center.

The two bottom panels show statistical tests for expansion. The bottom left panel shows the distribution of v_{out} values; if a system is expanding, then these will be predominantly positive. For Group D, the weighted median v_{out} value is $1.9 \pm 0.4 \text{ km s}^{-1}$, more than 3 standard deviations above 0, clearly demonstrating that the group is expanding. This expansion velocity is larger than any of those measured for the 28 systems investigated by Kuhn et al. (2019). The bottom right panel shows the mean expansion velocities for bins at different radii from the group center. Expansion velocity appears to increase with distance from the

⁹ In Section 7.3, we recalculate the O stars’ velocities with respect to Group D.

¹⁰ The classification of sources using proper motion in Section 3 restricts the allowable range of velocities, which truncates the tails of the velocity distributions and will slightly decrease the estimated σ_{1D} . We simulate this effect and find that it will produce a $<8\%$ effect on the results for velocity dispersions $<3 \text{ km s}^{-1}$, which is smaller than the estimated statistical uncertainty on the derived values.

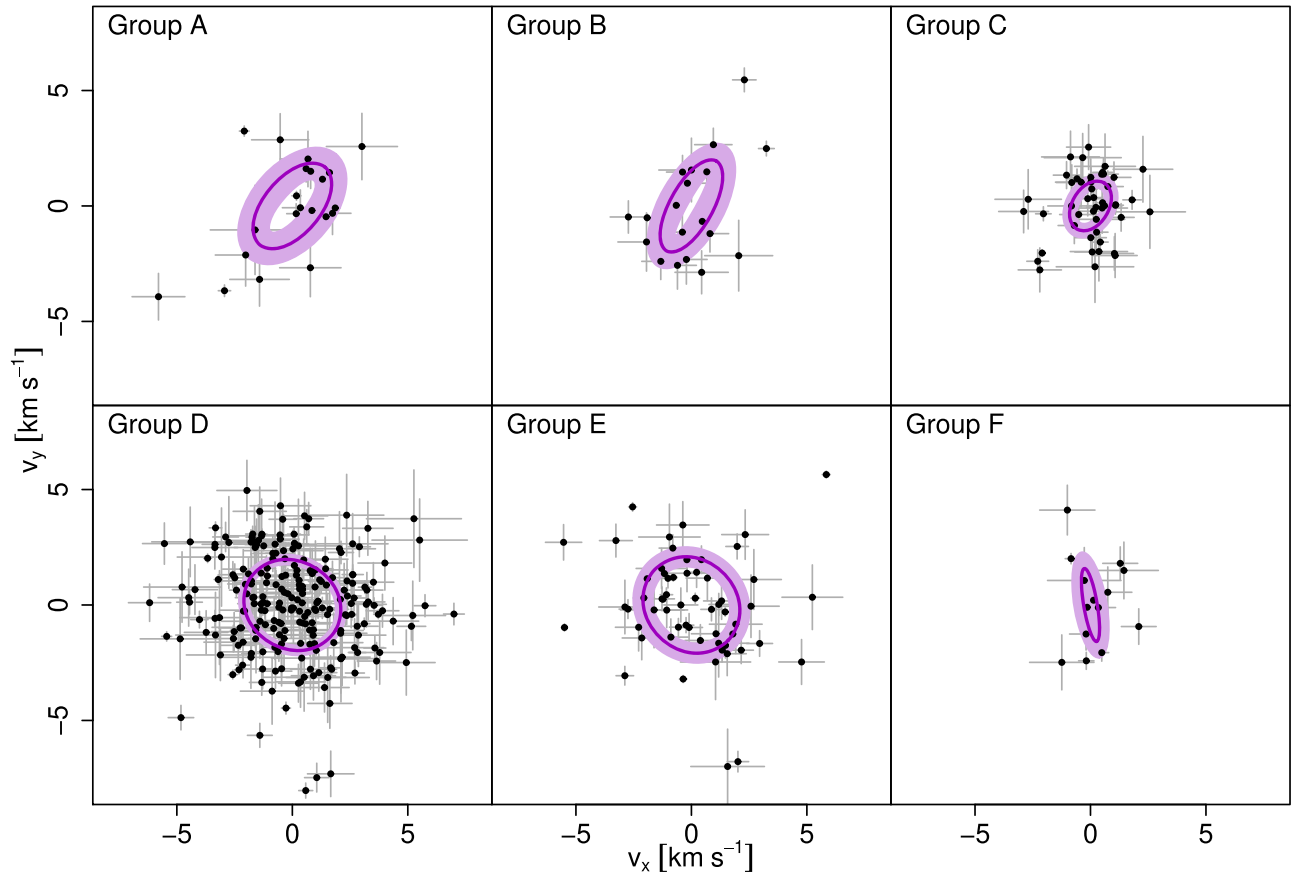


Figure 15. Scatter plots of star velocities. The magenta ellipses indicate the best-fit Gaussians; the ellipses are drawn at Mahalanobis distances of 1, and the shaded region shows the 95% confidence intervals on the ellipses.

Table 6
Internal Kinematic Properties of Stellar Groups

Group	Kendall's τ		Mean Motions		Linear Expansion Model Parameters				Total
	p_x	p_y	Mean v_{out} (km s^{-1})	Mean v_{az} (km s^{-1})	Gradient 1 ($\text{km s}^{-1} \text{pc}^{-1}$)	Gradient 2 ($\text{km s}^{-1} \text{pc}^{-1}$)	θ ($^\circ$)	σ_{scatter} (km s^{-1})	σ_{1D} (km s^{-1})
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A	>0.05	>0.05	0.7 ± 0.4	0.0 ± 0.5	1.8 ± 0.2
B	0.008	0.002	0.5 ± 0.7	-1.4 ± 1.0	1.7 ± 0.2
C	<0.001	>0.05	0.7 ± 0.6	0.6 ± 0.4	0.63 ± 0.11	0.04 ± 0.17	81 ± 12	0.8 ± 0.1	1.0 ± 0.1
D	<0.001	<0.001	1.9 ± 0.4	0.3 ± 0.2	0.50 ± 0.03	0.33 ± 0.03	98 ± 8	1.4 ± 0.1	2.0 ± 0.1
E	<0.001	>0.05	0.6 ± 0.6	-0.2 ± 0.4	0.33 ± 0.08	0.05 ± 0.07	87 ± 6	2.0 ± 0.1	2.1 ± 0.2
F	>0.05	>0.05	-0.1 ± 0.9	-0.1 ± 0.2	1.2 ± 0.3

Note. Column (1): stellar group. Columns (2)–(3): nonparametric tests for correlation between position and velocity in R.A. and decl. Columns (4)–(5): mean outward velocity and mean azimuthal velocity as calculated in Section 7.2. Columns (6)–(9): parameters from the linear model fit to the velocity from Section 7.3, including the velocity gradients along the direction of maximum gradient (Column (6)) and in the orthogonal direction (Column (7)), the position angle of the anisotropy (Column (8)), and the intrinsic velocity scatter left over after the velocity gradient is removed (Column (9)). Column (10): total velocity dispersion.

center, a trend that is seen in many of the expanding young star clusters and associations from Kuhn et al. (2019).

The other groups contain fewer stars than Group D, and the clear evidence for expansion is not found using these diagrams. For the other groups, arrow diagrams like those in the top two panels of Figure 16 are included in the figure set, but there are too few stars to make the other plots. For these groups, the weighted median v_{out} values are not significantly different from 0 (Table 6) owing to the large statistical scatter. Nevertheless, median v_{out} may not be the most sensitive test for expansion,

given that this statistic combines multiple dimensions and does not take advantage of the tendency, seen in many expanding regions, for velocity to increase with distance from the center.

We apply a further test for expansion, using the nonparametric Kendall's τ test to test for correlation between x and v_x and between y and v_y (Table 6). Statistically significant results would indicate a velocity gradient, which would imply expansion if the gradient is positive and contraction if negative. We find strong statistical evidence for expansion in both dimensions for Group D, moderate statistical evidence for

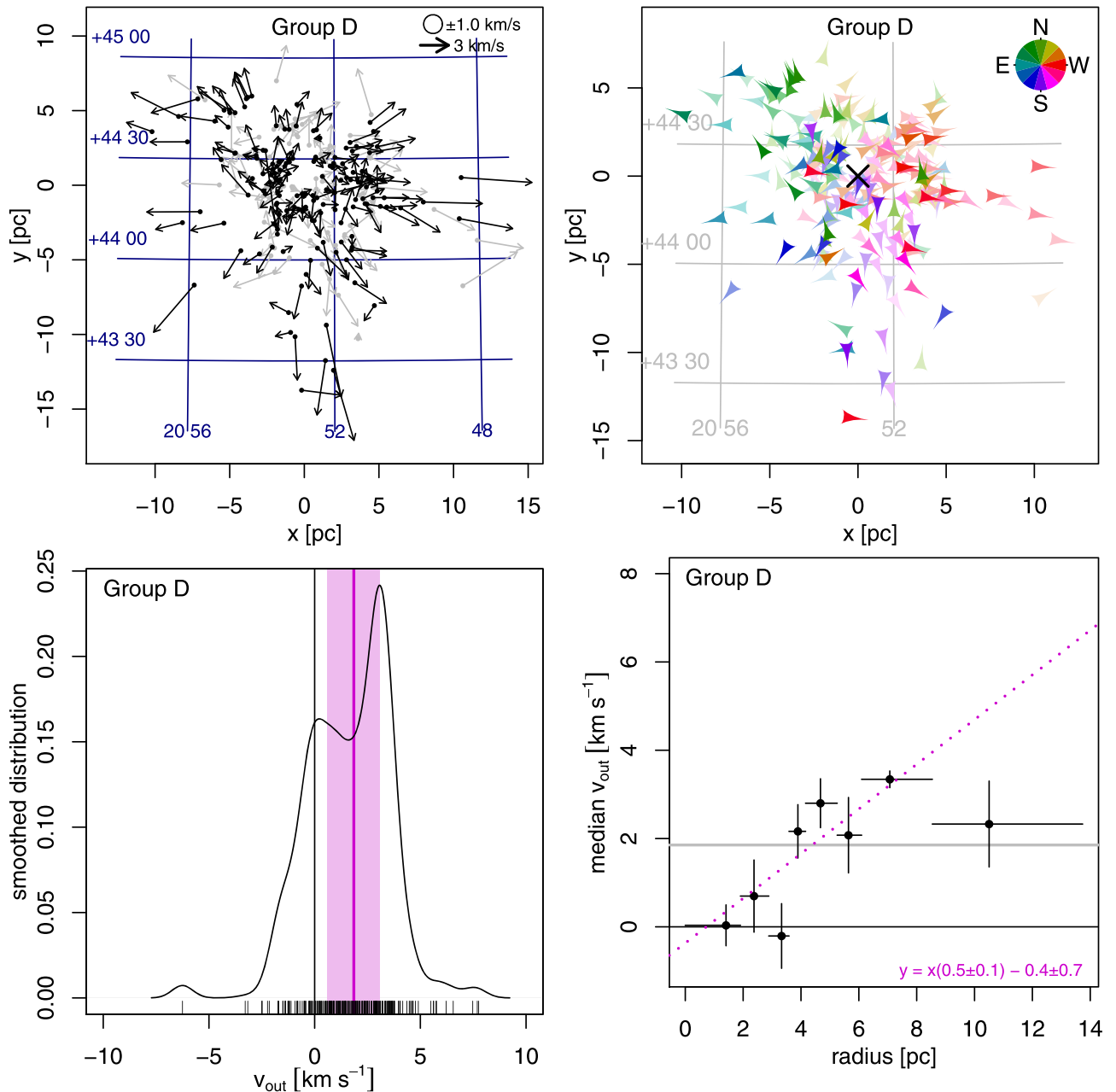


Figure 16. Plots of stellar kinematics for stars in Group D. Top left: vectors show motions of stars in the (x, y) plane relative to the center of the group. Stars with the most precise velocity estimates (< 1 km s^{-1}) are shown in black, while stars with less certain velocity estimates are shown in gray. Top right: arrow heads are color-coded by the direction of motion of a star as indicated by the color wheel. Bulk motions such as expansion, contraction, or rotation can be seen as gradients in color across the group. Bottom left: distribution of the outward velocity component v_{out} . The median value is indicated by the magenta line, and the 3σ confidence interval is shown by the shaded area. Bottom right: median value of v_{out} as a function of distance from the center of the region. An error-weighted least-squares regression fit is shown.

(The complete figure set (14 images) is available.)

expansion in both dimensions for Group B, and strong statistical evidence for expansion in only one dimension (x) for Groups C and E.

7.3. Velocity Gradients in Expanding Groups

Our results, along with the results from previous Gaia studies, suggest that stars in expanding young stellar groups often have velocities that are proportional to the distance from the groups' centers (Kuhn et al. 2019; Román-Zúñiga et al. 2019; Wright et al. 2019; Zamora-Avilés et al. 2019; Melnik &

Dambis 2020). However, Wright et al. (2019) point out that the expansion may be anisotropic. In Appendix D, we describe a linear model for velocity as a function of position, using a method that can account for anisotropy and is independent of choice of coordinate system. Table 6 reports several of the interesting properties from the models, including velocity gradients in the directions of maximum and minimum expansion, the position angle of anisotropy, and the intrinsic scatter that remains after the velocity gradient is accounted for.

We find statistically significant expansion for Groups C, D, and E, but no group had statistically significant rotation. Group

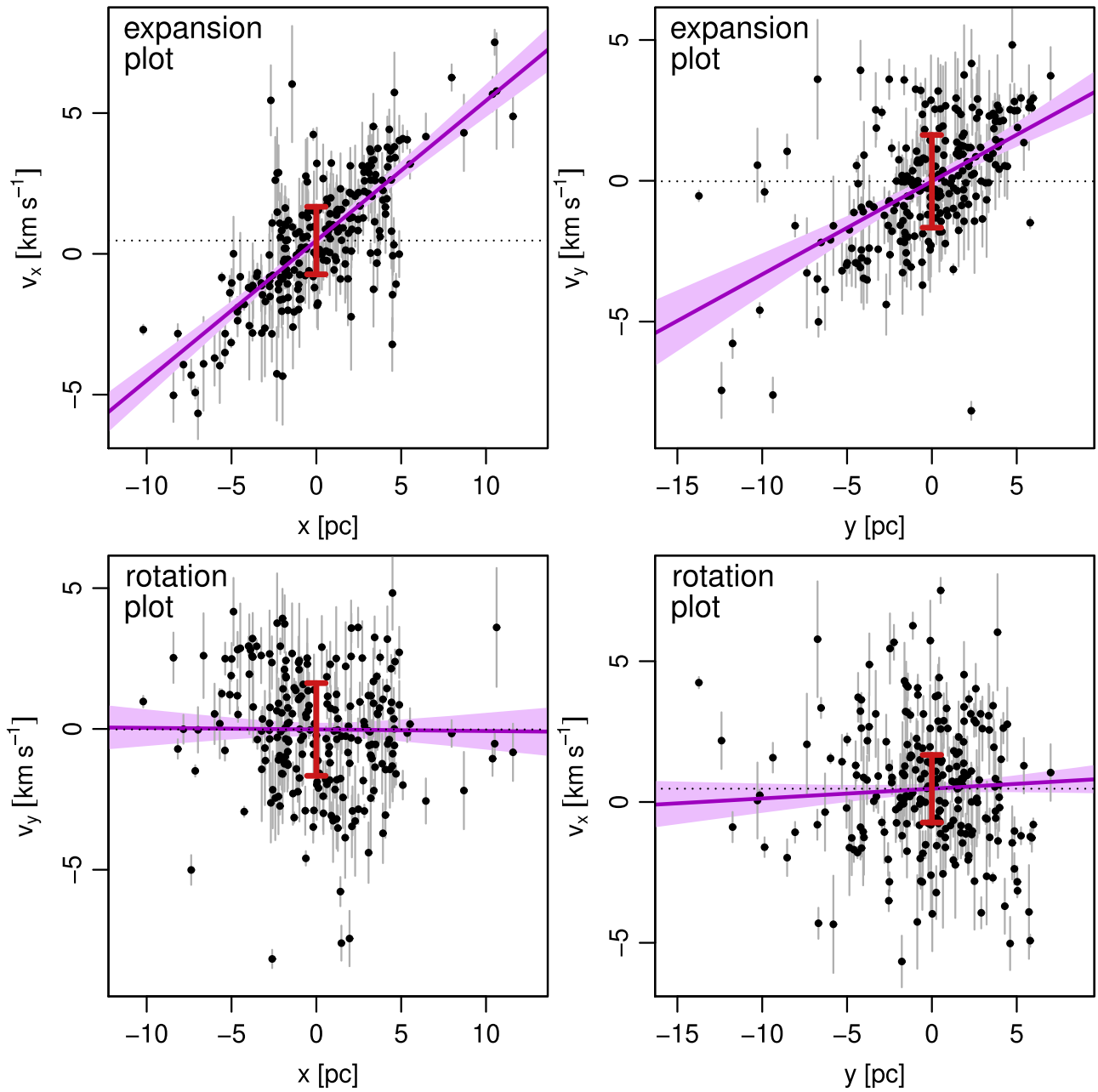


Figure 17. Plots of velocity coordinates v_x and v_y vs. position coordinates x and y for stars in Group D. The top row shows v_x vs. x and v_y vs. y , so expansion/contraction would produce correlation on these diagrams. The bottom row shows the combinations v_y vs. x and v_x vs. y , so correlation here would indicate rotation. We have overplotted graphical illustrations of our MCMC model fit for comparison to the data. On the top two plots we show lines indicating the relations $v_x = v_{x,0} + xA_{11}$ (left) and $v_y = v_{y,0} + yA_{22}$ (median and 95% credible region shown in magenta), while the bottom two plots show $v_y = v_{y,0} + xA_{21}$ (left) and $v_x = v_{x,0} + yA_{12}$ (right). The contribution from the intrinsic velocity dispersion, described by the covariance matrix Σ_{scatter} , is indicated by the red bars.

(The complete figure set (3 images) is available.)

D has the strongest relation between position and proper motion, with expansion in all directions, but with a larger gradient in the east–west direction ($0.5 \text{ km s}^{-1} \text{ pc}^{-1}$) than in the north–south direction ($0.3 \text{ km s}^{-1} \text{ pc}^{-1}$). In contrast, Groups C and E had nonzero expansion gradients only in the east–west direction, not the north–south direction. We note that the anisotropy position angles of all three groups are similar. The regression analysis and the calculation of the model parameters are described in greater detail in the appendix.

Figure 17 shows scatter plots of position versus velocity with the model regression lines overplotted. The graphs of v_x versus

x (top left) and v_y versus y (top right) exhibit positive correlations if there is expansion, while v_y versus x (bottom left) and v_x versus y (bottom right) exhibit correlations if there is rotation. The regression line and 95% credible interval from the linear model are shown in magenta. The intrinsic velocity scatter, applied equally to all points, is indicated by the red error bar (1σ), while the individual 1σ measurement uncertainties are indicated by the gray lines. For Group D, both the v_x versus x and v_y versus y plots show strong correlations, while for Groups C and E correlations are only strong on the v_x versus x plots owing to their $\theta \approx 90^\circ$ anisotropy position angles.

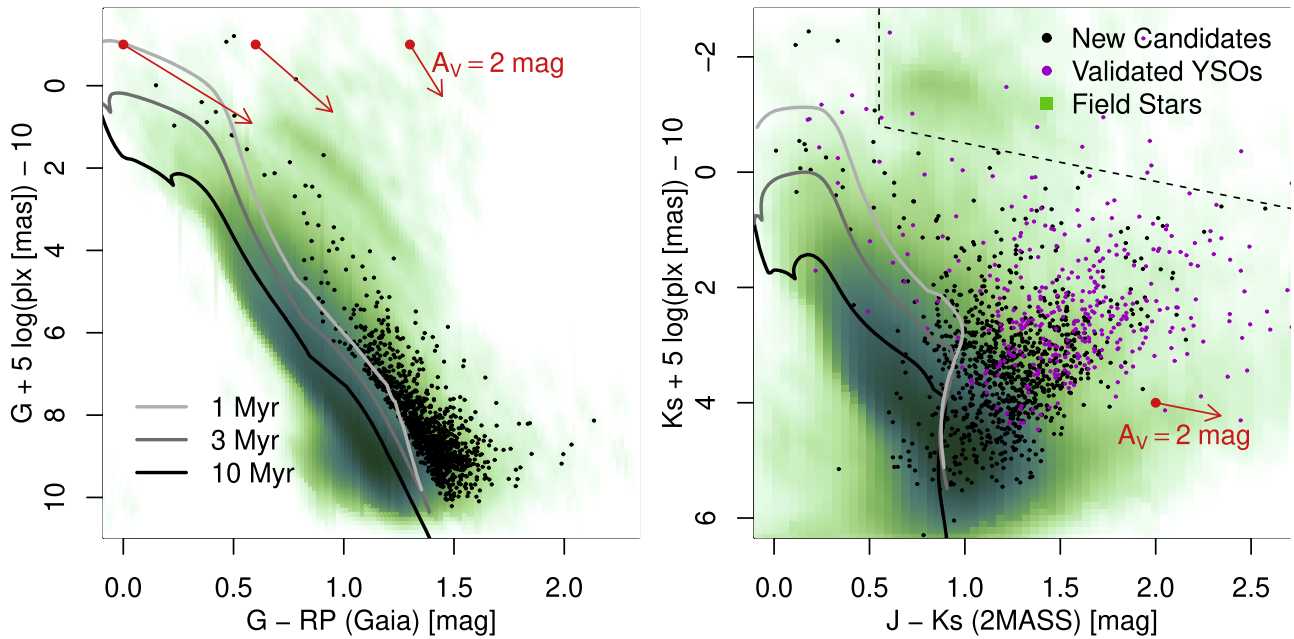


Figure 18. CMDs in the Gaia bands (left) and near-infrared (right) showing how the distributions of new candidate NAP members (black points) compare to other categories of sources. The green shaded regions show the density of Gaia sources with the same parallax range, with darker colors indicating higher number densities of sources. On both plots, field stars in the red clump and the giant branch can be seen near the top of the diagram. We use cuts on the near-infrared diagram (indicated by dotted lines) to remove sources from our candidate sample that have a high probability of being giant stars. On the near-infrared diagram, we also show YSOs from our Gaia-validated literature sample (magenta points). The distribution of these points is shifted to slightly higher K_s -band luminosities and redder $J - K_s$ colors. This may be the result of earlier studies selecting stars with disks, and thus more likely to have K_s -band excess emission, while the Gaia selection is independent of whether the object has a disk.

The O stars are both among the fastest-moving members of Group D. The Bajamar Star has a velocity of $6.6 \pm 0.5 \text{ km s}^{-1}$ relative to the center of the group, while HD 199579 has a velocity of $6.4 \pm 0.4 \text{ km s}^{-1}$. However, neither of these velocities is discrepant from the linear model. This suggests that whatever phenomenon accelerated these stars to their current velocities is also responsible for the expansion of the group as a whole. If we make the simplifying assumption that the velocities have been approximately constant, the Bajamar Star would have been nearest, in projection, to the center of Group D $\sim 1.5 \text{ Myr}$ ago, while HD 199579 would have been nearest $\sim 1.8 \text{ Myr}$ ago.

7.4. Effects of Spatial Covariances in Gaia Astrometric Errors

To ensure that our measurements of velocity gradients are robust, we examine how these would be affected by the systematic correlated errors in astrometry (Lindegren et al. 2018, their Section 14). They found that the covariance of quasar proper motion as a function of angular separation decreases from $\sim 4000 \mu\text{as}^2 \text{ yr}^{-2}$ at $0^\circ 07$ separation to a local minimum of $\sim 40 \mu\text{as}^2 \text{ yr}^{-2}$ at $0^\circ 43$ separation. This could induce an artificial proper-motion gradient of $\sim 0.17 \text{ mas yr}^{-1} \text{ deg}^{-1}$, which would look like a velocity gradient of $0.045 \text{ km s}^{-1} \text{ pc}^{-1}$ in a stellar association at a distance of 795 pc. Given that the velocity gradients that we detect are significantly larger than this value, they are most likely astrophysical.

8. New Member Candidates from Gaia

In addition to validating previously proposed YSOs in the NAP region, Gaia astrometry can be used to suggest new member candidates. In most earlier studies, candidates were identified based on criteria that required a star to possess a circumstellar disk (e.g.,

strong $H\alpha$ emission, infrared excess, large amplitude variability). However, many star-forming regions are dominated by pre-main-sequence stars for which disks are not detected (Broos et al. 2013). Thus, astrometric selection of candidates can reveal whether any populations of NAP members have remained undetected in previous analysis, which, in addition to their possible biases, are also incomplete in their selection of candidate members.

We start with a sample of all Gaia DR2 sources within 3° of the Bajamar Star that pass the same quality criteria from Section 2.2 and lie within the same parallax range and the (μ_ℓ, μ_b) region identified in Section 3. However, we apply further parallax cuts to select only objects that are consistent within 2 standard deviations of the median parallax of the groups, i.e., $\varpi + 2\sigma_\varpi \geq 1.06 \text{ mas}$ and $\varpi - 2\sigma_\varpi \leq 1.24 \text{ mas}$. These criteria greatly reduce the number of Gaia sources in the region, from $\sim 75,000$ Gaia sources in the original parallax range down to $\sim 10,000$ sources.

We further refine the sample to ensure that the selected objects are likely to be young, using an age limit of 3 Myr that matches the approximate maximum age found for members in Section 3. We identify Gaia sources whose absolute G -band magnitude and $G - RP$ color are 2 standard deviations above the 3 Myr Bressan et al. (2012) isochrone with $A_V = 1 \text{ mag}$ (Figure 18, left). However, remaining contaminants can include main-sequence stars with high extinction, as well as post-main-sequence stars. To reduce contamination from reddened main-sequence stars, we use the 2MASS $J - H$ versus $H - K_s$ diagram to estimate extinction using the reddening laws from Rieke & Lebofsky (1985). Typical reddening is $\Delta(J - H) \sim 0.3 \text{ mag}$ ($\approx 2.8 \text{ mag}$ in the V band) but ranges from 0 to 0.5 mag, with a tail out to 1.5 mag. Dereddening on the near-infrared color-color diagram can lead to degeneracies between low-mass and intermediate/high-mass stars. In most cases we pick the low-mass solution. However, when a star could

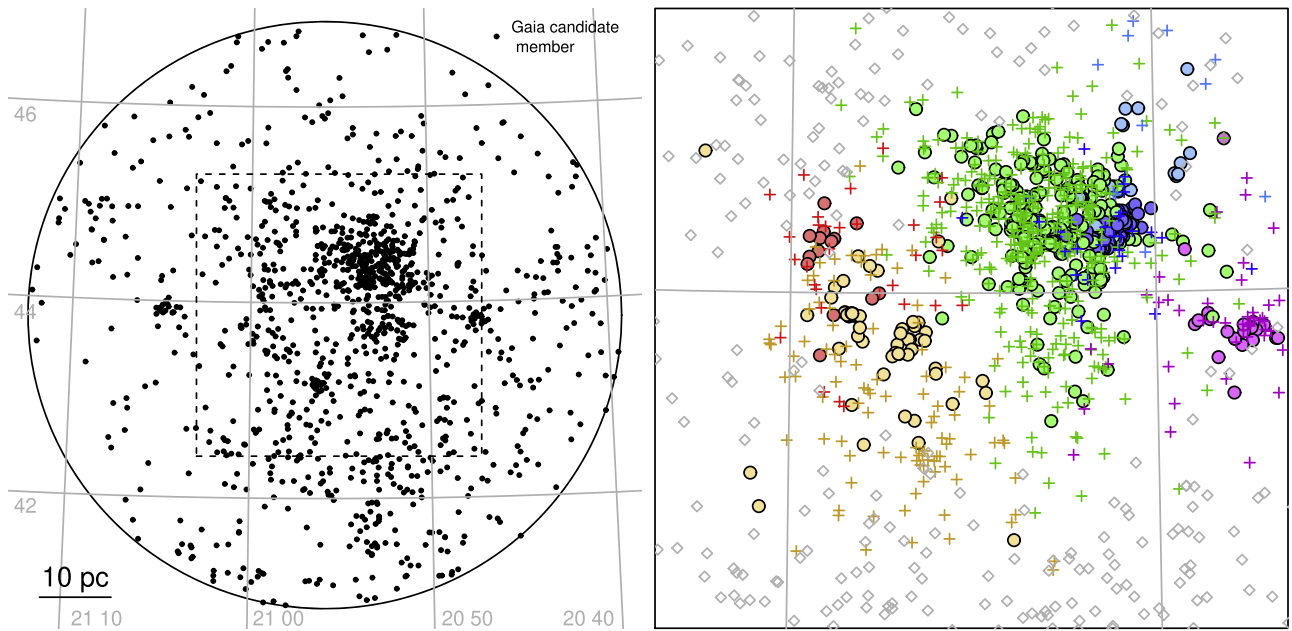


Figure 19. Spatial distributions of new candidate members. Left: all new candidates within 3° of the Bajamar Star. The dashed box outlines the blown-up central region shown in the other panel. Right: new candidates are shown as plus signs if they are assigned to existing groups or diamonds otherwise, while the validated YSOs from previous studies are shown as circles. Color-coding of group members is the same as in previous figures. This plot shows that most new candidates in the central region are distributed similarly to the distribution of the previously identified stars.

be low mass with $\Delta(J - H) < 0.3$ mag or intermediate mass with $\Delta(J - H) \geq 0.3$ mag, we pick the higher extinction to avoid classifying reddened intermediate-mass background field stars as pre-main-sequence members. Once we have estimated extinction, we require the star to still lie above the 3 Myr Gaia isochrone with the corresponding reddening. To reduce contamination by giants, we selectively remove objects from the region of 2MASS CMD where these objects are likely to dominate the sample: $J - K_s > 0.5$ mag and $M_{K_s} < 0.66 \times (J - K_s) - 0.8$ mag (Figure 18, right). This removes objects in the red clump and the upper giant branch. Finally, in addition to pre-main-sequence stars, we also include five additional stars of spectral type B as compiled by Skiff (2009). The final sample contains 1187 new candidates (Table 7).

The spatial distribution of the new candidates is shown in Figure 19. Candidates are spatially clustered in the NAP region as expected; however, several small clumps were not previously identified, as described below. There is also a distributed population of objects at angular separations of several degrees from the cluster (left panel). In position versus proper-motion space (not shown) the new candidates follow similar structures to those seen in Figures 5 and 6, but with more points scattered between the groups. Many of the distributed objects may be residual contaminants, which would be expected to be distributed approximately uniformly across the field. However, more data (e.g., spectroscopic follow-up observations) are necessary to determine the rate of contamination.

To assign candidates to groups A–F, as well as a “distributed” category, we used the k -nearest neighbor (k NN) algorithm implemented in the R package `class` (Venables & Ripley 2002). For the training examples, we used the 395 previously assigned members of groups A–F in addition to an equal number of randomly selected “distributed” stars from Gaia in the same parallax and proper-motion range that were not selected as new candidates. The k NN classifier is run using the normalized ℓ , b , μ_{ℓ^*} , and μ_b variables, and classifications are based on the five

Table 7
New Member Candidates from Gaia

Gaia DR2	α (ICRS)	δ (ICRS)	Group
2067063447700277504	20 50 06.05	+44 17 48.9	C
2166282204462982528	20 50 07.33	+45 49 22.4	distrib
2067060007428506112	20 50 07.69	+44 08 30.1	D
2163214738816798336	20 50 10.93	+44 51 22.0	D
2065821858551580288	20 50 12.03	+41 38 44.1	distrib
2163266832483920128	20 50 12.94	+45 32 43.1	distrib
2066624635182982144	20 50 14.35	+42 15 43.0	distrib
2067049428926311040	20 50 15.04	+43 58 11.8	D
2166282964674128256	20 50 15.47	+45 54 02.3	distrib
2067049532005529600	20 50 15.81	+43 58 59.9	A
2067060625914494080	20 50 16.82	+44 11 53.5	B

Note. Gaia DR2 sources selected as new member candidates. Columns are the same as in Table 2, with the addition of labels for Group G and distributed stars to the final column.

(This table is available in its entirety in machine-readable form.)

nearest neighbors. We assign 52% of the candidates to groups and 48% to the “distributed” category.

A previously unidentified group of ~ 17 new members is centered at 21:04:44+43:55:30. We call these stars Group G and list them in Table 7. This group is located next to the third-magnitude red supergiant star ξ Cyg. However, this star has a parallax of ~ 3.9 mas, meaning that it is much closer and therefore unrelated to the group. Nevertheless, the angular proximity to such a bright star may explain why the group was not previously identified.

Another small clump of previously unidentified stars lies just to the south of Group E. These stars continue the position–proper-motion trends seen for the stars in Group E, so we classify the new stars as a continuation of this group.

9. Discussion

Observations of the NAP complex reveal that it contains multiple groups of very young stars (~ 1 Myr) that are associated with distinct components of the molecular gas. This clumpy distribution of mass covers a region at least 30 pc in diameter, and Gaia parallaxes suggest that there are several conglomerations along the line of sight spanning up to 150–200 pc. The motions of the groups, as inferred from Gaia astrometry, appear to be approximately random relative to one another, and several of the groups are rapidly expanding.

9.1. Expanding Stellar Groups

Several mechanisms, including disruption by tidal forces, cloud dispersal, and sequential star formation, may play roles in the expansion of stellar groups in the NAP complex.

A clumpy mass distribution, as observed here, will generate internal tidal forces that affect stellar dynamics and influence the potential assembly of either bound clusters or unbound associations. Tidal fields would naturally produce an anisotropic velocity gradient similar to those observed in the NAP region. Furthermore, tidal forces would preserve clumpy substructure within a group (e.g., D and E) as it is stretched, while processes driven by dynamical relaxation would be more likely to erase these substructures. Tidal forces from giant molecular clouds have long been known to be a major cause of star cluster disruption (Spitzer 1958; Gieles et al. 2006), and these effects can occur in the star-forming environment itself, inhibiting the initial formation of a bound system (Kruijssen et al. 2011, 2012). Zamora-Avilés et al. (2019) have run numerical simulations in which gas clouds, driven away from the newly formed young stellar cluster by feedback, exert tidal forces on the cluster that pull it apart.

The tidal acceleration $a_{t,\text{axial}}$ per unit of displacement Δr along the separation axis due to a mass M at distance r is

$$a_{t,\text{axial}}/\Delta r = 2G\frac{M}{r^3}. \quad (12)$$

For Group D, the magnitude of the tidal acceleration from Cloud 2 ($2.1 \times 10^3 M_\odot$) located ~ 5 pc to the west of the center of the group (assuming that the displacement in the third dimension is small) is $0.15 \text{ km s}^{-1} \text{ pc}^{-1} \text{ Myr}^{-1}$. This acceleration would be enough to induce the observed $\sim 0.5 \text{ km s}^{-1} \text{ pc}^{-1}$ east–west velocity gradient in ~ 3 Myr. Other nearby clouds, notably Clouds 3, 4, and 8 (a potential remnant of Group D’s natal cloud), could also contribute to the tidal forces on this group. These clouds would also exert tidal forces on the other nearby, expanding group, C, located at the edge of Cloud 2. However, precisely calculating the effect of tidal forces is challenging owing to uncertainties in the third dimension, velocity differences between the groups and clouds, and systematic uncertainties in cloud mass.

Expansion of Group E is more difficult to explain with tides, since its separation from other parts of the NAP region by ~ 35 pc along the line of sight means that it would be much less strongly affected by tidal forces from other clouds. Thus, any tidal force would need to be generated by clumpy distributions of mass within Cloud 9 itself.

The ongoing dispersal of the molecular clouds due to O star winds and photoevaporation, combined with outflows from low-mass stars (Bally & Reipurth 2003; Bally et al. 2014), will

weaken the binding energy of the stellar groups, causing them to expand and/or disperse (e.g., Tutukov 1978; Adams 2000; Kroupa et al. 2001, and many others). In the northwest, the shell-like structure may be the outer rim of a bubble blown by the Bajamar Star, which would have been closer to this part of the cloud ~ 1.5 Myr ago. Cloud dispersal in this region may have affected the stellar groups that lie near the shell, particularly Groups C and D. In contrast, photoevaporation does not appear to be as significant in the Gulf of Mexico region that contains Group E. Nevertheless, Bally et al. (2014) found dozens of outflows here that are reprocessing a significant volume of the Gulf of Mexico cloud through supersonic shocks. Such outflows may dominate the injection of energy and momentum in the absence of massive stars (Li & Nakamura 2006; Nakamura & Li 2011) and may contribute to the cloud’s mass loss (Arce et al. 2007).

We examine the scenario in which Group D formed as an embedded cluster in virial equilibrium in a molecular cloud that was subsequently dispersed. Given Group D’s measured velocity dispersion of $\sigma_{1D} = 2 \text{ km s}^{-1}$ and assuming an initial size of ~ 1 pc, the corresponding virial mass of the system would have been $\sim 9 \times 10^3 M_\odot$. Thus, the natal cloud mass would have been ~ 6 times greater than the estimated mass of Cloud 8 and ~ 1.5 times the sum of the masses of all the clouds in the northwestern part of the NAP complex. Wright & Parker (2019) have demonstrated that expansion in this scenario could be anisotropic if it were preceded by collapse and bounce in an asymmetric gravitational potential, setting up an anisotropic velocity dispersion.

A third explanation for expanding groups is that they are produced by a second generation of star formation in an expanding shell of material. In regions where star formation has been attributed to material swept up in shells around expanding H II regions (e.g., Patel et al. 1998) the stars formed in this way would presumably inherit the velocities of the material in the expanding shells. In the NAP region, there is ongoing star formation in several of the clouds making up the shell structure, so it is plausible that such a scenario could be happening here too. In particular, Group C is associated with the bright rim cloud at the edge of Cloud 2. For several bright rim clouds in other star-forming regions, Getman et al. (2007, 2012) have shown that the stars in front of the cloud exhibit an age gradient, with the youngest stars nearest the cloud. In such a scenario, as a cloud is driven outward by an expanding shell, the sequential formation of stars could yield a velocity gradient.

Given the conditions in the NAP region, it seems likely that multiple scenarios operate simultaneously to induce the observed expansion patterns. For example, the lessening of the binding energy due to cloud dispersal would make groups more susceptible to disruption by interactions with neighboring clouds. Although theoretical work has shown that binary-star dynamics can preferentially eject O stars at high velocities ($> 30 \text{ km s}^{-1}$; Oh & Kroupa 2016), this mechanism does not appear to be required here because the O stars and low-mass stars follow the same velocity pattern.

The Gaia data show that individual groups are expanding, but the relative motions of the groups exhibit no clear pattern of either convergence or divergence (Figure 14, top left). This is similar to the situation found by Kuhn et al. (2019) in other large star-forming complexes like NGC 2264, NGC 6357, or the Carina Nebula and may help explain why expansion has been difficult to detect in young stellar associations analyzed as

a whole (e.g., Wright et al. 2016; Ward & Kruijssen 2018; Wright & Mamajek 2018). The velocity dispersion of all stars in the NAP complex is $\sigma_{1D} = 2.5 \text{ km s}^{-1}$ (Figure 14, bottom panels). This, combined with the separations of several to tens of parsecs between groups, means that the mass needed to gravitationally bind the system ($>10^5 M_{\odot}$) is orders of magnitude greater than the mass in stars. Thus, it is impossible for the system to coalesce as a bound cluster. Nevertheless, it may be possible in some groups for bound remnants to remain behind even after most stars have escaped (Baumgardt & Kroupa 2007).

9.2. Star Formation Scenarios

We can use our inferences about the dynamical state of the stars and gas, as well as the constraints on stellar ages, to test various theoretical scenarios for star formation as applied to the NAP region. These scenarios broadly break down into fast star formation, on the timescale of the freefall collapse of the molecular clouds (e.g., Elmegreen 2000; Hartmann et al. 2012), and slow star formation that persists over multiple freefall timescales (e.g., Tan et al. 2006; Krumholz & McKee 2020). In the former case, molecular clouds would have little pressure support (Vázquez-Semadeni et al. 2019), while in the latter case turbulent pressure support stabilizes the clouds against collapse and allows for an extended duration of star formation (Padoan & Nordlund 2011).

In the NAP region, we demonstrated that the velocity dispersion within individual clouds is consistent with the velocity dispersion that would be expected from either freefall collapse or virial equilibrium given the systematic uncertainties on the cloud masses (Section 6.3). When we consider the whole region (Figure 14), the combined velocity dispersion of all gas is larger than that of the individual clouds, with $\sigma_{\text{all}} = 4.2 \text{ km s}^{-1}$. In the complex, half the gas is contained within a projected radius of 9.4 pc, so the dynamical mass for the whole system calculated using Equation (9) would be $1.9 \times 10^5 M_{\odot}$, or about two and a half times the $7.2 \times 10^4 M_{\odot}$ mass estimated from ^{13}CO column density. Estimation depends on the virial parameter α , which we take to be 1 (virial equilibrium); however, a cloud with $\alpha = 2$ (freefall) would yield a dynamical mass estimate half as large, making the differences smaller. As earlier, the systematic uncertainties make distinguishing between these scenarios difficult. Nevertheless, in either case, the velocity differences between the clouds, as well as the velocity dispersions within the clouds, can be accounted for by gravitation.

The freefall timescales for the clouds are $\sim 1 \text{ Myr}$ (Table 5). This is similar to the $\sim 1 \text{ Myr}$ ages of the stars (Figure 2). This implies that, in each cloud, most star formation occurred within the last $1-2 \times \tau_{\text{ff}}$. There are two considerations that should be taken into account when interpreting freefall timescales. First, freefall timescales are typically calculated using the mean density of a cloud as we have done; a nonuniform density could change the mean τ_{ff} , but the correction factor is expected to be of order unity (Krumholz & McKee 2020). Second, we are calculating the present-day freefall timescales; however, in the past, when the stars observed today were forming, the freefall timescale would likely have been longer (Vázquez-Semadeni et al. 2019). Star cluster formation in a freefall time is expected for either clouds without turbulent support (Klessen et al. 1998) or clouds with decaying turbulence (Klessen 2000). Thus, the

observation that most of the stars are only one to several τ_{ff} old is consistent with the rapid star formation scenarios.

The spatial clustering of star formation also favors a scenario of rapid star formation in a turbulent molecular cloud. Overall, the NAP cloud complex has features expected for a giant molecular cloud sculpted by turbulence, including a clumpy, hierarchical density structure and velocity structure that yields a larger velocity dispersion for the whole system than within the individual clouds, broadly consistent with the Larson (1981) relation. Given that the major stellar groups follow the present-day cloud structure, this supports the view that the stars froze out on a timescale less than the crossing time for the complex (Elmegreen 2000).

The similarity in age ($\sim 1 \text{ Myr}$) of widely separated stellar groups raises the question of how star formation in the different clouds of the NAP complex became synchronized. Although many of the surveys preferentially selected stars that are young enough to still possess disks, even the X-ray-selected sample in the NAP region appears to follow the same age distribution (Appendix B). The sound crossing time for the complex is $t_s = 9.4 \text{ pc}/c_s = 40 \text{ Myr}$, meaning that the clouds would not be in thermal contact with each other on the collapse timescale, and we would not expect them to have similarly aged populations. A similar problem has been posed by Preibisch & Zinnecker (1999) for the Upper Scorpius OB association and by Herczeg et al. (2019) for the Serpens molecular clouds. A possible solution to this is an accelerating star formation rate, which is a feature of several models, including the conveyor belt model (Longmore et al. 2014) and global hierarchical collapse (Vázquez-Semadeni et al. 2017, 2019). With sufficiently high acceleration, most of the stars in star-forming clouds would have formed very recently, even if the first stars to form in those clouds formed at different times.

Although the NAP clouds appear to be forming stars on freefall timescales, fast star formation scenarios face challenges if applied to the entire Galaxy because they would imply a Galactic star formation rate much higher than observed (Zuckerman & Evans 1974; Krumholz et al. 2006). To overcome this, theoretical fast star formation scenarios invoke the quick disruption of the molecular clouds by stellar feedback to keep star formation inefficient and regulate the Galactic star formation rate (Chevance et al. 2020). In the NAP region, where the disruptive influences of the Bajamar Star and outflows from low-mass stars are already apparent, it appears likely that feedback processes could bring star formation to an end before a substantial fraction of the cloud mass is converted into stars.

10. Summary

In this study we have examined the kinematics of stars and gas in the NAP region using previously published candidate YSOs, Gaia astrometry, and a newly presented molecular gas map. Our main results are the following:

1. On the basis of Gaia parallax and proper motions, we identify a sample of 395 stars as high-probability members of the complex (estimated 3% residual contamination). We also reclassify hundreds of previously cataloged candidates as contaminants. In addition, ~ 2000 candidates remain ambiguous on the basis of their kinematics alone, though the majority show both

spatial clustering with the identified kinematic groups (Figure 3) and traditional signatures of the activity associated with stellar youth.

2. The locations of the confirmed members on the Gaia CMD suggest that nearly all stars are <3 Myr old and that most of them are ~ 1 Myr old or younger.
3. Most of the widely dispersed YSO candidates from previous studies are identified as contaminants, while the confirmed members tend to be more tightly clustered. We identify six groups using unsupervised cluster analysis in position and proper motion.
4. The NAP region is ~ 795 pc from the Sun. However, parallax distributions show slight differences in distances to the individual kinematic groups. Notably, the stars in the “Gulf of Mexico” region appear to be ~ 35 pc closer than the rest of the system, and a small group, containing the famous V1057 Cyg star, is ~ 130 pc farther away.
5. The locations of each of the stellar groups are spatially correlated with the main components of the molecular cloud. This implies that all parts of the NAP cloud complex, if sufficiently massive, are actively forming stars.
6. Most of the clouds have complex, multimodal velocity structures. We use the ^{13}CO map to estimate various cloud properties, including masses and velocity dispersions. We find high correlation between the masses estimated from integrated ^{13}CO column densities to be in remarkable agreement with dynamical mass estimates, suggesting a strong connection between gravitation and velocity. Mean freefall times for individual clouds are $\lesssim 1.5$ Myr.
7. Relative velocities of the different groups appear randomly oriented, showing no sign of either global expansion or contraction. The radial velocity of the “Gulf of Mexico” region implies that it is plunging inward. On the other hand, in the northwest of the NAP complex (“Atlantic” and “Pelican” regions), where the morphology of the CO gas suggests an expanding H II region, the relative motions of the stellar groups seem unaffected.
8. In contrast to the lack of global expansion, several stellar groups are individually rapidly expanding, with velocity gradients of $0.3\text{--}0.5$ km s $^{-1}$ pc $^{-1}$. The expansion gradients are anisotropic, and we argue that these could be, in part, attributed to tidal forces from within the clumpy molecular cloud complex.
9. The primary ionizing source in the region, the early O Bajamar Star, lies between two groups. We suggest that it likely originated as part of a group (which we call Group D) centered in the “Atlantic” part of the NAP region because its trajectory would trace back to this group. The star’s velocity (~ 6 km s $^{-1}$) is consistent with the expansion velocity seen for low-mass stars of the same group. Another O-type star, HD 199579, appears to be ejected in a different direction from the same group.
10. We identify >1000 new candidate members of the NAP region from the Gaia catalog, including a new seventh stellar group (Group G) located east of the complex. Slightly over half these new candidates are associated with Groups A–G, while the others are more broadly distributed throughout the 6° -diameter selection area of investigation. These objects would require follow-up observations to validate them.

Several lines of evidence suggest a scenario of rapid star formation in a freefall time (e.g., Elmegreen 2000; Hartmann et al. 2012; Vázquez-Semadeni et al. 2019) in the NAP complex. The structure of the stellar groups is spatially correlated with the structure of molecular clouds, stellar ages are similar to the freefall timescales of the star-forming clouds, and the gas velocity dispersions in the clouds are consistent with the velocities expected from gravitational collapse. Furthermore, an accelerating star-forming rate, as predicted by some of the models, would mean that, regardless of how long stars have been forming, the majority of stars would have been formed during the last few freefall timescales. This can explain why several distinct clouds in the NAP region that are sufficiently far apart to be out of direct thermal contact could all have very young stellar populations of nearly the same age.

Nevertheless, some important properties of the star-forming region remain relatively poorly determined. For example, the census of stars in the region is still highly incomplete even after our study, making it difficult to constrain the total stellar mass. While our sample of 395 Gaia-validated YSOs is useful for understanding the spatial and kinematic distributions of stars in this region, it is only representative of the optically brightest tail of the full stellar population.

We thank Min Fang for expert advice about the NAP region, Kevin Fogarty and Michael Grudić for helpful discussions, and the anonymous referee for useful comments. Development of YSO Corral was supported by NASA Award NNX17AF41G. A.R.A.M. was supported by Caltech’s Freshman Summer Research Institute (FSRI).

Facilities: FCRAO, Gaia, Keck:I (HIRES).

Software: astropy (Astropy Collaboration et al. 2013, 2018), BUGS (Lunn et al. 2009), class (Venables & Ripley 2002), JAGS (Plummer 2017), mclust (Scrucca et al. 2016), numpy (van der Walt et al. 2011), pandas (McKinney 2010; Pandas Development Team 2020), postgres (PostgreSQL development team 2020), R (R Core Team 2018), rjags (Plummer 2019), SAOImage DS9 (Joye & Mandel 2003), scipy (Virtanen et al. 2020), scikit-learn (Pedregosa et al. 2012), TOPCAT (Taylor 2005).

Appendix A

Parameters of the NAP Membership Classifier

Before applying the final step of the NAP membership classifier (i.e., classification using proper motions, described by the equations below), we first remove sources that either do not meet our Gaia quality criteria or have parallaxes outside the range $0.86\text{--}1.61$ mas (Section 3). For YSO candidates that pass these steps, membership probability, $p_{\text{mem}}(\mu_a, \mu_b)$, is calculated using the Gaussian mixture model (Figure 1, right panel) with the following parameters:

$$d_{\text{mem}}(\mu_i) = 0.64 \phi(\mu_i; \mu_1, \Sigma_1) \quad (\text{A1})$$

$$d_{\text{field}}(\mu_i) = 0.36 \phi(\mu_i; \mu_2, \Sigma_2) \quad (\text{A2})$$

$$p_{\text{mem}}(\mu_i) = d_{\text{mem}}(\mu_i) / [d_{\text{mem}}(\mu_i) + d_{\text{field}}(\mu_i)] \quad (\text{A3})$$

$$\mu_1 = (-3.35, -1.15) \quad (\text{A4})$$

$$\Sigma_1 = \begin{bmatrix} 0.48 & 0.06 \\ 0.06 & 0.52 \end{bmatrix} \quad (\text{A5})$$

$$\mu_2 = (-3.52, -1.79) \quad (\text{A6})$$

$$\Sigma_2 = \begin{bmatrix} 78.5 & -0.5 \\ -0.5 & 19.1 \end{bmatrix}, \quad (\text{A7})$$

where d represents the density of stars in proper-motion parameter space, ϕ denotes the bivariate normal distribution, $\mu = (\mu_{\ell^*}, \mu_b)$ are proper motions in Galactic coordinates in units of mas yr^{-1} , and Σ_i are the covariance matrices of the Gaussian components.

Appendix B Selection Effects Related to Stellar Ages

Selection of YSOs using features that are connected to disks and accretion means that pre-main-sequence stars that have lost their disks will be missed. This imposes the disk survival function, often modeled as an exponentially decreasing function with an e -folding timescale of 2–4 Myr (Mamajek 2009; Ribas et al. 2015; Richert et al. 2018), as a bias on the observed age distribution. In contrast, X-rays can identify pre-main-sequence

stars both with and without disks (Feigelson & Montmerle 1999; Feigelson 2018). Low-mass pre-main-sequence stars maintain high X-ray emission even after 10 Myr (e.g., Argiroffi et al. 2016; Preibisch et al. 2017), so X-ray selection should detect such a population if it exists.

The study by Damiani et al. (2017) provides an X-ray sample for the NAP region. In Section 3, we used the positions of the NAP members on the Gaia CMD as evidence that the NAP members are <3 Myr old. In Figure B1 we again show the same sources on CMDs, but this time with the X-ray sources marked. On the G versus $G - RP$ diagram, the X-ray sources lie within the distribution of stars selected by other methods. This suggests that both samples have similar age distributions. On the J versus $J - H$ diagram, relatively few X-ray sources have high $J - H$ values, indicating that the X-ray-selected sample is not as highly reddened as the disk-selected sample. Nevertheless, nearly all X-ray sources are above or to the right of the 1 Myr isochrone. Thus, we conclude

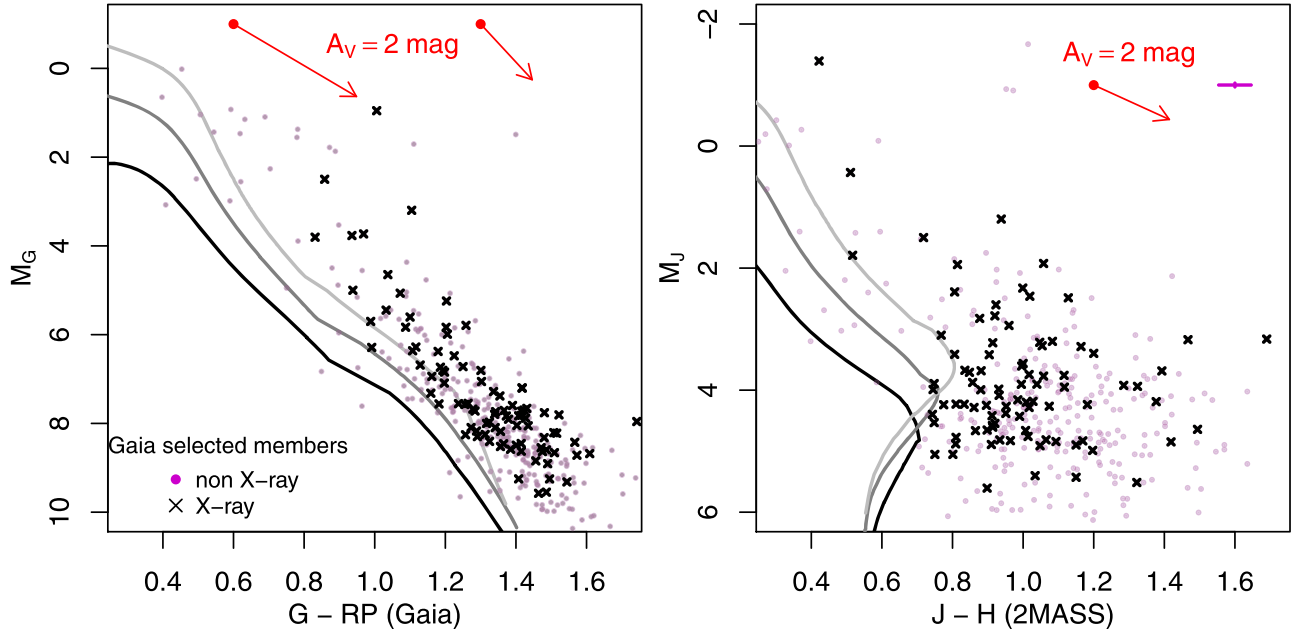


Figure B1. Optical (left) and near-infrared (right) CMDs of the astrometrically validated members, comparing stars selected using X-ray emission (black crosses) to stars selected by other criteria (magenta points). The X-ray candidates from Damiani et al. (2017) are some of the only stars in our literature-based sample selected with a method that is independent of whether a star has a disk. Nevertheless, the X-ray sample does not appear older than the other NAP members. We show the same isochrones and reddening laws as in Figure 2. The error bars illustrate typical 2MASS photometric uncertainties.

that there is no evidence for an older pre-main-sequence stellar population within the nebula.

Appendix C Star Formation within Cloud 8

Cloud 8 is superimposed near the center, and densest part, of the expanding stellar group D. Given this configuration, it appears that Cloud 8 could be the main remnant of the cloud responsible for producing this stellar group. However, it is alternatively plausible that this superposition is coincidental, and that the cloud is closer to us than the stellar group is. It is particularly difficult to distinguish between these two possibilities because our sample does not include many objects within Cloud 8, presumably due to difficulty detecting obscured YSOs. However, Cambr esy et al. (2002) identified a dense, embedded cluster (number 6 in their catalog; hereafter Cambr esy 6) near the center of this cloud.

Figure C1 shows the Spitzer/IRAC 3.6 μm image along with our Gaia members (green circles), Spitzer YSOs from Guieu et al. (2009) and Rebull et al. (2011) not included in our sample (yellow squares), the contours of Cloud 8 (outer white curves), and the location of Cambr esy 6 (white circle). A star cluster can clearly be seen in the Spitzer image at the location of Cambr esy 6, but almost none of the individual cluster members were identified as YSOs by any survey. This suggests that either the stars did not have infrared excess or the mid-infrared nebulosity in the region prevented reliable detection of infrared excess. Nevertheless, we suspect that Cambr esy 6 is young because, in XMM-Newton images (not shown), a group of spatially confused X-ray point sources can be seen at this location. The high absorption of this group and its spatial coincidence with the molecular gas suggest that it lies within Cloud 8 and that the cloud is actively forming stars. Nevertheless, the relation between Cambr esy 6 and Group D remains a matter for future studies.

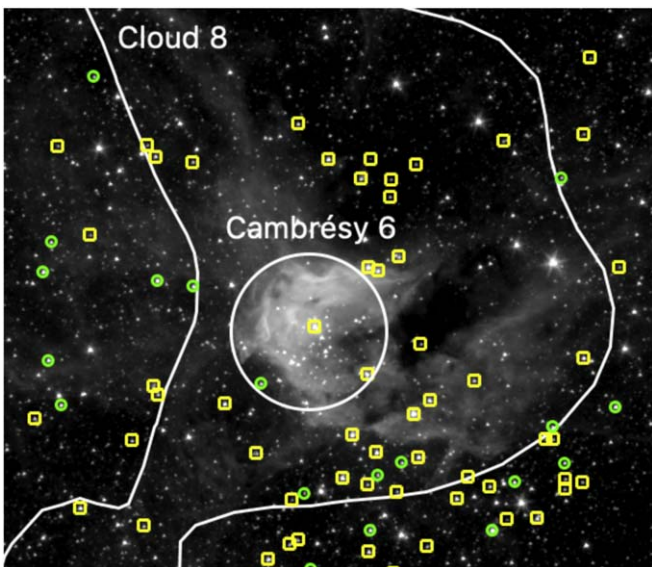


Figure C1. Spitzer/IRAC 3.6 μm image of the region around Cloud 8 (white contour). Identified YSO candidates are shown, including sources from Group D (green circles) and Spitzer candidates from Rebull et al. (2011) not included in our study (yellow squares). A $2\frac{1}{5}$ circle encompasses the cluster Cambr esy 6.

Appendix D Modeling Linear Expansion

Here we describe a linear model to relate the position of a star to its velocity. This model can account for expansion (homologous or anisotropic), contraction, and even some rotational effects.

We model the velocity of the i th star, $\mathbf{v}_i = (v_{x,i}, v_{y,i})$, as being related to its position, $\mathbf{x}_i = (x_i, y_i)$, by the linear regression equation

$$\mathbf{v}_i = \mathbf{v}_0 + M\mathbf{x}_i + \mathbf{v}_{\text{scatter},i}, \quad (\text{D1})$$

where \mathbf{v}_0 is a constant velocity shift, M is a 2×2 matrix, and the $\mathbf{v}_{\text{scatter},i}$ is a random vector drawn from a bivariate normal distribution with covariance matrix Σ_{scatter} . We use $\mathbf{v}_{\text{scatter}}$ to represent the intrinsic random deviations in velocity. In addition to the intrinsic scatter in velocity, in the observed velocities, $\mathbf{v}_{\text{obs},i}$, we must also account for Gaia’s measurement uncertainty. Thus, $\mathbf{v}_{\text{obs},i}$ is related to its actual velocity by the equation

$$\mathbf{v}_{\text{obs},i} = \mathbf{v}_i + \boldsymbol{\varepsilon}_i, \quad (\text{D2})$$

where measurement errors $\boldsymbol{\varepsilon}_i$ are drawn from Gaussian distributions with covariance matrices $\Sigma_{\boldsymbol{\varepsilon},i}$ obtained from the Gaia DR2 catalog. Putting this together gives us the likelihood equation,

$$\mathbf{v}_{\text{obs},i} \sim \mathcal{N}(\mathbf{v}_0 + M\mathbf{x}_i, \Sigma_{\text{scatter}} + \Sigma_{\boldsymbol{\varepsilon},i}), \quad (\text{D3})$$

from which we estimate the model parameters using a Bayesian approach.

For the Bayesian model fitting, we use Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution, which is implemented by the software package “Just another Gibbs sampler” (JAGS) version 4.3.0 (Plummer 2017), which is run using *rjags* (Plummer 2019) in R version 3.6.0. We use “noninformative” priors for the parameters, including uniform distributions for $v_{0,x}$ and $v_{0,y}$ from -2 to 2 km s^{-1} , and broad distributions for M and Σ_{scatter} that cover the reasonable values for these parameters.¹¹ The statistical model was implemented in the BUGS language. For each data set, three independent chains were run for 5000 iterations, retaining every fifth sample, after an initial burn-in of 1000 iterations. Convergence was assessed from inspection of the trace plots, as well as the Gelman–Rubin statistic < 1.001 (Gelman & Rubin 1992). Through experimentation with a variety of priors, we found that changes to the functional form for the priors have little effect on the results, provided that the priors are sufficiently broad. We also find that the results of the Bayesian analysis are approximately consistent with the results from numerical maximum likelihood estimation.

The 2×2 matrix M describes the dependence of velocity on position, encoding phenomena such as expansion, contraction, or rotation, which may be either radially symmetric or anisotropic. To examine these effects, we decompose M using

¹¹ The prior distribution for the matrix Σ_{scatter} is made up of χ^2_1 distributions, scaled by a factor of 9, for the diagonal entries, and uniform distributions from -1 to 1 for the correlation coefficients of the off-diagonal entries. For A , we base our priors on the decomposition of this matrix described later on: $D_{1,1}$ and $D_{2,2}$ use Gaussian priors with a mean of 0 and standard deviation of 1.5, and θ and φ use uniform priors from 0 to 2π .

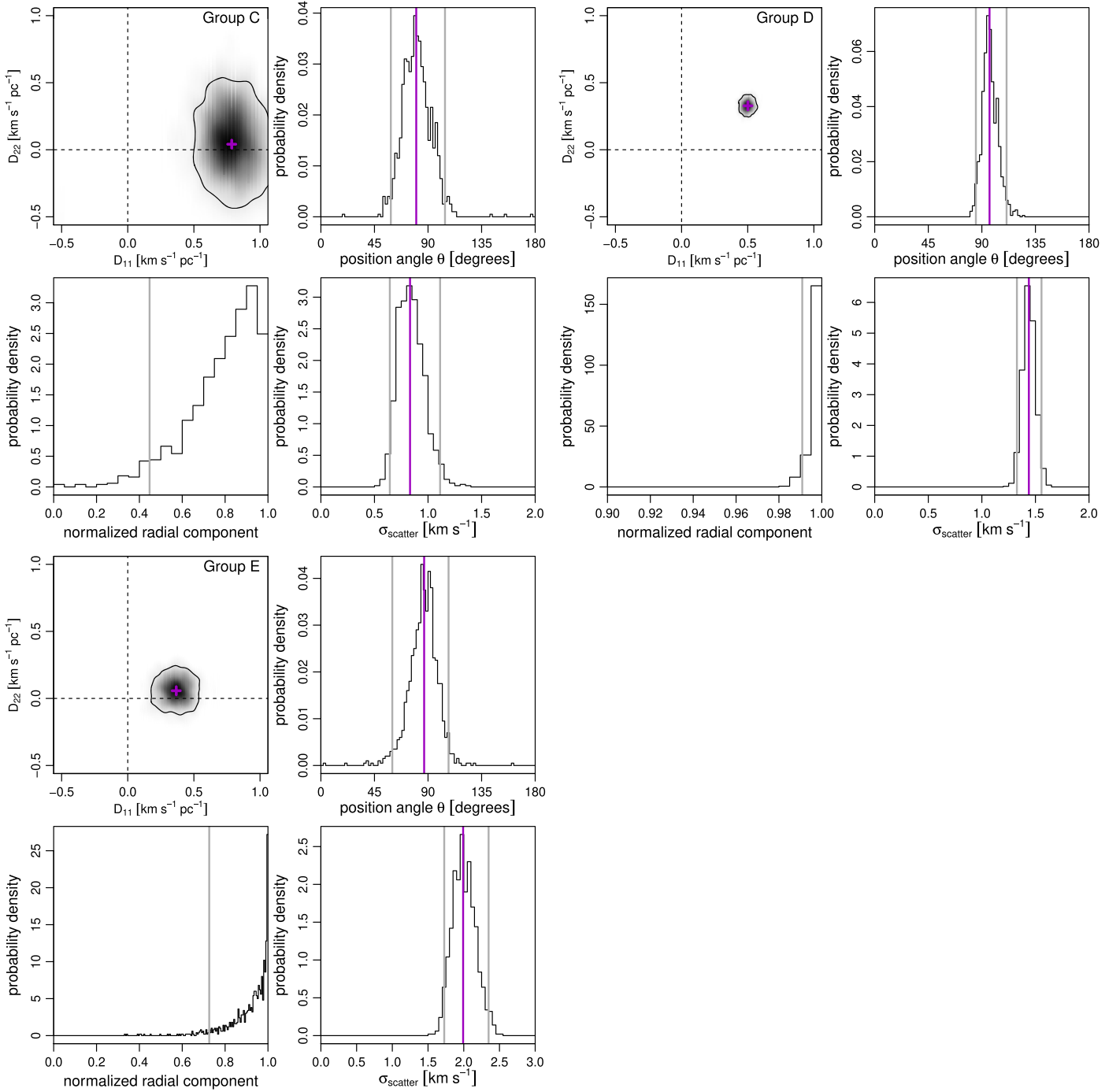


Figure D1. Marginal distributions for parameters of the linear velocity model in Groups C, D and E. These are the three stellar groups with enough sources to provide meaningful constraints on the model. For each fit, distributions are shown for the values of D (top left), the position angle θ of the anisotropy (top right), the radial component of the velocity gradient (bottom left), and the random velocity scatter (bottom right). The black contour lines (top left) and vertical gray lines (other graphs) enclose the regions of parameter space with 95% of the probability.

singular value decomposition (SVD) to write

$$M = UDV^*, \quad (\text{D4})$$

where U and V^* are unitary matrices and D is a diagonal matrix. By multiplying by the identity matrix, written as the product of two reflection matrices, without loss of generality, we can take U and V^* to be rotation matrices that rotate the coordinate systems by angles φ and θ , respectively, and D to be a diagonal matrix with entries $D_{1,1}$ and $D_{2,2}$, where the diagonal elements can be positive, negative, or zero, but $|D_{1,1}| \geq |D_{2,2}|$. Thus, M

first rotates the coordinate system by angle θ , the position angle of the velocity anisotropy; then applies the velocity gradient $D_{1,1}$, with units of $\text{km s}^{-1} \text{pc}^{-1}$, along the major axis of the anisotropy and $D_{2,2}$ along the minor axis; and finally rotates the coordinate system again by φ . The coordinate system rotations mean that the component of the velocity in the outward direction will be $\cos(\varphi + \theta)$ and the component of the velocity in the azimuthal direction will be $\sin(\varphi + \theta)$.

To interpret these results, we note that statistically significant nonzero values of $D_{1,1}$ or $D_{2,2}$ imply dependence of velocity on

position. If the velocity dependence is primarily in the radial direction (i.e., $\cos[\theta + \varphi] \approx 1$), then $D_{1,1}, D_{2,2} > 0$ implies expansion of the system, while $D_{1,1}, D_{2,2} < 0$ implies contraction. If the velocity dependence is primarily in the azimuthal direction (i.e., $\sin[\theta + \varphi] \approx \pm 1$), nonzero values of $D_{1,1}$ and $D_{2,2}$ imply rotation.

Figure D1 shows the marginal distributions for several of the interesting parameters in the linear models. For Group D, the allowed area of parameter space is very small, so constraints on the velocity gradients are tight. Both velocity gradients are positive, with $D_{1,1}$ being noticeably larger than $D_{2,2}$. For Groups C and E, the values of $D_{1,1}$ are constrained to be positive, but the uncertainties in $D_{2,2}$ are large enough that this parameter could equal zero. The marginal distributions of the position angles θ show that these are fairly tightly constrained for all three groups, C, D, and E. We also include marginal distributions of the normalized radial component defined as $\cos(\theta + \varphi)$. These distributions all favor values close to 1, showing that the velocity gradients imply expansion, not rotation.

Appendix E Nearby Clusters

Examination of the Gaia $\mu_{\alpha^*}-\mu_{\delta}$ distribution in the vicinity of the NAP region reveals two salient features that are not directly related to the NAP region. One of these, NGC 6997, is a star cluster superimposed on the “North America” region but generally thought to be at a different distance. This group of ~ 600 Gaia sources stands out clearly in proper-motion space because it has a smaller velocity dispersion, but it also has significantly different mean proper motions of $\mu_{\alpha^*} = -4.3$ mas yr $^{-1}$ and $\mu_{\delta} = -6.8$ mas yr $^{-1}$, meaning that confusion between its members and those of the NAP complex is minimal. The group has a parallax of 1.12 mas, meaning that it is ~ 100 pc farther than the main NAP complex.

The other feature is a group of ~ 50 Gaia sources centered at 21:01:53+45:12:00 with $\mu_{\alpha^*} = -1.7$ mas yr $^{-1}$, $\mu_{\delta} = -3.8$ mas yr $^{-1}$, and $\varpi_0 = 1.0$ mas.

ORCID iDs

Michael A. Kuhn  <https://orcid.org/0000-0002-0631-7514>

John M. Carpenter  <https://orcid.org/0000-0003-2251-0602>

References

- Adams, F. C. 2000, *ApJ*, **542**, 964
- Allen, L., Megeath, S. T., Gutermuth, R., et al. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 361
- Alves, J., Zucker, C., Goodman, A. A., et al. 2020, *Natur*, **578**, 237
- Arce, H. G., Shepherd, D., Gueth, F., et al. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil (Tucson, AZ: Univ. Arizona Press), 245
- Argiroffi, C., Caramazza, M., Micela, G., et al. 2016, *A&A*, **589**, A113
- Armond, T., Reipurth, B., Bally, J., & Aspin, C. 2011, *A&A*, **528**, A125
- Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
- Bally, J., Ginsburg, A., Probst, R., et al. 2014, *AJ*, **148**, 120
- Bally, J., & Reipurth, B. 2003, *AJ*, **126**, 893
- Bally, J., & Scoville, N. Z. 1980, *ApJ*, **239**, 121
- Baudry, J.-P., Raftery, A. E., Celeux, G., Lo, K., & Gottardo, R. 2010, *Journal of Computational and Graphical Statistics*, **19**, 332
- Baumgardt, H., & Kroupa, P. 2007, *MNRAS*, **380**, 1589
- Bertoldi, F., & McKee, C. F. 1992, *ApJ*, **395**, 140
- Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, *ARA&A*, **51**, 207
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, **427**, 127
- Broos, P. S., Getman, K. V., Povich, M. S., et al. 2013, *ApJS*, **209**, 32
- Cambřesy, L., Beichman, C. A., Jarrett, T. H., & Cutri, R. M. 2002, *AJ*, **123**, 2559
- Cánovas, H., Cantero, C., Cieza, L., et al. 2019, *A&A*, **626**, A80
- Carpenter, J. M. 2000, *AJ*, **120**, 3139
- Chevanec, M., Kruijssen, J. M. D., Hygate, A. P. S., et al. 2020, *MNRAS*, **493**, 2872
- Cohen, M., & Kuhl, L. V. 1979, *ApJS*, **41**, 743
- Comerón, F., & Pasquali, A. 2005, *A&A*, **430**, 541
- Corbally, C. J., Straižys, V., & Laugalys, V. 2009, *BaltA*, **18**, 111
- Da Rio, N., Tan, J. C., Covey, K. R., et al. 2017, *ApJ*, **845**, 105
- Damiani, F., Pillitteri, I., & Prisinzano, L. 2017, *A&A*, **602**, A115
- Elmegreen, B. G. 2000, *ApJ*, **530**, 277
- Elmegreen, B. G. 2002, *ApJ*, **577**, 206
- Erickson, N. R., Grosslein, R. M., Erickson, R. B., & Weinreb, S. 1999, *ITMTT*, **47**, 2212
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. 2011, *Cluster Analysis* (New York: Wiley)
- Feigelson, E. D. 2018, in *The Birth of Star Clusters*, ed. S. Stahler (Berlin: Springer), 119
- Feigelson, E. D., & Montmerle, T. 1999, *ARA&A*, **37**, 363
- Findeisen, K., Hillenbrand, L., Ofek, E., et al. 2013, *ApJ*, **768**, 93
- Fleming, G. D., Kirk, J. M., Ward-Thompson, D., & Pattle, K. 2019, arXiv:1904.06980
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, **616**, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, **595**, A1
- Galli, P. A. B., Loinard, L., Bouy, H., et al. 2019, *A&A*, **630**, A137
- Gelman, A., & Rubin, D. B. 1992, *StaSc*, **7**, 457
- Getman, K. V., Feigelson, E. D., Garmire, G., Broos, P., & Wang, J. 2007, *ApJ*, **654**, 316
- Getman, K. V., Feigelson, E. D., Kuhn, M. A., & Garmire, G. P. 2019, *MNRAS*, **487**, 2977
- Getman, K. V., Feigelson, E. D., Sicilia-Aguilar, A., et al. 2012, *MNRAS*, **426**, 2917
- Gieles, M., Portegies Zwart, S. F., Baumgardt, H., et al. 2006, *MNRAS*, **371**, 793
- Goldsmith, P. F., & Langer, W. D. 1999, *ApJ*, **517**, 209
- González, M., & Alfaro, E. J. 2017, *MNRAS*, **465**, 1889
- Gouliermis, D. A. 2018, *PASP*, **130**, 072001
- Großschedl, J. E., Alves, J., Meingast, S., et al. 2018, *A&A*, **619**, A106
- Grobieu, S., Rebull, L. M., Stauffer, J. R., et al. 2009, *ApJ*, **697**, 787
- Guszejnov, D., Grudić, M. Y., Offner, S. S. R., et al. 2020, *MNRAS*, **492**, 488
- Gutermuth, R. A., Pipher, J. L., Megeath, S. T., et al. 2011, *ApJ*, **739**, 84
- Hacar, A., Alves, J., Burkert, A., & Goldsmith, P. 2016, *A&A*, **591**, A104
- Hartmann, L., Ballesteros-Paredes, J., & Heitsch, F. 2012, *MNRAS*, **420**, 1457
- Herbig, G. H. 1958, *ApJ*, **128**, 259
- Herczeg, G. J., & Hillenbrand, L. A. 2015, *ApJ*, **808**, 23
- Herczeg, G. J., Kuhn, M. A., Zhou, X., et al. 2019, *ApJ*, **878**, 111
- Heyer, M., & Dame, T. M. 2015, *ARA&A*, **53**, 583
- Heyer, M. H., & Terebey, S. 1998, *ApJ*, **502**, 265
- Hillenbrand, L. A., & Baliber, N. 2015, *IAUGA*, **29**, 2253943
- Howarth, I. D., & van Leeuwen, F. 2019, *MNRAS*, **484**, 5350
- Joye, W. A., & Mandel, E. 2003, in *ASP Conf. Ser. 295, Astronomical Data Analysis Software and Systems XII*, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook (San Francisco, CA: ASP), 489
- Klessen, R. S. 2000, *ApJ*, **535**, 869
- Klessen, R. S., Burkert, A., & Bate, M. R. 1998, *ApJL*, **501**, L205
- Kounkel, M., Covey, K., Suárez, G., et al. 2018, *AJ*, **156**, 84
- Kroupa, P., Aarseth, S., & Hurley, J. 2001, *MNRAS*, **321**, 699
- Kruijssen, J. M. D., Maschberger, T., Moeckel, N., et al. 2012, *MNRAS*, **419**, 841
- Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., Portegies Zwart, S. F., & Icke, V. 2011, *MNRAS*, **414**, 1339
- Krumholz, M. R., Matzner, C. D., & McKee, C. F. 2006, *ApJ*, **653**, 361
- Krumholz, M. R., & McKee, C. F. 2020, *MNRAS*, **494**, 624
- Kuhn, M. A., & Feigelson, E. D. 2019, in *Applications in Astronomy*, ed. S. Fruhwirth-Schnatter, G. Celeux, & C. P. Robert (London, Boca Raton, FL: Chapman and Hall, CRC), 463
- Kuhn, M. A., Hillenbrand, L. A., Sills, A., Feigelson, E. D., & Getman, K. V. 2019, *ApJ*, **870**, 32
- Kutner, M. L., & Ulich, B. L. 1981, *ApJ*, **250**, 341
- Lada, C. J., Lombardi, M., Roman-Zuniga, C., Forbrich, J., & Alves, J. F. 2013, *ApJ*, **778**, 133

- Larson, R. B. 1981, *MNRAS*, **194**, 809
- Laugalys, V., Straizys, V., Vrba, F. J., et al. 2006, *BaltA*, **15**, 483
- Leung, H. W., & Bovy, J. 2019, *MNRAS*, **489**, 2079
- Li, Z.-Y., & Nakamura, F. 2006, *ApJL*, **640**, L187
- Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, **616**, A2
- Longmore, S. N., Kruijssen, J. M. D., Bastian, N., et al. 2014, in *Protostars and Planets VI*, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 291
- Luhman, K. L. 2018, *AJ*, **156**, 271
- Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N. 2009, *Statistics in Medicine*, **28**, 3049
- Lynds, B. T. 1962, *ApJS*, **7**, 1
- Maíz Apellániz, J., Sota, A., Arias, J. I., et al. 2016, *ApJS*, **224**, 4
- Mamajek, E. E. 2009, in *AIP Conf. Ser.* 1158, *Exoplanets and Disks: Their Formation and Diversity*, ed. T. Usuda, M. Tamura, & M. Ishii (Melville, NY: AIP), 3
- Mangum, J. G., & Shirley, Y. L. 2015, *PASP*, **127**, 266
- Marcy, G. W. 1980, *AJ*, **85**, 230
- McKinney, W. 2010, in *Proc. 9th Python in Science Conf.*, ed. S. van der Walt & J. Millman (Trieste: SISSA), 56
- Melnik, A. M., & Dambis, A. K. 2020, *MNRAS*, **493**, 2339
- Merrill, P. W., & Burwell, C. G. 1949, *ApJ*, **110**, 387
- Nakamura, F., & Li, Z.-Y. 2011, *ApJ*, **740**, 36
- Ogura, K., Sugitani, K., & Pickles, A. 2002, *AJ*, **123**, 2597
- Oh, S., & Kroupa, P. 2016, *A&A*, **590**, A107
- Orkisz, J. H., Pety, J., Gerin, M., et al. 2017, *A&A*, **599**, A99
- Padoan, P., & Nordlund, Å. 2011, *ApJ*, **730**, 40
- Pandas Development Team 2020, *pandas-dev/pandas: Pandas, v1.0.5*, Zenodo, doi:10.5281/zenodo.3898987
- Parker, R. J., Wright, N. J., Goodwin, S. P., & Meyer, M. R. 2014, *MNRAS*, **438**, 620
- Patel, N. A., Goldsmith, P. F., Heyer, M. H., Snell, R. L., & Pratap, P. 1998, *ApJ*, **507**, 241
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2012, arXiv:1201.0490
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, **500**, 501
- Plummer, M. 2017, *JAGS Manual*, <https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/>
- Plummer, M. 2019, *rjags: Bayesian Graphical Models Using MCMC*, <https://CRAN.R-project.org/package=rjags>
- PostgreSQL development team 2020, *PostgreSQL, Version 12.3*, <https://www.postgresql.org/>, Accessed on 2019 July 5
- Preibisch, T., Fleischlen, S., Gaczkowski, B., Townsley, L., & Broos, P. 2017, *A&A*, **605**, A85
- Preibisch, T., & Zinnecker, H. 1999, *AJ*, **117**, 2381
- R Core Team 2018, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>
- Rebull, L. M., Guieu, S., Stauffer, J. R., et al. 2011, *ApJS*, **193**, 25
- Reipurth, B., & Schneider, N. 2008, in *Handbook of Star Forming Regions, Volume I: The Northern Sky* ASP Monograph Publications, ed. B. Reipurth (San Francisco, CA: ASP), 36
- Renzo, M., Zapartas, E., de Mink, S. E., et al. 2019, *A&A*, **624**, A66
- Ribas, Á, Bouy, H., & Merín, B. 2015, *A&A*, **576**, A52
- Richert, A. J. W., Getman, K. V., Feigelson, E. D., et al. 2018, *MNRAS*, **477**, 5191
- Rieke, G. H., & Lebofsky, M. J. 1985, *ApJ*, **288**, 618
- Robitaille, T. P., Meade, M. R., Babler, B. L., et al. 2008, *AJ*, **136**, 2413
- Román-Zúñiga, C. G., Roman-Lopes, A., Tapia, M., Hernández, J., & Ramírez-Preciado, V. 2019, *ApJL*, **871**, L12
- Schoettler, C., de Bruijne, J., Vaher, E., & Parker, R. J. 2020, *MNRAS*, **495**, 3104
- Schwarz, G. 1978, *AnSta*, **6**, 461
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. 2016, *The R Journal*, **8**, 205
- Singh, A., Matzner, C. D., & Jumper, P. H. 2019, *ApJ*, **878**, 22
- Skiff, B. A. 2009, *yCat*, **1**, 2023
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, **131**, 1163
- Sota, A., Maíz Apellániz, J., Walborn, N. R., et al. 2011, *ApJS*, **193**, 24
- Spitzer, L. J. 1958, *ApJ*, **127**, 17
- Straizys, V., & Laugalys, V. 2008, *BaltA*, **17**, 143
- Tan, J. C., Krumholz, M. R., & McKee, C. F. 2006, *ApJL*, **641**, L121
- Taylor, M. B. 2005, in *ASP Conf. Ser.* 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Tobin, J. J., Hartmann, L., Furesz, G., Mateo, M., & Megeath, S. T. 2009, *ApJ*, **697**, 1103
- Tutukov, A. V. 1978, *A&A*, **70**, 57
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
- Vázquez-Semadeni, E., González-Samaniego, A., & Colín, P. 2017, *MNRAS*, **467**, 1313
- Vázquez-Semadeni, E., Palau, A., Ballesteros-Paredes, J., Gómez, G. C., & Zamora-Avilés, M. 2019, *MNRAS*, **490**, 3061
- Venables, W. N., & Ripley, B. D. 2002, *Modern Applied Statistics with S* (4th ed.; New York: Springer), <http://www.stats.ox.ac.uk/pub/MASS4>
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, **17**, 261
- Vogt, S. S., Allen, S. L., Bigelow, B. C., et al. 1994, *Proc. SPIE*, **2198**, 362
- Ward, J. L., & Kruijssen, J. M. D. 2018, *MNRAS*, **475**, 5659
- Welin, G. 1973, *A&AS*, **9**, 183
- Wendker, H. J., Benz, D., & Baars, J. W. M. 1983, *A&A*, **124**, 116
- Westerhout, G. 1958, *BAN*, **14**, 215
- Witham, A. R., Knigge, C., Drew, J. E., et al. 2008, *MNRAS*, **384**, 1277
- Wright, N. J., Bouy, H., Drew, J. E., et al. 2016, *MNRAS*, **460**, 2593
- Wright, N. J., Jeffries, R. D., Jackson, R. J., et al. 2019, *MNRAS*, **486**, 2477
- Wright, N. J., & Mamajek, E. E. 2018, *MNRAS*, **476**, 381
- Wright, N. J., & Parker, R. J. 2019, *MNRAS*, **489**, 2694
- Zamora-Avilés, M., Ballesteros-Paredes, J., Hernández, J., et al. 2019, *MNRAS*, **488**, 3406
- Zhang, S., Xu, Y., & Yang, J. 2014, *AJ*, **147**, 46
- Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2020, *A&A*, **633**, 51
- Zuckerman, B., & Evans, N. J. I. 1974, *ApJL*, **192**, L149