

# THE BOLOCAM GALACTIC PLANE SURVEY. XII. DISTANCE CATALOG EXPANSION USING KINEMATIC ISOLATION OF DENSE MOLECULAR CLOUD STRUCTURES WITH $^{13}\text{CO}(1-0)$

TIMOTHY P. ELLSWORTH-BOWERS<sup>1</sup>, ERIK ROSOLOWSKY<sup>2</sup>, JASON GLENN<sup>1</sup>, ADAM GINSBURG<sup>3</sup>, NEAL J. EVANS II<sup>4</sup>, CARA BATTERSBY<sup>5</sup>, YANCY L. SHIRLEY<sup>6,7</sup>, AND BRIAN SVOBODA<sup>6</sup>

<sup>1</sup> CASA, University of Colorado, UCB 389, Boulder, CO 80309, USA; [timothy.ellsworthbowers@colorado.edu](mailto:timothy.ellsworthbowers@colorado.edu)

<sup>2</sup> Department of Physics, 4-183 CCIS, University of Alberta, Edmonton, AB T6G 2E1, Canada

<sup>3</sup> European Southern Observatory, Karl-Schwarzschild-Straße 2, D-85748 Garching bei München, Germany

<sup>4</sup> Department of Astronomy, University of Texas, 2515 Speedway, Stop C1400, Austin, TX 78712, USA

<sup>5</sup> Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

<sup>6</sup> Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, AZ 85721, USA

Received 2014 April 22; accepted 2014 November 4; published 2015 January 13

## ABSTRACT

We present an expanded distance catalog for 1710 molecular cloud structures identified in the Bolocam Galactic Plane Survey (BGPS) version 2, representing a nearly threefold increase over the previous BGPS distance catalog. We additionally present a new method for incorporating extant data sets into our Bayesian distance probability density function (DPDF) methodology. To augment the dense-gas tracers (e.g.,  $\text{HCO}^+(3-2)$ ,  $\text{NH}_3(1,1)$ ) used to derive line-of-sight velocities for kinematic distances, we utilize the Galactic Ring Survey (GRS)  $^{13}\text{CO}(1-0)$  data to morphologically extract velocities for BGPS sources. The outline of a BGPS source is used to select a region of the GRS  $^{13}\text{CO}$  data, along with a reference region to subtract enveloping diffuse emission, to produce a line profile of  $^{13}\text{CO}$  matched to the BGPS source. For objects with a  $\text{HCO}^+(3-2)$  velocity,  $\approx 95\%$  of the new  $^{13}\text{CO}(1-0)$  velocities agree with that of the dense gas. A new prior DPDF for kinematic distance ambiguity (KDA) resolution, based on a validated formalism for associating molecular cloud structures with known objects from the literature, is presented. We demonstrate this prior using catalogs of masers with trigonometric parallaxes and H II regions with robust KDA resolutions. The distance catalog presented here contains well-constrained distance estimates for 20% of BGPS V2 sources, with typical distance uncertainties  $\lesssim 0.5$  kpc. Approximately 75% of the well-constrained sources lie within 6 kpc of the Sun, concentrated in the Scutum–Centaurus arm. Galactocentric positions of objects additionally trace out portions of the Sagittarius, Perseus, and Outer arms in the first and second Galactic quadrants, and we also find evidence for significant regions of interarm dense gas.

*Key words:* Galaxy: kinematics and dynamics – Galaxy: structure – ISM: clouds – methods: data analysis – stars: formation – submillimeter: ISM

*Supporting material:* machine-readable tables

## 1. INTRODUCTION

Continuum surveys of the Galactic plane at (sub)millimeter wavelengths (BGPS, Aguirre et al. 2011; Ginsburg et al. 2013; ATLASGAL, Schuller et al. 2009; Hi-GAL, Molinari et al. 2010), as well as all-sky cosmic microwave background missions (e.g., Planck Collaboration et al. 2011a, 2011b, 2014), have cataloged tens of thousands of dense molecular cloud cores and clumps; the possible precursors to stellar clusters, OB associations, or smaller stellar groups. Derived distances to and physical properties of these objects may answer several outstanding questions about massive star formation (Kennicutt & Evans 2012). What is the Galactic distribution of massive star formation in the Milky Way? What is the clump mass function and its relationship to the stellar initial mass function? What are the evolutionary processes of the dense interstellar medium?

While stellar and extragalactic studies may make use of standard(izable) candles, no intrinsic luminosity–distance relationship exists for molecular cloud structures. Distance estimates for these objects must, therefore, rely upon ancillary data. The distance probability density function (DPDF) formalism introduced by Ellsworth-Bowers et al. (2013, hereafter EB13) provides a means for combining an arbitrary number of ancillary data sets to derive distance estimates for molec-

ular cloud structures detected by continuum surveys. The primary input to this Bayesian method is a kinematic distance, whereby a measured line-of-sight velocity ( $v_{\text{LSR}}$ ) is combined with a Galactic rotation curve. Geometric considerations reveal that two heliocentric distances ( $d_{\odot}$ ) may correspond to a single  $v_{\text{LSR}}$  for Galactocentric locations within the solar circle, the so-called kinematic distance ambiguity (KDA). The DPDF formalism utilizes data from other wavelengths to place priors on the kinematic distance likelihood in an effort to resolve the KDA. This paper expands on the DPDF methodology by introducing a new source for deriving kinematic distance likelihoods and also a validated formalism for the association of detected molecular cloud structures with published catalogs of objects having robust distance measurements.

Other than being hindered by the KDA, the use of kinematic distances is limited only by the particulars of the chosen rotation curve and molecular line detection rates. Differences between rotation curves in the Galactic orbital velocity and the distance to the Galactic center can lead to  $\gtrsim 20\%$  differences in distances, or  $\gtrsim 40\%$  uncertainties in mass (Anderson et al. 2012). In terms of molecular line detection, the low- $J$  transitions of  $^{12}\text{CO}$ , while ubiquitous throughout the Galactic plane, are excited at low density, are generally optically thick, and do not provide a unique tracer for the dense molecular cloud structures detected in continuum surveys around  $\lambda = 1$  mm. Other molecular transitions excited only at higher density (e.g.,  $\text{HCO}^+(3-2)$ ,

<sup>7</sup> Adjunct Astronomer at the National Radio Astronomy Observatory.

NH<sub>3</sub>(1,1), CS(2–1); Evans 1999) have been targeted by pointed observations in an effort to derive single-component  $v_{\text{LSR}}$  for these structures (see Shirley et al. 2013; Wienen et al. 2012; Jackson et al. 2008). Detection rates using these transitions, however, can be somewhat poor ( $\sim 50\%$ – $85\%$ ), refocusing attention on CO for measuring  $v_{\text{LSR}}$ .

The largest single Galactic molecular line data set to date is the compiled <sup>12</sup>CO(1–0) surveys presented in Dame et al. (2001). With a resolution of several arcminutes, this mosaic revealed the basic structure of the Milky Way as seen in giant molecular clouds (GMCs). The combination of opacity with the low resolution means that structure visible in the Dame et al. data represents primarily the surfaces of GMCs, and not the denser interior clumps. The isotopologue <sup>13</sup>CO is  $\sim 50$ – $90$  times optically thinner (depending on Galactocentric location) and is a better tracer of structure within GMCs. In large-scale <sup>13</sup>CO(1–0) data, such as the Galactic Ring Survey (GRS; Jackson et al. 2006), it is simplest to identify and study GMC-scale structures (Roman-Duval et al. 2009, 2010), although it is possible to identify the clump-scale structures usually detected by continuum surveys (Rathborne et al. 2009). Because molecular cloud clumps exist within an envelope of more diffuse gas, disentangling the emission from a single continuum-detected object is challenging (Simon et al. 2006b; Battisti & Heyer 2014). The low effective density for <sup>13</sup>CO(1–0) excitation also significantly increases the detection of multiple cloud structures along a given line of sight. Direct  $v_{\text{LSR}}$  comparison of the brightest <sup>13</sup>CO peak along the line of sight toward continuum-detected objects with the  $v_{\text{LSR}}$  from a dense-gas tracer like HCO<sup>+</sup>(3–2) yields a  $\sim 81\%$  agreement rate (Shirley et al. 2013). In this work we present a technique for extracting a <sup>13</sup>CO  $v_{\text{LSR}}$  for molecular cloud structures using the morphology and flux density from continuum data as prior information to increase the agreement rate to  $\sim 95\%$ .

The kinematic distance likelihoods from <sup>13</sup>CO data expand the catalog of objects with detected  $v_{\text{LSR}}$  beyond the sample of dense-gas surveys, but the longitude and velocity range of the GRS implies all new sources will be subject to the KDA. To resolve the ambiguities, EB13 presented prior DPDFs based on eight-micron absorption features (EMAFs) and a simple model for the Galactic distribution of molecular gas; we expand upon that set here. With the recent publication of distance catalogs for large numbers of objects associated with sites of high-mass star formation it becomes economical to construct a mechanism for associating these distances with continuum-identified molecular cloud structures based on angular and velocity separations. We demonstrate this new prior with “gold-standard” trigonometric parallax measurements to CH<sub>3</sub>OH and H<sub>2</sub>O masers from the Bar and Spiral Structure Legacy (BeSSeL; Brunthaler et al. 2011) Survey and Japanese VLBI Exploration of Radio Astrometry (VERA) project<sup>8</sup> as well as robust KDA resolutions to H II regions using H I absorption techniques from the H II Region Discovery Surveys (HRDS; Bania et al. 2010, 2012).

The recent re-reduction and release of the BGPS version 2 data and associated catalog (Ginsburg et al. 2013, hereafter G13) provides a solid basis for derivation of posterior DPDFs for molecular cloud structures across the Galactic plane. We present a new distance catalog based on the BGPS V2.1 catalog, the methods presented in EB13, and those developed here. While molecular cloud clumps are the objects of interest for understanding the larger process of massive star formation

(McKee & Ostriker 2007), surveys like the BGPS are sensitive to scales ranging from cores that will from a single stellar system to GMCs themselves at the largest distances. We therefore use the term “molecular cloud structure” when discussing detected objects to account for this uncertainty. A forthcoming work will further discuss the nature of BGPS sources in terms of physical properties derived using DPDF distance estimates.

This paper is organized as follows. Section 2 describes the data used and the distance computation code. An overview of the DPDF formalism and relevant Galactic kinematics is given in Section 3. The estimation of molecular cloud clump  $v_{\text{LSR}}$  from <sup>13</sup>CO data is described in Section 4, while the new prior DPDFs are outlined in Section 5. Section 6 presents the results, and implications of this work are discussed in Section 7. A summary is presented in Section 8.

## 2. DATA

### 2.1. The Bolocam Galactic Plane Survey

The Bolocam Galactic Plane Survey version 2 (BGPS; Aguirre et al. 2011; G13), is a  $\lambda = 1.1$  mm continuum survey covering 192 deg<sup>2</sup> at 33'' resolution. It is one of the first large-scale blind surveys of the Galactic plane in this region of the spectrum, covering  $-10^\circ \leq \ell \leq 90^\circ$  with at least  $|b| \leq 0.5$ , plus selected regions in the outer Galaxy. For a map of BGPS V2.0 coverage and details about observation methods and the data reduction pipeline, see G13. From the BGPS V2.0 images, 8594 millimeter dust-continuum sources were identified using a custom extraction pipeline. BGPS V2.0 pipeline products, including image mosaics and the catalog, are publicly available.<sup>9</sup>

The BGPS data pipeline removes atmospheric signal using a principle component analysis technique that discards common-mode timestream signals correlated among bolometers in the focal plane array. The pipeline attempts to iteratively identify astrophysical signal and prevent that signal from being discarded with the atmospheric signal. This filtering may be characterized by an angular transfer function that passes scales from approximately 33'' to 6' (see G13 for a full discussion). The effective angular size range of detected BGPS sources therefore corresponds to anything from molecular cloud cores up to entire GMCs depending on heliocentric distance (Dunham et al. 2011b).

### 2.2. Molecular Line Spectroscopic Observations

Several spectroscopic follow-up programs have been conducted to observe BGPS version 1 sources in molecular transition lines with excitation effective density<sup>10</sup>  $n_{\text{eff}} \gtrsim 10^3$  cm<sup>-3</sup>. All 6194 BGPS V1 sources with  $\ell \geq 7.5$  were simultaneously observed in the  $J = 3-2$  rotational transitions of HCO<sup>+</sup> ( $\nu = 267.6$  GHz) and N<sub>2</sub>H<sup>+</sup> ( $\nu = 279.5$  GHz) with the Heinrich Hertz Submillimeter Telescope (HHT) on Mt. Graham, Arizona (Schlingman et al. 2011; Shirley et al. 2013, hereafter S13). Detectability in this HCO<sup>+</sup> line is a strong function of millimeter flux density, and only  $\approx 50\%$  of pointings produced a  $\geq 3\sigma$  detection. Detected sources, however, rarely (1%) have multiple velocity components, yielding a very robust kinematic catalog.

To complement the HHT survey, 555 BGPS V1 sources in the range  $29^\circ \leq \ell \leq 31^\circ$  were observed in the  $J = 2-1$  rotational

<sup>8</sup> <http://veraserver.mtk.nao.ac.jp>

<sup>9</sup> Available through IPAC at [http://irsa.ipac.caltech.edu/data/BOLOCAM\\_GPS](http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS).

<sup>10</sup> The effective density required to produce a line with a main beam brightness temperature of 1 K; may be up to several orders of magnitude smaller than the critical density (Evans 1999).

transition of CS ( $\nu = 97.98$  GHz) using the Arizona Radio Observatory 12 m telescope on Kitt Peak (Y. Shirley 2012, private communication). The lower  $n_{\text{eff}}$  of this transition led to CS(2–1) detections for 45% of sources not detected by the HHT survey in this region. A further study of this Galactic longitude range was observed in 2013 November at the HHT using the  $J = 2-1$  rotational transition of C<sup>18</sup>O ( $\nu = 219.6$  GHz). From these observations, 190 unique detections were extracted (M. Lichtenberger 2014, private communication).

Finally, a total of 686 BGPS V1 sources in the inner Galaxy (Dunham et al. 2011b) and Gemini OB1 molecular cloud complex (Dunham et al. 2010) were observed in the lowest inversion transition lines of NH<sub>3</sub> near 24 GHz with the Robert F. Byrd Green Bank Telescope (GBT). The (1,1) inversion is the strongest NH<sub>3</sub> transition at the cold temperatures of BGPS sources ( $T \approx 20$  K), and is the sole ammonia inversion transition used for velocity fitting. Detection rates for NH<sub>3</sub>(1,1) range from 61% in the Gemini OB1 molecular cloud to 72% in the inner Galaxy.

### 2.3. The Galactic Ring Survey

The GRS (Jackson et al. 2006) conducted large-scale observations of <sup>13</sup>CO(1–0) emission in the northern Galactic plane using the FCRAO 14 m telescope. The survey covers the longitude range  $18^\circ \leq \ell \leq 55^\circ 7'$ , with select regions down to  $\ell \approx 14^\circ$ , and spans  $|b| \leq 1^\circ$ , with angular resolution and sampling of  $46''$  and  $22''$ , respectively. Velocity coverage is  $-5 \text{ km s}^{-1} \leq v_{\text{LSR}} \leq 135 \text{ km s}^{-1}$  for  $\ell \leq 40^\circ$ , and  $-5 \text{ km s}^{-1} \leq v_{\text{LSR}} \leq 85 \text{ km s}^{-1}$  for  $\ell \geq 40^\circ$ , with spectral resolution of  $0.212 \text{ km s}^{-1}$ . Because the GRS aimed to study the Galactic molecular ring at  $R_{\text{gal}} \approx 4-5$  kpc, negative  $v_{\text{LSR}}$  in this region were omitted. The GRS data have an rms sensitivity of 0.13 K.

### 2.4. The distance-omnibus Code Base

The code used to compute DPDFs for this work, while developed for use with the BGPS, is easily configurable for use with other large (sub)millimeter surveys. It is written in the Interactive Data Language (IDL),<sup>11</sup> and is publicly available.<sup>12</sup> Like the theoretical framework for the DPDFs, the distance-omnibus code base is extremely flexible, and additional prior DPDF methods may be developed and implemented through the use of text configuration files. The user has control over which DPDFs are computed as well as Galactic properties such as rotation curve and  $R_0$ . The code is designed to run nearly autonomously and create the various intermediate data products on the fly given the disk locations of the required ancillary data sets. Presently available in the v1.0 release are the methods presented in EB13 and this paper.

## 3. DPDFS AND GALACTIC KINEMATICS

### 3.1. Overview of the DPDF Formalism

The DPDF formalism introduced in EB13 provides a powerful yet flexible technique for applying a diverse set of ancillary data to the problem of estimating distances to continuum-detected molecular cloud structures in the Galactic plane. In this Bayesian framework, the posterior DPDF for an object or position on the sky is given by

$$\text{DPDF}(d_\odot) = \mathcal{L}(v_{\text{LSR}}, l, b; d_\odot) \prod_i P_i(l, b; d_\odot), \quad (1)$$

where  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$  is the kinematic distance likelihood function,<sup>13</sup> and the  $P_i(l, b; d_\odot)$  are prior DPDFs based on ancillary data, principally used to resolve the KDA for objects in the inner Galaxy. The posterior DPDF is normalized to unit integral probability. The kinematic distance likelihood function is obtained from the marginalization over  $v_{\text{LSR}}$  of the product of the projected rotation curve along  $(\ell, b)$  as a  $f(d_\odot, v_{\text{LSR}})$  with the molecular line emission spectrum (itself a function of  $v_{\text{LSR}}$ ), and is generally double peaked for Galactocentric positions within the solar circle (i.e.,  $R_{\text{gal}} < R_0$ ). Any number of priors may be applied to resolve the KDA for a given molecular cloud structure, and any given prior may be restricted in its applicability (limited longitude coverage, etc.).

While the true power of the DPDF lies in being a continuous function of  $d_\odot$ , there are instances where a single-value distance estimate is desired. Two primary distance estimates may be derived from a DPDF: the maximum-likelihood distance ( $d_{\text{ML}}$ ), and the weighted-average first-moment distance ( $\bar{d}$ ). The  $d_{\text{ML}}$  is simply the distance corresponding to the largest probability in the posterior DPDF, while  $\bar{d}$  is computed via  $\bar{d} = \int_0^\infty d_\odot \text{DPDF } d(d_\odot)$ . If the DPDF is well constrained to have a single peak,  $d_{\text{ML}}$  and  $\bar{d}$  will be nearly equivalent, but in cases where the KDA resolution is not well constrained, these distance estimates may be substantially different and  $\bar{d}$  becomes a biased estimator of the distance while  $d_{\text{ML}}$  becomes less reliable. For well-constrained distances away from the tangent point we recommend  $d_{\text{ML}}$ , but  $\bar{d}$  represents a better distance estimate for objects near the tangent point due to the complex, multimodal structure of  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$  in these regions.

### 3.2. Galactic Kinematics

#### 3.2.1. The Galactic Rotation Curve

The choice of Galactic rotation curve can have a significant effect on derived kinematic distances. Over the years, various approaches have been used for deriving the rotation curve of the Milky Way, a decidedly complex endeavor given our location in the Galactic plane. Clemens (1985) computed a rotation curve using measured tangent velocities as a function of Galactic longitude from the Massachusetts–Stony Brook Galactic plane CO survey (Sanders et al. 1985). Similarly, Brand & Blitz (1993) used tangent-point data from 21 cm H I surveys and H II region distance and velocity measurements to construct a rotation curve. Both curves have been widely used to compute kinematic distances for molecular cloud clumps across the Galactic plane.

In addition to determining the spatial structure of the Milky Way, the BeSSeL survey has used their six-dimensional trigonometric parallax data (position, velocity) to fit Galactic rotation curve models (see Reid et al. 2009b). Unlike curves based on tangent-point data that require assumed values for  $R_0$  and  $\Theta_0$  (the radius of the solar circle, and rotational speed of that circular orbit, respectively), the fits to parallax data aim to independently measure these fundamental Galactic parameters. Reid et al. (2014) fit a relatively simple rotation curve with speed  $\Theta(R) = \Theta_0 + (d\Theta/dR)(R - R_0)$  to the maser parallax data, and applied different sets of priors on certain model parameters to reflect varying confidence in the prior information. The fit parameters allowed to vary are the three circular rotation curve parameters ( $R_0, \Theta_0, (d\Theta/dR)$ ), the values of the solar peculiar

<sup>11</sup> <http://www.exelisvis.com/ProductsServices/IDL.aspx>

<sup>12</sup> <https://github.com/BGPS/distance-omnibus>

<sup>13</sup> In EB13, this was referred to as  $\text{DPDF}_{\text{kin}}$ . We shift to the present notation to distinguish between the kinematic distance likelihood function and the prior DPDFs used to resolve the KDA.

**Table 1**  
Galactic Rotation Curve Parameters

Parameter	Reid et al. (2009b)	Used in EB13 <sup>a</sup>	Reid et al. (2014) Model A5	Used Here
$R_0$ (kpc)	$8.4 \pm 0.6$	8.51	$8.34 \pm 0.16$	8.34
$\Theta_0$ ( $\text{km s}^{-1}$ )	$254 \pm 16$	244	$240 \pm 8$	240
$\frac{d\Theta}{dR}$ ( $\text{km s}^{-1} \text{kpc}^{-1}$ )	0	0	$-0.2 \pm 0.4$	0
$U_\odot$ ( $\text{km s}^{-1}$ )	10.0	11.1	$10.7 \pm 1.8$	10.7
$V_\odot$ ( $\text{km s}^{-1}$ )	5.2	12.24	$15.6 \pm 6.8$	15.6
$W_\odot$ ( $\text{km s}^{-1}$ )	7.2	7.25	$8.9 \pm 0.9$	8.9
$\overline{U}_s$ ( $\text{km s}^{-1}$ )	0	0	$2.9 \pm 2.1$	2.9
$\overline{V}_s$ ( $\text{km s}^{-1}$ )	-15	-6	$-1.6 \pm 6.8$	0

**Note.** <sup>a</sup> Values from M. Reid (2011, private communication), Schönrich et al. (2010).

motion ( $U_\odot$ ,  $V_\odot$ ,  $W_\odot$ ), and the mean deviant (non-circular) motions of high-mass star forming regions (HMSFRs) toward the Galactic center ( $\overline{U}_s$ ) and in the direction of Galactic rotation ( $\overline{V}_s$ ). As discussed by those authors,  $\overline{U}_s$  and  $\overline{V}_s$  may be thought of as a first approximation of the kinematic effects (i.e., non-circular motion) of spiral structure.

We adopt model A5 of Reid et al. (2014) for the computation of kinematic distances in the DPDF formalism. This model was fit using a subset of the parallax measurements where outliers have been excluded, and includes loose priors for the  $V_\odot$  component of the solar motion and the mean deviant motions of HMSFRs, reflecting the large uncertainties surrounding measurements of these values.<sup>14</sup> The resulting fit parameters are listed in Table 1, along with the corresponding values used here and in previous works.

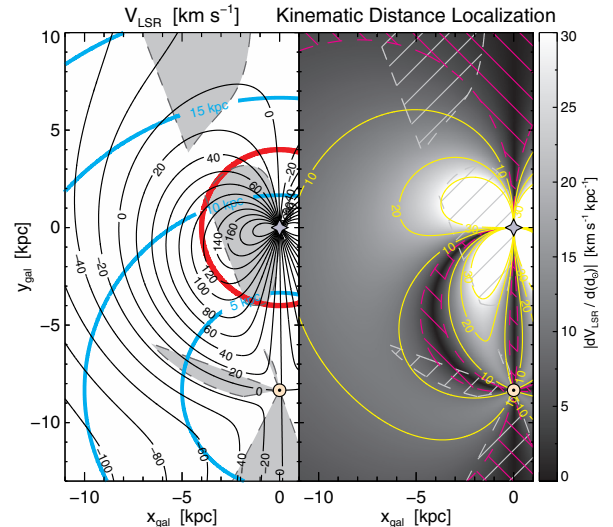
For simplicity, we apply a strictly flat rotation curve, consistent with the measured  $(d\Theta/dR) = -0.2 \pm 0.4 \text{ km s}^{-1} \text{ kpc}^{-1}$ , and we do not apply a correction for the counterrotation of HMSFRs ( $\overline{V}_s$ ), also consistent with the values from model A5. The nonzero mean source motion toward the Galactic center ( $\overline{U}_s$ ) from model A5 is included (see the derivation of the kinematic distance likelihood function in Appendix B), and results in a distance adjustment of  $\lesssim 0.2 \text{ kpc}$  for the vast majority of BGPS V2 sources. Contours in Figure 1 (left) illustrate the projected  $v_{\text{LSR}}$  derived from model A5 as a function of Galactocentric position for the northern Galactic plane.

### 3.2.2. Kinematic Avoidance Zones

Use of kinematic distances for molecular cloud structures is predicated on two principal assumptions: (1) the gas being observed is moving in a circular orbit around the Galactic center, and (2) that small deviant motions of the gas about the true  $v_{\text{LSR}}$  all map to a narrow set of heliocentric distances. Regions of the Galaxy that violate these assumptions must be excluded from kinematic distance calculations, as derived distances may be off by a factor of two or more (see Reid et al. 2009b). We label these regions “kinematic avoidance zones.”

Toward the center of the Milky Way, the long Galactic bar at  $R_{\text{gal}} \lesssim 3\text{--}4 \text{ kpc}$  (Fux 1999; Rodriguez-Fernandez & Combes 2008; Reid et al. 2014) introduces strong radial

<sup>14</sup> There has been some controversy in the literature about the values of the solar peculiar motion and the counterrotation ( $\overline{V}_s < 0$ ) of HMSFRs (see Reid et al. 2009b; McMillan & Binney 2010; Schönrich et al. 2010; Honma et al. 2012). While both  $\Theta_\odot + V_\odot$  and  $V_\odot - \overline{V}_s$  are highly constrained, the individual parameters are highly correlated (Reid et al. 2014).



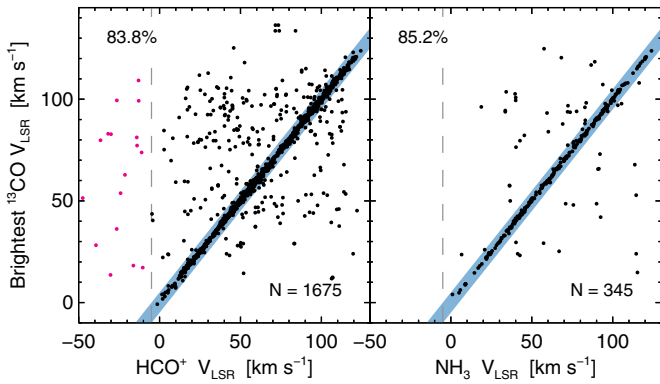
**Figure 1.** Galactic kinematics for distance determination. Left: contours mark lines of constant projected  $v_{\text{LSR}}$  based on the (flat) rotation curve of model A5 from Reid et al. (2014). Regions for which kinematic distances are not computed (kinematic avoidance zones) are shown in gray with dashed borders. The Sun (sun symbol) and Galactic center (star) are marked for reference. Cyan contours mark  $d_\odot$  in 5 kpc intervals, and the red circle identifies  $R_{\text{gal}} = 4 \text{ kpc}$ , inside which the Galactic bar dominates the kinematics. Right: the magnitude of the derivative of  $v_{\text{LSR}}$  with respect to  $d_\odot$  across the Galactic plane (grayscale). Magenta hashing identifies regions where  $|dv_{\text{LSR}}/d(d_\odot)| \leq 5 \text{ km s}^{-1} \text{ kpc}^{-1}$  (i.e., a projected peculiar motion of  $5 \text{ km s}^{-1}$  will yield a  $\geq 1 \text{ kpc}$  distance error). Kinematic avoidance zones are marked by gray hashing, and yellow contours mark constant values of  $|dv_{\text{LSR}}/d(d_\odot)|$ .

streaming motions of the gas, and care must be exercised to utilize kinematic information only for sources outside the influence of these orbits. This amounts to excluding much of  $|\ell| \lesssim 20^\circ$ . To quantify this exclusion, EB13 (their Figure 5) defined two regions in the longitude–velocity diagram for which  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$  is not computed. For  $7:5 \leq \ell \leq 21^\circ$ ,<sup>15</sup> the upper exclusion region is bounded by  $v_{\text{LSR}} = (3.33 \text{ km s}^{-1}) \times \ell^\circ + 15 \text{ km s}^{-1}$ , and includes the higher-velocity gas inside the bar. The lower region excludes the 3 kpc expanding arm, and is bounded by  $v_{\text{LSR}} = (2.22 \text{ km s}^{-1}) \times \ell^\circ - 16.7 \text{ km s}^{-1}$ . The effective regions in Galactocentric coordinates for these zones, computed for the rotation curve described above, are shown in Figure 1 (left) as gray areas toward the Galactic center. The allowed velocities in this longitude region correspond to the Scutum–Centaurus arm feature (see Dame & Thaddeus 2011).

Toward the Galactic anti-center ( $\ell = 180^\circ$ ), orbital circular motion is nearly perpendicular to the line of sight, making the projection of  $v_{\text{LSR}}$  onto the rotation curve highly subject to small peculiar motions of gas. We therefore define the region  $\ell = 180^\circ \pm 20^\circ$  as an additional kinematic avoidance zone (see Figure 1), and do not compute kinematic distance likelihoods for objects in this area.

Finally, the portions of the Galaxy nearly parallel to the Sun’s motion (i.e.,  $\ell = 90^\circ$  and  $270^\circ$ ) present fairly flat  $v_{\text{LSR}}(d_\odot)$  over the span of some 5 kpc. Peculiar motions are very likely to produce wildly inaccurate kinematic distances, as is the case for object G075.76+00.33 (see Section 5.1). The flatness of the projected rotation curve (i.e.,  $dv_{\text{LSR}}/d(d_\odot) \leq 5 \text{ km s}^{-1} \text{ kpc}^{-1}$ ;

<sup>15</sup> The lower longitude limit on this region is the lowest  $\ell$  for which BGPS spectroscopic observations were made. Significant blending of structures in the  $\ell - v$  diagram at lower longitude make use of any kinematic information challenging at best.



**Figure 2.** Comparison of the  $v_{\text{LSR}}$  of the brightest  $^{13}\text{CO}$  emission feature from the GRS data toward a BGPS object compared to the (left)  $\text{HCO}^+(3-2)$  velocities of Shirley et al. (2013) and (right)  $\text{NH}_3(1,1)$  velocities of Dunham et al. (2011b). The vertical gray dashed line in each panel represents the lower  $v_{\text{LSR}}$  bound of GRS observations. Objects depicted in pink ( $v_{\text{LSR}} < 5 \text{ km s}^{-1}$ ) would, therefore, not be detectable in the GRS data. The light blue region marks the  $|\Delta v_{\text{LSR}}| \leq 5 \text{ km s}^{-1}$  used to compute the matching statistic shown at the top of each panel.

magenta hashing in Figure 1 (right)) is not symmetric about  $\ell = 90^\circ$  and  $270^\circ$ , but instead follows the tangent circle toward the Galactic center. The expected virial motion within a HMSFR is  $\approx 5 \text{ km s}^{-1}$ , and distance errors of  $\gtrsim 1 \text{ kpc}$  due to virial motions are not desirable. At  $\ell \lesssim 70^\circ$ , this region around the tangent point becomes narrower and therefore acceptable. The rotation curve derivative is steep enough for  $\ell \gtrsim 100^\circ$  that the kinematic avoidance zone should only be defined over  $70^\circ \leq \ell \leq 100^\circ$ . In this range, however, the small  $dv_{\text{LSR}}/d(d_\odot)$  occurs only at small heliocentric distance (and consequently  $v_{\text{LSR}} \sim 0 \text{ km s}^{-1}$ ). Examination of the  $\ell - v$  diagram reveals that the Cygnus X region ( $\ell \approx 80^\circ$ ) extends from about  $-15 \text{ km s}^{-1} \leq v_{\text{LSR}} \leq 20 \text{ km s}^{-1}$  (falling in the avoidance zone), with the better-defined Perseus and Outer arms visible at more negative  $v_{\text{LSR}}$ . To fully encompass the Cygnus X region as a kinematic avoidance zone, we therefore limit the kinematic avoidance zone here to  $v_{\text{LSR}} \geq -15 \text{ km s}^{-1}$ , and the resulting region in Galactocentric coordinates is visible in Figure 1. The corresponding zone near  $\ell = 270^\circ$  would need to be defined based on gas kinematics near the Carina tangent.

#### 4. MOLECULAR CLOUD CLUMP $^{13}\text{CO}$ VELOCITY EXTRACTION

##### 4.1. Comparison of GRS $^{13}\text{CO}$ Spectra with Dense Gas Tracers

The most straightforward method for using CO data to assign  $v_{\text{LSR}}$  to molecular cloud clumps is to choose the brightest emission peak along the line of sight. This has been used by various groups (see Russeil et al. 2011; Eden et al. 2012) in lieu of observations of molecular line transitions with higher  $n_{\text{eff}}$  that trace the denser gas associated with molecular cloud clumps. While this process generally selects the dense material seen in dust continuum emission, areas of warmer diffuse gas or multiple molecular cloud clumps along a line of sight may lead to an incorrect  $v_{\text{LSR}}$  assignment.

To illustrate the challenges of simply choosing the brightest  $^{13}\text{CO}$  peak for BGPS objects, Figure 2 compares that velocity with available  $\text{HCO}^+(3-2)$  (S13) and  $\text{NH}_3(1,1)$  (Dunham et al. 2011b) observations. The extracted  $v_{\text{LSR}}$  from  $^{13}\text{CO}$  agrees with the dense-gas value to within  $5 \text{ km s}^{-1} \sim 85\%$  of the time. For the purposes of this comparison, the GRS data cubes were

averaged over the angular extent of the BGPS source to generate a composite  $^{13}\text{CO}$  spectrum; S13 show that using the single GRS spectrum at the location of peak  $\lambda = 1.1 \text{ mm}$  flux density yields an agreement rate  $\sim 81\%$ .

The GRS was designed to study the chemistry and kinematics of the large peak in molecular gas density about halfway between the Sun and the Galactic center, the so-called 5 kpc ring (Clemens et al. 1988). Because there is comparatively little CO emission on the far side of the Galaxy beyond the solar circle behind this feature in the first quadrant (see Clemens et al. 1988; Dame et al. 2001), the GRS did not observe negative  $v_{\text{LSR}}$ . Continuum surveys such as the BGPS detect the optically thin dust within molecular cloud clumps, and so are sensitive to objects on the far side of the Galaxy beyond the solar circle, embedded in what CO is present at those locations (see Section 6.1.3). Several such objects are illustrated in Figure 2 as pink dots with  $\text{HCO}^+(3-2)$   $v_{\text{LSR}} < -5 \text{ km s}^{-1}$ . Of the 1846 BGPS objects with detected dense gas  $v_{\text{LSR}}$  in the GRS overlap region, however, only 18 are beyond the solar circle.

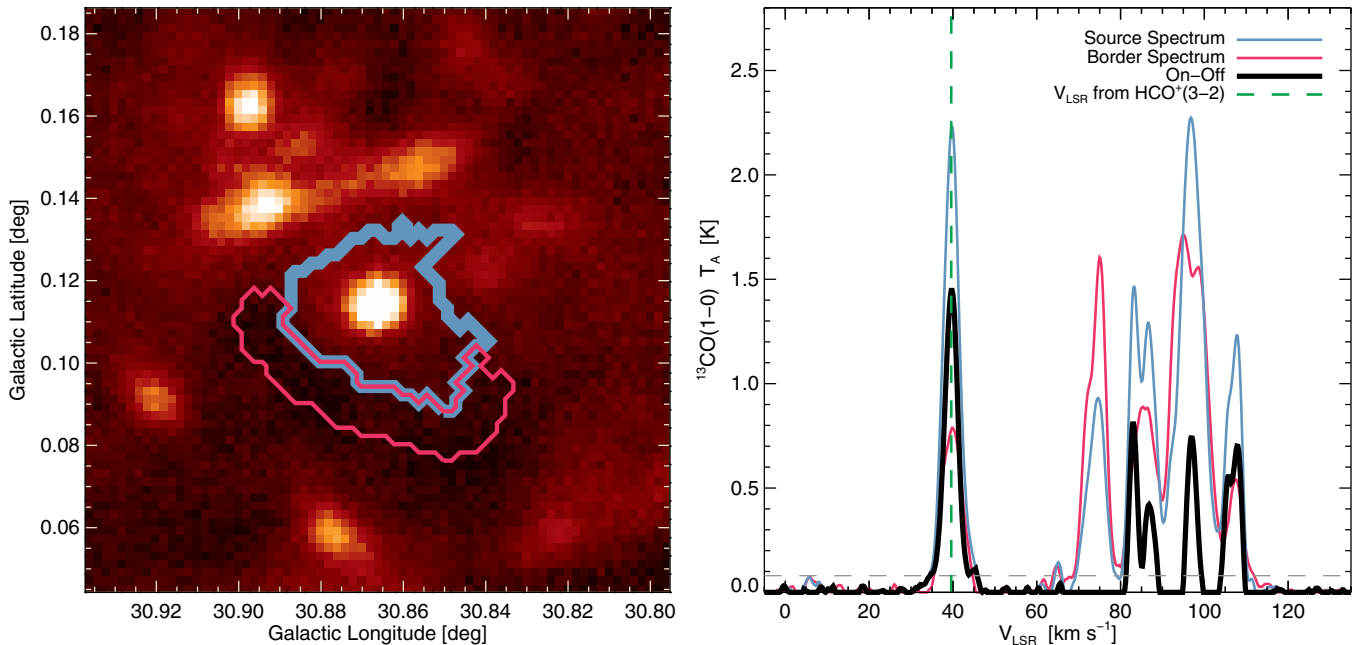
##### 4.2. Morphological Spectrum Extraction Technique

###### 4.2.1. General Description

While the  $\sim 85\%$  velocity matching rate for the brightest  $^{13}\text{CO}$  feature along a line of sight is encouraging, the low  $n_{\text{eff}}$  of  $^{13}\text{CO}(1-0)$  means that observed emission is not uniquely tied to molecular cloud clumps. A large patch of warm, diffuse gas or changes in the excitation temperature may produce an emission feature that outshines all others along a given line of sight. To mitigate this effect, we developed a technique for using millimeter continuum data as prior information for extracting a velocity spectrum for the molecular cloud structure of interest. The  $46''$  resolution of the GRS is similar enough to that of the BGPS that we may assume both surveys are sensitive to similar structures.

To leverage the continuum data's prior information, we use a morphological spectrum extraction technique to isolate the contribution to the  $^{13}\text{CO}$  emission from a particular molecular cloud structure from its more diffuse envelope (Rosolowsky et al. 2010b). The extracted spectrum is computed as  $T_{\text{source}}(v) = T_{\text{on}}(v) - T_{\text{off}}(v)$ , where  $T_{\text{on}}(v)$  is the average of the GRS spectra at the location of BGPS pixels within the Bolocat label masks (see Rosolowsky et al. 2010a), weighted by BGPS flux density. Each BGPS pixel is assigned the  $^{13}\text{CO}$  spectrum of the nearest  $(\ell, b)$  GRS pixel; the same  $^{13}\text{CO}$  spectrum ( $22''$  pixels) may therefore be assigned to more than one BGPS pixel ( $7/2$ ). The off-source spectrum is computed as the unweighted average of the GRS spectra assigned to BGPS pixels within one GRS resolution element of the Bolocat outline, *excluding* pixels assigned to another Bolocat object. This exclusion avoids subtracting relevant emission associated with the dense gas of neighboring catalog objects, leaving  $T_{\text{off}}(v)$  contributions only from lines of sight penetrating the more diffuse CO envelope. The  $46''$  width of the off region corresponds to  $\sim 1.1 \text{ pc}$  at  $d_\odot = 5 \text{ kpc}$ , well within the size of GRS cataloged clouds (Roman-Duval et al. 2010). A small handful of BGPS objects ( $N \sim 10$ ) are “landlocked” (i.e., completely surrounded by other Bolocat sources), and no  $T_{\text{off}}(v)$  may be computed. For these sources,  $T_{\text{on}}(v)$  is used in place of  $T_{\text{source}}(v)$ , and a flag is set.

We emphasize that the angular transfer functions of ground-based (sub)millimeter observations make this type of extraction possible, as atmospheric subtraction algorithms necessarily



**Figure 3.** Demonstration of the morphological spectrum extraction technique for example source G030.866+00.115. Left: the image depicts the BGPS V2 flux density on a linear scale from  $-0.10$  to  $1.40 \text{ Jy beam}^{-1}$ . The blue contour is the Bolocat source outline, as described in Rosolowsky et al. (2010a), and the pink contour depicts the off-source region, one GRS resolution element in width, excluding pixels associated with the continuum sources above and to the lower right. Right: the filtered  $^{13}\text{CO}(1-0)$  spectra for this object. Blue and pink spectra represent the  $T_{\text{on}}(v)$  and  $T_{\text{off}}(v)$ , respectively, while black represents  $T_{\text{source}}(v)$ , used to estimate the  $v_{\text{LSR}}$  of the denser gas associated with the molecular cloud clump. The vertical green dashed line marks the  $\text{HCO}^+(3-2)$   $v_{\text{LSR}}$  from S13.

remove large-scale diffuse dust emission, leaving well-defined sources to delineate on- and off-source regions. Application of this technique to space-based data (e.g., Hi-GAL) would likely require filtering of the data to remove Galactic cirrus emission, as the cirrus on scales larger than the BGPS sensitivity can contribute a factor of two to the derived dust column density for clump-scale objects (Battersby et al. 2011). To verify the correlation between BGPS-detected sources and physically meaningful structures, we compared the maximum angular scales recoverable by the BGPS against spectral-line mapping surveys, which recover emission at all spatial scales larger than the beam size. The  $\text{H}_2\text{O}$  Southern Galactic Plane Survey (HOPS; Walsh et al. 2011) observed several molecular transition lines around 24 GHz, mapping nearly 700 molecular cloud clumps in  $\text{NH}_3(1,1)$  (Purcell et al. 2012). Comparing the maximum solid angle the BGPS recovers ( $19.6 \text{ arcmin}^{-2}$ ;  $\theta_{\text{max}} \approx 300''$ ; G13) with the cloud solid angles from Purcell et al. shows that nearly 80% of the HOPS clumps would be fully recovered by the BGPS, with the remaining being somewhat truncated by the angular transfer function. Therefore, ground-based dust-continuum data selects nearly all of the region visible in  $\text{NH}_3(1,1)$  emission brighter than  $\approx 0.2 \text{ K}$  (the average rms noise temperature of the HOPS  $\text{NH}_3(1,1)$  spectra; Purcell et al. 2012), and is appropriate for using as a prior for extracting  $^{13}\text{CO}$  spectra for molecular cloud clumps.

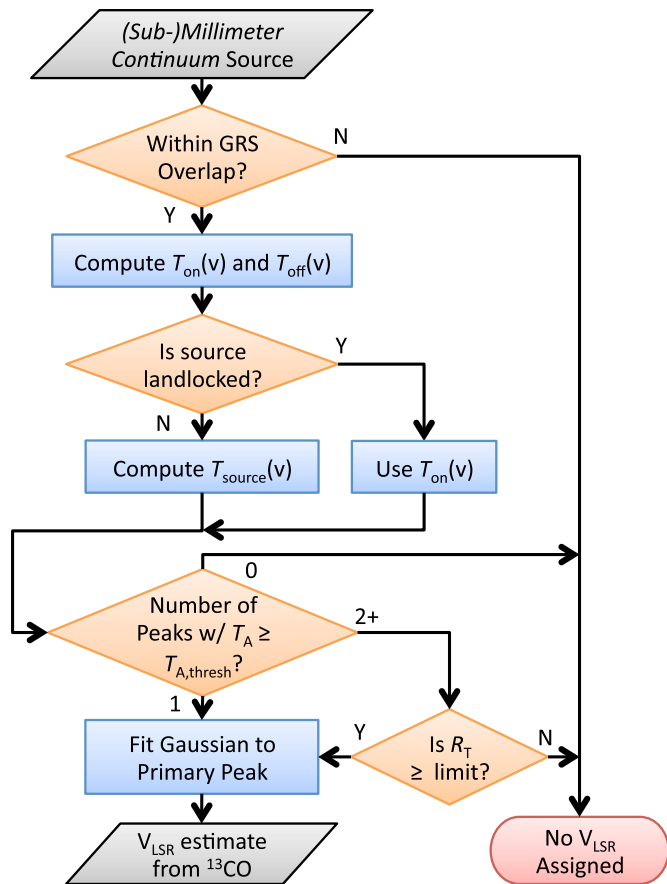
The averaged  $^{13}\text{CO}$  spectra were filtered using a kernel constructed to match the typical widths of  $^{13}\text{CO}$  features ( $\text{FWHM} \approx 2-5 \text{ km s}^{-1}$ ) to conserve the relative widths and heights of spectral line features while significantly reducing the noise. This allows for the detection of spectral features in  $T_{\text{source}}(v)$  with peak intensity less than the nominal GRS noise level. The resulting spectrum is clipped to contain only positive (emission) values. Figure 3 illustrates the morphological spectrum extraction process for object G030.866+00.115. In the left panel, the blue contour delineates the source, with the weighted-average

GRS spectrum shown in blue in the right panel. The “Off” region (pink outline) excludes BGPS pixels associated with the sources (pink) above and to the right, and the averaged GRS spectrum from this region is shown in pink in the right panel. While there is a general correspondence between the peaks of  $T_{\text{on}}(v)$  (blue) and  $T_{\text{off}}(v)$  (pink), the complicated nature of the emission in this region makes identifying a single or dominant peak from  $T_{\text{on}}(v)$  challenging at best.

#### 4.2.2. Extracting $v_{\text{LSR}}$ from $T_{\text{source}}(v)$

This new technique effectively yields a pointed catalog of “position-switched”  $^{13}\text{CO}(1-0)$  spectra, where the “reference position” is carefully chosen to include the diffuse envelope surrounding the molecular cloud clump. Like any such catalog, detection thresholds and flagging are required to produce a reliable set of  $v_{\text{LSR}}$  for kinematic distance computation. Two parameters control the quality and quantity of extracted spectra: the minimum  $T_A$  threshold for peaks in  $T_{\text{source}}(v)$  ( $T_{A,\text{thresh}}$ ), and the degree to which a single  $v_{\text{LSR}}$  peak dominates the final spectrum. We parameterize this latter quantity as the minimum ratio of the  $T_A$  of the primary peak to that of the secondary peak (when present), or  $R_T = T_{A(1)}/T_{A(2)}$ . For the example source in Figure 3, the primary peak in  $T_{\text{source}}(v)$  is near  $v_{\text{LSR}} = 40 \text{ km s}^{-1}$ , and the secondary peak near  $v_{\text{LSR}} = 83 \text{ km s}^{-1}$ , with  $R_T \approx 1.8$ .

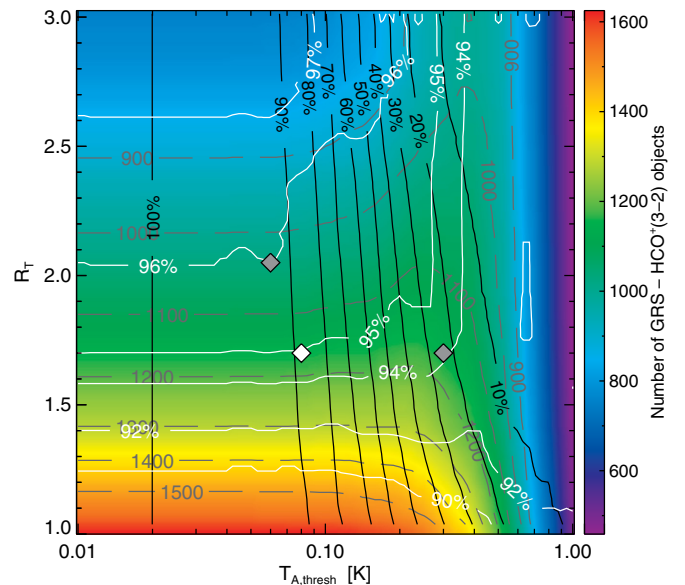
The algorithm for  $v_{\text{LSR}}$  extraction from  $T_{\text{source}}(v)$  is illustrated in the flow chart of Figure 4. For objects within the GRS coverage region,  $T_{\text{source}}(v)$  is examined for any peak above  $T_{A,\text{thresh}}$ . Next, the number of contiguous regions above  $T_{A,\text{thresh}}$  is counted; for two or more independent peaks,  $R_T$  is computed. There are three points in the process where a null  $v_{\text{LSR}}$  may be returned: (1) the molecular cloud structure is outside the limits of the GRS, (2) there are no peaks in  $T_{\text{source}}(v)$  above  $T_{A,\text{thresh}}$ , and (3) for multiply peaked spectra there is no clearly dominant peak (i.e.,  $R_T < \text{limit}$ ). For the remaining multiply peaked and all single-peaked spectra, a Gaussian is fit to the primary peak.



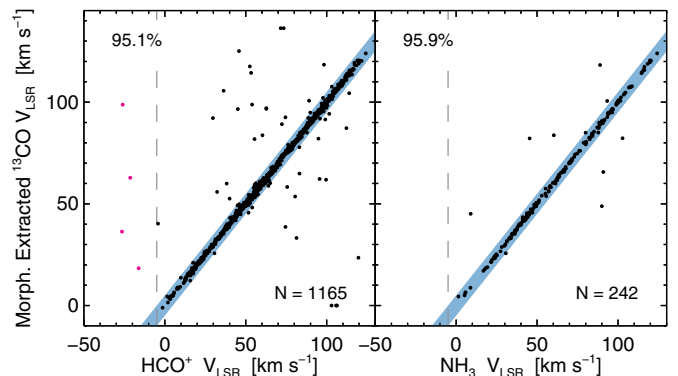
**Figure 4.** Flow chart for the assignment of  $v_{\text{LSR}}$  from the morphologically extracted  $^{13}\text{CO}$  spectrum for a given molecular cloud structure.

To determine the optimal values of these parameters for  $v_{\text{LSR}}$  extraction, we computed the following as functions of the two tunable parameters: (1) the number of objects ( $N_{\text{comp}}$ ) having both a valid  $^{13}\text{CO}$   $v_{\text{LSR}}$  and  $\text{HCO}^+$   $v_{\text{LSR}}$  from S13, (2) the fraction of  $N_{\text{comp}}$  whose  $^{13}\text{CO}$  and  $\text{HCO}^+$  velocities agree to  $\leq 5 \text{ km s}^{-1}$  (the approximate virial motion within HMSFRs), and (3) the fraction of  $N_{\text{comp}}$  whose  $^{13}\text{CO}$  spectra contain multiple peaks above  $T_{A,\text{thresh}}$ . Based on the topography of the parameter space, as shown in Figure 5, we chose values of  $T_{A,\text{thresh}} = 0.08 \text{ K}$ , and  $R_T = 1.70$  (white diamond) to best balance the number of BGPS V2 objects in the GRS– $\text{HCO}^+$  overlap with the fraction of those sources whose  $v_{\text{LSR}}$  agree to better than  $5 \text{ km s}^{-1}$ .

We note that at the selected values for  $T_{A,\text{thresh}}$  and  $R_T$ , nearly 90% of the morphologically extracted spectra are multiply peaked (black contours), yet the velocity-agreement fraction is virtually unchanged down to  $\sim 30\%$  multiplicity. Two discarded values are shown as gray diamonds for comparison. The diamond at  $(T_{A,\text{thresh}}, R_T) = (0.30 \text{ K}, 1.7)$  has a threshold at approximately twice the rms noise of the GRS data cubes, and follows the same  $N_{\text{comp}} \approx 1150$  contour the white diamond is on. The matching rate at this point, however, is 94%. A point with a higher matching percentage (96%) is the gray diamond at  $(T_{A,\text{thresh}}, R_T) = (0.06 \text{ K}, 2.05)$ . While this is one percent better than the white diamond, there are  $\approx 100$  fewer objects for which a morphological  $v_{\text{LSR}}$  may be extracted from the  $^{13}\text{CO}$  data. The white diamond was ultimately a choice made to balance  $N_{\text{comp}}$  with the matching percentage.



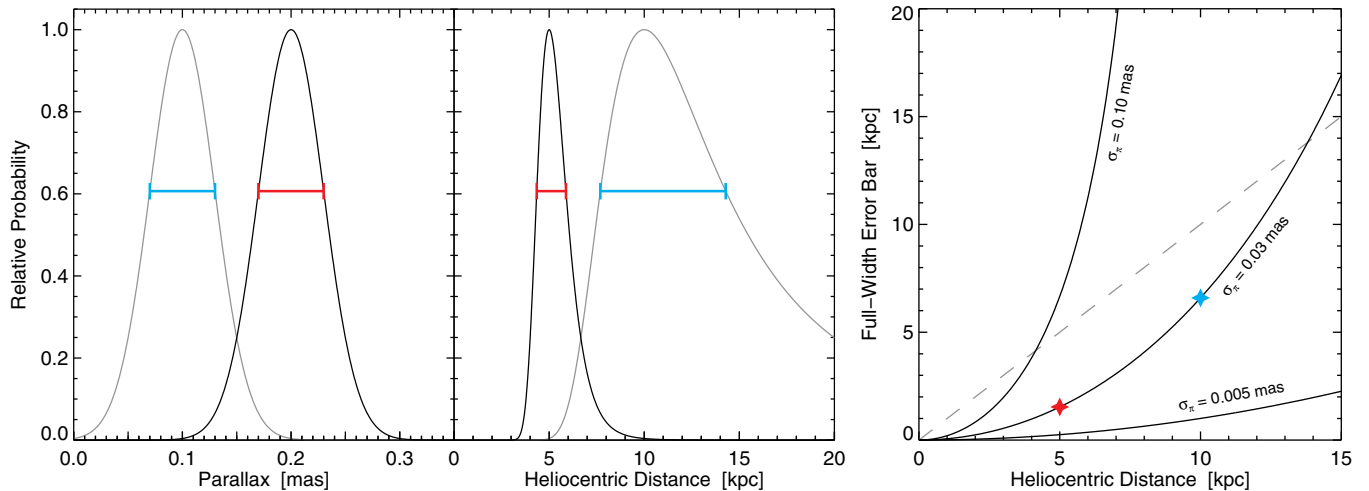
**Figure 5.** Optimization of the  $T_{A,\text{thresh}}$  for detection and  $R_T$  limit for multiply peaked spectra. Background color scale and gray dashed contours denote the number of BGPS V2 objects that possess both  $\text{HCO}^+(3-2)$  and morphologically extracted  $^{13}\text{CO}$  velocities ( $N_{\text{comp}}$ ). White contours mark the fraction of  $N_{\text{comp}}$  whose  $v_{\text{LSR}}$  from the two sources agree to better than  $5 \text{ km s}^{-1}$ . Black contours show the fraction of  $N_{\text{comp}}$  whose morphologically extracted spectrum contains more than one peak above  $T_{A,\text{thresh}}$ . The white diamond indicates the chosen values of  $T_{A,\text{thresh}}$  and  $R_T$  while the gray diamonds are discarded choices (see text).



**Figure 6.** Morphologically extracted  $^{13}\text{CO}$  spectrum  $v_{\text{LSR}}$  comparison with dense-gas tracers. Colors and lines as in Figure 2. Left: velocity comparison for the  $N = 1165$  BGPS V2 objects with a valid morphologically extracted  $^{13}\text{CO}$  spectrum and  $\text{HCO}^+(3-2)$   $v_{\text{LSR}}$  from Shirley et al. (2013). Right: velocity comparison for the  $N = 242$  BGPS V2 objects with a valid morphologically extracted  $^{13}\text{CO}$  spectrum and  $\text{NH}_3(1,1)$   $v_{\text{LSR}}$  from Dunham et al. (2011b).

#### 4.2.3. Comparison with $\text{HCO}^+(3-2)$

With the optimized parameter values for  $T_{A,\text{thresh}}$  and  $R_T$ , there are 1165 objects that possess both an  $\text{HCO}^+(3-2)$  detection and a valid morphologically extracted spectrum from the GRS  $^{13}\text{CO}$  data. The velocity comparison is shown in Figure 6, which echoes Figure 2, except that the correspondence has grown to  $\approx 95\%$ . As with Figure 2, pink circles at the left of the plot mark continuum-detected molecular cloud structures beyond the solar circle ( $R_{\text{gal}} > R_0$ ) that also have detectable  $^{13}\text{CO}$  emission in the foreground. It is likely these objects would have the correct  $v_{\text{LSR}}$  assigned (to 95% confidence) if the GRS had extended velocity coverage to negative  $v_{\text{LSR}}$ . The histogram of



**Figure 7.** Example of  $\text{DPDF}_{\text{px}}$  construction from trigonometric parallax measurements. Left: Gaussians in parallax with  $\mu_{\text{px}} = 0.20$  mas ( $\mu_{\text{px}} = 0.10$  mas) and  $\sigma_{\text{px}} = 0.03$  mas are plotted in black (gray). The full-width error bars (i.e.,  $\mu \pm \sigma$ ) are shown in red (cyan). Middle: interpolation of the parallax functions onto a linear distance scale, where  $d_{\odot} = 1/\pi_s$ . The corresponding full-width error bars are shown in red (cyan). Right: size of the full-width distance error bar as a function of parallax measurement uncertainty and heliocentric distance. The curves shown represent the lower bound on the uncertainty ( $\sigma_{\pi} = 0.005$  mas), a typical value ( $\sigma_{\pi} = 0.03$  mas), and a poorly constrained parallax measurement ( $\sigma_{\pi} = 0.10$  mas). The red and cyan stars identify the examples at left, and the gray dashed line marks the 1:1 relationship.

the velocity difference for all points in Figure 6 is well fit by a Gaussian with FWHM  $\approx 1.1$  km s $^{-1}$ , the channel width of the HCO $^{+}$  spectra (S13).

#### 4.3. The $^{13}\text{CO}$ Kinematic Distance Likelihood Function

Using the optimized parameters for computing a morphologically extracted spectrum from the GRS  $^{13}\text{CO}$  data, a kinematic distance likelihood function may be computed for BGPS sources with a valid  $^{13}\text{CO}$  spectrum. To maximize the use of kinematic information generated from this technique, we use the  $T_{\text{source}}(v)$  directly to compute  $\mathcal{L}(v_{\text{LSR}}, l, b; d_{\odot})$ , following the process described in EB13. Examination of Figure 3, however, reveals many small features in the extracted spectrum below the  $T_A$  threshold that are artifacts of the technique in addition to the stronger secondary peaks at  $80 \text{ km s}^{-1} \leq v_{\text{LSR}} \leq 110 \text{ km s}^{-1}$ . To eliminate these extraneous bumps, which would appear as spurious small peaks in the DPDF, we mask  $T_{\text{source}}(v)$  to include only the region  $v_{\text{LSR}} \pm 3\sigma$  from the Gaussian fit to the primary peak before multiplying it by the rotation curve probability density function. For the example source in Figure 3, this preserves the main peak, while eliminating probability associated with the secondary peaks.

## 5. CATALOG-BASED PRIOR DPDFs

As large-scale surveys of Galactic star formation have been published over the last several years, they are often accompanied by catalogs of distance estimates. While some methods are more robust and/or accurate than others, these catalogs offer additional anchor points for use with the DPDF formalism. With the relatively small present suite of prior DPDFs, the use of literature catalogs can expand the types of distance methods available for use with detected molecular cloud structures. In this section, we describe two literature catalogs of objects associated with sites of massive star formation (trigonometric parallax measurements of masers in Section 5.1, and robust KDA resolutions for H II regions in Section 5.2), and define a

means for associating distances from those catalogs with BGPS molecular cloud structures (Section 5.3).

### 5.1. Trigonometric Parallax Measurements

The BeSSeL survey has been conducting very long baseline interferometry (VLBI) observations of CH $_3$ OH and H $_2$ O masers associated with HMSFRs for the past several years (see Brunthaler et al. 2011; Reid et al. 2014 and references therein). The geometric distances returned by trigonometric parallaxes depend upon neither the choice of Galactic rotation curve nor other assumptions about the structure of the Galactic plane, offering an absolute distance benchmark. Parallax distances are especially useful within the kinematic avoidance zones described in Section 3.2.2. For instance, the  $\approx 16$  km s $^{-1}$  counterrotation of HMSFR G075.76+00.33 yields a kinematic distance of  $\approx 5.7$  kpc, apparently in the Perseus arm, but its parallax distance of  $3.5 \pm 0.3$  kpc places it in the Local arm (Xu et al. 2013).

To construct the prior  $\text{DPDF}_{\text{px}}$  for trigonometric parallax measurements ( $\pi_s$ ) associated with molecular cloud structures, we utilize the latest list of robust parallaxes from BeSSeL and VERA (Reid et al. 2014, Table 1), which contains measurements for 103 sources associated with HMSFRs throughout  $-12^{\circ} \leq \ell \leq 241^{\circ}$ . Source  $(\ell, b, v_{\text{LSR}})$  from the BGPS V2.1 catalog are compared with the association volumes for each maser in the parallax table, as discussed above. Because of the “gold standard” nature of trigonometric parallax measurements, there is no need to dilute  $\text{DPDF}_{\text{px}}$  to allow nonzero probability far from the measured distance. The prior DPDF is therefore created by constructing a Gaussian in parallax space using the value and uncertainty from Reid et al. (2014, Table 1) as the centroid and standard deviation, respectively. Next, that array is interpolated onto a linear distance array, where  $d_{\odot} = 1/\pi_s$ , with points spaced every 0.02 kpc to create a DPDF reflecting the asymmetric nature of parallax distance uncertainties. Example  $\text{DPDF}_{\text{px}}$  for  $\pi_s = 0.20 \pm 0.03$  (black) and  $\pi_s = 0.10 \pm 0.03$  (gray) are shown in Figure 7. The typical parallax uncertainty in the present sample is  $\approx 0.03$  mas, which translates to a distance uncertainty of  $^{+0.88}_{-0.65}$  kpc at  $d_{\odot} = 5$  kpc, or a



**Table 2**  
H II Region KDA Resolutions from the HRDS

Object Name	$\ell$ (°)	$b$ (°)	$v_{\text{LSR}}$ (km s <sup>-1</sup> )	$d_{\odot}$ (kpc)	KDA <sup>a</sup> Resol.	QF <sup>b</sup>	Ref.
U23.20+0.00a	23.200	0.000	22.4	14.00	F	B	1
U23.43-0.21	23.430	-0.210	101.1	6.00	N	A	1
U23.96+0.15	23.960	0.150	78.9	5.00	N	A	1
C24.30-0.15a	24.300	-0.150	55.5	11.70	F	A	1
C27.49+0.19	27.490	0.190	34.0	12.80	F	A	1
G032.272-0.226	32.272	-0.226	21.5	12.80	F	A	2
C33.42+0.00	33.420	0.000	76.5	9.40	F	A	1
G038.738-0.140	38.738	-0.140	60.9	9.20	F	B	2
G046.948+0.374	46.948	0.374	-44.4	16.20	F	A	2
G047.094+0.492	47.094	0.492	-54.5	17.50	F	...	3

**Notes.**

<sup>a</sup> N = near, F = far, T = tangent.

<sup>b</sup> Quality factor of the KDA resolution; see Anderson et al. (2012) for more details. Bania et al. (2012) does not assign a QF.

**References.** (1) Anderson & Bania 2009; (2) Anderson et al. 2012; (3) Bania et al. 2012.

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

full-width error bar (see Section 6.3) of 1.53 kpc. The right panel of Figure 7 illustrates the dependence of the distance uncertainty on heliocentric distance for three values of parallax uncertainty, showing the rapid increase in the fractional parallax uncertainty as a function of  $d_{\odot}$ .

## 5.2. H II Region KDA Resolutions

Galactic H II regions associated with HMSFRs offer an additional opportunity for generating prior DPDFs for molecular cloud structures. The HRDS (Bania et al. 2010, 2012) used the GBT and Arecibo Observatory to search for radio recombination lines indicative of ionized gas based on a candidate list compiled principally from mid-infrared and radio continuum data and catalogs. The resulting collection of previously known and newly discovered H II regions (448 from GBT and 37 from Arecibo) associated with massive star formation provide a sizable catalog from which to assign KDA resolutions for molecular cloud structures (Anderson & Bania 2009; Anderson et al. 2012; Bania et al. 2012).

KDA resolutions for HRDS sources rely upon the H I self-absorption (HISA) and H I emission/absorption (HIE/A) methods (see Anderson & Bania 2009; Roman-Duval et al. 2009). In short, both methods make use of cold H I in molecular clouds absorbing 21 cm emission from a backlighting source. A stable population of neutral atomic hydrogen is maintained even in the cold, dense regions of molecular cloud clumps through an equilibrium between H<sub>2</sub> formation and destruction by cosmic rays (Goldsmith et al. 2007). For HISA, cold gas at the near kinematic distance absorbs 21 cm emission from warm gas at the same  $v_{\text{LSR}}$  at the far kinematic distance. In the HIE/A method, 21 cm continuum emission from the H II region is examined for absorption features; absorption at  $|v_{\text{H II}}| < |v_{\text{absorption}}| < |v_{\text{tan}}|$  places the H II region at the far kinematic distance, and absorption only at  $|v_{\text{absorption}}| < |v_{\text{H II}}|$  places it at the near.

The set of 441 H II regions with strong KDA resolution (quality factor A or B; see Anderson et al. 2012) from the various HRDS publications are gathered in Table 2 as a reference for computing DPDF<sub>hrds</sub>. Since the HRDS Galactic longitude range is  $15^{\circ} \leq \ell \leq 67^{\circ}$ , all objects with positive  $v_{\text{LSR}}$  must

be assigned a KDA resolution (objects with  $v_{\text{LSR}} \leq 0$  km s<sup>-1</sup> are unambiguously beyond the solar circle). Because this prior is based on the physically relevant tangent distance ( $d_{\text{tan}}$ ), we construct DPDF<sub>hrds</sub> as a step function removing probability on the opposite side of  $d_{\text{tan}}$ . An analysis of posterior DPDFs suggests that objects within 1 kpc of  $d_{\text{tan}}$  should be given a “tangent” KDA resolution (Section 6.3), so a strict step function at  $d_{\text{tan}}$  would improperly bias the prior for these objects. We therefore model DPDF<sub>hrds</sub> with an error function possessing a rolloff width of 1 kpc for objects in Table 2 with a “N” or “F” resolution, and as a Gaussian centered on  $d_{\text{tan}}$  with  $\sigma = 1$  kpc for objects with a “T” resolution.

## 5.3. Associating Molecular Cloud Structures with Literature Catalog Objects

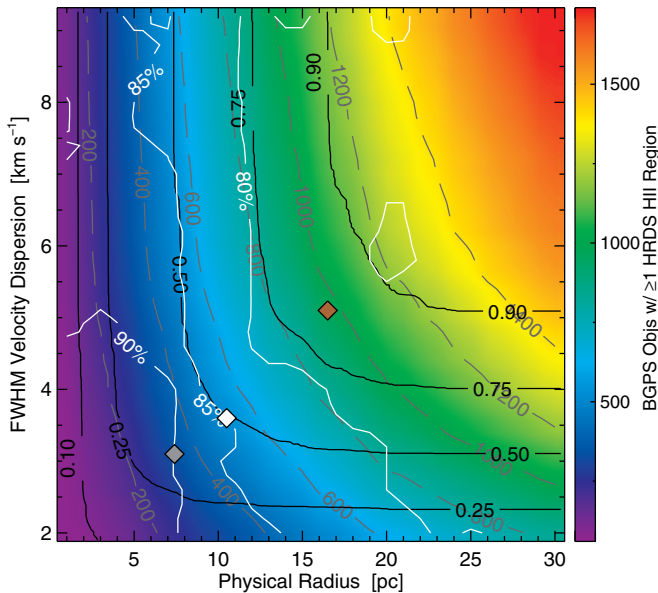
### 5.3.1. Definitions

Care must be taken when applying distance information from literature catalogs to create a prior DPDF. The clumpy nature of the interstellar medium and even within GMCs requires defining a volume around literature catalog objects for association with molecular cloud structures. Given that the three observational dimensions are plane of sky and velocity, we define a cylindrical association volume in coordinate-velocity ( $\ell, b, v_{\text{LSR}}$ ) space whose symmetry axis lies along the line of sight to the catalog object. Proper physical scaling of these cylinders should be based on typical coherent structures in the interstellar medium, namely GMCs. We use here the collection of GMCs cataloged by the GRS (Rathborne et al. 2009), which represent the coherent envelopes within which several or many molecular cloud clumps may reside. The physical properties of these GMCs (as computed by Roman-Duval et al. 2010) provide reasonable baselines for fixing the association volume.

The likelihood of a single molecular cloud structure lying within the association volume of more than one catalog object (e.g., H II regions) is nonzero. Distance assignments or KDA resolutions for multiple catalog entries may conflict, so a mechanism for combining information from multiple objects is required. Because a catalog object lying closest to a molecular cloud structure in coordinate-velocity space is more likely to assign the correct distance, we combine prior information from multiple catalog entries weighted by  $1/\xi$ , where the non-dimensional distance parameter is computed as

$$\xi = \left[ \left( \frac{\Delta\theta}{\theta_a} \right)^2 + \left( \frac{\Delta v_{\text{LSR}}}{\Delta v_a} \right)^4 \right]^{1/2}. \quad (2)$$

The quantities  $\Delta\theta$  and  $\Delta v_{\text{LSR}}$  are the angular separation on the sky and velocity separation, respectively, between the continuum-identified source and the catalog object, and  $\theta_a$  and  $\Delta v_a$  are the optimized association volume angular radius and velocity extent, respectively (determined below). The physical association volume is based on the physical radius of GRS clouds ( $R_c$ ), so  $\theta_a$  must be computed individually for each catalog object as  $R_c/d_{\odot}$ . The turbulent structure function dictates that  $\Delta v \sim R^{1/2}$  (Heyer et al. 2009), giving rise to the quartic velocity term in Equation (2). Since association is not considered outside the volume defined by  $\theta_a$  and  $\Delta v_a$ , the maximum possible value of  $\xi$  is  $\sqrt{2}$ . In the case of competing KDA resolutions for catalog objects with similar  $\xi$ , the resulting prior DPDF will be relatively unconstrained, accurately reflecting the distance uncertainty given the available information.



**Figure 8.** Optimization of the association volume for catalog-based priors, based on HRDS H II regions. Background color scale and gray dashed contours denote the number of BGPS V2 objects falling within the association volume of one or more HRDS H II region, given the  $R_a$  and  $\Delta v_a$  indicated. Black contours show the bivariate cumulative distribution function for  $R$  and  $\Delta v$  for 749 GRS clouds from Roman-Duval et al. (2010), where the values shown indicate the fraction of clouds whose physical properties are *both* smaller than the axes values at that point. White contours mark the fraction of BGPS sources whose EMAF-derived distances and H II region-derived distances agree to better than 1 kpc. The white diamond indicates the chosen values of  $R_a$  and  $\Delta v_a$  while the brown and gray diamonds are discarded choices (see text).

### 5.3.2. Optimization

As with the morphological spectrum extraction technique of Section 4, we must balance the number of objects to which this method may apply with some measure of the method’s accuracy over the parameter space formed by the physical association volume values ( $R_a, \Delta v_a$ ). The figure of merit for determining the optimization of this method is the agreement fraction between the heliocentric distance returned for BGPS catalog sources by the EMAF and H<sub>2</sub> priors (EB13) versus that of the HRDS prior described above. While the optimization over association volume is done with respect to GRS clouds (size scale of coherent structure), the EMAFs and H II regions whose distances are compared are most likely substructures of the larger clouds. Showing that these smaller objects can be associated across GMC-scale distances indicates the validity of this method.

The physical properties of the 749 GRS clouds studied by Roman-Duval et al. (2010) form the parameter space over which this optimization is performed, as they represent coherent molecular structures in  $(\ell, b, v_{\text{LSR}})$  space. The bivariate cumulative distribution function of these quantities for these GRS clouds is shown as black contours in Figure 8, where the value shown at any point represents the fraction of GRS clouds whose  $(R_a, \Delta v_a)$  are both smaller than that point. For example, 75% of GRS clouds have  $(R_a, \Delta v_a)$  smaller than the line marked 0.75. The bivariate distribution value is generally smaller for any given point in the plane than the marginalized cumulative distributions for the individual properties. For instance, the brown diamond marks the 90th percentile values for each of  $R_a$  and  $\Delta v_a$  individually, but the combination lies along the 83rd percentile contour.

Akin to the color scale of  $N_{\text{comp}}$  in Figure 5, the color scale and gray dashed contours in Figure 8 depict the number of BGPS V2 sources associated with one or more HRDS H II region (Section 5.2). The number of associated sources grows very rapidly with increasing association volume, and small shifts in  $(R_a, \Delta v_a)$  can have a factor of two effect on the number of sources included. While large association volumes would assign prior DPDFs to many molecular cloud structures, a regime is quickly reached where that volume far exceeds the physical sizes of GMCs. Conversely, small association volumes are assured to lie entirely within most GMCs, but the usefulness of the prior becomes limited.

To assess the accuracy of prior DPDFs in resolving the KDA, we compared the heliocentric distance derived using only  $\text{DPDF}_{\text{hrds}}$  (with association volume  $R_a, \Delta v_a$ ) as a prior on  $\mathcal{L}(v_{\text{LSR}}, \ell, b; d_{\odot})$  using Equation (1) with the distance derived using the priors  $\text{DPDF}_{\text{emaf}}$  and  $\text{DPDF}_{\text{H}_2}$  from EB13 in the same fashion. This comparison was done over the relevant range of  $(R_a, \Delta v_a)$  parameter space, given the GRS cloud properties from Roman-Duval et al. (2010). At each point in the  $(R_a, \Delta v_a)$  plane, the comparison was computed using the set of EMAF-identified BGPS sources associated with one or more H II regions given those values for the association volume. The rate at which the heliocentric distances from the two methods agree to better than 1 kpc is shown by the white contours in Figure 8. We used the  $\text{DPDF}_{\text{emaf}}$  as the basis for calibrating the literature catalog association volume prior because it provides a high-accuracy distance estimate compared to GRS cloud KDA resolutions ( $\approx 92\%$ ; EB13) and provides well-constrained distance estimates for more than 700 molecular cloud structures. With large physical (three-dimensional) separation between BGPS sources and cataloged H II regions, the distance agreement rate falls from  $\gtrsim 90\%$  to  $\approx 80\%$  while the number of possible sources increases many-fold. The fact that the distance-agreement rate does not fall below 80% likely indicates the expected existence of larger structure within the Galactic plane beyond the scale of GMCs. For comparison, if the objects with near and far KDA resolutions in the V2 distance catalog (see Section 6.4) are randomly reassigned and those at the tangent point remain unchanged, the distance agreement would be only  $\approx 64\%$ .

As with optimizing the morphological spectrum extraction technique of Section 4, choosing optimum values of  $(R_a, \Delta v_a)$  requires balancing distance resolution accuracy with the number of objects to which the method may apply. To encompass the physical properties of the bulk of GRS clouds, we focused on the 50th percentile contour of the bivariate cumulative distribution function, and settled on  $(R_a, \Delta v_a) = (10.5 \text{ pc}, 3.6 \text{ km s}^{-1})$  as the best balance between distance agreement rate and number of objects included (white diamond). With these values, 15% of the EMAF- and H II region-derived distances disagree with each other. The contrasting physical conditions in these tracers of star formation (cold, dense, starless gas versus hot bubbles around young stars) accounts for much of the difference, given the assumption of smooth Galactic  $8 \mu\text{m}$  emission in  $\text{DPDF}_{\text{emaf}}$ . Furthermore, assuming each prior has an independent distance-assignment success rate of 92% (EB13) regardless of physics, the comparison of distances should agree at a rate of  $(0.92)^2 \approx 0.85$ , the rate indicated by the white diamond.

Other diamonds in Figure 8 represent sub-optimal values. The brown diamond (mentioned above) has a relatively large association volume and encompasses nearly 1000 BGPS objects, but is larger than 83% of GRS clouds and has a distance matching

rate of less than 80%. The gray diamond at  $(R_a, \Delta v_a) \approx (7.5 \text{ pc}, 3 \text{ km s}^{-1})$  lies at the intersection of the marginalized 50th percentile levels for each parameter (and so is reasonably matched to GRS cloud properties) and has a 90% distance matching rate, but is useful for less than 400 BGPS objects.

## 6. RESULTS

### 6.1. Kinematics

#### 6.1.1. Dense Gas Velocity Catalogs and BGPS Version 2

The recent re-reduction and expansion of the BGPS data set (G13) has implications for the association of spectroscopic observations with the latest (V2.1)<sup>16</sup> continuum source catalog. BGPS-led spectroscopic surveys were conducted with earlier versions of Bolocat (either 0.7 or 1.0; see S13), and source positions and boundaries may have changed. These changes generally occur in crowded regions where decomposition is ambiguous or at low signal to noise (see G13 for a full discussion). It is therefore necessary to carefully associate extant  $v_{\text{LSR}}$  information with the V2.1 source catalog.

The Bolocat cataloging routine produces label maps that identify survey mosaic pixels belonging to each catalog entry (Rosolowsky et al. 2010a). To account for the finite solid angle encompassed by each spectroscopic pointing, any one  $v_{\text{LSR}}$  measurement must be associated with all catalog objects whose label mask lies within one beam of the center of the spectroscopic pointing. For the majority of sources, there is a one-to-one correspondence; however, one velocity pointing may be assigned to more than one catalog source in more crowded regions. In addition, some pointings are no longer associated with a BGPS V2 source and are ignored. Use of survey label maps also allows for the direct incorporation of spectroscopic observations not predicated on BGPS catalog positions (e.g., observations based on ATLASGAL or Hi-GAL sources; Wienen et al. 2012; Jackson et al. 2013).

If more than one spectroscopic pointing is assigned to a given BGPS source, the properties of those spectra are compared. First, multiple  $v_{\text{LSR}}$  of spectra within the same survey (i.e.,  $\text{HCO}^+(3-2)$ ) are compared. If the constituent velocities are more than  $5 \text{ km s}^{-1}$  discrepant, no  $v_{\text{LSR}}$  from that survey is assigned and a flag is returned, otherwise the spectral fit properties ( $v_{\text{LSR}}$ , linewidth, uncertainties) are combined, weighted by the peak temperature of each spectrum. Second, the fit properties of different species are compared (i.e.,  $\text{NH}_3(1,1)$  versus  $\text{HCO}^+(3-2)$ ), and discrepant velocities result in no  $v_{\text{LSR}}$  being assigned to that source. Fit properties of different species are combined weighted by signal-to-noise ratio.

A tabulation of the number of BGPS V2.1 catalog objects that have a valid  $v_{\text{LSR}}$  from each of the molecular line surveys used is presented in Table 3. As an example of the shift in kinematic information from version 1 of the BGPS catalog, the HHT surveys detected either  $\text{HCO}^+(3-2)$  or  $\text{N}_2\text{H}^+(3-2)$  emission associated with 3126 BGPS version 1 sources, but only 2676 version 2 sources have a valid (i.e., not conflicting)  $v_{\text{LSR}}$  from these data. For the  $\text{NH}_3$  surveys of Dunham et al., the number of associated sources went from 490 for V1 to 455 for V2. There is overlap between the dense-gas surveys, with some catalog objects having accordant velocity information from two or three different molecular line transitions. Therefore, although the values in Table 3 add up to a larger number, a total of 2925

**Table 3**

Number of BGPS V2 Sources for Each Velocity Tracer

Species	$N$	Ref.
$\text{HCO}^+(3-2)$	2604	1
$\text{N}_2\text{H}^+(3-2)$	69	1
$\text{CS}(2-1)$	256	2
$\text{NH}_3(1,1)$	453	3,4
$\text{C}^{18}\text{O}(2-1)$	141	5
$^{13}\text{CO}(1-0)$	2279	6

**References.** (1) Shirley et al. 2013; (2) Y. Shirley (2012, private communication); (3) Dunham et al. 2011b; (4) Dunham et al. 2010; (5) M. Lichtenberger (2014, private communication); (6) GRS (this work).

BGPS V2 sources have a valid  $v_{\text{LSR}}$  from one or more of the dense gas spectral surveys discussed in Section 2.2.

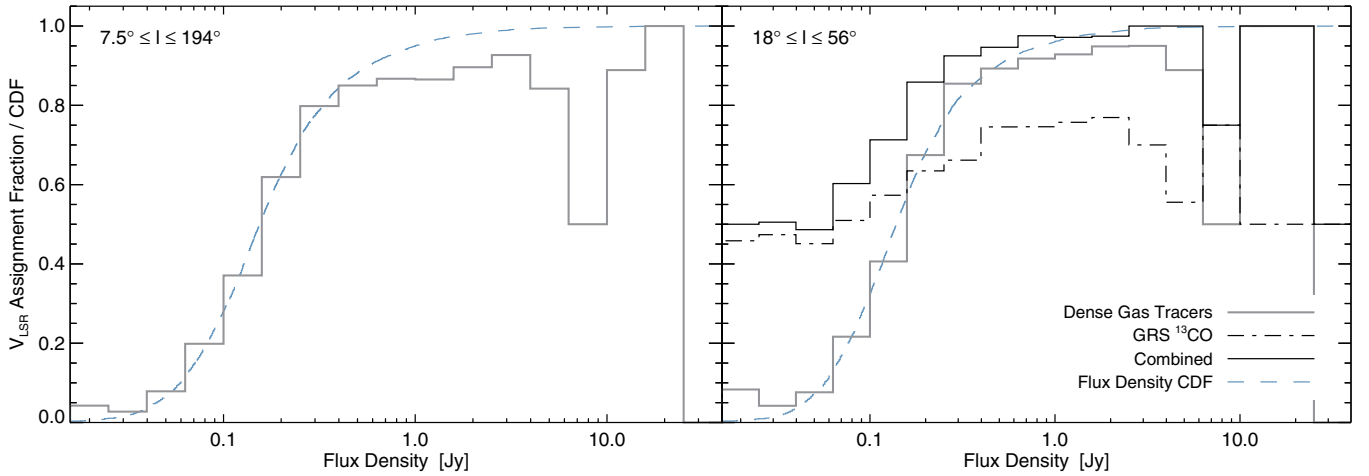
#### 6.1.2. Kinematic Catalog Expansion Using $^{13}\text{CO}$

Because  $^{13}\text{CO}(1-0)$  traces lower-density gas, the resulting kinematic distance likelihood sometimes represents a different velocity than one of the dense-gas tracers (see Figure 6). The molecular species described in Section 2.2 trace the dense environments of molecular cloud clumps and cores (Evans 1999), so we preferentially use that information over the  $^{13}\text{CO}$  data when it is available. New  $^{13}\text{CO}$  kinematic distances are therefore limited to those objects without an assigned  $v_{\text{LSR}}$  from the dense-gas tracers due to non-detection, detection of multiple velocity components with  $\Delta v \geq 5 \text{ km s}^{-1}$ , or sources not observed by spectroscopic surveys. As a result, while 2279 Bolocat objects have valid  $^{13}\text{CO}$   $v_{\text{LSR}}$  (as defined by the flow chart of Figure 4) only 958 have a  $^{13}\text{CO}$  velocity exclusively. The remainder overlap one or more dense gas tracer, and have velocities that agree with the other measurements 95% of the time. Combination of all available kinematic information yields a collection of 3900 BGPS sources (representing 45% of the entire V2 catalog) whose velocity information is included with the expanded distance catalog (Section 6.4).

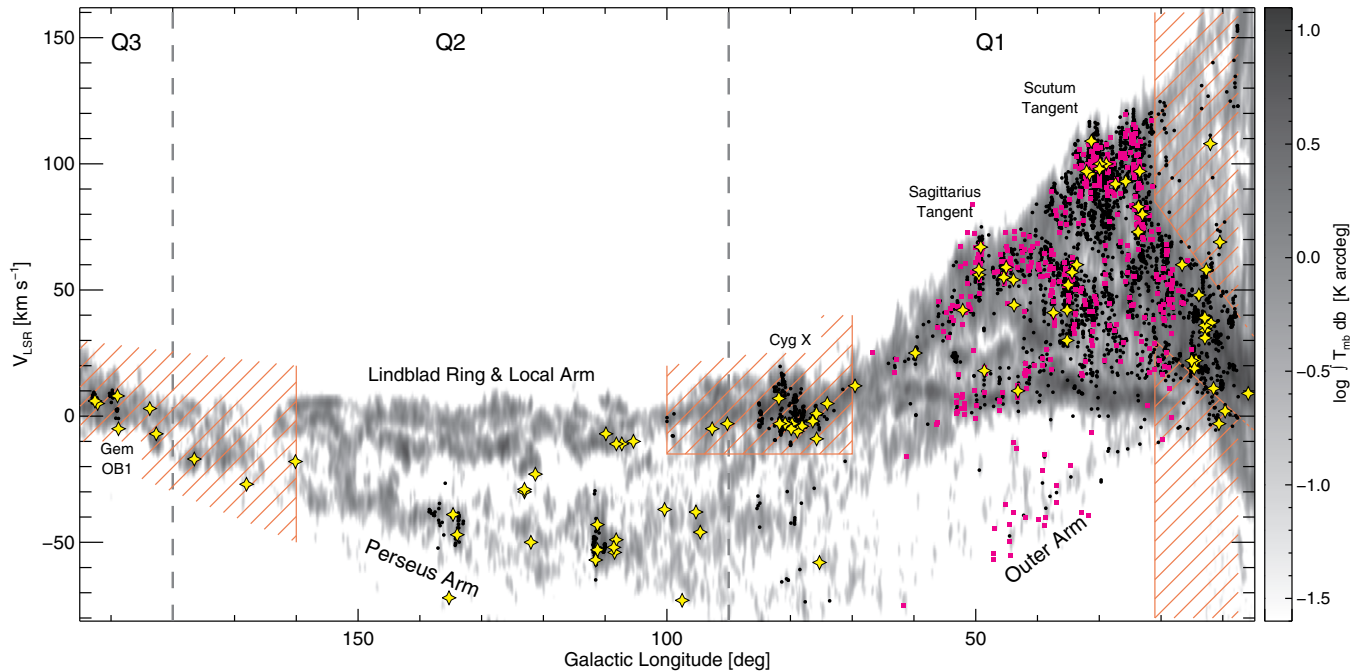
Shirley et al. (2013) showed a strong correlation between BGPS  $\lambda = 1.1 \text{ mm}$  flux density and detection fraction in  $\text{HCO}^+(3-2)$  or  $\text{N}_2\text{H}^+(3-2)$ . With the shift to BGPS V2 and the addition of  $^{13}\text{CO}(1-0)$  as an additional velocity tracer, we re-examine this relationship. The fraction of sources in logarithmic flux density bins that have an assigned  $v_{\text{LSR}}$  are shown in Figure 9; sources without such an assignment may be spectroscopic non-detections, have multiple velocity components, or be unobserved. The left panel of Figure 9 reflects the Galactic longitude range of the Shirley et al. survey, the most comprehensive sample to date. The dips in the  $v_{\text{LSR}}$  assignment rate at  $S_{1.1} \geq 4 \text{ Jy}$  are due to 11 catalog sources, 6 of which have multiple velocity components detected in  $\text{HCO}^+(3-2)$ , 2 are new sources at the edge of BGPS mosaics not identified in the V1 catalog, and 3 are new catalog entries created by the subdivision of V1 sources (see G13 for a discussion) that do not lie within one beam of a spectroscopic pointing. Aside from these sources, the S13 result that  $v_{\text{LSR}}$  assignment falls below 50% for  $S_{1.1} < 200 \text{ mJy}$  still holds.

The right panel of Figure 9 illustrates this comparison restricted to the Galactic longitude coverage of the GRS. In addition to the dense-gas tracers (solid gray), the velocity assignment based on morphologically extracted  $^{13}\text{CO}$  spectra (dot-dashed black) is shown, along with the combined (i.e., sources having a velocity from one or the other) assignment

<sup>16</sup> The V2.1 catalog corrects various cataloging errors and represents the definitive catalog of objects in the V2.0 images (G13).



**Figure 9.** Assignment of  $v_{\text{LSR}}$  as a function of  $\lambda = 1.1$  mm peak flux density. The blue dashed curve in each panel represents the cumulative distribution of flux densities for that subset of sources. Left: velocity assignment fraction for the collection of dense-gas tracers discussed in Section 2.2 over the longitude range of Shirley et al. (2013). The dips in assignment fraction at  $S_{1.1} \gtrsim 4$  Jy are due to multiple detected velocity components, new V2 sources on the edge of mosaic panels, or source boundary shifts between the V1 and V2.1 catalogs. Right: Galactic longitude further limited to the GRS coverage. Velocity assignment fractions of dense-gas tracers (solid gray) and morphologically extracted  $^{13}\text{CO}$  spectra (dot-dashed black) are shown, along with the combined velocity assignment fraction (solid black).



**Figure 10.** Longitude–velocity diagram for the BGPS. Background image is the latitude-integrated  $^{12}\text{CO}$  intensity of Dame et al. (2001). Black circles mark the locations of BGPS V2 sources, and magenta squares identify the H II regions from Table 2. Yellow stars mark the trigonometric parallax measurements presented in Table 1 of Reid et al. (2014). The kinematic avoidance zones discussed in Section 3.2.2 are shown as hashed regions. Various Galactic features are identified.

rate (solid black). Two features bear remark: (1) the velocity-assignment rate in  $^{13}\text{CO}$  never falls below 45% even at low flux density, and (2) the combined  $v_{\text{LSR}}$  assignment rate  $\gtrsim 95\%$  for  $S_{1.1} \geq 0.4$  Jy. The kinematic distance likelihood provided by  $^{13}\text{CO}$  is, therefore, a vital complement to the suite of dense-gas spectroscopic surveys.

### 6.1.3. Kinematics and Galactic Structure

The expanded BGPS kinematic catalog can be used to trace out prominent Galactic structure features in the first and second quadrants. Figure 10 illustrates the correlation between the more diffuse molecular component of the interstellar medium (latitude-integrated  $^{12}\text{CO}$  of Dame et al. 2001; grayscale back-

ground) and the dense gas seen in the BGPS continuum images (black circles). The molecular cloud structures identified by the BGPS strongly trace out the Scutum–Centaurus arm (whose tangent is at  $\ell \approx 30^\circ$ ) and the Sagittarius arm (tangent near  $\ell = 50^\circ$ ). Also prominent are the Cygnus X star-forming complex near  $\ell = 80^\circ$  and portions of the Perseus arm from  $\ell = 60^\circ$  all the way around to Gemini OB1 in the third quadrant. BGPS sources at negative  $v_{\text{LSR}}$  in the first quadrant are generally associated with CO emission identified as the Outer arm by Dame et al. (2001). The one feature in the  $^{12}\text{CO}$  image not traced by dense gas is the plethora of molecular gas near  $v_{\text{LSR}} = 0$  km s $^{-1}$  in the second quadrant, identified by Dame et al. as being in the Lindblad ring and Local arm.

Locations of HMSFRs, identified by HRDS H II regions or the maser emission targeted by the BeSSeL survey, are also indicated in Figure 10 and generally follow the same patterns as the BGPS molecular cloud structures. The H II regions from Table 2 are shown as magenta squares, and yellow stars mark the BeSSeL maser sources. While the HRDS only covered the inner Galaxy, its sources pick out the major spiral features including the Outer arm and Perseus arm in the range  $60^\circ \geq \ell \geq 40^\circ$ . The trigonometric parallax measurements of BeSSeL span the entire range of the BGPS kinematic catalog, and provide valuable heliocentric distance anchor points for objects in the hashed kinematic avoidance zones around the Galactic cardinal directions (Section 3.2.2).

## 6.2. Prior DPDFs

### 6.2.1. Priors from EB13

Two prior DPDFs were introduced in EB13 to resolve the KDA for BGPS sources. The first is based on the distribution of molecular gas in the Galactic disk to constrain sources at high Galactic latitude ( $|b| \gtrsim 0.4$ ) to the near kinematic distance. This constraint derives from the relatively narrow thickness of the disk’s molecular layer (HWHM  $\approx 60$  pc; Bronfman et al. 1988); lines of sight at these latitudes exit this layer before reaching the far kinematic distance for much of the inner Galaxy. We introduce here two limitations on the use of  $\text{DPDF}_{\text{H}_2}$  that were not relevant to the source sample in EB13. The model presented by Bronfman et al. (1988) sought to quantify molecular gas in the “5 kpc Ring” and outward through Galactic disk for  $R_{\text{gal}} \geq 0.2R_0$ . No component is included to model the Galactic central molecular zone. Additionally, in the outer Galaxy, there is no near/far kinematic discrimination required, and since the density of molecular gas rapidly diminishes as a function of distance in these regions, application of  $\text{DPDF}_{\text{H}_2}$  only skews the posterior DPDF to smaller  $d_\odot$ . We therefore limit the application of this prior to  $11^\circ \leq |\ell| \leq 70^\circ$ , with the upper limit corresponding to the start of the kinematic avoidance zone near  $\ell = 90^\circ$ .

The second prior from EB13 is based on the absorption of mid-infrared light by the dust seen in millimeter continuum surveys. Expanding on the concept of an infrared dark cloud (IRDC; see Simon et al. 2006a), EMAFs are visible as decrements in bright Galactic emission in the  $\lambda = 8 \mu\text{m}$  *Spitzer*/GLIMPSE images (Churchwell et al. 2009) at the locations of millimeter-detected molecular cloud structures. By comparing the column density derived from BGPS emission, EB13 placed each object within a numerical model of Galactic  $8 \mu\text{m}$  emission (Robitaille et al. 2012) to construct a  $\text{DPDF}_{\text{emaf}}$  based on the morphological comparison of synthetic infrared images with processed versions of the GLIMPSE mosaics. Application of the EMAF method in the present work remains virtually unchanged from EB13, but with the shift to BGPS V2 and expansion of the kinematic catalog, the collection of sources meeting the automated EMAF selection criteria were once again examined by eye to create the final list of rejected sources (see EB13 for details). Whereas there were 770 BGPS V1 sources for which  $\text{DPDF}_{\text{emaf}}$  was computed, 854 BGPS V2 sources met the final selection criteria, due in part to the expansion of the kinematic data set with GRS  $^{13}\text{CO}$  data. Of this set of EMAFs, well-constrained distance estimates exist for 679 (see Section 6.4), an improvement in itself over the V1 set of 618 well-constrained sources from EB13.

### 6.2.2. Trigonometric Parallax Measurements

Using the recently compiled list of 103 trigonometric parallax measurements from the BeSSeL Survey and VERA Project (Reid et al. 2014), we find a total of 292 BGPS V2 objects that can be associated with one or more maser sources within the confines of the association volume defined above. Due to the small uncertainty in the parallaxes themselves, 291 of these objects have well-constrained posterior DPDFs (see Section 6.3). The sole unconstrained BGPS source, G023.743–00.235, is associated with a maser identified by Reid et al. (2014) as being in the 4 kpc /Norma arm, and has a kinematic distance  $\sim 1.5$  kpc discrepant from the parallax distance. Because  $\text{DPDF}_{\text{px}}$  is offset from a kinematic distance peak, there still exists sizable probability in both kinematic peaks in the posterior DPDF. Otherwise, of this sample of BGPS V2 objects, only 223 have a  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$  despite all having a measured  $v_{\text{LSR}}$ ; the 69 sources lying in a kinematic avoidance zone therefore have well-constrained distance estimates owing exclusively to their association with a trigonometric parallax measurement.

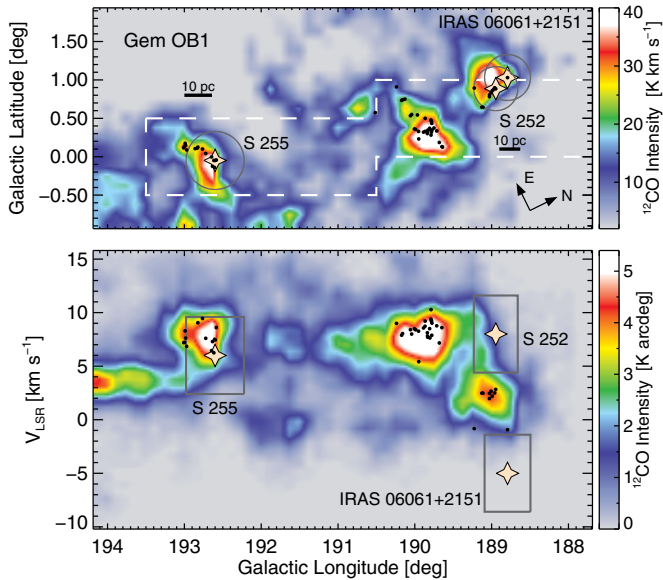
### 6.2.3. H II Regions

The combined HRDS catalog with robust KDA resolutions from Table 2 encompasses 441 H II regions, and a total of 525 BGPS V2 sources are associated with one or more of these regions. The plentiful nature of these objects (see Figure 10) means that given the association volume defined in Section 5.3.2, a BGPS V2 source will occasionally ( $\approx 14\%$ ) be associated with more than one HRDS object. Approximately 20% of this subset have an unconstrained  $\text{DPDF}_{\text{hrds}}$  based on the conflict between KDA resolutions for constituent H II regions, accounting for some 3% of the HRDS-associated BGPS sources. Overall, 95% of the HRDS-based DPDFs produce well-constrained distance estimates, with the remaining 2% being unconstrained sources arising from a disagreement between  $\text{DPDF}_{\text{hrds}}$  and  $\text{DPDF}_{\text{emaf}}$ .

### 6.2.4. Gemini OBI

In the outer Galaxy, star-forming regions tend to be in widely separated, easily distinguished, and coherent structures (Heyer et al. 1998; Brunt et al. 2003; Dunham et al. 2010). The Gemini OBI molecular cloud was one such region selected for detailed study by the BGPS (Dunham et al. 2010). Because it lies within a kinematic avoidance zone, its distance must be estimated without  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$ . Fortunately, there exist VLBI trigonometric parallax measurements of  $\text{H}_2\text{O}$  and  $\text{CH}_3\text{OH}$  masers for three sources in this molecular cloud (shown as stars in Figure 11). Two of these, S252 and IRAS 06061+2151, lie in the northernmost clump of gas, and have parallax distances of  $2.10^{+0.03}_{-0.03}$  kpc (Reid et al. 2009a) and  $2.02^{+0.53}_{-0.35}$  kpc (Niinuma et al. 2011), respectively. The parallax measurement of the H II region S255 (in the more southerly clump of gas) places it somewhat closer at  $1.59^{+0.07}_{-0.07}$  kpc (Rygl et al. 2010). The projections of the cylindrical association volumes for these sources are shown in both panels of Figure 11.

For the group around S255, most of the BGPS sources lie within the association volume of the parallax measurement, but a handful fall just outside. Based on the coherent nature of the  $^{12}\text{CO}(1-0)$  emission in Figure 11, we assign the 1.59 kpc distance of S255 to the remainder of the BGPS sources in the southerly group. Interestingly, while the maser locations for S252 and IRAS 06061+2151 are spatially coincident with the northernmost clump of gas and collection of BGPS sources, their velocities bracket the  $^{12}\text{CO}$  and dense-gas emission associated

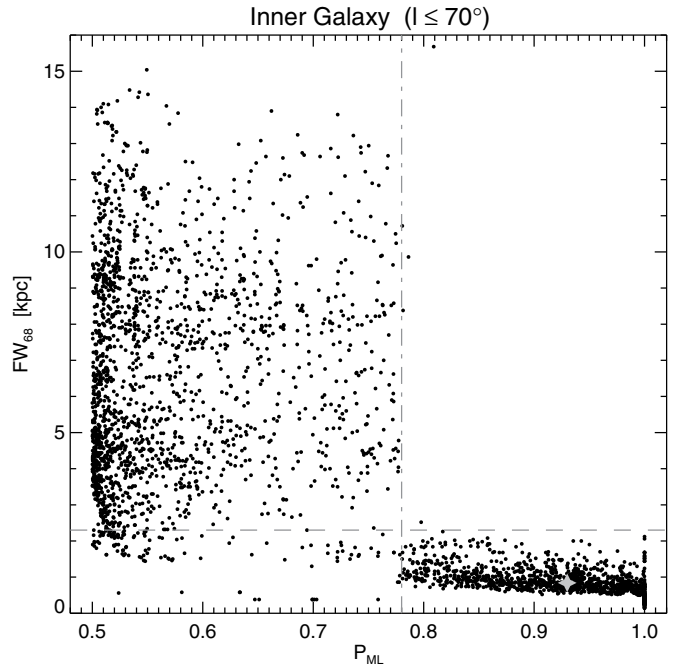


**Figure 11.** Gemini OB1 field. Top: sky-projected map of the Gemini OB1 molecular cloud. The background image is the velocity-integrated  $^{12}\text{CO}$  data of Stacy & Thaddeus (1991), Dame et al. (2001), and black circles mark the locations of BGPS V2 sources. Stars mark the locations of the masers with trigonometric parallax measurements, and the enclosing circles mark the projected association volumes, computed as in Section 5.3. The white dashed line marks the approximate boundary of the BGPS observations, and the scale bars indicate the projected distances for the two subregions. Bottom: the longitude–velocity integration over  $-1^\circ \leq b \leq 1^\circ$  of the same field. Both panels share a common longitude axis.

with the dust (Figure 11, bottom). The origins of these offsets are unclear, and it may be that these masers are associated with young stars that have been ejected from the northerly group of sources. The limited association volume for applying the parallax prior directly means that none of the BGPS sources in the more northerly group may be assigned a DPDF<sub>px</sub>. Following Dunham et al. (2010), therefore, we simply assign the 2.10 kpc distance of S252 to all of the BGPS sources in this complex at  $\ell \lesssim 191^\circ$ .

### 6.3. Analysis of “Well-constrained” Distance Estimates

The definition of a “well-constrained” distance estimate was characterized in EB13 for objects identified as EMAFs. With the introduction of additional kinematic distance likelihoods and prior DPDFs, we analyze the properties of the newly accumulated posterior DPDFs in the event different criteria are required for this definition. As a proxy for the strength of the KDA resolution,  $P_{\text{ML}}$  is the integrated posterior DPDF on the side of the tangent point ( $d_{\text{tan}}$ ) containing the maximum-likelihood distance ( $d_{\text{ML}}$ ). For sources well away from  $d_{\text{tan}}$  this includes the entirety of a single kinematic distance peak. Because the posterior DPDFs are normalized to unit integral probability and there are at most two possible kinematic distances,  $0.5 \lesssim P_{\text{ML}} \leq 1.0$ . The measure of constraint tightness on a distance estimate is computed as the full width of the region around  $d_{\text{ML}}$  containing at least 68.3% of the integrated probability and whose bounds occur at equal probability level ( $\text{FW}_{68}$ ). Sources near  $d_{\text{tan}}$  will have  $\text{FW}_{68} \approx 1\text{--}2$  kpc due to localization from  $d v_{\text{LSR}}/d(d_\odot)$  (Section 3.2.2). For single-peaked DPDFs,  $\text{FW}_{68}$  approximates the Gaussian  $\pm 1\sigma$  region around the centroid, but this approximation breaks down when one kinematic peak is not dominant over the other.



**Figure 12.** “Well-constrained” distance estimate criteria. Each BGPS V2.1 catalog object at  $\ell \leq 70^\circ$  is shown, plotted by its integrated posterior DPDF on the  $d_{\text{ML}}$  side of the tangent point ( $P_{\text{ML}}$ ) and the full-width error bar containing  $\geq 68.3\%$  of the integrated posterior DPDF around  $d_{\text{ML}}$  ( $\text{FW}_{68}$ ). The vertical dot–dashed line at  $P_{\text{ML}} = 0.78$  is shown to guide the eye, while the horizontal dashed line marks the  $\text{FW}_{68} \leq 2.3$  kpc boundary for a “well-constrained” distance estimate. The gray star identifies the median values of  $P_{\text{ML}}$  and  $\text{FW}_{68}$  for the well-constrained sample.

Choice of criteria for “well-constrained” distance estimates once again relies on the bivariate distribution of  $P_{\text{ML}}$  and  $\text{FW}_{68}$ , shown in Figure 12. As in EB13, the pattern whereby  $\text{FW}_{68}$  becomes large for  $P_{\text{ML}} \lesssim 0.78$  holds. This break is due to the geometry of  $\mathcal{L}(v_{\text{LSR}}, l, b; d_\odot)$ , where a single kinematic distance peak may contain  $\geq 68.3\%$  of the integrated posterior DPDF only if it sufficiently dominates over the other peak. For objects that have  $P_{\text{ML}} \gtrsim 0.78$ , the maximum value of  $\text{FW}_{68}$  is approximately 2.3 kpc. We again choose to define a “well-constrained” distance estimate as  $\text{FW}_{68} \leq 2.3$  kpc to include objects within  $\approx 1$  kpc of  $d_{\text{tan}}$  (lower left corner of Figure 12) because the near/far ambiguity for these sources leads to a distance discrepancy of  $\lesssim 2$  kpc. We note that while we chose  $\text{FW}_{68} = 2.3$  kpc to define “well-constrained,” it is clearly an upper limit, and the vast majority of these sources have  $\text{FW}_{68} \lesssim 1.5$  kpc, or a distance uncertainty of  $\lesssim 0.8$  kpc. The median values of  $P_{\text{ML}}$  and  $\text{FW}_{68}$  for the well-constrained sample are 0.928 and 0.84 kpc, respectively (gray star in Figure 12).

In the outer Galaxy, there is neither a tangent distance nor an ambiguity in the kinematic distance, so all distance estimates for objects in Quadrants II and III have  $P_{\text{ML}} = 1.0$  and  $\text{FW}_{68}$  well within the limit outlined above.

## 6.4. The V2 Distance Catalog and Posterior DPDFs

### 6.4.1. Catalog Description

A posterior DPDF for each BGPS V2.1 catalog object was computed using Equation (1) and the appropriate kinematic distance likelihood (dense gas or GRS  $^{13}\text{CO}$ ), and all applicable prior DPDFs. From the posterior DPDFs, the  $P_{\text{ML}}$  and  $\text{FW}_{68}$  statistics, distance estimates and KDA resolutions were determined. The resulting distance catalog includes relevant

**Table 4**  
KDA Resolutions for BGPS V2 Sources

KDA Resolution	Flag	$N_{\text{kin}}^{\text{a}}$	$N_{\text{tot}}^{\text{b}}$	$f_{\text{w.c.}}^{\text{c}}$ (%)
Near	N	984	984	57.5
Far	F	336	336	19.6
Tangent	T	193	193	11.3
Outer Galaxy	O	197	197	11.5
Unconstrained	U	1798	6487	...
Excluded <sup>d</sup>	X	0	397	...
Total		3508	8594	100

**Notes.**<sup>a</sup> Number of objects in the “kinematic sample.”<sup>b</sup> Number of objects from the full Bolocat V2.<sup>c</sup> Fraction of the well-constrained sources with this KDA resolution.<sup>d</sup> Object lies in a kinematic avoidance zone (Section 3.2.2).

information from the Bolocat V2.1, velocity information ( $v_{\text{LSR}}$ , survey), and heliocentric distance and Galactocentric position, if available.

With the expansion of the source list beyond  $\ell = 65^\circ$ , two new KDA resolution flags are introduced in this distance catalog. As in EB13, sources whose  $d_{\text{ML}}$  is within 1 kpc of  $d_{\text{tan}}$  are given the flag T, indicating they are at (or near enough) the tangent point. Sources in the inner Galaxy with “well-constrained” distance estimates with  $d_{\text{ML}} < (>) d_{\text{tan}}$  are again given the flag N (F). Outer Galaxy objects, for which there is no KDA, are assigned O if they have an associated  $v_{\text{LSR}}$  and lie outside a kinematic avoidance zone. Those objects throughout the Galactic plane lying inside a kinematic avoidance zone are given the flag X, specifying their kinematic information has been excluded. Objects in a kinematic avoidance zone that are associated with a trigonometric parallax measurement or are in Gemini OB1 may still be given a “resolved” KDA flag. The remaining sources that either have no kinematic information or whose posterior DPDF does not meet the criteria of Section 6.3 are

assigned the flag U. Table 4 lists the number of objects in the BGPS catalog with each KDA resolution flag. Column 3 lists objects in the “kinematic sample,” that is objects that possess either a kinematic distance from dense gas or  $^{13}\text{CO}$  spectra, an association with a trigonometric parallax measurement, or a location in Gemini OB1. The fourth column lists all BGPS V2 objects, while the final column indicates the fraction of well-constrained sources with each KDA resolution flag.

The BGPS distance catalog is presented in Table 5, which contains entries for each of the 3689 catalog objects in the kinematic sample. Objects with a well-constrained distance estimate (flags N/F/T/O) have the maximum-likelihood distance ( $d_{\text{ML}}$ ) listed, along with the associated error bars. Tangent point objects additionally list the first-moment distance ( $\bar{d}$ ), following the discussion in EB13. Object with flags U or X have no heliocentric distance information included. Galactocentric radius is computed for each object with a detected  $v_{\text{LSR}}$ , save those with KDA flag X, as  $R_{\text{gal}}$  is not subject to the KDA but is affected by the non-circular motions characterizing the kinematic avoidance zones. For objects with a well-constrained distance estimate, Galactocentric vertical position ( $z$ ) is also computed, subject to the coordinate transformation presented in Appendix C of EB13. The DPDFs for all 8594 objects in the BGPS V2.1 catalog are publicly available.<sup>17</sup>

DPDFs for example sources are shown in Figure 13 for several cases. First, the constituent DPDFs provide a well-constrained near distance for source G038.848–00.427 (panel (a)), with the DPDF<sub>emaf</sub> providing most of the resolving power. This source also demonstrates a  $^{13}\text{CO}$  velocity that agrees with the  $v_{\text{LSR}}$  from a dense gas tracer. Panel (b) illustrates a source (G043.124+00.029) associated with a trigonometric parallax measurement, W49 in this case. The DPDF<sub>px</sub> aligns well with the kinematic distance likelihood to produce a tightly constrained distance. An example of a marginal distance resolution is shown

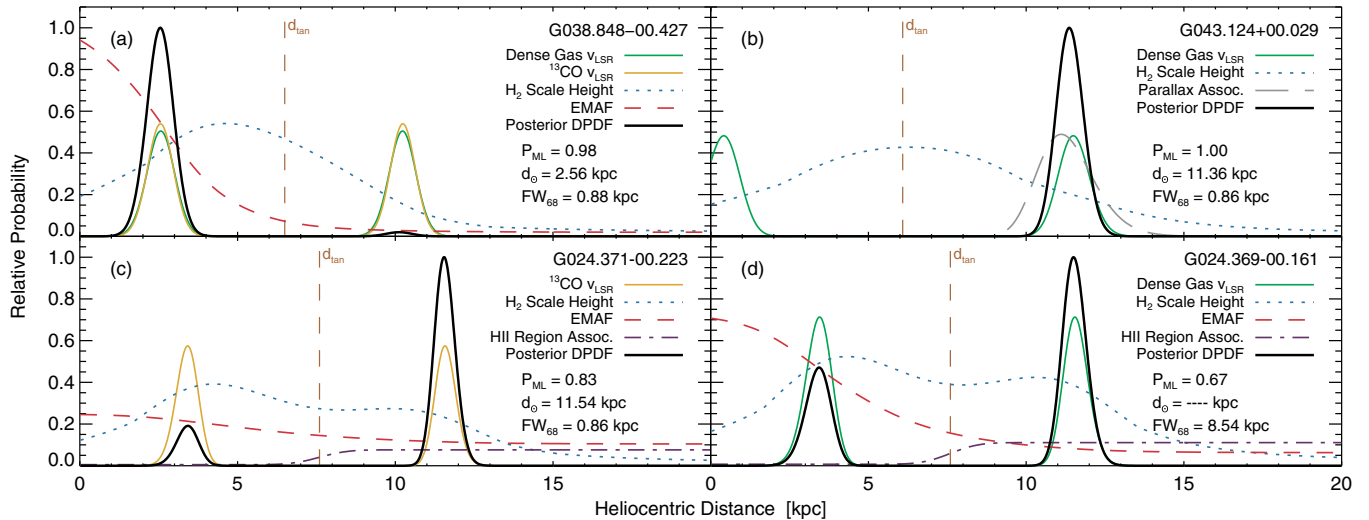
<sup>17</sup> Available through IPAC at [http://irsa.ipac.caltech.edu/data/BOLOCAM\\_GPS](http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS)

**Table 5**  
BGPS V2 Distance Catalog

V2.1 Catalog Properties				Velocity		Heliocentric Distance			Galactocentric Position		
Catalog Number	$\ell$ (°)	$b$ (°)	$S_{\text{int}}$ (Jy)	$v_{\text{LSR}}$ (km s <sup>-1</sup> )	Ref.	KDA Resol.	$P_{\text{ML}}^{\text{a}}$	$d_{\text{ML}}$ (kpc)	$\bar{d}$ (kpc)	$R_{\text{gal}}^{\text{b}}$ (kpc)	$z^{\text{c}}$ (pc)
3475	20.441	-0.030	0.72(0.17)	76.0	3	F	0.99	11.00 <sup>+0.26</sup> <sub>-0.24</sub>	...	4.32 <sup>+0.20</sup> <sub>-0.17</sub>	-12(5.1)
4061	24.171	0.063	0.56(0.11)	110.4	6	N	0.78	5.60 <sup>+0.44</sup> <sub>-0.42</sub>	...	3.96 <sup>+0.23</sup> <sub>-0.20</sub>	16(5.0)
4969	29.279	0.096	1.39(0.14)	87.0	5	N	0.79	4.82 <sup>+0.58</sup> <sub>-0.56</sub>	...	4.76 <sup>+0.31</sup> <sub>-0.27</sub>	20(5.0)
5197	30.464	0.033	1.60(0.18)	106.0	1, 2	U	0.51	...	...	4.46 <sup>+0.14</sup> <sub>-0.14</sub>	...
5560	31.715	0.502	0.31(0.09)	53.6	6	U	0.65	...	...	5.85 <sup>+0.23</sup> <sub>-0.23</sub>	...
6196	35.043	-0.474	5.14(0.34)	51.2	1	F	0.79	10.38 <sup>+0.60</sup> <sub>-0.58</sub>	...	5.96 <sup>+0.38</sup> <sub>-0.33</sub>	-86(8.1)
6711	43.816	-0.120	0.39(0.14)	46.0	6	T	0.65	6.08 <sup>+0.18</sup> <sub>-0.20</sub>	6.09(0.20)	5.77 <sup>+0.01</sup> <sub>-0.00</sub>	-0.9(5.1)
6964	53.591	-0.244	1.31(0.24)	59.7	1, 3	T	0.54	4.86 <sup>+0.78</sup> <sub>-0.78</sub>	4.86(0.77)	6.71 <sup>+0.05</sup> <sub>-0.00</sub>	-4.4(6.8)
7612	110.047	-0.019	0.14(0.05)	-50.7	1	O	1.00	4.32 <sup>+0.68</sup> <sub>-0.62</sub>	...	10.63 <sup>+0.47</sup> <sub>-0.41</sub>	28(5.0)
8210	192.584	-0.043	13.93(0.95)	8.6	1, 4	O	1.00	1.60 <sup>+0.06</sup> <sub>-0.06</sub>	...	9.91 <sup>+0.06</sup> <sub>-0.06</sub>	28(5.0)

**Notes.** Errors are given in parentheses.<sup>a</sup> The integrated posterior DPDF on the  $d_{\text{ML}}$  side of the tangent point; used as a proxy for the quality of the distance constraint.<sup>b</sup>  $R_{\text{gal}}$  is unambiguous and is computed from  $v_{\text{LSR}}$  and assumed uncertainty of 7 km s<sup>-1</sup> (Reid et al. 2009b) for sources without a well-constrained distance estimate.<sup>c</sup> Errors include contributions from variations in  $z$  along the line of sight over the range  $d_{\text{O}} \pm \sigma_d$  and the  $\pm 5$  km s<sup>-1</sup> uncertainty in the solar offset above the Galactic midplane (Jurić et al. 2008), added in quadrature.**References.** (1) Shirley et al. 2013; (2) Y. Shirley (2012, private communication); (3) Dunham et al. 2011b; (4) Dunham et al. 2010; (5) M. Lichtenberger, 2014, private communication; (6) GRS (this work).

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



**Figure 13.** Collected DPDFs for example sources. (a) Source G038.848–00.427 has a  $^{13}\text{CO}$  velocity agreeing with the dense gas  $v_{\text{LSR}}$ , and is placed at the near kinematic distance by  $\text{DPDF}_{\text{emaf}}$  with high probability. (b) Source G043.124+00.029 is associated with the trigonometric parallax measurement of W49 N. (c) Source G024.371–00.223 has kinematic information from  $^{13}\text{CO}$  and also has conflicting  $\text{DPDF}_{\text{emaf}}$  and  $\text{DPDF}_{\text{hrds}}$ , but the latter contributes more strongly to marginally place this source at the far kinematic distance. (d) Source G024.369–00.161 is associated with the same H II region as the source in c, but the stronger contribution from  $\text{DPDF}_{\text{emaf}}$  renders the distance to this source ambiguous.

in panel (c), where source G024.371–00.223 has conflicting  $\text{DPDF}_{\text{emaf}}$  and  $\text{DPDF}_{\text{hrds}}$ . In this case, the HRDS prior contributes more strongly to the posterior DPDF, placing enough probability in the far kinematic distance peak to constrain the distance. A neighboring source (G024.369–00.161, panel (d)), however, has a stronger EMAF morphological match that leaves the posterior DPDF without sufficient probability in a single peak for a constrained distance.

#### 6.4.2. Source Properties

From the pool of sources in Table 5, 1710 (49%) have a well-constrained distance estimate, representing a substantial population of molecular cloud structures for which physical properties may be derived (T. Ellsworth-Bowers, in preparation). A summary of the quantities in Table 5 is presented in Figure 14 and discussed here.

The distribution of Galactic longitude (top left panel) picks out concentrations of sources along the Galactic plane. The black histogram represents the set of well-constrained distances, while cyan represents the remainder of the BGPS V2.1 catalog. Gray hashing delineates the longitude-projected kinematic avoidance zones (Section 3.2.2), although only the region around  $\ell = 180^\circ$  applies to all  $v_{\text{LSR}}$ . Any well-constrained objects within these zones either lie at an allowed  $v_{\text{LSR}}$  or are associated with a trigonometric parallax measurement (Section 5.1). The red histogram describes the distribution of spectroscopic observations from S13, illustrating the regions for which distances could possibly be derived (e.g., S13 did not observe  $\ell \geq 200^\circ$ , as this was a new region in the BGPS V2 release; G13).

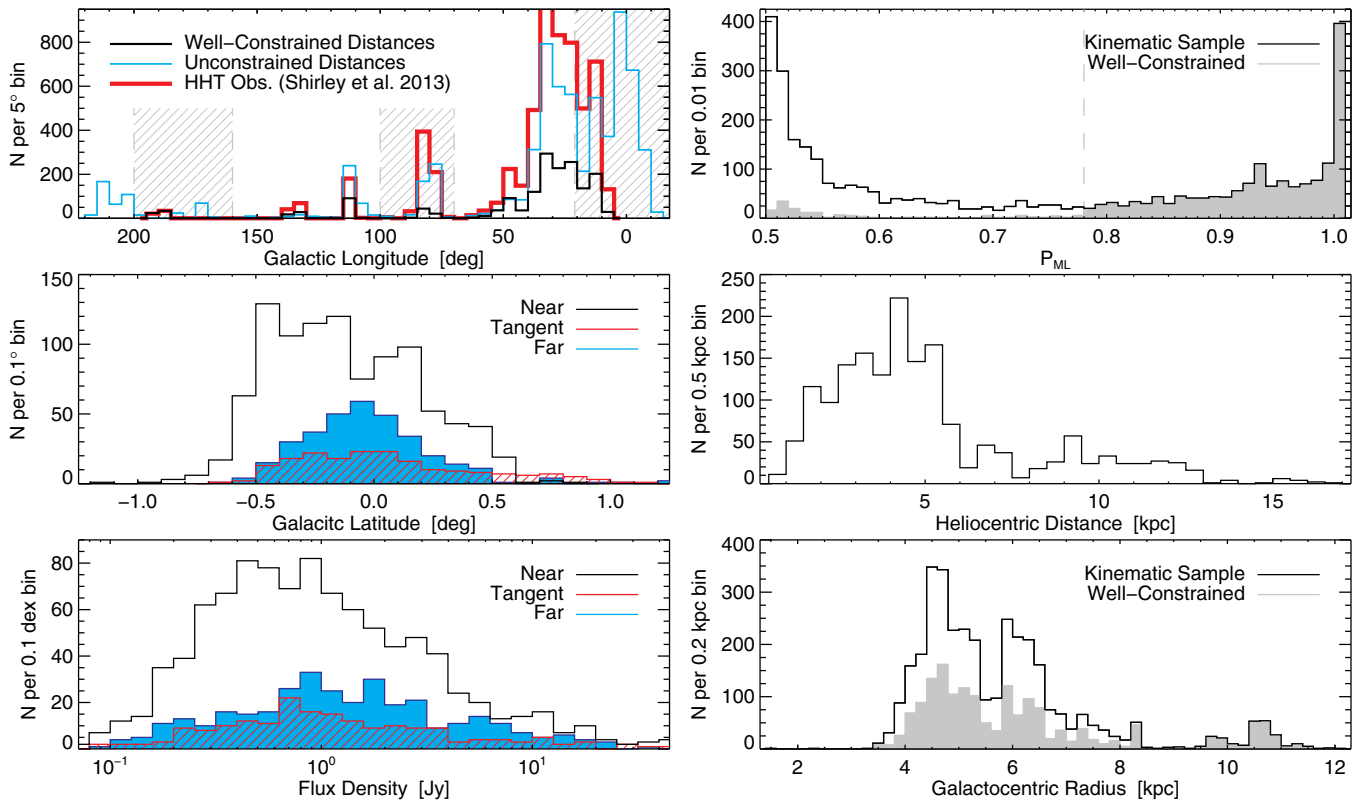
The histograms of Galactic latitude and  $\lambda = 1.1$  mm flux density are shown for sources at  $\ell \leq 90^\circ$  in the middle and bottom left panels of Figure 14, respectively, for each of the three KDA resolutions to illustrate the systematic effects of sources at different distances. Outer Galaxy sources are excluded to minimize the effects of Galactocentric radius on the analysis. In the middle left panel, objects at the near kinematic distance or tangent point subtend larger swaths across the width of the Galactic plane ( $\text{FWHM} \approx 0.8^\circ$  and  $\approx 0.9^\circ$ , respectively) than the

far kinematic group ( $\text{FWHM} \approx 0.6^\circ$ ). This is to be expected as objects at  $|b| \gtrsim 0.4^\circ$  are generally assigned the near kinematic distance by  $\text{DPDF}_{\text{H}_2}$ . A Kolmogorov–Smirnov (K-S) test finds that even the far and tangent groups come from different underlying distributions at a 99.7% confidence level, while the near group is further divergent. In terms of the millimeter flux density, the median values for the near and tangent groups are  $\approx 0.8$ – $0.9$  Jy, and that of the far group is 1.1 Jy. The slightly greater median for the far group implies that the fainter objects detected by the BGPS tend to be nearby. A future release of Bolocat will utilize algorithms for separating filamentary versus compact emission, and will be able to investigate whether fainter objects tend to be more filamentary, and therefore not resolvable or detectable at great distance. A K-S test shows that while the near and far groups are drawn from different distributions (at the 99.998% confidence level), the tangent group seems to straddle the fence, as it cannot be ruled out that it differs from the near or the far groups at the 95% confidence level.

The right panels in Figure 14 illustrate the distance resolution aspects of Table 5. The distribution of  $P_{\text{ML}}$  for the kinematic sample and subset of well-constrained sources is plotted in the top right panel, with the overlap being complete for  $P_{\text{ML}} \geq 0.78$ . The well-constrained sources with  $P_{\text{ML}} \approx 0.5$  are near the tangent point, as the peaks in the posterior DPDFs straddle  $d_{\text{tan}}$ . The middle right panel shows the heliocentric distance distribution of well-constrained sources, with 1237/1710 (72%) of sources nearer than 5.5 kpc. Since Galactocentric radius is not subject to the KDA, the bottom right panel shows the distributions for both the kinematic sample and the well-constrained subset. There is a strong break at  $R_{\text{gal}} \approx 5.5$  kpc, which may represent the division between the Scutum–Centaurus and Sagittarius arms or it may be an artifact of the chosen flat rotation curve.<sup>18</sup> Kinematic

<sup>18</sup> Persic et al. (1996) derived a “universal” spiral galaxy rotation curve from the observed rotation curves of over 1000 galaxies that shows a clear downturn in the circular velocity in the inner several kiloparsecs. This type of downturn is consistent with the measured parallaxes and proper motions for Galactic HMSFRs of Reid et al. (2014), indicating that a flat rotation curve is not valid for  $R_{\text{gal}} \lesssim 5$  kpc.





**Figure 14.** Summary of source properties from Table 5. Top left: comparison of the Galactic longitude distributions for objects with well-constrained (black) vs. unconstrained (cyan) distances, with the red histogram showing the distribution of spectroscopic observations of Shirley et al. (2013). The gray hashed regions mark the longitude-projected kinematic avoidance zones (Section 3.2.2). Middle left: distributions of Galactic latitude for sources with well-constrained distance estimates at  $\ell \leq 90^\circ$ . Colors represent near (black), far (blue), and tangent (red) KDA resolutions. Bottom left: as above, but showing the distributions of  $\lambda = 1.1$  mm flux density. Top right: distributions of  $P_{ML}$  for the entire kinematic sample (black) and sources with well-constrained distance estimates (gray). Middle right: heliocentric distance distribution for the well-constrained subset. Bottom right: galactocentric radius distributions for the entire kinematic sample (black) and sources with well-constrained distance estimates (gray).

distances are unambiguous for  $R_{gal} > R_0$ , so all objects in the kinematic sample beyond the solar circle have well-constrained distance estimates. The marked gap at  $R_{gal} = 8.5\text{--}9.5$  kpc is the result of the only spiral feature (Perseus arm) within the BGPS coverage region with appreciable gas in this Galactocentric radius range lying within a kinematic avoidance zone.

### 6.5. Galactocentric Positions

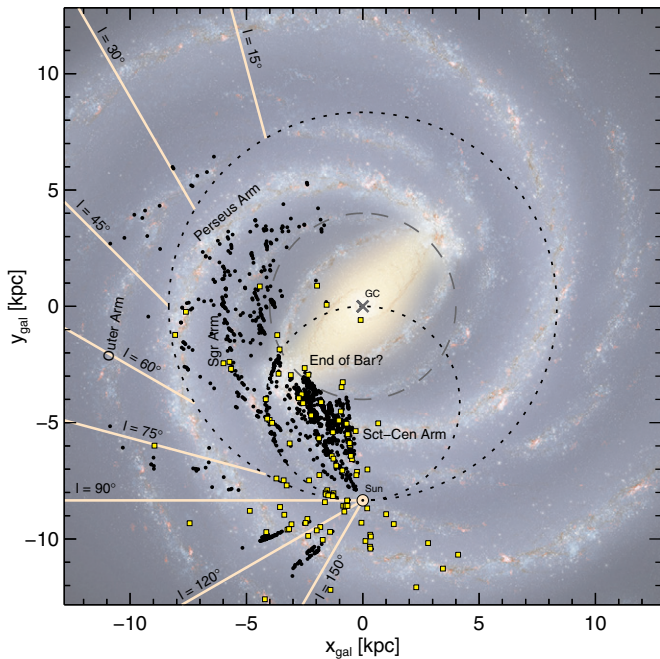
One important application of a large collection of well-constrained distance estimates for molecular cloud structures is the elucidation of Galactic structure in terms of the dense molecular gas that hosts star formation. Galactocentric positions may be derived using the  $(\ell, b, d_\odot) \rightarrow (R_{gal}, \phi, z)$  conversion matrix from Appendix C of EB13, which accounts for the  $\approx 25$  pc vertical offset of the Sun above the Galactic midplane (Humphreys & Larsen 1995; Jurić et al. 2008).

#### 6.5.1. Face-on View of the Milky Way

The face-on map of the Milky Way from the north Galactic pole is shown in Figure 15, with the maximum-likelihood distance (or  $\bar{d}$  for sources near  $d_{tan}$ ) for each well-constrained source plotted atop an illustration of the Galaxy derived from *Spitzer* near-infrared stellar data (R. Hurt: NASA/JPL-Caltech/SSC), scaled to the  $R_0$  from Table 1. There are two key attributes of this figure that bear mentioning. The first is the spread in heliocentric distance of sources along any given line of sight

(most noticeable around  $\ell = 30^\circ$  and  $\ell = 110^\circ$ ). Error bars are not shown in Figure 15 for clarity, but with a typical uncertainty of  $\approx 0.5$  kpc, object positions within various complexes are self-consistent. In addition to the uncertainty inherent in the posterior DPDF, deviant motions of the gas away from circular motion around the Galactic center can offset  $v_{LSR}$  and shift derived distances away from their true position, thereby creating an apparent dispersion along the line of sight and smoothing the underlying structure.

The second notable attribute is the placement of BGPS objects in regions of the background image that appear devoid of stars in the model (i.e.,  $\ell \approx 30^\circ \pm 10^\circ$ , or  $(x_{gal}, y_{gal}) \approx (-4, 0)$  kpc). There are two possible interpretations. The first is that the background image is a “best guess” only, based on stellar distributions from *Spitzer* data. Robust distance measurements for molecular cloud clumps may well be telling a different story of the locations of spiral arms and Galactic structure. For example, Egusa et al. (2011) found a significant population of molecular gas “downstream” of spiral arms in M51, nearly spanning the interarm region and coincident with H II regions identified in near-infrared images. The second interpretation is these are objects incorrectly placed at the far kinematic distance by the set of prior DPDFs currently implemented. In this case, it is likely that as the suite of data-driven prior DPDFs grows, sources will shift away from these vacant regions in the *Spitzer* model. The  $\approx 80$  BGPS sources in this area are primarily associated with HRDS H II regions; a (future,



**Figure 15.** Face-on view of the Milky Way for sources with well-constrained distance estimates (black circles), plotted atop an artist’s rendering of the Milky Way (R. Hurt: NASA/JPL-Caltech/SSC) viewed from the north Galactic pole. Yellow squares mark the locations of masers with trigonometric parallaxes (Reid et al. 2014, Table 1). The image has been scaled to match the  $R_0$  used for calculating kinematic distances. The outer dotted circle marks the solar circle, and the inner dotted circle the tangent point as a function of longitude. The dashed circle at  $R_{\text{gal}} = 4$  kpc outlines the region influenced by the long Galactic bar where the assumed flat rotation curve breaks down (Benjamin et al. 2005; Reid et al. 2014). Various suggested Galactic features are labeled. For clarity, distance error bars are not shown.

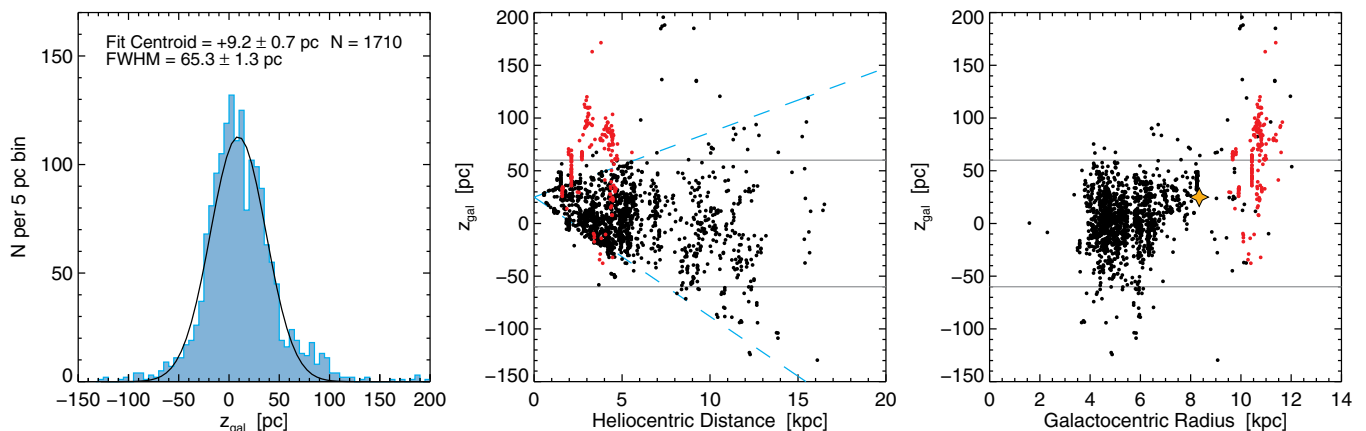
undeveloped) self-consistent HI absorption prior DPDF may solve this mystery.

Notwithstanding uncertainties in source location in Figure 15, several prominent Galactic features begin to suggest themselves based on the BGPS V2 distance catalog. The most significant is the end of the Galactic bar near  $\ell = 30^\circ$  and the start of the Scutum–Centaurus arm moving to smaller longitude. Next is the general outline of the Sagittarius arm, visible from  $(x_{\text{gal}}, y_{\text{gal}}) \approx (-3, 3)$  kpc counterclockwise around to its

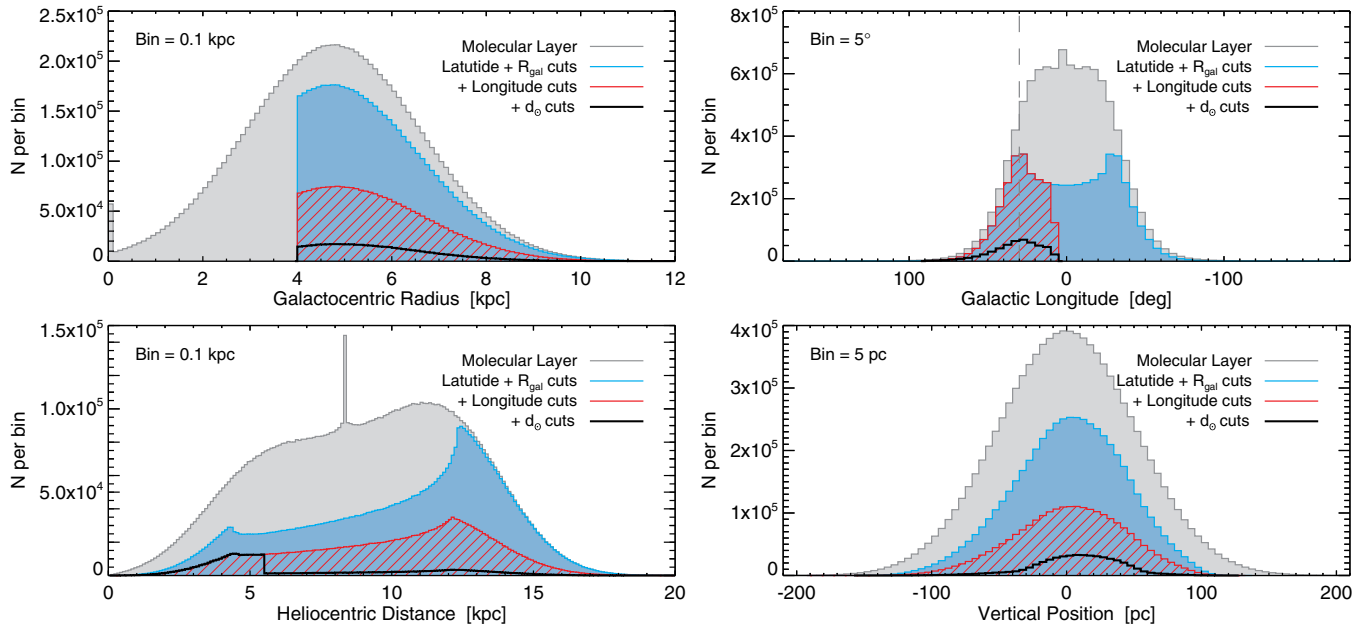
tangency near  $\ell = 50^\circ$ . Portions of the Perseus arm are traceable in the  $\ell = 40^\circ$ – $50^\circ$  region and again in the outer Galaxy. Finally, the BGPS detects 23 objects in the Outer arm beyond the solar circle in the  $\ell = 20^\circ$ – $80^\circ$  range, at a heliocentric distance of  $\approx 10$ – $15$  kpc.

### 6.5.2. Vertical Distribution of Star Formation

In addition to the face-on map of the Milky Way, well-constrained distance estimates permit study of the vertical distribution ( $z$ ) of sources about the Galactic midplane (Figure 16). The errors tabulated in the last column of Table 5 include contributions from variations in  $z$  along the line of sight over the range  $d_\odot \pm \sigma_d$  and the  $\pm 5$  km s $^{-1}$  uncertainty in the solar offset above the Galactic midplane (Jurić et al. 2008), added in quadrature. The left panel depicts the histogram of  $z$ , which may be fit by a Gaussian with a centroid at  $+9.2 \pm 0.7$  pc, a FWHM of  $65.3 \pm 1.3$  pc, and a reduced  $\chi^2_{\text{red}} = 1.8$ . The centroid being at slight positive  $z$  should not be confused with a centroid at slight positive Galactic latitude. In the middle panel, however, it is apparent that the width and centroid of the distribution may be slightly misleading owing to the nominal  $|b| \leq 0.5$  limit of BGPS coverage. The cyan dashed lines in that panel mark this limit at  $\ell = 30^\circ$  (these limits rotate to more positive values at larger longitude owing to the Sun’s vertical displacement above the  $z = 0$  plane). In both the middle and right panels, red circles mark BGPS sources in the outer Galaxy ( $\ell > 90^\circ$ ) where survey coverage was neither blind nor uniform, but rather focused on known regions of star formation. The gray dot–dashed lines mark the FWHM of the Galactic molecular layer ( $=120$  pc; Bronfman et al. 1988). The BGPS does not probe the full width of the molecular layer until  $d_\odot \gtrsim 6$  kpc, whereas the bulk of the distance catalog ( $\approx 76\%$ ) is closer than this point. The FWHM of the distribution in the left panel, therefore, should be viewed as a lower limit on the scale height of dense star-forming gas in the Galactic plane (see Section 7.1 for a discussion of the observational effects of the BGPS on the derived vertical position distribution). The rightmost panel in Figure 16 illustrates the relationship between Galactocentric radius and vertical position; the orange star marks the Sun’s location. Visible here is a warp in the molecular disk beyond the solar circle.



**Figure 16.** Vertical distribution of sources about the Galactic midplane. Left: histogram of  $z$  with Gaussian fit overplotted. Center: vertical position as a function of heliocentric distance, with cyan dashed lines showing approximate boundaries of BGPS coverage ( $|b| \leq 0.5$ ) at  $\ell = 30^\circ$ . Sources plotted in red are at  $\ell > 90^\circ$ . The gray dot–dashed lines mark the 60 pc scale height of molecular gas (Bronfman et al. 1988). Right: vertical position as a function of Galactocentric radius. Red sources and gray dot–dashed lines as in the middle panel. The star marks the Sun’s location.



**Figure 17.** Analysis of the observational effects on the measured vertical scale height of dense molecular cloud structures. In each panel, the histograms represent (gray) sources spread throughout the entire molecular layer, (blue) sources meeting the  $|b| \leq 0.5$  and  $R_{\text{gal}} \geq 4$  kpc limits of the BGPS, (red) sources additionally meeting the  $90^\circ \geq \ell \geq 7.5$  longitude limits of the kinematic sample in the inner Galaxy, and (black) sources additionally meeting heliocentric distance criteria to match the observed sample (see text). The panels show the distributions of sources in Galactocentric radius (top left), Galactic latitude (top right), heliocentric distance (bottom left), and vertical position about the Galactic midplane (bottom right).

## 7. DISCUSSION

### 7.1. On the Vertical Distribution and Galactic Census of Dense Molecular Cloud Structures

The vertical distribution of dense molecular cloud structures from the present subset of BGPS sources seems to imply that the star-forming structures in the interstellar medium have a smaller scale height than that measured for the molecular gas. As noted in Section 6.5.2, however, the nominal  $|b| \leq 0.5$  limit of the BGPS may impose an observational bias on this result. Here we investigate the connection between the observed distribution of sources with well-constrained distances and the Galactic distribution of molecular gas.

We conducted a series of simulations whereby objects were randomly distributed according to the axisymmetric  $\text{H}_2$  distribution of Wolfire et al. (2003, derived from the observations of Bronfman et al. 1988), then “observed” through a series of increasingly restricting criteria. The  $(R_{\text{gal}}, \phi, z)$  positions of 40,000 simulated sources were drawn randomly from the distributions of each coordinate from the Wolfire et al. model, and then converted to the observational  $(\ell, b, d_\odot)$  coordinates using the inverse of the matrix operation from Appendix C of EB13. This number of objects was chosen so that the final subset was comparable in number to the set of well-constrained distances for BGPS objects presented in this paper.

Since the molecular hydrogen  $R_{\text{gal}}$  distribution is modeled as a Gaussian with sufficient width to have non-zero probability at  $R_{\text{gal}} = 0$ , any randomly drawn negative value was assigned to the Galactic center. This mathematical convenience also somewhat simulates the observed spike in source number in the Central Molecular Zone. The distributions of Galactocentric position ( $R_{\text{gal}}$  and  $z$ ) and locally viewed coordinates ( $\ell$  and  $d_\odot$ ) for the entire simulated sample are shown as the filled gray histograms in Figure 17. The vertical position distributions (bottom right

**Table 6**  
Vertical Distribution Fits for Subsets

Set/Cut	Centroid (pc)	FWHM (pc)	$N$
Molecular layer	$0.0 \pm 0.3$	$120.0 \pm 0.6$	$40000 \pm 0$
Latitude + $R_{\text{gal}}$ cuts	$4.1 \pm 0.4$	$98.8 \pm 0.7$	$21318 \pm 92$
+ Longitude cuts	$4.3 \pm 0.5$	$97.4 \pm 1.1$	$9180 \pm 77$
+ $d_\odot$ cuts	$10.9 \pm 0.8$	$67.5 \pm 1.5$	$1987 \pm 33$

panel) may be fit with a Gaussian to determine the centroid and FWHM for inter-comparison. The histograms in Figure 17 represent the sum of 250 independent realizations of 40,000 simulated sources distributed randomly. The mean and standard deviation of the fits to the vertical position distributions of the 250 independent realizations are shown in Table 6 along with the number of objects remaining after each successive cut.

To simulate the observation function of the BGPS, we created several cuts to narrow the range of objects toward that which would be detected by the survey. First, all sources with  $|b| > 0.5$  and  $R_{\text{gal}} < 4$  kpc were removed (resulting in the blue filled histograms in Figure 17). The latitude cut matches the BGPS survey limits, and the Galactocentric radius cut eliminates sources in the Galactic bar, whose kinematic distances are unreliable. These criteria alone remove nearly half of the objects and narrow the vertical position distributions (see Table 6). Next, Galactic longitude was restricted to the contiguous portion of the BGPS survey also observed in  $\text{HCO}^+(3-2)$  with the HHT (S13), or  $90^\circ \geq \ell \geq 7.5$ . This removes effects of incomplete BGPS coverage in the outer Galaxy and lack of distances (due to velocity information) in the innermost Galaxy. The cumulative effect of these two cuts leaves just under one quarter of the original sources, and the distributions are shown as red hashed histograms in Figure 17.

This set of sources encompasses the direct observational restrictions on the molecular layer, but the red histogram of heliocentric distance does not resemble the  $d_{\odot}$  distribution of BGPS well-constrained distances, shown in Figure 14 (middle right). As was discussed in EB13, the EMAF-based prior DPDF strongly selects sources at the near kinematic distance, as distant sources are less likely to have a discernible mid-infrared contrast for comparison with the *Spitzer*/GLIMPSE mosaics. Because the distribution in Figure 14 drops sharply for  $d_{\odot} \gtrsim 5.5$  kpc, we modeled this by randomly removing 90% of the sources in the red histogram in Figure 17 beyond  $d_{\odot} = 5.5$  kpc. This removal fraction approximates the ratio of sources per bin at 4 kpc to that at  $\sim 11$  kpc in Figure 14.

The final subset contains a little under 2000 objects, comparable with the 1710 BGPS objects with well-constrained distance estimates. Two points bear mention. The first is the vertical position distribution of the final subset, whose FWHM ( $67.5 \pm 1.5$  pc) and centroid ( $+10.9 \pm 0.8$  pc) are consistent with the measured values ( $65.3 \pm 1.3$  pc and  $+9.2 \pm 0.7$  pc, respectively) from Section 6.5.2. The observed narrow vertical position distribution of BGPS sources is therefore consistent with the molecular layer as probed by CO (Bronfman et al. 1988) mitigated by the observational restrictions of the BGPS. Second, in order to create an appropriately sized subset ( $N \sim 2000$ ) that meets the observational parameters of the well-constrained distance sources of Table 5, an initial sample of 40,000 dense molecular cloud structures is required. This value represents an approximate census of the total expected number of these objects throughout the Galactic plane.

### 7.2. $^{13}\text{CO}$ as a Molecular Cloud Clump Tracer

The ubiquitous nature of CO in the Galactic plane has made it an invaluable tool for studying large-scale Galactic structure (see Dame et al. 2001). The low excitation density of the optically thinner isotopologue  $^{13}\text{CO}$  explains the many velocity components identified in surveys such as the GRS along any given line of sight, especially toward the crowded inner Galaxy. The catalog of molecular cloud clumps presented in Rathborne et al. (2009) implies that  $^{13}\text{CO}$  emission is concentrated in denser regions surrounded by an envelope of more diffuse emission, represented by the larger GMCs also cataloged by Rathborne et al. Utilizing this emission enhancement for assigning a velocity to dust-continuum-identified molecular cloud structures has usually meant looking for the brightest emission peak along the line of sight.

Comparison between the  $v_{\text{LSR}}$  extracted directly from the  $^{13}\text{CO}$  spectra of the GRS and that from a dense gas tracer (see Section 2.2), however, yields only a  $\sim 85\%$  matching rate. This implies that while the  $^{13}\text{CO}$  associated with molecular cloud clumps does produce enhanced emission, it may be outshone by expanses of less dense gas not associated with the molecular cloud clump in question, whether by quantity of gas or changes in excitation temperature due to the local environment. Extraction of a spectrum utilizing the morphology of millimeter dust continuum emission as a prior increases the agreement rate with  $\text{HCO}^+(3-2)$  to 95%, with approximately one in eight of the disagreeing sources having a  $v_{\text{LSR}}$  outside the velocity range of the GRS data. By identifying molecular cloud structures through millimeter dust continuum emission, it is therefore possible to convert a molecular transition line with low  $n_{\text{eff}}$  into a powerful dense gas tracer.

The morphological spectrum extraction technique is able to assign a  $v_{\text{LSR}}$  to molecular cloud structures at a high rate,

**Table 7**  
Application of DPDFs to BGPS V2 Sources

DPDF	$N$	$N_{\text{wc}}$	Well-constrained Sources Only		
			Fraction of Method (%)	Fraction of w.c. <sup>a</sup> (%)	Fraction of Full V2.1 (%)
Dense Gas $v_{\text{LSR}}$	2432	1254	51.6	73.3	14.6
$^{13}\text{CO}$ $v_{\text{LSR}}$	958	338	35.3	19.8	3.9
H <sub>2</sub> Scale Height	4474	1384	30.9	80.9	16.1
EMAF	854	679	79.5	39.7	7.9
Parallax Assoc.	292	291	99.7	17.0	3.4
H II Region Assoc.	525	498	94.9	29.1	5.8
Gemini OB1	49	49	100	2.9	0.6
Well Constrained	1710	1710	...	100	19.9

**Note.** <sup>a</sup> Fraction of the set of well-constrained sources.

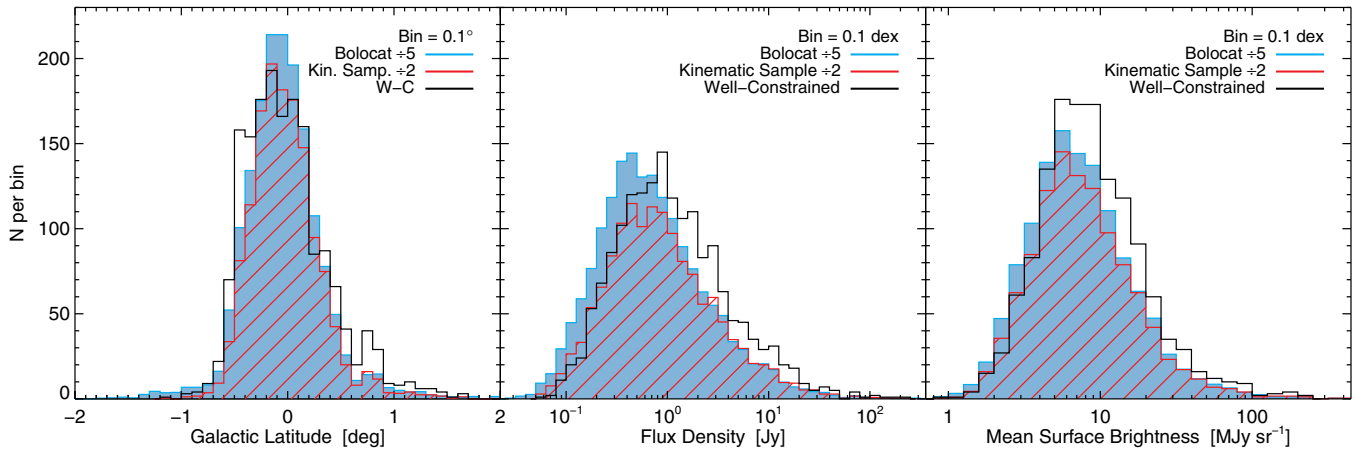
exceeding 80% for  $S_{1.1\text{mm}} \gtrsim 0.4$  Jy, and does significantly better than dense gas tracers for  $S_{1.1\text{mm}} \lesssim 0.1$  Jy, where the detection rate for other molecules falls below 20%. Especially due to the limited Galactic longitude range of the GRS, this technique does not add kinematic information for a significant number of bright BGPS sources, but it is almost exclusively the tracer of choice for low flux density objects.

While this technique was developed using the GRS with its northern coverage, it is directly applicable to upcoming large CO surveys. The Mopra Southern Galactic Plane CO Survey (Burton et al. 2013) will observe the  $J = 1-0$  transitions of  $^{12}\text{CO}$ ,  $^{13}\text{CO}$ , and  $\text{C}^{18}\text{O}$  over the range  $305^\circ \leq \ell \leq 345^\circ$  and  $|b| \leq 0.5^\circ$  with  $\sim 0.7$  K rms sensitivity. With  $35''$  angular and  $0.1 \text{ km s}^{-1}$  spectral resolution, this survey will complement the GRS and provide a symmetric view of the molecular Galaxy into the fourth quadrant. Additionally, the northern hemisphere JCMT/HARP  $^{12}\text{CO}(3-2)$  survey (Dempsey et al. 2013) will cover  $10^\circ \leq \ell \leq 65^\circ$  and  $|b| \leq 0.5^\circ$  with  $\sim 1$  K rms sensitivity and (smoothed)  $16''$  angular and  $1 \text{ km s}^{-1}$  spectral resolution. The improved angular resolution of the HARP and Mopra surveys will provide CO observations more closely matched to the ATLASGAL data set. For use with these surveys, the two tunable parameters in the morphological spectrum extraction technique (threshold  $T_A$  for detection, and the primary-to-secondary peak  $T_A$  ratio) will require calibration, as in Figure 5.

### 7.3. Inter-comparison of Prior DPDF Methods

The performance of each DPDF method presented here is shown in Table 7. The first two columns of numbers describe how many objects with a DPDF from the indicated method are in the whole catalog and the well-constrained subset, respectively. To evaluate the power of each method, the remaining columns use the number of well-constrained sources ( $N_{\text{wc}}$ ) as the numerator to indicate the constituent fractions of various larger sets. The “fraction of method” column ( $N_{\text{wc}}/N$ ) indicates how well that method produces well-constrained distance estimates, while the “fraction of w.c.” column describes how powerful the method is for constraining distances to molecular cloud structures. For instance, parallax association has a very high well-constrained rate (99.7%), but its sources constitute only 17% of the well-constrained set. The final column in Table 7 specifies the fraction of the full V2.1 catalog for each method; the values are simply scaled down from the previous column.

The greatest utility of the Bayesian DPDF formalism used here is the application of multiple prior DPDFs for KDA



**Figure 18.** Comparison of observable quantities for the entire BGPS catalog (divided by 5, blue), the kinematic sample (divided by 2, red), and set of well-constrained distance estimates (black). Left: the Galactic latitude distributions. Middle: source  $\lambda = 1.1$  mm flux density distributions. Right: source-averaged surface brightness, computed using Equation (3).

**Table 8**  
Well-constrained Fraction for Overlapping Priors

	H <sub>2</sub> Scale Height	EMAF	Parallax Assoc.	H II Region Assoc.	Gemini OB1
H <sub>2</sub> Scale Height	0.309	0.789	0.995	0.949	...
EMAF		0.795	1.000	0.855	...
Parallax Assoc.			0.997	1.000	...
H II Region Assoc.				0.949	...
Gemini OB1					1.000

resolution. The use of mostly orthogonal priors allows for the maximum number of possible KDA resolutions, but there is some overlap between priors. To illustrate the interaction of priors, Table 8 computes the fraction of sources in the intersection of prior DPDFs (i.e., having both) that have well-constrained distance estimates. For clarity, only the upper portion of the symmetric matrix is shown in the table, and the diagonal elements represent the “fraction of method” column in Table 7. There is no intersection with the Gemini OB1 prior, as this was defined to assign a distance only to objects in that molecular cloud complex that could not be associated with one of the trigonometric parallax measurements.

The distribution of molecular gas (H<sub>2</sub> scale height) does not offer significant leverage for KDA resolution by itself, but its intended utility is limited to high-latitude objects. For instance, of the 4474 objects with DPDF<sub>H<sub>2</sub></sub>, some 3094 (69%) are in the kinematic sample, placing an upper limit on the KDA resolution rate. Of this kinematic pool, only 1384 (45%) have a well-constrained distance estimate.

The prior DPDF based on EMAFs has a  $\approx 80\%$  success rate for deriving well-constrained distance estimates, nearly identical to the rate from EB13. The intersection of the EMAF sample with either of the catalog-based priors is small. Maser emission and H II regions are both associated with the later stages of star formation (see Battersby et al. 2010; Dunham et al. 2011a), whereas EMAFs correspond to the earlier stages (EB13). It was also noted in EB13 that bright PAH emission near  $\lambda = 8 \mu\text{m}$  excited by UV photons from H II regions breaks the assumption of smooth Galactic emission against which EMAFs are seen, and can skew the distance estimate returned by DPDF<sub>emaf</sub>. So, while the EMAF-HRDS overlap is small, the high fraction (86%) of well-constrained distance estimates implies that the KDA resolutions of the individual priors tend to agree, as opposing KDA resolutions would

result in unconstrained posterior DPDFs. This agreement rate is evidence that some level of localized bright  $8 \mu\text{m}$  emission from H II regions does not significantly affect the KDA resolutions of EMAF-identified molecular cloud structures.

The new prior DPDFs introduced in this paper both show very high rates of well-constrained distance estimates. The nearly 100% rate for DPDF<sub>px</sub> should be expected, as the precision of recent VLBI trigonometric parallaxes produces very narrow distance uncertainties. The sole DPDF<sub>px</sub> outlier is kinematically aberrant, and the DPDF formalism rightly returned an unconstrained distance estimate. The H II region prior produces well-constrained distance estimates for 95% of the associated sources, which is slightly tempered in the EMAF-HRDS overlap, as discussed above. The high distance-assignment rate should be viewed as an indication that the priors thusly assigned do not collide with other prior DPDFs because the coordinate–velocity association volume of Section 5.3 does not exceed the physical association scale length of HMSFRs in the Galactic plane.

#### 7.4. A Representative Sample?

The 1710 sources forming the BGPS V2 distance catalog provide a sizable sample for making comprehensive inferences about the Galactocentric distributions and physical properties of molecular cloud structures. The strength of such inferences, however, hinges on how representative this sample is of the BGPS as a whole. The relationships between the entire Boloccat, the kinematic sample of 3508 sources, and the set of well-constrained distance estimates are shown in Figure 18 for the distributions of Galactic latitude, flux density, and mean surface brightness. The entire Boloccat (divided by 5) is shown in blue, the kinematic sample (divided by 2) in red, and the well-constrained distances are plotted in black. Mean surface

brightness is computed via

$$SB = \frac{S_{1.1}}{\pi \theta_R^2} = 12.4 \text{ MJy sr}^{-1} \left( \frac{S_{1.1}}{1 \text{ Jy}} \right) \left( \frac{\theta_R}{33''} \right)^{-2}, \quad (3)$$

where  $S_{1.1}$  is the BGPS integrated flux density and  $\theta_R$  is the deconvolved radius of the catalog source (Rosolowsky et al. 2010a). Surface brightness is related to the source column density, but is a strictly observable quantity not requiring assumptions about the physics of the dust or environment.

The Galactic latitude distributions for the three sets of sources all have a Gaussian centroid at  $b \approx -0^\circ.1$ , and a FWHM  $\approx 0^\circ.7$ , suggesting that the prior DPDFs currently implemented do not bias our sampling of Galactic latitude. A set of two-sample K-S tests, however, shows that the distributions are not identical at the 99.5% confidence level or greater. In terms of source flux density, the median values for the entire catalog, kinematic sample, and well-constrained distances are 0.62 Jy, 0.76 Jy, and 0.96 Jy, respectively. This skew for the kinematic sample is accounted for in the kinematic detection rate as a function of flux density (Figure 9). The further skew to higher flux density in the well-constrained sample is likely due in part to the flux–density selection effects of the EMAF prior (EB13). For the distributions of  $SB$ , the differences are smaller, with median values of 6.8, 7.1, and 8.0 MJy sr<sup>-1</sup> for the full catalog, kinematic sample, and well-constrained sample, respectively. Another set of two-sample K-S tests reveal that the well-constrained sample is not drawn from the same parent population as the others at the 99.97% confidence level or greater, but the full catalog and kinematic samples cannot be distinguished at the 95% confidence level. Although the flux–density distributions for these two sets are significantly different, the tendency for brighter sources to subtend larger solid angles (the Spearman rank correlation between  $S_{1.1}$  and  $\theta_R$  is 0.75–0.78 for all three groups; see Equation (3)) balances this effect. The implication is that the averaged surface brightness of sources is not a significant factor in the kinematic detection rate, but may slightly influence the ability of the currently implemented prior DPDFs (such as  $\text{DPDF}_{\text{emaf}}$ ) to produce well-constrained distance estimates.

The comparison of these observable quantities suggests that the present distance catalog is biased toward brighter and denser objects than either the Bolocat as a whole or even the kinematic sample of objects with detected  $v_{\text{LSR}}$ . Physical source properties derived from this catalog, therefore, are not entirely representative of the dimmer and/or more diffuse objects detected by the BGPS. As a further check on the representative nature of this distance catalog, we compared our derived distances with those of Dunham et al. (2011b), who studied a subset of BGPS V1 sources using NH<sub>3</sub>(1,1). Of the 456 objects presented in that paper, 169 lie within 60'' of the peak location of a V2.1 source with a well-constrained distance estimate. Distances derived here agree with those of Dunham et al. at a rate of 70%, where their KDA resolutions relied heavily upon H I absorption techniques evaluated by eye. The future development of an automated prior DPDF based on H I absorption will undoubtedly resolve some of these discrepancies.

## 8. SUMMARY

We expanded the DPDF formalism for molecular cloud structures originally presented in EB13 by including a new kinematic distance likelihood method and additional prior

DPDFs for KDA resolution. For the specific case of the BGPS, we present an expanded distance catalog corresponding the recently released Version 2 maps and source catalog (G13).

As a primary foundation, the DPDF formalism uses a kinematic distance likelihood within its Bayesian framework. Molecular transition line surveys that probe dense gas (e.g., HCO<sup>+</sup>(3–2); S13) have a  $\sim 50\%$ – $85\%$  detection rate when pointed at catalog positions from continuum surveys of molecular cloud structures. To increase the fraction of sources with kinematic information, we developed a technique for morphologically extracting a spectrum from the <sup>13</sup>CO data of the GRS for molecular cloud structures using the dust–continuum data as prior information. The low effective density for excitation of <sup>13</sup>CO(1–0) requires that the emission spectrum from the more diffuse envelope around a millimeter catalog source be subtracted in order to return a single detectable  $v_{\text{LSR}}$ . When compared to using the  $v_{\text{LSR}}$  of the brightest <sup>13</sup>CO peak along the line of sight, this morphological spectrum extraction technique increases the velocity–matching rate between <sup>13</sup>CO  $v_{\text{LSR}}$  and HCO<sup>+</sup>(3–2) from  $\approx 85\%$  to  $\approx 95\%$ . The additional velocities derived from <sup>13</sup>CO provide kinematic information for half again as many sources as the dense gas tracers alone, and this method is directly extensible to upcoming large CO surveys across the Galactic plane.

We also introduce a new set of prior DPDFs based on the association of molecular cloud structures with literature catalog objects with reliable distance estimates or robust KDA resolutions. Molecular cloud structures from the BGPS are associated with these catalog entries based on a coordinate–velocity volume derived from the cumulative distributions of GMC physical properties from the GRS (Roman-Duval et al. 2010) projected to the distance of the reference source. The combined list of trigonometric parallax measurements of masers associated with HMSFRs from the BeSSeL Survey and VERA Project (tabulated in Reid et al. 2014;  $N = 103$ ) offers a treasure trove of gold-standard distances independent of kinematic assumptions. The high precision of these parallax measurements translates into small heliocentric distance uncertainties. Consequently, all but one of the 292 BGPS V2 sources associated with one or more maser parallax measurement have well-constrained distance estimates from the posterior DPDFs. In parallel, the list of 441 H II regions from the HRDS with robust KDA resolutions provides a strong constraint on distances to BGPS objects; nearly 95% of sources lying within the association volume of one or more HRDS H II region have well-constrained distance estimates. In the outer Galaxy, the kinematic avoidance zone around  $\ell = 180^\circ$  prohibits the use of kinematic distances. The BGPS-studied Gemini OB1 molecular cloud, however, lies in this region, and we assign distances to objects in this complex based on the trigonometric parallaxes to objects S252 (2.10 kpc; Reid et al. 2009a) and S255 (1.59 kpc; Rygl et al. 2010).

We present the V2 distance and velocity catalog for the BGPS, where all molecular transition-line spectroscopic observations have been aligned with the updated source catalog of G13. The catalog and computed DPDFs are publicly available. Of 3508 BGPS V2 sources with kinematic information, 1710 (49%) now have well-constrained distance estimates. The Galactocentric positions of these objects trace out various Galactic features, including portions of the Sagittarius, Perseus, and Outer arms, as well as the Scutum–Centaurus arm between the Sun and the Galactic center. The vertical distribution of BGPS molecular cloud structures with well-constrained distance estimates is narrower than the measured distribution of CO in the disk, but

**Table 9**  
Glossary of Acronyms

Term	Definition
BeSSeL	The Bar and Spiral Structure Legacy Survey (Brunthaler et al. 2011)
BGPS	The Bolocam Galactic Plane Survey (Aguirre et al. 2011, G13)
DPDF	Distance probability density function
EMAF	Eight-micron absorption feature
GMC	Giant molecular cloud
GRS	The Galactic Ring Survey (Jackson et al. 2006)
HHT	Heinrich Hertz Submillimeter Telescope (Mt. Graham, AZ)
HISA	H I self absorption
HIE/A	H I emission/absorption
HMSFR	High-mass star forming region
HRDS	H II Region Discovery Survey (Bania et al. 2010, 2012)
IRDC	Infrared dark cloud
KDA	Kinematic distance ambiguity
VERA	The Japanese VLBI Exploration of Radio Astrometry project
VLBI	Very long baseline interferometry

an analysis of the observational biases in the BGPS reveals that our measurement is consistent with the CO layer as measured by Bronfman et al. (1988). Furthermore, in order to obtain the  $\sim 2000$  objects in the present distance catalog, there must be  $\sim 40,000$  molecular cloud clumps throughout the Galactic plane. This represents the first approximate census of the total expected number of these objects. Finally, while the collection of well-constrained distances represents only 20% of the BGPS catalog, the distributions of observable quantities are generally representative of the entire catalog, but are biased toward higher-density sources due to the suite of prior DPDFs currently available.

This work was supported by the National Science Foundation through NSF grant AST-1008577. The BGPS project was supported in part by NSF grant AST-0708403, and was performed at the Caltech Submillimeter Observatory (CSO), which was supported by NSF grants AST-0540882 and AST-0838261. The CSO was operated by Caltech under contract from the NSF. ER is supported by a Discovery Grant from NSERC of Canada. NJE is supported by NSF grant AST-1109116. This publication makes use of molecular line data from the Boston University-FCRAO Galactic Ring Survey (GRS). The GRS is a joint project of Boston University and Five College Radio Astronomy Observatory, funded by the NSF under grants AST-9800334, AST-0098562, AST-0100793, AST-0228993, & AST-0507657.

## APPENDIX A

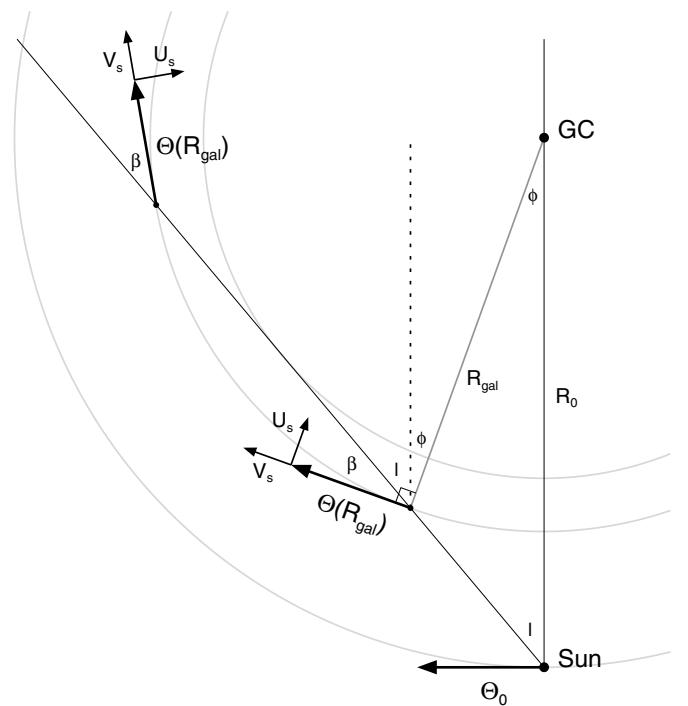
### GLOSSARY

This work contains many acronyms; in Table 9 we present a glossary of terms for reference.

## APPENDIX B

### DERIVATION OF KINEMATIC DISTANCE LIKELIHOODS FROM THE REID ET AL. (2014) ROTATION CURVE

The rotation curve of Reid et al. (2014) used here includes parameters describing the mean peculiar motions of HMSFRs away from the standard circular orbits about the Galactic center. In this appendix, we derive the kinematic distance likelihood  $\mathcal{L}(v_{\text{LSR}}, l, b; d_{\odot})$  from Equation (1) using the full features of the Reid et al. curve. The relevant geometry for this derivation is shown in Figure 19.



**Figure 19.** Galactic geometry for the computation of kinematic distances. Circular arcs mark orbits of constant  $R_{\text{gal}}$ , the outermost being the solar circle. Of the two inner orbits shown, the orbit with larger  $R_{\text{gal}}$  intersects the line of sight along  $\ell$  twice, each with equal projected  $v_{\text{LSR}}$ . The innermost orbit is tangent to the line of sight, and represents the smallest  $R_{\text{gal}}$  visible along the illustrated  $\ell$ . Various velocities, angles, and distances are indicated, which are described in the text.

The circular rotation velocity as a function of  $R_{\text{gal}}$  is given by

$$\Theta(R_{\text{gal}}) = \Theta_0 + (d\Theta/dR_{\text{gal}})(R_{\text{gal}} - R_0), \quad (\text{B1})$$

where  $\Theta_0$ ,  $R_0$ , and  $d\Theta/dR_{\text{gal}}$  are taken from Table 1. By geometry, the observed line-of-sight velocity to an object along longitude  $\ell$  and latitude  $b$  with Galactocentric radius  $R_{\text{gal}}$  is

$$v_{\text{obs}} = (\Theta(R_{\text{gal}}) + \overline{V}_s) \cos \beta \cos b + \overline{U}_s \sin \beta \cos b - \Theta_0 \sin \ell, \quad (\text{B2})$$

where  $\beta$  is the angle between the circular motion of the source and the line of sight, and  $\overline{V}_s$  and  $\overline{U}_s$  are the mean peculiar

motions of HMSFRs, as discussed in Section 3.2.1. Using the fact that

$$\beta = \frac{\pi}{2} - \ell - \phi \quad (\text{B3})$$

and the trigonometric relationships for  $\phi$  (the Galactocentric azimuth) are

$$\sin \phi = \frac{d_{\odot} \sin \ell}{R_{\text{gal}}} \quad (\text{B4})$$

$$\cos \phi = \frac{R_0 - d_{\odot} \cos \ell}{R_{\text{gal}}}, \quad (\text{B5})$$

Equation (B2) may be written as

$$v_{\text{obs}} = (\Theta(R_{\text{gal}}) + \bar{V}_s) \left[ \frac{R_0}{R_{\text{gal}}} \sin \ell \right] \cos b + \bar{U}_s \left[ \frac{R_0}{R_{\text{gal}}} \cos \ell - \frac{d_{\odot}}{R_{\text{gal}}} \right] \cos b - \Theta_0 \sin \ell. \quad (\text{B6})$$

To remove  $R_{\text{gal}}$  and frame this relationship solely in terms of velocities and the distance ratio ( $d_{\odot}/R_0$ ), we utilize the relationship

$$R_{\text{gal}} = \sqrt{R_0^2 + d_{\odot}^2 - 2R_0 d_{\odot} \cos \ell} \quad (\text{B7})$$

to recast the observed velocity as

$$v_{\text{obs}} = \left[ \frac{(\Theta(R_{\text{gal}}) + \bar{V}_s) \cos b}{\sqrt{1 + (d_{\odot}/R_0)^2 - 2(d_{\odot}/R_0) \cos \ell}} - \Theta_0 \right] \sin \ell + \frac{\bar{U}_s \cos b}{\sqrt{1 + (d_{\odot}/R_0)^2 - 2(d_{\odot}/R_0) \cos \ell}} \left[ \cos \ell - \frac{d_{\odot}}{R_0} \right]. \quad (\text{B8})$$

Equation (B8) describes the velocity of the source as a function of  $d_{\odot}$  relative to the solar circle orbit with velocity  $\Theta_0$ . Comparison with  $v_{\text{LSR}}$  reported at the telescope from molecular transition line observations, however, requires correction for the solar peculiar motion used by the telescope. Rather than convert the reported velocities to match the solar peculiar motion fit as part of the rotation curve (Reid et al. 2014), we chose to transform the velocity of Equation (B8) into the (old) IAU solar motion frame. The proper heliocentric velocity is computed from

$$v_{\text{helio}} = v_{\text{obs}} - (U_{\odot} \cos \ell + V_{\odot} \sin \ell) \cos b - W_{\odot} \sin b, \quad (\text{B9})$$

where ( $U_{\odot}$ ,  $V_{\odot}$ ,  $W_{\odot}$ ) are the components of the solar peculiar motion from Table 1. Next, add back in the IAU value of 20 km s<sup>-1</sup> in the direction of 18<sup>h</sup> right ascension and +30° declination (B1950).

The correction of Equation (B8) to the telescope-reported frame results in a  $v_{\text{tel}}(\ell, b, d_{\odot})$ . We compute the kinematic distance likelihood  $\mathcal{L}(v_{\text{LSR}}, l, b; d_{\odot})$  over an array of heliocentric distance out to 20 kpc every 20 pc. To encapsulate all information contained in the spectral line observations, we use a Gaussian fit to the spectrum,  $v_{\text{LSR}}$ . The likelihood function is then computed via

$$\mathcal{L}(v_{\text{LSR}}, l, b; d_{\odot}) = \int \exp \left[ -\frac{(v_{\text{tel}}(\ell, b, d_{\odot}) - v')^2}{2\sigma_{\text{vir}}^2} \right] v_{\text{LSR}} dv', \quad (\text{B10})$$

where  $v'$  is the dummy integration variable over the range of valid velocities, and  $\sigma_{\text{vir}}$  is the magnitude of expected virial motions within regions of massive star formation, accounting for peculiar motions of individual molecular cloud clumps (= 7 km s<sup>-1</sup>; Reid et al. 2009b).

## REFERENCES

- Aguirre, J. E., Ginsburg, A. G., Dunham, M. K., et al. 2011, *ApJS*, 192, 4
- Anderson, L. D., & Bania, T. M. 2009, *ApJ*, 690, 706
- Anderson, L. D., Bania, T. M., Balser, D. S., & Rood, R. T. 2012, *ApJ*, 754, 62
- Bania, T. M., Anderson, L. D., & Balser, D. S. 2012, *ApJ*, 759, 96
- Bania, T. M., Anderson, L. D., Balser, D. S., & Rood, R. T. 2010, *ApJL*, 718, L106
- Battersby, C., Bally, J., Ginsburg, A., et al. 2011, *A&A*, 535, A128
- Battersby, C., Bally, J., Jackson, J. M., et al. 2010, *ApJ*, 721, 222
- Battisti, A. J., & Heyer, M. H. 2014, *ApJ*, 780, 173
- Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2005, *ApJL*, 630, L149
- Brand, J., & Blitz, L. 1993, *A&A*, 275, 67
- Bronfman, L., Cohen, R. S., Alvarez, H., May, J., & Thaddeus, P. 1988, *ApJ*, 324, 248
- Brunt, C. M., Kerton, C. R., & Pomerleau, C. 2003, *ApJS*, 144, 47
- Brunthaler, A., Reid, M. J., Menten, K. M., et al. 2011, *AN*, 332, 461
- Burton, M. G., Braiding, C., Glueck, C., et al. 2013, *PASA*, 30, 44
- Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, *PASP*, 121, 213
- Clemens, D. P. 1985, *ApJ*, 295, 422
- Clemens, D. P., Sanders, D. B., & Scoville, N. Z. 1988, *ApJ*, 327, 139
- Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, *ApJ*, 547, 792
- Dame, T. M., & Thaddeus, P. 2011, *ApJL*, 734, L24
- Dempsey, J. T., Thomas, H. S., & Currie, M. J. 2013, *ApJS*, 209, 8
- Dunham, M. K., Robitaille, T. P., Evans, N. J., II, et al. 2011a, *ApJ*, 731, 90
- Dunham, M. K., Rosolowsky, E., Evans, N. J., II, Cyganowski, C., & Urquhart, J. S. 2011b, *ApJ*, 741, 110
- Dunham, M. K., Rosolowsky, E., Evans, N. J., II, et al. 2010, *ApJ*, 717, 1157
- Eden, D. J., Moore, T. J. T., Plume, R., & Morgan, L. K. 2012, *MNRAS*, 422, 3178
- Egusa, F., Koda, J., & Scoville, N. 2011, *ApJ*, 726, 85
- Ellsworth-Bowers, T. P., Glenn, J., Rosolowsky, E., et al. 2013, *ApJ*, 770, 39 (EB13)
- Evans, N. J., II. 1999, *ARA&A*, 37, 311
- Fux, R. 1999, *A&A*, 345, 787
- Ginsburg, A., Glenn, J., Rosolowsky, E., et al. 2013, *ApJS*, 208, 14 (G13)
- Goldsmith, P. F., Li, D., & Krčo, M. 2007, *ApJ*, 654, 273
- Heyer, M., Krawczyk, C., Duval, J., & Jackson, J. M. 2009, *ApJ*, 699, 1092
- Heyer, M. H., Brunt, C., Snell, R. L., et al. 1998, *ApJS*, 115, 241
- Honma, M., Nagayama, T., Ando, K., et al. 2012, *PASJ*, 64, 136
- Humphreys, R. M., & Larsen, J. A. 1995, *AJ*, 110, 2183
- Jackson, J. M., Finn, S. C., Rathborne, J. M., Chambers, E. T., & Simon, R. 2008, *ApJ*, 680, 349
- Jackson, J. M., Rathborne, J. M., Foster, J. B., et al. 2013, *PASA*, 30, 57
- Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, *ApJS*, 163, 145
- Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, *ApJ*, 673, 864
- Kennicutt, R. C., & Evans, N. J. 2012, *ARA&A*, 50, 531
- McKee, C. F., & Ostriker, E. C. 2007, *ARA&A*, 45, 565
- McMillan, P. J., & Binney, J. J. 2010, *MNRAS*, 402, 934
- Molinari, S., Swinyard, B., Bally, J., et al. 2010, *PASP*, 122, 314
- Niinuma, K., Nagayama, T., Hirota, T., et al. 2011, *PASJ*, 63, 9
- Persic, M., Salucci, P., & Stel, F. 1996, *MNRAS*, 281, 27
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011a, *A&A*, 536, A22
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011b, *A&A*, 536, A23
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014, arXiv:1406.5093
- Purcell, C. R., Longmore, S. N., Walsh, A. J., et al. 2012, *MNRAS*, 426, 1972
- Rathborne, J. M., Johnson, A. M., Jackson, J. M., Shah, R. Y., & Simon, R. 2009, *ApJS*, 182, 131
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2009a, *ApJ*, 693, 397
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *ApJ*, 783, 130
- Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009b, *ApJ*, 700, 137
- Robitaille, T. P., Churchwell, E., Benjamin, R. A., et al. 2012, *A&A*, 545, A39
- Rodriguez-Fernandez, N. J., & Combes, F. 2008, *A&A*, 489, 115
- Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, *ApJ*, 723, 492
- Roman-Duval, J., Jackson, J. M., Heyer, M., et al. 2009, *ApJ*, 699, 1153
- Rosolowsky, E., Dunham, M. K., Ginsburg, A., et al. 2010a, *ApJS*, 188, 123



- Rosolowsky, E., Ginsburg, A., & Ellsworth-Bowers, T. 2010b, in ASP Conf. Ser. 438, *The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey*, ed. R. Kothes, T. L. Landecker, & A. G. Willis (San Francisco, CA: ASP), 76
- Russeil, D., Pestalozzi, M., Mottram, J. C., et al. 2011, *A&A*, 526, A151
- Rygl, K. L. J., Brunthaler, A., Reid, M. J., et al. 2010, *A&A*, 511, A2
- Sanders, D. B., Scoville, N. Z., & Solomon, P. M. 1985, *ApJ*, 289, 373
- Schlingman, W. M., Shirley, Y. L., Schenk, D. E., et al. 2011, *ApJS*, 195, 14
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, *A&A*, 504, 415
- Shirley, Y. L., Ellsworth-Bowers, T. P., Svoboda, B., et al. 2013, *ApJS*, 209, 2 (S13)
- Simon, R., Jackson, J. M., Rathborne, J. M., & Chambers, E. T. 2006a, *ApJ*, 639, 227
- Simon, R., Rathborne, J. M., Shah, R. Y., Jackson, J. M., & Chambers, E. T. 2006b, *ApJ*, 653, 1325
- Stacy, J. A., & Thaddeus, P. 1991, in ASP Conf. Ser. 16, *Atoms, Ions and Molecules: New Results in Spectral Line Astrophysics*, ed. A. D. Haschick & P. T. P. Ho (San Francisco, CA: ASP), 197
- Walsh, A. J., Breen, S. L., Britton, T., et al. 2011, *MNRAS*, 416, 1764
- Wienen, M., Wyrowski, F., Schuller, F., et al. 2012, *A&A*, 544, A146
- Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, *ApJ*, 587, 278
- Xu, Y., Li, J. J., Reid, M. J., et al. 2013, *ApJ*, 769, 15